ISSN: 1065-514X VLDEEZ 8(1–4) 1–572 (1998)

Editor In Chief
George W. Zobrist
University of Missouri, Rolla

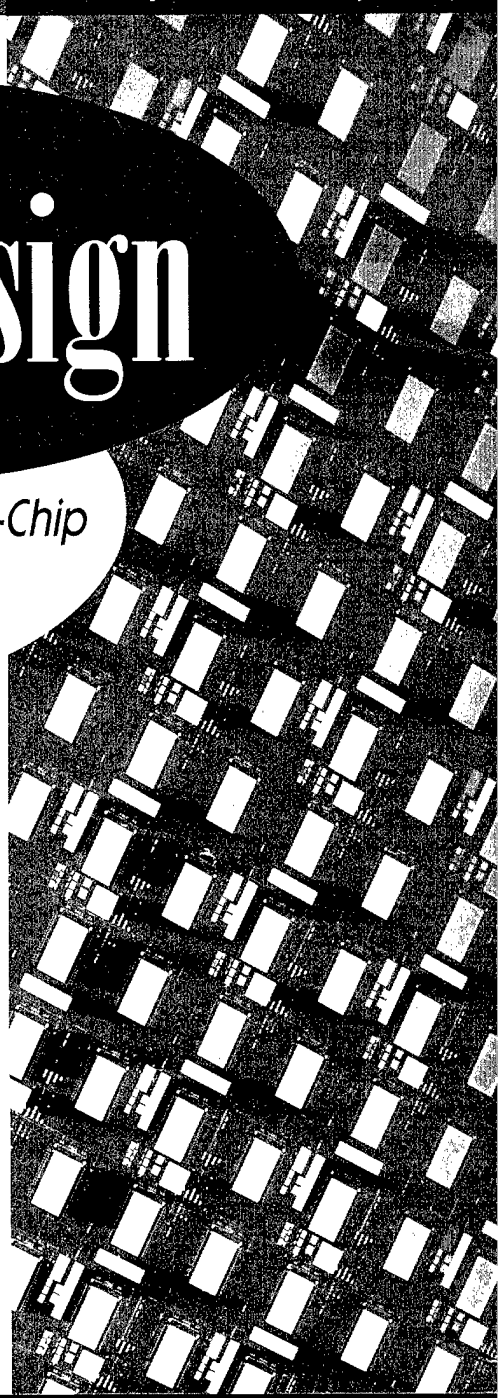
Special Issue on Computational Electronics

Papers presented at The Fifth International Workshop on Computational Electronics

Guest Editor

WOLFGANG POROD
University of Notre Dame

VOL. 8, NOS. 1–4 (1998)



VLSI design

An International Journal of Custom-Chip Design, Simulation, and Testing

Gordon and Breach Science Publishers

VLSI Design

An International Journal of Custom-Chip Design, Simulation, and Testing

Editor In Chief

Dr. George W. Zobrist
Department of Computer Science
University of Missouri-Rolla
1870 Miner Circle
Rolla, MO 65409
Phone: 573-341-4492
Fax: 573-341-4501
email: zobrist@umr.edu
URL: http//www.cs.umr.edu/VLSI

Regional Editors

Dr. Magdy Bayoumi, USA Editor
Center for Advanced Computer Studies
University of Southwestern Louisiana
Lafayette, Louisiana 70504
FAX: (318) 231-5771
email: mab@cacs.usl.edu

Dr. Sunil R. Das, Canadian Editor
Department of Electrical Engineering
University of Ottawa
Ottawa, Ontario
Canada K1N 6N5
FAX: 613-562-5175
email: das@trix.genie.uottawa.ca

Dr. Heaki Kobayashi, USA Editor
Department of Electrical and Computer Engineering
University of South Carolina
Columbia, South Carolina 29208
FAX: (803) 777-8045
email: kobayash@ece.scarolina.edu

Dr. Amiya Nayak, Canadian Editor and Book Review Editor
School of Computer Science
Carleton University
Ottawa, Ontario
Canada K1S 5B6
FAX: 613-520-4334
email: nayak@scs.carleton.ca

Dr. Lalit Patnaik, Southeast Asia Editor
Department of Computer Science/Automation
Microprocessor Applications Laboratory Indian Institute of Science Bangalore, India 560 012
FAX: 91-0812-341683
email: lalit@micro.iisc.ernet.in

Dr. C. P. Ravikumar, Indian Editor
Department of Electrical Engineering
Indian Institute of Technology
Hauz Khas
New Delhi, India 110 016
FAX: +91-11-686-2037
email: rkumar@ee.iitd.ernet.in

Dr. Hiroto Yasuura, Japanese Editor
Department of Information Systems
Interdisciplinary Graduate School of Engineering/Science
Kyushu University
Kasuga Koen, 6-1
Kasuga, Fukuoka 816 Japan
FAX: +81-92-591-2716
email: yasuura@is.kyushu-u.ac.jp

EDITORIAL ADVISORY BOARD

Dr. Ramon Acosta, International Software Systems, Inc., Austin, Texas • **Dr. Don Bouldin**, University of Tennessee, Knoxville • **Dr. Melvin A. Breuer**, University of Southern California, Los Angeles • **Dr. Ralph K. Cavin, III**, North Carolina State University, Raleigh • **Dr. Narsingh Deo**, University of Central Florida, Orlando • **Dr. Srinivas Devadas**, MIT, Cambridge, MA • **Dr. Fausto Distante**, Politecnico di Milano, Italy • **Dr. Hideo Fujiwara**, Advanced Institute of Science and Technology, Nara, Japan • **Dr. Huey-Liang Hwang**, National Tsing-Hua University, Hsin-Tsu, Taiwan • **Dr. Marwan Jabri**, University of Sydney, Australia • **Dr. Osamu Karatsu**, NTT-LSI Laboratories, Atsugi-shi, Japan • **Dr. Amar Mukherjee**, University of Central Florida, Orlando • **Dr. Saburo Muroga**, University of Illinois, Urbana • **Dr. Stephan Olariu**, Old Dominion University, Norfolk, Virginia • **Dr. C. V. Ramamoorthy**, University of California, Berkeley • **Dr. N. Ranganathan**, University of South Florida, Tampa • **Dr. Andrew Sage**, George Mason University, Fairfax, Virginia • **Dr. Majid Sarrafzadeh**, Northwestern University, Illinois • **Dr. X. Song**, University of Montreal, Canada • **Dr. L. Spaanenburg**, University of Groningen, Netherlands • **Dr. Victoria Stavridou**, University of London, Surrey, UK • **Dr. Leon Stok**, IBM T. J. Watson Research Center, Yorktown Hights, NY • **Dr. Earl Swartzlander**, University of Texas, Austin • **Dr. Philip A. Wilsey**, University of Cincinnati, Ohio• **Dr. Si-Qing Zheng**, Louisiana State University, Baton Rouge

AIMS AND SCOPE

This journal, **VLSI Design**, will present state-of-the-art papers in computer-aided VLSI design, analysis, implementation, simulation, and testing. Many contributions will discuss VLSI theory; others will address practical engineering applications, often presenting the theory behind applications and implementations. The journal will present papers on intelligent design, interfacing, new testing techniques, system integration, and other important areas of this fast-paced field. **VLSI Design** will provide a dynamic, high quality international forum for original papers and tutorials by academic, industrial, and other scholarly contributors in VLSI design, testing, implementation, systems, simulation, and analysis. Its scope also includes papers that address technical trends, problem areas, and other pressing issues within the field of VLSI.

Notes for contributors can be found at the back of the journal.

© 1998 OPA (Overseas Publishers Association) N. V. Published by license under the Gordon and Breach Science Publishers imprint. All rights reserved. Except as permitted under national laws or under the photocopy license described below, no part of this publication may be reproduced or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise, or stored in a retrieval system of any nature, without the advance written permission of the Publisher.

Cover photo courtesy of SEMATECH. SEMATECH is a nonprofit research and development consortium of U. S. semiconductor manufacturers in partnership with the Department of Defense. Based in Austin, Texas, SEMATECH's mission is to solve the technical challenges required to keep the U.S. number one in the global semiconductor industry.

A **CD-ROM ARCHIVE** version (which may include visual data such as 3-D animations and video clips) is available for this journal at a nominal extra charge to subscribers.

(Continued on inside back cover)

UNIVERSITY OF NOTRE DAME

Department of Electrical Engineering

September 7, 1998

Defense Technical Information Center
8725 John J. Kingman Road
STE 0944
Ft. Belvoir, Virginia 22060-6218

Subject: Grant Number N00014-97-1-0036

Dear Madam or Sir:

Please find enclosed the required copy of the **Proceedings** for the ***Fifth International Workshop on Computational Electronics*** which took place on the Notre Dame campus May 28-30, 1997.

Sincerely,

Wolfgang Porod

Wolfgang Porod
Professor

Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556-5637

Phone: (219) 631-6376
FAX: (219) 631-4393
e-mail: Wolfgang.Porod@ND.edu

DTIC QUALITY INSPECTED 1

Volume 8, Numbers 1–4 (1998)
Pages 1–572

Special Issue on Computational Electronics

Papers Presented at The Fifth International Workshop on Computational Electronics (IWCE-5)

Notre Dame, Indiana
May 28–30, 1997

Guest Editor
Wolfgang Porod

Sponsored by
National Science Foundation
Office of Naval Research
Army Research Office
Univ. Notre Dame College of Engineering
Univ. Notre Dame Department of Electrical Engineering
Technical Co-Sponsorship, IEEE Electron Devices Society

19980911 051

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

VLSI DESIGN

Volume 8, Numbers 1–4 (1998)

CONTENTS

Guest Editorial xiii
IWCE-5 Committees xv

Device Simulation I

Two-dimensional Carrier Transport in Si MOSFETs 1
Shin-Ichi Takagi

Monte Carlo Simulations of Impact Ionization Feedback in MOSFET Structures 13
Jeff D. Bude

Device Simulation II

Inclusion of Quantum Confinement Effects in Self-Consistent Monte Carlo Device Simulations 21
R. W. Kelsall and A. J. Lidsey

Two-dimensional Modelling of HEMTs Using Multigrids with Quantum Correction 29
Eric A. B. Cole, Tobias Boettcher and Christopher M. Snowden

MOMENTS: The Modular Monte Carlo Environment for Charge Transport Simulation, Overview and Applications 35
Mark Peskin and Christine Maziar

High-Field Hole Transport in Strained Si and SiGe by Monte Carlo Simulation: Full Band Versus Analytic Band Models 41
F. M. Bufler, P. Graf and B. Meinerzhagen

Modeling for ULSI

Interface Roughness Effects in Ultra-Thin Tunneling Oxides 47
D. Z.-Y. Ting, Erik S. Daniel and T. C. McGill

Thermal Effects

Modeling of Thermal Effects in Semiconductor Structures 53
Christopher M. Snowden

Carrier Thermal Conductivity: Analysis and Application to Submicron-Device Simulation 59
A. Greiner, L. Varani, L. Reggiani, M. C. Vecchi, T. Kuhn and P. Golinelli

Molecular Structures

Molecular Wire Interconnects: Chemical Structural Control, Resonant Tunneling and Length Dependence 65
Mathieu Kemp, Vladimiro Mujica, Adrian Roitberg and Mark A. Ratner

Ionic Channels in Biological Membranes: Natural Nanotubes Described by the Drift-Diffusion Equations 75
Bob Eisenberg

Software Development

Writing Research Software in a Large Group for the NEMO Project 79
Gerhard Klimeck, Dan Blanks, Roger Lake, R. Chris Bowen, Chenjing L. Fernando, Manhua Leng, William R. Frensley, Dejan Jovanovic and Paul Sotirelis

Optoelectronics

Theory and Modeling of Lasing Modes in Vertical Cavity Surface Emitting Lasers 87
Benjamin Klein, Leonard F. Register, Karl Hess and Dennis Deppe

Device Simulation III

Cellular Automation Study of Time-Dynamics of Avalanche Breakdown in IMPATT Diodes 93
G. Zandler, R. Oberhuber, D. Liebig, P. Vogl, M. Saraniti and P. Lugli

3D Parallel Finite Element Simulation of In-Cell Breakdown in Lateral-Channel IGBTs 99
A. R. Brown, A. Asenov and J. R. Barker

Comparison of Iteration Schemes for the Solution of the Multidimensional Schrödinger-Poisson Equations 105
A. Trellakis, A. T. Galick, A. Pacelli and U. Ravaioli

Cellular Automata Studies of Vertical Silicon Devices 111
M. Saraniti, G. Zandler, G. Formicone and S. Goodnick

High Performance Computing

X3D Moving Grid Methods for Semiconductor Applications 117
Andrew Kuprat, David Cartwright, J. Tinka Gammel, Denise George, Brian Kendrick, David Kilcrease, Harold Trease and Robert Walker

Multilevel Algorithms for Large-scope Molecular Dynamics Simulations of Nanostructures on Parallel Computers 123
Aiichiro Nakano, Rajiv K. Kalia and Priya Vashishta

Ensemble Monte Carlo and Full-Wave Electrodynamic Models Implemented Self-Consistently on a Parallel Processor Using Perfectly Matched Layer Boundary Conditions 129
Ik-Sung Lim, Robert O. Grondin and Samir El-Ghazaly

Mathematical Models

Applicability of the High Field Model: An Analytical Study Via Asymptotic Parameters Defining Domain Decomposition 135
Carlo Cercignani, Irene M. Gamba, Joseph W. Jerome and Chi-Wang Shu

Smooth Quantum Hydrodynamic Model Simulation of the Resonant Tunneling Diode 143
Carl L. Gardner and Christian Ringhofer

Spherical Harmonic Modeling of a 0.05 μm Base BJT: A Comparison with Monte Carlo and Asymptotic Analysis 147
C.-H. Chang, C.-K. Lin, N. Goldsman and I. D. Mayergoyz

Numerical Examination of Photon Recycling as an Explanation of Observed Carrier Lifetime in Direct Bandgap Materials 153
Joseph W. Parks Jr., Kevin F. Brennan and Arlynn W. Smith

Electronic Structure Calculations Using An Adaptive Wavelet Basis 159
D. A. Richie, P. von Allmen, K. Hess and Richard M. Martin

Quantum Simulation I

Open Problems in Quantum Simulation in Ultra-Submicron Devices 165
D. K. Ferry and J. R. Barker

Theory of Electron Transport in Small Semiconductor Devices Using the Pauli Master Equation 173
M. V. Fischetti

Quantum Transport in Open Nanostructures 179
I. V. Zozoulenko and K.-F. Berggren

Application of the Wigner-Function Formulation to Mesoscopic Systems in Presence of the Electron-Phonon Interaction 185
C. Jacoboni, A. Abramo, P. Bordone, R. Brunetti and M. Pascoli

Quantum Simulation II

Quantum Networks: Dynamics of Open Nanostructures 191
Günter Mahler and Rainer Wawer

A Generalized Monte Carlo Approach for the Analysis of Quantum-Transport Phenomena in Mesoscopic Systems: Interplay Between Coherence and Relaxation 197
Fausto Rossi, Stefano Ragazzi, Aldo Di Carlo and Paolo Lugli

Coherent Control of Light Absorption and Carrier Dynamics in Semiconductor Nanostructures 203
Walter Pötz and Xuedong Hu

A New Computational Approach to Photon-Assisted Tunneling in Intense Driving Fields Based on a Fabry-Perot Analogy 209
Mathias Wagner

Phase Space Boundary Conditions and Quantum Device Transport 215
H. L. Grubin, J. R. Caspar and D. K. Ferry

Quantum Structures

Single-Electron Memories 219
Christoph Wasshuber, Hans Kosina and Siegfried Selberherr

Electron-LA Phonon Interaction in a Quantum Dot 225
T. Ezaki, N. Mori and C. Hamaguchi

Self-Consistent Calculations of the Ground State and the Capacitance of a 3D Si/SiO_2 Quantum Dot 231
A. Scholze, A. Wettstein, A. Schenk and W. Fichtner

An Interband Tunnel Oscillator: Intrinsic Bistability and Hysteresis of Trapped Hole Charge in a Double-Barrier Structure 237
F. A. Buot

Tunneling Between Multimode Stacked Quantum Wires 247
M. Macucci, A. T. Galick and U. Ravaioli

POSTER PRESENTATIONS

Monte Carlo Simulation of Non-Local Transport Effects in Strained Si on Relaxed $Si_{1-x}Ge_x$ Heterostructures 253
F. Gámiz, J. B. Roldán and J. A. López-Villanueva

A β-SiC MOSFET Monte Carlo Simulator Including Inversion Layer Quantization 257
F. Gámiz, J. B. Roldán and J. A. López-Villanueva

Development of a Method for Determining the Dependence of the Electron Mobility on the Longitudinal-Electric Field in MOSFETs 261
J. B. Roldán, F. Gámiz and J. A. López-Villanueva

Quantum Distribution-function Transport Equations in Non-normal Systems and in Ultra-fast Dynamics of Optically-excited Semiconductors 265
F. A. Buot

Applicability of the High Field Model: A Preliminary Numerical Study 275
Carlo Cercignani, Irene Gamba, Joseph W. Jerome and Chi-Wang Shu

Simulation of Bistable Laser Diodes with Inhomogeneous Excitation 283
Gang Fang and Ting-Wei Tang

Intersubband Relaxation in Step Quantum Well Structures 289
J. P. Sun, H. B. Teng, G. I. Haddad, M. A. Stroscio and G. J. Iafrate

Resonances in Conductance Through Tunable Attractors 295
Yong S. Joe and Ronald M. Cosby

Convergence Properties of the Bi-CGSTAB Method for the Solution of the 3D Poisson and 3D Electron Current Continuity Equations for Scaled Si MOSFETs 301
D. Vasileska, W. J. Gross, V. Kafedziski and D. K. Ferry

Wave Function Scarring Effects in Open Ballistic Quantum Cavities 307
R. Akis and D. K. Ferry

Complete RF Analysis of Compound FETs Based on Transient Monte Carlo Simulation 313
S. Babiker, A. Asenov, N. Cameron, S. P. Beaumont and J. R. Barker

Monte Carlo Calibrated Drift-Diffusion Simulation of Short Channel HFETs 319
A. Asenov, S. Babiker, S. P. Beaumont and J. R. Barker

RF Performance of Si/SiGe MODFETs: A Simulation Study 325
S. Roy, A. Asenov, S. Babiker, J. R. Barker and S. P. Beaumont

Ab-initio Coulomb Scattering in Atomistic Device Simulation 331
C. R. Arokianathan, J. H. Davies and A. Asenov

Numerical Evaluation of Iterative Schemes for Drift-diffusion Simulation 337
Mahesh B. Patil, Umberto Ravaioli and Thomas Kerkhoven

Simulation of Si-MOSFETs with the Mutation Operator Monte Carlo Method 343
Jürgen Jakumeit, Amanda Duncan, Umberto Ravaioli and Karl Hess

A New HEMT Breakdown Model Incorporating Gate and Thermal Effects 349
Lutfi Albasha, Christopher M. Snowden and Roger D. Pollard

Rate Equation Modelling of Nonlinear Dynamics in Directly Modulated Multiple Quantum Well Laser Diodes 355
Stephen Bennett, Christopher M. Snowden and Stavros Iezekiel

Temperature Dependence of the Electron and Hole Scattering Mechanisms in Silicon Analyzed through a Full-Band, Spherical-Harmonics Solution of the BTE 361
Susanna Reggiani, Maria Cristina Vecchi and Massimo Rudan

Monte Carlo Simulation of Intersubband Hole Relaxation in a GaAs/AlAs Quantum Well 367
R. W. Kelsall

Bi-Dimensional Simulation of the Simplified Hydrodynamic and Energy-Transport Models for Heterojunction Semiconductors Devices Using Mixed Finite Elements 375
A. Marrocco and Ph. Montarnal

Semiconductor Device Noise Computation Based on the Deterministic Solution of the Poisson and Boltzmann Transport Equations 381
Alfredo J. Piazza and Can E. Korman

Consistent Hydrodynamic and Monte-Carlo Simulation of SiGe HBTs Based on Table Models for the Relaxation Times 387
B. Neinhüs, S. Decker, P. Graf, F. M. Bufler and B. Meinerzhagen

Additive Decomposition Applied to the Semiconductor Drift-Diffusion Model 393
Elizabeth J. Brauer, Marek Turowski and James M. McDonough

Monte Carlo Simulations of High Field Transport in Electroluminescent Devices 401
Manfred Dür, Stephen M. Goodnick, Martin Reigrotzki and Ronald Redmer

Modeling of Radiation Fields in a Sub-Picosecond Photo-Conducting System 407
K. A. Remley, A. Weisshaar, V. K. Tripathi and S. M. Goodnick

New "Irreducible Wedge" for Scattering Rate Calculations in Full-Zone Monte Carlo Simulations 413
John Stanley and Neil Goldsman

Formulation of A Self-Consistent Model for Quantum Well *pin* Solar Cells: Dark Behavior 419
S. Ramey and R. Khoie

Hydrodynamic (HD) Simulations of *N*-Channel MOSFET's with a Computationally Efficient Inversion Layer Quantization Model 423
Haihong Wang, Wei-Kai Shih, Susan Green, Scott Hareland, Christine M. Maziar and Al F. Tasch Jr.

Study of Electron Velocity Overshoot in NMOS Inversion Layers 429
Wei-Kai Shih, Srinivas Jallepalli, Mahbub Rashed, Christine M. Maziar and AL. F. Tasch Jr.

Study on Possible Double Peaks in Cutoff Frequency Characteristics of AlGaAs/GaAs HBTs by Energy Transport Simulation 437
T. Okada and K. Horio

Shell-Filling Effects in Circular Quantum Dots 443
M. Macucci and Karl Hess

Modeling of Shot Noise in Resonant Tunneling Structures 449
G. Iannaccone, M. Macucci and B. Pellegrini

Impact Ionization and Hot-Electron Injection Derived Consistently from Boltzmann Transport 455
Paul Hasler, Andreas G. Andreou, Chris Diorio, Bradley A. Minch and Carver A. Mead

Inclusion of Bandstructure and Many-Body Effects in a Quantum Well Laser Simulator 463
F. Oyafuso, P. von Allmen, M. Grupen and K. Hess

Optical and Electronic Properties of Semiconductor 2D Nanosystems: Self-consistent Tight-binding Calculations 469
Andrea Reale, Aldo Di Carlo, Sara Pescetelli, Marco Paciotti and Paolo Lugli

Transient Phenomena in High Speed Bipolar Devices 475
Michael S. Obrecht, Edwin L. Heasell, J. Vlach and Mohamed I. Elmasry

Transverse Patterns in the Bistable Resonant Tunneling Systems Under Ballistic Lateral Transport 481
V. A. Kochelap, B. A. Glavin and V. V. Mitin

SPIN – A Schrödinger-Poisson Solver Including Nonparabolic Bands 489
H. Kosina and C. Troger

Advantages of Semiconductor Device Simulator Combining Electromagnetic and Electron Transport Models 495
S. M. Sohel Imtiaz, Samir M. El-Ghazaly and Robert O. Grondin

Quantum Transport and Thermoelectric Properties of InAs/GaSb Superlattices 501
J.-F. Lin and D. Z.-Y. Ting

Multiband Quantum Transmitting Boundary Method for Non-Orthogonal Basis 507
G.-C. Liang, Y. A. Lin, D. Z.-Y. Ting and Y.-C. Chang

Calibration of a One Dimensional Hydrodynamic Simulator with Monte Carlo Data 515
O. Muscato, S. Rinaudo and P. Falsaperla

Hyperbolic Hydrodynamical Model of Carrier Transport in Semiconductors 521
Angelo Marcello Anile, Vittorio Romano and Giovanni Russo

A Hydrodynamic Model for Transport in Semiconductors without Free Parameters 527
P. Falsaperla and M. Trovato

Modeling of Poly-Silicon Carrier Transport with Explicit Treatment of Grains and Grain Boundaries 533
Edwin C. Kan and Robert W. Dutton

Formulation of the Boltzmann Equation as a Multi-Mode Drift-Diffusion Equation 539
K. Banoo, F. Assad and M. S. Lundstrom

A 3D Nonlinear Poisson Solver 545
Gyula Veszely

An Alternative Geometry for Quantum Cellular Automata 549
Paul G. Krause, Rachel M. Mueller, P. Douglas Tougaw and Janelle M. Weidner

Electrostatic Formation of Coupled Si/SiO_2 Quantum Dot Systems 555
Per Hyldgaard, Henry K. Harbury and Wolfgang Porod

Electron Transport in One-Dimensional Magnetic Superlattices 559
Zhen-Li Ji and D. W. L. Sprung

Boundary Condition for the Modeling of Open-circuited Devices in Non-equilibrium 567
Joseph W. Parks Jr. and Kevin F. Brennan

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), p. xiii
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Guest Editorial

The Fifth International Workshop on Computational Electronics (IWCE-5) was held on the campus of the University of Notre Dame, May 28–30, 1997. As in previous IWCE meetings, the workshop covered all aspects of advanced simulation of electronic transport in semiconductors and semiconductor devices, especially those which use large computational resources. In particular, IWCE-5 focused on the following three major themes: (i) Device Simulation, (ii) Optoelectronics, and (iii) Quantum Simulation. The scientific program, organized in a single-session format, consisted of invited lectures, contributed talks, and poster presentations. IWCE-5 attracted some 170 registered participants, and featured 15 invited speakers, 40 oral presentations, and 75 poster papers. As a novelty, IWCE-5 included an evening session with hands-on software demonstrations.

The present workshop is the fifth in a series of international meetings which evolved from earlier workshops held under the auspices of the NSF-sponsored National Center for Computational Electronics (NCCE) at the Beckman Institute of the University of Illinois at Urbana-Champaign. The first International Workshop on Computational Electronics was held there on May 28–29, 1992. IWCE-2 was held at the University of Leeds (UK) on August 11–13, 1993. The third workshop took place in Portland, Oregon, on May 18–20, 1994, and IWCE-4 was held in Tempe, Arizona, on October 30–November 1, 1995.

We are grateful for financial sponsorship of IWCE-5 by the National Science Foundation, the Office of Naval Research, the Army Research Office, and the National Center for Computational Electronics, as well as local support by the Notre Dame Department of Electrical Engineering and the College of Engineering. Technical co-sponsorship was provided by the IEEE Electron Devices Society.

We gratefully acknowledge the conference committees for their help in the workshop organization. The International Advisory Committee provided guidance and gave shape to IWCE-5 through the nomination of session topics and invited speakers. The Program Committee selected contributed and poster papers from among the submitted abstracts.

Finally, I would like to thank all the participants for their contributions and for helping to make IWCE-5 a success!

Wolfgang Porod
University of Notre Dame
May 1997

IWCE-5 COMMITTEES

CHAIRMAN

Wolfgang Porod, Univ. Notre Dame

PROGRAM COMMITTEE

Jeff Bude,	Lucent
Bob Dutton,	Stanford Univ.
Chihiro Hamaguchi,	Osaka Univ., Japan
Steve Goodnick,	Arizona State Univ.
Steve Laux,	IBM
Craig Lent,	Univ. Notre Dame
Mark Lundstrom,	Purdue Univ.
Hiroshi Mizuta,	Hitachi Cambridge, UK
Umberto Ravaioli,	Univ. Illinois
Chris Ringhofer,	Arizona State Univ.
Chris Snowden,	Univ. Leeds, UK

ADVISORY COMMITTEE

G. Baccarani,	Univ. Bologna, Italy	Gerald Iafrate,	ARO
Herb Bennett,	NIST	Joseph W. Jerome,	Northwestern Univ.
Felix Buot,	NRL	Thomas Kerkhoven,	Univ. Illinois
Larry Cooper,	ONR	Tom McGill,	Cal Tech
W. M. Coughran, Jr.	Lucent	Donald J. Rose,	Duke Univ.
David Ferry,	Arizona State Univ.	Kent Smith,	Lucent
Wolfgang Fichtner,	ETH Zurich, Switzerland	Christopher Stanton,	Univ. Florida
Max Fischetti,	IBM	Mike Stroscio,	ARO
Bill Frensley,	Univ. Texas at Dallas	Al F. Tasch,	Univ. Texas at Austin
Masao Fukuma,	NEC, Japan	Peter Vogl,	TU Munich, Germany
Hal Grubin,	SRA	Kiyoyuki Yokoyama,	NTT, Japan
Karl Hess,	Univ. Illinois	Naoki Yokoyama,	Fujitsu, Japan

DEVICE SIMULATION I

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 1–11
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Two-dimensional Carrier Transport in Si MOSFETs

SHIN-ICHI TAKAGI

Advanced Semiconductor Devices Research Laboratories, Toshiba Corporation, 8, Shinsugita-cho, Isogo-ku, Yokohama, Japan 235

The importance of 2-dimensional (2D) features of carriers in Si MOSFETs on the device performance is re-examined experimentally and theoretically from the viewpoint of low-field mobility, velocity in high tangential fields and the inversion-layer capacitance. It is confirmed that low-field mobility and inversion-layer capacitance can be understood well in terms of the 2D subbands and the 2D carrier transport. In order to obtain fully-quantitative understanding of low-field mobility, however, it is still necessary to more accurately determine the amount of the scattering parameters in the inversion layer. On the other hand, saturation velocity is considered to be less influenced by the 2D quantization, while it is found experimentally that saturation velocity is slightly dependent on surface carrier concentration.

According to the knowledge of 2-dimensional carrier transport in Si inversion layer, an effective way to have higher current drive is to increase the occupancy of the 2-fold valleys, which have lower conductivity mass, on a (100) surface. From this viewpoint, two device structures, strained Si MOSFETs and SOI MOSFETs with ultra-thin SOI films, are introduced and the behavior of low-field mobility is analyzed through the calculations of the subband structures and phonon-limited mobility.

Keywords: MOSFET, Si, carrier transport, mobility, carrier velocity, inversion layer, 2-dimensional carrier gas, subband structure, strain, SOI

1. INTRODUCTION

It is well recognized that 2-dimensional (2D) features of carriers significantly affect the carrier transport in the inversion layer of Si, which determines the drain current of MOSFETs. The influence of the 2D quantization on the I-V characteristics of MOSFETs appears typically in the following ways. (1) Mobility is modulated through the change of the scattering probability due to the differences in the density-of-states, the form factor of the envelope function of 2D carriers and the scattering parameters. (2) Gate capacitance is decreased by the inversion-layer capacitance, which is determined by the finite thickness of the envelope function. (3) The threshold voltage of MOSFETs is increased by the existence of the subband energy. The influences of (1) and (3) are pronounced more with an increase in the substrate impurity concentration of MOSFETs and the influence of (2) is pronounced more with an decrease in the gate oxide thickness. Thus, the accurate

consideration of the 2D quantization is quite important to quantitatively understand the electrical properties of scaled MOSFETs.

Many theoretical and experimental works [1] have so far been done on the characterization of the 2D carrier transport and, as a result, the understanding of the 2D quantization effects on the MOSFETs performance has been basically obtained, particularly at low temperatures. It seems, however, that the direct experimental evidence to verify the contribution of the 2D quantization at room temperature is still lacking and the quantitative description has not been fully obtained yet. In this paper we re-examine the influence of the 2D quantization on the basic physical quantities to determine I-V characteristics of MOSFETs, the mobility and the saturation velocity, v_{sat}, in the inversion layer, and the inversion-layer capacitance, C_{inv}, through the systematic experiments. In addition, the analysis based on the self-consistent subband calculations is performed to explain the experimental results. The present understanding and the questions still unsolved are presented with comparing the self-consistent subband calculations.

Moreover, based on the knowledge of the 2D carrier transport in the inversion layer of Si, a subband engineering scenario to obtain the higher performance of Si MOSFETs is introduced.

2. SUBBAND STRUCTURE OF Si MOS INTERFACE

Figure 1 shows the schematic diagram of the subband structure for inversion-layer electrons on a (100) surface, on which MOSFETs are commonly fabricated. Electrons confined near Si/SiO_2 interface by normal electric field are quantized in the direction perpendicular to the interface and grouped into electronic subbands as 2D Electron Gas (2DEG). A characteristic point of the subband structure on (100) is that electrons in the inversion layer are classified into two electronic systems, the 2-fold degenerate valleys and the 4-fold degenerate valleys, as shown in Figure 1.

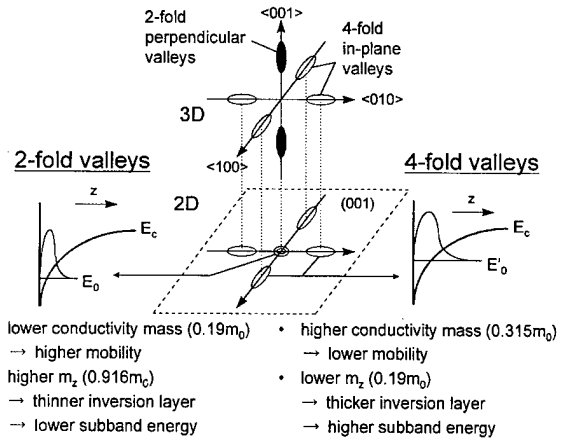


FIGURE 1 Schematic diagram of the subband structures for electrons on a (100) surface and the characteristics of the 2-fold and the 4-fold valleys.

Electrons in the 2-fold valleys have higher mobility due to smaller conductivity mass and thinner inversion-layer thickness due to higher effective mass perpendicular to the Si/SiO_2 interface than those in the 4-fold valleys. As a result, the electrical properties of inversion-layer electrons on (100) need to be understood as the weighed combination over these two electronic systems.

3. IMPACT OF 2D CARRIER GAS ON CARRIER TRANSPORT PROPERTIES IN MOSFETs

3.1. Low Field Mobility

Low field mobility in Si MOSFETs has so far been studied most extensively as the 2-dimensional carrier transport from both the theoretical and the experimental viewpoints. Recently, the inversion-layer mobility has come to be explained in terms of the scattering theory for the 2D carrier gas [2–6], even at room temperature. Figure 2 shows the experimental relationship between electron mobility on (100) and the normal effective field, E_{eff}, as a parameter of the substrate impurity concentration, N_A. E_{eff} is defined by $q(N_{dpl}+\eta N_s)/\epsilon_{Si}$, where N_{dpl} is the surface concentration of the space charge in the depletion region, N_s is the surface carrier concentration and ϵ_{Si} is the

permittivity of Si. One evidence of the 2D quantization effects on the inversion-layer mobility appears as its E_{eff} dependence, which is the origin of the "universal curve". The universality of the inversion-layer mobility for E_{eff}, typically seen in Figure 2, has been verified experimentally over a wide range of the substrate impurity concentration for n- and p-MOSFETs [7–11] fabricated on several surface orientations [12]. The E_{eff} dependence in low and moderate E_{eff} region, which is roughly proportional to $E_{\text{eff}}^{-0.3}$, is thought to be attributable to phonon scattering for 2DEG, while the stronger E_{eff} dependence in higher E_{eff} region is influenced by surface roughness scattering. The mobility limited by intravalley acoustic phonon scattering for 2DEG is represented under the approximation of the single subband occupation [12, 13] by

$$\mu_{\text{ac}} = \frac{q}{m_{\text{c}}} \cdot \frac{\hbar^3 \rho s_{\text{l}}^2}{m_{\text{d}} n_{\text{v}} D_{\text{ac}} k_{\text{B}} T} \cdot W$$

$$\approx \frac{q}{m_{\text{c}}} \cdot \frac{\hbar^3 \rho s_{\text{l}}^2}{m_{\text{d}} n_{\text{v}} D_{\text{ac}} k_{\text{B}} T} \cdot \left(\frac{256 \epsilon_{\text{Si}} \hbar^2}{81 m_3 q^2} \right)^{1/3} \cdot \left(N_{\text{dpl}} + \frac{11}{32} N_{\text{s}} \right)^{-1/3} \propto E_{\text{eff}}^{-1/3} \tag{1}$$

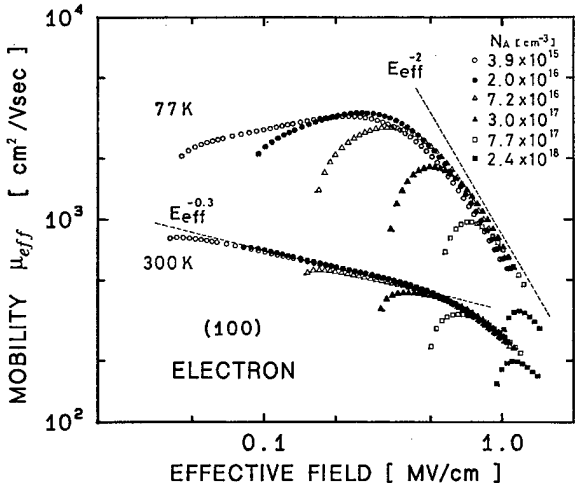


FIGURE 2 Relationship between inversion-layer electron mobility on (100) and the effective normal field, E_{eff}, as a parameter of the substrate impurity concentration. E_{eff} is defined by $q(N_{\text{dpl}} + N_{\text{s}}/2)/\epsilon_{\text{Si}}$ ($\eta = 1/2$), where N_{dpl} is the surface concentration of the space charges and N_{s} is the surface carrier concentration.

m_{c} is conductivity mass, m_{d} is density-of-states mass, n_{v} is the valley degeneracy, D_{ac} is the deformation potential of acoustic phonon, s_{l} is sound velocity. This E_{eff} dependence originates in the fact that the mobility is in proportion to the inversion-layer thickness, which is proportional to $E_{\text{eff}}^{-1/3}$ under η of 11/32. Since the inverse of the inversion-layer thickness determines the energy band width of acoustic phonon that can couple with 2DEG, the thinner inversion layer leads to the higher scattering rate with acoustic phonon.

In order to more quantitatively describe mobility at room temperature, however, it is necessary to take the contributions of all the subbands and the interactions with intra- and intervalley phonons into consideration. At present, the amount of the phonon-limited mobility and its E_{eff} dependence seems not to have been perfectly represented by the theoretical calculations [3, 4, 6]. Figures 3 and 4 show the E_{eff} dependence and the temperature dependence of the experimental and the calculated phonon-limited electron mobility, respectively. The

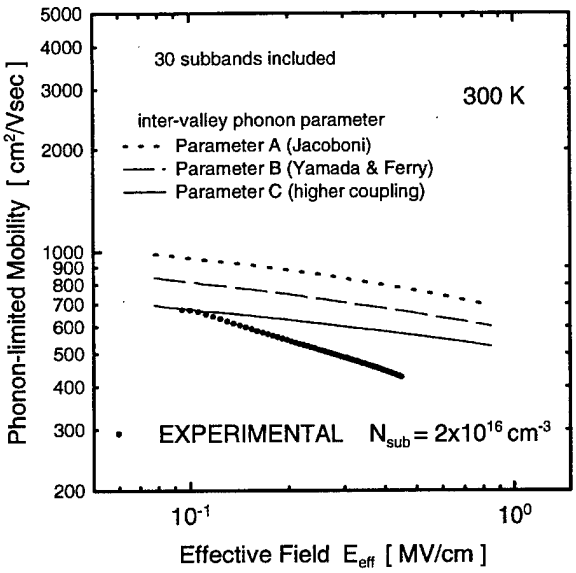


FIGURE 3 Calculated phonon-limited mobility in the inversion layer at room temperature as a function of E_{eff}. Three parameters sets of intervalley scattering, set A, B and C, which are listed in Table I, are compared with the experimental results, which are shown as closed circles.

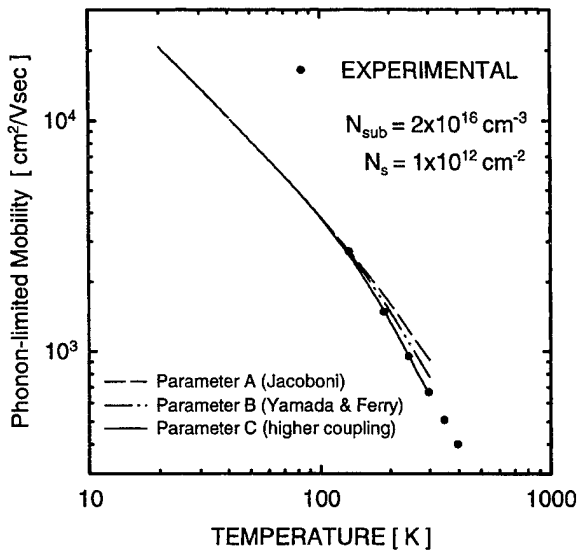


FIGURE 4 Calculated phonon-limited mobility in the inversion layer at N_s of 1×10^{12} cm^{-2} as a function of temperature. Three parameters sets of intervalley scattering, set *A*, *B* and *C*, which are listed in Table I, are compared with the experimental results, which are shown as closed circles.

mobility calculation was carried out under the relaxation time approximation, based on the results of the self-consistent subband calculations [14] including 20 subbands. Three different parameter sets of intervalley phonon scattering, which are listed in Table I, are used in the calculations. It is found that the amount of the mobility calculated under the bulk phonon scattering parameters (parameter set *A* [15] and *B* [16]) is higher than that of the experimental one and the calculated temperature dependence is weaker than the experimental one.

One possible origin of this discrepancy might be the difference in the scattering parameters in the inversion layer from those in bulk, which might be partly due to the difference of phonons relevant to intervalley scattering between 3DEG and 2DEG, coming from the selection rule in the wave-vector conservation. As one example of the different scattering parameters, a parameter set including larger deformation potential of intervalley f-phonon scattering (parameter set *C*) was used for the mobility calculation. The results are also shown in Figures 3 and 4. It is found that the higher coupling with intervalley phonons than in bulk can provide better agreement with the experimental results for both the magnitude of mobility and its temperature dependence. Furthermore, as described later, the analysis of the mobility enhancement in strained Si MOSFETs also suggests the higher coupling between 2DEG and intervalley phonons. The mobility enhancement factor, the ratio of the mobility in strained Si MOSFETs to that in conventional Si MOSFETs, is also explained well by assuming the higher coupling with intervalley phonons (see Fig. 8). In order to more quantitatively obtain the values of the scattering parameters in the inversion layer, other experimental evidences through the direct evaluation of the scattering parameters are required. On the other hand, even if the parameter set *C* is used, the calculated E_{eff} dependence is still

TABLE I Parameters for intervalley phonon scattering models. E_k and D_k are the values of phonon energy and deformation potential, respectively. The parameters of models *A* and *B* were taken from [15] and [16], respectively

parameter model	type of intervalley scattering	E_k[meV]	D_k[$\times10^8$[eV/cm]
		19.0	0.3
model *A*	f	47.5	2.0
		59.1	2.0
(Jacoboni)		12.1	0.5
	g	18.6	0.8
		62.2	11.0
model *B*	f	59.0	8.0
(Yamada and Ferry)	g	63.0	8.0
model *C*	f	59.0	11.0
(this work)	g	63.0	8.0

weaker than the experimental one. Further refinements on the phonon scattering model including the effect of the anisotropic deformation potential [4, 17, 18] are also expected to provide better agreement.

Another unclear point regarding the universal mobility curve is the origin of η in E_{eff}. It has already been reported that the value of η is 1/2 for electrons on (100) [7–9, 11] and 1/3 for holes on (100) [10, 11] at room temperature. Furthermore, it has also been found [12] that the value of η is 1/3 for electrons on (110) and (111), suggesting that the subband structure can affect the value of η significantly. Although, according to (1), η is roughly 1/3, the only one subband is assumed for this equation. Thus, the detailed analysis of η including all the subbands and the contributions of intra- and inter-valley scattering [19] are required to clarify the origin.

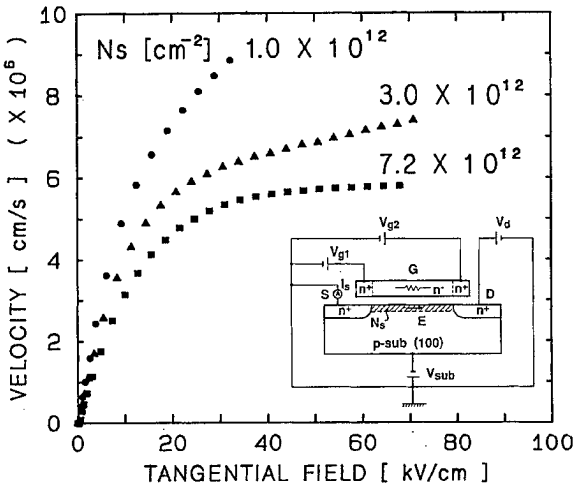


FIGURE 5 Experimental results of the electron velocity–electric field curves in the inversion layer at room temperature as a function of N_s. The direction of the electric field is parallel to ⟨110⟩ axis. The inset figure shows the schematic cross section of resistive gate MOSFETs to evaluate carrier velocity in the inversion layer. In order to realize the uniform N_s along the channel, the voltage of $V_{g1} + V_d$ is applied to the terminal V_{g2}.

3.2. Saturation Velocity

Saturation velocity, v_{sat}, in the inversion layer is a quite important physical quantity in determining the drain current in ultra-short channel MOSFETs. It seems, however, that the magnitude of v_{sat} in the inversion layer and the dependences on the physical parameters such as E_{eff}, N_s and T have not been fully established yet, partly because the large variation is seen in the measured values among the references [20–22]. While several data have suggested that v_{sat} in the inversion layer is lower than that in bulk, it has not been clarified whether the 2D quantization at the Si/SiO_2 interface has any influence on v_{sat} in the inversion layer. Thus, the relationship between the carrier velocity in the inversion layer and the tangential electric field has been evaluated, using MOSFETs with the resistive gate [20], in order to study the effect of the 2D quantization on v_{sat}. Figure 5 shows the experimental v-E curves as a parameter of N_s. The schematic cross section of the device structure is shown as the inset of Figure 5. It is observed that v_{sat} is dependent on N_s and decreases from the value in bulk Si, 1×10^7 cm/s, with increasing N_s, though the saturation of v is not obtained at N_s of 1×10^{12} cm^{-2}.

One of the most important points in experimentally evaluating v_{sat} in the inversion layer is to realize the uniform distribution of N_s and the electric field along the channel of MOSFETs. Note that the velocity determined from the drain current in conventional MOSFETs is inaccurate because of the non-uniformity of N_s along the channel, unless MOSFETs with sufficiently thick gate oxides or SOI MOSFETs with sufficiently thick buried oxides are used [22]. The uniformity of N_s along the channel in the high-resistive gate MOSFETs shown in Figure 5 has been examined [23], using device simulations. As a consequence, it was found that, although a distribution of N_s along the channel is observed, the N_s dependence of v_{sat} in Figure 5 cannot be explained only by this non-uniformity of N_s, but this N_s dependence is provided by some feature inherent to carriers in the channel of MOSFETs.

In order to examine whether the lowering of v_{sat} with an increase in N_s is attributable or not to the 2-dimensional properties of carriers, the substrate

bias dependence of the velocity was measured. It should be noted that the N_s dependence of v_{sat} in Figure 5 can include both effects of the normal electric field and the carrier concentration, because V_g was simply changed to control N_s. The application of both V_g and V_{sub} allows to control N_s and E_{eff} separately. If the 2D quantization plays an important role in the lowering of v_{sat}, E_{eff} is expected to determine v_{sat} in the inversion layer. Figure 6 shows the v-E curves in the inversion layer with and without the substrate bias. The gate voltage in the curve B was adjusted so as to make E_{eff} in the curve B identical to that in the curve C, where the substrate bias of -6 V was applied. On the other hand, the curve A and the curve C have the same N_s. It is found in Figure 6 that the velocity in the curve A is almost the same as in the curve C and, thus, the velocity is determined not by E_{eff} but by N_s itself. This fact means that the lowering of v_{sat} with an increase in N_s might be attributable not to the 2-dimensional properties of carriers, but to an effect inherent to N_s or carrier concentration itself like carrier–carrier scattering.

One possible mechanism responsible for the N_s dependence of v_{sat} might be plasmon scattering [24]. If plasmon decays into electrons through the emission of phonons, this process could become the energy dissipation mechanism of hot carriers under the high electric field. Note that the plasmon energy with the carrier concentration of 5×10^{18} cm^{-3}, which is typically seen in the inversion layer, is 40 meV, which is almost the same as the energies of intervalley phonons seen in Table I. As the carrier concentration increases, the plasmon energy increases and the possibility of the energy loss due to the phonon emission might also increase. Further experimental evidence and theoretical verification on the relevance of plasmon with the energy loss of hot carriers in the inversion layer are strongly required.

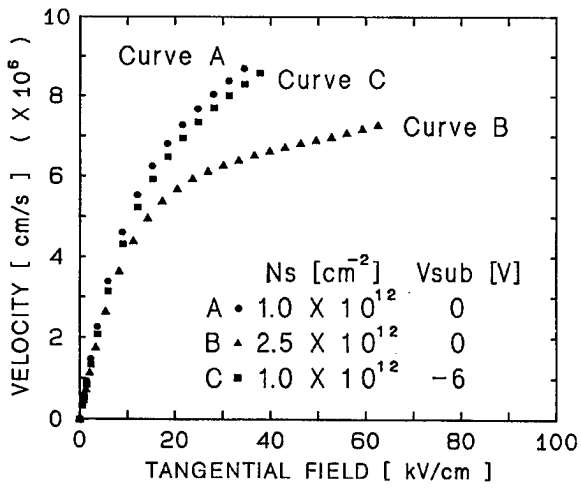


FIGURE 6 Electron velocity–electric field curves with (Curve C) and without (Curve A and B) the substrate bias, V_{sub}. While the value of N_s is in Curve C is the same as in Curve A, the values of E_{eff} in Curve C is the same as in Curve B.

4. INVERSION-LAYER CAPACITANCE

One of the most crucial limitations in the miniaturization of MOSFETs is that the gate capacitance cannot be sufficiently increased by thinning the gate oxide, because of the existence of the inversion-layer capacitance, C_{inv} [25–27]. Since the total gate capacitance, C_{tot}, is described by $C_{ox}/(1+C_{ox}/C_{inv})$, the influence of C_{inv} becomes more serious with reducing the oxide thickness. Thus, it is quite important to clarify the origin of C_{inv} and to quantitatively evaluate C_{inv}. Figure 7 shows the schematic diagram of C_{inv} as a function of N_s under the simple analytical models [27], suggesting that the effect of the 2D quantization on C_{inv} is dominant in higher N_s region. The reason why the 2D quantization is important on C_{inv} is that the finite inversion-layer thickness, which effectively works as the series capacitance to the oxide capacitance, is much thicker in 2D carrier gas than in bulk carriers. In the model of the single subband occupation, C_{inv} due to the finite inversion-layer thickness under the 2D quantization can be simply described by

$$C_{inv}^{thickness} = \frac{\epsilon_{Si}}{Z_{inv}} = \left(\frac{4\epsilon_{Si}^2 q^2 m_3}{9\hbar^2}\right)^{1/3}\left(N_{depl}+\frac{11}{32}N_s\right)^{1/3}$$
$$\approx \left(\frac{11\epsilon_{Si}^2 q^2}{72\hbar^2}\right)^{1/3}\cdot m_3^{1/3}\cdot N_s^{1/3} \qquad (2)$$

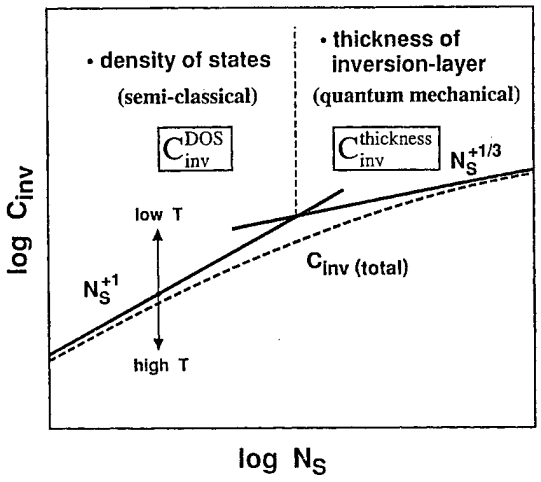


FIGURE 7 Schematic diagram of the physical origins of C_{inv} and the N_s dependence of C_{inv} under the simple analytical models [27].

Here, Z_{inv} is the average thickness of 2DEG, h is the Plank constant and m_3 is the effective mass of Si perpendicular to the Si/SiO_2 interface. According to this formulation, C_{inv} should have the dependence on m_3. Figure 8 shows the experimental C_{inv} at room temperature as a function of N_s for (100), (110) and (111) surfaces, which have the different values of m_3. It is found that the measured values of C_{inv} become smaller in the order of (100), (110) and (111). Actually, it is confirmed that the surface orientation dependence of C_{inv} is explained quantitatively by the $m_3^{1/3}$ dependence. This result is the direct experimental evidence for the fact that C_{inv} at room temperature is determined by the quantum mechanical inversion-layer thickness, because the surface orientation does not make any difference for 3D carrier gas. This result also means that a larger value of m_3 leads to the suppression of the degradation of the gate capacitance due to C_{inv}.

Figure 9 shows the experimental and the calculated C_{inv} for inversion-layer electrons on (100) at room temperature. The calculated C_{inv} was determined directly from $d\Psi_s/dV_g$ through the self-consistent subband calculations without using any definition of the inversion-layer thickness [27]. It is confirmed that C_{inv} is accurately represented by the calculation, meaning that C_{inv} is quantitatively understood even at room temperature in terms of the 2D subband.

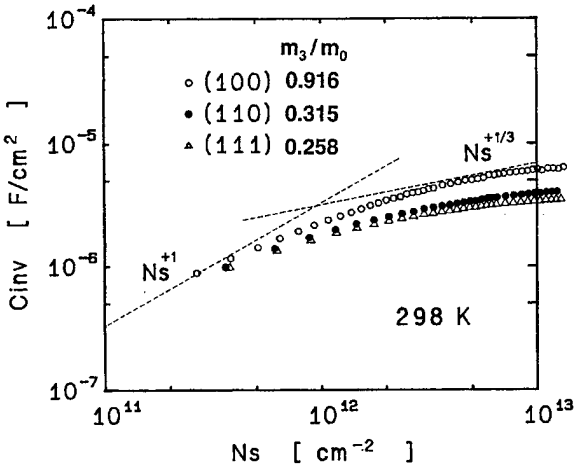


FIGURE 8 Experimental $C_{inv} - N_s$ curves for electrons in the inversion layer on (100), (110) and (111) surfaces at 298 K. The value of effective mass perpendicular to the interface, m_3, is described for each surface orientation.

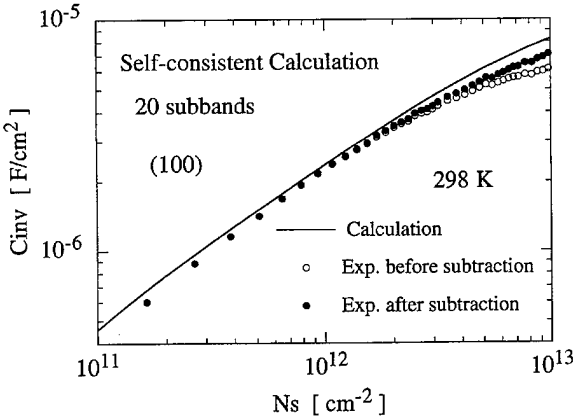


FIGURE 9 Comparison in C_{inv}-N_s curves on (100) surface at 298 K between the self-consistent calculation and the experimental results. The open circles and the closes circles represent the results before and after the influence of the depletion layer in the poly-Si gate was subtracted from the experimental data. In the self-consistent subband calculation, 20 subbands (10 subbands for each subband ladder) were taken into account.

5. SUBBAND ENGINEERING FOR HIGHER PERFORMANCE MOSFETs

In order to obtain the higher current drive in MOSFETs, higher mobility and higher inversion-layer capacitance are required. In terms of the effective mass, these requirements mean lighter effective mass parallel to Si/SiO_2 interface, which increases mobility, and heavier effective mass perpendicular to the interface, which maximizes C_{inv}. From this viewpoint, the 2-fold valleys on a (100) surface are an optimum electronic system in the inversion layer of Si, as summarized in Figure 1. In conventional MOSFETs, however, the energy difference in the lowest subband energy of the 2-fold valleys, E_0, and that of the 4-fold valleys, E_0', is small. As a result, the occupancy of the 2-fold valleys in conventional MOSFETs is not sufficiently large at room temperature. Therefore, an effective strategy to obtain the higher current drive is to increase the energy difference in the lowest subband, $E_0' - E_0$. There are two possible ways to realize such a modulation in the subband structures. One way is to apply the tensile strain in Si substrate. It is known that tensile biaxial strain parallel to the interface causes the band splitting between the 2-fold valleys and the 4-fold valleys. This structure corresponds to strained Si MOSFETs. The other way is to utilize the size effect due to built-in confinement potential such as the band discontinuity at hetero-interfaces. For example, when SOI films in SOI MOSFETs is thinner than the inversion layer in bulk MOSFETs, the subband structures can be significantly modified and the resultant occupancy of the 2-fold valleys can increase by the size effect of the SOI film itself. The effectiveness of this subband engineering is examined through the subband and mobility calculations.

5.1. Strained Si MOSFETs

It is known that tensile strain in Si, which is typically seen in Si grown on relaxed SiGe, causes the band splitting between the 2- and the 4-fold valleys, which leads to the increase in $E_0 - E_0'$ and the resulting higher occupation in the 2-fold valleys. As shown in Figure 10, it has been confirmed experimentally [28], [29] that the inversion-layer mobility in strained Si MOSFETs at room temperature increases up to around twice as high as in conventional MOSFETs with an increase in tensile strain, which is controlled by the Ge content of SiGe substrates. The calculated phonon-limited mobility for 2DEG in strained Si is also shown in Figure 10. Good agreement with the experimental mobility ratio is obtained by using the higher coupling constants with intervalley phonons (parameter set C). This is another evidence for the higher coupling of intervalley phonons with 2DEG in the inversion layer than with bulk electrons, as described in 3.1. The calculated results have also revealed that, in addition to the preferential occupation of the 2-fold valleys, the suppression of intervalley scatter-

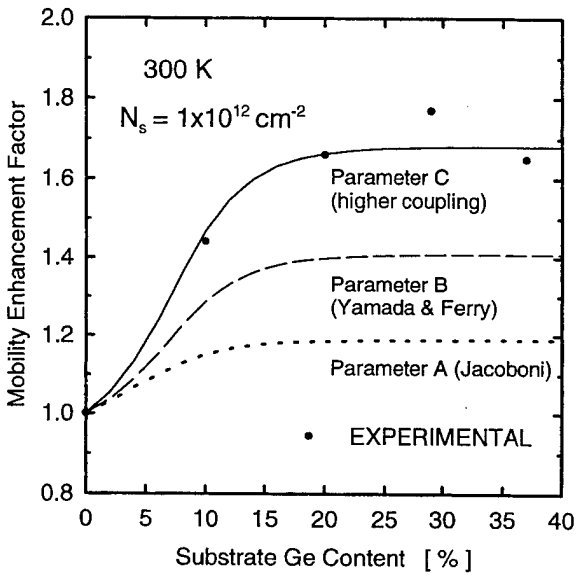


FIGURE 10 Mobility enhancement factor (the ratio of the mobility in strained Si MOSFETs to that in unstrained (conventional) Si MOSFETs) versus the Ge content of a SiGe substrate on which the strained layer is formed. The amount of strain in strained Si is in proportion to the Ge content. The experimental results were taken from [28]. The dotted, dashed and solid lines represent the calculated results using the intervalley phonon scattering parameter sets A, B and C.

ing due to the band splitting leads to the mobility enhancement in strained Si MOSFETs [6, 30]. It has also been found that C_{inv} in the inversion layer of Si slightly increases with tensile strain.

5.2. Ultra-Thin Film SOI MOSFETs

In order to examine the possibility of the modulation of the subband energy and the resultant occupancy of the 2-fold valleys in SOI MOSFETs with SOI films thinner than the inversion layer in bulk MOSFETs, the calculations of the subband structures and the phonon-limited mobility were performed [31]. Figure 11 shows the calculated phonon-limited mobility as a function of the SOI thickness, T_{SOI}. With decreasing T_{SOI} the mobility decreases slightly at first, and increases with decreasing T_{SOI} from 5 nm to 3 nm. With decreasing T_{SOI} from 3 nm, the mobility decreases again. It is found, moreover, that the mobility with the SOI thickness of around 3 nm can be higher than that in conventional MOSFETs. This enhancement is attributable to the fact that, with decreasing the SOI thickness, the electron occupation of the 2-fold valleys becomes higher, because of the increase in E_0'. The calculated occupancy of the 2-fold and the 4-fold valleys is shown as a function of T_{SOI} in Figure 12. It is confirmed that the occupancy of the 2-fold valleys increases with a decrease in T_{SOI} and becomes almost 100% around 3 nm of T_{SOI}. The reason why E_0' increases more rapidly than E_0 with a decrease in T_{SOI} is that the inversion layer thickness of the 4-fold valleys is thicker than that of the 2-fold valleys and, thus, E_0' is more sensitive to the SOI physical thickness than E_0.

It has been also confirmed that C_{inv} in SOI MOSFETs with T_{SOI} thinner than 5 nm is significantly improved, because the inversion-layer thickness is determined by the SOI thickness itself [32]. These results demonstrate that, if the SOI MOSFETs that have the perfectly flat interfaces with the gate oxide and buried oxide can be realized, the performance of SOI MOSFETs with T_{SOI} of 3 nm to 5 nm can be much higher than that of bulk MOSFETs.

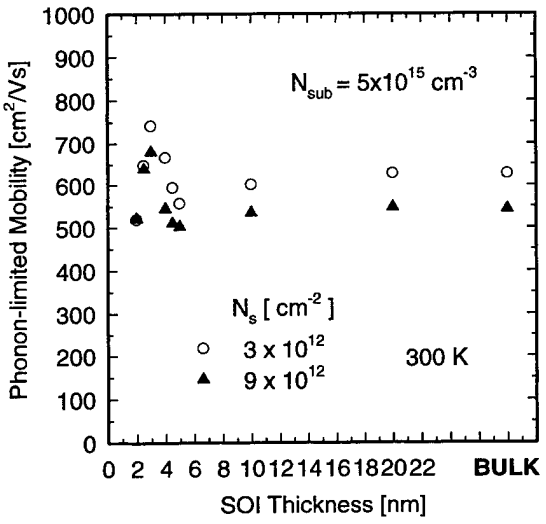


FIGURE 11 Calculated phonon-limited mobility in SOI MOSFETs as a function of SOI thickness at N_s of 3×10^{12} cm^{-2} and 9×10^{12} cm^{-2}. The substrate impurity concentration and the buried oxide thickness are 5×10^{15} cm^{-3} and 100 nm, respectively.

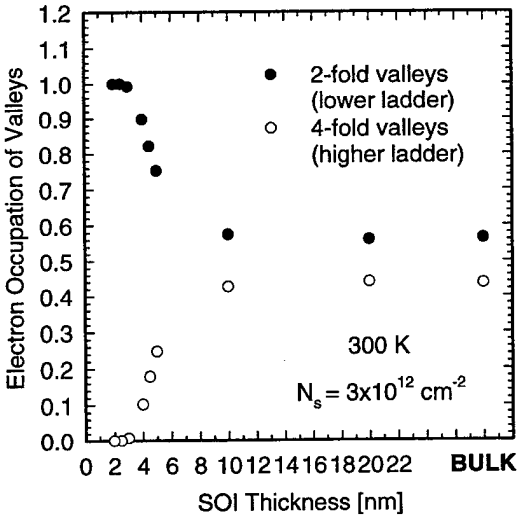


FIGURE 12 Calculated electron occupancy of the 2-fold and the 4-fold valleys as a function of the SOI thickness at N_s of 3×10^{-12} cm^{-2}.

6. CONCLUSION

The importance of the 2D features of the inversion layer electrons on the current drive of MOSFETs at room temperature has been re-examined through the systematic experiments and the self-consistent subband calculations. Inversion-layer mobility and inversion-layer capacitance can be understood well from the viewpoint of the 2D subband and the 2D carrier transport. However, further refinements in the scattering parameters and models are still needed to have fully-quantitative description of inversion-layer mobility. On the other hand, saturation velocity in the inversion layer is considered to be much less affected by the 2D features of carriers.

Based on the knowledge of the Si subband structures, an effective strategy for higher MOSFET performance, the enhancement of the occupancy of the 2-fold valleys on a (100) surface, has been presented. Higher mobility of strained Si MOSFETs and SOI MOSFETs with ultra-thin SOI films, both of which are the typical examples of this subband engineering, has been confirmed by the theoretical calculations.

Acknowledgements

I would like to thank A. Toriumi for his continuous support and helpful suggestion throughout this work. I am also grateful to J. Koga, M. Iwase and H. Satake for their support in the device fabrication and stimulating discussions and to K. Matsuzawa and N. Shigyo for their support in the device simulation. I would also wish to thank J. L. Hoyt, J. J. Welser, K. Rim and Professor J. F. Gibbons, Stanford University, CA, for their helpful suggestions on strained Si MOSFETs. Finally, I acknowledge H. Tango, M. Yoshimi, T. Wada, K. Natori, H. Okano and H. Hara for their continuous encouragement.

References

[1] Ando, T., Fowler, A. B. and Stern, F. (1982). *Rev. Mod. Phys.*, **54**, 437.
[2] Masaki, K., Hamaguchi, C., Taniguchi, K. and Iwase, M. (1989). *Jpn. J. Appl. Phys.*, **28**, 1856.
[3] Jungemann, C., Emunds, A. and Engl, W. L. (1993). *Solid State Electron*, **32**, 1529.
[4] Fischetti, M. V. and Laux, S. E. (1993). *Phys. Rev.*, **B48**, 2244.
[5] Lee, K., Choi, J. S., Sim, S. P. and Kim, C. K. (1991). *IEEE Trans. Electron Devices*, **ED-38**, 1905.
[6] Takagi, S., Hoyt, J. L., Welser, J. J. and Gibbons, J. F. (1996). *J. Appl. Phys.*, **80**, 1567.
[7] Sabnis, A. G. and Clemens, J. T. (1979). In *IEDM Tech. Dig.*, p. 18.
[8] Sun, S. C. and Plummer, J. D. (1980). *IEEE Trans. Electron Devices*, **ED-27**, 1497.
[9] Arora, N. D. and Gildenblat, G. S. (1987). *IEEE Trans. Electron Devices*, **ED-34**, 89.
[10] Watt, J. T. and Plummer, J. D. (1987). In *Symposium on VLSI Tech. Dig.* p. 81.
[11] Takagi, S., Toriumi, A., Iwase, M. and Tango, H. (1994). *IEEE Trans. Electron Devices*, **ED-41**, 2357.
[12] Takagi, S., Toriumi, A., Iwase, M. and Tango, H. (1994). *IEEE Trans. Electron Devices*, **ED-41**, 2363.
[13] Price, P. J. (1981). *Ann. Phys.* (N.Y.), **133**, 217.
[14] Stern, F. (1972). *Phys. Rev. B*, **5**, 4891.
[15] Jacoboni, C. and Reggiani, L. (1983). *Rev. Mod. Phys.*, **55**, 645.
[16] Yamada, T., Zhou, J.-B., Miyata, H. and Ferry, D. K. (1994). *IEEE Trans. Electron Devices*, **ED-41**, 1513.
[17] Mizuno, H., Taniguchi, K. and Hamaguchi, C. (1993). *Phys. Rev. B*, **48**, 1512.
[18] Fischetti, M. V. and Laux, S. E. (1996). *J. Appl. Phys.*, **80**.
[19] Vasileska, D., Bordone, P., Eldridge, T. and Ferry, D. K. (1996). *Physica B*, **227**, 333.
[20] Takagi, S. and Toriumi, A. (1992). In *IEDM Tech. Dig.*, p. 711.
[21] References in [20].
[22] Assaderaghi, F., Sinitsky, D., Borkor, J., Ko, P. K., Gaw, H. and Hu, C. (1997). *IEEE Trans. Electron Devices*, **ED-44**, 664.
[23] Matsuzawa, K., Takagi, S., Suda, M., Oowaki, Y. and Shigyo, N. (1997). In International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), *Tech. Dig.*, p. 217.
[24] Fischetti, M. V. (1991). *Phys. Rev. B*, **44**, 5527.
[25] Baccarani, G. and Worderman, M. R. (1983). *IEEE Electron Devices*, **ED-30**, 1295.
[26] Liang, M.-S., Choi, J. Y., Ko, P.-K. and Hu, C. (1986). *IEEE Electron Devices*, **ED-33**, 409.
[27] Takagi, S. and Toriumi, A. (1995). *IEEE Electron Devices*, **ED-42**, 2125.
[28] Welser, J., Hoyt, J. L. and Gibbons, J. F. (1994). *IEEE Electron Device Lett.*, **EDL-15**, 100.
[29] Welser, J., Hoyt, J. L., Takagi, S. and Gibbons, J. F. (1994). In *IEDM Tech. Dig.*, 373.
[30] Vogelsang, T. and Hofmann, K. R. (1993). *Appl. Phys. Lett.*, **63**, 186.
[31] Takagi, S., Koga, J. and Toriumi, A. (1997). In International Conference on Solid State Devices and Materials (SSDM), *Tech. Dig.*, p. 154.
[32] Takagi, S. and Toriumi, A. (1994). In *Device Research Conference (DRC) Tech. Dig.*, IVA-6.

Authors' Biography

Shin-ichi Takagi was born in Tokyo, Japan, on August 25, 1959. He received the B.S., M.S. and

Ph.D. degrees in electronic engineering from the University of Tokyo, Tokyo, Japan, in 1982, 1984 and 1987, respectively. He joined the Toshiba Research and Development Center, Kawasaki, Japan, in 1987, where he has been engaged in the research on the device physics of Si MOSFETs. From 1993 to 1995, he was a Visiting Scholar at Stanford University, Stanford, CA, where he studied the Si/SiGe hetero-structure devices. Currently, he is with Toshiba Advanced Semiconductor Devices Research Laboratories, Yokohama. He has served on the program committee of the International Electron Device Meeting since 1996. He is a member of the IEEE Electron Device Society, and the Japan Society of Applied Physics.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 13–19
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Monte Carlo Simulations of Impact Ionization Feedback in MOSFET Structures

JEFF D. BUDE

Bell Labs, Lucent Technologies, 600 Mountain Avenue, Murray Hill, NJ 07974, USA

Although impact ionization feedback is recognized as an important current multiplication mechanism, its importance as a carrier heating mechanism has been largely overlooked. This work emphasizes the inclusion of impact ionization feedback in Monte Carlo device simulations, and its implications for carrier heating in sub-micron CMOS and EEPROM technologies.

Keywords: Impact ionization, feedback, flash EEPROM, secondaries, distribution function tails

INTRODUCTION

Monte Carlo (MC) transport simulation is a widely recognized tool capable of accurately obtaining the high energy tail of the electron energy Distribution Function (DF) responsible for impact ionization, oxide degradation and gate currents, I_G, in MOSFETs. MC simulations and measurements of these hot electron effects have shown that the oxide interface DF in MOSFETs is composed of two parts [1]: the conventional channel electron DF, and a secondary electron DF coming from Impact Ionization (II) FeedBack (FB) [2]. The channel electron DF is enhanced by Electron–Electron (EE) scattering, and at low biases is principally populated by this effect. Although the secondary electron DF, generated by the coupled impact ionization of electrons and holes is usually ignored, it can dominate the channel DF tail. The relative importance of the two depends on the channel and substrate doping, drain junction depth, oxide thickness and most importantly, the substrate bias, V_{BS}. This work emphasizes the inclusion of impact ionization feedback in Monte Carlo device simulation, and its implications for carrier heating in sub-micron MOSFETs. In addition, results from a new class of EEPROM devices based on the II FB effect will be discussed.

HEATING BY II FEEDBACK IN MOSFETs

In the following, simulations have been performed using the full-band MC transport simulator, SMC [3], as a post-processor for the device simulator PADRE [4]. Doping profiles and device geometry are computed using the 2D process simulator,

PROPHET [4], and electric fields are computed for a given bias condition by PADRE. Then, given these fields, SMC solves for the device DFs in a manner similar to [5]. For more details, see [6] and references therein. Also, in this section, in order to clarify the essential elements of the II FB process in MOSFETs, EE scattering will not be considered.

Figure 1 illustrates the general phenomenon of II feedback in an *n*MOSFET [2]. Channel electrons, e_1, are injected into the drain where they II forming low energy electron-hole pairs with current II multiplication M_1. The secondary electrons, e_2 leave through the drain while the secondary holes, h_2, diffuse to the Drain-Substrate Junction (DBJ), are heated by its fields and are injected into the substrate where they II again with multiplication M_2 forming e_3 and h_3. The h_3 holes leave through the substrate, but the e_3 electrons fall back through the DBJ and vertical gate controlled potential drops reaching the oxide interface. This process continues with e_3 ionizing leading to a series of pair productions alternating between electrons and holes (II feedback) with multiplications M_3, M_4, etc. Here, $I_B = I_S(1 + M_1 + M_1M_2 + M_1M_2 M_3 + \ldots)$. The DBJ is in breakdown when this series diverges, but in the following, the devices are not in breakdown and all the $M_i < 1$.

A $L_{CH} = 0.25\,\mu$m device with strong II FB effects has been simulated to quantify the feedback effect.

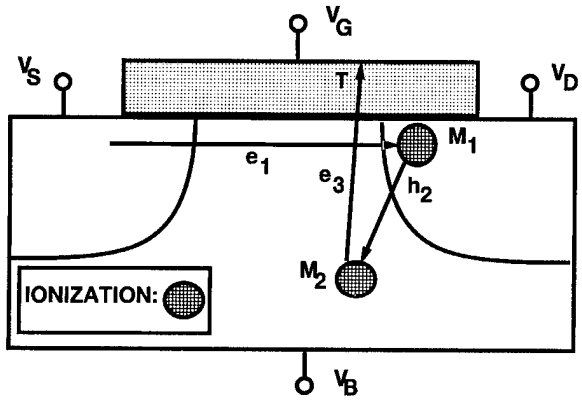


FIGURE 1 Diagram illustrating II feedback in MOSFETs.

Figure 2 shows the potential energy along the channel (from A to B in Fig. 4) and from the drain to the substrate (B to C) with $V_{GS} = 2$ V, $V_{DS} = 3$ V. Figure 3 shows the DF integrated over the device domain. The e_1 DF (no II FB) shows a rapid decay for energies above the pinch-off potential energy drop, $V_p = 2.5$ eV. There are no channel electrons above the oxide conduction band discontinuity of about $\Delta_{OX} = 3.1$ eV, but there are many at the II rate threshold of about 1.1 eV, so $I_B > 0$. The e_3 contribution is also shown in Figure 3 (II FB, full $\delta\varepsilon_{sec}$). The channel electrons ionize with $M_1 = 0.065$, initiating the FB process. The h_2 holes ionize with $M_2 = 0.02$, so that the current carried by the secondary e_3 electrons is quite small: $I_{e_3}/I_{e_1} = M_1M_2 = 0.0013$. Although the DBJ is far from breakdown, there is a broad tail extending to high energies $> \Delta_{OX}$, dominating the e_1 DF above V_p. It is sufficient to consider only the e_3 secondary electron DF since the DBJ is not in breakdown.

The energies of e_3 electrons reaching the oxide interface can be as great as $E_3^{max} = qV_{DB} + qV_{bi} + \delta_{sec}$, where $\delta\varepsilon_{sec}$ is the energy of formation of the e_3 electrons by the II of the h_2 holes, and qV_{bi} is the

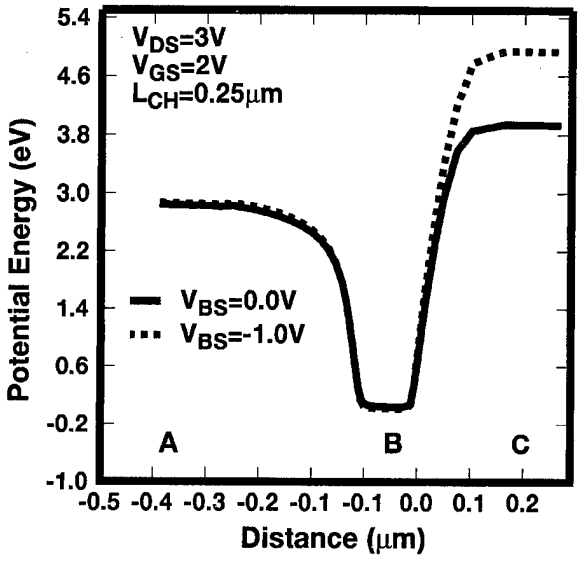


FIGURE 2 Potential energy through $L_{CH} = 0.25\,\mu$m device of a 0.25 μm process from A–B, B–C (Fig. 1).

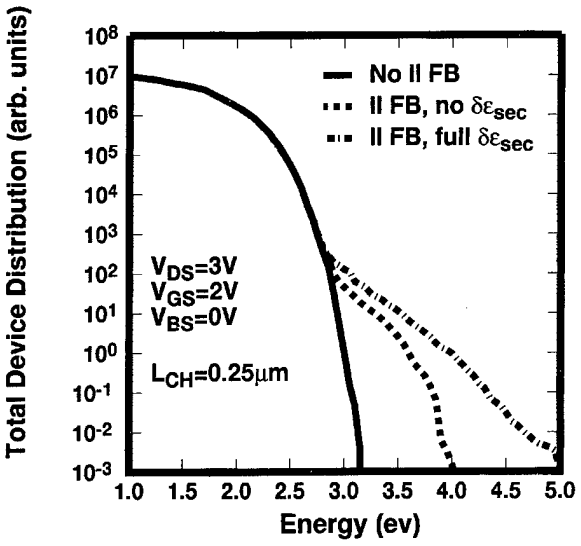


FIGURE 3 DFs integrated over the device domain for conditions and device of Figure 2.

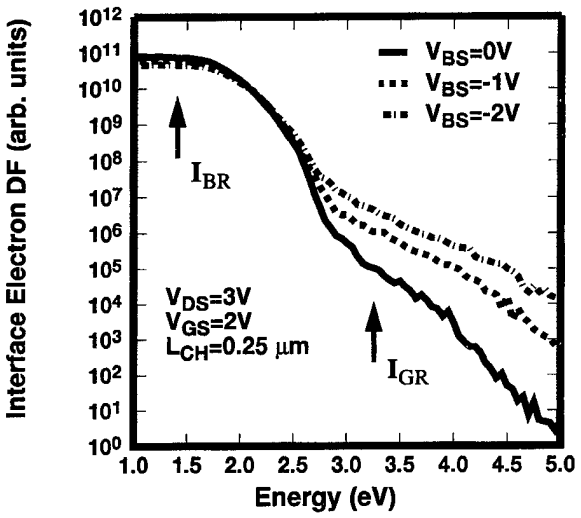


FIGURE 4 DFs integrated at the oxide interface for conditions of Figure 2. Approximate energy thresholds for substrate currents (I_{BR}) and gate currents (I_{GR}) are labelled with arrows.

built-in junction potential. Typically, $qV_{bi} \approx 0.7$ to 0.8 eV. The importance of $\delta\varepsilon_{sec}$ (a table generated during the II rate calculation [7] from the full Si band structure) is clear comparing the full $\delta\varepsilon_{sec}$ curve to that computed with $\delta\varepsilon_{sec}=0$ (Fig. 3). The full curve extends to high energies uniformly, whereas the curve with no energy of creation begins to drop rapidly for energies near $= qV_{DB} + qV_{bi} = 4$ eV.

GATE CURRENTS BY II FEEDBACK

It is clear from above that V_{BS} should have a large effect on the tail of the DF. When a negative substrate bias, $V_{BS} < 0$, is applied, changes in the channel potential are small, but qV_{DB} changes rigidly with V_{BS} (see Fig. 2). Therefore, e_1 heating and I_B are not strongly affected, whereas e_3 heating and I_G should change exponentially. Figure 4 shows the electron DF integrated along the oxide interface for $V_{BS}=0$, -1 and -2 V. The low energy channel DF (below 2.5 eV) changes very little (< 2 X) whereas the secondary tails responsible for I_G change exponentially ($\approx$ 1000 X) as V_{BS} goes from 0 to -2 V. This is confirmed experimentally in Figure 5 which shows $I_{GR}=I_G/I_S$ versus $I_{BR}=I_B/I_S$ for $V_{DS}=V_{GS}$ at various V_{BS}. For a given value of I_{BR} (which represents an equivalent channel heating) I_{GR} increases exponentially as V_{BS} is made more negative. This result contradicts the "effective temperature models" which predict a direct correlation between I_{GR} and I_{BR} [8] based on the assumption that both currents are generated by one thermalized channel DF. Figure 5 also shows a comparison between simulated and measured I_{GR} and I_{BR}. Qualitative trends are clearly reproduced by the simulation. Considering the approximate treatment oxide injection and transport, the remarkably good quantitative agreement with experiment is more questionable.

In addition to enhancing the high energy tail, II FB also changes the spatial distribution of the hot carriers. Figure 6 shows the number of hot secondary electrons with energies above the oxide barrier (3.2 eV) as a function of position along the oxide interface (conditions of Fig. 4). As V_{BS} is made more negative, the hot carriers spread out into the channel, enhancing L_G when V_{GS} is low. For I_G generation, the fields in the oxide at the

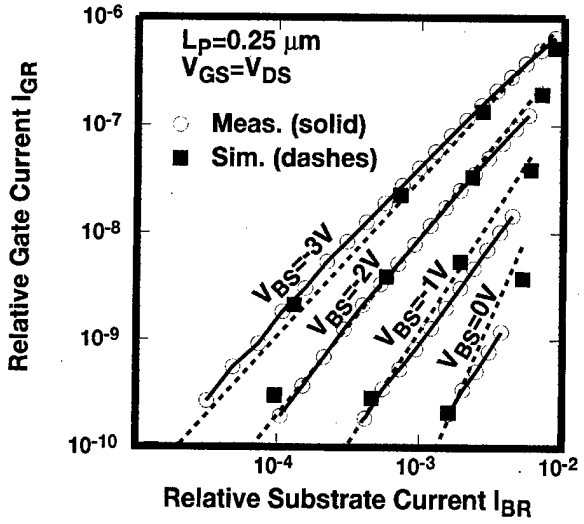


FIGURE 5 Measured and simulated I_{GR} versus I_{BR} for different V_{BS} on a $L_{CH}=0.25\,\mu m$ device. Squares, 0.5 V steps and circles, 0.1 V steps down from $V_{DS} = V_{GS} = 4\,V$.

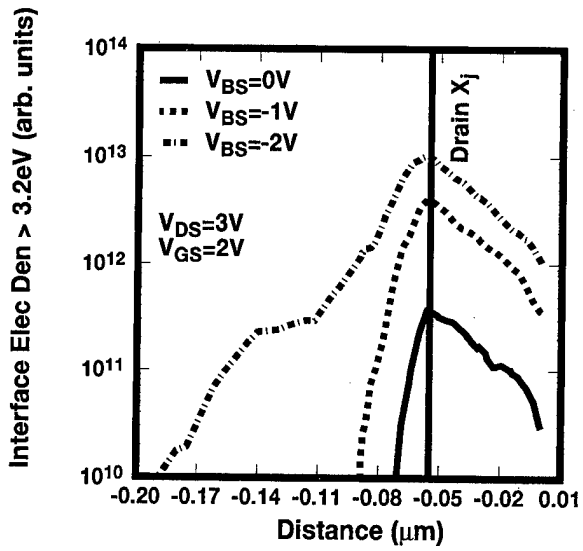


FIGURE 6 DFs integrated at the oxide interface for conditions of Figure 2.

point of injection must accelerate the electrons to the gate electrode. When $V_{GD}<0$, accelerating fields only exist in the channel back towards the source. The secondaries, coming from the substrate, provide oxide injection at these points. This figure suggests that I_G enhancement by V_{BS} is stronger the smaller V_{GS} is compared to V_{DS}. This is confirmed in Figure 7 which shows measured gate injection efficiency, I_{GR} versus V_{GS} as a function of V_{BS} for a 0.25 μm device [9]. Note also that for $V_{BS}=-3\,V$ there is strong gate injection for V_{GS} all the way down to the threshold voltage of the device, 1.3 V. For V_{GS}, 1.3 V, I_G will go to zero with the sub-threshold drain current.

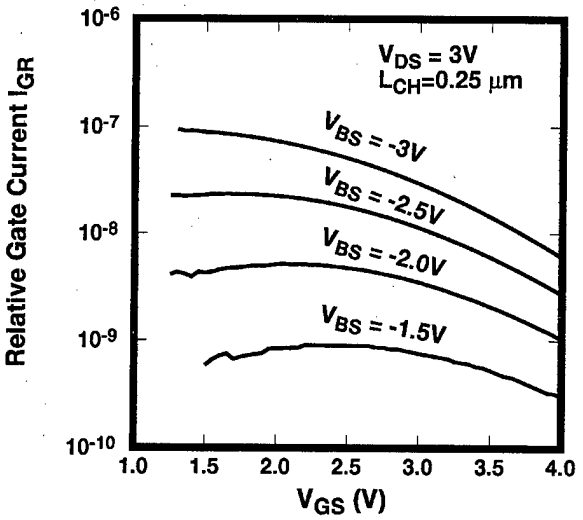


FIGURE 7 Measured I_{GR} for different V_{BS} as a function of V_{GS} on a $L_{CH}=0.25\,\mu m$ device.

TWO-PART DISTRIBUTION FUNCTIONS

As noted in the introduction, EE scattering can extend the channel DF to higher energies than V_p and V_{DS} [1, 10–11]. This EE enhanced channel DF will compete with the II FB at high energies. The relative importance of the two depends primarily on M_2 and T, factors which are controlled primarily by the sharpness of the DBJ junction and the strength of the gate field. M_2 and T become large for shallow junction devices with high channel or junction doping, and as shown above, negative V_{BS}. One easy way to determine the importance of the two is through sensitivity to V_{BS}. The channel DF is fairly independent of V_{BS},

even when EE scattering is included, but the II FB tails are strongly affected by it. In general, for a hot electron effect with a threshold energy of E_{TH}, there will always exist a V_{BS}^* such that for $V_{BS} < V_{BS}^*$, the secondaries will dominate the channel DF. For I_G, E_{TH} is 3.2 eV for points on the interface where gate fields are favorable for injection. When $V_{BS} < V_{BS}^*$, dI_G/dV_{BS} will become large. In devices of older technology generations in which II FB effect tend to be weak, $|V_{BS}^*|$ can be as high as -7 V [1]. Figure 8 shows I_G versus V_{BS} measured on two devices. The curves have been normalized by the same factor for both devices. DEV1 is a 0.25 μm device with strong FB effects similar to that of Figure 3, and DEV2 is a $L_{CH} = 0.32$ μm device optimized for high I_G (see below). The measured data was obtained by measuring floating-gate V_{TH} shifts, and the simulations were performed with both II FB and EE scattering. In DEV1, the FB DF begins to dominate I_G for $V_{BS} < -0.5$ V as shown by both measurement and simulation. In DEV2, the FB DF controls I_G even at $V_{BS} = 0$ V. Oxide DFs for DEV1 in Figure 8 are shown in Figure 9.

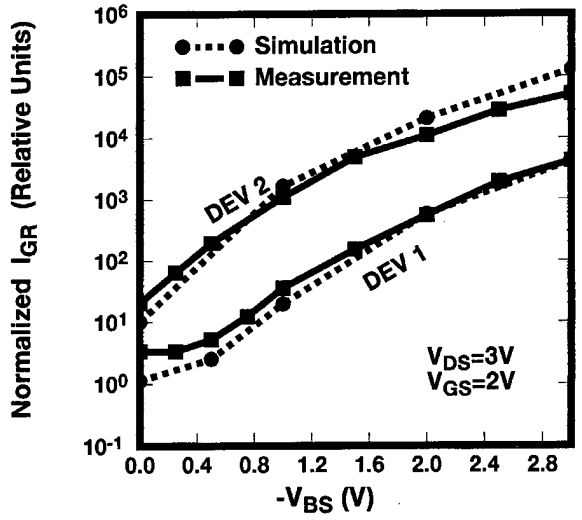


FIGURE 8 Measured and simulated I_G for two different devices: DEV1, similar to that in Figure 2, and DEV2 a $L_{CH} = 0.32$ μm device optimized for high I_G.

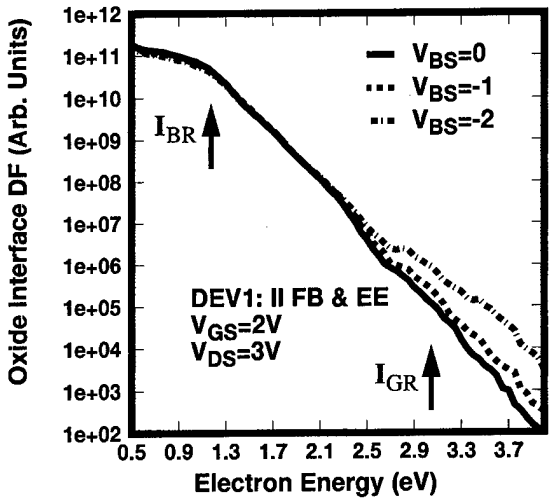


FIGURE 9 Simulated interface DFs for different V_{BS} on DEV1 for Figure 7 including II FB and EE scattering. Approximate I_B and I_G thresholds are labelled as in Figure 4.

THE CHISEL FLASH EPROM CELL

Nonvolatile memories (Flash EPROMs) use hot carrier gate injection for programming. Previous devices have relied on channel hot electron injection for programming which requires high V_{GS} (10 V or more) and V_{DS} (5–6 V) to achieve adequate programming times of 100 μs or faster. MC simulations and measurements like those shown above suggested that the II FB effect could be utilized to greatly improve programming performance for low V_{DS} and V_{GS}, leading to a lower power, better scaled EPROM cell. Based upon the physical understanding provided by these simulations, Flash EPROM devices and a writing methodology were designed to optimize the II FB process to produce a CHISEL (channel initiated secondary electron) EPROM cell [9].

Simulation optimization of the cell concentrated on the following: heavily doped junctions increase substrate current, hole multiplication, M_3 and e_3 electron transmission to the gate, T; shallow junctions enhance M_1 and T; and, $V_{BS} < 0$ enhances M_2 and T. CHISEL memory devices (Fig. 10) were fabricated by forming a stacked gate on nominal 0.25 μm nMOS devices exhibiting thin oxides and

shallow junctions (tox = 6 nm, xj = 75 nm). Because this device is based on a fully scaled CMOS technology, it is ideal for scaling into the deep submicron regime. Boron halos were added to increase junction doping, hence, M_2 and T. Most importantly, a $V_{BS} < 0$ writing scheme was adopted.

Figure 11 shows CHISEL cell programming transients for $V_{DS} = -V_{BS} = 2.5$ and 3 V and a control gate bias of 4 V; efficient programming for times less than 100 μs can be attained with biases much lower than standard Flash devices require. Note also that the cell has strong programming until it turns itself (high I_G with floating gate potentials down to the cell threshold voltage – Fig. 7). Figure 12 highlights the strong writing time dependence on V_{BS} as anticipated.

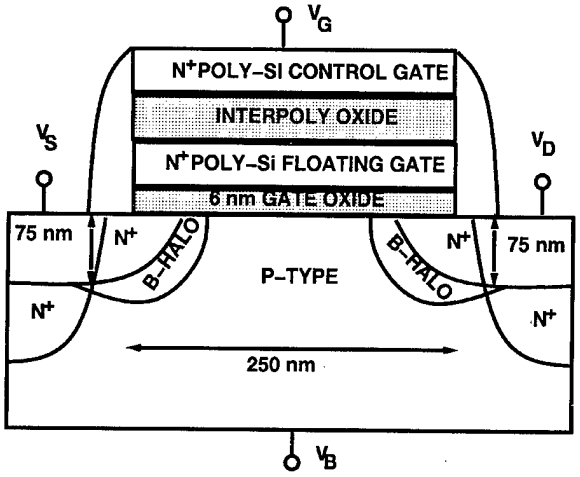


FIGURE 10 Stacked-gate CHISEL EEPROM device structure.

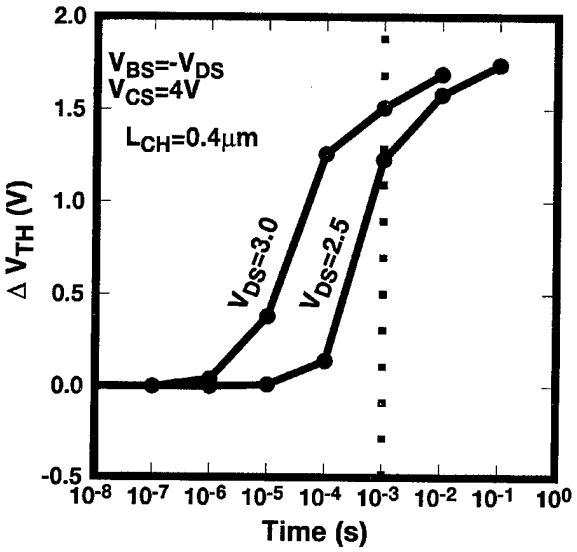


FIGURE 11 Measured programming transients for a CHISEL EEPROM cell.

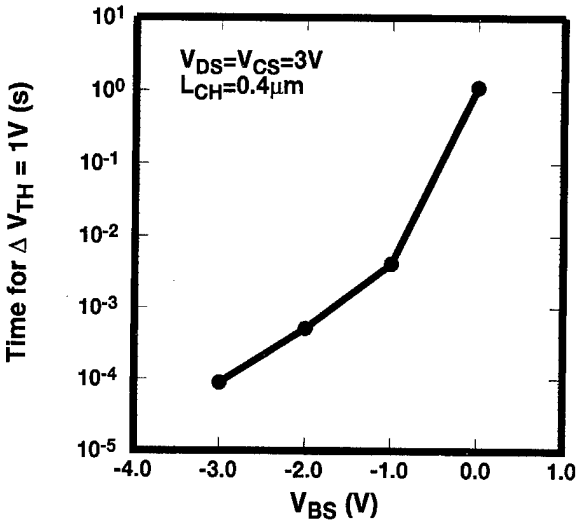


FIGURE 12 Time to reach a vth change of 1 V during programming of a CHISEL EEPROM cell.

CONCLUSIONS

Physically based MC transport simulation have clarified the importance of II FB and the two-part nature of the high energy DF tails in sub-micron MOSFETs, showing good agreement with experiment. This knowledge has led to the creation of a new class of scalable low voltage, low power EPROM devices.

References

[1] Bude, J. D. *et al.* (1996). *1996 IEDM Tech. Digest*, p. 865.
[2] Bude, J. D. (1995). *1995 Symposium on VLSI Technology*, p. 125.
[3] Bude, J. D. and Smith, R. K. (1994). In *Semiconductor Science and Technology*, HCIS 8, 9 5S, p. 840.

[4] Pinto, M. R. *et al.* (1992). *IEDM Tech. Digest*, p. 923.
[5] Higman, J. M. *et al.* (1989). *IEEE Trans. Elec. Dev.*, **ED-36**, p. 930.
[6] Bude, J. D. (1995). *Proc. of 1995 Int. Conf. on Solid State Devices and Materials*, pp. 228–230.
[7] Bude, J. D. and Mastrapasqua, M. (1995). *IEEE Elec. Dev. Lett.*, **16**, pp. 439–441.
[8] Tam, S. *et al.* (1984). *IEEE Trans. Elec. Dev.*, p. 1116.
[9] Bude, J. D. *et al.* (1995). *1995 IEDM Tech. Digest*, p. 989.
[10] Abramo, A. *et al.* (1995). *IEDM Tech. Digest*, p. 301.
[11] Fischetti, M. V. and Laux, S. E. (1995). *IEDM Tech. Digest*, p. 305.

Authors' Biography

Jeff Bude was born in St. Louis, Missouri on July 26, 1966. He received the B.S. degree with highest honors in Electrical Engineering at the University of Illinois, Urbana, in 1987. He received the M.S. degree in 1989 and the Ph.D. degree in 1992, both from the University of Illinois, Urbana. In 1992 he joined Bell Laboratories in Murray Hill, NJ as a member of the technical staff in the ULSI Technology Research Department. He has been involved in ULSI device simulation and research emphasising hot carrier effects and reliability issues.

Address: Bell Labs, Lucent Technologies, Room 2D-307B 600 Mountain Avenue, Murray Hill, NJ 07974 Phone (voice): (908) 582-3615 Phone (FAX): (908) 582-6000 e-mail: bude@lucent.com.

DEVICE SIMULATION II

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 21–27
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Inclusion of Quantum Confinement Effects in Self-Consistent Monte Carlo Device Simulations

R. W. KELSALL* and A. J. LIDSEY

Institute of Microwaves and Photonics, School of Electronic and Electrical Engineering, University of Leeds, Leeds LS2 9JT, UK

The design of Monte Carlo FET simulations is discussed, with specific attention to the methods used to describe quantum confinement effects. A new model is presented, which employs self-consistent coupling of Schrodinger, Poisson and Monte Carlo algorithms, and explicit calculation of the scattering rates between confined and unconfined states. Comparisons between the new model and a standard semi-classical Monte Carlo model are presented for a 0.1 μm gate-length $In_{0.52}Al_{0.48}As/In_{0.53}Ga_{0.47}As/InP$ MODFET. Whilst the quantum model yields minor corrections in the predicted output characteristics, it is found that these results can be achieved without repeated iterations of the Schrodinger equation.

Keywords: Monte Carlo, HEMT, MODFET, quantum

1. INTRODUCTION

The methodology for developing semiconductor device simulations using Monte Carlo algorithms is now reasonably mature. The scheme of coupling an Ensemble Monte Carlo particle simulation to a 2-dimensional (2D) Poisson Solver has been successfully applied to both FETs and HBTs, and now represents the standard method in this subject area [1–2]. Numerous enhancements to the basic model have been developed to obtain greater accuracy–in particular the use of exact [3] or parameterised [4] 'full bandstructure' models.

However, the standard Monte Carlo device simulation is semi-classical: electrons and holes are modelled as discrete particles, whilst scattering between states is described using quantum mechanical perturbation theory. It has long been appreciated that in certain situations – such as the MOSFET inversion layer and the MODFET channel layer – quantum confinement effects may be significant. Inclusion of such effects requires, in principle, self-consistent solution of the 2D Poisson and Schrodinger equations, and the Monte Carlo transport algorithm; a task which appears far too computationally expensive for practical

* Corresponding author.

application. Consequently, there have been very few attempts to incorporate quantum confinement. In this paper, we review briefly some previous Monte Carlo FET models in which the problem has been addressed, and then discuss our own Monte Carlo/Poisson/Schrodinger MODFET simulator.

2. REVIEW OF MONTE CARLO FET MODELS

Quantum confinement effects were actually considered in one of the earliest Monte Carlo MODFET models. Ravaioli and Ferry [5] calculated the first two quantised energy levels in the MODFET channel using a triangular well approximation. This yielded an analytic solution parameterised by the perpendicular (gate-substrate) electric fields thus, the energy level values could be easily updated when the Poisson equation was re-solved, and the effect of potential variation along the channel could be included. The scattering rates between quantum confined states were parameterised by energy level separation and carrier density, and hence could also be updated. Clearly, the main shortcoming of this otherwise elegant approach is the validity of the triangular well approximation these devices.

Park and Brennan [6] developed a MODFET simulator which also included solutions of the Schrodinger equation for the first two subbands in the channel. The eigenvalues were obtained from a numerical solution of the 1-dimensional (1D) Schrodinger equation, whilst the scattering rates were obtained using approximate variational wavefunctions. However, the Schrodinger equation was solved assuming an average 2D carrier density in the channel, hence no position dependence was included.

A key aspect of including quantum confinement in FET simulations is the method used for describing transitions between (quantum) confined and unconfined (classical or bulk-like) states. This problem is, itself, and artifact of the simulation model: in reality *all* states can be described using quantum mechanics; however, the use of Monte Carlo particle algorithm leads inevitably to a classical model of electron kinematics. In both the above models, electrons outide the channel were modelled as semi-classical particles. Electrons in the channel layer which acquired a kinetic energy in excess of the barrier height were also 'converted' into classical particles.

A similar scheme was used by Fischetti and Laux, in their simulation of inversion layer MOSFETs [7]. Their simulation code includes Monte Carlo, 2D Poisson and 1D Schrodinger algorithms, which are solved self-consistently. The scattering rates between confined states are calculated using the exact numerical wavefunctions. However, Fischetti and Laux also conclude, as above, that it is not practical to describe all electrons in the device by Schrodinger eigenfunctions. Hence their model, too, includes transitions between two alternative descriptions of electronic states. Scattering rates were calculated according to the nature of the initial state: for unconfined electrons, rates for transitions between pairs of bulk-like states were used, whereas, for electrons confined in the channel, rates for transitions between pairs of quantum confined states were used, irrespective of the energy of the final state. Such a scheme again necessitates 'conversion' of confined electrons into unconfined, and vice versa. These 'conversions' violate energy and momentum conservation, because the quantised spectrum of states in the channel has fewer degrees of freedom than the continuum to bulk states. Also, the unconfined 'conversion' process contained no physics to specify in which subband the 'converted' electron should reside.

3. A NEW SELF-CONSISTENT SCHRODINGER-POISSON-MONTE CARLO MODEL

Our MODFET model also includes self-consistently coupled Schrodinger, Monte Carlo and Poisson algorithms. However, in an attempt to overcome some of the problems associated with

conversion between the confined and unconfined descriptions of electrons states, we have designed an alternative model of the scattering and kinematics involved in electron transfer in/out of the channel.

The Schrodinger equation is solved numerically in a series of 1D slices between source and drain, but self-consistency between the electrostatic potential and quantum-mechanical charge distribution is maintained. The scattering rates between confined states in the MODFET channel are calculated *at each position along the channel* using the exact numerical wavefunctions, and are recalculated periodically throughout the simulation.

Electrons outside the channel layer are modelled as classical particles. Within the channel layer, both particle-like electrons and quantum confined electrons are permitted. Quantum confined states are calculated for the first n subbands in the channel: higher subbands are assumed to merge into a continuum, characterised by bulk-like states. Transitions between confined and continuum states are described solely by scattering processes. For confined-continuum scattering, the final state is described by a wavepacket which has a single **k**-value, a constant amplitude within the channel layer, and zero amplitude outside. A bulk-like density of final states is assumed. After such a scattering event, the final state is modelled as a classical particle with a well defined energy and wavevector: its real space co-ordinate (in the direction prependicular to the heterolayers) is chosen randomly within the channel.

Conversely, for any particle which enters the channel there is a non-zero probability of scattering into a confined state. Transition rates are calculated using, for the *initial* state, a wavepacket as described above, and a Schrodinger eigenfunction for the final state. The final density of states is that appropriate to the quantum confined subband.

Transitions between confined and continuum states are mediated by all possible interactions: acoustic and polar optical phonon, and alloy disorder scattering. Whilst this approach avoids some of the difficulties with energy and momentum conservation discussed above, we have clearly deployed new approximations, such as the assumption of a contiunuous spectrum of states and the form of the continuum wavepacket. Our choice of wavepacket function appears to best reflect the bulk-like form of incoming and outgoing particles. No confinement effects are included for the X and L valleys, and scattering rates from confined Γ states to the (unconfined) X and L valleys are calculated explicitly. In practice we have included just the lowest three subbands in the channel: a fourth subband can be included without too much computational expense, but its population is invariably small, and thus will have little effect on the simulation results.

4. SIMULATION PERFORMANCE

To test our simulation, we chose for simplicity a 3-layer MODFET geometry. The materials system was lattice matched $In_{0.52}Al_{0.48}As/In_{0.53}Ga_{0.47}As/InP$, which has large conduction band offsets and is therefore more likely to be affected by quantum confinement. No cap layer was included in the simulation: injection and collection were modelled using ohmic source and drain regions which spanned both the supply and channel layers (see Fig. 1).

Our first objective was to ascertain the importance of quantum confinement by comparing results of our new code with those from our semi-classical Monte Carlo model. On a microscopic level, we may expect two principal differences: in the distribution of charge across the channel width, and in the electron scattering rates. Figure 2 shows a comparison of the scattering

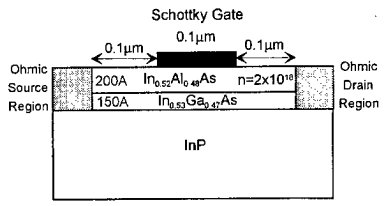


FIGURE 1 Schematic diagram of the simulated 0.1 μm gate-length InP MODFET.

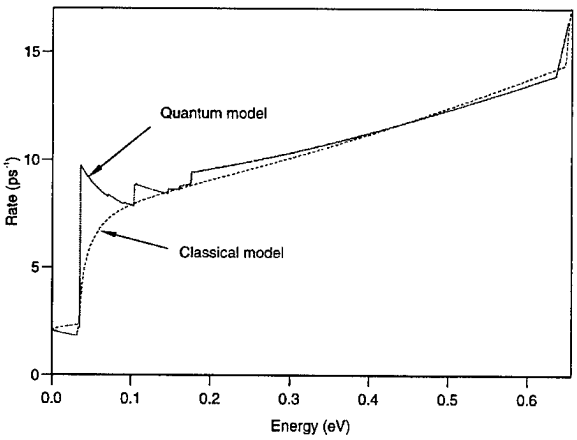


FIGURE 2 Full curve: total (phonon plus alloy) scattering rate for an electron in the lowest subband in the channel. (The impurity scattering rate was not included in this diagram since, although the rates (at low electron energies) are large, the effect of impurity scattering is relatively small, as it gives no energy relaxation and relatively little momentum relaxation). The electron wavefunctions were calculated at the mid-point of the channel, for gate and drain biases of $V_G = 0.2\,V$, $V_D = 1.0\,V$. Broken curve: scattering rate calculated in the standard semi-classical model.

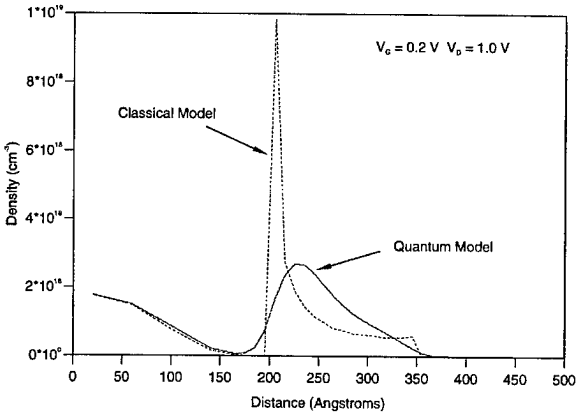


FIGURE 3 Electron distribution across a vertical slice in the source region of the channel. The horizontal scale measures distance from the top of the InAlAs supply layer downwards towards the substrate, with the channel located between $d = 200 - 350$ Å.

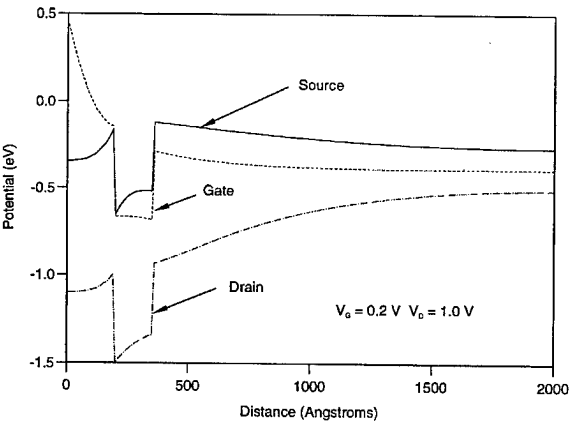


FIGURE 4 Potential (conduction band-edge) profiles along vertical slices taken in the source, gate and drain regions respectively. Horizontal scale as in Figure 2.

rates for the two models. The rates are very similar at high energies (the shoulder at $E \approx 0.65$ eV marks the onset of intervalley scattering, which has identical rates in both models). At low energies the total rate in the quantum model exceeds that in the classical model, due to the difference in the 2D and 3D densities of states.

Figure 3 shows the charge distrubution at the source end of the channel in the quantum and classical simulations. The delocalised nature of the electron wavefunction results in a broader charge distribution in the quantum model, with much less charge accumulation at the channel/supply layer interface ($d = 200$ Å) – where the channel potential is lowest.

In Figures 4–6 we examine the extent of channel non-uniformity. Figure 4 shows typical potential profiles taken from the source, gate and drain regions of the device. Figure 5 shows the wavefunctions for the first eigenstate at the same three locations. There is very little difference in the wavefunctions in the source and drain regions, despite the much lower potential at the drain end of the device, but the peak in the wavefunction under the gate is shifted, due to the altered shape of the channel potential well. However, Figure 6 shows that the shift in wavefunction position leads to only a few percent difference between the confined–confined scattering rates at different positions along the channel. For this bias configuration the same observations apply to scattering from higher subbands too: thus it appears that channel non-uniformity has relatively little effects on scattering.

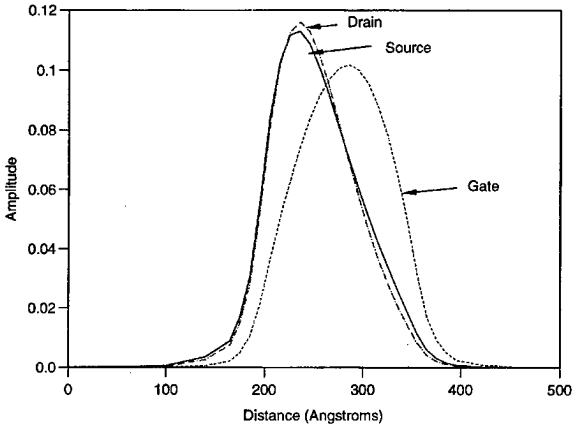


FIGURE 5 Wavefunctions of the lowest eigenstate in the source, gate and drain regions of the device: $V_G = 0.2\,V$ $V_D = 1.0\,V$.

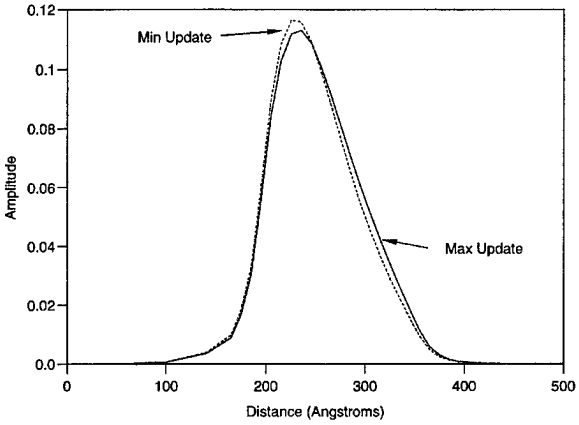


FIGURE 7 Ground-state wavefunction at the source end of the channel, obtained using minimum and maximum updates (see text). $V_G = 0.2\,V$, $V_D = 1.0\,V$.

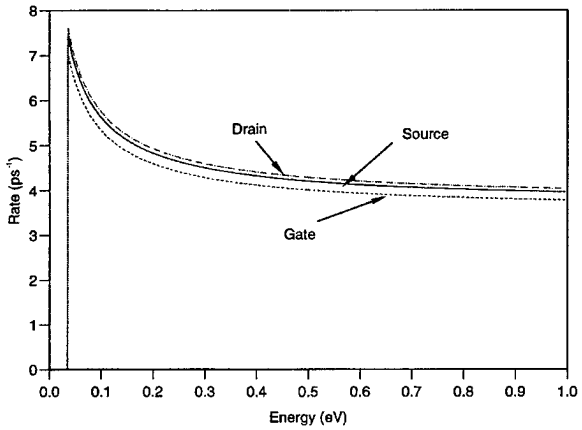


FIGURE 6 Confined–confined (intrasubband) polar optical phonon scattering rate for an electron in the lowest subband, in the source, gate and drain regions of the channel. $V_G = 0.2\,V$, $V_D = 1.0\,V$.

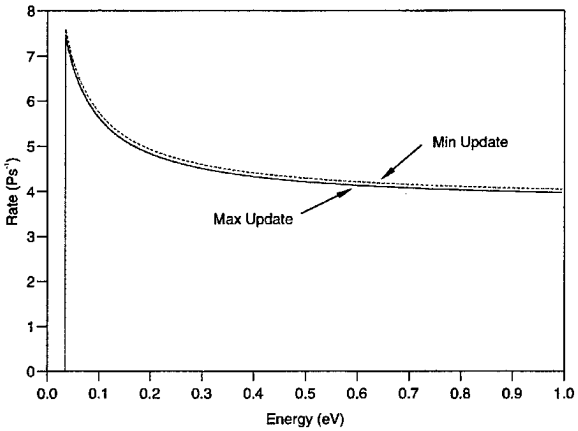


FIGURE 8 Confined–confined (intrasubband) polar optical phonon scattering rate for an electron in the lowest subband, at the source end of the channel, obtained using minimum and maximum updates. $V_G = 0.2\,V$, $V_D = 1.0\,V$.

The main computational overhead in a self-consistent simulation is the frequency at which the relevant parameters must be re-calculated. In Figures 7 and 8 we compare wavefunctions and scattering rats from simulations with minimum and maximum self-consistency. In both cases the simulation was run using the classical model, until a steady state was attained. In the minimum update case, the Schrodinger equation was then solved once, at each position along the channel, using a time-averaged potential. The channel scattering rates were calculated using the Schrodinger wavefunctions. The simulation was then continued, using the quantum model, but with no re-calculation of the channel eigenstates or scattering rates. In the maximum update case, the simulation was continued, re-solving the Schrodinger equation at every Poisson timestep. The channel scattering rates were calculated using the original set of eigenstates, and then on a further five occasions during the simulation, using time-averaged eigenstates in each case.

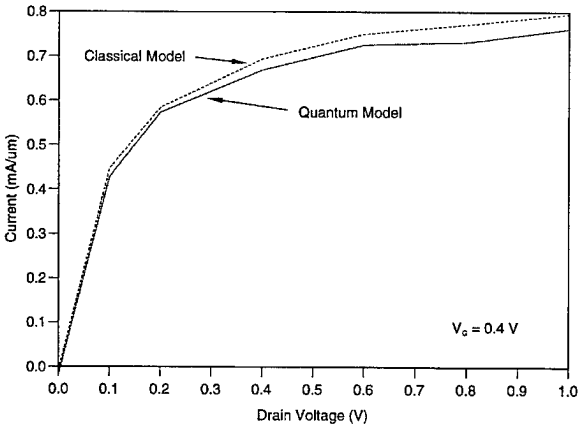


FIGURE 9 Simulated output characteristics for the 0.1 μm gate-length InP MODFET using the classical and quantum models.

In practice, we found surprisingly little difference between the steady-state wavefunctions for the two cases. Figure 7 is typical of all the comparisons we examined, for wavefunctions at different positions and for different subbands. Correspondingly, the scattering rates show little variation with update frequency; again, Figure 8 is typical of the various rates we examined for this device geometry and biasing.

Figure 9 shows that the quantum model predicts a reduced drain current throughout the bias range: this is caused by a reduction in the predicted electron velocities, which is partly due to the greater phonon scattering (as discussed above). However, it must also be appreciated that the quantum model includes remote impurity scattering, a concept which does not feature in the classical description, and which provides some additional momentum relaxation at low energies. Despite the difference in charge distribution in the quantum and classical models, the overall electron density in the channel is very similar.

5. CONCLUSIONS

Our aim in this work has been to examine the importance of including quantum confinement effects in Monte Carlo MODFET simulations. Clearly, we have yet to cover the full parameter space for such devices: however, our results to date imply that the inclusion of confined electron states, and of quantum confined scattering rates, has only a modest effect on predictions of device performance, relative to those obtained from standard semi-classical Monte Carlo models. Whilst the quantum model predicts, in general, a broader charge density distribution across the channel than does the calssical model, the overall charge density in the channel is largely unchanged. Furthermore, it appears that the corrections predicted by the quantum model can be obtained without recourse to a fully self-consistent Schrodinger/Poisson/Monte Carlo scheme: once a steady state solution is obtained from the classical model, the quantum correction can be obtained by a single calculation of the channel eigenstates and corresponding wavefunctions. This represents a substantial saving in CPU time.

References

[1] Hockney, W. and Eastwood, J. W. (1988). *Computer Simulation Using Particles* IOP Publishing Ltd., Bristol.
[2] Jacoboni, C. and Lugli, P. (1989). *The Monte Carlo Method for Semiconductor Device Simulation* Springer-Verlag, Vienna.
[3] Fischetti, M. V. and Laux, S. E. (1988). "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects", *Phys. Rev. B.*, **3**, 9721–9745.
[4] Venturi, F., Sangiorgi, E., Brunetti, R., Quade, W., Jacoboni, C. and Ricco, B. (1991). "Monte Carlo simulations of high-energy electrons and holes in Si-*n*-MOSFETs", *IEEE Trans. Computer Aided Design*, **11**, 1276–1286.
[5] Ravaioli, U. and Ferry, D. K. (1986). "MODFET Ensemble Monte Carlo model including the quasi-two-dimensional electron gas", *IEEE Trans. Electron Devices*, **33**, 677–681.
[6] Park, D. H. and Brennan, K. F. (1990). "Monte Carlo simulations of 0.35 um gate length GaAs and InGaAs HEMTs", *IEEE Trans. Electron Devices*, **37**, 618–628.
[7] Fischetti, M. V. and Laux, S. E. (1993). "Monte Carlo study of electron transport in silicon inversion layers", *Phys. Rev. B.*, **48**, 2244.

Authors' Biography

Robert W. Kelsall was born in Rotherham, England in 1964. He received the B.Sc. and Ph.D. degrees from the University of Durham,

UK, in 1985 and 1989 respectively; the latter for research in Monte Carlo modelling of electronic transport in GaAs quantum wells.

From 1989 to 1993 he worked as a research assistant, developing self-consistent Monte Carlo simulations of HEMTs and multilayer MOSFETs. He was appointed to a lectureship at the University of Leeds, UK in 1993, and is a member of the newly formed Institute of Microwaves and Photonics at Leeds, where he is working on semiconductor device simulations for hot-carrier degradation studies in MOSFETs, quantum confinement effects in HEMTs, and far-infra-red intersubband lasers.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 29–34
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Two-dimensional Modelling of HEMTs Using Multigrids with Quantum Correction

ERIC A. B. COLE [a,*], TOBIAS BOETTCHER [a,b] and CHRISTOPHER M. SNOWDEN [b]

Centre for Nano-Device Modelling, University of Leeds, Leeds LS2 9JT, UK.
[a] *Department of Applied Mathematics, University of Leeds, Leeds LS2 9JT, UK;*
[b] *School of Electronic and Electrical Engineering,*
University of Leeds, Leeds LS2 9JT, UK

The two-dimensional multi-layered HEMT is modelled isothermally by solving the Poisson and current continuity equations consistently with the Schrödinger equation. A multigrid method is used on the Poisson and current continuity equations while the electron density is calculated at each level by solving the Schrödinger equation in one-dimensional slices perpendicular to the layer structure. A correction factor is introduced which enables relatively accurate solutions to be obtained using a low number of eigensolutions. A novel method for discretising the current density which can be generalised to the non-isothermal case is described. Results are illustrated using a two layer AlGaAs-GaAs HEMT.

Keywords: HEMT, Schrödinger, Poisson, current-continuity, energy transport, Bernoulli function, multigrid, GaAs, AlGaAs

1. INTRODUCTION

This paper addresses the problem of solving the modelling equations for High Electron Mobility Transistors (HEMTs) in which the Schrödinger equation is solved explicitly. A basic 4-layer structure consisting of layers of $Al_uGa_{1-u}As$ with different proportions of mole fraction u of aluminium content is shown in Figure 1 with the contacts arranged along the x-axis and with the y-axis perpendicular to the layer boundaries. Such devices have been shown to operate up to 213 GHz [1]. Fast operation is achieved because the electrons near the layer interfaces (the "2-dimensional" electrons) are confined in narrow potential wells and suffer little scattering. The equations governing the modelling of the device are the Schrödinger, Poisson, current continuity, and the energy transport equations. In this paper we will apply some new techniques to the isothermal case, although they may be easily generalised to the non-isothermal case.

Many attempts have been made to obtain robust algorithms and rapid numerical solutions of the

* Corresponding author. Tel.: 0113 233 5117. Fax: 0113 242 9925. E-mail: amt6eac@amsta.leeds.ac.uk.

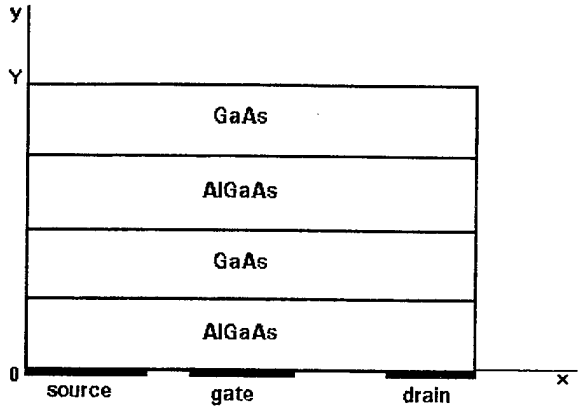


FIGURE 1 A cross-section of a four-layer HEMT showing the positions of the contacts and the layer structure.

equations which model the devices [1–13]. A special function (the *C*-function) has been developed [2] which allows both the current continuity and energy transport equations to be discretised in a consistent manner. A multigrid method [3] has been developed and applied to a one dimensional model, and a more efficient method of including the eigensolutions of the Schrödinger equation has been developed [4].

Figure 2 shows the band offset structure between different layers. This structure will depend on the proportion u of Al in the layer $Al_uGa_{1-u}As$. The equations to be solved are:

(i) The Poisson equation

$$\nabla \cdot (\varepsilon_0 \varepsilon_r \nabla \psi) = -q(N_D^+ - n - T^-) \qquad (1.1)$$

for the solution of the electrostatic potential ψ.

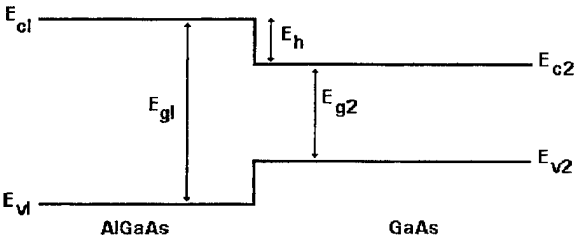


FIGURE 2 The band offset structure between different layers.

Here, q is the magnitude of the electron charge, n is the total electron density, and ε_0 is the permittivity of free space. Writing $E_F \equiv -q\phi$ as the quasi Fermi level, ϕ is fixed at the bias values on the contacts. The values of ψ on the source and drain are given by putting the electron density equal to the doping density using the Fermi function expressions described below. We put $\psi = E_h - \phi_b - \phi$ on the gate where $\phi_b = -0.7$ V is the built-in potential.
(ii) In the steady state we take

$$\nabla \cdot \mathbf{J} = 0, \qquad \mathbf{J} = -q\mu n \nabla \phi \qquad (1.2)$$

for the current density **J**. Generally the mobility μ may depend on the field.
(iii) The Schrödinger equation

$$-\frac{\hbar^2}{2}\nabla \cdot \left(\frac{1}{m}\nabla \xi_i\right) + (V_{xc} + E_h - q\psi)\xi_i = \lambda_i \xi_i \qquad (1.3)$$

where ψ is the electrostatic potential and ξ_i and λ_i ($i = 0, 1, 2, \ldots$) are the energy eigenfunctions (normalised) and eigenvalues respectively. The forms of the exchange correlation energy V_{xc}, effective mass m, relative permittivity ε_r, conduction band discontinuity E_h, N_D^+ and T^- are taken from Adachi [5]. The form of the kinetic energy operator $(-\hbar^2/2)\nabla \cdot (1/m)\nabla$ is taken for its hermitian qualities [14, 15]. The full two-dimensional solution of the Schrödinger equation is very time consuming, and in practice is solved in one dimensional slices perpendicular to the layer structure (that is, in the y-direction). We therefore solve the one dimensional equation

$$-\frac{\hbar^2}{2}\left(\frac{d}{dy}\left(\frac{1}{m}\frac{d\xi_i}{dy}\right)\right) + (V_{xc} + E_h - q\psi)\xi_i = \lambda_i \xi_i \qquad (1.4)$$

by imposing the boundary conditions $\xi_i(0) = \xi_i(Y) = 0$ ($i = 0, 1, 2, \ldots$).

The electron density n inside the quantum wells will be given in terms of the eigensolutions of this equation. The total electron density is given by

$n = n_2 + n_3$ where n_2 (for the "2-dimensional" electrons) is the contribution from the sub-bands given by the solution of the Schrödinger equation and n_3 (for the "3-dimensional" electrons) is the bulk electron density. We first choose a maximum number L of eigensolutions of the Schrödinger equation with which to work. Outside the potential well defined by $\lambda_{L-1} < E_c$ we have $n_2 = 0$ and

$$n_3 = N_{c3} F_{\frac{1}{2}}\left(\frac{1}{kT}(E_F - E_c)\right) \tag{1.5}$$

where $N_{c3} \equiv 2(2\pi mkT/h^2)^{1.5}$.

Inside the well we must avoid the double counting of contributions, and thus we take (for the one dimensional Schrödinger solution only)

$$n_2 = N_{c2} \sum_{i=0}^{L-1} |\xi_i(y)|^2 \ln(1 + e^{(1/kT)(E_F - \lambda_i)}) \tag{1.6}$$

where $N_{c2} \equiv 4\pi mkT/h^2$, and

$$n_3 = \frac{1}{2}\pi\left(\frac{8m}{h^2}\right)^{\frac{3}{2}} \int_{\lambda_{L-1}}^{\infty} \frac{(E - E_c)^{\frac{1}{2}}}{1 + \exp((1/kT)(E - E_F))} dE$$
$$= N_{c3} \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{(x + ((1/kT)(\lambda_{L-1} - E_c))^{\frac{1}{2}}}{1 + \exp(x - (1/kT)(E_F - \lambda_{L-1}))} dx. \tag{1.7}$$

Inside the well the integral in Equation (1.7) involves two parameters $(1/kT)(\lambda_{L-1} - E_c)$ and $(1/kT)(E_F - \lambda_{L-1})$, and its evaluation requires considerable computing time. An approximation to this integral has been developed [4] which speeds up the calculation by allowing a relatively small number of eigensolutions to be used. Outside the well the electron density is given in terms of the Fermi function $F_{(1/2)}$ only, and the approximation of Bednarczyk and Bednarczyk [16] is used for the evaluation.

In the following section we briefly describe this correction process and introduce a new method of writing the current density **J** so that the Bernoulli function method may be applied to the solution of the current-continuity equation. The final section describes the results of a two-dimensional simulation.

2. SOLUTION OF THE EQUATIONS

On writing $a \equiv (1/kT)(E_F - \lambda_{L-1})$ and $b \equiv (1/kT)(\lambda_{L-1} - E_c)$, Equation (1.7) becomes

$$n_3 = N_{c3} \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{(x + b)^{\frac{1}{2}}}{1 + e^{x-a}} dx. \tag{2.1}$$

It is this integral that we wish to approximate. Now define a function $c(a, b)$ such that

$$n_3 = N_{c3}\left(F_{\frac{1}{2}}(a) + c(a, b) \ln(1 + e^a)\right). \tag{2.2}$$

It has been shown [4] that, within the working parameter range of the variables given typically by $-8.0 \le a \le 8.0$, $0 < b \le 8.0$, the function $c(a, b)$ can be approximated by

$$c(a, b) = (0.782 - 0.023a)\sqrt{b}. \tag{2.3}$$

Hence the value n_3 inside the well may be rapidly calculated using Equation (2.2) with the value of c given by Equation (2.3). It was also shown that if this correction factor is used (rather than taking its value to be zero) then the number of eigensolutions needed for the calculation of n_2 may be reduced to give good accuracy. We now describe a novel discretisation of the current-continuity equation. In the isothermal case we may write

$$\mathbf{J} = -q\mu n \nabla \phi \tag{2.4}$$

where $E_F = -q\phi$, and the electron density is a function of ψ, ϕ, x and y: $n \equiv n(\psi, \phi, x, y)$. Firstly consider the x-component of Equation (2.4). Keeping y constant we have

$$\left[\frac{\partial n}{\partial x}\right]_y = \left[\frac{\partial \psi}{\partial x}\right]_y \left[\frac{\partial n}{\partial \psi}\right]_{\phi,x,y} + \left[\frac{\partial \phi}{\partial x}\right]_y \left[\frac{\partial n}{\partial \phi}\right]_{\psi,x,y} + \left[\frac{\partial n}{\partial x}\right]_{y,\psi,\phi}$$

which we write more simply as

$$\frac{\partial n}{\partial x} = \frac{\partial \psi}{\partial x} n_\psi + \frac{\partial \phi}{\partial x} n_\phi + n_x$$

when it is clear which variables are being kept constant. Substitution of the expression for $\partial\phi/\partial x$

into Equation (2.4) gives the first component of **J** as

$$J_{(x)} = q\mu\frac{n_\psi}{n_\phi}\left(\frac{\partial\psi}{\partial x}+\frac{n_x}{n_\psi}\right)n - q\mu\frac{n}{n_\phi}\frac{\partial n}{\partial x}. \quad (2.5)$$

In the interval $(x_i \le x \le x_{i+1}, y = y_j)$ define the quantity

$$f(x, y_j) \equiv \psi + \left(\frac{n}{n_\psi}\right)_{i+\frac{1}{2},j} \ln n$$

Then in this interval we can approximate Equation (2.5) by

$$J_{(x)} = \alpha\frac{\partial f}{\partial x}n + \beta\frac{\partial n}{\partial x}$$

where $\alpha \equiv q\mu n_\psi/n_\phi$ and $\beta \equiv -q\mu n/n_\phi$. We invoke the Bernoulli function approach [2] to this discretisation by assuming that $\beta, \alpha/\beta$ and $\partial f/\partial x$ are constants in the interval. Then

$$J_{(x)i+\frac{1}{2},j} = \frac{\beta_{i+\frac{1}{2},j}}{h_i}\left[B\left(-\left(\frac{\alpha}{\beta}\right)_{i+\frac{1}{2},j}(f_{i+1,j}-f_{i,j})\right)n_{i+1,j} - B\left(\left(\frac{\alpha}{\beta}\right)_{i+\frac{1}{2},j}(f_{i+1,j}-f_{i,j})\right)n_{i,j}\right] \quad (2.6)$$

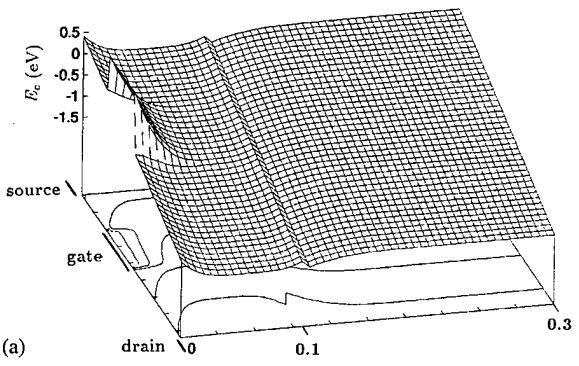
where $B(x) \equiv x/(e^x-1)$ and $h_i \equiv x_{i+1}-x_i$. The y-component may be similarly discretised by introducing a function $g(x_i, y) \equiv \psi + (n/n_\psi)_{i,j+\frac{1}{2}} \ln n$ which is defined in the interval $(x = x_i,\ y_j \le y \le y_{j+1})$. Assumptions similar to the above have been derived elsewhere [2] and as a brief indication of their validity, note that in the non-degenerate case we have $\alpha/\beta = -q/kT$ when n is given in purely exponential terms.

The above formulation is general in that the exact functional form of the electron density is not specified. Once the specification is given then the derivatives n_ψ, n_ϕ, n_x and n_y may be calculated. In the non-isothermal case, Equation (2.4) will not apply and the generalisation of Equation (2.5) will contain extra terms $\partial n/\partial T$. The solution will then be given using the C-function approach [2] which generalises the Bernoulli function approach.

In order to reduce the computational effort in reaching a solution, a multigrid approach has been developed [3] in which the Poisson and current-continuity equations are solved using multigrids while the Schrödinger equation is solved by a non-multigrid scheme at each level to provide the expression for the electron density n. This has increased the speed of the solution by almost five times in some cases.

3. SIMULATION RESULTS

The model problem is that of a two-layer $Al_{0.3}Ga_{0.7}As$-GaAs HEMT. The ends of the source, gate and drain were at $x = 0.0\,\mu m$, $0.1\,\mu m$, $0.4\,\mu m$, $0.7\,\mu m$, $1.3\,\mu m$ and $1.4\,\mu m$, the total depth of the device was $Y = 0.3\,\mu m$ with the layer interface at $y = 0.1\,\mu m$. The doping of the AlGaAs and GaAs was taken as $5\times10^{23}\ m^{-3}$ and $1\times10^{19}\ m^{-3}$ respectively. High doping regions of $1.2\times10^{24}\ m^{-3}$ were taken around the source an drain to simulate ohmic contacts, and a depletion value of $8\times10^{16}\ m^{-3}$ was taken on the contact edges between source and gate and gate and drain to pin E_c to a value at which the quantum well could form. The electron temperature was taken as a constant 300 K. The mobilities of the



(a)

FIGURE 3 Results of the simulation with $V_{gs} = -0.1$ V and $V_{ds} = 2.0$ V, with all distances in μm, showing (a) the conduction band E_c, (b) the electron density n, and c) the second energy eigenfunction ξ_1.

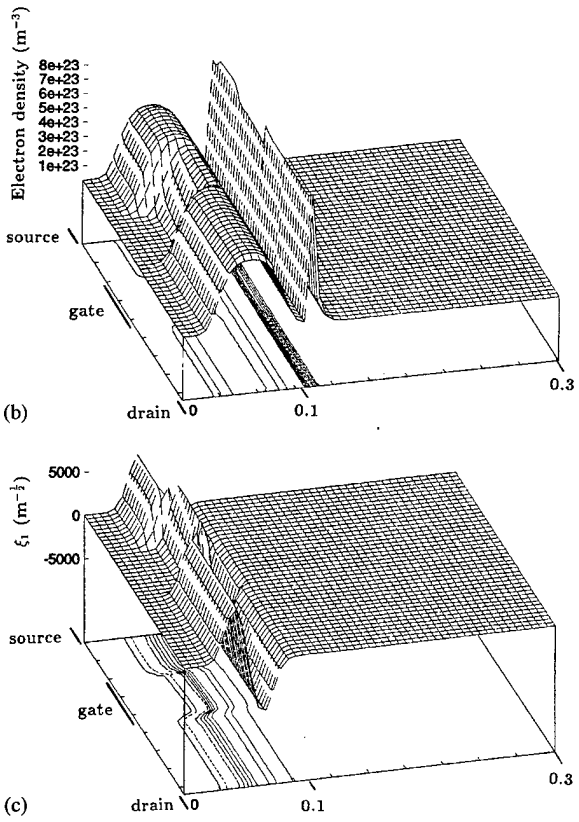


FIGURE 3 (Continued).

AlGaAs and GaAs were taken as $0.2\ m^2V^{-1}s^{-1}$ and $0.8\ m^2V^{-1}s^{-1}$ respectively.

Results shown in Figure 3 were derived for the case $V_{gs} = -0.1$ V, $V_{ds} = 2.0$ V. A multigrid method was used to solve the Poisson and current continuity equations. The Schrödinger equation was solved on each grid by a non-multigrid method to provide the electron density. The eigenvalues were found using a QL algorithm with implicit shifts, ordered, and the first L eigenfunctions found by back substitution. The multigrid method described in [3] was used. Four uniform grids were taken with the coarsest grid being 40×25 – this was found to be the minimum size for rapid calculation of the exact solution on the coarsest grid without losing the quantum well structure in the y-direction. The coarse grid solution was found using Newton iteration with automatic stopping based on the sizes of the residuals of the Poisson and current continuity equations. It was found that an initial number of 20 iterations had to be performed on this solution without quantum correction to allow the quantum well structure to become established. Ten pre- and post-smoothing iterations were made on the finer grids. As found earlier [4] it was found that using only three eigenvalues ($L=2$) with the correction factor of Equation (2.3) provided comparable results with those of the benchmark case $L=9$ with no correction used, thus speeding the numerical solution. Further, solutions could be found for fixed values of V_{gs} and V_{ds} without having to approach them in small increments.

References

[1] Rosenbaum, S. E., Kormanyos, B. K., Jelloian, L. M., Matloubian, M., Brown, A. S., Larson, L. E., Nguyen, L. D., Thompson, M. A., Katehi, L. P. B. and Rebeiz, G. M. (1995). *IEEE Trans. Microwave Theory and Techniques*, **43**(4), 927–932.

[2] Cole, E. A. B. and Snowden, C. M. (1995). *Int. J. Num. Modelling: Electronic Networks, Devices and Fields*, **8**, 13.

[3] Cole, E. A. B., Snowden, C. M. and Boettcher, T. (1997). *Int. J. Num. Modelling: Electronic Networks, Devices and Fields*, **10**, 121.

[4] Cole, E. A. B., Boettcher, T. M. and Snowden, C. M. (1997). *Semicond. Sci. Technol.*, **12**, 100.

[5] Adachi, S. (1985). *J. Appl. Phys.*, **58**, R1.

[6] Cole, E. A. B. (1993). In Compound Semiconductor Device Modelling, C. M. Snowden and R. E. Miles, Eds., Springer Verlag, New York, p. 1.

[7] Tang, T.-W. (1984). *IEEE Trans. Electronic Devices*, **ED-31**, 1912.

[8] Snowden, C. M. and Loret, D. (1987). *IEEE Trans. Electronic Devices*, **ED-34**, 212.

[9] Feng, Y.-K. and Hintz, A. (1988). *IEEE Trans. Electronic Devices*, **ED-35**, 1419.

[10] McAndrew, C. C., Heasell, E. L. and Singhal, K. (1988). *Semicond. Sci. Technol.*, **3**, 758.

[11] Zhou, J.-R. and Ferry, D. K. (1995). *VLSI Design*, **3**, 159.

[12] Ando, Y. and Itoh, T. (1988). *IEEE Trans. Electron Devices*, **35**, 2295.

[13] Ng, S.-H. and Khoie, R. (1991). *IEEE Trans. Electron Devices*, **38**, 852.

[14] Stern, F. (1984). *Phys. Rev. B.*, **30**, 840.

[15] von Roos, O. (1983). *Phys. Rev. B.*, **27**, 7547.

[16] Bednarczyk, D. and Bednarczyk, J. (1978). *Phys. Lett.*, **64A**, 409.

Authors' Biographies

Eric A. B. Cole received the B.Sc. and Ph.D. degrees from the University of Wales at Cardiff. He is a Senior Lecturer in the Department of Applied Mathematical Studies at the University of Leeds and is a past Chairman of the School of Mathematics at Leeds. He is currently the Director of the Centre for Nano-Device Modelling at Leeds. His main research interests include semiconductor device modelling, and multitemporal relativity.

Tobias M. Boettcher received his B.Sc., in Mathematical Physics from the University of Sussex, and his M.Sc., in Nonlinear Mathematical Models from the Universities of Edinburgh and Heriot-Watt. He has recently successfully completed his Ph.D. in the Department of Applied Mathematical Studies and the School of Electronic and Electrical Engineering at the University of Leeds.

Christopher M. Snowden received the B.Sc., M.Sc. and Ph.D. degrees from the University of Leeds. He holds the Chair of Microwave Engineering in the Institute of Microwaves and Photonics and is currently Head of the School of Electronic and Electrical Engineering. He is a Fellow of the IEEE and a Fellow of the IEE, and is a Distinguished Lecturer for the IEEE Electron Devices Society. He is co-Chairman of the IEEE MTT-1 Committee and a Member of the 1997 IEEE MTT-S Technical Program Committee. Main research interests include compound semiconductor device modelling, microwave, terahertz and optical nonlinear subsystems design and advanced semiconductor devices.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 35–40
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

MOMENTS: The Modular Monte Carlo Environment for Charge Transport Simulation, Overview and Applications

MARK PESKIN and CHRISTINE MAZIAR

Microelectronics Research Center, The University of Texas at Austin, Austin, TX 78712, USA

We present MOMENTS, a newly developed software library for Monte Carlo simulation of semiconductor devices. This library uses object-oriented design principles to provide a flexible, extensible toolset that allows rapid development of a wide variety of Monte Carlo simulation applications. It allows concurrent simulation of multiple particle species (e.g. electrons and holes) with arbitrary user-defined interactions between species (e.g. generation – recombination and carrier – carrier scattering) in arbitrary geometries using either analytic or numerical bandstructure representations. The modular design allows virtually all simulation parameters to be freely varied across the simulation domain. MOMENTS also takes advantage of the parallelism inherent in the ensemble Monte Carlo approach, employing a scheme that can support a wide variety of parallel architectures with active load balancing. To demonstrate some of the library's capabilities, we also present preliminary results from a GaAs avalanche photodiode (APD) simulator based on MOMENTS.

Keywords: Simulation, Monte Carlo, object-oriented, semiconductors, transport, parallel, avalanche photodiode

INTRODUCTION

Conventionally, applications of the Monte Carlo method to semiconductor device simulation have come in the form of highly specialized codes designed specifically to address a relatively constrained and well-defined set of research questions. The scope of these specialized tools has been further limited by the tremendous computational demands of Monte Carlo simulation. However, as both conventional silicon and optoelectronic device designs continue to scale dimensionally, the Monte Carlo method, with its unmatched ability to accurately characterize the nonlocal transport phenomena that dominate the behavior of devices with very small geometries, will become an increasingly attractive simulation option. In order to satisfy the demands of the TCAD community, future Monte Carlo simulators should possess three key properties: *flexibility* – allowing the code to be easily adapted to explore a variety of devices, geometries, and materials; *extensibility* – permitting the user to add new models and capabilities to the simulation package as needed without requiring modification of existing code; and *performance* – so that the simulator can be used cost effectively on a day-to-day basis.

We have developed a new Monte Carlo simulation environment, MOMENTS, that attempts to address these three key requirements by exploiting

advanced object-oriented and parallel software design in C++. Rather than attempting the (probably) impossible task of creating a single application capable of addressing every anticipated research need, we have instead developed MOMENTS as a library of objects that can be quickly and easily assembled to create a variety of simulation applications. Thus MOMENTS can be thought of as a toolbox or application programming interface (API) for Monte Carlo simulation. The MOMENTS API is designed to be easy to use but has many sophisticated features and is general enough to subsume most existing Monte Carlo models, allowing those models to be leveraged for use in new problems.

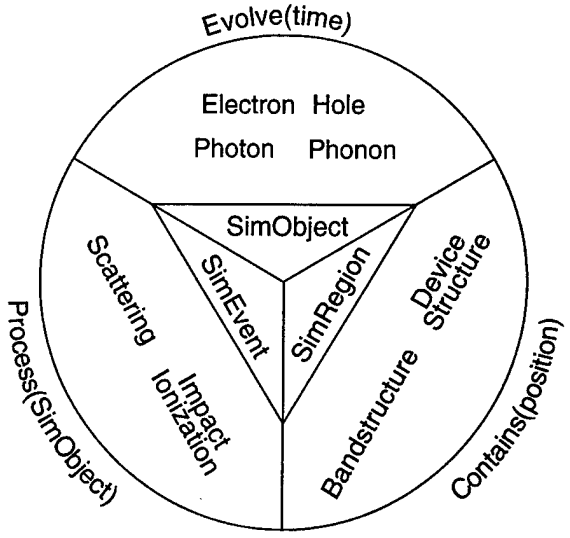


FIGURE 1 Schematic showing the three root functional classes of MOMENTS, examples of simulation elements they are designed to represent, and their primary methods.

DESIGN PRINCIPLES

MOMENTS comprises a collection of C++ classes designed to represent the functional components of a Monte Carlo simulation. These classes are organized hierarchically, making full use of the inheritance properties of object oriented software. The root level classes have intentionally simple interface specifications. This promotes maximum flexibility, since any child class which supports these very general interfaces can be easily integrated into an existing simulation. Simulations are defined by the interactions between objects based on three primary root level classes: SimObject, SimRegion, and SimEvent (see Fig. 1). The SimObject class is the heart of MOMENTS. Each SimObject represents a self-contained entity that, based on its surroundings, knows how to evolve its own state over a fixed time interval. Classes to represent electrons and holes are descendants of SimObject. In addition, the user may define other particle species based on the SimObject class as long as their time-evolving behavior can be defined in the context of the simulation. Possible examples include photons, phonons, excitons, or mobile ions. This makes MOMENTS useful in a wide variety of simulation applications, including process simulation. Since SimObjects are self-contained, several different particle species can easily be simulated concurrently and new instances of simulated particles can be created "on-the-fly". This allows direct study of multi-species interactions, including direct simulation of generation-recombination processes.

The SimRegion class is designed to encapsulate spatial structure data in the most general way possible; the only requirement for SimRegion objects is that they must determine whether or not they contain a given spatial coordinate. The SimRegion class also supports a hierarchy in which each SimRegion object can contain one or more sub-regions. This allows structures of arbitrary complexity and dimensionality to be represented. It also allows the programmer to create anything from bulk up through full 3-D simulations without altering the underlying models in the code. SimRegion objects, which comprise both real-space and k-space structures (bandstructures), also serve as containers for data (e.g. material data, electric fields, etc.) using the SimRegistry mechanism described below. This data storage is also hierarchical. If a SimObject requests data from a sub-region for which that data has not been

specified, the request will automatically be passed to that sub-region's parent region.

The SimEvent class represents stochastically occurring events that alter the state of a SimObject (e.g. scattering mechanisms). SimEvents related to a given SimObject class are collected in SimEventGroups, which can automatically handle event selection using an efficient adaptive self-scattering technique. Different SimEventGroups can be specified in different regions of the simulation domain. SimEvents can also be loaded into and unloaded from SimEventGroups at will, even while the simulation is running. This allows transport simulators based on MOMENTS to adaptively select the most appropriate scattering models for a given portion of the simulation. It is fairly straightforward to create new scattering models or modify the behavior of existing ones using C++ inheritance.

MOMENTS derives its flexibility from this modular design. Classes belonging to each of the three functional groups described above can be mixed and matched as needed, even during the course of a running simulation. MOMENTS is also fully extensible; it is relatively easy for users to create new SimObject, SimRegion, and SimEvent subclasses, using the inheritance properties of C++, and integrate them into existing codes. The modular MOMENTS structure encourages collaboration and simplifies code maintenance. It also provides potential performance benefits. Since virtually every aspect of the simulation, including scattering models, self-scattering rates, bandstructure representations, and even interpolation schemes, can be varied freely across different regions of the simulation domain, the user can focus the greatest degree of computational effort in regions of the domain where the greatest accuracy is desired.

PARALLELISM IN MOMENTS

MOMENTS addresses performance concerns primarily by exploiting the opportunities for parallelism inherent in the Monte Carlo method. In the MOMENTS system, the collections of SimObjects necessary for ensemble Monte Carlo simulations are collected in one or more SimPopulation objects. The members of a SimPopulation can then be distributed to one or more "contexts" – independent subgroups of SimObjects that evolve in parallel under independent threads of control. The members of each context are synchronized and allowed to interact at predetermined intervals in a manner analogous to Bulk Synchronous Parallelism [1] (see Fig. 2). While contexts are busy handling the evolution of their respective SimObjects, the main thread is free to perform subsidiary tasks such as re-calculation of electric fields using a Poisson solver or I/O. The MOMENTS context scheme is applicable to a wide variety of parallel architectures, including heterogeneous distributed systems, and supports active load balancing. Such parallelization of the Monte Carlo method raises several interesting issues. For example, the indeterminate timing of calls to a given random number generating sequence by different contexts would tend to prevent reproduction of a given sequence of events – a useful debugging technique. MOMENTS addresses this

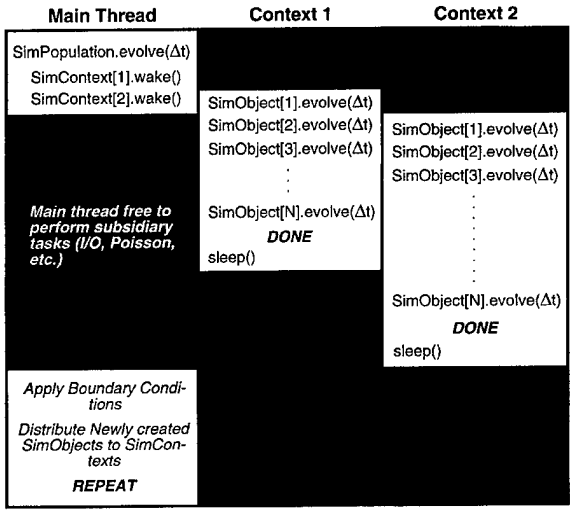


FIGURE 2 Illustration of a simulation cycle for a SimPopulation object with two attached contexts.

problem by providing each context with its own random number generating sequence. Calls to the SimRand() function are automatically directed to the generating sequence for the calling thread, shielding the user from this complexity. Another problem common to all parallel codes is data contention by multiple threads of control. Many key data structures in MOMENTS have been carefully designed to avoid problems caused by data-contention among multiple contexts.

SUPPORT CLASSES

In addition to the three main class hierarchies described above, MOMENTS includes a number of support classes. Grids can be defined using the SimGrid, SimGridPoint, and SimGridCell classes. These grid classes are very general and can support all types of structured and unstructured grids in one, two, or three dimensions. Boundary conditions are handled by the SimSource and SimSink classes, which can emulate all types of conventional boundaries. In addition, these classes can be used to simulate absorption (recombination) or emission (photoemission, thermal generation, etc.) throughout entire regions of the simulation domain. The SimRecorder class can be used to automatically record the characteristics of one or more SimObjects at specified intervals. MOMENTS is also equipped with a variety of objects to handle thread control in parallel implementations, including a number of thread-safe data structures.

One of the key support classes is the SimRegistry – a construct that allows members of a "client" class to transparently create storage in every member of a SimRegistry "host" class. The client can then address the "registered" data item for a given object of the host class using an array-style notation. Storage for this registered data is automatically created in any new objects based on the host class. The SimRegistry concept is important because it allows users to extend the functionality of existing host classes without actually having to modify the source code of the host class. Also, any number of client classes can use the same host class for storage. SimRegion, SimGridPoint, SimBandStruct, and SimThread are all examples of SimRegistry host classes.

GAAS APD SIMULATION

As an example of the capabilities of MOMENTS, we present preliminary results from a GaAs avalanche photodiode (APD) simulator based on the MOMENTS library. This simulator will be used to study the anomalous multiplication noise-suppression recently detected in III – V compound separate absorption and multiplication (SAM) APDs with very thin ($\leq$ 200 nm) multiplication regions [2, 3]. Conventional APD theory relates the multiplication noise to the ratio of hole to electron ionizations (k) in the multiplication region of the device. Smaller values of k are associated with less noise. The lower observed noise values in devices with very thin multiplication regions may be associated with phenomena related to the nonlocality of the impact ionization process. One such phenomenon is known as the "dead space effect", referring to the fact that charge carriers with low initial energies must be accelerated through a high field for a sufficient distance to acquire enough energy to ionize.

Obtaining a detailed understanding of the dead space effect and its relationship to APD device geometry and noise figures requires simultaneous simulation of both electrons and holes, including the ability to simulate new carriers as they are generated by impact ionization events. Moreover, accurate characterization of transport and impact ionization in high-field regions entails the use of realistic numerically calculated bandstructure data for both carrier types. The capabilities of MOMENTS make it ideally suited for the study of this complex problem. Concurrent simulation of multiple carrier species is trivial in MOMENTS, thanks

to the SimObject abstraction. The library's ability to generate new carriers on the fly allows a rigorous simulation of the impact ionization process. Finally, the modular MOMENTS design simplifies incorporation of numerical bandstructure data based on nonlocal pseudopotential calculations. The MOMENTS package was used to create a simple 1-D GaAs APD simulator. The simulator employs realistic nonlocal pseudopotential bandstructure representations for both carrier types. In addition to impact ionization, intra and inter-valley scatterings by both optical and acoustic phonons are included. The simulator, assembled from MOMENTS components, requires less than a page of computer code.

Preliminary results from this new simulator are encouraging. Figure 3 is a graph of gain and k for several multiplication region thicknesses with a fixed field of 700 kV/cm in the multiplication region. The k values demonstrate the same rapidly decreasing trend for thinner multiplication regions seen in experimental devices. Figure 4 shows the relative ionization rate for each carrier type as a function of position along a 100 nm multiplication layer. In this graph, electrons are injected at the right boundary ($x=0$) and accelerate to the left, while generated holes are accelerated back to the right. There is a clear "dead space" for each carrier type, indicated by the shaded regions. We believe this is the first quantification of dead space in APDs. Note that the dead space region for holes is approximately 3 – 4 times longer than that for electrons, indicating that as the multiplication layer shrinks, hole ionizations will be preferentially suppressed. Thus it appears that dead space may indeed explain the anomalously low noise levels observed in thin multiplication layer APDs, although further study will be necessary for confirmation.

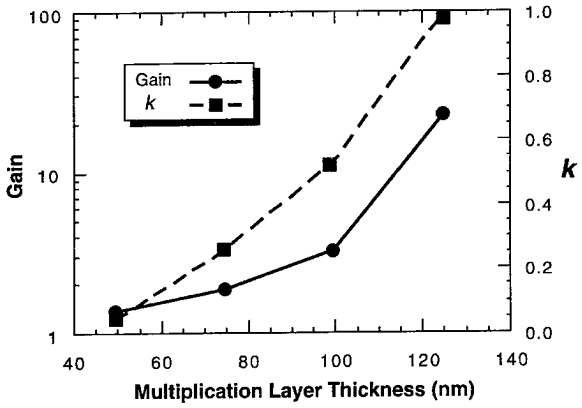


FIGURE 3 Simulated gain and k-ratio for several GaAs APD multiplication layer thicknesses with an applied field of 700 kV/cm.

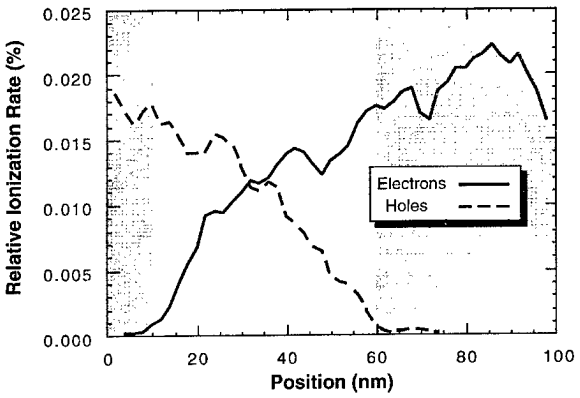


FIGURE 4 Simulated relative ionization rates as a function of position across a 100 nm multiplication layer with an applied field of 700 kV/cm.

CONCLUSION

In conclusion, we have developed MOMENTS, a modular library of C++ objects that serves as a framework for development of high-performance Monte Carlo simulation applications. The library supports parallel execution in an architecture-independent manner, and provides tremendous flexibility, extensibility, maintainability, and ease-of-use – properties lacking in most conventional Monte Carlo simulators. To demonstrate the advanced capabilities of MOMENTS, we have presented preliminary results from a GaAs APD simulator based on the MOMENTS library. To the best of our knowledge this Monte Carlo tool is the first to concurrently simulate the transport of

both electrons and holes using realistic numerical bandstructure representations for both carrier species with direct simulation of the carrier generation process. We believe that MOMENTS will help make the Monte Carlo method much more accessible to the TCAD research community as a whole. This should be a tool that will be especially attractive to researchers working on novel device structures or in novel material systems.

Acknowledgements

This work was supported, in part, the Joint Services Electronics Program (F49620-95-C-0045), the Texas Advanced Technology Program (TATP) and Motorola. The United States Government is authorized to distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

References

[1] Valiant, L. G., "A Bridgioeng Model for Parallel Computation", *Communications of ACM*, August 1990.
[2] Anselm, K. A. *et al.* (1996). "A Resonant Cavity, Separate-Absorption-and-Multiplication, Avalanche Photodiode with Low Excess Noise Factor", *IEEE Electron Device Lett.*, **17**(3).
[3] Hu, C. *et al.* (1996). "Noise Characteristics of Thin Multiplication Region GaAs Avalanche Photodiodes", *Appl. Phys. Lett.*, **69**(24).

Authors' Biographies

Mark Peskin received a B.A. in Physics (Magna cum Laude) from Wesleyan University in 1992 and an M.S.E. in Electrical Engineering in 1995 from the University of Texas at Austin, where he is currently completing a Ph.D. His research interests include semiconductor device physics, Monte Carlo and Hydrodynamic simulation, Monte Carlo parameter extraction and model development, optoelectronic simulation, nonlinear optimization techniques, parallel simulation systems, and advanced simulation software development.

Christine Maziar received the B.S.E.E. (with Highest Distinction), the M.S.E.E. and the Ph.D., degrees from Purdue University in 1981, 1984, and 1986. In January of 1987, she joined the faculty of the University of Texas at Austin, where she is a Professor of Electrical and Computer Engineering and Vice Provost. She holds the Archie W. Straition Endowed Teaching Fellowship in Electrical and Computer Engineering. Her research interests include modeling of charge transport in high performance device structures, semiconductor device physics, and device simulator enhancement. She has authored or coauthored over 100 journal or conference publications. She is a member of IEEE, ASEE, APS, Eta Kappa Nu, Tau Beta Pi and Sigma Pi Sigma.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 41–45
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

High-Field Hole Transport in Strained Si and SiGe by Monte Carlo Simulation: Full Band Versus Analytic Band Models

F. M. BUFLER*, P. GRAF** and B. MEINERZHAGEN

Institut für Theoretische Elektrotechnik und Mikroelektronik, FB 1, Postfach 33 04 40, Universität Bremen, D-28334 Bremen, Germany

Monte Carlo results are presented for the velocity-field characteristics of holes in (i) unstrained Si, (ii) strained Si and (iii) strained SiGe using a full band model as well as an analytic nonparabolic and anisotropic band structure description. The full band Monte Carlo simulations show a strong enhancement of the drift velocity in strained Si up to intermediate fields, but yield the same saturation velocity as in unstrained Si. The drift velocity in strained SiGe is also significantly enhanced for low fields while being substantially reduced in the high-field regime. The results of the analytic band models agree well with the full band results up to medium field strengths and only the saturation velocity is significantly underestimated.

Keywords: Monte Carlo simulation, analytic and full band structures, strained Si and SiGe

1. INTRODUCTION

The progress in epitaxial growth techniques of unstrained and strained SiGe layers have led to intensified efforts to explore the potential performance enhancements in SiGe based devices. In particular, the practical usefulness of *p*-MOSFETs with a channel consisting of strained Si [1] or strained SiGe [2] has been recently demonstrated. Since field effect devices operate in the low-field and in the high-field regime, reliable modeling of hole transport is important for both cases. However, previous publications on hole transport in strained Si and SiGe covered only the low field regime [3] or were restricted to strained SiGe and electric field strengths below 20 kV/cm [4]. Hence, there is a clear need for investigations of high-field effects like velocity saturation where the consideration of the full band structure is often necessary for accurate results. On the other hand, for devices with realistic Germanium profiles full band Monte Carlo simulations still involve an unmanageable computational burden (e.g. prohibitive memory requirements) and analytic band structure

*Corresponding author. Present address: Institut für Integrierte Systeme, Gloridstrasse 35, CH-8092 Zürich, Switzerland.
**Present address: Intermetall, Hans-Bunte-Str. 19, D-79108 Freiburg, Germany.

approximations have to be used instead. The aim of this paper is therefore twofold: On one hand, we perform for the first time full band Monte Carlo simulations for strained Si and SiGe in the high-field regime. On the other hand, we present a simple analytic hole band model and evaluate its range of validity.

2. MODEL DESCRIPTION

The full band model for strained Si or SiGe is obtained by nonlocal empirical pseudopotential calculations including spin-orbit interaction [9]. For the analytic band structure we neglect the warping of the three valence bands $v = 1, 2, 3$ and use a simple parametrization according to

$$E(1+\alpha_v E) = \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_{\parallel,v}} + \frac{k_y^2}{m_{\parallel,v}} + \frac{k_z^2}{m_{\perp,v}}\right) \qquad (1)$$

with $E = \varepsilon - \varepsilon_{0,v}$ because of the feasibility of this formula for applications. The scattering mechanisms included are optical phonons and acoustic phonons in the isotropic and elastic equipartition approximation. In SiGe both Si-type and Ge-type phonons are considered and alloy scattering is taken into account with the alloy scattering potential adjusted to drift mobility measurements in unstrained SiGe [10]. Exactly the same coupling constants are used with the full band and the analytic band model. The parameters α_v, $m_{\parallel,v}$ and $m_{\perp,v}$ are adapted to the full band structure for the purpose of transport applications. The starting point is therefore the expression for the Ohmic drift mobility which involves for the scattering processes used only the Density Of States (DOS) and the square of the group velocity averaged over an energy surface $\mathcal{V}^2$ [11]. Then a parabolic expression is used for each band to determine the masses m_{DOS} and m_{cond} by adjusting the DOS and $\mathcal{V}^2$, respectively, up to about 40 meV above the band edge to the respective full band results. For a good transport description the mobility of the analytic band model in Eq. (1) must equal the mobility obtained with the parabolic fits to the full band model. This condition yields for $\alpha = 0$ in the unstrained case ($m_{\parallel} = m_{\perp} \equiv m$)

$$m = m_{\mathrm{cond}}^{2/5} m_{\mathrm{DOS}}^{3/5} \qquad (2)$$

Finally, using this mass m the nonparabolicity factor α is obtained from fitting the DOS up to 1 eV. A similar procedure applies in the strained case.

3. VERIFICATION

In Figures 1, 2, 3 and 4 drift mobilities and drift velocities resulting from the full band and the analytic band model are compared with experimental data in the case of unstrained Si and Ge because an accurate reproduction of the experiments is essential in view of the increased importance of details of phonon and band models in the strained case. Overall good agreement is achieved. Especially the full band model in Figure 2 reproduces accurately the anisotropy of the velocity-field characteristics as well as the saturation drift velocity of Ref. [7]. Within the isotropic band approximation (unstrained case) also the

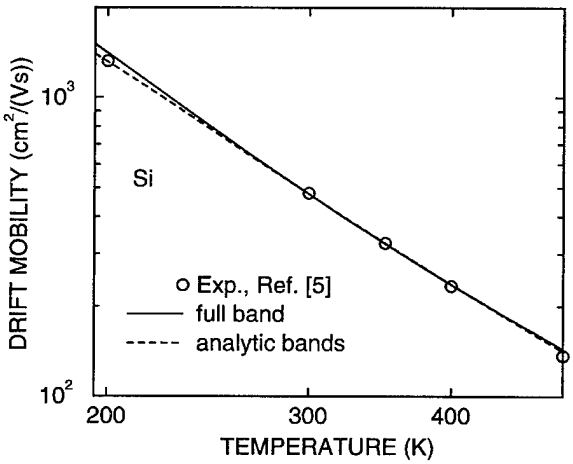


FIGURE 1 Temperature dependence of Ohmic drift mobility for holes in unstrained Si: comparison of full band model, analytic band model and experimental results.

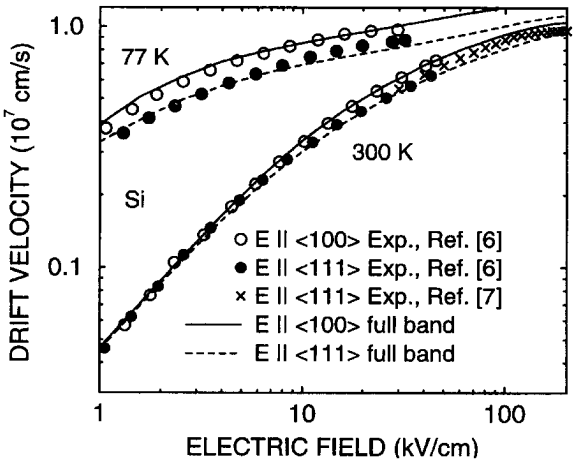


FIGURE 2 Velocity-field characteristics of the full band model for unstrained Si at 77 and 300 K in comparison with experimental results.

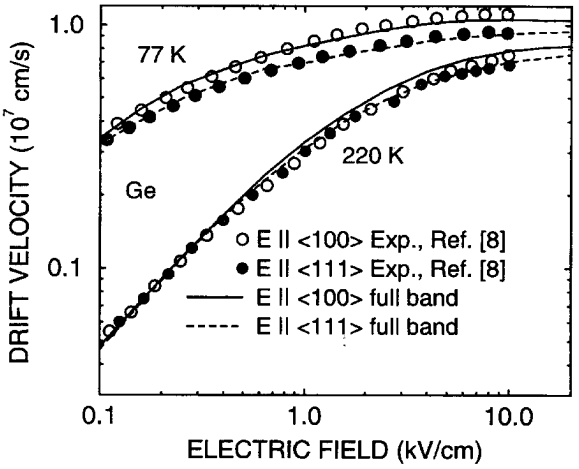


FIGURE 3 Velocity-field characteristics of the full band model for unstrained Ge at 77 and 220 K in comparison with experimental results.

analytic band model yields surprisingly good results and only significantly underestimates the drift velocity above 50 kV/cm.

4. RESULTS

In Figures 5 and 6 the high-field results for strained Si and SiGe are shown. While the value of the saturation drift velocity in unstrained Si is

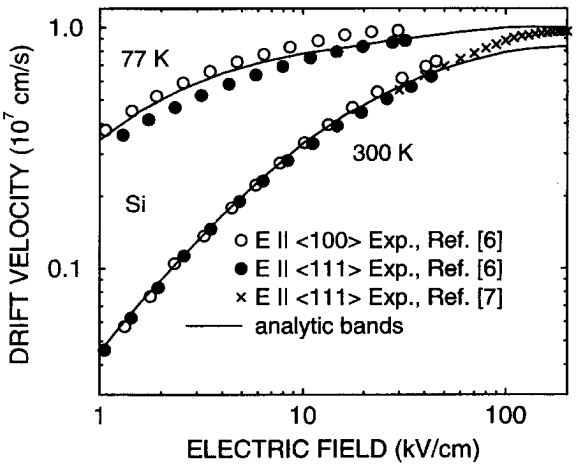


FIGURE 4 Velocity-field characteristics of the analytic band model for unstrained Si at 77 and 300 K in comparison with experimental results.

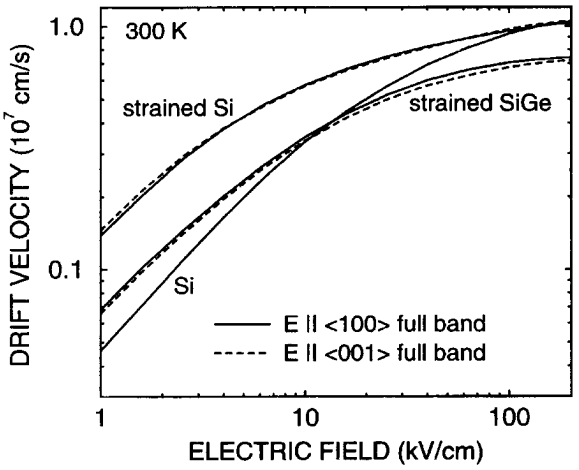


FIGURE 5 Results of the full band model for the velocity-field characteristics at 300 K in strained Si grown on a $Si_{0.7}Ge_{0.3}$ substrate, in unstrained Si and in strained $Si_{0.6}Ge_{0.4}$ grown on a Si substrate.

retained in strained Si, the drift velocity at lower fields is considerably improved due to the enhanced population of the light hole band. In contrast, the saturation velocity is reduced in strained SiGe, but there is still a substantial improvement up to intermediate fields. But please keep in mind that no realistic estimate of the corresponding device performance can be based on

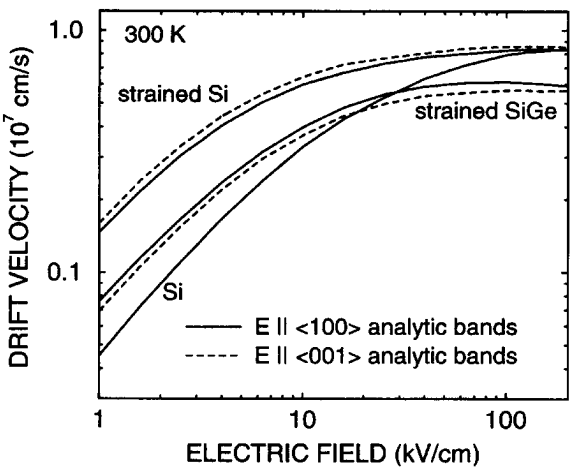


FIGURE 6 Results of the analytic band model for the velocity-field characteristics at 300 K in strained Si grown on a $Si_{0.7}Ge_{0.3}$ substrate, in unstrained Si and in strained $Si_{0.6}Ge_{0.4}$ grown on a Si substrate.

Figure 5 alone because advantages like the possibility of modulation doping have to be considered for this purpose as well. The analytic band model again underestimates the drift velocity above 50 kV/cm and somewhat overestimates anisotropy.

Acknowledgements

This work was supported in part by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie under contract no. 01 M 2416 A. The authors would like to thank P. Vogl (TU München) for discussions, M. M. Rieger (University of Cambridge, U.K.) for calculating the full band structures as well as K. Tittelbach-Helmrich and P. Gaworzewski (Frankfurt/Oder) for providing their experimental mobility data prior to publication.

References

[1] Nayak, D. K., Goto, K., Yutani, A., Murota, J. and Shiraki, Y. (1996). "High-mobility strained-Si PMOSFET's", *IEEE Trans. Electron Devices*, **43**, 1709–1716.

[2] Risch, L., Fischer, H., Hofmann, F., Schäfer, F., Eller, M. and Aeugle, T. "Fabrication and electrical characterization of Si/SiGe *p*-channel MOSFETs with a delta doped boron layer", in *Proc. ESSDERC*, Bologna, 1996, **26**, 465–468.

[3] Fischetti, M. V. and Laux, S. E. (1996). "Band structure, deformation potentials, and carrier mobility in strained Si, Ge and SiGe alloys", *J. Appl. Phys.*, **80**, 2234–2252.

[4] Hinckley, J. M. and Singh, J. (1990). "Hole transport theory in pseudomorphic $Si_{1-x}Ge_x$ alloys grown on Si(001) substrates", *Phys. Rev. B.*, **41**, 2912–2926.

[5] Green, M. A. (1990). "Intrinsic concentration, effective densities of states, and effective mass in silicon", *J. Appl. Phys.*, **67**, 2944–2954.

[6] Canali, C., Ottaviani, G. and Alberigi-Quaranta, A. (1971). "Drift velocity of electrons and holes and associated anisotropic effects in silicon", *J. Phys. Chem. Solids*, **32**, 1707–1720.

[7] Smith, P. M. and Frey, J. (1981). "High-field transport of holes in silicon", *Appl. Phys. Lett.*, **39**, 332–333.

[8] Reggiani, L., Canali, C., Nava, F. and Ottaviani, G. (1977). "Hole drift velocity in germanium", *Phys. Rev. B.*, **16**, 2781–2791.

[9] Rieger, M. M. and Vogl, P. (1993). "Electronic-band parameters in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates", *Phys. Rev. B.*, **48**, 14276–14287.

[10] Tittelbach-Helmrich, K. and Gaworzewski, P. (1997). Unpublished.

[11] Fu, Y., Grahn, K. J. and Willander, M. (1994). "Valence band structure of Ge_xSi_{1-x} for hole transport calculation", *IEEE Trans. Electron Devices*, **41**, 26–31.

Authors' Biographies

Fabian M. Bufler studied physics at the TU Braunschweig and RWTH Aachen (both Germany) including an academic year at the Université de Grenoble I (France) with a scholarship of the Studienstiftung des deutschen Volkes and received the Dipl.-Phys. degree in 1992. Since then he was working as a research assistant at the Institut für Theoretische Elektrotechnik, RWTH Aachen, and moved in 1995 together with the group of Prof. Bernd Meinerzhagen to the Institut für Theoretische Elektrotechnik und Mikroelektronik, Universität Bremen (Germany). His current research interests are concerned with transport theory and full band Monte Carlo simulation for SiGe based devices.

Peter Graf received the Diploma in physics from the RWTH Aachen (Technical University of Aachen), Germany, in 1991. At the Institut für Theoretische Elektrotechnik, RWTH Aachen, he then worked on the Monte-Carlo simulation of homogeneous systems and its benefits for hydrodynamic charge transport parameters. After join-

ing the Institut für Theoretische Elektrotechnik und Mikroelektronik, Universität Bremen, he dealt with algorithms for Monte-Carlo device simulation and their implementation into the institute's Monte-Carlo heterojunction device simulator.

Bernd Meinerzhagen received the Dipl.-Ing. degree in electrical engineering in 1977, the Dipl.-Math. degree in mathematics in 1981, the Dr.-Ing. degree in electrical engineering in 1985 and the "venia legendi" in 1995 all from the RWTH Aachen (Germany). During 1986 he was a Member of Technical Staff at AT&T Bell Laboratories in Allentown (USA) and, after returning to the RWTH Aachen in 1987, became head of the silicon technology modeling and simulation (TCAD) group. In 1995 he was appointed Professor at the University of Bremen, where his current research interests include TCAD and the theory of electromagnetic fields and networks.

MODELING FOR ULSI

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 47–51
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Interface Roughness Effects in Ultra-Thin Tunneling Oxides

D. Z.-Y. TING[a,b,*], ERIK S. DANIEL[b] and T. C. McGILL[b]

[a] *Department of Physics, National Tsing Hua University, Hsinchu, Taiwan 300, ROC;*
[b] *Thomas J. Watson, Sr., Laboratory of Applied Physics, California Institute of Technology Pasadena, California 91125, USA*

Advanced MOSFET for ULSI and novel silicon-based devices require the use of ultra-thin tunneling oxides where non-uniformity is often present. We report on our theoretical study of how tunneling properties of ultra-thin oxides are affected by roughness at the silicon/oxide interface. The effect of rough interfacial topography is accounted for by using the Planar Supercell Stack Method (PSSM) which can accurately and efficiently compute scattering properties of 3D supercell structures. Our results indicate that while interface roughness effects can be substantial in the direct tunneling regime, they are less important in the Fowler-Nordheim regime.

Keywords: Ultrathin, oxide, SiO_2, tunneling, interface roughness

1. INTRODUCTION

The continued scaling of Metal-Insulator-Semiconductor (MIS) device structures has brought much attention to ultrathin oxides. Normal operation of a MOSFET with 1.5 nm direct-tunneling gate oxide has been reported [1]. Tunneling through oxide barriers, as a mechanism for leakage currents, is of particular interest. Typical theoretical analysis models the oxide layer as a 1D barrier with an effective barrier height and an effective mass. The barrier height may be obtained experimentally or treated as a fitting parameter, while the effective mass is normally used as a parameter for fitting measured current-voltage (I-V) characteristics [2]. Tunneling coefficients can be calculated using the well-known WKB approximation. Integration of tunneling coefficient curves, with the appropriate Fermi factors describing carrier statistics, then yields an analytical I-V curve formula [3] which can be used conveniently for comparison with experimental data. A somewhat similar treatment uses multiple scattering theory instead of the WKB approximation to compute tunneling coefficients to provide clarification of mechanisms for leakage currents through ultrathin oxides [4]. A still more advanced treatment solves Poisson and Schrodinger equa-

* Corresponding author.

tions self-consistently for accumulated layers in metal-oxide-semiconductor devices to calculate tunneling currents [5].

An important aspect typically not treated in these models is oxide non-uniformity. Cundiff and co-workers [6] showed experimental evidence that roughness at the Si/SiO_2 interface increases with decreasing oxide layer thickness. This makes interface roughness particularly important in ultrathin oxides. Interface roughness can have dramatic effects on the current-voltage characteristics of MIS structures. It has been shown that constant current stressing of MIS structures in the Fowler-Nordheim tunneling regime can induce non-uniformities at the Si/SiO_2 interface [7]. This quasi-breakdown in the oxide can lead to dramatic increases in direct tunnel current. In this paper we treat the MIS tunnel structure using a 3D model which is similar to the standard models, but allows for 3D potential variations associated with Si/SiO_2 interfacial non-uniformity.

2. METHOD

Standard treatment uses a one-dimensional potential to describe the oxide barrier. With interfacial non-uniformity, we need to use a three-dimensional description. In principle, the variations in the non-uniform potential extend indefinitely in the directions along the interface. In practice, we do not perform computation for an infinite domain, but use instead a quasi-3D supercell geometry to approximate the physical problem. We treat the problem of tunneling through a non-uniform barrier using the Planar Supercell Stack Method (PSSM) [8]. The device structure treated by PSSM consists of an active layer sandwiched between two semi-infinite flat band electrode regions. Let the z axis be the direction of current flow. Then the active region is composed of 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. A one-band nearest-neighbor tight-binding Hamiltonian is used to describe the potential over this volume of interest. Our model is formally equivalent to the one-band effective mass equation.

$$-\frac{\hbar^2}{2}\nabla \cdot \frac{1}{m(\mathbf{x})}\nabla\psi + V(\mathbf{x})\psi = E\psi, \qquad (1)$$

discretized over a Cartesian grid, and subject to periodic boundary conditions (with supercell periodicity) in the x- and y-directions, and open boundary conditions in the z-direction. PSSM solves the quantum mechanical scattering problem exactly for the 3D geometry described by the planar supercell stack, and allows us to compute transmission coefficients with a high degree of numerical accuracy and efficiency. Note that even though the supercell geometry imposes an artificial periodicity to make computations tractable, the use sufficiently large supercells can minimize supercell artifacts and yield excellent descriptions of the physical problem.

3. RESULTS AND DISCUSSION

Our model of the MOS tunnel structure consists of a metal (or poly-Si) electrode, followed by a pure oxide layer, then a rough interfacial layer, and finally a silicon electrode. We assume that the rough interfacial layer consists of a 50%–50% mixture of oxide and Si in random configurations. The Si sites, and the oxide sites, for that matter, may aggregate and form patches. We will call these silicon patches islands, and characterize them by their lateral extent (island size) and the thickness of the interfacial layer (island height). For convenience, we also assume flat-band conditions in the metal and silicon electrodes, and let all the voltage drop occur in the oxide and the rough interface. The potential barrier height and the effective mass of the oxide barrier is taken to be 3.22 eV, and 0.35 m_0, respectively. We use a cubic mesh with discretization distance of 0.13575 nm, and 64×64 planar supercells in our simulations.

Figure 1 shows transmission coefficient spectra for three MOS tunnel structures with rough Si/

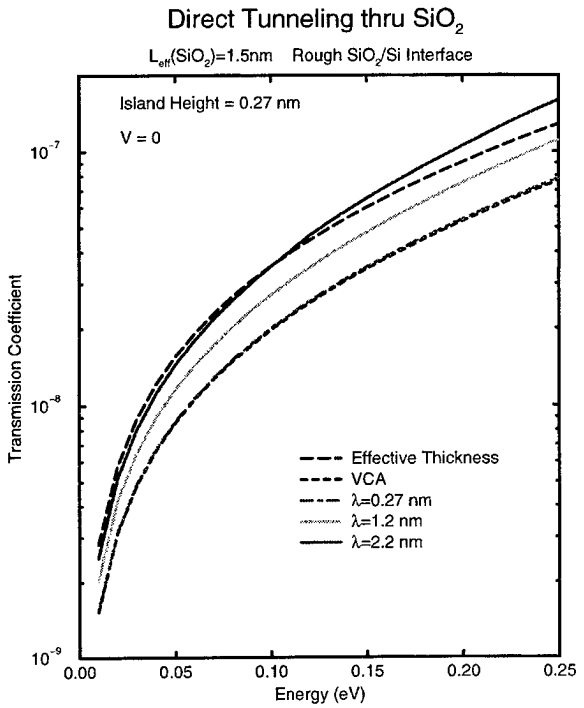


FIGURE 1 Transmission coefficient spectra for MOS tunnel structures with rough Si/SiO_2 interfaces characterized by different island sizes, at zero applied bias. For comparison, results for 1D models using effective barrier thickness and effective barrier height are also shown.

SiO_2 interfaces characterized by island sizes of λ = 0.27, 1.2 and 2.2 nm. The pure oxide tunnel barrier layer is 1.36 nm thick, and the height of the rough interfacial layer is 0.28 nm. Note that tunneling probabilities increase considerably with island size. Results of two 1D models are also plotted for comparison. The first treats the oxide and the interfacial layer as having a combined effective thickness of 1.5 nm (the thickness of the pure oxide layer, plus half the thickness of the interfacial layer); the second treats the interfacial layer as having a potential height which is the average of the oxide and the silicon potentials (i.e., the virtual crystal approximation, or VCA potential). Note that the effective thickness model does not offer a good description of the rough interfaces. The VCA result agrees well only with that obtained for the sample with the small island size. The reason for the agreement is because when island sizes are smaller than the deBroglie wavelength of the incoming electron, the random potential at the interface is only seen by the electron in an averaged (VCA) sense. In structures where island sizes become larger than the electron deBroglie wavelength, an electron, in a sense, finds its way through the "softer" spots (portions with lower potential, i.e., the silicon islands) of the interfacial layer, and thereby enhance its tunneling probability.

This is demonstrated in Figure 2. The top panel shows the transmission coefficient spectra of the λ = 0.27 nm and λ = 2.2 nm structures again for reference, while the bottom panel shows the curves of island transmission fraction, defined as the ratio of the sum of probability densities over all silicon sites in the interfacial layer divided by the total probability density in the interfacial layer. Since the interfacial layer consists of 50% silicon sites and 50% oxide sites, if an electron shows no preference for tunneling through the silicon islands

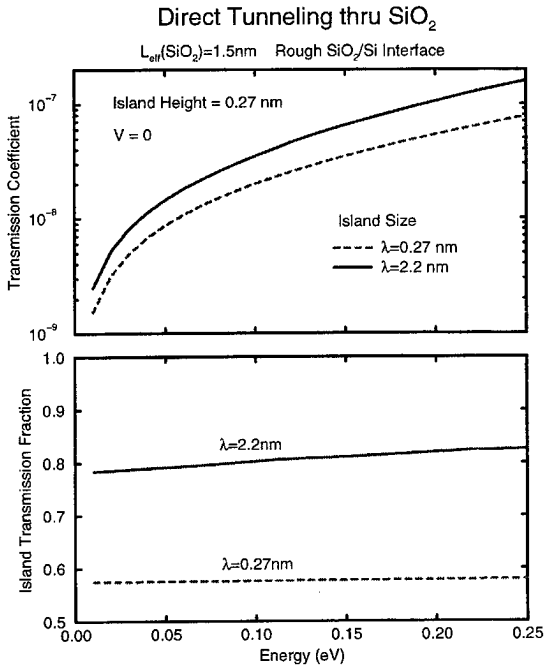


FIGURE 2 Transmission coefficients and island transmission fractions for two MOS structures with rough interfaces characterized by different island sizes, at zero applied bias.

or the oxides, the island transmission fraction would be 0.5; a fraction greater than 0.5 indicates a preference for tunneling through the silicon islands. Our results show that in the $\lambda = 0.27$ nm (small island) structure, there are essentially no preferential pathways through the roughness layer. However, in the $\lambda = 2.2$ nm (large island) structure, there is a definite preference for tunneling through the silicon islands. In fact, in this case the island transmission fraction increases with electron energy, or decreasing deBroglie wavelength.

We next examine the effect of interface roughness on tunneling properties under applied bias. We use the $\lambda = 2.2$ nm structure as in the previous figures. Figure 3 shows a set of transmission coefficient spectra for this structure under various applied biases. Transmission spectra for the corresponding VCA structure are also shown for comparison. Note that while there is a considerable difference between the supercell (with interface roughness) and VCA (without) results at low biases, the difference diminishes at high biases. This indicates that the role of interface roughness becomes less

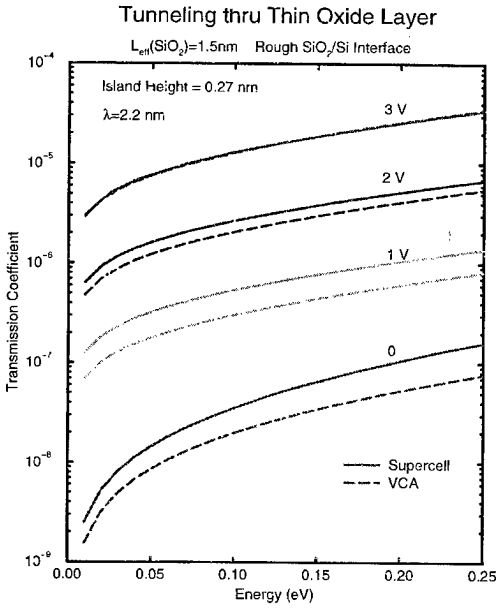


FIGURE 3 Transmission coefficient spectra for a MOS tunnel structure with rough Si/SiO_2 interfaces at several applied biases. Results for the corresponding VCA structure is also shown.

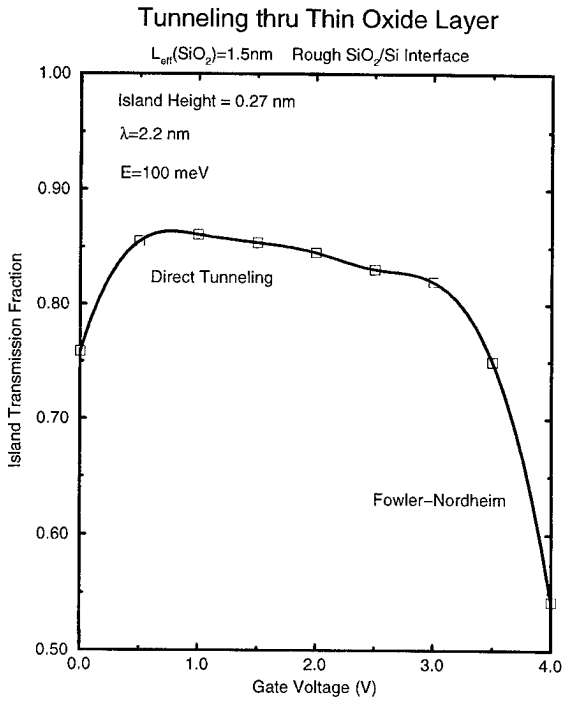


FIGURE 4 Island transmission fraction for the MOS tunnel structure depicted in Figure 3.

important as high biases. Figure 4 illustrates this further by showing the island transmission fraction for the same $\lambda = 2.2$ nm structure as a function of applied bias. Note that while there is a preference for tunneling through silicon islands at low biases, the preference drops rapidly as the applied bias exceeds 3 volts. The reason for this turns out to be rather simple. At low biases, the device structure operates under the direct tunneling regime. An electron going through the rough interfacial layer sees the oxide portions as energy barriers and the silicon islands as open pathways. At sufficiently high biases, the device operation transitions to the Fowler-Nordheim tunneling regime, where the conduction band edge at the trailing interface of the barrier (in our case, the rough interfacial layer) is biased below the incoming electron energies. Thus the oxide portions of the rough interfacial layer no longer act as electron barriers, putting them on more equal footing with the silicon islands. In other words, in the low-bias direct tunneling

regime, an electron traverses through the oxide portions of the interfacial layer with evanescent characteristics, and the silicon portions with propagating characteristics. In the high-bias Fowler-Nordheim tunneling regime, an electron traverses through both types of the interfacial layer as a propagating state, hence there would be less reasons for preferential pathways.

In summary, we have studied tunneling through MOS structures with rough Si/SiO_2 interfaces using a 3D supercell simulation. We find that interfacial non-uniformities make oxides more permeable in the direct tunneling regime, especially in structures with larger islands. In the Fowler-Nordheim regime, however, non-uniformity effects on tunneling characteristics are less pronounced.

Acknowledgment

The authors would like to thank O. J. Marsh for helpful discussions. This work was supported by the U.S. Office of Naval Research (ONR) under Grant No. N00014-92-J-1845, and by the ROC National Science Council under Grant No. NSC 86-2112-M-007-001.

References

[1] Momose, H. S., Ono, M., Yoshotomi, T., Ohguro, T., Nakamura, S., Saito, M. and Iwai, H. (1996). "1.5 nm direct-tunneling gate oxide Si MOSFETs", *IEEE Transactions on Electron Devices.*, **43**, 1233–1242.

[2] Brar, B., Wilk, G. D. and Seabaugh, A. C. (1996). "Direct extraction of the electron tunneling effective mass in ultrathin SiO_2", *Appl. Phys. Lett.*, **69**, 2728–2730.

[3] Simmons, J. G. (1963). "Generalized formula for the electrical tunnel effect between similar electrodes separated by a thin insulating film", *J. Appl. Phys.*, **34**, 1793–1803.

[4] Nagano, S., Tsukiji, M., Ando, K., Hasegawa, E. and Ishitani, A. (1994). "Mechanism of leakage current through the nanoscale SiO_2 layer", *J. Appl. Phys.*, **75**, 3530–3535.

[5] Rana, F., Tiwari, S. and Buchanan, D. A. (1996). "Self-consistent modeling of accumulation layers and tunneling currents through very thin oxides", *Appl. Phys. Lett.*, **69**, 1104–1106.

[6] Cundiff, S. T. *et al.* (1997). "Si/SiO_2 interface roughness: Comparison between surface second harmonic generation and *X*-ray scattering", *Appl. Phys. Lett.*, **70**, 1414–1416.

[7] Hirose, M., Alay, J. L., Yoshida, T. and Miyazaki, S. (1996). "Electronic density of states at the ultrathin SiO_2/Si interfaces", in The Physics and Chemistry of SiO_2 and the $Si\text{-}SiO_2$ interface-3. H. Z. Massoud, E. H. Poindexter and C. R. Helms, Ed., *Proc.*, **96–1**, 485–496, The Electrochemical Society, Pennington, NJ.

[8] Ting, D. Z.-Y., Kirby, S. K. and McGill, T. C. (1993). "Three-dimensional simulations of quantum transport in semiconductor nanostructures", *J. Vac. Sci. Technol. B*, **11**, 1738–1742.

Authors' Biographies

David Z.-Y. Ting is an Associate Professor of Physics at the National Tsing Hua University in Hsinchu, Taiwan, ROC, and a Visiting Research Associate at the California Institute of Technology. His research activities include theoretical studies of electronic and optical properties of semiconductor alloys and heterostructures, quantum transport in nanostructures and tunne devices, and optical simulations.

Erik S. Daniel received the B.A. degree in Physics and Mathematics from Rice University in Houston, TX in 1992. He is currently a doctoral candidate in Physics at the California Institute of Technology (Caltech) in Pasadena. His current research includes theoretical modeling, fabrication, and testing of semiconductor quantum effect devices.

Thomas C. McGill is the Fletcher Jones Professor of Applied Physics at the California Institute of Technology. His research interests are in the broad area of microstructures for modern electronics, semiconductor physics and devices.

THERMAL EFFECTS

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 53–58
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Modeling of Thermal Effects in Semiconductor Structures*

CHRISTOPHER M. SNOWDEN

Institute of Microwaves and Photonics, School of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT, UK

A fully coupled electro-thermal hydrodynamic model is described which is suitable for modelling active devices. The model is applied to the non-isothermal simulation of pseudomorphic high electron mobility transistors (pHEMTs). A large-scale surface temperature model is described which allows thermal modelling of semiconductor devices and monolithic circuits. An example of the application of thermal modelling to monolithic circuit characterization is given.

Keywords: Thermal, modeling, semiconductor, electro-thermal, pseudomorphic, HEMT, transistors

INTRODUCTION

Until relatively recently very little effort has been devoted to thermal considerations associated with most semiconductor devices, yet in many cases the temperature of the device and self-heating have a very strong impact on their performance. The high power densities associated with modern microwave transistors makes this class of active device particularly sensitive to temperature effects and dependent on good thermal management. Compound semiconductors such as GaAs generally have a far lower thermal conductivity than Si, which can in turn lead to high operating temperatures and poor thermal stability.

Thermal analysis of semiconductor devices can be considered in two broad categories-solutions in the proximity of the active device (usually restricted purely to a local hydrodynamic treatment for a cross-section of a few square microns) [1, 2], and solutions which encompass the whole die, which may have many active devices or elements that are interacting thermally [3, 4, 5]. Most simulations that follow the latter large-scale thermodynamic approach deal with steady-state heat sources and do not encompass the active device aspects. In practice, it is desirable to link the behaviour on the microscopic device level to the macroscopic scale, since this would allow a self-consistent thermal solution to be obtained.

* Invited paper.

Contemporary semiconductor devices often have complex geometries, incorporating air bridges and heat shunts in a highly three-dimensional structure. There is generally very limited control over the local device temperature and it is often very difficult to measure. Experimental characterization usually draws on temperature-sensitive liquid crystal paints for large-scale devices and infra-red scanning for more accurate results. The thermal time constants associated with most practical semiconductor devices lie in the range 10 μs to 10 ms. These relatively long time-constants are in contrast to the time-scales over which most carrier dynamics occur (typically in the range 0.1ps to 1 ns). A consequence of the large difference in time constants is that it is not usually practicable to follow to the steady-state the evolution of full time domain electro-thermal simulations of semiconductor devices, where the time-dependent heat flow equation is solved self-consistently with the carrier transport equations.

THERMAL AND ELECTRO-THERMAL MODELING

This paper addresses the modeling of thermal effects in semiconductor structures, examining the basis electro-thermal simulation and then considers the impact on transport modelling and finally large-scale thermal modelling of complete structures (e.g. a full die). Self-heating has a significant impact on parameters such as mobility, gneration-recombination and trap occupancy. A rigorous electro-thermal solution is described here with a multi-cell coupled thermal and transport model. This is achieved by solving a set of hydrodynamic equations derived from the Boltzmann transport approximation (utilizing four moments), coupled to an accurate solution of the heat flow equation.

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\mathbf{v}) = 0 \tag{1}$$

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = \frac{q}{m^*}\mathbf{E} - \frac{2}{3m^* n}\nabla(nw) + \frac{1}{3n}\nabla(nv^2) - \frac{\mathbf{v}}{\tau_p} \tag{2}$$

$$\frac{\partial w}{\partial t} + \mathbf{v} \cdot \nabla w = q\,\mathbf{v} \cdot \mathbf{E} - \frac{2}{3n}\nabla \cdot \left[n\mathbf{v}\left(w - \frac{m^*}{2}v^2 \right) \right] - \frac{1}{n}\nabla \cdot \mathbf{Q} - \frac{w - w_0}{\tau_w} \tag{3}$$

$$c_L \rho_L \frac{\partial T}{\partial t} = \nabla \cdot (\kappa_L \nabla T) + H_S \tag{4}$$

$$H_S = \mathbf{J} \cdot \mathbf{E} + qE_g G \tag{5}$$

Here q represents the charge on an electron and $\nabla \cdot \mathbf{Q}$ represents the energy flow, n is the electron density, $\mathbf{v}$ the electron velocity, E electric field, m^* effective mass, τ_p and τ_w momentum and energy relaxation times respectively, and w average electron energy, c_L specific heat, ρ_L density, κ_L lattice thermal conductivity, T lattice temperature, and G generation-recombination rate. The transport parameters are assumed to be temperature-dependent and are obtained from bulk Monte Carlo simulations. The average electron energy is expressed in terms of kinetic energy and electron temperature T_e, as,

$$w = \frac{1}{2}m^* v^2 + \frac{3}{2}k\,T_e \tag{6}$$

where T_e is the average electron temperature and k is Boltzmann's constant. The energy assumptions behind this model are discussed in [6].

The numerical solution of the heat flow equation (equation 4) requires careful consideration to obtain an accurate solution, with a third-order boundary condition. Ghione *et al.*, have suggested that to obtain accurate results for MESFETs, the simulation domain for analysis should be extended horizontally for up to three times the source-drain contact spacing and to a depth of up to ten times the active layer thickness of the device [2].

Thermal modelling of semiconductor devices requires that the transport and material parameters are well characterized as a function of temperature. In the case of GaAs the following set of temperature-dependent material parameters are required:

$$\text{Low field mobility} \quad \mu_0(T) = \mu_0(300)\left(\frac{300}{T}\right)^{2.3} \tag{7}$$

$$\text{Permittivity} \quad \varepsilon_r(T) = \varepsilon_r(300)[1 + B_\varepsilon(T - 300)] \tag{8}$$

$$\text{Schottky barrier height} \quad V_{Bi}(T) = V_{Bi}(T_0) - k_T(T - T_0) \tag{9}$$

$$\text{Thermal conductivity} \quad \kappa_L = 108.0(T - 273.15)^{-0.26} \tag{10}$$

where $\mu_0(300)$ is the low field mobility at 300 K, $\varepsilon_r(300)$ is the permittivity at 300 K, k_T is the temperature coefficient of the barrier height (in the range -0.26 to -1.6×10^{-3} VK^{-1} depending on temperature), T_0 is the reference temperature, B_ε is the permittivity temperature coefficient (in the range 0.9 to $1.2 \times 10^{-4}\,\mathrm{K}^{-1}$).

Local iso-thermal and non-isothermal models have been developed for a wide variety of semiconductor devices including BJTs, HBTs, MESFETs and HEMTs (for example [7, 8, 9]). The quasi-two-dimensional pHEMT model described in [8, 9] and illustrated in Figure 1 is well suited to electro-thermal modelling because of its very high speed of solution. Examples of isothermal and non-isothermal DC characteristics obtained using this approach are shown in Figure 2 for a 0.2 micron gate length pHEMT. It should be noted that this pHEMT model also includes a self-consistent solution of the Schrödinger equation with the Poisson and transport equations. A comparison of measured and simulated non-isothermal DC characteristics for a similar pHEMT are shown in Figure 3. Differences between the two DC characteristics are largely attributable to uncertainties in the exact density of the pulse-doped layer in the device (there is no fitting of the data between measured and simulated results).

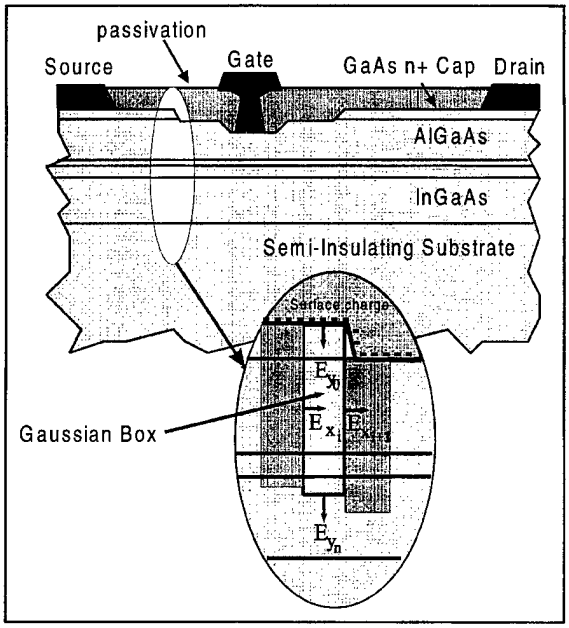
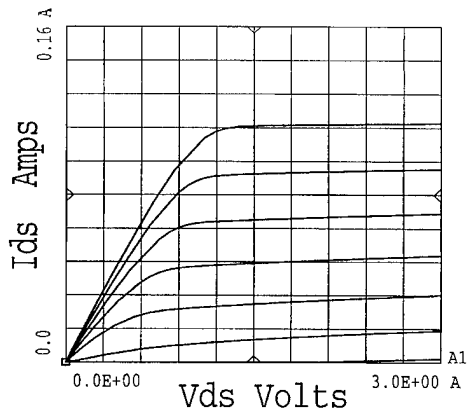
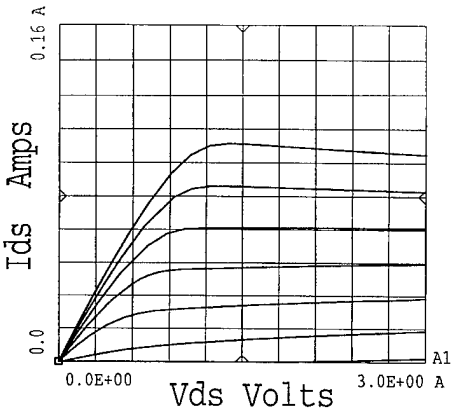


FIGURE 1 Quasi-two-dimensional model for a pseudo-morphic high electron mobility transistor, illustrating the Gaussian box technique as described in [8] and [9].

A three-dimensional thermal analysis technique has been developed which allows full-scale analysis of large device and die structures, Figure 4. The steady-state heat flow equation $\nabla \cdot \kappa(T)\nabla T = 0$ is solved for a three-dimensional temperature distribution $T(x, y, z)$ using the method of Liou [4] and Gao *et al.* [5]. Since in practice most semiconductor devices are located very close to the surface of most die, it is generally only required to obtain the temperature at or close to the surface. In these circumstances it is sufficient to calculate only the surface temperature $T(x, y, 0)$. Liou and Gao's method uses a double Fourier expansion method to speed up the solution. An initial estimate of the temperature T', is first obtained by assuming that the temperature-dependent



(a)



(b)

V_{GS}: 0.5 to -2.5 V steps of -0.5 V

FIGURE 2 Simulated DC characteristics for (a) isothermal and (b) non-isothermal pHEMT models, implemented using the quasi-two-dimensional model.

thermal conductivity $\kappa(T)$ is equal to the thermal conductivity κ_0 at the mounting (heat sink) temperature T_0,

$$T'(x,y,0) = \sum_{i=1}^{N} \left\{ \frac{Q_i t \Delta x \Delta y}{\kappa_0 L_x L_y} + \sum_{m,n=0}^{\infty} \left[C_{m,n} Q_i \frac{\tanh(\gamma_{m,n} t)}{\kappa_0 \gamma_{m,n}} \cos\frac{m\pi x_i}{L_x} \cos\frac{n\pi y_i}{L_y} \right] \cos\frac{m\pi x}{L_x} \cos\frac{n\pi y}{L_y} \right\} + T_0. \tag{7}$$

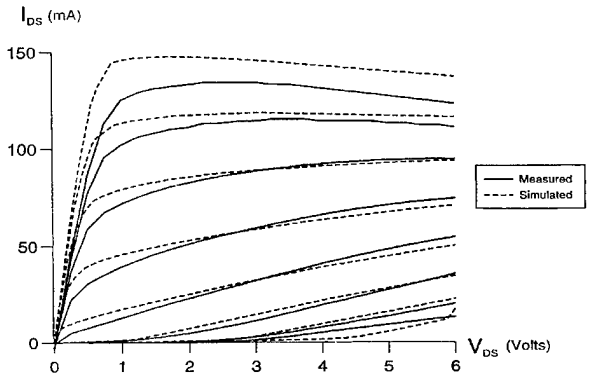


FIGURE 3 Measured and simulated DC non-isothermal characteristics for a 0.2 micron gate length pHEMT.

where,

$$\gamma_{m,n} = \pi \left[\left(\frac{m}{L_x}\right)^2 + \left(\frac{n}{L_y}\right)^2 \right]^{0.5} \tag{8}$$

$$C_{m,n} = \frac{16}{mn\pi^2} \sin\frac{m\pi\Delta x}{2L_x} \sin\frac{m\pi\Delta y}{2L_y}, \quad m \neq 0, n \neq 0 \tag{9}$$

$$C_{m,n} = \frac{4\Delta y}{m\pi L_y} \sin\frac{m\pi\Delta x}{2L_x}, \quad m = 0, n = 0 \tag{10}$$

$$C_{m,n} = \frac{4\Delta x}{n\pi L_x} \sin\frac{n\pi\Delta y}{2L_y}, \quad m = 0, n \neq 0 \tag{11}$$

where t is the thickness of the substrate, m and n the Fourier coefficients. The active region of the die where the devices are located is subdivided into regions of area $\Delta x \Delta y$ and are treated at heat sources where $Q_i \Delta x \Delta y$ is the heat generated at the i^{th} unit area located at (x_i, y_i). N is the number of unit areas on the surface. The upper limit of the coefficients m and n must be as large as possible and in practice values in excess of 70 ensure a local accuracy of better than 1 K. The heatsink is assumed to be attached to the bottom of the substrate (at $T = T_{\text{mount}}$), whilst all other surfaces except the heat sources (the active region of the semiconductor devices) are assumed to be adiabatic. The temperature T' is corrected for the temperature-dependent thermal conductivity using Kirchhoff's transform, yielding,

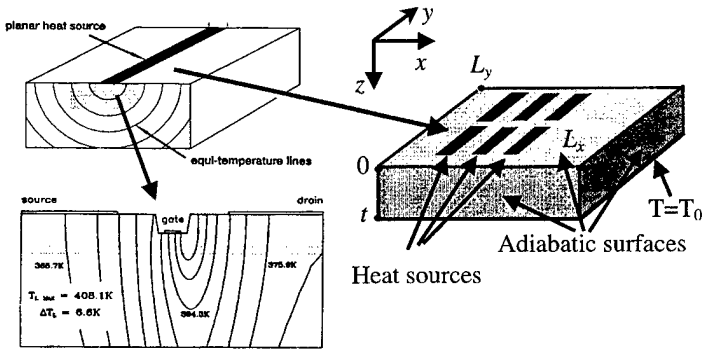


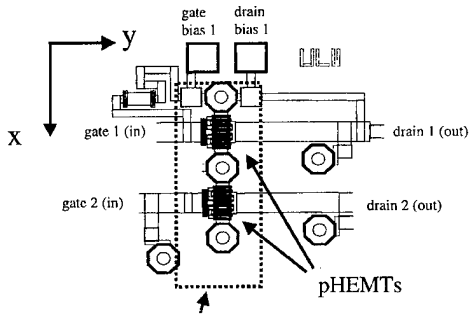
FIGURE 4 Three-dimensional temperature distribution and model based on 'line' heat sources associated with the local heat generation within the active channel of the device.

$$T(x,y,0)=\left[\frac{1}{T_0^{b-1}}-\frac{(b-1)(T'-T_0)}{T_0^{b}}\right]^{-(1/b-1)} \tag{12}$$

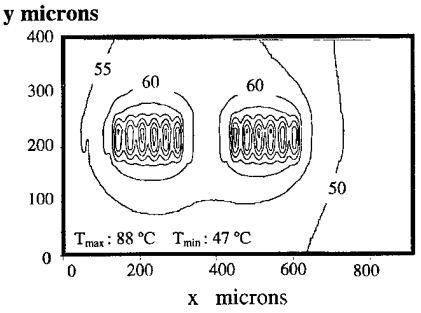
which assumes that the thermal conductivity is proportional to $(T/T_0)^{-b}$. In the case of GaAs $b=1.22$.

The heat sources may be treated locally using the coupled electro-thermal model, and the solution is iterated between the local transport model and the full die model. The temperature obtained from the 3D die model may be used as an initial estimate in the local transport model. The power density obtained from the solution of the transport model can in turn then be used to set Q_I in the 3D die model. It should be noted that it is necessary to minimise the number of cells since a separate transport model must be solved for each cell. This does not present a significant computational burden since the quasi-two-dimensional models can be solved in less than 0.01 cpu seconds on a typical workstation and in practice the surface temperature calculation time is the dominant factor.

An example of the application of thermal modelling is shown in Figure 5, where the surface temperature distribution is calculated for a pHEMTs 38 GHz monolithic power amplifier design. A partial view of the monolithic die layout in the vicinity of the pHEMTs is shown in Figure 5(a) and the region chosen for thermal analysis is outlined (an area of 800 × 400 microns). The die is



(a)



(b)

FIGURE 5 Application of thermal modelling to monolithic pHEMT amplifier design. (a) partial view of layout of a balanced 38 GHz monolithic pHEMT power amplifier (b) surface temperature distribution for the area defined by the dashed line.

400 microns thick and the mounting temperature at the back of the GaAs substrate is 40 °C. The results of the thermal simulation are shown in

Figure 5(b), where the higher temperatures surrounding the pHEMT finger regions are evident. The 3D thermal simulator utilized over 20,000 unit cells in the x, y plane to obtain the surface temperature distribution. This technique has also been applied to the coupled electro-thermal modelling of heterojunction bipolar transistors [10]. These results have been compared with infra-red measurements and were found to show good agreement (better than 6% over the whole die).

CONCLUSIONS

An electro-thermal model has been presented for semiconductor devices capable of describing the interaction between the lattice temperature, carrier transport and the observed electrical behaviour. A coupled hydrodynamic-thermal model has been used to represent the carrier transport in the vicinity of the active region of the device. Large-scale assessment of surface temperature has been achieved using a steady-state electro-thermal model. The significance of thermal effects in small-scale devices has been illustrated for the case of the short gate-length pHEMT both as a discrete device and integrated into a practical monolithic circuit.

Acknowledgements

The author would like to acknowledge Chris Morton, Rob Johnson, Rob Drury, José Santos and John Atherton for their contributions to this research. HP-EEsof, M/A-COM and the UK EPSRC supported some aspects of this work.

References

[1] Ghione, G., Golzio, P. and Naldi, C. U. (1987). "Thermal Analysis of Power GaAs MESFETs", *Proc. NASECODE V Conf.*, Ed. J. J. H. Miller, pp. 195–200.

[2] Ghione, G., Golzio, P. and Naldi, C. U. (1988). "Self-Consistent Thermal Modelling of GaAs MESFETs", *Alta Frequenza*, **LVII**(7), pp. 311–319.

[3] Liou, L. L., Ebel, J. L. and Huang, C. I. "Thermal Effects on the Characteristics of AlGaAs/GaAs Heterojunction Bipolar Transistors Using Two-Dimensional Numerical Simulation", *IEEE Trans. Electron Devices*, **ED-40**(1), pp. 35–42, January 1993.

[4] Liou, L. L. and Bayraktaroglu, B. "Thermal Stability Analysis of AlGaAs/GaAs Heterojunction Bipolar Transistors With Multiple Emitter Fingers", *IEEE Trans. Electron Devices*, **ED-41**(5), pp. 629–635, May 1994.

[5] Gao, G-B, Whang, M-Z, Gui, X. and Morkoç H. "Thermal Design Studies of High-Power Heterojunction Bipolar Transistors", *IEEE Trans. Electron Devices*, **ED-36**(5), pp. 854–863, May 1989.

[6] Lundstrom, M. (1990). *Fundamentals of Carrier Transport.* Addison-Wesley Publishing Co.

[7] Snowden, C. M. and Pantoja, R. R. "Quasi-two-dimensional MESFET simulations for CAD" *IEEE Trans. Elec. Dev.*, **36**(9), pp. 1564–1574, September 1989.

[8] Morton, C. G., Atherton, J. S., Snowden, C. M. Pollard, R. D. and Howes, M. J., "A Large-Signal Physical HEMT Model", *Proc. IEEE MT T-S, San Francisco*, pp. 1759–1762, June 1996.

[9] Drury, R. and Snowden, C. M. "A Quasi-Two-Dimensional HEMT Model for Microwave CAD Applications", *IEEE Trans. Electron Devices*, **42**(6), pp. 1026–1032, June 1995.

[10] Snowden, C. M. Large-Signal Microwave Characterization of AlGaAs/GaAs HBT's based on a Physics-based Electrothermal Model", *IEEE Trans. MTT*, **45**(1), pp. 58–71, January 1997.

Authors' Biography

Christopher M. Snowden received the B.Sc, M.Sc and Ph.D. degrees from the University of Leeds. He currently holds the Chair of Microwave Engineering in the Institute of Microwaves and Photonics at the University of Leeds where he is also Head of the School of Electronic and Electrical Engineering. He was a Visiting Research Associate at the California Institute of Technology in 1987 and during the period 1990–91 he represented M/A-COM Inc. as Senior Staff Scientist. He is a Member of the MIT Electromagnetics Academy and is a Fellow of the IEEE and a Fellow of the IEE. He is a Distinguished Lecturer for the IEEE (Electron Devices Society).

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 59–64
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Carrier Thermal Conductivity: Analysis and Application to Submicron-Device Simulation

A. GREINER[a], L. VARANI[b], L. REGGIANI[c], M. C. VECCHI[d], T. KUHN[e] and P. GOLINELLI[f]

[a] *Institut für Mikrosystemtechnik, Albert-Ludwigs-Universität Freiburg, Am Flughafen 17, 79110 Freiburg, Germany;*
[b] *CEM2 (CNRS UMR 5507), Université Montpellier II, 34095 Montpellier Cedex 5, France;*
[c] *INFM, Dipartimento di Scienza dei Materiali, Università di Lecce, Via Arnesano, 73100 Lecce, Italy;*
[d] *Dipartimento di Elettronica Informatica e Sistemistica, Università di Bologna, viale Risorgimento 2, 40136 Bologna, Italy;*
[e] *Institut für Theoretische Physik II, Westfälische Wilhelms-Universität, Wilhelm--Klemm-Str. 10, 48149 Münster, Germany;*
[f] *INFM, Dipartimento di Fisica, Università di Modena, Via Campi 213/A, 41100 Modena, Italy*

Within a correlation-function (CF) formalism, the kinetic coefficientsof charge carriers in semiconductors are studied under different conditions. For the case of linear response in equilibrium, thetransitions from the non-degenerate to the degenerate regimes as wellas from ballistic to diffusive conditions are discussed within ananalytical model. Generalizing the method to high-field transport innondegenerate semiconductors, the CFs are determined by Monte Carlo (MC) calculations for bulk silicon from which the appropriate thermalconductivity has been obtained and included into the hydrodynamic code HEIELDS. For an n^+nn^+ submicron structure the temperatureand velocity profiles of the carriers have been calculated with HFIELDS.

Keywords: Kinetic coefficients, hydrodynamic transport equations, correlation functions, Monte Carlo

1. INTRODUCTION

Thermal conductivity of charge carriers is of fundamental interest in describing transport phenomena in bulk materials as well as electronic devices. To provide a microscopic theory for this coefficient, the CF approach represents a very effective method [1]. As a consequence of the fluctuation-dissipation theorem, the carrier transport coefficients may be determined by the spectrum of the fluctuations in the system. Results for the thermal conductivity available in literature are mainly based on relaxation-time approximations [2, 3]. A weighting of the single relaxation time with a power of the energy yields a generalization ofthe Wiedemann-Franz law (WFL) given by

$$\kappa = \left(\frac{5}{2} + c\right)\frac{k_B^2}{q}\mu n T, \tag{1}$$

where c is the so called power law exponent, μ is the mobility and T the electron temperature. This WFL together with the energy relaxation time is usually introduced within hydrodynamic approaches [4–8].

In this paper we present the appropriate set of CFs for the cases of linear response around thermodynamic equilibrium in the ballistic and

diffusive regime under different degeneracy conditions [9], as well as the hot-carrier regime in the classical-diffusive condition [10, 11] with its applications to the simulation of an n^+nn^+-structure.

2. LINEAR RESPONSE REGIME

In the linear regime, generalized fluxes are the response of the material to generalized externally applied forces mediated by the kinetic (or Kelvin-Onsager) coefficients $L_{\mu\nu}$ according to

$$\mathbf{j}_\mu = \sum_\nu L_{\mu\nu}\mathbf{X}_\nu \tag{2}$$

with μ, $\nu = 1$, 2, $\mathbf{j}_1$ and $\mathbf{j}_2$ denoting the electrical current and the energy-flux densities, respectively, while $\mathbf{X}_1$ and $\mathbf{X}_2$ are the generalized driving forces. The kinetic coefficients $L_{\mu\nu}(\omega)$ in general depend on angular frequency ω and are defined by means of the Fourier-Laplace transforms $S_{\mu\nu}(\omega)$ of the respective CFs $C_{\mu\nu}(t)$ from [12]:

$$L_{\mu\nu}(\omega) = \frac{V S_{\mu\nu}(\omega)}{k_B T} \tag{3}$$

where V is the volume, k_B is the Boltzmann constant, T the absolute temperature and:

$$S_{\mu\nu}(\omega) = \int_0^\infty d\omega C_{\mu\nu}(t)\exp(-i\omega t). \tag{4}$$

Here $C_{\mu\nu}(t)$ is the symmetrized CF at thermal equilibrium

$$C_{\mu\nu}(t) = \frac{1}{2}\langle \delta\mathbf{J}_\mu(0)\delta\mathbf{J}_\nu(t) + \delta\mathbf{J}_\nu(0)\delta\mathbf{J}_\mu(t)\rangle \tag{5}$$

where the $\delta\mathbf{J}_\nu(t) = \mathbf{J}_\nu - \langle\mathbf{J}_\nu\rangle$ denote the operators of the respective observed fluctuating quantities, and $\langle\ldots\rangle$ indicates averaging with respect to the stationary statistical operator. The thermal conductivity κ is given as a function of the $L_{\mu\nu}$

$$\kappa(\omega) = \frac{L_{11}(\omega)L_{22}(\omega) - L_{12}(\omega)L_{21}(\omega)}{TL_{11}(\omega)}. \tag{6}$$

For a finite one-dimensional conductor of length ℓ limited by ideal (i.e. completely absorbing and thermalizing) contacts, if scattering is treated within the relaxation time approximation, the thermal conductivity κ can be calculated analytically showing the transitions both from ballistic to diffusive (with increasing ℓ) and from degenerate to non-degenerate (with increasing T) conditions. The static (i.e. $\omega = 0$) limiting cases are given by [9]

$$\kappa^{ndg,bl} = \frac{nk_B^2 T}{m}\tau_T^{ndg} = \frac{n\ell k_B^{3/2} T^{1/2}}{(2\pi m)^{1/2}}, \tag{7a}$$

$$\kappa^{dg,bl} = \frac{\pi^2}{6}\frac{nk_B^2 T}{m}\tau_T^{dg} = \frac{2\pi^2}{3}\frac{\ell k_B^2 T}{h}, \tag{7b}$$

$$\kappa^{ndg,df} = \frac{3}{2}\frac{nk_B^2 T}{m}\tau_c, \tag{7c}$$

$$\kappa^{dg,df} = \frac{\pi^2}{3}\frac{nk_B^2 T}{m}\tau_c. \tag{7d}$$

Here $\tau_T^{ndg/dg}$ is the carrier transit time in the ballistic (bl) case under non-degenerate and degenerate conditions, respectively, while in the diffusive (df) cases τ_c denotes the relaxation time. Figure 1 shows the behavior of $\kappa(0)$ at fixed carrier density as function of ℓ and T. The maximum value is reached in the ballistic degenerate case.

3. FAR FROM EQUILIBRIUM (HOT-CARRIER) CASE

The CFs of microscopic fluxes are calculated with respect to stationary values at the given bias point and the thermodynamic temperature T is replaced by the noise-temperature spectrum $T_n(E, \omega = 0)$ [13] associated with velocity fluctuations at the given field E, where we restrict to the stationary case ($\omega = 0$). The thermal conductivity parallel to the applied electric field, $\kappa(E, \omega = 0)$, generalized to hot-carrier conditions is thus given by Eq. (6) with the following replacement:

$$T \Rightarrow T_n(E,\omega) = \frac{V\,\mathrm{Re}[S_{11}(E,\omega)]}{k_B\mathrm{Re}[\sigma_d(E,\omega)]}, \tag{8}$$

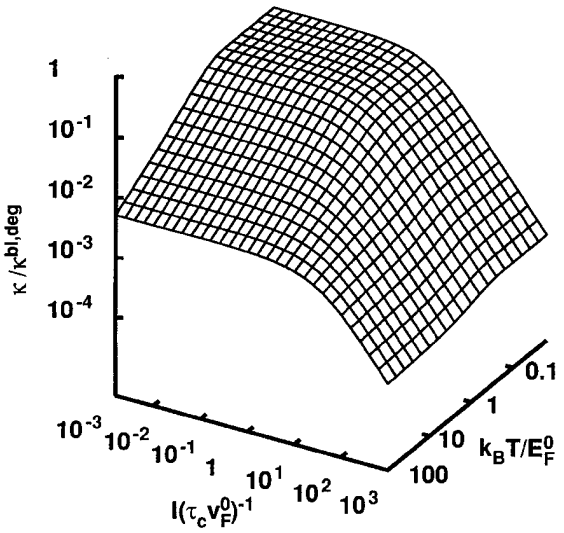


FIGURE 1 Thermal conductivity at zero frequency (normalized to its value in the ballistic degenerate case) at a fixed density n as a function of device length and temperature. v_F^0 ($E_F{}^0$) denote the Fermi velocity (Fermi energy) at $T=0$ K for the given density n and τ_c is the relaxation time.

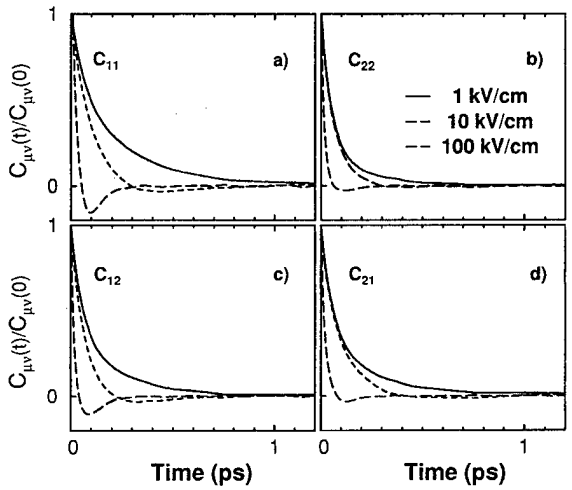


FIGURE 2 Longitudinal CFs normalized to their initial value for extrinsic n-Si with doping concentration $n=10^{17}\,cm^{-3}$ at 300 K and increasing electric-field strengths of 1, 10, and 100 kV/cm.

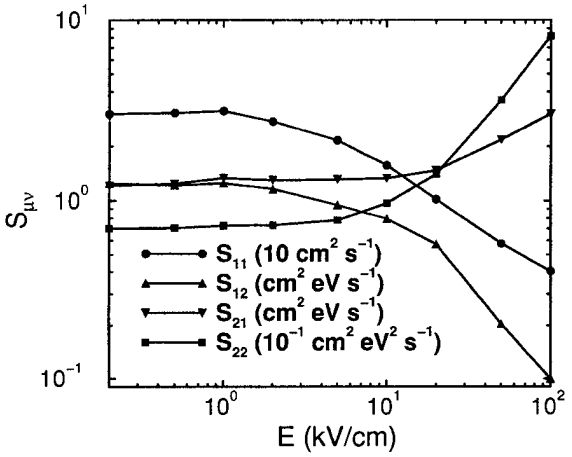


FIGURE 3 Spectral densities of the CFs of Figure 2; at $\omega=0$ as a function of electric-field strength.

where σ_d is the differential conductivity. Under thermal-equilibrium conditions (i.e. $E=0$) $T_n(\omega)=T$, and standard linear-response formalism is recovered [14]. The CFs and conductivities entering the definition of κ are calculated using MC simulations [15]. We have considered the case of extrinsic n-Si, doping concentration $10^{17}\,cm^{-3}$, at $T=300$ K. Figure 2 from a) to d) reports the four longitudinal CFs $C_{\mu\nu}(t)$ (fluctuations of flux variable along the electric field direction) at increasing values ofthe electric field normalized to their respective initial values. At increasing fields the correlations decay faster. $C_{22}(t)$ shows the fastest decay, which indicates that its decay-rate is the sum of momentum and energy rates. Thus we see that even at thermal equilibrium an approach based on a single time-scale is very poor. We observe that the variances $C_{11}(0)$, $C_{22}(0)$ and $C_{12}(0)=C_{21}(0)$ increase systematically with increasing electric field,which is due to hot-carrier effects. Figure 3 reports thelow frequency spectral densities $S_{\mu\nu}(E, \omega=0)$ of the respective CFs as a function of the electric field. Their different behaviors are the result of the competitive effect between shortening time-scales and increasing variances of the corresponding CFs at increasing fields. The carrier thermal-conductivity at low frequency calculated from Eq. (6) is shown in Figure 4. It is compared to the values obtained from a parameterized WFL (1), for different values of the power law exponent. As can be noticed, the

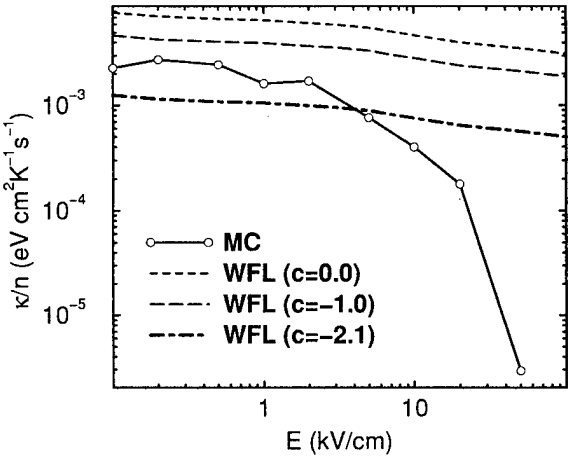


FIGURE 4 Thermal conductivity per unit concentration in *n*-Si at 300 K as a function of electric-field strength. The different curves are obtained by setting $c=0.0$, $c=-1.0$ and $c=-2.1$ in the WFL and by using the CF approach together with MC.

strong decrease at high fields may not be reproduced by a parameterized WFL.

4. DEVICE SIMULATIONS

To study the role of thermal conductivity in device csimulation, the transport in a Si submicron n^+-n n^+ structure has been analyzed by means of the hydrodynamic device simulator HFIELDS [16]. The simulated structure is the same n^+-n-n^+ diode analyzed in [17]: the high doped n^+-regions have length 0.1 μm and doping concentration $5\times10^{17}\,\mathrm{cm}^{-3}$, the lower doped *n*-region has length 0.4 μm and doping concentration $2\times10^{15}\,\mathrm{cm}^{-3}$.

The electron mobility $\mu(\varepsilon)$ and energy-relaxation time $\tau_\varepsilon(\varepsilon)$ as a function of the mean electron energy ε, required in the hydrodynamic simulator, are calculated from the velocity and energy versus field curves obtained from MC simulations.

For the thermal conductivity both the models obtained from the results of MC simulations and from the parameterized WFL are adopted. In Figure 5 the velocity profile along the device obtained with different models of the thermal conductivity are shown. When the WFL (1) is applied, the spike in the electron velocity curve can be remarkably reduced by setting the value of the power law exponent to $c=-2.1$; at the same time the velocity profiles is smoothed with respect to the MC solution presented in [17] for the same structure. This can be ascribed to the slow decrease of the WFL thermal conductivity with increasing electric field. The CF approach yields to a stronger dependence of the thermal conductivity on the electric field, which allows to obtain a better comparison with MC data. The carrier temperature profile is shown in Figure 6: alsoin this case the WFL approach tends to smooth the

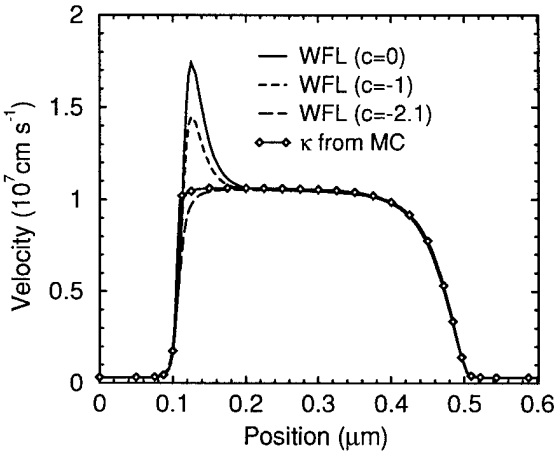


FIGURE 5 Velocity profile of the n^+-n-n^+ device calculated with four different thermal conductivities as shown in Figure 4.

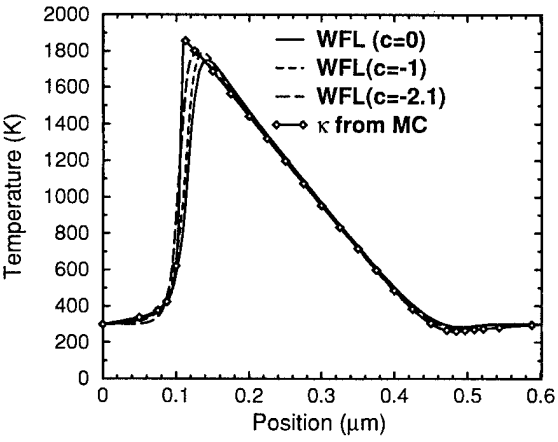


FIGURE 6 Temperature profile of the n^+-n-n^+ device for the cases as shown in Figure 5.

profile with respectto the case when the results of the CF method are applied.

5. CONCLUSION

We have shown that the CF formalism provides an effective method forcalculating electronic transport parameters in different transportregimes. These parameters, calculated in the case of non-equilibriumand incorporated into the hydrodynamic simulatorhfields, allowfor a description of the electronic behavior of submicron structuresin agreement with the results of more sofisticated microscopic approaches.

Acknowledgments

Support of this work by the Commission of European Community through the contract ERBCHRXCT920047 is gratefully acknowledged. Authors gratefully thank Drs. R. Brunetti and M. Rudan for fruitful discussions.

References

[1] Kubo, R., Toda, M. and Hashitsume, N. (1991). *Statistical Physics II* (Springer, Berlin).
[2] Jones, W. and March, N. H. (1973). *Theoretical Solid State Physics* (Dover New York), **2**.
[3] Kireev, P. S. (1974). *Semiconductor Physics*, 2 ed. (Mir Publishers, Moscow).
[4] Blotekjaer, K. (1970). *IEEE Trans. Electron. Devices*, ED-**17**, 38.
[5] Anile, A. and Pennisi, S. (1992). *Phys. Rev.*, **B46**, 13186.
[6] Stettler, M., Alam, M. and Lundstrom, M. (1993). *IEEE Trans. Electron. Devices*, **ED-40**, 733.
[7] Jyegal, J. and Massa, T. D. (1994). *Journal of Applied Physics*, **76**, 4413.
[8] Rudan, M., Vecchi, M. and Ventura, D., In *Mathematical problems in semiconductors physics*, **340** *of Pitman Research Notes in Mathematics Series*, edited by, P. Marcati, P. Markowich and R. Natalini (Longman, 1995), pp. 186–214.
[9] Greiner, A., Reggiani, L., Varani, L. and Kuhn, T. (1997). *Phys. Rev. Lett.*, **78**, 1114.
[10] Golinelli, P. *et al.* (1996). In *Hot Carriers in Semiconductors*, edited by , K. Hess, , J. Leburton and U. Ravaioli (Plenum Press, New York), p. 405.
[11] Golinelli, P., Varani, L. and Reggiani, L. (1996). *Phys. Rev. Lett.*, **77**, 1115.
[12] Mohling, F. (1982). *Statistical Mechanics* (Wiley, New York).
[13] Price, P. (1965). In *Fluctuation Phenomena in Solids*, edited by, R. Burgess (Academic, New York).
[14] Kubo, R. (1959). In *Lectures in Theoretical Physics*, edited by, W. Brittin and L. Dunham (Interscience Publs., New York), **1**, p. 120.
[15] Brunetti, R. *et al.* (1981). *Journal of Applied Physics*, **52**, 6713.
[16] Forghieri, A. *et al.* (1988). *IEEE Trans. on CAD of ICAS* **CAD-7**, 231.
[17] Gnudi, A., Odeh, F. and Rudan, M. (1990). *European Trans. on Telecommunications and Related Technologies*, **1**, 307.

Authors' Biographies

Andreas Greiner born 1960 in Ludwigsburg (GERMANY), graduated in physics in 1988 and received the Ph.D. in physics in1992 from Stuttgart University. His main research activity is in carrier transport simulation of semiconductor devices.

Luca Varani born 1963 in Carpi (ITALY), graduated in physics in1989, the Ph.D. degree in physics in 1993 from Modena University and inelectronics in 1996 from Montpellier University. He is researcher at the Centre d'Electronique et de Micro-optoelectronique de Montpellier(CNRS UMR 5507). His main research activity is in the field of electronic transport and noise in semiconductors.

Lino Reggiani, born 1941 in Modena (ITALY), received the Dr. in Physics degree in 1968 and the Diploma di Perfezionamento in Physics in 1972 from Modena University. Since 1994 he is full Professor at the Material Science Department of the Lecce University. His main research activity is in the field of electron transport and fluctuations in semiconductors.

Maria Cristina VECCHI, born 1964 in Avezzano (ITALY), graduated in electrical engineering in 1990 and received the Ph.D. in 1995 from Bologna University. She is a post-doctoral fellow of the DEIS at University of Bologna. Her research activity is in the field of semiconductor device simulations.

Tilmann Kuhn, born 1959 in Alpirsbach (GERMANY), graduated in physics in 1984 and received the Ph.D. in physics from the University

of Stuttgart. Since 1996 is full professor at the University of Münster. His research interest is in the field of carrier dynamics and fluctuations in semiconductors far from equilibrium on ultrashort length or time scales.

Paola Golinelli, born 1969 in Rivoli (ITALY), received the Dr. in Physics degree in 1995 from Modena University. Her main research activity has been the generalization of charge carrier thermal conductivity to hot-carrier conditions.

MOLECULAR STRUCTURES

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 65–74
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Molecular Wire Interconnects: Chemical Structural Control, Resonant Tunneling and Length Dependence

MATHIEU KEMP[a], VLADIMIRO MUJICA[b], ADRIAN ROITBERG[c] and MARK A. RATNER[a,*]

[a] *Department of Chemistry and Materials Research Center, Northwestern University, Evanston, Illinois 60208-3113;* [b] *Universidad Central de Venezuela, Caracas, Venezuela;* [c] *National Institute of Standards and Technology, Gaithersburg, MD*

Molecular wires have several promising features, that would appear to make them ideal for advanced interconnects in nanoscale electronic devices. We discuss several aspects of the linear and nonlinear conductance of molecular wire interconnects. Topics include energy dependence of molecular conductance, resonant tunneling behavior, control of conductance by molecular structure and geometry, length dependence including the tunneling regime energetics. Design rules using molecular interconnects will differ substantially from those with more standard, lithographically structured silicon interconnects. In particular, the dissipation mechanisms will differ, both tunneling and ballistic regimes should be available, coulomb blockade and staircase behavior will be observed (but under differing conditions) and fabrication of gate electrodes is a challenge.

Keywords: Molecular wire, interconnect, resonant tunneling, molecular electronics

1. INTRODUCTION

Early work in molecular electronics was largely speculative, because the appropriate fabrication and characterization methods were not available on the nanometer distance scale characteristic of molecular structures [1–7]. Feynman's remarks concerning fabrication from the bottom up (by molecular or atomic assembly) rather than from the top down (by masking/lithograph techniques) is the intellectual antecedent for much of this work [8].

The advent of molecular surface functionalization and self-assembly methods [9], coupled with synthetic organic and inorganic chemical methods for producing molecular wire structures reproducibly and in good yield [10, 11], has improved candidate structures for molecular wires. Development of both scanning probes (especially STM) and nanofabricated measurement structures such as mechanically controllable break junctions [12], nanopore enclosures [13] and assembled nanodot arrays [14] has permitted structural investigation

* Corresponding author. Tel.:847/491-5652, fax: 847/491-7713, e-mail: ratner@chem.nwu.edu.

of individual molecules or assemblies at electrode interfaces, and deduction of conductances for single molecules [1, 12, 15, 16].

Within the past 18 months, individual molecules have been shown to exhibit resonant tunneling behavior [17], coulomb blockade and staircase behavior at room temperature [12], gate control of electronic currents [17, 18], long distance (micron) current carrying capability [17, 18], and rectification properties [19]. Coupled with related work on inelastic electron tunneling spectroscopy [20], on Langmuir-Blodgett film conductivity [21], and on the much larger area of conductive organic materials [22], this recent work constitutes an impressive knowledge base suggesting that the use of molecular wires as interconnects is practical and promising.

Important advantages of molecular wires include:

1. Size (characteristic radii of molecular wires are subnanometer, while length can vary from subnanometer to micron).
2. Preparation: synthetic chemistry allows a myriad of candidate structures, with differing rigidity, optical properties, band gaps, electronic couplings, vibronic couplings, chemical reactivity, optical behavior, etc. [11, 22].
3. Low Cost: most polymer prices are quoted per pound or per ton.
4. Tailorability and functionalization possibility: molecular wire structures can be varied using synthetic chemistry; attachment to electrodes can vary in bond strength and electronic coupling [11, 22].

While much of the interest in molecular electronics has been in device structures [1–7] (molecular rectifiers [19, 23, 24], molecular logic gates and switches [25], molecular two-state devices or *n*-state devices, molecular recognition sensors [26], molecular optoelectronic gates), it is probably the interconnect structure of molecules that is most promising for hybrid molecular/solid state electronics.

2. MOLECULAR WIRE CONDUCTANCE

Electron transfer through molecular wires is closely related to intramolecular electron transfer, in which a molecular bridge is used to facilitate transfer of a single electron between terminal donor and acceptor molecular species [27, 28]. The difference between these two situations is that the continuum, that causes linear transport (rate constant in the intramolecular electron transfer situation or conductance in the molecular wire) will differ: for intramolecular electron transfer, the continuum is vibronic, while for molecular wires, the continuum is the electronic energy level continuum of the electrodes [27].

A characteristic molecular wire structure is shown in Figure 1A, in which a thiol terminated aromatic molecular bridge is linked to gold

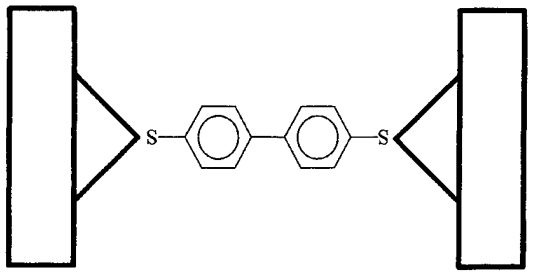


FIGURE 1A Schematic of a molecular wire group bridge between two continuum metal electrodes.

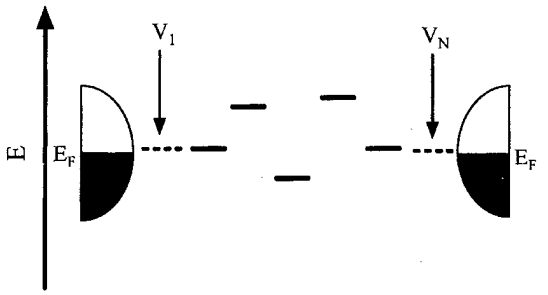


FIGURE 1B Energy level schematic diagram of a molecular wire structure like that in Figure 1a. The hemispherical structures represent the energy bands of the metal; E_F is the Fermi energy. The discrete energy levels are those of the site orbitals in the molecule, while V_1 and V_N are the matrix elements coupling the terminal sites in the molecular wire to the metallic interface.

electrodes; in the sketch, the electrode link is indicated as a single interaction, but its actual nature will vary depending on the chemical binding or (in less favorable cases) physisorption interaction between bridge and electrode. The model that we will use for computing the conductance is sketched in Figure 1B. This is an energy-level diagram showing the continuum electrode levels, the fermi energy, the discrete energy levels corresponding to the atomic sites in the molecular wire, and the characteristic tunneling strengths (electronic mixing matrix elements) V_1 and V_N at the first and Nth sites of the wire.

Several formalisms for calculating the conductance are available. Lang has developed [29] the density functional approach using a jellium model, and has demonstrated its ability to deal with atomic wire structures. Joachim and collaborators [30] have used an approach in which the electrode is represented as a tight binding, finite metal block. Datta and collaborators [31], and our group, have utilized a formalism in which the interaction between the discrete states of the molecular wire and the continuum states of the electrode is represented by a Newns/Anderson formalism [32]. Details are given elsewhere [33], but the important point is that, due to the bridge, the first and last sites in the molecule acquire a self-energy, denoted Σ. That is,

$$\begin{aligned} H_{11} &\rightarrow H_{11} + \sum_1(E) \\ H_{NN} &\rightarrow H_{NN} + \sum_N(E) \end{aligned} \tag{1}$$

Molecular site energies are unchanged, except for the first and last sites, each of which acquires a real part (energy shift) and an imaginary part (lifetime). Equation 2 indicates the two parts of the self-energy.

$$\sum_k(E) = \wedge_k(E) - i\Delta_k(E) \qquad K = 1, N. \tag{2}$$

Using the results of this discrete/continuum interaction representation, and the Bardeen tunneling formula [34], for conductance, one can derive formulas for the conductance g or current I both in the linear regime (current assumed proportional to voltage) and in the more general situation. These results are given by, respectively,

$$g = \frac{2e^2}{\pi\hbar}\Delta_1(E_F)\Delta_N(E_F)|G_{IN}(E_F)|^2 \tag{3}$$

$$I = \frac{2e}{\pi\hbar}\int_{E_F - eW}^{E_F} dE\Delta_1(E)\Delta_N(E + eW)|G_{IN}(E, W)|^2 \tag{4}$$

Here E_f is the electrode Fermi level (injection level), W is the applied voltage, 4γ is the bandwidth, Δ_N and Δ_1 are the imaginary parts of the self-energy from Equation 2, given by

$$\Delta_k(E) = V_k^2/\gamma\begin{cases} \sqrt{1 - E^2/4\gamma^2} & |E/2\gamma| < 1 \\ 0 & |E/2\gamma| > 1 \end{cases} \tag{5}$$

Finally, G_{1N} is the $(1,N)$ matrix element of the electronic Green's function, described by

$$G^{-1}(E) = E + i\eta + \sum(E) - H; \tag{6}$$

Here H is the molecular electronic Hamiltonian, and η is a positive infinitesimal.

The physical interpretation of Equation 3 is straightforward: the wire conductance is given by the atomic unit of conductance times the electrode interaction at the first and N^{th} sites of the bridge, times the Green's function term that describes the electronic mixing between the orbitals at the two terminal sites of the molecular wire. Equation 4 generalizes this to the dependence upon applied voltage, and is applicable to the situation in which the source/drain current/voltage curve is not linear. This equation, thus, permits understanding of staircase behavior [33]. That requires, though, proper solution to the Poisson equation that describes the potential dependence of the molecular site energetics. This problem has not really been addressed yet in the published literature.

Although descriptions of the staircase behavior using Equation 4 have been given [33] we limit ourselves here to the linear conductance behavior, as the simplest characterization of the interconnect ability of these molecular wires.

3. STRUCTURAL AND ENERGETIC DEPENDENCE OF WIRE CONDUCTANCE

Molecular wires will exhibit substantial sensitivity to the details of the molecular electronic structure. Understanding the dependence of conductance on laboratory parameters such as construction of the wire itself, temperature, chemical linkage between wire and electrode and injection energy requires a combination of chemical concepts for the wire and continuum concepts for the interface.

3.1. Conductance Variation with Bridge Chemical Structure and Injection Energy

Three continuous variables characterize the conductance of the molecular wire interconnect: the injection or Fermi energy, the temperature, and the potential difference between the two electrodes; for a linear conductance as given by Equation 3, the latter variable can be ignored. Temperature dependence is a complicated problem; one expects statistical factors in the electrodes and various shape modification of the barrier by the applied field (essentially the Fowler-Nordheim treatment) [35] and temperature dependent behaviors in the wire itself, including possible electron trapping behavior, dephasing, and structural changes. For simplicity, we neglect temperature effects (limit of low temperatures).

The only important continuous variable is then the injection energy. Figure 2 shows the predicted energy dependence for a fully conjugated (assumed planar) diphenyl bisthiol structure with gold electrodes. The notations σ and π refer to the angular momentum symmetries. The effective injection has both σ and π in planar geometries these are good symmetry quantum numbers, so that the channels can be separated. Several interesting aspects of Figure 2 can be noted: there is resonant tunneling behavior characterized by negative differential conductance – that is, as the

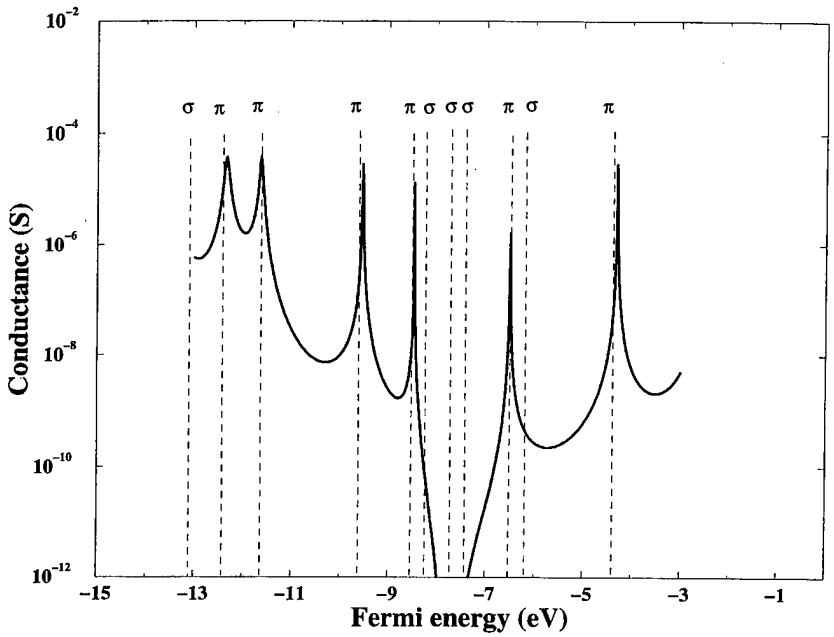


FIGURE 2 Conductance of the molecular structure in Figure 1A, calculated using the linear approximation formula of Equation 3. Note the resonances (--- lines represent molecular orbital eigenvalues). Resonant tunneling occurs whenever the π level is degenerate with the Fermi energy. Conductance maxima are very close to the atomic unit of conductance, corresponding to efficient resonant tunneling.

injection energy passes through one of the resonances corresponding to a molecular orbital of the bridging structure, the conductance exhibits a local maximum, subsequently decreasing with increased injection energy. The self-energy terms of Equation 2 become important here: divergence of the conductance is avoided because of the continuum structures of the electrodes, that provide imaginary (damping) behavior at the terminal sites of the molecular bridge. The finite conductance will be reduced compared to the atomic conductance $g_0 = (12.9\,K\Omega)^{-1}$, by effects of molecular mixing, of the interfacial covalent link, and of vibrational dissipation that are not included in this discussion [40–42]. In general, the π-type contributions are larger than σ: this is in accordance with a huge amount [36] of organic chemical wisdom, that delocalization through the π system is much more effective than σ.

For actual conductance computation, it is important to know the injection energy. This problem is not yet solved, although it seems reasonable to assume [31, 33] that the injection energy will be fixed at mid-gap. This assumption has been made in several computations, and agreement with experiment is generally very good. Actual proof requires self-consistent solution to the Poisson equation that describes actual electron density/potential structures in the inhomogeneous interconnect situation.

Since π-delocalization is generally stronger than σ, one might expect that twisting of the molecular structure to impede the π conjugation would decrease conductance. This is shown in Figure 3 for triple bonded and double bonded systems respectively, and the differences are striking. The angular dependence is much smaller in the triple bond case, essentially because the triple bond is cylindrically symmetric. Conversely, the double bond exhibits mirror symmetry in the plane, so that twisting by $\pi/2$ around it results in localization and strongly reduced conductance. The actual

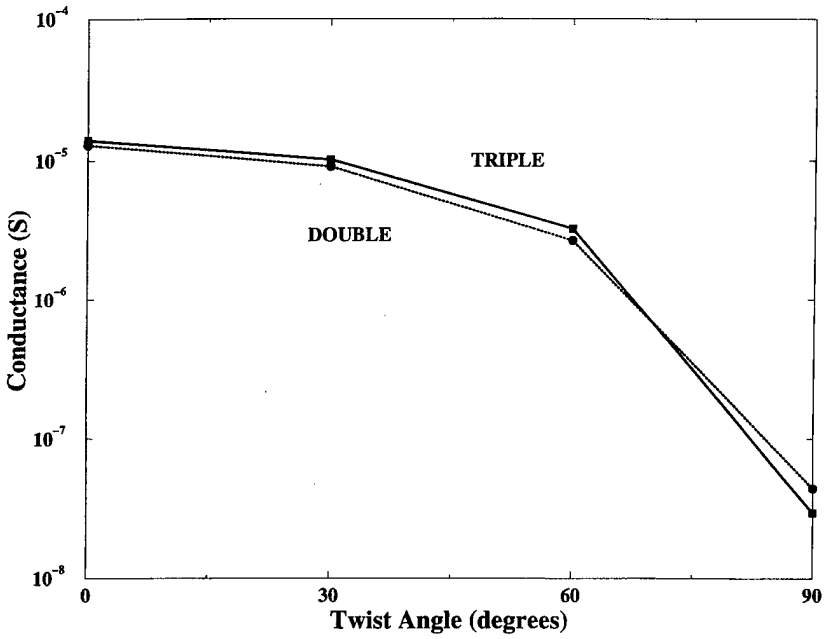


FIGURE 3 Computed conductance, as a function of twist angle, for molecules of the same structure as in Figure 1a, but with double and triple bonds inserted between the rings. When the π systems become orthogonal (90 degree twist angle), the conductance is substantially reduced. The functional dependences on the double or triple bonds differ, because of the different symmetries involved.

conductance of the double bonded structures are larger than that of the triple bonded bridges, because of bond alternation effects that result in stronger electron localization for the triple bonded (alkyne) than the double bonded (alkene) structures. Alkyne linkages are of interest in connection with molecular photonic wires [37] and nonlinear optical devices [38].

3.2. Length Dependence – Tunneling and Ohmic Behaviors

Molecular wires generally are not ohmic, except under special conditions. For an ohmic linear wire, one expects the conductance to scale as the inverse wire length. Exponentially-decaying tunneling has been well documented in intramolecular electron transfer reactions [28, 29], and is expected on a series of theoretical grounds to characterize molecular wires in the absence of dissipation, dephasing, trapping, and defect formation effects. Accordingly, Figure 4A shows the predicted conductance with thiol (sulfur) first and last chain sites and the indicated bridging units. Several characteristic behaviors expected on the basis of organic chemistry are apparent here: the π systems are better conductors than the σ systems, bond-length alternation effects make the alkynes less good than the alkenes, changing local bond geometries and so-called cis and trans linkages change the conductance of the polyalkyne structures.

Decay is exponential with length in all cases. The reason is that the electron is never actually localized on the molecular bridge, but rather undergoes bridge-assisted coherent tunneling between electrodes. This arises because in our formal approach, dissipation mechanisms, that would allow the electron actually to trap on the bridge, are not included.

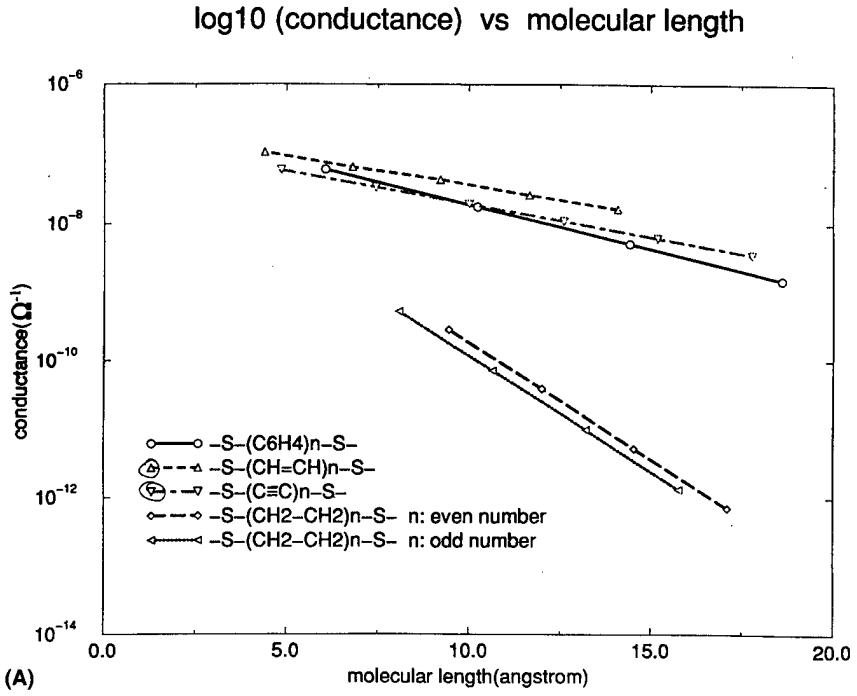


FIGURE 4A Logarithm of the conductance versus molecular length, for typical double bonded and triple bonded extended systems (such as those in Figure 1A, but with double bond or triple bond chains introduced between the rings). Notice the characteristic exponential decay with distance, and the different decays of the triple and double bonded systems.

Inclusion of dissipation has been discussed for intramolecular electron transfer [40–42] (the rate constant for intramolecular electron transfer is expected to scale generally as the conductance for molecular wires, although inclusion of dissipation in real molecular wire calculations has not yet been presented). Figure 4B presents the length dependence of the intramolecular transfer rate if a tight binding model is assumed, and the matrix element V_b characterizes the mixing between the sites of the molecular bridge of Figure 1. The gap, ω, is the energy required to place an electron in an affinity level (that is, to reduce chemically) of the molecular bridge. For fixed mixing matrix element $V_1 = V_N = V_b$, there is substantial energy variation of the rate constant. Indeed, when the gap becomes comparable to the mixing matrix element, there is essentially no decay of the rate constant with the length of the bridge (since the continuum is included, resonance tunneling does not result in a divergence of the rate constant). Physically, dissipation mechanisms will always occur, so that no true length independence can occur. In weak dissipation, however, the length decay may be very small, as has recently been observed [17, 18] in very low temperature measurements on carbon nanotube structures.

The energy gap ω of Figure 4B is related to the injection energy of Figure 2; vanishing ω corresponds to injection exactly at one of the resonances in Figure 2.

Functional dependence of the rate constant (conductance) can be formulated simply if V_b/ω is much less than one, and the tight binding model

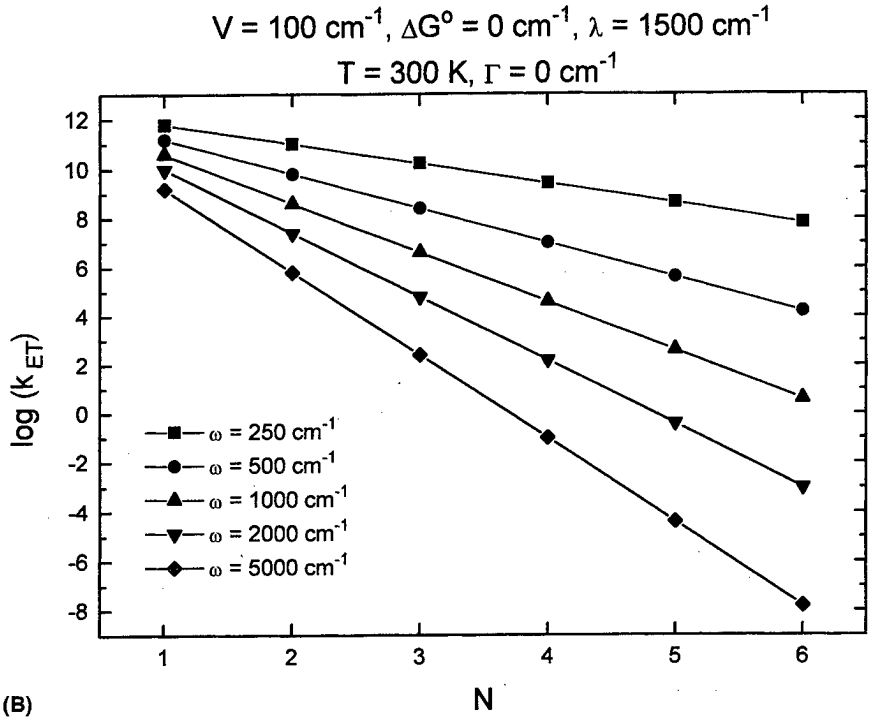


FIGURE 4B The calculated intramolecular electron transfer rate (units of s^{-1}). The parameters are those of a simple tight binding model, as in Equation 7; V is the matrix element between neighboring sites, λ the reorganization energy, and ω the gap between donor or acceptor levels and bridge levels. Note that the characteristic decay parameter (the β of Equation 9) indeed strongly depends on the gap size.

(to chemists, the Huckel model) is assumed for the molecular bridge. Under these conditions, the Green's function will scale like [42] ($\hbar \equiv 1$)

$$G_{IN} \sim \frac{V_1 V_N}{\omega} (V_B/\omega)^{N-1} \tag{7}$$

for an N-site bridge, and then the conductance behavior (or rate constant) is expected to follow Equation 8.

$$g(N) = g(1) \exp\{-\beta(N-1)R_0\} \tag{8}$$

This exponential decay is then characterized by the constant β, which in the simplest (super exchange) model is given as

$$\beta = \frac{-2}{R_0} \ln(V_B/\omega) \tag{9}$$

Here R_0 is the repeat unit length on the molecular bridge. Exponential behavior has been very well studied for intramolecular electron transfer rates [28, 43].

4. REMARKS

Molecules are of substantial interest as possible device components in advanced electronics and optoelectronics; their most obvious and promising short-term application is in the area of interconnects. Electron transfer can, then, be used directly to pass current among active device components (or, more speculatively, to provide field switching possibilities) [46].

We have analyzed several of the characteristic behaviors to be expected for single molecule interconnects. Even in relatively unfavorable cases, single molecules can carry a current of order 0.1 nA for applied fields of one volt; [14, 15], this is large enough for applications, and can be increased in more favorable cases such as single-walled nanotubes [18] or strongly conjugated oligomers.

There are, essentially, two regimes in the conductance of characteristic molecular wires. At low temperatures, resonant tunneling structures are expected, with substantial increases in the conductance (approaching the atomic unit of conductance), near molecular resonances, suggesting possible resonant tunneling applications. With increasing temperature, dephasing processes will become important and one will eventually find ohmic behavior, with the conductance scaling inversely with length [42]. Because molecules are substantially smaller than the usual quantum dot structures, coulomb blockade and staircase effects can be seen [12] not only at low temperatures but also at room temperature, suggesting the possibility of single electronics under ambient conditions.

The work on molecular wires to date is at the proof of concepts stage: problems such as the relative resistance of particular molecular wire structures [30, 31], the specific attachment of molecular wires to electrode surfaces, understanding the temperature and length dependence of the conductance [33], and engineering the geometry of particular molecular wire interconnects [11] have been discussed. It is clear that use of molecular wires as circuit elements still requires overcoming several major obstacles, of which the most serious is probably addressing individual molecular wires, and anchoring them in a specific geometry on assembled chip structures.

Acknowledgements

We are grateful to the **NSF** and **CONICIT** for support through a collaborative **NSF/CONICIT** international research program. **MR** is also grateful to the chemistry divisions of the National Science Foundation and the Office of Naval Research for support. We are indebted to Mark Reed, Chongwu Zhou, Abe Nitzan, Bill Davis, Yi Mao and Mike Wasielewski for very helpful suggestions and enjoyable collaboration.

References

[1] F. L. Carter, ed. (1982). *Molecular Electronic Devices*, New York: Dekker.
[2] F. L. Carter, R. K. Siatkowski and H. Wohltgen, eds. (1988). *Molecular Electronic Devices*, Amsterdam: Elsevier.

[3] R. R. Birge, ed. *Molecular and Biomolecular Electronics*, (ACS, Washington, DC, 1994).
[4] R. M. Metzger, P. Day and C. G. Papavassiliou, eds., (1991). *Lower-Dimensional Systems and Molecular Electronics*, New York: Plenum.
[5] A. Aviram, ed. (1992). *Molecular Electronics–Science and Technology*, Washington, DC: AIP.
[6] J. Jortner and M. A. Ratner, Eds., *Molecular Electronics* (Blackwells, London, 1997).
[7] M. C. Petty, M. R. Bryce and D. Bloor, ed. *Introduction to Molecular Electronics*, (Oxford, New York, 1995).
[8] Feynman, R. P. (1960). *Eng. and Science*, **23**, 22 ; in H. D. Gilbert, Ed., *Miniaturization* (Reinhold, New York, 1961).
[9] Mirkin, C. A. and Ratner, M. A. (1992). *Ann. Rev. Phys. Chem.*, **43**, 719.
[10] Wu, C. G. and Bein, T. R. (1994). *Science*, **264**, 157.
[11] Tour, J. M. (1996). *Chem. Revs.*, **96**, 537.
[12] Zhou, C., Muller, C. J., Reed, M. A., Burgin, T. P. and Tour, J. M., in ref. 6; Zhou, C., Reed, M. A., Muller, J., Burgin J. P. and Tour, J. M. submitted to *Science*.
[13] Zhou, C. and Reed, M. A., unpublished.
[14] Dorogi, M., Gomez, J., Osifchin, R., Andres, R. P. and Reifenberger, R. (1995). *Phys. Rev. B.*, **52**, 9071; Andres, R. P., Bein, T. and Dorogi M. *et al.* (1996) *Science*, **272**, 1323.
[15] Joachim, C., Gimzewski, J. K., Schlitter, R. D. and Chavy, C. (1995). *Phys. Rev. Lett.*, **74**, 2102; Joachim, C. and Gimzewski, J. K. (1995). *Europhys. Lett.*, **30**, 409.
[16] Tao, N. J. (1996). *Phys. Rev. Lett.*, **76**, 4066.
[17] Bockrath, D. M., Cobden, D. H., McEuen, R. L., Chopra, N. G., Zettl, A., Thess, A. and Smalley, R. A. (1997). *Science*, **275**, 1922 .
[18] Tans, S. J., Devoret, M. H., Dai, H., Thess, A., Smalley, R. J., Geerligs, L. J. and Dekker, C. (1997). *Nature*, **386**, 474.
[19] Dhirani, A., Lin, P. H., Guyot-Sionnest, P., Zehner, R. W. and Sita, L. A. (1997). *J. Chem. Phys.*, **106**, 5249 .
[20] Hipps, K. W. and Mazur, U. (1994). *J. Phys. Chem.*, **98**, 5824; Mazur, U. and Hipps, K. W. (1994). *J. Phys. Chem.*, **98**, 8169.
[21] Fischer, C. M., Burghard, M., Roth, S. and Klitzing, K. V. (1994). *Europhys. Lett.*, **28**, 129; Fischer, C. M., Burghard, M. and Roth, S., in Ref. 6.
[22] W. R. Salaneck, I. Lundstrom and B. Ranby, ed. (1991). *Conjugated Polymers and Related Compounds*, London: Oxford University Press, 1993; J. L. Bredas and R. Silbey, eds. *Conjugated Polymers.*, Dordrecht: Kluwer, 1991.
[23] Aviram, A. and Ratner, M. A. (1974). *Chem. Phys. Lett.*, **29**, 277.
[24] Metzger, R. M., in *Molecular and Biomolecular Electronics*, edited by R. R. Birge (ACS, Washington, DC, 1994), p. 81 a very complete overview.
[25] Walz, J., Ulrich, K., Port, H., Wolf, H. C., Wonner, J. and Effenberger, F. (1993). *Chem. Phys. Lett.*, **213**, 322; K. Schaumburg in W. Gopel and C. Ziegler, eds. (1992). *Nanostructures Based on Molecular Materials*, Weinheim: VCH, 156.
[26] Mirkin, C. A., Ulrich, K., Mucic, R. C. and Storhoff, J. J. (1996). *Nature*, **382**, 607; Alivisatos, A. P., Johnson, K. P., Peng, X., Wilson, P. E., Loweth, C. J., Bruchez, M. P. and Schultz, P. G. (1996). *Nature*, **382**, 609.
[27] Ratner, M. A. and Jortner, J., in Ref.6.
[28] Marcus, R. A. and Sutin, N. (1985). *Biochim. Biophys. Acta.*, **811**, 265; Newton, M. D. and Sutin, N. (1984). *Annu. Rev. Phys. Chem.*, **35**, 437; Barbara, P. F., Meyer, T. J. and Ratner, M. A. (1996). *J. Phys. Chem.*, **100**, 13148.
[29] Lang, N. D. (1996). *Phys. Rev. B.*, **52**, 5335.
[30] Joachim, C. and Vinuesa, J. F. (1996). *Europhys. Lett.*, **33**, 635.
[31] Datta, S. (1995). *Electronic Transport in Mesoscopic Systems*, Cambridge: Cambridge University Press; Samanta, M. P., Tian, W., Datta, S., Henderson, J. I. and Kubiak, C. P. (1996). *Phys. Rev. B.*, **53**, 7626.
[32] Newns, D. M. (1969). *Phys. Rev.*, **178**, 1123; Anderson, P. W. (1961). *Phys. Rev.*, **124**, 41.
[33] Mujica, V., Kemp, M. and Ratner, M. A. (1994). *J. Chem. Phys.*, **101**, 6849; Mujica, V., Kemp, M. and Ratner, M. A. (1994). *J. Chem. Phys.*, **101**, 6856; Kemp, M., Mujica, V. and Ratner, M. A. (1993). *J. Chem. Phys.*, **101**, 5172; Mujica, V., Kemp, M., Roitberg, A. and Ratner, M. A. *Condensed Matther Theories* II, edited by E. V. Luduena, P. Vashishta and R. F. Bishop (Nova Science, Commack, 1996) (in press); Mujica, V., Kemp, M., Roitberg, A. and Ratner, M. (1996). *J. Chem. Phys.*, **104**, 7296.
[34] Landauer, R. (1957). *IBM J. Res. Dev.*, **1**, 223; Landauer, R. (1981). *Phys. Lett.*, **85a**, 91.
[35] Zhou, C. and Reed, M. A., Private Communication.
[36] Pope, M. and Swenberg, C. E., *Electronic Processes in Organic Crystals* (Oxford, 1982).
[37] Lindsey, J. M. (1996). *J. Am. Chem. Soc.*, **11**, 3996.
[38] Priyadarshy, S., Therien, M. J. and Beratan, D. N. (1996). *J. Am. Chem. Soc.*, **8**, 1504.
[39] Cave, R. and Newton, M. D., in ref. 6.
[40] Skourtis, S. and Mukamel, S. (1995). *Chem. Phys.*, **197**, 367.
[41] Felts, A. K., Pollard, W. T. and Friesner, R. A. (1995). *J. Phys. Chem.*, **99**, 2929.
[42] Davis, W., Wasielewski, M., Mujica, V., Nitzan, A. and Ratner, M. A. *J. Phys. Chem.*, in press.
[43] Risser, S., Beratan, D. N. and Onuchic, J. N., in Ref. 6.

Authors' Biographies

Mathieu Kemp was born in Montreal, Canada in 1963. He received his Ph.D. in Physics from the University of North Carolina in 1992. His research interests include: molecular electronics, nanoelectronics, and electronic transport in amorphous semiconductors. He is working with Mark Ratner at Northwestern University.

Adrian Roitberg was born in Buenos Aires, Argentina in 1962. He received his Ph.D. in Chemistry form the University of Illinois at Chicago in 1992. His research interests include: protein modeling, electron transfer, and molecular electronics. He is a guest research chemist at the National Institute for Standards and Technology.

Vladimiro Mujica was born in Caracas, Venezuela in 1954. He received his Ph.D. in Quantum Chemistry form Uppsala University in 1985. His

research interests include: mesoscopic devices, nonlinear optics and density functional theory. He is a professor at Universidad Central de Venezuela.

Mark A. Ratner was born in Cleveland, Ohio in 1942. He received his Ph.D. in Chemistry from Northwestern University in 1969. His research interests include: nonlinear optics, electron transfer, molecular electronics, protein modelling, polymer electrolyte transport, and SCF methods. He is a professor at Northwestern University.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 75–78
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Ionic Channels in Biological Membranes: Natural Nanotubes Described by the Drift-Diffusion Equations

BOB EISENBERG*

Dept. of Molecular Biophysics and Physiology, Rush Medical Center, Chicago IL 60612

An important class of biological molecules–proteins called ionic channels–conduct ions (like Na^+, K^+, Cl^-) through a narrow tunnel of fixed charge ('doping'). Ionic channels are the main pathway by which substances move into cells and so are of great biological and medical importance: a substantial fraction of all drugs used by physicians act on channels. Channels can be studied in the tradition of computational electronics. Drift diffusion equations form an adequate model of IV relations of 6 different channel proteins in ~10 solutions over ±150 mV. Ionic channels can also be studied with the powerful techniques of molecular biology. Atoms can be modified one at a time and the location of every atom can be determined. Ionic channels are natural nanotubes that can be controlled more precisely and easily than physical nanostructures but biologists need help if realistic simulations are to be done atomic detail.

Keywords: Biological membranes, ionic channels, drift diffusion equations, Gummel iteration

Drift-diffusion equations, combined with Poisson's equation (*PDD*), are widely used in physical sciences to describe the flux of charge carriers through systems containing fixed charge (doping) [28]. An important class of biological molecules–proteins called ionic channels–conduct ions (like Na^+, K^+, Cl^-), and thus current, through a narrow tunnel of fixed charge ('doping') formed by the polar residues of the protein [36] although electron flow plays no direct role in their conduction of current. These proteins can be studied with the full power of molecular biology [1]; for example, they can be modified one atom at a time with the techniques of molecular genetics. Thus, these natural nanotubes are a natural 'hole in the wall' that can be controlled more precisely and easily than many physical nanostructures.

Ionic channels open and close ('gate') to give currents that are a random telegraph signal [34]. The properties of gating are complex and the structure(s) and mechanism(s) that produce gating are not known [23], but the flow of ions through open channels is much simpler, and obeys the *PDD* equations, as we shall see [10–13, 19, 37].

Channels are the main pathway by which substances move in and out of cells and so are of

* Invited by and submitted to VLSI Design (Proceedings of the Fifth International Workshop on Computational Electronics).

great biological and medical importance: they are responsible for signaling in the nervous system; for coordination of muscle contraction – including the coordination of cardiac muscle that allows the heart to function as a pump – and they are involved in transport in every cell and organ, for example, in the kidney, intestine and endocrine glands [1, 36]. A substantial fraction of all drugs used by physicians act directly or indirectly on channels [35].

Channels are studied one molecule at a time in hundreds, if not thousands of laboratories every day [14–16, 31], using Neher & Sakmann's patch clamp method [6, 32, 34] (for which they received the Nobel Prize). The concentrations of ions outside channels (that carry current through the channel) can be directly controlled and the shape of current voltage (IV) relations can be manipulated. In this way, a wide range of IV behavior can be measured from the single doping profile of one type of channel and so much can be inferred about the doping profile from IV measurements (if they are taken in many different (pairs of) concentrations of current carriers). For all these reasons, channels are a popular object for experimentation: thousands (!) of abstracts describing their properties are presented each year at the annual meeting of the Biophysical Society and hundreds of papers are published about them, chiefly in the Biophysical Journal and the Journal of Physiology (London).

Channels are also an appealing and important object for theoretical analysis and numerical simulation. Open channels are probably the simplest protein structures of general biological importance. Unlike many other subjects of biophysical investigation, ionic channels are a general biological system with importance for every organ, tissue, and cell in an animal and plant. Indeed, they are probably just as important for subcellular organelles. Ionic channels are well defined biological systems that can be investigated both with the techniques of molecular biology and of biophysical chemistry, helped substantially by the techniques and insights of semiconductor physics, I believe.

Ionic movement plays an important role in the function of all proteins – e.g., enzymes – and so a model that describes ionic movement in channels is likely to give important insight into protein function in general. Indeed, the closely related Poisson-Boltzmann theory [18, 24] has been of considerable help already, even though it is a strictly equilibrium theory that does not permit flux at any time or location.

Theories of physical chemistry [4] and electrochemistry [30] certainly should be able to predict the movement of ions through a tunnel of fixed charge – a hole in the wall – on the biological time scale of $100\,\mu\text{sec} - 10\,\text{sec}$, but they need to be supplemented by the theories and simulations of carrier transport in general, e.g., in semiconductors [28]. The physical chemical tradition has not often dealt with flux [2, 3, 5, 8, 9, 21, 22, 25, 29] in such a system of fixed charge, whereas the flux of charge carriers has been the main subject of semiconductor physics and computational electronics for many years, if not decades.

Five laboratories have shown that the *PDD* of semiconductor physics form an adequate model of IV relations of 6 different channel proteins in ~ 10 pairs of solutions (containing different concentrations of the charge carriers Na^+, K^+, Cl^-) in the range ± 150 mV [10–13, 37]. The IV relations are *qualitatively* different in different types of channels – some are linear, some sublinear and some superlinear – because different channel proteins have qualitatively different profiles of fixed charge arising from their different sequences of amino acids.

The structures of two of these proteins (porin and its mutant G-119D) are known from the standard methods of molecular biology: the location of every atom has been determined by X-ray diffraction [17, 26, 27, 33, 38] with an accuracy of ~ 0.1 Å. The mutant has one extra negative charge. Measurements of IV relations from a single molecule of porin [37] allow the *PDD* model to estimate the additional charge as $-0.97\,e$, although this estimate will undoubtedly change as more work is done. (I hasten to add that no information

about the proteins is used in the analysis except the length and diameter of the channel; parameters were not adjusted in any way.)

We conclude that the *PDD* equations seem to be an adequate model open channels. It is surprising that the *PDD* equations work as well as they do, given their evident inadequacies. I imagine they work this well because the fixed charge density of channels is large ($\sim 3\times 10^{21}\,\mathrm{cm}^{-3}$) compared to the concentration of ions outside the channel (2×10^{19} to $1\times 10^{21}\,\mathrm{cm}^{-3}$); because the biological range of voltages is quite limited (± 200 mV); and because the *PDD* model uses effective parameters. Eisenberg, Chen, and Schuss have recently shown how the *PDD* equations can be derived in single file systems like channels that conduct one ion at a time. The *PDD* equations, or equations quite like them, describe the mean properties of ensembles of Langevin equations, each of which specifies the motion of a single ion (of a particular type moving from a given side of the channel), each of which is coupled to its own reaction field described by a Poisson equation and boundary conditions.

The *PDD* equations are just a first, low resolution description of open channels. More realistic models (using Monte Carlo simulations called molecular dynamics in the world of proteins [7, 8, 20]) are needed to provide insight with atomic resolution. It is likely that many critical functions of channels and enzymes will be best understood this way – by atomic resolution simulations that include flux-but biologists cannot do the simulations themselves: their simulations of atomic resolution have been confined strictly to equilibrium as have those of most chemists. Much help is needed from the community of computational electronics if flux, and electrical potentials at the electrodes (i.e., boundaries) of the system, are to be included in chemical simulations of atomic detail.

Ionic channels are so important biologically, but so well defined physically, that they are an ideal object of biophysical investigation. The techniques of computational electronics and molecular biology can be joined together to determine how ionic channels work. Perhaps the same will prove true of many other chemical and biological systems, but it is wise to try the simple ones first. Nothing is likely to be simpler physically than a hole in the wall.

Acknowledgment

The adventure of the open channel has been shared Duan Chen every step of the way. I hope he has enjoyed our work together as much as I have. We are both grateful to Mark Lundstrom and Joe Jerome who introduced us to the community of computational electronics.

References

[1] Alberts, B. *et al.* (1994). *Molecular Biology of the Cell* Third ed. New York, Garland.

[2] Allen, M. P. and Tildesley, D. J. (1987). Computer Simulation of Liquids, New York, Oxford.

[3] Allen, M. P. and Tildesley, D. J. (1990). Computer Simulation in Chemical Physics, Series C: *Mathematical and Physical Sciences*, **397**, Boston, Kluwer.

[4] Berry, S. R., Rice, S. A. and Ross, J. (1980). Physical Chemistry, New York, John Wiley & Sons.

[5] Binder, K. (1995). The Monte Carlo Method in Condensed Matter Physics Second ed. *Topics in Applied Physics*, **71**, New York, Springer.

[6] Boulton, A. A., Baker, G. B. and Walz, W. (1995). *Patch Clamp Applications and Protocols*, Totowa, NJ, Humana Press.

[7] Brooks, B. R. *et al.* (1983). CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.*, **4**, 187–217.

[8] Brooks, C. L., Karplus, M. and Pettitt, B. M. (1988). *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, New York: John Wiley & Sons.

[9] Catlow, F. R. A., Parker, S. C. and Allen, M. P. (1990). Computer Modelling of Fluids, Polymers, and Solids. Series C: *Mathematical and Physical Sciences*, **293**, Boston: Kluwer.

[10] Chen, D. *et al.* (1995). Hydrodynamic model of temperature change in open ionic channels, *Biophysical J.*, **69**, 2304–2322.

[11] Chen, D. *et al.* (1997). Permeation through the calcium release channel (CRC) of cardiac muscle, *Biophys. J.*, **72**, A108.

[12] Chen, D. P., Lear, J. and Eisenberg, R. S. (1997). Permeation through an open channel. Poisson-Nernst-Planck theory of a synthetic ionic channel, *Biophys. J.*, **72**, 97–116.

[13] Chen, D. P., Nonner, W. and Eisenberg, R. S. (1995). PNP theory fits current-voltage (IV) relations of a neuronal anion channel in 13 solutions, *Biophys. J.*, **68**, A370.

[14] Conley, E. C. (1996). *The Ion Channel Facts Book. I. Extracellular Ligand-gated Channels*, **1**, New York, Academic Press.

[15] Conley, E. C. (1996). *The Ion Channel Facts Book. II. Intracellular Ligand-gated Channels*, **2**, New York, Academic Press.

[16] Conley, E. C. (1997). *The Ion Channel Facts Book. III. Inward Rectifier & Intercellular Channels*, **3**, New York, Academic Press.
[17] Cowan, S. W. *et al.* (1992). Crystal structures explain functional properties of two E coli porins, *Nature*, **358**, 727–733.
[18] Davis, M. E. and McCammon, J. A. (1990). Electrostatics in biomolecular structure and dynamics, *Chem. Rev.*, **90**, 509–521.
[19] Eisenberg, R. S. (1996). Computing the field in proteins and channels, *J. Membrane Biol.*, **150**, 1–25.
[20] Elber, R. *et al.* (1993). MOIL: A molecular dynamics program with emphasis on conformational searches and reaction path calculations., in *Statistical Mechanics, Protein Structure and Protein-Substrate Interactions*, Plenum Press, New York.
[21] Evans, D. J. and Morriss, G. P. (1990). *Statistical Mechanics of Nonequilibrium Liquids*, New York, Academic Press.
[22] Haile, J. M. (1992). *Molecular Dynamics Simulation*, New York, John Wiley & Sons.
[23] Hille, B. (1992). *Ionic Channels of Excitable Membranes* 2nd ed. Sunderland, Sinauer Associates Inc.
[24] Honig, B. and Nichols, A. (1995). Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.
[25] Hoover, W. G. (1991). *Computational Statistical Mechanics*, New York, Elsevier.
[26] Jeanteur, D. *et al.* (1994). Structural and functional alterations of a colicin-resistant mutant of ompF porin from Eschericia coli, *Proc. Natl. Acad. Sci. USA*, **91**, 10675–10679.
[27] Lou, K.-L. *et al.* (1996). Structural and functional characterization of ompF porin mutants selected for large pore size. I. Crystallographic analysis, *J. Biol. Chem.*, **271**, 20669–20675.
[28] Lundstrom, M. (1992). *Fundamentals of Carrier Transport*, NY, Addison-Wesley.
[29] Mareschal, M. and Holian, B. L. (1992). *Microscopic Simulations of Complex Hydrodynamic Phenomena*, New York, Plenum Press.
[30] Newman, J. S. (1991). *Electrochemical Systems* 2nd ed. Englewood Cliffs, NJ, Prentice-Hall.
[31] Peracchia, C. (1994). *Handbook of Membrane Channels*, New York, Academic Press.
[32] Rudy, B. and Iverson, L. E. (1992). Ion Channels, *Methods in Enzymology*, **207**, New York, Academic Press.
[33] Saint, N. *et al.* (1996). Structural and functional characterization of ompF porin mutants selected for large pore size. II. Functional characterization, *J. Biol. Chem.*, **271**, 20676–20680.
[34] Sakmann, B. and Neher, E. (1995). *Single Channel Recording*, Second ed. New York, Plenum.
[35] Schultz, S. G. *et al.* (1996). *Molecular Biology of Membrane Disorders*, New York, Plenum.
[36] Stryer, L. (1995). *Biochemistry* Fourth ed. New York, W. H. Freeman.
[37] Tang, J. *et al.* (1997). Permeation through porin and its mutant G119D. *Biophysical Journal*, **72**, A108.
[38] Weiss, M. S. and Schulz, G. E. (1992). Structure of porin refined at 1.8 Å resolution. *J. Mol. Bio.*, **227**, 493–509.

Authors' Biography

Bob Eisenberg was born in Brooklyn, NY, in 1942, and educated at Horace Mann School, then Harvard. He was trained as a biochemist by John Edsall, as a biophysicist by Bernard Katz, as a physiologist by Andrew Huxley, and (from afar) by Alan Hodgkin. Working at UCLA from 1968–1976, he learned some applied mathematics from Julian Cole and Victor Barcilon while studying current flow in cells, chiefly muscle. Moving to Rush Medical College, Chicago, in 1976, he became the Bard Professor and Chairman of their Department of Molecular Biophysics and Physiology. He has been working on self–consistent models of ionic channels for some ten years.

SOFTWARE DEVELOPMENT

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 79–86
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Writing Research Software in a Large Group for the NEMO Project

GERHARD KLIMECK[a,*], DAN BLANKS[a], ROGER LAKE[a], R. CHRIS BOWEN[b], CHENJING L. FERNANDO[b], MANHUA LENG[c], WILLIAM R. FRENSLEY[c], DEJAN JOVANOVIC[a] and PAUL SOTIRELIS[a]

[a] *Applied Research Laboratories, MS 134, Raytheon TI Systems, Dallas, TX 75265;*
[b] *Semiconductor Process and Device Center, MS 3702, Texas Instruments Incorporated, Dallas, TX 75265;*
[c] *School of Engineering, The University of Texas at Dallas, Richardson, TX 75083*

The nanoelectronic modeling (NEMO) program is the result of a three-year development effort involving four universities and the former Corporate Research and Development Laboratory of Texas Instruments, now Applied Research Laboratory, Raytheon TI Systems, to create a comprehensive quantum device modeling tool for layered semiconductor structures. Based on the non-equilibrium Green function formalism, it includes the effects of quantum charging, bandstructure and incoherent scattering from alloy disorder, interface roughness, acoustic phonons, and polar optical phonons. NEMO addresses the diverse needs of two different types of users: (i) the engineer/experimentalist who desires a black-box design tool and (ii) the theorist who is interested in a detailed investigation of the physics. A collection of models trade off physical content with speed and memory requirements. Access to this comprehensive theoretical framework is accommodated by a Graphical User Interface (GUI) that facilitates device prototyping and *in situ* data analysis. We describe a hierarchical software design that allows rapid incorporation of theory enhancements while maintaining a user-friendly GUI, thus satisfying the conflicting criteria of ease of use and ease of development. The theory and GUI modules share data structures that define the device structure, material parameters, and simulation parameters. These data structures may contain general data such as integer and real numbers, option lists, vectors, matrices and the labels for both batch and GUI operation. NEMO generates the corresponding GUI elements at run-time for display and entry of these data structures.

Keywords: NEMO, nanoelectronics-, semiconductor-device-models, quantum-device-simulation, quantum-transport-simulation, Green's-function-methods, quantum-interference-devices, resonant-tunneling-diode, quantum-mechanical-model

* Corresponding author. gekco@rtis.ray.com.

INTRODUCTION

At the inception of NEMO in 1993, quantum transport programs were in their infancy as compared to the physical comprehensiveness of semiclassical Monte Carlo simulators. The existing software only addressed a limited range of quantum devices and quantum transport theory. The NEMO project was conceived to create a single software package that would simultaneously include bandstructure effects, self-consistent charging effects, and incoherent scattering effects, and be flexible enough to model a wide variety of device designs.

Texas Instruments and the collaborating universities evaluated a number of different formalisms (see Fig. 1) and chose the Non-equilibrium Green Function (NEGF) approach as the most general and flexible alternative. The multiple sequential scattering algorithm was incorporated within the NEGF formalism and the other approaches were eventually dropped. The theory [1, 2] implemented in NEMO and simulation results [3–5] generated with NEMO can be found in other publications.

This paper addresses the software engineering aspects of writing NEMO in terms of the basic software design and some of the most important software methods we used to facilitate the tasks of both the programmer and the user. The first section reviews the fundamental program design. This is followed by a discussion of the event-driven methodology used to construct the simulation parameters user interface. The next section reviews the method used to transfer data between NEMO data structures, files, and the graphical user interface. The concluding section summarizes the development environment and software tools used to construct NEMO.

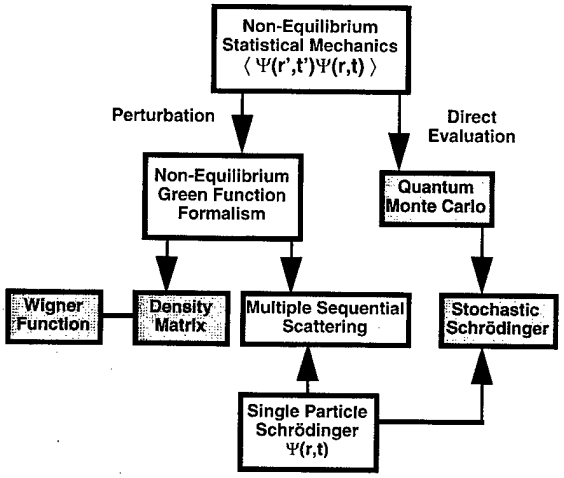


FIGURE 1 Quantum transport theories pursued during the NEMO project. The light gray background indicates the incorporation into the NEGF approach. The dark background indicates approaches that were dropped.

FUNDAMENTAL PROGRAM DESIGN

The basic content of the NEMO software is shown in Figure 2 . At the heart lies the NEGF formalism and open system boundary conditions. The next shell lists the physical models which may be included within several different approximations. The outer shell lists the software algorithms and features that support the interior calculation code.

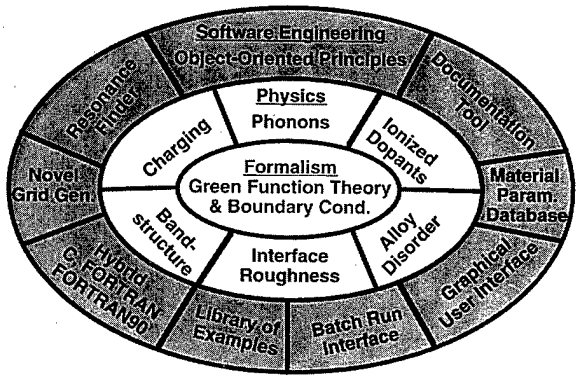


FIGURE 2 Content of the NEMO software. Non-equilibrium Green function theory and novel quantum boundary conditions are at the core. Various physical phenomena can be blended into the theory on a firm footing. Software engineering enables the efficient numerical simulation of the theory and provides user friendly access.

BATCH VS. GUI PROGRAM DESIGN

In its original form, NEMO was a command-line program. This "NEMO-Batch" version was later augmented with a Graphical User Interface (GUI) that provides graphical methods for entry of input parameters, plots calculation results both during and after the calculation, and allows the user to compare data and calculations all in the same program. Unlike NEMO-Batch, which performs the simulation without interruption until completion, the NEMO-GUI program allows the user to start, interrupt, and abort calculations at any time.

Despite the obvious advantages of the GUI interface, NEMO-Batch was still very useful. Since it excludes all GUI-related functionality, NEMO-Batch uses less memory, runs faster, and is easier to port to different computer platforms than NEMO-GUI. Consequently, we devised a method to retain the NEMO-Batch code as the heart of the NEMO-GUI program. To completely understand how NEMO operates, we must examine the design of both the batch and GUI versions of NEMO.

Figure 3(a) illustrates the fundamental operation of NEMO-Batch. The Input Deck is a text file that contains the input parameter required for the calculation. The user calls NEMO-Batch on a command line using the filename of the Input Deck as the sole argument. NEMO-Batch reads and processes the Input Deck, runs the computation, stores the results in output data files, and then exits. A NEMO-Batch run can perform the calculation for a single bias point or a sequence of bias points. While a current vs. voltage calculation is a standard output, many other outputs are available, including charge density, band profile, and transmission coefficients.

As shown in Figure 3(b) the NEMO-GUI operation includes most of the NEMO-Batch code. This design requires a separation of the GUI and batch code into different modules. The NEMO-GUI compilation includes all of the program modules whether they are batch or GUI-related. The NEMO-Batch compilation excludes purely GUI modules from the executable. There are some cases where a mixture of batch and GUI code is unavoidable, particularly in message routines and the functions that plot and store calculation results during the calculation. For those situations, a preprocessor-defined flag effectively deletes GUI-related code from any routines with a mix of batch-GUI code.

Both NEMO-Batch and NEMO-GUI use the same verification procedures to check the validity of the input parameters. The validation routine determines the parameters required for the requested calculation and checks for their presence in the input deck. If any material or model parameters are missing, NEMO provides default values from its internal database. In this manner, users can either accept the default parameters or input their own parameters. This feature addresses the needs of the device physicist who wants detailed control over the calculation process and the device engineer who only wants to determine general device behavior as quickly as possible.

Even though they use the same validation routines, NEMO-Batch and NEMO-GUI have very different means of responding to validation errors. For NEMO-Batch, the validation routine only needs to check parameters read from the Input Deck. If an error occurs in the verification procedure, NEMO-Batch must exit and report the error to a log file.

NEMO-GUI not only checks the Input Deck values, but also validates parameters as the user enters them into the GUI. If an error occurs for either case, NEMO-GUI displays an error message with advice on why the error occurred and how to remedy the problem. The user must input the correct information before starting the calculation or saving the input parameters. This procedure ensures that an Input Deck created using NEMO-GUI will be valid for NEMO-Batch operation. Consequently, even when using NEMO-Batch for the calculation, NEMO-GUI provides the most convenient method to construct the input deck.

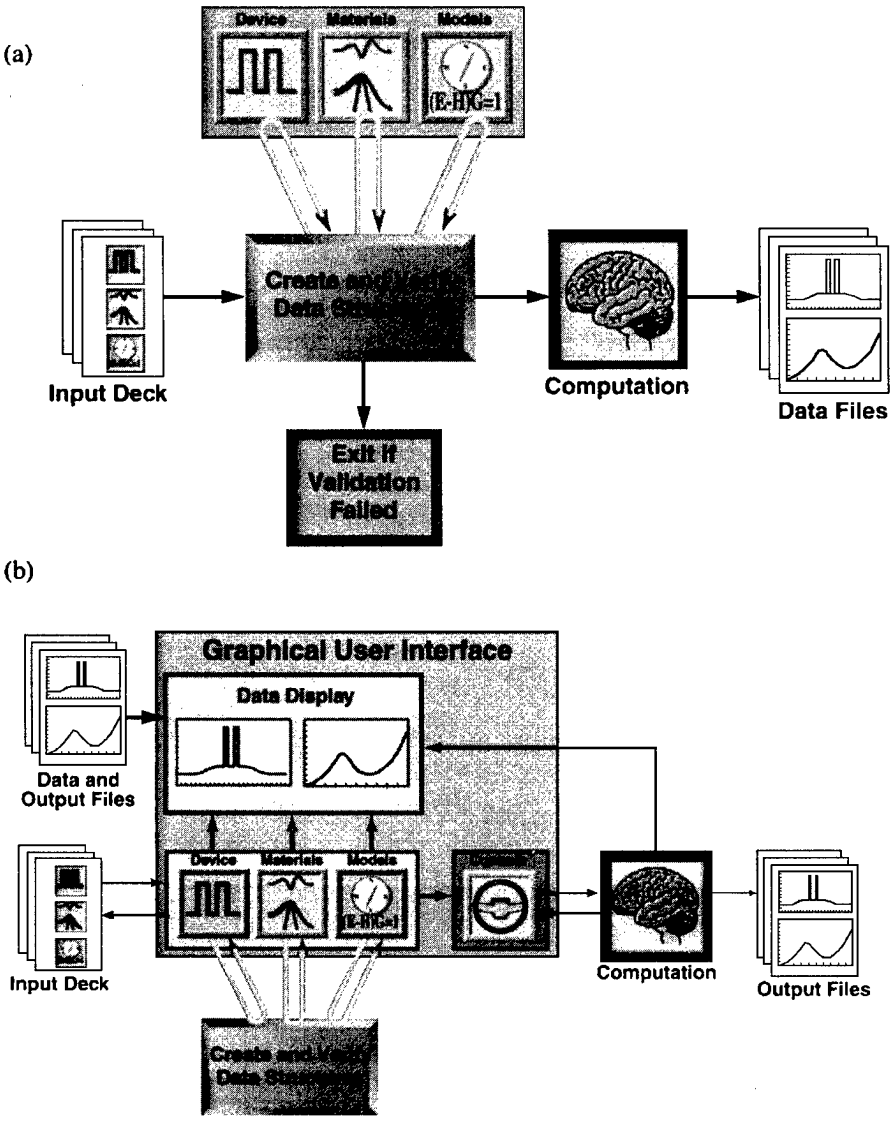


FIGURE 3 Basic design of NEMO program for (a) batch and (b) GUI operation. Input parameters are classified as device structure parameters, material parameters, and model parameters. Device structure parameters specify the device geometry, layer composition, and any other external physical parameters such as the terminal connections and temperature. Material parameters include the band structure information such as effective mass, energy gaps, and band offsets. Model parameters specify the theroretical models used for the calculation such as band structure, charge self-consistency, scattering models, simulation domains, and grids.

SIMULATION PARAMETER INPUT METHODOLOGY

NEMO incorporates a wide spectrum of models with varying tradeoffs between accuracy and speed. This range of models is essential given the large range of devices we wish to model for different scientific and engineering applications.

The downside of this flexibility is that NEMO has over 100 simulation parameters, many of

which interact in complex ways. In traditional simulators, the user must comprehend all of the simulation parameters, carefully select the essential parameters for a particular study, and ensure that all parameter values are self-consistent. This can be a daunting task for even the simplest simulation.

It is equally difficult for the programmer to deal with for two reasons: (1) New models and parameters added to the simulator must be reflected in the GUI as quickly as possible, and (2) simulation parameters should only be defined in a single module accessed by both the theory and the GUI to ease the code maintenance.

Our approach to this problem was to create the GUI interface at run-time based on a hierarchical system of simulation parameters. These simulation parameters are defined in a single module shared by the GUI and theory code. A key advantage to this method is that it insulates changes in the simulation parameter set from changes in the GUI and vice versa. The next section reviews the method used to create the run-time GUI interface.

The most basic level of the hierarchy directs the choice of fundamental models such as the potential, band structure, and scattering models. NEMO determines the data structures required by these models and creates a new GUI interface. The GUI displays default choices that the user can either accept or change.

When the user prompts NEMO to accept these parameters, NEMO tests for valid user input and enforces consistency between interacting parameters. For example, if one parameter value must be larger than another parameter, NEMO forces adherence to this rule.

After validating the selected options, NEMO determines the data structures required by the next level on the hierarchy, creates and displays a new GUI interface, and repeats the process until the parameter set is sufficient to run the simulation. If the user goes back to any portion of the hierarchy to select new options, parameters at higher levels of the hierarchy remain intact, while parameters at the lower levels are added and subtracted as appropriate.

This event-driven procedure only presents the parameters needed by the selected models. All other parameters are either hidden or are disabled from user entry. Limiting the parameter display in this manner greatly reduces the sheer number of simulation parameters presented to the user. Since NEMO provides default values for all simulation parameters, the user can often specify a few of the top-level parameters and then prompt NEMO to use the default values for the remainder of the hierarchy. In most cases, this approach generates a reasonable simulation. Nevertheless, the user has the option of adjusting all parameters down to the lowest level of the hierarchy, thus fulfilling a primary goal of NEMO to give the user full control over all aspects of the simulation.

DATA TRANSFER AND RUN-TIME GUI CREATION

The NEMO program must transfer large amounts of information between data structures, files, and the GUI. For a small program, it is sufficient to write customized routines for these operations, but this approach became impractical as NEMO grew in size. To resolve this problem, we designed a generic algorithm for transferring information to and from data structures. In addition to transferring information, this same algorithm creates portions of the GUI at run-time for parameter entry. A detailed description of this algorithm is beyond the scope of this paper, but due to its importance to NEMO and its utility to both the user and the programmer, we will highlight the basic features.

We refer to this process as the MemberDescriptor algorithm. We defined a MemberDescriptor data structure that contains all of the information needed to translate values between a data structure member and its destination. The MemberDescriptor includes a character string identifier, the

memory location of a data structure member relative to the data structure address, and an enumerated flag that defines the data type (number, character string, option list, etc.). Each data structure has at least one associated MemberDescriptor array. Each element of the MemberDescriptor array corresponds to one of the data structure members. To transfer information from the data structure, the translation routine uses the base address of the data structure and the MemberDescriptor array to derive the values of each data structure member. The inverse of this procedure transfers values from either the GUI or a file into each data structure member.

For example, assume we wish to transfer the value of an integer data structure member to a file. The translation routine adds the relative position of the integer data structure to the base address of the data structure. To determine the actual integer value, the translator casts the contents of the memory location to an integer. The derived integer value is stored in the file in the general format:

$$< \text{identifier} > = < \textit{value} >$$

where the <identifier> is the character string identifier set in the MemberDescriptor data structure and <*value*> is the integer number. Reversing this process inputs the same information from the file and stores the integer value into the data structure.

This procedure is used in many programs as a generic method for file input and output [6]. For NEMO, we extended the algorithm to transfer information between the GUI and the internal data structures. Figure 4 illustrates how data values are transferred between data structures and a GUI panel consisting of an option list, a text box, a check box, and a graphics plot with adjustable cursors.

For the simulation parameter input method discussed in the last section, we use the MemberDescriptor algorithm to create the GUI interface for each hierarchy. This dynamic design vastly enhances the usability of the GUI compared to a traditional static design that would require a fixed set of GUI elements throughout the program operation.

The MemberDescriptor approach saves a great deal of time and trouble for the programmer. Adding a new data structure no longer requires customized code to transfer values between data structures, files, and the GUI. Only a new MemberDescriptor array is needed. In fact, even software team members who are totally ignorant of GUI programming techniques can effectively "program" the GUI in this manner.

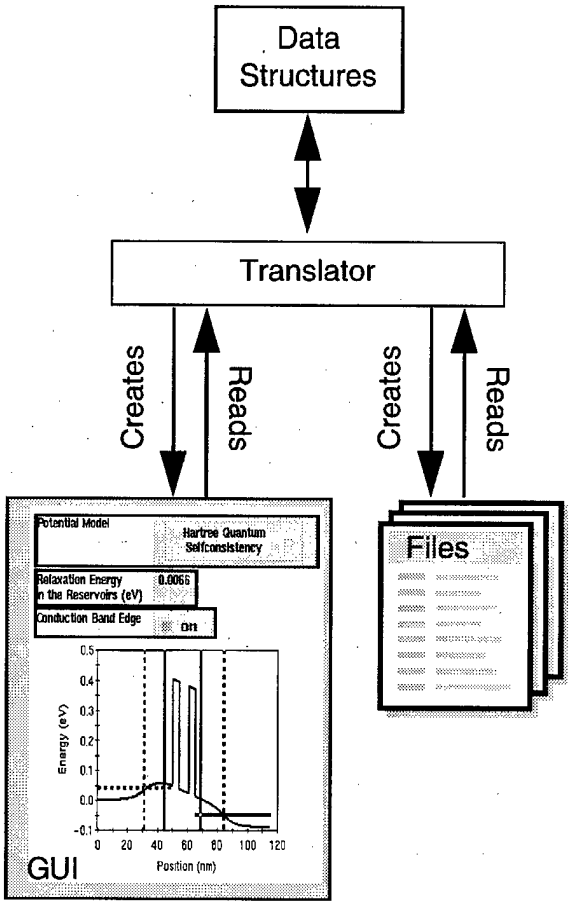


FIGURE 4 Generic data transfer between NEMO data structures, the GUI, and files. The structure and representations of the simulation parameters in the GUI and the files are fully specified in the data structures.

DEVELOPMENT ENVIRONMENT

The following summarizes key elements of the environment used to create the NEMO simulator:

Programming Language Most of the program was coded using ANSI C with some FORTRAN 77 and FORTRAN 90 code used in modules that required the greatest speed.

Development Platforms Code development occurred on Unix workstations either using HP Apollo or SGI platforms. We also ported NEMO to the Sun, IBM, and Linux Unix platforms.

Program Library System To facilitate code sharing and revision updates, we utilized a standard program library system called RCS (Revision Control System). The RCS system manages a check-out and check-in system that only allows one programmer to edit a file at a time. The RCS system also stores changes between code versions and can recover any previous version of the project.

Graphical User Interface The NEMO GUI was written in Motif which uses C-callable routines to create and control GUI objects on a wide variety of Unix platforms.

GUI Tools The static elements of NEMO were designed using the XDesigner software tool. The spreadsheet table and the 2D, 3D, and contour plot tools were programmed using the 3rd party XRT Widget set.

Run-time Debugger NEMO relies heavily on dynamic memory creation techniques that are extremely difficult to debug with conventional debuggers. The Purify debugger provided the means to find memory violations incurred at run-time and greatly simplified the task of tracking down these errors.

Parallelization NEMO-Batch has been parallelized using MPI (Message Passing Interface) at the outer loop of current-voltage characteristics. Individual bias points are computed in parallel.

CONCLUSION

We have presented highlights of the NEMO software project germane to construction of a large-scale device simulator. It is hoped that these methods will prove useful to other groups engaged in similar types of software projects, whether or not they pertain to quantum device simulations.

References

[1] Lake, R., Klimeck, G., Bowen, R. C. and Jovanovic, D. (1997). *J. Appl. Phys.*, **81**, 7845.
[2] Roblin, P. and Liou, W. (1993). *Phys. Rev.*, B **47**, 2146.
[3] Klimeck, G. *et al.* (1995). *Appl. Phys. Lett.*, **67**, 2539.
[4] Klimeck, G. *et al.* In the *1995 53rd* Annual Device Research Conference Digest, p. 52; *1997 55th Annual Device Research Conference Digest*, p. 92; R. Lake *et al. In the 1995 54th Annual Device Research Conference Digest*, p. 174 (IEEE, Inc., NJ).
[5] Bowen, R. C. *et al.* (1197). *J. Appl. Phys.*, **81**, 3207.
[6] Nye, A. and O'Reilly, T. (1993). X Toolkit Intrinsics Programming Manual OSF/Motif 1.2 edition, **4** of O'Reilly X-Windows Series, Chapt. 10.

Authors' Biographies

Gerhard Klimeck is a Member of the Technical Staff at Raytheon TI Systems. He is the NEMO program manager and his present research interest is the efficient simulation of quantum devices.

Dan Blanks is a Member of the Technical Staff at Raytheon TI Systems. His present research interest is the design and coding of the graphical user interface of the NEMO program.

Roger Lake is a Member of the Technical Staff at Raytheon TI Systems. His present research interests include scattering, disorder, and optical effects in RTD's and quantum wells.

R. Chris Bowen is a Member of the Technical Staff at Texas Instruments Incorporated. His current research interest is full-band modeling of quantum electron transport in Silicon and III–V structures.

Chenjing L. Fernando is a Member of the Technical Staff at Texas Instruments Incorporated. Her current research interest is process synthesis.

Manhua Leng has switched from computational electronics to computational finance on Wallstreet.

William R. Frensley is a Professor in the Electrical Engineering Dept. at the U. of Texas at Dallas. His current research interest is modeling of

quantum electron transport in nanoelectronic devices.

Dejan Jovanovic is a Staff Scientist at Motorola Inc. His research interests include the quantum mechanical simulation of ultra-scaled FET's.

Paul Sotirelis is the Computational Electronics and Nanoelectronics On-Site Lead at the ASC-MSRC, one of four large DoD high performance computing centers.

OPTOELECTRONICS

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 87–91
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Theory and Modeling of Lasing Modes in Vertical Cavity Surface Emitting Lasers

BENJAMIN KLEIN[a,*], LEONARD F. REGISTER[a], KARL HESS[a] and DENNIS DEPPE[b]

[a] *Beckman Institute and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801;*
[b] *Microelectronics Center, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084*

The problem of obtaining the lasing modes and corresponding threshold conditions for vertical cavity surface emitting lasers (VCSELs) is formulated as a *frequency-dependent* eigenvalue problem in required *gain amplitudes* and corresponding fields. Both index and gain guiding are treated on an equal footing. The complex gain eigenvalues define necessary but not sufficient conditions for lasing. The actual lasing frequencies and modes that the VCSEL can support are then determined by matching the gain necessary for the optical system in both magnitude and phase to the gain available from the laser's electronic system. Examples are provided.

Keywords: VCSELs, modes, lasing, modeling, gain eigenvalues

1 INTRODUCTION

In the simulation of edge emitting lasers, the optical field can usually be modeled simply by a predetermined set of Fox-Li quasimodes of the passive cavity [1]. The threshold condition can then be expressed as mode gain times mode lifetime equals unity. No laser previously has presented the challenge to optical simulation now presented by vertical cavity surface emitting lasers (VCSELs).

The challenge to VCSEL simulation is not just a much more complicated optical cavity geometry for analysis – although this problem alone is significant – but that quasi-mode analysis, itself, can no longer be relied upon. The mirrors are distributed and lateral confinement may be produced by a combination of index and gain guiding. (See, for example, Ref. [2].) Thus the VCSEL optical cavity boundaries, along with the passive cavity modes and conventional parameters such as the photon lifetime, are often poorly defined.

2 THEORY

At the threshold for lasing, the optical field will be self-supporting in the presence of the gain supplied

* Corresponding author.

by the laser. Consider an open VCSEL optical cavity with tensor electric susceptibility $\overleftrightarrow{\chi}(\vec{r},\omega) = \overleftrightarrow{\chi}_{\text{cav}}(\vec{r},\omega) + \overleftrightarrow{\chi}_g(\vec{r},\omega)$ where $\overleftrightarrow{\chi}_g(\vec{r},\omega)$ is the necessarily complex susceptibility representing the laser gain inside the active region provided under bias, and $\overleftrightarrow{\chi}_{\text{cav}}(\vec{r},\omega)$ is the potentially complex susceptibility representing the rest of the VCSEL. Within the semiclassical approximation, lasing requires [3]

$$\vec{E}(\vec{r},\omega) = \int d\vec{r}\,' \overleftrightarrow{G}_{\text{cav}}(\vec{r},\vec{r}\,',\omega) \cdot [j\omega\varepsilon_o \overleftrightarrow{\chi}_g(\vec{r}\,',\omega) \cdot \vec{E}(\vec{r}\,',\omega)], \tag{1}$$

where $\vec{E}(\vec{r},\omega)$ is the electric component of the lasing field, $G_{\text{cav}}(\vec{r},\vec{r}\,',\omega)$ is the tensor Green's function for radiation from a current source within the open VCSEL cavity defined by $\overleftrightarrow{\chi}_{\text{cav}}$, and $j\omega\varepsilon_o \overleftrightarrow{\chi}_g(\vec{r}\,',\omega) \cdot \vec{E}(\vec{r}\,',\omega)$ acts as an equivalent current source.

If the spatial distribution of the gain is known to good approximation, then Eq. (1) reduces to an eigenvalue problem in the necessary gain amplitudes for lasing and the corresponding lasing fields. That is, if $\overleftrightarrow{\chi}_g(\vec{r},\omega) = \kappa(\omega)\overleftrightarrow{\chi}_g^{(0)}(\vec{r})$ where $\overleftrightarrow{\chi}_g^{(0)}(\vec{r})$ is known, then Eq. (1) gives

$$\frac{1}{\kappa(\omega)}\vec{E}(\vec{r},\omega) = \hat{\mathcal{G}}(\omega)\vec{E}(\vec{r},\omega), \tag{2}$$

where $\hat{\mathcal{G}}(\omega)$ is an integral operator defined by

$$\hat{\mathcal{G}}\vec{E}(\vec{r}) \equiv j\omega\varepsilon_o \int d\vec{r}\,' \overleftrightarrow{G}_{\text{cav}}(\vec{r},\vec{r}\,') \cdot [\overleftrightarrow{\chi}_g^{(0)}(\vec{r}\,') \cdot \vec{E}(\vec{r}\,')]. \tag{3}$$

Note that this eigenvalue problem need only be evaluated over regions for which $\overleftrightarrow{\chi}_g^{(0)}(\vec{r})$ is significant, that is, over the active region of the laser. The solutions to this eigenvalue problem will be complex gain eigenvalues and the corresponding fields as a continuous function of frequency, in contrast to the more familiar case of finding the discrete real frequency eigenvalues and modes of a closed passive cavity.

It is also possible to repartition the VCSEL susceptibility to take part of $\overleftrightarrow{\chi}_{\text{cav}}$ out of the Green's function and put it in Eq. (2) as a second source term. This may assist in the solution of problems for which the Green's function is not easily obtained. Extending the formulation in this way changes the ordinary eigenvalue problem of Eq. (2) into a generalized eigenvalue problem.

The solutions to Eq. (2) establish necessary but not sufficient conditions for the VCSEL to lase. The frequencies at which the VCSEL can actually lase are those for which the complex gain susceptibility necessary for the optical system can be matched to the complex gain susceptibility available from the laser electronic system in both magnitude and phase.

3 EXAMPLE

To illustrate the use of this formulation, lasing in the VCSEL cavity diagrammed in Figure 1 was considered [2]. This cavity allows for relatively easy analysis while allowing much of the essential physics of VCSELs to be modeled. $\chi_{\text{cav}}(\vec{r},\omega)$was taken as that of the planarly layered structure including the nominally lossy quantum well layer, but with no lateral confinment. Note, for simplicity, all susceptibilities were treated as scalars. The spatial distribution of the gain susceptibility was

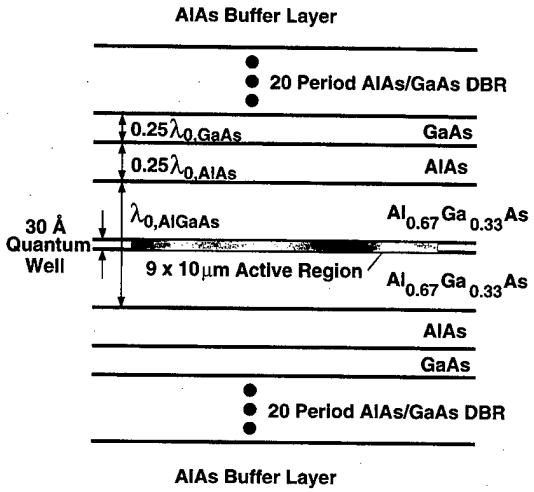


FIGURE 1 Model VCSEL cavity. $\lambda_o = 980\,\text{nm}$.

approximated as uniform inside the active region and zero elsewhere. Thus for later convenience we set $\chi_g^{(0)}(\vec{r}) \equiv -j$ within the $9\,\mu m \times 10\,\mu m \times 3\,nm$ quantum well active region, and $\chi_g^{(0)}(\vec{r}) \equiv 0$ elsewhere.

The Green's functions for such a planarly layered structure are well known [4], and only those with both source and field coordinates within the thin quantum well active layer were required. The Green's functions were obtained in Fourier-space where the only portion of the calculation that requires a computer is the calculation of the distributed-Bragg-reflector (DBR) reflectivities vs. frequency, incident angle and polarization, which may be performed using several equivalent methods [5, 6]. Of course, calculation of the Green's function for more complex cavity structures could require computationally intensive numerical methods, while taking advantage of the cylindrical symmetry in many VCSELs could significantly reduce the computational load. However, while of obvious practical importance, how the Green's function is obtained or what coordinate system is used is of no conceptual importance in this formulation. Note that the potentially difficult problem of finding the Green's function would only need to be solved once for all bias conditions, as the Green's functions are independent of bias as long as effects such as thermal expansion of the cavity are ignored.

Once the Green's function was calculated, the gain eigenvalue problem of Eq. (4) was discretized and solved. Using a moment method, the field, approximated as constant over the width of the well, was expanded in rectangular testing functions with unknown complex coefficients in the $(x-y)$ plane of the well, i.e.

$$\vec{E}(x,y) = \sum_{\vec{n}} \vec{E}(\vec{n})\,\mathrm{rect}\left(\frac{x-n_x d_x}{d_x}\right)\mathrm{rect}\left(\frac{y-n_y d_y}{d_y}\right), \tag{4}$$

where $\mathrm{rect}[(\eta-\eta_o)/d]$ is a rectangular pulse of amplitude one and full width d centered at η_o, and the discrete variable $\vec{n} \equiv (n_x, n_y)$ labels the grid sites. (For notational convenience, the various dependencies on frequency are no longer indicated explicitly.) Then the inner products of both sides of Eq. (4) were taken with the same rectangular testing functions (Galerkin's method [7]). The result was a finite set of linear equations for the unknown field coefficients $\vec{E}(\vec{n})$ of the same form as Eq. (2),

$$\sum_{\vec{n}'} j\omega\varepsilon_o \chi_g^{(0)}(\vec{n}) \overleftrightarrow{G}_{\mathrm{cav}}(\vec{n},\vec{n}') \cdot \vec{E}(\vec{n}') = \frac{1}{\kappa}\vec{E}(\vec{n}). \tag{5}$$

The corresponding discretized Green's function was obtained from the continuous Green's function in Fourier-space by

$$\overleftrightarrow{G}_{\mathrm{cav}}(\vec{n},\vec{n}') = \int\int dk_x dk_y e^{jk_x d_x(n_x-n_x')} e^{jk_y d_y(n_y-n_y')} \times \frac{4\sin^2(k_x d_x/2)\sin^2(k_y d_y/2)}{\pi^2 k_x^2 k_y^2 d_x d_y} \overleftrightarrow{G}_{\mathrm{cav}}(k_x,k_y) \tag{6}$$

for these rectangular testing functions. For this work, a commercially available (IMSL) routine for generalized complex eigenvalue problems was used to solve Eq. (5).

Figure (2) shows the first three gain eigenvalues $\kappa(\omega)$ (two are essentially degenerate) plotted at

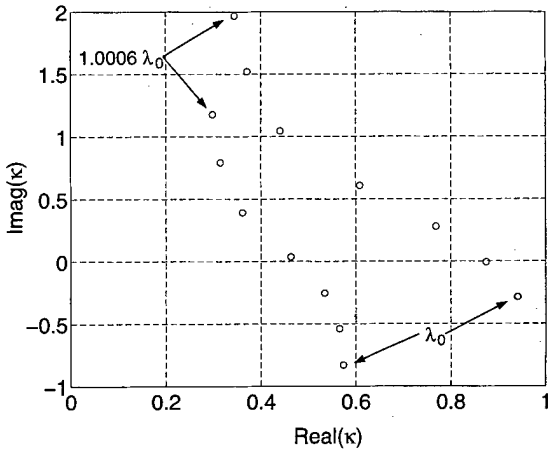


FIGURE 2 First three complex gain eigenvalues at discrete intervals in frequency; two are essentially degenerate. Lasing is possible only where the eigenvalue curves crosses the real axis, with the values of the gain eigenvalues at these crossings defining the threshold condition for lasing.

constant intervals in frequency. Approximating the gain susceptibility available from the electronic system of the laser under bias as purely imaginary, lasing is only possible where one of the gain eigenvalue curves crosses the real axis, such that $\chi_g(\vec{r},\omega) = \kappa(\omega)\chi_g^{(0)}(\vec{r})$ is also purely imaginary. The gain eigenvalue at the crossing frequency defines the threshold gain susceptibility. Approximating the gain susceptibility as purely imaginary is done here for illustrative purposes only. In reality the actual lasing modes of the system would be determined by the intersection of the gain eigenvalues with a curve describing the Kramers-Kronig relation between the real and imaginary parts of the gain susceptibility. The large area active region considered in this example produces a very tight spacing of gain eigenvalues in frequency, but the gain amplitudes required for lasing are significantly different because of the different overlaps of the self-consistently calculated lasing modes with the active region. Figure (3) shows the field patterns of the first and third lasing modes inside the gain region.

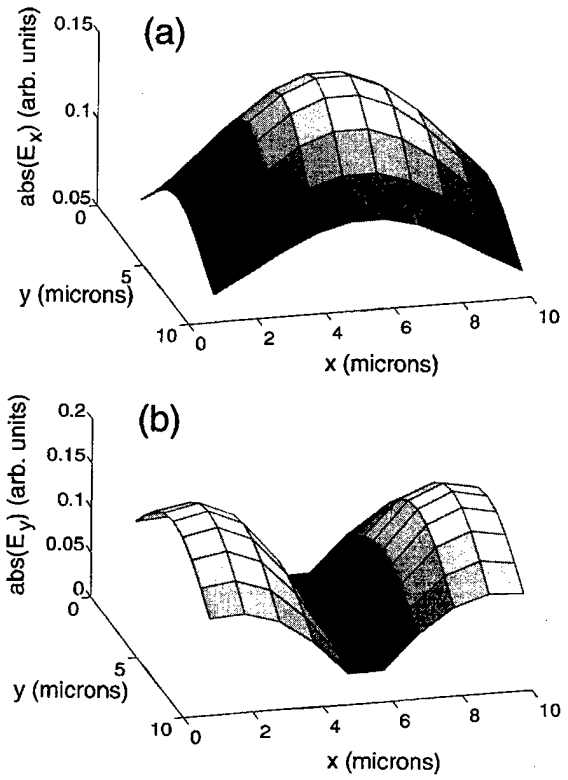


FIGURE 3 Lasing fields at (a) $\lambda \simeq 1.0003\ \lambda_o$ and (b) $\lambda \simeq 1.0001\ \lambda_o$ as a function of position within the gain region.

4 CONCLUSION

The problem of obtaining the lasing modes and corresponding threshold conditions for VCSELs has been formulated as a *frequency-dependent* eigenvalue problem in required *gain amplitudes* and corresponding fields. Index and gain guiding are treated on an equal footing. The complex gain eigenvalues define necessary but not sufficient conditions for lasing. The actual lasing frequencies and modes that the VCSEL can support are then determined by matching the gain necessary for the optical system in both magnitude and phase to that available from the laser electronic system. With this formulation the lasing modes and corresponding threshold conditions are well defined even when the optical cavity boundaries and conventional cavity parameters such as the photon lifetime are not.

This formulation also has the practical advantage that the problem of obtaining the self-supporting lasing fields of a VCSEL cavity is separated into two distinct and already thoroughly studied problems: that of finding the optical Green's functions for a fixed radiation source in an open cavity and that of solving a complex eigenvalue problem. Thus, this formulation provides a framework for the application of a large preexisting knowledge base to the relatively new challenge of modeling VCSELs. For example, in the specific implementation used for this work, the cavity Green's function was obtained by textbook methods [4–6] and the generalized complex eigenvalue problem was solved using a commercially available numerical routine.

Acknowledgements

This material is based upon work supported by the Office of Naval Research.

References

[1] Grupen, and Hess, invited paper IWCE, Tempe Az, October 1995, to be published in VLSI Design.
[2] Lin, C. C. and Deppe, D. G. (1995). *J. Lightwave Tech.*, **13**, 575.
[3] Chow, W., Koch, S. and Sargent, M. *III, Semiconductor Laser Physics* (Springer-Verlag, Berlin, 1994) p. 55.
[4] Chew, W. C. *Waves and Fields in Inhomogeneous Media* (Van Nostrand Reinhold, New York, 1990) pp. 57–79, 375–418.
[5] Chew, W. C. pp. 45–56.
[6] Born, M. and Wolf, E. *Principles of Optics, Sixth Edition (corrected)* (pergamon Press, Oxford, 1993) pp. 51–66.
[7] Balanis, C. A. *Advanced Engineering Electromagnetics* (John Wiley and Sons, New York, 1989) p. 692.

Authors' Biographies

Benjamin Klein received his B.S. and M.S. degrees in Electrical Engineering from the University of Wisconsin at Madison in 1994 and 1996, respectively. He is currently pursuing his Ph.D. in Electrical Engineering at the University of Illinois at Urbana-Champaign. His thesis research is the numerical simulation of vertical cavity surface emitting lasers.

Leonard F. Register is a Research Scientist in the Beckman Institute and Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign. His current research interests include dissipative quantum transport, laser simulation, and leakage currents in thin oxides.

Karl Hess has dedicated the major portion of his research career to the understanding of electronic current flow in semiconductors and semiconductor devices with particular emphasis on effects pertinent to device miniaturization. His theories and use of large computer resources are aimed at complex problems with clear applications and relevance to miniaturization of electronics. He is currently the Swanlund Professor of Electrical and Computer Engineering, Professor of Physics, Adjunct Professor for Supercomputing Applications and a Research Professor in the Beckman Institute working on topics related to Molecular and Electronic Nanostructures. He has received numerous awards, for example the IEEE David Sarnoff Field Award for electronics in 1995.

Dennis G. Deppe is an Associate Professor in the Electrical and Computer Engineering Department at The University of Texas at Austin. He received the Ph.D. degree in 1988 from the University of Illinois at Champaign-Urbana, where he studied atomic diffusion in III–V semiconductor heterostructures, and semiconductor laser fabrication and characterization. After receiving the Ph.D. in 1988 he worked as a Member of Technical Staff at AT&T Bell Laboratories in Murray Hill, New Jersey researching semiconductor lasers, and in 1990 accepted a position as Assistant Professor at The University of Texas at Austin. He was promoted to Associate Professor in 1994, and holds the Robert and Jane Mitchell Endowed Faculty Fellowship in Engineering. His research specialties include III–V semiconductor device fabrication, optoelectronics, and laser physics. He has published over 120 technical journal articles and holds 6 U.S. patents, and has received the National Science Foundation Presidential Young Investigator and Office of Naval Research Young Investigator Awards.

DEVICE SIMULATION III

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 93–98
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Cellular Automaton Study of Time-Dynamics of Avalanche Breakdown in IMPATT Diodes

G. ZANDLER[a], R. OBERHUBER[a], D. LIEBIG[a], P. VOGL[a,*], M. SARANITI[b] and P. LUGLI[c]

[a] *Walter Schottky Institute, Technical University of Munich, 85748 Garching, Germany;*
[b] *Electrical Engineering Department, Arizona State University, Tempe, AZ 85287-6206;*
[c] *Electrical Engineering Department, University of Rome, Tor Vergata, 00133 Rome, Italy*

Employing a recently developed efficient cellular automaton technique for solving Boltzmann's transport equation for realistic devices, we present a detailed study of the carrier dynamics in GaAs avalanche *p-i-n* (IMPATT) diodes. We find that the impact ionization in reverse bias *p-i-n* diodes with ultrathin (less than 50 nm) intrinsic regions is triggered by Zener tunneling rather than by thermal generation. The impact generation of hot carriers occurs mainly in the low-field junction regions rather than in the high field intrinsic zone. The calculations predict significantly more minority carriers on the *n*-side than on the *p*-side.

Keywords: Cellular Automata, IMPATT diodes, zener tunneling, impact ionization, avalanche breakdown

1. INTRODUCTION

The carrier dynamics in modern nanometer size devices is dominated by hot carrier effects and nonlocal transport phenomena such as tunneling and impact ionization. A realistic prediction and understanding of these effects in real devices that operate at room temperature requires at least the solution of the full semiclassical Boltzmann equation. A few years ago, we have developed a novel method, the Cellular Automaton (CA) approach namely, that provides a computationally efficient scheme for solving Boltzmann's equation in position and momentum phase space [1–2]. Since then, we have significantly refined this approach [3–4] and enhanced its speed and robustness; a detailed review of the present status of the CA method has been given very recently [5]. The present paper focuses on the application of this scheme to a physically intriguing situation where the carrier dynamics is highly complex and very far from equilibrium. We present a micro-

* Corresponding author: phone: +49-89-289-12750, FAX +49-89-289-12737, email: vogl@wsi.tu-muenchen.de.

scopic analysis of the carrier dynamics near avalanche breakdown in GaAs and AlGaAs IMPATT diodes.

2. THE CA METHOD

We briefly summarize the highlights of the CA method [1–5]. A cellular automaton consists of a lattice with a finite number of states attached to each lattice site that can be interpreted as pseudoparticles. Their dynamics is governed by a limited set of *local* transition rules. The perhaps most crucial factor that sets cellular automata apart from standard finite differencing methods is the reduction of all physical variables to a finite set of discrete values. In contrast to finite differencing methods, the size of this set can be kept relatively small due to an optimized phase space discretization. This discretization utilizes the fact that the final carrier distribution function changes on a much coarser scale in phase space than typical changes of phase space variables within on time step. In our present implementation, we use a two-dimensional hexagonal lattice in real space. Attached to each lattice site are of the order of 10^5 momentum states that form a hexagonal closed packed structure. Transition rules among these states represent the quantum mechanical collisions as well as the drift and diffusion terms in the Boltzmann equation. In the simulation, the particle dynamics consists mainly in look-up operations of pre-calculated and hierarchical scattering tables. This results in a numerically very efficient algorithm for the particle dynamics that is typically faster than the standard Monte Carlo scheme by a factor of 30–50.

From a physics point of view, in particular concerning scattering mechanisms, the CA and Monte Carlo scheme are equivalent. Indeed, we have employed both methods in obtaining the results of this paper, mostly to check the accuracy and consistency of the CA results.

3. CARRIER DYNAMICS IN IMPATT DIODES

Recently, GaAs and GaAlAs based IMPact Avalanche Transit Time (IMPATT) diodes have been fabricated with a high power output at frequencies up to 200 GHz [6–8]. The avalanche zone of such a diode consists of a *p-i-n* diode with highly doped *n* and *p* regions and an ultrathin intrinsic zone of 20 to 50 nm. The build-up of the carrier avalanche under high reverse bias is a subtle interplay between thermal generation, interband (Zener) tunneling, impact ionization, and other scattering mechanisms [9].

3.1. Scattering Rates

The present simulations incorporate all relevant standard scattering mechanisms for carriers, such as ionized impurity, plasmon, intra- and intervalley phonon scattering, alloy scattering, and thermal generation, and nonparabolic electron and hole bands [10]. Impact ionization is accounted for by invoking the model of Kane [11], using density of states that have been calculated with the empirical pseudopotential method. The $k \cdot p$ model of Krieger [12] is employed to calculate the interband tunneling rate for direct transitions as a function of electric field. The absolute magnitude of the impact ionization rate of electrons and holes is chosen as to reproduce the experimental ionization coefficients in bulk GaAs and AlGaAs [13–15].

In Figures 1 and 2, we show several scattering rates for electrons and holes in GaAs. This figure reveals that the impact ionization scattering rate for holes is markedly lower than for electrons. An interesting consequence of this difference is illustrated in Figure 2. It shows the average distance an electron or hole travels in bulk GaAs between impact ionization scattering events when the field is 1 MV/cm. In contrast to the quasi-ballistically moving electrons, the slower holes suffer many other scattering processes before they are able to

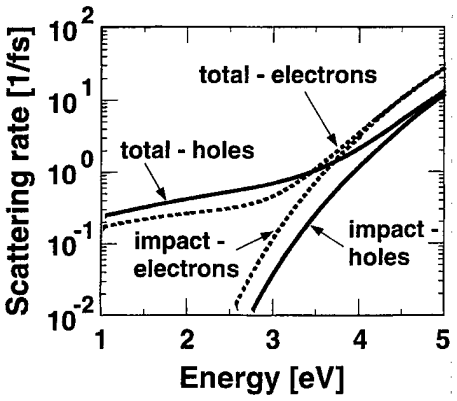


FIGURE 1 Impact ionization scattering rates (labeled by "impact") and total scattering rates ("total"), in units of 1/fs, for electrons in the Γ valley and heavy holes, respectively, in bulk GaAs at room temperature and a doping level of $2\times10^{18}\,\mathrm{cm}^{-3}$ as a function of energy in eV.

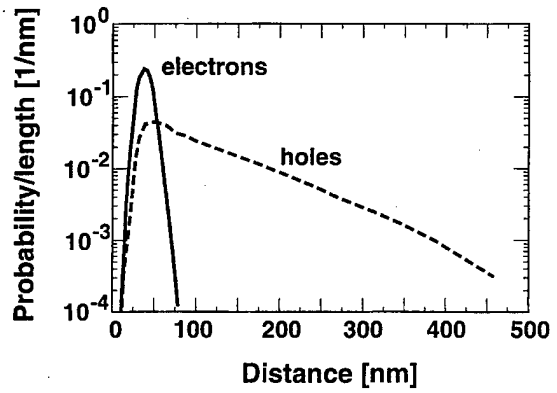


FIGURE 2 Probability per distance for an electron (respectively, hole) in bulk GaAs to initiate an impact ionization along its trajectory of a length specified by the abscissa in nm. The electric field is 1 MV/cm.

impact ionize. This leads to the broad distribution of hole trajectories in Figure 2.

3.2. The Avalanche Generation Process: Qualitative Picture

We now consider a GaAs *p-i-n* diode with a 20 nm intrinsic zone and a carrier concentration of $n=p=2\times10^{18}\,\mathrm{cm}^{-3}$. The density in the intrinsic zone is set to $n=2\times10^{15}\,\mathrm{cm}^{-3}$. In Figure 3, we illustrate schematically the time development of

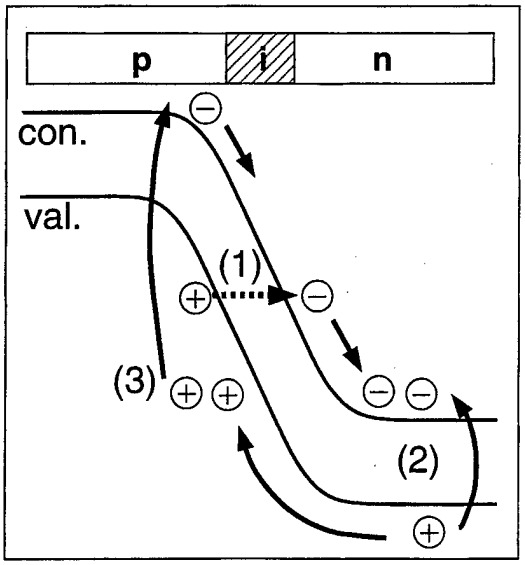


FIGURE 3 Schematic picture of carrier dynamics in a highly doped IMPATT diode with ultrathin *i*-zones under reverse bias that is close to breakdown. The relevant dynamical processes are Zener tunneling, labeled by (1), impact ionization in the depletion zone (labeled by (2) and only shown on the *n*-side for simplicity), and impact ionization in the high field region, labeled by (3).

the carrier distribution in the diode after applying a reverse bias of 5.5 eV.

For a reverse bias between 4 and 5 V, the resulting high field of approximately 1 MV/cm initiates electron hole pair generation by interband tunneling (step (1) in Fig. 3). The thermal generation rate is found to be negligible compared to the tunneling rate for this device geometry. The generated electrons on the *n*-side get rapidly accelerated by the electric field, move away from the band edge and gain a significant amount of energy. The key point is that the electrons dissipate their excess energy efficiently by impact ionization within the low field depletion region rather than in the high field zone (step (2)). The holes that are generated via these impact ionization processes get accelerated by the field back into and through the whole intrinsic zone. This allows the holes to gain sufficient excess energy so that they induce impact ionization processes already within the *i*-zone. This ignites the avalanche process (step (3)).

Thus, the carrier generation in GaAs IMPATT diodes with ultrathin *i*-zones is triggered by Zener tunneling in the high field region, whereas the generation process itself starts within the depletion layers. This leads to a pronounced dark-space effect in the impact ionization rates. For applied voltages below 5.7 V, this carrier generation cycle does not lead to breakdown. However, it takes almost 30 ps after switch-on before the current is fully stationary when the applied bias is close to this breakdown threshold value.

Since the impact ionization rate and the drift velocity of electrons is higher than that of holes, the electrons get more efficiently cooled on the *n*-side of the intrinsic zone than the holes do on the opposite side. This leads to a slightly higher average energy of the holes, as well as to a higher density of holes on the *p*-side than electrons on the *n*-side. In the avalanche regime, the higher hole density effectively screens the electric field in the intrinsic zone. This leads to an interesting effect for short *i*-zones of less than 50 nm. Once the reverse bias exceeds the threshold value for avalanche multiplication, the holes are able to flood the whole *i*-zone. This screens the electric field and effectively impedes the avalanche breakdown.

3.3. Quantitative Results and Comparison with Data

To substantiate this qualitative picture, Figures 4(a)–4(d) characterize the carrier distribution for a reverse bias of 5.5 V in the 20 nm *p-i-n* diode (i.e., just below breakdown) quantitatively. The spatial field profile is drawn as grey inset in these figures in order to show the spatial extent of the depletion zone. The electric field is constant within the intrinsic zone and amounts to 1.1 MV/cm.

The energy of electrons (Fig. 4(a)) reaches a maximum at the center of the high field zone. There, the electrons loose energy predominantly by impact ionization. Electrons are able to fly a longer distance than holes before they loose energy and pick up a given amount of energy on a shorter trajectory. Therefore, the average electron energy

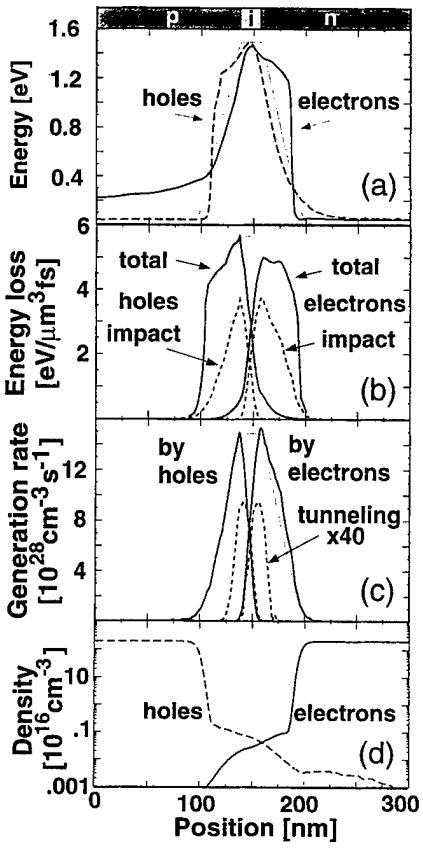


FIGURE 4 The figure refers to a 20 nm PIN diode with a doping level of $2\times10^{18}\,cm^{-3}$ in the *n*- and *p*-regions, respectively, and an applied reverse bias of 5.5 V. The shape of the electric field is shown in grey to indicate the width of the depletion zone. Its magnitude is 1.1 MV/cm. (a) Calculated average energy of electrons and holes as a function of position. (b) Calculated energy loss rate of electrons and holes. Shown are the total energy losses and the contributions from impact ionization. (c) Calculated impact ionization generation rate of electron hole pairs induced by electrons and holes, respectively (full lines). Zener generation rate of electrons and holes, respectively (dashed lines). (d) Calculated density of electrons and holes.

decreases more slowly and extends farther into the depletion region on the *n*-side than the hole energy does on the *p*-side.

The same effect can be deduced from the spatially resolved energy loss (Fig. 4(b)). The energy loss by impact ionization is higher for electrons than for holes and remains large throughout the depletion region on the *n*-side. Since electrons gain energy more rapidly than

holes, they are more likely to impact ionize repeatedly. The slower holes, on the other hand, remain hotter which leads to their higher total energy loss. Particularly near the end of the *p*-sided depletion region, their energy loss is dominated by phonon and plasmon scattering. These results are consistent with the higher integral of the electron induced electron-hole pair generation rate that is shown in Figure 4(c). This figure also includes the number of generated electrons and holes by Zener tunneling. The latter generation process gives a negligible contribution to the current within 0.2 V of breakdown.

The higher electron-induced impact ionization rate generates more holes near the *i-n* junction than the opposite process generates electrons near the *p-i* region. Therefore, the density of minority carriers is higher on the *n*-side than on the *p*-side (Fig. 4(d)). This effect is enhanced by the higher mobility of electrons that causes them to diffuse out of the *p*-zone more rapidly. The asymmetry of carrier densities within the high field region that can be seen in Figure 4(d) is another consequence of the higher impact scattering rate for electrons. The holes that are generated by impact ionization on the *n*-side get accelerated through the high field region towards the *p*-zone, leading to the excess density of holes on the *p*-side.

Figure 5 compares the present calculations with the measured current-voltage characteristics for a 20 nm and 50 nm pin diode, respectively. The agreement is seen to be very good. In Figure 6, we predict the I-V characteristics of 30 nm AlGaAs *p-i-n* structures. Since the energy gap is higher than in GaAs, the impact ionization rate is lower which causes the breakdown voltage to increase and the current density to decrease with increasing Al concentration. The lower breakdown voltage compared to the 20 nm diode discussed above is caused by the higher doping level of the *n* and *p*-regions.

The work has been partially supported by SIEMENS and by the Deutsche Forschungsgemeinschaft (SFB 384).

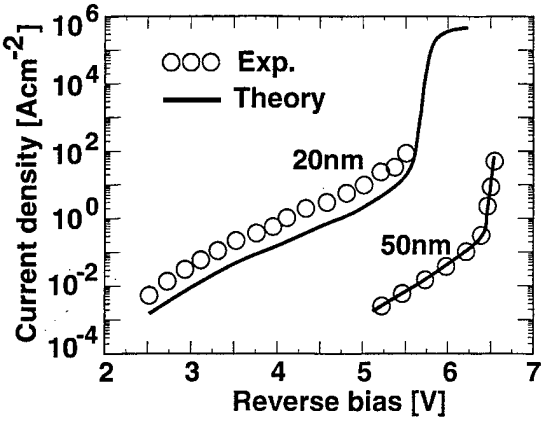


FIGURE 5 Calculated (full curves) and experimental current densities (stars), in A/cm^2, versus applied reverse bias in V for a 20 nm and 50 nm GaAs *p-i-n* diode. The doping concentration in the *n* and *p*-layers amounts to 2×10^{18} cm^{-3}. The experimental data are from [6].

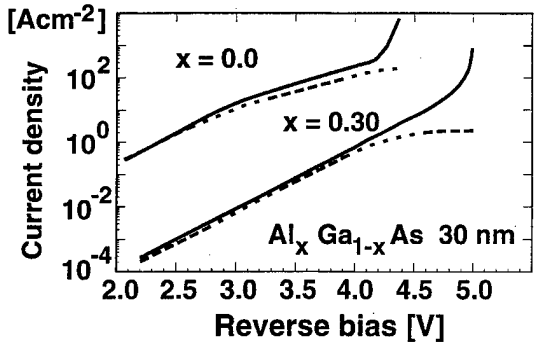


FIGURE 6 Predicted current densities versus applied reverse bias for a 30 nm GaAs and $Ga_{0.7}Al_{0.3}As$ *p-i-n* diode. The doping concentration in the *n* and *p*-layers is 5×10^{18} cm^{-3}. The dotted lines show the Zener tunneling contribution to the current density.

References

[1] Kometer, K., Zandler, G. and Vogl, P. (1992). "Lattice-gas cellular-automaton method for semiclassical transport in semiconductors", *Phys. Rev. B.*, **46**, 1382–1394.

[2] Zandler, G., Di Carlo, A., Kometer, K., Lugli, P., Vogl, P. and Gornik, E. (1993). "A comparison of Monte Carlo and cellular automaton approaches for semiconductor device simulations", *IEEE Electron Dev. Letters*, **14**, 77–79.

[3] Rein, A., Zandler, G., Saraniti, M., Lugli, P. and Vogl, P. (1994). "Cellular Automaton Simulations of Planar Doped Barrier Field Effect Transistor", *Proceedings of the 3rd International Workshop on Computational Electronics* (S. M. Goodnick, ed.), 7–10, Oregon State University Press, Corvallis.

[4] Zandler, G., Saraniti, M., Rein, A. and Vogl, P. (1996). "Cellular Automata for Device Simulation É Concepts and Applications", *Proceedings of 1996 Int. Conf. on Simulation of Semiconductor Processes and Devices*, 39–42, (Business Center for Academic Societies Japan), Tokyo, ISBN 0-7803-2745-4.
[5] Vogl, P., Zandler, G., Rein, A. and Saraniti, M. (1997). "Cellular Automaton approach for semiconductor transport", in *Theory of Transport Properties of Semiconductor Nanostructures* (E. Schoell, ed.), Chapman and Hall, London, in print.
[6] Gaul, L., Huber, S., Freyer, J. and Claassen, M. (1991). "Determination of tunnel-generation rate from GaAs PIN-structures", *Solid State Electron.*, **34**, 723–726.
[7] Poebl, M., Bogner, W. and Gaul, L. (1994). "CW GaAs MITATT Source on Copper Heatsink up to 160 GHz", *Electronic Letters*, **30**, 1316–1317.
[8] Tschernitz, M. and Freyer, J. (1995). "140 GHz GaAs Double-Read IMPATT Diodes", *Electronic Letters*, **31**, 582–583.
[9] Liebig, D., Lugli, P., Vogl, P., Claasen, M. and Harth, W. (1992). "Tunneling and ionization phenomena in GaAs PIN diodes", *Microelectronic Engineering*, **19**, 127–130.
[10] Canali, C., Pavan, P., Di Carlo, P., Lugli, P., Malik, R., Manfredi, M., Neviani, A., Vendrame, E., Zanoni, E. and Zandler, G. (1996). "Experimental and Monte Carlo analysis of impact-ionization in AlGaAs/GaAs HBT's", *IEEE Trans. on Elec. Dev.*, **43**, 1769–1777.
[11] Kane, E. O. (1967). "Electron scattering by pair production in silicon", *Phys. Rev.*, **159**, 624–631.
[12] Krieger, J. B. (1996). "Theory of tunneling in semiconductors with dgenerate band structure", *Ann. Phys.*, **36**, 1–60.
[13] Ando, H. and Kanbe, H. (1981). "Ionisations coefficient measurements in GaAs by using multiplication noise characteristics", *Solid State Electron.*, **24**, 629–634.
[14] Bulman, G. E., Robbins, V. M., Brennan, K. F., Hess, K. and Stillman, G. E. (1983). "Experimental determination of Impact ionization coefficients in (100) GaAs", *IEEE Electron Dev. Lett.*, **4**, 181–185.
[15] Robbins, V. M., Smith, S. C. and Stillman, G. E. (1988). "Impact ionization in $Al_xGa_{1-x}As$ for $x=0.1-0.4$", *Appl. Phys. Lett.*, **52**, 296–298.

Authors' Biographies

Günther Zandler is research associate at the Walter Schottky Institute of the Technical University of Munich and deals with high field transport and ultrafast phenomena.

Ralph Oberhuber is doing his Ph.D. thesis at the Technical University in Munich on new device concepts.

Dennis Liebig is doing his Ph.D. thesis at the Technical University in Hamburg-Harburg on device modeling.

Peter Vogl is professor of physics at the Technical University of Munich. He is interested in the theory of electronic structure and carrier transport in solids.

Marco Saraniti is post-doctoral fellow at ASU in Tempe. He focuses on numerical aspects of the simulation of electron devices.

Paolo Lugli is professor of optoelectronics at the 2nd University of Rome and focuses on semiconductor device modeling and ultrafast phenomena.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 99–103
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

3D Parallel Finite Element Simulation of In-Cell Breakdown in Lateral-Channel IGBTs

A. R. BROWN*, A. ASENOV and J. R. BARKER

Device Modelling Group, Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow G12 8LT, Scotland, UK

In this paper we describe the use of 3D parallel finite element solution of the Poisson equation to calculate the in-cell breakdown voltage in lateral channel IGBTs. The solver is based on topologically rectangular grids, and uses a domain decomposition approach to partition the problem on an array of mesh connected processors. A parallel BiCGSTAB solver has been developed to solve the Possion equation. Hole and electron ionisation integrals are calculated to determine the breakdown voltage. The effects of varying the doping concentration in the n^- base region and stopper surface concentration are investigated.

Keywords: 3D simulation, parallel, finite element, breakdown, IGBT

INTRODUCTION

Insulated Gate Bipolar Transistors (IGBTs) [1] are among the leading devices in the power semiconductor market, combining a low on-state voltage typical of bipolar devices, with MOS gated switching. Modern IGBTs have a cellular structure such as that shown in Figure 1. The optimisation of the cell design is an ongoing issue. The current density, for example, can be increased by reducing the cell-to-cell separation. Unfortunately in lateral channel IGBTs the minimum cell-to-cell separation is restricted by cell-to-cell pinch-off which limits the current density. Although vertical-channel non pinch-off devices have been demonstrated [2] the lateral channel IGBTs with stopper implantation are still the preferred technological choice. The stopper, however, should not compromise the overall device breakdown. The calculation of the in-cell breakdown in the stopper design process requires an accurate 3D solution of the Poisson equation [3] with fine resolution around the metallurgical *p-n* junction.

In this paper we describe the calculation of the in-cell IGBT breakdown using a parallel finite element 3D solution of the Poisson equation on an array of processors. The 3D discretisation is based on a topologically rectangular finite element (FE) grid which facilitates the partitioning of the domain over arrays of processors. Results for the

* Corresponding author. Tel.: +44 141 330 4792. Fax: +44 141 330 4907. E-mail: A. Brown@elec.gla.ac.uk.

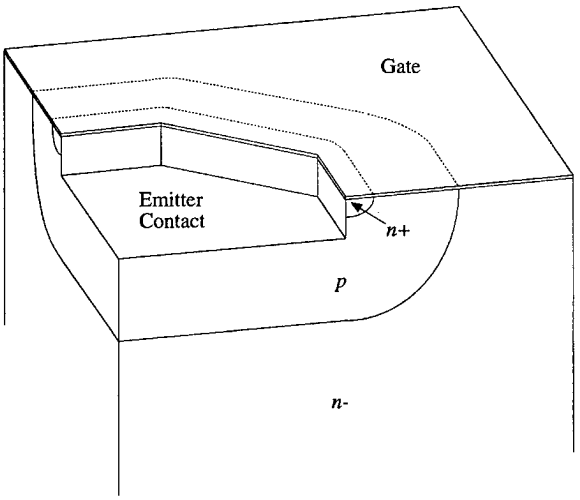


FIGURE 1 Schematic view of the top of an octagonal cell IGBT.

dependence of the breakdown voltage on the device geometry and the doping profile are presented.

PARALLEL IMPLEMENTATION

The Poisson solver used to study the in-cell IGBT breakdown is part of a general purpose device simulator in development [4]. This simulator is designed for parallel MIMD architectures based on a 2D array of mesh connected processors. The current version runs on an 8 node PowerPC Parsytec X-plorer under Parix. This provides the necessary computing power for computationally intensive 3D simulation.

The parallel implementation is based on the spatial domain decomposition approach, whereby the solution domain is partitioned across the 2D array of processors. Each processor stores information and updates the solution only for the nodes which belong to its subdomain. The partitioning is greatly simplified by the use of a topologically rectangular grid which is described later. Both a parallel 4-colour block Newton SOR solver, and a parallel BiCGSTAB solver have been developed and used for the solving the Poisson equation. While the SOR solver is efficient and works well at low applied voltage, convergence problems at the high voltages require for breakdown calculations were encountered. The BiCGSTAB solver in combination with Newton iterations, and appropriate damping, works well in the high voltage range.

Due to the large problem size, there is not sufficient memory to run the program on an individual processor of the Parsytec system and to have a proper speed-up assesment. However the speed-up obtained by running on the 8×90 MHz PPC601 Parsytec system, compared to a single 170 MHz UltraSparc Workstation is approximately 2.25.

GRID GENERATION

In order to obtain the required accuracy in the solution of the Poisson equation it is essential to have a finite element grid which conforms to the shape of the metallurgical *p-n* junctions and to the cell topology. The use of unstructured finite element grids can present serious problems when the parallel implementation is based on domain decomposition, and can lead to very low efficiency of the parallel code. To avoid these difficulties we have developed a topologically rectangular FE grid [5].

The structured topologically rectangular grid retains the number of grid nodes in any particular index direction. This makes the partitioning of the grid over an array of mesh connected processors a straight forward procedure. The partitioning is done by dividing the number of grid nodes in each index direction by the number of processors available in the processor array in the same direction [6]. For such partitioning only nearest-neighbour communications are required in the design of iterative solvers, because adjacent nodes in adjacent partition subdomains appear on neighbouring processors.

We have implemented two different approaches for the generation of topologically rectangular

grids in the discretisation of the IGBT cell (Fig. 2). The most straightforward method is to deform a Cartesian grid to fit key contours in the device structure, such as the *p-n* junctions. This method was used in the generation of the grid in the vertical walls of the IGBT cell. The deformation algorithm has been designed to be as general as possible to allow the use of arbitrarily shaped defining contours described in a parameterised form. This allows the generation of complex grids. A different method was used in the design of the grid for the top plane of the IGBT. This method again uses key contours such as the emitter contact edge and surface *p-n* junctions. These contours are used to generated additional guiding contours, the spacing of which controls the grid density. The correct number of nodes are placed on each guiding contour to satisfy the criteria for a topologically rectangular grid. This method is more flexible, but has the restriction that the grid must have equal numbers of nodes in each direction.

The full 3D grid is generated by extending the grid for the top surface vertically, and then deforming this 3D grid by the deformations calculated for the side walls of the cell. Figure 3 shows the full 3D grid for the region down to the p-n^- junction, clearly illustrating the internal curvature of the grid, and the shape of the *p-n* junction itself.

BREAKDOWN CALCULATION AND RESULTS

The calculation of breakdown voltage is based on the evaluation of the ionisation integrals for holes and electrons. From the solution of the Poisson equation the electric field in each element is calculated, and the point of maximum electric field strength is determined. The integration path for the ionisation integrals passes through this point and is traced along the electric field vector in both directions until either a boundary is encountered, or the electric field strength becomes negligible. The ionisation integrals along this path are then calculated. The hole ionisation integral was found to be more sensitive to changes in applied voltage and was used in the procedure for tracing the breakdown voltage. However we check that both hole and electron integrals become unity at the breakdown voltage.

The point of highest electric field occurs near the corner of the p-n^- junction which corresponds to

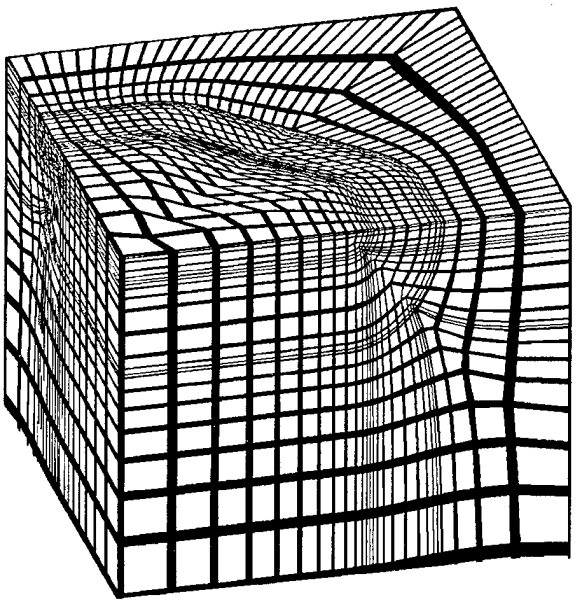


FIGURE 2 Topologically rectangular 3D grid showing top and sides of the 3D grids.

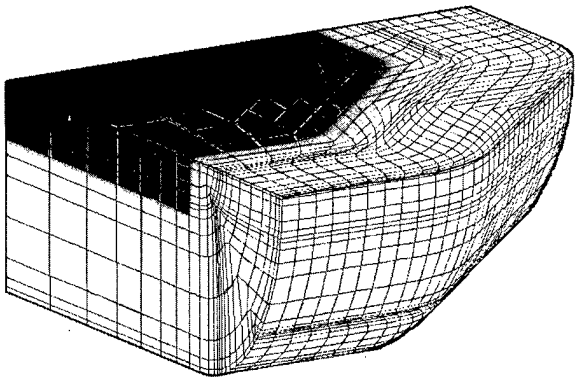


FIGURE 3 The grid at the metallurgical p-n^- junction showing junction curvature.

the corner of the octagonal emitter contact. This is the region where the junction exhibits the highest curvature. The effect of the doping in the n^- base region on the breakdown voltage is shown in Figure 4. The in-cell breakdown is compared to the breakdown voltage for a one-sided planar junction. As expected, the increasing doping concentration of the base region reduces the breakdown voltage significantly. Although the smaller inter-cell distance and the gate potential significantly reduce the curvature of the depletion layer edge, and the strength of the fringing field, the IGBT cell shows a noticeable reduction in breakdown voltage compared to the planar junction.

The introduction of a surface stopper implantation reduces the channel length, prevents the pinch-off of the vertical JFET, and lowers the on-state voltage drop. The stopper concentration, however, affects the in-cell breakdown. Figure 5 shows the effect that increasing the surface concentration of the implantation has on the breakdown voltage of a typical IGBT rated at 600 V. For a surface concentration above $10^{16}\,cm^{-3}$ the breakdown voltage falls sharply and the stopper concentration limits the blocking capabilities of the device.

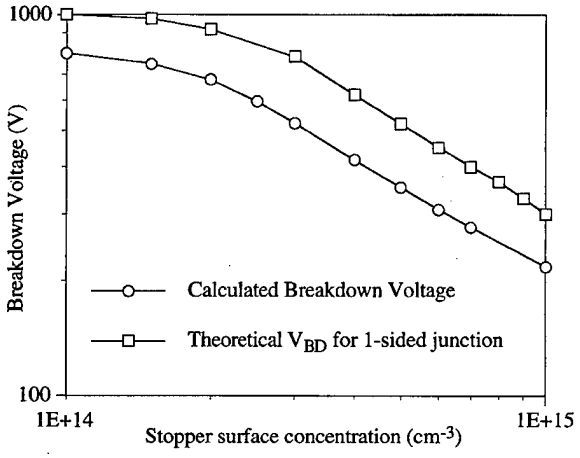


FIGURE 4 Breakdown voltage vs n^- base region doping concentration.

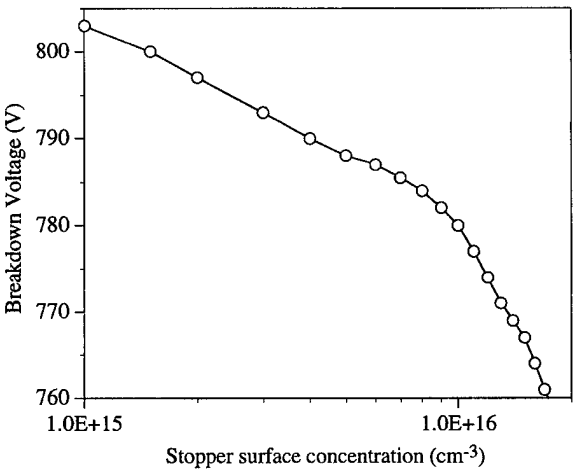


FIGURE 5 Breakdown voltage vs stopper implant surface concentration.

CONCLUSIONS

In this paper we have described 3D parallel finite element simulation of in-cell breakdown voltages in cellular IGBTs. The simulator is based on a topologically rectangular grid which reflects the geometry of the device, and simplifies the partitioning of the solution domain in the domain decomposition based parallel implementation. The results of the breakdown calculations illustrate the importance of realistic 3D simulation in the design of cellular IGBTs.

References

[1] Baliga, B. J., Adler, M. S., Love, R. P., Gray, P. V. and Zommer, N. D. (1984). "The Insulated Gate Transistor: A New Three-Terminal MOS-Controlled Bipolar Power Device", *IEEE Trans. Electron Dev.*, **ED-31**(6), 821–828.

[2] Baliga, B. J. (1996). "Trends in Power Semiconductor Devices", *IEEE Trans. Electron Dev.*, **43** (10), 1717–1731.

[3] Asenov, A., Barker, J. R., Brown, A. R. and Lee, G. L. (1996). "Scalable Parallel 3D Finite Element Nonlinear Poisson Solver", *Journal of Simulation Practice and Theory*, **4**, 155–168.

[4] Brown, A. R., Asenov, A., Roy, S. and Barker, J. R. (1995). "Parallel 3D Finite Element Power Semiconductor Device Simulator Based on Topologically Rectangular Grid", *Simulation of Semiconductor Devices and Processes*,

6, Eds. H. Ryssel and P. Pichler, Springer Verlag, 336–339.

[5] Asenov, A., Brown, A. R., Roy, S. and Barker, J. R. "Topologically Rectangular Grids in the Parallel Simulation of Semiconductor Devices", *VLSI Design*, in press.

[6] Asenov, A. M., Reid, D. and Barker, J. R. (1995). "Speed-up of Scalable Iterative Linear Solvers Implemented on an Array of Transputers", *Parallel Computing.*, **21**, 669–682.

Authors' Biographies

Andrew Brown is a research assistant in the Department of Electronics and Electrical Engineering at the University of Glasgow. He received his B.Eng. (Hons) in Electronics and Electrical Engineering from the University of Glasgow in 1992. He is currently working on the 3D parallel finite element simulation of Insulated Gate Bipolar Transistors. His research interests include device modelling, parallel computing, and visualisation.

Asen Asenov had 10 years industrial experience as a head of the Process and Device Modelling Group in IME-Sofia, developing one of the first integrated process and device CMOS simulators IMPEDANCE. He was visiting professor at the Physics Department of TU Munich, and is currently a Reader in the Department of Electronics and Electrical Engineering, Glasgow University. As a leader of the Device Modelling Group he has contributed to the development of 2D and 3D device simulators and their application in the design of FETs, SiGe MOSFETs and IGBTs. He also investigates the design of parallel algorithms.

John Barker is Professor of Electronics in the Department of Electronics and Electrical Engineering. He has a long standing interest in computational methods, device modelling and transport theory. From 1970–85 he was a member of the Theory group in the Dept. of Physics, University of Warwick, aside from 1978–79 when he worked at IBM T. J. Watson Laboratory, North Texas State University and Colorado State University. From 1987–89 he was academic director of the IBM UK/Glasgow University Kelvin Project on Numerically Intensive Parallel Computing. He is academic director of the Parallel Processing Centre at the University of Glasgow.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 105–109
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Comparison of Iteration Schemes for the Solution of the Multidimensional Schrödinger-Poisson Equations

A. TRELLAKIS[a,*], A. T. GALICK[a], A. PACELLI[b] and U. RAVAIOLI[a]

[a] *Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801;*
[b] *Dipartimento di Elettronica e Informazione and CEQSE-CNR, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy*

We present a fast and robust iterative method for obtaining self-consistent solutions to the coupled system of Schrödinger's and Poisson's equations in quantum structures. A simple expression describing the dependence of the quantum electron density on the electrostatic potential is used to implement a predictor–corrector type iteration scheme for the solution of the coupled system of differential equations. This approach simplifies the software implementation of the nonlinear problem, and provides excellent convergence speed and stability. We demonstrate the algorithm by presenting an example for the calculation of the two-dimensional bound electron states within the cross-section of a GaAs-AlGaAs based quantum wire. For this example, six times fewer iterations are needed when our predictor–corrector approach is applied, compared to a corresponding underrelaxation algorithm.

Keywords: Semiconductor, simulation, quantum wire, iteration, predictor–corrector, Schrödinger-Poisson

1 INTRODUCTION

As electronic device dimensions approach nanometer scale, quantum effects are expected to dominate their electronic properties. Consequently there is considerable interest in the efficient numerical simulation of such structures, not only for exploring novel device architectures but also for maintaining reliability for reduced present-day devices. In this paper we present a fast and robust iterative method for solving the two-dimensional Schrödinger-Poisson equations and obtaining the electron states in the cross-section of a quantum wire, and we compare the efficiency of our method with the standard underrelaxation algorithm.

The physical model used to describe the quantum wire consists, in the effective mass approximation, of Schrödinger's equation

$$-\frac{\hbar^2}{2}\nabla\left[\frac{1}{m^*}\nabla\psi_l\right] + [V_h - e\phi + V_{xc}(n) - E_l]\psi_l = 0 \tag{1}$$

* Corresponding author. Computational Electronics Group, 3223 Beckman Institute 405 N. Mathews Urbana, IL 61801, USA. E-mail: trell@ceg.uiuc.edu. Tel.: (217)-244-1919. Fax: (217)-244-4333.

coupled with a nonlinear Poisson equation

$$\nabla(\varepsilon\nabla\phi) = -e(-n + p + N_D^+ - N_A^-), \quad (2)$$

where the unknowns are the wavefunctions ψ_l belonging to energy levels E_l, and the electrostatic potential ϕ. Here, m^* is the electron effective mass, ε the dielectric constant, n and p are the electron and hole concentrations, N_D^+ and N_A^- the ionized donor and acceptor concentrations, V_h the heterojunction step potential, and finally V_{xc} the exchange correlation potential in the local density approximation.

The quantum electron density n is obtained from the eigenpairs (E_l, ψ_l) of Schrödinger's equation as

$$n = \sum_l N_l |\psi_l|^2, \quad (3)$$

where

$$N_l = \left(\frac{2m^* k_B T}{\pi^2 \hbar^2}\right)^{1/2} \mathcal{F}_{-1/2}\left(\frac{E_F - E_l}{k_B T}\right) \quad (4)$$

denotes the occupancy of the l-th eigenstate, and T is the Temperature, k_B Boltzmann's constant, E_F the Fermi level, and $\mathcal{F}_{-1/2}$ the complete Fermi-Dirac integral of order $-1/2$.

We avoid the expense of solving Schrödinger's equation in a very large grid by using the quantum formulation for n (3) only within the region of the quantum wire itself, while in the rest of the simulation domain, where the electron density is small, we replace it by its classical expression. The hole concentration p is very small throughout the quantum wire, but needs to be included into Poisson's equation to ensure a correct description of the band bending in the substrate [1].

2 SOLUTION BY UNDERRELAXATION

This system of coupled nonlinear differential equations is usually solved by an iteration between Poisson's and Schrödinger's equation [2, 3]. Since a plain iteration by itself does not converge, it is necessary to underrelax in the electron density n using an adaptively determined relaxation parameter $\omega^{(k)}$. This underrelaxation approach can be briefly outlined as follows:

1. Solve nonlinear Poisson equation using the old electron density $n^{(k-1)}$ to obtain electrostatic potential $\phi^{(k)}$.
2. Solve Schrödinger's equation using $\phi^{(k)}$ and $V_{xc}(n^{(k-1)}$ to obtain a new set of eigenpairs $(E_l^{(k)}, \psi_l^{(k)})$.
3. Calculate an intermediate electron density $n_{int}^{(k)}$.
4. Choose an appropriate relaxation parameter $\omega^{(k)}$ to obtain new density $n^{(k)}$, and repeat outer iteration until n becomes stationary.

The weakness of this method is the inherent instability of the outer iteration which is controlled by the underrelaxation procedure only. The relaxation parameter $\omega^{(k)}$ is not known in advance and needs to be dynamically readjusted during the course of the iteration. If $\omega^{(k)}$ is too large, the total quantized charge $\int n\,dx$ oscillates from one iteration step to the other without reaching convergence, while if $\omega^{(k)}$ is too small, too many iteration steps are necessary to achieve convergence.

3 SOLUTION BY A PREDICTOR–CORRECTOR TYPE APPROACH

To address this problem we have to modify the underrelaxation algorithm in a way that partially decouples both partial differential equations and damps the oscillations in the total electric charge. We can achieve this goal by incorporating a modified expression for the quantum electron density $\tilde{n}(\phi)$ into Poisson's equation, which approximates the implicit dependency of the quantum electron density n on the electrostatic potential ϕ due to Schrödinger's equation.

A suitable expression of this type is provided by quantum mechanical perturbation theory. Using its formalism one can show [4, 5], that for a small perturbation of the electrostatic potential $\delta\phi$ the

occupancies N_i within the electron density (3) become ϕ-dependent as

$$N_l(\delta\phi) \approx \tilde{N}_l(\delta\phi) \tag{5}$$

$$= \left(\frac{2m*k_BT}{\pi^2\hbar^2}\right)^{1/2} \mathcal{F}_{-1/2}\left(\frac{E_F - E_l + e\delta\phi}{k_BT}\right). \tag{6}$$

In our predictor–corrector approach we utilize this result by applying it to arbitrary large changes in ϕ:

We start the k-th iteration step by solving a modified Poisson equation

$$\nabla(\varepsilon\nabla\phi^{(k)}) = e[\tilde{n}(\phi^{(k)}) - p - N_D^+ + N_A^-], \tag{7}$$

where the potential-independent quantum electron density n is replaced by the potential dependent predictor $\tilde{n}$,

$$\tilde{n}(\phi^{(k)}) = \sum_l \tilde{N}_l^{(k-1)} \left(\phi^{(k)} - \phi^{(k-1)}\right) |\psi_l^{(k-1)}|^2, \tag{8}$$

with occupancies $\tilde{N}_l^{(k-1)}$ defined as in (6) and the superscript $k-1$ indicating the use of $E_l^{(k-1)}$. The electrostatic potential $\phi^{(k)}$ obtained from solving Poisson's equation (7) together with the predicted value for the electron density $\tilde{n}(\phi^{(k)})$ is then used within Schrödinger's equation (corrector)

$$-\frac{\hbar^2}{2}\nabla\left(\frac{1}{m^*}\nabla\psi_l^{(k)}\right) + (V_h - e\phi^{(k)} + V_{xc}(\tilde{n}^{(k)}) - E_l^{(k)}\Big)\psi_l^{(k)} = 0 \tag{9}$$

to calculate a corrected update $n^{(k)}$ to the exact electron density

$$n^{(k)} = \sum_l N_l^{(k)} |\psi_l^{(k)}|^2. \tag{10}$$

It is essential to insert the latest predictor value $\tilde{n}^{(k)}$ into the exchange correlation in Schrödinger's equation (9), since the numerical experiment shows that the use of $n^{(k-1)}$ within V_{xc} is detrimental to convergence. The use of $\tilde{n}^{(k)}$ also follows the rule that the latest predictor value should enter the corrector.

The numerical experiment also shows, that, unlike in the standard algorithm, convergence is achieved without underrelaxation, so the outer iteration can be reduced to a simple alternation between Poisson's and Schrödinger's equation until the quantum electron density n becomes stationary:

$$\text{Residual}(n) = \|n^{(k)} - n^{(k-1)}\|_2 \leq \varepsilon. \tag{11}$$

4 NUMERICAL RESULTS

We compared the efficiency of both iteration schemes by computing the bound states for the model device shown in Figure 1 [2]. The quantum wire is based on a GaAs-AlGaAs modulation doped structure, and delimited by two metal gates with a distance of 400 nm. The gate voltages were chosen as $V_G = 1.3$ V with respect to the substrate for both contacts; additionally we assumed a Schottky barrier of approximately 1 eV between metal and semiconductor material. The Al concentration was set equal to 26% throughout the AlGaAs layer, and we also included a fixed surface charge density of 1.6×10^{12} cm^{-2} at the GaAs-air interface. All calculations were done at a temperature of $T = 4.2$ K.

Both Schrödinger's equation as well as Poisson's equation are discretized by a box integration finite difference method to take material discontinuities into account. Since the quantum wire covers only a small part of the entire simulation domain, we employ a non-uniform rectangular mesh concentrated around the wire region to minimize computational cost while retaining high accuracy within the region of interest. For the model device discussed above a 99×88 grid was found to yield satisfactory spatial resolution and short program run-times on HP100 workstations.

After discretization, Schrödinger's equation becomes a sparse eigenvalue problem, which is

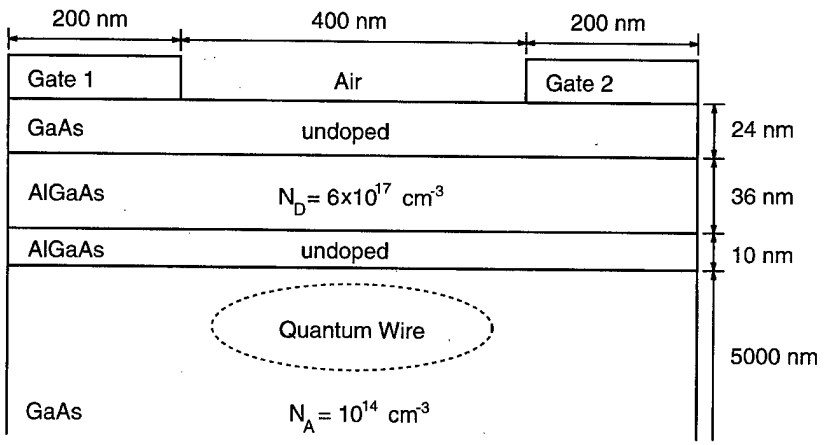


FIGURE 1 The model device structure used in the calculations.

solved by Chebyshev-Arnoldi iteration [6]. This method allows us to compute only the physically relevant lowest-energy states. The nonlinear Poisson equation is solved by Newton-Raphson iteration with inexact line search. The necessary solution of a sparse linear system at each iteration step is accomplished by a version of the preconditioned conjugate gradient method involving red-black reordering. The Fermi-Dirac integrals, which need to be evaluated repeatedly at each grid point, are computed using rational function approximations, which combine speed of execution with high accuracy [4, 7].

We compare the convergence speed of the predictor-corrector method with a fast adaptive underrelaxation scheme (an adaptive nonlinear version of the standard Gauss-Seidel algorithm) developed by one of the authors [6], which uses a heuristic method to adjust the relaxation parameter $\omega^{(k)}$. Using the residual in the quantum electron density (11) as measure to quantify convergence we find that our predictor–corrector approach makes the outer iteration stable. In fact, the residual decreases by close to one order of magnitude each step (Fig. 2), which is a 6-fold increase in convergence speed compared to underrelaxation. Additionally we find that the residual decreases uniformly from one step to the next,

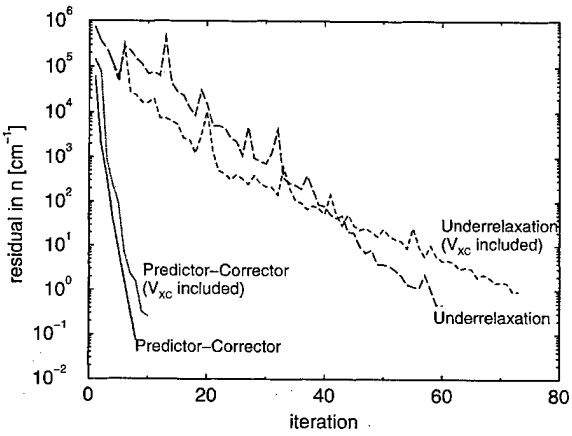


FIGURE 2 Residual of the quantum electron density using a 99×88 grid.

while for the heuristic underrelaxation algorithm occasional increases are possible.

The inclusion of exchange correlation V_{xc} results in an additional slowdown of convergence for the underrelaxation scheme and also decreases its stability. However, our predictor–corrector method remains stable and efficient, if the last available predictor value $\tilde{n}^{(k+1)}$ is inserted into V_{xc} as outlined above.

We verified the rapid uniform convergence of our predictor–corrector method for a wide range of temperatures, gate voltages and geometries, and

found it in all cases clearly superior to adaptive underrelaxation. This result indicates a fairly wide range of applicability, and considering the ease of implementation of our method we strongly recommend its use even for existing codes.

Acknowledgements

This work was supported by the National Science Foundation grants ECS 95-09751 and ECD 89-43166, Nato Collaborative Grant CRG.950753, and by a graduate fellowship (A.T.) of the Computational Science and Engineering program at the University of Illinois.

References

[1] Kerkhoven, T., Raschke, M. W. and Ravaioli, U. " Self-consistent simulation of quantum wires in periodic heterojunction structures", *J. Appl. Phys.*, **74**(2), 1199–1204, July 1993.

[2] Laux, S. E. "Numerical Methods for Calculating Self-Consistent Solutions of Electron States in Narrow Channels", *Proceedings of the Fifth International NASECODE Conference*, Boole Press, Dublin, Ireland, 1987, pp. 270–275.

[3] Kerkhoven, T., Galick, A. T., Ravaioli, U., Ahrends, J. H. and Saad, Y. "Efficient numerical simulation of electron states in quantum wires", *J. Appl. Phys.*, **68**(7), 3461–3469, October 1990.

[4] Trellakis, A., Galick, A. T., Pacelli, A. and Ravaioli, U. "A new iteration scheme for the solution of the two-dimensional Schrödinger-Poisson equations in quantum structures", *J. Appl. Phys.*, **81**(12), 7880–7884, June 1997.

[5] Pacelli, A. "Self-consistent solution of the Schrödinger equation in quantum wells by implicit iteration", *IEEE Trans. Electron Devices*, **44**(7), 1169–1171, July 1997.

[6] Galick, A. (1993). "Efficient solution of large sparse eigenvalue problems in microelectronic simulation", *Ph.D. Thesis*, University of Illinois at Urbana-Champaign, United States.

[7] Trellakis, A., Galick, A. T. and Ravaioli, U. (1997). "Rational Chebyshev Approximation for the Fermi-Dirac Integral $\mathcal{F}_{-3/2}(x)$", *Solid-State Electron.*, **41**(5) 771–773.

Authors' Biographies

Umberto Ravaioli is a professor in the Department of Electrical and Computer Engineering of the University of Illinois at Urbana-Champaign. His fields of professional interest are semiconductor device physics and simulation, numerical methods, and high performance computing.

Andrea Pacelli was a visiting scholar with the Beckman Institute of the University of Illinois at Urbana-Champaign, and he is now a Ph.D. student at the Politecnico di Milano, Italy. His research interests include the physics, modeling and characterization of MOS devices, hot-electron effects and Monte Carlo simulation.

Albert T. Galick is a Postdoctoral Researcher in the Computational Electronics Group in the Beckman Institute of the University of Illinois at Urbana-Champaign. His research focuses on the solution of large scale eigenvalue problems as typical for the simulation of semiconductor devices and dielectric wave-guides.

Alexandros Trellakis is a Ph.D. student at the University of Illinois at Urbana-Champaign. His areas of interest include semiconductor device simulation, simulation of optical cavities, and large-scale simulation of quantum systems.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 111–115
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Cellular Automata Studies of Vertical Silicon Devices

M. SARANITI [a], G. ZANDLER [b], G. FORMICONE [a] and S. GOODNICK [a,*]

[a] *Electrical Engineering Department, Arizona State University, Tempe (AZ) 85287-6206;*
[b] *Physik Department and Walter Schottky Institut, TU-München, Garching, Germany*

We present systematic theoretical Cellular Automata (CA) studies of a novel nanometer scale Si device, namely vertically grown Metal Oxide Field Effect Transistors (MOSFET) with channel lengths between 65 and 120 nm. The CA simulations predict drain characteristics and output conductance as a function of gate length. The excellent agreement with available experimental data indicates a high quality oxide/semiconductor interface. Impact ionization is shown to be of minor importance. For inhomogeneous *p*-doping profiles along the channel, significantly improved drain current saturation is predicted.

Keywords: Cellular automaton, submicron MOSFET, Silicon, simulation, semiconductor device

1 INTRODUCTION

As a numerically very efficient discrete variant of the Monte Carlo (MC) technique [1], the CA [2] method significantly reduces the gap in computational speed between simulation techniques based on solving the full Boltzmann equation, and moment-based approaches such as the hydrodynamic method. The CA method is therefore particularly suited to simulate highly nonlinear charge transport, typical for the ultra-short devices described in this paper.

2 VERTICAL MOSFET

Molecular Beam Epitaxy (MBE) and Chemical Vapor Deposition (CVD) allow the realization of vertical devices with characteristic channel lengths well below 0.1 μm. Recently, CVD epitaxy was utilized to grow vertical MOSFET's with ultra short channels (see [3, 4]). The epitaxially grown layer structures are selectively etched to a depth of 0.8 μm and thermally oxidized to grow the SiO_2 gate dielectric with a thickness of 5 nm. The quality of the oxide is comparable to planar structures. Figure 1 shows the cross section of such a vertical MOS [4]. The simulated geometry is shown on the left side of the figure, and corresponds to one half of the real device.

Devices with different doping levels and channel length were fabricated. The doping concentration in the source and drain n^+ regions are in the range $10^{18}\,cm^{-3} \div 2 \times 10^{19}\,cm^{-3}$, while the *p*-buffer layer contains an acceptor concentration of

* Corresponding author: Tel.: (602)965 2030, Fax: (602)965 8118 e-mail: saraniti@asu.edu.

$2 \times 10^{18}\,cm^{-3}$. These devices were grown with an effective channel length varying from 65 to 170 nm.

To investigate the influence of nonlinear transport on the electrical characteristics of such ultra short devices, we performed systematic calculations for all the realized structures and compare our results to experimental data. We further explore the influence of inhomogeneous *p*-doping along the channel in improving the drain current saturation behavior.

2.1 Current-Voltage Curves

As a critical test for the nominal geometry settings like oxide thickness and *p*-doping level, we have compared the experimental and calculated sub-threshold behavior of the devices described above. Excellent agreement was obtained for several devices with different geometries and doping concentrations, which indicates that the nominal and technologically realized values of the geometry agree well with each other. This is exemplified in Figure 2 for the 65 nm CVD vertical MOSFET structure. There, the applied drain voltage is 0.5 V.

Based on these geometries, we investigated the drain current saturation behavior of the vertical MOSFET for different channel lengths. Figure 3 shows the output current versus drain voltage for three different channel lengths of 170 nm, 120 nm and 65 nm, respectively. The bias of the gate was fixed to the value of 2 V. Clear drain current saturation behavior is found even for the ultra-short, 65 nm device. The increasing output current

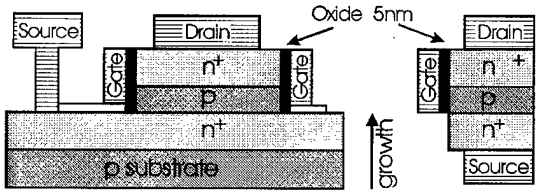


FIGURE 1 Layout of the vertical CVD-EPI MOS. The growth direction is shown, while the equivalent simulated region is on the right side.

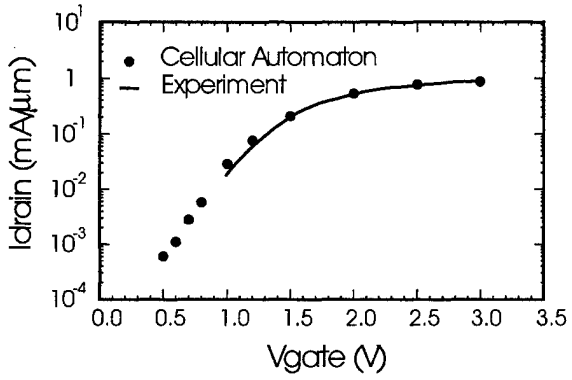


FIGURE 2 Sub-threshold output current of the simulated device versus gate voltage when a low ($V_D = 0.5$ V) drain bias is applied.

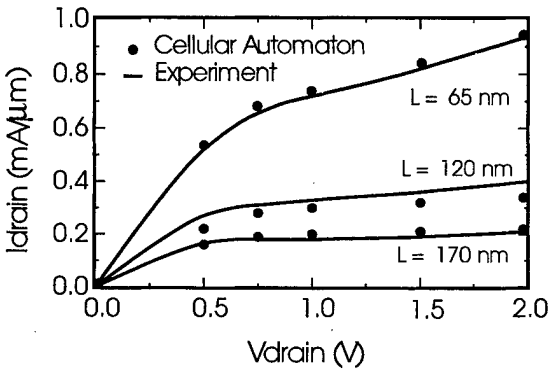


FIGURE 3 Computed (points) and experimental (full lines) current-voltage characteristics with decreasing channel length at $V_G = 2$ V gate bias.

and output conductance with decreasing channel length reflect a strong barrier reduction at the source junction. Nevertheless, even for high bias condition, no parasitic charge transport occurs in the bulk parallel to the conduction channel. The increase of drain current and output conductance due to the electrostatic effects mentioned above are further enhanced by velocity overshoot in the shorter channels. This velocity overshoot is driven by a high, inhomogeneous electric field along the channel. For the 65 nm n-p-n^+ structure biased at $V_D = V_G = 2$ V, we find a peak value of 800 kV/cm at the n^+p drain junction.

For this bias condition, the average electron velocity in the conduction channel of the 65 nm device is shown in Figure 4. Pronounced velocity overshoot is evident all along the conduction channel, the maximum velocity being reached at the drain end of the *p*-buffer, where it peaks to a value of three times the saturation velocity in homogeneous Si.

The overall agreement for the complete current–voltage characteristics between the CA results and experimental data is exemplified for the 70 nm device in Figure 5. For this device, simulations were performed up to a bias value of $V_D = 2.5$ V, to investigate the occurrence of impact ionization at higher bias. Up to the highest bias point shown, negligible contributions due to impact ionization were found. However, this process becomes important at higher drain bias. Analogous results were found for the 65 nm device.

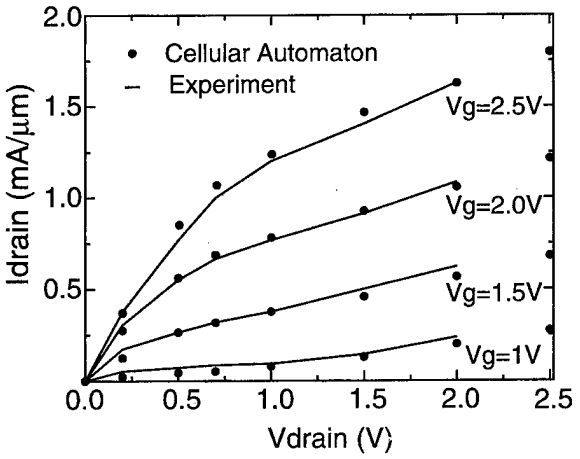


FIGURE 5 Computed (points) and experimental (full lines) current–voltage characteristics of the 70 nm vertical MOSFET.

2.2 Bulk Effects

The potential barriers due to the *np* (source) and *pn* (drain) junctions prevent a parasitic bulk current from flowing parallel to the channel inversion layer. At the same time, the electric field due to these junctions can reach high values, giving rise to impact ionization or even field breakdown. Optimization of such devices must account for the electrostatic behavior of the bulk region in order to improve saturation and avoid impact ionization at a given operating bias. Figure 6 (left) shows the concentration of carriers in the bulk 65 nm n^+pn system when a bias of 2 V is applied between the *n* and n^+ regions corresponding to the source and drain of the device shown in Figure 1. The fact that the potential barrier due to the *npn* junction (Fig. 6) is still intact at this drain bias ensures that no parasitic bulk current is flowing.

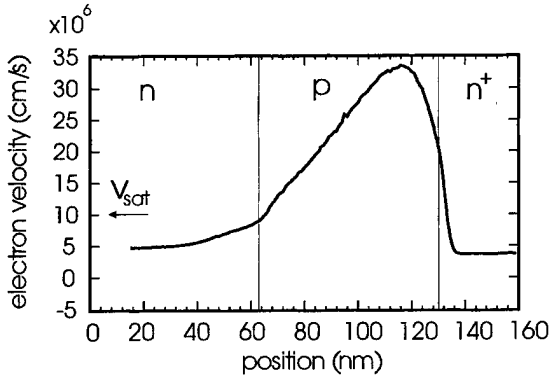


FIGURE 4 Average drift-velocity of electrons in the CVD-EPI MOSFET of Figure 1; a pronounced velocity overshoot is evident in the *p*-buffer.

2.3 Design Optimization

In order to improve the saturation behavior of the vertical MOSFET, an additional 20 nm p^+-buffer was introduced in the central *p*-region. The chosen

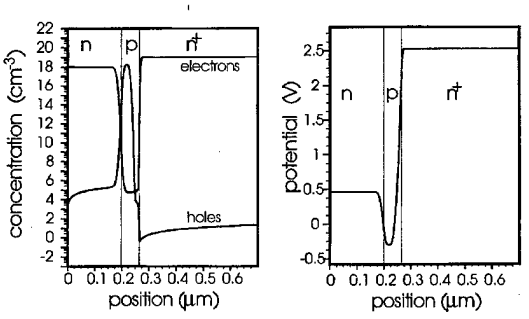


FIGURE 6 Carrier concentration (left) and potential profile (right) in the bulk *n-p-n* structure present in the vertical 65 nm MOSFET.

doping concentration in this p^+-buffer is $3\times10^{18}\,cm^{-3}$. The buffer was placed 10 nm from the *n-p* source junction. The buffer extension and position were optimized to reduce the influence of the drain potential on the output current, thus improving the saturation behavior. Its separation from the *n-p* source junction keeps the value of the junction field low enough to avoid impact ionization. Additionally, punch-through action is reduced due to the higher doping, which is crucial for devices with shorter channels. The drain current reduction due to the presence of the buffer was compensated by reducing the oxide thickness to 4 nm.

Simulation results for the device with the p^+-buffer are compared in Figure 8 with measurements made on a transistor with a homogeneously doped channel. The saturation behavior is remarkably improved compared to the characteristics shwon in Figure 5, and the reduced oxide thickness completely compensates the current reduction due to the p^+-buffer.

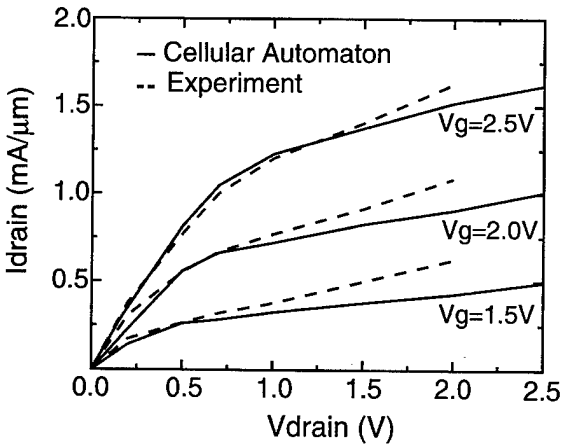


FIGURE 8 Improved saturation behavior of the simulated MOSFET after adding an additional 20 nm p^+ buffer. The simulation results (full lines) are compared with experimental data of Figure 5.

3 CONCLUSIONS

The capability of the CA approach to accurately predict highly nonlinear transport behavior and the resulting electrical characteristics in real nanostructured semiconductor devices was demonstrated. A novel family of vertical MOSFET's was investigated, giving results which are in excellent agreement with the experiments. The crucial role of the bulk parasitic conduction was stressed, and an optimized device design was proposed which improves the overall performance of the device.

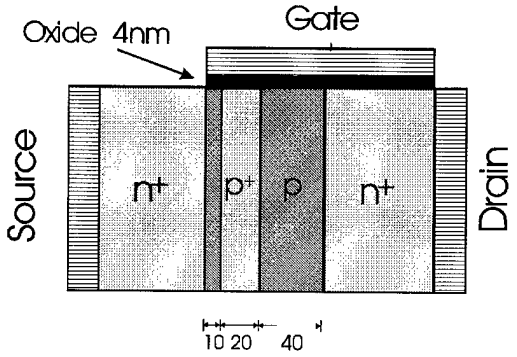


FIGURE 7 Cross-section of the saturation optimizing geometry. A $3\times10^{18}\,cm^{-3}$ p^+-buffer is included within the central *p* region.

Acknowledgements

The authors would like to thank Peter Vogl for his valuable suggestions. This work was partially supported by the NSF, grant ECS-9312240, and by the Siemens project SFE.

References

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

[2] Kometer, K., Zandler, G. and Vogl, P. (1992). "Lattice-gas cellular-automation method for semiclassical transport in semiconductors", *Physical Review B.*, **46**, 1382–1394, July.

[3] Hofmann, F., Krautschneider, W., Risch, L. and Schäfer, H. (1995). "CVD-EPI MOS transistor with a 65 nm vertical channel", in *Extended Abstracts of the 1995 International Conference on Solid State Devices and Materials*, Osaka, 46–48.

[4] Risch, L., Krautschneider, W., Hofmann, F., Schäfer, H., Aeugle, T. and Rösner, W. "Vertical MOS transistors with 70 nm channel length", *IEEE Transactions on Electron Devices*, **43**, 1495–1498, September 1996.

Authors' Biographies

Macro Saraniti received on April 1991 the Laurea degree in Physics from the University of Modena (Italy). In 1996 he received his Ph.D. in Physics from the Technical University of München (Germany) with a thesis on "Development of efficient numerical techniques for semiconductor device simulations". He is currently Faculty research associate of the Electrical Engineering Department of Arizona State University. His research interests are related to the numeric aspects of semiconductor devices simulation, as well as to the analysis and design of Si and GaAs ultra-short devices.

Günther Zandler received his Ph.D. from the University of Innsbruck, Austria, in 1989. In 1988, he became research associate at the Institute for Experimental Physics at the University of Innsbruck. In 1989 he joined the Physics Department of the Technical University of Munich as research associate for theoretical semiconductor physics. His research interests focuse on the simulation of electron devices, high field transport and ultrafast phenomena in bulk and nano-structured semiconductors.

Gabriele Formicone received the Laurea in Physics from the University of Roma (Italy) "La Sapienza" in June 1994, and the M.S. Degree in Electrical Engineering from Arizona State University in December 1996. He is also enrolled in the Ph.D. Degree at ASU, where he has been Graduate Research Assistant from June '94 to June '97. Currently he is Simulation Engineer at the Communication Products Laboratories, Motorola. His interests are devoted to the many aspects of physics and transport in submicron Si/SiGe devices and heterojunctions.

Stephen Goodnick received his B.S. degree in Engineering Science from Trinity University in 1977, and his M.S. and Ph.D. in Electrical Engineering from Colorado State University in 1979 and 1983 respectively. He was an Alexander von Humboldt Fellow at the Technical University of Munich, Germany and the University of Modena, Italy in 1985 and 1986. He was the Melchor visiting chair at the University of Notre Dame in 1991. He was a faculty member from 1986 to 1997 in the Department of Electrical and Computer Engineering at Oregon State University, Corvallis, Oregon, where he held the rank of Professor. He is presently Chair and Professor of Electrical Engineering at Arizona State University. Dr. Goodnick is a senior member of IEEE, a member of the American Physical Society, Sigma Xi, Eta Kappa Nu and Blue Key National Honor Society. He has co-authored over 70 journal articles and book chapters related to transport in semiconductor devices and microstructures.

HIGH PERFORMANCE COMPUTING

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 117–121
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

X3D Moving Grid Methods for Semiconductor Applications*

ANDREW KUPRAT†, DAVID CARTWRIGHT, J. TINKA GAMMEL, DENISE GEORGE, BRIAN KENDRICK, DAVID KILCREASE, HAROLD TREASE and ROBERT WALKER

Los Alamos National Laboratory

The Los Alamos 3D grid toolbox handles grid maintenance chores and provides access to a sophisticated set of optimization algorithms for unstructured grids. The application of these tools to semiconductor problems is illustrated in three examples: grain growth, topographic deposition and electrostatics. These examples demonstrate adaptive smoothing, front tracking, and automatic, adaptive refinement/derefinement.

Keywords: Adaptive mesh smoothing, multimaterial grids, moving grids, moving surfaces, unstructured grids, moving finite elements

INTRODUCTION

3D grain growth modeling and 3D topographic simulation have in common the requirement of accurately representing time dependent surface motion in a 3D volume. Problems involving fixed surfaces such as parasitic parameter extraction for interconnect modeling can benefit from adaptive grid methods because they substantially reduce solution error from iteration to iteration. The Los Alamos X3D grid toolbox provides a set of capabilities including initial grid generation of complex multimaterial geometries and grid optimization that preserves material interfaces. X3D data structures and toolbox methods are designed as objects in order to be user accessible and extensible. X3D commands can be issued from within an application driver program, and the example applications use this feature to perform, as needed, grid reconnection, node merging and smoothing. Additionally, all X3D data structures are available to the application driver via calls to utility routines. Thus when an application detects a non-routine event such as a topological change in a material region under deformation, a special purpose user routine can easily be incorporated into the system. These design features promote the separation of the physical based simulation from

*Work supported by the U.S. Department of Energy.
†Corresponding author.

the grid maintenance chores, but allow full access to the grid data structures when required.

Gradient Weighted Moving Finite Elements Applied to Metallic Grain Growth

In our first application involving the X3D toolbox [1], we use Gradient Weighted Moving Finite Elements [2] (GWMFE) to move a multiply-connected network of triangles to model the annealing of 3-D metallic grains. We assume evolution of grain interfaces obeys the simple equation

$$v_n = \mu K,$$

where v_n is the normal velocity of the interface, μ is the mobility, and K is the local mean curvature [3]. Gradient Weighted Moving Finite Elements minimizes

$$\int (v_n - \mu K)^2 \, dS$$

over all possible velocities of the interface vertices. (The integral is over the surface area of the interfaces.) This leads to system of $3N$ ODE's:

$$\mathbf{C}(\mathbf{y})\dot{\mathbf{y}} = \mathbf{g}(\mathbf{y}),$$

where $\mathbf{y}$ is the $3N$-vector containing the x, y, and z coordinates of all N interface vertices, $\mathbf{C}(\mathbf{y})$ is the matrix of inner products of finite element basis functions, and $\mathbf{g}(\mathbf{y})$ is the right-hand side of inner products involving surface curvature. The ODE's are integrated using an implicit variable time step integrator.

As an example, we evolve a 5 grain microstructure in the confined geometry of an aluminum interconnect on a semiconductor chip, and in Figure 1 we show the smooth surfaces of the grains at an intermediate time in their evolution under mean curvature. (The initial state for the time evolution was obtained using Monte Carlo annealing of a discrete effective model on a fixed lattice [4] and is not shown.) Visible in this exploded view

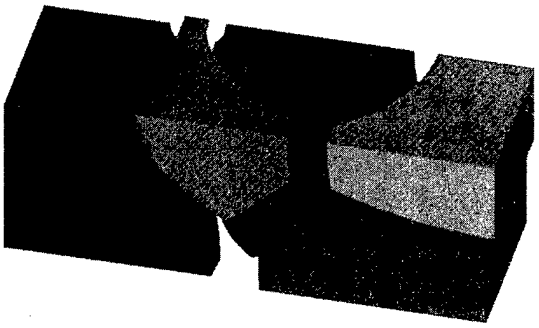


FIGURE 1 5 grain microstructure evolved under mean curvature motion using GWMFE.

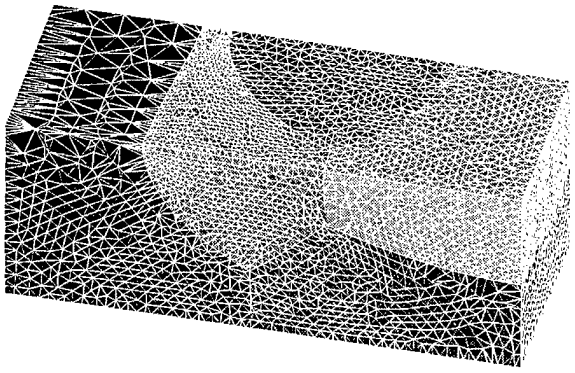


FIGURE 2 Evolved 5 grain microstructure showing surface grid. Evolution has caused imbalance in grid density.

are triple points on the surface of the interconnect, as well as a triple line and tetrahedral point in the interior. The corresponding surface grid is shown in Figure 2. It is clear that for this simulation to continue, some of triangles must be annihilated to prepare for topological events such as the "pinching off" of a grain. That is, a front tracking simulation must involve a nontrivial amount of grid manipulation in order to successfully run to completion. In Figure 3, we show the effect of **massage** which is a grid manipulation command in the X3D toolbox. As seen in the figure, the **massage** command has derefined the unstructured mesh without significantly damaging the shapes of the grains. Indeed, the command takes as user input a "damage" tolerance which gives the maximum acceptable amount of grain shape deformation allowable in the derefinement pro-

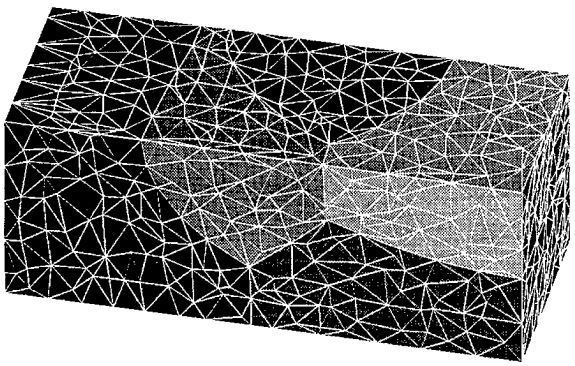


FIGURE 3 5 grain evolved grid after X3D **massage** command. The grid has been derefined and refined according to desired edge length criteria while preserving material interfaces and external boundaries.

cess. Also called is the **recon** command which allows changing of connectivity of the associated tetrahedral mesh in the interior of the volume. Still a work in progress, we anticipate that with the availability of these X3D grid manipulation tools we will be able to run the annealing process to steady-state. These microstructures will then provide high quality three-dimensional models for electromigration reliability simulation.

Finite Volume Electrostatic Calculations on a Solution Adapted Mesh

Problems involving fixed surfaces can also benefit from adaptive grids. To illustrate this, we show the effects of solution adaptive grid generation for calculation of electric fields in nontrivial geometry – as occurs when attempting to extract parasitic parameters for interconnect modeling. Figure 4 shows the interior of a box with a conical intrusion. We solve Laplace's equation on this three-dimensional domain using a finite volume solver. (Only the surface triangles are shown in the figure; the calculation is done on unshown tetrahedra that conform to the surfaces and fill the volume.) As is well known, the electric field becomes arbitrarily large near the tip of the cone. Displayed at the tip of the cone is an isosurface for the component of electric field aligned with the

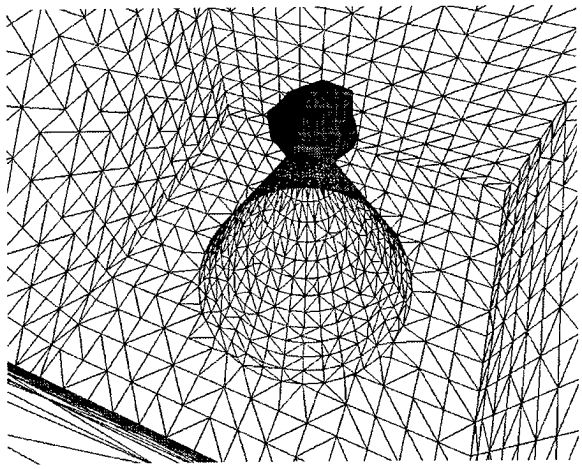


FIGURE 4 Surface triangles from an unadapted grid used to solve Laplace's equation on a unit cube with the bottom surface pierced by a sharp cone. Displayed at the tip of the cone is an isosurface of the component of the electric field aligned with the axis of the cone.

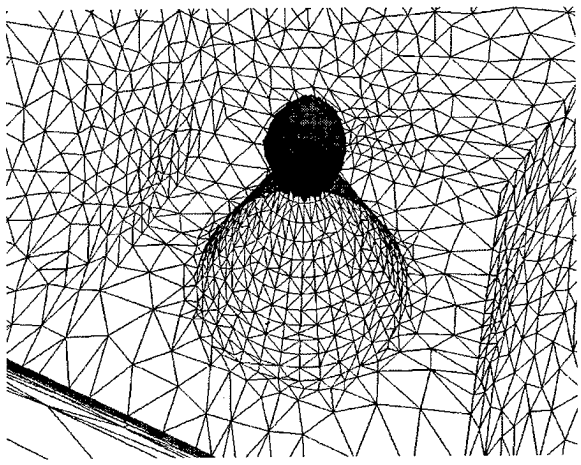


FIGURE 5 Surface triangles after error-dependent grid adaption. Note improvement in electric field isosurface. Adaption was turned off on the surface of the cone to prevent too much grid from disappearing into the tip of the cone where the electric field is infinite.

axis of the cone. As can be seen, this isosurface is nonsmooth due to lack of resolution in this area of highest solution error. In Figure 5 we show the same view and electric field component isosurface for the solution on a grid that has been adapted to an *a posteriori* error estimate. The isosurface is smoother, indicating that adaptive error-based

smoothing has successfully attracted the solution elements to the critical area near the tip of the cone. (However, to prevent the entire mesh from disappearing into the singularity at the tip of the cone, node movement on the surface of the cone was turned off.)

The solution adaption was accomplished using the X3D **smooth** command which adapts a mesh to minimize the $\mathcal{L}^2$ norm of the gradient of solution error, as calculated using an *a posteriori* error estimate based on estimated second derivatives of the solution [5]. In order for the finite volume method to work, a Delaunay grid is required. Thus, after each adaptive smoothing iteration, the X3D **recon** command is called to adjust the mesh connectivity to restore the Delaunay condition on the tetrahedral mesh. We note that the maximum electric field on the adapted mesh of Figure 5 is approximately ten times higher than that on the unadapted mesh of Figure 4. Thus, without increasing the number of nodes used, the solution adaption using X3D toolbox commands was automatically able to increase the quality of the solution in a critical area.

3D Topographic Simulation

In simulating topographic etch and deposition, TopoSim3D separates the chemical and physical processes from grid generation and maintenance operations. The flux calculation, source characterization and chemical reaction mechanisms have access to the needed X3D data structures and to the X3D geometry services. X3D is used to build the initial 3D tetrahedral mesh that represents the wafer and to accurately track the evolution of the material interfaces as they change with time. As material is deposited or etched away from interfaces, those tetrahedral faces (interface triangles) that lie on the boundary between two materials move in time. From a computation of the flux of materials arriving at the surface of the interface triangles, we move triangle nodes after apportioning the area of each triangle to each of the nodes it shares. The velocity of each node is then computed from an area-weighted vector average of the contributions from each of the triangles that share the node.

In low pressure simulations, the transport of material between the boundary plane and the surface of a wafer is represented by straight-line trajectories of particles. The flux of materials arriving at a triangle on the surface interface depends on the source emanation rate, the distance between the source and the surface element, the relative orientation of the source and the surface element, and the visibility of the surface as viewed from the source. In the simplest model, all the material arriving at the wafer surface stays there, implying that the only source elements are those that lie on the boundary plane. However, if some vapor phase materials arriving at the surface do not react there, they (or other species) will be re-emitted, and in such cases, all interface triangles serve as potential source elements. By specifying the surface chemistry (using the ChemKin and Surface ChemKin reaction software libraries), TopoSim3D incorporates the calculation of "sticking coefficients" in a natural way. If the rate of a surface reaction is slower on a surface element

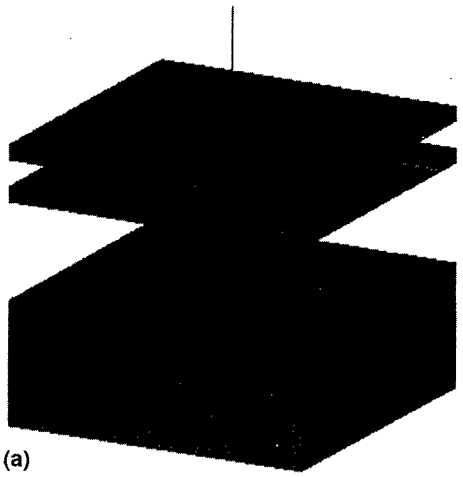


FIGURE 6 Topographic deposition onto two materials with different 'sticking coefficients'. View *a* shows the solid model of the wafer. View *b* gives the initial surface grid of the hole with the front surface of the hole removed. View *c* shows an intermediate time step in the deposition. View *d* shows the final time step in the deposition. Note the differences in upper and lower material 'sticking coefficients'.

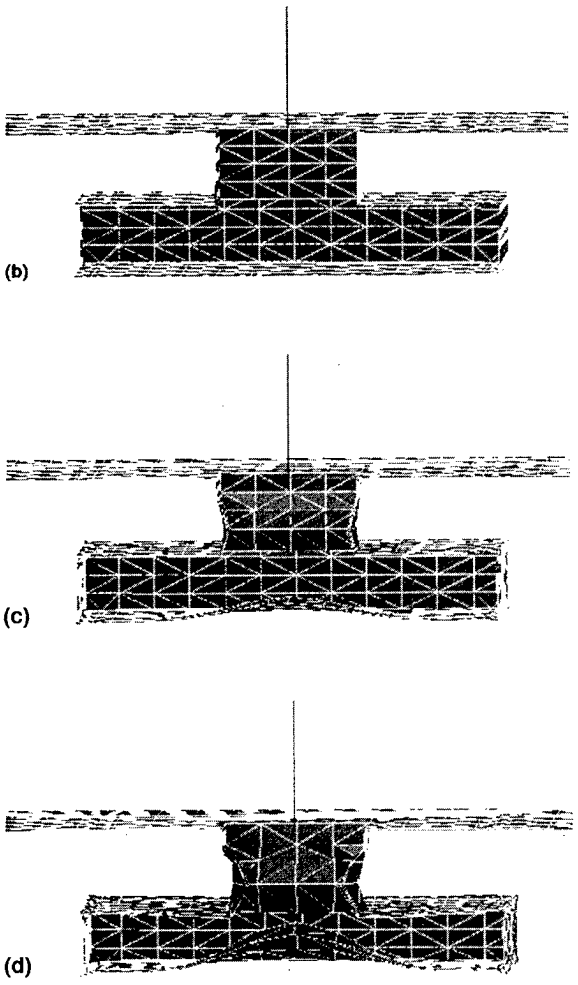


FIGURE 6 (Continued).

than the arrival rate of reactants, mass balance requires the excess material to be re-emitted from that interface triangle, which in turn becomes an effective source element for other surface elements.

As in grain growth modeling, the X3D grid maintenance calls are used to refine, reconnect and smooth the interface surface during the time evolution. TopoSim3D uses the X3D data structures to help avoid folding problems and detect topological events such as a pinch-off. Figure 6 shows deposition on an overhang structure with material-dependent sticking coefficients.

CONCLUSION

As shown in the three examples presented, X3D grid optimization techniques are well suited to solving time-dependent and geometrically challenging problems occurring in semiconductor applications.

References

[1] George, D. C. (1995). *X3D User's Manual*, Los Alamos National Lab Report, LA-UR-95-3608.
[2] Miller, K. (1997). *A Geometrical-Mechanical Interpretation of Gradient-Weighted Moving Finite Elements, SIAM J. Num. Anal.*, **34**, 67–90.
[3] Porter, D. A. and Easterling, K. E. (1988). *Phase Transformations in Metals and Alloys*, Great Britain: Van Nostrand Reinhold, pp. 130–136.
[4] Kuprat, A. and Gammel, J. T. *Modeling Metallic Microstructure Using Moving Finite Elements*, preprint, available at http://xxx.lanl.gov/abs/physics/9705041.
[5] Bank, R. E. and Smith, R. K. (1997). *Mesh smoothing using a posteriori error estimates, SIAM J. Num. Anal.*, **34**, 979–997.

Author Biography

Andrew P. Kuprat received his Ph.D. in Mathematics from the University of California at Berkeley and is currently a staff member at Los Alamos National Laboratory. He is interested in adaptive mesh algorithms applied to problems in semiconductor process and device modeling, as well as to other convection/diffusion problems in physics. He is also interested in mesh generation and computational geometry.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 123–128
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Multilevel Algorithms for Large-scope Molecular Dynamics Simulations of Nanostructures on Parallel Computers

AIICHIRO NAKANO*, RAJIV K. KALIA and PRIYA VASHISHTA

Concurrent Computing Laboratory for Materials Simulations, Department of Computer Science, Department of Physics and Astronomy Louisiana State University, Baton Rouge, LA 70803-4020

Molecular Dynamics (MD) is a powerful tool for the atomistic understanding of long-range stress-mediated phenomena, phonon properties, and mechanical failure of nanostructures. For realistic modeling of nanostructures, however, the scope of simulations must be extended to large system sizes, long simulated times, and complex realism. We have developed new multilevel algorithms and physical models encompassing multiple levels of abstraction: i) space-time multiresolution schemes; ii) adaptive curvilinear-coordinate load balancing; iii) hierarchical dynamics via a rigid-body/implicit-integration/normal-mode approach; iv) variable-charge MD based on electronegativity equalization; and v) multilevel preconditioned conjugate gradient method. Fuzzy clustering is used to facilitate the seamless integration of the multiple levels of abstraction.

Keywords: Molecular dynamics, parallel computing, multilevel algorithms, nanostructures

As new fabrication technologies for quantum nanostructures emerge (such as strain-induced self-organized growth [1] and substrate encoded size-reduced epitaxy [2]), there is growing need for hybrid atomistic/mesoscopic computer simulations. Molecular Dynamics (MD) [3] is a powerful tool for the atomistic understanding of long-range stress-mediated phenomena, phonon properties, and mechanical failure of nanostructures. For realistic modeling of nanostructures, however, the scope of simulations must be extended to larger system sizes, longer simulated times, and more complex realism than what has been feasible until recently. In this paper we describe various new multilevel algorithms for large-scale, long-time MD simulations.

In MD simulations, a system is represented by a set of atomic coordinates, $\{\mathbf{x}_i | \, i = 1, \ldots, N\}$, where N is the number of atoms. Time evolution of the system is governed by Newton's second law of motion [3],

$$m_i \frac{d^2\mathbf{x}_i}{dt^2} = \mathbf{g}_i(\mathbf{x}_i), \tag{1}$$

where m_i and $\mathbf{g}_i(\{\mathbf{x}_i\}) = -\partial V/\partial \mathbf{x}_i$ are the mass and force for the i-th atom. The potential energy

* Corresponding author.

function $V(\{\mathbf{x}_i\})$ consists of a sum over atomic pairs and triples [4].

The most prohibitive computational problem in **MD** simulations is associated with the Coulomb potential. Because of its long range, each atom interacts with all the other atoms in the system. Therefore the evaluation of the Coulomb potential requires $O(N^2)$ operations. An **MD** algorithm is developed based on multiresolutions in both space and time [5]. The long-range Coulomb interaction is computed with the Fast Multipole Method (**FMM**) [6, 7]. The **FMM** uses the truncated multipole expansion and local Taylor expansion for the Coulomb potential field (see Fig. 1). By computing both expansions recursively on a hierarchy of cells, the Coulomb potential is computed with $O(N)$ operations. Short- and medium-range non-Coulombic interactions are computed with the Multiple Time-Scale (**MTS**) approach [8, 9]. The **MTS** method is based on the fact that the farther the distance between particles the slower is the time variation of forces. Therefore different time steps are used to computer forces for different interparticle separations. To implement this MultiResolution Molecular Dynamics (**MRMD**) algorithm on parallel computers, we use spatial decomposition. Processors are logically organized as a cubic array of dimensions $\mathbf{P}_x \times \mathbf{P}_y \times \mathbf{P}_z$, and we partition the simulation system into subsystem into subsystems of equal volume accordingly (Fig. 2). For a 4.2 million-atom SiO_2 system, one **MD** step takes only 4.8 seconds on the 512-node Intel Touchstone Delta machine [5]. The memory-bound parallel efficiency [10] of the program is 0.92 and the communication overhead is 8%.

Simulation of nanostructures is often characterized by irregular atomic distribution. One practical problem in simulating such irregular systems on parallel computers is that of load imbalance [11]. Because of the irregular distribution of atoms, the uniform spatial decomposition results in unequal partition of workloads among processors. As a result the parallel efficiency is degraded significantly.

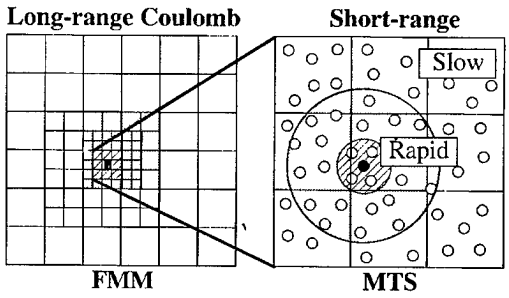


FIGURE 1 Schematic representation of spatial multiresolution for a two-dimensional system. (Left) A hierarchy of cells in the fast multipole method. (Right) The direct forces on a particle (solid circle) are due to the primary (open circles within the hatched area), secondary (open circles within the shaded area), and tertiary (the other open circles) neighbor atoms.

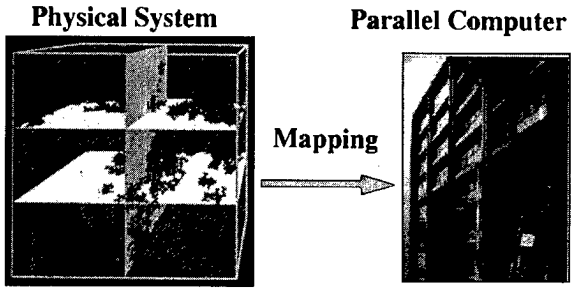


FIGURE 2 Regular spatial decomposition scheme for parallel computing. (Left) The physical system is divided into subsystems of equal volume. Spheres and planes represent atoms and subsystem boundaries, respectively. (Right) Each subsystem is mapped onto a computing node in a parallel computer.

To avoid this problem, a dynamic-load-balancing capability is added to the **MRMD** program [12]. The new load-balancing scheme introduces a curvilinear coordinate system [13], ξ, which is related to the atomic coordinate, $\mathbf{x}$, by a mapping,

$$\xi = \mathbf{x} + \sum_{Q} \mathbf{x}_Q \exp(i\mathbf{Q}\mathbf{x}). \qquad (2)$$

Workloads are partitioned with a uniform 3-dimensional mesh in the curvilinear coordinate system. The variational parameters $\{\mathbf{x}_Q\}$ are chosen to minimize the load-imbalance and communication costs. Simulated annealing is used to solve the optimization problem. For an irregular

nanocluster-assembled material, the load-balancing scheme sped up simulations by a factor of 4.2 [12].

Many important material processes (e.g., sintering and sol-gel processes) are characterized by time scales that are many orders-of-magnitude larger than atomic time scales (10^{-15} sec). A new algorithm is developed for large-scale, long-time **MD** simulations by combining a hierarchy of subdynamics (Fig. 3) [14]. Equation (1) is numerically integrated using a reference system $\mathbf{r}_i$, which represents two types of essential dynamics, i.e., global conformational changes and fast atomic oscillations. The reference system is thus defined as a superposition, $\mathbf{r}_i = \mathbf{r}_{0i} + \mathbf{r}_{hi}$, where $\mathbf{r}_{0i}$ and $\mathbf{r}_{hi}$ are the collective and harmonic parts, respectively.

The collective part of the reference system represents the rigid-body motion of clusters. We use the quaternion formulation of rigid-body dynamics in order to avoid the numerical singularity associated with angular coordinates [15]. A large time step (10^{-12} sec) is used for the numerical integration. The harmonic part $\mathbf{r}_{hi}$ of the reference system represents the fast oscillation of each atom around the local potential minimum, and its equation of motion can be integrated analytically in terms of trigonometric functions [16]. The residual system, defined as $\mathbf{z}_i = \mathbf{x}_i - \mathbf{r}_i$, is expected to vary slowly, since the rapidly oscillating harmonic motions have been subtracted. Therefore its equation is integrated by an implicit integration scheme using Δt which is much larger than atomic time scales [17]. The integration scheme is stable for an arbitrarily large Δt, and it is also symplectic [17]. Symplectic integrators preserve the phase-space volume, and this is essential for the long-time stability of orbitals.

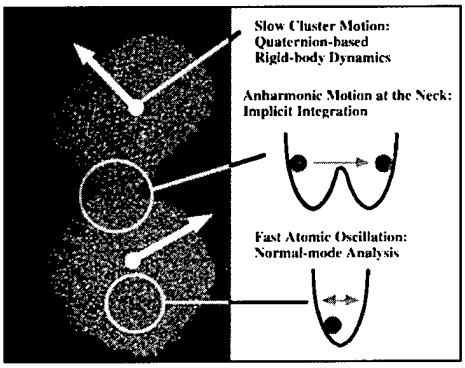


FIGURE 3 Various physical processes involved in the sintering of nanoclusters. i) Relative rotation of clusters is included by rigid-body dynamics with fuzzy clustering; ii) anharmonic atomic motions lead to surface diffusion and the growth of the neck between clusters, and these motions are included by implicit integration of Newton's equations; iii) thermal atomic motions assist the above diffusion process, and these high-frequency motions are dealt with through the normal-mode analysis.

The greatest challenge, however, is to integrate these heterogeneous abstraction levels into a seamless, unified scheme. To facilitate such integration, we find it useful to introduce the concept of fuzzy clustering [18, 19]. We introduce a membership function $P(i \in c)$ which describes the degree of association between atom i and cluster c. The principle of maximum entropy is used to determine $P(i \in c)$. The fuzzy-body/Implicit-integration/Normal-mode (**FIN**) scheme sped up a simulation of nanocluster sintering by a factor of 28 over a conventional explicit integration scheme, without loss of accuracy. A parallel implementation of the scheme achieves an efficiency of 0.94 for a 12.7 million-atom nanocrystalline solid on 64 nodes of an **IBM SP2** computer [14].

Conventional interatomic potential functions used in **MD** simulations are often fitted to bulk solid properties, and they are not transferable to systems containing defects, cracks, surfaces, and interfaces. In these systems, the partial charges on the atoms vary dynamically according to the change in the local environment. This environment-dependent charge distribution is crucial for the physical properties of these systems including the fracture toughness. Transferability of interatomic potentials is greatly enhanced by incorporating variable atomic charges which dynamically adapt to the local environment. Atomic charges can be determined by equalizing electronegativity [20].

However, the increased physical realism in the variable-charge **MD** is accompanied by increased computational cost for minimizing the electrostatic energy at every **MD** step. This minimization is equivalent to the electronegativity equalization condition that the chemical potentials be equal for all the atoms. This condition leads to a linear equation system for atomic charges, $\{q_i\}$:

$$\sum_j M_{ij} q_j = \mu - \chi_i, \tag{3}$$

where M_{ij} denotes the Coulomb-interaction matrix, χ_i is the electronegativity, and the Lagrange's multiplier μ is determined from the charge-neutrality constraint.

A Multilevel Preconditioned Conjugate-Gradient (**MPCG**) method is developed for this minimization problem by splitting the Coulomb-interaction matrix into short- and long-range components: $M = M_s + M_l$ [21]. The short-range matrix is the contributions from atomic pairs (i,j) within the nearest neighbor leaf cells used in the **FMM** [5–7]. The sparse short-range matrix M_s is used as a preconditioner to improve the spectral property of the linear system and thereby accelerating the convergence [22]. For α-Al_2O_3 crystal, the preconditioner reduces the execution time to achieve the same convergence level by 20% [21]. Numerical tests involving up to 26.5 million atoms are performed on an **IBM SP2** computer. Figure 4 shows the parallel efficiency (solid lines) and communication overhead (dashed lines) as a function of the number of atoms. The results with and without preconditioning are denoted by circles and squares, respectively. For the largest system, the preconditioning improves the parallel efficiency from 0.92 to 0.95 [21]. The communication overhead of the **MPCG** scheme is 5% of the total execution time for the largest system. The multilevel preconditioning scheme enhances the locality of computation by extensively using the short-range interaction matrix M_s, and consequently the program runs efficiently on parallel platforms.

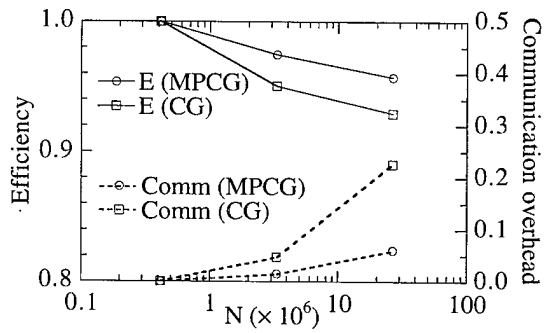


FIGURE 4 Memory-bound parallel efficiency of the Multilevel Preconditioned Conjugate Gradient (MPCG) program (solid lines) as a function of the number of atoms. Open circles and open squares are the results for the MPCG and the Conjugate Gradient (CG) methods, respectively. Communication overheads of the same program are shown by the dashed lines.

In summary, we have developed various parallel multilevel algorithms for multiscale phenomena in nanostructures. Using these algorithms, multi-million-atom **MD** simulations are being performed for: i) nanocluster-assembled Si_3N_4 [23, 24], SiO_2, SiC, and Al_2O_3; ii) Si/Si_3N_4 and Al/Al_2O_3 interfaces; and iii) GaAs stepped surfaces and mesas. Mechanical properties including fracture, long-range stress-mediated phenomena, and phonon properties in these nanostructures are being investigated.

Acknowledgements

This work was supported by Army Research Office (Grant No. DAAH04-96-1-0393), NSF (Grant No. DMR-9412965 and a CAREER Award, ASC-9701504), DOE (Grant No. DE-FG05-92ER45477), AFOSR (Grant No. F 49620-94-1-0444), USC-LSU Multidisciplinary University Research Initiative (Grant No. F 49620-95-1-0452), Petroleum Research Fund (Grant No. 31659-AC9), and Louisiana Education Quality Support Fund (Grant No. LEQSF(96-99)-RD-A-10 and DOD/LEQSF(96-99)-03). Computations were performed at Maui High Performance Computing Center (**MHPCC**), Argonne National

Laboratory, Caltech Concurrent Supercomputing Facility, and the Concurrent Computing Laboratory for Materials Simulations (**CCLMS**) at Louisiana State University.

References

[1] Nötzel, R., Temmyo, J. and Tamamura, T. (1994). "Self-organized Growth of Strained InGaAs Quantum Disks", *Nature*, **369**, 131.

[2] Madhukar, A. (1993). "Growth of Semiconductor Heterostructures on Patterned Substrate: Defect Reduction and Nanostructures", *Thin Solid Films*, **231**, 8.

[3] Allen, M. P. and Tildesley, D. J. (1987). Computer Simulation of Liquids, Oxford: Oxford Univ. Press, 1987.

[4] Vashishta, P., Kalia, R. K., Rino, J. P. and Ebbsjö, I. (1990). "Interatomic Potential for SiO_2: A Molecular-dynamics Study of Structural Correlations", *Phys. Rev. B.*, **41**, 12197.

[5] Nakano, A., Kalia, R. K. and Vashishta, P. (1994). "Multiresolution Molecular Dynamics Algorithm for Realistic Materials Modeling on Parallel Computers", *Comput. Phys. Commun.*, **83**, 197.

[6] Greengard, L. and Rokhlin, V. (1987). "A Fast Algorithm for Particle Simulations", *J. Comput. Phys.*, **73**, 325.

[7] Toukmaji, A. Y. and Board, J. A. (1996). "Ewald Summation Techniques in Perspective: A Survey", *Comput. Phys. Commun.*, **95**, 73.

[8] Streett, W. B., Tildesley, D. J. and Saville, G. (1978). "Multiple Time-step Methods in Molecular Dynamics", *Mol. Phys.*, **35**, 639.

[9] Tuckeman, M. E., Berne, B. J. and Martyna, G. J. (1991). "Molecular Dynamics Algorithm for Multiple Time Scales: Systems with Long Range Forces", *J. Chem. Phys.*, **94**, 6811.

[10] Sun, X.-. and Gustafson, J. L. (1991). "Toward a Better Parallel Performance Metric", *Par. Comp.*, **17**, 1093.

[11] Fox, G. C., Williams, R. D. and Messina, P. C. (1994). Parallel Computing Works, San Francisco: Morgan Kaufmann, Chap. 11.

[12] Nakano, A. and Campbell, T."An Adaptive Curvilinear-coordinate Approach to Dynamic Load Balancing of Parallel Multiresolution Molecular Dynamics", *Par. Comp.*, in press.

[13] Gygi, F. (1993). "Electronic-structure Calculations in Adaptive Coordinates", *Phys. Rev. B.*, **48**, 11692.

[14] Nakano, A. (1993). "Fuzzy Clustering Approach to Hierarchical Molecular Dynamics Simulation of Multiscale materials Phenomena", *Comput. Phys. commun.* in press.

[15] Evans, D. J. and Murad, S. (1977). "Singurarity Free Algorithm for Molecular Dynamics Simulation of Rigid Polyatomics", *Mol. Phys.*, **34**, 327.

[16] Zhang, G. and Schlick, T. (1993). "LIN: A New Algorithm to Simulate Dynamics of Biomolecules by Combining Implicit-integration and Normal Mode Techniques", *J. Comput. Chem.*, **14**, 1212.

[17] Skeel, R. D., Zhang, G. and Schlick, T. (1997). "A Family of Symplectic Integrators: Stability, Accuracy, and Molecular Dynamics Applications", *SIAM J. Sci. Comput.*, **18**, 203.

[18] Zadeh, L. A. (1965). "Fuzzy Sets", *Inform. Control*, **8**, 338.

[19] Rose, K. and Fox, G. C. (1990). "Statistical Mechanics and Phase Transitions in Clustering", *Phys. Rev. Lett.*, **65**, 945.

[20] Streitz, F. H. and Mintmire, J. W. (1994)."Electrostatic Potentials for Metal-oxide Surfaces and Interfaces", *Phys. Rev. B.*, **50**, 11996.

[21] Nakano, A. (1997). "Parallel Multilevel Preconditioned Conjugate-gradient Approach to Variable-charge Molecular Dynamics", *Comput. Phys. Commun.*, **104**, 59.

[22] Saad, Y. (1996). Iterative Methods for Sparse Linear Systems, Boston: **PWS**.

[23] Kalia, R. K., Nakano, A., Tsuruta, K. and Vashishta, P. (1997). "Morphology of Pores and Interfaces and Mechanical Behavior of Nanocluster-assembled Silicon Nitride Ceramic", *Phys. Rev. Lett.*, **78**, 689.

[24] Kalia, R. K., Nakano, A., Omeltchenko, A., Tsuruta, K. and Vashishta, P. (1997). "Role of Ultrafine Microstructures in Dynamic Fracture in Nanophase Silicon Nitride", *Phys. Rev. Lett.*, **78**, 2144.

Authors' Biographies

Aiichiro Nakano is an Assistant Professor in the Department of Computer Science at Louisiana State University. He obtained his Ph.D. in Physics from University of Tokyo in 1989. His research interests are: Computer aided nanomaterials design; computational nanoelectronics; parallel multilevel algorithms; and visualization and network technologies for computational sciences.

Rajiv K. Kalia is a Professor in the Department of Physics and Astronomy and the Department of Computer Science at Louisiana State University. He is a co-founder (with Dr. Priya Vashishta) of the Concurrent Computing Laboratory for Materials Simulations (CCLMS). He obtained his Ph.D. in Physics from Northwestern University in 1976. His research interests include computational materials science and high performance computing, in particular: i) atomistic simulations of glasses, nanophase ceramics, and microporous materials; ii) quantum transport in nanodevices; iii) new materials-simulation techniques based on wavelets and multiresolution approaches; and iv) parallel algorithms for materials simulations.

Priya Vashishta is a Cray Research Professor of Computational Sciences in the Department of Physics and Astronomy and the Department of

Computer Science at Louisiana State University. Prior to LSU, he was a Senior Scientist at Argonne National Laboratory, where he was the Director of the Solid State Science Division during 1979 and 1982. His research interests include parallel algorithms for multi-million-particle materials simulations of fracture in amorphous and crystalline ceramic films and coatings.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 129–133
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Ensemble Monte Carlo and Full-Wave Electrodynamic Models Implemented Self-Consistently on a Parallel Processor Using Perfectly Matched Layer Boundary Conditions

IK-SUNG LIM [a], ROBERT O. GRONDIN [b,*] and SAMIR EL-GHAZALY [b]

[a] *Motorola Inc., 2200 W. Broadway, Mesa AZ 85202;*
[b] *Department of Electrical Engineering, Arizona State University, Tempe AZ 85287*

We have been using a self-consistent formulation of full-wave electromagnetic solvers and ensemble Monte Carlo techniques to model ultrafast photoconductivity. Our simulations are running on a MasPar machine. This paper will address aspects of this simulation which may interest workers who are simulating not only photoconductive systems but other systems as well which involve electrodynamics, waves and wave phenomena and ensemble Monte Carlo transport models. In particular, we will report on the inclusion of perfectly matched layer approaches to absorbing boundary conditions for electromagnetic waves. These have in the past several years become widely used in computational electromagnetics codes because they reduce error due to spurious numerical wave reflection off of an absorbing boundary by several orders of magnitude. We will also address the issue of computational cost and show that a full-wave electromagnetic approach is more competitive with a Poisson's equation approach than one might believe. Lastly, our system has the feature that the active portion where the electrons and holes lie is in fact a small fraction of the total experimental system's volume. Unless care is exerted one either has a very significant load imbalance problem or high communications overhead. We compare two different tradeoffs between load imbalance and communications overhead.

Keywords: Monte Carlo, electrodynamic, full-wave, parallel processor, PML, wave-absorbing boundaries

INTRODUCTION

Semiconductor device models always incorporate electromagnetic forces, usually by a quasistatic field calculation. The electric field is allowed to vary in time but its spatial derivatives are assumed to be electrostatic in nature, that is

$$\nabla \bullet \mathbf{E} = \rho/\varepsilon \qquad (1)$$

* Corresponding author: Tel.: (602)-965-5954, Fax: (602) 965-8325, e-mail: Bob. Grondin@asu.edu.

and

$$\nabla \times \mathbf{E} = 0. \tag{2}$$

The irrotational electric field in equation (2) is the essential approximation being made. In contrast, one can use an electrodynamic or full-wave electromagnetic model [1–8] in which one solves Maxwell's curl equations

$$\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t \tag{3}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \partial \mathbf{D}/\partial t. \tag{4}$$

In this paper we will discuss this issue using a system where the transport model is an ensemble Monte Carlo model. We will discuss load imbalance/communications overhead problems associated with the parallelization of such a system, show that a full-wave electromagnetic approach is more cost competitive than one might believe and will report on the inclusion of perfectly matched layer approaches to absorbing boundary conditions for electromagnetic waves.

TRADEOFFS BETWEEN QUASI-STATIC AND FULL-WAVE MODELS

The physical motivation for switching to a full-wave model of a semiconductor device may be either internal or external to the device. Our devices are often incorporated into a larger system. In some cases, e.g. when computing broadband response in the millimeter-wave and submillimeter-wave regions, electrodynamic models of the external components are used. Incorporating full-wave techniques into a device model is then quite natural.

The internal physical issue is when does a quasistatic approach to the internal field calculation break down? The typical quasi-static device model solves Poisson's equation

$$\nabla^2 V = -\rho/\varepsilon \tag{5}$$

to determine a scalar potential function V. Fields and forces are then computed by taking the gradient of this scalar potential i.e.

$$\mathbf{E} = -\nabla V. \tag{6}$$

The boundary value problem associated with Poisson's equation is computationally demanding. The gradient operation of equation (6) on the other hand is computationally trivial. Poisson's equation can always be used, even in a fully electrodynamic setting, if one makes the appropriate selection of gauge [9, 10]. What fails in electrodynamic setting is the computationally trivial gradient operation of equation (6). Instead, we must additionally solve for the vector potential and compute the electric field using

$$\mathbf{E} = -\nabla V - \partial \mathbf{A}/\partial t. \tag{7}$$

Equation (6) implies the irrotational field of equation (2) but equation (7) does not.

Another part of the tradeoff between quasistatic and full-wave approaches is the issue of computational cost. Before dealing with this issue, a more detailed discussion of our computational system is needed. We have been implementing this system on a MasPar MP-2 machine. This machine has 8192 Processing Elements (PE) connected in a mesh. It is a SIMD machine with two routes for data transfer between elements. The fastest communication is between nearest neighbor PEs on a network called the Xnet. The slower route is a global router which communicates between any two PEs. As described before [1–3], we will couple two different calculations together. We self-consistently couple an Ensemble Monte Carlo (EMC) model with a finite difference solution of Maxwell's curl equations. These two calculations involve different types of data. The EMC calculation has data which is associated with a specific particle while the electromagnetic calculation has data which is associated with a location in space. There are natural methods for parallelizing both computations but these methods differ in their utilization of the PE array due to this difference in data type. In the EMC calculation one wishes to evenly distribute carries over the PE array while in

the field computation one assigns adjoining regions of space to nearest neighbors PEs. Our problem is to simultaneously parallelize both solutions.

Two basic strategies for this simultaneous parallelization will be compared here. In both algorithms, we compute the fields using the adjacent PE approach mentioned above. (The initial condition for the fields is produced by solving Poisson's equation using a red/black SOR technique [11]). In algorithm 1, an equal number of particles is assigned to each processor thus alleviating any load imbalance in the EMC calculation but a high price is paid for communication between the PEs during the process where we switch between the two calculations. In algorithm 2, a region of space is assigned to a single processor even during the EMC calculation. This processor handles the particles found in this region. This minimizes communication overhead but introduces a significant load imbalance as the EMC region is only a small part of the simulation domain. In order to alleviate this load imbalance, the field information for just the EMC region is duplicated throughout the PE array so that carriers can be located on any PE which has the appropriate field information.

We apply this to the physical system shown in Figure 1. A nonuniform grid was used. The grid used a 2 micron spacing in all three dimensions in the passive region and a 0.1 micron spacing in the EMC region. There is a gradual transition between the two extremes. The run times of these two algorithms are shown in Table I. They are broken down into entries for the field calculation, the MC calculation and the coupling of these two. As expected, algorithm 1 spent less time on the EMC calculation due to better load balance but more time in coupling the two calculations due to the communication overhead. Over all, algorithm 1 was somewhat superior to algorithm 2 for this particular example. The performance of the algorithms though is sufficiently close that no conclusion should be reached that algorithm 1 is always superior to algorithm 2.

Now, how does a quasistatic computation compare in numerical cost with a full wave approach? Since we have both a Poisson solver and a full-wave model here, we can compare the two. We duplicate the computation just described only we proceed quasistatically. (Of course, a quasistatic solution of a transmission line transient response in unphysical but our goal is to compare computational costs). This run-time data is also shown in Table I along with the number of iterations required per time step. The full-wave model uses 1 iteration per time step while the number of iterations per timestep in the Poisson's solver depends on the convergence criteria. If too many iterations are required per timestep, the quasi-static model loses any cost advantage it may have had over a full-wave solution. In fact, for a very tight convergence criteria, the quasistatic model is signficantly more expensive than the full-wave!.

Examination of Figure 1 shows that we are simulating an open system with electromagnetic waves propagating away towards infinity. As our grid is finite in size we need to minimize the numerical reflection of these outgoing waves at the

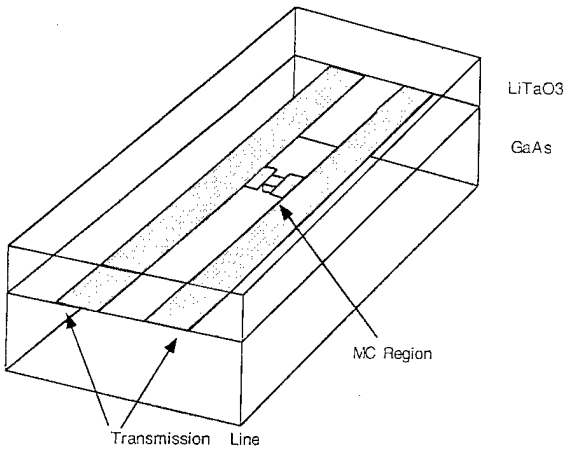


FIGURE 1 The system simulated here is a microstrip line structure on a GaAs substrate. A dc bias is applied and the system is which will be excited by a subpicosecond pulse in the Monte Carlo or MC region shown. An electro-optic material is place over the top of the system. The grid size for the electromagnetics code was 127 × 63 × 54. 8192 electrons and 8192 holes were included in the Monte Carlo calculations.

TABLE I Computational run times for various solution techniques

Method	Average Iteration	Field (sec/1000iter)	Coupling (sec/1000iter)	MC (sec/1000iter)	Total (sec/1000iter)
Dynamic algorithm 1	1	589.4	814.6	40.1	1444.1
Dynamic algorithm 2	1	588.8	335.8	823.6	1748.2
Static error: 1E-3V	1.98	354.5	380.7	35.8	771.0
Static error: 1E-4V	11.8	975.6	378.6	40.6	1394.8
Static error: 1E-5V	65.4	4375.8	379.7	39.5	4795.0

end of our numerical grid. There has been noticeable advance in such absorbing boundary conditions for electromagnetic waves in the past several years, usually associated with the phrase "Perfectly Matched Layer" or PML [12–18]. These techniques have reduced the numerical error associated with spurious numerical reflection by several orders of magnitude and are particularly advantageous for the case of oblique incidence. We use the technique of Berenger [12–14] although other techniques have been developed as well [15–18]. Here we will outline the conceptual nature of these techniques and refer the reader to the references for details.

The essential idea is to add a fictitious "Perfectly Matched Layer" on the outside of all surfaces where we have to implement a wave absorbing boundary. We then extend our numerical grid out into this fictitious layer, terminating it at the end of this layer. Since this is a purely artifical layer, we are free to pick its material properties to meet two constraints. First, the reflection coefficient associated with the interface between the valid solution space and the fictitious absorbing space must be zero for all frequencies and all angles of incidence. This will eliminate numerical reflection off of this interface. However, accomplishing this goal just moves the original problem to the new truncation of the grid at the far end of the fictitious absorbing layer. We will have a reflection off of this truncation. Therefore our second constraint on the PML is that it must be a lossy medium which attenuates electromagnetic waves. The techniques described in the references succeed in meeting both constraints.

In summary, one can proceed with an electrodynamically based approach in the modeling a semiconductor device. For the cases where one is already expending significant computational resources e.g. three dimensional device models, quasistatic and electrodynamic techniques do not necessarily differ significantly in computational cost. If one does so, the use of a PML type of absorbing boundary is recommended. Care however may be required if one implements the system on a massively parallel processor array as the field calculation and the transport model may employ very different types of data and additionally, the device itself may be a comparatively small portion of the simulation domain.

References

[1] Lu, Y., Joshi, R., El-Ghazaly, S. and Grondin, R. O. (1989). *Solid State Electronics*, **32**, 1297.
[2] El-Ghazaly, S., Joshi, R. and Grondin, R. O. (1990). *IEEE Trans. Microwave Theory and Techniques*, **38**, 629.
[3] Connolly, K., El-Ghazaly, S., Grondin, R., Joshi, R. P. and Ferry, D. K. (1992). *Semiconductor Science and Technology*, **7**, B199.
[4] Goodnick, S. M., Pennatu, S. S., Ranawake, U. A., Lenders, P. M. and Tripathi, V. K. (1995). *Inter. Journal Numerical Modeling: Electronic Networks, Devices and Fields*, **8**, 205.
[5] Megahed, M. A. and El-Ghazaly, M. (1995). *IEEE Trans. Microwave Theory Tech.*, **43**, 2590.
[6] Alsunaidi, M. A., Imtiaz, S. M. and El-Ghazaly, S. M. (1996). *IEEE Trans. Microwave Theory Tech.*, **44**, 799.
[7] Sohel Imtiaz, S. M., El-Ghazaly, S. M. and Grondin, R. O., these *proceedings*.
[8] Sano, E. and Shibata, T. (1990). *IEEE J. Quantum Electronics*, **26**, 372.

[9] Jackson, J. D. (1962). *Classical Electrodynamics*, John Wiley & Sons, New York.
[10] Mahan, G. D. (1981). *Many Particle Physics*, Plenum Press, New York.
[11] Brandt, A. (1981). "Multigrid Solvers on Parallel Computers" in *Elliptic Problem Solvers*, M. Schultz Ed., Academic Press, New York.
[12] Berenger, J. O. (1994). *J. Computational Physics*, **114**, 185.
[13] Berenger, J. P. (1996). *IEEE Trans. Antennas Prop.*, **44**, 110.
[14] Berenger, J. P. (1996). *J. Computational Physics*, **127**, 363.
[15] Mittra, R. and Pekel, U. (1995). *IEEE Trans. Microwave and Guided Wave Letters*, **5**, 84.
[16] Pekel, U. and Mittra, R. (1995). **5**, 258.
[17] Rappaport, C. M. (1995). *IEEE Microwave and Guided Wave Letters*, **5**, 90.
[18] Sackes, Z. S., Kingsland, D. M., Lee, R. and Lee, J. (1995). *IEEE Trans. Antennas and Propagation*, **43**, 1460.

Authors' Biographies

Robert Grondin was born and raised in Michigan. He attended the University of Michigan obtaining the B.S., M.S. and Ph.D. degrees. From 1981 to 1983 he was a post-doctoral research associate at Colorado State university. In 1983 he joined the faculty of the Department of Electrical Engineering at Arizona State University where he is presently an associate professor. He is a senior member of the IEEE and a member of the AAAS.

Simir M. El-Ghazaly received the Ph.D. degree, in Electrical Engineering, from the University of Texas at Austin, Texas, in 1988. He joined Arizona State University as Assistant Professor in August 1988, and became an Associate Professor in 1993. Dr. El-Ghazaly is a senior member of IEEE, an elected member of Commissions A and D of URSI, a member of Tau Beta Pi, Sigma Xi, and Eta Kappa Nu. He is the secretary of US National Committee of URSI, Commission A. He is also the Chairman of the Chapter Activities Committee of the IEEE Microwave Theory and Techniques Society.

Ik-Sung Lim received the B.S., degree from Sogang University, Seoul Korea and the Masters degree in electrical engineering from Arizona State University and is presently completing the Ph.D. degree at Arizona State University. Since 1987 he has worked as an engineer in Motorola Semiconductor Products Sector in Arizona.

MATHEMATICAL MODELS

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 135–141
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Applicability of the High Field Model: An Analytical Study Via Asymptotic Parameters Defining Domain Decomposition

CARLO CERCIGNANI [a], IRENE M. GAMBA [b], JOSEPH W. JEROME [c,*] and CHI-WANG SHU

[a] *Politecnico di Milano, 20133 Milano, Italy;*
[b] *Courant Institute, New York University, New York, NY 10012;*
[c] *Department of Mathematics, Northwestern University, Evanston, IL 60208;*
[d] *Division of Applied Mathematics, Brown University, Providence, RI 02912*

In this paper, we present a mesoscopic-macroscopic model of self-consistent charge transport. It is based upon an asymptotic expansion of solutions of the Boltzmann Transport Equation (BTE). We identify three dimensionless parameters from the BTE. These parameters are, respectively, the quotient of reference scales for drift and thermal velocities, the scaled mean free path, and the scaled Debye length. Such parameters induce domain dependent macroscopic approximations. Particular focus is placed upon the so-called high field model, defined by the regime where drift velocity dominates thermal velocity. This model incorporates kinetic transition layers, linking mesoscopic to macroscopic states. Reference scalings are defined by the background doping levels and distinct, experimentally measured mobility expressions, as well as locally determined ranges for the electric fields. The mobilities reflect a coarse substitute for reference scales of scattering mechanisms. See [9] for elaboration.

The high field approximation is a formally derived modification of the augmented drift-diffusion model originally introduced by Thornber some fifteen years ago [25]. We are able to compare our approach with the earlier kinetic approach of Baranger and Wilkins [5] and the macroscopic approach of Kan, Ravaioli and Kerkhoven [20].

Keywords: Asymptotic parameters, mesoscopic-macroscopic model, augmented drift-diffusion, high field model, domain decomposition

1. INTRODUCTION

This presentation is motivated by the search for more reliable macroscopic models of high field transport in submicron structures which are computationally efficient. This is a research topic of major interest to the micro-electronics industry.

Transport in submicron structures differs from transport in bulk material in many ways. It includes far from equilibrium situations created

* Corresponding author.

by large electric fields in small structures; ballistic electrons, when dimensions of the drift-forcing term are of the same order as the mean free path; and new effects, introduced by the close proximity to the boundaries of regions where the density gradients are (relatively) rapidly changing.

Most of the previous work on transport in submicron semiconducting structures used several corrections to drift-diffusion equations in which the current is a sum of drift and diffusion terms, where both have field-dependent mobility and diffusion coefficients, respectively. References on models and computational methods for $N^+ - N - N^+$ structures of Si or GaAs can be found in the literature of the last decades [24, 23, 16, 21, 25, 20]. However, the computations have not yielded results which have been genuinely satisfactory, and the main reason has been the lack of understading of the accuracy of the modeling in many heterogeneous structures. An alternative to the drift-diffusion equations has been Monte Carlo simulation, a very costly computational procedure.

An approach intermediate between drift-diffusion and Monte Carlo simulation was generated by the hydrodynamic like theory, using velocity moments of the Boltzmann equation (see [6, 2, 3, 15]) and realizing closure by imposing constitutive relationships between the state macroscopic variables. Extensive numerical modeling of $N^+ - N - N^+$ Si and GaAs structures under several geometric constraints can be found in [12, 17, 18, 19]. For the derivation of these models, the distribution function is often assumed to be a Maxwellian distribution. However, this assumption has been shown to be questionable in lightly doped submicron structures, in part because of the presence of ballistic electrons.

However, it has been mathematically proved that hydrodynamic models relax to energy transport models. The latter have been derived as macroscopic limits associated with a particular choice of space-time scale that makes some collision mechanisms dominant (see [10] for a rigorous proof in the case of the relaxation of the full hydrodynamic model). These models, however, cannot describe situations where the drift-forcing term and the resulting scaled mean free path are of the same order, but they are asymptotically correct in regions where the predefined background is very slowly varying (i.e., almost constant doping). A new approach attempting to fill in the middle ground among drift-diffusion, hydrodynamic energy transport, and Monte Carlo simulation (see [4, 5, 26]) was the use of direct, numerical solution of the Boltzmann equation within the relaxation time approximation. The use of the Boltzmann equation for non-electron transport is justified in the above references, as well as in [7, 13]. The relaxation time approximation within the Boltzmann-Poisson system approach, positing that all scattering processes can be characterized by a few scattering rates setting up the scale of relaxation time constants, may not be strictly valid for GaAs [11].

Our goal, however, is to present a full solvable model for an $N^+ - N - N^+$ structure, as an alternative to all previous calculations, which will produce comparable results with greater efficiency than attained by [4, 5, 20]. In particular, we incorporate into the modeling the macroscopic limit associated with the choice of space, time, potential drop, and local electric field scales, combined with the relaxation time scale, which makes the drift-forcing term and collisions both dominant, and of the same scale set up by the local scaled mean free path.

As a consequence of this approach, we incorporate a level of modeling that takes into account mesoscopic-macroscopic multiscales, and corresponding limit derivations, according to different scales becoming dominant in different regions of the $N^+ - N - N^+$ device, depending on the spatial inhomogeneity of the lightly doped region.

2. PRELIMINARIES OF THE HIGH FIELD MODEL

We identify three dimensionless parameters, viz.,

- The ratio η of drift and free velocity (the latter usually taking on the thermal velocity);

- The scaled mean free path ε and
- The scaled Debye length γ from the electrostatic potential equation of a self-consistent model.

The dimensionless parameters induce domain dependent macroscopic approximations, which are valid within regimes that do not change the scale of the Debye length. This is of paramount importance in the macroscopic derivation associated with the BTE under a self-consistent field: the scaling of the forcing term depends on the scaled external and internal fields, and must not override the limiting regime of the macroscopic derivation. Therefore, this model incorporates kinetic transition layers linking mesoscopic and macroscopic states.

Reference scalings are defined by the background doping levels and distinct, experimentally measured mobilities which serve to set up the scale of the relaxation time approximation of the scattering mechanisms. In addition, scales take into account locally determined ranges for the electric fields. Numerical experiments of this proposed model are also presented by the authors in [8]. Although this is a first approach to domain decomposition models, based on a mesoscopic-macroscopic linking, our results seem to give an improvement (see also [13]) of computational efficiency, relative to the kinetic computations of Baranger and Wilkins [5]. Moreover, the model itself seems to be an improvement over the augmented drift-diffusion model used by Kan, Ravaioli and Kerkhoven [20]. This model, based on a model introduced empirically by Thornber [25], employs necessary corrections to the usual macroscopic approach to high field models.

3. BOLTZMANN-POISSON SYSTEM SCALES

Finally, we concentrate on the derivation of the macroscopic models for semiconductors, based on asymptotic expansions of multiscale dimensionless BTE-Poisson systems. For the sake of simplicity we shall consider only electrons. The semiclassical Boltzmann-Poisson system, within a parabolic band-relaxation time approximation, may be written:

$$\partial_t F + v \cdot \nabla_x F - \frac{e}{m} E(x,t) \cdot \nabla_v F = \frac{\langle F \rangle M - F}{\tau},$$
$$E = -\nabla\phi, \quad \nabla \cdot (\epsilon \nabla\phi) = e(\langle F \rangle - N_d(x)), \qquad (1)$$

where $F(x, v, t)$ is the density function for the electron at position x, velocity v and time t. The constants e, m, τ and ϵ represent the electric charge constant, its mass, the relaxation time reference scale, and permittivity of the material, respectively. The functions ϕ and E of (x, t) represent the electrostatic potential and its field. The bracket $\langle\cdot\rangle$ denotes the usual average of the distribution function with respect to the velocity variable. Finally $M = M(v, x)$ denotes the Maxwellian centered at zero velocity with $\theta = \theta(x)$ representing the background temperature of the lattice, measured in units of specific energy. Derivations of mesoscopic limits associated with system (1) depend on a particular choice of space-time scales, collision mechanism scales, and drift velocity scales induced by self-referencing field scales. These regimes, as we shall see, can make some terms of the BTE dominant. The idea that follows then, is to *expand the solution of the BTE about the solution of the* **ground** *equation corresponding to the dominant terms*. Of course, this will depend on the form of scattering mechanisms that set up the scales of the collision terms. Lately, a very nice survey of a hierarchy of macroscopic models for semiconductors under limit regimes that correspond to dominant collision mechanisms has been presented in [1]. These macroscopic models correspond to classical drift-diffusion and so-called energy transport models. They arise from expansion about kernels of the collision operators, which are field independent, so they do not incorporate the scales of the external, or even self-induced electric field. Diffusions are field independent, and consequently they cannot approximate the ballistic electron distribution func-

tions that have been observed by Barenger and Wilkins [4, 5]. After the work initiated by [14] and [22] on strong forcing scaling for external fields, where the dominant term in the scaled Eq. (1) is given by the balance of the force-drift (e/m) $E(x,t)\nabla_v F$ and the collision terms, the following 3-scale dimensionless formulation has been formulated in [9] in ℓ Euclidean dimensions:

$$\eta F_t + v\cdot\nabla_x F + \frac{\eta}{\varepsilon}\nabla_x\phi\cdot\nabla_v F = \frac{1}{\varepsilon}\{\langle F\rangle M - F\},$$
$$\Delta\phi = \gamma\left(\frac{1}{\eta^{\ell}}\langle F\rangle - N_d(x)\right), \qquad (2)$$

where $\varepsilon = \frac{\tau\bar{\theta}^{1/2}}{L}$ is the scaled mean free path for a length scale L and relaxation time τ, and $\bar{\theta}$ is the reference scale of the background energy. The dimensionless constant $\eta = U/\bar{\theta}^{1/2}$, where U, in units of velocity, is given by $\frac{\tau e}{m}\frac{[\phi]}{L}$, takes into account the reference scale of the local electric field. Finally, the dimensionless constant $\gamma = \frac{\bar{\rho}eL^2}{\epsilon[\phi]}$ represents the scaled Debye length, where $\bar{\rho}$ scales the density of the fixed background $N_d(x)$.

Now, a strong force regime corresponds to values of $\eta = O(1)$, $\gamma = O(1)$ and $\varepsilon < L$. In this case, system (2) is obtained, where the distribution function F is scaled with the drift velocity U (instead of the low field scalings that use the thermal velocity scale for F). The time scale is then fixed at LU^{-1}.

4. ASYMPTOTIC LIMITS AND THE HIGH FIELD MODEL

Clearly, any asymptotic expansion is justified under the assumption that the scaling does not break down. Thus, any expansion of the BTE-Poisson system (2) requires that the scale of the internal field defined by the Poisson equation (i.e., depending on the value of γ) does not override the scaling of the BTE.

$\bar{\rho}$ is the reference scale of the predefined $N_d(x)$ and will change scale whenever $N_d(x)$ does. Accordingly, so will the local field. Therefore, as an immediate consequence of the fact that $\gamma = \gamma(\bar{\rho})$ can be easily checked for GaAs (as in [5, 20]), the dimensionless quantities vary in the N^+ regions from the N regions. Thus, in N^+ regions, $\eta = \varepsilon = o(1)$. Because of scaling times and the distribution function F, accordingly, (2) yields the Drift-Diffusion-Energy Transport regime as the relative electric field remains moderate and the collision mechanisms become dominant.

However, in the N region it is computed by the authors in [8] that $\varepsilon \approx (1/7)$ and $\eta = O(1)$. Thus, while γ remains of order $O(1)$, it has been shown that densities and the electric field satisfy the equations,

$$\rho_t + \nabla\cdot J = 0,$$
$$J = -\tau[\nabla_x(\rho(\theta + \mu^2 E\otimes E))] - \mu\rho E$$
$$+\tau\mu E\rho\nabla\cdot(\mu E) - \tau\mu\rho\frac{e}{\epsilon}(\mu\rho E - w),$$
$$\nabla\cdot w = 0, \quad \operatorname{curl} w = \mu E\otimes\nabla\rho,$$
$$E = -\nabla\phi, \quad \nabla\cdot(\epsilon\nabla\phi) = (\rho - N_d(x)), \qquad (3)$$

in their dimensionalized formulation. Here μ is the mobility. The new variable w can be interpreted as the curl of a magnetic field, associated with the high electric current. This model is a correction of the one proposed by Thornber [25], where the diffusion and transport coefficients were given empirically.

We point out that in this relaxation time approximation, the mobility coefficient is given by $\mu = \frac{e\tau}{m}$, so clearly the saturation of the mobility depends on how the scattering rates scale for τ changing with respect to the electric field. The standard assumption based on curve fitting is to take $\mu = C_1/(1 + C_2|E|^2)^{1/2}$ (see [16]).

We stress that this regime is valid only for the strong force scaling. The transition regime from weak to strong forcing scaling, and conversely, ought to be done by carefully considering the corresponding kinetic transition layers. The implementation and solution of this transition problem require solution of boundary value problems at the kinetic and macroscopic level as well.

5. DOMAIN DECOMPOSITION AND TRANSITION LAYERS

Appropriate boundary data at the kinetic level and the corresponding limiting data for the mesoscopic approximation are presented in [9]. For kinetic level distribution functions there are prescribed exact neutral space charge conditions at contact boundaries and standard insulating boundary conditions at insulating walls. Then, these conditions, together with the asymptotic approximations, yield Dirichlet boundary conditions at the macroscopic level, with a correction term of the order of the asymptotic parameter (i.e., the order of the scaled mean free path ε). The contact condition, at the kinetic level, on a space boundary section Γ reads [9]:

$$F_- = \left(N_d(x) - \int_{v\cdot n>0} F_+ dv\right) \frac{P}{\int_{v\cdot n<0} P_+ dv}, \quad \text{on } v\cdot n < 0, \tag{4}$$

where n denotes the outer normal with respect to the boundary section Γ, and P denotes the lowest order term in the expansion for either weak force scaling (i.e., P is a Maxwellian) or strong force scaling (i.e., P peaks asymmetrically and is electric field dependent).

The corresponding macroscopic condition for the inner boundary can be computed explicitly [9] and yields

$$\rho(x) = N_d(x) + o(\varepsilon)K(x),$$
$$K(x) = \frac{\langle vF_-^{(1)}\rangle \cdot n\langle F_-\rangle}{\langle v\cdot F_-^{(0)}\rangle \cdot n}. \tag{5}$$

Here,

$$F \simeq F^{(0)} + \varepsilon F^{(1)} + O(\varepsilon^2), \quad F^{(0)} = \langle F\rangle P,$$
$$\langle f_-\rangle = \int_{v\cdot n<0} f\, dv.$$

We finally remark that the Ohmic contacts on N^+ regions correspond to a local scaled mean free path ε of much smaller order than in the N regions. Hence, at boundary ohmic contacts on N^+ regions, condition (5) is almost a neutral space charge condition at the macroscopic level, as historically has been, and still is typically prescribed. However, at the transition layer from the N^+ to the N regime, the transfer of the data entails a kinetic computation using condition (4); then, one prescribes an exact background density, and thus delivers, at the next level, for the strong force model, a condition for the density that is an $O(\varepsilon)$ derivation from the background density. This fact can actually be observed in the computations of Barenger and Wilkins [4, 5].

Numerical implementations of strong-weak forcing decomposition are presented by the authors [8] in this same issue. However, we use there a matching with hydrodynamic computations of well accepted performance. In Figure 3 of [8], we present a graph of η, which clearly correlates with the suggested domain decomposition. Theoretical justifications and alternative numerical methods of these domain decomposition ideas are underway and will be presented in future work.

Acknowledgements

The second author is supported by the National Science Foundation under grant DMS-9623037. The third author is supported by the National Science Foundation under grant DMS-9424464. The fourth author is supported by the National Science Foundation under grants ECS-9214488 and ECS-9627849, and the Army Research Office under grant DAAH04-94-G-0205.

References

[1] Ben Abdallah, N. and Degond, P. (1996). On a hierarchy of macroscopic models for semiconductors. *J. Math. Phys.*, **37**, 3306–3333.

[2] Anile, A. M. and Muscato, O. (1995). Improved hydrodynamical model for carrier transport in semiconductors. *Phys. Rev. B*, **51**, 16728–16740.

[3] Baccarani, G. and Wordeman, M. R. (1985). An investigation of steady-state velocity overshoot effects in Si and GaAs devices. *Solid State Electr.*, **28**, 407–416.

[4] Barenger, H. U. and Wilkins, J. W. (1984). Ballistic electrons in an inhomogeneous submicron structure: Thermal and contact effects. *Phys. Rev. B*, **30**, 7349–7351.
[5] Barenger, H. U. and Wilkins, J. W. (1987). Ballistic structure in the electron distribution function of small semiconducting structures: General features and specific trends. *Phys. Rev. B*, **36**, 1487–1502.
[6] Blotekjaer, K. (1970). Transport equations for electrons in two-valley semiconductors. *IEEE Trans. Electron Devices*, **17**, 38–47.
[7] Cercignani, C. (1987). *The Boltzmann Equation and its Application*. Springer-Verlag, New York.
[8] Cercignani, C., Gamba, I. M., Jerome, J. W. and Shu, C.-W., Applicability of the high field model: A preliminary numerical study. This volume.
[9] Cercignani, C., Gamba, I. M. and Levermore, C. D., (1997). High field approximations to Boltzmann-Poisson system boundary conditions in a semiconductor. *Appl. Math. Lett.*, **10**, 111–117.
[10] Chen, G.-Q., Jerome, J. and Zhang, B., Existence and the singular relaxation limit for the inviscid hydrodynamic energy model. *Proceedings, Workshop on Modeling and Computation for Applications in Mathematics, Science and Engineering*, Oxford University Press, to appear.
[11] Conwell, E. (1982). *Handbook of Semiconductors*, **1** (W. Paul, ed.), p. 513, North–Holland.
[12] Fatemi, E., Jerome, J. and Osher, S. (1991). Solution of the hydrodynamic device model using high-order non-oscillatory shock capturing algorithms. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, **10**, 232–244.
[13] Ferry, D. K. and Grondin, R. O. (1991). *Physics of Submicron Devices*. Plenum Press, New York.
[14] Frosali, G., van der Mee, C. V. M. and Paveri Fontana, S. L. (1989). Conditions for run-away phenomena in the kinetic theory of particle swarms. *J. Math. Phys.*, **3**, 1177–1186.
[15] Gnudi, A., Odeh, F. and Rudan, M. (1988). An efficient discretization scheme for the energy continuity equation in semiconductors. In *Proceedings of SISDP*, 387–390.
[16] Jerome, J. W. (1996). *Analysis of Charge Transport; A Mathematical Theory of Semiconductor Devices*. Springer.
[17] Jerome, J. W. and Shu, C.-W. (1994). Energy models for one-carrier transport in semiconductor devices. In *Semiconductors, Part II*, **59** of *IMA Volumes in Mathematics and its Applications* (W. M. Coughran, J. Cole, P. Lloyd and J. K. White, eds.), pp. 185–207. Springer-Verlag, New York.
[18] Jerome, J. W. and Shu, C.-W. (1995). The response of the hydrodynamic model to heat conduction, mobility, and relaxation expressions. *VLSI DESIGN*, **3**, 131–143.
[19] Jerome, J. W. and Shu, C.-W. (1995). Transport effects and characteristic modes in the modeling and simulation of submicron devices. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, **14**, 917–923.
[20] Kan, E. C., Ravaioli, U. and Kerkhoven, T. (1991). Calculation of velocity overshoot in submicron devices using an augmented drift-diffusion model. *Solid-State Electr.*, **34**, 995–999.
[21] Markowich, P. A., Ringhofer, C. A. and Schmeiser, C. (1990). *Semiconductor Equations*. Springer-Verlag, Vienna.
[22] Poupaud, F. (1992). Runaway phenomena and fluid approximation under high fields in semiconductor kinetic theory. *Z. Angew. Math. Mech.*, **72**, 359–372.
[23] Selberherr, S. (1984). *Analysis and Simulation of Semiconductor Devices*. Springer-Verlag, New York.
[24] Sze, S. M. (1981). *Physics of Semiconductor Devices*. John Wiley, New York.
[25] Thornber, K. K. (1983). Current equations for velocity overshoot. *IEEE Electron Device Lett.*, **3**, 69–71.
[26] Trugman, S. A. and Taylor, A. J. (1986). Analytic solution of the Boltzmann equation with applications to electron transport in inhomogeneous semiconductors. *Phys. Rev. B*, **33**, 5575–5584.

Authors' Bigraphies

Carlo Cercignani received his Univ. Milano Laurea Physics in 1961 and his Univ. Milano Laurea Mathematics in 1963. From 1963 to 1966 he was Assistant and Associate Professor of Plasma Physics at the University of Milano. He was a Visiting Associate Professor of Applied Mathematics at M.I.T. during 1966–67. He came to Politecnico di Milano in 1968 and has been Professor in Theoretical Mechanics since 1975. His scientific specialties include the kinetic theory of gases and its applications to rarefied gas dynamics, neutron transport and semiconductors; nonequilibrium statistical mechanics, fluid dynamics and singular integral equations. His current research interests are analytical and numerical treatment of kinetic and transport equations for gases and semiconductors. He is a member of Accademia Nazionale dei Lincei, Istituto Lombardo and Académie des Sciences de Paris. In addition he is a member of the National Committees for Mathematics of the Italian Research Council (C.N.R., President), Scientific Committee of the National Group for Mathematical Physics (G.N.F.M.) of C.N.R., a member of General Assembly of IUTAM, General Assembly of IMU, Congress Committee of IUTAM, Scientific Committee of 1st European Mathematical Congress, and Advisory Committee for Symposia on Rarefied Gas Dynamics. He serves on the following editorial boards (1993): Fluid Dynamics Research, Transport Theory and Statistical Physics, MECHANICS Research Communications, Annali di Matematica Pura e Applicata, Applied Mathematics Letters, European Journal of Me-

chanics B (Ass. Ed.), Nonlinear Differential Equations and Applications, Surveys in Mathematics for Industry, Continuum Mechanics and Thermodynamics. He has authored or coauthored seven scientific books and more than two hundred scientific publications. He is the recipient of the Gold Medal for Mathematics of the Accademia dei XL (1982), Prize "Citta' di Cagliari" for Applied Mathematics (1992), Docteur Honoris Causa of the University Pierre et Marie Curie (Paris VI) (1992) and the Humboldt Prize (Academic Year 1995–1996).

Irene M. Gamba received the Ph.D. degree in Mathematics from The University of Chicago in 1989. From 1990 to 1991 she was at Purdue University as a Visiting Assistant Professor and Research Associate. She was an Assistant Professor at The College of New Jersey, Trenton, NJ, during 1991–92. She came to the Courant Institute of Mathematical Sciences, New York University, in 1992. There she was an NSF Post doctoral Fellow during 1992–94, an Assistant Professor from 1994–96 and Associate Professor during 1996–97. She has accepted a position as Professor of Mathematics at the University of Texas at Austin beginning in August, 1997. Her research interests include nonlinear analysis applied to gas dynamics and charged-particle transport systems in the mathematical modeling of microelectronic devices.

Joseph W. Jerome received the Ph.D. degree in Mathematics from Purdue University in 1966. He was visiting Assistant Professor at the Mathematics Research Center, University of Wisconsin, during 1966–68 and was Assistant Professor at Case Western Reserve University during 1968–70. He joined Northwestern University in 1970, where he has been Professor of Mathematics and Applied Mathematics since 1976 . He has held sabbatical positions at Oxford University, England, 1974–75, University of Texas, 1978–79 and Bell Laboratories, Murray Hill, 1982–83. He was visiting scholar at the University of Chicago in 1985. He received the Distinguished Alumnus Award from Purdue University's School of Science in 1996. His research interests include applied analysis, numerical analysis, computational electronics, and ion transport in biology. The most recent of his three books, Analysis of Charge Transport, was published by Springer in 1996.

Chi-Wang Shu received the B.S. degree in Mathematics from the University of Science and Technology of China (USTC) in 1982 and the Ph. D. degree in Applied Mathematics from the University of California at Los Angeles (UCLA) in 1986. He was a post doctoral fellow at the Institute for Mathematics and Its Applications (IMA), University of Minnesota, during 1986–1987 and came to Brown University in 1987, where he has been Professor of Applied Mathematics since 1996. He was a co-recipient of the Chinese Academy of Science Award in Numerical Analysis and Scientific Computing in 1995. His research interests include numerical analysis, scientific computing and computational physics. He currently serves on the editorial boards of Mathematics of Computation, SIAM Journal on Numerical Analysis and Journal of Computational Mathematics.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 143–146
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Smooth Quantum Hydrodynamic Model Simulation of the Resonant Tunneling Diode

CARL L. GARDNER* and CHRISTIAN RINGHOFER†

Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804

Smooth quantum hydrodynamic (QHD) model simulations of the resonant tunneling diode are presented which exhibit enhanced negative differential resistance (NDR) when compared to simulations using the original $O(\hbar^2)$ QHD model. At both 300 K and 77 K, the smooth QHD simulations predict significant NDR even when the original QHD model simulations predict no NDR.

Keywords: Quantum hydrodynamic model, resonant tunneling diode

The original $O(\hbar^2)$ quantum hydrodynamic (QHD) equations have been remarkably successful in simulating the effects of electron tunneling through potential barriers including single and multiple regions of negative differential resistance and hysteresis in the current-voltage curves of resonant tunneling diodes. However, the model relies on an ad hoc replacement of derivatives of the potential with derivatives of the logarithm of the electron density in order to avoid infinite derivatives at heterojunctions.

Refs. [1] and [2] present an extension of the QHD model that is mathematically rigorous for classical potentials with discontinuities—as are present at heterojunction barriers in quantumsemiconductor devices. The stress tensor in this "smooth" QHD model actually cancels the leading singularity in the classical potential at a barrier and leaves a residual smooth effective potential with a lower potential height in the barrier region (see Fig. 1).

The smooth QHD equations have the same form as the classical hydrodynamic equations:

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x_i}(nu_i) = 0 \tag{1}$$

$$\frac{\partial}{\partial t}(mnu_j) + \frac{\partial}{\partial x_i}\left(mnu_iu_j - P_{ij}\right) = -n\frac{\partial V}{\partial x_j} - \frac{mnu_j}{\tau_p} \tag{2}$$

$$\frac{\partial W}{\partial t} + \frac{\partial}{\partial x_i}\left(u_iW - u_jP_{ij} + q_i\right) = -nu_i\frac{\partial V}{\partial x_i} - \frac{\left(W - \frac{3}{2}nT_0\right)}{\tau_w} \tag{3}$$

* The research of C.L.G. is supported in part by the U.S. Army Research Office under grant DAAH04-95-1-0122 and by the National Science Foundation under grant DMS-9706792. Corresponding author.
† The research of C.R. is supported in part by the National Science Foundation under grant DMS-9706792.

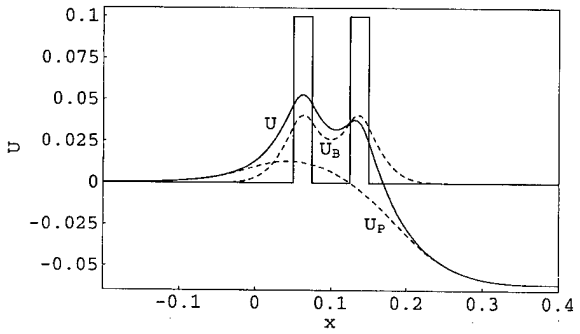


FIGURE 1 Smooth effective potentials U, U_B, and U_P for 0.1 eV double barriers at 300 K. x is in Å

where repeated indices are summed over and where n is the electron density, $\boldsymbol{u}$ is the velocity, m is the electron mass, P_{ij} is the stress tensor, V is the classical potential energy, W is the energy density, $\boldsymbol{q}$ is the heat flux, and T_0 is the ambient temperature. Collision effects are modeled by the relaxation time approximation, with momentum and energy relaxation times τ_p and τ_w. These transport equations are coupled to Poisson's equation for the electric potential.

Quantum effects enter through the expressions for the stress tensor and the energy density:

$$P_{ij} = -nT\delta_{ij} - \frac{\hbar^2 n}{4mT}\frac{\partial^2 \bar{V}}{\partial x_i \partial x_j} \tag{4}$$

$$W = \frac{3}{2}nT + \frac{1}{2}mnu^2 + \frac{\hbar^2 n}{8mT}\nabla^2 \bar{V} \tag{5}$$

where $T = 1/\beta$ is the electron temperature and the "quantum" potential $\bar{V}$ is

$$\bar{V}(\beta, \boldsymbol{x}) = \frac{1}{\beta}\int_0^\beta d\beta' \left(\frac{\beta'}{\beta}\right)^2 \int d^3X' \left(\frac{2m\beta}{\pi(\beta-\beta')(\beta+\beta')\hbar^2}\right)^{3/2} \times \exp\left\{-\frac{2m\beta}{(\beta-\beta')(\beta+\beta')\hbar^2}(\boldsymbol{X}'-\boldsymbol{x})^2\right\} V(\boldsymbol{X}'). \tag{6}$$

The quantum correction to the classical stress tensor and energy density is valid to all orders of $\hbar^2$ and to first order in βV, and involves both a smoothing integration of the classical potential over space and an averaging integration over temperature. The barrier height $\mathcal{B}$ is incorporated into the QHD transport equations (1)–(3) by replacing $V \to V + \mathcal{B}$. (Poisson's equation is not changed).

We define the 1D smooth effective potential in the momentum conservation equation (2) as the most singular part of $V - P_{11}$:

$$U = V + \frac{\hbar^2}{4mT}\frac{d^2\bar{V}}{dx^2}. \tag{7}$$

The double integration over both space and inverse temperature provides sufficient smoothing so that the P_{11} term in the smooth effective potential cancels the leading singularity in the classical potential at a barrier.

The 1D steady-state smooth QHD equations are discretized [3] using a conservative upwind method adapted from computational fluid dynamics. The discretized equations are then solved by a damped Newton method.

There are two contributions to the quantum potential $\bar{V}$: the double barrier potential and the "self-consistent" electric potential from Poisson's equation. Note that second derivatives of $\bar{V}$ appear in the stress tensor and energy density, which then are differenced in the smooth QHD transport equations. Thus we compute

$$\bar{V}'' = \bar{V}_B'' + \bar{V}_P''. \tag{8}$$

$\bar{V}_B''$ is just computed once since it only depends on the barriers and not on the applied voltage or state variables (n, u, T, V_P). In computing $\bar{V}_P''$, we first use Poisson's equation to obtain

$$\bar{V}_P''(\beta, x) = \frac{e^2}{\varepsilon}\int_0^\beta \frac{d\beta'}{\beta}\left(\frac{\beta'}{\beta}\right)^2 \int dX' \left(\frac{2m\beta}{\pi(\beta-\beta')(\beta+\beta')\hbar^2}\right)^{3/2} \tag{9}$$

$$\times \exp\left\{ -\frac{2m\beta}{(\beta-\beta')(\beta+\beta')\hbar^2} X'^2 \right\}$$
$$(N_D(X'+x) - n(X'+x))$$

where N_D is the density of donor ions. Then to efficiently compute the convolution (9), we take advantage of properties of the Fourier transform.

We present simulations of a GaAs resonant tunneling diode with $Al_xGa_{1-x}As$ double barriers at 300 K (77 K). The barrier height $\mathcal{B}$ is set equal to 0.1 (0.05) eV. The diode consists of n^+ source (at the left) and drain (at the right) regions with the doping density $N_D = 10^{18}$ cm^{-3}, and an n channel with $N_D = 5 \times 10^{15}$ cm^{-3}. The channel is 200 (250) Å long, the barriers are 25 (50) Å wide, and the quantum well between the barriers is 50 Å wide. Note that the device has 50 Å spacers between the barriers and the contacts. We have chosen parameters to highlight differences between the original and smooth QHD models.

Figure 1 illustrates the smooth effective potentials U_B (for the barriers), U_P (the Poisson contribution), and U (barrier plus Poisson contributions) for the resonant tunneling diode at 300 K at the voltage $V = 0.056$ where the $I-V$ curve peaks in Figure 2.

Smooth QHD simulations of the resonant tunneling diode exhibit enhanced negative differential resistance when compared to simulations using the original $O(\hbar^2)$ QHD model. The current-voltage curve for the resonant tunneling diode at 300 K is plotted in Figure 2 and at 77 K is plotted in Figure 3. It is interesting that the original $O(\hbar^2)$ QHD model (see Refs. [4, 3] and references therein) predict very different $I-V$ curves—in fact, at both 300 K and 77 K the original $O(\hbar^2)$ QHD model fails to produce negative differential resistance for these devices.

Simulations of the resonant tunneling diode using the Wigner-Boltzmann/Poisson equations are planned to determine which of the QHD models gives better agreement with the more complete quantum kinetics. In these comparisons, we will use Fokker-Planck collision terms with a relaxation time τ in the Wigner-Boltzmann equation, which then implies momentum and energy relaxation times $\tau_p = \tau$ and $\tau_w = \tau/2$ in the QHD models.

We will then be able to answer the question: In what parameter range and how accurately do the smooth QHD solutions (first three moments) reflect the solutions to the full Wigner-Boltzmann equation? There should be a technologically important range of parameters (device size, ambient temperature, potential barrier height, applied voltage, semiconductor material, etc.) in which the smooth QHD model gives solutions and $I-V$ curves that are very close to those given by the full Wigner-Boltzmann/Poisson system.

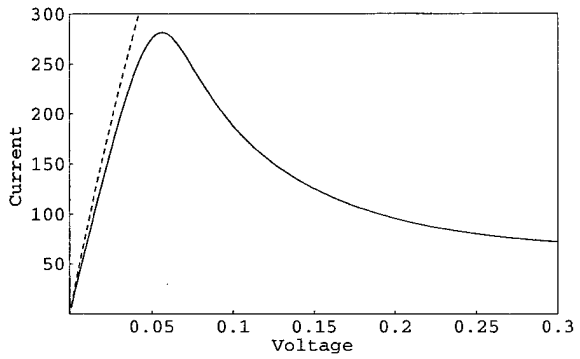


FIGURE 2 Current density in kiloamps/cm^2 vs. voltage for the resonant tunneling diode at 300 K. The solid curve is the smooth QHD computation and the dotted line is the $O(\hbar^2)$ computation. The barrier height is 0.1 eV.

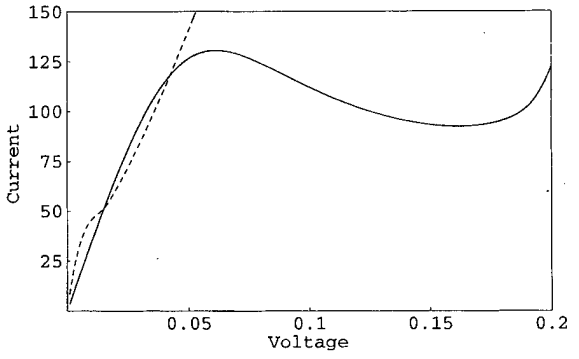


FIGURE 3 Current density in kiloamps/cm^2 vs. voltage for the resonant tunneling diode at 77 K. The solid curve is the smooth QHD computation and the dotted line is the $O(\hbar^2)$ computation. The barrier height is 0.05 eV.

References

[1] Gardner, C. L. and Ringhofer, C. (1996). "Smooth quantum potential for the hydrodynamic model," *Physical Review*, **E 53**, pp.157–167.

[2] Gardner, C. L. and Ringhofer, C. (1998). "Approximation of thermal equilibrium for quantum gases with discontinuous potentials and application to semiconductor devices," *SIAM Journal on Applied Mathematics*, accepted for publication.

[3] Gardner, C. L. (1994). "The quantum hydrodynamic model for semiconductor devices", *SIAM Journal on Applied Mathematics*, **54**, pp. 409–427.

[4] Grubin, H. L. and Kreskovsky, H. L. (1989). "Quantum moment balance equations and resonant tunnelling structures", *Solid-State Electronics*, **32**, pp. 1071–1075.

Authors' Bigraphies

Carl L. Gardner is Professor of Mathematics at Arizona State University. His current research interests lie in classical and quantum semiconductor device simulation and computational fluid dynamics.

Christian Ringhofer is Professor of Mathematics at Arizona State University. His current research interests include classical and quantum transport equations and moment models for semiconductor device modeling.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 147–151
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Spherical Harmonic Modeling of a 0.05 μm Base BJT: A Comparison with Monte Carlo and Asymptotic Analysis

C.-H. CHANG, C.-K. LIN, N. GOLDSMAN* and I. D. MAYERGOYZ

Department of Electrical Engineering, University of Maryland, College Park, MD 20742

We perform a rigorous comparison between the Spherical Harmonic (SH) and Monte Carlo (MC) methods of solving the Boltzmann Transport Equation (BTE), on a 0.05 μm base BJT. We find the SH and the MC methods give very similar results for the energy distribution function, using an analytical band-structure, at all points within the tested devices. However, the SH method can be as much as seven thousand times faster than the MC approach for solving an identical problem. We explain the agreement by asymptotic analysis of the system of equations generated by the SH expansion of the BTE.

Keywords: Device Modeling, Simulation, Spherical Harmonic, Boltzmann, Monte Carlo

1. INTRODUCTION

The Spherical Harmonic (SH) method for solving the BTE has given rise to considerable controversy. Researchers developing the method have reported promising results [1–3]. Other researchers have questioned whether the SH method can indeed provide the distribution function [4]. Still others have called for a study of various methods, including the SH, especially under conditions of rapid spatial variations [5].

In this paper, we respond to these questions by performing a rigorous comparison between the SH and MC methods of solving the Boltzmann equation, while using a 0.05 μm base BJT as a test vehicle. *We find the SH and the MC methods give very similar results for the energy distribution function, using an analytical band-structure, at all points within the tested devices. However, the SH method can be as much as seven thousand times faster than the MC approach for solving an identical problem.*

We explain the agreement between SH and MC by analysis of the SH system of equations. Asymptotic analyses show that, due to high scattering rates, high order SH terms decay quickly enough to have minimal effect on the energy distribution function, even for abrupt large spatial variations. Additionally, boundary conditions further restrict the population of higher

* Corresponding author.

order terms, thereby facilitating use of a low order SH expansion.

2. OVERVIEW OF SPHERICAL HARMONIC METHOD

Since the main focus of this paper is to demonstrate and explain why the SH method can give relatively good agreement with MC simulations, we begin with an overview of the SH approach. First, we write the BTE below:

$$\frac{1}{\hbar}\nabla_k \varepsilon \cdot \nabla_r f(\vec{r},\vec{k}) + \frac{q}{\hbar}\nabla_r \phi(\vec{r}) \cdot \nabla_k f(\vec{r},\vec{k}) = \frac{V}{(2\pi)^3}\sum_i \int [f(\vec{r},\vec{k}')S_i(\vec{k}',\vec{k}) - f(\vec{r},\vec{k})S_i(\vec{k},\vec{k}')]d^3\vec{k}' \tag{1}$$

where $\phi(\vec{r})$ is the potential, $f(\vec{r},\vec{k})$ is the electron distribution function; $\vec{k}$ is the electron wave-vector; $\vec{r}$ is the position vector; $\hbar$ is Planck's constant, and the sum is over the various collision processes represented by the transition rates S_i. Now, we express the distribution function in terms of the following SH expansion:

$$f(\vec{r},\vec{k}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} f_l^m(\vec{r},\varepsilon)Y_l^m(\theta,\psi) \tag{2}$$

here $Y_l^m(\theta,\psi)$ are the spherical harmonic basis functions, θ and ψ are the polar and azimuthal angles of the wave-vector, respectively. $f_l^m(\vec{r},\varepsilon)$ represents the expansion coefficients. The eventual goal of the approach is to determine the coefficients and thereby the distribution function for the device. We showed previously that by substituting the spherical harmonics expansion for the distribution function into the BTE, while taking advantage of the recurrence and orthogonality relationships between spherical harmonics the LHS of the BTE can be transformed into an infinite set of identical expressions for the coefficients. We also showed the if we take scattering to be either isotropic or elastic (which is a reasonable approximation in silicon), and make use of the addition theorem, the collision integral can be transformed into a relatively simple expression [1, 2]. Putting the LHS and the collision integral together gives the following equation for the zero order coefficient f_0^0:

$$\sum_{i=1}^{2}\left\{v(\varepsilon)\left[\frac{\partial}{\partial x_i} - qE_i(\vec{r})\left(\frac{\partial}{\partial\varepsilon} - \frac{l-1}{2}\frac{\gamma'(\varepsilon)}{\gamma(\varepsilon)}\right)\right]\hat{a}_i^+\right\}f_0^0 = \frac{f_0^0(\vec{r},\varepsilon \mp \Delta\varepsilon)}{\tau_{0,i}^{\text{in}\mp}(\varepsilon)} - \frac{f_0^0(\vec{r},\varepsilon)}{\tau_{0,i}^{\text{out}\mp}(\varepsilon)} \tag{3}$$

While the expressions for the coefficients for $l>0$ are

$$\sum_{i=1}^{2}\left\{v(\varepsilon)\left[\frac{\partial}{\partial x_i} - qE_i(\vec{r})\left(\frac{\partial}{\partial\varepsilon} - \frac{l-1}{2}\frac{\gamma'(\varepsilon)}{\gamma(\varepsilon)}\right)\right]\hat{a}_i^+ + v(\varepsilon)\left[\frac{\partial}{\partial x_i} - qE_i(\vec{r})\left(\frac{\partial}{\partial\varepsilon} + \frac{l+2}{2}\frac{\gamma'(\varepsilon)}{\gamma(\varepsilon)}\right)\right]\hat{a}_i^-\right\}f_l^m = -\frac{f_l^m(\vec{r},\varepsilon)}{\tau_{l,i}^{\text{out}\mp}(\varepsilon)} \tag{4}$$

where $v(\varepsilon) = \sqrt{2m\gamma(\varepsilon)}/m\gamma'(\varepsilon), \gamma'(\varepsilon) = d\gamma(\varepsilon)/d\varepsilon$, $\gamma(\varepsilon)$ represents the dispersion relation, and $E_i(\vec{r})$ is the electric field in the i direction. The sum over i represents the Cartesian directions of a 2-D device cross-section in real-space. The raising and lowering operators $\hat{a}_i^{\pm}$ which we developed to relate the coefficients are [1]:

$$\hat{a}_1^- f_l^m \equiv \frac{1}{2}\{-\alpha_{l-1}^m\alpha_l^m f_{l-1}^{m-1} + \alpha_{l-1}^{-m}\alpha_l^{-m} f_{l-1}^{m+1}\}$$

$$\alpha\hat{a}_1^+ f_l^m \equiv \frac{1}{2}\{\alpha_{l+1}^{-m+1}\alpha_l^{-m+1} f_{l+1}^{m-1} - \alpha_{l+1}^{m+1}\alpha_l^{m+1} f_{l+1}^{m+1}\}$$

$$\hat{a}_2^- f_l^m \equiv \alpha_{l-1}^{-m+1}\alpha_l^m f_{l-1}^m; \quad \hat{a}_2^+ f_l^m \equiv \alpha_{l+1}^m\alpha_l^{-m+1} f_{l+1}^m;$$

$$\alpha_l^m = \sqrt{\frac{l+m}{2l+1}}$$

where

$$\frac{1}{\tau_{l,i}^{\text{out}\mp}(\varepsilon)} = \frac{V}{(2\pi)^3}\int S_{l,i}^{\mp}(\varepsilon,\varepsilon')h(\varepsilon')d\varepsilon' : \quad \frac{1}{\tau_{l,i}^{\text{in}\mp}(\varepsilon)} = \frac{V}{(2\pi)^3}\int S_{l,i}^{\mp}(\varepsilon',\varepsilon)h(\varepsilon')d\varepsilon' \tag{5}$$

In (5) we have introduced $S_{l,i}$, which is l'th coefficient of the Legendre polynomial expansion of the i'th type of scattering mechanism [2], as well as the density of states $h(\varepsilon) = 4\sqrt{2}\pi/\hbar^3(m^{*3/2}\gamma^{1/2}(\varepsilon)\gamma'(\varepsilon))$. The superscript *out* refers to a process where the electron is scattered out of the state $\vec{k}$ to another state $\vec{k}'$, while the superscript *in* represents scattering into $\vec{k}$. The superscripts $\mp$ in (5) refer to absorption (−) and emission (+) of energy respectively. (For elastic processes $\mp$ have been eliminated, since there is no energy emission or absorption. In cases where there is no process for energy emission (absorption), the superscript + (−) has been eliminated).

While a system of equations has been obtained to arbitrarily high order, we must truncate the system to actually obtain a solution. Since the scattering rate in silicon is relatively high, and silicon is a covalent semiconductor, a relatively low order truncation should be applicable. (This truncation will be verified later in the paper). We therefore use an expansion which includes the first four spherical harmonics ($Y_0^0, Y_1^{-1}, Y_1^0, Y_1^0$). After truncating, the total energy H is used as the independent energy variable as opposed to the kinetic energy ε, ($H=\varepsilon-q\phi$) [3]. We can then arrive at the following tractable expression for the symmetrical part of the distribution function throughout the device [1, 2]:

$$
\begin{aligned}
-\nabla\cdot\left[\frac{\gamma^{3/2}(H+q\phi)\tau(H+q\phi)}{\gamma'(H+q\phi)}\nabla F_0^0(x,y,H)\right] &= \frac{3(\varrho-1)D_n^2 m^{*5/2}}{2\sqrt{2}\pi\hbar^3\rho\omega_n}\left(\frac{\gamma^{1/2}(H+q\phi)\gamma'(H+q\phi)}{e^{\hbar\omega_n/K_0T}-1}\right)\\
&\{\sqrt{\gamma(H+q\phi-\hbar\omega_n)}\gamma'(H+q\phi-\hbar\omega_n)\\
&\times[F_0^0(x,y,H-\hbar\omega_n)-e^{\hbar\omega_n/K_0T}F_0^0(x,y,H)]\\
&+\sqrt{\gamma(H+q\phi+\hbar\omega_n)}\\
&\times\gamma'(H+q\phi+\hbar\omega_n)[e^{\hbar\omega_n/K_0T}F_0^0(x,y,H+\hbar\omega_n)\\
&-F_0^0(x,y,H)]\} \qquad (6)
\end{aligned}
$$

In (6) D_n = optical phonon deformation potential; ω_n = optical phonon vibrational frequency; F_0^0 = symmetrical part of electron distribution; $1/\tau$ is the total phonon scattering rate.

3. BJT SIMULATION RESULTS AND DISCUSSION OF TRUNCATION

The main point of this paper is to demonstrate that a low order truncation is indeed suitable for determining the energy distribution function in semiclassical silicon devices under the spherical band approximation. To show this, we used the above formulation to simulate a 0.05 μm base BJT. We chose the BJT because it represented a very aggressive challenge, with electric fields pointing in both directions, a narrow base of 50 nm, with fields varying as rapidly as 200 kV/cm over distances as small as 20 nm. As a benchmark, we also simulated the same device using the Monte Carlo approach. The input parameters (analytical band structure, intervalley, optical and acoustic phonon scattering) were identical for both the SH and MC calculations. Since the main objective was to determine if the low-order SH method could respond to the rapid field variations of the BJTs, we performed both the Monte Carlo and the Spherical Harmonic calculations as post-processor simulations.

We show some example results below. In Figures 1a and 1b we show the doping profile and resulting electric field with an applied potential of $V_{BE}=1.0$ V and $V_{CB}=3.0$V. In Figures 2a and 2b we show the 3-dimensional result for the energy distribution function calculated along the device. From a qualitative point of view, both distributions look alike, although the one generated by the SH approach is populated over the complete range of the calculation whereas the MC results does not have data for high energies due to the time consuming stochastic nature of the MC process. In Figures 3a and 3b we show the energy distribution in more detail by plotting its values at regular intervals along the device. In the plots, we compare the SH results (solid lines) with the MC results (open circles). Clearly, the SH and MC

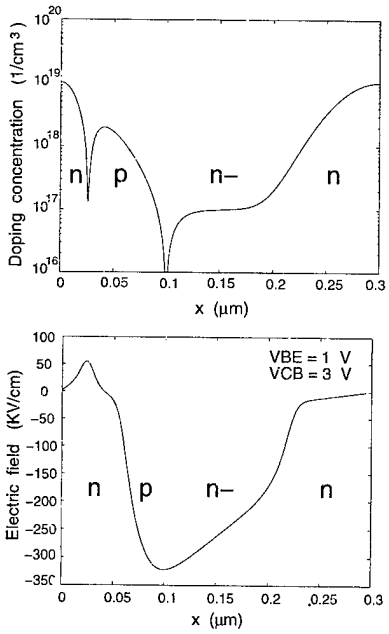


FIGURE 1 The doping concentration (top) and electric field (bottom) for the 0.05 μm *p*-base Gaussian doped BJT with $V_{BE} = 1.0$ V and $V_{CB} = 3.0$ V.

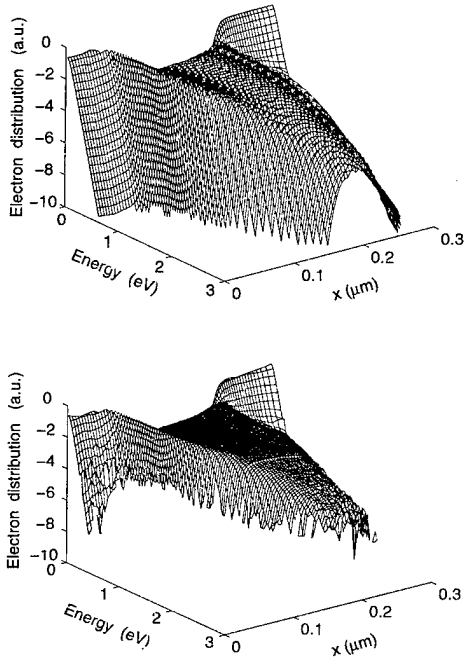


FIGURE 2 The calculated electron distribution function by SH-BTE (top) and MC (bottom) for the 0.05 μm *p*-base Gaussian doped BJT.

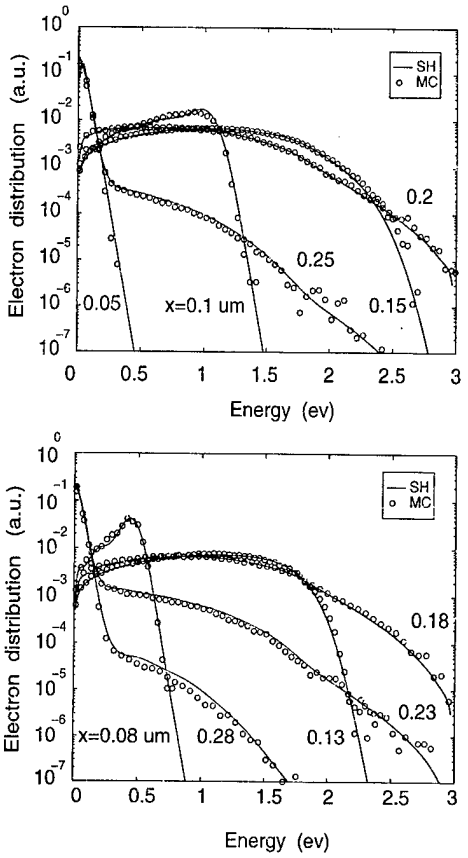


FIGURE 3 Agreement of the electron distribution function between SH-BTE and MC for the 0.05 μm *p*-base Gaussian doped BJT.

results are in very good agreement for the entire distribution function. We find this to be an extremely encouraging result, considering that the SH method required only 10 sec to evaluate, while the MC approach took more than 10 hours on a DEC Alpha 266 MHz workstation. The large difference in CPU time is attributable to the deterministic nature of the SH approach, while the MC method requires considerable time to accumulate reasonable statistics for the high energy tail and to surmount the large potential barrier at the emitter-base junction. (The MC could probably be optimized but our attempts at statistical enhancement proved unreliable when subjected to scrutiny).

We explain the agreement between SH and MC by asymptotic analysis. First of all, for very large values of SH index l, $l+1 \approx l$, and $F_{l+1} \approx F_l$. Under these conditions, Eq. (4) reduce to the following relatively simple expression.

$$v(H+q\phi)\frac{\partial F_n(x,H)}{\partial x} = -\frac{F_l(x,H)}{\lambda(H+q\phi)} \qquad (7)$$

where $\lambda(H+q\phi) = v(H+q\phi)\tau(H+q\phi)$ is a mean free path, the value of which depends on energy. For large energy λ becomes approximately constant, while for lower energies λ can be approximated as the following piecewise linear expression: $\lambda(H+q\phi) = \alpha \cdot (H+q\phi) + \beta$ within $[x_0, x]$ for a constant H. By integrating (7) over position space from x_0 to x for a constant H, using the chain rule $dx = \frac{dx}{d\varepsilon} d\varepsilon$, as well as approximating $E(x)$ as constant in the small interval, we can obtain

$$F_l(x,H) = F_l(x_0,H)\left[\left(\frac{\lambda(H+q\phi)}{\lambda(H+q\phi_0)}\right)^{\frac{1}{\alpha q E(x)}}\right] \qquad (8)$$

which simplifies to the following expression in the high energy region where λ is constant.

$$F_l(x,H) = F_l(x_0,H)\exp\left(-\frac{x-x_0}{\lambda}\right) \qquad (9)$$

In high energy, (9) shows that F_l will decay exponentially in x-space with a characteristic length of λ. Depending on the transport model one is using, λ ranges from approximately 2 nm to 5 nm for energies greater than 1.5 eV in silicon. This indicates that high-energy spherical harmonics decay very quickly in comparison with typical dimensions in deep submicron devices. For regions of low energy and low electric field, (8) ensures that F_l also decays very rapidly with distance. (While this is not immediately obvious, it is due to the relationship between ϕ and E, and the fact that λ is monotonically decreasing, ensures the exponent in (8 becomes very large, and the term in the square brackets is always less than unity). For low to moderate energies, where there is a transition from a low electric field to a high one, or visa-versa, the monotonic nature of λ again insures that the square bracketed term in (8) will be small.

In addition to the asymptotic analysis for l, physical boundary conditions also suppress the magnitude of high order terms. For example, the physical requirement that the distribution function be single valued at the energy $\varepsilon = 0$ (is not dependent on angle at zero energy), leads to the requirement that all SH components with $l > 0$ must vanish at zero energy. Furthermore, the requirement that the distribution function is close to Maxwellian at the device contacts implies that higher order SH terms are negligible at these boundaries. Also, since in silicon the scattering rate is fairly high, and since it is a covalent semiconductor, we usually take the major scattering rates to be isotropic. This leads to scattering which tends to give rise to a distribution function which is largely spherical in nature. Finally, as one examines the generalized SH system given by Eqs. (3) and (4), it is clear that the equations are only coupled to nearest neighbors in l. It is therefore difficult to propagate information from low order terms to higher order ones, thereby minimizing the relative importance of higher order SH coefficients.

References

[1] Hennacy, K. A., Wu, Y.-J., Goldsman, N. and Mayergoyz, I. D. (1995). "Deterministic MOSFET Simulation using a Generalized Spherical Harmonic Expansion of the Boltzmann Equation", *Solid-State Electronics*, **38**, 1498–1495.

[2] Liang, W., Goldsman, N., Mayergoyz, I. and Oldsiges, P. (1997). "2-dMOSFET modeling including surface effects and impact ionization by self-consistent solution of the Boltzmann, Poisson and hole-continuity equations", *IEEE Transactions on Elec. Dev.*, **44**, 257–276.

[3] Gnudi, A., Ventura, D. and Baccarani, G. (1993). "Modeling Impact Ionization in a BJT by Means of Spherical Harmonics Expansion of the Boltzmann Transport Equation", *IEEE Transactions on CAD*, **12**, 1706.

[4] Bude, J., Fischetti, M. and Laux, S. (1996). Independent private communications.

[5] Dutton, R., Hess, K. and Ravaioli, U. (1996). Independent private communications.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 153–157
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Numerical Examination of Photon Recycling as an Explanation of Observed Carrier Lifetime in Direct Bandgap Materials

JOSEPH W. PARKS Jr.[a], KEVIN F. BRENNAN[a,*] and ARLYNN W. SMITH[b]

[a] *Microelectronics Research Center, Georgia Institute of Technology, Atlanta, GA 30332;*
[b] *ITT Night Vision, Roanoke, VA 24019*

Photon recycling is examined as an explanation for the observed large carrier lifetimes in an InP/InGaAs photodiode. This effect extends the effective carrier lifetime within a device by re-absorbing a fraction of the photons generated through radiative band-to-band recombination events. In order to predict the behavior of this carrier generation, photon recycling has been added to our two-dimensional macroscopic device simulator, STEBS-2D. A ray-tracing preprocessing step is used to map all of the possible trajectories and absorption of various wavelengths of emitted light from each originating node within the device. The macroscopic simulator uses these data to determine the spatial location of the re-absorbed radiation within the geometry of the device. By incorporating the ray tracer results with the total quantity and spectral content of recombined carriers at each node within the simulation, the recycled generation rate can be obtained. A practical application of this model is presented where the effects of photon recycling are used as a possible explanation of the discrepancy between the theoretically predicted and experimentally observed radiative recombination rate in a double heterostructure photodetector.

Keywords: Photon recycling, reabsorption, self-excitation, macroscopic simulation

1. INTRODUCTION

It is often observed that radiative band-to-band transitions are the dominant recombination mechanism in many high quality direct gap semiconductors. Those carriers which recombine radiatively conserve energy through the liberation of a photon with a characteristic energy near the semiconductor bandgap. Photon recycling is the reabsorption of this emitted light elsewhere in the device to produce additional electron-hole pairs. Thus, photon recycling can be described as an additional generation mechanism which is tightly coupled to the total amount of radiative recombination over the range of the device. The effectiveness of photon recycling depends on many factors such as the dominance of

* Corresponding author.

the radiative recombination mechanism and the light confinement properties of the structure.

The influence of photon recycling on the performance of many devices has been investigated [1 – 3], and in several cases it has been shown that neglecting this effect leads to improper analysis of the carrier transport. Therefore, a general model including the influence of photon recycling should be included within advanced macroscopic device simulators. Unfortunately, many of the models used to analyze photon recycling have relied on either one-dimensional or analytical approximations to estimate the distribution of the generated carriers, and general multi-dimensional models don't exist.

The primary impetus behind this study is to develop a fully numerical, general two-dimensional photon recycling model which can be directly coupled into the STEBS-2D [4] macroscopic device simulator. To achieve the desired result, ray tracing techniques are employed to monitor the trajectories and absorption of light propagating through the device geometry. Since the optical properties of the devices of interest are spatially inhomogeneous, and since the generated light from the recombination events has a non-uniform spectral content, the ray tracer must be used to map the absorption of light originating from any point in the device over a range of wavelengths.

As an example of the utility of the model, an InP/InGaAs double heterostructure photodiode [5] is examined to determine whether the photon recycling phenomenon can be used to explain the discrepancy between theoretically predicted and experimental lifetimes. Experimentally, the lifetime is measured by monitoring the magnitude of the collected current as a light source is moved across the top of the device. Clearly from the geometry of the structure and the nature of the experiment, one-dimensional or analytical models are insufficient and a two-dimensional analysis is necessary.

2. NUMERICAL MODEL

The overall goal in the implementation of photon recycling is the mapping of the emission of photons at one point to their absorption at all other points in the simulation domain. This requires a discrete approach at the positional as well as the spectrum level. The spatial discretization of the simulation domain follows that of the electrical model, while the continuum of emitted wavelengths is broken into several discrete energy ranges representative of the wavelengths which could originate from a given material.

Ray tracing techniques are used here to model the propagation of light through the crystal. A general purpose, three dimensional, ray tracing algorithm based on the idea of constrained volumes is used to provide the mapping of light emission to absorption [6]. The emitted rays are tracked from volume to volume, and the distance through each is used to determine the total absorption at each node. As the rays move between volumes of differing material, they are refracted and secondary reflected rays are generated. Implicit in this model are features such as wavelength dependent refraction, critical angle calculation, wavelength and angle dependent transmission coefficients, and various surface reflection models including antireflection coatings.

The ray tracing itself is fairly straight forward. From each node, a large number of rays are emitted in random directions. The path of each ray is followed until either the ray is attenuated or is coupled out of the device without possibility of return. For each initiating point, the detailed absorption within the device is catalogued. This process must be repeated for each simulated wavelength owing to the differing absorption and refraction indices. These data are arranged as a set of cross-referenced lists relating the initiating point of emission, the points contributing to the absorption, and the amount of attenuation.

The information from the ray tracing is incorporated into the source term of the current continuity equations of the macroscopic simulator as seen in the flowchart given in Figure 1. Within each iteration, the spectrum of the spontaneously emitted light is obtained for each node, (x, y). The energy dependent spontaneous emission rate, R, is estimated through the use of the van-Roosbroeck Shockley relation given by Eq. (1)

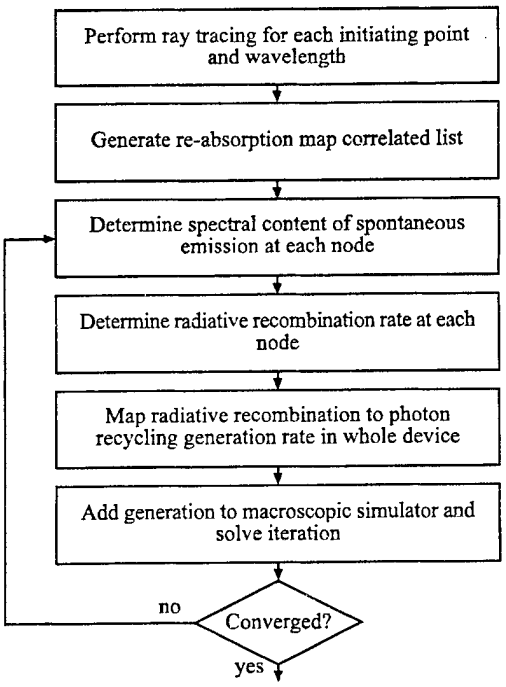


FIGURE 1 Flowchart depicting steps within the applied photon recycling model.

$$R_{(x,y)}(v)dv = \frac{8\pi(hv)^2 n(v)^2 \alpha(v)}{h^2c^2\left[\exp\left(\frac{hv-\Delta E_{f(x,y)}}{kT}\right)-1\right]} \quad (1)$$

where ΔE_f is the current estimate of the difference between the electron and hole quasi-Fermi levels, and α and n are the wavelength dependent absorption and index of refraction [7]. Next, an estimate of the total radiative recombination rate at the node is determined by:

$$R_{\mathrm{RAD}(x,y)} = B(n_{(x,y)}p_{(x,y)} - n_i^2) \quad (2)$$

where the radiative constant B is taken to be the integral of Eq. (1) divided by n_i^2. The normalized spectrum produced from Eq. (1) is divided into non-overlapping energy ranges centered around the discrete wavelengths simulated by the ray tracer. The fraction, $f_{(x,y,\lambda)}$, of the overall spontaneous rate for each wavelength is taken to be the integral of Eq. (1) over each energy range normalized to the integral of the total spectrum. Using the output from the ray tracer, the total photon recycling generation rate at any point, (l, k) can be determined by:

$$G_{(l,k)} = \sum_{x,y}^{\Omega} R_{\mathrm{RAD}(x,y)} \frac{V_{(x,y)}}{V_{(l,k)}} \left[\sum_{\lambda} f_{(x,y,\lambda)} r_{(x,y,l,k,\lambda)}\right] \quad (3)$$

where r is the ray tracer results describing the fraction of the absorbed power at (l, k) due to emission at (x, y) for wavelength λ and $V_{(x,y)}$ and $V_{(l,k)}$ are the control volume areas for the originating and absorbing elements.

The macroscopic model used in this study is the drift-diffusion subset of the hydrodynamic simulator. Here, Poisson's equation is coupled to the current continuity equations for inhomogeneous materials [4]. By including the generation rate from Eq. (3) into the continuity equations and iterating until a self consistent solution is obtained, the effects from photon recycling may be fully included into the device simulation.

3. SIMULATION RESULTS

The aforementioned model, included in the macroscopic device simulator, is used to examine the affect of photon recycling upon the performance of the device sketched in Figure 2. For this device, excess carriers are produced from a 1.3 micron monochromatic light source ten microns in diameter moved laterally across the top surface. This light is absorbed exclusively in the InGaAs with an internal quantum efficiency of 97 percent. For the device in question, the measured minority carrier lifetime is approximately three times larger than that predicted from the theoretical radiative recombination rate. Photon recycling is conjectured here to be a possible cause of this discrepancy.

The two-dimensional device simulator cannot accurately capture the inherent three-dimensional nature of the diffusion of excess carriers in the original experiment. Therefore an assumption

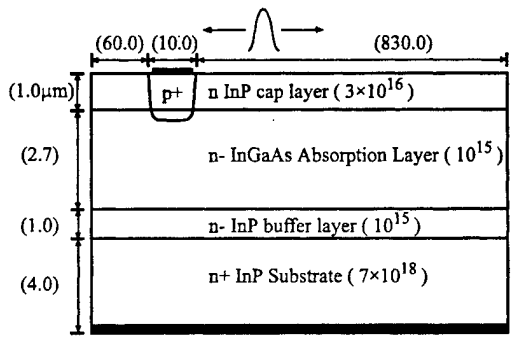


FIGURE 2 Illustration of device geometry and doping profiles for the photodiode under investigation. The device is nominally reverse biased to five volts and carrier excitation is through photogeneration from the laterally moving light source. All dimensions are in microns and doping in cm^{-3}.

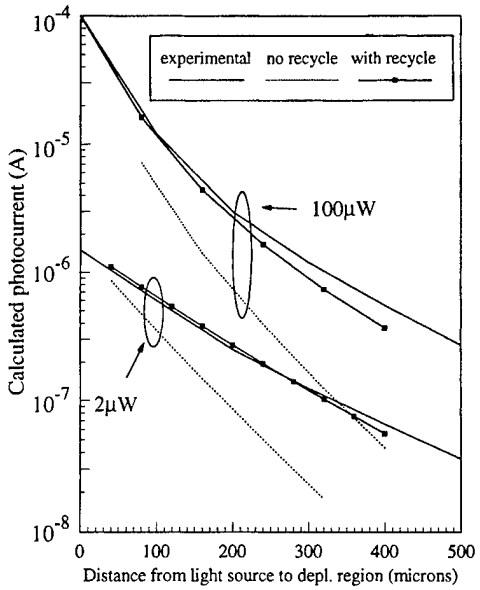


FIGURE 3 Calculated and experimental collection currents for the double heterostructure photodiode under 2 μW and 100 μW incident optical power. Each case illustrates the calculated current both with and without the inclusion of photon recycling.

must be made upon the value of the non-simulated third dimension of the device. The calculated output current is first made to agree with the reported results using the experimentally obtained lifetime by adjusting the level of carrier injection by an appropriate choice of length for the third dimension. Once this control is established, the simulation is again performed using the lower, theoretically predicted, lifetime. The results of this simulation are shown by the dashed curve in Figure 3 for two different light intensities. Comparison of the experimental results to the calculations made neglecting photon recycling with the theoretical lifetime clearly show that the output current is greatly underestimated. The simulation is also performed with the inclusion of photon recycling. The curve marked with the solid squares in Figure 3 shows the calculated photocurrent including photon recycling assuming the theoretical carrier lifetime. Comparison of the experimental results to the calculations including photon recycling shows that the inclusion of the additional photon recycling generation rate increases the magnitude of the collected current to a value clearly resembling that of experiment. From these results, it is observed that photon recycling provides the mechanism to describe the disparity between the experimental and theoretical lifetimes observed in this device.

4. CONCLUSIONS

This paper describes a general purpose model used to include the effects of photon recycling into a macroscopic device simulator. A three-dimensional ray tracer is used to predict how spontaneously emitted light propagates and is absorbed within a semiconductor device. These data are then coupled to a two-dimensional macroscopic device simulator and are used in conjunction with the quantity and spectral content of the radiative recombination rate to predict the total carrier generation due to photon recycling. As an example of the model, the effect of photon recycling is examined in a double heterostructure InP/InGaAs photodiode where it is observed that the inclusion of photon recycling within the InGaAs restores the

agreement between the experimentally and theoretically obtained values of carrier lifetime.

References

[1] Durbin, S. M. and Gray, J. L., "Numerical Modeling of Photon Recycling in Solar Cells", *IEEE Trans. Elec. Dev.*, **41**(2), pp. 239–245, Feb. 1994.

[2] Asbeck, P., "Self-Absorption Effects on the Radiative Lifetime in GaAs-GaAlAs Double Heterostructures", *J. Appl. Phys.*, **48**(2), pp. 820–822, Feb. 1977.

[3] Sieg, R. M. and Ringel, S. A., "Reabsorption, Band-gap Narrowing, and the Reconciliation of Photoluminescence Spectra with Electrical Measurements for Epitaxial n-InP", *J. Appl. Phys.*, **80**(1), pp. 448–458, July. 1996.

[4] Parks, J. W., Smith, K. F., Brennan, K. F. and Tarof, L. E., "Theoretical Study of Device Sensitivity and Gain Saturation of Separate Absorption, Grading, Charge and Multiplication InP/InGaAs Avalanche Photodiode", *IEEE Trans. Elec. Dev.*, **43**(12), pp. 2113–2121, Dec. 1996.

[5] Gallant, M. and Zemel, A., "Long Minority Hole Diffusion Length and Evidence for Bulk Radiative Recombination Lifetime in InP/InGaAs Double Heterostructures", *Appl. Phys. Lett.*, **52**(20), pp. 1686–1688, May 1988.

[6] Smith, A. W., Rohatgi, A. and Neel, S. C. (1990). "TEXTURE: A Ray Tracing Program for the Photovoltaic Community", *Proc.* 21st *IEEE Photovoltaic Spec. Conference.*

[7] Pankove, J. I. (1971). *Optical Processes in Semiconductors*, Prentice-Hall, Englewood Cliffs, N. J.

Authors Biographies

Joseph W. Parks Jr. was born in Oak Ridge, TN on May 21, 1970. He received his B.S. degree in electrical engineering from the University of Tennessee, Knoxville in 1992 and his M.S.E.E. from the Georgia Institute of Technology, Atlanta, GA, in 1993. He is presently working on his Ph.D. degree also in electrical engineering at the Georgia Institute of Technology. His research work involves the numerical modeling of semiconductor devices with emphasis in the drift-diffusion and hydrodynamic simulation of photodetectors and avalanche photodiodes.

Kevin F. Brennan received the B.S. degree in physics from the Massachusetts Institute of Technology, Cambridge, in 1978, and the M.S. degree in physics and Ph.D. degree in electrical engineering from the University of Illinois, Urbana-Champaign, in 1984.

He is currently Institute Fellow and Professor, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta. His current research interests include the physics and modeling of semiconductor devices. Of particular interest are the physics and modeling of avalanche photodiodes, confined state ionization devices, high field effects in semiconductors, photoconductors, and high speed transistors.

Dr. Brennan was the recipient of a Presidential Young Investigator Award through the National Science Foundation.

Arlynn W. Smith received the B.S. degree in ceramic engineering from Alfred University, Hornell, NY, in 1984, the M.S. degree from the Georgia Institute of Technology, Atlanta, GA, in 1987, and the Ph.D. degree in electrical engineering from Georgia Tech. in 1992.

From 1992 to 1993, he held a post-doctoral fellowship at the Georgia Tech Research Institute, where he developed a three-dimensional simulation code for modeling the chemical vapor infiltration of ceramic matrix composites. From 1993 to 1995, he performed a post-doctoral fellowship with the microelectronics research center at Georgia Tech, where he developed a non-parabolic hydrodynamic simulation code for heterojunction acoustic charge transport, charge transport devices, and avalanche photodiodes. Currently, Dr. Smith is working for ITT Night Vision in Roanoke, VA developing advanced cathode structures for night vision applications.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 159–163
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Electronic Structure Calculations Using An Adaptive Wavelet Basis

D. A. RICHIE[a,*], P. VON ALLMEN[a], K. HESS[a] and RICHARD M. MARTIN[b]

[a] *Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana, Illinois 61801;*
[b] *Department of Physics, University of Illinois, Urbana, Illinois 61801*

The use of a wavelet basis can lead to efficient methods for performing *ab initio* electronic structure calculations of inherently localized structures. In this work wavelets are used to construct an adaptive basis which is optimized dynamically throughout the calculation. The computational effort of such a method should scale linearly with the number of basis functions. The adaptive basis is tested for the case of bulk Si using only a local s-pseudopotential.

Keywords: Wavelets, multiresolution analysis, electronic structure, density functional theory, adaptive basis

INTRODUCTION

Much progress has been made in the area of *ab initio* electronic structure calculations using the local density approximation (LDA) of density functional theory (DFT). Within DFT the total energy is considered to be a functional of the electron density. The electronic ground state is found by minimizing this total energy, leading to the solution of the Kohn-Sham single-particle Schrödinger equation.

Despite its advantages, the commonly used plane wave (PW) basis possesses serious drawbacks when applied to localized systems. Examples of such systems include those with deep atomic pseudopotentials, e.g., oxygen, or structures of a localized nature, e.g., molecules, surfaces and interfaces. In such cases a prohibitively high energy cutoff is required to provide the necessary resolution. Due to the delocalized nature of the PW basis functions, this increase in resolution is applied globally even though the increased resolution is needed in an essentially localized region of space.

As a result of these problems various alternatives to the PW method have been proposed. Most can be classified as real-space methods. These include the use of adaptive curvilinear coordinates [1], finite difference [2] and multigrid [3] methods as well as wavelet bases. It is the use of a wavelet basis which will be the topic of this paper.

*Corresponding author.

The use of a wavelet basis for electronic structure calculations was first proposed by Cho *et al.* [4] where non-orthonormal wavelets were used. The work of Wei and Chou [5] used the compact orthonormal wavelets constructed by Daubechies [6]. The same Daubechies wavelets are used in our work. In addition, both Ref. [4] and Ref. [5] used a basis that was predetermined by simple physical arguments and remained fixed throughout the calculation. In our work a truly adaptive basis is used which does not rely on any *a priori* assumptions, but instead allows the calculation to dynamically determine the optimum basis for each electronic orbital. Our work is also unique in that the entire calculation is performed in a wavelet basis. This creates the possibility for a general method where both the execution time and storage scale linearly with the number of basis functions used to represent each electronic orbital.

WAVELET BASIS

A detailed discussion of wavelets can be found in Ref. [6]. A wavelet basis forms a multiresolution analysis (MRA). Such an analysis represents a function as the limit of increasingly finer approximations, with each approximation being the result of smoothing, or averaging, the function over some fixed length scale. The result is a series of embedded function spaces V_m such that

$$\cdots V_{m+2} \subset V_{m+1} \subset V_m \subset V_{m-1} \subset V_{m-2} \cdots$$

A basis spanning each function space, or approximation space, can be constructed from the translations of some localized function $\phi(x)$ known as the scaling function. Thus, a basis for the approximation space V_m can be written $\{\phi_{mn}(x) = 2^{m/2}\phi(2^m x - n),\ n \in \boldsymbol{Z}\}$. The corresponding characteristic length scale is $\sim 2^{-m}$. The MRA can be seen to be generated by this 'mother scaling function' $\phi(x)$. The choice of the mother scaling function is in fact the beginning of the MRA which generates a wavelet basis.

Next, consider the projection of a function onto two successive approximation spaces V_m and V_{m+1}. Information is lost when going from the finer approximation to the coarser one. This leads to the introduction of an additional function space W_m defined to be the orthogonal complement of V_m in V_{m+1}, i.e., $V_m \oplus W_m = V_{m+1}$ and $V_m \perp W_m$. This function space, or detail space, W_m contains the information which is lost when making the coarser approximation. Conversely, the finer approximation of a function can be constructed from its projection onto the coarser approximation space plus its projection onto the associated detail space.

A basis spanning each detail space can be constructed from the translations of some localized function $\psi(x)$, known as a wavelet. Thus, W_m may be spanned by $\{\psi_{mn} = 2^{m/2}\psi(2^m x - n),\ n \in \boldsymbol{Z}\}$. Here the function $\psi(x)$ is known as the 'mother wavelet' and is directly related to the associated mother scaling function used to generate the MRA. It is the combination of scaling functions and wavelets which are commonly referred to as a wavelet basis.

There exist a wide variety of wavelets constructed to have different properties. In this work we use two types of wavelets, specifically the wavelet known as Daubechies-6 ($\mathcal{D}_6$) and the Haar wavelet. In Figure 1 and Figure 2 the mother scaling function and mother wavelet are shown for $\mathcal{D}_6$. Both wavelets are compact, complete and orthonormal. In addition, the $\mathcal{D}_6$ wavelet has a continuous first and second derivative.

The expansion of some function $f(x)$ onto a wavelet basis will therefore have the form of an expansion onto the translations of some coarse scaling function plus the detail of the expansion onto the translations and dilations of the associated wavelet,

$$f(x) = \sum_{n=-\infty}^{+\infty} \bar{f}_{0n}\phi_{0n}(x) + \sum_{m=0}^{+\infty}\sum_{n=-\infty}^{+\infty} f_{mn}\psi_{mn}(x). \quad (1)$$

The infinite sums over all translations can be removed by imposing periodic boundary conditions, which leads to periodized wavelets. A three

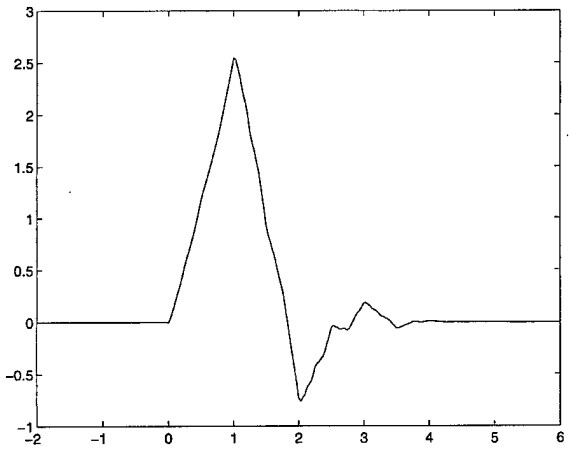


FIGURE 1 Real space plot of the Daubechies-6 mother scaling function.

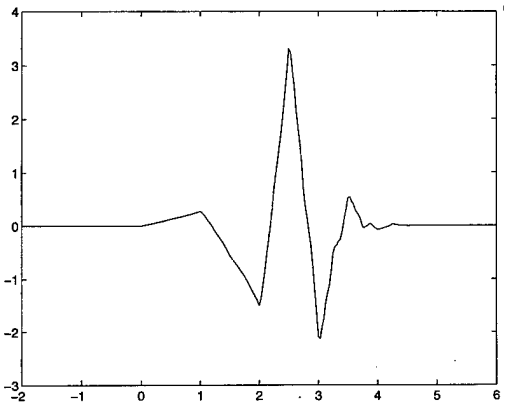


FIGURE 2 Real space plot of the Daubechies-6 mother wavelet.

dimensional basis can be constructed from tensor products of the one dimensional basis.

One of the key properties of a wavelet basis is its ability to represent the localized detail of a function using a small number of wavelets. This is a result of the fact that the wavelet expansion coefficients, f_{mn} in Eq. (1), will decay rapidly with increasing $m > M$, where the *local* smoothness of the function can be characterized by the length scale $\sim 2^{-M}$. *If the magnitude of the expansion coefficient is less than some threshold which is small compared to the norm of the function, then no further detail is needed in that region.* This leads to the idea of a 'compressed basis', the structure of which is determined by the above criteria, forming a minimal basis set representation for a specific function. Such a compressed basis will be used to represent the electronic orbitals in the electronic structure calculation.

CALCULATION

The electronic structure of the system is determined here using the local density functional theory neglecting the spin. Therefore we determine the electronic ground state by solving the Kohn-Sham equation[1],

$$H_{\mathrm{KS}}\,\psi_i(r) = E_i\,\psi_i(r) \tag{2}$$

where the electron density is given by $\rho(r) = 2\sum_{i=1}^{N} |\psi_i(r)|^2$ for a 2N-electron system since each state $\psi_i(r)$ is doubly-occupied [7]. Details of this equation and methods of solution can be found in Ref. [8].

A wavelet based Laplacian operator is needed for the kinetic energy term of the Kohn-Sham Hamiltonian as well as in the solution of Poisson's equation. This Laplacian operator is performed by matrix multiplication. The elements of the differentiation matrix are pre-calculated using the method described in Ref. [9] The solution of Poisson's equation in a wavelet basis is by itself an interesting topic. The efficiency of the wavelet basis in this problem is well established [10].

The operation involving the multiplication of the wave function by the local potential is performed using the following algorithm. First, both the wave function and the local potential are mapped from a $\mathcal{D}_6$ wavelet representation to a

[1]Atomic units are used throughout.

Haar wavelet representation using a transform. Then, the two functions are multiplied in their Haar representations. The product is then mapped back to a $\mathcal{D}_6$ wavelet representation using an inverse to the transform mentioned above. Thus, the multiplication is performed entirely in a wavelet basis.

The electronic ground state of the system is found by solving Eq. (2) iteratively. The wavelet basis is held fixed until convergence is achieved. The basis is then compressed, removing unnecessary basis functions. Then basis functions are added to effectively double the resolution globally. This is necessary to allow the basis to 'grow' to the optimum structure. With this new basis, the total energy is again minimized and the process is repeated.

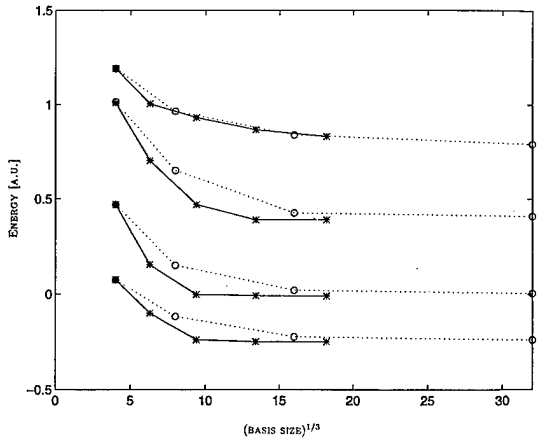


FIGURE 3 Eigenvalues of the Kohn-Sham Hamiltonian for a supercell containing 8 Si atoms. A local s-pseudopotential is used. The eigenvalues are plotted vs. (basis size)$^{1/3}$. Comparision is between the adaptive and non-adaptive calculations indicated by the solid and dashed lines, respectively.

RESULTS

We have tested this method by applying it to a supercell containing 8 Si atoms located at their bulk positions and using a local s-pseudopotential. First, a simpler calculation is performed where the multiplication of the electronic orbital with the local potential is performed by first mapping both functions to real space via a fast wavelet transform, multiplying the two functions and then mapping back to wavelet space. This procedure is similiar to that used in Ref. [5]. In this calculation there is no compression since the full cubic grid must be used. This "full grid" method is performed for increasing grid sizes of $4^3, 8^3, 16^3$ and 32^3. The results are shown in Figure 3. The correct degeneracies (1-6-6-3) of the 16 eigenvalues are achieved consistent with the symmetry of the problem.

Next, the same calculation is performed with an adaptive wavelet basis. Here, the multiplication is performed using the algorithm, described above, where the entire calculation is performed in a wavelet basis. The correct degeneracies are again achieved to within 10^{-4} a.u. The adaptive basis converges to the correct eigenvalues. For the three lowest levels the convergence requires less basis functions.

CONCLUSION

We have demonstrated that it is possible to minimize the total energy in an electronic structure calculation while simultaneously optimizing the wavelet basis set used to represent each electronic orbital. This general method should lead to efficient, potentially linearly scaling electronic structure calculations. Furthermore, if the atoms are allowed to move as in a molecular dynamic simulation, the adaptive basis should adjust to the needs of the structure. Presently, the various corrections resulting from exchange-correlation and non-local effects need to be implemented to perform calculations precise enough to compare with other more established methods. The advantages of an adaptive wavelet basis should lead to efficient methods when applied to larger, localized structures.

Acknowledgements

This work was supported by the Office of Naval Research and the Army Research Office.

References

[1] Gygi, F. and Galli, G. (1995). *Phys. Rev.*, **B52**, R2229.
[2] Chelikowsky, J. R., Troullier, N. and Saad, Y. (1994). *Phys. Rev.*, **B50**, 11355.
[3] Briggs, E. L., Sullivan, D. J. and Bernholc, J. (1996). *Phys. Rev.*, **B54**, 14362.
[4] Cho, K., Arias, T. A., Joannopoulos, J. D. and Lam, Pui K. (1993). *Phys. Rev. Lett.*, **71**, 1808.
[5] Wei, S. and Chou, M. Y. (1996). *Phys. Rev. Lett.*, **76**, 2650.
[6] Daubechies, I. (1992). Ten Lectures on Wavelets (*SIAM*, Philadelphia).
[7] Kohn, W. and Sham, L. J. (1965). *Phys. Rev.*, **A140**, 1133.
[8] Stich, I., Car, R., Parrinello, M. and Baroni, S. (1989). *Phys. Rev.*, **B39**, 4997.
[9] Beylkin, G. (1992). *SIAM J. Numer. Anal.*, **6**, 1716.
[10] Goedecker, S. and Ianov, O. V., unpublished.

Author's Biographies

David A. Richie received his B.S. degree in electrical engineering from the University of Wisconsin-Madison. He is currently a doctorate student at the University of Illinois at Urbana-Champaign where he is working towards a Ph.D. in physics. He was previously a student fellow at Sandia National Laboratories where his research involved the study of vertical cavity surface emitting lasers. He is currently studying the use of adaptive wavelet bases in electronic structure calculations and molecular dynamics simulations.

Paul von Allmen obtained his B.S. and Ph.D. in Physics from the Swiss Federal Institute of Technology at Lausanne, Switzerland. He joined the Zurich IBM Research Laboratory as a Postdoctoral Research Associate in 1990 and was an Invited Scholar at the Beckman Institute, University of Illinois at Urbana-Champaign from 1992 to 1997. He presently is a Senior Scientist at the Motorola Phoenix Corporate Research Laboratories. His interests include subband structures and many-body effects in confined electron systems and dynamical properties of nanostructures and surfaces. He has performed *ab initio* simulations of hydrogen and deuterium desorption from a silicon surface. He presently investigates the field emission properties of carbon materials.

Karl Hess has dedicated the major portion of his research career to the understanding of electronic current flow in semiconductors and semiconductor devices with an emphasis on effects pertinent to device miniaturization. He is currently the Swanlund Professor of Electrical and Computer Engineering, Professor of Physics, Adjunct Professor for Supercomputing Applications and a Research Professor in the Beckman Institute working on topics related to Molecular and Electronic Nanostructures. He has received numerous awards including the IEEE David Sarnoff Field Award for electronics in 1995.

Richard M. Martin is Professor of Physics at the University of Illinois, Urbana-Champaign. His area of research is theoretical condensed matter physics, with emphasis on the electronic structure of materials and quantum devices. The primary goals of his research are to predict the properties of real systems by simulating the many-body quantum system of electrons using density functional and quantum Monte Carlo methods and advanced computational methods. He received his Ph.D. in physics from the University of Chicago and has worked at Bell Laboratories and the Xerox Palo Alto Research Center. He is a fellow of the American Physical Society and the American Association for the Advancement of Science.

QUANTUM SIMULATION I

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 165–172
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Open Problems in Quantum Simulation in Ultra-Submicron Devices*

D. K. FERRY[a,†] and J. R. BARKER[b]

[a]*Center for Solid State Electronics Research, Arizona State University, Tempe, AZ 85287-6206;*
[b]*Department of Electrical and Electronic Engineering, University of Glasgow, Glasgow G12 8QQ, United Kingdom*

Quantum transport is becoming more significant as device size shrinks. For example, as device sizes are scaled below 0.1 μm, the number of impurities becomes quite small, so that they are no longer homogeneously distributed throughout the device volume and the carriers are localized into *quantum boxes*, in that self-energy corrections produced by locally high carrier densities will lead to *quantum dot* formation. This leads to the need to discuss transport through an array of such *quantum structures*. Here, we discuss several issues which must be considered in treating the transport through such devices.

Keywords: Quantum transport, device modeling, impurities

1. INTRODUCTION

The demand for enhanced performance in VLSI circuits pushes semiconductor devices to miniaturization limits, at which point leading edge devices will employ 0.07 μm gate lengths. There is evidence that we can go beyond this size, and many laboratories have produced experimental devices in the 20–30 nm gate length range. Thus, one can approach sizes that are only 100× the atomic spacing in the semiconductor. We must treat transport in a quantum mechanical manner, though we would like to extrapolate from present classical approaches, and retain the appealing framework of the latter formalism.

One major problem is caused by the impurity atoms. For these devices, the active region will contain so few dopant atoms that fluctuations in this number will affect device performance. The impurity profiles determine the device configuration, and most assumptions presume that the impurities are randomly dispersed with a uniform doping profile. However, the fluctuation in the number of total particles in the channel for small devices is comparable to the actual number itself. In addition, it can be expected that the devices will exhibit noticeable conductance variations. *It is clear that the carrier density in the channel is very inhomogeneous* and this will lead to carrier localization in regions approximately 10–20 nm

*Work supported in part by ONR and DARPA.
†Corresponding author.

on a side, because the exchange energy leads to band-gap narrowing in this region. Consequently, these carriers sit in small 3D *quantum boxes* containing 10–20 electrons per box [1]. Yet most approaches to classical device simulation do not adequately treat this fluctuation in the carrier density and doping, nor indeed the strong Coulomb correlations.

Several issues must be considered in treating the transport through such devices, among which are:

(1) Phase coherence within the array of quantum boxes;
(2) The transition between semi-classical and fully quantum transport, including the ability or inability of the present heuristic methods based on a quantum potential to properly describe charge bunching and tunneling;
(3) The role of the contacts, vis-à-vis the fabricated boundaries, and the actual versus internal boundaries;
(4) Electron correlation due to the small capacitances between such dots, and the possibility of single-electron effects in fluctuations of the overall device;
(5) The possible need to provide decoherence at the source or drain.

Here, we address some of these problems. To begin, we discuss a general quantum formulation, as well as providing an approach that will allow studies of the above questions. We discuss a prototypical quantum dot, and its experimental behavior, before considering the quantum trajectories and possible extension of device descriptions such as the quantum hydrodynamic approach and the quantum potential.

2. COUPLING THE DEVICE TO THE SYSTEM

Let us consider an isolated quantum device (e.g., one quantum dot in a real transistor), which is embedded in its environment. The principles are described by the isolated device, its input and output contacts, and the coupling between these parts. The device is described by its Hamiltonian H_d and wave function Ψ_d, while the input and output contacts are described by the Hamiltonians H_u and H_y and wave functions Ψ_u and Ψ_y, respectively. The Hilbert space for the total wave function is a summation space, rather than a tensor product space which is often used to describe system-environment interactions. That is, there is a complete Hilbert space for each of the wave functions, and the total space is the summation of these three spaces. This choice of space allows us to talk about wave functions localized in the device, or in the contact, or hybridized between them using wave functions isolated in one part of the overall Hilbert space. It is assumed that the Hamiltonian remains *separable* into parts representing each space plus coupling terms. Hence, Ψ_d, Ψ_u and Ψ_y represent members of the complete Hilbert spaces for those isolated parts of the total system. From some simple manipulations, we can arrive at the transmission matrix from a pure input state to a pure output state

$$t = (H_d + AH_u^{-1}R)^{-1}CH_u^{-1}A. \qquad (1)$$

Here, A, C and R are the coupling matrices from the input contact to the device, from the device to the output contact, and from the device back to the input contact, respectively. There are at least three different cases of interest [1].

The case that is most relevant to the present discussion is particular to quantum dots in which the contacts are open quantum point contacts and we cannot ignore $AH_u^{-1}R$. The nature of these quantum point contacts is that the modes in the contact region are eigenmodes of the effective waveguide, and this produces significant collimation of the entering particle beam [2]. There is then a significantly different behavior in the dot, and the response is governed by the semiclassical trajectories that are excited in the dot by the waveguide modes. This behavior is clearly seen in both square dots [3] and in stadium (oval) dots that are classically chaotic [4, 5], as may be seen in Figure 1. In these dots, in which only a small

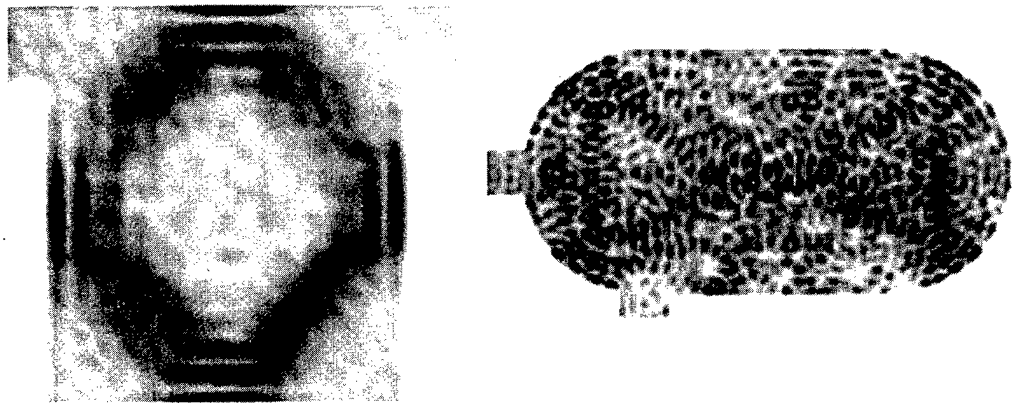


FIGURE 1 Collimation of the input electrons by the quantum point contact leads to the excitation of heavily scarred wave functions, representative of the *regular* trajectories in the ballistic quantum dots. Typical scars are shown for a square dot (a) and an oval, or stadium, dot.

family of trajectories is excited by the collimated beam, the spectral density is still evolving from a series of delta functions, and the regular paths give oscillatory contributions to the spectral density. Indeed, measurements of the phase breaking time have clearly shown the 2D to 0D transition in the dot as the temperature is lowered (below 130 mK in a 1.0 micron dot) [6]. The periodic orbits combine coherently to produce much stronger oscillations than those of an isolated orbit of a chaotic system. This is important, as *the coherence of the phase integrals leads to quite long and periodic orbits.* It is the periodic nature of these long orbits, which contributes a large number of terms to the spectral density, corresponding to returns to the *input* contact after different numbers of traversals. The interference among these waves of successive circuits gives rise to the set of quantized levels that appear in the dot.

The role of the contacts can also be seen in the actual phase-breaking time measured in the dots. In Figure 2, we show measurements on a set of nominally square dots [7]. Here, the phase-breaking time is found to be relatively independent of the dot area, but does depend upon the opening of the QPCs. If the overall resistance is less than 12 kΩ, the phase-breaking time is about 30 ps. On the other hand, if the overall resistance is greater than 15 kΩ, the phase-breaking time increases by an order of magnitude. We interpret this as an additional phase-breaking process, external to the dot, that is coupled to the dot by a sufficient opening of the QPCs.

What we have found in the above discussion correlates well to what we know from our under-

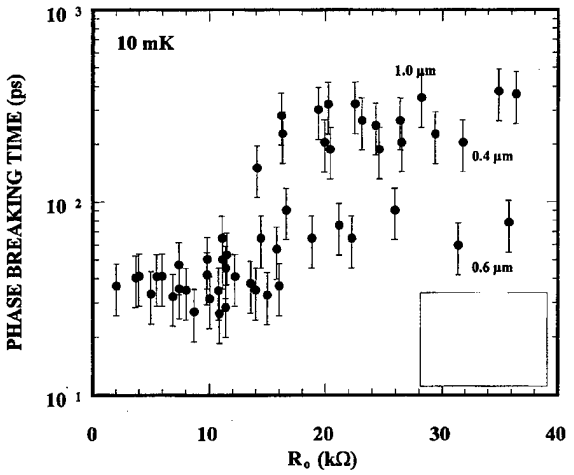


FIGURE 2 The phase breaking time measured experimentally for three different sizes of ballistic quantum dot. The important aspect is that each shows a transition to a much smaller value of the phase breaking time as the quantum point contacts are opened (smaller resistance) so that the dot is more strongly coupled to the environment.

standing of classical device simulation: The nature of the contacts can dramatically affect, and even dictate, device performance. This point is often missed in discussions of quantum transport which are derived from *equilibrium* quantum statistical mechanics. The role of the boundaries cannot be overlooked, and usually is the dominant control mechanism in device performance. Without proper treatment of these boundary conditions, no adequate quantum device simulation can be attained.

3. TRAJECTORIES, PATHS AND POTENTIALS

Let us now return to the question of the DOS, as measured by the transport in the quantum dot. In square quantum dots, we have found from comparison between theory and experiment that the transport behavior is generally *regular*. By regular, it is meant that the quasi-ballistic motion scatters elastically off the walls of the dot. In the fully quantum limit of a zero-dimensional dot, the density of states (at $T = 0$) is simply a sum of delta functions, located at the energy levels of the dot, although the DOS must change to the uniform two-dimensional form as these delta functions are broadened. For a classically chaotic system, it has been thought that opening the quantum billiard (to interactions with the outside world) will introduce chaotic behavior in the level statistics and in the transport. *Only recently have the measurements of magnetotransport in fully quantized dots* ($\Delta \gg k_B T$) *appeared that show the residual classically regular transport, ruling out the normally accepted chaos theory.*

For regular motion, only one family of periods may be observed, but this family will be composed of the fundamental and several harmonics, as seen in the experimental results. We are led to the assertion that *the study of transport in "regular" billiards exposes the fundamental quantum structure of the dots themselves, and provides important information as to how the transport will exhibit itself in future electron devices.*

The understanding of how classical *flows* evolve into quantum mechanical *currents* is still an open issue. There has been considerable controversy over the description of phase space trajectories in quantum mechanics. The incorporation of a *quantum potential* is supposed to provide quantum mechanical deviations from the classical orbits. The most common approach to the quantum potential is through the decomposition of the wave function as

$$\Psi(r, t) = \sqrt{\rho(r, t)}\, \exp(iS(r, t)/\hbar), \qquad (2)$$

where $\rho(r, t)$ is the density (ρ also will be used as the density *matrix* later, but this is not its meaning here), r and t are the space-time manifold, and $S(r, t)\hbar$ is the phase associated with the motion.

An unambiguous construction of *quantum* trajectories, associated with pure states (in both configuration space and phase space) is provided by the Bohm-de Broglie *interpretation* of quantum mechanics [8, 9]. Here, the Schrödinger equation for a carrier describes the deterministic, causal evolution of a complex field, but is augmented by the postulate that the carrier may be considered as a classical point particle with position vector $\boldsymbol{x}(t)$ and momentum vector $\boldsymbol{p}(t)$ embedded within the pilot field $\Psi(r, t)$, given by (2). At any time t_0, the carrier is found with probability density $\rho(r, t_0)$ to have position $\boldsymbol{x}(t_0) = r$ and $\boldsymbol{p}(t_0) = \nabla \boldsymbol{S}(r, t_0)$. Thereafter, the particle trajectory in configuration space and phase space is followed *deterministically* under the influence of the pilot field via the *parameterization*

$$\begin{aligned} \boldsymbol{x}(t) &= \boldsymbol{x}(t_0) + \int_{t_0}^{t} \boldsymbol{p}(t')dt', \\ \boldsymbol{p}(t) &= \nabla S[r(t), t]. \end{aligned} \qquad (3)$$

This picture is *not* a transformation to a particle picture which can replace the quantum description of the system, but instead provides a description in a kinematical sense of the behavior of the system. The evolution of S, and hence $\boldsymbol{p}$, is provided by the Schrödinger equation, which is a separate field

theory. Classically, the action S is determined *completely* by the underlying particle dynamics in the system, viz. the Hamilton-Jacobi equations. The quantum "action" S is also claimed to be determined by a Hamilton-Jacobi-like equation, which differs from the classical version by coupling to the continuity equation through the quantum potential V_Q. The latter determines the probability density of locating the particle within quantum theory, as noted previously, and satisfies the continuity equation. However, S is not determined entirely by the local dynamics, but must satisfy a crucial topological condition which follows from its origin as the phase of an independent complex field.

One must be sure to properly incorporate the single-valuedness of the wave function, otherwise the equation of motion for S will not be properly constructed. In particular, we note that S is defined modulo h (Planck's constant). Since, the *velocity* is normally described from $\boldsymbol{v} = \nabla S/m$, the periodic nature of S leads to the quantization

$$\oint \boldsymbol{v} \cdot d\boldsymbol{r} = lh/m, \tag{4}$$

where l is an integer. In classical hydrodynamics, the integral on the left is known as the velocity circulation, and is related through Stoke's theorem to the *vorticity* of the motion. Most approaches to the use of dynamical equations ignore the quantization within the system. The quantization must be put in quite early in the process, as it will not evolve from a classical distribution function.

If we consider any closed contour for which (4) holds, we can write at any point $S' = S + Nh$. Let us suppose that $N \neq 0$ and that somewhere within the area enclosed by the contour, S goes to zero. This must occur where $\sqrt{\rho} = 0$, i.e., at the nodes which are required to preserve the continuity of the wave function itself. At these points, both the real and the imaginary parts of the wave function vanish, and this describes the nodal lines in space. In addition to the quantized orbits (which can be extended to any multiply-connected region), the nodal lines are regions in which the motion is no longer irrotational.

It is important to point out, that closed orbits in a ultrasmall device can be induced not only by reflections between two barriers (the resonant level in the double-barrier tunneling device), but also by multiple scattering from impurities (the source of weak localization and universal conductance fluctuations), or from any combination of these factors. The latter is important for universal conductance fluctuations and, indeed, has been suggested for the observation of weak localization in a single quantum point contact.

One can make a minimalist improvement on the Bohm-deBroglie theory, which is normally described by [10]

$$\langle \delta(\boldsymbol{x}(t) - r) \rangle = \rho(r, t), \quad \boldsymbol{p}_0 = \nabla S(r, t). \tag{5}$$

We assume that $\boldsymbol{x}(t)$ is a stochastic variable, such that the mean momentum is given by the gradient of the phase as in (5):

$$\langle \boldsymbol{p} \rangle = \int d^3x \rho \nabla S(r, t). \tag{6}$$

To account for the stochastic nature, we introduce a momentum fluctuation, which is defined through $\rho \boldsymbol{p}_1 = \hbar \nabla \rho / 2$. Then, we find a new result for the mean square momentum as

$$\langle p^2 \rangle = \int d^3x \{ \rho [p_0^2 + p_1^2] \}. \tag{7}$$

This is the true quantum mechanical expectation value, as compared to the Bohm picture, which omits the term in p_1^2. As noted in [11], this omission leads to a violation of Heisenberg's uncertainty relation. In this extended picture, the equations for continuity and momentum are given by

$$-\frac{\partial S}{\partial t} = \frac{p_0^2 + p_1^2}{2m} + V - \frac{\hbar}{2m\rho} \nabla \cdot (\rho \boldsymbol{p}_1), \tag{8}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \left(\frac{\rho \boldsymbol{p}_0}{m} \right) = 0. \tag{9}$$

The traditional momentum equation is now found by taking the divergence of (8), but it is clear

that the quantum potential now splits into two components

$$V_Q = \frac{p_1^2}{2m} - \frac{\hbar}{2m\rho}\nabla\cdot(\rho \boldsymbol{p}_1), \tag{10}$$

where the leading the term is related to the kinetic energy associated with the quantum fluctuations. The second term in (10) involves the divergence of the fluctuation current and its spatial average over the position probability density vanishes. This is a crucial point, and restores the inherent statistical nature of quantum mechanics to the apparent deterministic form of the Bohm picture. We find that $\boldsymbol{p}_1$ satisfies identically a *stationary* Fokker-Planck equation

$$\nabla\cdot\left(\frac{\rho \boldsymbol{p}_1}{m}\right) = \frac{\hbar}{2m}\nabla^2\rho, \tag{11}$$

reinforcing our view that $\boldsymbol{p}_1$ describes the fluctuations. The above equations are an alternative and more acceptable semantic deconstruction [12] of the Schrödinger equation (which must be augmented by the phase boundary condition (4) as before), and points out that there are many different definitions of a "quantum potential".

4. PROBLEMS OF THE QUANTUM POTENTIAL

Over and above the above problems of the uniqueness of a quantum potential, the more normally used one has its own problems with any sort of fluctuations. Using the wave function (2) for a pure state, we may easily write two continuity equations for the probability density

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho v) = 0, \tag{12}$$

and for the action

$$\frac{\partial S}{\partial t} + \frac{(\nabla S)^2}{2m} + V + V_Q = 0, \tag{13}$$

where

$$V_Q = -\frac{\hbar^2}{2m}\frac{\nabla^2\sqrt{\rho}}{\sqrt{\rho}} \tag{14}$$

is the *more usual* quantum potential. Taking the gradient of (13) produces an Euler equation which includes the additional force arising from the quantum potential, and this has been the justification for much of the usage of the heuristic quantum hydrodynamic equations for mixed state systems. One must be careful to note the multivalued nature of S, quantization of the orbits, and the boundary conditions, scattering, and the continuity of the wave function.

To extend the above approach to a mixed state, one can use the same equations above, but this incorporates errors and approximations. Instead, let us write the density matrix as

$$\begin{aligned}\rho(r, r', t) &= \sum_{n=1}^{N} P_n|\psi_n\rangle\langle\psi_n| = \sum_{n=1}^{N} P_n\rho_n,\\ v_n &= \frac{\nabla S_n}{m}, \quad \sum_{n=1}^{N} P_n = 1,\end{aligned} \tag{15}$$

where, for generality, we do not specify any functional form *a priori* for the probabilities P_n for the states ψ_n. We may now introduce an average velocity through the prescription

$$\rho v_a \equiv \sum_{n=1}^{N} = P_n\rho_n v_n, \tag{16}$$

for which the continuity Eq. (12) is recovered for the average velocity rather than that of any pure state. We find that a single quantum potential does not immediately appear in the problem, and

$$\begin{aligned}m\frac{\partial v_a}{\partial t} + m\Bigg[\sum_{n=1}^{N} P_n[v_n\nabla\cdot(\rho_n v_n) + \rho_n(v_n\cdot\nabla)v_n]\\ - v_a\nabla\cdot(\rho v_a)\Bigg] = -\nabla V - \sum_{n=1}^{N} P_n\rho_n\nabla V_{Qn}.\end{aligned} \tag{17}$$

Clearly, the "energy" term is quite complex, and not simply related to a single *temperature*, and even the diagonal terms contain significant fluctuations beyond those related to the velocity fluctuations. *The quantum potential is not that of the pure state.* Suppose we have a small fluctuation around some steady value, in which we can define the parameters such that

$$\rho_n = \rho + \Delta\rho_n \quad \langle \ldots \rangle = \sum_{n=1}^{N} P_n \ldots, \quad \rho = \langle \rho_n \rangle. \tag{18}$$

In this case, the leading correction term is of second order, but we can write the new quantum potential as

$$\sum_{n=1}^{N} P_n \rho_n \nabla V_{QN} \rightarrow \rho \nabla V_Q (1 - \alpha \langle \Delta\rho_n \Delta\rho_n \rangle) + \ldots \tag{19}$$

The term in the angular brackets represents a significant contribution to the overall density-density correlation function, which is known to produce weak localization and universal conductance fluctuations in mesoscopic devices. Hence, we see that there is a modification of the quantum potential in the mixed state due to these fluctuations, which are readily induced by scattering processes. We can understand the source of the correction term by rewriting (14) as

$$V_Q = -\frac{\hbar^2}{2m} \frac{\nabla^2 \sqrt{\rho}}{\sqrt{\rho}} = -\frac{\hbar^2}{2m} \left[\frac{\nabla^2 \rho}{2\rho} - \left(\frac{\nabla \rho}{2\rho} \right)^2 \right]. \tag{20}$$

The second term is directly related to the zero-point fluctuations, and it is clear that the additional fluctuations introduced by scattering, boundaries, etc., will produce a major contribution to the density–density correlation function through this term. It is the corrections arising from this term that dominantly lead to (19) and, indeed, it is the properties of the second term that are important in introducing the fluctuations of the last section. It is this second term that inspires the introduction of $\boldsymbol{p}_1$ in the previous section.

Finally, let us return to the crucial topological condition on the phase of individual quantum states, as expressed through condition (4). This condition achieves enormous complexity for mixed state descriptions, and leads to a condition on the average velocity which takes the form

$$\Gamma_a \oint \boldsymbol{v}_a \cdot d\boldsymbol{r} = \oint d\boldsymbol{r} \cdot \sum_{n=1}^{N} P_n \frac{\rho_n(r) \boldsymbol{v}_n(r)}{\rho(r)}, \tag{21}$$

subject to the N conditions

$$\oint \boldsymbol{v}_n \cdot d\boldsymbol{r} = l_n h/m. \tag{22}$$

Because the quantum numbers l_n are positive or negative integers, there is the possibility that the circulation Γ_a, which is a superposition of terms involving the l_n, will "phase mix" to zero. If this were exact, the quantum hydrodynamic equations would indeed be simple extensions of classical hydrodynamic theory apart from the quantum potential terms. However, to our knowledge this point has never been explored.

5. DISCUSSION

Why has our emphasis been on the concept of defining a quantum kinetic (trajectory based) picture, or even that of a quantum hydrodynamic (streamlines based) picture? As discussed above, experiments are now available on ballistic quantum dots in which one can study both the quantum transport and the classical ballistic transport in a manner that good agreement is achieved between the two and with experiment. Moreover, hydrodynamic approaches have been extensively pursued for "quantum" simulation of small devices. Here, however, we have demonstrated significant limitations and problems in the latter.

Our motivation is to find a true quantum kinetic, or quantum hydrodynamic, picture that

properly allows for the fluctuations, and recognizes the particular problems which derive from the explicit non-locality of the problem and properly accounts for the topological phase boundary conditions. The success of Monte Carlo suggests that one might analogously find a Monte Carlo, trajectory-based, non-deterministic picture for the Schrödinger equation and/or the density-matrix Liouville equation.

To this end, we can return to the ballistic quantum dots, as the regular trajectories that exist in these structures provide an ideal testbed for studying the evolution of the classical trajectories into the quantum kinetic picture. Moreover, these structures are explicitly dependent upon the details of the confining boundary potentials and the initial conditions defined by the contacts. They are an ideal structure to study quantum kinetics without the important quantum effects being masked by an overlaying role of inelastic processes, which ultimately destroy much of the quantization.

Acknowledgement

The authors would like to express their appreciation to Jon Bird for making the data of Figure 3 available prior to publication, and to D. Vasileska, S. M. Goodnick, and A. Asenov for helpful discussions.

References

[1] Ferry, D. K., Akis, R., Udipi, S., Vasileska, D., Pivin, D. P. Jr., Connolly, K. M., Bird, J. P., Ishibashi, K., Aoyagi, Y., Sugano, T. and Ochiai, Y. (1997). "Carrier Transport in Nanodevices," *Jpn. J. Appl. Phys.*, **36**, 1841–5.

[2] Akis, R., Ferry, D. K. and Bird, J. P. (1996). "Magneto-transport fluctuations in regular semiconductor ballistic quantum dots", *Phys. Rev. B*, **54**, 17705–15.

[3] Bird, J. P., Ferry, D. K., Akis, R., Ochiai, Y., Ishibashi, K., Aoyagi, Y. and Sugano, T. (1996). "Periodic conductance fluctuations and stable orbits in mesoscopic semiconductor billiards", *Euro-Phys. Lett.*, **35**, 529–34.

[4] Akis, R., Ferry, D. K. and Bird, J. P., "Wave-function scarring effects in stadium-shaped quantum dots", *Phys. Rev. Lett.*.

[5] Okubo, Y., Ochiai, Y., Vasileska, D., Akis, R., Ferry, D. K., Bird, J. P., Ishibashi, K., Aoyagi, Y. and Sugano, T. (1997). "*Stability of Regular Orbits in Ballistic Quantum Dots*", *Phys. Lett.* A, in press.

[6] Bird, J. P., Ishibashi, K., Ferry, D. K., Ochiai, Y., Aoyagi, Y. and Sugano, T. (1995). "Phase breaking in ballistic quantum dots: Transition from two- to zero-dimensional behavior", *Phys. Rev. B*, **51**, 18037–40.

[7] Bird, J. P., private communication.

[8] Bohm, D. (1952). "A suggested interpretation of the quantum theory in terms of 'hidden' variables, I and II", *Phys. Rev.*, **85**, 166–193.

[9] Holland, P. R. (1993). *The Quantum Theory of Motion* (Cambridge University Press, Cambridge).

[10] Barker, J. R. (1995). "Trajectories in Quantum Transport", in *Quantum Transport in Ultrasmall Devices*, Eds. by D. K. Ferry, H. L. Grubin, C. Jacoboni and A.-P. Jauho (Plenum Press, New York), pp. 171–180.

[11] Barker, J. R. (1992). "Fundamental Aspects of Quantum Transport Theory", in *Basic Properties of Semiconductors*, Vol. 1 of *Handbook of Semiconductors*, revised, Ed. by P. T. Landsberg (North-Holland, Amsterdam), pp. 1079–1127.

[12] Barker, J. R. (1994). "On the pilot field representation of quantum transport theory", *Semicond. Sci. Technol.*, **9**, 911–917.

Authors' Biographies

David K. Ferry is Regents' Professor of Electrical Engineering at Arizona State University, where he has been since 1983. He is a fellow of the American Physical Society and the Institute of Electrical and Electronics Engineers.

John R. Barker is Professor of Electronics at the University of Glasgow. He is a Fellow of the Royal Society of Edinburgh.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 173–178
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Theory of Electron Transport in Small Semiconductor Devices Using the Pauli Master Equation

M. V. FISCHETTI*

IBM Research Division, Thomas J. Watson Research Center, P. O. Box 218, Yorktown Heights, NY 10598, USA

It is argued that the Pauli master equation can be used to simulate electron transport in very small electronic devices under steady-state conditions. Written in a basis of suitable wavefunctions and with the appropriate open boundary conditions, this equation removes some of the approximations which render the Boltzmann equation unsatisfactory at small length-scales. The main problems consist in describing the interaction of the system with the reservoirs and in assessing the range of validity of the equation: Only devices smaller than the size of the electron wavepackets injected from the contacts can be handled. Two one-dimensional examples solved by a simple Monte Carlo technique are presented.

Keywords: Master equation, electron transport, small devices

1. INTRODUCTION

As noted by Frensley discussing quantum transport in open systems [1], the Pauli master equation [2], (PME) could constitute an intuitive description of electron transport in semiconductor devices of size comparable to the electron wavelength: One could capture the wavelike nature of transport, lost in the Boltzmann-transport-equation (BTE) picture, so bypassing the weak-field, long devices limitations. In retaining the weak-scattering and completed-collision limits one would not reach a full-fledged quantum description, as required to study very fast time transients [3, 4]. Yet, as devices approach the sub-50 nm length-scale, the advantage of avoiding the concept of point-like electrons appears overwhelming.

Although arguments have been raised against the correctness of the PME [1], it is applicable to very *small devices* and to *steady-state* phenomena: *Small devices*, because the PME rests on the absence of off-diagonal elements of the density matrix ρ. In turn, their absence implies that the contacts inject 'plane waves', which corresponds physically to the injection of spatially highly delocalized wavepackets. Hence, the device must be smaller than the 'dephasing length' λ_ϕ in the contacts [5]. For heavily doped Si, this is of the order of 50 nm, which is of technological interest. *Steady state*, since any nontrivial time transient

* Corresponding author.

(e.g., relaxation of photo-excited carriers, fast switching of a device) would excite those off-diagonal terms of ρ which the PME cannot handle. Indeed, as also note by Frensley [1], the PME is inconsistent with current continuity in this case.

2. MASTER EQUATION AND CONTACTS

Let's consider a semiconductor region unbounded in the (x, y)-plane, homogeneous in this plane, in contact with two reservoirs at $z=0$ and $z=L$. Let's write the single-particle Hamiltonian as:

$$-\frac{\hbar^2\nabla^2}{2m^*} - eV(z) + H_{\text{int}} + H_{\text{res}} = H_0 + H_{\text{int}} + H_{\text{res}}, \tag{1}$$

within the envelope/effective-mass approximation, m^* being the effective mass. H_{int} describes the electron-phonon interaction or interactions with impurities, etc. H_{res} represents the interaction between the device and the external reservoirs. The time evolution of the system is given by the 'transport' equation for the 'reduced' (electronic) density matrix ρ:

$$\frac{\partial\rho}{\partial t} = \frac{i}{\hbar}[\rho, H] + \left(\frac{\partial\rho}{\partial t}\right)_{\text{res}}. \tag{2}$$

The last term on the right-hand-side describes the effect of the reservoirs. Assuming $H_{\text{int}} = \alpha H'$, we can derive a transport equation from Eq. (2), by considering only terms to the leading order in α, and following the 'standard' derivation of Kohn and Luttinger [7], but now using as basis functions not plane waves, but the eigenstates $|\mu\boldsymbol{K}\rangle$ of H_0, such that $H_0|\mu\boldsymbol{K}\rangle = [E_\mu + \hbar^2K^2/(2m^*)]|\mu\boldsymbol{K}\rangle = E_\mu(\boldsymbol{K})|\mu\boldsymbol{K}\rangle$. This simplifyies Eq. (2), since the commutator now contains only the interaction term H_{int}. The wave-functions are simply of the form $\zeta_\mu(z)\exp(i\boldsymbol{K}\cdot\boldsymbol{R})/(2\pi)$, where $\boldsymbol{K}$ and $\boldsymbol{R}$ are the wavevector and position on the (x, y)-plane, and the functions $\zeta_\mu(z)$ obey the one-dimensional Schrödinger equation for the potential $V(z)$, subject to the boundary conditions $V(0)=V_L$ and $V(L)=V_R$. Taking $-eV_L > -eV_R$, for any given energy $E_\mu > -eV_L$, two independent solutions represent waves incident from the left (+) or right (−), and partially reflected and transmitted, with reflection and transmission amplitudes $r^\pm$ and $t^\pm$. For $-eV_L > E_\mu > -eV_R$, the nondegenerate solutions represent non-current-carrying states. In the following the index $\sigma = \pm$ will label left- and right-traveling waves in the energy range $E_\mu > eV_L$, and $\langle z|\mu\sigma\rangle = \zeta_\mu^\sigma(z)$.

To the leading order in α, the diagonal elements, $\rho_{\mu\boldsymbol{K}\sigma}$ (symbol used in place of $\rho_{\mu\boldsymbol{K}\sigma\mu\boldsymbol{K}\sigma}$) of the density matrix obey the Eqs. [7 – 9]:

$$\frac{\partial\rho_{\mu\boldsymbol{K}\sigma}}{\partial t} = \sum_{\nu\boldsymbol{K}'\sigma'\neq\mu\boldsymbol{K}\sigma}\left[W_{\mu\boldsymbol{K}\sigma;\,\nu\boldsymbol{K}'\sigma'}\rho_{\nu\boldsymbol{K}'\sigma'} - W_{\nu\boldsymbol{K}'\sigma';\mu\boldsymbol{K}\sigma}\rho_{\mu\boldsymbol{K}\sigma}\right] + \left(\frac{\partial\rho_{\mu\boldsymbol{K}\sigma}}{\partial t}\right)_{\text{res}}. \tag{3}$$

here $W_{\nu\boldsymbol{K}'\sigma';\mu\boldsymbol{K}\sigma}$ represents the probability per unit time of a transition from the state $|\mu\boldsymbol{K}\sigma\rangle$ to the state $|\nu\,\boldsymbol{K}'\sigma'\rangle$ using Fermi golden rule. While the derivations provided in [7–9] can be followed very easily, now one must also assume that the reservoirs do not alter the ratio α between off-diagonal and diagonal matrix elements. Van Hove [8] has shown the validity of Eq. (3) in describing approach to equilibrium, in the case of closed systems (for which the term $(\partial\rho/\partial t)_{\text{res}}$ is absent), provided the initial state is chosen as 'quasi-diagonal': At $t=0$ we must have $\rho_{\mu\boldsymbol{K}\sigma\nu\boldsymbol{K}'\sigma'}=0$ for $E_\mu(\boldsymbol{K}) - E_\nu(\boldsymbol{K}') > \delta$, where $\delta/\hbar$ is of the order of the relaxation rate, $1/\tau$. Similarly, Eq. (3) is valid as long as the contacts do not 'inject' off-diagonal elements, that is $(\partial\rho_{\mu\boldsymbol{K}\sigma\nu\boldsymbol{K}'\sigma'}/\partial t)_{\text{res}} = 0$ for $E_\mu(\boldsymbol{K}) - E_\nu(\boldsymbol{K}') > \delta$. This is the case for very delocalized wavepackets entering the contact, plane waves being the ideal limiting case. In this limit, the effect of the contacts can be modeled phenomenologically by assuming that the left reservoir attempts to restore charge neutrality by injecting a flux of

particles of any $\boldsymbol{K}$ into the right-bound component of the states associated to waves incident from the left, ζ_μ^+, fixed by the Fermi level at the cathode, E_F^L. Accounting also for the flux out of the device via either contact, one can write:

$$\left(\frac{\partial \rho_{\mu+}}{\partial t}\right)_{\text{res}} = [f^{(L)}(E_\mu) - \rho_{\mu+} - T_\mu^- \rho_{\mu-}] \frac{\hbar k_\mu^-}{m^*} |A_\mu^+|^2. \tag{4}$$

where $\rho_{\mu\sigma} = \sum_{\boldsymbol{K}} \rho_{\mu \boldsymbol{K}\sigma}$, $T_\mu^\mp = |A_\mu^\mp t_\mu^\mp|^2 / |A_\mu^\pm|^2$, $\hbar k_\mu^\pm = [2m^*(E_\mu + eV_{R/L})]^{1/2}$, and the A's are normalization constants. Similar expressions also hold for states traveling from the right. The occupation of the states at thermal equilibrium, as determined by the 'left' reservoir, has been expressed in terms of the equilibrium Fermi function (integrated over the 'transverse' wavevectors $\boldsymbol{K}$), $f^{(th,L)}(E_\mu)$, possibly shifted in $\boldsymbol{k}$-space, in order to account for the open nature of the contacts [1, 6]. By analogy, a similar expression (with the terms $\rho_{\mu+} - T_\mu^- \rho_{\mu-}$ replaced by ρ_μ) is also used to express the linear relaxation of the of states with energy $-eV_R < E_\mu < -eV_L$. Finally, the net current flowing through the device is given by:

$$J = -e \sum_\mu \left\{ \rho_{\mu+} \frac{\hbar k_\mu^+}{m^*} |t_\mu^+ A_\mu^+|^2 - \rho_{\mu-} \frac{\hbar k_\mu^-}{m^*} |t_\mu^- A_\mu^-|^2 \right\}, \tag{5}$$

where obviously only the doubly-degenerate states with $E_\mu > -eV_L$ contribute to the sum.

Criticisms have been raised against Eq. (3) based on the violation of continuity [1]. Yet, it should be noted that at steady state the reservoirs and the reverse scattering-induced transitions provide the 'missing' current and current continuity is trivially satisfied. However, continuity is indeed violated when using the PME to describe 'fast' time-dependent phenomena: For example, any time dependent perturbation $\delta V(t)$ will cause an electron in $|\mu \boldsymbol{K}\sigma\rangle$ at $t = 0$ to have nonzero amplitude also in a state $|\nu \boldsymbol{K}'\sigma'\rangle$. At long times the requirement that this amplitude be negligible translates into the condition $|e\partial\delta V/\partial t| \ll \hbar/\tau_\phi^2$, expressing the fact that the applied bias cannot change appreciably over a relaxation time, so that the PME is unable to handle nontrivial time-dependent phenomena.

3. MONTE CARLO SOLUTION OF ONE DIMENSIONAL EXAMPLES

In order to implement Eq. (3) in one-dimensional situations, starting from a given potential $V(z)$, such as a Thomas-Fermi solution, the Schrödinger equation is solved with vanishing Neumann boundary conditions. The eigenvalues $\{E_\mu\}$ will be spaced densely enough (to accurately reproduce the bulk density of states in the reservoirs) if the length L of the device is selected appropriately. The extended states can be finally decomposed into left- and right-traveling waves following the 'quantum transmitting boundary method' by Lent and Kirkner [10]. The levels $|\mu\rangle$ are populated following Pötz [6] and the Poisson equation can be solved again, using well-known potential-damping schemes to improve convergence. Iterating this Poisson/Schrödinger scheme yields a solution in the ballistic case. Scattering is introduced by calculating all transition rates $W_{\nu \boldsymbol{K}'\sigma';\mu \boldsymbol{K}\sigma}$ between the levels, and adding the PME as a third equation over which one must iterate. A Monte Carlo algorithm is used here. Scattering and injection/exit from contacts are treated stochastically in order to determine the new populations $\rho_{\mu \boldsymbol{K}\sigma}$. Typically, 300,000 'electrons' are employed, using a time step $\delta t \approx 10^{-16}$ s in the Monte Carlo algorithm. After about 100–500 Monte Carlo steps, the Poisson and the Schrödinger equations are solved again until convergence is obtained. Finally, the current flowing through the device is obtained from Eq. (5), while information about carrier density, kinetic energy, and velocity as a function of position inside the device can be obtained from the expectation values of the corresponding operators as traces over their products with the density matrix.

As a first example, let's consider a simple *nin* resistor at 300 K consisting of a 300 nm-long

region between $z = 0$ and $z = L$ with n-type regions 100 nm-long n-type doped to 10^{17} cm^{-3}, and an intrinsic region also 100 nm long. A uniform effective mass of $m^* = 0.32\ m_0$ (where m_0 is the free electron mass) and a dielectric constant $\varepsilon = 11.7\ \varepsilon_{vac}$ (where ε_{vac} is the permittivity of vacuum) are used. Nonpolar scattering with optical phonons is included using a photon energy 60 meV, an optical deformation potential of 5×10^8 eV/cm, a crystal density $\rho_x = 2.33$ g/cm^3. These parameters imply a 'dephasing length' λ_ϕ in reservoirs of the same material of the order of 785 nm, larger than the active region of the device, about 100 nm long.

Figure 1 shows the potential, carrier density, kinetic energy, and velocity at the end of the iteration. The dashed lines are the corresponding

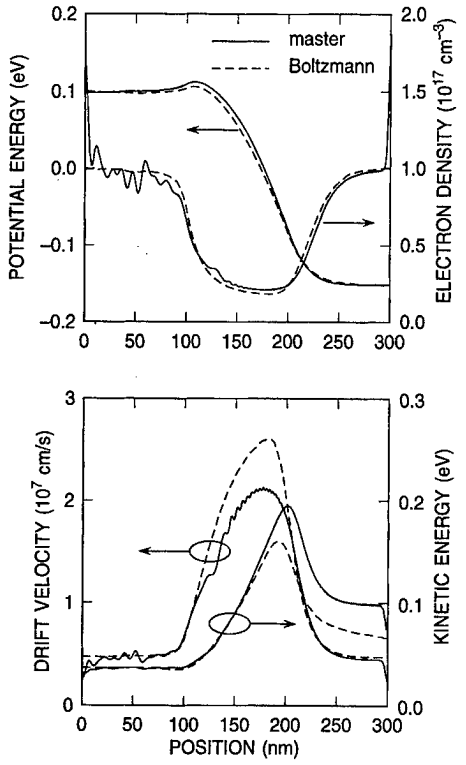


FIGURE 1 Potential and charge density (top frame), kinetic energy and drift-velocity (bottom frame) as a function of position z inside an *nin* device calculated using the master (solid lines) and Boltzmann (dashed lines) equation. The charge accumulation at the device boundaries is an artifact caused by boundary conditions employed to solve the Schrödinger equation.

quantities obtained from a conventional Monte Carlo solution of the BTE using identical parameters. The current density is about 6.78×10^4 A/cm^2, about 10% smaller than the value obtained using the BTE, 7.58×10^4 A/cm^2. This difference originates mainly from a slightly higher built-in barrier at the cathode/intrinsic-region junction seen in the 'quantum' results, caused by the penetration of wavefunctions into the barrier itself. The penetration of charge into classically forbidden region is also the origin of the redistribution of charge seen in the top frame of Figure 1: The charge density obtained using the PME is larger in the intrinsic region, as wavefunctions penetrate into the gap beyond the classical turning point. This charge redistribution is the major factor responsible for the differences seen in average kinetic energy and drift velocity.

The second example is a simple GaAs/$Al_{0.3}Ga_{0.7}As$ resonant tunneling structure. Two 150 nm thick layers of GaAs-like semiconductor with n-type doping (2×10^{17} cm^{-3}) are separated by two AlGaAs tunnel barriers, each 2.8 nm thick, enclosing a 4.5 nm wide semiconductor well. Two 3 nm-thick undoped 'spacer' layers, in turn, enclose the double-barrier region. 'Conventional' parameters for GaAs and $Al_{0.3}Ga_{0.7}As$ are used [12], also including nonpolar scattering with acoustic phonons and polar scattering with longitudinal optical phonons. A lattice temperature of 300 K is used. The contact mobility used to displace the distributions injected by the reservoirs is taken to be $\mu_{el} = 5000$ cm^2/Vs, corresponding to a dephasing length in the reservoirs, $\lambda_\phi \approx 130$ nm, larger than the active region of the device. Since even at room temperature scattering is not sufficiently strong to fully populate the states in the accumulation layer, as discussed by Frensley [1], the curves in Figure 2 show that a large fraction of the applied bias drops over the cathode region to the left of the double-barrier region. Although similar results have been obtained by Frensley [1] and Kluksdahl *et al.* [11], they appear poorly credible, since the potential profile clearly depends on the size of the simulated region.

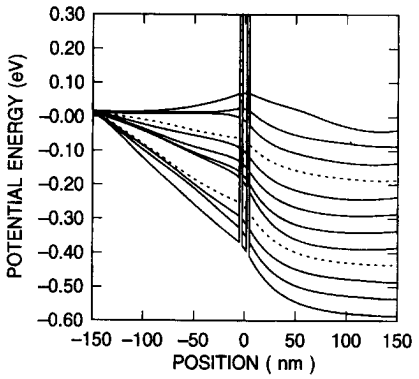


FIGURE 2 Potential energy *vs.* position in a resonant tunnel diode calculated including electron-phonon scattering. The curves are parametrized by the applied bias from 0.0 to 0.6 V in steps of 0.05 V. Note the 'gaps' at the left of the barriers around 0.2 and 0.45 V. Convergence could not be obtained close to these biases, because of an instability caused by scattering, as discussed in the text. The dotted solutions for biases of 0.2 and 0.45 V are only representative of the average potential during the iteration.

Figure 3 shows the current-voltage characteristics: As the resonance is approached (at about 0.2 V), the associated build-up of charge in the well screens the cathode field, increasing the field across the collector barrier. A critical point is soon reached at which the rate at which scattering feeds charge into the states localized in the well becomes smaller than the rate at which the well-charge leaks to the right. Charge is removed from the well, and the potential profile now 'snaps', a larger voltage now dropping at the left of the barriers. The population of the current-carrying scattering states, and so the current itself, now drops. As the applied bias is increased further, another 'snapping point' is reached, at about 0.45 V. These critical bias points can be seen as 'gaps' in the cathode potential-profile in Figure 2.

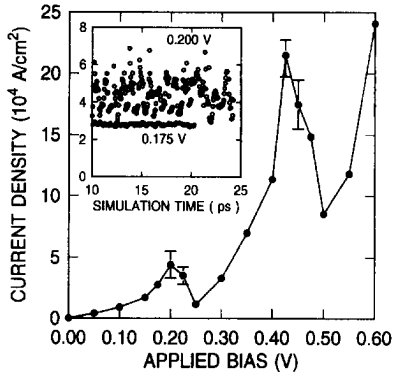


FIGURE 3 Current density in the GaAs/AlGaAs resonant tunneling structure with the inclusion of electron-phonon scattering. Note the 'snap back' of the current at biases of 0.2 and 0.45 V, corresponding to the instability regions of Figure 2. The error bars are used to illustrate the range of the oscillation of the current. The inset shows the Calculated current density *vs.* simulation time at two bias points, slightly away from (0.175 V) and close to (0.200 V) resonance. Note the instability of the current at the bias point near resonance.

As loosely indicated by the error bars in Figure 3, for biases close to and slightly beyond resonance the situation is unstable: As 'resonant wavefunctions' leak out to the anode, scattering to non-current-carrying states in the collector region is enhanced, thanks to a larger overlap (form) factor. The resulting redistribution of charge alters the potential profile, driving the system away from resonance. As this happens, the charge in the anode side of the barriers drops once more, and the system is now driven back towards a resonant condition, and the cycle repeats once more. The inset of Figure 3 illustrates some features of this instability. The good convergence properties of the computed current density for a bias slightly away from resonance (0.175 V) are compared to the unstable behavior for a bias close to the first resonance (0.200 V). In the latter case, with some imagination one may detect an oscillatory behavior between two states [11, 13].

4. SUMMARY AND CONCLUSIONS

The main conclusion of this paper is that there is an interesting class of problems for which the Pauli master provides a solution of the electron transport problem overcoming some of the strongest limitations of the semiclassical Boltzmann equation: For devices with active regions smaller than a few hundreds of nm, the master equation makes it possible to account for the wave nature of the electrons, although only in the weak scattering

limit and in steady state or slow (adiabatic) time transients. The major price one pays is the inability to handle fast time-transients and strong scattering situations, collisional broadening and coherent transport in the femtosecond time scale being the most notable examples. The advantages consist in the ability to treat inelastic scattering processes, arbitrarily strong electric fields (including intra-collisional field effects) tunneling phenomena, elastic and not, and, in general, all those effects which depend on the wavelike nature of the electrons.

Acknowledgements

The author is deeply indebted to W. R. Fresnley, A. Kumar, R. Landauer, S. E. Laux, P. J. Price, N. Sano and F. Stern for many discussions, suggestions, and criticisms, and to L. Reggiani for having drawn attention to the important work by van Hove.

References

[1] Frensley, W. R. (1990). "Boundary conditions for open quantum systems driven far from equilibrium", *Rev. Mod. Phys.*, **62**, 745–791.
[2] Pauli, W. (1928). In "*Festschrift zum 60. Geburtstage A. Sommerfeld*" (Hirzel, Leipzig), p. 30.
[3] Jacoboni, C. (1992). "Comparison between quantum and classical results in hot-electron transport", *Semicond. Sci. Technol.*, **7**, B6–B11.
[4] See, for instance, Khun, T. and Rossi, F. (1992). "Monte Carlo simulation of ultrafast processes in photoecited semiconductors: Coherent and incoherent dynamics", *Phys. Rev. B*, **46**, 7496–7514.
[5] Thanks are due to R. Landauer for making this suggestion, with the strong warning, however, that no formal proof has ever been given that this is indeed true.
[6] Pötz, W. (1989). "Self-consistent model of transport in quantum well tunneling structures", *J. Appl. Phys.*, **66**, 2458–2466.
[7] Kohn, W. and Luttinger, J. M. (1957). "Quantum Theory of Electrical Transport Phenomena", *Phys. Rev.*, **108**, 590–611.
[8] Léon, Van Hove (1955). "Quantum-mechanical perturbations giving rise to a statistical transport equation", *Physica*, **XXI**, 517–540.
[9] Jones, W. and March, N. H. (1973). "*Theoretical Solid State Physics, Volume 2: Non-equilibrium and Disorder*" (Wiley-Interscience, Bristol), p. 736 ff. and Appendix A6.1.
[10] Lent, G. S. and Kirkner, D. J. (1990). "The quantum transmitting boundary method", *J. Appl. Phys.*, **67**, 6353–6359.
[11] Kluksdahl, N. C., Kriman, A. M., Ferry, D. K. and Ringhofer, C. (1989). "Self-consistent study of the resonant-tunneling diode", *Phys. Rev. B*, **39**, 7720–7735.
[12] Fischetti, M. V. (1991). "Monte Carlo Simulation of Transport in Technologically Significant Semiconductors of the Diamond and Zinc-Blende Structure", *IEEE Trans. Electron Devices*, **38**, 634–649.
[13] Buot, F. A. and Rajagopal, A. K. (1995). "Theory of novel nonlinear quantum transport effects in resonant tunneling structures", *Mat. Sci. Eng. B*, **35**, 303–317.

Author Biography

Massimo V. Fischetti is a Research Staff Member at the IBM T. J. Watson Research Center in Yorktown Heights, New York. His research interests include electronic transport in semiconductors and insulators and the simulation of semiconductor devices. He is a Fellow of the American Physical Society.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 179–184
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Quantum Transport in Open Nanostructures

I. V. ZOZOULENKO* and K.-F. BERGGREN

Department of Physics and Measurement Technology, Linköping University, S-581 83 Linköping, Sweden

Electron transport was studied in an open square quantum dot with a dimension typical for current experiments. A numerical analysis of the probability density distribution inside the dot was performed which enabled us to unambiguously map the resonant states which dominate the conductance of the structure. It was shown that, despite of the presence of dot openings, transport through the dot is effectively mediated by just a few (or even a single) eigenstates of the corresponding closed structure. In a single-mode regime in the leads, the broadening of the resonant levels is typically smaller than the mean energy level spacing, Δ. On the contrary, in the many-mode regime this broadening typically exceeds Δ and has an irregular, essentially non-Lorentzian, character. It was demonstrated that in the latter case eigenlevel spacing statistics of the corresponding closed system are not relevant to the averaged transport properties of the dot. This conclusion seems to have a number of experimental as well as numerical verifications.

Keywords: Quantum dots, resonant states, conductance fluctuations

In nanoscaled semiconductor quantum dots, electron motion is confined in all spatial dimensions and the lateral shape of the dot can be controlled by an applied gate voltage [1–6]. In high quality samples at low temperatures electron transport is ballistic, i.e., large-angle elastic scattering events occur only at the boundaries of the structure and the phase coherence length well exceeds the dimension of the device. During recent years a great deal of effort has been focused on the transport properties of ballistic microstructures. In particular, both the statistical properties of the conductance fluctuations [1–4, 7–10] as well as geometry-specific, non-averaged features of the magnetoresistance of quantum dots [1, 4, 5, 9, 11] have been extensively studied.

Transport properties of open microstructures are often analyzed on the basis of the known characteristics of the spectrum of the corresponding isolated system. However, when the dot becomes open, eigenlevels interact and acquire a finite broadening due to the possibility for electrons to escape from the dot via lead openings. With several propagating modes in the leads, this broadening might well exceed the mean energy level separation, Δ, resulting in an overlap of a vast number of resonances. Also, the presence of dot openings may cause a significant distortion of

* Corresponding author.

corresponding eigenstates. Under these circumstances it is not *a priori* evident whether a discussion of transport through an open dot on the basis of the properties of the Hamiltonian of the closed structure is still meaningful. To the best of our knowledge, up to now no direct theoretical calculations on the actual broadening of the resonant levels for the open dots in the transmissive regime are available.

Besides, in the current literature there exists a number of conflicting reports on the effects of leads on the character of electron dynamics in open systems (chaotic vs regular). In particular, [12] shows that the statistics of the spectra for open dots are exactly the same as those of the corresponding closed system. At the same time, results [9, 10, 13] suggest that the leads attached to the dot may change the level statistics, so that transition to chaos can occur in a nominally regular system. On the contrary, Wang *et al.* [8] conclude that the openness of the dot makes chaotic scattering non-essential.

In this paper, on the basis of our direct mapping of resonant states performed for an open square dot, we hope to contribute to the clarification of some fundamental issues in this context.

The system under investigation is a relatively large square dot with the side $L = 1\,\mu\text{m}$ which is typical for current experiments. It is connected to reservoirs by quantum point contact (QPC)-like openings (leads), see Figures 1–3. For the sake of simplicity, hard wall confinement and a flat potential profile inside the dot are assumed which seems to be a good approximation for large dots [11]. We disregard effects of the soft impurity potential due to remote donors as well as inelastic scattering events.

Conductance through the dot in perpendicular magnetic field B at finite temperature is calculated within the Landauer-Büttiker formalism [14], where a transmission probabilities and wave functions were computed by making use of the recursive Green function technique [15]. Analyzing the probability density distribution inside the dot we are in a position to identify resonant energy

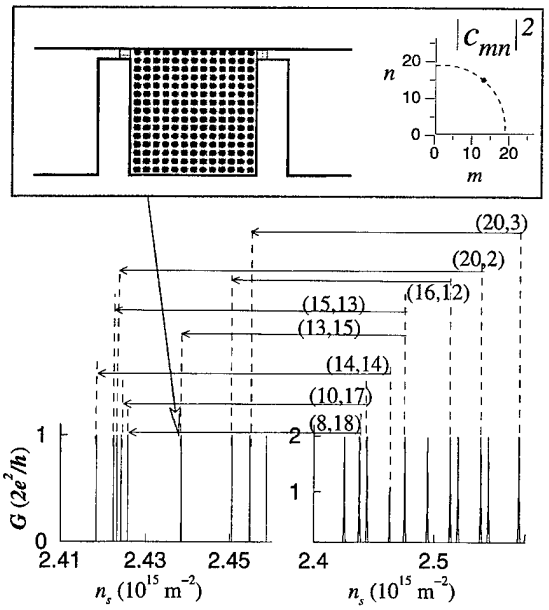


FIGURE 1 Lower left: The conductance of the square dot (schematically depicted in the upper panel) in the tunneling regime as a function of the sheet electron density $E_F m^*/\pi\hbar^2$. Temperature $T=0$. Shadow regions in the leads represent tunneling barriers with the height exceeding the Fermi energy. Lower right: Eigenenergy levels of the isolated dot. Height of the peaks represents the degree of degeneracy (1 or 2). Upper panel shows the calculated probability density distribution $|\Psi_{mn}|^2$ inside the dot for one of the tunneling peaks (left) and the corresponding numerical results for the coefficients $|c_{mn}|^2$ calculated on the basis of Eq. (1) (right). Dashed lines indicate the circle with the radius $r = k_F L/\pi$. A similar analysis has been done for the rest of the peaks and the correspondence between eigenstates of the isolated square and resonant levels of the dot is indicated by the arrows. Quantum numbers of the resonant states, (m, n), are shown in the parenthesis. In the case under consideration the side of the dot was chosen to be $L = 0.5\,\mu\text{m}$.

states which effectively mediate transport through the dot at a given E_F. To do this, we numerically expand the solution of the scattering problem in the open dot, $\Psi(x, y; E)$, in the set of eigenstates of the closed dot, $\psi_{mn} = \frac{2}{L}\sin\frac{\pi m x}{L}\sin\frac{\pi n y}{L}$ (with eigenenergies $\varepsilon_{mn} = \hbar^2/2m^*(k_m^2 + k_n^2); k_m = \frac{\pi m}{L}$, $k_n = \frac{\pi n}{L}$)

$$\Psi(x, y; E) = \frac{2}{L}\sum_m\sum_n c_{mn}(E)\sin\frac{\pi m x}{L}\sin\frac{\pi n y}{L}. \quad (1)$$

Coefficients c_{mn} represent the contributions of the eigenstates m, n in the total wave function.

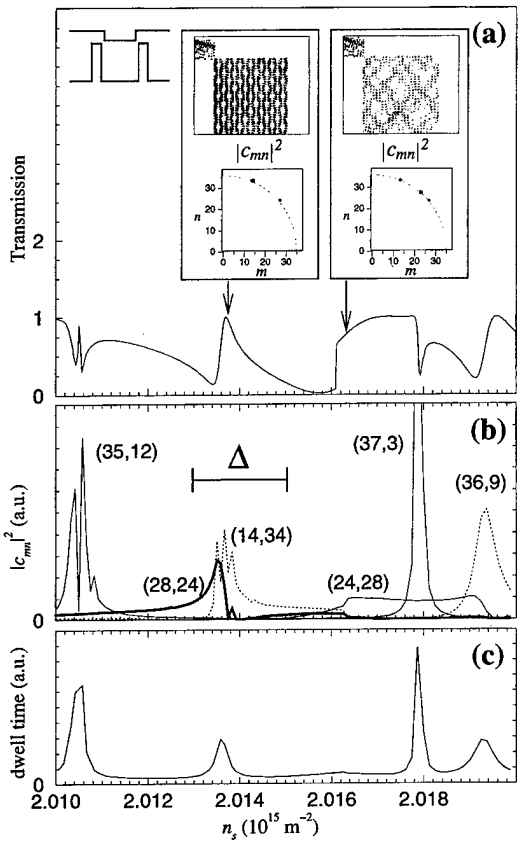


FIGURE 2 The conductance of the square dot (schematically depicted in the inset) as a function of the sheet electron density $n_s = E_F m^*/\pi\hbar^2$. The side of the dot L = 1 μm; temperature $T=0$. The lead openings support one propagating mode. Insets show calculated $|\Psi_{mn}|^2$ inside the dot for two representative values of n_s and the corresponding numerical results for the coefficients $|c_{mn}|^2$ calculated on the basis of Eq. (1). Dashed lines indicate the circle with the radius $r = k_F L/\pi$. (b) Dependence of the coefficients $|c_{mn}|^2$ identifying the dominant resonant states with the quantum numbers (m, n) on the sheet electron density n_s. A horizontal bar indicates the mean energy level spacing Δ. (c) Dwell time of the dot.

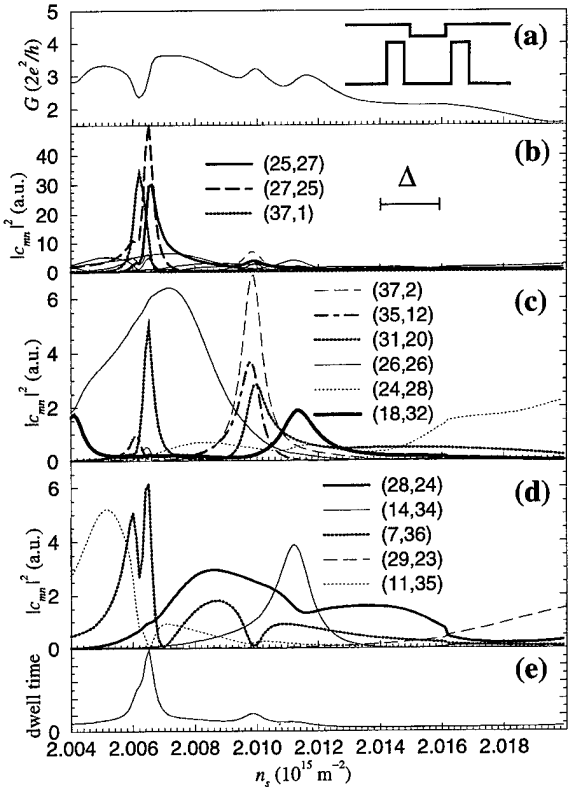


FIGURE 3 The conductance of the square dot (see inset) in a many-mode regime in the lead openings, $N=5$. The side of the dot $L=1$ μm; temperature $T=0$. (b) Coefficients $|c_{mn}|^2$ identifying the dominant resonant states with the quantum numbers (m, n). A horizontal bar indicates the mean energy level spacing Δ. (c), (d) The same as (b) but in a refined scale. (The contribution from the three dominant states indicated in (b) is not shown). (e) Dwell time of the dot.

In what follows we focus on the three different transport regimes, namely the tunneling, the single-mode, and the many-mode regimes.

Tunneling regime Figure 1(a) shows conductance of the square dot in a regime when the dot is weakly coupled to the leads. Each conductance peak corresponds to an excitation of *one single* resonant energy level which effectively mediates transport at the given Fermi energy. Near its maximum, each peak is characterized by the Lorentzian shape, in accordance to the Breit-Wigner formalism. The positions of the peaks are shifted with respect to the corresponding eigenenergy levels of the isolated square (cf. Figs. 1(a) and (b)).

Single-mode regime Figure 2(a) shows the conductance of the square dot where lead openings are adjusted to support one propagating mode. Here the pattern of the probability density distribution $|\Psi_{mn}|^2$ exhibits a complicated structure, where eigenstates of the isolated square are not easily recognized. Thus, a numerical analysis on the basis of Eq. (1), in contrast to the tunneling case, is essential. Calculating the expansion coefficients

c_{mn}, we find that, at the given k_F, only those coefficients associated with the circle in the k-space with the radius $R = k_F \approx \sqrt{2m^*\varepsilon_{mn}}/\hbar = \pi\sqrt{m^2 + n^2}/L$ are distinct from zero, see Figure 2(a). (Note that non-vanishing contributions from other coefficients would indicate that eigenstates of the isolated dot are essentially distorted by the lead openings such that a discussion of the transport in open structure in terms of eigenstates of the isolated dot does not make sense). Typically, we find that only a few (and sometimes even a single) coefficients give a dominant contribution. A broadening of the resonant levels due to the effect of the dot openings in the k-space is less than the distance between neighboring eigenstates whose quantum numbers differ by one, $\Delta k = |k_n - k_{n\pm1}| = |k_m - k_{m\pm1}| = \pi/L$. Therefore, we conclude that *despite the presence of dot openings, transport through the structure is still effectively mediated by a few eigenstates of the corresponding closed dot with the eigenenergies lying in close proximity to the Fermi energy, $\varepsilon_{mn} = E_F$.*

Calculating the coefficients c_{mn} as a function of the Fermi energy, we extract information about the lead-induced broadening of the energy levels of the dot. A contribution of the dominant resonant energy states is shown in Figure 2(b). The mean energy level spacing, $\Delta = 2\pi\hbar^2/m^*L^2 \sim 70$ mK, is indicated by a horizontal bar. In contrast to the tunneling regime, the lineshape of $|c_{m,n}(E)|^2$ can be non-Lorentzian. Moreover, different states are characterized by different broadenings and they may overlap with each other. However, a broadening (half-width) of the resonant energy levels is typically less than Δ. Therefore, *transport measurements at very low temperatures in a single-mode regime in the leads may probe a single resonant energy level of the dot.*

A comparison between Figures 2(a) and (b) shows that features in the conductance of the dot are related to excitations of the particular eigenstates of the square. However, this correspondence is rather complicated: different eigenstates can be responsible for opposite features in the dot conductance (dips and peaks).

Many-mode regime As the lead openings become wider, a number of the resonant states excited in the dot increases. Nevertheless, like in the single-mode regime, at the given Fermi energy, *a non vanishing contribution comes only from the coefficients which lie in the closest proximity to the circle with the radius k_F in the k-space.* Therefore, even in a many-mode regime, transport through an open structure is still effectively mediated by eigenstates of the corresponding closed dot. A broadening of the resonant energy levels increases with an increase of the lead openings. In Figure 3 wave function patterns are analyzed on the basis of Eq. (1) and contributions from dominant states are shown in a representative interval of the Fermi energy. Typically, several states dominate transport at a given E_F. A broadening of the energy levels has, as a rule, a complicated essentially non-Lorentzian character with half-width being different for different states. In contrast to the single mode regime, a half-width of the resonant energy levels is typically larger than the mean energy level spacing, Δ. Comparing the conductance of the dot, Figure 3; and the dependence $c_{mn} = c_{mn}(E_F)$, one can trace a certain correspondence between the two. However, because many eigenstates typically contribute to the conductance at a given E_F, a detailed explanation of the features of the dot conductance is not possible.

In what follows we critically examine several approaches [9, 12] to the analysis of the statistics of the spectra of open dots. A statistical analysis of the distribution of the energies at resonances of conduction fluctuations in chaotic stadium and regular circular billiards has been performed by Ishio [9]. In both billiards the statistics follows Wigner-type distribution which was taken as an indication of the transition to chaos. (The Wigner-type statistics of the nearest energy level spacing distribution is characteristic of the classically chaotic closed billiards). With regard to this analysis, the question immediately arises "does the resonance energy statistics of the conductance fluctuations reproduce the corresponding statistics of the isolated system?" We have shown above

that even in a single mode regime, resonant energies in the conductance fluctuations only occasionally correspond to resonant eigenstates of the isolated dot. Instead, in most cases resonant energies are related to those energies when more than one state is simultaneously excited in the dot such that their mutual interference leads to the resonance behavior of the transmission coefficient. Therefore, in our opinion, Wigner-type statistics of the spacing of the conductance fluctuation resonances cannot be taken as an indication of the transition to chaos in a nominally regular but open system.

Wang *et al.* [12] analyzed statistics of the open system on the basis of the calculated electron dwell time, $\tau \sim \int ds|\Psi_{mn}|^2$, which identifies the time an electron spends inside the dot; in the above definition an integration is performed within the dot area. Statistics of the spectra were found to be exactly the same as that of the corresponding closed system. Again, a similar question arises "does the dwell time maxima (which is integrated characteristic) unambiguously identify individual resonant states of the open dot?".

Figures 2(c) and 3(c) show the calculated dwell time in the square dot in the single- and many-mode regimes respectively. In a single-mode regime all maxima in the dwell time *do correspond* to the resonant eigenstates, (cf. Figs. 2(b) and (c)). However, dwell time does not identify resonant states unambiguously because some eigenstates are overlooked by this analysis. Nevertheless, since the number of missing states is usually small our analysis, as far as a single mode regime is concerned, tends to support the conclusion [12] that the statistics of the dwell time spectra for open dots are the same as that of the corresponding closed systems.

In a many-mode regime, our analysis strongly suggest that the resonant level spacing statistics become ill-defined. This is not only because the broadening of resonant states typically exceeds mean level spacing (for Lorentzian broadening one can still define statistics of the spacings between peak maxima). This is due to the fact that the broadening itself in many cases is essentially non-Lorentzian, see Figure 3(c). For most of the resonant states the concept of statistics of the spectra does not make any sense, since it is not possible to introduce any reasonable definition of the spacing between resonances. Therefore, we conclude that *for the quantum dots strongly coupled to the leads with several modes available in the lead openings, eigenlevel spacing statistics of the corresponding closed system are not relevant to the averaged transport properties of the structure.*

This conclusion seems to have a number of experimental as well as numerical verifications. The difference between statistical properties of the conductance oscillations in a chaotic (stadium) and a regular (circular) dots has been studied by Marcus *et al.* [1]. Corresponding averaged autocorrelation functions are almost identical over the two decades of decay, although they exhibit quantitative distinctions in the tail. The data, from the similar studies of Berry *et al.* [2] for chaotic (circular with a bar) and regular (circular) dots, does not show any significant discrepancy over the four orders of magnitude in power. Numerical studies of autocorrelations functions for a chaotic (stadium) and a regular (circular) dots [8] in a many-mode regime show the similar behavior, which is in accordance with our arguments.

To conclude, despite of the presence of dot openings, transport through the open dot is effectively mediated by just a few eigenstates of the corresponding closed structure. In a single-mode regime in the leads the broadening of the resonant levels is typically smaller than the mean energy level spacing, Δ. On the contrary, in the many-mode regime the broadening exceeds Δ and has essentially a non-Lorentzian character.

Acknowledgement

I.V.Z. acknowledges a grant from the Royal Swedish Academy of Sciences.

References

[1] Marcus, C. M., Rimberg, A. J., Westervelt, R. M., Hopkins, P. F. and Gossard, A. C., "Conductance Fluctuations and Chaotic Scattering in Ballistic Microstructures", *Physical Review Letters*, July 1992.

[2] Berry, M. J., Katine, J. A., Westervelt, R. M. and Gossard, A. C., "Influence of Shape on Electron Transport in Ballistic Quantum Dots", *Physical Review B*, December 1994.

[3] Chang, A. M., Baranger, H. U., Pfeiffer, L. N. and West, K. W., Weak Localization in Chaotic versus Nonchaotic Cavities: A Striking Difference in the Line Shape", *Physical Review Letters*, October 1994.

[4] Persson, M., Pettersson, J., von Sydow, B., Lindelof, P. E., Kristensen, A. and Berggren, K.-F., "Conductance Oscillations Related to the Eigenenergy Spectrum of a Quantum Dot in Weak Magnetic Field", *Physical Review B*, September 1995.

[5] Bird, J. P., Ferry, D. K., Akis, R., Ochiai, Y., Ishibashi, K., Aoyagi, Y. and Sugano, T., "Periodic Conductance Oscillations and Stable Orbits in Mesoscopic Semiconductor Billiards", *Europhysics Letters*, September 1996.

[6] Zozoulenko, I. V., Schuster, R., Berggren, K.-F. and Ensslin, K., "Ballistic Electrons in an Open Square Geometry: Selective Probing of Resonant-Energy Levels", *Physical Review B*, April 1997.

[7] Jalabert, R. A., Baranger, H. U. and Stone, A. D., "Conductance Fluctuations in the Ballistic Regime: A Probe of Quantum Chaos?", *Physical Review Letters*, November 1990.

[8] Wang, Y., Wang, J., Gou, H. and Roland, C., "Tunneling Through Quantum-Dot Systems: A Study of the Magneto-Conductance Fluctuations", *J. Physics: Condensed Matter*, August 1994.

[9] Ishio, H., "Quantum Transport and integrability in Open Billiards", *J. Statistical Phyics*, January 1996.

[10] Albeverio, S., Haake, F., Kurasov, P., Kuś, M. and Šeba, P., "S-matrix, resonances and wave functions for transport through billiard with leads", *J. Math. Phys.*, October 1996.

[11] Reiman, S. M., Persson, M., Lindelof, P. E. and Brack, M., "Shell Structure of a Circular Quantum Dot in a Weak Magnetic Field", *Z. Phys. B*, October 1996.

[12] Wang, Y., Zhu, N., Wang, J. and Guo, H., "Resonance States of Open Quantum Dots", *Physical Review B*, June 1996.

[13] Berggren, K.-F. and Ji, Z.-L., "Quantum Chaos in Nano-sized Billiards in Layered Two-Dimensional Semiconductor Structures", *CHAOS*, November 1996.

[14] Datta, S., *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, 1995).

[15] Zozoulenko, I. V., Maaø, F. A. and Hauge, E. H., "Coherent Magnetotransport in Confined Arrays of Antidots", *Physical Review B*, March 1996.

Authors' Biographies

Igor Zozoulenko is a senior scientist at the Department of Physics at Linköping University. His current research deals with electronic aspects of low-dimensional semiconductor structures like quantum dots, wells and related systems. During his career he has awarded fellowships from Royal Norwegian Council for Scientific and Industrial Research and Royal Swedish Academy of Sciences, and has held positions at Institute for Theoretical Physics, Kiev.

Karl-Fredrik Berggren is Professor of Theoretical Physics at the Department of Physics at Linköping University and head of the Laboratory of Theoretical Physics. In general terms his research is focused on condensed matter physics. He is a member of the Swedish Council for High Performance Computing, a member of the Computational Physics Board of the European Physical Society, and a member of the Condensed Matter Theory Committee at the Nordic Institute of Theoretical Physics (NORDITA), Copenhagen.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 185–190
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Application of the Wigner-Function Formulation to Mesoscopic Systems in Presence of Electron-Phonon Interaction

C. JACOBONI, A. ABRAMO, P. BORDONE*, R. BRUNETTI and M. PASCOLI

Istituto Nazionale per la Fisica della Materia, Dipartimento di Fisica, Università di Modena, Via Campi 213/A, 41100 Modena, Italy

A theoretical and computational analysis of the quantum dynamics of charge carriers in presence of electron-phonon interaction based on the Wigner function is here applied to the study of transport in mesoscopic systems. Numerical applications are shown for a) a wave packet scattering with phonons while crossing a potential profile and b) electrons scattering with phonons in a finite device with open boundary conditions.

Keywords: Quantum transport, Wigner function, mesoscopic systems, boundary conditions

1. INTRODUCTION

A rigorous quantum transport theory of electrons in semiconductor structures including both coherent propagation and scattering mechanisms is still lacking even though its need has been recognized since many years. In particular we miss a theory feasible for numerical calculations to be compared with experimental results and/or with results of the semiclassical theory with the purpose of enlightening the different predictions of the two models.

In the literature different approches have been used [1–3]. The Wigner function (WF) [1] provides a rigorous quantum-mechanical tool and constitues a direct link between quantum and classical descriptions of the evolution of the system in phase space. In this approach however electron-phonon interaction has been treated using either relaxation time [4, 5] or a classical Boltzmann collision operator [6]. In this paper several advancements in the effort to solve the above problem exactly are presented.

The definition of the WF can be extended to include electrons interacting with phonons as follows [7]:

$$f_w(\boldsymbol{r},\boldsymbol{p},n_q,n'_q,t)=\frac{1}{h^3}\int d\boldsymbol{r}'e^{-i\frac{\boldsymbol{p}}{\hbar}\boldsymbol{r}'}\rho(\boldsymbol{r}+\boldsymbol{r}'/2,n_q;\boldsymbol{r}-\boldsymbol{r}'/2,n'_q) \tag{1}$$

* Corresponding author.

where $\boldsymbol{r}$ and $\boldsymbol{p}$ are the electron position and momentum, ρ is the density matrix for the system formed by an electron and the phonon gas, and n_q is the occupation number of the phonon mode $\boldsymbol{q}$. Trace over the phonon variables will lead to the traditional electron WF.

In this paper we discuss an integral equation for the above WF and apply it to study quantum electron transport in open systems in presence of scattering. In particular suitable boundary conditions can be used which restrict the analysis of the WF inside a finite domain. Wigner trajectories have been identified also for the case of electron-phonon interaction and constitute a guiding criterion for replacing initial conditions with boundary conditions for the integral equation.

2. INTEGRAL EQUATION

Let us consider a system described by the Hamiltonian

$$\boldsymbol{H} = \boldsymbol{H}_0 + \boldsymbol{H}' \tag{2}$$

where $\boldsymbol{H}_0$ is the "unperturbed" Hamiltonian, containing the description of the free electrons, free phonons and, in the present approach, also the potential profile inside the system (device). $\boldsymbol{H}'$ contains the scattering agents, i.e., the electron-phonon interaction.

The time evolution of the density matrix is described by the Liouville-von Neumann equation. Starting from this equation we derived the following integral equation for the WF:

$$\begin{aligned} f_w(\boldsymbol{r},\boldsymbol{p},n_q,n'_q,t) &= h^3 \sum_{n,n'} f_{nn'}(\boldsymbol{r},\boldsymbol{p}) \\ & e^{-i(\omega(n,n_q)-\omega(n',n'_q))(t-t_0)} \int d\boldsymbol{r}' \int d\boldsymbol{p}' f^*_{nn'}(\boldsymbol{r}',\boldsymbol{p}') \\ & \times f_w(\boldsymbol{r}',\boldsymbol{p}',n_q,n'_q,t_0) + h^3 \sum_{n,n'} f_{nn'}(\boldsymbol{r},\boldsymbol{p}) \\ & \int_{t_0}^{t} dt' e^{-i(\omega(n,n_q)-\omega(n',n'_q))(t-t')} \end{aligned} \tag{3}$$

$$\begin{aligned} & \times \sum_{mm_q} \int d\boldsymbol{r}' \int d\boldsymbol{p}' \{ \mathcal{H}'(nn_q,mm_q) \\ & f^*_{mn'}(\boldsymbol{r}',\boldsymbol{p}') f_w(\boldsymbol{r}',\boldsymbol{p}',m_q,n'_q,t') \\ & - f^*_{nm}(\boldsymbol{r}',\boldsymbol{p}') f_w(\boldsymbol{r}',\boldsymbol{p}',n_q,m_q,t') \mathcal{H}'(mm_q,n'n'_q) \} \end{aligned}$$

In the above equation $\mathcal{H}' = \boldsymbol{H}'/i\hbar$, and $\mathcal{H}'(nn_q, mm_q)$ are its matrix elements on the basis $\{|\phi_l, n_q\rangle\}$ of $\boldsymbol{H}_0$. The coefficients f_{lm} $(\boldsymbol{r}, \boldsymbol{p})$ are given by

$$f_{lm}(\boldsymbol{r},\boldsymbol{p}) = \frac{1}{h^3} \int d\boldsymbol{r}' e^{-i(p/\hbar)r'} \left\langle \boldsymbol{r} + \frac{\boldsymbol{r}'}{2} | \phi_l \right\rangle \left\langle \phi_m | \boldsymbol{r} - \frac{\boldsymbol{r}'}{2} \right\rangle \tag{4}$$

They allow to move from the WF to the density-matrix representation and viceversa. Finally $\hbar\omega(n, n_q)$ indicates the total unperturbed energy of the electron-phonon system when the electron is in the n-th eigenstate and the phonon bath is in a state with n phonons in mode $\boldsymbol{q}$.

In Eq. (3) the first term in the r.h.s. describes the ballistic coherent propagation from the initial time t_0 to the observation time t. The second term describes the contribution of the unknown WF to the "last" interaction vertex at any time t' between t_0 and t followed by the free propagation from t' to t. Even though the equation is written for the entire system and contains also the phonon variables, it has been shown that a Monte Carlo solution allows for a proper average over the phonon coordinates, as long as hot-phonon effects are neglected [8].

3. WIGNER TRAJECTORIES

It is well known that a point-like electronic WF in phase space cannot represent a real physical system, since it violates the uncertainty principle. We may however consider the evolution in time of a δ-like contribution to the WF inside Eq. (3). For free electrons its trajectory in phase space is the same as for semiclassical particles. In fact, for a plane-wave basis the coefficients f_{lm} are given by

$$f_{kk'}(\boldsymbol{r},\boldsymbol{p}) = \frac{1}{(2\pi)^3} e^{i(k-k')r} \delta^3 \left(\boldsymbol{p} - \frac{\hbar}{2}(\boldsymbol{k}+\boldsymbol{k}') \right) \tag{5}$$

and the ballistic evolution of the WF as given by the first term on the rhs of Eq. (3) yields:

$$f_w(\boldsymbol{r},\boldsymbol{p},n_q,n'_q,t) = f_w\left(\boldsymbol{r}-\frac{\boldsymbol{p}}{m}(t-t_0),\boldsymbol{p},n_q,n'_q,t_0\right) e^{-i(\omega(n_q)-\omega(n'_q))(t-t_0)} \quad (6)$$

where m is the electron effective mass. The first factor and the exponential factor describe the free trajectory of the electron and the time evolution of the free-phonon bath, respectively. This result is not surprising since in absence of scattering and up to quadratic potentials (harmonic oscillator) a differential equation can be written for the WF that coincides with the Boltzmann equation, so that the dynamical evolution of the WF follows the same trajectories of the representative points of a classical gas. This implies a deformation of the WF while evolving in time due to the fact that, as time increases, the higher-momentum components move faster than the lower-momentum components.

If phonon scattering is included, similar trajectories exist, with very interesting properties, as long as a single interaction diagram is considered without time integrations. In order to prove this result Eq. (3) must be substituted into itself to obtain the second-order term corresponding to one scattering event. Then assuming a constant potential and using again plane waves as basis set, the second-order correction, e.g., for the terms corresponding to phonon emission, results to be

$$\Delta f_w(z,p_z,n_q,n_q,t) = 2\mathcal{R}\left\{\sum_q F^2(q)\langle n_q\rangle \int_{t_0}^{t} dt' \int_{t_0}^{t'} dt'' e^{-iq\frac{p_z+\hbar q/2}{m}(t'-t'')} e^{i\omega_q(t'-t'')} \times f_w\left(z-\frac{p_z}{m}(t-t')-\frac{p_z+\hbar q/2}{m}(t'-t'') - \frac{p_z+\hbar q}{m}(t''-t_0), p_z+\hbar q, n_q-1, n_q-1, t_0\right)\right\} \quad (7)$$

where for simplicity a one-dimensional case is considered and $F(q)$ is the coupling function entering the electron-phonon hamiltonian [9].

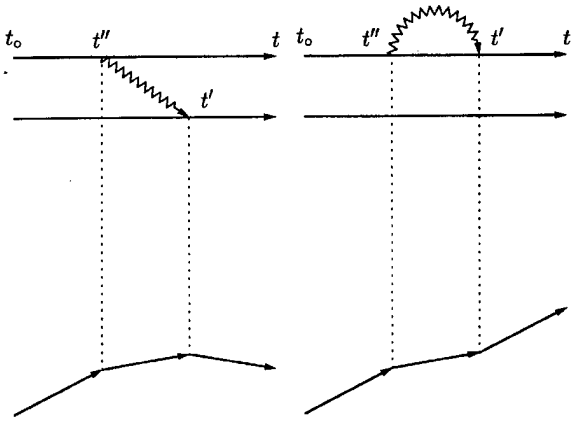


FIGURE 1 Example of the Wigner trajectories that contribute to the result in Eq. (3) for real and virtual emissions of a mode q.

The trajectories described by the above equation are represented in the diagrams contained in Figure 1. They correspond to the semiclassical trajectories where $q/2$ is transferred to the electron at each vertex. As a result of the time integrations, however, a sum over infinite trajectories gives the WF correction corresponding to a single scattering event.

4. INITIAL AND BOUNDARY CONDITIONS

Guided by the considerations of the previous section we face the problem of finding an equation equivalent to Eq. (3) for the WF in a finite region inside given boundaries. The problem is not trivial owing to the nonlocality of the interaction and to the integration from $-\infty$ to $+\infty$ in the integral over z'. We have proved that for the solution of the integral equation it is possible to substitute the knowledge of the WF over all space at a given (initial) time t_0 with the knowledge of the same function inside the region of interest at t_0 and at all times at the boundary for "entering" momenta. The integral equation for such a situation results to be

$$f_w(z,p_z,n_q,n'_q,t) = h\sum_{nn'} f_{nn'}(z,p_z)e^{-i(\omega(nn_q)-\omega(n'n'_q))(t-t_0)}\int_{-A}^{A} dz'$$

$$\int dp'_z f^*_{nn'}(z',p'_z) f_w(z',p'_z,n_q,n'_q,t_0)$$

$$+ h\sum_{nn'} f_{nn'}(z,p_z)\int_{t_o}^{t} dt' e^{-i(\omega(nn_q)-\omega(n'n'_q))(t-t')}$$

$$\left\{\left[\int_0^{+\infty} dp'_z \frac{p'_z}{m} f^*_{nn'}(-A,p'_z)\right.\right.$$

$$\times f_w(-A,p'_z,n_q,n'_q,t') - \int_{-\infty}^{0} dp'_z \frac{p'_z}{m} f^*_{nn'}(A,p'_z)$$

$$\left.f_w(A,p'_z,n_q,n'_q,t')\right]$$

$$+ \int_{-A}^{A} dz' \int dp'_z \sum_{mm_q} [\mathcal{H}'(nn_q,mm_q) f^*_{mn'}(z',p'_z)$$

$$f_w(z',p'_z,m_q,n'_q,t') - f^*_{nm}(z',p'_z)$$

$$\left.f_w(z',p'_z,n_q,m_q,t')\mathcal{H}'(mm_q,n'n'_q)]\right\}. \qquad (8)$$

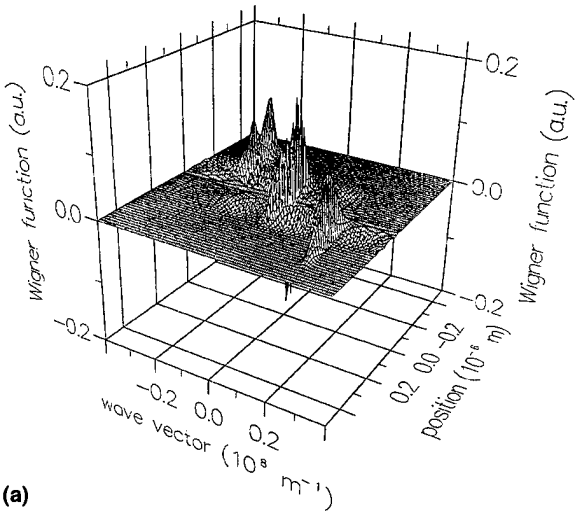
where again for simplicity a one-dimensional case has been considered, and $-A$ and A are the device boundaries.

5. APPLICATIONS

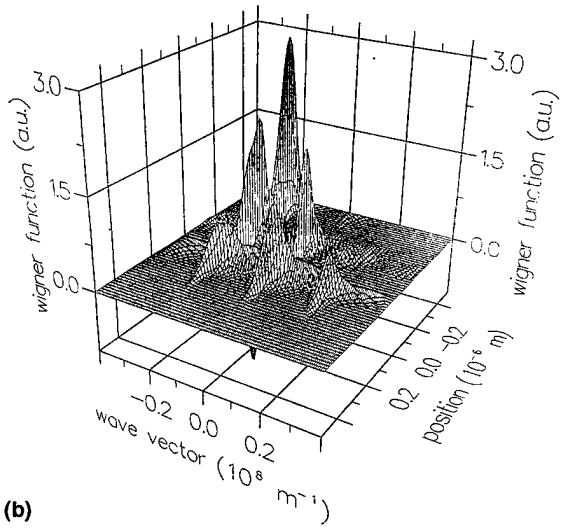
As a case study we have first analysed a single electron-phonon scattering event for a wave packet crossing a potential profile [10]. In this case the numerical procedure is simplified by the fact that the whole system can be enclosed in a large, but finite box. Thus the basis set of functions is discrete and can be used without finite-difference approximations in k-space. Three cases have been considered for a single electron-phonon scattering event:

a) The effect of a quantum collision over the free propagation,
b) Intra-collisional field effect when the phonon interaction takes place during a propagation through a region with a constant electric field, and
c) Scattering during the otherwise resonant propagation through a double barrier.

As an example Figure 2 shows the scattering and scattered WF at $t = 1.5$ ps after the initial condition for an electron crossing a double barrier in resonant conditions and emitting an optical phonon. The interaction is switched on when the packet hits the double barrier, namely at $t = 0.5$ ps after the initial condition. A complex quantum-



(a)



(b)

FIGURE 2 Unperturbed Wigner function (above) and second-order correction due to a quantum phonon emission (below) for an electron crossing a double-barrier potential profile in resonant conditions at different times. The interaction is active from $t = 0.5$ ps after the initial condition.

dynamical interference between potential profile and scattering is observed at successive times [10].

In order to verify the correctness of the use of the boundary conditions for an open mesoscopic system as described in Section 4 we have first evaluated the ballistic evolution of the first term in Eq. (8) assuming a step potential profile, $f_w = 0$ at $t = t_0$ inside the device, and a Maxwellian local-equilibrium incoming distribution at the left and right boundaries. Results are shown in Figure 3 at different times after the initial condition.

At the longest times (not shown in the figure for space reasons) we recover for f_w the same result which is obtained without using boundary conditions for a stationary infinite system, where f_w is evaluated by means of an equilibrium density matrix diagonal over the scattering states.

In order to describe quantum transport across a very small device in presence of scattering the integral in Eq. (8) is expanded perturbatively through iterative substitutions. The second-order

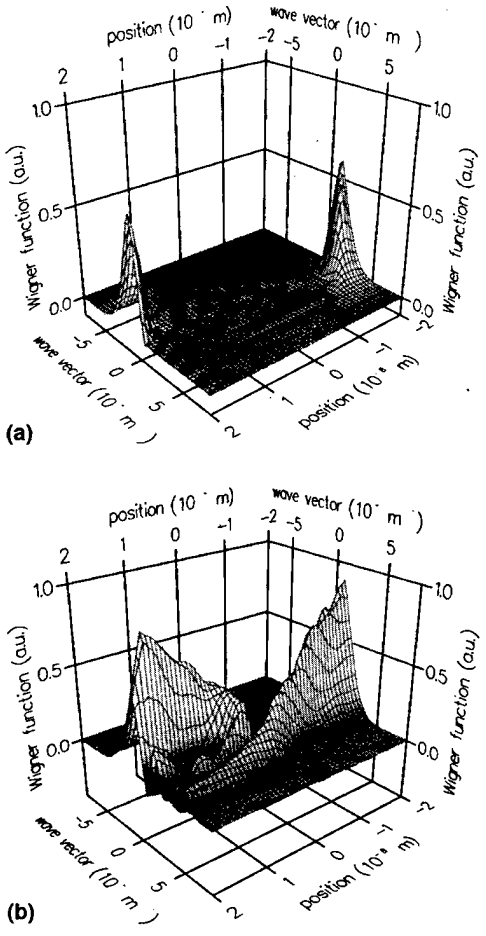


FIGURE 3 Ballistic evolution of the first term in Eq. (8) at two times (t = 10 fs and 80 fs) for electrons entering from the boundaries into a region with a potential step.

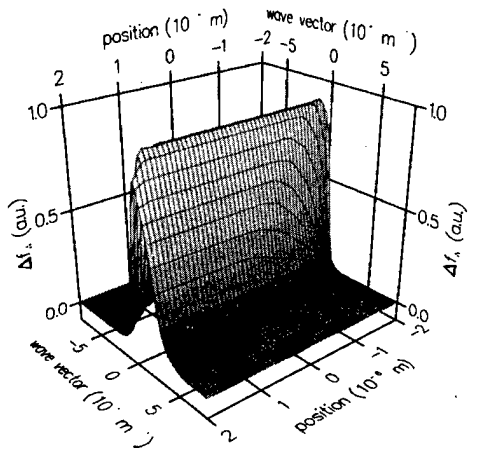


FIGURE 4 In-emission term summed over all phonon modes q at t = 50 fs after the electron-phonon interaction has been switched on.

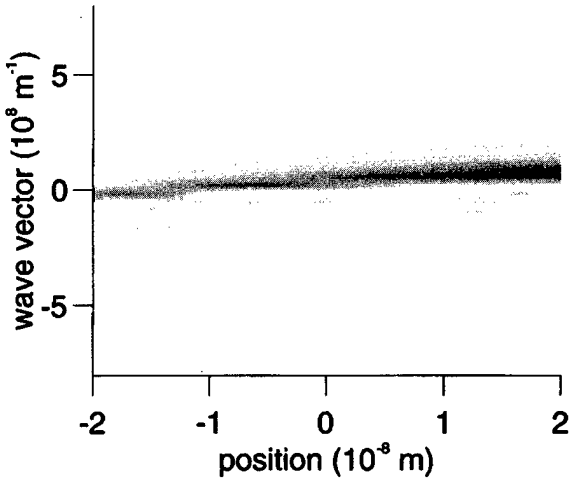


FIGURE 5 Contribution to the Wigner function due to real emission of a mode $q = 1.66 \cdot 10^8$ m^{-1} at 200 fs after the electron-phonon interaction has been switched on. Only electrons entering with positive p are able to emit. At long times classical conservation is recovered.

contribution is the sum of four terms: real and virtual absorptions and emissions. As an example Figure 4 shows the result for the in-emission term summed over all phonon modes q which compares very well with the classical result.

An interesting difference instead emerges when only a single mode q is considered in the quantum calculation, as shown in Figure 5. At short times right after the entrance energy non-conservation modifies the in-scattering term.

Acknowledgement

Partially supported by ARO and ONR through ERO through Contract n. N68171-96-C-9089.

References

[1] Frensley, W. R. (1990). Boundary conditions for open quantum systems driven far from equilibrium. *Rev. Mod. Phys.*, **62**(3), 745.
[2] Datta, S. (1990). A simple kinetic equation for steady-state quantum transport. *J. Phys. Condens. Matter*, **2**, 8023.
[3] Rossi, F., Ragazzi, S., Di Carlo, A. and Lugli, P. (1997). A generalized Monte Carlo approach for the analysis of quantum-transport phenomena in mesoscopic systems: Interplay between coherence and relaxation. *To be published in VLSI Design.*
[4] Frensley, W. R. (1987). Wigner-function model of a resonant-tunneling semiconductor device. *Phys. Rev. B*, **36**, 1570.
[5] Kluksdahl, N. C., Kriman, A. M., Ferry, D. K. and Ringhofer, C. (1989). Self-consistent study of the resonant-tunneling diode. *Phys. Rev. B*, **39**(11), 7720.
[6] Frensley, W. R. (1988). Quantum Transport Modeling of Resonant-Tunneling Devices. *Solid State Electron.*, **31**, 739.
[7] Rossi, F., Jacoboni, C. and Nedjalkov, M. (1994). A Monte Carlo solution of the Wigner transport equation. *Semicond. Sci. Technol.*, **9**, 934.
[8] Brunetti, R., Jacoboni, C. and Rossi, F. (1989). Quantum Theory of Transient Transport in Semiconductors: A Monte Carlo approach. *Phys. Rev. B*, **39**, 10781–10790.
[9] Brunetti, R., Jacoboni, C., Venturi, F., Sangiorgi, E. and Riccò, B. (1989). A many-band silicon model for hot-electron transport at high energies. *Solid State Electron.*, **32**, 1663–1667.
[10] Brunetti, R. and Jacoboni, C. (1997). Wave packet analysis of electron-phonon interaction in the Wigner formalism. *To be published.*

Authors' Biographies

Carlo Jacoboni (Ph.D. in Physics at Purdue University, Indiana, in 1969) is Full Professor of Theoretical Physics and Dean of the Faculty of Sciences.

Antonio Abramo (Ph.D. in Electrical Engineering at the University of Bologna, Italy in 1995) is Post-Doc at the University of Modena.

Paolo Bordone (Ph.D. in Physics at the University of Modena, Italy in 1992) is Post-Doc at the University of Modena.

Rossella Brunetti (Ph.D in Physics at the University of Modena, Italy in 1987 is Research and Teaching Assistant at the Physics Department of the University of Modena.

Marco Pascoli (Laurea in Physics at the University of Modena, Italy in 1996).

The research activity of the group is mainly related to Monte Carlo applications to semiclassical and quantum transport and electron-device modeling.

QUANTUM SIMULATION II

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 191–196
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Quantum Networks: Dynamics of Open Nanostructures

GÜNTER MAHLER and RAINER WAWER*

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

The superposition principle makes quantum networks behave very differently from their classical counterparts: We discuss how local and non-local coherence are generated and how these may affect the function of composite systems. Numerical examples concern quantum trajectories, quantum noise and quantum parallelism.

Keywords: Coherence, entanglement, quantum trajectories, quantum parallelism

1. GENERAL BACKGROUND

Beyond some 15 years from now any realistic roadmap for nanoelectronics [1] tends to become rather vague: The present trend towards miniaturization is expected to continue till then, but as physics is not scale-invariant, fundamental limitations will be piling up, all of which are eventually related to quantum features of some sort: there are limits of manufacturing, limits of control, limits of simulation capabilities, limits of classical system theory, and limits of classical models of computation, to name but a few.

We will not be able to discuss these issues here at any depth; instead, we intend to address some pertinent questions based on "toy-models" rather than sophisticated device simulations. Nevertheless, this approach should be able to demonstrate in what sense quantum networks will differ from classical ones.

1.1. How Can We Design Quantum Systems?

Stable structure, though eventually a quantum phenomenon in its own right, enters nanophysics as a "classical" design tool: Confinement of electron-(or photon-) fields allows to discretize the state space, at least in some energy range of interest. This "quantum-size-effect" considerably simplifies the specification and observation of the non-classical features to be discussed below, as it significantly reduces the number of "subsystems" to be considered.

1.2. What is a Subsystem?

It turns out that the only viable definition of a subsystem is operational: A subsystem has to be separated out by the classical environment (experimental set-up) in terms of measurement modes. A quantum dot or an atom (within an array) will be a

* Corresponding author.

subsystem if and only if it is selectively "contacted". Note that if this condition is not fulfilled, the array has to be treated as a unit; though this will not change the total state space, it will certainly change the type of possible observations.

1.3. How do we Describe Subsystems?

We restrict ourselves to subsystems μ characterized by n different states $|p_\mu\rangle$, each. The number of independent observables (represented by a set of orthonormalized operators $\hat{\lambda}_j(\mu)$) is then restricted to n^2, i.e., $j = 0$ (denoting the unit-operator: "no action"), $1,\ldots,n^2-1$. Any operator can be written as some linear combination of these. We may choose the $\hat{\lambda}_j$ such that the expectation-values $\langle\hat{\lambda}_j\rangle$ form a vector [2], [3], for $n = 2$ the so-called Bloch-vector.

1.4. How do we Describe Networks?

Any operator (for a network composed of N subsystems, say) can be written as a linear combination of the n^{2N} "cluster-operators"

$$\hat{Q}_{jk\ldots r} = \hat{\lambda}_j(1)\hat{\lambda}_k(2)\ldots\hat{\lambda}_r(N) \qquad (1)$$

These $\hat{Q}$, in general, do not commute; they come in different classes, specified by the number $m \leq N$ of subsystems which are actually affected (which equals the number of indices unequal zero): $\hat{Q}_{00\ldots0}$ is the unit-operator, $\hat{Q}_{j0k\ldots0}$ an operator acting on $\mu = 1, 3$ (compact notation $\hat{\lambda}_j(1)\hat{\lambda}_k(3)$), etc. It is important to realize that our definition of subsystems guarantees that expectation values $\langle\hat{Q}\rangle$ of any such m-cluster-operator can be obtained from a series of respective coincidence experiments (ensemble- or time-averaged).

1.5. Can we Reconstruct any State?

The density matrix is uniquely specified once the expectation values of this full hierarchy of cluster-operators are given [3]. As the number of independent state parameters grows exponentially with N, complete state reconstruction [4] tends to be severely limited. So, N needs to be small, as incomplete measurements are likely to miss non-classical features all together. (This is why our world appears so classical).

1.6. What is a "Classical" State?

Within an entirely quantum mechanical treatment the notion of a "classical" state is somewhat ambiguous. For a definition one usually refers to a complete local basis as the eigenstates of a local operator (or set of those): For the two-level system studied here we use the local basis operators, $\hat{\lambda}_3(\mu)$. Products of their eigenstates, $|p_1p_2,\ldots,p_N\rangle$, $p_i = 1, 2,\ldots,n$, then define a complete basis set for the whole network, taken as our "classical" reference.

1.7. What are "Non-classical" States?

Superpositions of those classical states are now non-classical by definition: There are superpositions *within* a single subsystem only (local coherence, described here by the expectation values of $\hat{\lambda}_1(\mu), \hat{\lambda}_2(\mu)$) and superpositions involving *more than one* subsystem (entanglement, described by appropriate expectation values of $m \geq 2$ – cluster – operators). The fact that the subsystems are separable does not imply that the *state* of the network is separable into these constituents: The implies a kind of "quantum fuzziness" of the subsystems: Entangled states are states which, allowing for any local [5] unitary transformation, cannot be written as a simple product state. A useful entanglement measure should thus be invariant under such local transformations. (There is a generalization for mixed states).

1.8. How can we Quantify Entanglement?

Entanglement has been introduced as a property of a wavefunction not to factor into a product. This can be tested by means of appropriate

expectation values. The matrix-elements [3]

$$M_{ij\ldots}^{\mu\nu\ldots} = \langle(\hat{\lambda}_i(\mu) - \langle\hat{\lambda}_i(\mu)\rangle)(\hat{\lambda}_j(\nu) - \langle\hat{\lambda}_j(\nu)\rangle)\cdots\rangle \tag{2}$$

are of the form of covariances known from classical statistical physics; non-zero terms indicate deviations from the product state form [6]. Convenient measures, which are much less detailed but turn out to be invariant under any local unitary transformations, are (for $n = 2$)

$$\beta^{\mu\nu} = \frac{1}{3}\sum_{ij} M_{ij}^{\mu\nu} M_{ij}^{\mu\nu}, \quad \beta^{\mu\nu\sigma} = \frac{1}{4}\sum_{ijk} M_{ijk}^{\mu\nu\sigma} M_{ijk}^{\mu\nu\sigma} \tag{3}$$

These measures have upper bounds; here they are normalized to 1. For a network with $N > 2$ subsystems there is not just one but a whole hierarchy of such entanglement measures. In addition to these we also introduce a measure for local coherence; with the eigenstates of $\hat{\lambda}_3$ as the local reference ($n = 2$), a possible choice is

$$\alpha^{\mu} = \langle\hat{\lambda}_1(\mu)\rangle^2 + \langle\hat{\lambda}_2(\mu)\rangle^2 \tag{4}$$

Of course, α^{μ} cannot be invariant under local unitary transformations.

1.9. What is the Origin of Entanglement?

Perhaps the best known source of entanglement is the permutation symmetry requirement for indistinguishable particles. However, such Fermi (Bose-) states are not characterized by specific entanglement types. Eigenstates of global operators like total angular momentum, may also exhibit entanglement. A *dynamical* source of entanglement, finally, can be traced back to any two-subsystem interaction (like Coulomb interaction): The corresponding Hamiltonian generates a unitary time evolution which is non-local and will thus, in general, change the entanglement.

1.10. What are Observable Effects of Entanglement?

While coherence physics can do with appropriate local measurements ($m = 1$), entanglement shows up not before $m \geq 2$-cluster-operators are involved: Typically this applies to the total Hamiltonian itself; well-known effects are energy shifts and (exchange-) splittings for Fermi-systems.

More specifically, entanglement-physics relates to the study of a single network through a number of selective modes or contacts. We will discuss three qualitatively different examples based on the same type of model.

2. NUMERICAL SIMULATIONS

2.1. Outline of Model

The model we are considering could be an array of $N = 3$ quantum dots, here simplified as two-level subsystems. In order these to represent real subsystems, they need to be addressable separately. We assume they are spectroscopically different, i.e., they can be distinguished in frequency space and thus be separately driven by a classical external electromagnetic field (coupling strength g^{μ}, detuning δ^{μ}). Their mutual coupling, $C_R^{\mu\nu}$, is of the Ising-spin type. This coupling modifies the energy level spacing of one system depending on the state of its neighbors. The eigenstates of this interacting network are still product states, though. Local coherence is introduced by the external driving fields, which is then transformed into entanglement by means of the interaction. In this way there is an entanglement source that can be switched on and off [3].

2.2. Quantum Trajectories

In the first two examples all three subsystems "play the same role", i.e., all the parameters g, δ, C_F are independent of μ; they are chosen such that only the transitions from the ground state $|111\rangle$ to $|112\rangle$, $|121\rangle$ and $|211\rangle$ are in resonance. In addition

there are 3 local damping channels, again all with the same damping rate W. The dynamics is described by the appropriate master equation.

As we are interested in the detailed simulation of individual networks, the theoretical approach of choice is stochastic unraveling (see [2] and references therein): In this method the master equation is interpreted as a rule for a modified non-unitary continuous evolution of a pure state, interrupted by stochastic jumps to another pure state. Incoherent driving is described as a selective rate-process (transition rates $(g^{\mu})^2\pi/2$) implying $\alpha^{\mu} = \beta^{\mu\nu} = \beta^{123} = 0$ at all times ("classical case"). While the latter (Fig. 1a) is confined to the definite alternatives of being in the lower or upper state ("telegraph signal"), the trajectories in the non-classical domain (Fig. 1b) are strongly correlated (they practically coincide here) and show a quite different behavior: Between the jumps there is a continuous motion, as coherence and entanglement allow to explore additional regions of state space: Figure 2 shows the time-averaged entanglement measures as they depend on g/W.

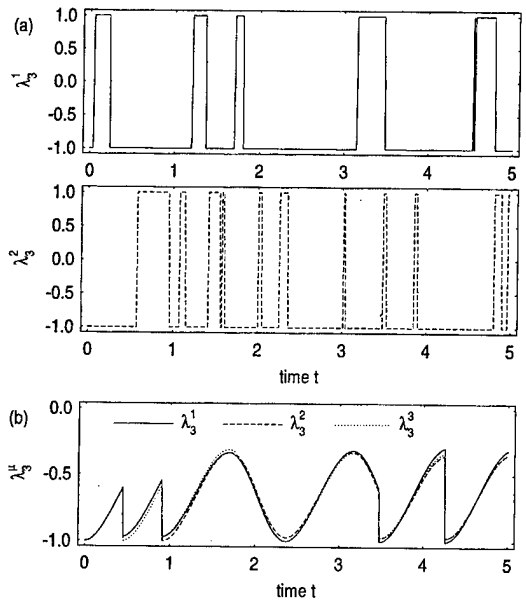


FIGURE 1 Quantum trajectories of a driven and damped $N = 3$-network ($g/W = 2.5$): (a) classical limit (incoherent pumping rate $= g^2\pi/2$), (b) non-classical case (coherent driving).

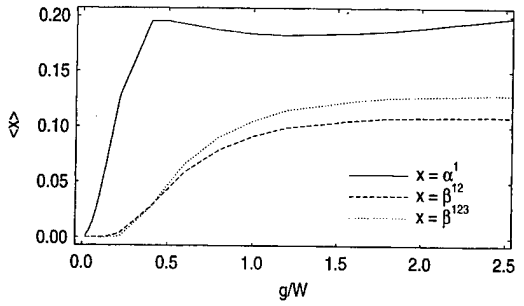


FIGURE 2 Averaged coherence-, α, and entanglement-measures, $\beta^{\mu\nu}$, β^{123} as a function of the coherent driving field strength, g, over the damping rate, W.

2.3. Quantum Noise

The quantum trajectories as such do not constitute a measurement, let alone a complete one. However, assuming the damping of each subsystem to be due to radiative decay only, the trajectories of the above model directly connect to specific luminescence signals: Each photon detection event in one of the distinguishable frequency channels (photon counts n^{μ} per sampling time T) is related to a jump in the trajectory. The simulated detection traces may then be statistically analyzed in terms of the counting covariances [7].

$$N^{\mu\nu\cdots} = \overline{(n^{\mu} - \overline{n^{\mu}})(n^{\nu} - \overline{n^{\nu}})\cdots} \qquad (5)$$

As shown in Figure 3, the noise properties, indeed, reflect the fact that the network state is (on average) non-classical and follows non-classical trajectories. The result for incoherent driving is

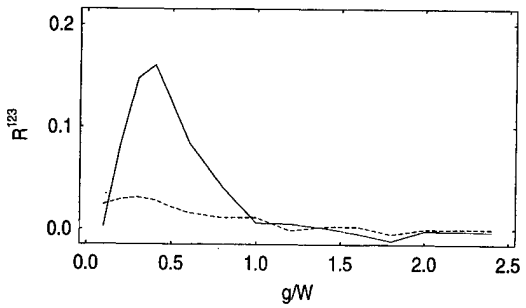


FIGURE 3 Photon-counting covariance of the luminescence signals: Shown is $R^{123} = N^{123}/(N^{11}N^{22}N^{33})^{1/2}$ with $N^{\mu\nu\cdots}$ as defined in Eq. (5). The sampling time is $T \gg W^{-1}$.

included for comparison (broken line); there is virtually no correlation N^{123} in this case.

2.4. Quantum Parallelsim

Contrary to the two preceding examples we turn now to a closed system, with the sub-systems playing different roles: These roles are implemented by means of pulsed changes of the parameters controlling the Hamilton model: Well-known examples are π-pulses to invert the occupation of 2-level systems. Interacting networks of the type studied here allow to implement conditioned π-pulses, i.e., the action of the light pulse on subsystem $\mu = 1$ is a π-pulse only, if the neighbor $\mu = 2$ is in state $|1\rangle$, say; no action otherwise [8]. A continuously driven system 2 in contact at time t_1 with such a "memory-gate" 1 (pulse duration is assumed negligible), subsequently at t_2 in contact with a gate 3 will generate entanglement as discussed before. However, the "logic" of these gate operations make system 2 emulate an ensemble that was actually measured at those two times (Fig. 4): System 2 represents in parallel all *potential trajectories* going through the two measurement steps! This quantum parallelism exploits entanglement in a surprising way. So-called quantum-algorithms would make sure that final measurements, which invariably pick an actual trajectory and thus deviate from ensemble behavior, indicate desired problem solutions [9]. Their reliance on large-N-quantum networks makes them a challenging target indeed for future developments.

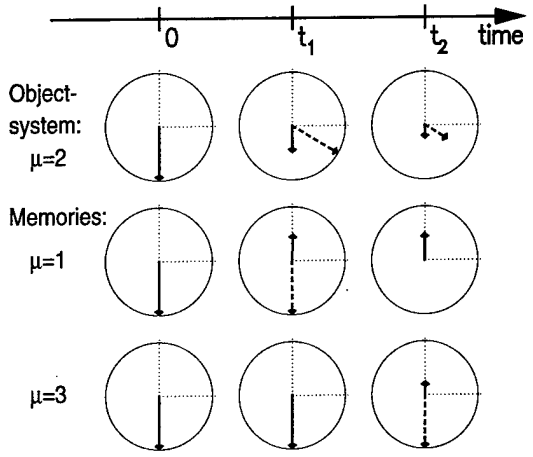


FIGURE 4 Single closed network: Subsystem-Bloch-vectors (in $\langle\hat{\lambda}_2\rangle$, $\langle\hat{\lambda}_3\rangle$-plane) at 3 instatnts of time. Subsystem $\mu = 2$ is continuously driven (rotates counter-clockwise) and is brought into contact with a "gate"-subsystem $\mu = 1$ at time t_1, and with a "gate" $\mu = 3$ at time t_2 (broken lines: before gate operation). Each gate is seen to "project" subsystem 2 on the $\langle\hat{\lambda}_3\rangle$-axis, as in an ensemble measurement. Note that a single system measurement would result in either up or down.

Acknowledgement

Financial support by the Deutsche Forschungsgemeinschaft and by the Oregon Center of Optics is gratefully acknowledged. One of us (G.M.) thanks Howard Carmichael and Michael Raymer for valuable discussions.

References

[1] See, e.g., document from U.S. semiconductor industry, http://www.sematech.org/public/roadmap/doc.
[2] Keller, M. and Mahler, G. (1994). "Nanostructures, entanglement, and the physics of quantum control", *J. Mod. Optics*, **41**, 2537; Mahler, G. and Weberruß, V. A. (1995). "*Quantum Networks: Dynamics of Open Nanostructures*", Springer, Berlin.
[3] Mahler, G., Keller, M. and Wawer, R. (1997). "Quantum networks: master equation and local measurements", *Z. Phys. B*, **104**, 153.
[4] Leonhardt, U., Munroe, M., Kiss, T., Richter, Th. and Raymer, M. G. (1996). "Sampling of photon statistics and density matrix using homodyne detection", *Optics Commun.*, **127**, 144.
[5] A *local unitary transformation* can be written as a product of transformations on individual subsystems.
[6] cf. also the so-called *Bell-inequalities*: Bell, J. S. (1964). *Physics (N.Y)*, **1**, 195.
[7] Mahler, G. and Wawer, R. (1997). "Quantum dot arrays: preparation and detection of non-classical states", *Superlattics and Microstructures*, **21**, 7.
[8] See e.g., Barenco, A., Deutsch, D. and Ekert, A. (1995). "Conditional quantum dynamics and logic gates", *Phys. Rev. Letters*, **74**, 4083.
[9] Ekert, A. and Jozsa, R. (1996). *Rev. mod. Physics*, **68**, 733.

Authors' Biography

Prof. Dr. Günter Mahler studied physics at the Universities of Frankfurt, Munich, and Regens-

burg (Germany). He received his Ph.D. in 1972 and was appointed Professor for Theoretical Physics by the University of Stuttgart in 1978. He has been visiting Professor with the University of Strasbourg, France, the Arizona State University, the Santa Fe Institute, and the University of Oregon, USA. His main fields of interest concern semiconductor transport, molecular electronics, quantum stochastics, quantum information, and quantum optics.

Rainer Wawer studied physics at the University of Stuttgart, where he received his Diploma in Theoretical Physics. He is presently working in the field of quantum measurement models and quantum trajectories in partial fulfillment of Ph.D. requirements.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 197–202
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

A Generalized Monte Carlo Approach for the Analysis of Quantum-Transport Phenomena in Mesoscopic Systems: Interplay Between Coherence and Relaxation

FAUSTO ROSSI[a,*], STEFANO RAGAZZI[b], ALDO DI CARLO[b] and PAOLO LUGLI[b]

[a] *Istituto Nazionale Fisica della Materia (INFM) and Dipartimento di Fisica, Universitá di Modena, I-41100 Modena, Italy;*
[b] *INFM and Dipartimento di Ingegneria Elettronica, Università di Roma "Tor Vergata", I-00133 Roma, Italy*

A theoretical investigation of quantum-transport phenomena in mesoscopic systems is presented. In particular, a generalization to "open systems" of the well-known Semiconductor Bloch equations is proposed. Compared to the conventional Bloch theory, the presence of spatial boundary conditions manifest itself through self-energy corrections and additional source terms in the kinetic equations, which are solved by means of a generalized Monte Carlo simulation.The proposed numerical approach is applied to the study of the scattering-induced suppression of Bloch oscillations in semiconductor superlattices as well as to the analysis of quantum-transport phenomena in double-barrier structures.

Keywords: Quantum, transport, simulation, mesoscopic, coherence, dissipation

1. INTRODUCTION

The Monte Carlo (MC) method, which has been applied for more than 25 years for calculation of semiclassical charge transport in semiconductors, is the most powerful numerical tool for micro-electronics device simulation [1]. However, it is well known that present-day technology pushes device dimensions toward limits where the traditional semiclassical transport theory can no longer be applied and a more rigorous quantum transport theory is required [2].

In this paper, a generalized MC approach for the analysis of hot-carrier transport and relaxation phenomena in quantum devices is proposed. The method is based on a generalized MC solution of the set of kinetic equations governing the time evolution of the single-particle density matrix. Our approach can be regarded as an extension to open systems of the generalized MC method recently proposed for the analysis of the coupled coherent and incoherent carrier dynamics in photoexcited semiconductors [3]. Compared to more academic quantum-kinetic approaches [4] – whose applica-

* Corresponding author.

tion is often limited to highly simplified physical models and conditions–, the proposed simulation scheme can be applied to realistic cases, allowing on the one hand a proper description of quantum-interference phenomena induced by the potential profile and on the other hand maintaining all the well known advantages of the conventional MC method.

As a first application, we will discuss the scattering-induced suppression of Bloch oscillations in semiconductor superlattices. As a second example, we will study the strong interplay between coherence and relaxation within a double-barrier structure.

2. THEORETICAL APPROACH

In order to properly describe carrier-transport phenomena in mesoscopic structures, let us consider a generic electron-phonon system, whose Hamiltonian can be schematically written as:

$$\boldsymbol{H} = (\boldsymbol{H}_c + \boldsymbol{H}_p) + (\boldsymbol{H}_{cc} + \boldsymbol{H}_{cp}) = \boldsymbol{H}_0 + \boldsymbol{H}'. \quad (1)$$

The single-particle term $\boldsymbol{H}_0$ is the sum of the noninteracting carrier and phonon Hamiltonians $\boldsymbol{H}_c$ and $\boldsymbol{H}_p$ while the many-body contribution $\boldsymbol{H}'$ includes carrier-carrier as well as carrier-phonon interactions. More specifically, the single-particle Hamiltonian $\boldsymbol{H}_c$ describes the noninteracting carrier system within the potential profile of our mesoscopic structure (including possible external fields). By denoting with $\phi_\alpha(\boldsymbol{r}) = \langle \boldsymbol{r}|\alpha\rangle$ the wavefunction of the single-particle state α and with ε_α the corresponding energy, the noninteracting-carrier Hamiltonian reads:

$$\boldsymbol{H}_c = \sum_\alpha \varepsilon_\alpha a_\alpha^\dagger a_\alpha. \quad (2)$$

Here, the usual second-quantization picture in terms of creation ($a_\alpha^\dagger$) and destruction (a_α) operators has been employed.

The basic quantity in our theoretical approach is the single-particle density matrix

$$\rho_{\alpha\beta} = \langle a_\beta^\dagger a_\alpha \rangle. \quad (3)$$

Its diagonal elements correspond to the usual distribution functions of the semiclassical Boltzmann theory ($f_\alpha = \rho_{\alpha\alpha}$) while the off-diagonal terms ($\alpha \neq \beta$) describe the degree of quantum-mechanical phase coherence between states α and β. Starting from the Heisenberg equations of motion for the operators a_α it is possible to derive a set of equations of motion for the density-matrix elements $\rho_{\alpha\beta}$, called semiconductor Bloch equations (SBE) [5], whose general structure is given by:

$$\frac{d}{dt}\rho_{\alpha\beta} = \frac{d}{dt}\rho_{\alpha\beta}\bigg|_{\boldsymbol{H}_0} + \frac{d}{dt}\rho_{\alpha\beta}\bigg|_{\boldsymbol{H}'}. \quad (4)$$

The time evolution induced by the single-particle Hamiltonian $\boldsymbol{H}_0$ can be evaluated exactly yielding:

$$\frac{d}{dt}\rho_{\alpha\beta}\bigg|_{\boldsymbol{H}_0} = \frac{1}{i\hbar}\sum_{\alpha'\beta'} \mathcal{E}_{\alpha\beta,\alpha'\beta'}\rho_{\alpha'\beta'} = -i\omega_{\alpha\beta}\rho_{\alpha\beta}. \quad (5)$$

with $\omega_{\alpha\beta} = (\varepsilon_\alpha - \varepsilon_\beta)/\hbar$. On the contrary, the time evolution due to the many-body Hamiltonian $\boldsymbol{H}'$ involves phonon-assisted as well as higher-order density-matrices; Thus, approximations are needed in order to "close" our set of equations of motion (with respect to our kinetic variables). In particular, the "mean-field" approximation together with the Markov limit approximation allow us to derive a set of closed equations of motion which is still local in time:

$$\frac{d}{dt}\rho_{\alpha\beta}\bigg|_{\boldsymbol{H}'} = \sum_{\alpha'\beta'} \Gamma'_{\alpha\beta,\alpha'\beta'}\rho_{\alpha'\beta'}. \quad (6)$$

The matrix $\Gamma'_{\alpha\beta,\alpha'\beta'}$ is, in general, a complex quantity: its real part describes a sort of generalized scattering rate while its imaginary part is related to energy-renormalization phenomena. Recently, a full quantum-mechanical approach

has been proposed [6], which overcomes the usual Markov limit in describing the carrier-phonon interaction. However, due to the huge amount of CPU time required, its applicability is still limited to short time-scales and extremely simplified situations.

The analysis presented so far is typical of a so-called "closed" system, i.e., a physical system defined over the whole coordinate space. However, this is not the case of interest for the analysis of quantum-transport phenomena in mesoscopic devices, where the properties of the carrier subsystem are strongly influenced by spatial boundary conditions. In order to better understand this crucial point, let us reconsider our theoretical scheme in terms of real-space coordinates.The proper quantum-mechanical description is then based on the so-called Wigner function [7], given by the following Weyl-Wigner transformation of the single-particle density matrix:

$$f^W(\boldsymbol{r},\boldsymbol{k}) = \sum_{\alpha\beta} \rho_{\alpha\beta} u_{\alpha\beta}(\boldsymbol{r},\boldsymbol{k}) \tag{7}$$

with

$$u_{\alpha\beta}(\boldsymbol{r},\boldsymbol{k}) = (2\pi)^{-\frac{3}{2}} \int d\boldsymbol{r}' \phi_\alpha\left(\boldsymbol{r}+\frac{1}{2}\boldsymbol{r}'\right) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}'} \times \times \phi^*_\beta\left(\boldsymbol{r}-\frac{1}{2}\boldsymbol{r}'\right). \tag{8}$$

By applying the above Weyl-Wigner transformation to the non-interacting SBE (5), we obtain:

$$\left.\frac{d}{dt} f^W(\boldsymbol{r},\boldsymbol{k},t)\right|_{H_0} = \int d\boldsymbol{r}' d\boldsymbol{k}' L_0(\boldsymbol{r},\boldsymbol{k};\boldsymbol{r}',\boldsymbol{k}') \times \times f^W(\boldsymbol{r}',\boldsymbol{k}',t) \tag{9}$$

where

$$L_0(\boldsymbol{r},\boldsymbol{k};\boldsymbol{r}',\boldsymbol{k}') = -i \sum_{\alpha\beta} u_{\alpha\beta}(\boldsymbol{r},\boldsymbol{k}) \omega_{\alpha\beta} u^*_{\alpha\beta}(\boldsymbol{r}',\boldsymbol{k}') \tag{10}$$

is the single-particle Liouville operator in the $\boldsymbol{r}$, $\boldsymbol{k}$ representation.

For the case of a closed system, the Wigner function f^W is defined for any value of the real-space coordinate $\boldsymbol{r}$ and its time evolution is fully determined by its initial condition.

On the contrary, for the case of an open system, f^W is defined only within a given region Ω of interest and its time evolution is determined by the initial condition f_0^W inside such region as well as by its values on the boundary r_b of the domain Ω at any time $t' > t_0$. More specifically, by applying the Green's function theory [8] to Eq. (9) we have:

$$f^W(\boldsymbol{r},\boldsymbol{k};t) = \int_\Omega d\boldsymbol{r}' \int d\boldsymbol{k}' G_0(\boldsymbol{r},\boldsymbol{k};\boldsymbol{r}',\boldsymbol{k}';t-t_0) \times \times f^W(\boldsymbol{r}',\boldsymbol{k}';t_0) + + \int d\boldsymbol{r}_b \int d\boldsymbol{k}' \int_{t_0}^t dt' G_0(\boldsymbol{r},\boldsymbol{k};\boldsymbol{r}_b,\boldsymbol{k}';\ t-t') \times \times f_b^W(\boldsymbol{r}_b,\boldsymbol{k}',t') v(\boldsymbol{k}'), \tag{11}$$

where

$$G_0(\boldsymbol{r},\boldsymbol{k};\boldsymbol{r}',\boldsymbol{k}';\tau) = \sum_{\alpha\beta} u_{\alpha\beta}(\boldsymbol{r},\boldsymbol{k}) e^{-i\omega_{\alpha\beta}\tau} u^*_{\alpha\beta}(\boldsymbol{r}'\boldsymbol{k}') \tag{12}$$

is the single-particle evolution operator, while $v(\boldsymbol{k})$ is the component of the carrier group velocity normal to the boundary surface. We clearly see that the value of f^W is obtained from the single-particle propagation of the initial condition f_0^W inside the domain Ω plus the propagation of the boundary values f_b^W from the points of the surface at any time t' to the point $\boldsymbol{r}$, $\boldsymbol{k}$ of interest.

Let us now come back to the density-matrix formulation by introducing the inverse Weyl-Wigner transform:

$$\bar{\rho}_{\alpha\beta} = \int_\Omega d\boldsymbol{r} \int d\boldsymbol{k} u^*_{\alpha\beta}(\boldsymbol{r},\boldsymbol{k}) f^W(\boldsymbol{r},\boldsymbol{k}). \tag{13}$$

The above density matrix $\bar{\rho}$ provides a description of our open system equivalent to the Wigner-function one. By applying the above transforma-

tion to Eq. (11) and then performing its time derivative, we finally obtain:

$$\left.\frac{d}{dt}\bar{\rho}_{\alpha\beta}\right|_{H_0} = \frac{1}{i\hbar}\sum_{\alpha'\beta'}\bar{\mathcal{E}}_{\alpha\beta,\alpha'\beta'}\bar{\rho}_{\alpha'\beta'} + \bar{S}_{\alpha\beta}, \qquad (14)$$

where

$$\bar{\mathcal{E}}_{\alpha\beta,\alpha'\beta'} = \left[U\mathcal{E}U^{-1}\right]_{\alpha\beta,\alpha'\beta'} \qquad (15)$$

is the single-particle self-energy tensor in Eq. (5) "dressed" by the transformation

$$U_{\alpha\beta,\alpha'\beta'} = \int_\Omega d\boldsymbol{r}\int d\boldsymbol{k}u^*_{\alpha\beta}(\boldsymbol{r},\boldsymbol{k})u_{\alpha'\beta'}(\boldsymbol{r},\boldsymbol{k}), \qquad (16)$$

while

$$\bar{S}_{\alpha\beta} = \sum_{\alpha'\beta'}U_{\alpha\beta,\alpha'\beta'}\int d\boldsymbol{r}_b\int d\boldsymbol{k}u^*_{\alpha'\beta'}(\boldsymbol{r}_b,\boldsymbol{k})v(\boldsymbol{k})\times \times f_b^W(\boldsymbol{r}_b,\boldsymbol{k}) \qquad (17)$$

is a source term induced by our spatial boundary conditions.

Eq. (14) is the desired generalization to the case of open systems of the standard single-particle SBE in Eq. (5). In addition to the source term (17) (whose explicit form depends on the particular values of f_b^W on the surface of our domain), the presence of spatial boundary conditions induces modifications on the self energy $\mathcal{E}$ of the system via the transformation U in Eq. (16).

3. SOME SIMULATED EXPERIMENTS

The theoretical approach presented so far is the starting point of our MC simulation. The generalized SBE (4) are solved by means of the same MC simulation scheme described in [3]. The method is based on a time-step separation between coherent and incoherent dynamics (see Eq. (4)). The former accounts in a rigorous way for all quantum phenomena induced by the potential profile of the device as well as for the proper boundary conditions (see Eq. (14)). The latter, described within the basis given by the eigenstates α of the potential profile (see Eq. (6)), accounts for all the relevant scattering mechanisms by means of a conventional "ensemble" MC simulation [1].

The above numerical approach has been applied to the study of the scattering-induced suppression of Bloch oscillations in semiconductor superlattices and to the analysis of quantum-transport phenomena in a double-barrier structure.

In the first case, a biased GaAs/AlGaAs multi-quantum-well (MQW) has been considered (15 periods of a 20 Å barrier and a 100 Å well with a barrier height of 0.5 eV). The electric field in the MQW region is about 12 kV/cm. The ultrafast dynamics of an injected wavepacket (injection energy $E = 50$ meV) has been simulated. Figure 1 shows the position mean value of the wavepacket as a function of time. For the scattering-free case, the expected Bloch oscillations in real space are well reproduced. They originate from the coherent motion of our wavepacket through the superlattice structure. In the presence of scattering mechanisms, such oscillations are strongly damped. This is mainly due to carrier-LO phonon scattering, whose typical time-scale is comparable with the

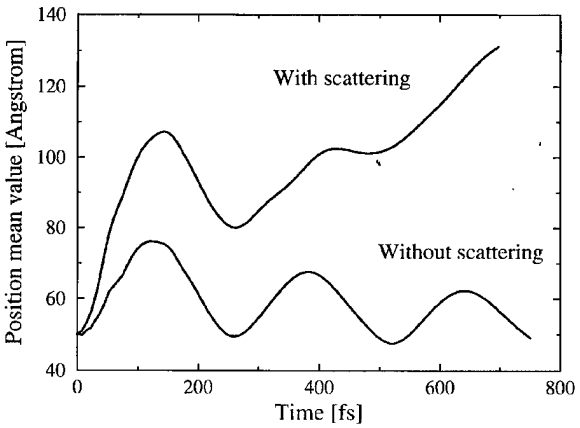


FIGURE 1 Position mean value of an electronic wavepacket within a biased semiconductor super lattice as a function of time with and without scattering (see text).

sub-picosecond time-scale of our Bloch oscillations.

We have than carried out the simulation of an electronic wavepacket entering (from left) a double-barrier structure. The injected carriers are energetically in resonance with the second level of the well. In Figure 2 the charge distribution across the double-barrier structure is shown at different times with and without scattering. As expected, for the scattering-free case (Fig. 2b) we obtain the well-established resonance scenario: the wavepacket enters in resonance with the second energy level of the structure and after some time-delay a part of it is transmitted and a part is reflected. On the contrary, in the presence of scattering events (Fig. 2a), the resonance process is strongly influenced by a charge transfer from the second to the first energy level. This effect is clearly evidenced by the charge distribution inside the well region, which for the free-scattering case exhibits the typical second-level symmetry while in the presence of scattering it approaches the symmetry of the ground level.

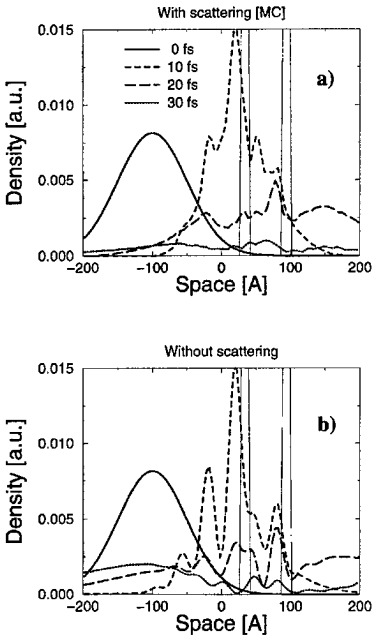


FIGURE 2 Charge distribution across a double-barrier structure at different times corresponding to an electron wavepacket injected in resonance with the second level (see text). The scattering rates have been artificially increased by a factor 10 with respect to the GaAs values in order to emphasize the relaxation to the first level.

4. CONCLUSIONS

In this paper, we have presented a generalization to open systems of the well-known semiconductor Bloch equations. The proposed approach, based on the density-matrix formalism, allows a proper description of the strong coupling between coherent and incoherent dynamics in mesoscopic systems. These generalized SBE are numerically solved by means of the same MC simulation scheme used already for the study of ultrafast carrier dynamics in photoexcited semiconductors [3]. Our simulated experiments clearly show the failure of any pure coherent or incoherent approach in describing such quantum-transport regime.

Acknowledgements

We are grateful to Carlo Jacoboni, Tilmann Kuhn and Elisa Molinari for stimulating and fruitful discussions. This work was supported in part by the EC Commission through the Network "ULTRAFAST". One of us (ADC) acknowledges TRACS support.

References

[1] Jacoboni, C. and Lugli, P. (1989). *The Monte Carlo Method for Semiconductor Device Simulations* (Springer, Wien).

[2] Rossi, F., Brunetti, R. and Jacoboni, C. (1992). in *Hot Carriers in Semiconductor Nanostructures: Physics and Applications*, edited by J. Shah (Academic Press Inc., Boston), p. 153.

[3] Kuhn, T. and Rossi, F. (1992). *Phys. Rev. Lett.*, **69**, 977; Rossi, F., Haas, S. and Kuhn, T. (1994). *Phys. Rev. Lett.*, **72**, 152; Haas, S., Rossi, F. and Kuhn, T. (1996). *Phys. Rev. B*, **53**, 12855.

[4] Mclennan, M. J., Lee, Y. and Datta, S. (1991). *Phys. Rev. B*, **43**, 13846; Kluksdahl, N. C., Kriman, A. M., Ferry, D. K., Ringhofer, C. (1989). *Phys. Rev. B*, **39**, 7720; Biegel, B. A. *et al.* (1996). *Phys. Rev. B*, **54**, 8070.

[5] Haug, H. and Koch, S. W. (1994). *Quantum Theory of the Optical and Electronic Properties of Semiconductors*, 3rd Edn. (World Scientific, Singapore).

[6] Brunetti, R., Jacoboni, C. and Price, P. J. (1994). *Phys. Rev. B*, **50**, 11872
[7] Frensley, W. (1990). *Rev. Mod. Phys.*, **62**, 3.
[8] Morse, P. M. and Feshbach, H. (1953). *Methods of theoretical physics*, New York: McGraw-Hill.

Author's Biographies

Fausto Rossi born in 1962 in Carpi (Italy), Ph.D. in Physics in 1993 from Modena University. He is currently part of the INFM research unit of the Modena University. His research activity is mainly devoted to the analysis of non-equilibrium phenomena in semiconductors, namely ultrafast dynamics of photoexcited carriers and quantum-transport simulation.

Stefano Ragazzi graduated in Electronic Engineerings in 1997 at the El. Eng. Dept. on the University of Rome "Tor Vergata", he is currently enrolled in the Italian Navy.

Aldo Di Carlo graduated in Physics at the University of Roma, Italy, in 1991, obtained the Ph.D. degree in Physics in 1995 at the Walter Schottky Institute of the Technical University of Munich (Germany). He is currently researcher at the El. Eng. Dept. of the University of Roma, "Tor Vergata", where his work is mainly focused on the theoretical study of optical and transport processes in semiconductor nanostructures, devices and polymers.

Paolo Lugli born in Carpi (Italy) in 1956, obtained his Laurea in Physics at the University of Modena in 1979 and his Ph.D. in El. Eg. at Colorado State University in 1985. He is currently Full Professor of Optoelectronics at the University of Roma "Tor Vergata". His main field of activity is the theoretical study and numerical simulation of semiconductor nanostructures and devices.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 203–207
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Coherent Control of Light Absorption and Carrier Dynamics in Semiconductor Nanostructures

WALTER PÖTZ* and XUEDONG HU

University of Illinois at Chicago, Physics Department, Chicago, IL 60607, USA

We present two examples of coherent control of inter(sub)band transitions in a semiconductor double well by coherent light sources. Accounting for the upper hole subband and two lowest electron subbands, a microscopic theoretical analysis shows that electron-hole pair generation by a sub-picosecond pump pulse can be controlled by the intensity and the phase of a dc microwave field which resonantly couples the two electron subbands. Light absorption can be either enhanced or reduced. Secondly, it is shown that proper combination of two *pulsed* laser fields allows control of electron inter(sub)band transitions and final-state population, i.e., the formation of indirect versus direct excitons.

Keywords: Coherent control, semiconductor, quantum well, optics, theory

Coherent control of final-state population and chemical reactions has long been pursued in atomic and molecular physics [1–5]. More recently, improvements in ultrafast spectroscopy have allowed induction and observation of coherent phenomena in semiconductors, in form of coherent charge oscillations in double wells, Bloch oscillations, and coherent control of photocurrent, to name some of the highlights of progress in recent years [6–10]. In this paper we investigate coherent control of inter(sub)band transitions, absorption, and final-state population theoretically. Specifically we discuss a scheme which allows *phase controlled* light absorption and control of final-state population in semiconductor heterostructures by the interplay of two coherent light sources. As a specific example we give results for an asymmetric 145 Å/25 Å/100 Å GaAs-AlGaAs-GaAs double well whose electron subband splitting is controlled by a static electric field and which is exposed to a tunable sub-picosecond pump pulse which generates electron-hole pairs (excitons) across its main energy gap. Calculations are done within the framework of a microscopic theory in form of Boltzmann-Bloch equations which account for the carrier–carrier Coulomb interaction [11]. In the present study, subband splittings and time scales are such that LO phonon

* Corresponding author.

scattering may be neglected. Light pulses are of Gaussian shape and the light-matter coupling is treated for a classical light field *including* the counter-rotating part. The peak of the pump pulse arrives at time zero, relative to which we define the phase ϕ of the microwave (MW) field $\vec{E}_{\mathrm{MW}}(t) = \vec{E}_0(t)\cos(\omega_{\mathrm{MW}}t + \phi)$. A comparison to the rotating-wave approximation (RWA), which has been used in an earlier study, is made [12].

When a coherent dc MW field resonantly couples the two electron subbands, the system undergoes Rabi oscillations between its uncoupled eigenstates when originally prepared in one of the latter. In the situation depicted in Figure 1, the dipole matrix element between hole $|1\rangle$ and upper electron subband $|2\rangle$ is much stronger than hole and lower electron subband $|3\rangle$. Hence, if a pump pulse is applied at resonance with the direct exciton maximum absorption is obtained. However, application of a MW field reduces the admixture of left-well eigenfunction $|W\rangle$ in $|2\rangle$ and transfers it to $|3\rangle$. Hence, the MW field reduces absorption at the direct exciton peak [13]. Conversely, if the pump pulse is tuned near the indirect exciton, application of a MW field enhances absorption. The *phase* of the coherent dc MW field enters the coupling between the electron subbands and hence the complex electron interband polarization. When the pump pulse duration is shorter than the inverse of the MW-induced Rabi frequency, the phase influences the absorption process from a third level, here, the top hole subband $|1\rangle$. This is demonstrated in Figure 2 for the case where a 80 fs pump pulse of Gaussian profile is tuned near resonance with the indirect excitons of the DW with a subband splitting of 20 meV. The MW intensity is about 2 MWcm^{-2}, corresponding to a Rabi period near 100 fs. Clearly, the presence of the MW field increases absorption by about 15 percent for phase $\pi/2$. This effect is more pronounced at the direct exciton peak [13] The RWA underestimates the importance of second harmonics in the carrier dynamics induced by the MW field. Therefore, shorter pump pulses than predicted by within the RWA may be required to display the phase dependence of absorption.

For the present three-level system and within the RWA for transitions *between conduction and valence bands*, the pump pulse photon Boltzmann equation in the presence of coherence in the carrier

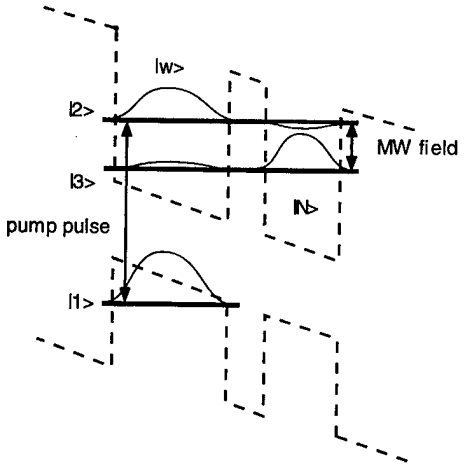


FIGURE 1 Illustration of a biased asymmetric GaAlAs-GaAs double well. $|1\rangle$, $|2\rangle$ and $|3\rangle$ denote double-well eigenstates; $|W\rangle$ and $|N\rangle$ are the lowest wide-well and narrow-well eigenstates, respectively. The shape of wave functions is indicated by thin solid lines on top of the corresponding energy levels.

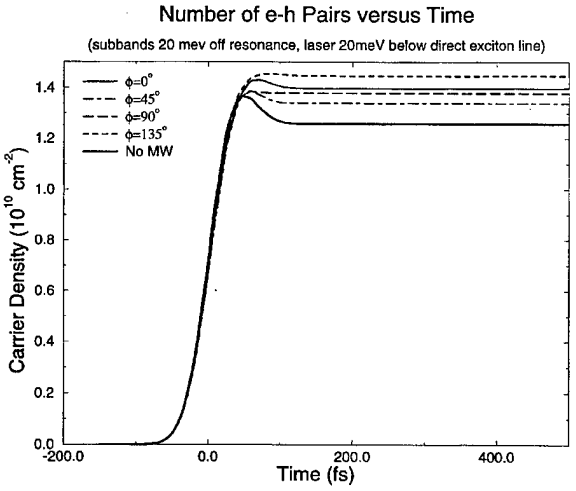


FIGURE 2 Density of photo-generated electron-hole pairs as a function of time and MW phase Φ for a pump pulse photon energy which is 20 meV below the direct exciton peak.

system is of the structure [11].

$$\frac{dN(q,t)}{dt} = \frac{2\pi}{\hbar^2} \sum_{k,\alpha,\alpha'=2,3} M_{1\alpha'}(q) M^*_{1\alpha}(q) \delta_E \{f_{\alpha\alpha'}(k,t)(1-f_{11}(k,t))(N(q,t)+1) - f_{11}(k,t)(1-f_{\alpha\alpha'}(k,t))N(q,t)\},$$

$$\approx \frac{2\pi}{\hbar^2} N(q,t) \sum_{k,\alpha,\alpha'=2,3} M_{H\alpha'}(q) M^*_{H\alpha}(q) \delta_E \{f_{\alpha\alpha'}(k,t) - f_{11}(k,t)\}, \text{ for } N(q,t) \gg 1,$$

where $f_{\alpha\alpha'}(k, t)$ are carrier distribution functions, for $\alpha = \alpha'$, and interband polarizations, for $\alpha \neq \alpha'$. $N(q, t)$ is the photon occupation number and $M_{\alpha\alpha'}(q)$ is the matrix element for coupling subband α to subband α' via photon q. δ_E is the appropriate energy-conserving delta function. This equation shows that electron interband polarization influences photon absorption. Figure 3, in turn, shows how the phase of the MW field determines the sign of the real part of the electron interband polarization during the presence of the pump pulse to allow coherent control of the absorption process.

A possible scheme to control the final-state population in semiconductors resembles previous work on atoms for which adiabatic switching has been demonstrated to allow transitions between molecular levels which are dipole-forbidden by means of optical coupling to a third level [4, 5]. In this scheme, which is sketched in Figure 1, two temporally and spatially overlapping light pulses couple three levels of the system. The initial state of the system is $|1\rangle$, the desired final state is $|3\rangle$, and the intermediate state is $|2\rangle$. If the light pulse coupling $|2\rangle$ and $|3\rangle$, with amplitude $a(t)$ and duration T_{MW}, arrives and ends before arrival and, respectively, end of the pulse coupling $|1\rangle$ and $|2\rangle$, with amplitude $b(t)$ and duration T_p, and the pulse amplitudes change sufficiently slowly (relative to the characteristic frequencies of the driven three-level system), stimulated Raman adiabatic passage (STIRAP) occurs and the system undergoes an adiabatic transition from level $|1\rangle$ to level $|3\rangle$. [4] The condition for an adiabatic transition for $T = T_p = T_{\mathrm{MW}}$ is

$$T \cdot \left[\frac{\delta}{2} \pm \sqrt{\frac{\delta^2}{4} + |a|^2 + |b|^2}\right] \gg \hbar,$$

where δ is the light field detuning.

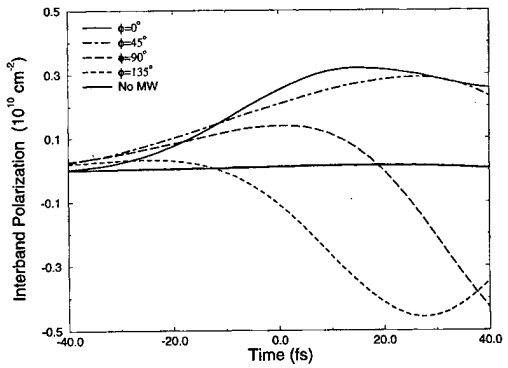


FIGURE 3 Real part of $|W\rangle - |N\rangle$ interband polarization as a function of time and MW phase.

However, in semiconductors resonant coupling between subbands calls for MW pulses, which may be difficult to generate at sufficiently high intensity. Moreover, due to characteristic decoherencing times for (free) carriers in semiconductors passage times can not be much longer than a picosecond. Therefore, the STIRAP process may be impractical, if not impossible, to be adopted to semiconductor nanostructures. Hence, we have investigated sub-picosecond *nonadiabiatic* transfer based on the three-subband scheme in Figure 1, where subband $|1\rangle$ is the top hole subband of a heterostructure, $|2\rangle$ and $|3\rangle$ are the two lowest electron subbands separated by 25 meV, in the present case. A 100fs subpicosecond pump pulse is used to generate electron-hole pairs (direct excitons) associated with levels $|1\rangle$ and $|2\rangle$. Its duration must be *shorter* than the inverse Rabi frequency between subband $|2\rangle$ and $|3\rangle$ which is established by a concurrent MW pulse. For excitation densities of about 10^{10} carriers per cm^2 this ensures MW-induced charge oscillations between subband $|2\rangle$ and $|3\rangle$. Intensity and

duration of the MW field are adjusted so that the latter permits one half of a Rabi oscillation between $|2\rangle$ and $|3\rangle$, such that the electrons get trapped in state $|3\rangle$ ($\approx |N\rangle$) after a single tunneling process. Here, the MW pulse resonantly couples the two electron subbands ($\delta = 0$) and has a duration of $T_{MW} = 320$ fs, corresponding to about two MW periods, and a peak intensity of about 1 MW cm^{-2}. Zero time delay between the two pulses and phase π of the MW field relative to the peak of the pump pulse were found to give best results.

In Figure 4 we show the number of electrons in the left well versus the number of electrons in the right well, and the total number of carriers (holes). Calculations including the counter-rotating field contributions, thick lines, are compared to those within the RWA, thin lines. It is clearly evident that the presence of the MW field reverses the tendency for direct (solid lines) versus indirect exciton (dot-dashed lines) formation. Without MW field and owing to the shape of wave functions, predominantly direct excitons are formed initially. Figure 4 also shows that the RWA gives almost quantitatively correct results for the final state population in the present case. However, it predicts simple (damped) harmonic charge oscillations, whereas the full calculation shows a more complicated dynamics.

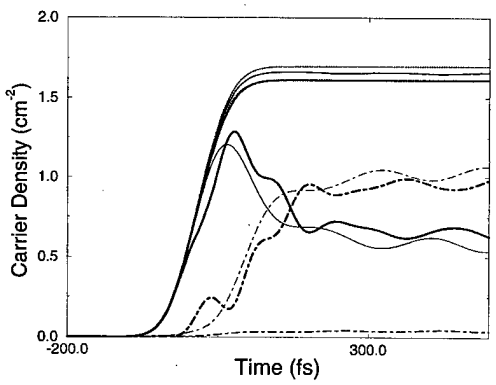


FIGURE 4 Carrier densities in the double well versus time. Solid lines: wide well (direct excitons); dot-dashed lines: narrow well (indirect excitons); dashed lines: holes (total number of excitons); thick lines: with microwave field (phase π) regular lines: no MW field; thin lines: MW field in RWA approximation.

In summary, we have given theoretical results which indicate that coherent control of intersubband transitions on a subpicosecond time scale is possible by means of coherent light sources. In particular, we have investigated coherent control of light absorption and final-state population in semiconductor double wells. It should be pointed out that the latter should also be achievable by interference between single and triple photon absorption, in analogy to experiments on molecules [3], and coherent current control via interference between single and two-photon absorption [8]. Details of these findings will be published elsewhere [12]. The rotating-wave approximation, which allows analytic solution of the coupled three-level system, is found to give nearly quantitatively correct results for the non-adiabatic final-state-control processes. However, for the coherent control process of absorption discussed here, it merely gives qualitatively correct results. Specifically, it fails to give an accurate account of charge oscillations induced by the MW field and, consequently, the correct value for the phase of optimal coupling.

Acknowledgements

We thank Prof. R. J. Gordon and Prof. W. A. Schroeder for helpful discussions. This work has been supported by the US Army Research Office.

References

[1] Tannor, D. J. and Rice, S. A. (1988). *Adv. Chem. Phys.*, **70**, 441.
[2] Brumer, P. and Shapiro, M. (1989). *Accts. Chem. Res.*, **22**, 407.
[3] Lu, S., Park, S. M., Xie, Y. and Gordon, R. J. (1992). *J. Chem. Phys.*, **96**, 6613.
[4] Oreg, J., Hioe, F. T. and Eberly, J. H. (1994). *Phys. Rev. A*, **29**, 690.
[5] Gaubatz, U., Rudecki, P., Schiemann, S. and Bergmann, K. (1990). *J. Chem. Phys.*, **92**, 5363; Bergmann, K. and Shore, B. W. (1995). in *Molecular Dynamics and Spectroscopy by Stimulated Emission Pumping*, Eds. H.-L. Dai and R. W. Field (World Scientific, Singapore), p. 315.
[6] Heberle, A. P., Baumberg, J. J. and Köhler, K. (1995). *Phys. Rev. Lett.*, **75**, 2598.

[7] Citrin, D. S. (1996). *Phys. Rev. Lett.*, **77**, 4596.
[8] Hachè, A., Kostoulas, Y., Atanasov, R., Hughes, J. L. P., Sipe, J. E. and van Driel, H. M. (1997). *Phys. Rev. Lett.*, **78**, 306.
[9] Leo, K., Shah, J., Göbel, E. O., Damen, T. C., Schmitt-Rink, S., Schäfer, W. and Köhler, K. (1991). *Phys. Rev. Lett.*, **66**, 201.
[10] Waschke, C., Roskos, H. G., Schwedler, R., Leo, K., Kurz, H. and Köhler, K. (1993). *Phys. Rev. Lett.*, **70**, 3319.
[11] Pötz, W. (1996). *Phys. Rev. B*, **54**, 5647; Pötz, W. (1996). *Appl. Phys. Lett.*, **68**, 2553; Pötz, W. and Hohenester, U., unpublished.
[12] Pötz, W. (1997). *Appl. Phys. Lett.*, **71**, 395.
[13] Pötz, W. (1997). *Phys. Rev. Lett.*, **79**, 3262.

Authors' Biographies

Walter Pötz got his Ph.D. from the University of Graz, Austria, in 1982. His research area is the theory of semiconductors, with emphasis on optical and transport phenomena in nanostructures.

Xuedong Hu received his Ph.D. from the University of Michigan, Ann Arbor, in 1996. His research is in the area of squeezed phonon states and, more recently, coherent phenomena in semiconductors.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 209–214
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

A New Computational Approach to Photon-Assisted Tunneling in Intense Driving Fields Based on a Fabry-Perot Analogy

MATHIAS WAGNER

Hitachi Cambridge Laboratory, Madingley Road, Cambridge CB3 0HE, United Kingdom

Quantum transport in multiple-barrier systems exposed to intense laser fields is investigated by employing an analogy with the Fabry-Perot interferometer of classical optics. This analogy is exact if no further approximations are introduced. The Fabry-Perot approach is shown to be quite robust: Approximating the barriers by δ-functions with equivalent dc transmission and reflection amplitudes still yields excellent agreement with "exact" orthodox transfer-matrix calculations. The advantage of the former is its much reduced computational demands as well as ease of use. 73.40.Gk, 73.50.Pz, 72.40.+w, 42.50.Hz.

Keywords: Tunneling, quantum transport, time-dependent fields, laser, Fabry-Perot

INTRODUCTION

It is commonly agreed that ultimately, with the ever increasing miniaturization of devices, one will have to resort to quantum-transport theories for a proper modeling of device characteristics. On the other hand, it is equally clear that today's quantum-transport theories are less mature than semi-classical approaches [1] and demand huge computer resources in order to give only reasonably accurate results. Moreover, most quantum-transport theories currently available are restricted to steady-state operations or a small-signal analysis. Although strict theoretical treatments of non-equilibrium Green's functions with arbitrary time-dependence and initial many-particle correlations do exist [2], they are generally too complicated to be easily implemented. Nevertheless, serious efforts are being made in this direction (See, for instance, the NEMO project described in [3]).

The issue we want to address in this paper is that of quantum transport in the presence of intense external time-dependent driving fields where the usual small-signal analysis fails. The subject of (in) coherent quantum transport in strongly driven heterostructures has recently gained much momentum with the advent of free-electron lasers, which are capable of providing intense radiation [4]. Here we report on a new approach to the notoriously complex problem of a theoretical description of quantum transport in intense laser fields that is based on an analogy with the Fabry-

Perot interferometer in optics. In previous studies, a transfer-matrix method had been employed to calculate the reflection and transmission amplitudes of scattering states in driven heterostructures [5–10]. Due to the large size of the transfer matrix even for a one-dimensional system (the size being proportional to the highest order of photon processes one wishes to take into account), these calculations demand large computer resources when aimed at studying the regime of strong driving. In contrast, the Fabry-Perot approach is rather modest in its computational demands and moreover, much faster for not too complex heterostructures. Our aim here is not to develop a fully-fledged theory for realistic device simulation that includes all possible scattering mechanisms, but rather to put forward a useful tool for gaining a good understanding of the basic physics involved in the quantum-transport in strongly driven systems. To this end we seek to solve the time-dependent single-particle Schrödinger equation but refrain from making any assumptions whatsoever about the strength of the driving field. The additional degree of freedom introduced by an explicit time dependence of the Hamiltonian opens up the possibility for new effects to be explored, only a few of which have already been discovered and studied in some detail, such as the "dynamic localization" of minibands in driven superlattices [11, 12] and the "quenching of tunneling" in driven tunneling diodes [6, 8].

FABRY-PEROT THEORY

The textbook approach to calculating the probability for an electron to tunnel through a multiple-barrier structure is to solve for the corresponding scattering state with the proper boundary conditions, which consist of a traveling incoming wave and reflected as well as transmitted outgoing waves. A convenient way of doing this is to set up and solve a transfer-matrix problem [6, 8]. There is, however, an alternative approach that takes advantage of an analogy with optics [13]: Rather than solving a large system of simultaneous equations, one can track the electron in a manner resembling ray tracing: An electron incident on the first barrier has some probability to tunnel. If it does tunnel, it will travel to the next barrier, where it may be reflected or again tunnel. However, the electron reflected at the second barrier is not lost, it still has a chance to tunnel through this barrier. All it needs to do is to travel back to the first barrier, get reflected there and come back to the second barrier to have a second try. It is clear that in a multiple-barrier structure the number of possible paths an electron can take and still finally arrive at the far side is enormous. However, all these paths can be summed up in a few geometric series as we will demonstrate in the following.

Before proceeding any further, however, let us introduce the notation pertinent to tunneling in driven systems. The main difference compared to ordinary tunneling in static structures is that in a harmonically driven system the electron can pick up or lose integer units of photon quanta $\hbar\omega$ from the driving field. Hence, the electron is not restricted to its original energy E, but rather has a finite probability of being found in one of the *photonic sidebands* or "channels" at energies $E + n\hbar\omega$. As a consequence, *partial* transmission probabilities will need to be calculated for all channels individually. Obviously, the number of sidebands to take into account for this will depend on the strength of the driving field.

Let us denote the partial transmission amplitude of a single driven barrier from channel j (i.e., with incident energy E_j) to channel l as $t_{lj}^{l,r}$, where the superscripts l, r indicate from which side the electron impinges on the barrier (see Fig. 1). A similar notation is used for the reflection: $r_{lj}^{l,r}$. Furthermore, the "optical" path across a quantum well of width d carries the phase factor $Q_{lj} = \delta_{lj} \exp(ik_l d)$ with $k_l = \sqrt{2m(E_0 + l\hbar\omega)}/\hbar$ being the wave vector of sideband l.

For a double-barrier diode in particular, the transmission amplitudes for an electron incident,

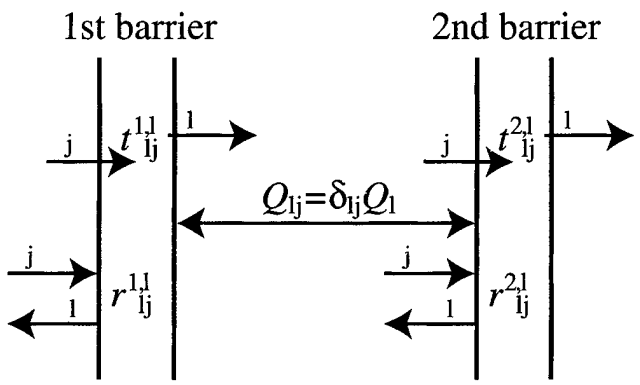


FIGURE 1 Schematics of the notations used for the transmission and reflection amplitudes and the optical path Q in the quantum well that underlie the Fabry-Perot approach.

say, from the left-hand side can be summed up in a geometric series, which in a shorthand matrix notation can be written as

$$t = t^{2,l}(I + R + R^2 + \cdots)Qt^{1,l} = t^{2,l}(I - R)^{-1}Qt^{1,l}, \tag{1}$$

with I being the identity matrix and the kernel $R = Qr^{1,r}\,Qr^{2,l}$ describing one reflection cycle in the quantum well starting and ending at the left wall of the right barrier. Note that while in the static case R is a scalar, it is a matrix of infinite rank for finite ac driving. In a similar fashion, the reflection amplitudes at the double-barrier diode are found as

$$r = r^{1,l} + t^{1,r}Qr^{2,l}(I - R)^{-1}Qt^{1,l}. \tag{2}$$

To give the reader an impression of the complexity to expect when treating more than two barriers, we also present the formula for the transmission amplitude through a *triple* barrier diode,

$$t = t^{3,l} \times \frac{1}{I - \frac{1}{(I-R_2)}Q_2t^{2,l}\frac{1}{(I-R_1)}Q_1r^{1,r}Q_1t^{2,r}Q_2r^{3,l}} \times \frac{1}{(I-R_2)}Q_2t^{2,l}\frac{1}{(I-R_1)}Q_1t^{1,l}, \tag{3}$$

where the reciprocal is understood to denote the matrix inverse.

As usual, the transmission *probability* is obtained from the transmission *amplitude* via $T_{jl} = |k_j/k_l|m_l/m_j|t_{jl}|^2$. The main advantage of a formulation along these lines is that the basic building blocks of reflection and transmission amplitudes through single barriers may be determined by other means, either approximately or even exactly. For instance, assuming the photon energy to be small compared to the energy of the incident electron, the transmission and reflection amplitudes through a single δ-function barrier, driven by a uniform ac field $F\cos\omega t$, are approximately given by [13]

$$\begin{aligned} t_n &= t_{dc}i^n\delta_{n0}, \\ r_n &= r_{dc}(-i)^nJ_n(2a), \end{aligned} \tag{4}$$

where $a = eFk_0/m\omega^2$, with k_0 being the wave vector of the incident electron, J_n the nth Bessel function of the first kind and t_{dc} and the r_{dc} the transmission and reflection amplitudes, respectively, for the *static* barrier. In this approximation, the kernel R of Eqs. (1) and (2) describing the multiple reflections in the quantum well is evaluated as

$$\begin{aligned} R_{jl} &= r^{1,r}_{dc}r^{2,l}_{dc}\sum_n \exp(ik_jd)i^{j-l-n}J_{j-l-n}(2a) \\ &\quad \times \exp(ik_{l+n}d)(-i)^nJ_n(2a) \\ &\approx r^{1,r}_{dc}r^{2,l}_{dc}\exp\{ik_0d[2 + v(3j + l)/4]\} \\ &\quad \times J_{j-l}(|4a\sin(k_0dv/4)|), \end{aligned} \tag{5}$$

where $v = \hbar\omega/E_0$ is the dimensionless photon energy.

RESULTS FOR A DOUBLE-BARRIER DIODE

For the remainder of this paper we shall focus on results for double-barrier diodes driven by a spatially uniform ac field $F\cos\omega t$. Although the Fabry-Perot method by itself is exact, the subsequent approximations made above such as the assumption of δ-barriers etc. may affect the final result. However, our findings indicate that the agreement with the standard transfer-matrix technique [10] is still excellent. Figure 2a presents typical results of the Fabry-Perot approach, based on Eqs. (1) and (4–5). Shown is the probability for an electron incident at the energy E_0 of the undriven resonance, to transverse a symmetric double-barrier structure and emerge at the far side in sideband n, i.e., at energy $E_0 + n\hbar\omega$. To leading order, the sideband amplitudes depend on the dimensionless parameter $a = eFk_0/m\omega^2$. For comparison, we also present in Figure 2b the results of a calculation using transfer matrices. The agreement is very good, with minor discrepancies appearing in the asymmetry of the "mirror" sidebands at $n = \pm|n|$, which are due to the approximations introduced in Eqs. (4–5). Apart from being computationally faster by one or two orders of magnitude compared to the transfer-matrix method, a major virtue of the Fabry-Perot approach lies in the fact that for generating Figure 2a we have only used the *dc transmission and reflection amplitudes of a single, undriven barrier.* This is a substantial advantage, as it allows us to "plug in" the transmission and reflection amplitudes obtained by using other approaches. Incidentally, the fact that the entire effect of the driving ac field can be subsumed in the Bessel function factors of Eq. (4) is a rather surprising result by itself. And last but not least, the Fabry-Perot approach is more amenable to introducing a phase-breaking mechanism by randomizing the phase in the optical-path matrix Q traversing the quantum well [14].

The agreement between the Fabry-Perot and the transfer-matrix method is just as good for other energies of the incident electron. In Figure 3 we assumed all parameters to be the same as in Figure 2, except for the electron's energy, which we took to be equal to that of the $n = -1$ sideband, i.e., $E = E_0 - \hbar\omega$. Most noteworthy in this Figure is the exact reproduction of the splitting of the $n = 0$ and $n = 2$ channels for small values of $eFk_0/m\omega^2$, which cannot be accounted for in a simpler product ansatz proposed earlier [6].

The pronounced dips in the transmission probability seen in Figures 2 and 3 at particular values of the driving field are a manifestation of the

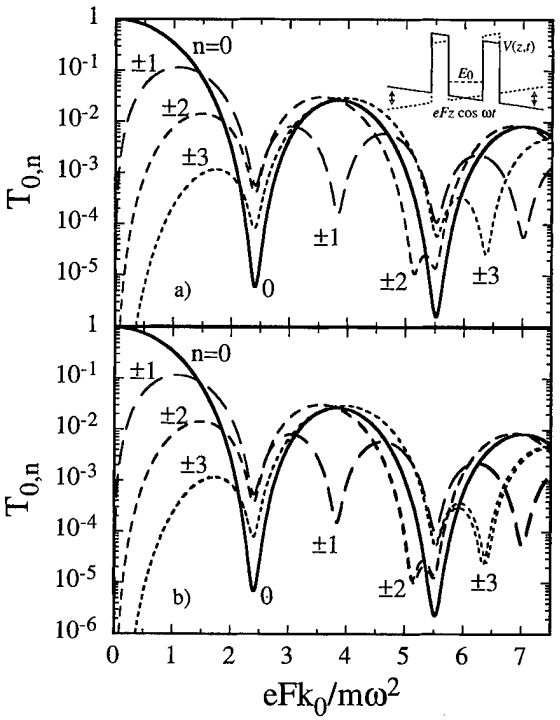


FIGURE 2 Transmission probability in the lowest few photonic sidebands through a driven double-barrier diode (see inset in top panel) as a function of driving strength for an electron incident at energy E_0, the energy of the lowest resonance in the static quantum well: a) Fabry-Perot, b) transfer-matrix method.

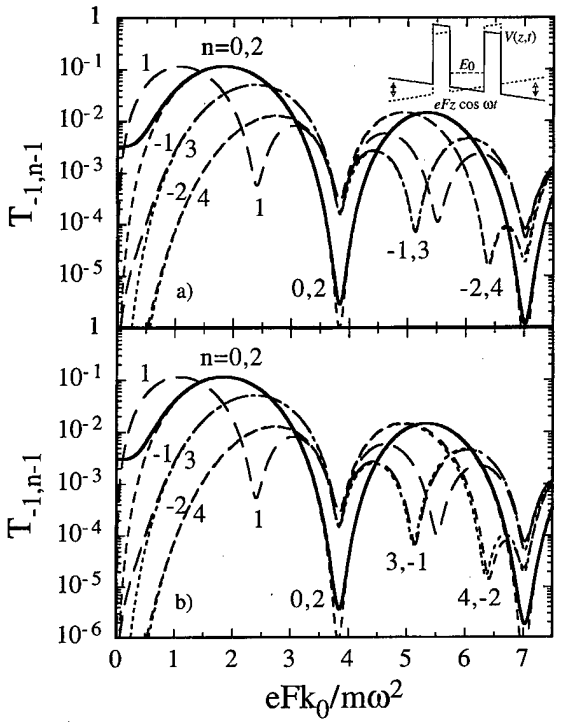


FIGURE 3 Same as Figure 2 , but for an electron incident at energy $E_0 - \hbar\omega$, i.e., one photon quantum below the quantum-well resonance.

"quenching of tunneling" predicted for strongly driven systems [6]. Basically, in order to traverse the double-barrier structure, an incident electron first has to couple into the driven resonance in the quantum well. The efficiency for this to happen is proportional to the overlap of the electronic wave function with that of the driven resonance. A similar overlap integral occurs when the electron exits the quantum well. Evidently, the electron cannot couple to the entire wave function of the driven resonance, but rather interacts only with its energy component or "spectral weight" *at the energy of the electron*. This spectral weight turns out to be a strong function of the driving field [9] and in particular may have pronounced minima under suitable driving conditions, which leave their fingerprints as characteristic minima in the transmission probability.

CONCLUSIONS

In conclusion, a Fabry-Perot approach has been used to reduce the problem of coherent transmission and reflection in strongly driven complex tunneling structures to that of *single* driven barriers. This approach gives valuable insight into the physics involved and is easy to use at the same time. It is even possible to go one step further by replacing the finite-height barriers with δ barriers having the same effective dc transmission and reflection amplitudes. It then turns out that the ac driving field affects only the *relative* amplitudes of the photonic sidebands, while the dc transmission and reflection probabilities determine the *absolute* magnitude summed over all sidebands – even for finite ac fields.

References

[1] Ferry, D. (1997). "Open problems in quantum simulation in ultradeep-submicron devices", Fifth International Workshop on Computational Electronics, Notre Dame, Indiana, May 28–30.

[2] Wagner, M. (1991). "Expansions of non-equilibrium Green's functions", *Phys. Rev. B*, **44**, 6104.

[3] Klimeck, G. *et al.* (1995). "Quantum device simulation with a generalized tunneling formula", *Appl. Phys. Lett.*, **67**, 2539.

[4] Keay, B. J. *et al.* (1995). "Photon-assisted electric field domains and multiphoton-assisted tunneling in semiconductor superlattices", *Phys. Rev. Lett.*, **75**, 4098.

[5] Liu, H. C. (1988). "Time-dependent approach to double-barrier quantum well oscillators", *Appl. Phys. Lett.*, **52**, 453.

[6] Wagner, M. (1994). "Quenching of resonant transmission through an oscillating quantum well", *Phys. Rev. B*, **49**, 16544.

[7] Iñarrea, J., Platero, G. and Tejedor, C. (1994). "Coherent and sequential photoassisted tunneling through a semiconductor double-barrier structure", *Phys. Rev. B*, **50**, 4581.

[8] Wagner, M. (1995). "Photon-assisted transmission through an oscillating quantum well: A transfer-matrix approach to coherent destruction of tunneling", *Phys. Rev. A*, **51**, 798.

[9] Wagner, M. (1996). "Strongly driven quantum wells: An analytical solution to the time-dependent schrödinger equation", *Phys. Rev. Lett.*, **76**, 4010.

[10] Wagner, M. (1996). "Coherent destruction of tunneling in resonant tunneling diodes driven by a strong laser field", in *Quantum Coherence and Decoherence, Foundations of Quantum Mechanics in the Light of New Technology*, edited by K. Fujikawa and Y. A. Ono (Elsevier, Amsterdam), p. 147.

[11] Holthaus, M. (1992). "The quantum theory of an ideal superlattice responding to far-infrared laser radiation", *Z. Phys. B*, **89**, 251.
[12] Holthaus, M. (1992). "Collapse of minibands in far-infrared irradiated superlattices", *Phys. Rev. Lett.*, **69**, 351.
[13] Wagner, M. and Zwerger, W. (1997). "Characteristic scaling parameters for tunneling in strong time-dependent electric fields", *Phys. Rev. B*, **55**, 10217R.
[14] Hu, Y. (1988). "An optical model of double-barrier resonant tunneling oscillators", *J. Phys. C*, **21**, L23.

Author Biography

Mathias Wagner received his Diploma in 1988 and his Dr. rer. nat. in 1990 in Physics, both from Hamburg University, Germany, where he was also awarded the membership of the *Studienstiftung des Deutschen Volkes*. Except for 1 1/2 years at the Central Research Laboratory, Hitachi Ltd., Tokyo, he has since been working in the Hitachi Cambridge Laboratory, UK, where his main interests include quantum transport in driven systems and correlated electronic systems. He is a member of the German Physical Society.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 215–217
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Phase Space Boundary Conditions and Quantum Device Transport

H. L. GRUBIN[a,*], J. R. CASPAR[a] and D. K. FERRY[b]

[a]*Scientific Research Associates, Inc., P. O. Box 1058 Glastonbury, CT 06033;*
[b]*Center for Solid State Electronics Research, Arizona State University, Tempe AZ 85287*

Device-device interaction is described within the framework of the quantum Liouville equation. Approximations to treating this interaction, within the framework of alterations to phase space boundary conditions, are illustrated.

Keywords: Liouville equation, phase space boundary conditions, quantum devices, thermal deBroglie wavelength, device-device interactions

INTRODUCTION

Classical descriptions of device-device interactions involve simple boundary conditions, often through an external circuit, or in some cases through electromagnetic interactions. Quantum device-device interactions are richer, as there are several critical lengths. Device-device interactions on a scale of, or smaller than, a thermal deBroglie wavelength, $\lambda_{\text{DeBroglie}} = \sqrt{\hbar^2/2mk_BT}$, introduce coherence, and two barriers separated by distances of the order of $\lambda_{\text{deBroglie}}$ result in resonant tunneling. Debye length-scales are a second critical length. But the critical length-scale of interest is an **inter-device length,** $\lambda_{\text{inter-device}}$, which is the minimum inter-device separation before changes in one device affect the properties of an adjacent device. The concept of an inter-device separation is more general than that associated with either the thermal deBroglie wavelength or the Debye length, as its value arises from phase space boundaries and involves at least two coordinates. In this paper we illustrate the concept of an inter-device length for a pair of double barriers.

We first examine the *form* of the Liouville potential in the coordinate representation: $V(x, x') \equiv V(x - Vx')$, x and x' being phase space coordinates. Based on the *form* of $V(x, x')$ we identify both non-interacting and interacting diodes. We then solve the quantum Liouville equation for a **surrogate** problem where changes in the charge distribution arise from changes in the phase space boundary conditions that, we argue, represent the device-device interaction.

* Corresponding author.

The case of non-interacting diodes is identified in Figure 1, where we display $V(x, x')$ for a pair of double barrier structures. The coordinate divisions represent multiples of an isolated device, and for a double barrier RTD of length 200 nm, the number 2 corresponds to 400 nm. In the following we will refer to two regions: region 1, where $0 \leq x \leq 1, 0 \leq x' \leq 1$, and region 3: $1 \leq x \leq 2, 1 \leq x' \leq 2$. The pairs of double barriers are centered in regions 1 and 3, and the absence of any curvature away from the double barrier Liouville potential indicates that these two structures are not interacting. Figure 2 shows a modification of the Liouville potential for a perturbation centered on the diagonal in region 1 at a point $x = x' = 0.75$. The distortion, which spills into region '3', as shown in Figure 3, will alter the density matrix in region 3. On the basis of Figure 3 and diagrams similar to it, we argue that introducing variations in the phase space boundary conditions can approximate the effect of device-device interaction on the behavior of any given device.

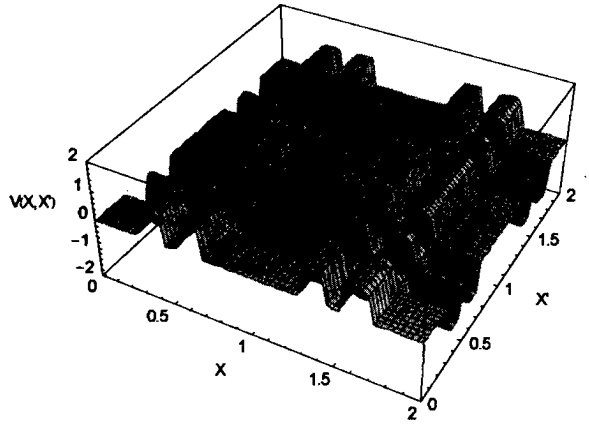


FIGURE 1 The Liouville potential for a pair of non-interacting double barrier diodes centered in regions 1 and 3.

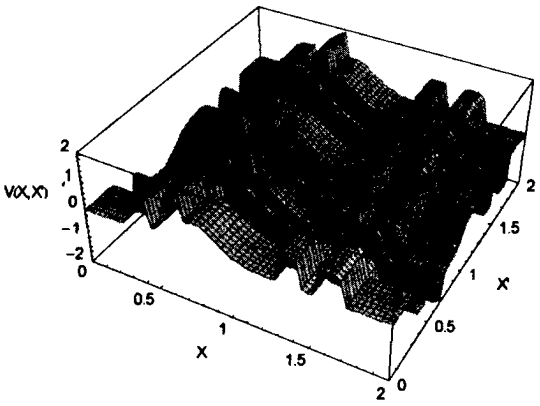


FIGURE 2 The Liouville potential for a perturbation in region 1, that spills over into region 3. These structures are interacting.

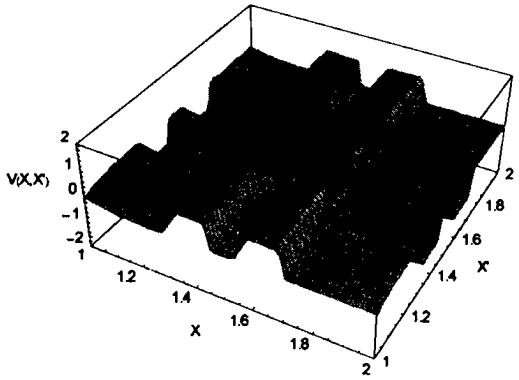


FIGURE 3 A display, within region 3, of a section of the distorted-Liouville potential of Figure 2. Note the distortion of the Liouville potential near the origin of region 3.

THE ROLE OF PHASE SPACE BOUNDARY CONDITIONS AND CONCLUSIONS

While solutions to the quantum Liouville equation provide the means for examining device-device interactions on a large scale, as a practical matter this approach is premature, as computational times would be excessive. Rather, we solve the Liouville equation subject to boundary conditions that represent the effects of device-device interaction. In particular we have solved the quantum Liouville equation [1]:

$$i\hbar\left(\frac{\partial\rho(x,x')}{\partial t}+\frac{\partial\rho(x,x')}{\partial t}\Big]_{\text{dissipation}}\right)$$
$$=-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial x'^2}\right)\rho(x,x')+V(x,x')\rho(x,x'),$$

for a 200 nm square region for symmetric and asymmetric double barrier structures, and symmetric quantum wells. For noninteracting structures the density matrix boundary conditions along the line $x' = 200$nm, was exp $[-((x-x') / 4 \lambda_{deBroglie})^2]$. For calculations representing interacting structures the boundary condition is $A(x-x')$ exp$[-((x-x') / 4 \lambda_{deBroglie})^2]$. The modulating function was chosen arbitrarily. Figure 4 displays the boundary variation invoked in this study, and Figure 5 displays the effects of this boundary variation on the charge density within the symmetric RTD. Figure 5 also displays the potential energy. The calculation with the smallest charge within the barrier occurs for a modulating function that is set to unity. For the change in the modulating function shown in Figure 4, there is significant charge accumulating within the barriers. Very similar results emerge from the asymmetric double barrier structures. *Significantly*, the form of the perturbation affects the form of the charge density alterations. Additional calculations reveal that the effect of the boundary is most dramatic for perturbations close to the origin. Because of the Gaussian decay of the boundary condition, the further the perturbation is away from the origin the weaker the effect. The results also suggest that the shorter the inter-device separation the greater the interaction. Supplementary calculations indicate that double barrier diodes with cladding regions doped to $10^{24}/m^3$, require a minimum separation of approximately 180 nm to be considered as isolated.

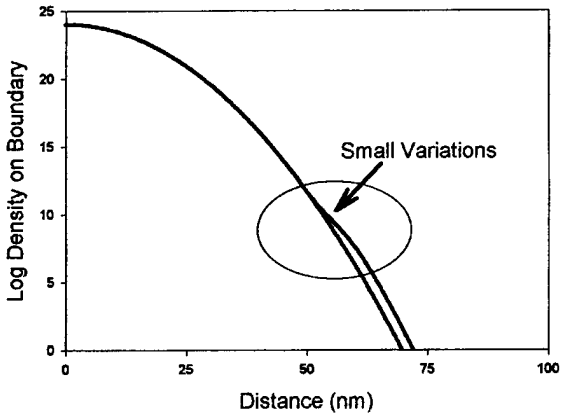


FIGURE 4 Variations of the density matrix boundary condition along $x' = 200$nm for a double barrier structure.

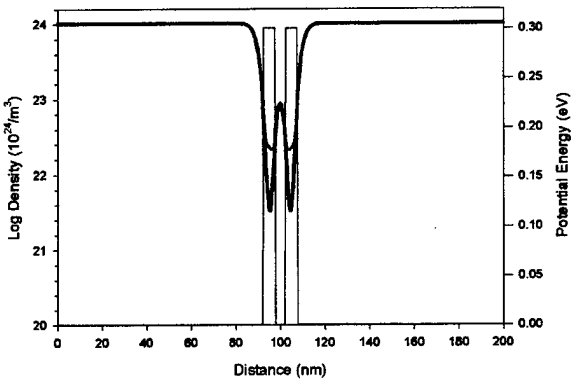


FIGURE 5 The density distributions for a non-self consistent solution to the quantum Liouville equation subject to the variations of Figure 4.

Acknowledgements

This work was sponsored by DARPA.

References

[1] Grubin, H. L. (1994). Density Matrix Simulations of Semiconductor Devices: "Quantum Transport in Ultra Small Devices", Eds. D. K. Ferry, H. L. Grubin and A. P. Jauho, Plenum Press, N. Y.

QUANTUM STRUCTURES

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 219–223
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Single-Electron Memories

CHRISTOPH WASSHUBER*, HANS KOSINA and SIEGFRIED SELBERHERR

Institute for Microelectronics, TU-Vienna Gusshausstrasse 27-29/E360, A-1040 Wien, Austria

One of the most promising applications of single-electronics is a single-electron memory chip. Such a chip would have orders of magnitude lower power consumption compared to state-of-the-art dynamic memories, and would allow integration densities beyond the tera bit chip.We studied various single-electron memory designs. Additionally we are proposing a new memory cell which we call the *T*-memory cell. This cell can be manufactured with state-of-the-art lithography, it operates at room temperature and shows a strong resistance against random background charge.

Keywords: Memory, single-electron tunneling, simulation, random background charge

1. INTRODUCTION

Over the last years many applications of the Coulomb blockade and of Coulomb oscillations, so called single-electron devices, have been proposed. The most fundamental device is the single-electron transistor [1] which was soon followed by pumps, turnstiles, supersensitive electro-meters, memories and logic gates (see for example [2]). Many of these devices have been fabricated as laboratory prototypes with various manufacturing processes, and have been operated usually at temperatures at or below liquid Helium (4.2 K). Some effects could be seen and some prototypes could be operated up to liquid nitrogen temperature (77 K) [3] and only very few functioned at room temperature (300 K) [4]. But almost all of these prototypes suffer from the extraordinary charge sensitivity of single-electron devices which makes them strongly vulnerable to uncontrollable impurities, traps and charge transport in other parts of the circuit. Additionally all employed manufacturing processes have not been proven to be applicable for mass production. That is, the challenge for the next years is to find reliable and reproducible mass production methods for robust room temperature single-electron devices.

All results and conclusions are based on computer simulations which have been performed with our single-electron device and circuit simulator SIMON. For more information about SIMON please refer to [5, 6].

* Corresponding author.

2. DIFFERENT SINGLE-ELECTRON MEMORY CELLS

In the following we are describing briefly the function of different single-electron memory cells. Some of them have already been extensively treated in other publications.

2.1. Flip-Flop

One design possibility is to mimic conventional CMOS architecture with single-electron devices. A static single-electron memory cell design or flip-flop, based on this approach was proposed by A. Korotkov *et al.* [7].

2.2. Electron Trap

Nakazato and Ahmed [8] proposed a dynamic memory cell pushed to its extreme limit. A small number of electrons is stored on a single quantum dot. The central island of a single-electron transistor which is extremely charge sensitive, may be used as sensor for the logic states. However exactly this charge sensitivity makes these devices prone to random background charge. The more tunnel junctions are used the less likely it is that electrons are co-tunneling to ground. That is why this memory is a dynamic memory, since the charges can not be trapped infinitely. Nevertheless, the refresh rate can easily extend to the second or minute range.

2.3. Ring Memory

Another idea which is a generalization of the bistable quantum cell for cellular automata by Lent *et al.* [9] is the ring memory cell. An even number n of tunnel junctions are connected to a ring, and $n/2$ electrons are inserted into the ring. These electrons will repel each other and can form two stable configurations. Applying voltage pulses will switch the state of the ring to either one of the stable configurations.

2.4. Q_0 -Independent Memory

The first random background charge or Q_0-independent memory was proposed by Likharev and Korotkov [10]. The basic idea is the following. Electrons are stored on an island or floating gate. A single-electron transistor which is very charge sensitive on its gate, is used to sense the changes of charge on the floating gate. The trick to achieve the Q_0-independence is not to sense any absolute charges, but to sense the current oscillations in the single-electron transistor, which are caused by changes in the charge on the gate. In other words, the charge change on the floating gate induces current oscillations. These oscillations occur at any background charge. Only the phase, not the amplitude is background charge dependent. The cell can only be read destructively by discharging the floating gate. If current oscillations are detected, the floating gate was charged with electrons. If no oscillations are detected, no charge was stored on the floating gate.

2.5. Discussion of Simulation Results

We simulated the above memory cells and especially studied their operation temperature, complexity and random background charge dependence. Due to lack of space we can not discuss the details of our simulation results. But all of the above designs have at least one flaw. Many are random background charge dependent (flip-flop, electron trap, ring memory). Some are quite complex and show a relative low operation temperature compared to other designs with the same tunnel junction parameters (flip-flop, ring memory).

3. THE THREE CHALLENGES

Thus we can name three challenges that have to be addressed:

- Room temperature operation,
- Industrial mass production,

- Random background charge independence for reliable operation.

Coulomb blockade effects are only observed if the charging energy associated with the tunneling of a single electron is bigger than the thermal energy kT, with k the Boltzmann-constant, and T the absolute temperature. To produce devices which operate at room temperature a resolution limit of < 10 nm is necessary. This means, that for industrial mass production only naturally formed tunnel junctions (poly-silicon, granular films) are feasible today. It is not possible with todays industrial lithography to produce a single tunnel junction with dimensions smaller 10 nm. The resolution limit of optical lithography is more than a magnitude larger than 10 nm. Clearly, the next decade will see new high resolution lithography methods (X-ray, e-beam, near-field, nano-imprint, STM, AFM, ...), but until they are not available for industrial mass production only naturally formed tunnel junctions in granular films can reach resolutions in the nano-meter regime and thus allow room temperature operation.

4. A NEW MEMORY CELL

Combining todays lithography with granular film processes one can manufacture what we call a T-memory cell. Two granular film batches are arranged in a T-shape (see Fig. 1). It is a combination of a multi-electron trap with a multi-island transistor for read-out. The cross-bar of the T is the multi-island transistor which is controlled by a gate electrode which stores either a number of electrons or a number of holes. Writing a '1' or '0' is done by applying a positive or negative voltage pulse at the word-line. This will charge the gate electrode which has a similar function as the floating gate in a flash memory cell. Thus in our example a positive write pulse on the word-line will store some holes on the gate electrode representing '1' and a negative write pulse will store some electrons on the gate electrode representing '0'. Destructively reading the memory cell is done by applying for example a negative voltage pulse at the word-line and sensing current oscillations of I_{out}. If I_{out} oscillates then the cell held a '1'. If no oscillations are picked up then the cell held a '0'. Accordingly the contents of the cell has to be restored.

The T-cell can be viewed as a combination of the electron trap [11] and the Q_0-independent memory cell [10], with the change that we use tunnel junction arrays for both elements, the trap and the read-out transistor. This change makes the T-memory cell much easier to manufacture. In fact,

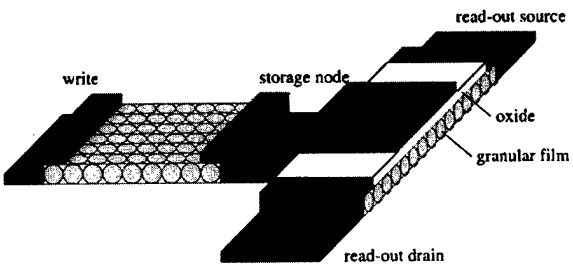


FIGURE 1 T-memory cell.

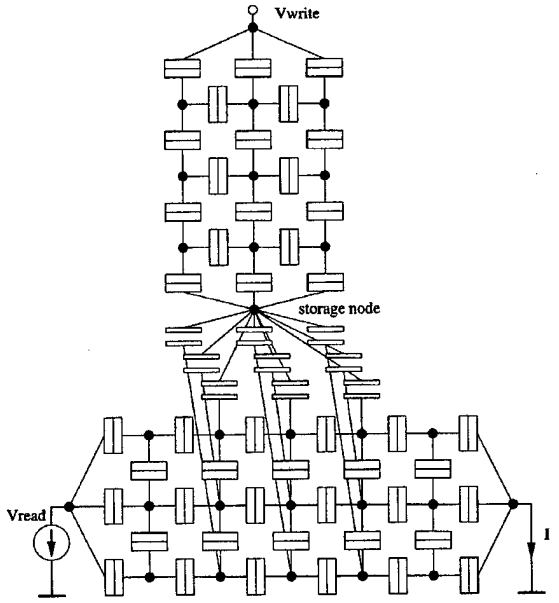


FIGURE 2 T-memory cell circuit.

since the granular film batches can have dimensions in the 100 nm regime, state-of-the-art optical lithography is sufficient to produce such memory cells.

Yano *et al.* [12] manufactured memory cells where the storage dots and read-out circuit are located in the same granular film, a poly-silicon film. The disadvantage of this design is, that it is much more difficult or even impossible to tune process parameters in order to change the characteristics of storing and reading separately. In our design, the granular films for the storing part and the read-out part could be produced with entirely different process parameters, thus optimizing each performance characteristics.

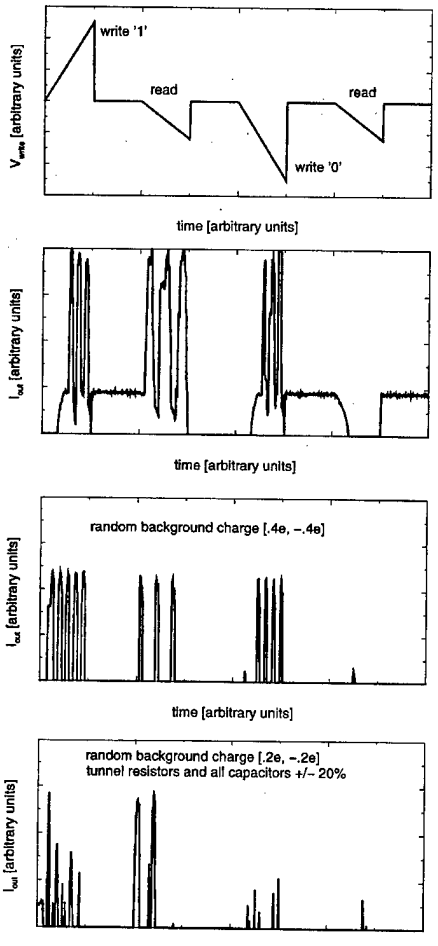


FIGURE 3 Writing and reading the *T*-memory cell.

We simulated the cell of Figure 2 and similar ones. In Figure 3 we give a typical write-read cycle for a '1' and a '0'. We simulated the same write-read cycle for different random background charges (Q_0 = 0, [0.2e, −0.2e], [0.4e, −0.4e]) and randomly changed tunnel resistors and capacitors. As can be clearly seen, the Coulomb oscillations are only present when a '1' was stored in the memory, independent of background charge.

5. CONCLUSION

We simulated and studied various single-electron memory designs. By analyzing proposed memories (flip-flop, electron trap, Q_0-independent memory, ...) we came up with a new memory cell, our so called *T*-memory cell. It is manufacturable with todays production methods, operates at room temperature, shows a strong background charge resistance, and is down-scalable to atomic dimensions.

References

[1] Fulton, T. A. and Dolan, G. J., Observation of single-electron charging effects in small tunnel junctions. *Physical Review Letters*, **59**(1), 109–112, July 1987.

[2] Averin, D. V. and Likharev, K. K. (1992). Possible applications of the single charge tunneling. In H. Grabert and M. H. Devoret, editors, *Single Charge Tunneling – Coulomb Blockade Phenomena in Nanostructures*, Chapter 9, 311–332. Plenum Press and NATO Scientific Affairs Division, New York and London.

[3] Chen, W.,. Ahmed, H. and Nakazato, K., Coulomb blockade at 77 K in nano-scale metallic islands in a lateral nano-structure. *Applied Physics Letters*, **66**(24), 3383–3384, June 1995.

[4] Yano, K., Ishii, T., Hashimoto, T., Kobayashi, T., Murai, F. and Seki, K. (1994). Room-temperature single-electron devices. In *Extended Abstracts of the International Conference on Solid State Devices and Materials*, pp. 325–327, Yokohama.

[5] Wasshuber, C. and Kosina, H. (1997). A single-electron device and circuit simulator. *Superlattices and Microstructures*, **21**(1), 37–42.

[6] Web-page of SIMON. http://members.magnet.at/cats-meow/ and http://www.iue.tuwien.ac.at/software/simon.html.

[7] Korotkov, A. N., Chen, R. H. and Likharev, K. K., Possible performance of capacitively coupled single-electron transistors in digital circuits. *Journal of Applied Physics*, **78**(4), 2520–2530, August 1995.

[8] Nakazato, K. and Ahmed, H., The multiple-tunnel junction and its application to single-electron memory and logic circuits. *Japanese Journal of Applied Physics*, **34**(Part1, 2B), pp. 700–706, February 1995.
[9] Lent, C. S., Tougaw, P. D. and Porod, W. A. Bistable quantum cell for cellular automata. In *Int. Workshop on Computational Electronics*, pp. 163–166, May 1992.
[10] Likharev, K. K. and Korotkov, A. N. (1995). Analysis of Q_0-independent single-electron systems. In *Int. Workshop on Computational Electronics*, p. 42.
[11] Nakazato, K. Blaikie, R. J. and Ahmed, H., Single-electron memory. *Journal of Applied Physics*, **75**(10), 5123–5134, May 1994.
[12] Yano, K., Ishii, T., Sano, T., Mine, T., Murai, F. and Seki, K. (1996). Single-electron-memory integrated circuit for giga-to-tera bit storage. *IEEE International Solid-State Circuits Conference*, pp. 266–267.

Author's Biography

Christoph Wasshuber received the *Diplomingenieur* and the Ph.D. degrees in micro-electronics and semiconductor-electronics from the Vienna University of Technology, Austria, in 1993 and 1997, respectively. He was for one and a half years with the *Asada Laboratory at the Tokyo University, Japan* studying single-electronics, and joined then the *Institut für Mikroelektronik* at the Vienna University of Technology. His current interests include modeling and simulation of single-electron devices, quantum effects in semiconductor devices, and computer aided engineering in VLSI technology in general.

Hans Kosina was born in Haidershofen, Austria, in 1961. He received the *Diplomingenieur* and the Ph.D. degrees in electrical engineering from the Vienna University of Technology in 1987 and 1992, respectively. He was one year with the *Institut für flexible Automation* and joined then in 1988 the *Institut für Mikroelektronik* at the Vienna University of Technology. Currently he is employed as an assistant professor heading the device modeling group.

His current interests include physics and technology of solid state devices and integrated circuits.

Siegfried Selberherr was born in Klosterneuburg, Austria, in 1955. He received the degree of *Diplomaingenieur* in Control Theory and Industrial Electronics from the Vienna University of Technology in 1978. Since that time he joined the *Institut für Allgemeine Elektrotechnik und Elektronik*, previously called the *Institut für Physikalische Elektronik*, at the Vienna University of Technology. He finished his thesis on "Two Dimensional MOS Transistor Modeling" in 1981. Dr. Selberherr has been holding the *venia docendi* on computer-aided design since 1984. He is the head of the *Institut fur Mikroelektronik* since 1988. Hiscurrent topics are modeling and simulation of problems for microelectronics engineering. He authored and coauthored more than 200 publications in journals and conference proceedings. Furthermore, he wrote a book *Analysis and Simulation of Semiconductor Devices*. Dr. Selberherr is member of the *Association for Computing Machinery* (1979), the *Society of Industrial and Applied Mathematics* (1980) and the *Verband deutscher Elektrotechniker*.

VLSI DESIGN
1998, Vol. 8, Nos. (1-4), pp. 225-230
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Electron-LA Phonon Interaction in a Quantum Dot

T. EZAKI*, N. MORI and C. HAMAGUCHI

Department of Electronic Engineering, Osaka University, Japan 2-1 Yamada-oka, Suita, Osaka 565, Japan

Relaxation time due to electron-longitudinal-acoustic (LA) phonon interaction is calculated in a GaAs quantum dot (QD) with N electrons (from $N=1$ to 4), where electrons in a narrow quantum well are confined by a parabolic confining potential, by using the exact eigen states of electrons. Although the energy levels become dense with increasing the number of electrons, the modification of the relaxation time is found to be not so strong, which attributes the fact that many electron eigen states consist of only a few dominant single electron states which limit the electron relaxation. By comparing relaxation process via intermediate states with the direct process, several fastest processes are found to be realized by relaxation through intermediate states between the initial state and the ground state. The effect of change in the quantum well width is also discussed.

Keywords: Electron-phonon interaction, quantum dot, phonon bottleneck, exact diagonalization method, Coulomb interaction

1. INTRODUCTION

Low dimensional semiconductor structures such as quantum dots (QDs) have attracted much attention because of their unique features in electronic properties [1-3] and their applications to future electron devices [4, 5]. Although their quantized atomic like energy levels are expected to have advantages for optical devices, energy relaxation via electron-optical-phonon interaction is diminished in QDs for energy level separation of order 1 meV because the mismatch between the energy level separation and the optical-phonon energy of, for example, 36 meV in the bulk GaAs [6-10], which is often referred as "phonon bottleneck". Several theoretical and experimental approaches have been devoted to the study of the phonon bottleneck and pointed out a possibility to avoid it by utilizing Auger processes [11], multiphonon processes [12] or the role of excited energy levels of the exciton [13].

In the process of light emission, electrons and holes are first created in higher energy states and relax down to the ground state in cascade via

* Correpsonding author.

Preceding Page Blank

phonon emissions and finally recombine to emit light. Energy relaxations from higher energy levels to sublevels near the ground state are expected to be fast because optical-phonon interaction plays a dominant role in the electron relaxation. At the final stage of electron relaxation toward the ground state, electron-longitudinal-acoustic (LA) phonon interaction is expect to be important because of the small energy separation compared to the optical-phonon energy. In a QD containing a few electrons with dot size comparable to the effective Bohr radius, many electron eigen states are considerably modified by the Coulombic interaction between the electrons. In order to investigate electron-phonon interaction in such a QD, we therefore take into account the exact N-electron eigen states (for N up to 4) by numerically diagonalizing the complete N particle Hamiltonian [14–16] including the Coulombic interaction between electrons and then calculate electron-LA phonon relaxation rates toward the ground state using the N particle eigen states. Since there are many intermediate states between the initial state and the ground state, we calculate relaxation time via intermediate states in order to find the fastest process.

2. EIGEN STATES OF A MODEL QD

We consider a QD fabricated on a AlGaAs/GaAs/AlGaAs quantum well structure as a model system. The confining potential of a circular QD in x-y plane parallel to the hetero-interface is modeled by an isotropic and parabolic potential $\frac{1}{2}m^*\omega_0(x^2+y^2)$ with $\hbar\omega_0$ being the single electron confining energy. Throughout this paper we assume a typical value of $\hbar\omega_0 = 1$ meV for the parabolic potential. The Hamiltonian of N electrons including the Coulomb interaction between the electrons can be written as

$$\mathcal{H} = \sum_{i=1}^{N} \mathcal{H}_{0i} + \sum_{i<j} \frac{e^2}{4\pi\varepsilon |r_i - r_j|} \qquad (1)$$

$$\mathcal{H}_{0i} = -\frac{\hbar^2}{2m^*}\nabla_i^2 + \frac{1}{2}m^*\omega_0^2(x_i^2 + y_i^2) + V(z_i) \qquad (2)$$

where m^* is the effective mass of an electron. Along the z direction, we assume an infinite square well potential of the width W and consider only the ground state associated with the quantized z motion, because electrons are frozen out into the lowest subband in experimentally realized dots.

In order to solve the N particle Hamiltonian, we use N particle Slater determinants as the basis sets which can be composed by the solutions of the single particle Hamiltonian $\mathcal{H}_{0i}$. The many particle eigen functions are then obtained by diagonalizing the many particle Hamiltonian. Figure 1 shows the lowest 40 eigen energies of a QD with single, two and three electrons. For a single electron QD, equally separated energy levels are seen, which reflects the single-electron confining energy $\hbar\omega_0 = 1$ meV. When more electrons are added into the QD, total ground state energy of the system moves up and becomes greater than the simple sum of the single-electron energy. Moreover the energy level structure becomes more dense with increasing the number of electrons. From this analysis the Coulomb interaction is expected to play an very important role in determining the energy spectra in a few electron system in a QD [1, 16].

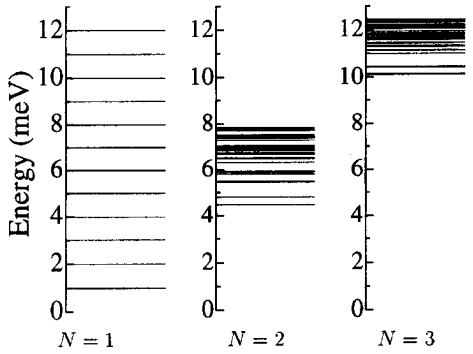


FIGURE 1 Total eigen energies in a QD which contains single, two and three electrons. In the model used in the present study, energy level separation for single electron QD is $\hbar\omega_0 = 1$ meV. For a QD with two and three electrons, energy level spectra become to form more complex and dense structures due to the Coulombic interaction.

3 ELECTRON-LA PHONON INTERACTION

Since we are interested in energy relaxation of electrons in QDs at low temperatures, we take into account only the phonon emission process and thus assume the lattice temperature $T=0$K for simplicity. Since the maximum energy separation in our model system is the order of 1 meV, electron-optical-phonon relaxation can be ignored and the electron-LA phonon interaction dominates the relaxation processes. The electron-LA phonon scattering rate $W_{ii'}$ derived from the first-order perturbation theory using the Fermi's golden rule is described by the following relation

$$W_{ii'} = \frac{2\pi}{\hbar}\sum_{q} |\langle i'|H_{e-ph}|i\rangle|^2 (n_q+1)\delta(E_{i'}-E_i+\hbar\omega_q) \tag{3}$$

where n_q is the phonon occupation number which is equal to 0 in the present calculation. The electron-LA phonon interaction Hamiltonian H_{e-ph} is given by the following equation

$$H_{e-ph} = \sum_{ii'}\sum_{q} V_{ii'}(q) C_{i'}^{\dagger} C_i (a_q + a_{-q}) \tag{4}$$

$$V_{ii'} = D\sqrt{\frac{\hbar}{2d\omega_q}}\langle i'|qe^{iqr}|i\rangle \tag{5}$$

where $C_i^{\dagger}(C_i)$ is an operator to create (annihilate) an electron of state, i and $a_q^{\dagger}(a_q)$ is an operator for creation (annihilation) of a phonon with momentum q. In the present calculation we assume that the electrons interact with bulk LA phonons and thus the deformation potential for electron-LA phonon is given by the deformation potential of the bulk GaAs, $D=7$ eV.

4. RESULTS AND DISCUSSION

Figure 2 shows the relaxation times for transition from the excited states to the ground state in a QD with $\hbar\omega_0 = 1$ meV and $W=50$ Å. In the relaxation processes we have to take into account scattering

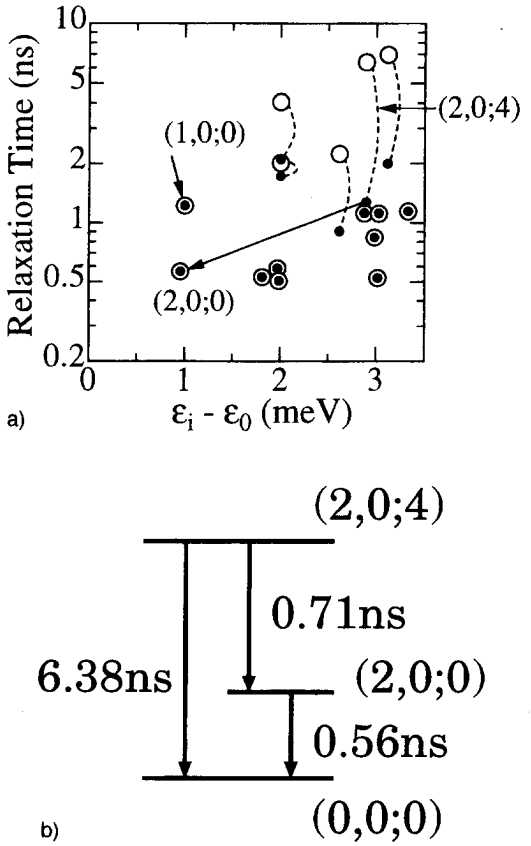


FIGURE 2 Relaxation time as a function of the energy difference between the initial state and the ground state for a two-electron QD (a). Large open circles and small solid circles denote the direct relaxation and the fastest relaxation, respectively. A schematic diagram for the direct and the fastest relaxations from the initial state (2, 0; 4) is also plotted (b).

from the excited state to the ground state through various intermediate states in addition to the direct transition. In Figure 2(a) we show the calculated relaxation time toward the ground state in a two-electron QD where open circles denote direct processes and the fastest relaxations are plotted by the solid circles. The indices $(\mathcal{M},\mathcal{S};\mathcal{N})$ shown in Figure 2 are quantum numbers of eigen states, indicating $\mathcal{N}$-th state of N-electrons system with a total angular momentum $\hbar\mathcal{M}$ and a total spin $\hbar\mathcal{S}/2$. The connected pairs of an open circle and a solid circle indicate the relaxation from the same initial states and mean the existence of faster indirect process, whereas the open circles with the

solid circles indicate that the fastest process is the direct process and the indirect process is very slow because of the symmetry of the wave functions. For electrons with $(\mathcal{M},\mathcal{S};\mathcal{N}) = (2, 0; 4)$, the fastest indirect transition arises from the scattering process via the state $(2,0;0)$ as shown by the arrow. A schematic diagram of the relaxations for the $(2, 0; 4)$ state is illustrated in Figure 2(b), where the energy separations from the ground state and the relaxation times are given for the corresponding eigen states in Figure 2(a). The relaxation time for the direct process is about 6.38 ns, whereas the relaxation time for the indirect process is given by 1.27 ns which is given by the sum of 0.71 ns for the scattering from $(2,0;4)$ to $(2,0;0)$ and 0.56 ns for the scattering from $(2,0;0)$ to the ground state $(0,0;0)$. In Figure 3 we plotted relaxation times for $N=3$ and 4 electrons in a QD with $\hbar\omega_0 = 1$ meV and $W=50$ Å, where the electrons at higher energy states have longer relaxation time due to the difficulty in satisfying the energy and momentum conservation. Although energy levels become dense with increasing the number of electrons, modification of relaxation times are not so strong because many electron eigen states consist of only a few dominant single electron states which limit the electron relaxation for many electron states.

In Figure 4 we compare the fastest relaxation times in two QDs with different well widths $W=50$ Å (solid circles) and 100 Å (open triangles), where two cases are considered, one for 2 electrons (a) and the other for 3 electrons (b) in the QDs. The change in the well width W results in the change in the confining energy in z direction and in no change in the eigen energy in x-y direction. The change in the confinement in the z direction changes the maximum phonon wave vector $q_{\max} \sim 2\pi/W$ which gives the maximum phonon energy $\hbar c_s q_{\max}$ [8] where c_s is the sound velocity. Maximum phonon energy is 4.3 meV for $W=50$ Å

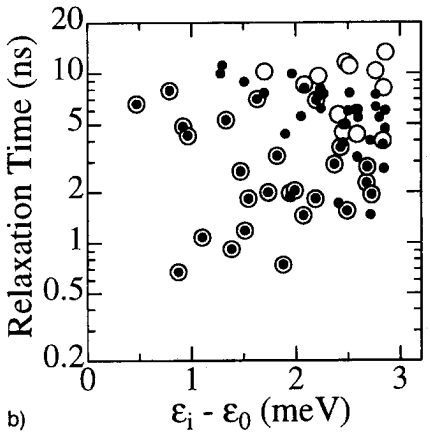
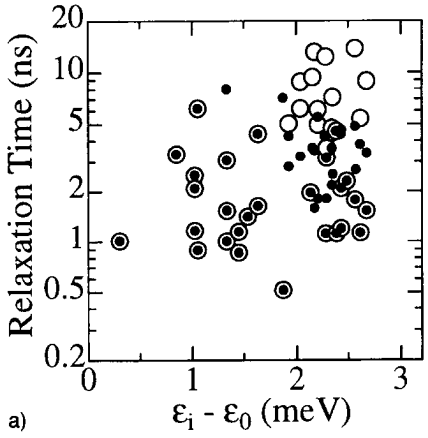


FIGURE 3 Relaxation time as a function of the energy difference between the initial state and the ground state for three (a) and four (b) electrons' QDs. Large open circles and small solid circles denote the direct relaxation and the fastest relaxation, respectively.



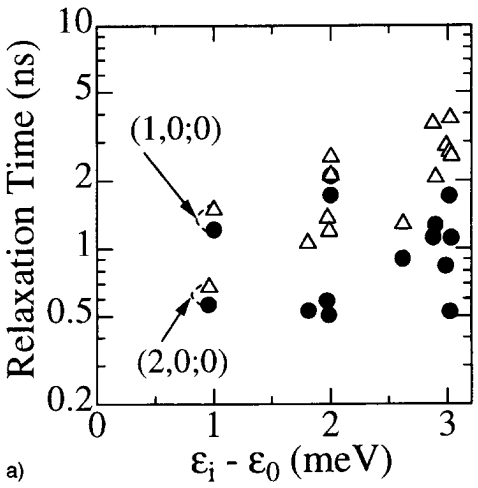


FIGURE 4 Electron relaxation rate in a QD containing the number of electrons $N=2$(a) and 3(b) with different well widths of $W=50$ Å (solid circles) and 100 Å (open triangles).

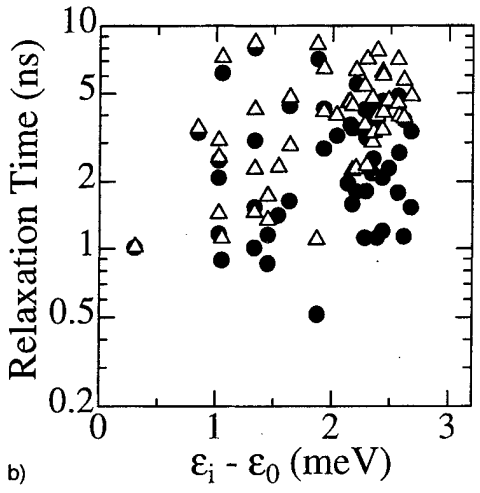


FIGURE 4 (Continued).

and is 2.2 meV for $W = 100$ Å. It is seen in Figure 4 that the relaxation time for a QD with $W = 100$ Å begins to increase at the energy separation of 2 meV. From the present analysis it may be concluded that the change of the confining energy in the z direction results in a change in the electron relaxation process even in a QD with several electrons.

5. CONCLUSION

Electron relaxations toward the ground state via LA-phonon scatterings in a QD with interacting N electrons (from $N = 1$ up to 4) are calculated using N electrons' eigen states obtained by the exact diagonalization method. We find that energy levels become dense with increasing the number of electrons and that the modification of relaxation times is not so strong because many electron eigen states consist of only a few dominant single electron states which limit the electron relaxation. Several fastest processes are realized by relaxation through intermediate states between the initial state and the ground state.

We find that the confinement in the z direction limits the maximum phonon energy and relaxation time becomes longer when an energy separation between the initial state and the final state exceeds maximum phonon energy.

References

[1] Ashoori, R. C., Stormer, H. L., Weiner, J. S., Pfeiffer, L. N., Baldwin, K. W. and West, K. W., "*N*-Electron Ground State Energies of a Quantum Dot in Magnetic Field", *Phys. Rev. Lett.*, July 1993.

[2] Weinmann, D., Häusler, W. and Kramer, B., "Spin Blockades in Linear and Nonlinear Transport through Quantum Dots", *Phys. Rev. Lett.*, February 1995.

[3] Tarucha, S., Austing, D.G., Honda, T., van der Hage, R. J. and Kouwenhoven, L. P., "Shell Filling and Spin Effects in a Few Electron Quantum Dot", *Phys. Rev. Lett.*, October 1996.

[4] Bazán, G., Orlov, A. O., Snider, G. L. and Bernstein, G. H., "Charge detector realization for AlGaAs/GaAs quantum-dot cellular automata", *J. Vac. Sci. Technol. B*, Nov./Dec. 1996.

[5] Buks, E., Schuster, R., Heiblum, M., Mahalu, D., Umansky, V. and Shtrikman, H., "Measurement of Phase and Magnitude of the Reflection Coefficient of a Quantum Dot", *Phys. Rev. Lett.*, November 1996.

[6] Benisty , H., Sotomayor-Torrès, C. M. and Weisbuch, C., "Intrinsic mechanism for the poor luminescence properties of quantum-box systems", *Phys. Rev. B*, November 1991.

[7] Bockelmann, U. and Bastard, G., "Phonon scattering and energy relaxation in two-, one-, and zero-dimensional electron gases", *Phys. Rev. B*, November 1990.

[8] Benisty, H., "Reduced electron-phonon relaxation rates in quantum-box systems: Theoretical analysis," *Phys. Rev. B*, May 1995.

[9] Sotomayor Torres, C. M. (1996). "Energy Relaxation in quantum dots: recent developments on the phonon bottleneck", *Hot carriers in Semiconductors*, New York, Plenum Press.

[10] Sawaki, N., Niwa, S., Taya, M., Murakami, T. and Suzuki, T. (1996). "Reduction of energy relaxation rate of photo-excited hot electrons in quasi-one and zero dimensional structures", *Hot carriers in Semiconductors*, New York, Plenum Press.

[11] Bockelmann, U. and Egeler, T., "Electron relaxation in quantum dots by means of Auger processes", *Phys. Rev. B*, December 1992.

[12] Inoshita, T. and Sakaki, H., "Electron relaxation in a quantum dot: Significance of multiphonon processes", *Phys. Rev. B*, September 1992.

[13] Bockelmann, U., "Exciton relaxation and radiative recombination in semiconductor quantum dots", *Phys. Rev. B*, December 1993.

[14] Maksym, P. A. and Chakraborty, T., "Quantum Dots in a Magnetic Field: Role of Electron-Electron Interactions", *Phys. Rev. Lett.*, July 1990.

[15] Maksym, P. A. and Chakraborty, T., "Effect of electron-electron interactions on the magnetization of quantum dots", *Phys. Rev. B*, January 1992.

[16] Yang, R. and Ruden, P. P., "Electron-electron interaction in three-dimensional model quantum box", *J. Appl. Phys.*, August 1995.

Authors' Biographies

Tatsuya Ezaki is a graduated student at Department of Electronic Engineering, Osaka University, Osaka, Japan and expected to receive Ph. D. degree in March 1998. He has been involved with research work on electron transport in quantum wires and infra-red detectors utilizing intersubband transition in multi-quantum wells. His current interests include electronic states in quantum dots and electron-phonon interaction in quantum dots. He is a memeber of Japan Society of Applied Physics.

Nobuya Mori is an Assistant Professor at Department of Electronic Engineering, Osaka Unviersity, Osaka, Japan. His research interests include quantum transport in semiconductor nano-structures and magnetotransport in semiconductors. He was a British Council Fellow at Department of Physics, University of Nottingham, Nottingham, England, from August 1994 to March 1996. He is a member of Institute of Physics, a memebr of Physical Society of Japan, and a memebr of Japan Society of Applied Physics.

Chihiro Hamaguchi is a Professor of Electronic Engineering at Osaka Univesity, Osaka, Japan. His research interests include electron transport in semiconductor quantum structures, magnetotransport in semiconductors, modualtion spectroscopy and device simulation. He is a Fellow of the IEEE, a Fellow of the Americal Physical Society, Chairman of Electron Devices Society of IEEE Tokyo Chapter, Editorial Board of Semiconductor Science and Technology. He has served the chairman of International Conference on Hot Carriers, vice-chariman of International Symposium on New Phenomena in Mesoscopic Structures and committee member of many international conferences.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 231–235
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Self-Consistent Calculations of the Ground State and the Capacitance of a 3D Si/SiO_2 Quantum Dot

A. SCHOLZE*, A. WETTSTEIN, A. SCHENK and W. FICHTNER

Swiss Federal Institute of Technology, Integrated Systems Laboratory, Gloriastrasse 35, CH-8092 Zürich, Switzerland

We perform self-consistent electronic structure calculations in the framework of inhomogeneously and anisotropically scaled local density functional theory of a fully 3D modeled Si/SiO_2 quantum dot. Electrons are laterally confined in the semiconductor/oxide heterojunction by a metallic gate atop of the device. Total charge densities, total free energies, chemical potentials for different numbers of electrons in the dot, and the differential capacitances for various dot sizes are calculated. We observe shell filling effects in the differential capacitance. The *magic*-numbers are governed by the six valley bandstructure of silicon, which leads to four fold degenerated single particle levels in the dot.

Keywords: Si/SiO_2 quantum dots, capacitance, self-consistent calculations, shell-filling effects

1. INTRODUCTION

With VLSI/ULSI production technology rapidly approaching the 100 nm range, simulation of quantum effects plays an increasingly important role in nowadays device simulation efforts. For most of the state of the art device simulation tools available today, work is already under way to include the effects of one dimensional quantum confinement in silicon MOSFET channels [1]. Typically, the Schrödinger-Poisson equation is solved and corrections to the charge density, which enters the semiclassical device equations are given. However, with device features below the 100 nm range in silicon and zero dimensional confinement in quantum dots dominating the device operation, this concept is bound to fail to give accurate descriptions.

Throughout the last decade there has been intense research focusing on so called single electron devices. These devices utilize effects as Coulomb blockade and single electron charging, which arise with zero dimensional confinement. Various concepts are pursued which encompass for instance single electron transistors (SET) [2] or two state switching cells in an adiabatic switching

* Corresponding author. Present address: Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana, Illinois 61801.

paradigm [3]. However, all these concepts suffer from severe limitations. The demands on the purity of materials to avoid random offset charges will be difficult to meet. Phase coherent switching as in the Lent/Porod cells [3] requires virtually defect-free materials over a considerably large spatial extent. Other limitations arise from the very device principles. SETs need to have capacitances in the aF range in order to be operated at room temperature. Therefore, the switching power to achieve reasonable response times will be unsuitably large for large scale integration [4].

A different concept was put forward by IBM researchers J. J. Welser and S. Tiwari [5]. A quantum dot is embedded in the gate oxide of a field effect transistor (FET). The dot is charged with a small but finite number of electrons and changes the threshold voltage of the FET according to the electron number. The device was already shown to have stable room temperature operation by various groups [6], and hence, is a very promising candidate for further silicon based memory devices.

One of the key requirements for device simulation arising from all these concepts is the ability to model zero dimensional confinement. Here, we present a concept and first results of self-consistent calculations of quantum dots in Si/SiO_2 heterojunctions. The model dot is a somewhat artificial device, however, it already incorporates features that will arise with more realistic structures.

2. MODEL QUANTUM DOT

Our model consists of a 90 nm thick undoped Si layer and a 10 nm thick SiO_2 layer. The lateral dimensions of the supercell are 50 nm × 50 nm with a metallic gate on the bottom covering the whole bottom side. Atop of the structure we modeled another metallic square gate, the size of which was varied between 10 nm × 10 nm and 20 nm × 20 nm. The gate voltage is applied between the top and the bottom gate. The size of the top gate defines the lateral size of the quantum dot in the Si/SiO_2 heterojunction.

3. NUMERICAL METHOD

We used a nonuniform tensor product mesh with Dirichlet boundary conditions at the gates and Neumann boundary conditions elsewhere. The operators are discretized using a finite difference approach. We solve the coupled Kohn-Sham (Schrödinger)-Poisson equation system self-consistently including exchange and correlation in the local density approximation. Weak coupling of the dot to the reservoir (leads) is assumed. This constraint imposes vanishing wavefunctions at the boundaries of the supercell. The six valley bandstructure of silicon is taken into account via the kinetic operator

$$T = \frac{1}{2}\sum_{i,j} w_{ij} p_i p_j, \tag{3.1}$$

where $p_j = -i\hbar(\partial/\partial x_j)$. w_{ij} is the reciprocal effective mass tensor. In terms of the transformation matrix a_{ij} from the principal axes of a constant energy ellipsoid we write

$$p_j = \sum_k a_{jk} p'_k \tag{3.2}$$

$$w_{ij} = \sum_k a_{ik} a_{jk} w'_{kk} = w_{ji} \tag{3.3}$$

where $w'_{kk} = 1/m'_k$ are the principal reciprocal effective masses of the semiconductor.

Total charge densities, total free energies (F), chemical potentials (μ) for different numbers of electrons in the dot, and the differential capacitances (C_d)

$$F(N) = kT \ln \sum_\alpha \exp[-F_\alpha/kT]\delta_{N\alpha}, N \tag{3.4}$$

$$\mu(N) = F(N) - F(N-1) \tag{3.5}$$

$$C_d(N) = \frac{q^2}{\mu(N+1) - \mu(N)} \tag{3.6}$$

are calculated. Using these results, we are able to determine the number of electrons in the dot

$$\bar{N} = \frac{\sum_\alpha N_\alpha \exp[-(F_\alpha - \mu N_\alpha)/kT]}{\sum_\alpha \exp[-(F_\alpha - \mu N_\alpha)/kT]} \tag{3.7}$$

as the thermodynamical average over all configurations of occupation numbers $\alpha = \{n_i\}$. Here, μ denotes the chemical potential in the reservoir (leads).

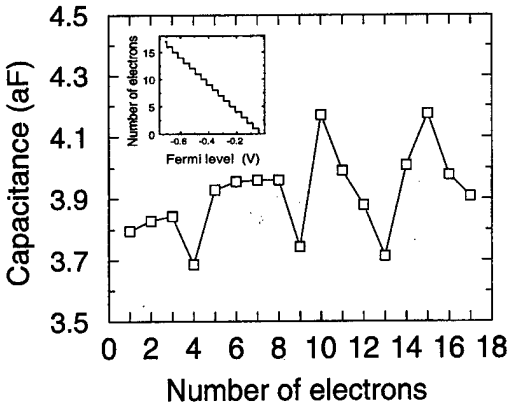


FIGURE 1 Si/SiO_2 quantum dot: differential capacitance vs number of electrons for a gate size of 10 nm × 10 nm. The gate voltage is 2 V. The inset shows a plot of the number of electrons vs the Fermi level in the leads.

4. RESULTS

We assumed the gate to be perpendicular to the ⟨100⟩ substrate orientation. Applying the transformations of equations Eqs. (3.1), (3.2) and (3.3) we obtain three different Hamiltonian operators, which differ in their kinetic parts by different effective masses perpendicular to the interface. However, they contain the same effective potential. The single particle levels are four-fold degenerate. This is in contrast to the spherical total energy surface of GaAs where only spin degeneracy is present. In silicon we have an additional $(-k, +k)$ degeneracy, since two equivalent valleys occur in the $-k$ and the $+k$ directions. All calculations are performed for a low temperature of 4.2 K, neglecting intervalley splitting effects.

The self-consistent single particle potential has shown to be quasi parabolic. Consequently, the single particle eigenvalue spectrum exhibits a shell structure similar to the harmonic oscillator spectrum.

Following the arguments of Macucci *et al.* [7] for GaAs systems, we discuss the differential capacitance curves displayed in Figures 1 and 2. When adding electrons to the dot, the free energy increases almost linearly when adding to the same shell, however, the slope increases from shell to shell. Therefore, a minimum in the capacitance

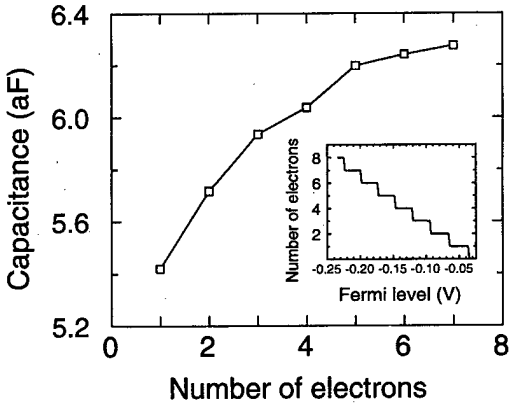


FIGURE 2 Si/SiO_2 quantum dot: differential capacitance vs number of electrons for a gate size of 20 nm × 20 nm. The gate voltage is 2 V. The inset shows a plot of the number of electrons vs the Fermi level in the leads.

occurs, when a shell is filled. In contrast to GaAs, the shell-filling in silicon is governed by the four-fold degenerate single particle levels. Since the confinement in the z-direction, perpendicular to the Si/SiO_2 interface, is stronger than the in-plane confinement due to the top gate, only eigenstates from the Hamiltonian with the longitudinal effective mass in z-direction contribute to the total charge at low electron filling. We observe a minimum of the capacitance at the 4th electron, which is due to the complete filling of the first four-

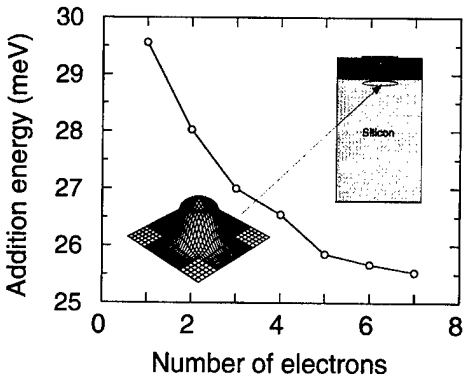


FIGURE 3 Si/SiO_2 quantum dot: addition energy vs number of electrons for a gate size of 20 nm ×20 nm. The inset shows the principal device structure and a plot of the charge density for 12 electrons (cut parallel to the interface).

fold degenerate eigenstate. The minimum is more pronounced with the smaller gate (Fig. 1) which leads to stronger confinement in the x- and y-directions. The larger gate, and therefore, weaker confinement (Fig. 2) leads to an almost linear increase of the capacitance. Shell-filling effects are suppressed, and the dot electrons behave more like a classical circular charge distribution, that increases its capacitance with increasing radius as $C=8\varepsilon_0\varepsilon_r r$. The classical value for the capacitance of this dot is 8 aF ($r=10$ nm, $\varepsilon_r=11.8$), which means a slight overestimation of our value of 6 aF ($N=4$). However, the effective size of the dot underneath the gate is somewhat smaller than the actual gate size.

The simple picture of harmonic oscillator-like shells is not valid when higher electron numbers occur. States from the other Hamiltonians with different symmetries of the wavefunctions then constitute almost arbitrary shells. This is especially evident for a stronger confinement in the x- and y-direction due to a smaller gates. Some shell-filling at the 13th electron is indicated (Fig. 1).

5. CONCLUSIONS

We showed that shell-filling effects occur in silicon quantum dots. Due to the larger effective masses, however, these effects are not as dominant as in GaAs dots of the same size. However, since room temperature operation of quantum dot devices requires a considerable level separation compared to the thermal level broadening, overall smaller device features are needed with silicon and shell-filling effects will dominate again.

Magic-numbers are governed by the six valley bandstructure of silicon. Four-fold degenerate single particle levels are occupied. This again is in contrast to GaAs, where only spin degeneracy is present. The three principal directions of the constant energy surface lead to a system of three different Schrödinger equations, which have to be solved self-consistently with respect to the effective single particle potential. The problem of inter-valley coupling has not been tackled in this paper, however, it may play a role. Further investigation in this problem is needed.

References

[1] Wettstein, A., Schenk, A., Scholze, A., Garreton, G. and Fichtner, W. (1997). "Charge Carrier Quantization Effects in Double-Gated SOI MOSFETs", *Proc. of 6th Int. Symposium on ULSI Science and Technology*.

[2] Haug, R. J. and Klitzing, K. V. (1995). "Prospects for research on Quantum Dots and Single-Electron Transistors", *FED Journal*, **6**, 5.

[3] Tougaw, P. D., Lent, C. S. and Porod, W. (1997). "Bistable saturation in coupled quantum-dot cells", *J. Appl. Phys.*, **74**, 3558.

[4] Mooij, J. E. (1993). Single Electronics: Status and Prospects, "*Proc. of Int. Conf. on Solid State Devices and Materials* ".

[5] Geppert, L., "Solid State: Technology 1997 Analysis and Forecast ", *IEEE Spectrum*, January 1997.

[6] Guo, L., Leobandung, E. and Chou, S. Y. (1997). "A Silicon Single-Electron Transistor Memory Operating at Room Temperature", *Science*, **275**, 649; Nakajima, A., Futatsugi, T., Kosemura, K., Fukano, T. and Yokoyama, N. (1997). "Room temperature operation of Si single-electron memory with self-aligned floating dot gate", *Appl. Phys. Lett.*, **70**, 1742.

[7] Macucci, M., Hess, K. and Iafrate, G. J. (1993). "Electronic energy spectrum and the concept of capacitance in quantum dots", *Phys. Rev. B*, **44**, 17 354.

Authors' Biographies

Andreas Scholze was born in 1969. He received the Dipl. Phys. degree from the University of Jena in

1995. After graduation he joined the Integrated Systems Laboratory of ETH as a research assistant, where he now works with the group of Prof. W. Fichtner in the fields of nanodevice simulation and novel device principles. He is currently working as a visiting research associate at the Beckman Institute for Advanced Science and Technology at the University of Illinois in Urbana-Champaign (UIUC) with the group of Prof. K. Hess.

Andreas Wettstein was born 1969. He received the Dipl. Phys. degree from the Technical University in Karlsruhe in 1995. Since then he is working with the group of Prof. W. Fichtner at the Integrated Systems Laboratory of ETH. His main research interest is on quantum effects in MOSFET inversion layers and general silicon device physics.

Andreas Schenk was born in 1957. He received the Dipl. Phys. degree and the Ph.D. from Humboldt University in Berlin (HUB) in 1981 and 1987, respectively. From 1987 till 1991 he was working on various aspects of the physics and simulation of optoelectronic devices. In 1991 he joined the Integrated Systems Laboratory of ETH working as a senior research/teaching assistant, where he qualified to give lectures at university in 1997 for "Physics and Modeling of Microelectronic Devices". His main activities include physics-based models for advanced simulation of submicron silicon devices and their application in the TCAD software released by ISE AG Zurich.

Wolfgang Fichtner received the Dipl. Ing. degree in physics and the Ph.D. degree in electrical engineering from TU Vienna, Austria, in 1947 and 1978, respectively. From 1979 through 1985, he was member of the technical staff at AT&T Bell Laboratories, Murray Hill, NJ. Since 1985 he is Professor and Head of the Integrated Systems Laboratory at the Swiss Federal Institute of Technology (ETH Zurich). In 1993, he founded ISE Integrated Systems Engineering AG, a company in the field of technology CAD. Wolfgang Fichtner is a Fellow of the IEEE and member of the Swiss National Academy of Engineering.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 237–245
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

An Interband Tunnel Oscillator: Intrinsic Bistability and Hysteresis of Trapped Hole Charge in a Double-Barrier Structure

F. A. BUOT

Naval Research Laboratory, Washington, D. C. 20375-5320

We introduced a novel high-frequency source based on interband tunneling. A polarization-induced oscillation of trapped-hole-charge occurs in an AlGaSb/InAs/AlGaSb resonant tunneling device. Rate equations for Zener tunneling, polarization, and electron-hole recombination is used to analyze the nonlinear dynamics of this device structure. The nonoscillatory state is unstable against the limit-cycle operation. The amplitude of trapped hole oscillation increases with bias, but the time-averaged values can be approximated by a step function. These lead to the hysteresis of the averaged trapped hole charge in AlGaSb barrier, and to the experimental intrinsic bistability in AlGaSb/InAs/AlGaSb resonant tunneling device. Large-scale time-dependent simulation of quantum transport with interband-tunneling dynamics is needed for the design optimization of this novel class of oscillator useful for high-bandwidth applications.

Keywords: Limit cycle, hysteresis, trapped holes, intrinsic bistability, zener effect, resonant tunneling devices, self-oscillation

1. INTRODUCTION

The 'hetero' junction has become the basic building block of most of the advanced high-speed devices for electronic, microwave, and optoelectronic applications [1, 2]. Moreover, tunneling devices exhibit autonomous oscillation, similar to Gunn effect devices but at much higher frequencies in nanometric sizes. For conventional resonant tunneling devices (RTD) [3–6], this occurs when the device is operating in the negative-differential-resistance (NDR) region, just after the resonant current peak. The oscillation addressed in this paper occurs before the resonant current peak, based on interband tunneling in RTD with staggered band-gap alignment.

A staggered band-edge alignment can be realized by using InA/AlSb hetero junctions, Figure 1a. In a simple implementation of a novel interband tunnel high-frequency source, a deeper quantum-well-for-holes is desirable which can support a localized hole state; this is obtained by using In As/AlGaSb heterojunctions, Figure 1b. Unless otherwise specified, quantum well refers to the conduction band edge and conduction-band electrons.

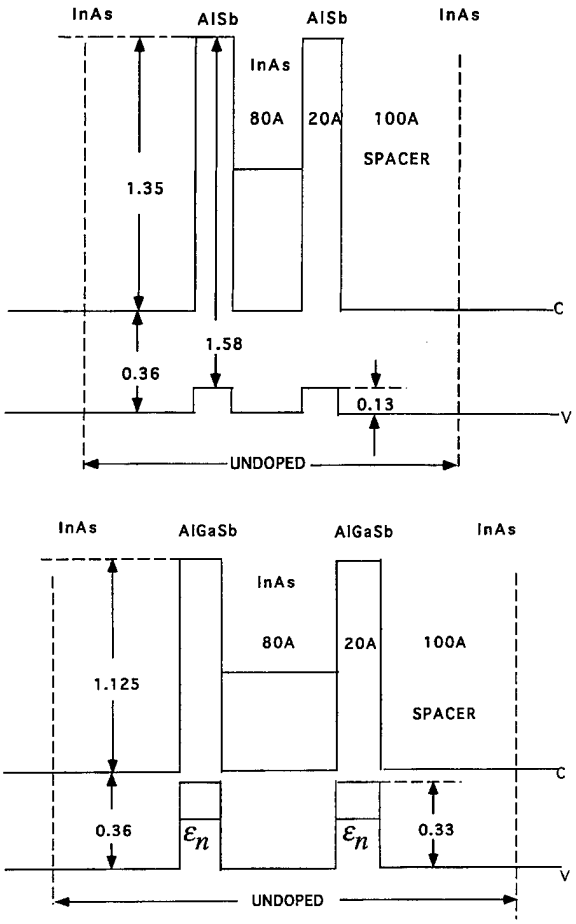


FIGURE 1 (a) Energy band edge (EBE) alignment of RTD using InAs/AlSb heterojunction. (b) Energy band edge alignment of RTD using InAs/AlGaSb heterojunction. Approximated band-edge offsets are indicated in electron volts.

The new mechanism of modulating the resonant energy level in the quantum well with respect to the energy distribution of the electrons from the emitter can simply be described through the oscillatory build-up and decay of the polarization pairing between electrons in the quantum well and trapped holes in the barrier. This modulation is controlled by trapped holes (similar to base charges of a bipolar transistor). Thus, for the first time we realize an autonomous control of a significant current by an interband process.

We refer to the polarization pair as a *duon*, since this Coulomb-correlated e–h pair only moves in the transverse direction, in contrast to *exciton*. This is in analogy to the use of *trion* in referring to a correlated exciton and electron in adjacent quantum-well heterostructure [11]. In Sec. 3, we introduce the physics of the *duon* dynamics. The limit cycle solution leads to an oscillatory voltage drop between the quantum well and the barrier. Since common experimental techniques are incapable of investigating these oscillations [3, 4, 8], the current-voltage ($I-V$) characteristic is also calculated in Sec. 4. The results agree with the experiment. In Sec. 5, we draw some conclusions, as well as give a summary of this paper.

2. HIGH-FREQUENCY OPERATION

Under bias, when the localized electrons in the AlGaSb barrier of Figure 1b see the available states these electrons tunnel to the drain by Zener transition, leaving behind holes in a discrete 'longitudinal' energy level, ε_n. This is initiated when a matching of ε_n with available conduction-band states in the drain first occurs, at $k_z^{D^2} \approx k_F^{D^2}$ in Figure 2. Figure 2 serves to define several quantities used in the calculations of Sec. 4. The drain acts as a sink due to unoccupied states above k_F^D that could satisfy the conservation of transverse crystal momentum associated with ε_n.

As hole charging occurs, Figure 3 (1), the polarization Figure 3 (2) creates a high-field domain, at the expense of the potential drop between the contact and the barrier. When the situation shown in Figure 3 (3) is reached, the onset of other mechanisms for hole discharging may also occur, namely, thermal activation of the valence electrons in the continuum to recombine with localized holes, or loss of any bound hole states in the barrier. The 'hole leakage' will restore the large voltage drop between the barrier edge and the right contact. The situation shown in Figure 3 (1) is revisited, after which the process repeats. Therefore, oscillations of the hole charging of the AlGaSb barrier can occur at high frequency by virtue of the nanometric features of

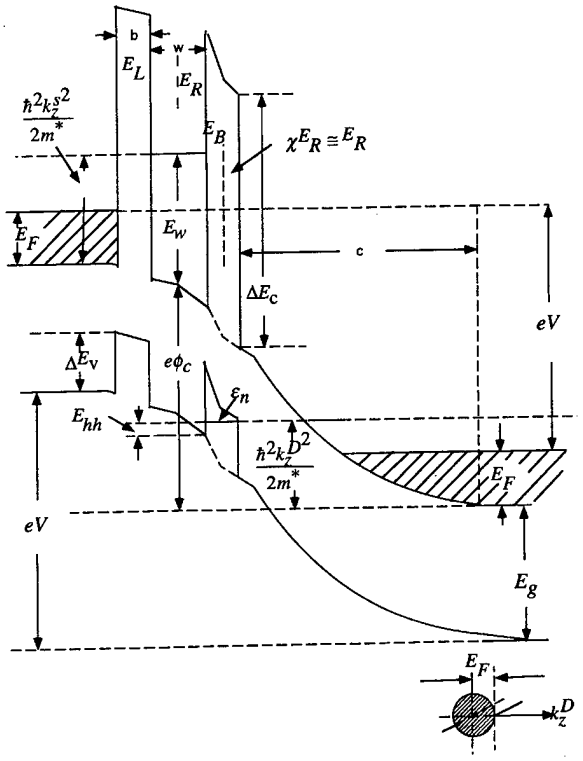


FIGURE 2 Schematic averaged EBE profile showing the various quantities used in the calculations of $I-V$ plot. The shaded regions in the lower right-hand corner indicate the occupied transverse and longitudinal momentum states in the drain.

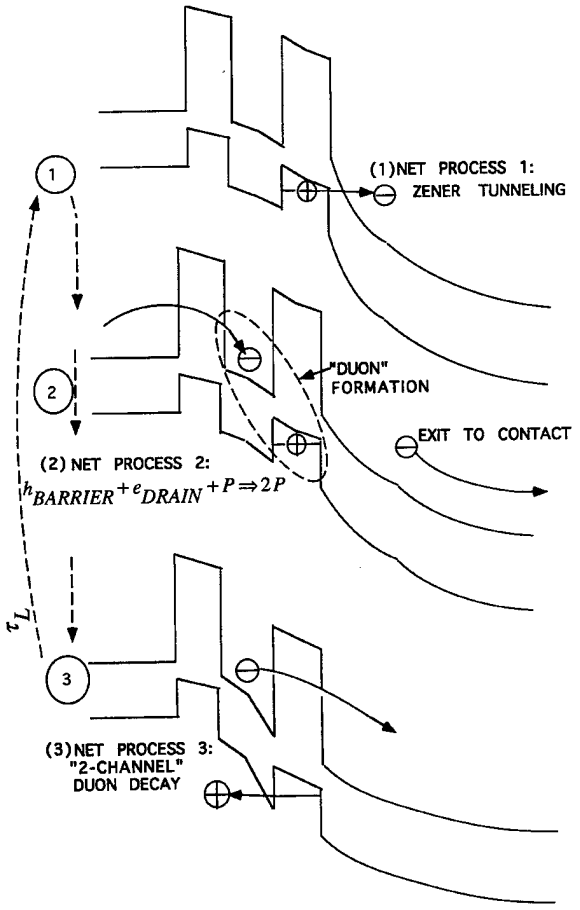


FIGURE 3 (1) e–h generation by Zener tunneling. (2) *duon* generation is through an *autocatalytic* process. (3) three mechanisms for hole discharging are mentioned in the text.

the device. We estimate the charging time, which is the dominant time scale of the problem, to be about 200–1000 femtoseconds for a heavy-hole state in AlGaSb with indirect-gap interband tunneling through Keldysh effect.

The above dynamical process limits the amount of hole charge that can be trapped in the barrier as a function of bias. The interband recombination process can not compete with the conduction-band electron tunneling process through the barrier [3]. The criteria for either detailed balance or oscillatory behavior are governed by the two characteristic times, namely, the polarization-charge build-up time, τ_B, and the charge-leakage time, τ_L. If $\tau_B > \tau_L$, then the charging process will be lagging behind the discharging process and oscillations will result. This holds true in conventional RTD and single-electron devices [10]. It is here estimated that $\tau_B > \tau_L$, by virtue of several possible fast hole-discharge channels mentioned above. The 'book-keeping equations' ('chemical kinetics' modeling) of the device operation follows.

3. POLARIZATION PAIRING DYNAMICS

We derive here the coupled rate equations for the processes in Figure 3. The duon formation is via the polarization-induced transport of conduction electrons from the emitter to the quantum well,

coupled with a succeeding e–h generation by Zener tunneling. In process (2), the Zener-tunneled electron flows to the metallic contact of the drain, and this is substituted by the tunneled conduction electron from the emitter to form a *duon*. We observe here an 'autocatalytic' or positive-feedback aspect of the *duon* formation.

Let G be the maximum rate of e–h generation by Zener tunneling. Note that G is a direct measure of the applied bias. The duon generation rate with 'three interacting components' can be expressed as $\tilde{\Delta}\mathcal{N}_B^2 P$, where $\mathcal{N}_B$ is the concentration of 'unpaired' holes which is equal to the concentration of 'exiting' electrons produced by Zener tunneling, P is the concentration of *duons*, and $\tilde{\Delta}$ is a parameter which is expected to acquire, in appropriate ranges, a nonlinear dependence on P as discussed below. We can now write the 'effective' generation rate of unpaired trapped holes in the barrier as

$$\frac{\partial}{\partial t}\mathcal{N}_B = G - \tilde{\Delta}\mathcal{N}_B^2 P. \tag{1}$$

The total concentration of trapped holes in the barrier, Q_B, at any time is given by $Q_B = \mathcal{N}_B + P$.

For large P the transfer of conduction electrons from the emitter to the quantum well becomes more efficient since it is approaching the resonance peak and $\tilde{\Delta}$ increases. For very small P, the concentration of resident electrons already existing in the quantum well (refer to Fig. 4) will also render the polarization pairing to be much more efficient, and hence a larger $\tilde{\Delta}$, than for the intervening ranges between small P and large P. This acquired nonlinearity of the parameter $\tilde{\Delta}$ is important in establishing a limit cycle operation of the device, since it limits the growth of the solution from the unstable focus. Indeed, we shall see that the 'linear' criterion for unstable stationary operation is that $F(G) > 4\tilde{\Delta}$ where $F(G)$ is a constant for a fixed bias.

The decay rate for P is expected to saturate for very large P. In all foreseeable cases, we may express the decay rate of the *duon* concentration as

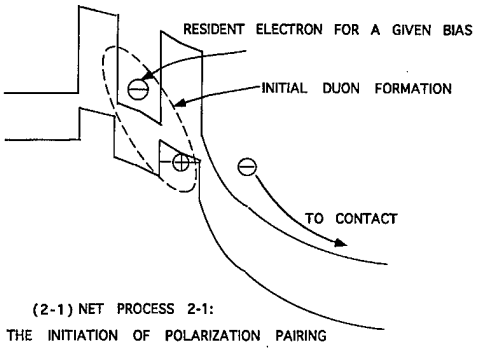


FIGURE 4 More efficient generation of initial polarization pairs or *duons* is due to the initial resident excess electron in the quantum well for a given applied bias.

$= \alpha P/(1+\beta P)$, where $1/\beta$ represents the sum of available states in the valence band of the quantum well region and the states participating in thermal recombination; otherwise it represents the actual concentration of hole states in the barrier in the case of the loss of bound hole state. This is similar to the decay law ubiquitous in chemical kinetics [12]. The parameter α is the decay rate constant and α/β is the value of the saturated decay rate of *duons*. Therefore, we can now write the *duon* generation rate as

$$\frac{\partial}{\partial t}P = \tilde{\Delta}\mathcal{N}_B^2 P - \frac{\alpha P}{1+\beta P}. \tag{2}$$

As seen in Eq. (4) below, the physical situation corresponds to $\alpha/\beta > G$. Indeed, we can estimate that $1/\tau_B \approx G \approx G$ and $1/\tau_L \approx \alpha/\beta$. Therefore $\alpha/\beta > G$ implies that $\tau_B > \tau_L$.

3.1. Stability Analysis

The stationary solution for a fixed bias (implying a fixed G) to Eqs. (1) and (2) is given by

$$G = \tilde{\Delta}\mathcal{N}_B^2 P = \frac{\alpha P}{1+\beta P}. \tag{3}$$

The total stationary trapped hole concentration, Q_B, is thus given by

$$Q_B = \mathcal{N}_B^0 + P^0 = \frac{G}{\alpha - \beta G} + \left(\frac{\alpha - \beta G}{\tilde{\Delta}}\right)^{1/2}, \quad (4)$$

which is a sum of an increasing and a decreasing function of bias. The more accurate average value under a limit cycle oscillation is shown in Sec. 3.2 to be approximately independent of bias. Since Q_B and P^0 are constants, the *duon* production rate is via the transfer of conduction electrons from the emitter to the quantum well and the *duon* decay rate is via transfer of conduction electrons from the quantum well to the drain. Thus, the d.c. operation no longer involves interband processes, as schematically shown in Figure 5.

Complete analysis of the stability of the stationary solution as well as the full derivation of the limit cycle solution will be given elsewhere. By transforming to dimensionless variables: $\Pi = \beta P$, $\mathcal{Q} = \beta \mathcal{N}_B$, $\Delta = \tilde{\Delta}(1/\beta)^2/\alpha$, $\mathcal{G} = G/(\alpha/\beta) < 1.0$, the criterion for unstable stationary solution is that $(1 - \mathcal{G})^3 > 4\Delta$. Following a perturbation technique using multiple time scales [13, 14], we have obtained to second order in the smallness parameter, $\varepsilon = (\{\Delta - \Delta_c\}/\Delta_2)^{1/2}$, where $\Delta_c = ((1 - \mathcal{G})^3/4)$ and Δ_2 comes from the expansion of Δ near Δ_c in powers of ε, the limit cycle solution. This is given as

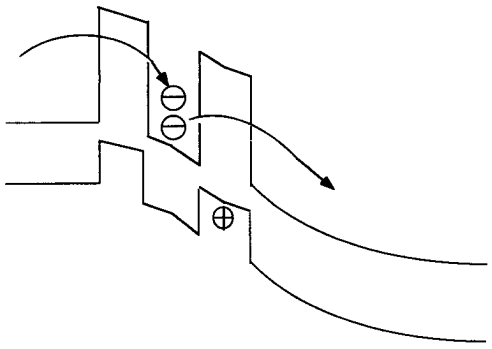


FIGURE 5 At steady state the two conduction-electron-mediated *duon* decay and generation processes are balanced resulting in steady-state current across the double barrier structure.

$$\begin{pmatrix} \Pi \\ \mathcal{Q} \end{pmatrix} = \begin{pmatrix} \Pi^0 \\ \mathcal{Q}^0 \end{pmatrix} + \varepsilon \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} + \varepsilon^2 \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} + O(\varepsilon^3), \quad (5)$$

The two column vectors $\binom{P_0}{q_0}$ and $\binom{P_1}{q_1}$ contain the oscillating factors associated with first and second orders, respectively. We note that $\binom{P_1}{q_1}$ and higher-order terms also contain higher-order time-independent terms which increases with bias. Therefore, the average value of $\binom{\pi}{\mathcal{Q}}$ is given by

$$\begin{pmatrix} \pi \\ \mathcal{Q} \end{pmatrix}_{\text{average}} = \begin{pmatrix} \pi^0 \\ \mathcal{Q}^0 \end{pmatrix} + \text{higher-order corrections}, \quad (6)$$

where the leading higher-order corrections comes from these time-independent terms.

3.2. Dependence with Bias

Based on a finite limit of the amplitude (which only depends on the slow time scale), the limit cycle solution is found to occur within the range of values of the parameter Δ where the 'linear' criterion for unstable focus $(1 - \mathcal{G})^3 > 4\Delta$ still holds. This is analogous to the numerically simulated limit cycle of AlGaAs/GaAs/AlGaAs double-barrier heterostructure operating in the NDR region [7]. For values of Δ near the critical point, the rate of change with respect to $\mathcal{G}$ is approximately zero. Indeed, we have from Eq. (4),

$$\frac{d\beta Q_B(\mathcal{G})}{d\mathcal{G}} = \frac{1}{(1 - \mathcal{G})^2} - \left(\frac{1}{4\Delta(1 - \mathcal{G})}\right)^{1/2} \leq 0, \quad (7)$$

where the equality is obtained at $\Delta = \Delta_c = (1 - \mathcal{G})^3/4$. Since $\mathcal{G}$ is our measure of the applied voltage applied, we conclude from Eq. (7) and by taking into account the higher-order correction terms which increases with $\mathcal{G}$ that the average total hole charge trapped in the barrier is approximately independent of bias.

Denoting the leading time-dependent part of $Q_B(t)$ as $\delta Q_B(t)$, the total trapped hole charge in the barrier oscillates with amplitude that increases with bias and is given by

$$\delta Q_B(t) = \frac{4(|\Delta - \Delta_c|)^{1/2}}{\beta} \left[\frac{\mathscr{G}}{(1-\mathscr{G})^4 |\{-8-17\mathscr{G}+36\mathscr{G}^2-11\mathscr{G}^3\}|}\right]^{1/2} [\cos\Omega t - \sin(\Omega t + \Phi)] \tag{8}$$

where the frequency Ω consists of a function of $\mathscr{G}$ plus higher-order terms, and $\Phi = \{\mathscr{G}/(1-\mathscr{G})\}^{1/2} > 0$. The oscillation amplitude grows in response to the increasing maximum electric field in the depletion region with the applied bias, since the maximum e–h generation rate, $\mathscr{G}$, by Zener tunneling [15] increases with bias.

4. INTRINSIC BISTABILITY IN InAs/$Al_xGa_{1-x}Sb$ RTD

The time-averaged hole charge in the barrier is referred to as $Q(\text{AlGaSb}) = eQ_B$, where e is the positive unit charge. This value is approximately independent of bias in Sec. 3, after an abrupt increase at $k_z^{D^2} \approx k_F^{D^2}$ in Figure 2. The self-consistency of the potential alone, in Figure 2, demands that the polarization and hence Q(AlGaSb) increases monotonically with bias. We shall see that the simultaneous solution to these two requirements, plus the continuity condition, leads to a 'parallelepiped' hysteresis of the trapped hole charge in the barrier.

We only need three field parameters to include a concave EBE profile in the barrier region. The inflection point is assumed to have a measure zero as far as the integration of the fields to obtain the total voltage drop across the device is concerned. For nonzero average value of $\mathscr{N}_B$, which is the concentration of "unpaired" trapped holes, we also expect a nonzero superposed polarization between the barrier and spacer layer to be affecting the potential profile, as indicated by a simple 'kink' in the spacer region of Figure 2. We estimate the fourth field parameter in the second half of the barrier as proportional to E_R, as the figure suggests with proportionality factor, $\chi(V) \leq 1.0$, and still maintain the physical requirement of concave EBE profile in this region. This accounts for nonzero average $\mathscr{N}_B$.

From Figure 2, the trapped hole charge in the second barrier is given by the expression: $\chi E_R - E_B = Q(\text{AlGaSb})/\varepsilon$, from the Poisson equation. We estimate the proportionality factor χ is close to unity and positive. From the requirement of faster voltage drop in the barrier region in Figure 2, we must have E_B more negative than χE_R, thus we obtain $Q(\text{AlGaSb}) > 0$ consistent with the trapped hole charge in the second barrier.

4.1. Hysteresis of Trapped Hole Charge

The positive applied bias, V, in Figure 2, is given by the following expression, $eV = e|E_L|(b + w/2) + e|E_R|(w/2) + E_g + (\hbar^2 k_z^{D^2}/2m^*) - E_{hh}$, where w and b are the width of the quantum well and barrier, respectively. We use the Poisson equation to eliminate $|E_R|$ in terms of $|E_L|$. Since all fields on the average have negative sign for positive applied voltage, we may also write Poisson equation as $|E_R| - |E_L| = |Q_w|/\varepsilon$, $|E_B| - \chi|E_R| = Q(\text{AlGa Sb})/\varepsilon$. Therefore, we obtain the following expression for the trapped hole charge,

$$Q(\text{AlGaSb}) = \frac{2\varepsilon\chi}{ew}\left[(E_g - E_{hh}) - \left(eV - e|E_L|\left(b + \frac{w}{2}\right) - e|E_B|\frac{w}{2\chi}\right) + \frac{\hbar^2 k_z^{D^2}}{2m^*}\right]. \tag{9}$$

The trapped hole charge is an increasing function of $k_z^{D^2}$ in Eq. (9), since the 'polarization',

without the constraint of quantum transport nonlinear dynamics, should increase monotonically with bias.

The independence with bias beyond a threshold value of the trapped holes in Sec. 3 is expressed here by a step function

$$Q(\text{AlGaSb}) = Q_h \Theta(k_z^{D^2} - k_F^{D^2}). \qquad (10)$$

The simultaneous solutions of Eqs. (9) and (10) is shown graphically in Figure 6, Eq. (9) for $Q(\text{AlGaSb})$ vs. $k_z^{D^2}$ is approximated by positive sloping lines. Upon applying the continuity condition, open circles and solid circles are solutions for the increasing voltage sweep and decreasing voltage sweep, respectively. A 'parallelepiped' hysteresis of trapped hole charge as a function of bias is clearly indicated.

4.2. Hysteresis in the $I-V$ Characteristics

This is obtained by describing the whole length of the device by three independent fields, namely, E_L, E'_R and E'_B. Note that the field χE_R used before is only valid in the right-half of the barrier region, by virtue of nonzero average concentration of 'unpaired' trapped holes, $\mathcal{N}_B$, as indicated in Figure 2. The field E_L is as defined before, whereas E'_R is defined by the relation: $E'_R(b/2 + w/2) = E_R(w/2) + E_B(b/2)$, and E'_B is the constant field approximation for the rest of the device of dimension $[(b/2) + c]$. As a consequence, we also have the following relation: $|E'_R| - |E_L| = \alpha' |Q_w|/\varepsilon$ and $|E'_R| - |E'_B| = \beta' |Q(\text{AlGaSb})/\varepsilon$, where α' and β' proportionality constants. Thus we may write $eV = e|E_L|(b + w/2) + e|E'_R|(w+b)/2 + e|E'_B|(b/2+c)$. We then expressed $|E'_R|$ and $|E'_B|$ in terms of $|Q_w|$, $Q(\text{AlGaSb})$, and $|E_R|$. We also use the relation: $e|E^L| = e(V - \phi_c)/(2b + w)$, where $e\phi_c = eV - E_w + \hbar^2 k_z^{s^2}/2m^*$ to obtain the result

$$|Q_w| = \frac{2\varepsilon}{\alpha'\left((b+c) + \frac{w}{2}\right)} \left\{ \frac{\hbar^2 k_z^{s^2}}{2m^*} - E_w + \frac{V}{2\xi} + \frac{\beta' Q(\text{AlGaSb})}{2\xi\left(\frac{\varepsilon}{b/2+c}\right)} \right\}, \qquad (11)$$

where we have $\xi = (2b + w + c)/(2b + w)$. The quantum transport requirement for Q_w was given by Buot and Rajagopal [9,10] as

$$|Q_w| = \frac{em^*}{\pi\hbar^2\beta} \ln\left\{1 + \exp\beta\left(E_F - \frac{\hbar^2 k_z^{s^2}}{2m^*}\right)\right\} \left(\frac{\tau_d}{\tau_e}\right) \Theta(k_z^2), \qquad (12)$$

which is zero for $k_z^{s^2} \leq 0$, $1/\tau_d = 1/\tau_e + 1/\tau_c$, where $1/\tau_c$ is the effective rate of decay of Q_w into unoccupied collector states and $1/\tau_e$ is equal to the rate of supply of electrons from the emitter to the quantum well. The simultaneous solution of Eqs. (11) and (12) is also graphically obtained as shown in Figure 7. In Figure 7a, Eq. (11) is approximated by parallel sloping lines. The values for $Q(\text{AlGaSb})$ are solutions obtained from Figure 6 which create an offset in the sloping lines of Eq. (11), leading to higher values of Q_w as indicated by the dotted arrows. For the increasing voltage sweep, the solutions for Q_w are given by the intersection points $S_1, S_2, S_3^F, S_4^F, S_5^F, S_6^F(\text{low}), S_6(\text{high}), S_7$ and S_8. For the reverse voltage sweep, the corresponding solution points are $S_8, S_7, S_6, S_5^R, S_4^R, S_3^R, S_2^R$ [high], S_2[low], and back to S_1.

The RTD current can be approximated by Q_w/τ_c. The resulting $I-V$ has all the salient features of

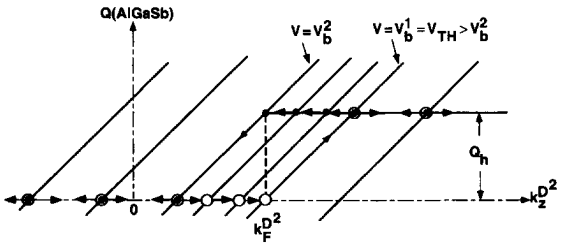


FIGURE 6 Graphical solution of Eqs. (9) and (10).

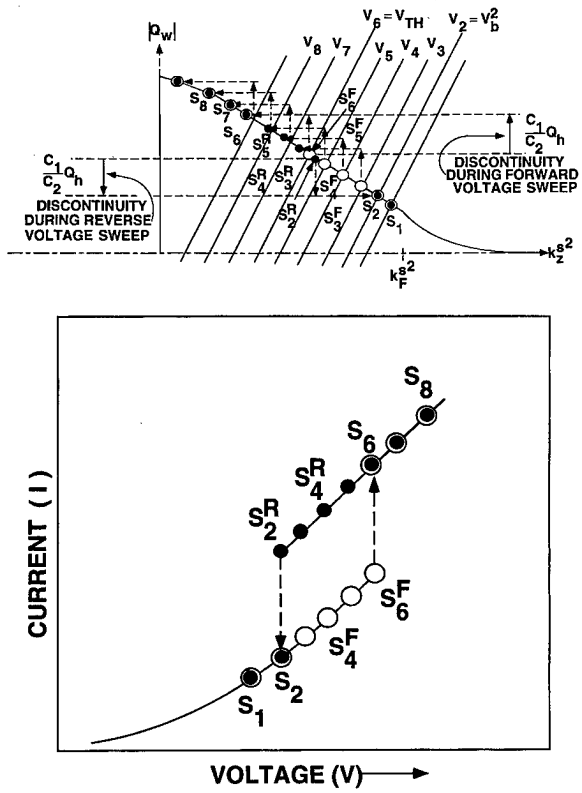


FIGURE 7 (a) Graphical solution of Eqs. (11) and (12). (b) The solution for the $I-V$ characteristic showing 'parallelepiped' hysteresis occurring before the RTD current peak, in agreement with the experiment.

the experimental results [9], shown in Figure 7b with corresponding solution points indicated. This result serves as indirect experimental evidence on the ability of this device to function as a high-frequency source.

5. CONCLUSIONS

The nonlinear dynamics of coupled systems of *duons* and 'unpaired' trapped-holes in RTD with staggered band-gap alignment provides a fundamental physical basis of the experimental intrinsic bistability of AlGaSb/InAs/AlGaSb RTD. Zener tunneling, stimulated formation of *duons*, and the self-oscillation of the trapped hole charge provide the mechanism behind a new high-frequency source introduced in this paper. We note that the RTD current peak occurs when the quantum-well energy level aligns with the bottom of the conduction band of the emitter. In some emitter designs, the Fermi level adjacent to the barrier increases as this alignment is approached, leading to a high-transconductance with the high-field domain acting as a self-gate. Thus, the transconductance of this 'self-gated'-transistor oscillator can be made large, yielding a novel high-frequency source with a usable power.

The 'chemical kinetics' modeling used above to analyze the device operation adds to the store of analytical tools for characterizing complex devices. However, large-scale time-dependent simulation is needed for further research and design optimization. Besides potential applications in communications, defect engineering would also make it as a triggering element in semiconductor lasers.

Acknowledgements

The author is thankful to Dr. A. K. Rajagopal for helpful comments on the manuscript. He gratefully acknowledge partial support from the Office of Naval Research.

References

[1] Yu, E. T., McCaldin, J. O. and McGill, T. C. (1992). In *Solid-State Physics, Advances in Research and Applications*, San Diego: Academic Press.

[2] Ting, D. Z. Y., Yu, E. T. and McGill, T. C. (1992). "Multiband treatment of quantum transport in interband tunnel devices", *Phys. Rev.*, **B45**, 3583–3592.

[3] Jensen, K. L. and Buot, F. A. (1991). "Numerical simulation of intrinsic bistability and high-frequency current oscillations in resonant tunneling structures", *Phys. Rev. Lett.*, **66**, 1078–108.

[4] Biegel, B. A. and Plummer, J. D. (1996). "Comparison of self-consistency iteration options for Wigner function method of quantum device simulation", *Phys. Rev.*, **B54**, 8070–8082.

[5] Buot, F. A. and Rajagopal, A. K. (1993). "High-frequency behavior of quantum-based devices: equivalent circuit,

nonperturbative response, and phase-space analyses", *Phys. Rev.*, **48**, 17217–17232.
[6] Buot, F. A. and Jensen, K. L. (1991). "Intrinsic high-frequency oscillations and equivalent circuit model in the negative differential resistance region of resonant tunneling devices", *Int. J. Comp. Math. Elec. Electron. Eng.*, COMPEL **10**, 241–253.
[7] Woolard, D. L., Buot, F. A., Rhodes, D. L., Lu, X. L., Lux, R. A. and Perlman, B. S. (1996). "On the different roles of hysteresis and intrinsic oscillations in resonant tunneling structures", *J. Appl. Phys.*, **79**, 1515–1525.
[8] Sollner, T. C. L. G., Goodhue, W. D., Tannewald, P. E., Parker, C. D. and Peck, D. D. (1983). "Resonant tunneling through quantum wells at frequencies up to 2.5 THz", *Appl. Phys. Lett.*, **43**, 588–590.
[9] Buot, F. A. and Rajagopal, A. K. (1994). "Hysteresis of trapped charge in AlGaSb barrier as a mechanism for the current bistability in AlGaSb/InAs/AlGaSb double-barrier structures", *Appl. Phys. Lett.*, **64**, 2994–29964; 1994 experimental measurement of $I-V$ was reported by D. H. Chow and J. N. Schulman, "Intrinsic bistability in In As/$Al_xGa_{1-x}Sb$ resonant tunneling devices", *Appl. Phys. Lett.*, **64**, 76–78.
[10] Buot, F. A. and Rajagopal, A. K. (1995). "Theory of novel nonlinear quantum transport effects in resonant tunneling structures", *Mat. Sci. Eng.*, **B35**, 303–317.
[11] Yoon, H. W. and Pfeiffer, L. N. (1996). "Spatial transport of free trions in mixed-type GaAs/AlAs quantum wells", *Bull. Am. Phys. Soc.*, **41**, 239.
[12] See e.g., Paddeu, S., Erokhin, V. and Nicolini, C. (1996). *Thin Sol. Films*, **284–285**, 854.
[13] Pimpale, A., Landsberg, P. T., Bonilla, L. L. and Velarde, M. G. (1981). "Limit cycle in a bound exciton recombination model in non-equilibrium semiconductors", *J. Phys. Chem. Solids.*, **42**, 873–881.
[14] Nayfeh, A. H. (1981). *Introduction to Perturbation Techniques*, NY: John Wiley.
[15] Zener, C. (1934). "Non-adiabatic crossing of energy levels", *Proc. Roy. Soc. London*, **A137**, 696–702, 1932. See also C. Zener, "A theory of dielectric breakdown of solid dielectrics", *Proc. Roy. Soc. London*, **A145**, 523–529.

Author Biography

Dr. Felix A. Buot has served on the research staff of the University of London, ICTP (Trieste, Italy), McGill, St. Francis Xavier, Stanford, and Cornell University. He is a Fellow of the Washington Academy of Sciences, and President of the Philippine-American Academy of Science and Engineering. He is a member of the Editorial Board, *Transport Theory and Statistical Physics*. He was a UNDP Consultant, 1993 and 1996. His interests include device performance and reliability, optoelectronics, nonequilibrium quantum theory, multiband dynamics, and physics of computation. He is a Research Physicist at the U.S. Naval Research Laboratory.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 247–252
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Tunneling Between Multimode Stacked Quantum Wires

M. MACUCCI[a,*], A. T. GALICK[b] and U. RAVAIOLI[b]

[a] *Dipartimento di Ingegneria dell'Informazione, Via Diotisalvi, 2 I-56126 Pisa, Italy;*
[b] *Beckman Institute, 405 N Mathews Urbana, IL 61801*

Tunneling between vertically stacked quantum wires has been investigated. The wires are assumed to have the dimension perpendicular to the tunneling barrier much smaller than the other transverse dimension, so that only the lowest mode in such direction is to be taken into account, while many modes in the other direction are filled. A model with hard-wall confinement has been used for the investigation of the transport problem, and the tunneling conductance has been computed, via a recursive Green's functions procedure.

Keywords: Quantum wire, tunneling, stacked 2DEGs, green's functions

1. INTRODUCTION

The possibility of fabricating heterostructures containing two weakly coupled and independently contacted 2DEGs (two-dimensional electron gases) has originated several proposals for possible devices exploiting controlled tunneling between such 2DEGs. The basic idea, proposed by Sakakibara *et al.*[1], consists in fabricating two stacked 2DEGs with different thicknesses, so that in normal conditions the ground states for the two quantum wells are not aligned. By applying a potential to a metallic gate obtained on top of the heterostructure, it is possible to alter the potential landscape and to line up the two ground states, so that tunneling will be possible. If we want to observe interference effects dependent on the length of the metallic gate, the situation is, however, more complex than what is discussed in Ref.[1]. While in a quantum wire the number of transverse modes is limited or we can even achieve monomode propagation, in a 2DEG the number of transverse modes is not limited and the continuous distribution of wavevectors may tend to wash out interference effects. In particular, the dependence of the tunneling conductance on the gate length will differ from that reported in Ref.[1] whenever the number of propagating modes is significantly larger than one.

This happens because each pair of modes contributing to conductance will be associated with a different transfer length, corresponding to

* Corresponding author.

the length over which complete transfer from one waveguide to the other occurs. Therefore, as a function of the coupling length, the tunneling conductance will not exhibit a periodic behavior, but, rather, a superposition of different periodicities. In the limiting case of two laterally unconfined 2DEGs the number of occupied transverse subbands is practically infinite and any periodic pattern as a function of gate length should disappear. However, if the transverse dimension is limited by etching, thus defining two stacked quantum wires instead of two stacked 2DEGs, it is possible to limit the number of transverse modes contributing to conduction, and therefore to preserve some sort of periodic or quasi-periodic behavior. This is the case investigated in the present paper: we study the tunneling conductance between two quantum wires with rectangular cross-section (with a large aspect ratio, greater than 10) separated by a barrier along the longer side.

2. MODEL

Our model consists of two vertically stacked quantum wires with a rectangular cross section, as shown in Figure 1. At the outer ends, the two quantum wires are separated by hard walls and are therefore completely independent, so that the electron wave functions are completely localized in each of them, while in the "coupling region" they are separated by a tunneling barrier of finite height. The central section of the coupling region corresponds to the area covered by the metallic gate used to shift the potentials of the two quantum wires. This section is connected to the outer sections by means of a region with graded potentials, with a length corresponding to that over which the gate potential is effectively screened. In order to reduce the computer time used for the simulations and considering that most of the relevant physics is preserved unchanged, we have included only one intermediate section between the central region and the outer ends, with a potential landscape corresponding to an average of those relative to the adjacent sections. The gate bias is assumed to tilt the potential linearly along the vertical direction. In order to solve the transport problem along this structure, we use the quasi-3D approach that we developed for the investigation of electron waveguide couplers [2, 3]: the structure is divided, along the longitudinal direction, into a number of slices, each of which has a constant potential profile. The Schrödinger equation is solved within each slice and the eigenvalues and eigenfunctions thus obtained are used to evaluate the tunneling conductance by means of a recursive Green's function approach [2].

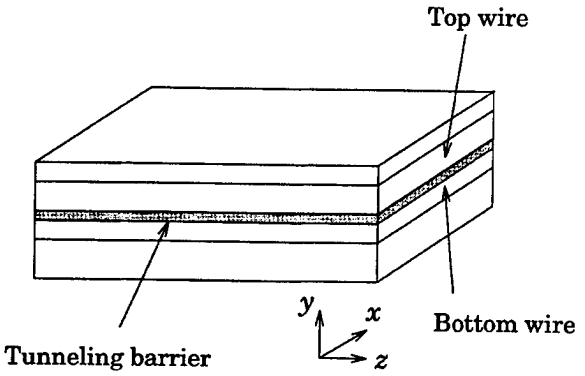


FIGURE 1 Sketch of the stacked quantum wire structure.

In the case of the present structure we are faced with an additional problem: the effective mass in the region corresponding to the barrier dividing the two quantum wells is different from that in the wells themselves. Let us assume a reference system with y along the vertical direction, perpendicular to the tunneling barrier, z along the longitudinal direction, parallel to current flow, and x along the horizontal transverse direction (see Fig.1). The Schrödinger equation in each slice reads:

$$\frac{-\hbar^2}{2m^*(y)}\frac{\partial^2\psi}{\partial x^2} - \frac{\hbar^2}{2}\frac{\partial}{\partial y}\left[\frac{1}{m^*(y)}\frac{\partial\psi}{\partial y}\right] - \frac{\hbar^2}{2m^*(y)}\frac{\partial^2\psi}{\partial z^2} + V(x,y)\psi = E\psi. \tag{1}$$

We can try a solution of the type $\psi = \chi(x, y)$ exp (ikz) (more generally it would be $\psi = \chi(x, y)$ $[A \exp(ikz) + B \exp(-ikz)]$): the transverse equation (on the $x-y$ plane) can then be written as

$$\frac{-\hbar^2}{2m^*(y)}\frac{\partial^2\chi(x,y)}{\partial x^2} - \frac{\hbar^2}{2}\frac{\partial}{\partial y}\left[\frac{1}{m^*(y)}\frac{\partial\chi(x,y)}{\partial y}\right] - \frac{\hbar^2}{2m^*(y)}(-k^2)\chi(x,y) + V(x,y)\chi(x,y) = E\chi(x,y). \quad (2)$$

If we define $\Delta m^*(y) = m^*(y) - m_0^*$, where m_0^* is the effective mass in the quantum wells, and $E_z = \hbar^2/(2m_0^*)k^2$, we get:

$$\frac{-\hbar^2}{2m^*(y)}\frac{\partial^2\chi(x,y)}{\partial x^2} - \frac{\hbar^2}{2}\frac{\partial}{\partial y}\left[\frac{1}{m^*(y)}\frac{\partial\chi(x,y)}{\partial y}\right] - E_z\frac{\Delta m^*(y)}{m_0^* + \Delta m^*(y)}\chi(x,y) + V(x,y)\chi(x,y) = (E - E_z)\chi(x,y). \quad (3)$$

We notice that if the term $E_z\Delta m^*(y)/((m_0^*) + \Delta m^*(y))$ can be neglected compared to $V(x, y)$, this is just a standard 2D Schrödinger equation with variable effective mass. Such term is nonzero only within the barrier, where $\Delta m^*(y)/(m_0^* + \Delta m^*(y))$ is, however, always smaller than 0.5 (0.44 for GaInAs/AlInAs and 0.264 for GaAs/AlGaAs). Furthermore it is clearly $E_z < E_f$, where E_f is the Fermi level. Thus, as long as we choose values of the Fermi level that are less than one half of the barrier height, the transverse problem can be reduced to that of solving a conventional 2D Schrödinger equation, without introducing too large an error.

3. NUMERICAL PROCEDURE AND RESULTS

The Schrödinger equation in each transverse slice is discretized on a uniform finite-difference mesh, and the eigenvalue problem is solved with an efficient iterative Chebishev-Arnoldi technique. The typical size of the mesh varies between 20 × 50 and 20 × 350 points, and the lowest eigenvalues up to the 50th are evaluated. The presence of a varying effective mass along the y direction is treated with the standard formalism ensuring the continuity of the wave function and of its normal derivative at the interface [4], as shown in Eq. (1). The transmission coefficients between the two waveguides are then computed with the modified recursive Green's function method [3], and finally averaged over the derivative of the Fermi function, in order to take into account the effects of temperature [5] (all of our calculations have been performed for a temperature of 1.6 K).

We have taken into consideration the parameter values for two different material systems: a) GaInAs/AlInAs and b) GaAs/AlGaAs. For material system a) we have considered an effective mass $m^* = 0.042\ m_0$ in GaInAs and $m^* + \Delta m^* = 0.075\ m_0$ in the barrier region, m_0 being the mass of the free electron. A value of 0.5 eV has been chosen for the conduction band discontinuity. For material system b) an effective mass $m^* = 0.067\ m_0$ has been considered in GaAs and $m^* + \Delta m^* = 0.091\ m_0$ in the barrier region. The conduction band discontinuity has been assumed to be 0.3 eV.

In all the cases we have studied, the thicknesses of the heterostructure layers have been kept constant, while the width along x of the quantum wires has been varied. The thickness of the upper wire along y is 9 nm, that of the lower wire 7 nm and that of the barrier 4 nm. All the energies are indicated with reference to that of the lowest edge of the bottom quantum wire.

In Figure 2 we report the tunneling conductance between the input of one wire and the output of the other, in units of the conductance quantum $G_0 = 2\ e^2/h$, for a GaInAs/AlInAs structure with widths of 50 nm (dotted line), 100 nm (dashed line) and 200 nm (solid line). The tunneling conductance is plotted versus the potential shift applied to the upper edge of the top quantum wire (the potential, as discussed before, is assumed to vary linearly in the vertical direction). We have considered only positive values of such shift, since

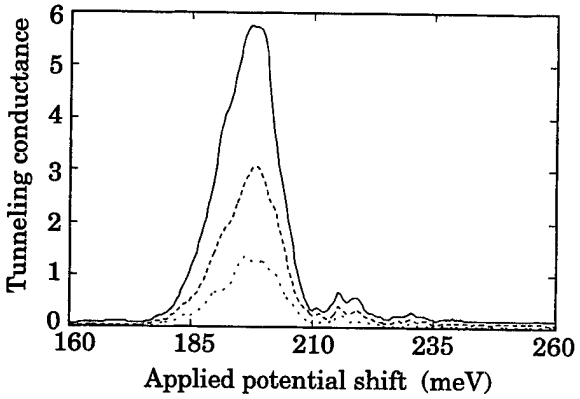


FIGURE 2 Tunneling conductance (in units of the conductance quantum) for the GaInAs/AlInAs structure versus the applied potential shift, with wire widths of 50 nm (dotted line), 100 nm (dashed line) and 200 nm (solid line). The Fermi level has been set at 0.24 eV.

the top wire is thicker than the bottom one and thus it needs to be raised to a higher potential in order to line up the energy levels in the two wires. The coupling length is 200 nm.

For the results shown in this figure, a Fermi level of 0.24 eV has been considered and up to 50 transverse modes have been included in the calculation. A clear peak, reaching a maximum for a bias of approximately 198 meV, is visible, corresponding to the line-up of the energy levels associated with the subbands in the two wires. At this point it is useful to draw a comparison between the situation for devices in which the coupled wires are obtained by lateral confinement [6] in the same 2DEG and the present situation. In the former case all the subbands line up for the same bias condition only if the two wires are identical and, even in such a case, only the highest ones give a significant contribution to the tunneling current, because they have the largest transverse wave vectors, in the direction orthogonal to the potential barrier. In the latter case, instead, the spacing between the subbands is determined by the width of the wire along x, which is the same for both wires, and therefore all of them line up at the same time. More importantly, all the modes give a similar contribution to the tunneling current, because the wave vector orthogonal to the tunneling barrier is the same for all of them. The result is that if the width of the wire increases, the height of the conductance peak also increases, due to the contribution of a larger number of modes, but its shape or position does not vary significantly.

In Figure 3 we report the tunneling conductance results for the GaAs/AlGaAs material systems, for structures 50 nm (thin line) and 100 nm (thick line) wide. Also in this case the coupling length (corresponding to the length of the biasing gate) is 200 nm, while the Fermi level has been set at 0.15 eV. The peak is located approximately at 95 meV: a value much lower than that in Figure 2, because of the significant difference in the effective mass value.

We have then investigated the behavior of the tunneling conductance as a function of the coupling length. If only one mode were contributing to the tunneling current, we would expect the sinusoidal oscillation predicted also by the coupled mode theory used for the analysis of optical waveguides; in this case, however, there are many contributing modes, each of which is associated with a different transfer length. As a result, the tunneling conductance exhibits some complex structure, depending on the number of modes

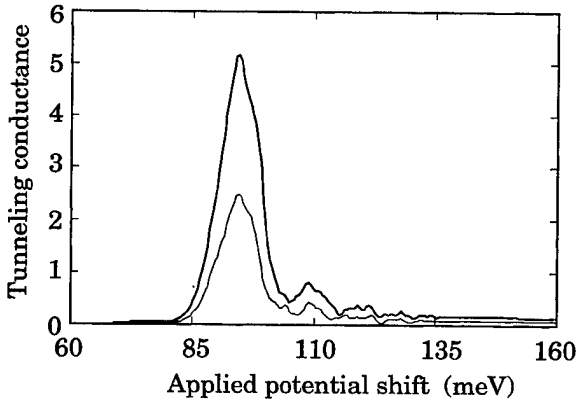


FIGURE 3 Tunneling conductance (in units of the conductance quantum) for the GaAs/AlGaAs structure versus the applied potential shift, with wire widths of 50 nm (thin line) and 100 nm (thick line). The Fermi level has been set at 0.15 eV.

and on the relationships between the related transfer lengths.

In Figure 4 the tunneling conductance is plotted versus the length of the coupling region, for the GaInAs/AlInAs material system and for three different choices of the parameters. Two curves have been obtained for a width of 50 nm and a Fermi level of 0.22 eV (dashed line) and of 0.24 eV (thin line). The third curve (thick line) is for a width of 200 nm and a Fermi level of 0.24 eV. In all three cases the applied potential shift is 196 meV. A rather complex structure can be observed, due to the simultaneous presence of several pairs of coupled modes, each of which is characterized by a different transfer length. For the lowest value of the Fermi level we have an almost monomode propagation and the tunneling conductance oscillates sinusoidally.

Results for the GaAs/AlGaAs material system are reported in Figure 5, where the thick curve and the thin curve have been obtained for a Fermi level of 0.16 eV and wire widths of 200 nm and of 50 nm, respectively. The dashed curve is instead for a wire width of 50 nm and a Fermi level of 0.13 eV. The applied potential shift is 100.5 meV. Also for this material system we observe a transition from a quasisinusoidal dependence of the tunneling current on the coupling length (when just one pair of modes contributes) to a rather complex behavior.

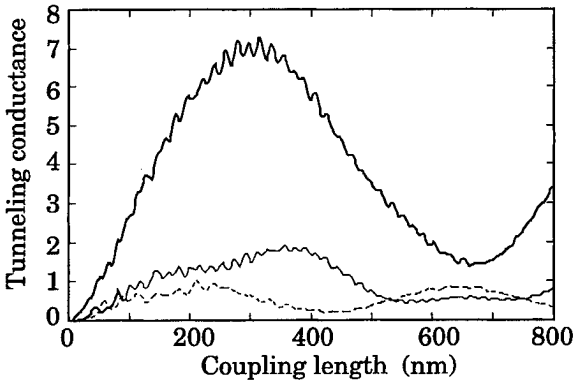


FIGURE 4 Tunneling conductance (in units of the conductance quantum) for the GaInAs/AlInAs structure versus the coupling length, with a wire width of 50 nm and a Fermi level set at 0.24 eV (thin line) or at 0.22 eV (dashed line). The thick line represents the results for a wire width of 200 nm and a Fermi level set at 0.24 eV. For all three cases the applied bias potential is 196 meV.

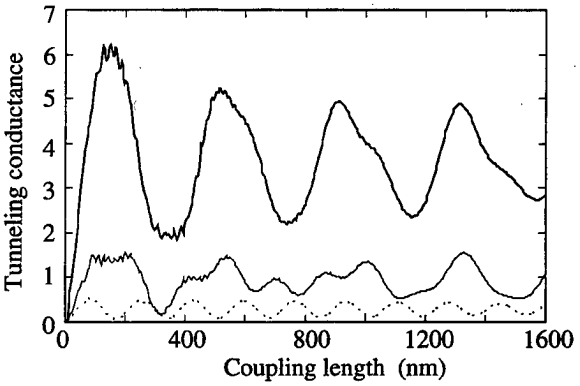


FIGURE 5 Tunneling conductance (in units of the conductance quantum) for the GaAs/AlGaAs structure versus the coupling length, with a wire width of 50 nm and a Fermi level set at 0.16 eV (thin line) or at 0.13 eV (dashed line). The thick line represents the results for a wire width of 200 nm and a Fermi level set at 0.16 eV. For all three cases the applied bias potential is 100.5 meV.

4. CONCLUSIONS

We have investigated the tunneling conductance between two stacked quantum wires, as a function of an applied transverse electric field and of the coupling length. The conductance peak that is observed when the subbands in one wire line up with those in the other wire has a height depending on the total number of occupied subbands. Thus a relatively low "on" resistance can be achieved, which makes the investigated structure suitable for operation as a controlled switch.

Acknowledgement

This work has received partial support from the NATO Collaborative Research Grant n. 950753 and NSF grant ECS 95-26127. One of the authors (M. M.) acknowledges useful discussion with Michele Governale, Freek Prins and Heiko Fresser, and support from the Italian Ministry of the University and Scientific Research.

References

[1] Sakakibara, H., Noguchi, M., Thornton, T. J., Hirakawa, K. and Ikoma, T., "Electron Wave Switch with Distributed Coherent Resonant Tunneling Coupling", in the *Proceedings of the International Electron Device Meeting IEDM* 93-411.

[2] Macucci, M., Ravaioli, U. and Kerkhoven, T. (1992). "Analysis of Electron Transfer between Parallel Quantum Wires", *Superlattices and Microstructures*, **12**, 509.

[3] Macucci, M., Galick, A. and Ravaioli, U. (1995). "Quasi-three-dimensional Green's-function simulation of coupled electron waveguides", *Phys. Rev. B*, **52**, 5210.

[4] Ando, Y. and Itoh, T. (1987). "Calculation of transmission tunneling currents across arbitrary potential barriers", *Journal of Appl. Phys.*, **61**, 1497.

[5] Wees, J., van, Kouwenhoven, L. P., Willems, E. M. M., Harmans, C. J. P. M., Mooji, J. E., Houten, H. van, Beenakker, C. W. J., Williamson, J. G. and Foxon, C. T. (1991). "Quantum ballistic and adiabatic electron transport studied with quantum point contacts", *Phys. Rev. B*, **43**, 12431.

[6] del Alamo, J. A. and Eugster, C. C. (1990). "Quantum field-effect directional coupler", *Appl. Phys. Lett.*, **56**, 78.

Authors' Biographies

Massimo Macucci received the Laurea in Electronic Engineering in 1987 from the University of Pisa, Italy, and the Ph.D. in Electrical Engineering from the University of Illinois, in 1993. He is currently serving on the faculty of the Electrical Engineering Department at the University of Pisa, Italy. His research interests include quantum-interference and single-electron devices, Coulomb Blockade phenomena, modeling and measurements of noise in electron devices.

Albert Galick received the B.S. degree in mathematics from the Massachusetts Institute of Technology in 1980, the M.S. degree in Mathematics in 1984 and the Ph.D. Degree in Computer Science in 1993. He worked as programmer with AT&T in 1980 and from 1984 to 1986. Since 1993 he is a post-doctoral research associate at the Beckman Institute of the University of Illinois.

Umberto Ravaioli received the Laurea in Electronic Engineering in 1980 and the Laurea in Physics in 1982, from the University of Bologna, Italy, and the Ph.D. in Electrical Engineering from Arizona State University, in 1986. He is now a Professor of Electrical and Computer Engineering at the University of Illinois, Urbana-Champaign.

POSTER PRESENTATIONS

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 253–256
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Monte Carlo Simulation of Non-Local Transport Effects in Strained Si on Relaxed $Si_{1-x}Ge_x$ Heterostructures

F. GÁMIZ *, J. B. ROLDÁN and J. A. LÓPEZ-VILLANUEVA

Departamento de Electrónica y Tecnología de Computadores. Universidad de Granada. Facultad de Ciencias. Avd. Fuentenueva s/n. 18071 Granada (Spain)

Electron transport properties of strained-Si on relaxed $Si_{1-x}Ge_x$ channel MOSFETs have been studied using a Monte Carlo simulator. The steady- and non-steady-state high-longitudinal field transport regimes have been described in detail. Electron-velocity-overshoot effects are studied in deep-submicron strained-Si MOSFETs, where they show an improvement over the performance of their normal silicon counterparts. The impact of the Si layer strain on the performance enhancement are described in depth in terms of microscopic magnitudes.

Keywords: Strained Si layer, electron velocity overshoot, conduction effective mass reduction, intervalley scattering rate reduction, Monte Carlo simulation

1. INTRODUCTION

Recently, both theoretical and experimental works have shown important electron mobility enhancement when silicon is grown pseudomorphically on relaxed $Si_{1-x}Ge_x$ at different temperatures. The strain causes the six-fold degenerate valleys of the silicon conduction band minimum to split into two groups: two lowered valleys with the longitudinal effective mass axis perpendicular to the interface, and four raised valleys with the longitudinal mass axis parallel to the interface. This splitting reduces the intervalley phonon scattering rate compared with that of unstrained silicon. In addition, in the lowered valleys, which are more populated in the strained case, electrons show a smaller conduction effective mass (transverse mass) in transport parallel to the interface. The combination of the light effective mass and reduced intervalley scattering gives rise to higher electron mobility [1]. Moreover, the lower intervalley-scattering rates make the energy relaxation times higher, originating important electron velocity overshoot. These advantages can be used to improve MOSFETs parameters, taking advantage both of the higher carrier mobility and the higher electron velocity overshoot, thus greatly improving short channel MOSFET transconductance.

* Corresponding author.

2. MONTE CARLO SIMULATOR

We have developed a strained-Si-*n*-type Monte Carlo simulator by adapting a previous unstrained-Si one which includes inversion-layer quantization and a nonparabolic band model [2]. The quantization effects are included by solving the Poisson equation coupled with the one-dimensional Schrodinger equation all along the channel. The value of the conduction band offset for the four-fold in-plane bands over the value for the two-fold out-of-plane bands was $0.67x$ eV, where x was the Ge mole fraction [3] and the valley shape was not modified by the strain. The two-dimensional Poisson equation is solved throughout the MOSFET to account for short-channel effects. Phonon, surface-roughness and Coulomb scattering rates have been evaluated at each point of the channel [3]. The electron energy for all our simulations was always under 0.5 eV, and so a nonparabolic simplified band structure can be used to accurately describe the Si band structure [4].

3. RESULTS

We have simulated strained-Si on relaxed $Si_{1-x}Ge_x$ long-channel MOSFETs in order to obtain the steady-state electron velocity and energy curves versus the longitudinal-electric field (see Fig. 1). After having fixed the transverse-electric field we increased the longitudinal field step by step, recording the most important transport magnitudes when the steady-state was reached. The saturation velocities obtained were almost the same for all Ge mole fractions: 1.1×10^7 cm/s at 300 K and 1.4×10^7 cm/s at 77 K, although they were a bit higher ($\approx1-2\%$) for $x > 0$.

The energy- and momentum-relaxation times can be calculated making use of the data shown in Figure 1. Different relaxation times are obtained as x changes. The momentum- and energy-relaxation times, for different Ge mole fractions and temperatures are shown in Figure 2. The momentum-relaxation times are smaller than the energy ones. The difference between the momentum- and energy-relaxation times produces nonlocal electron transport effects such as electron-velocity overshoot [5]. These effects are expected to occur on a time scale shorter than the energy-relaxation time. It is foreseeable that the higher energy-relaxation times observed in Figure 2 as x rises lead to increased electron velocity-overshoot effects and therefore higher MOSFET transconductances as channel dimensions are reduced. The relaxation times are lower at low temperature than at room temperature and therefore electron velocity overshoot effects are higher at low temperature as it shown in Figure 3.

We have studied velocity-overshoot effects by applying a sudden longitudinal-electric field of 2×10^5 V/cm to a steady-state electron distribution achieved under the influence of a longitudinal-electric field of 1×10^4 V/cm. The time evolution of

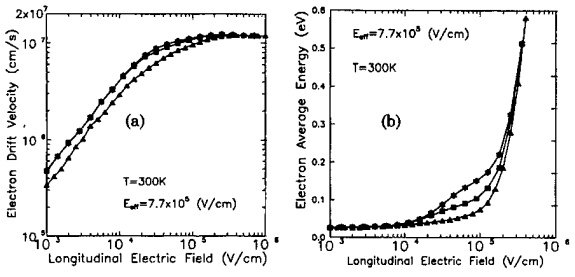


FIGURE 1 Steady-state drift-electron-velocity curves (a) and electron-average energy (b) versus longitudinal-electric field for $E_{eff}=7.7\times10^5$ V/cm at $T=300$ K. (▲: $x=0$; ■: $x=0.2$; ∗: $x=0.4$). The high effective field chosen is typical of very short channel MOSFETs where high doping profiles and thin oxides are used.

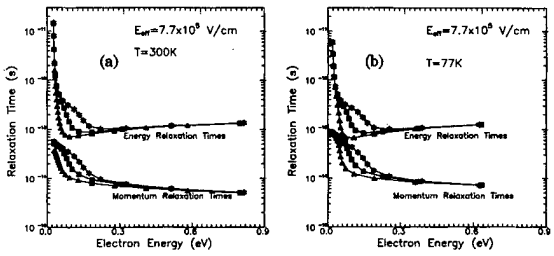


FIGURE 2 Energy- and momentum-relaxation times versus electron energy for $E_{eff} = 7.7\times10^5$ V/cm at $T=300$ K (a) and $T=77$ K (b). (▲: $x=0$; ■: $x=0.2$; ∗ :$x=0.4$).

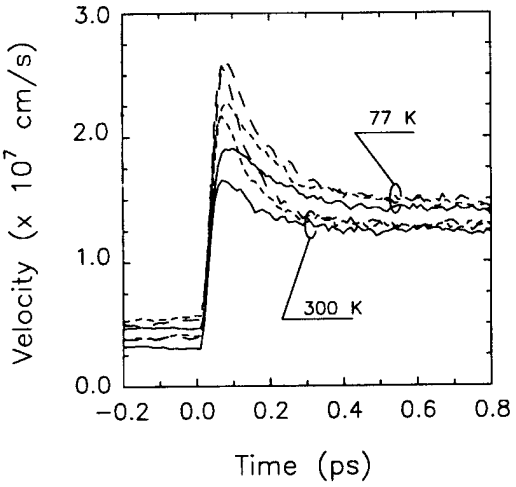


FIGURE 3 Transient overshoot velocity with a sudden application of the field 2×10^5 V/cm at room and low temperature for $E_{\text{eff}} = 7.7\times10^5$ V/cm. Unstrained-silicon (solid line), $x=0.2$ (short dashed line), $x=0.4$ (long dashed line).

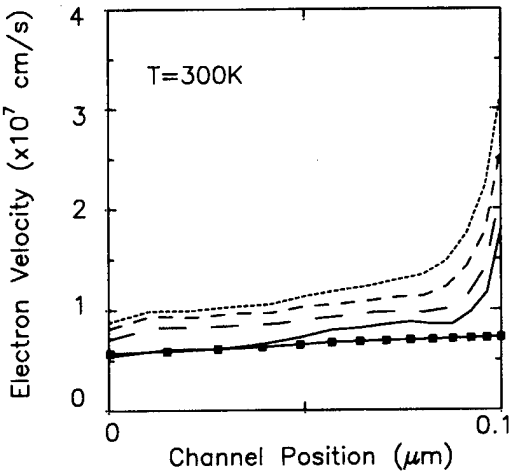


FIGURE 4 Electron velocity versus channel position at $T=300$ K. Velocity distribution obtained in an 0.1 μm MOSFET with $V_{GS}=1.3$ V and $V_{DS}=0.5$ V for $x=0$ (solid line), $x=0.1$ (long dashed line), $x=0.2$ (medium dashed line), $x=0.3$ (short dashed line). Low longitudinal-electric-field velocity corrected using Thornber's expression and the longitudinal field distribution obtained for the previous MOSFET with $x=0$ (Squares).

the electron velocity is shown in Figure 3. It is clear that the time taken to reach the steady-state velocity corresponding to a longitudinal-electric field of 2×10^5 V/cm increases as x rises. This result is coherent with the energy-relaxation times shown in Figure 2. It is clear that the increase of the energy-relaxation times as x rises causes the time taken to reach the steady-state to be longer and hence enhances the velocity-overshoot effects.

We have simulated several 0.1 μm channel length MOSFETs for $x=0, 0.1, 0.2$ and 0.3 (the thickness of the strained Si layer is 4.6 nm). The MOSFET external bias was $V_{DS}=0.5$ V, $V_{GS}=1.5$ V, and $V_{SB}=0$. The velocity distribution along the channel obtained for each Ge mole fraction is plotted in Figure 4 at room temperature. The electron velocity is higher than the saturation velocity for strained and unstrained-Si channel MOSFETs near the drain edge. This effect is due to the high longitudinal electric-field gradient the carriers face as they travel toward the drain. This gradient makes the electrons overshoot the velocity they would have if they were subject to steady-state transport.

4. CONCLUSIONS

A Monte Carlo simulator has been used to study the electron transport properties of strained-Si on relaxed $Si_{1-x}Ge_x$ channel MOSFETs and the performance improvement of these devices at high-longitudinal fields. Similar saturation velocities are obtained no matter the value of the Ge mole fraction, however the electron velocity overshoot effects increase as the Ge mole fraction rises.

References

[1] Nelson, S. F., Ismail, K., Chu, J. O. and Meyerson, B. S. (1993). "Room-temperature electron mobility in strained Si/SiGe heterostructures", *Appl. Phys. Lett.*, **63**, 367.

[2] Gámiz, F., López-Villanueva, J. A., Jiménez-Tejada, J. A., Melchor, I. and Palma, A. (1994). "A comprehensive model for Coulomb scattering in inversion layers", *J. Appl. Phys.*, **75**, 924.

[3] Gámiz, F., Roldán, J. B., López-Villanueva, J. A. and Cartujo, P. (1996). "Coulomb scattering in strained silicon inversion layers on $Si_{1-x}Ge_x$ substrates", *Appl. Phys. Lett.*, **69**, 797.

[4] Fischetti, M. V. and Laux, S. E. (1993). "Monte Carlo study of electron transport in silicon inversion layers", *Phys. Rev.*, **B48**, 2244.

[5] Sai-Halasz, G. A., Wordeman, M. R., Kern, M. R., Rishton, S. and Ganin, E. (1988). "High transconductance and velocity overshoot in NMOS devices at the 0.1 μm gate-length level", *IEEE Elec. Dev. Let.*, **EDL-9**, 464.

Authors' Biographies

Francisco Gámiz graduated with a degree in physics in 1991, and received the Ph.D. in 1994 from the University of Granada. Since 1991 he has been working on the characterization of scattering mechanisms and their influence on the transport properties of charge carriers in semiconductor heterostructures. His current research interest includes the effects of many-carriers on the electron mobility and the interpretation of the influence of high longitudinal electric fields have on MOS transistors. Current interest are also related to SiGe and SiC, and SOI devices, and quantum transport. He has coauthored several papers in all these subjects. He is an Associate Professor at the University of Granada.

Juan B. Roldán graduated with a degree in physics in 1993, and received the Ph.D. in 1997 from the University of Granada. Since 1993 he has been working on the MOS device including 2D transport, non-local effects and Monte Carlo simulations. Current interests are also related to SiGe and SiC devices. He is a Teaching Assistant at the University of Granada.

Juan A. López-Villanueva graduated in 1984, Ph.D. in 1990 (University of Granada) with a thesis on the degradation of MOS structures by Fowler-Nordheim tunneling. Since 1985 he has been working on deep-level characterization and, mainly, MOS device physics, including Fowler-Nordheim and direct tunneling, quantum effects, 2D transport, effects of nonparabolicity, scattering mechanisms and Monte Carlo simulation of charge transport. He has coauthored several papers in all these subjects. His current research interest includes, simulation and modelling of electron devices. His educational activities also include analog systems for electronic instrumentation and power electronics. He is an Associated Professor at the University of Granada.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 257–260
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

A β-SiC MOSFET Monte Carlo Simulator Including Inversion Layer Quantization

F. GÁMIZ, J. B. ROLDÁN * and J. A. LÓPEZ-VILLANUEVA

Departamento de Electrónica y Tecnología de Computadores, Universidad de Granada, Facultad de Ciencias, Avd. Fuentenueva s/n, 18071 Granada (Spain)

Electron transport properties in SiC quantized inversion layers have been studied by means of a Monte Carlo procedure. It has been observed that the contribution of polar-optical phonon scattering produces a significant influence of the effective-electric field on the high longitudinal field transport regime, this being the main difference of SiC with respect to standard Si inversion layers. The energy- and momentum-relaxation times have been calculated and the results suggest that electron velocity overshoot effects are less important than in Si MOSFETs. The electron mobility is not very different from their silicon counterparts, but the saturation velocity is higher.

Keywords: Silicon carbide, inversion layer, high field transport properties, electron mobility, polar-optical phonon scattering

1. INTRODUCTION

Cubic silicon carbide (β-SiC) is thought to be a material well-suited for electronic devices operating under harsh conditions such as high temperature, high power, and high frequency due to its high saturated drift velocity, wide band gap, and high thermal conductivity [1–2]. The fabrication of high temperature metal-oxide-semiconductor field-effect transistors (MOSFETs), junction field-effect transistors, and metal-semiconductor field-effect transistors has already been demonstrated [3].

As a consequence of this interest, theoretical analyses, and simulations on the electronic properties of silicon carbide have begun to appear in the last few years. However, up to now, numerical simulations related to the electron transport properties of silicon carbide inversion layers have not considered electron quantization. The wide experience obtained in silicon and other semiconductor material inversion layers over many years suggests that for a complete evaluation of the performance potential of silicon carbide MOSFETs, a detailed analysis of the transport properties of quantized silicon carbide inversion layers is essential.

* Corresponding author: Tel.: 34-58-246145. Fax. 34-58-2432300. E-mail: paco@gcd.ugr.es or juamba@gcd. ugr.es.

2. MONTE CARLO SIMULATOR

Electron transport properties such as electron velocity and mobility at room and higher temperatures were studied by using a Monte Carlo simulator we have developed [4, 5]. The position of the subband minima and the electron concentration of each of them were obtained by the self-consistent solution of the Poisson and Schroedinger equations for each value of the effective-transverse-electric field. A non-parabolic band model was assumed taking $\alpha = 0.323\,\text{eV}^{-1}$ [2], which limits our study to low-electron energies (below 1 eV). We have allowed the electron to travel in six subbands and to move between them. Acoustic, intervalley-optical and polar-optical phonons, surface-roughness and Coulomb scattering due to both bulk impurities and interface-trapped and oxide-trapped charge have been taken into account.

3. RESULTS

We have simulated a SiC MOSFET with an oxide thickness of 150 Å and a substrate impurity concentration $N_A = 1\times10^{16}\,\text{cm}^{-3}$. We have paid special attention to the role that polar-optical phonon scattering plays in the total scattering rates. We have studied the ratio R between the total scattering rates with and without the inclusion of polar-optical phonon scattering mechanisms. It has been plotted in Figure 1 for the lowest energy subbands versus electron kinetic energy. As can be observed, the lower the effective-electric field the higher the ratio R for a wide range of electron kinetic energy. Therefore, the energy relaxation times are dominated by polar-optical phonon scattering at low effective-electric fields.

We have simulated long-channel β-SiC MOSFETs to obtain velocity curves versus longitudinal-electric field (Fig. 2). As can be seen, the drift velocity curves reach a maximum and then tend to decrease as the drift-electric field increases. There is a strong dependence of the velocity curves on the

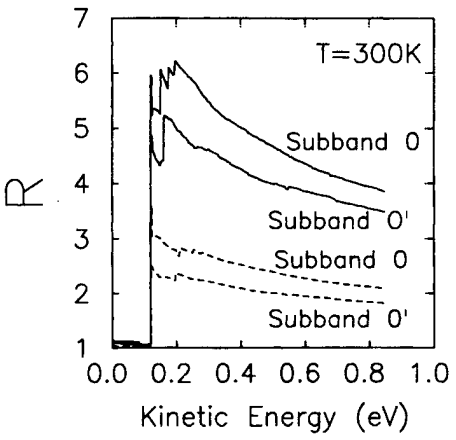


FIGURE 1 Ratio R between the total scattering rates taking and not taking into account polar-optical phonon scattering for the first primed (0) and unprimed (0′) subbands at $T = 300\,\text{K}$ for two different effective-electric fields: $E_{EFF} = 10^5$ V/cm (solid line) and $E_{EFF} = 8\times10^5$ V/cm (dashed line).

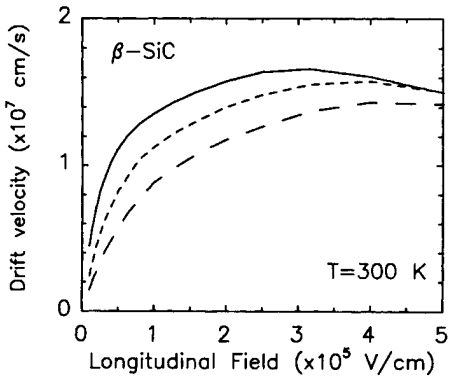


FIGURE 2 Homogeneous steady-state electron drift velocity curves in β-SiC at room temperature for different transverse-effective fields. (solid: $E_{EFF} = 1\times10^5$ V/cm, short dashed: $E_{EFF} = 5\times10^5$ V/cm, long dashed: $E_{EFF} = 1\times10^6$ V/cm) for $T = 300\,\text{K}$.

effective-electric field, this fact is connected with the results of Figure 1 (the lower the transverse field the higher the saturation velocity). The saturation velocity is higher than in silicon, which is one of the main reasons explaining the potential importance of silicon carbide as an electron device material, in addition to its wide band gap. Unfortunately there are no available experimental data to compare with our simulation data.

Making use of the steady-state velocity and energy plots we have calculated (Fig. 3) the energy-

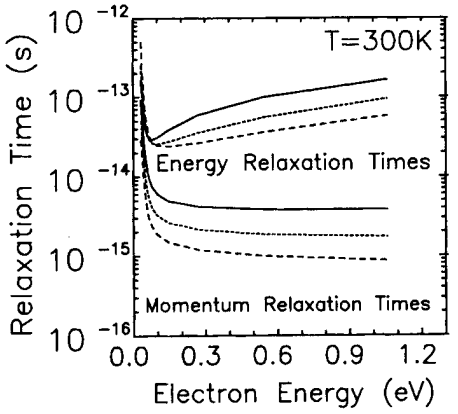


FIGURE 3 Momentum- and energy-relaxation times versus electron energy for a β-SiC MOSFET at $T=300$ K (solid line) $E_{EFF}=10^5$ V/cm, (short dashed line) $E_{EFF}=6.7\times10^5$ V/cm, (long dashed line) $E_{EFF}=10^6$ V/cm.

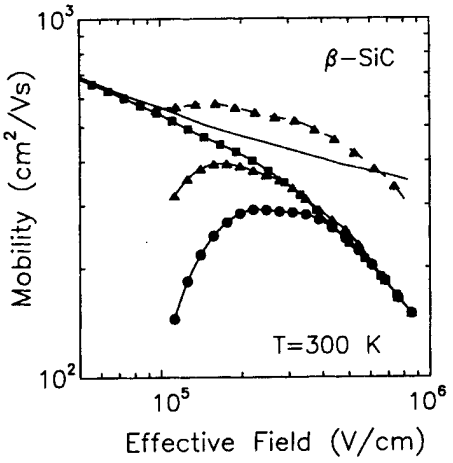


FIGURE 4 Electron mobility curves in β-SiC versus transverse effective field at room temperature: (a) only phonon scattering (solid line), (b) phonon and surface-roughness scattering (solid line and full squares), (c) Coulomb scattering due only to substrate doping impurities (solid line and full triangles), and (d) assuming, in addition to what is included in (c), a typical interface charged layer of $N_{it}=1\times10^{11}$ cm^{-2} (solid line and full circles). Electron mobility curve in Si (dashed line and full triangles).

and momentum-relaxation times at room temperature by means of the analytical expressions given in Ref. 6. This calculation has been done by using an average effective mass taking into consideration the fraction of electrons in the primed and unprimed subbands for each effective field. The SiC relaxation times are approximately one order of magnitude lower than the silicon ones, implying that the electron-velocity-overshoot effects are going to be less important than in silicon MOSFETs. At this point, it is worth highlighting the importance of the effective-electric field on the high longitudinal-electric-field transport properties in SiC MOSFETs. There exists a strong dependence of the saturation velocity (homogeneous transport in the channel) and of the relaxation times (inhomogeneous transport in the channel, velocity overshoot) on the effective-electric field, which is more important than in standard silicon MOSFETs.

We have simulated and plotted in Figure 4 the mobility curves versus the transverse-effective field at room temperature assuming (a) only phonon scattering (solid line), (b) phonon and surface-roughness scattering (solid line and closed squares), (c) Coulomb scattering due only to substrate doping impurities (solid line and triangles), and (d) assuming, in addition, a typical interface charged layer of $N_{it}=1\times10^{11}$ cm^{-2} (closed circles). A mobility curve for a Si MOSFET corresponding to case (c) is included in dashed line and triangles. When only phonon scattering is taken into account, the electron mobility at low transverse-electric fields approaches the SiC bulk-mobility value (1000 cm^2/Vs). The reduction of the average distance between the inversion charge and the oxide interface increases surface-roughness and Coulomb scattering. In this respect, surface-roughness, Coulomb, and phonon-scattering probabilities (the last one is higher in SiC due to the additional contribution of polar-optical phonons), which are the three main scattering mechanisms that limit electron mobility in the channel of a MOSFET, are higher in SiC than in Si. It is clear that SiC MOSFETs are by no means candidates to substitute conventional Si ones at room temperature. However, at high temperature, where Si MOSFETs can no operate well, SiC devices play their major role.

4. CONCLUSIONS

Electron transport properties have been evaluated in SiC quantized inversion layers. It has been shown that the influence of the effective-electric field on the high longitudinal field transport regime is significant, mainly due to the contribution of polar-optical phonon scattering which is essential at low effective fields for low and room temperatures. The electron-velocity-overshoot effects are less important than in Si MOSFETs. The electron mobility is not too different from their silicon counterparts, but the saturation velocity is nevertheless higher. So, this transport feature, in addition to a wide band gap, and high thermal conductivity, make SiC a promising candidate for high-temperature/high-power applications.

References

[1] Ivanov, P. A. and Chelnokov, V. E. (1992). "Recent development in SiC single-crystal electronics", *Semicond. Sci. Technol.*, **7**, 863–880.

[2] Tsukioka, K., Vasileska, D. and Ferry, K. (1993). "An ensemble Monte Carlo study of high-field transport in β-SiC", *Physica B.*, **185**, 467

[3] Palmour, J. W., Kong, H. S. and Davis, R. F. (1988). "Characterization of device parameters in high-temperature metal-oxide-semiconductor field-effect transistors in β-SiC thin films", *J. Appl. Phys.*, **64**, 2168–2177.

[4] Gámiz, F., Roldán, J. B., López-Villanueva, J. A. and Carceller, J. E. (1996). "Electron velocity saturation in quantized silicon carbide inversion layers", *Appl. Phys. Lett.*, **69**, 2219–2221.

[5] Roldán, J. B., Gámiz, F., López-Villanueva, J. A. and Cartujo, P. (1997). "Electron transport properties of quantized silicon carbide inversion layers", *J. Electr. materials*, **26**, 203–207.

[6] Bordelon, T. J., Wang, X. L., Maziar, C. M. and Tasch, A. F. (1991). "An evaluation of energy transport models for silicon device simulation", *Solid-State Electron.*, **34**, 617.

Authors' Biographies

Juan B. Roldán graduated with a degree in physics in 1993, and received the Ph.D. in 1997 from the University of Granada. Since 1993 he has been working on the MOS device physics including 2D transport, non-local effects and Monte Carlo simulations. Current interests are also related to SiGe and SiC devices. He is a Teaching Assistant at the University of Granada.

Francisco Gámiz graduated with a degree in physics in 1991, and received the Ph.D. in 1994 from the University of Granada. Since 1991 he has been working on the characterization of scattering mechanisms and their influence on the transport properties of charge carriers in semiconductor heterostructures. His current research interest includes the effects of many-carriers on the electron mobility and the interpretation of the influence of high longitudinal electric fields have on MOS transistors. Current interest are also related to SiGe and SiC, and SOI devices, and quantum transport. He has coauthored several papers in all these subjects. He is an Associate Professor at the University of Granada.

Juan A. López-Villanueva graduated in 1984, Ph.D. in 1990 (University of Granada) with a thesis on the degradation of MOS structures by Fowler-Nordheim tunneling. Since 1985 he has been working on deep-level characterization and, mainly, MOS device physics, including Fowler-Nordheim and direct tunneling, quantum effects, 2D transport, effects of nonparabolicity, scattering mechanisms and Monte Carlo simulation of charge transport. He has coauthored several papers in all these subjects. His current research interest includes, simulation and modelling of electron devices. His educational activities also include analog systems for electronic instrumentation and power electronics. He is an Associate Professor at the University of Granada.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 261–264
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Development of a Method for Determining the Dependence of the Electron Mobility on the Longitudinal-Electric Field in MOSFETs

J. B. ROLDÁN *, F. GÁMIZ and J. A. LÓPEZ-VILLANUEVA

Departamento de Electrónica y Tecnología de Computadores, Universidad de Granada, Facultad de Ciencias, Avd., Fuentenueva s/n, 18071 Granada Spain

A new experimental method for determining the dependence of the electron mobility on the longitudinal-electric field has been developed. The development, validation and explanation of this new method has been carefully carried out. We have applied this procedure to standard submicron MOSFETs and after having obtained the mobility dependence on both the transverse- and longitudinal-electric fields we reproduced the experimental output curves. The saturation velocity has also been calculated using the mobility curves obtained by this new method.

Keywords: Electron mobility, longitudinal-electric field, saturation velocity, output curves, drift-diffusion simulation

1. INTRODUCTION

The electron-mobility dependence on the longitudinal-electric field, oriented along the channel, has been studied experimentally. A "closed-loop" investigation has been performed on the validity and accuracy of a new method of determining the reduction of the channel mobility connected to the increase of the longitudinal field. This method is applicable to standard transistors. An estimation of the saturation electron velocity in the channel has been accomplished comparing a widely used expression [1]. The data obtained with this procedure can be used to develop local models to account for the mobility dependence on the longitudinal-electric field in drift-diffusion simulators.

2. DESCRIPTION OF THE METHOD

The basic concept of idea of this method is the comparison of experimental drain current versus drain voltage curves $I_D - V_{DS}$, measured in short-channel transistors to the results produced by a two-dimensional simulator [2, 3] in which an

* Corresponding author.

accurate low-field mobility is used and the mobility reduction caused by the longitudinal-electric field is included. The total channel length L is divided in N subchannels with lengths L_i, $i=1, 2, \ldots, N$, respectively. Instead of fixing the L_i values, the channel potential values at the ends of the i-th subchannel, V_{i-1} and V_i, are defined as the separation between the quasifermi levels at these points, $V_0 = V_S$ and $V_N = V_D$ being the potentials at the total channel ends. Therefore, the separation into subchannels is obtained by dividing $V_D - V_S$ into very small intervals and fixing the channel potential at the ends of each subchannel. The number of subchannels is high enough and the difference $V_i - V_{i-1}$ in each subchannel is low enough that the longitudinal-electric field in each subchannel can be considered to be uniform. The application of the method is as follows:

1) Starting from a very small V_{DS}, for which one subchannel is enough, the gate-to-source voltage is varied and the low-field mobility as well as its dependence on the transverse-electric field is obtained.
2) V_{DS} is increased in very small steps and the experimental result is compared to the simulated result calculated by using the low-field mobility obtained in point (1). New subchannels are introduced if necessary. When the experimental and computed currents deviate more than a prefixed (also very small) limit, δI, a mobility reduction is apparent at the subchannel placed at the drain end of the channel. Then, the mobility in this subchannel is multiplied by a factor of $\lambda < 1$ in order to reproduce the experimental current again. The values for the transverse- and longitudinal-electric fields in this subchannel are introduced in an array: E_{trans} (1) and E_{long} (1), respectively.
3) V_{DS} is increased again. In all the subchannels for which the longitudinal-electric field is greater than E_{long} (1), the low-field mobility is multiplied by λ and the current thus obtained is compared to the experimental one. When the deviation between the computed and experimental currents is again greater than δI, a further reduction is assumed in the drain end of the channel and the mobility is once more multiplied by λ, thus obtaining a value equal to the low-mobility multiplied by λ^2. The transverse- and longitudinal-electric fields in this subchannel are introduced as a new element of the array: E_{trans} (2) and E_{long} (2), respectively.

At the end of this process, the mobility values for a set of pairs (E_{trans}, E_{long}) is obtained. This is schematized with the vertical solid line plotted in Figure 1.

3. VALIDATION PROCEDURE

Once we developed the algorithm to obtain the dependence of the mobility on the longitudinal-electric field we proceeded to validate it. To do so, we simulated several output curves, using our drift-diffusion simulator for a MOSFET and a supposed mobility dependence, after applying our method we were able to reproduce this dependence, confirming the correctness of the procedure. (Fig. 2).

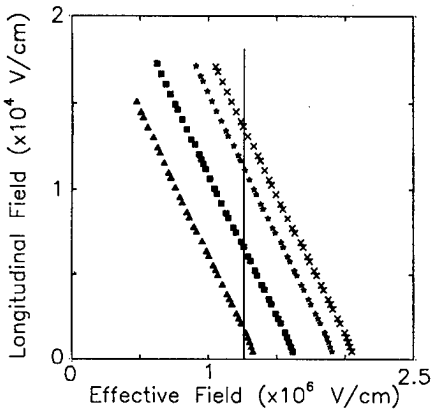


FIGURE 1 Tables ($E_{\text{trans}(i)}$, $E_{\text{long}(i)}$) used to account for the mobility reduction as the longitudinal field rises. They correspond to four different output curves. The solid line represents a common chosen effective field for all the tables. This value will be employed to obtain the mobility versus longitudinal field curves.

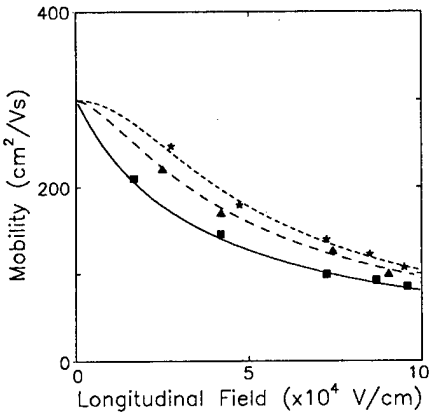


FIGURE 2 Mobility curves versus longitudinal-electric field obtained with Equation 1 to simulate output curves and validate this method. The v_{SAT} chosen was $v_{SAT} = 1.1\times10^7$ cm/s, $\beta = 1$ (solid line), $\beta = 1.5$ (long dashed line), $\beta = 2$ (short dashed line). The results of our method are plotted in squares for $\beta = 1$, triangles for $\beta = 1.5$ and stars for $\beta = 2$.

4. EXPERIMENTAL RESULTS AND DISCUSSION

After having validated the accuracy of our method, we applied it to Si MOSFETs fabricated by Lucent Technologies with $L_{EFF} = 0.6\,\mu m$, $W_{EFF} = 20\,\mu m$, $T_{ox} = 157$ Å, $N_A = 1.2\times10^{17}\,cm^{-3}$ and $R_{SD} = 130\,\Omega$. We have applied our technique to a set of measured output curves with $\delta I = 10^{-3}$ mA and $\lambda = 0.96$. The results obtained can be observed in Figure 3, where the mobility versus effective and longitudinal field surface is shown. Expression 1 was used to obtain the β parameter and the saturation velocity v_{SAT}.

$$V(E_{\parallel}, E_{\perp}) = \frac{\mu_0(E_{\perp})E_{\parallel}}{\left(1+\left(\frac{\mu_0(E_{\perp})\times E_{\parallel}}{v_{sat}}\right)^{\beta}\right)^{\frac{1}{\beta}}} \qquad (1)$$

The values that gave a better fit were $\beta = 1$ and $v_{SAT} = 1.1\times10^7$ cm/s. These values were independent of the effective field, confirming the result by Modelli and Manzini [4]. Using Thornber's expression with these parameters we have tried to reproduce the experimental output characteristics. The result is shown in Figure 4.

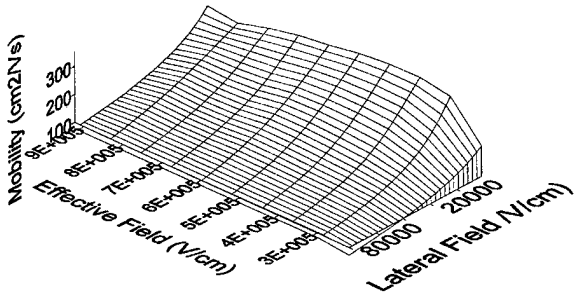


FIGURE 3 Experimental electron mobility as a function of the effective- and longitudinal-electric fields at room temperature.

The data provided by this new procedure are very useful in order to obtain local mobility models for drift-diffusion simulators, where the dependence of the longitudinal-electric field is included. The influence of the series resistance and its bias dependence can be taken into account easily. It is also possible to assess the influence of electron velocity overshoot effects in very short channel MOSFETs by using a augmented mobility model [5].

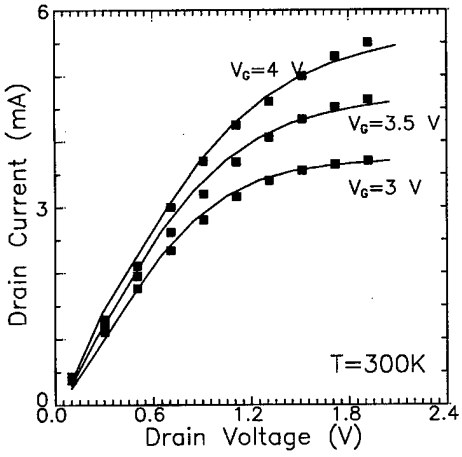


FIGURE 4 Experimental and simulated output curves at $T = 300$ K for a $L_{EFF} = 0.6\,\mu m$ MOSFET. The simulated curves have been obtained by using the mobility shown in Figure 3, whose dependency on the longitudinal-electric field was obtained with this new method.

5. CONCLUSIONS

A new experimental method for determining the dependence of the electron mobility on the longitudinal-electric field has been developed, validated and applied to standard submicron MOSFETs. This method can be easily implemented in a 2D drift-diffusion simulator. All the different dependencies of the electron mobility on the longitudinal-electric field can be taken into account at the same time by means of this method. A saturation-velocity value of 1.1×10^7 cm/s was calculated by comparing the mobility curves to Thornber's expression.

Acknowledgements

We would like to thank José López Serrano from Lucent Technologies (Spain) for providing the MOSFETs (0.9 μm technology) we have used in our study. This work has been carried out within the framework of research project TIC 95-0511, supported by the Spanish Government (CICYT).

References

[1] Thornber, K. K. (1980). "Relation of drift velocity to low-field mobility and high-field saturation velocity", *J. Appl. Phys.*, **51**, 2127.

[2] Roldán, J. B., Gámiz, F., López-Villanueva, J. A. and Carceller, J. E. (1995). "Monte Carlo simulation of a submicron MOSFET including inversion layer quantization", *4th IWCE*, p. 37.

[3] Roldán, J. B., Gámiz, F., López-Villanueva, J. A. and Carceller, J. E. (1997). "Dependence of the electron mobility on the longitudinal electric field in MOSFETs", *Semicond. Sci. Technol.*, **12**, 321–330.

[4] Modelli, A. and Manzini, S. (1988). "High-field drift velocity of electrons in silicon inversion layers", *Solid-State Electron.*, **31**, 99.

[5] Roldán, J. B., Gámiz, F., López-Villanueva, J. A. and Carceller, J. E. (1997). "Modeling effects of electron velocity overshoot in a MOSFET", *IEEE Transc. on Electron Devices*, **44**, 841–846.

Authors' Biographies

Juan B. Roldán graduated with a degree in physics in 1993, and received the Ph.D. in 1997 from the University of Granada. Since 1993 he has been working on the MOS device physics including 2D transport, non-local effects and Monte Carlo simulations. Current interests are also related to SiGe and SiC devices. He is a Teaching Assistant at the University of Granada.

Francisco Gámiz graduated with a degree in physics in 1991, and received the Ph.D. in 1994 from the University of Granada. Since 1991 he has been working on the characterization of scattering mechanisms and their influence on the transport properties of charge carriers in semiconductor heterostructures. His current research interest includes the effects of many-carriers on the electron mobility and the interpretation of the influence of high longitudinal electric fields have on MOS transistors. Current interest are also related to SiGe and SiC, and SCI devices, and quantum transport. He has coauthored several papers in all these subjects. He is an Associate Professor at the University of Granada.

Juan A. López-villanueva graduated in 1984, Ph.D. in 1990 (University of Granada) with a thesis on the degradation of MOS structures by Fowler-Nordheim tunneling. Since 1985 he has been working on deep-level characterization and, mainly, MOS device physics, including Fowler-Nordheim and direct tunneling, quantum effects, 2D transport, effects of nonparabolicity, scattering mechanisms and Monte Carlo simulation of charge transport. He has coauthored several papers in all these subjects. His current research interest includes, simulation and modelling of electron devices. His educational activities also include analog systems for electronic instrumentation and power electronics. He is an Associate Professor at the University of Granada.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 265–273
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Quantum Distribution-function Transport Equations in Non-normal Systems and in Ultra-fast Dynamics of Optically-excited Semiconductors

F. A. BUOT

Naval Research Laboratory, Washington, D.C. 20375-5320

The derivation of the quantum distribution-function transport equations combines the Liouvillian super-Green's function technique and the lattice Weyl-Wigner formulation of the quantum theory of solids. A generating super-functional is constructed which allows an algebraic and straightforward application of quantum field-theoretical techniques in real time to derive coupled quantum-transport, condensate, and pair-wavefunction equations. In optically-excited semiconductors, quantum distribution-function transport equations are given for phonons, plasmons, photons, and electron-hole pairs and excitons by transforming the Bethe-Salpeter equation into a multi-time evolution equation. The virtue of quantum distribution function is that it allows easy application of 'device-inflow' subsidiary boundary conditions for simulating femtosecond device-switching phenomena.

Keywords: Quantum distribution function, quantum transport, ultra-fast optics, excitons, non-equilibrium superconductivity, nonequilibrium superfluidity, lattice Weyl transform, Liouvillian, super Green's function

1. INTRODUCTION

There is a need for generalized quantum distribution-function transport equations, valid for non-normal, non-uniform, and ultra-fast systems, as bases for large-scale computer simulations. This becomes urgent with advances in material science, ultra-fast laser probes, nanofabrication, and the development of more powerful energy beams. The drive to produce systems which are functionally more dense and have wider bandwidths will lead nanostructure devices to atomic-scale dimensions with different materials: insulators, semiconductors, metals, and superconductors.

The nonequilibrium quantum transport theory including pairing dynamics is formulated in terms of the Liouville-space (L-space) quantum-field theory [1–2] and lattice Weyl transform technique [3, 4]. For normal systems, this reduces to the nonequilibrium Green's function technique of Schwinger [5], Kadanoff and Baym [6], and Keldysh [7], coupled with the lattice Weyl-Wigner formulation of the quantum theory of solids [3, 4]. Several new results are derived with the present approach.

This L-space approach has provided the action principle for a multi-variable functional theory of nonequilibrium condensed-matter systems [8–10]. Thus, the method set forth here may open doors for the investigation of ultra-fast dynamics in quantum nanostructures. So far, only the distribution-function approach has characterized, in time-domain, a highly-nonlinear and highly-nonequilibrium quantum behavior [11–14].

2. QUANTUM DYNAMICS IN LIOUVILLE SPACE

The density-matrix equation of quantum statistical dynamics in Hilbert space (H-space) becomes a super-Schrodinger equation for the super-state vector in L-space as

$$i\hbar\frac{\partial}{\partial t}|\rho(t)\rangle\rangle = \mathscr{L}|\rho(t)\rangle\rangle. \tag{1}$$

$\rho(t)$ is the density-matrix operator for the whole many-body system in H-space, and $|\rho(t)\rangle\rangle$ is the corresponding super-state vector in L-space. The super-operator $\mathscr{L}$ corresponds to the commutator $[\mathscr{H}, \rho]$, and is referred to in this paper as the Liouvillian. Thus, we may write the Liouvillian $\mathscr{L}$ as $\mathscr{L} = \hat{\mathscr{H}} - \tilde{\mathscr{H}}$, which define $\hat{\mathscr{H}}$ and $\tilde{\mathscr{H}}$. These have the property that $\hat{\mathscr{H}}|\rho(t)\rangle\rangle = |\mathscr{H}\rho(t)\rangle\rangle$, and $\tilde{\mathscr{H}}|\rho(t)\rangle\rangle = |\rho(t)\mathscr{H}^{\dagger}\rangle\rangle$. These relations are valid for fermions and bosons. For number-conserving fermion operator $\mathscr{H}$, $\tilde{\mathscr{H}} = \tilde{\mathscr{H}}^{\dagger}$. The quantum field super-operators, $\hat{\psi}(\tilde{\psi})$ and $\hat{\psi}^{\dagger}(\tilde{\psi}^{\dagger})$, are defined through their commutation relations in H-Space.

3. SUPER S-MATRIX THEORY IN L-SPACE

The "transition probability" in L-space is given by the following relation

$$\langle\langle\Lambda(t)|\rho(t)\rangle\rangle = \langle\langle 1|\bar{S}(\infty, -\infty)|\rho_{\mathrm{eq}}\rangle\rangle \tag{2}$$

where $|1\rangle\rangle$ is a unit super-vector. We have

$$|\rho(t)\rangle\rangle = T\exp\left\{\frac{-i}{\hbar}\int_0^t \mathscr{L}_0(t')dt'\right\}\bar{S}(t, -\infty)|\rho_{\mathrm{eq}}\rangle\rangle$$

$$|\Lambda(t)\rangle\rangle = T^{\mathrm{ac}}\exp\left\{\frac{-i}{\hbar}\int_t^0 \mathscr{L}_0(t')dt'\right\}\bar{S}(t, \infty)|\mathscr{O}\rangle\rangle \tag{3}$$

where T^{ac} denotes anti-chronological time ordering and

$$\bar{S}(t, -\infty) = T\exp\left\{\frac{i}{\hbar}\int_0^t \mathscr{L}_0(t')dt'\right\} T\exp\left\{\frac{-i}{\hbar}\int_{-\infty}^t \mathscr{L}(t')dt'\right\} T\exp\left\{\frac{-i}{\hbar}\int_0^t \mathscr{L}_0(t')dt'\right\}$$

The "transition probability" obeys the equality

$$\langle\langle 1|\bar{S}(\infty, -\infty)|\rho_{\mathrm{eq}}\rangle\rangle = \exp\frac{i}{\hbar}W \tag{4}$$

where W is identified as the effective action. It can be shown that

$$W = \int_{-\infty}^{\infty}\langle\langle\Lambda(t)|i\hbar\frac{\partial}{\partial t} - \mathscr{L}|\rho(t)\rangle\rangle\, dt \tag{5}$$

This relation forms the basis of a time-dependent functional theory of condensed matter discussed by Rajagopal and Buot in a series of papers [8–10].

4. GENERALIZED QUANTUM DISTRIBUTION FUNCTIONS

We introduce a 4-component second quantized quantum field super-operators given by

$$\Psi(1)^T = (\hat{\psi}(1) \quad \tilde{\psi}^{\dagger}(1) \quad \hat{\psi}^{\dagger}(1) \quad \tilde{\psi}(1)) \tag{6}$$

This is a generalization of the multi-component quantum field operator first introduced by Nambu [15], and by De Dominicis and Martin [16]. In

terms of the Ψ fields, a system Liouvillian $\mathscr{L}$ can in general be expressed as

$$\mathscr{L} = \sum_{N=1}^{M} v(1,2,3,\ldots,N)\,\Psi(1)\Psi(2)\Psi(3)\ldots\Psi(N) + \mathscr{L}_{\text{ext}} \tag{7}$$

where the Schwinger source term is given by $\mathscr{L}_{\text{ext}} = \hat{\mathscr{H}}_{\text{ext}} - \tilde{\mathscr{H}}_{e,\text{ext}} = u(1)\Psi(1) + u(12)\Psi(1)\Psi(2)$. The field super-operator averages can be written in terms of the $\bar{S}$-matrix, e.g.,

$$\mathscr{G}(1,2) = \frac{(i\hbar)^2 \frac{\delta^2\langle\langle 1|\bar{S}(\infty,-\infty)|\rho_{\text{eq}}\rangle\rangle}{\delta u(2)\delta u(1)}}{\langle\langle 1|\bar{S}(\infty,-\infty)|\rho_{\text{eq}}\rangle\rangle} \tag{8}$$

We obtained the following ($e=1$ for bosons, $e=-1$ for fermions),

$$\mathscr{G}(1,2) = i\hbar\begin{pmatrix} F(1,2) & G(1,2) \\ eG^T(1,2) & {}^{\dagger}F(1,2) \end{pmatrix} \tag{9}$$

where the superscript T indicates the taking of the transpose, and,

$$\begin{aligned} F(1,2) &= \begin{pmatrix} g^c_{hh} & g^<_{hh} \\ g^>_{hh} & g^{ac}_{hh} \end{pmatrix}(1,2), \\ {}^{\dagger}F(1,2) &= \begin{pmatrix} g^c_{ee} & eg^<_{ee} \\ eg^>_{ee} & g^{ac}_{ee} \end{pmatrix}(1,2) \\ G(1,2) &= \begin{pmatrix} G^c & eG^< \\ G^> & eG^{ac} \end{pmatrix}(1,2), \\ eG^T(1,2) &= \begin{pmatrix} eG^{c^T} & eG^{>^T} \\ G^{<^T} & G^{ac^T} \end{pmatrix}(1,2) \end{aligned} \tag{10}$$

$G(1,2)$ corresponds to the Keldysh nonequilibrium Green's function. We refer to the $\mathscr{G}(1,2)$ simply as moment quantum distribution function. Moments are defined for time-ordered quantum field super-operators. We also define quantum correlation functions or quantum cumulants, $\mathscr{K}$, analogous to the classical statistical theory. This distinction is important in treating the quantum transport of superfluids and quanta of real classical fields. We will also refer to both as generalized quantum distribution functions (GQDF).

In the application of Wick's theorem (for superfluid Bose system it is assumed that $\mathscr{L}_0$ is expanded about the condensate for Wick's theorem to be applicable) to evaluate the $\mathscr{G}$'s in terms of diagrams or graphs, $\mathscr{K}(1,2,\ldots n)$ represents the topologically distinct 'connected' subset of graphs. We have

$$\mathscr{K}(1,2,\ldots,n) = (i\hbar)^n \frac{\delta^n \ln\langle\langle 1|\bar{S}(\infty,-\infty)|\rho_{\text{eq}}\rangle\rangle}{\delta u(n)\,\delta u(n-1)\ldots\ldots\delta u(1)} \tag{11}$$

Similar functional derivative relations can be obtained between GQDF with even number of indices by using the variation with respect to the external Schwinger source term $u(1,2)$. Since $u(1,2)$ is an ordinary c-number, the order of the $u(i,j)$'s is not critical in taking the functional derivatives.

4.1. Self-Consistent Equations for Generalized Quantum Distribution Functions

We have

$$\left[\mathscr{G}^0(12)^{-1} - \breve{\Sigma}(12)\right]\left\{\frac{\mathscr{K}(23)}{i\hbar}\right\} = \delta(13) \tag{12}$$

where $\mathscr{G}^0(\xi'',2)^{-1} = (\tau_4^e)^{-1}i\hbar\,\delta(\xi'',2)\partial/\partial t_2 - \mathscr{V}(\xi'',2)$, where the (τ_4^e)-matrix arise from the commutation relation of the $\Psi(i,j)$'s, $\mathscr{V}(\xi'',2)$ is a one-body potential matrix, and $\breve{\Sigma}(12)$ is the particle super self-energy matrix. For fermions, $\mathscr{G}(1,2) \equiv \mathscr{K}(1,2)$. $\breve{\Sigma}(12)$ is expressed in terms of $\mathscr{K}$'s (up to second-order cumulants) and vertex functions. These vertex functions obey equations similar to the Dyson equation, involving functional derivative of the self-energy with respect to second-order GQDF hence decoupling the BBKGY hierarchy. The self-energy due to e-e interaction includes the electron-plasmon vertex function.

5. QUANTUM TRANSPORT EQUATIONS

The time-evolution equation for $\mathscr{K}(i,j)$ is obtained. We write the resulting equations for

the 2×2 matrix elements of $\mathscr{K}(i, j)$ as

$$G^{0^{-1}}\tilde{G} = \delta + \Sigma\tilde{G} + \Delta^{\dagger}\mathfrak{J} \tag{13}$$

$$G^{0^{-1^T}}\tilde{G}^T = \delta + \Sigma^T\tilde{G}^T + {}^{\dagger}\Delta\mathfrak{J} \tag{14}$$

$$G^{0^{-1^T}}\,{}^{\dagger}\mathfrak{J} = \Sigma^T\,{}^{\dagger}\mathfrak{J} + e^{\dagger}\Delta\tilde{G} \tag{15}$$

$$G^{0^{-1}}\mathfrak{J} = e\Delta\,\tilde{G}^T + \Sigma\mathfrak{J} \tag{16}$$

where Δ is the pair potential or gap function, and $G^{0^{-1}}$ is a diagonal matrix with elements proportional to $i\hbar\delta(12)\,\partial/\partial t_2 - \bar{v}(12)$ with $\bar{v}(12)$ proportional to a one-body external potential. These equations were also given by Aronov, *et al.* [17a] and formally resemble the well-known Gorkov equations [17b] for superconductors at thermal equilibrium.

Solving quantum transport problems [11] centers on the evolution of $\rho^{<}(12) \equiv ei\hbar\,G^{<}(12)$, which happens to be one of the matrix elements of the nonequilibrium matrix Green's function of Eq. (9). $\tilde{G}^{>,<}$ contains all information about the statistical aspects of the field intensity. This is coupled to the 'advanced' and 'retarded' propagators often directly related to the experiment and contain all the energetics and dynamical information of the system. We obtain the transport equation for $\tilde{G}^{>,<}$ given by the following expressions.

5.1. Nonequilibrium Superconductivity

We have the following expressions for superconductors,

$$\begin{aligned} i\hbar\left(\frac{\partial}{\partial t_1} + \frac{\partial}{\partial t_2}\right)G^{>,<}(12) &= [\bar{v} + \mathrm{Re}\Sigma^r,\, G^{>,<}] \\ &- [\mathrm{Re}\,G^r,\, \Sigma^{>,<}] \\ &- \frac{i}{2}\{\Gamma,\, G^{>,<}\} + \frac{i}{2}\{A,\, \Sigma^{>,<}\} \\ &+ \{\Delta^r_{hh}\, g^{>,<}_{ee} + g^{>,<}_{hh}\Delta^r_{ee}\} \\ &+ \{\Delta^{>,<}_{hh}\, g^a_{ee} + g^a_{hh}\Delta^{>,<}_{ee}\} \end{aligned} \tag{17}$$

$$\begin{aligned} i\hbar\left(\frac{\partial}{\partial t_1} + \frac{\partial}{\partial t_2}\right)G^r(12) &= [(\bar{v} + \Sigma^r),\, G^r] \\ &+ [\Delta^r_{hh}g^r_{ee} - g^r_{hh}\Delta^r_{ee}] + \{\Delta^{<}_{hh}\, g^r_{ee} + g^r_{hh}\Delta^{>}_{ee}\} \\ &- \{\Delta^{<}_{hh}g^r_{ee} + g^r_{hh}\Delta^{<}_{ee}\} + [g^{>}_{hh} - g^{<}_{hh}]\Delta^r_{ee} - [g^{>}_{hh} \\ &- g^{<}_{hh}][\Delta^{>}_{ee} - \Delta^{<}_{ee}] \end{aligned} \tag{18}$$

Equations for the pair wavefunctions, $g^{>,<}_{ee}$, g^r_{ee}, $g^{>,<}_{hh}$, g^r_{hh}, etc, are also obtained. Using lattice Weyl transformation, we can transform the above total time evolution equations, Eqs. (17) – (18) into quantum transport equations in (**p**, **q**, E, t) phase space for superconductive systems.

5.2. Nonequilibrium Bose Superfluids

The expressions for $i\hbar(\partial/\partial t_1 + \partial/\partial t_2)\,\tilde{G}^{>,<}(12)$, $i\hbar(\partial/\partial t_1 + \partial/\partial t_2)\tilde{G}^r(12)$, and pair-wavefunction equations, are similarly obtained by taking $e = 1$ in Eqs. (13) – (16) and applying the self-consistency condition for the boson-particle self-energies. We can also transform the total time evolution equations into quantum transport equations in (**p**, **q**, E, t) phase space for Bose superfluid systems [18]. This will not be given in this paper.

6. ULTRA-FAST DYNAMICS OF EXCITED SEMICONDUCTORS

The physics of highly-excited semiconductor heterostructures has been of continuing research interest [19–23]. The electron–hole (e–h) pair theory is formally identical to the theory of superconductivity for extremely high-density of e–h pairs. However, in the low density limit real electron–hole pair bound states do occur, these are localized composite e–h elementary excitations called 'excitons' [23]. The *device physics* [24] of interacting matter and radiation requires the consideration of all ranges of e–h densities and short-time dynamics.

The multi-component (in the "hat" and "tilde" indices) quantum field super-operators for electro-

magnetic field (transverse and longitudinal) and lattice vibrations are given, respectively, by

$$A(\zeta)=\frac{1}{\sqrt{4\pi c}}\begin{pmatrix}\hat{A}(\zeta)\\ \tilde{A}(\zeta)\end{pmatrix},\quad \Theta(\zeta)=\begin{pmatrix}\hat{\vartheta}(\zeta)\\ \tilde{\vartheta}(\zeta)\end{pmatrix},\quad \text{and } U(\xi)=\begin{pmatrix}\sqrt{m_k}\hat{u}(\xi)\\ \sqrt{m_k}\tilde{u}(\xi)\end{pmatrix}. \tag{19}$$

We used a "composite field operator" as the fourth field, $\Psi(1,2)$, since coupling to fermions only occur through bilinear product of fermion fields. This is made up of $\mathscr{F}(12)$, $\Re(12)$, $e\Re^T(12)$, and ${}^\dagger\mathscr{F}(12)$, consisting of different combination of the "hat" and "tilde" indices. We have,

$$\Psi(1,2)^T=\left(\mathscr{F}(1,2)\ \ \Re(1,2)\ \ e\Re^T(1,2)\ \ {}^\dagger\mathscr{F}(1,2)\right) \tag{20}$$

The time-ordered averages of the components give the familiar quantities, namely,

$$\langle T\Psi(1,2)\rangle=\mathscr{K}(1,2)=\begin{pmatrix}F(1,2) & G(1,2)\\ eG^T(1,2) & {}^\dagger F(1,2)\end{pmatrix} \tag{21}$$

The scalar potential field super-operator $\Theta(\zeta)$ is not an independent field [25].

7. ELECTRON–HOLE AND EXCITON TIME EVOLUTION EQUATION

The super propagator for electron-hole and exciton is defined by

$$\mathscr{K}(\bar{2}\bar{3};\,12)=\frac{\delta\mathscr{K}(\bar{2}\bar{3})}{\delta f^{\text{ext}}(12)} \tag{22}$$

$f^{\text{ext}}(12)$ includes the matrix element of the dipole moment containing the selfconsistent transverse electromagnetic field in the microcavity, for example. In the matrix definition of $\mathscr{K}(12;34)$, the nonequilibrium pair propagator is identified as one of the matrix elements, namely, the '(1,4th)'-component.

7.1. Schrödinger–Wannier Equation for the Exciton Wavefunction

It is helpful to derive from the Bethe-Salpeter equation, the Schrödinger–Wannier equation for the electron-hole pair wavefunction. Neglecting the self-energy terms and retaining only the Coulomb interaction terms, we obtain after setting $t_1=t_2=t$

$$\frac{i}{\hbar}\frac{\partial}{\partial t}\varphi(12)=\left[-\left\{\frac{\hbar^2}{2m_1}\nabla_1^2+\frac{\hbar^2}{2m_2}\nabla_2^2\right\}-\frac{e^2}{|r_1-r_2|}\right]\varphi(12) \tag{23}$$

This is the familiar Schrödinger equation for a two-particle system consisting of an electron and a hole. Using a dielectric function to screen the potential, the resulting bound states correspond to exciton states [23].

7.2. Time Evolution Equations for $\mathscr{K}$ (12; 34)

The total time derivative of $\mathscr{K}(12;34)$ is obtained from the Bethe-Salpeter equation, the details will be published elsewhere. The '(1,4th)' element describing the electron–hole pair propagator is denoted by $\mathscr{G}_{\text{exciton}}(12;\,34)$, which in turn contains in its '(1,4)th' element the e–h pair density matrix $\mathscr{G}^<_{\text{exciton}}(12;\,34)$.

7.3. Transport Equation for the Nonequilibrium Pair Propagator

What needs to be done is to single out the equation for the electron-hole pair density matrix, which we denote by $\mathscr{G}^<_{\text{exciton}}(12;\,34)$. The result is very long and will be published elsewhere. Interested readers are encouraged to contact the authors for this detail. The final transport equation is obtained by setting $t_1=t_2$, and $t_3=t_4$ and taking the lattice Weyl transform of the above transport equation [in general double lattice Weyl transform in spatial variables]. To demonstrate this, let us take the simplest approximation of this equation, which we

write us

$$i\hbar\left(\frac{\partial}{\partial t_1}+\frac{\partial}{\partial t_3}\right)\mathscr{G}^{<}_{\text{exciton}}(12,\, t_1;\, 34,\, t_3)$$
$$=\left\{\begin{array}{c}[\bar{\text{v}}(12'')]\mathscr{G}^{<}_{\text{exciton}}(2''2,\, t_1;\, 34,\, t_3)\\ -\mathscr{G}^{<}_{\text{exciton}}(12'',\, t_1;\, 34,\, t_3)[\bar{\text{v}}^{T}(2''2)]\end{array}\right\}$$
$$-\left\{\begin{array}{c}[\bar{\text{v}}(33'')]\mathscr{G}^{<}_{\text{exciton}}(12,\, t_1;\, 3''4,\, t_3)\\ -\mathscr{G}^{<}_{\text{exciton}}(12,\, t_1;\, 34'',\, t_3)[\bar{\text{v}}^{T}(4''4)]\end{array}\right\} \quad (24)$$

Upon taking the lattice Weyl transform and taking the gradient expansion the result is

$$i\hbar\frac{\partial}{\partial t}\mathscr{G}^{<}_{\text{exciton}}(p_1,\, q_1;\, p_2,\, q_2;\, E,\, t)$$
$$=-i\hbar\Big\{\nabla_{p_1}\in(p_1)\bullet\nabla_{q_1}$$
$$\mathscr{G}^{<}_{\text{exciton}}(p_1,\, q_1;\, p_2,\, q_2;\, E,\, t)\Big\}$$
$$+i\hbar\Big\{\nabla_{p_2}\in(p_2)\bullet\nabla_{q_2}$$
$$\mathscr{G}^{<}_{\text{exciton}}(p_1,\, q_1;\, p_2,\, q_2;\, E,\, t)\Big\}$$
$$\frac{1}{(h^3)}\int dp_1' K^{s}_{\bar{\text{v}}}(p_1,\, q_1;\, p_1',\, q_1)$$
$$\mathscr{G}^{<}_{\text{exciton}}(p_1',\, q_1;\, p_2,\, q_2;\, E,\, t)$$
$$-\frac{1}{(h^3)}\int dp_2' K^{s}_{\bar{\text{v}}}(p_2,\, q_2;\, p_2',\, q_2)$$
$$\mathscr{G}^{<}_{\text{exciton}}(p_1,\, q_1;\, p_2',\, q_2;\, E,\, t) \quad (25)$$

The last equation is the appropriate Wigner distribution transport equation of a two-particle distribution corresponding to the Schrödinger–Wannier equation of Eq. (23)

8. TRANSPORT EQUATIONS FOR PHONONS

The "reduced density matrix" for phonons is defined by

$$S^{<}_{ij}(\xi\xi';\, tt')=\frac{i}{\hbar}\langle T\hat{u}_i(\xi,\, t)\tilde{u}_j(\xi',\, t')\rangle \quad (26)$$

The lattice Weyl transform can be cast in the form

$$-iS^{<}_{ij}(k,\, k';\, p'',\, l_c;\, \omega,\, t_c)$$
$$=A_{ij}(k,\, k';\, p'',\, l_c;\, \omega,\, t_c)\; n(p'',\, l_c;\, \omega,\, t_c) \quad (27)$$

Similarly, we can write the phonon "hole" distribution as

$$-iS^{>}_{ij}(k,\, k';\, p'',\, l_c;\, \omega,\, t_c)$$
$$=A_{ij}(k,\, k';\, p'',\, l_c;\, \omega,\, t_c)\;[1+n(p'',\, l_c;\, \omega,\, t_c)] \quad (28)$$

8.1. Phonon Transport Equations

We obtain the following expressions

$$\left(\frac{\partial^2}{\partial t^2}-\frac{\partial^2}{\partial t'^2}\right)S^{>,<}$$
$$=\left[\text{Re}\prod{}^{r}-d^{(2)},\; S^{>,<}\right]+\left[\prod{}^{>,<},\,\text{Re}\,S^{r}\right]$$
$$+i\left\{\text{Im}\prod{}^{r},\, S^{>,<}\right\}-i\left\{\prod{}^{>,<},\,\text{Im}\,S^{r}\right\} \quad (29)$$
$$=\left[\text{Re}\prod{}^{r}-d^{(2)},\; S^{>,<}\right]+\left[\prod{}^{>,<},\,\text{Re}\,S^{r}\right]$$
$$\pm\frac{1}{2}\left\{\prod{}^{>,<},\, S^{<,>}\right\}\mp\tfrac{1}{2}\left\{\prod{}^{<,>},\, S^{>,<}\right\}$$

$$\left(\frac{\partial^2}{\partial t^2}-\frac{\partial^2}{\partial t'^2}\right)S^{r,a}=-\left[d^{(2)},\, S^{r,a}\right]+\left[\prod{}^{r,a},\, S^{r,a}\right] \quad (30)$$

where we have used the definition for the force constants

$$d^{(2)}_{ij}(\xi\xi')=\frac{\phi^{(2)}_{ij}(\xi\xi')}{\sqrt{m_k m_{k'}}} \quad (31)$$

and the following identities

$$i\,\text{Im}\,S^{r}=\frac{S^{>}-S^{<}}{2},\quad i\,\text{Im}\,\Pi^{r}=\frac{\Pi^{>}-\Pi^{<}}{2} \quad (32)$$

What then take the lattice Weyl transform of Eqs. (29) and (30) to obtain the transport equations in $(\vec{p},\, \vec{X}_{l_c},\, \omega,\, t)\equiv(\vec{p},\, \vec{q},\, E,\, t)$ phase-space.

8.2. The Phonon Boltzmann Equation

More revealing equations can be seen by neglecting off-diagonal or "inter branch" terms, and expanding the equations in terms of the gradients. We defined a renormalized kinematic frequency by $\Omega^2_{\lambda\lambda}(p,q,E,t) = \omega^{o}_{\lambda\lambda}{}^2 - \mathrm{Re}\prod^r_{\lambda\lambda}(p,q,E,t)$. The Boltzmann equation for the distribution function of phonons, $n_{\lambda\lambda}(p, q, E, t)$, from vibration branch λ readily follows by neglecting leading quantum corrections. We obtain the familiar interpretation [22] relating $(i/2)\{\hbar\Pi^<_{\lambda\lambda}(p,q,E,t)/E\}$ as the scattering-out rate and $(i/2)\{\hbar\Pi^<_{\lambda\lambda}(p,q,E,t)/E\}$ as the scattering-in rate. By taking the renormalized frequency to be given by the solution of the equality $\Omega_{\lambda\lambda}(p, q, \omega\, t) = \omega$, with this solution denoted by $\omega_\lambda(p, t)$, we finally obtain

$$\frac{\partial}{\partial t}n_\lambda(p,q,\omega_\lambda,t) + \nabla_p\hbar\omega_\lambda(p,t)\bullet\nabla_q n_\lambda(p,q,\omega_\lambda,t) = -\frac{i\prod^>_\lambda(p,q,\omega_\lambda,t)}{2\omega_\lambda(p)}n_\lambda(p,q,\omega_\lambda,t) + \frac{i\prod^<_\lambda(p,q,\omega_\lambda,t)}{2\omega_\lambda(p)}\left[1+n_\lambda(p,q,\omega_\lambda,t)\right] \tag{33}$$

The renormalized group velocity [26] emerges since phonons do not diffuse freely but interact with the environment as well as collide with each other. Similar situation arises in deriving the plasmon and photon Boltzmann equations. We expect the leading term of $\mathrm{Re}\prod^r_{\lambda\lambda}(p,q,E,t)$ to be independent of time. An example of the RHS of Eq. (33) for phonon–phonon interaction can be found in the work of the author [27] and in the study of phonon hydrodynamics and second sound [28].

9. TRANSPORT EQUATIONS FOR PLASMONS

9.1. The Plasmon Boltzmann Equation

The plasmon super-propagator is given by $D_o(\zeta,\zeta') = \delta\langle\Theta(\zeta)\rangle/\delta\rho^{\mathrm{ext}}(\zeta')$. The plasmon Boltzmann equation is obtained by means of gradient expansion. We have

$$\frac{\partial}{\partial t}D^{>,<}_o(p,q,E,t) + \left[\frac{\frac{2}{\hbar^2}p - \nabla_p\mathrm{Re}\prod^r_o(p,q,E,t)}{\frac{\partial}{\partial E}\mathrm{Re}\prod^r_o(p,q,E,t)}\right]\bullet\nabla_q D^{<,>}_0(p,q,E,t) = \mp\frac{i}{2}\left\{\frac{\prod^{>,<}_o(p,q,E,t)}{\hbar\frac{\partial}{\partial E}\mathrm{Re}\prod^r_o(p,q,E,t)}D^{<,>}_0(p,q,E,t)\right\} \pm\frac{i}{2}\left\{\frac{\prod^{<,>}_o(p,q,E,t)}{\hbar\frac{\partial}{\partial E}\mathrm{Re}\prod^r_o(p,q,E,t)}D^{>,<}_o(p,q,E,t)\right\} \tag{34}$$

where we have left out terms involving diffusion in momentum and energy space. We would like to point out the notable use of the renormalized plasmon group velocity given by $\left(\frac{2}{\hbar^2}p - \nabla_p\mathrm{Re}\prod^r_o(p,q,E,t)\right)/\frac{\partial}{\partial E}\mathrm{Re}\prod^r_o(p,q,E,t)$. The use of the renormalized diffusion velocity arises from the same reason as that given in the derivation of the Boltzmann equation for phonons. The right-hand side are the familiar collision terms. The omitted terms describe the kinematics of the dynamical motion of plasmons.

10. TRANSPORT EQUATIONS FOR PHOTONS

We will also neglect the off-diagonal terms in polarization indices. The result to leading order is given by

$$\frac{\partial}{\partial t}F^{>,<}_{\lambda\lambda}(p,q,E,t) + \frac{\Omega_{\lambda\lambda}(p,q,E,t)}{\omega}\nabla_p\hbar\Omega_{\lambda\lambda}(p,q,E,t)\bullet\nabla_q F^{>,<}_{\lambda\lambda}(p,q,E,t) = -\frac{i}{2}\left\{\frac{P^>_{\lambda\lambda}(p,q,E,t)}{\omega}\right\}F^<_{\lambda\lambda}(p,q,E,t) + \frac{i}{2}\left\{\frac{P^<_{\lambda\lambda}(p,q,E,t)}{\omega}\right\}F^>_{\lambda\lambda}(p,q,E,t) \tag{35}$$

where $\Omega^2_{\lambda\lambda}(p,q,E,t) = c^2p^2 - \mathrm{Re}P^r_{\lambda\lambda}(p,q,E,t)$ and we have neglected the leading quantum corrections on the right-hand side. We note that P^r is related to the transverse dielectric tensor [29]. We may

consider $\Omega^2_{\lambda\lambda}(p,q,E,t) > 0$ to constitute the validity of the Boltzmann equation for photons with diffusive term. The use of renormalized group velocity corresponds to the use of refractive index well-known in optics.

A self-consistent Boltzmann transport equation follows by setting, $\Omega^2_{\lambda\lambda}(p,q,\omega,t) = c^2p^2 - \mathrm{Re}P^r(p,q,\omega,t) = \omega^2 > 0$. Upon substituting the solution for ω, which we will denote by $\omega_\lambda\ (p,\ t\ )$, we obtain

$$\frac{\partial}{\partial t}F^{>,<}_{\lambda\lambda}(p,q,\omega_\lambda,t) + \nabla_p\hbar\omega_\lambda(p,t)\bullet\nabla_qF^{>,<}_{\lambda\lambda}(p,q,\omega_\lambda,t)$$
$$= -\frac{i}{2}\left\{\frac{P^{>}_{\lambda\lambda}(p,q,\omega_\lambda,t)}{\omega_\lambda}\right\}F^{<}_{\lambda\lambda}(p,q,E,t)$$
$$+\frac{i}{2}\left\{\frac{P^{<}_{\lambda\lambda}(p,q,\omega_\lambda,t)}{\omega_\lambda}\right\}F^{>}_{\lambda\lambda}(p,q,E,t) \qquad (36)$$

11. CONCLUDING REMARKS

A major challenge in analyzing the ultra-fast dynamics of semiconductor gain material for investigating microcavity lasers is the fact that the electron–hole system properties are strongly affected by many-particle Coulomb interactions, and strong coupling to the light and crystal lattice-phonon fields [24]. In device physics of highly excited semiconductor systems, involving all ranges of e–h pair densities, the exciton, e–h Cooper pairs, and e–h plasma energetics and their accompanying distributions greatly affect the polaritons, phonons, biexcitons, and higher-order 'pairing' (excitonic molecules) distributions, their wavefunctions and energies. For the bosons of real fields, the diffusion velocity is not equal to the group velocity of bare excitations but is defined only by its renormalized value. Quanta of these classical fields interact with the environment or collide with each other as it diffuse in space. The condensate and normal excitation energetics and their accompanying distribution also influence the gap function and energy gap. The full transport equations will be published elsewhere.

Acknowledgement

The author is grateful to Dr. A. K. Rajagopal for helpful discussions. This work is supported in part by the Office of Naval Research.

References

[1] Schmutz, M. (1978). "Real-time Green's functions in many-body problems", *Z. Physik*, **B30**, 97–106; Buot, F. A. and Rajagopal, A. K. (1994). "Quantum transport using Liouvillian quantum-field dynamics and functional approach to self-consistent many-body and scattering effects", in *Proc. Third Int. Workshop on Comp. Electronics*, Corvallis: Oregon State University, pp. 183–186.

[2] Arimatsu, T. and Umezawa, H. (1987). "General structure of non-equilibrium thermo field dynamics", *Prog. Theor. Phys.*, **77**, 53–66; Suzuki, M. (1985). "Thermo field dynamics in equilibrium and non-equilibrium quantum systems", *J. Phys. Soc. Jpn.*, **54**, 4483–4485.

[3] Buot, F. A. (1974). "Method for calculating *Tr* $\mathscr{H}^n$ in solid state theory", *Phys. Rev.*, **B10**, 3700–3705; "Weyl transform and the magnetic susceptibility of relativistic Dirac electron gas", *Phys. Rev.*, **A8**, 1570–1581, 1973 & *Phys. Rev*, **A9**, 211, 1974; "Formalism of distribution-function method in impurity screening", *Phys. Rev*, **B14**, 977–989, 1976; "Magnetic susceptibility of interacting free and Bloch electrons", *Phys. Rev.*, **B14**, 3310–3328, 1976; "Real-space tight-binding and discrete phase-space many-body quantum transport", Superlattices and Microstructures **11**, 103–111, 1992 . See also P. Kasperkovitz,"Wigner–Weyl formalisms for toroidal geometries", *Ann. Phys.*, **230**, 21, 1994 (Note that for crystals with inversion symmetry, there is an odd number of lattice points obeying the Born-von Karman periodic boundary condition).

[4] Buot, F. A. and Jensen, K. L. (1990). "Lattice Weyl–Wigner formulation of exact many-body quantum transport theory and applications to novel quantum-based devices", *Phys. Rev.*, **B42**, 9429–9457; Buot, F. A. "Exact integral operator form of the Wigner distribution function equation in many-body quantum transport theory", *J. Stat. Phys.*, **61**, 1223–1256, 1990.

[5] Schwinger, J. (1961). "Brownian motion of a quantum oscillator", *J. Math Phys.*, **2**, 407–432.

[6] Kadanoff, L. P. and Baym, G. (1962). *Quantum statistical Mechanics*. New York: Benjamin.

[7] Keldysh, L. V. (1964). "Diagram technique for nonequilibrium processes" , *Zh. Eksp. Theor. Phys.*, **47**, 1515, [1965, *Sov. Phys.-JETP* **20**, 1018]. See also V. Korenman, "Nonequilibrium quantum statistics: application to lasers", *Annals Phys.*, **39**, 72–126, 1966; P. Danielewicz, "Quantum theory of nonequlibrium processes, I", *Annals Phys.*, **152**, 239–304, 1995.

[8] Rajagopal, A. K. and Buot, F. A. (1965). "A nonequilibrium time-dependent functional theory based on Liouvillian quantum field dynamics", *Int. J. Quantum Chem.*, **56**, 389–397.

[9] Rajagopal, A. K. and Buot, F. A. (1995). "Nonequilibrium time-dependent functional theory for coupled

interacting fields", *Phys. Rev.*, **B51**, 1883–1897; "Time-dependent functional theory of superconductors", *Phys. Rev.*, **B52**, 6769–6774, 1995.

[10] Rajagopal, A. E. and Buot, F. A. (1996). "Generalized functional theory of interacting coupled Liouvillian quantum fields of condensed matter", *Topics in Current Chemistry 181 Density functional theory II*. New York: Springer Verlag.

[11] Buot, F. A. (1993). "Mesoscopic physics and nanoelectronics: nanoscience and nanotechnology", *Phys. Reports*, **234**, 73–174.

[12] Jensen, K. L. and Buot, F. A. (1991). "Numerical simulation of intrinsic bistability and high-frequency current oscillations in resonant tunneling structures", *Phys. Rev. Lett.*, 66, 1078–1081; Biegel B. A. and Plummer, J. D. "Comparison of self-consistency iteration options for the Wigner function method of quantum device simulation", *Phys. Rev. B*, **54**, 8070–8082, 1996.

[13] Buot, F. A. and Jensen, K. L. (1991). "Intrinsic high-frequency oscillations and equivalent circuit model in the negative differential resistance region of resonant tunneling devices", *COMPEL*, **10**, 241.

[14] Buot, F. A. and Rajagopal, A. K. (1993). "High-frequency behavior of quantum-based devices: equivalent circuit, nonperturbative response, and phase-space analyses", *Phys. Rev.*, **B48**, 17217–17232. See also, "Theory of novel nonlinear quantum transport effects in resonant tunneling structures", *Materials Sc. and Engng.*, **B35**, 303–317, 1995.

[15] Nambu, Y. "Quasi-particles and gauge invariance in the theory of superconductivity", *Phys. Rev.*, **117**, 648–663, 1960 ; see also P.W. Anderson, "Random-phase approximation in the theory of superconductivity", *Phys. Rev.*, **112**,1900, 1958 ; A. K. Rajagopal, "Spin waves in an interacting electron gas", *Phys. Rev.*, **142**, 152, 1966.

[16] De Dominicis, C. and Martin, P. C. (1964). "Stationary entropy principle and renormalization in normal and superfluid systems. I. Algebraic formulation", *J. Math. Phys.*, **5**, 14–30.

[17a] Aronov, A. G., Gal'perin, Yu M., Gurevich, V. L. and Kozub, V. I. in Langenberg, D. N. and Lakin, A. I. "Kinetic approach", Eds., *Nonequilibrium Superconductivity*, New York: North Holland, 1986, pp. 325–376.

[17b] For a discussion on Gorkov equations, see for example A. L. Fetter and J. D. Walecka, Quantum Theory of Many-Particle Systems , New York: McGraw Hill, 1971, p. 183.

[18] Kirkpatrick, T. R. and Dorfman, J. R. (1985). "Transport in dilute but condensed nonideal Bose gas: kinetic equations" , *J. Low Temp. Phys.*, **58**, 301–331.

[19] Yokoyama, H. (1992). "Physics and device applications of optical microcavities", *Science*, **256**, 66.

[20] Keldysh, L. V. (1992). "Coherent excitonic molecules", *Solid St. Comm.*, **84**, 37–43; L. V. Keldysh, "Macroscopic coherent states of excitons in semiconductors", in *Bose-Einstein Condensation*, eds. A. Griffin, D.W. Stoke, and S. Stringari , Cambridge: Univ. Press, 1995

[21] Schmitt-Rink, S., Chemla, D. S. and Haug, H., (1988). "Nonequilibrium theory of the optical Stark effect and spectral hole burning in semiconductors", *Phys. Rev.*, **B37**, 941.

[22] Koinov, Z. G. and Glinskii, G. F. (1988). "A new approach to the theory of polaritons in semiconductors at finite temperatures: local-field effects and crystal optics approximation", *J. Phys. A: Math. Gen.*, **21**, 3431–3450; see also Z. G. Koinov (1990). "Self-consistent approach to the theory of Wannier excitons in polar semiconductors", *J. Phys. Condens. Matter.*, **2**, 6507–6518.

[23] Knox, R. (1963). "Excitons", *Solid State Phys. Suppl.*, **5**, 2.

[24] Jahnke, F. and Koch, S. W. (1995) "Many-body theory for semiconductor microcavity lasers", *Phys. Rev.*, **A52**, 1712–1727.

[25] DuBois, D. F. (1967). "Nonequilibrium statistical mechanics of plasmon and radiation", in *Lectures in Theoretical Physics*, New York: Gordon & Breach, pp. 469–620.

[26] Horie, C. and Krumhansl, J. A. (1964). "Boltzmann equation in a phonon system", *Phys. Rev.*, **A136**, 1397–1407.

[27] Buot, F. A. (1972). "On the relaxation rate spectrum of phonons", *J. Phys. C: Solid State Phys.*, **5**, 5–14.

[28] Benin, D. (1975). "Phonon viscosity and wide-angle phonon scattering in superfluid helium", *Phys. Rev.*, **B11**, 145–149; H. Beck, "On the temperature behavior of second sound and Poiseuille flow", *Z. Physik*, **B 20**, 313–322, 1975 ; B. Perrin, "Sound propagation and vibrational relaxation in molecular crystals", *J. Chemie Phys.*, **82**, 191–197, 1985

[29] Agronovich, V. M. and Konobeev, Yu. V. (1964). "Theory of the dielectric permittivity of crystals", *Sov. Phys.-Solid state*, **5**, 1858.

Authors' Biography

Dr. Felix A. Buot has served on the research staff of the University of London, ICTP (Trieste, Italy), McGill, St. Francis Xavier, Stanford, and Cornell University. He is a Fellow of the Washington Academy of Sciences, and President of the Philippine-American Academy of Science and Engineering. He is a member of the Editorial Board, *Transport Theory and Statistical Physics*. He was a UNDP Consultant, 1993 and 1996. His interests include device performance & reliability, optoelectronics, nonequilibrium quantum theory, multiband dynamics, and physics of computation. He is a Research Physicist at the U.S. Naval Research Laboratory.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 275–282
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Applicability of the High Field Model: A Preliminary Numerical Study

CARLO CERCIGNANI [a], IRENE M. GAMBA[b], JOSEPH W. JEROME[c,*] and CHI-WANG SHU[d]

[a]*Politecnico di Milano, 20133 Milano, Italy;* [b]*Courant Institute, New York University, New York, NY 10012;*
[c]*Department of Mathematics, Northwestern University, Evanston, IL 60208;*
[d]*Division of Applied Mathematics, Brown University, Providence, RI 02912*

In a companion presentation, we have discussed the theory of a mesoscopic/macroscopic model, which can be viewed as an augmented drift-diffusion model. Here, we describe how that model is used. The device we consider for this presentation is the one dimensional GaAs n^+-n-n^+ structure of length 0.8 μm. First, a full Hydro-Dynamic (HD) model, proven reliable when compared with Monte Carlo simulations, is used to simulate the device via the ENO finite difference method. As applied to the full device, the new model is not necessarily superior to traditional Drift-Diffusion (DD). Indeed, when we plot the quantity $\eta = \mu_0 E/\sqrt{kT_0/m}$, where μ_0 is the mobility constant and $E = -\phi'$ is the electric field, we verify that the high field assumption $\eta > 1$, required for the high field model, is satisfied only in an interval given approximately by [0.2, 0.5]. When we run both the DD model and the new high field model in this restricted interval, with boundary conditions of concentration n and potential ϕ provided by the HD results, we demonstrate that the new model outperforms the DD model. This indicates that the high field and DD models should be used only in parts of the device, connected by a transition kinetic regime. This will be a domain decomposition issue involving interface conditions and adequate numerical methods.

Keywords: Augmented drift-diffusion, high field model, domain decomposition, ENO algorithm

1 INTRODUCTION

In previous work, we have demonstrated the robustness of an algorithm (ENO: Essentially Non-Oscillatory) designed for the simulation of the hydrodynamic model for semiconductors over a wide range of parameters. In [5] and [6], n^+-n-n^+ diodes in one dimension and MES-FETS in two dimensions were simulated. The present paper deals with the high field model introduced in [3], and appearing elsewhere in these proceedings [2].

*Corresponding author: Tel.: (847) 491-5575, Fax: (847) 491-8906, e-mail: jwj@math.nwu.edu.

2 THE MODELS

The new model is an extension of the Drift-Diffusion (DD) model. Hence we first describe the DD model in a context for comparison.

2.1 Drift-Diffusion Model in One Dimension with μ Depending On E

The DD model is well documented (see, for example, [4]). It is given by:

$$n_t + J_x = 0, \tag{2.1.1}$$

where the representation in terms of hyperbolic and viscous components is given by

$$J = J_{\text{hyp}} + J_{\text{vis}}, \tag{2.1.2}$$

and

$$\begin{aligned} J_{\text{hyp}} &= -\mu n E, \\ J_{\text{vis}} &= -\tau(n\theta)_x. \end{aligned} \tag{2.1.3}$$

Here, n denotes carrier concentration and J denotes current defined per unit charge modulus. We have separated current components in anticipation of the high field model to follow. Also, the electric field is denoted by E, so that in terms of the electrostatic potential ϕ,

$$E = -\phi_x, \quad (\varepsilon\phi_x)_x = e(n - n_d), \tag{2.1.4}$$

where we use the customary expressions, for doping n_d, dielectric ε, and charge modulus e. Set

$$\tau = \frac{m\mu}{e}, \quad \theta = \frac{k_b}{m} T_0. \tag{2.1.5}$$

τ denotes the relaxation time, m denotes effective mass and ambient temperature is denoted by T_0, with θ given in energy units. Of course, μ is mobility, and we take μ to be dependent upon the electric field E, using the formula (2.42) in [4]:

$$\mu(E) = 2\mu_0 \Big/ \left[1 + \sqrt{1 + 4(\mu_0|E|/v_d)^2}\right]. \tag{2.1.6}$$

Here, v_d has the interpretation of the saturation velocity and μ_0 is the low field mobility. The velocity for the DD model is a derived quantity computed by $v = J/n$.

The device we consider for this presentation is the one dimensional GaAs $n^+ - n - n^+$ structure of length 0.8 μm. The device used is as follows: $x \in [0, 0.8]$; the doping is defined by $n_d(x) = 10^5/\mu\text{m}^3$ in $0 \le x \le 0.1$ and in $0.5 \le x \le 0.8$, and by $n_d(x) = 2 \times 10^3$ in $0.15 \le x \le 0.45$, with a smooth intermediate transition. The boundary conditions are: fixed n at both ends, and fixed ϕ at both ends (with a difference = v bias). Simulations are performed for v bias = 0, 0.5 and 1.0 V, but results are shown for v bias = 1.0 only to save space. Other parameters: $m = 0.065 \times 0.9109$ (10^{-30} Kg), $e = 0.1602$ (10^{-18} C), $k_b = 0.138046 \times 10^{-4}$ (10^{-18} J/Kelvin), $\varepsilon = 13.2 \times 8.85418$ (10^{-18} F/μm).

In the definition of μ, we take $\mu_0 = 4.0$ μm^2/V/ps. We consider $T_0 = 300$ K for which $v_d = 0.6$ μm/ps, which is taken to be the maximum of the velocity in the HD run with $v$bias = 1.0.

The results are shown in Figure 1. They do not match as well with those of the HD model (with a doping n_d dependent μ), as in the silicon case which is not presented here. The oscillations in the velocity near the left junction are due to the numerical differentiation of rapidly changing quantities.

2.2 The High Field Model

The model can be written as follows:

$$n_t + J_x = 0, \tag{2.2.1}$$

where the representation in terms of hyperbolic and viscous components is now given by

$$J = J_{\text{hyp}} + J_{\text{vis}}, \tag{2.2.2}$$

and

$$\begin{aligned} J_{\text{hyp}} &= -\mu n E + \tau\mu \left(\frac{e}{\varepsilon}\right) n(-\mu n E + \omega), \\ J_{\text{vis}} &= -\tau[n(\theta + 2\mu^2 E^2)]_x + \tau\mu E(\mu n E)_x \end{aligned} \tag{2.2.3}$$

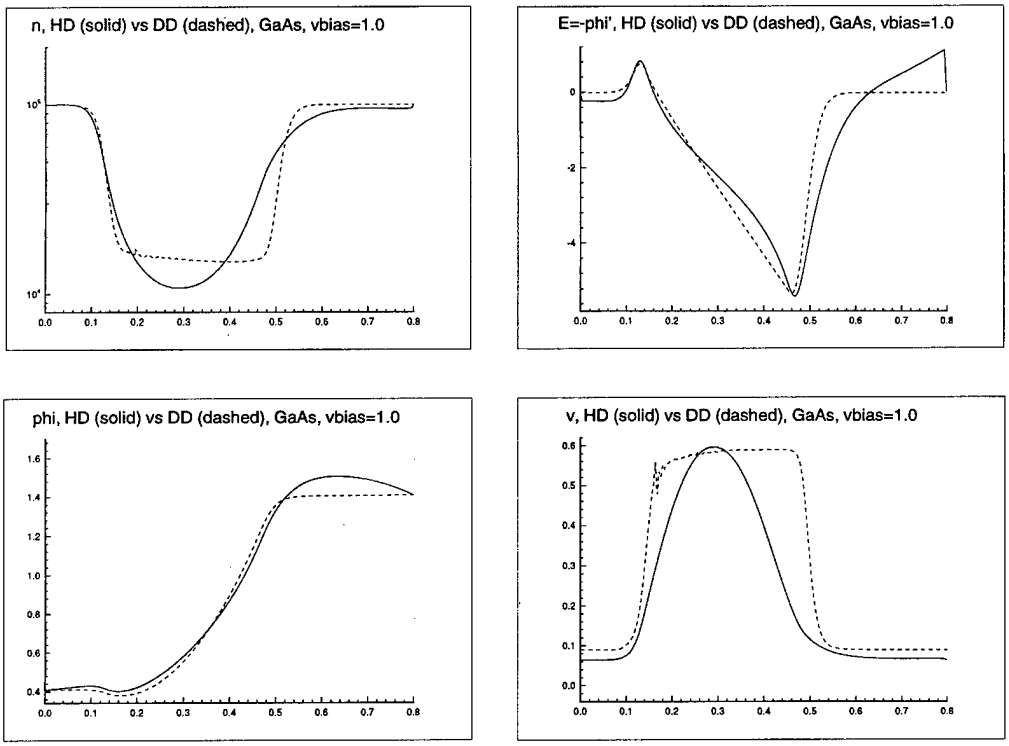


FIGURE 1 Comparison of the results of the HD model with a doping n_d dependent μ (solid line) and that of the DD model with an electric field E dependent μ (dashed line). GaAs, $T_0 = 300$, vbias = 1.0. Top left: concentration n in μm^{-3}; top right: the electric field $E = -\phi_x$ in volts/μm; bottom left: the potential ϕ in volts; bottom right: the velocity v in μm/ps.

Also, the electrostatic equation is satisfied as in the DD model, and τ and θ are defined similarly. Furthermore, ω is taken to be a constant:

$$\omega = (\mu n E)|_{x=0}, \tag{2.2.4}$$

and the velocity is again derived quantity.

We again consider the same GaAs device as before. The results are shown in Figure 2. Not much improvement, if any, is observed over the DD results. The spikes in the velocity near the junctions are due to the derived nature of that quantity; particularly the numerical differentiation of rapidly changing quantities.

3 RESTRICTION TO CHANNEL

We plot the quantity $\eta = \mu_0 E/\sqrt{(kT_0/m)}$ where μ_0 is the mobility constant and $E = -\phi'$ is the electric field (obtained from the HD simulation), in Figure 3.

We can clearly see from Figure 3 that the high field assumption $\eta > 1$ is satisfied only in an interval given approximately by [0.2, 0.5]. We thus run both the DD model and the new high field model in this restricted interval, with boundary conditions of concentration n and potential ϕ provided by the HD results. We can see in Figure 4 that the new model decisively outperforms the DD model.

This indicates that the high field and DD models should be used only in parts of the device, connected by a transition kinetic regime. This will be a domain decomposition issue involving interface conditions and adequate numerical methods, and is currently under investigation.

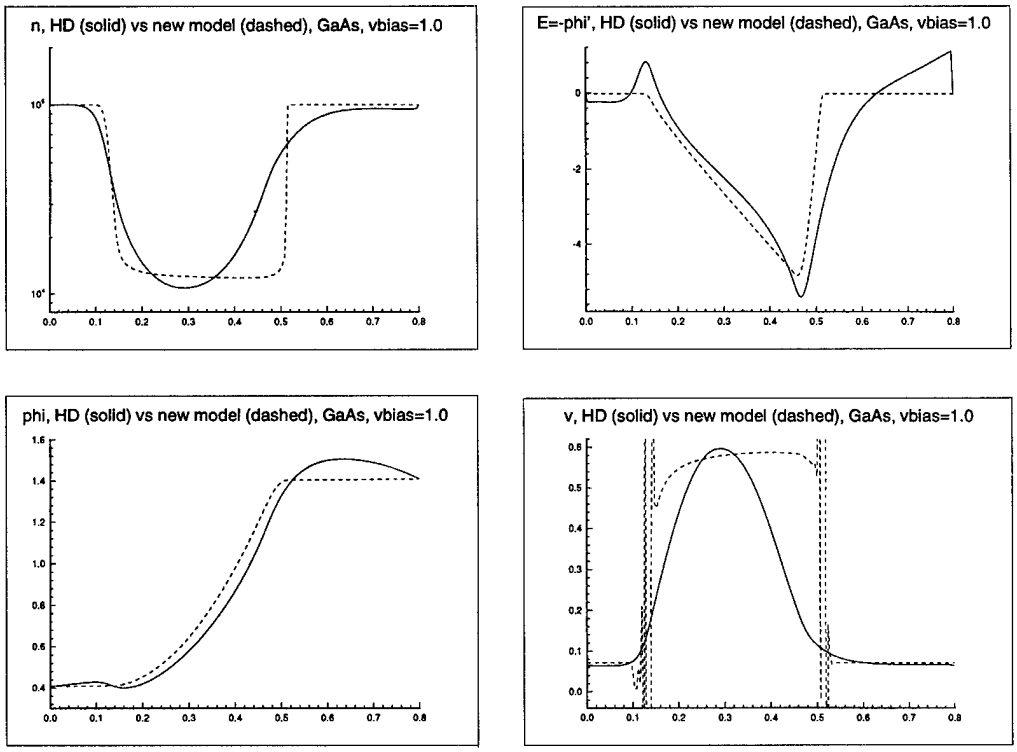


FIGURE 2 Comparison of the results of the HD model (solid line) and that of the new model (dashed line). GaAs, $T_0 = 300$, $v$bias $= 1.0$. Top left: concentration n in μm^{-3}; top right: the electric field $E = -\phi_x$ in volts/μm; bottom left: the potential ϕ in volts; bottom right: the velocity v in μm/ps.

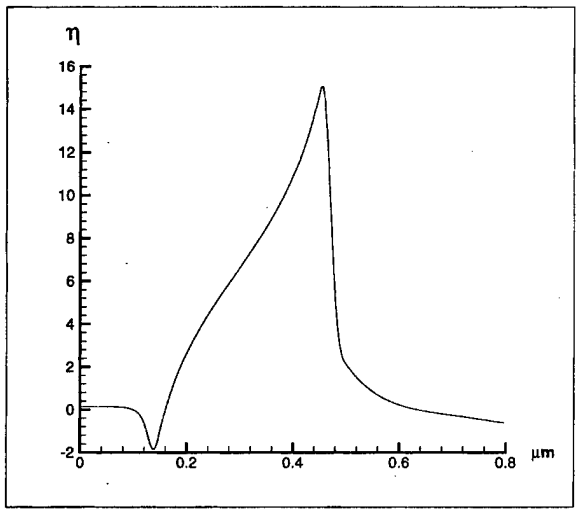


FIGURE 3 The quantity $\eta = \mu_0 E/\sqrt{(kT_0/m)}$, where $E = -\phi'$ is the electric field obtained from the HD simulation.

4 COMPARISON WITH KINETIC AND AUGMENTED DRIFT-DIFFUSION MODELS

The high field model may be thought of as a form of an augmented DD model. In this section, we briefly compare the results of our simulations with those of [1] and [7]. The results of [7] deal primarily with tracking the carrier drift velocity in Silicon. The model is very similar to that originally employed by Thornber [9]. We note that the perturbation terms of the high field model introduced here include more than simply differentiated electric field terms. Comparisons are made in [7] among various models, including DD and augmented DD, as well as the hydrodynamic model and Monte-Carlo simulation. In these studies, the authors of [7] tended to find that standard DD understated drift velocity, even in GaAs. In our own comparisons, when using

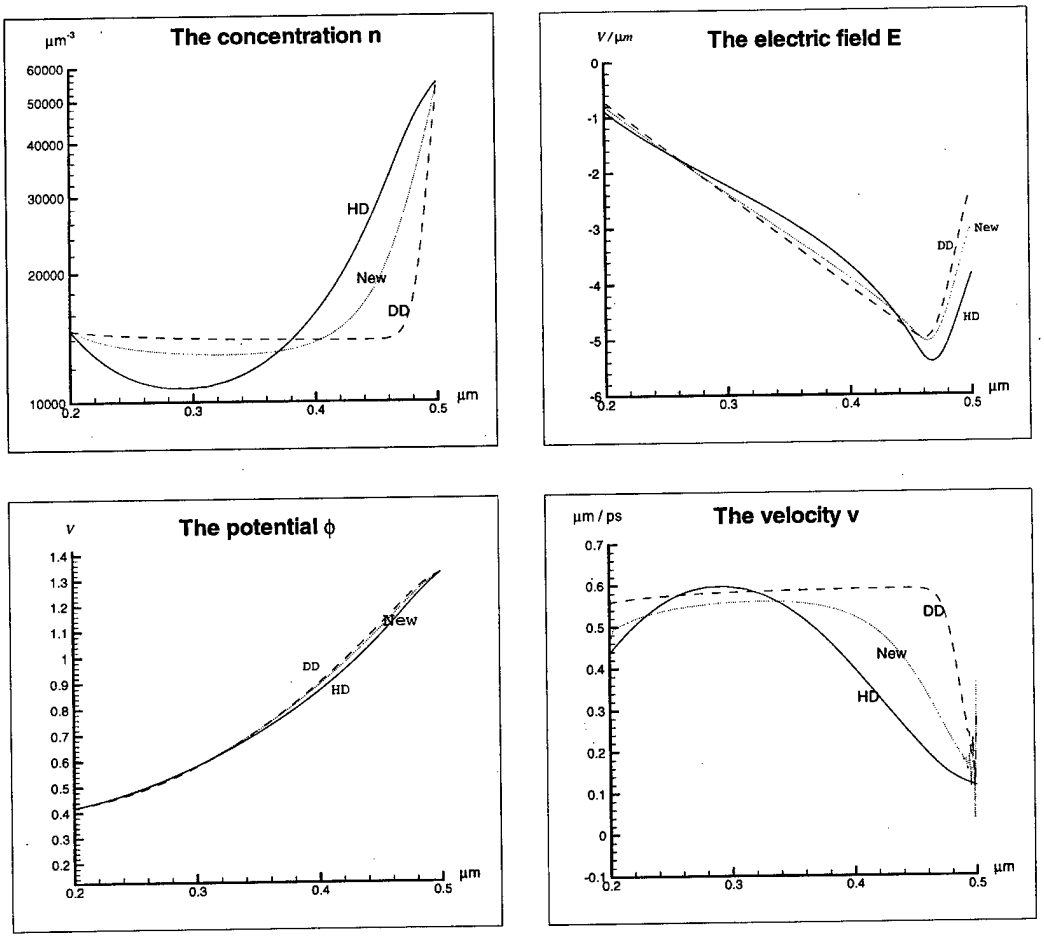


FIGURE 4 The comparison of the Drift-Diffusion (DD) model (dashed line), the new high field model (dotted line), and the HydroDynamic (HD) model (solid line), in the sub-interval [0.2, 0.5] using the HD results as boundary conditions. GaAs, $T_0 = 300$, $v$bias = 1.0. Top left: concentration n in μm^{-3}; top right: the electric field $E = -\phi_x$ in volts/μm; bottom left: the potential ϕ in volts; bottom right: the velocity v in μm/ps.

derived velocity, and a field dependent mobility in DD, we did not experience such a pronounced understatement. We have presented the results of the simulations in Figure 1 for comparison with standard DD, and in Figure 2 for comparison with the high field model. And, remarkably, we find that in the channel, as a problem posed only there, DD overstates the derived velocity! Moreover, the high field model predicts a velocity intermediate between the DD model and the hydrodynamic model.

Paper [1] is difficult to correlate directly because emphasis is placed upon frequency distributions. Nonetheless the average velocity and electric field calculations made there compare favorably with our own. Issues of heating and cooling discussed there are not raised in the present paper, but we intend to consider the effect of energetics in future work. The model has already been developed by the second author.

5 ALGORITHM

We shall only briefly describe the algorithm used in this paper, namely the ENO scheme developed in [8]. The ENO scheme is designed for a system of hyperbolic conservation laws of the form,

$$u_t = f(u)_x = g(u, x, t), \tag{5.1}$$

where $u = (u_1, \ldots, u_m)^T$, and the hyperbolicity condition,

$\frac{\partial f}{\partial u}$ is diagonalizable, with real eigenvalues,

holds. An initial condition is adjoined to (5.1).

For systems of conservation laws, local field by field decomposition is used, to resolve waves in different characteristic directions. For this purpose, analytical expressions are needed for the eigen-values and eigenvectors of the Jacobian matrix. This reduces the determination of the scheme to the case of a single conservation law. Thus, to describe the schemes, consider the scalar one dimensional problem, and a conservative approximation of the spatial operator given by

$$L(u)_j = -\frac{1}{\Delta x}(\hat{f}_{j+\frac{1}{2}} - \hat{f}_{j-\frac{1}{2}}). \tag{5.2}$$

Here, the numerical flux $\hat{f}$ is assumed consistent:

$$\hat{f}_{j+\frac{1}{2}} = \hat{f}(u_{j-1}, \ldots, u_{j+k}); \quad \hat{f}(u, \ldots, u) = f(u). \tag{5.3}$$

The conservative scheme (5.2), which characterizes the $\hat{f}$ divided difference as an approximation to $f(u)_x$, suggests that $\hat{f}$ can be identified with an appropriate function h satisfying

$$f(u(x)) = \int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} h(\xi)d\xi. \tag{5.4}$$

If H is any primitive of h, then h can be computed from H'. H itself can be approximated by polynomial interpolations using Newton's divided difference method, beginning with differences of order one, since the constant term is arbitrary. The necessary divided differences of H, of a given order, are expressed as constant multiples of those of f of order one lower. The main ingredient of the ENO method is the adaptive choice of stencil: it begins with a starting point to the left or right of the current "cell" by means of upwinding, as determined by the sign of the derivative of a selected flux (or the eigenvalue of the Jacobian in the system case); as the order of the divided differences is increased, the divided differences themselves determine the stencil: the "smaller" divided difference is chosen from two possible choices at each stage, ensuring a smoothest fit.

Acknowledgments

The second author is supported by the National Science Foundation under grant DMS-9623037. The third author is supported by the National Science Foundation under grant DMS-9424464. The fourth author is supported by the National Science Foundation under grants ECS-9214488 and ECS-9627849, and the Army Research Office under grant DAAH04-94-G-0205. Computation is supported by the Pittsburgh Supercomputer Center.

References

[1] Barenger, H. U. and Wilkins, J. W. (1987). Ballistic structure in the electron distribution function of small semiconducting structures: General features and specific trends. *Physical Review B*, **36**, 1487–1502.

[2] Cercignani, C., Gamba, I. M., Jerome, J. W. and Shu, C.-W. Applicability of the high field model: An analytical study via asymptotic parameters defining domain decomposition, this volume.

[3] Cercignani, C., Gamba, I. M. and Levermore, C. L. (1997). High field approximations to a Boltzmann-Poisson system and boundary conditions in a semiconductor. *Appl. Math. Lett.*, **10**, 111–117.

[4] Jerome, J. W. (1996). *Analysis of Charge Transport; A Mathematical Theory of Semiconductor Devices*, Springer.

[5] Jerome, J. W. and Shu, C.-W. (1994). Energy models for one-carrier transport in semiconductor devices. In *IMA Volumes in Mathematics and Its Applications*, **59**, Springer, 185–207.

[6] Jerome, J. W. and Shu, C.-W. (1995). Transport effects and characteristic modes in the modeling and simulation of submicron devices. *IEEE Trans. CADICAS*, **CAD-14**, 917–923.

[7] Kan, E. C., Ravaioli, U. and Kerkhoven, T. (1991). Calculation of velocity overshoot in submicron devices using an augmented drift-diffusion model, *Solid-State Electr.*, **34**, 995–999.

[8] Shu, C.-W. and Osher, S. J. (1989). Efficient implementation of essentially non-oscillatory shock capturing schemes, II. *J. Comp. Physics*, **83**, 32–78.

[9] Thornber, K. K. (1983). Current equations for velocity overshoot, *IEEE Electron Device Lett.*, **3**, 69–71.

Authors' Biographies

Carlo Cercignani received his Univ. Milano Laurea Physics in 1961 and his Univ. Milano Laurea Mathematics in 1963. From 1963 to 1966 he was Assistant and Associate Professor Of Plasma Physics at the University of Milano. He was a Visiting Associate Professor of Applied Mathematics at M.I.T. during 1966–67. He came to Politecnico di Milano in 1968 and has been Professor in Theoretical Mechanics since 1975. His scientific specialties include the kinetic theory of gases and its applications to rarefied gas dynamics, neutron transport and semiconductors; nonequilibrium statistical mechanics, fluid dynamics and singular integral equations. His current research interests are analytical and numerical treatment of kinetic and transport equations for gases and semiconductors. He is a member of Accademia Nazionale dei Lincei, Istituto Lombardo and Académie des Sciences de Paris. In addition he is a member of the National Commitees for Mathematics of the Italian Research Council (C.N.R., President), Scientific Committee of the National Group for Mathematical Physics (G.N.F.M.) of C.N.R., a member of General Assembly of IUTAM, General Assembly of IMU, Congress Committee of IUTAM, Scientific Committee of 1st European Mathematical Congress, and Advisory Committee for Symposia on Rarefied Gas Dynamics. He serves on the following editorial boards (1993): Fluid Dynamics Research Transport Theroy and Statistical Physics, MECHANICS Research Communications, Annali di Matematica Pura e Applicata, Applied Mathematics Letters, European Journal of Mechanics B (Ass. Editor), Nonlinear Differential Equations and Applications, Surveys in Mathematics for Industry, Continuum Mechanics and Thermodynamics. He has authored or coauthored seven scientific books and more than two hundred scientific publications. He is the recipient of the Gold Medal for Mathematics of the Accademia dei XL (1982), Prize "Citta' di Cagliari" for Applied Mathematics (1992), Docteur Honoris Causa of the University Pierre et Marie Curie (Paris VI) (1992) and the Humboldt Prize (Academic Year 1995–1996).

Irene M. Gamba received the Ph.D. degree in Mathematics from The University of Chicago in 1989. From 1990 to 1991 she was at Purdue University as a Visiting Assistant Professor and Research Associate. She was an Assistant Professor at The College of New Jersey, Trenton, NJ, during 1991–92. She came to the Courant Institute of Mathematical Sciences, New York University, in 1992. There she was an NSF Postdoctoral Fellow during 1992–94, an Assistant Professor from 1994–96, and Associate Professor during 1996–97. She has accepted a position as Professor of Mathematics at the University of Texas at Austin beginning in Augest, 1997. Her research interests include nonlinear analysis applied to gas dynamics and charged-particle transport systems in the mathematical modeling of microelectronic devices.

Joseph W. Jerome received the Ph.D. degree in Mathematics from Purdue University in 1966. He was visiting Assistant Professor at the Mathematics Research Center, University of Wisconsin, during 1966–68, and was Assistant Professor at Case Western Reserve University during 1968–70. He joined Northwestern University in 1970, where he has been Professor of Mathematics and Applied Mathematics since 1976. He has held sabbatical positions at Oxford University, England, 1974–75, University of Texas, 1978–79, and Bell Laboratories, Murray Hill, 1982–83. He was visiting scholar at the University of Chicago in 1985. He received the Distinguished Alumnus Award from Purdue University's School of Science in 1996. His research interests include applied analysis, numerical analysis, computational electronics, and ion transport in biology. The most recent of his three books, Analysis of Charge Transport, was published by Springer in 1996.

Chi-Wang Shu received the B.S. degree in Mathematics from the University of Science and Technology of China (USTC) in 1982 and the Ph.D. degree in Applied Mathematics from the University

of California at Los Angeles (UCLA) in 1986. He was a postdoctoral fellow at the the Institute for Mathematics and Its Applications (IMA), University of Minnesota, during 1986–1987, and came to Brown Uinversity in 1987, where he has been Professor of Applied Mathematics since 1996. He was a co-recipient of the Chinese Academy of Science Award in Numerical Analysis and Scientific Computing in 1995. His research interests include numerical analysis, scientific computing and computational physics. He currently serves on the editorial boards of Mathematics of Computation, SIAM Journal on Numerical Analysis, and Journal of computational Mathematics.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 283–287
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Simulation of Bistable Laser Diodes with Inhomogeneous Excitation

GANG FANG and TING-WEI TANG*

Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003

A comprehensive time-dependent 1-D computer model is developed for the simulation of multi-section bistable laser diodes. Analysis of a Fabry-Perot (FP) cavity laser diode indicates that the height and shape of the optical input pulse as well as the wavelength play important roles in the set/reset operation of optical switching.

Keywords: Transient simulation, bistable laser diodes, optical switching, set/reset operation, undershoot switching

1. INTRODUCTION

In recent years, BiStable Laser Diodes (BSLDs) have attracted much attention because of their potential applications in optical switching and wavelength conversion. The simulation of BSLDs involves large-signal analysis, which currently existing LD simulators (such as CLADISS [1]) cannot perform. Most studies of BSLDs are conducted using a set of rate equations that are expressed as a function of time only [2]. Such an approach has two main limitations. First, an approximation of distributed loss is required. Second, the spatial dependence of the gain saturation inside the cavity is not taken into account. Most existing models use a single variable to represent the total photon density without considering the difference in wavelength. Moreover, these models often do not include all pertinent physical processes which compete with each other to influence the dynamic characteristics of the device. Therefore a more sophisticated simulator is needed to assist the design as well as the understanding of the physics of BSLDs.

In this work, a comprehensive time-dependent one-dimensional computer model is developed for the simulation of multi-section bistable laser diodes. Longitudinal spatial hole-burning, nonlinear gain, carrier induced refractive index change (frequency chirp), and thermal effects are rigorously accounted for.

* Corresponding author. Fax: 413-545-4611. E-mail: ttang@ecs.umass.edu.

2. PHYSICAL MODEL

A typical longitudinal view of a multi-section laser diode is shown in Figure 1. The laser consists of two gain sections and a saturable absorption section in between.

The optical field $\tilde{E}(z,t)$ can be separated into the forward (+) and backward (−) propagating parts, satisfying the traveling-wave rate equations,

$$\pm\frac{\partial\tilde{E}^{\pm}}{\partial z}+\frac{1}{v_g}\frac{\partial\tilde{E}^{\pm}}{\partial t}=\left(\frac{2\pi}{\lambda}n_r+i\frac{1}{2}G\right)\tilde{E}^{\pm}, \quad (1)$$

where

$$G(\lambda,n(z))=\Gamma g(\lambda,n(z))-\Gamma\alpha_{ac}-(1-\Gamma)\alpha_{cl},$$
$$n_r(\lambda,n(z))=n_{eff}+\Gamma\Delta n_r(\lambda,n(z)),$$

and Γ is the confinement factor, n_{eff} is the effective refractive index determined by waveguide structure, g is the gain function, α_{ac} and α_{cl} represent the absorption coefficients in the active layer and cladding layers, respectively, and Δn_r is the carrier induced change of refractive index.

If we assume $\tilde{E}^{\pm}(z,t)=E^{\pm}(z,t)e^{i\phi^{\pm}(z,t)}$, after substituting into equation (1) and separating the real and imaginary parts, we obtain

$$\pm\frac{\partial E^{\pm}}{\partial z}+\frac{1}{v_g}\frac{\partial E^{\pm}}{\partial t}=\frac{1}{2}G(\lambda,n(z))E^{\pm}, \quad (2)$$

$$\pm\frac{\partial\phi^{\pm}}{\partial z}+\frac{1}{v_g}\frac{\partial\phi^{\pm}}{\partial t}=\frac{2\pi}{\lambda}n_r(\lambda,n(z)). \quad (3)$$

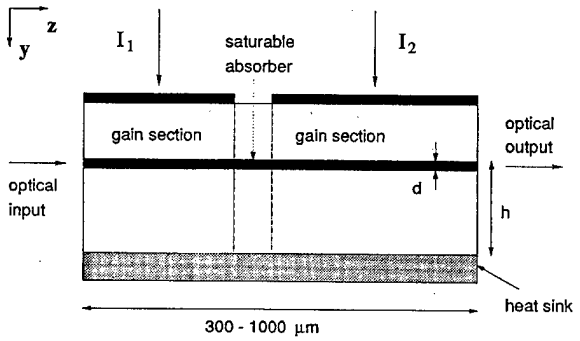


FIGURE 1 Longitudinal structure of a three-section laser diode.

For each longitudinal mode (λ_m), the amplitude and phase satisfy the following boundary conditions at $z = 0$ and $z = L$:

$$E_m^+(0,t)=r_1E_m^-(0,t), \quad (4)$$

$$\phi_m^+(0,t)=\phi_m^-(0,t)+2m\pi, \quad (5)$$

$$E_m^-(L,t)=r_2E_m^+(L,t), \quad (6)$$

$$\phi_m^-(L,t)=\phi_m^+(L,t), \quad (7)$$

where m is an integer, r_1 and r_2 are reflective coefficients of the front and rear facets, respectively. However, for an incident optical field (at $z = 0$) of an arbitrary wavelength (λ_{in}), since the incident optical field $t_1E_{ex}(t)$ interferes with the reflected field $r_1\tilde{E}^-(0,t)$ just inside the left (front) facet, the boundary condition at $z = 0$ becomes

$$\tilde{E}_{in}^+(0,t)=r_1\tilde{E}_{in}^-(0,t)+t_1E_{ex}(t), \quad (8)$$

and at $z = L$, it is given by equations (6) and (7). Under the steady-state condition, equation (3) can be rewritten as

$$\phi^-(0,t)-\phi^+(0,t)=\Delta\Phi, \quad (9)$$

where

$$\Delta\Phi\equiv\frac{2\pi}{\lambda}\cdot\left[2n_{eff}L+2\Gamma\int_0^L\Delta n_r(\lambda,n(z))dz\right] \quad (10)$$

is the round-trip phase change. For the light in the m^{th} mode, $\Delta\Phi_m$ must be equal to $2m\pi$. Thus we can calculate wavelength change (frequency chirp) by solving this nonlinear equation. For the incident light, since its wavelength is determined by the external light source, $\Delta\Phi_{in}$ will change with the carrier density in the active region. Since the round-trip time ($2L/v_g\sim 10\,ps$) is typically much shorter than the response time ($>0.2\,ns$) for devices under the study, we can still use the boundary condition (8) for the time-dependent problems, instead of solving the rate equations for the phases.

If we define P_m (Q_m) as the photon density in m^{th} mode traveling in the $+z$ ($-z$) direction,

$$P_m = P_m(z,t) = \frac{1}{v_g}|E_m^+(z,t)|^2/\hbar\omega_m, \tag{11}$$

$$Q_m = Q_m(z,t) = \frac{1}{v_g}|E_m^-(z,t)|^2/\hbar\omega_m, \tag{12}$$

and take into consideration the spontaneous emission, the traveling-wave rate equations for the photon density become,

$$\frac{\partial P_m}{\partial t} + v_g\frac{\partial P_m}{\partial z} = v_g \cdot G_m(n) \cdot P_m + \frac{1}{2}\beta_{sp} \cdot R_{sp}(n), \tag{13}$$

$$\frac{\partial Q_m}{\partial t} - v_g\frac{\partial Q_m}{\partial z} = v_g \cdot G_m(n) \cdot Q_m + \frac{1}{2}\beta_{sp} \cdot R_{sp}(n), \tag{14}$$

and the corresponding boundary conditions are,

$$P_m(0,t) = R_1 Q_m(0,t), \tag{15}$$

$$Q_m(L,t) = R_2 P_m(L,t), \tag{16}$$

where $R_1 = |r_1|^2$ and $R_2 = |r_2|^2$.

For the incident light (P_{in}, Q_{in}), the rate equations remain the same, while the boundary conditions are different, given by

$$\sqrt{P_{in}(0,t)}\cos(\phi_0^+) = \sqrt{R_1 Q_{in}(0,t)}\cos(\phi_0^-) + \sqrt{T_1 P_{ex}(t),(t)}, \tag{17}$$

$$\sqrt{P_{in}(0,t)}\sin(\phi_0^+) = \sqrt{R_1 Q_{in}(0,t)}\sin(\phi_0^-), \tag{18}$$

$$\phi_0^- - \phi_0^+ = \Delta\Phi_{in}(n), \tag{19}$$

$$Q_{in}(L,t) = R_2 P_{in}(L,t), \tag{20}$$

where $\phi_0^+(t) = \phi_{in}^+(0,t)$, $\phi_0^-(t) = \phi_{in}^-(0,t)$, $T_1 = |t_1|^2$, and $P_{ex}(t) = (1/v_g)E_{ex}^2(t)/\hbar\omega_{in}$.

We assume that all sections are current controlled and that, due to the leakage, only a fraction η of the injected carrier is captured by the active layer where the carriers distribute uniformly in the lateral and transverse directions. No doping is assumed and hence it follows from the charge neutrality condition that the hole density equals the electron density. The rate equation for carrier density $n(z, t)$ in the active layer is given by

$$\frac{\partial n(z,t)}{\partial t} = \frac{\eta J(z)}{qd} - \frac{n}{\tau_{nr}} - R_{sp}(n) - C_{Aug}n^3 - v_g\Gamma \cdot \left\{\sum_m g_m(n)(P_m + Q_m) + g_{in}(n)(P_{in} + Q_{in})\right\}. \tag{21}$$

In our model, the linear gain is approximated by

$$g_L(\lambda,n) = \begin{cases} g_A \cdot (n - n_0), & n < n_{10} \\ g_2 \cdot n^2 + g_1 \cdot n + g_0, & n_{20} > n > n_{10} \\ g_G \cdot (n - n_0). & n > n_{20} \end{cases}$$

where g_A and g_G are the differential gain coefficients in the gain and loss region, n_0 is the transparency carrier density. These parameters can be determined by theoretical calculation or experiment. It should be emphasized that different values of g_A, g_G, and n_0 are used for each longitudinal mode and optical input.

The stimulated emission becomes nonlinear whenever large optical intensities exist in the active layer. To account for this effect a simplified gain model is used, i.e.,

$$g_{NL}(\lambda,n) = \left[1 - \varepsilon_{NL}\sum_{m,in}(P+Q)\right] \cdot g_L(\lambda,n), \tag{22}$$

where ε_{NL} is the material gain compression factor. Then the carrier-induced change of refractive index is assumed to be given by

$$\Delta n_r(\lambda,n) = -\alpha_R\frac{g_L(\lambda,n)}{2k}, \tag{23}$$

where α_R is assumed to be a constant.

3. NUMERICAL SOLUTION METHODS

In this work, we have improved the spatial discretization scheme described in [3]. A first-order backward finite difference formula is applied to $P(z, t)$ and a forward difference applied to $Q(z, t)$. The photon density is defined at the nodes while the carrier density is defined between two adjacent nodes. The change in the carrier density in a region now depends on the average number of photons surrounding that region. A trapezoidal rule/backward-differentiation-formula composite method [4] has been implemented for the time integration. The system of discretized nonlinear equations is solved simultaneously using the full Newton method. The simulation result shown in Figure 4 can be obtained within 20 minutes (CPU time) on a DECstation 5000/200.

4. SIMULATION RESULTS

A two-electrode FP cavity InGaAs/InP laser diode [5] has been simulated. An example of set-reset operation (undershoot switching) using a single wavelength ($\lambda_{in} = 1557.6\,nm$) light is shown in Figure 2. The bias current is chosen to be $I_1 = 18.4\,mA$. The intensities of pulses for set and reset are 0.874 mW and 13.12 mW, respectively.

The mechanism of undershoot switching is believed to be dependent on different changing rates of carrier densities in the gain and absorption regions under a variation of the optical excitation. We have found that the shape of the pulses plays an important role in undershoot switching, as shown in Figures 3 and 4.

From our simulation results, a subnanosecond switch-on (rise) time is observed, while the switch-off (fall) time of the device is about 3 ns. We also found that an input pulse with a higher intensity and faster fall time would reduce the switch-off time somewhat but still no less than 2 ns.

The wavelength of the injected optical pulses is also crucial for the undershoot switching. The minimum set/reset powers will be modulated by the resonant characteristic of the Fabry-Perot cavity as shown in Figure 5. The two curves are almost parallel. Both have minima near the resonant wavelength.

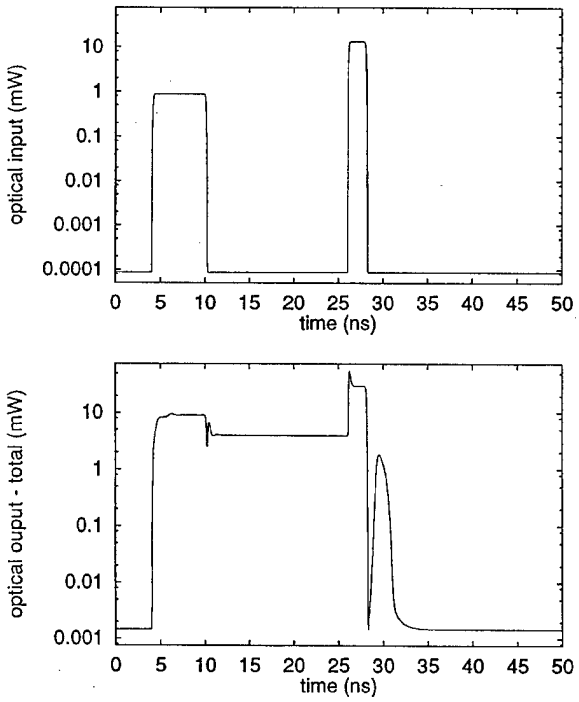


FIGURE 2 The set-reset operation (undershoot shooting): optical input and optical output.

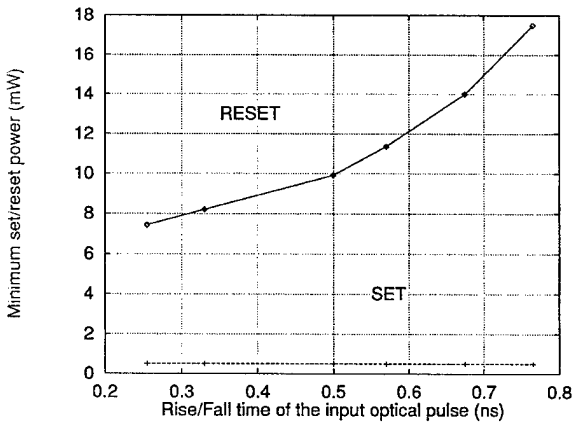


FIGURE 3 Influence of the shape (rise/fall time) of the input optical pulses.

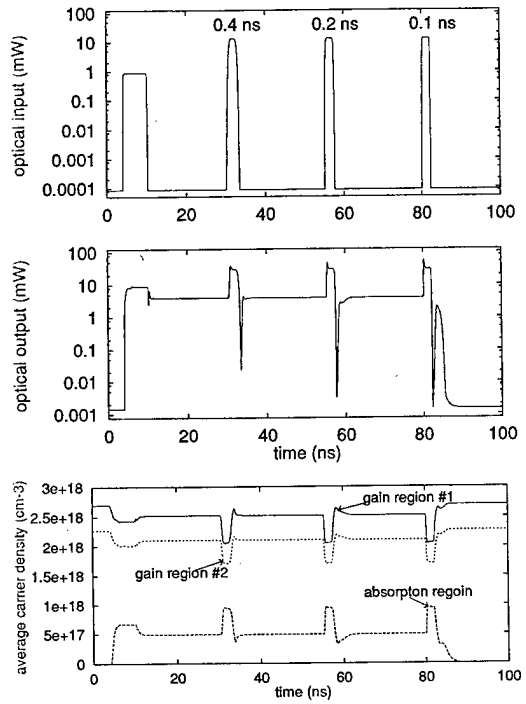


FIGURE 4 Examples of the set-reset operation: the effects of pulse shape on undershooting for $W_{in} = 13.12$ mW, $\lambda_{in} = 1557.6$ nm.

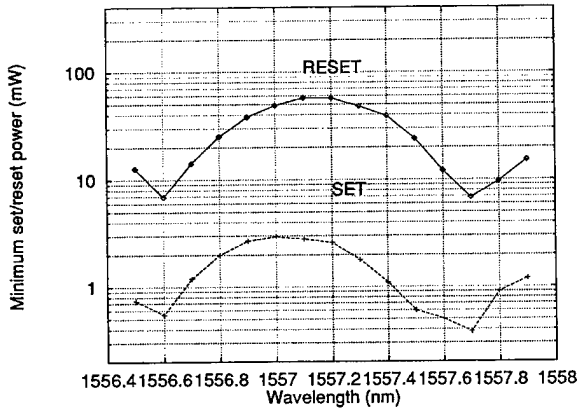


FIGURE 5 Influence of the wavelength of the optical signals on the set-reset operation.

5. CONCLUSION

In this paper, a comprehensive transient 1-D computer model for the simulation of multi-section bistable laser diodes is presented. Traveling wave rate equations are solved for each longitudinal mode and optical input. All the important effects have been included in this model.

Analyses of two-electrode FP cavity InGaAsP/InP laser diodes incorporating an absorber region are carried out to illustrate the capability of the simulator. Special attention is paid to the optical set-reset operation (undershoot switching). Both the height and shape (rise/fall time) as well as the wavelength of the optical input pulse are shown to play an important roles in undershoot switching.

References

[1] Vankwikelberge, P., Morthier, G. and Baets, R. (1990). "CLADISS, a longitudinal, multimode model for the analysis of the static, dynamic and stochastic behaveiour of diode lasers with distributed feedback", *IEEE J. Quantum Electron.*, **26**, 1728–1741.

[2] Kawaguchi, H. (1987). "Progress in optical functional devices using two-section laser diodes/amplifiers" *Optical and Quantum Electron.*, **19**, S1–S36.

[3] Wong, Y. L. and Carroll, J. E. (1987). "A travelling-wave Rate Equation Analysis For Semiconductor Laser", *Solid-State Electron.*, **30**, 13–19.

[4] Bank, B. E., Coughan, W. M., Fichtner, W., Grosse, E. H., Rose, D. J. and Smith, R. K. (1985). "Transient Simulation of Silicon Devices and Circuit", *IEEE Trans. on CAD/ICAS*, CAD-4, 436–451.

[5] Okada, M., Takizawa, H., Kikuchi, H. and Fujikake, H. (1993). "The effect of a Detuned Optical Input on Bistable Laser Diodes with Inhomogeneous Current Injection", *IEEE J Quantum Electron.*, **29**, 109–119.

Authors' Biographies

Gang Fang received his M.S.E.C.E. from University of Massachusetts. He is now working on his doctoral degree at Carnegie Mellon University. His research interests include physical layout automation, circuit analysis, and technology CAD.

Ting-wei Tang is Professor of Electrical and Computer Engineering at University of Massachusetts, Amherst. His research interests include semiconductor device physics and numerical simulation.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 289–293
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Intersubband Relaxation in Step Quantum Well Structures

J. P. SUN[a,*], H. B. TENG[a], G. I. HADDAD[a], M. A. STROSCIO[b] and G. J. IAFRATE[b]

[a] *Solid State Electronics Laboratory, Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, Michigan, 48109-2122;*
[b] *U.S. Army Research Office, P.O. Box 12211, Research Triangle Park, North Carolina, 27709-2211*

Intersubband relaxation due to electron interactions with the localized phonon modes plays an important role for population inversion in quantum well laser structures designed for intersubband lasers operating at mid-infrared to submillimeter wavelengths. In this work, intersubband relaxation rates between subbands in step quantum well structures are evaluated numerically using Fermi's golden rule, in which the localized phonon modes including the asymmetric interface modes, symmetric interface modes, and confined phonon modes and the electron–phonon interaction Hamiltonians are derived based on the macroscopic dielectric continuum model, whereas the electron wave functions are obtained by solving the Schrödinger equation for the heterostructures under investigation. The sum rule for the relationship between the form factors of the various localized phonon modes and the bulk phonon modes is examined and verified for these structures. The intersubband relaxation rates due to electron scattering by the asymmetric interface phonons, symmetric interface phonons, and confined phonons are calculated and compared with the relaxation rates calculated using the bulk phonon modes and the Fröhlich interaction Hamiltonian for step quantum well structures with subband separations of 36 meV and 50 meV, corresponding to the bulk longitudinal optical phonon energy and interface phonon energy, respectively. Our results show that for preferential electron relaxation in intersubband laser structures, the effects of the localized phonon modes, especially the interface phonon modes, must be included for optimal design of these structures.

Keywords: Localized phonons, relaxation rates, intersubband laser

1 INTRODUCTION

In recent years, novel tunneling injection lasers using structures based on semiconductor quantum wells have provided the means for achieving lasing wavelengths from the mid-infrared to the submillimeter wave region [1, 2, 3]. These lasers incorporate narrow quantum well regions that

* Corresponding author.

must have thicknesses as small as 30–50 Å and have transition level separations engineered to one unit or a multiple of bulk Longitudinal-Optical (LO) phonon energy. It is now well known that the shape and energies of LO phonon modes are modified by quantum wells. Specifically, LO phonons in quantum wells may be described by the dielectric continuum model in terms of confined and interface modes that have properties different from those of bulk phonons [4, 5, 6]. In this paper, the interface and confined LO phonon modes of the continuum model [7] are used with Fermi's golden rule to calculate relaxation rates between quasi bound states in step quantum well structures. We show that the electron relaxation rates are dominated by the interface phonon scattering, and that the symmetries of phonon modes induce inherent selection rules of transitions among well levels. In addition, the sum rule for the form factors [6] is verified numerically for the step quantum well structures.

2 INTERFACE PHONON ASSISTED TRANSITION

The well structure in this study is shown in Figure 1, denoting the innermost GaAs to be layer 1, the next $Al_{0.25}Ga_{0.75}As$ layer 2 and the outermost $Al_{0.6}Ga_{0.4}As$ layer 3. There are six types of phonon modes due to the double interfaces between layers 1 and 2: (1) symmetric interface modes, (2) antisymmetric interface modes, (3) confined LO modes, (4) confined TO modes, and (5) half-space LO modes. With the eigenmodes given by [6], we can obtain the electron-phonon interaction from the Fröhlich continuum model

$$H_{e-p} = e\int d^2r \int dz\, \rho(\mathbf{r}, z)\phi(\mathbf{r}, z), \qquad (1)$$

where $\rho(\mathbf{r}, z)$ is the generalized electron density operator and $\phi(\mathbf{r}, z)$ the electrostatic potential created by the LO phonons. Since wavefunctions of the first two levels have opposite parities,

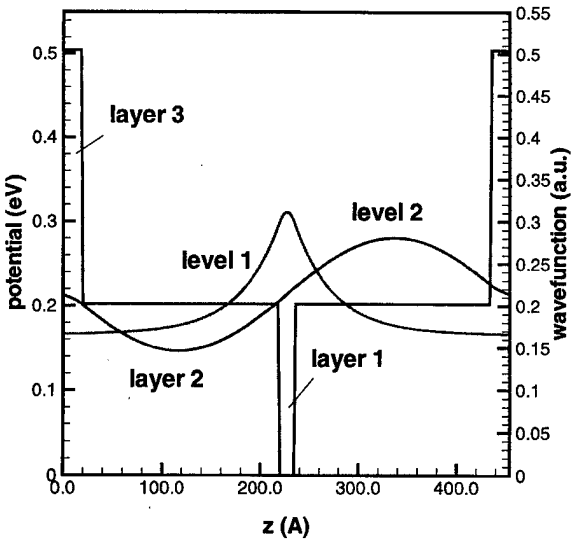


FIGURE 1 Well potential and wavefunctions of the first two levels.

only the antisymmetric modes may contribute significantly to electron transition rates between these two levels. For the antisymmetric phonon modes,

$$H_{A^{\pm}} = \Sigma_q \left[\frac{\hbar\omega_{A^{\pm}}e^2}{2\varepsilon_0 L^2}\right]^{\frac{1}{2}} \left[\beta_1^{-1}(\omega_{A^{\pm}})\coth\left(\frac{1}{2}qa\right) + \beta_2^{-1}(\omega_{A^{\pm}})\right]^{-\frac{1}{2}} \times \frac{1}{\sqrt{2q}} e^{iqr} f_A(q, z)[a_{A^{\pm}}(q) + h.c.], \qquad (2)$$

where q is the two dimensional phonon wavevector parallel to the heterostructure interface,

$$\beta_n(\omega) = \left[\frac{1}{k_{\infty n}} - \frac{1}{k_{0n}}\right]\frac{\omega_{Ln}^2}{\omega^2}\left[\frac{\omega^2 - \omega_{Tn}^2}{\omega_{Ln}^2 - \omega_{Tn}^2}\right]^2, \quad n = 1, 2 \qquad (3)$$

where ω_{Tn} and ω_{Ln} are the frequencies of the longitudinal optical (LO) phonon and the transverse-optical (TO) phonon, respectively. $\omega_{A^{\pm}}$ are the antisymmetric mode frequencies and are given

by solutions of the following equation,

$$\varepsilon_1(\omega)\coth\left(\frac{1}{2}Qa\right)+\varepsilon_2(\omega)=0. \tag{4}$$

For long wavelengths they approach ω_{L1} and ω_{T2}. f_A is defined as

$$f_A(qz)=\begin{cases}-e^{q(z+\frac{1}{2}a)}, & z\le-\frac{1}{2}a\\ \sinh(qz)/\sinh(\frac{1}{2}qa), & -\frac{1}{2}a\le z\le\frac{1}{2}a\\ e^{q(z-\frac{1}{2}a)}. & z\ge\frac{1}{2}a\end{cases} \tag{5}$$

The electron-phonon interaction induces electron transition between the two well levels. Using Fermi's golden rule, the rates of transitions assisted by various modes of the interface phonon and bulk phonon are calculated as functions of electron in-plane energy. In Figure 2, numerical results of the antisymmetric and bulk phonon assisted transitions are given at a well level energy separation of $E_2-E_1=47$ meV, which shows that the antisymmetric interface phonon assisted transition dominates at a level separation close to the antisymmetric phonon frequency. Contributions from the other phonon modes are generally negligible.

The maximum transition rates occur when the electron in-plane energy compensates the difference between the mode frequency and energy level separation. Figure 3 presents a comparison of the maximum transition rates of the bulk mode and antisymmetric mode at various values of E_2-E_1, which are obtained by varying the quantum well width. A device structure which is engineered with an energy level separation of the antisymmetric interface mode frequency would have a relaxation rate several times larger than that with a bulk mode frequency separation, as shown in Figure 3.

For the structure under study, the upper antisymmetric mode plays the most important role in the transition between the energy level 2 and level 1. For other transitions (e.g. between level 3 and level 1) the symmetric modes may become important. The selection rule is due to well

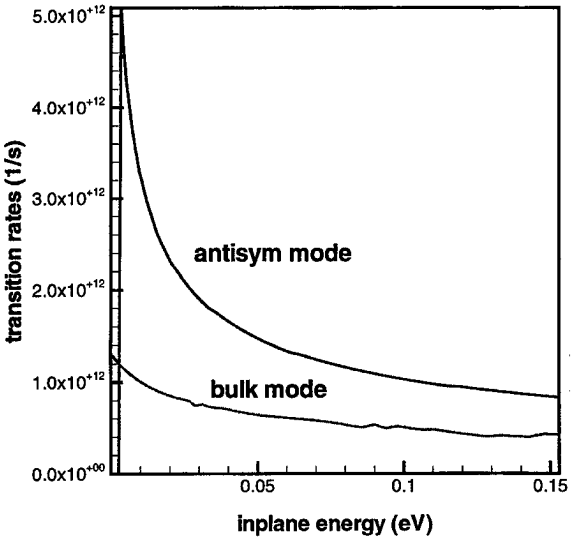


FIGURE 2 Transition rates of antisymmetric and bulk modes.

defined parities of the antisymmetric and symmetric modes. In most cases, it is also applicable to non-symmetric well structures to a certain degree, but the specific phonon modes differ in different device structures and will have to be derived for a specific structure.

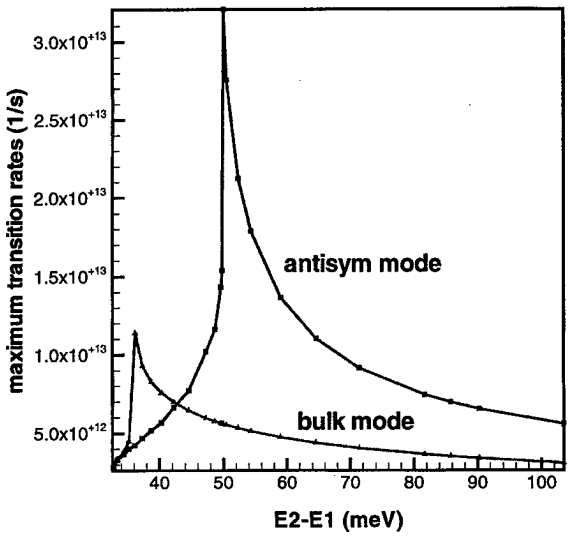


FIGURE 3 Resonant transition rates.

3 SUM RULE OF FORM FACTORS

The form factors [6] can be defined through effective electron–electron interaction due to exchange of optical phonons. The form factors are independent of material parameters but depend on the spatial shape of the phonon modes and subband electron wavefunctions. In this double heterostructure, the form factors for a subband and different phonon modes are given by

$$
\begin{aligned}
F_S(q) &= \frac{1}{1+\tanh(\frac{1}{2}qa)}\left[\int_{-\infty}^{\infty} dz|\psi(z)|^2 f_S(qz)\right]^2,\\
F_A(q) &= \frac{1}{1+\coth(\frac{1}{2}qa)}\left[\int_{-\infty}^{\infty} dz|\psi(z)|^2 f_A(qz)\right]^2,\\
F_C(q) &= \int_{-\frac{a}{2}}^{\frac{a}{2}} dz \int_{-\frac{a}{2}}^{\frac{a}{2}} dz'|\psi(z)|^2|\psi(z')|^2\\
&\quad\left[e^{-q|z-z'|} - \frac{1}{1+\tanh(\frac{1}{2}qa)}\frac{\cosh(qz)\cosh(qz')}{\cosh^2(\frac{1}{2}qa)}\right.\\
&\quad\left. - \frac{1}{1+\coth(\frac{1}{2}qa)}\frac{\sinh(qz)\sinh(qz')}{\sinh^2(\frac{1}{2}qa)}\right],\\
F_{H+}(q) &= \int_{\frac{a}{2}}^{\infty} dz \int_{\frac{a}{2}}^{\infty} dz'|\psi(z)|^2|\psi(z')|^2\\
&\quad\left[e^{-q|z-z'|} - \left[e^{-q|z+z'-a|}\right],\right.\\
F_{H-}(q) &= \int_{-\infty}^{-\frac{a}{2}} dz \int_{-\infty}^{-\frac{a}{2}} dz'|\psi(z)|^2|\psi(z')|^2\\
&\quad\left[e^{-q|z-z'|} - \left[e^{-q|z+z'+a|}\right].\right.
\end{aligned}
\tag{6}
$$

where f_S is given by

$$
f_S(qz) = \begin{cases} e^{q(z+\frac{1}{2}a)}, & z \le -\frac{1}{2}a \\ \cosh(qz)/\cosh(\frac{1}{2}qa), & -\frac{1}{2}a \le z \le \frac{1}{2}a \\ e^{q(z-\frac{1}{2}a)}. & z \ge \frac{1}{2}a \end{cases}
\tag{7}
$$

These form factors always satisfy the sum rule [6]

$$
\begin{aligned}
F_B(q) &= F_S(q) + F_A(q) + F_C(q)\\
&\quad + F_{H+}(q) + F_{H-}(q),
\end{aligned}
\tag{8}
$$

which is a direct consequence of the complete orthonormality of the interface eigenmodes. Verifying the sum rule provides a justification of our numerical calculations. The sum rule between the various form factors has been verified for different well levels of the step quantum well structure. As an example, in Figure 4, the calculated form factors for level 2 are plotted. The sum of form factors of the symmetric mode and half space modes is indeed equal to that of the bulk mode. The form factors for the confined LO mode and antisymmetric mode in this case are negligible compared to those for the symmetric and half space ($H+$ and $H-$) modes, although this is not always true for other levels.

4 CONCLUSION

By examining and comparing electron transition rates induced by bulk and interface phonon modes, our calculations for step quantum well structures designed for intersubband lasers have shown that, to facilitate preferential electron relaxation, it is important to consider the transition induced by the interface phonon modes and have electron subbands engineered with their separation close to the interface phonon energy in the device structures.

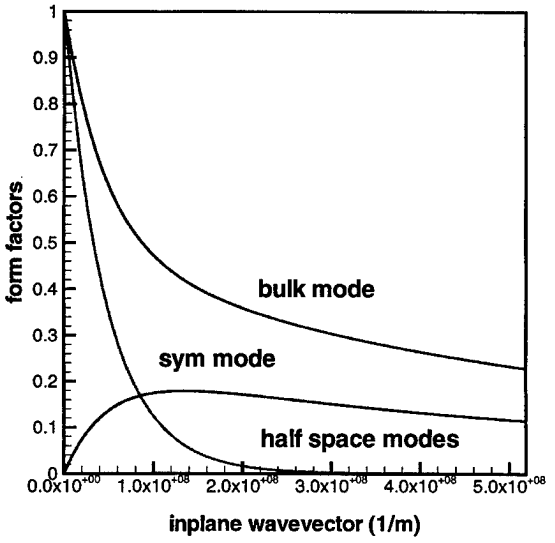


FIGURE 4 Sum rule of form factors.

References

[1] Sun, H. C., Davis, L., Sethi, S., Singh, J. and Bhattacharya, P. (1993). *IEEE Photonics Tech. Lett.*, **5**, 870.

[2] Bhattacharya, P., Singh, J., Yoon, H., Gutierrez-Aitken, A. L., Jambunathan, R., Davis, L. and Sun, H. C. (1994). *IEEE LEOS Newsletter*, 20, August 1995; Davis, L., Sun, H. D., Yoon, H. and Bhattacharya, P. *Appl. Phys. Lett.*, **64**, 3222.

[3] Faist, J., Capasso, F., Sivco, D. L., Sirtori, C., Hutchinson, A. L. and Cho, A. Y. (1994). *Science*, **264**, 553.

[4] Fuchs, R. and Kliewer, K. L. (1966). *Phys. Rev.*, **140**, A2076 (1965); Kliewer, K. L. and Fuchs, R. *Phys. Rev.*, **144**, 495.

[5] Licari, J. J. and Evrard, R. (1977). *Phys. Rev.*, **B15**, 22345.

[6] Mori, N. and Ando, T. (1989). *Phys. Rev.*, **B40**, 6175.

[7] Stroscio, M. A. (1996). *J. Appl. Phys.*, **80**, 6864.

Authors' Biographies

J. P. Sun received his B.S. degree from Nanjing Institute of Technology, China, and M.S. and Ph.D. degrees from the University of Michigan, all in Electrical Engineering. He has worked at Nanjing Solid State Devices Institute, China, Electronics Research Laboratory, University of California at Berkeley, and Solid State Electronics Laboratory, The University of Michigan, where he is now an Assistant Research Scientist. His current research interests include modeling of electronic and photonic devices based on quantum well structures.

H. B. Teng received his B.S. degree from University of Science and Technology of China and M.S. degree from Institute of Theoretical Physics, Chinese Academy of Science, both in physics, and his M.S. degree from the University of Michigan in Electrical Engineering. He is currently a Ph.D. student of The University of Michigan working on electronic and photonic device modeling.

G. I. Haddad received the B.S.E., M.S.E. and Ph.D. degrees in electrical engineering from the University of Michigan. He is currently the Robert J. Hiller Professor and Chairman of the Electrical Engineering and Computer Science Department and Director of the Center for High Frequency Microelectronics, The University of Michigan. His current research areas are microwave and millimeter-wave solid-state devices and monolithic integrated circuits, microwave-optical interactions and optoelectronic devices and integrated circuits. Among numerous awards he received, he is the recipient of the 1996 IEEE-MTT Distinguished Educator Award. He is a member of the American Society for Engineering Education, and the American Physical Society. He is a Fellow of the IEEE and a member of the National Academy of Engineering.

Dr. Michael A. Stroscio has published widely in the field of physics and in recent years has specialized in quantum transport and phonon effects in dimensionally-confined structures. He has published about 140 journal articles. He is the Principal Scientist at he U.S. Army Research Office. Dr. Stroscio is a Fellow of the IEEE and a Fellow of the Yale Science and Engineering Association. He is an Adjunct Professor at Duke University and at the North Carolina State University.

Dr. Gerald J. Iafrate has published widely in the field of solid-state electronics and has made seminal contributions in quantum transport theory. He has published about 140 journal articles in these fields. He is the former Director of the U.S. Army Research Office and the first U.S. Army Senior Scientist of modern times; he was also the Principal Research Scientist at the U.S. Army electronics Technology and Device Laboratory. Dr. Iafrate is a Fellow of the IEEE, a Fellow of the APS, and a Fellow of the AAAS. Dr. Iafrate is a Fellow of the Polytechnic University and he is now Professor of Electrical Engineering Department at Notre Dame University.

VLSI DESIGN
1998, Vol. 8, Nos. (1-4), pp. 295-299
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Resonances in Conductance Through Tunable Attractors

YONG S. JOE* and RONALD M. COSBY

Department of Physics and Astronomy, Ball State University, Muncie, IN 47306, USA

We present the effects of resonant tunneling through doubly-coupled attractors in series in a quantum nanosystem. It is found that multiple split-tunneling peaks arise from resonant tunneling through multiple nondegenerate quasi-bound-states of the coupled attractors. We show that the tunability of coupling is achieved by modulating the strength of the attractor for a *fixed* separation length between the attractors, and by adjusting the geometrical separation length for a *fixed* intervening barrier strength.

Keywords: Nanostructures, Semiconductors, Tunneling

1. INTRODUCTION

Current state-of-the-art techniques result in surface-independent probes or contacts that may be biased to introduce tunable scattering potentials ("impurities") in quantum nanosystems [1]. Many theoretical and experimental works then have studied the elastic scattering by impurities in these systems. In particular, for a quantum structure containing a *finite-size* attractive impurity, called an 'attractor', it is shown that an attractor introduces multiple quasi-bound-states (QBSs) in the channel for a sufficiently strong attractor, and these states give rise to multiple resonant peaks *before* the first plateau in the conductance [2]. These conductance peaks appear due to resonant tunneling through the QBSs.

In this paper, we study the splitting of multiple resonant peaks before the first plateau in the conductance for a quantum nanosystem with two identical attractors in series. The separation of the split-conductance peaks through both the first and the second QBSs depends strongly on the coupling between the two attractors, which can be tuned by modulating the intervening barrier strength [3] and/or changing the geometrical separation length.

2. MODEL AND METHOD

A sketch of the model two-dimensional nanosystem with two identical attractors in series is illustrated in Figure 1(a), where relevant para-

*Corresponding author. Tel.: (765) 285-8879, Fax: (765) 265-5674, e-mail: ysjoe@bsu-cs.bsu.edu.

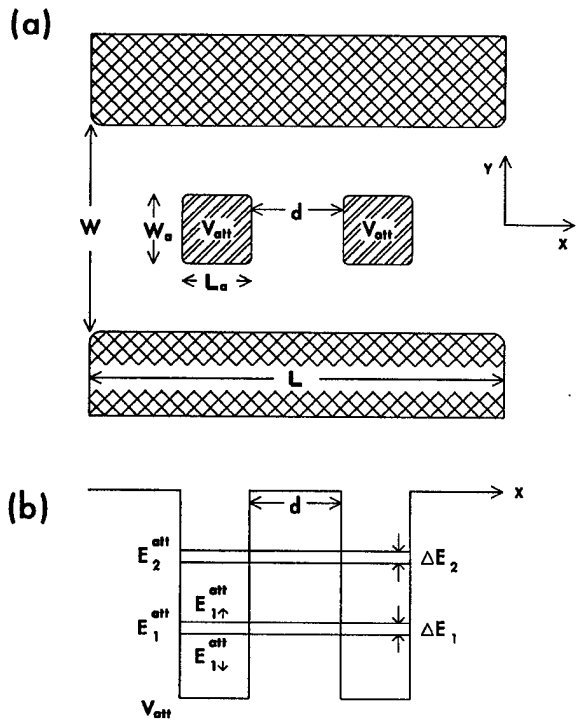


FIGURE 1 (a) Quantum structure studied: Two identical finite-size attractors are coupled with the coupling length d along the transport direction. (b) A schematic potential profile of double attractors along the channel, where two-nondegenerate sets of QBSs are shown in a potential well.

meters are defined. We present results for the case where $W/W_a = 2$, $W_a/L_a = 1$, and the attractors are centered in the channel. We note that similar types of the structure modeled here, using several different techniques, have been realized experimentally [4]. For instance, Geim *et al.* [5] studied resonant tunneling through "donor pair-molecules" with random separation in quantum devices. Figure 1(b) illustrates a schematic potential profile of double attractors along the channel, where two-nondegenerate sets of QBSs are shown in a potential well.

Our calculations of the conductance are based on a nearest-neighbor tight-binding (TB) Hamiltonian and a recursive Green's function technique. Electrons in this system are free to move in the x-direction but are confined in the y-direction, as described by the Schrödinger equation $\{-\hbar^2/2m \nabla_y^2 + V_c(y) - E_n\}\psi_n(y) = 0$, where $V_c(y) = \infty$ for the regions defined by hard walls [shown as cross-shaded regions in Figure 1(a)], and $V_c(y) = -V_{att}$ in the region with the attractors for $|2y| \leq W_a$ [shown as shaded regions], and otherwise $V_c(y) = 0$.

By discretizing the system spatially with lattice constant a, the resulting TB Hamiltonian is exactly solved by using the recursive Green's function method [6]. The Green's function $\mathbf{G}$ of a system in the presence of a perturbation $\hat{\mathbf{V}}$ is related to the Green's function $\mathbf{G}^\circ$ in the absence of $\hat{\mathbf{V}}$. The necessary recursion relations are obtained by taking the appropriate matrix elements of Dyson's equation, $\mathbf{G} = \mathbf{G}^\circ + \mathbf{G}^\circ\hat{\mathbf{V}}\mathbf{G}$. The conductance is then obtained using the Landauer equation [7], $G = (2e^2/h)\,\Sigma_{\mu\nu}|t_{\mu\nu}|^2 = (2e^2/h)\,\mathrm{Tr}\,(\mathbf{t}\mathbf{t}^\dagger)$, where $\mathbf{t}$ is the transmission matrix.

3. RESULTS AND DISCUSSIONS

We investigate conductance peak splitting due to resonant tunneling through multiple QBSs in a quantum system with two attractors in series. Figure 2 shows the conductance G of the structure as a function of the Fermi energy for different attractor strengths. (Here, the Fermi energy is

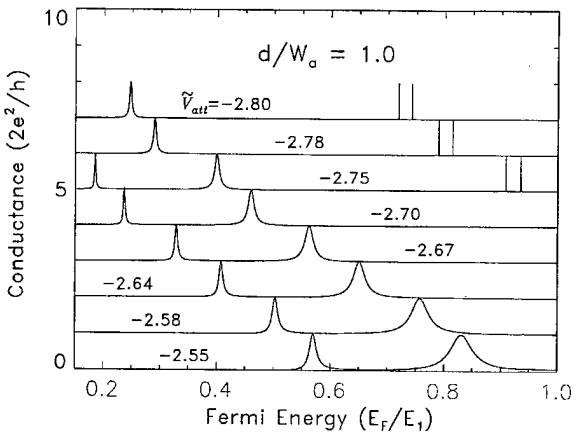


FIGURE 2 Splitting of conductance peaks before the first plateau as a function of the Fermi energy for different attractor strengths $\tilde{V}_{att}$ and a constant coupling length.

normalized with respect to the first subband in the channel W, $E_1 = \hbar^2\pi^2/2mW^2$). These curves are vertically offset one with respect to the other by one conductance unit for clarity. It is clearly seen that resonant tunneling peaks [two for $\tilde{V}_{att}$ $(= \sqrt{V_{att}/E_1})$ in the range from -2.55 to -2.70 and four for $\tilde{V}_{att}$ from -2.75 to -2.80] appear before the first plateau ($E_F/E_1 = 1$) in the conductance. The appearance of these peaks is a consequence of both the formation of multiple QBSs in the attractor [2] [when an incident energy of the electron becomes the resonant energy E_F^{res}, the longitudinal momenta $p_{n,x}$ are enhanced because the QBSs E_n^{att} have negative values $(p_{n,x} = \sqrt{2m(E_F^{res} - E_n^{att})})$], and the overlap of these QBSs due to the coupling between the two attractors [3].

First, we note that the resonance widths of each peak through E_1^{att} and E_2^{att} are very different; a qualitative explanation follows. While E_F varies from 0 to E_1, all subbands E_m in the narrow channel without an impurity are evanescent modes. The mode coupling between E_m and E_n^{att} plays an important role in determining the interaction matrix element, which is a vital factor in the calculation of the transmission coefficient and the conductance. Intermode couplings between E_m and E_n^{att} for (m, n) combinations with even and odd modes yield negligible contributions because the modes have opposite parities. The resonant tunneling through E_1^{att} is mainly due to the coupling between E_1 and E_1^{att}. Similarly, the resonant peak through E_2^{att} is due to the coupling between E_2 and E_2^{att}. However, since the evanescent mode E_2 is always energetically further away from the Fermi energy than E_1, the coupling between E_2 and E_2^{att} is weaker than that between E_1 and E_1^{att}. Therefore, when an electron moves from E_2 to the E_2^{att} state, it stays a longer time in E_2^{att} and barely tunnels through it because of weak coupling. This indicates that the mean lifetime is larger and the width of the tunneling peak is narrower.

Second, we discuss the splitting of tunneling peaks in the presence of coupling between attractors. For two identical series attractors, the single resonant peak through the first QBS of E_1^{att} splits into two well-separated peaks, shown in Figure 2 in the lower range of the Fermi energy for most values of $\tilde{V}_{att}$ displayed. [Here, the first half of the split peaks is not shown for $\tilde{V}_{att} = -2.78$ and -2.80]. In addition to these peaks, a conductance peak through the second QBS of E_2^{att} also splits into two sharp spikes for $\tilde{V}_{att} = -2.75, -2.78$ and -2.80. A more careful examination shows that each tunneling peak through E_2^{att} has a Lorentzian shape centered around the resonance energy [2]. We attribute this peak splitting to the elimination of the energy degeneracy due to the coupling between two attractors. In other words, if the separation between identical attractors is small enough, then the aligned, uncoupled QBSs in each attractor overlap and produce new nondegenerate states of the coupled attractor system. For a simple two-attractor model, one coupled state is slightly higher $(E_{n,\uparrow}^{att})$ and the other slightly below $(E_{n,\downarrow}^{att})$ the uncoupled energy (E_n^{att}) [see Fig. 1(b)], with a perfect tunneling rate indicated (i.e., transmittance through these states reaches one). Therefore, multiple split-conductance peaks arise from the resonant tunneling through multiple nondegenerate QBSs of the coupled attractors.

Next, we discuss the effect of the coupling between the attractors on conductance peak splitting. It is expected that the separation of the split-tunneling peaks in conductance is reduced, as the coupling between two attractors becomes weaker. This result is observed in the double and triple quantum dots, where a coupling (i.e., tunnel barrier) between the dots is controlled by separate gates [8]. Here, we show that the tunability of coupling can also be achieved by modulating (1) the strength of the attractor $\tilde{V}_{att}$ for a *fixed* separation length between the attractors, and (2) the coupling length d (i.e., geometrical separation) for a *fixed* intervening barrier strength.

As the attractor depth $|\tilde{V}_{att}|$ increases, in Figure 2, the separation of the split peaks through both the first and the second QBS becomes smaller. This is clearly seen in Figure 3, where the energy

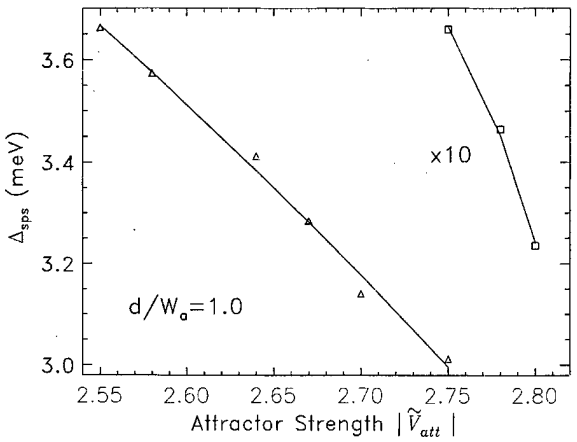


FIGURE 3 The separation of peak splitting Δ_{sps} as a function of $|\tilde{V}_{att}|$ is plotted, for a fixed coupling length $d/W_a = 1.0$, indicating that weak coupling between the attractors can be achieved by increasing $\tilde{V}_{att}$ resulting in strong localized QBSs in the attractor.

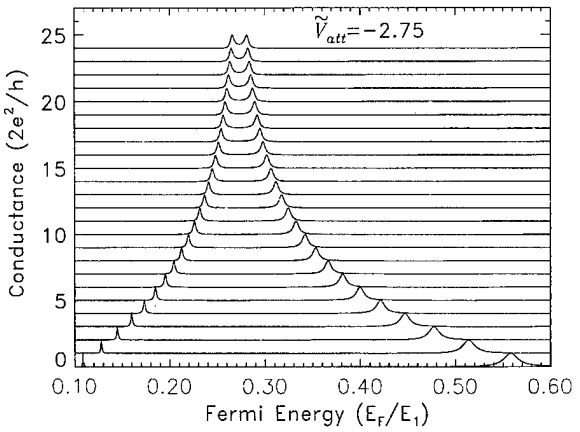


FIGURE 4 The doubly split-conductance peaks through the *first* two-nondegenerate QBSs for different coupling length. The geometrical separation length varies from $d/W_a = 0.5$ (bottom curve) to $d/W_a = 3.0$ (top curve) in steps of 0.1, and the curves are vertically shifted for clarity.

separation of peak splitting, Δ_{sps}, as a function of $|\tilde{V}_{att}|$ is plotted. Here, Δ_{sps} through both non-degenerate states E_1^{att} (triangles) and E_2^{att} (squares; multiplied by a factor of 10 for clarity) decreases, albeit with different slopes. This can be explained by the fact that for a constant geometrical coupling between the attractors, a larger $|\tilde{V}_{att}|$ makes the QBSs lower in energy and more localized in the attractor, and therefore results in less interactions or overlaps between them. As a consequence, the separation of peak splitting in conductance decreases as $|\tilde{V}_{att}|$ increases. We expect that the peak splitting will physically disappear when $|\tilde{V}_{att}|$ is large enough to produce strong localized QBSs in each attractor, i.e., for an extremely weak coupling case.

Figure 4 shows the calculated conductance as a function of the Fermi energy for various coupling lengths. In order to see the effect of the geometrical coupling between the attractors on tunneling peak splitting, we focus on doubly split-tunneling peaks through the *first* two-nondegenerate QBSs ($E_{1,\uparrow}^{att}$ and $E_{1,\downarrow}^{att}$) for a *fixed* $\tilde{V}_{att} = -2.75$. The geometrical separation length varies from $d/W_a = 0.5$ (bottom curve) to $d/W_a = 3.0$ (top curve) in steps of 0.1, and the origin of the curve for different coupling length is shifted vertically in order to make a clear comparison. As d/W_a increases (i.e., the attractors are weakly coupled), the level separation ΔE_1 ($= E_{1,\uparrow}^{att} - E_{1,\downarrow}^{att}$) becomes narrower and hence the separation of the split-conductance peaks through these states is reduced in Figure 4. It is expected that when the attractors are far away from each other, each one has its original QBS E_1^{att}, and these degenerate states give rise to a single resonant peak before the first plateau. Notice that as d/W_a increases, the width of the peak through $E_{1,\uparrow}^{att}$ becomes narrower because the QBS $E_{1,\uparrow}^{att}$ is more bound and localized around the attractor. As expected, this feature can be seen in the opposite way through the QBS $E_{1,\downarrow}^{att}$.

Figure 5 shows how the separation of the split peaks changes with increasing coupling length d/W_a. It appears that the energy separation of peak splitting, Δ_{sps}, decreases exponentially as coupling length d/W_a increases for $\tilde{V}_{att} = -2.75$. Therefore, for the split-conductance peaks through both E_1^{att} and E_2^{att}, the coupling length d, in general, is related to Δ_{sps} as $d \approx -W \ln(\Delta_{sps})$ for different $\tilde{V}_{att}$.

4. CONCLUSIONS

In summary, we have studied the effects of resonant tunneling through doubly-coupled at-

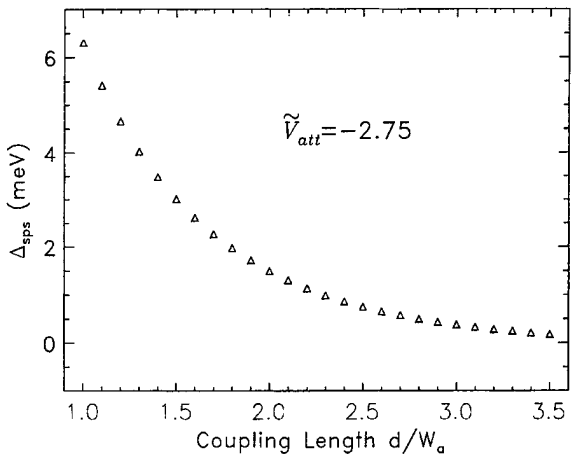


FIGURE 5 A relationship between the separation of the split peaks Δ_{sps} and coupling length d/W_a, showing that Δ_{sps} decreases exponentially as d/W_a increases for $\tilde{V}_{att} = -2.75$.

tractors in series in a quantum nanosystem. We have found that the doubly split-conductance peaks arise from the resonant tunneling through two nondegenerate QBSs of the coupled attractors, and the separation of the split-tunneling peaks are strongly affected by the coupling length and the intervening barrier strength.

Acknowledgements

This work was partially supported by the National Center for Supercomputing Applications under Grant Number DMR930007N and utilized the Convex Exemplar at the University of Illinois at Urbana-Champaign. RMC acknowledges partial support from the Center for Energy Research/Education/Service at Ball State University.

References

[1] Feng, Y., Sachrajda, A. S., Taylor, R. P., Adams, J. A., Davies, M., Zawadzki, P., Coleridge, P. T., Landheer, D., Marshall, P. A. and Barber, R. (1993) *Appl. Phys. Lett.*, **63**, 1666.

[2] Joe, Y. S. and Cosby, R. M. (1997) *J. Appl. Phys.*, **81**, 6217.

[3] Joe, Y. S. and Cosby, R. M. (1997) *Solid State Commun.*, **101**, 731.

[4] Yamada, S. and Yamamoto, M. (1996) *J. Appl. Phys.*, **79**, 8391; Savchenko, A. K., Kuznetsov, V. V., Woolfe, A., Mace, D. R., Pepper, M., Ritchie, D. A. and Jones, G. A. C. (1995) *Phys. Rev. B.*, **52**, R17021; Eugster, C. C., del Alamo, J. A., Melloch, M. R. and Rooks, M. J. (1992) *Phys. Rev. B.*, **46**, 10146; McEuen, P. L., Alphenaar, B. W., Wheeler, R. G. and Sacks, R. N. (1990) *Surf. Sci.*, **229**, 312; Faist, J., Guéret, P. and Rothuizen, H. (1990) *Phys. Rev. B.*, **42**, 3217.

[5] Geim, A. K., Foster, T. J., Nogaret, A., Mori, N., McDonnell, P. J., La Scala, N. Jr., Main, P. C. and Eaves, L. (1994) *Phys. Rev. B.*, **50**, 8074.

[6] Joe, Y. S., Khatun, M. and Cosby, R. M. (1995) *J. Appl. Phys.*, **78**, 7120; Sols, F., Macucci, M., Ravaioli, U. and Hess, K. (1989) *J. Appl. Phys.*, **66**, 3892.

[7] Landauer, R. (1989) *J. Phys. C.*, **1**, 8099.

[8] Waugh, F. R., Berry, M. J., Mar, D. J., Westervelt, R. M., Campman, K. L. and Gossard, A. C. (1995) *Phys. Rev. Lett.*, **75**, 705; (1996) *Phys. Rev. B.*, **53**, 1413.

Authors' Biographies

Young S. Joe received his Ph. D. in Condensed Matter Physics in 1992 from Ohio University, Athens, OH. Following an appointment as a Visiting Scholar at the National Research Council of Canada, he joined the faculty in the Department of Physics and Astronomy at Ball State University in 1993.

Ronald M. Cosby earned the Ph. D. in Solid State Physics at the University of Kentucky in 1971 and holds masters' degrees in physics and computer science. He is a professor of physics at Ball State University, having served on the faculty of the Department of Physics and Astronomy since 1970.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 301–305
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Convergence Properties of the Bi-CGSTAB Method for the Solution of the 3D Poisson and 3D Electron Current Continuity Equations for Scaled Si MOSFETs

D. VASILESKA[a,*], W. J. GROSS[b], V. KAFEDZISKI[a] and D. K. FERRY[a]

[a] *Department of Electrical Engineering Arizona State University, Tempe, AZ 85287-5706, USA;*
[b] *Intel Corp., Chandler, AZ*

As semiconductor technology continues to evolve, numerical modeling of semiconductor devices becomes an indispensible tool for the prediction of device characteristics. The simple drift-diffusion model is still widely used, especially in the study of subthreshold behavior in MOSFETs. The numerical solution of these two equations offers difficulties in small devices and special methods are required for the case when dealing with 3D problems that demand large CPU times. In this work we investigate the convergence properties of the Bi-CGSTAB method. We find that this method shows superior convergence properties when compared to more commonly used ILU and SIP methods.

Keywords: Semiconductor devices, device modeling, Bi-CGSTAB method, discrete impurities, subthreshold conduction

1. INTRODUCTION

The demand of having more transistors on a single chip continues to push MOSFETs toward shorter channel lengths. In order to maintain the long-channel behavior and suppress undesirable short-channel effects, channel doping levels must increase, whereas oxide thickness, junction depths and isolation spacing decrease. To make final threshold voltage adjustments and to eliminate the punch-through effect, ion implantation is used at several steps in the fabrication process of these state-of-the-art devices. The stochastic nature of ion implantation across the wafer leads to significant deviation in both the number of dopants and the arrangement of these dopant atoms from device to device [1, 2]. While a continuum doping model may be a valid assumption for large devices, it fails in the case of deep-submicrometer devices. For example, in MOSFETs with effective channel lengths and widths on the order of 0.1 μm and channel doping of $5\times10^{17}\,\mathrm{cm}^{-3}$, the number of

*Corresponding author.

impurity atoms in the inversion layer is less than 100. Thus, differences in the arrangement of dopant atoms, as well as fluctuations in the number of dopants under the gate, lead to significant variation of the off-state leakage current in these state-of-the-art devices.

The high doping levels used in submicrometer devices lead to large gradients in the potential and charge within the device, which requires a very fine mesh. This makes some of the commonly used iterative methods for the solution of both 3D Poisson and 3D electron current continuity equations rather inefficient because of poor convergence properties. To overcome this, we use a variant of the Conjugate Gradient Squared method (CGS), the so-called Bi-CGS Stabilized (Bi-CGSTAB) method [3]. The convergence properties of this method are found to be superior to those of the Incomplete Lower-Upper (ILU) factorization method [4]. A variant of the ILU method, Stone's [5] Strongly Implicit Procedure (SIP) was also investigated, but found to be comparable to ILU. Finally, we discuss our 3D simulations of the subthreshold characteristics of 0.1 μm gate-length MOSFETs with various gate-widths.

2. APPROACH

The geometrical and structural parameters of the device used in our numerical analysis are illustrated in Figure 1. The depth of the discrete doping region is 0.14 μm which, for substrate doping $N_a = 5\times10^{17}\,\mathrm{cm}^{-3}$ and oxide thickness $t_{ox} = 3\,\mathrm{nm}$, corresponds to more than twice the depth of the depletion region. The 3D Poisson and electron current continuity equation are solved assuming Fermi-Dirac (degenerate) statistics with D_n and μ_n constant.

The boundary conditions for φ, n and p are as follows: Contact regions are assumed to be charge neutral so the electrostatic potential φ and the carrier concentrations n and p at the contacts are prescribed by the usual Dirichlet conditions. At the free surfaces and artificial boundaries, the

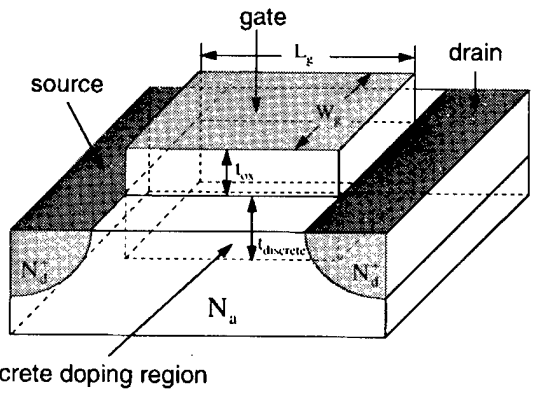


FIGURE 1 Schematic view of the device structure. The box underneath the gate shows the spatial extension of the discrete doping region.

perpendicular field and the perpendicular current are set to zero. We use the decoupled scheme [6] in which, the linearized Poisson equation is repeatedly solved for the electrostatic potential φ. The resulting values of φ are then used to obtain the improved values for n and p. This loop corresponds to one Gummel cycle. The above process is repeated until self-consistent values of φ, n and p are obtained.

The application of finite-difference techniques to both 3D Poisson and 3D electron current continuity equations leads to algebraic equations having a well-defined structure, represented by $\mathbf{Ax} = \mathbf{b}$. The most suitable methods for solution of these two equations are direct methods, but the computational cost becomes prohibitive as the number of equations increases. This has led to iterative procedures that utilize the structure of the coefficient matrix. The ILU methods, and the use of preconditioning together with conjugate gradients, provide a significant increase in the power of iterative methods.

Within incomplete factorization schemes [4], the original matrix **A** is decomposed into a product of lower and upper triangular matrices **L** and **U** respectively. This is achieved by modifying matrix **A** through the addition of a small matrix **N**. Thus, one solves the modified system $\mathbf{LUx} = \mathbf{b} + \mathbf{Nx}$ by solving successively the matrix equations $\mathbf{LV} = \mathbf{b} + \mathbf{Nx}$ and $\mathbf{V} = \mathbf{Ux}$, where **V** is an auxiliary matrix.

The superfluous terms of **N** affect the convergence of the ILU method. Stone [5] suggested the introduction of partial cancellation which minimizes the influence of these additional terms and accelerates the rate of convergence. This partial cancellation procedure is implicitly included in the calculation of **L** and **U**. To provide rapid convergence, the fraction that is canceled is varied on successive pairs of iterations.

The basic CG algorithm loses its applicability when the resulting system of equations is not SPD, as it is the case with the electron current continuity equation. In such circumstances, the best alternatives are Lanczos-type algorithms. In recent years, the CGS method due to Sonneveld [7] has been recognized as an attractive transpose-free variant of the Bi-Conjugate Gradient (Bi-CG) iterative method. The Bi-CGSTAB method due to Van der Vorst [3] is a variant of the CGS algorithm which avoids squaring of the residual polynomial. The convergence behavior of this method is smoother because it produces more accurate residual vectors and, therefore, more accurate solutions. To reduce the spectral conditioning number and improve the distribution of the smallest eigenvalues, one applies preconditioning techniques together with conjugate gradient solvers. Successful preconditioning matrix can be obtained by using ILU factorization [8]: If **L** and **U** are the strictly-lower and the strictly-upper triangular parts of **A**, then the preconditioning matrix is

$$\mathbf{K}_{\mathrm{ILU}} = (\mathbf{L} + \tilde{\mathbf{D}})\tilde{\mathbf{D}}^{-1}(\mathbf{U} + \tilde{\mathbf{D}}), \tag{1}$$

where $\mathrm{diag}(\mathbf{K}_{\mathrm{ILU}}) = \mathrm{diag}(\mathbf{A})$. Once the diagonal $\tilde{\mathbf{D}}$ is computed, one scales the original matrix **A**

$$\tilde{\mathbf{A}} = \tilde{\mathbf{D}}^{-1/2}\mathbf{A}\tilde{\mathbf{D}}^{-1/2} = \mathrm{diag}(\tilde{\mathbf{A}}) + \tilde{\mathbf{L}} + \tilde{\mathbf{U}}. \tag{2}$$

The preconditioning matrix for this symmetrically scaled matrix is of the form

$$\tilde{\mathbf{K}} = (\tilde{\mathbf{L}} + \mathbf{I})(\mathbf{I} + \tilde{\mathbf{U}}), \tag{3}$$

where **I** is the identity matrix. The Bi-CGSTAB method is now applied to the preconditioned system

$$(\tilde{\mathbf{L}} + \mathbf{I})^{-1}\tilde{\mathbf{A}}(\mathbf{I} + \tilde{\mathbf{U}})^{-1}\tilde{\mathbf{x}} = (\tilde{\mathbf{L}} + \mathbf{I})^{-1}\tilde{\mathbf{b}}, \tag{4}$$

where $\tilde{\mathbf{b}} = \tilde{\mathbf{D}}^{-1/2}\mathbf{b}$ and $\mathbf{x} = \tilde{\mathbf{D}}^{-1/2}(\mathbf{I} + \tilde{\mathbf{U}})^{-1}\tilde{\mathbf{x}}$. In the calculation of the product $(\tilde{\mathbf{L}} + \mathbf{I})^{-1}\tilde{\mathbf{A}}(\mathbf{I} + \tilde{\mathbf{U}})^{-1}\tilde{\mathbf{p}}$, significant amount of extra work is avoided by using Eisenstat's trick [9].

In [1], the implantation process was simulated in the following way: n impurity atoms were placed in a region V_{tot} which was 8000 times larger than the discrete doping region V_{disc}. Only those atoms (k) that fell within the discrete doping region were retained. In this way, one generates a binomial distribution [10], which in the limit $n = N_{\mathrm{a}}V_{\mathrm{tot}} \to \infty$, $p = V_{\mathrm{disc}}/V_{\mathrm{tot}} \to 0$ and $np \to a$ (where $a = N_{\mathrm{a}}V_{\mathrm{disc}}$ is the mean value of the process), approaches a Poisson distribution

$$\mathrm{P}(k) \to \frac{a^k}{k!}\mathrm{e}^{-a}. \tag{5}$$

This observation suggests that one can mimic ion implantation by drawing a random number k from the Poisson distribution itself and uniformly distribute these k impurity atoms within the discrete doping region of the device by using triplets of independent uniformly distributed random numbers. This simplification is essential when dealing with devices with wider gates and very high substrate doping.

3. RESULTS AND CONCLUSIONS

The test device used for our convergence studies has 0.1 μm gate length and 0.2 μm gate width. The nonuniform finite-difference grid used in these simulations has 55×40×55 grid points along the x (channel length), y (depth) and z (width) axis. The predetermined threshold for self-consistency for the electrostatic potential is 10^{-5} thermal voltages.

In Figure 2, we compare the convergence properties of various Poisson's equation solvers

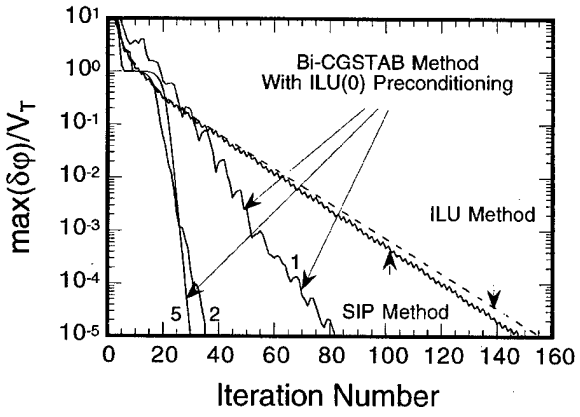


FIGURE 2 Convergence properties of the 3D Poisson equation solvers.

for $V_S = V_D = V_{BS} = 0\,V$ and $V_G = 1\,V$. From the results shown, it follows that preconditioned Bi-CGSTAB method is more economical than the ILU and SIP methods. It gains very little when using more than two internal loops. The use of the ILU preconditioner reduced the number of iterations by approximately a factor of two. The slope of these curves is another indication of the performance of each solver.

The convergence behavior of our continuity equation solvers is shown in Figure 3. The applied bias used in these simulations is $V_S = V_{BS} = 0\,V$, $V_D = 10$ mV and $V_G = 1\,V$. Again , the preconditioned Bi-CGSTAB method shows superior convergence behavior. Gummel's decoupling scheme converged in 21 (13) cycles when using Bi-CGSTAB (ILU) method. However, the computational work within each Gummel cycle for the ILU method is more than 200 iterations, so that the total simulation time for achieving self-consistency using this method is significantly longer.

We also studied the width dependence of the threshold voltage V_{th} in scaled Si MOSFETs. Here, we use 0.1 μm gate-length and 0.05, 0.1, 0.15 and 0.2 μm gate-widths. Statistical averaging of the results obtained over 25 different devices was performed. The doping profile of each device is evaluated using our simplified procedure. Using the bisection method, we determine the value of the gate voltage $V_G = V_{th}$ for which the drain current equals $I_D = I_{off} = 10^{-10}$ A. A set of 20 transfer characteristics for devices with $W_g = 0.2$ μm which have different number and distribution of impurity atoms under the gate is shown in Figure 4. The continuum doping model results are also shown in this figure. For the devices with discrete doping distribution under the gate, the subthreshold slope does not vary significantly from device to device, and is 78.06±1.75 mV/decade. Even though the gate-width of these devices is 4 times larger than that used in [1], we still observe a significant spread of the transfer

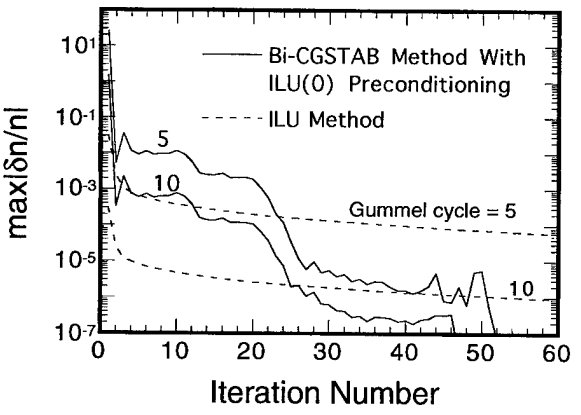


FIGURE 3 Convergence properties of the 3D electron current continuity equation solvers.

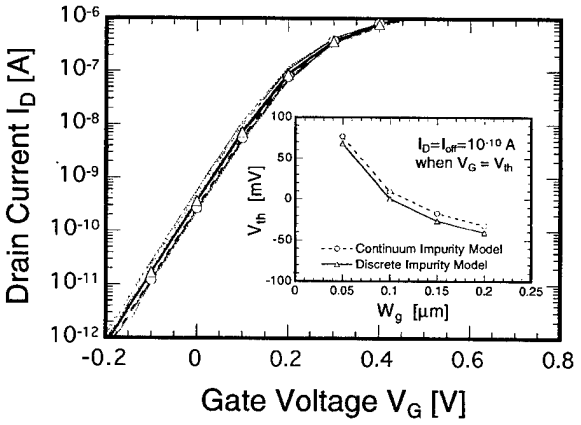


FIGURE 4 MOSFETs subthreshold transfer characteristics. In the inset we show the width dependence of the threshold voltage for subthreshold conduction.

characteristics along the gate axis due to the nonuniformity of the potential barrier which allows for early turn-on of some parts of the channel. This observation is supported by the fact that the averaged threshold voltage is always shifted about 10 mV to lower voltages. As shown in the inset of Figure 4, this threshold voltage shift is essentially independent of the width of the gate.

In conclusion, for solving very large systems of algebraic equations derived from 3D problems, iterative methods remain the only viable solution technique. The greatly improved convergence rates achieved by preconditioned conjugate gradient methods means that the solution of such problems is more attainable now than ever before. For both 3D Poisson and 3D electron current continuity equations, the Bi-CGSTAB method showed relatively smooth and superior convergence behavior when compared to the ILU methods.

Acknowledgments

The authors would like to thank C. Ringhofer, C. Gardner and H. Mittelman, for valuable discussions and comments. This work was supported by the Office of Naval Research.

References

[1] Wong, H.-S. and Taur, Y. (1993). "Three-Dimensional Atomictic Simulation of Discrete Random Dopant Distribution Effects in Sub-0.1 μm MOSFET's", *IEDM Technical Digest*, 705–708.

[2] Zhou , J.-R. and Ferry, D. K. (1995). "3D Simulation of Deep-Submicron Devices", *IEEE Computational Science and Engineering*, **2**, 30–37, summer.

[3] Van der Vorst, H. A. (1992). "Bi-CGSTAB: A Fast and Smoothly Convergent Variant of Bi-CG for the Solution of Nonsymmetric Linear Systems", *SIAM J. Sci. Stat. Comput.*, **13**, 631–644.

[4] Gadiyak, G. V. and Obrecht, M. S. (1986). "The Use of Factorization Methods for Solving the Charge Transfer Equations in Semiconductor Devices", *Simulation of Semiconductor Devices and Processes: Proceedings of the Second International Conference*, 147–156, Ed. by. K. Board and D. R. J. Owen, Swansea UK: Pineridge Press.

[5] Stone, H. L. (1968). "Iterative Solution of Implicit Approximations of Multidimensional Partial Differential Equations", *SIAM J. Numer. Anal.*, **5**, 530–558.

[6] Gummel, H. K. (1964). "A Self-Consistent Iterative Scheme for One-Dimensional Steady State Transistor Calculations", *IEEE Trans. Electron Devices*, **ED-11**, 455–465.

[7] Sonneveld, P. "CGS, a Fast Lanczos-Type Solver for Nonsymmetric Linear Systems", *SIAM J. Sci. Stat. Comput.*, **10**, 36–52, January 1989.

[8] Van der Vorst, H. A. (1989). "High ,Performance Preconditioning", *SIAM J. Sci. Stat. Comput.*, **10**, 1174–1185.

[9] Eisenstat, S. C. (1981). "Efficient Implementation of a Class of Preconditioned Conjugate Gradient Methods", *SIAM J. Sci. Stat. Comput.*, **2**, 1–4.

[10] Picinbono, B. (1993). *Random Signals and Systems*, Englewood Cliffs, New Jersey: Prentice Hall.

Authors' Biographies

Dragica Vasileska is Assistant Professor at Arizona State University. She received her Ph.D. degree in Electrical Engineering from Arizona State University in 1995. Her research interests include semiconductor device physics, 3D device simulations and transport theory.

William J. Gross received the B.S. degree in microelectronic engineering from R.I.T. in 1990 and the M.S. degree in electrical engineering from Arizona State University in 1994. He is presently pursing his Ph.D. in electrical engineering at ASU in the simulation of deep-submicrometer MOSFETs.

Venceslav Kafedziski is currently working toward his Ph.D. degree at Arizona State University. His research interests include digital communications, the applications of information and estimation theory and digital signal processing to wireless communications.

David K. Ferry is Regents' Professor of Engineering at Arizona State University. He is a Fellow of both the American Physical Society and the Institute of Electrical and Electronics Engineers. He pursues the physics and modeling of nanostructured devices.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 307–312
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Wave Function Scarring Effects in Open Ballistic Quantum Cavities

R. AKIS and D. K. FERRY*

Nanostructures Research Group, Center for Solid State Electronics, Arizona State University, Tempe, AZ 85287

Numerical simulations have been carried out of the magneto-transport and corresponding wave functions of two quantum dot structures, namely the stadium and the Sinai billiard. In our simulations, the Schrödinger equation is mapped onto a tight-binding lattice by replacing derivatives by finite differences, and the conductance is calculated via the Landauer formula following the application of an iterative technique to translate across the structure. In both structures, many of the resonance features in the transport show scarring, that is, the amplitude of the corresponding wave functions is highly concentrated along underlying periodic classical orbits. Our analysis indicates that certain periodicities evident in the magneto-transport can be associated with particular orbits.

Keywords: Transport, numerical simulation, quantum simulation, chaos, periodic orbits, finite differences

1. INTRODUCTION

Semiconductor billiards have recently attracted considerable interest as a novel probe of transport and quantum chaos. The devices examined typically consist of a central mesoscopic cavity, connected to external reservoirs by a pair of quantum point contacts through which current is measured. The lithographic size of these dots is usually made much smaller than the mean free path, so that transport is ballistic. While it has been suggested that multiple billiard scattering from cavity walls will induce chaotic behavior [1], recent studies of rectangular billiards have established the basic regular nature of the orbits [2]. In stadium billiards, it is still thought that chaotic behavior is induced, even in open structures [3, 4]. However, recent experiments on stadium shaped dots suggest that their electrical properties are dominated by a characteristic *periodicity* in their magnetoconductance [3, 5]. In the earlier study, it was tentatively suggested that this periodicity arises from the remnants of regular, semi-classical orbits within the dots [3]. In order to investigate

*Corresponding author. Tel.: 602-965-2570, Fax: 602-965-8058, e.mail: ferry@asu.edu.

this effect, we have modeled the dots by solving the quantum mechanical problem on a discrete lattice using a numerically stabilized variant of the transfer matrix approach [6]. The details of this method are summarized in the next section. The section following presents our theoretical results for the stadium as well as a comparison with experiment. Simulating dots of comparable size, we reproduce the periodicity of the fluctuations found experimentally. Moreover, we find that the wave functions of these structures can be heavily scarred by periodic orbits, with certain scars recurring periodically in magnetic field, in good correspondence with selected peaks in the power spectra. In the next section, we consider the Sinai billiard, which consists of a nominal square ballistic quantum dot with a circular anti-dot in the center and is another structure whose classical analog is chaotic. Recent experiments [7] have shown self-similar (fractal) behavior in the magnetoconductance of such structures, appearing in a series of peaks in the resistance. Not only do our simulations show indications of self-similarity on scales similar to that observed experimentally, but that this behavior can in fact *coexist* with the periodic scarring effect associated with regular orbits. The paper concludes with a summary.

2. METHOD OF CALCULATION

As displayed in Figure 1, the typical situation is one in which ideal quantum wires, which extend outward to $\pm\infty$ are connected to the quantum dot, a stadium in this case. This problem can be solved by using an iterative method [6] applied to the discretized version of the Schrödinger equation, obtained by keeping terms up to first order in the approximation of the derivative:

$$(E_F - \mathbf{H}_j)\psi_j + \mathbf{H}_{j,j-1}\psi_{j-1} + \mathbf{H}_{j,j+1}\psi_{j+1} = 0 \quad (1)$$

where ψ_j is a M-dimensional vector containing the amplitudes of the j th slice. The problem is solved on a square lattice of lattice constant a with the

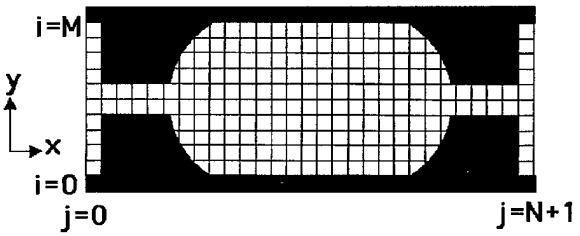


FIGURE 1 The geometry of one of the quantum dots considered in this study. The grid represents the underlying mesh on which the calculations are performed, though in practice the grid is much finer that that shown here.

wires extending M lattice sites across in the x direction and the region of interest being broken down into a series of slices along the y direction. In this equation, the $\mathbf{H}_j$ matrices represent Hamiltonians for individual slices and the matrices $\mathbf{H}_{j,j-1}$ and $\mathbf{H}_{j,j+1}$ give the inter-slice coupling. By approximating the derivative, the kinetic energy terms of Schrödinger's equation get mapped onto a tight-binding model with $t = -\hbar^2/2m^*a^2$ representing nearest neighbor hopping. The potential simply adds to the on site energies. Assuming a flux incident from the left, this equation can be used to derive a transfer matrix which allows us to translate across the system and thus calculate the transmission coefficients which enter the Landauer-Büttiker formula to give the conductance. Transfer matrices however are made unstable due to exponentially growing and decaying contributions of the evanescent modes. This difficulty can be overcome by performing some clever matrix manipulations and calculating the transmission by a iterative procedure rather than just multiplying transfer matrices together [6]. The amplitudes of the wave functions at specific values of x and y can be found easily by backward substitution after the iteration is performed.

3. THE STADIUM

The correspondence between periodicities in the conductance and scarring effects was first suggested as a possibility in the context of stadium

shaped quantum dots by Marcus *et al.* [3]. Provided an ensemble average was *not* performed, they found that a Fourier analysis of their conductance fluctuations revealed the presence of strong peaks at a few discrete frequencies. The stadium is of particular interest, since its closed, classical analog is well known to be chaotic (it should be noted that a closed *quantum* stadium structure has also been achieved [7]). As such, one might expect the behavior to reflect ergodicity, or phase space filling. Such behavior is only recovered after ensemble averaging over *many* conductance traces, each obtained by altering the gate voltage [3, 4]. Importantly, such averaging has the effect of returning the dot, and its point contact leads, to the classical regime, since the quantization is averaged out, and the proper quantum behavior is masked.

The results of a Fourier analysis of a single conductance trace are shown in Figure 2(a) as curve (a), taken from Ref. [3]. In the inset is a micrograph of the actual dot on which the experiments were performed. As shown, the input and output leads were at right angles to each other. The dot size was approximately 0.4×1.0 μm and the electron density was 3.6×10^{11} cm^{-2}. Curve (b) is the Fourier power spectrum obtained from our simulation of a similar sized dot (after accounting for depletion). Both the experiment and the simulation show strikingly similar harmonic content, with well defined peaks at f $\sim$ 20, 33 and 65 T^{-1} in both cases. For the simulation, four modes were allowed to propagate in the leads. In addition to the periodicity, we see scarring effects. Figure 2(b) shows a representative example, a wave function scarred by an underlying "bow-tie" shaped classical orbit, which occurs at $B=0.229$ T. Unfortunately, while other scars do appear, it is difficult to establish a connection between any scar and a particular power spectrum peak, as similar looking scars can be resolved only at a very few other values of field.

Figure 3 shows an example in which the stadium has centrally *aligned* leads, dimensions of 0.4 μm × 0.8 μm and an electron density of 4×10^{11} cm^{-2}.

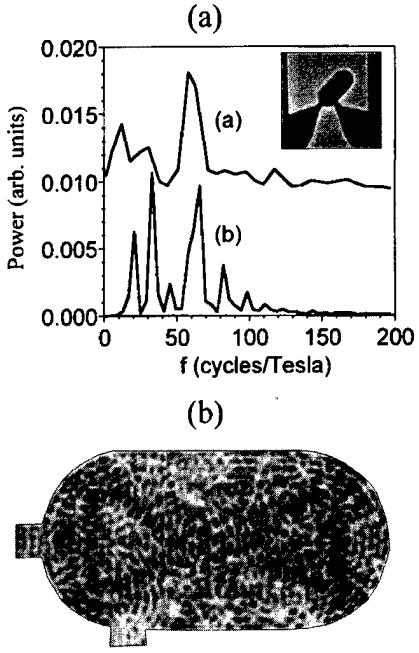


FIGURE 2 In panel (a), the Fourier power spectrum of the conductance fluctuations is plotted. The curve labeled (a) is for the experimental stadium dot shown in the inset (from Ref. [3], with permission). Curve (b) is the result of the quantum simulation. Panel (b) show a "bow-tie" scar that appears in the stadium wave function. Probability density ($|\psi(x,y)|^2$) versus x and y plotted here, with higher amplitude corresponding to darker shading.

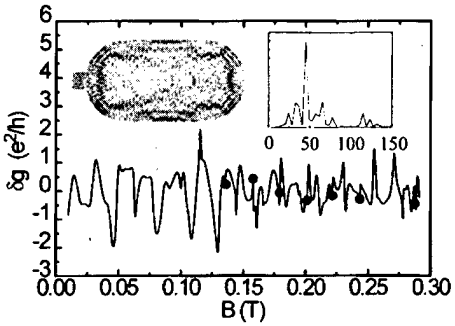


FIGURE 3 Conductance fluctuations, δg, are plotted as a function of magnetic field for a stadium quantum dot with aligned leads. The right inset shows the corresponding power spectrum. The left inset shows a "whispering gallery" scar that recurs at the points indicated by the solid circles.

Here we plot the conductance fluctuations for this configuration and the corresponding power spectrum is shown in the left inset. A single peak at 47

T^{-1}clearly dominates. The corresponding magnetic field period is 0.021 T, in good agreement with the spacings between successive minima in the fluctuations. The solid circles correspond to values of B where a "whispering-gallery" scar was observed in the wave function. An example of such a scar is shown in the left inset, which corresponds to $B=0.288$ T. The quantization of modes in the contacts causes the electrons to be injected in collimated beams directed at well defined angles [2], and it is this effect that is responsible for exciting the orbit seen here. The "whispering-gallery" scars appear at very nearly periodic intervals, with the period also being 0.021 T, thus a strong correspondence can be made between a particular power spectrum peak and a periodic orbit in this case.

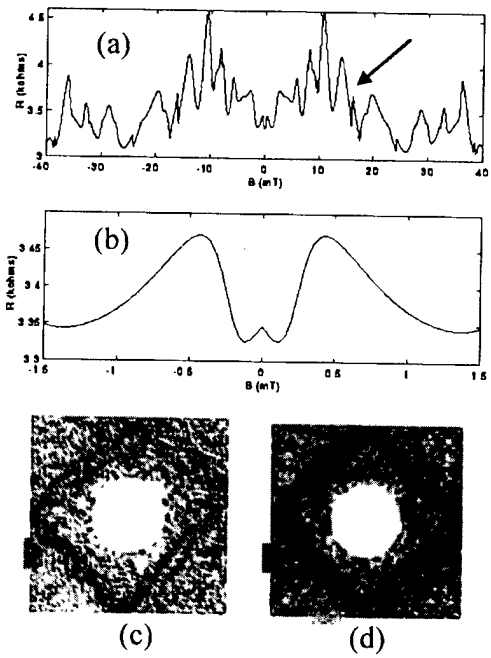


FIGURE 4 (a) A resistance trace obtained from a simulation of a Sinai billiard, with input and output leads are in the bottom left hand corner of the structure (these are seen clearly in (d)). (b) A blow-up of the central region of shown in (a). (c) The scarred wave function at $B=16$ mT, which corresponds to the resonance indicated by the arrow in the top frame. (d) Another diamond scar from the periodic series, this one at $B=49$ mT.

4. THE SINAI BILLIARD

As mentioned in the introduction, the Sinai billiard is another example in which the classical dynamics is expected to be chaotic, at least in the closed structure. The experimental observation of "self-similarity" in the magnetoresistance of such structures has been recently reported [8]. Self-similar or fractal behavior has long been associated with chaos. We have performed simulations of Sinai billiard structures, with configurations similar to the experimental ones, with a quantum dot 1.0 μm square, an anti-dot in the center, two leads in the bottom left-hand size corner (refer to Fig. 4(c)) and an electron density of 2.3×10^{11} cm^{-2}. Figure 4(a) shows a resistance trace obtained from such a simulation, with the leads adjusted to support four propagating modes, and the central anti-dot approximately 0.2 μm in radius. We show this example because, first of all, it shows indications of the self-similar behavior seen experimentally. A blow-up of the central region of Figure 4(a), is displayed in Figure 4(b). Note that the basic form of two broad humps occurs on both scales. The peak of the hump in (a) occurs at ~10 mT and in (b) at ~0.5 mT, indicating a scaling factor of ~20, which is in agreement with that observed experimentally [8]. Secondly, despite this apparent self-similar behavior, this example also shows periodic scarring effects similar to that observed for stadium. Figure 4(c) shows the scarred wave function $B=16$ mT, which corresponds to the resonance indicated by the arrow in Figure 4(a). Note the distorted diamond-shaped periodic orbit reflected by the wave function. Similar scarred wave functions appear at intervals of $\sim$ 8 mT. Figure 4(d) shows another diamond scar from this periodic series, this one at $B=49$ mT. As in the stadium, the excitation of this particularly periodic orbit can be attributed in large part due to the collimation effect on the input leads. In fact, similar scars have been observed in simple square quantum dots [2]. What is interesting however in the Sinai billiard is that, the

circular anti-dot in the center can in fact cause a "shadow" effect, which can make configuration appear more regular than the simple regular dot. This is most apparent when the collimated beam exiting the left contact just grazes the anti-dot and is indicated by the periodic appearance of many well defined scars. For example, in a 0.3 μm square dot, four examples of a diamond shaped orbit were observed at periodic intervals, with the periodic scarring effects becoming much weaker as the dot is made larger [2]. In contrast, with a much larger Sinai billiard dot, we have seen the diamond pattern recur as many as six times. We shall give a much fuller account of these results elsewhere.

5. SUMMARY

Despite the fact that the closed, classical analogs of the structures we have studied are know to be chaotic, we have found that regular periodic orbits strongly influence the magneto-transport of both the stadium and Sinai billiard structures. Evidence for these orbits is seen in scarred wave functions, which can recur periodically in magnetic field. In both cases, the quantization of modes in the leads is crucial in exciting the particular orbits evident in the scarred wave functions, as this quantization causes the electrons to be injected in collimated beams directed at well defined angles. For the stadium in particular, the scarring periodicities show good correspondence with power spectrum peaks, providing strong evidence that the periodicities evident in the magneto-conductance fluctuations are in fact tied to specific periodic orbits. Moreover, we obtain power spectra that agree well with the experimental observations. In the Sinai billiard, we see indications of the self-similar behavior observed experimentally coexisting with the periodic scarring behavior. The actual origin of the self-similar behavior in this structure requires further study. However, in this regard, it should be mentioned that the results of a semi-classical analysis have lead to the prediction of fractal conductance fluctuations in chaotic cavities, notably for cases in which there is a mixed (chaotic and regular) phase space [9].

Acknowledgements

The authors would like to express their appreciation to C.M. Marcus for allowing us to use his data in this paper and R. P. Taylor for useful discussions. The work was supported by ONR and DARPA.

References

[1] Jalabert, R. A., Baranger, H. U. and Stone, A. D. (1990). "Conductance Fluctuations in the Ballistic Regime: a Probe of Quantum Chaos?", *Phys. Rev. Lett.*, **65**, 2442.
[2] Akis, R., Ferry, D. K. and Bird, J. P. (1996). "Magneto-transport Fluctuations in Regular Semiconductor Ballistic Quantum Dots", *Phys. Rev.* B., **54**, 17705.
[3] Marcus, C. M., Rimberg, A. J., Westervelt, R. M., Hopkins, P. F. and Gossard, A. C. (1992). "Conductance Fluctuations and Chaotic Scattering in Ballistic Microstructures", *Phys. Rev. Lett.*, **69**, 506 .
[4] Chang, A. M., Baranger, H. U., Pfeiffer, L. N. and West, K. W. (1994). "Weak Localization in Chaotic versus Nonchaotic Cavities: a Striking Difference in Lineshape", *Phys. Rev. Lett.*, **73**, 2111.
[5] Okubo, Y., Ochiai, Y., Bird, J. P., Ferry, D. K., Ishibashi, K., Aoyagi, Y. and Sugano, T. (1997). "Magnetically Induced Supression of Phase Breaking in Ballistic Mesoscopic Billiards", *Phys. Rev. B.*, **55**, 1368.
[6] Usuki, T., Saito, M., Takatsu, M., Kiehl, R. A. and Yokoyama, N. (1995). "Numerical Analysis of Ballistic Electron Transport in Magnetic Fields by Using a Quantum Point Contact and a Quantum Wire", *Phys. Rev. B.*, **52**, 8244.
[7] Harbury, Y. and Porod, W. (1996). "Elastic Scattering Theory for Electronic Waves in Quantum Corrals", *Phys. Rev. B.*, **53**, 15455.
[8] Taylor, R. P. *et al.* (1997). "Self-Similar Magnetoresistance of a Semiconductor Sinai Billiard", *Phys. Rev. Lett.*, **78**, 1952.
[9] Ketzmerick, R. "Fractal Conductance Fluctuations in Generic Chaotic Cavities", *Phys. Rev. B.*, in press.

Authors' Biographies

Richard Akis received his bachelor's degree in physics from Lakehead University, Thunder Bay, Canada in 1986, where he won the Science Gold Medal. He received a Ph.D. in Physics in 1991 from McMaster University, Hamilton, Canada, in the area of superconductivity. For his postdoctoral

work, he switched fields, and has been studying semiconductor nanostructures, initially at Simon Fraser University, Burnaby, Canada, where he was a NSERC postdoctoral fellow and subsequently at Concordia University, Montreal, Canada, where he was a research associate. Since 1995, he has been a faculty research associate at Arizona State University.

David K. Ferry is Regents' Professor of Engineering at Arizona State University. Prior to joining ASU in 1983, he was a faculty member at Colorado State University (1977–83) and Texas Tech University (1967–73), and worked at the Office of Naval Research (1973–77). He received his bachelor's and master's degrees in electrical engineering from Texas Technological College in 1962 and 1963, respectively, and the Ph.D. in electrical engineering from the University of Texas in 1966. He was a NSF postdoctoral fellow at the University of Vienna (1966–67).

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 313–317
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Complete RF Analysis of Compound FETs Based on Transient Monte Carlo Simulation

S. BABIKER, A. ASENOV*, N. CAMERON, S. P. BEAUMONT and J. R. BARKER

Department of Electronics and Electrical Engineering, Nanoelectronics Research Centre, Glasgow University, Glasgow G12 8LT, Scotland, UK

In this paper we described a complete methodology to extract the RF performance of 'real' compound FETs from time domain Ensemble Monte-Carlo (EMC) simulations which can be used for practical device design. The methodology is based on transient finite element EMC simulation of realistic device geometry. The extraction of the terminal current is based on the Ramo-Shockley theorem. Parasitic elements like the gate and contact resistances are included in the RF analysis at the post-processing stage. Example of the RF analysis of pseudomorphic HEMTs illustrates our approach.

Keywords: Monte-Carlo, RF analysis, compound FETs, simulation

1. INTRODUCTION

The remarkable RF performance of compound FETs such as GaAs MESFETs and InGaAs HEMTs with channel lengths down to 0.1 μm is due to well pronounced velocity overshoot. The use of simulation for predictive analysis and design of such devices require in many cases the employment of full scale EMC technique [1–3]. However, most of the published EMC studies of compound FETs consider simplified device geometry and focus mainly on the transport physics and the effect of the enhanced channel velocity on the DC device characteristics. Far more important for the proper design of modern short channel compound FETs is the RF performance which is determined not only by the high field transport but also by the device geometry and the surface effects. The T- or Γ-shaped gate, the gate recess and the passivation in such devices critically affect the device parasitics and the overall RF device performance.

In this paper we describe a methodology based on the EMC simulation to investigate the RF performance of FETs. The terminal currents are estimated using the Ramo-Shockley theorem. The device parasitics are included through the proper finite-element description of the gate and recess shapes. The external parasitics are included in the post-processing stage.

* Corresponding author.

2. TRANSIENT CURRENT

The Heterojunction 2D Finite Element simulator (H2F) and its Monte-Carlo module are described in detail elsewhere [4, 5]. In the MC simulation the total transient terminal current in response to a step change in the applied voltages required for the y-parameter extraction is the sum of the particle current and the displacement current. According to the Ramo-Shockley theorem [6] the instantaneous transient current in electrode i due to N discrete moving charges within the device is given by the sum of the current $I_i'(t)$ solely contributed by the movement of the N charged particles with fixed potentials at electrodes and the current $I_i''(t)$ induced due to the time-varying potentials of the electrodes through the capacitive coupling through the electrodes. The current $I_i'(t)$ is given by

$$I_i'(t) = -\sum_{j=1}^{N} q_j v_j(t) \cdot \nabla f_i \qquad (1)$$

where q_j is the charge of the super-particle, $v_j(t)$ is the velocity of the particle and f_i is the solution of the Laplace equation $\nabla\cdot(\varepsilon\nabla f_i)=0$ with a unit voltage applied to electrode i, while all other electrodes are grounded.

The current $I_i''(t)$ associated with the time varying potential is calculated from the capacitance matrix components C_{ij} associated with the electrodes of the simulated device. The capacitance matrix components are obtained from the solution of the Laplace equation as $C_{ij} = \Delta Q_i/\Delta V_j$ where ΔQ_i is the change in the electrode charge in response to a change in the potential of a particular electrode. When a step perturbation ΔV_j of the terminal voltage is applied $I_i''(t)$ flows only during the first time step Δt and is given by:

$$I_i''(t) = C_{ij}\frac{\Delta V_j}{\Delta t} \qquad (2)$$

The displacement current during the remaining part of the transient is related to the charges induced on the electrode by the moving particles in the device associated with the redistribution of the mobile charge and is accounted for by Eq. (1).

3. RF ANALYSIS

The flow diagram of the complete time domain RF EMC analysis is given in Figure 1. The y-parameters are calculated by Fourier decomposition of the current transients obtained in response to step perturbations in the terminal voltages [7]. The cut-off frequency of the simulated device f_T^{sim} is extracted by solving $\log[G_c^{\text{sim}}(\log f)] = 0$ where $G_c = dI_d/dI_g$ is the current gain expressed as a function of y-parameters.

In order to extract the maximum frequency of oscillation of the simulated device $f_{\max}^{\text{sim}}$ the y-parameters are transformed into s-parameters S^{sim}. The maximum frequency of oscillation $f_{\max}^{\text{sim}}$

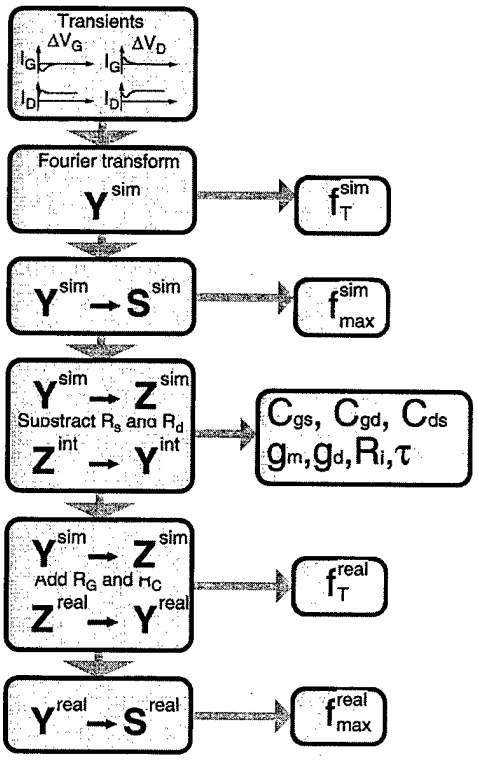


FIGURE 1 Flow diagram of the RF EMC analysis.

is then extracted by solving $\log[MAG^{sim}(\log f)]=0$ where MAG^{sim} is the maximum available gain.

For typical MESFET and HEMT simulation domains, we have adopted the equivalent circuit model presented in Figure 2. The gate resistance, the contact resistances and any external inductive components are excluded from the EMC simulation. The source R_{s1} and the drain R_{d1} resistances in Figure 2 represent the resistance of the regions between the gate and the source and drain ohmic contacts respectively. To extract accurately the small signal equivalent circuit the y-parameters of the simulated device $\mathbf{Y}^{sim}$ are transformed into z-parameters $\mathbf{Z}^{sim}$. The estimated source and drain resistances R_{s1} and R_{d1} are subtracted from $\mathbf{Z}^{sim}$ to obtain the z-parameters $\mathbf{Z}^{int}$ of the 'intrinsic' device. Finally $\mathbf{Z}^{int}$ are transformed back into y-parameters $\mathbf{Y}^{int}$ from which the components of the 'intrinsic' small signal circuit can be analytically extracted [7]:

In order to evaluate the cut-off frequency f_T^{real} and the maximum frequency of oscillations f_{max}^{real} of the 'real' device the gate resistance R_g, the contact resistances R_c and eventually the inductive components L_g, L_s and L_d first have to be incorporated in the z-parameters of the real device $\mathbf{Z}^{real}$ which are then transformed into $\mathbf{Y}^{real}$. From $\mathbf{Y}^{real}$ and S^{real} the figures of merit of the 'real' device we can estimate f_T^{real} and f_{max}^{real}. The intrinsic minimum noise figure is also evaluated from the two-port y-parameters and the current traces.

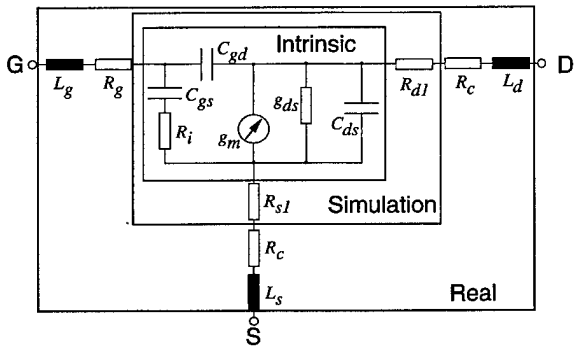


FIGURE 2 Small signal equivalent circuit of a real FET. Figure also shows the 'intrinsic' and 'simulated' device equivalent circuit.

4. RESULTS

We apply the described RF analysis in the simulation of a 120 nm T-gate InGaAs channel pHEMT with 22 nm gate to channel separation [8]. The y-parameters extracted from the Fourier decomposition of the transients are shown in Figures 3 a,b. Table I summarises the small signal

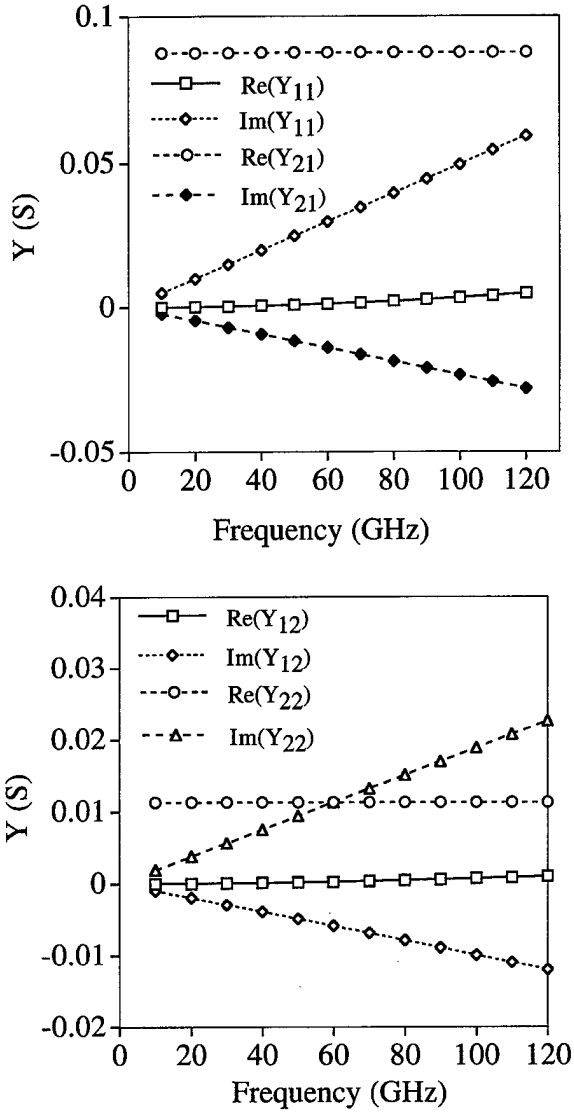


FIGURE 3 Y-parameters as a function of frequency extracted from the Fourier decomposition of the current transients at $V_g=-0.2$ and $V_d=1.5$ V.

TABLE I Small signal equivalent circuit components calculated from the EMC simulations, $V_g = -0.2$ and $V_d = 1.5$ V. Given also are the experimental values. fT and f_{max} are given at $R_g = R_c = 5\,\Omega$

	C_{gs}(fF)	C_{gd}(fF)	C_{ds}(fF)	g_{mo}(ms)	g_{ds}(fF)	$R_i(\Omega)$	τ(pS)	fT(GHz)	f_{max}(GHz)
MC	75.5	14.2	15.1	87.6	11.2	2.3	0.17	142	164
Exp.	77.8	8.73	14	67	8.65	3.75	0.32	114	150

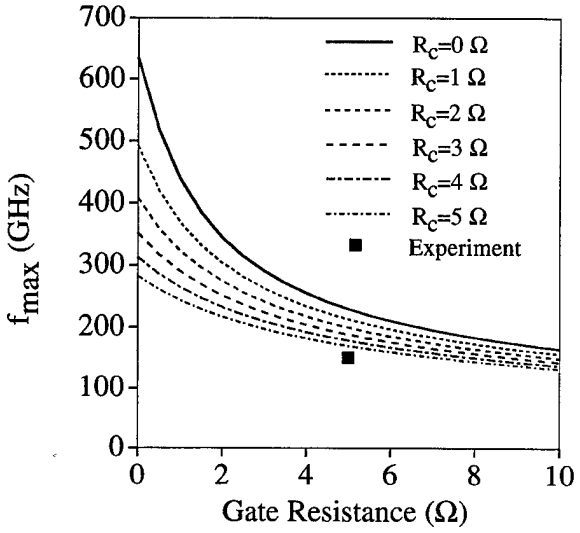


FIGURE 4 Effect of gate and contact resistances on f_{max}.

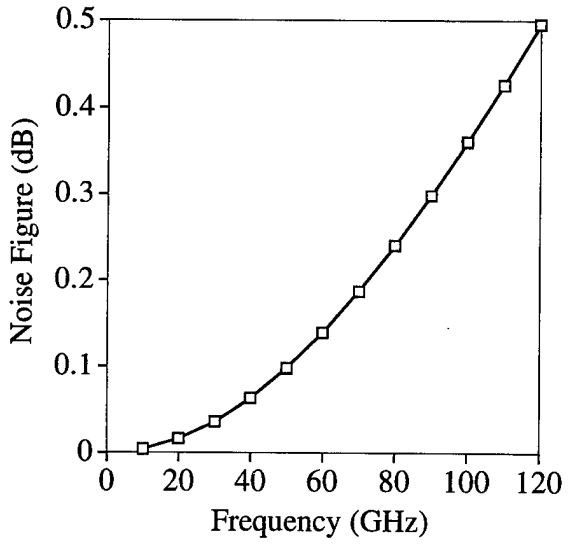


FIGURE 5 Intrinsic noise figure for the 120 nm pHEMT extracted from the current traces at $V_g = -0.2$ and $V_d = 1.5$ V.

equivalent circuit parameters extracted from the EMC simulations, $V_g = -0.2$ and $V_d = 1.5$ V, together with the experimental values. The higher simulation g_{mo} and f_T values is due to the fact that the MC simulations overestimate the velocity overshoot in the channel. The significant effect the gate and contact resistances have on the maximum frequency of oscillation is depicted in Figure 4. The intrinsic noise figure extracted from the transients is shown in Figure 5.

5. CONCLUSION

In this paper we have presented a comprehensive methodology for the RF analysis of FETs. The methodology is based on the transient EMC simulations of the intrinsic device followed by a post-processing stage during which the parasitic elements are included. It allows realistic estimation of the RF performance from MC simulations. The capabilities of the scheme are illustrated in example simulations for the 120 nm gate length pHEMT.

References

[1] Park, D. H. and Brennan, K. F. (1990). "Monte Carlo simulation of 0.35-μm gate-length GaAs and InGaAs HEMT's", *IEEE Trans. Electron Devices*, **37**, 618–628, .

[2] Kizilyally, I. C., Artaki, M., Shah, N. J. and Chandra, A. (1991). "Scaling properties and short-channel effects in submicrometer AlGaAs/GaAs MODFET's: A Monte Carlo study", *IEEE Trans. Electron Devices*, **40**, 234–249.

[3] Kim, K. W., Tian, H. and Litteljohn, M. A. (1991). "Analysis of delta doped and uniformly doped AlGaAs/GaAs HEMT's by ensemble Monte Carlo simulation", *IEEE Trans. Electron Devices*, **38**, 1737–1742.

[4] Babiker, S., Asenov, A., Barker, J. R. and Beaumont, S. P. (1996). "Finite element Monte Carlo simulation of recess gate compound FETs", *Solid State Electronics*, **39**, 629–635.

[5] Asenov, A., Reid, D., Barker, J. R., Cameron, N. and Beaumont, S. P. (1993). "Finite element simulation of

recess gate MESFETs and HEMTs. The Simulator H2F", in *Simulation of Semiconductor devices and processes*, S. Selberherr, H. Stippel, E. Strasser, eds., Wien: Springer Verlag, **5**, 265–268.
[6] Kim, H., Min, H., Tang, T. and Park, Y. (1992). "An Extended Proof of the Ramo-Shockley Theorem", *Solid State Electronics*, **34**(11), 1251–1253.
[7] González, T. and Pardo, D. (1990). "Monte Carlo determination of the intrinsic small-signal equivalent circuit of MESFET's," *IEEE Trans. Electron Devices*, **42**, 605–611.
[8] Cameron, N., Taylor, M. R. S., Mclelland, H., Holland, M., Thayne, I., Elgaid, K. and Beaumont, S. P. (1995). "A high performance, high yield, dry etched pseudomorphic HEMT for W-band use", *IEEE Trans. Microwave Theory and Techniques Symposium Digest*, Orlando, FL, 435–438.

Authors' Biographies

Sharief Babiker has studied Electrical Engineering in Khartoum University Sudan (1979–1984). During the period 1985–90 he worked at the Electrical Engineering Department of Khartoum University. In 1994 he was awarded the Ph.D. degree for work on the theory and modelling of single-electronic devices and systems. His current research interests concentrate on the simulation and modelling of FETs with emphasis on sub-micron gate length InGaAs channel pHEMTs for RF applications.

Asen Asenov had 10 years industrial experience as a head of the Process and Device Modelling Group in IME–Sofia, developing one of the first integrated process and device CMOS simulators IMPEDANCE. He was visiting professor at the Physics Department of TU Munich, and is currently a Reader in the Department of Electronics and Electrical Engineering, Glasgow University. As a leader of the Device Modelling Group he has contributed to the development of 2D and 3D device simulators and their application in the design of FETs, SiGe MOSFETs and IGBTSs.

Nigel Cameron graduated from the University of Bath in 1984 and joined British Telecom's research laboratories to work on plasma process research and development for silicon fabrication. In March 1988 he moved to the Nanoelectronics Research Centre at the University of Glasgow where his responsibilities include the development of ultra-fast transistors in compound semiconductors and their integration into manufacturable processes for millimetre-wave integrated circuits (MMICs). His interests extend across a broad range of semiconductor technologies encompassing materials and device physics; circuit design and test; manufacturing; and dc and high frequency characterisation. Nigel Cameron is the author or co-author of more than 35 technical papers.

Steven Beaumont was educated at the University of Cambridge and has been with the Department of Electronics and Electrical Engineering at the University of Glasgow since 1978. He became Head of Department in 1995 and he convenes the Nanoelectronics Research Centre's management committee. His research interests lie in the field of nanometre-scale fabrication and its application to electronic and optoelectronic devices. He has over 100 publications on electron beam nanolithography, dry etching, short-gate III–V based transistors, quantum transport devices, the optical properties of quantum dots, and single electron devices. Latterly he has become involved with the issue of manufacturability of mm-wave circuits and the use of nanometre-scale fabrication techniques coupled with technology-based device simulations to forecast performance and yield with the minimum of process iterations.

John Barker is Professor of Electronics in the Department of Electronics and Electrical Engineering. He has a long standing interest in computational methods, device modelling and transport theory. From 1970–85 he was a member of the Theory group in the Dept. of Physics, University of Warwick, aside from 1978–79 when he worked at IBM T. J. Watson Laboratory, North Texas State University and Colorado State University. From 1987–89 he was academic director of the IBM UK/Glasgow University Kelvin Project on Numerically Intensive Parallel Computing. He is academic director of the Parallel Processing Centre at the University of Glasgow.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 319–323
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Monte Carlo Calibrated Drift-Diffusion Simulation of Short Channel HFETs

A. ASENOV*, S. BABIKER, S. P. BEAUMONT and J. R. BARKER

Nanoelectronics Research Centre, Department of Electronics and Electrical Engineering, Glasgow University, Glasgow G12 8LT, Scotland, UK

In this paper we present a methodology to use drift diffusion (DD) simulations in the design of short channel heterojunction FETs (HFETs) with well pronounced velocity overshoot. In the DD simulations the velocity overshoot in the channel is emulated by forcing the saturation velocity in the field dependent mobility model to values corresponding to the average velocity in the channel obtained from Monte Carlo (MC) simulation. To illustrate our approach we compare enhanced DD and MC simulation results for a pseudomorphic HEMTs with 0.12 μm channel length, which are in good agreement. The usefulness of the described methodology is illustrated in a simulation example of self aligned gamma gate pseudomorphic HEMTs. The effect of the gamma gate shape and the self aligned contacts on the overall device performance has been investigated.

Keywords: Numerical simulation, overshot, heterojunction FETs, drift-diffusion approach

1. INTRODUCTION

Commercial device simulators like MEDICI [1] and BLAZE [2] are flexible, fast and work in a user friendly environment. Employing finite element approach they can describe accurately the complex geometry of modern short recess gate heterojunction FETs (HFETs). Interface charge, surface states, and deep levels can be eventually included in the simulations and the self heating can be treated through coupling to the heat flow equation. The transient algorithms are stable and allow for large time steps. Frequency domain analysis is readily available. External circuit elements like contact and gate resistances, pad inductances and capacitances can be included in the simulations and in the *rf* analyses. Unfortunately the drift-diffusion (DD) approach, which is at the heart of such simulators, cannot predict the velocity overshoot responsible for the high performance of many short channel HEFTs [3]. The hydrodynamic (HD) options offered as extensions to the DD engines of the above simulators deal with the overshoot but slow down the simulations

* Corresponding author. Tel: +44 141 330 5233, Fax: +44 141 330 4907, E-mail: A. Asenov@elec.gla.ac.uk.

and often have convergence and parameter indentification problems. The ensemble Monte Carlo (MC) method, usually implemented in in-house software [4], is still computationally expensive and cannot match all the features of the DD commercial simulators. particularly their speed.

In this paper we describe a methodology for using MC calibrated DD simulation with enhanced channel velocity in the design and optimisation of short channel HFETs including HEMTs and strained Si channel SiGe MODFETs. The DD simulations are calibrated with respect to our finite element Monte Carlo simulator H2F [5]. The usefulness of this approach is illustrated in simulation examples of self aligned gamma gate pseudomorphic HEMTs.

2. EMULATION OF THE OVERSHOOT

The enhanced DD approach is based on the observation from MC simulations that in many short channel (0.1–0.2 μm gate) HFETs the velocity overshoot extends along the whole high field channel region. The average velocity profile in the channel obtained from MC simulation can be emulated by increasing the saturation velocity v_s in a simple, three parameter, silicon-type field dependent mobility model [6] with velocity v is given by

$$v = \frac{\mu_0 E}{\left[1 + \left(\frac{\mu_0 E}{v_s}\right)^b\right]^{\frac{1}{b}}} \qquad (1)$$

where E is the electric field, μ_0 is the low field mobility and b is a model parameter. The calibration is done by adjusting the three parametes μ_0, v_s and b in order to match the MC velocity profile in the source and in the channel region for a bias point in the middle of the uselful part of the device characteristics. In most of the cases this is enough to achieve satisfactory agreement between the MC and the DD simulation results over the whole range of applied voltages of interest.

Figure 1 illustrates the velocity along the channel of a 0.12 μ gate length pseudomorphic HEMT with 22 nm gate-to-channel separation, 50 nm recess offset and delta doping in the supply layer. The device is described in more details elsewhere [7]. The velocity profile obtained from DD MEDICI simulation with v_s forced to 2.8×10^7cm/s, $\mu_0 = 5000$ $cm^2v^{-1}s^{-1}$and $b=2$ matches well with the MC velocity profile in the source region and in the channel. In the drain region, where most of the particle in the MC simulation are in the L-valley, the DD simulation overestimates the velocity and hence underestimates the drain resistance. This however does not affect seriously the simulated *rf* performance since the drain resistance has usually a weak influence on the measured and extracted *s*-parameters.

3. COMPARISON OF DD AND MC DC AND RF RESULTS

In order to justify the use of the enhanced DD approach in practical short channel HFET device simulation and design we compere the *dc* and the *rf* results obtained from both DD and MC simulations. The comparison is based on the same

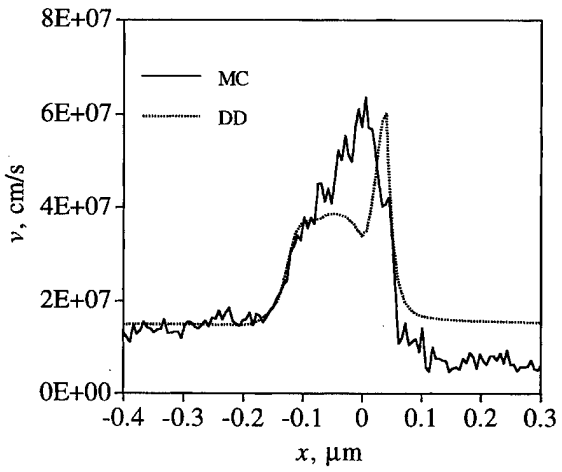


FIGURE 1 Average velocity in the channel of a 0.12 μm pHEMT obtained from DD and MC simulation at $V_G = 0$ V and $V_D = 1.5$ V.

pseudomorphic HEMTs described in Section 2. Exactly the same simulation geometry is used in both the DD MEDICI's and the MC H2F's simulations. The source and drain contact resistances which cannot be directly included in the MC simulation are also excluded from the DD simulations. The velocity profile is calibrated at one *dc* bias point corresponding to $V_G = 0\,\text{V}$ and $V_D = 1.5\,\text{V}$. The *dc* ouput characteristics are compared in Figure 2 and are in remarkably good agreement in both the low and in the high drain volatge regions. This is partially due to the fact that the deconfinement which cannot be properly treated in the DD simulation does not affect seriously the operation of the simulated pseudomorphic HEMTs with relatively deep ($In_{0.3}Ga_{0.7}As$) channels.

The comparison of the *rf* results obtained form DD and MC simulation is based on time domain transient technique which is the only available choice in the MC case [8]. Complex two port *y*-parameters are extracted by Fourier transforming the gate and drain current transients in response to small changes in the gate and drain voltages. The frequency dependence of the real and the imaginary parts of y_{11} and y_{12} is plotted in Figure 3. The *y*-parameters extracted from the DD and MC transient simulations are in good agreement. This

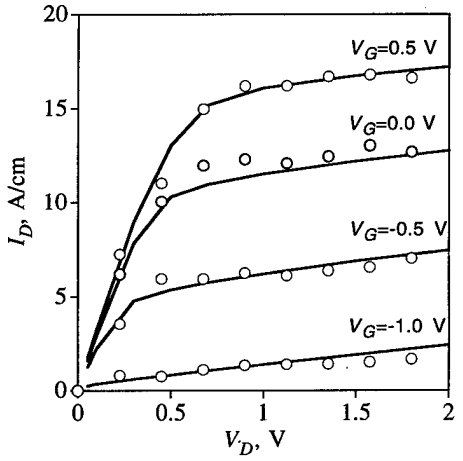


FIGURE 2 Comparison between DD and MC simulated output characteristics of a 0.12 μm pHEMT.

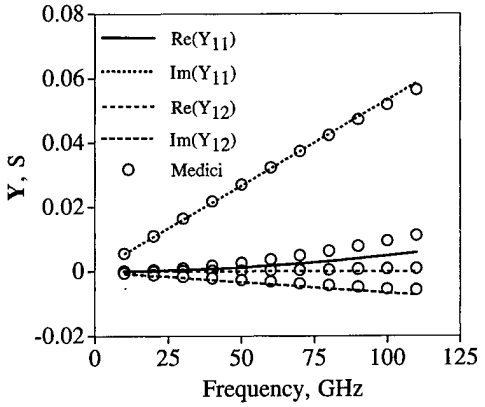


FIGURE 3 Comparison between *Y* parameters extracted from DD and MC simulation of a 0.12 μm pHEMT at $V_G = -0.2$ V and $V_D = 1.5$ V. The channel width is 100 μm.

agreement holds also for the small signal circuit elements and the cut-off frequency f_T extracted from the *y*-parameters and for the maximum frequency of oscillations $f_{\max}$ extracted after transforming the *y*-parameters into *s*-parameters.

4. SELF ALIGNED GAMMA GATE EXAMPLE

Calibrated DD simulations can be used confidently to investigate device geometry effects which do not affect strongly the lateral field profile in the channel but are important for the *rf* device performance including the gate and contact shapes, self-alignment, cap layer modification, recess design etc. When the lateral field profile is altered, for example as a result of channel length scaling, the DD simulations have to be recalibrated with respect to a new MC calculation.

We illustrate the described DD approach in the simulation of self aligned gamma gate pHEMTs. The vertical layer structure of these devices is the same as the structure described in Section 2. The profile of the gamma gate and the corresponding solution domain are outlined in Figure 4. In the DD simulation we investigate the effect of the gate overlap d_{c-r} on the drain side of the channel on f_T and $f_{\max}$. The results are presented in Figure 5.

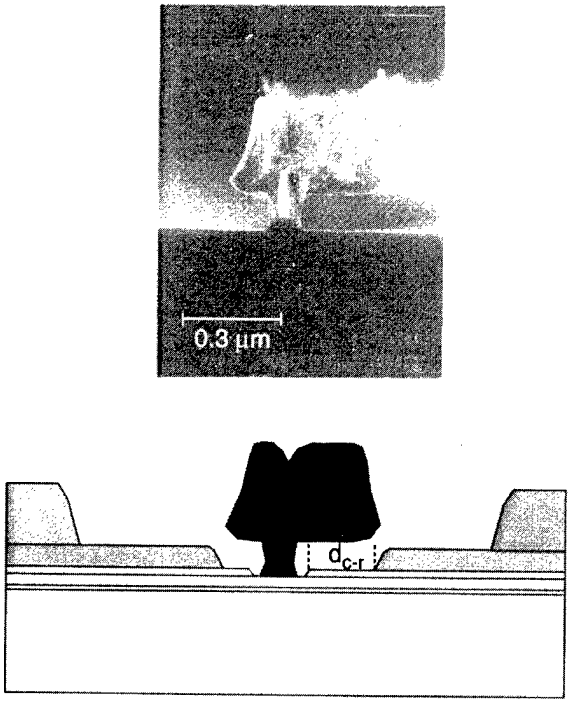


FIGURE 4 Self aligned gamma gate pHEMT.

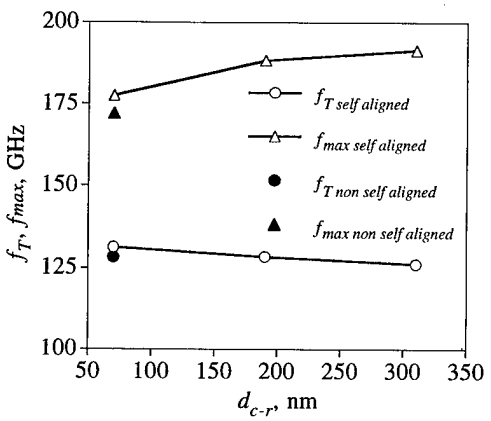


FIGURE 5 Dependence of f_T and f_{max} on the gate overlap d_{c-r} for the self aligned gamma gate pHEMT illustrated in Figure 4 with channel width is 100 μm, contact resistance 1 Ω and gate resistance 5 Ω. DD simulations.

In the self aligned technology a trade-off between the reduced series resistances and the increased gate-to-contacts parasitic capacitances determines the over all device performance. The gamma gate design itself is a trade off between the reduced gate resistance and the increased gate-to-drain parasitic capacitance.

5. CONCLUSIONS

The velocity overshoot in the channel of modern short gate length HFETs can be satisfactory emulated in DD simulations by increasing the saturation velocity in a simple silicon type field dependent mobility model. Usuallly calibration of the DD simulations to one properly selected bias point of the corresponding MC simulations leads to good agreement in the DD and MC *dc* and *rf* characteristics over a wide range of applied voltages. Calibrated DD simulations can be used to speed up the design of modern short channel HFETs where geometry related parasitic effects complete with the enhanced channel transport in determining the overall device performance.

Acknowledgements

The work is supported by EPSRC through the Nanolelectronics Rolling Grant RG/6601BS.

References

[1] TMA MEDICI,. Version 2.3, Technology Modelling Associates, Inc., 1996.
[2] Silvaco, BLASE. Version 4.0.0.R, Silvaco International, 1996.
[3] Babiker, S., Cameron, N., Asenov, A. and Beaumont, S. P. (1996). 'New evidence for velocity overshoot in a 200 nm Pseudomorphic HEMT', *Microelectronics Journal*, **27**(8), 785–793.
[4] Park, D. H. and Brennan, K. F. (1990). 'Monte Carlo simulation of 0.35 μm gate-length GaAs and InGaAs HEMT', *IEEE Trans. Electron Devices*, **37**(3), 618–628.
[5] Babiker, S., Asenov, A., Barker, J. R. and Beaumont, S. P. (1996). 'Finite element Monte Carlo simulation of recess gate compound FETs', *Solid State Electronics*, **39**(5), 629–635.
[6] Caughey, D. M. and Thomas, R. E. (1967). 'Carrier mobilities in silicon empirically related to doping and field', *IEEE Trans. Electron Devices*, **55**, 2192–2193.
[7] Cameron, N. I., Murad, S., McLelland, H., Asenov, A., Taylor, M. R. S., Holland, M. C. and Beaumont, S. P. (1996). 'Gate Recess Engineering of Pseudomorphic

$In_{0.30}GaAs/GaAs$ HEMTs'. *Electronics Letters*, **32**(8), 770–772.

[8] González, T. and Pardo, D. (1995). 'Monte Carlo determination of the intrinisic small-signal equivalent circuit of MESFET's', *IEEE Trans. Electron Devices*, **42**(4), 605–611.

Authors' Biographies

Asen Asenov had 10 years industrial experience as a head of the Process and Device Modelling Group in IME-Sofia, developing one of the first integrated process and device CMOS simulators IMPEDANCE. He is currently a Reader in the Department of Electronics and Electrical Engineering, Glasgow University. As a leader of the Device Modelling Group he is involved in the development of 2D and 3D device simulators and their application in the design of compound FETs, SiGe MOSFETs and IGBTs. Other interest include design of parallel algorithms.

Sharief Babiker studied Electrical Engineering in Khartoum University Sudan (1979–1984). During the period 1985–90 he workded at the Electrical Engineering Department of Khartoum University. In 1994 he was awarded a Ph.D. degree for work on the theory and modelling of single-electronic devices and systems. His current research interests concentrate on the simulation and modeling of FETs with emphasis on submicron gate length InGaAS channel pHEMTs for RF applications.

John Barker is Professor of Electronics in the Department of Electronics and Electrical Engineering. He has a long standing interest in computational methods, device modelling and transport theory. From 1970–85 he was a member of the Theory group in the Dept. of Physics, University of Warwick, aside form 1978–79 when he worked at IBM T. J. Watson Laboratory, North Texas State University and Colorado State University. From 1987–89 he was academic director of the IBM UK/Glasgow University Kelvin Project on Numerically Intensive Parallel Computing. He is academic director of the Parallel Processing Centre at the University of Glasgow.

Steve Beaumont was educated at the University of Cambridge and has been with the Department of Electronics and Electrical Engineering at the University of Glasgow since 1978. He became Head of Department in 1995 and he convenes the Nanoelectronics Research Centre's management committee. His research interests lie in the field of nanmetre-scale fabrication and its application to electronic and optoelectronic devices. He has over 100 publications on electron beam nanolithography, dry etching, short-gate III–V based transistors, quantum transport devices, the optical properties of quantum dots, and single electron devices. Latterly he has become involved with the issue of manufacturability of mm-wave circuits and the use of nanometre-scale fabrication techniques coupled with technology-based device simulations to forecast performance and yield with the minimum of process iterations.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 325–330
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

RF Performance of Si/SiGe MODFETs: A Simulation Study

S. ROY, A. ASENOV*, S. BABIKER , J. R. BARKER and S. P. BEAUMONT

Nanoelectronics Research Centre, Department of Electronics and Electrical Engineering, Glasgow University, Glasgow G12 8LT, Scotland, UK

The microwave performance potential of Si/SiGe pseudomorphic MODFETs are studied, in comparison to state of the art InGaAs pseudomorphic HEMTs. Both devices have equivalent structures corresponding to a physical HEMT used for calibration. We use an RF analysis technique based on transient Monte Carlo simulations to estimate the intrinsic noise figures, the RF figures of merit f_T and f_{max}, and the effect of contact and gate resistances. Both devices exhibit velocity overshoot below the gate region. It is shown that the difference in noise figures and f_T values can be mainly attributed to differences in device channel velocity. f_{max} exhibits a strong dependence on device contact resistance, eroding some of the performance advantage of the pseudomorphic HEMT.

Keywords: Si/SiGe, heterostructures, Monte Carlo, microwave performance, RF analysis

INTRODUCTION

Recent theoretical and experimental studies show that low field mobility and velocity overshoot are enhanced in Si layers grown pseudomorphically on relaxed SiGe substrates. Induced strain breaks the six-fold degeneracy of the Si conduction band, resulting in an improved band offset for the two conduction valleys whose transverse effective mass is in the plane of the heterojunction. This increases the in plane effective mobility and reduces inter-valley scattering in the Si layer [1]. Modulation doped field effect transistors (MODFETs) based on this material system have been demonstrated, and show significant potential for RF applications [2, 3]. Although measured mobilities are lower than those reported in optimal III–V based devices, the compatibility between SiGe and conventional Si processing technology makes such MODFETs attractive for Si MMICs design and microwave signal processing applications integrated on conventional Si chips. Therefore, comparison of strained Si channel devices with well established members of the III–V family will

* Corresponding author. Tel.: ++44 141 330 5233, Fax: ++44 141 330 4907, e-mail: A. Asenov@elec.gla.ac.uk.

provide greater insight into their potential performance and limitations.

As part of this goal, we here study the microwave performance potential of a Si/SiGe pseudomorphic MODFET in comparison with a state of the art InGaAs channel pseudomorphic HEMT, using an RF analysis technique based on transient Monte Carlo simulations.

DEVICE STRUCTURES AND PERFORMANCE ANALYSIS

To allow a fair comparison similar device layer structures are considered (Figs. 1a,b). Both devices have δ-doping separated by a 2.5 nm spacer from the channel, a T shape recess gate with 50 nm recess offset and heavily doped (n-type $10^{18}\,cm^{-3}$) cap layers. An effective δ-doping of $5\times10^{12}\,cm^{-2}$ is considered in the both cases. Total gate-to-channel separation is 22 nm and the gate length is 0.12 μm. p-type background substrate doping of $10^{14}\,cm^{-3}$ is considered. This corresponds to the dimensions of real pseudomorphic HEMT fabricated at the Glasgow Nanoelectronics Research Centre and used for validation and calibration of the RF Monte Carlo analysis technique (described in an accompanying paper [4]). We realise that for the Si/SiGe MODFET this may pose unresolved growth problems. Specifically we conjecture improvements in growth technology and low temperature processing to allow formation of a well defined As δ-doping supply layer. For simplicity of Monte Carlo modelling a uniform SiGe substrate is assumed, instead of a 'virtual substrate' with graded Ge concentration. The 'etch stop' region of the Si/SiGe device is also considered as (undoped) SiGe. Simple 1-D Poisson calculations indicate that if a strained Si 'etch stop' region is included, it will exhibit negligible parallel conductance at a d.c. bias of $V_G = -0.25$ V.

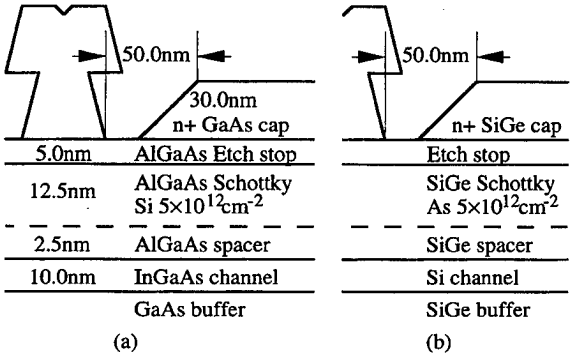


FIGURE 1 Vertical layer structure of (a) a 0.12 μm pseudomorphic HEMT and (b) a 0.12 μm pseudomorphic MODFET. Both devices have the same vertical dimensions.

RF analysis of HEMT and MODFET performance is based on Monte Carlo simulation of device transient response. We follow Yamada's [1] treatment of the effect of the strain on the Si channel band structure, with conduction band splitting of $\Delta E_c = 0.67x$ (eV) and band gap $E_g = 1.11 - 0.74x$ (eV). However the six phonon scattering model of Jacoboni [5] is implemented, instead of the four phonon model in [1]. Acoustic and ionised impurity scattering modes are also included, with ionised impurity scattering calculated from the Brooks-Herring model [6]. Bulk velocity-field characteristics obtained from the model are in agreement with experimental unstrained Si data at 77 K and 300 K and previous results for strained Si [1, 7]. The transient Monte Carlo simulations begin by following 5×10^4 superparticles for 2 ps settling time at d.c. bias, then a further 2.0→2.6 ps during which device statistics are recorded at 1 fs intervals. A step change $\Delta V_G = 0.2$ V, or $\Delta V_D = 0.3$ V is then consecutively applied, and the transient response measured for a further 2 ps. It is found that structure and doping dependant THz oscillations in device drain current (possibly due to plasma oscillations in the channel, or in the heavily doped cap layer) may mask the detailed form of the transients, and so a number of traces are averaged to define the response. Complex y-parameters are derived by Fourier transforming these terminal current transients, and used to extract the small signal equivalent circuit, intrinsic noise figures, and estimate the RF performance figures of merit f_T (cutoff frequency) and f_{max} (maximum frequency of oscillation). Finally, the small signal equivalent

circuit is augmented by the addition of external impedance, and thus the effect of contact and gate resistance on the 'real' device operation estimated. A detailed description of the analysis process is given in [4].

RESULTS AND DISCUSSION

The average velocity in the channel of both devices is compared in Figure 2 for a d.c bias of $V_D = 1.5$ V and $V_G = -0.25$ V. In the pseudomorphic HEMT this corresponds to the region of maximum transconductance, while in the pseudomorphic MODFET it is the region of maximum transconductance achieved while constraining parallel conduction in the δ-doped region to less than 10% of total conduction. The gate extends from $x = -0.12 \rightarrow 0.0\,\mu$m. Both devices show distinct overshoot below the gate region. Peak velocity in the pseudomorphic MODFET, at the drain end of the strained Si channel, approaches twice the saturation velocity of bulk unstrained Si. The ratio between peak velocity in the MODFET compared to that of the pseudomorphic HEMT InGaAs channel is 2.8.

Figures 3–5 and Table I characterise the RF performance of the two devices. The intrinsic noise figures NF (in decibels) are shown in Figure 3. From the definition of NF we postulate that the significant difference between the devices will be in their transconductance $g_{mo} = Y_{21}$. All other terms - y-parameters and current fluctuations $<\delta I^2>$ – we therefore represent by an approximate constant C.

$$NF = 1 + \left[\frac{Y_{11}}{Y_{21}}\right]^2 \frac{<\delta I_d^2>}{<\delta I_g^2>} \approx 1 + \left(\frac{C}{g_{mo}}\right)^2 \quad (1)$$

From Figure 3 and equation (1) the ratio of transconductances g_{HEMT}/g_{MODFET} is indeed found to be approximately constant below 60 GHz, with value 2.6. The transconductance values directly obtained from the Monte Carlo simulations are $g_{HEMT} = 87$ and $g_{MODFET} = 31$, with ratio 2.8. It can be concluded that the bulk of the

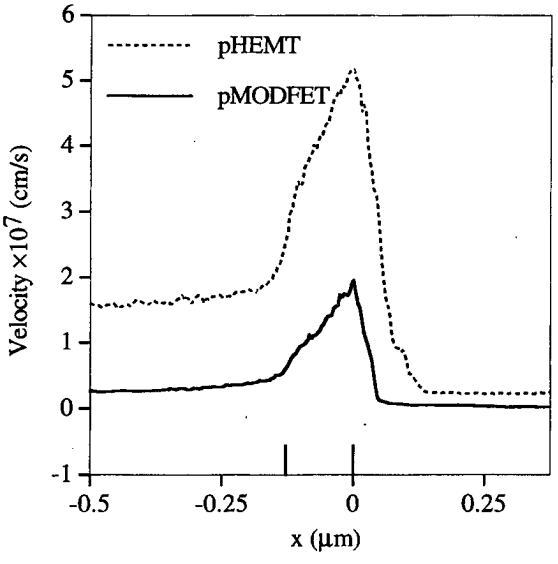


FIGURE 2 Comparison between average velocity in the channel of 0.12 μm pseudomorphic HEMT and MODFET. $V_D = 1.5$ V, with $V_G = -0.25$ V corresponding to maximum transconductance.

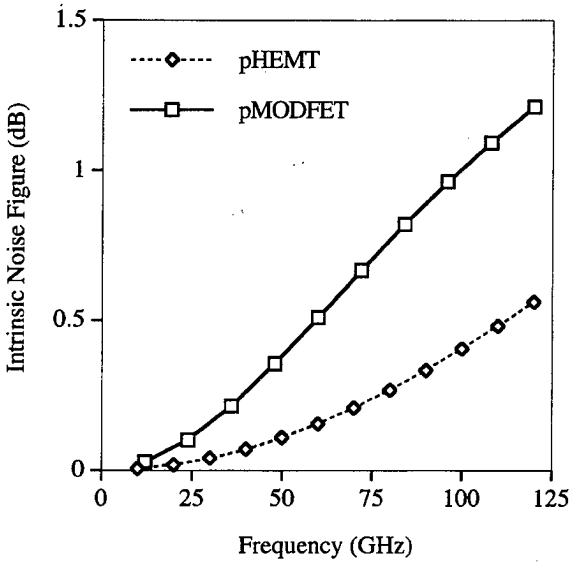


FIGURE 3 Comparison between minimum intrinsic noise figures for pseudomorphic HEMT and MODFET as a function of frequency. $V_D = 1.5$ V, with $V_G = -0.25$ V.

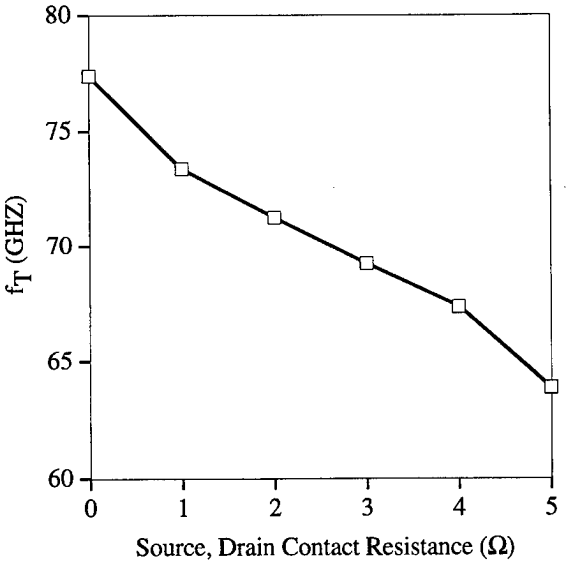


FIGURE 4 Variation of f_T in a pseudomorphic MODFET as a function of source and drain contact resistance (f_T invariant to changes in gate resistance).

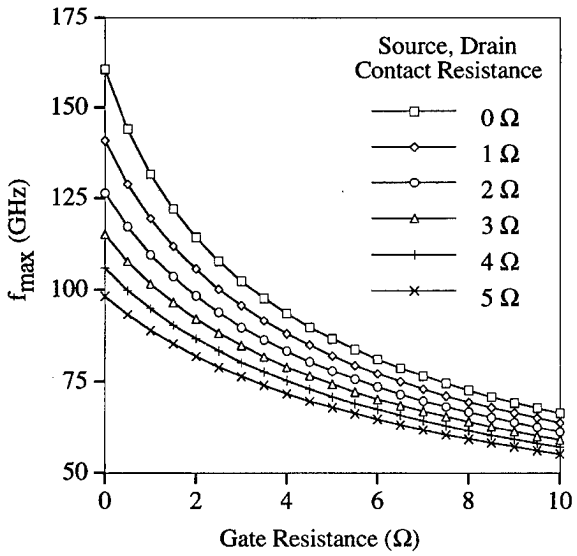


FIGURE 5 Variation of f_{max} in a pseudomorphic MODFET as a function of gate and contact resistances.

difference in noise figures between the two devices can be attributed to the differences in transconductance – and thus to differences in channel velocity.

TABLE I Calculated RF figures of merit for pseudomorphic HEMT and MODFET at bias $V_D = 1.5\,V$, $V_G = -0.25\,V$. Device width of 100 μm is assumed. f_T, f_{max} calculated for both negligible external gate and drain contact resistance, and for more realistic contact resistances of 5Ω

RF figure of merit (GHz)	Pseudomorphic HEMT	Pseudomorphic MODFET
with negligible contact resistance		
f_T	192	77.4
f_{max}	634	161
gate, source & drain resistances all of 5 Ω		
f_T	156	63.9
f_{max}	169	68.0

Table I lists the RF figures of merit extracted from the transient response of each device. The ratio of cut-off frequencies f_T for the pseudomorphic HEMT and MODFET is approximately 2.5 and independent of external resistance. This implies that the value of f_T is primarily governed by the channel velocity of the respective device. However, the ratio of f_{max} for the pseudomorphic HEMT and MODFET is 3.9 for intrinsic devices, and drops to 2.5 when external contact resistances of 5Ω are applied. A more realistic inclusion of device parasitics in this case reduces the advantages of the pseudomorphic HEMT. Finally, Figures 4 and 5 detail the effect of varying contact resistance on both figures of merit for the pseudomorphic MODFET, and show clearly the strong dependence of f_{max} on device parasitics.

CONCLUSIONS

The microwave performance potential of a Si/SiGe pseudomorphic MODFET was studied, in comparison to a state of the art InGaAs channel pseudomorphic HEMT. We used an RF analysis technique based on transient Monte Carlo simulations, calibrated to a physical HEMT. The difference in device noise figures was attributed mainly to differences in device transconductance, and thus to differences in channel velocity. The cut-off frequency f_T was also shown to be governed by channel velocity. However the max-

imum frequency of oscillation f_{max} was seen to exhibit a strong dependence on device contact resistance, eroding some of the advantage of the pseudomorphic HEMT over the Si/SiGe MODFET.

References

[1] Yamada, T., Zhou, J., Miyata, H. and Ferry, D. K. (1994). "In-Plane Transport Properties of $Si/Si_{1-x}Ge_x$ Structure and its FET Performance by Computer Simulation", *IEEE Trans. Elec. Devices*, **41**(9), 1513.

[2] Ismail, K., Rishton, S., Chu, J., Chan, K. and Meyerson, B. (1993). "High-performance Si/SiGe *n*-type modulation-doped transistors", *IEEE Electron Device Letters*, **14**(7), 348–350.

[3] Welser, J., Hoyt, J. L. and Gibbons, J. F. (1994). "Electron Mobility Enhancement in Strained-Si N-Type Metal-Oxide-Semiconductor Field-Effect Transistors", *IEEE Electron Device Letters*, **15**(3), 100.

[4] Babiker, S., Asenov, A., Cameron, N., Beaumont, S. P. and Barker, J. R. "Complete RF analysis of compound FETs based on transient Monte Carlo simulation", *this issue*.

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

[6] Brooks, H. (1951). "Scattering by Ionized Impurities in Semiconductors", *Phys. Rev.*, **83**, 879.

[7] Roldén, J. B., Gámiz, F., López-Villanueva, J. A. and Carceller, J. E. (1996). "A Monte Carlo study on the electron-transport properties of high-performance strained-Si on relaxed SiGe channel MOSFETs", *J. Appl. Phys.*, **80**(9), 5121.

Authors' Biographies

Scott Roy is a researcher in the Department of Electronics and Electrical Engineering at the University of Glasgow, from where he received a Ph.D. in 1994 for investigations into the engineering and architectural aspects of extended single electronic systems. His interests include high performance computing (he represented Glasgow University in ESPRIT Project ZEUS; Zentren Europáischen Supercomputings) and its application to device modelling. He is presently developing codes to simulate and optimise *n*-channel SiGe FETs for VLSI and RF applications – and designing, in collaboration with Motorola UK, multiprocessor systems on which they will run.

Sherief Babiker studied Electrical Engineering in Khartoum University Sudan (1979–1984). During the period 1985–90 he worked at the Electrical Engineering Department of Khartoum University. In 1994 he was awarded a Ph.D. degree for work on the theory and modelling of single-electronic devices and systems. His current research interests concentrate on the simulation and modelling of FETs with emphasis on submicron gate length InGaAs channel pHEMTs for RF applications.

Asen Asenov had 10 years industrial experience as a head of the Process and Device Modelling Group in IME – Sofia, developing one of the first integrated process and device CMOS simulators IMPEDANCE. He was visiting professor at the Physics Department of TU Munich, and is currently a Reader in the Department of Electronics and Electrical Engineering, Glasgow University. As a leader of the Device Modelling Group he has contributed to the development of 2D and 3D device simulators and their application in the design of FETs, SiGe MOSFETs and IGBTs. He also investigates the design of parallel algorithms.

John Barker is Professor of Electronics in the Department of Electronics and Electrical Engineering. He has a long standing interest in computational methods, device modelling and transport theory. From 1970–85 he was a member of the Theory group in the Dept. of Physics, University of Warwick, aside from 1978–79 when he worked at IBM T.J. Watson Laboratory, North Texas State University and Colorado State University. From 1987–89 he was academic director of the IBM UK/Glasgow University Kelvin Project on Numerically Intensive Parallel Computing. He is academic director of the Parallel Processing Centre at the University of Glasgow.

Steve Beaumont was educated at the University of Cambridge and has been with the Department of Electronics and Electrical Engineering at the University of Glasgow since 1978. He became Head of Department in 1995 and he convenes the Nanoelectronics Research Centre's management committee. His research interests lie in the field of

nanometre-scale fabrication and its application to electronic and optoelectronic devices. He has over 100 publications on electron beam nanolithography, dry etching, short-gate III – V based transistors, quantum transport devices, the optical properties of quantum dots, and single electron devices. Latterly he has become involved with the issue of manufacturability of mm-wave circuits and the use of nanometre-scale fabrication techniques coupled with technology-based device simulations to forecast performance and yield with the minimum of process iterations.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 331–335
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Ab-initio Coulomb Scattering in Atomistic Device Simulation

C. R. AROKIANATHAN, J. H. DAVIES and A. ASENOV*

Device Modelling Group, Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow G12 8QQ, Scotland, UK

As devices shrink to dimensions below 0.1 μm, it becomes essential to treat impurities and carriers as individual charges. We describe some approaches to ionised impurity scattering where the potential of the impurities is included directly in the dynamics of a Monto Carlo type simulation rather than as a scattering rate. The divergence in the Coulomb potential creates difficulties for mesh-based solutions of Poisson's equation, which we have compared with more accurate Ewald summation. However, we find that the mesh does not introduce significant errors, and reproduces well the expected mobility as a function of doping. Highly accurate integration of the equation of motion is needed for free carriers, and the initial distribution is problematic. In contrast, a simple treatment of phonon scattering by Brownian dynamics is more tolerant of errors because it tends to restore the system to equilibrium.

Keywords: Semiconductors, discrete charges, numerical simulation, Monte Carlo, mobility

1. INTRODUCTION

As semiconductor devices shrink to dimensions below 0.1 μm, their properties begin to differ from larger structures as the atomistic nature of the charges begins to exert its influence [1]. The simulation of such devices requires a full-scale 3D treatment, including both impurities and carriers as individual particles [2, 3]. This presents a severe computational load, both in the solution of Poission's equation and because the equation of motion must be integrated through a complex potential landscape. Rapid variation of the potential in space may create difficulties with mesh-based calculations, both for the dynamics and Poisson's equation.

In this paper we discuss some issues associated with the *ab-initio* treatment of ionised impurity scattering, where the potential of the impurities is included directly in the dynamics of a Monte Carlo simulation rather than being introduced as a scattering rate. We have studied several methods for the solution of Poisson's equation and the integration of the equation of motion to evaluate

* Corresponding author.

their efficiency and accuracy. For Poisson's equation we compared a mesh-based solution with a near-exact potential based on Ewald summation [4]. The equation of motion was integrated either with a high-order adaptive Runge-Kutta scheme, for free carriers and impurities alone, or with a low-order Euler method within Brownian dynamics [5], which also includes phonons and proves to be more forgiving.

These methods were tested by calculating the mobility of electrons in a slab with randomly distributed impurities. Their application is also illustrated with the simulation of a 0.08 μm channel length dual gate GaAs MESFET.

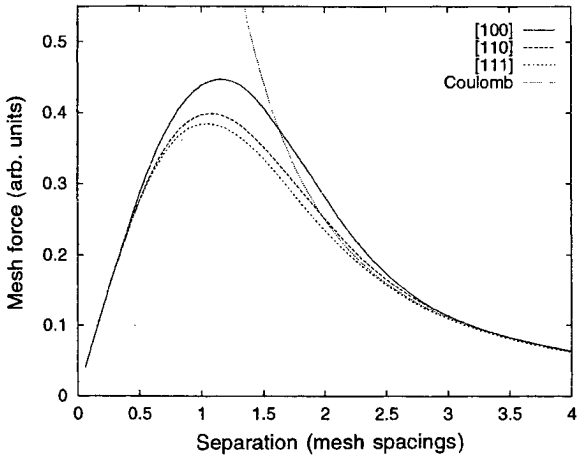


FIGURE 1 The forces between two point charges on along the main axes of symmetry of a cubic mesh cell. The true Coulomb force has been shown for comparison.

2. COMPUTATIONAL ASPECTS

The simplest approach to include individual impurity charges in a particle-based device simulation is to ascribe them to a mesh and to solve Poisson's equation as a boundary value problem. This method is efficient and works for devices with arbitrary boundaries. Unfortunately, difficulties arise close to charges where the $1/r$ Coulomb potential diverges. The errors in calculating the interparticle forces from a mesh-based solution of Poisson's equation along the major axis of symmetry of a cubic mesh cell are illustated in Figure 1. The errors are significant when the distance between the charges becomes smaller than 2–3 mesh spacings. It is practical to correct the short-range part of the interaction estimated from the mesh analytically. For a 3D solution of Poisson's equation on an uniform mesh we use a simple linear interpolation for the corrections which reflects the symmetry of the mesh cell.

Direct analytic evaluation of the Coulomb forces between point charges can give accurate results but becomes to costly when large numbers of particles are involved in the simulation, or where the boundaries of the device are complex. For simple cases, with periodic boundary conditions, the method of Ewald summation [4] can be used. We have used a polynomial approximation to the Ewald sums with fitted coefficients and compared the results with mesh-based solution of Poisson's equation.

Integration of the equations of motion in the complex potential landscape associated with the individual charges poses problems because the magnitude of the Coulomb force varies rapidly near the impurities. This necessitates the use of high-order integration schemes and small time-steps for free carriers. In this case we use an adaptive Runge-Kutta approach. Errors accumulate during the simulation and tight tolerances are therefore required.

Phonon scattering can be included using Brownian dynamics [5]. This is essentially an ensemble Monte Carlo method with a greatly simplified scattering term. Low-order Euler integration of the equation of motion is satisfactory in this case because the interaction with the phonons built into Brownian dynamics tends to erase errors in the integration of forces.

3. IONISED IMPURITY SCATTERING

The methods discussed above have been tested by calculating the mobility of electrons in bulk GaAs at room temperature. The crystal was represented

as a periodic array of cubic cells and ionised impurity scattering was introduced through the potential of randomly distributed impurities in each cell. A uniform background charge was added to maintain overall neutrality. Poisson's equation was solved using a multigrid technique for the mesh-based methods. A typical pontential distribution in a $(50\ \text{nm})^3$ cell is given in Figure 2. The same positions of impurities were used in the Ewald summation approach.

The mesh spacing must not exceed the distance at which the Coulomb interaction equals the thermal energy – about 2.8 nm at room temperature. In our calculations we used a 1 nm mesh. It is natural to assume that the cell size has to be larger than the mean free path of the carriers. For doping concentration $N_D = 10^{24}\ \text{m}^{-3}$ the mean free path is approximately 50 nm and a mesh with more than 50^3 points is therefore needed. This introduces a problem for lower concentrations where the mesh has to be even larger. However, experiments for $N_D = 10^{24}\ \text{m}^{-3}$ with meshes ranging from 10^3 to 200^3 points gave a very small spread in the calculated mobility, $\mu_n = 0.417 \pm 0.008\ \text{m}^2/\text{Vs}$. The history of 65000 particles was followed in the simulation and the diffusion coefficient was calculated from the variance in the range using a least-squares fit. The mobility was then deduced from the Einstein relation.

The results are summarised in Figure 3 and compared with the most commonly used analytical formulas. Phonon scattering was excluded from the Brownian simulations using Mattheisen's rule where necessary. The Brownian dynamics simulation with a mesh-based solution of Poisson's equation is in good agreement with the Conwell–Weisskopf formula over the whole range of doping. Brownian dynamics in combination with the more accurate Ewald sum for Poisson's equation gives lower mobility than the mesh-based results. The results for the Ewald summation and free carrier dynamics show a marked reduction in mobility which is more pronounced as the doping level is increased. The most likely cause of this phenomenon is the trapping of carriers in the potential well at the site of each impurity. The lack of phonon scattering means that is no mechanism for the electrons to exchange energy with the crystal and reach thermal equilibrium. It is therefore important to impose the correct initial distribution of carriers. The use of an unmodified Maxwell–Boltzmann distribution to initiate the simulation would pose severe problems because it predicts that all carriers are trapped in the Coulomb wells.

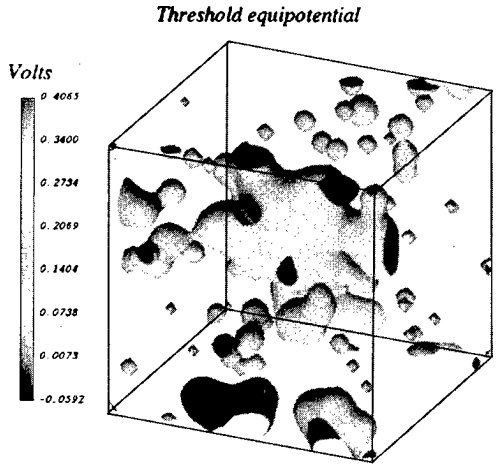


FIGURE 2 Random potential in a uniformly doped slab near the percolation threshold.

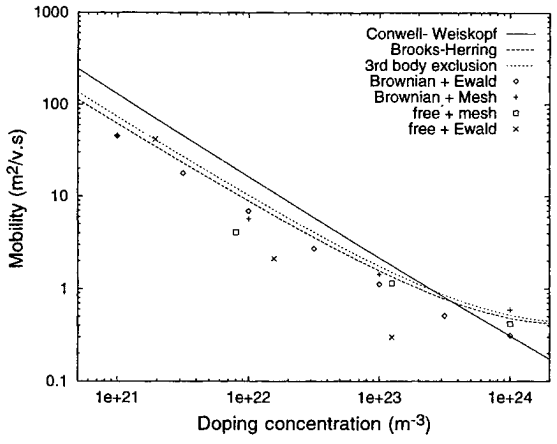


FIGURE 3 Mobility as a function of the doping concentration.

4. DUAL GATE MESFET EXAMPLE

The Brownian method on a mesh was used in a simple simulation of an 80 nm dual-gate MESFET with $N_D = 5\times10^{23}\,m^{-3}$ in the channel. Figure 4 illustrates the potential in the channel near threshold; fluctuations due to the random donors are evident. Figure 5 shows that the results of this atomistic simulation agree well with drift-diffusion results from MEDICI, at low source-drain voltage. The slope of the $I_D(V_G)$ characteristic is proportional to the mobility, and shows that impurity scattering is properly included in the Brownian simulation.

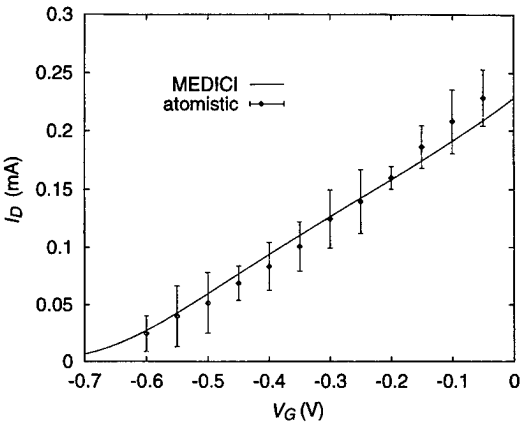


FIGURE 5 Transfer characteristics of an 80 nm channel dual gate GaAs MESFET, comparing the "atomistic" Brownian simulations and drift-diffusion MEDICI simulations at low drain voltage ($V_D = 0.05\,V$).

5. CONCLUSIONS

We have considered different approaches to the solution of both Poisson's equation and the equation of motion in a simulation where ionized impurities are treated as discrete random charges. We conclude that the simpler mesh-based solution of Poisson's equation does not introduce significant errors, and reproduces well the expected mobility as a function of doping. Highly accurate integration of the equation of motion is needed for free carriers, but the interaction with phonons in the Brownian method makes it more tolerant of errors.

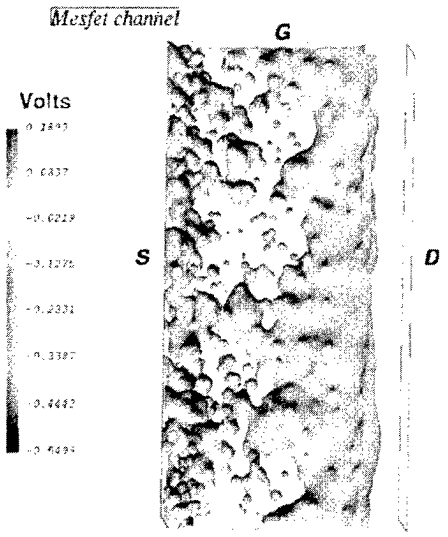


FIGURE 4 Equipotential surfaces in the channel of 80 nm channel dual gate GaAs MESFET at threshold.

References

[1] Wong, H.-S. and Taur, Y. (1993). "Three-dimensional 'atomistic' simulation of discrete random dopant distribution effects in sub-0.1 μm MOSFETs", *IEDM Tech. Digest*, 705–708.

[2] Ferry, D. K., Kriman, A. M., Kann, M. J. and Joshi, R. P. (1991). "Molecular dynamics extension of Monte Carlo simulation in semiconductor device modeling", *Computer Physics Communication*, **67**, 119–134.

[3] Tarnay, K., Masszi, F., Poppe, A., Verhas, P., Kocsis, T. and Kohari, Z. (1993). "A 3D Monte Carlo semiconductor device simulator for submicron MOS transistors", *Software Applications in Engineering*, **CH-46**, 345–358.

[4] Adams, D. J. and Dobey, G. (1987). *J. Comp. Physics*, **72**, 156.

[5] Arokianathan, C. R., Asenov, A. and Davies, J. H. (1996). "An approach based on Brownian motion for the simulation of ultrasmall semiconductor devices", *J. Applied Physics*, **80**(1), 1–7.

Authors' Biographies

Clinton Arokianathan graduated from Glasgow University with a B.Eng. in Physics and Electronic

Engineering in 1993. He then undertook a Ph.D. at the Department of Electronics and Electrical Engineering, where he developed parallel and serial code for the molecular dynamics simulation of submicron devices. He now works on recognition systems and vision processing for Roke Manor Research in Romsey.

John Davies gained his Ph.D. from Cambridge University on aspects of electronic stages in amorphous semiconductors. He later worked on the quantum theory of high-field transport in semiconductors at Cornell. Since coming to Glasgow in 1986, much of his research has been concerned with the modelling of ultra-small structures used in quantum transport. In particular, long-ranged fluctuations from ionized donors were shown to have a serious effect on transport in wires and quantum point contacts. Other interests include resonant tunnelling and effects of stress in nanostructures.

Asen Asenov had 10 years industrial experience as a head of the Process and Device Modelling Group in IME–Sofia, developing one of the first integrated process and device CMOS simulators IMPEDANCE. He is currently a Reader in the Department of Electronics and Electrical Engineering, Glasgow University. As a leader of the Device Modelling Group he is involved in the development of 2D and 3D device simulators and their application in the design of compownd FETs, SiGe MOSFETs and IGBTs. Other interests include design of parallel algorithms.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 337–341
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Numerical Evaluation of Iterative Schemes for Drift-diffusion Simulation

MAHESH B. PATIL [a], UMBERTO RAVAIOLI [b,*] and THOMAS KERKHOVEN [c]

[a] *Department of Electrical Engineering, Indian Institute of Technology at Kanpur, Kanpur, U.P. 208016, India;* [b] *Beckman Institute, University of Illinois at Urbana-Champaign, 405 N. Mathews Avenue, Urbana, IL 61801, USA;* [c] *Department of Computer Science, University of Illinois at Urbana-Champaign, 1304 W. Springfield Avenue, Urbana, IL 61801, USA*

We introduce an iterative scheme to solve the drift-diffusion device simulation problem, which combines the Gummel iteration with the "pointwise iteration", and then we compare its convergence behavior with other iteration strategies, for different test cases. Comparisons are made with the standard Gummel approach and a nonlinear multigrid iteration. The combined "Gummel-Pointwise" iteration has significantly better convergence characteristics than the Gummel iteration in all cases. The cost of the pointwise iteration varies only linearly with the number of grid points. While in terms of CPU time for the solution the pointwise iteration is not always faster in 2-D, in 3-D the combined technique becomes more and more advantageous as the grid size increases. It is found that the nonlinear multigrid scheme is overall not as effective as the combined iteration.

Keywords: Drift-diffusion, iterative methods, gummel iteration

1 INTRODUCTION

A major problem for the development of drift-diffusion solvers for 3-D applications is the difficulty in adapting existing 2-D schemes. The Gummel and Newton iteration methods are most commonly used to solve the nonlinear steady-state problem arising from the three semiconductor device equations [1], (Poisson equation, electron continuity and hole continuity, in the order)

$$\nabla^2\psi = -\frac{q}{\varepsilon}(p - n + N_d^+ - N_a^-) \tag{1}$$

$$\vec{\nabla}\cdot\mathbf{J}_n - qR = 0 \tag{2}$$

$$\vec{\nabla}\cdot\mathbf{J}_p + qR = 0 \tag{3}$$

where all symbols have their usual meaning.

* Corresesponding author. Tel.: (217) 244-5765, Fax: (217) 244-4333, e-mail: ravaioli@uiuc.edu.

In the case of Gummel iteration, the equations are solved in a decoupled manner. Poisson equation is solved at all grid points, followed by the electron continuity equation, and then by the hole continuity equation, repeating the procedure until convergence. In Newton's method the three device equations are solved simultaneously using a linearization procedure. The Gummel iteration always maintains the coupling between grid points, arising from discretization, while the equations are loosely coupled through successive updates of the variables ψ, n and p. In the Newton iteration, both the coupling between neighbouring grid points and that between the three equations is maintained during each iteration.

In practice, the Gummel method is found to be efficient if the coupling between the equations is weak; otherwise, the convergence may become quite slow. The Newton method is more robust, and converges in relatively few iterations, provided that the initial guess is sufficiently close to the solution. The major drawback of Newton's method is that it involves a matrix of size $3N \times 3N$ (where N is the number of grid points), which increases both memory requirements and CPU time at each iteration. These constraints are particularly severe in three dimensions if a direct method for matrix solution is to be used.

2 NUMERICAL APPROACH

Several strategies have been reported to circumvent the limitations of Gummel's and Newton's iterations [2–7], such as approaches for acceleration of convergence in Gummel's scheme, reduction of memory requirements in Newton's scheme, and so on. The purpose of this paper is to explore alternative iteration schemes to solve the device equations, and to evaluate their performance through numerical experiments.

A complementary approach to Gummel iteration is represented by the so called pointwise iteration. Here, the three equations are solved simultaneously at a given grid point, treating the variables at adjacent points as constants, and this procedure is repeated for all grid points. Thus, the coupling between the equations is maintained at each grid point, whereas the grid points are coupled to each other only loosely through the update of values at the neighbors.

The pointwise iteration does not need a large matrix to be stored as the equations are solved simultaneously one point at a time; this makes the iteration very attractive from the memory point of view. This procedure has been used as the smoothing iteration in nonlinear multigrid solution of the device equations [9–12]. Also, in the Alternate Block Factorization [7] and ILU/Knot [6] methods, the idea of coupling the three equations at each grid point has been used, although not explicitly as an iteration.

In our numerical experiments we have used rectangular grids, and it is convenient to adopt a *red-black* reordering as in [12], where the mesh points are arranged as red and black squares on a checker board. The three equations are solved for all red points first, followed by all black points. At each grid point, we have three equations in three unknowns (ψ, n, and p). In our specific implementation, we first perform a few Gummel-type iterations. If the residuals are not smaller than a specified tolerance, at the end of the Gummel-type iterations, we switch to Newton's method on the 3×3 system to solve the equations at that grid point.

Since the Gummel and the Pointwise iterations have complementary aspects, it is interesting to combine the two approaches to examine the effect on the convergence rate. This combined iteration, which we call the "GP_M" iteration consists of the following: (a) perform one Gummel iteration, (b) perform M times the pointwise iteration. As we will show later through numerical examples, this scheme can significantly accelerate the convergence.

To complete the comparison, we have also implemented a NonLinear MultiGrid method (NLMG) as described in [12]. Here, the pointwise iteration described above is used for smoothing,

full weighting for the restriction of residuals, injection for the restriction of solution, and bilinear interpolation for transferring error from coarse to fine level. We have implemented V-cycle with $\nu_1=2$ and $\nu_2=1$ where ν_1 and ν_2 are the number of pre-smoothing and post-smoothing steps respectively. For the NLMG iteration to converge, the coarse level cannot be arbitrarily coarse as pointed out in [12]. For our examples with a limited grid size, we found that the iteration did not converge with three levels, therefore, only two levels have been used. Two approaches can be used to solve the problem on the coarsest level: (a) The three equations may be solved simultaneously at all grid points using Newton's method as in [12]. (b) A large number (typically 50 to 100) of smoothing iterations may be performed to solve the coarse-grid problem approximately. This approach is also employed in [10] although, in that work, the grid points are scanned in a different order. Both approaches have been tried here. We will refer to one V-cycle as one Non-Linear MultiGrid (NLMG) iteration, NLMG with Newton's method at the coarse level as NLMG (N) and NLMG with m smoothing iterations at the coarse level as NLMG (mS).

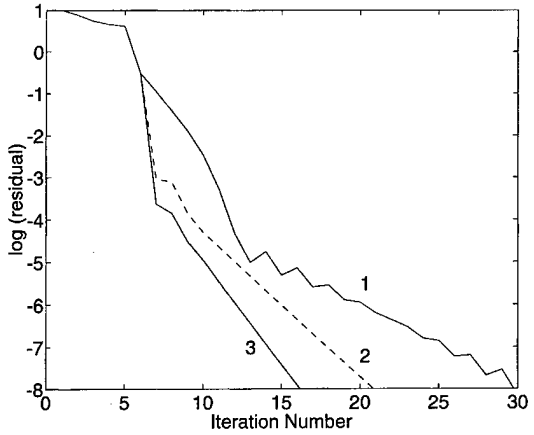


FIGURE 1 2-norm of residual versus iteration number for 2D p^+n device with a forward bias of 0.8 V and the solution for 0.6 V as the initial guess. (a) 1: Gummel, 2: GP_1, 3: GP_{25}. Damping was employed for the first 5 iterations. (b) 1: NLMG (N), 2: NLMG (50 S), 3: NLMG (100 S).

3 SIMULATION RESULTS

Because of space limitations, we present here only two examples of 2-D simulation, involving a silicon p^+n junction with a uniform substrate n-doping of 10^{17} cm^{-3}) and a Gaussian diffusion for the p^+-region with peak density 5×10^{18} cm^{-3} at the contact, and a 3-D example with a n^+n with the same doping levels. To facilitate the comparison, we use here a uniform grid with 33×33 points in 2-D and $17\times17\times17$ in 3-D. Similar tests conducted with non-uniform rectangular grids produced very similar behavior.

For the 2-D simulation we consider a situation of large forward bias (0.8 V) where the solution for 0.6 V forward bias is used as the initial guess (Figs. 1(a) and 1(b)), and also a reverse bias of 0.5 V with the equilibrium solution as the initial guess (Fig. 2). It is clearly seen that the GP_M iteration converges in significantly fewer iterations than the Gummel iteration in all cases. For the forward bias case, GP_1 and NLMG (N), respectively, are the most efficient in terms of number of iterations required for convergence. Note also that the convergence of the NLMG (mS) iteration is

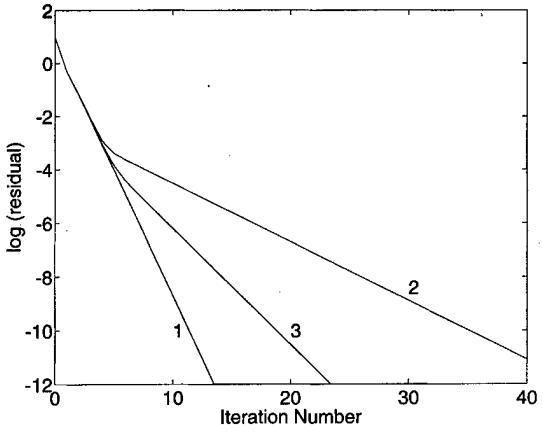


FIGURE 2 2-norm of residual versus iteration number for 2D p^+n device with a reverse bias of 0.5 V and equilibrium initial guess. Damping was employed for the first 5 iterations. 1: Gummel, 2: GP_1.

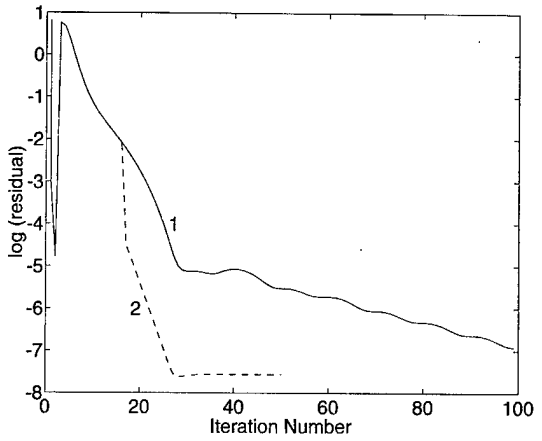


FIGURE 3 2-norm of residual versus iteration number for 3D n^+n device with 0.5 V and equilibrium initial guess.1: Gummel, 2: GP_{25}. Damping was employed for the first 15 iterations.

slow even for $m = 100$ which was found to be very effective instead in other simulations, not shown here, involving a forward biased $n^+ - n$ junction. For reverse bias, our simple NLMG iteration did not converge. We found that the coarse-grid problem for this case does not represent the fine-grid problem accurately enough because the depletion region on the n side is only two or three grid points wide. A locally adapted multigrid method [10] is more appropriate for this situation. In terms of CPU time for the example of Figure 1(a), the cost per iteration on the 33×33 grid was found to be: 0.155 s (Gummel); 0.118 s (Pointwise); 0.714 (NLMG(N)). This translates into total solution times of: 6.1 s (Gummel); 6.8 s (GP_1); 32.0 s (GP_25); 8.8 s (NLMG(N)); 24.0 s (NLMG(50S)); 22.0 s (NLMG(100s)). This shows that in 2-D problems there is little computational advantage in using alternative iteration schemes.

However, as drift-diffusion problems are pushed into 3-D, the combined Gummel-Pointwise iteration has immediately a marked advantage over the traditional Gummel iteration, since the computational cost of the pointwise iteration only grows linearly with the number of points. As a final example, we consider a 3D n^+n structure, with a $17 \times 17 \times 17$ grid which is uniform in each direction, with a voltage of 0.5 V applied to the n^+ contact. The equilibrium solution is used as the initial guess. For this example, Poisson's equation was solved using the multigrid method (correction scheme), and the continuity equations were solved using YSMP [8]. The residual for the Gummel and the GP_{25} iterations are shown in Figure 2. Again, the GP iteration converges in fewer iterations than Gummel. But while GP_{25} requires much more CPU time to solve the problem for a similar 2-D n^+n structure (83.0 s, as opposed to 29.0 s for Gummel iteration), in the 3-D case the Gummel solution requires 1065.0 s, and the GP_{25} solution only 595.0 s.

We conclude that the proposed Gummel-Pointwise iterative scheme in general improves the convergence behaviour of the Gummel scheme significantly. In terms of CPU time, the new scheme is advantageous for certain 2D cases; and for 3D simulation, it is expected to be always more effective than Gummel. It is suggested that implementation of the pointwise iteration on a parallel-architecture computer will enhance its effectiveness. The red black pointwise iteration can be naturally parallelized, as all red (or black) points can be processed in parallel and independently. This will drastically reduce the cost of the pointwise iteration, and the GP iteration will then be only about as expensive as Gummel iteration but with much better convergence characteristics.

Acknowledgements

The authors are thankful to Dr. Edwin Kan and Dr. Steve Laux for helpful comments. This work was supported by the National Science Foundation grants NSF ECS 95-26127 (NCCE) and ECS 95-09751.

References

[1] Selberherr, S. (1984). *Analysis and Simulation of Semiconductor Devices*, New York, Springer-Verlag.
[2] Toyabe, T., Masuda, H., Aoki, Y., Shukuri, H. and Hagiwara, T. (1985). "Three-dimensional device simulator CADDETH with highly convergent matrix solution algorithms", *IEEE Trans. Electron Devices*, **32**(10), 2038–2044.
[3] Wu, K.-C., Chin, G. R. and Dutton, R. W. (1991). "A STRIDE towards practical 3-D device simulation", *IEEE Trans. Computer-Aided Design*, **10**(9), 1132–1140.

[4] Pommerell, C. and Fichtner, W. (1991). "New developments in iterative methods for device simulation", *Proc. SISDEP 4 Conf.*, pp. 243–248.
[5] Simon, C., Sadkane, M. and Mottet, S. (1993). "Newton-GMRES method for coupled nonlinear systems arising in semiconductor device simulation", *Proc. SISDEP 5 Conf.*, pp. 243–248.
[6] Rafferty, C. S., Pinto, M. R. and Dutton, R. W. (1985). "Iterative methods in semiconductor device simulation", *IEEE Trans. Electron Devices*, **32**(10), 2018–2027.
[7] Bank, R. E., Chan, T. F., Coughran, W. M. Jr. and Smith, R. K. (1989). "The alternate-block factorization procedure for systems of partial differential equations", *BIT*, **29**, 938–954.
[8] Ellman, H. C. (1982). *Iterative methods for large, sparse,* nonsymmetric systems of linear equations, Yale University Computer Science Department Research Report 229.
[9] Constapel, R. and Berger, M. (1989). "A multigrid approach for device simulation using local linearization", in *Proc. NASECODE 6 Conf.*, pp. 355–359.
[10] Molenaar, J. and Hemker, P. W. (1990). "A multigrid approach for the solution of the 2D semiconductor equations", *Impact of Computing in Science and Engineering*, **2**, 219–243.
[11] De Zeeuw, P. M. (1992). "Nonlinear multigrid applied to a one-dimensional stationary semiconductor model", *SIAM J. Sci. Stat. Comput.*, **13**(2), 512–530.
[12] Joppich, W. and Mijalkovic, S. (1993). *Multigrid methods for process simulation*, New York, Springer-Verlag.

Authors' Biographies

Mahesh B. Patil received the B.Tech. degree from Indian Institute of Technology, Bombay, India, in1984, the M.S. dregree from the University of Southern California, Los Angeles, in 1987, and the Ph.D. degree from the University of Illinois at Urbana-Champaign in 1992, all in electrical engineering. In 1993 he was a visiting scientist with the Hitachi Central Research Laboratories, Japan. He is currently an Associate Professor of electrical engineering with Indian Institute of Technology, Kanpur, India.

Umberto Ravaioli received the Laurea in Electronics Engineering in 1980 and the Laurea in Physics in 1982, from the University of Bologna, Italy, and the Ph.D. in Electrical Engineering from Arizona State University, in 1986. He is now a Professor of Electrical and Computer Engineering at the University of Illinois, Urbana-Champaign.

Thomas Kerkhoven received the Drs. degree in Physics from the University of Amsterdam, Nederlands, in 1981, and the Ph.D. in numerical analysis from the Department of Computer Science of Yale University in 1985. He is now a Professor of Computer Science at the University of Illinois, Urbana-Champaign.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 343–347
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Simulation of Si-MOSFETs with the Mutation Operator Monte Carlo Method

JÜRGEN JAKUMEIT[a], AMANDA DUNCAN[b], UMBERTO RAVAIOLI[c,*] and KARL HESS[d]

[a] *II. Phys. Inst., University of Köln, Zülpicher Str. 77, D-50937 Köln, Germany;*
[b] *Intel Corporation, RA1-309, 5200 N.E. Elam Young Parkway, Hillsboro, OR 97124-6497;*
[c] *Beckman Institute, University of Illinois at Urbana-Champaign, 405 N. Mathews Avenue, Urbana, IL 61801, USA;*
[d] *Beckman Institute, University of Illinois at Urbana-Champaign, 405 N. Mathews Avenue, Urbana, IL 61801, USA*

The Mutation Operator Monte Carlo method (MOMC) is a new type of Monte Carlo technique for the study of hot electron related effects in semiconductor devices. The MOMC calculates energy distributions of electrons by a physical mutation of the distribution towards a stationary condition. In this work we compare results of an one dimensional simulation of an 800 nm Si-MOSFET with full band Monte Carlo calculations and measurement results. Starting from the potential distribution resulting from a drift diffusion simulation, the MOMC calculates electron distributions which are comparable to FBMC-results within minutes on a modern workstation. From these distributions, substrate and gate currents close to experimental results can be calculated. These results show that the MOMC is useful as a post-processor for the investigation of hot electron related problems in Si-MOSFETs. Beside the computational efficiency, a further advantage of the MOMC compared to standard MC techniques is the good resolution of the high energy tail of the distribution without the necessity of any statistical enhancement.

Keywords: Monte Carlo simulation, hot electrons

1 INTRODUCTION

During the last two years we have developed a new approach to hot electron related problems in Si-MOSFETs, which is based on a mixture of Evolutionary optimization algorithms (EA) [1] and Monte Carlo (MC) technique [2]. The EA is used to optimize energy distributions of electrons towards an agreement with measurement results like substrate or gate currents. A major problem in this quasi-backward calculation of electron distributions is that many different distributions fit to one measured current value. It is very difficult or even impossible to force the EA towards physical correct distributions only by restrictions. We overcome this problem by using a Physical

* Corresponding author. Tel.: (217) 244-5765, Fax: (217) 244-4333. E-mail: ravaioli@uiuc.edu.

Mutation Operator (PMO) within the EA. The PMO forces the EA to converge to distributions which are in agreement with the physical model of the PMO. The PMO itself is the basis for a novel type of Monte Carlo technique, the Mutation Operator Monte Carlo method (MOMC), which applies the PMO to an initial distribution repeatedly, until a stationary condition is reached. The effectiveness of the EA depends greatly on the quality of the PMO, which should be at the same time accurate and inexspensive. The present work focuses on specific details and applications of the MOMC itself. The approach is built on a relatively simple transport model which is calibrated with a traditional Monte Carlo simulation. We would like to stress that since the MOMC was conceived as the first approximation step for an EA approach, it cannot be as accurate in all details as the full band Monte Carlo. But when the parameters of the MOMC are well calibrated from Monte Carlo simulation results, as done here, the MOMC can then be used as stand-alone method when one needs a rapid approximate evaluation of hot carrier distribution. Ideally, this method should be used as a first order test, to select conditions for which a more costly Monte Carlo simulation should be run.

2 MUTATION OPERATOR MONTE CARLO METHOD

The main part of the MOMC is the PMO, which mutates an energy distribution of electrons by a Monte-Carlo-like simulation of the electron transport (see Fig. 1). The distribution in a small interval around a randomly choosen energy E_{start} and position x_{start} is associated to a MC-electron. The drift and scattering processes of this electron are simulated in the same way as in a standard MC-simulation, until a final energy E_{final} and position x_{final} are reached, and the electron density represented by the MC-electron is returned to the distribution. For the MC-simulation of the electron movement a simple one-valley effective mass

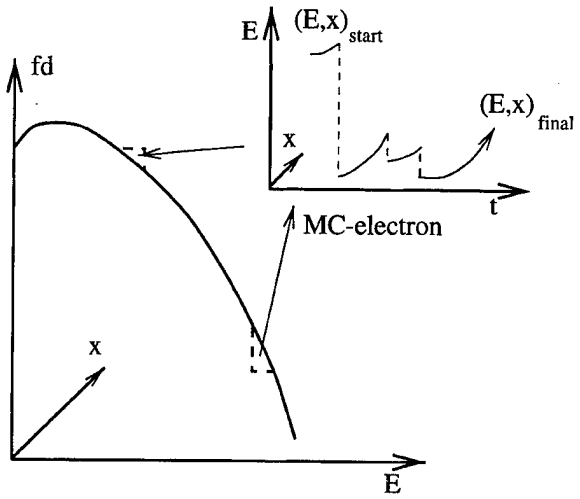


FIGURE 1 Principle of the physical mutation operator.

approach is used, taking the effective mass m_{eff} as a free parameter. Scatterings due to phonon absorption, phonon emission and impact ionization are included in the model. The phonon scattering rates are calculated by a full band approach [3]. Instead of many rates for scattering between different valleys one averaged scattering rate for absorbtion and one for emission is used (see [2] for details). For impact ionization the rate computed in [4] was taken (assuming zero field and no collision broadening). The phonon scattering and the impact ionization rates are then multiplied by the factors f_{phonon} and f_{impact}, respectively. Introduction of these two parameters gives us more flexibilty to calibrate the simple model by comparison with full band approaches.

The Monte Carlo simulation in the PMO needs the carrier density and electric field or potential profile in the simulation region as input (along the channel in case of Si-MOSFETs).

Since our main interest is to use the MOMC to calculate the high energy tail of the distribution, which has only a small impact on the total carrier distribution, in this approach the electric field and the carrier density are taken as inputs and Poisson's equation is not solved self-consistently.

From the electron distributions calculated by the MOMC substratecurrents I_{sub} are calculated

assuming that I_{sub} is determined by the amount of holes created by impact ionization:

$$I_{sub} = e \cdot \int_{channel} \int_0^\infty fd(E,x) W_{ii}(E) dE\, dx,$$

where e is the electron charge, $W_{ii}(E)$ is the impact ionization rate and $fd(E,x)$ the energy distribution of electrons integrated over the depth of the simulated region (with electron distribution $fd(E, x)$ we denote the distribution function times density of states).

The gate current I_{gate} is estimated using a WKB approach with Schottky lowering of the barrier [5]:

$$I_{gate} = K_{hit} \cdot e \cdot \int_{channel} \int_0^\infty fd(E,x) P_{trans}(E,x) dE\, dx,$$

where $P_{trans}(E, x)$ is the probability to cross the SiO_2-layer. We assumed that the rate at which electrons trying to cross the Si/SiO_2 interface is proportional to the carrier density. The proportional constant $K_{hit} = 7.5 \cdot 10^{-9}\,s^{-1}$ was calibrated by a comparison with FBMC and measurement results.

3 RESULTS AND DISCUSSION

In this work we present results from an one dimensional simulation of a 0.8 μm Si-MOSFET, for which experimental data were published in the classic paper by Ng *et al.* [6]. The input for the MOMC, i.e. the carrier density and electric field profile, was calculated by a Full Band Monte Carlo (FBMC) program developed at the University of Illinois and by a HydroDynamic (HD) model (ISE AG, Zürich).

Figure 2 shows the electron distribution given by MOMC and FBMC for a drain voltage $Vd = 6\,V$ and a gate voltage $Vg = 7\,V$ in the region important for substrate and gate current generation, i.e. the crossover from channel to drain (the drain region starts at location $x = 1$ μm; source and substrate contacts are held at 0 V bias). MC-MOMC denotes the MOMC results, which were

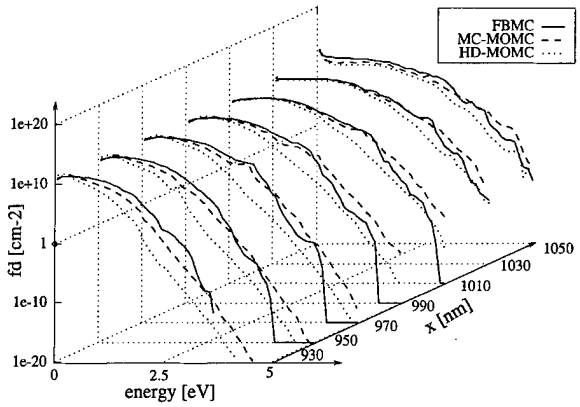


FIGURE 2 Energy distribution of electrons in a 800 nm Si-MOSFET($V_{drain} = 6\,V$, $V_{gate} = 7\,V$) at the crossover from channel to drain. MOMC results calculated from FBMC input (dashed line, MC-MOMC) and HD input (dotted line, HD-MOMC) are compared with FBMC simulation results (solid line).

obtained using the FBMC results for carrier density and electric field as input, and HD-MOMC the MOMC results calculated on the basis of HD-model input. MOMC results calculated from FBMC input are in fairly good agreement with the FBMC results, especially at the crossover between channel and drain at 1 μm. The HD-MOMC distribution does not fit the FBMC results as well but is reasonably close for $x < 1$ μm.

The difference between HD-MOMC distributions and FBMC results for $x < 1$ μm can be explained by looking at Figure 3. Here, the electric field and drift velocity are plotted as functions of x. The HD-model gives a smaller rise of the electric field for $x < 0.9$ μm and then a steeper increase up to the peak at 1.01 μm, leading to a later heating of the electrons. This explains the lower electron distribution calculated from HD-MOMC input for $x < 1$ μm. The drift velocityprofiles from FBMC and MC-MOMC calculations are comparable. Both simulations yield a broad peak between 0.92 μm and 1.02 μm. The HD-MOMC simulation leads to a sharper peak at $x = 0.99$ μm reflecting the sharper peak of the electric field.

The good resolution of the high energy tail by the MOMC becomes obvious by looking at the

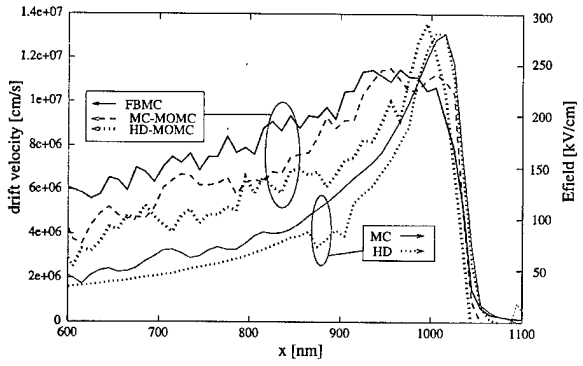


FIGURE 3 Drift velocity and electric field profile along the channel for an 800 nm Si-MOSFET ($V_{drain}=6$ V, $V_{gate}=7$ V) calculated by FBMC simulation (solid lines) and MOMC simulation using FBMC results (dashed lines) or HD-results (dotted lines) as input.

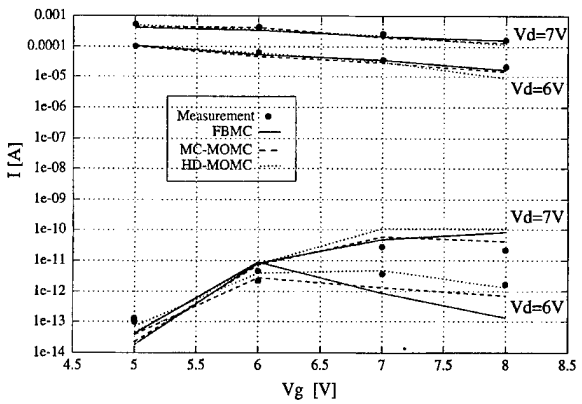


FIGURE 4 Comparison of substrate (top curves) and gate currents (bottom curves) calculated from FBMC, MC-MOMC and HD-MOMC results with experimental data from Ng *et al.* [5].

distributions above 4 eV (Fig. 2). The FBMC result drops to zero when the carrier density goes below 1^{-10} cm^{-2}. The MOMC resolves this regime with the same resultion as the low energy part. Theoretically, the resolution of the MOMC is independent of the order of magnitudes displayed. This could be proven by bulk simulations up to 80 orders of magnitude. A statistical enhancement, which was used in the FBMC simulation, is not necessary. Therefore, the MOMC could also be used as a post-processor for standard MC simulations, in order to avoid the computational complexity connected with statistical enhancement. Figure 4 compares substrate and gatecurrents calculated from the three different approaches with the measurement results of Ng *et al.* Obviously FBMC, MC-MOMC and HD-MOMC give an excellent agreement between measured and calculated substrate currents. For gate currents the agreement is not as good, showing differences up to one order of magnitude. Interestingly the gate currents calculated from FBMC distributions are not closer to the measurement results than those from MC-MOMC and HD-MOMC. It is not possible to judge from the comparison with measurement, which simulation approach yields the more realistic electron distributions.

The above results demonstrate that the MOMC gives energy distributions of electrons which are close to FBMC results and in agreement with measurement results even when HD-model calculations are used as input. For each (*Vd*, *Vg*)-pair in Figure 4, the MOMC calculates the corresponding distribution within several minutes on a HP735 workstation. Thus, the computational cost is comparable to HD-simulation and significantly less than FBMC calculations. This makes the MOMC a very interesting post processor for HD-simulations to gain insight on information like the energy distribution of carriers, without increasing the computational expense. The discussion above on the difference between FBMC and HD-MOMC distributions demonstrates also that such information is important for the interpretation of simulation results like the effect of slight changes in the electric field profile on the electron distribution and related hot electron effects.

4 CONCLUSIONS

A comparison of MOMC results with FBMC calculations and measurement results for a 800 μm Si-MOSFET shows that the MOMC can be used as a fast post processor in device simulation. The method is based on a physical mutation operator which changes (mutates) electron distributions by a Monte Carlo like simulation of the electron

movement using a simple effective mass model including electron-phonon scattering and impact ionization. Energy distributions of electrons calculated by the MOMC from hydrodynamic drift diffusion calculations were found to be comparable to FBMC results. The computational cost for the MOMC calculations is comparable to or smaller than the cost for hydrodynamic calculations. A next step is the application of the MOMC to short channel MOSFETs in the sub 100 nm range. Further improvements of the physical model, a nonparabolic approach to the bandstructure and additional scattering rates are planned.

Acknowledgements

This work was partially supported by the Deutsche Forschungsgemeinschaft contract No. JA853/1-1 (J.J.), the Semiconductor Research Corporation, Contract 96-CJ-816 (U.R and K.H.), and by fellowships from the US DOE Computational Science Graduate Fellowship Program and the Intel Foundation (A.D.).

References

[1] Jakumeit, J. (1995). "Genetic algorithms: A new approach to energy balance equations", *Appl. Phys. Lett.*, **66**(14), 1812–1814.

[2] Jakumeit, J., Ravaioli, U. and Hess, K. (1996). "Calculation of hot electron distributions in silicon by means of an evolutionary algorithm", *J. Appl. Phys.*, **80**(9), 5061–5066.

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

[4] Bude, J., Hess, K. and Iafrate, G. J. (1992). "Impact ionization in semiconductors: Effects of high electric fields and high scattering rates", *Phys. Rev. B.*, **45**(19), 10958–10964.

[5] Huang, C., Wang, T., Chen, C. N., Chang, M. C. and Fu, J. (1992). "Modeling hot-electron gate current in Si MOSFET's using a coupled drift-diffusion and Monte Carlo method", *IEEE Trans. on Electron Dev.*, **39**(11), 2562–2568.

[6] Ng, K. K. and Taylor, G. W. (1983). "Effects of hot carrier trapping in *n*- and *p*-channel MOSFET's", *IEEE Trans. on Electron Dev.*, **30**(8), 871–876.

Authors' biographies

Jürgen Jakumeit studied physics at the University of Cologne and finished his dissertation in 1994. His research concerned quantum size effects in semiconductor structures. Through two half year stays at the University of Michigan, Ann Arbor, MI, and the University of Illinois, Urbana-Champaign, IL, he got engaged in the field of device simulation. At the moment he works as Postdoc at the University of Cologne on the simulation of silicon MOSFETs and the investigation of quantum interference effects.

Amanda Duncan obtained the Ph.D. in Electrical Engineering at the University of Illinois, Urbana-Champaign, in 1996, with a dissertation on full band Monte Carlo simulation of Si devices. She is now a Senior Component Device Engineer with Intel Corporation, Hillsboro, Oregon.

Umberto Ravaioli is a Professor in the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign. His main research interests are in the areas of semiconductor device physics and simulation, numerical methods, and high performance computing.

Karl Hess is a Swanlund Chair Professor in the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign. His main research interests are in the areas of semiconductor device physics, quantum transport, and opto-electronics device simulation.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 349–353
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

A New HEMT Breakdown Model Incorporating Gate and Thermal Effects

LUTFI ALBASHA*, CHRISTOPHER M. SNOWDEN and ROGER D. POLLARD

Institute of Microwaves and Photonics, School of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT, UK

This paper presents a comprehensive physical model for the breakdown process in HEMTs. The model is integrated into in a fast quasi-two-dimensional HEMT physical simulator. The work is based on a full study of the complex interactions between the different breakdown mechanisms and the influence of design parameters. The model takes account of tunnelling effects in the region of the gate metallization, and of the thermal effects in the active channel under the gate region.

Keywords: Gate tunneling, thermal modelling, breakdown, quasi-two-dimensional modelling

1. INTRODUCTION

Microwave circuits such as power amplifiers operate under large-signal conditions. Their ability to perform efficiently is limited by the devices' breakdown characteristics, which limit the transistor's performance and power output. The accuracy of the large-signal design relies on the availability of suitable breakdown models. Popular breakdown theories have not been adequate to independently explain the full picture of the breakdown process in HEMTs. The effects of the gate leakage and substrate conduction in HEMTs on this have not been simulated. Analytical models for avalanche breakdown, such as Frensley's [1], were based on physical simulations of the active channel around the gate. The effects of the gate leakage and substrate conduction on this have not been included. This paper considers breakdown more comprehensively including the effects of reverse gate conduction, thermal fluctuations within the active region and substrate conduction [2].

2. MODEL DESCRIPTION

The Quasi-Two-Dimensional (Q2D) model used here is based on the earlier work of Snowden [3].

*Corresponding author. Tel.: +44 113 2332082, Fax: +44 113 2332032, email: l.albasha@elec-eng.leeds.ac.uk.

The Q2D method is based on the assumption that the fundamental driving force for electron transport is the x-directed electric field. The HEMT Q2D model important features have been reported previously [4]. In contrast to MESFETs, spurious substrate current occurs in HEMT buffer layers due to the lateral E_x field component. This current is drawn around the depleted channel and reduces the magnitude of the field. It is assumed in this paper that, as a result of the reduction in the electric field, an increase in the breakdown voltage is possible. Figure 1 shows a schematic diagram of a HEMT device.

The computational interpretation of the breakdown model presented in this paper is based on the interactions between the avalanche and gate leakage mechanisms. The relation that both processes simultaneously have with the device design parameters and power dissipation inside the device is included. The flow chart of the model is shown in Figure 2. The gate leakage, conventionally termed 'soft breakdown', is assumed to always occur in devices prior to avalanche. The flow of electrons from the gate into the semiconductor would then influence the impact ionisation process. Adding the leaked electrons to the channel electrons constituting the increasing drain current imposes this effect. The leakage is allowed in the three pinchoff stages.

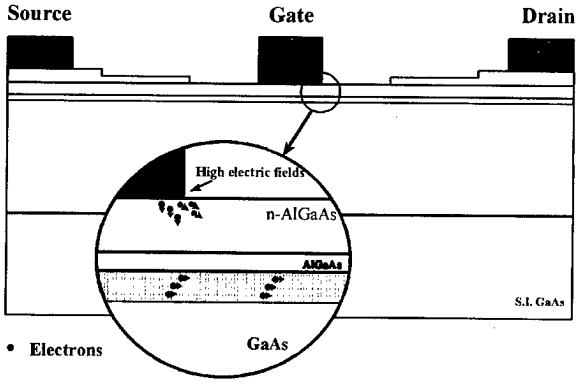


FIGURE 1 Schematic diagram of a HEMT device showing electrons leaking form gate.

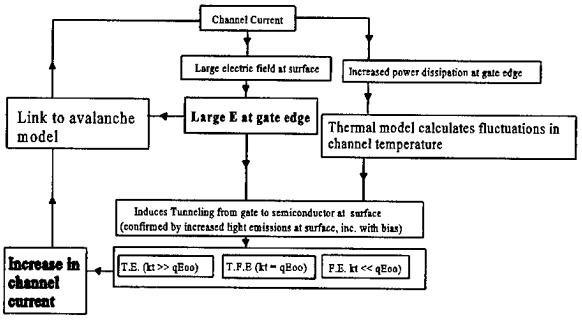


FIGURE 2 Flow chart of the gate breakdown model.

3. THERMALLY DRIVEN GATE LEAKAGE CURRENT

The gate leakage model is principally based on a combination of Padovani [5] and Rideout [6] equations. These equations involve complex functions of temperature, barrier height and semiconductor parameters. The form that these equations are presented in the above references is quite complicated. In order to maintain the numerical efficiency of the physical device simulator, new approximations are introduced in this paper, which simplify the tunneling functions and maintain the accuracy of the solution. Tunneling through the gate metal-semiconductor barrier becomes significant in the reverse direction than in the forward direction because the bias voltages involved are usually greater. This cause the potential barrier to become thin enough such that tunneling is dominant. The Thermionic (T), Thermionic-Field (T-F) and Field Emission (FE) gate leakage currents are dependent on the lattice temperature and material specifications. What determines the current mechanism is the temperature of the channel and the applied bias. A new thermal model [7] incorporated into the Q2D physical model computes the instantaneous temperature of the active channel. The output of this model is linked with the breakdown model. This assists the dynamic update of the gate leakage current mechanism.

3.1. Thermionic-field Tunneling Current

This mechanism is the major form of tunneling. The tunneling current is defined by the equation

$$I_{T-F} = AJ_s \exp(E/\varepsilon') \tag{1}$$

Where A is the gate area, E is applied energy calculated from qV_r, V_r being the applied reverse bias voltage given a positive sign throughout this work. q is the electron charge and ε' is an energy term defined as:

$$\varepsilon' = E_{oo}\left[\frac{E_{oo}}{KT} - \tanh\left(\frac{E_{oo}}{KT}\right)\right]^{-1} \tag{2}$$

The term E_{oo} quantifies the diffusion potential from metal into semiconductor. It has two equivalent definitions given in [5] and [6]. The latter definition is adopted in this work, however, slight numeric alteration is needed for consistency of units:

$$E_{oo} = 18.57 \times 10^{-15}\left(\frac{N}{m_r \varepsilon_r}\right)^{1/2} \tag{3}$$

N is the doping density per m^{-3}. In order to use equation (3) in (2), E_{oo} must be converted back by multiplying by q, the electron charge. The type of HEMT used connects the gate metalisation directly to the N-doped AlGaAs layer.

In equation (1), J_s is defined as the saturation current given in [5] as:

$$J_s = \frac{R(\pi E_{oo})^{1/2}}{KT}\left[qV_r + \frac{q\phi_b}{\cosh^2\left[\frac{E_{oo}}{KT}\right]}\right]^{1/2} \times \exp\left(\frac{-q\phi_b}{E_o}\right) \tag{4}$$

where ϕ_b is the Schottky barrier height. E_o is a term defined as:

$$E_o = E_{oo}\coth\left(\frac{E_{oo}}{KT}\right) \tag{5}$$

It was observed during the course of this work that in the saturation current equation (4), the square root term was dominated by the first term. Hence a simpler approximate numerical expression for the saturation current was deduced:

$$J_s = \frac{R}{kT}[\pi E_{oo} q V_r]^{1/2} \exp\left(\frac{-q\phi_b}{E_o}\right) \tag{6}$$

which is integrated into the model. An empirical difference limit between equations (4) and (6) was reached after some experimentation's beyond which if this limit is exceeded, the solution obtained from equation (6) was observed to affect the numerical accuracy of the tunneling current. The model then switches to the more stringent expression of equation (4). The thermionic and field emission currents are calculated using similar equations. Threshold equations are incorporated which, according to the thermal status of the device, the appropriate leakage mechanisms is invoked.

4. SIMULATION RESULTS

Figure 3 shows the effect of varying the temperature on the tunnelling currents at various

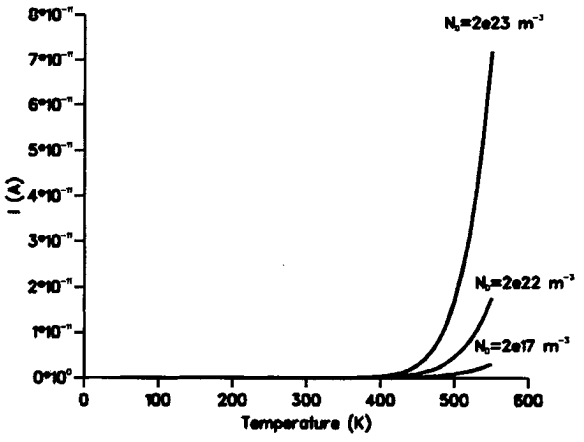


FIGURE 3 Tunneling current versus channel temperature at various doping concentrations.

doping levels. An increase in the T-F current is observed with an increase in the doping concentration. Figure 4 compares between the thermionic-field and field emission currents with respect to temperature. The T-F current is clearly of more prominent effect. The simulation of the DC I-V characteristics of a HEMT device enabling gate current tunnelling mechanism is shown in Figure 5. The soft breakdown mechanism is influenced by the design parameters and applied bias. It can be practically and numerically observed when the stimulating conditions are met. Comparisons with measured data in [8] shows good agreement between simulated and measured results of a MESFET case.

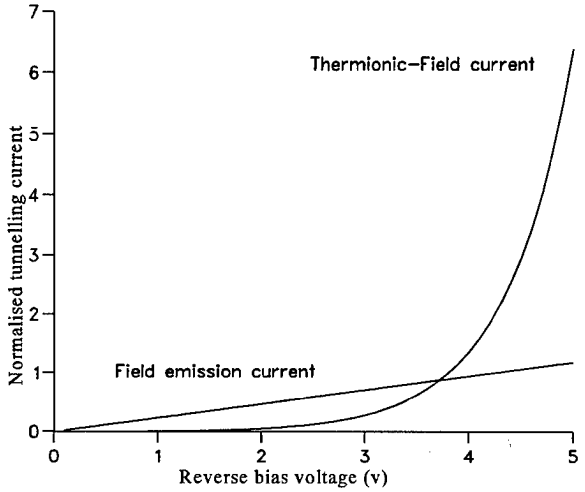


FIGURE 4 Comparison between the thermionic-field and field-emission leakage currents.

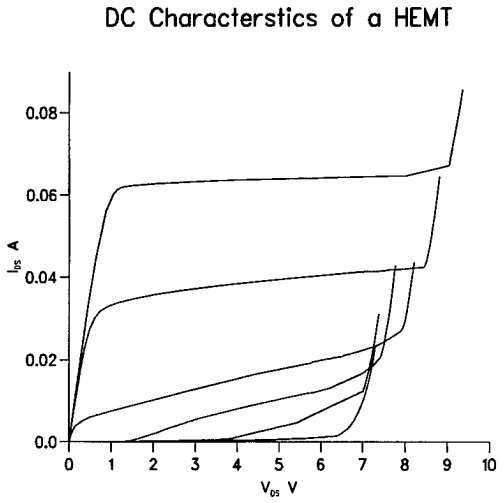


FIGURE 5 I-V characteristics of a HEMT device incorporating the gate model.

5. CONCLUSION

The Q2D physical device simulation of HEMTs has been expanded to include soft thermal breakdown effects using tunnelling current mechanisms. The influence of soft breakdown mechanism on the DC characteristics is influenced by the design parameters and applied bias. The thermal effects in the active channel under the gate region are included. This is conducted using an incorporated thermal model, which calculates the channel temperature and updates the thermal breakdown model.

References

[1] Frensley, W. R. (1981). "Power-limiting breakdown effects in MESFETs, " *IEEE Trans.*, **ED-28**(8), 962–970.

[2] Morton, C. G., Atherton, J. S., Snowden, C. M., Pollard, R. D. and Howes, M. J. "A large-signal physical HEMT model ," *Int. Microwave Symp. MTT-S*, **3**, 1759–1762, San Francisco.

[3] Snowden, C. M. and Pantoja, R. R. (1989). "Quasi-two-dimensional MESFET simulation for CAD", *IEEE Trans.*, **ED-36**(9), 1564–1574.

[4] Morton, C. G. and Snowden, C. M. "HEMT physical model for MMMIC CAD", *Proceedings of 25th European Microwave Conference*, Sept. 1995.

[5] Padovani, F. A. and Stratton, R. (1966). "Field and thermionic-field emission in schottky barriers", *Solid-State Electronics*, **9**, 695–707.

[6] Crowell, C. R. and Ridout, V. L. (1969). "Normalised thermionic-field (T-F) emission in metal-semiconductor (schottky) barriers", *Solid-State Electronics*, **12**, 89–105.

[7] Johnson, R. G., Snowden, C. M. and Pollard, R. D. (1997). "A Physics Based Electro-Thermal Model For Microwave And Millimetre Wave HEMTs", *International microwave symposium*, MTT-S, Denver, USA.

[8] Albasha, L., Snowden, C. M. and Pollard, R. D. "Breakdown Characterization of HEMTs and MESFETs Based on A New Thermally Driven Gate Model", *International Conference on Simulation of Semiconductors processes and Devices SISPAD*, 8–10 Sep. 1997, Cambridge, MA, USA (To be published).

Authors' Biographies

Lutfi Albasha received his B.Eng. (Hons.) degree in Electronic and Electrical Engineering from the

University of Leeds in 1990. In 1991 he obtained an M.Sc. (with distinction) in RF communications. In 1992, He joined the Microwave and Terahertz Technology Group at the University of Leeds and obtained his Ph.D. in 1995. His thesis was on the electromagnetic modelling of microwave circuits and its feasibility as a CAD tool. In 1994 he was employed by Filtronics Components Ltd., UK. Currently he is a research fellow at the University of Leeds conducting new studies on the high frequency breakdown modelling in HEMTs and its measurement techniques at W-band frequencies. His research interests include electromagnetic numerical modelling and the design and measurements of microwave circuits and devices.

Christopher M. Snowden received the B.Sc., M.Sc. and Ph.D. degrees from the University of Leeds. After graduating in 1977 he worked as an Applications Engineer for Mullard, Mitcham. His Ph.D. studies were conducted in association with Racal-MESL and were concerned with the large-signal characterisation of MESFET microwave oscillators. In 1982 he was appointed Lecturer in the Department of Electronics at the University of York. He joined the Microwave Solid State Group in the Department of Electrical and Electronic at the University of Leeds in 1983. He now holds the Chair of Microwave Engineering in the Microwave and Terahertz Technology Research Group and is also currently Head of the Department of Electronic and Electrical Engineering. During 1987 he was a Visiting Research Associate at the California Institute of Technology. He has been a Consultant to M/A-COM Inc., Corporate Research and Development since 1989, where he was on sabbatical leave during the period 1990–91. During this year he represented M/A-COM as Senior Staff Scientist. He was Chairman of the 1995 international Microwaves and RF Conference. He is a Member of the MIT Electromagnetics Academy. He is also a Top Scientist at the International Research Centre for Telecommunications-Transmission and Radar, Delft University of Technology, Netherlands. Professor Snowden is a Fellow of the IEEE and a Fellow of the IEE. He is a Distinguished Lecturer (1996/7) for the IEEE (Electron Devices Society). He is co-Chairman of the MTT-1 Committee and a Member of the 1997 IEEE MTT-S Technical Program Committee. His main research interests include compound semiconductor device modelling, microwave, terahertz and optical nonlinear subsystem design and advanced semiconductor devices. He has written 7 books and over 190 papers.

Roger D. Pollard was born in London, England in 1946. He received his technical education, graduating with the degrees of BSc and PhD in Electrical and Electronic Engineering, at the University of Leeds, Leeds, UK.

He holds the Hewlett-Packard Chair in High Frequency Measurements in the School of Electronic and Electrical Engineering at the University of Leeds where he has been a faculty member since 1974. He is Deputy Director of the Institute of Microwaves and Photonics which has over 40 active researchers, a strong graduate program and has made contributions to microwave passive and active device research. The activity has significant industrial collaboration as well as a presence in continuing eduction through its Microwave Summer School. Professor Pollard's personal research interests are in microwave network measurements, calibration and error correction, microwave and millimetre-wave circuits and large-signal and nonlinear characterization. He has been a consultant to the Hewlett-Packard Company, Santa Rosa, CA since 1981. He has published over 100 technical articles and hold 3 patents.

Roger Pollard is a Chartered Engineer, a member of the Institution of Electrical Engineers (UK) and a Fellow of the IEEE. He is 1998 President of the IEEE Microwave Theory and Techniques Society where he is serving his second term as an elected member Administrative Committee.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 355–360
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Rate Equation Modelling of Nonlinear Dynamics in Directly Modulated Multiple Quantum Well Laser Diodes

STEPHEN BENNETT [a,*], CHRISTOPHER M. SNOWDEN [b] and STAVROS IEZEKIEL [b]

[a] *Department of Electronic and Electrical Engineering, University College London, London, WCIE 7JE, UK;*
[b] *Department of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT, UK*

A theoretical (using rate equations) and experimental study of the nonlinear dynamics of a distributed feedback multiple quantum well laser diode is presented. The analysis is performed under direct modulation. Period doubling and period tripling are identified in both the measurements and simulations. Period doubling is found over a wide range of modulation frequencies in the laser. Computational results using rate equations show good agreement with the experimental results.

Keywords: MQW laser diodes, rate equations, nonlinear dynamics, period doubling, period tripling

1. INTRODUCTION

Multiple Quantum Well (MQW) laser diodes are widely used in many applications because of their superior performance characteristics over bulk laser diodes. These include higher modulation bandwidths and lower threshold currents, both of which are desirable qualities for high-speed optical fibre links. Directly modulated laser diodes are operated under conditions that can lead to a wide variety of nonlinear behaviour, including period doubling [1] and chaos [2]. It is therefore vital to have an accurate model of nonlinear behaviour that is numerically efficient, and for the most part rate equations are used by most workers. However, in contrast to a large volume of work using rate equations for nonlinear modelling of bulk structures, little such work exists for MQW lasers. This paper presents a detailed theoretical and experimental study of the nonlinear dynamics of an InGaAsP/InGaAs MQW $\lambda/4$ shifted Distributed Feedback (DFB) laser diode with 16 quantum wells [3] under direct modulation.

* Corresponding author.

2. MQW LASER DIODE MODEL

To account for the carrier dynamics in quantum well lasers, a rate equation model based on that proposed by Nagarajan *et al.* [4] is used in our numerical simulations. The rate equations for the carrier density in the quantum wells (N) and the barriers (N_B) and the photon density in the optical cavity (S) can be written as,

$$\frac{dN_B}{dt} = \frac{\Gamma_q I}{eV} - \frac{N_B}{\tau_c} + \frac{\Gamma_q N}{\tau_e} \tag{1}$$

$$\frac{dN}{dt} = \frac{N_B}{\Gamma_q \tau_c} - N\left[\frac{1}{\tau_n} + \frac{1}{\tau_e}\right] - v_g G(N,S)\,S \tag{2}$$

$$\frac{dS}{dt} = \left[\Gamma v_g G(N,S) - \frac{1}{\tau_p}\right] S + \Gamma \beta B N^2 \tag{3}$$

where Γ_q is the fraction of the MQW region filled by the quantum wells, N_B/τ_c is the loss rate of carriers from the barriers to the quantum wells and N/τ_e is the loss rate of carriers from the quantum wells to the barriers. τ_p is the photon lifetime, V is the active region volume, Γ is the mode confinement factor, and v_g is the group velocity. The dependence of the carrier lifetime (τ_n) on N is modelled as,

$$\frac{1}{\tau_n} = A + BN + CN^2 \tag{4}$$

where the terms BN and CN^2 model bimolecular and Auger recombination respectively. The output power per facet P is related to S via,

$$P = \frac{SVh\nu v_g}{2\Gamma L} \ln\left(\frac{1}{R}\right) \tag{5}$$

where R is the facet reflectivity (≈ 0.32) and L is the cavity length. The injection current (I) can be expressed $I = I_{DC} + I_{RF} \sin(2\pi f t)$, where I_{DC} is the bias current, I_{RF} is modulation current amplitude and f is the modulation frequency. Using an isolator between the microwave source and the laser to minimise impedance mismatch problems enables I_{RF} to be approximated, $I_{RF} \approx \sqrt{2P_{RF}/Z_0}|\Gamma_L - 1|$, where P_{RF} is the incident RF input power, Γ_L is the reflection coefficient of the laser and Z_0 is the characteristic impedance. The quantum wells have one conduction subband so that the relationship between the carrier density and the junction voltage V_j can be approximated [5],

$$N \approx kT \frac{m_c}{\pi \hbar^2 L_z} \ln\left[1 + \exp\left(\frac{eV_j - E_{ph}}{kT}\right)\right] \tag{6}$$

where m_c is the electron effective mass, L_z is the quantum well width and E_{ph} is the photon energy. The MQW material gain $G(N,S)$ is modelled by [6]

$$G(N,S) = \frac{G_0}{(1+\varepsilon S)} \ln\frac{N}{N_0} \tag{7}$$

where ε is the nonlinear gain coefficient, N_0 is the transparent carrier density and G_0 is a constant dependant on well structure.

It is known that the fraction of the spontaneous emission coupled into the lasing mode (β factor) plays a large part in determining the nonlinear dynamics of laser diodes [2, 7–9] and it has recently been found that bulk DFB laser diodes can exhibit a reduced β factor [2] compared to their FP counterparts (8×10^{-7} compared to 10^{-5} in typical FP lasers). The value of β used in our simulations was therefore measured using the same technique described in [2]. Other parameters which were not material or dimensional in nature were

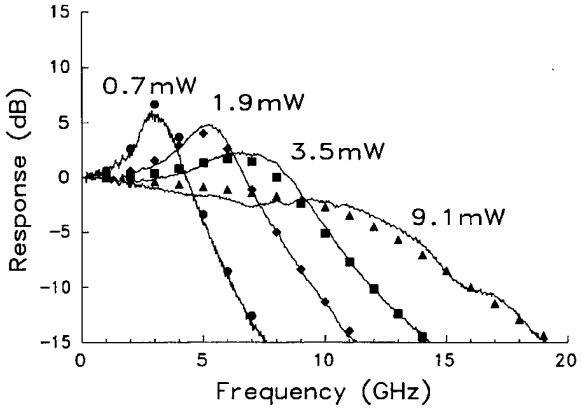


FIGURE 1 Measured (solid lines) and simulated (points) small-signal modulation frequency response of the laser with increasing output power.

extracted from the small-signal measurements shown in Figure 1. The parameter values used in the simulations were: $V = 5.1 \times 10^{-17}\,\mathrm{m}^3, \Gamma = 0.22$, $\Gamma_q = 0.66, \tau_p = 1.3\,\mathrm{ps}, \beta = 10^{-6}, G_0 = 141107\,\mathrm{m}^{-1}$, $N_0 = 2.41 \times 10^{24}\,\mathrm{m}^{-3}, \varepsilon = 3.24 \times 10^{-23}\,\mathrm{m}^3, A = 10^8\,\mathrm{s}^{-1}$, $B = 10^{-16}\,\mathrm{m}^3\mathrm{s}^{-1}$, $C = 3 \times 10^{-41}\,\mathrm{m}^6\mathrm{s}^{-1}$, $\tau_c = 20\,\mathrm{ps}$, and $\tau_e = 191\,\mathrm{ps}$. Simulations were performed by solving (1)–(3) using a fourth order Runge-Kutta scheme.

3. NONLINEAR DYNAMIC ANALYSIS

3.1. Results

So as to study the nonlinear dynamics resulting from direct modulation of the laser alone, optical feedback into the laser was minimised by the use of an optical isolator. The experimental setup used to analyse the nonlinear dynamics of the lasers is shown in Figure 2. The incident RF input power to the laser (P_{RF}) was monitored through a directional coupler and the problem of laser mismatch to 50 Ω was minimised through the use of a microwave isolator.

For a range of output power levels* (P_{out}) and modulation frequencies (f), P_{RF} was swept between −5 and 22 dBm and the regions where nonlinear behaviour occurred recorded. These regions are shown in Figure 3 along with the simulated results. There were both regions of period doubling and period tripling. Period tripling occurred when the laser was biased with

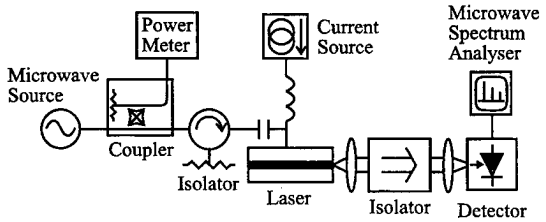
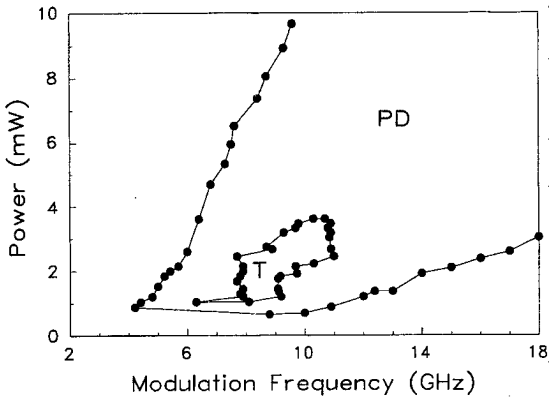


FIGURE 2 Experimental setup used to investigate the nonlinear dynamics of the laser.



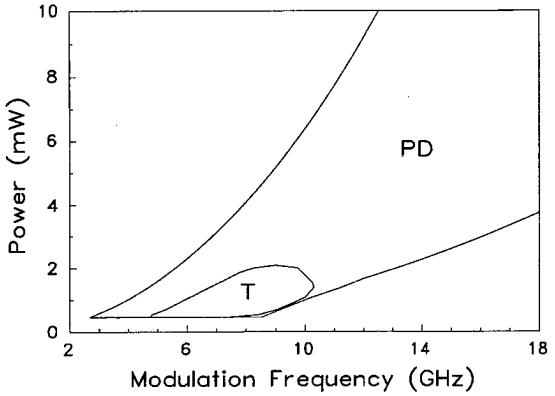


FIGURE 3 Regions of where nonlinear behaviour were recorded for (a) the measured laser, (b) the simulated laser. PD denotes the regions where period doubling were recorded and T the region where period tripling was observed.

P_{out} levels of between 0.7 mW and 3.6 mW. In the region where period tripling occurred, as P_{RF} was increased, the route to period tripling was always via period doubling. Figure 4 shows the measured and simulated output frequency spectrum of the laser under conditions of single period, period doubling and period tripling behaviour.

There were features of the period doubling which were common to both simulated and measured lasers. Namely the upper and lower modulation frequency limits of the period doubling regions seemed to follow one and two times

* Corresponding to a range of bias currents.

the relaxation frequency of the laser (f_R) respectively. For a fixed ratio of modulation frequency to relaxation frequency the value of P_{RF} required to produce period doubling increased almost linearly with P_{out}. These features are illustrated in Figure 5 where the calculated values of P_{RF} required for period doubling to occur in the laser are displayed versus P_{out} and the modulation frequency normalised to f_R. The measured and simulated values of P_{RF} required for period doubling to occur in the laser versus P_{out} are displayed in Figure 6. When obtaining the data displayed in Figure 6 the modulation frequency was adjusted at each P_{out} to keep the ratio of modulation frequency to relaxation frequency constant, which for Figure 6 was 1.5. It is expected that this relationship is due to the corresponding increase in relaxation damping

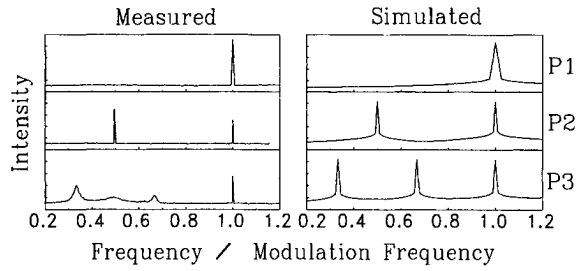


FIGURE 4 Measured and simulated frequency spectrum of the laser under conditions of single period (P1), period doubling (P2) and period tripling (P3) behaviour.

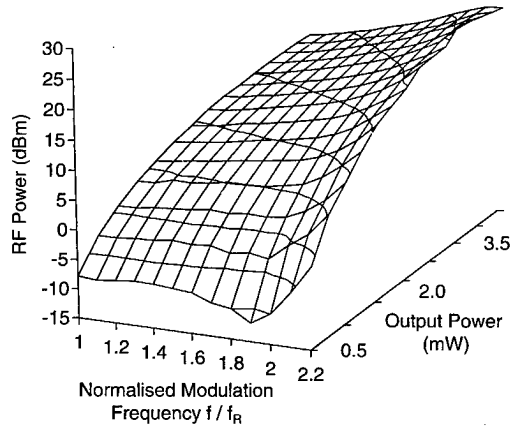


FIGURE 5 Three dimensional plot displaying the calculated values of P_{RF} required for period doubling to occur in the laser versus P_{out} and the modulation frequency normalised to f_R.

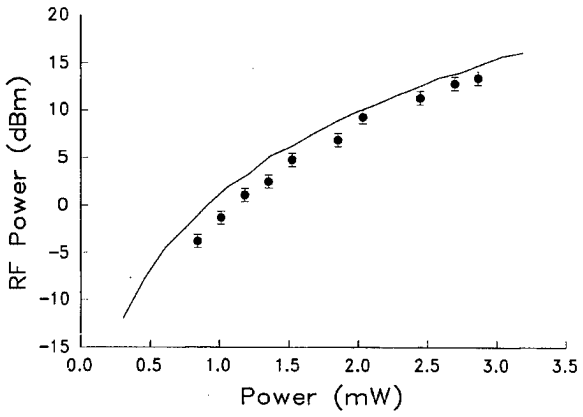


FIGURE 6 Measured (points) and simulated (solid line) RF input power required to produce period doubling versus output power per facet. The modulation frequency to relaxation frequency ratio is kept constant at 1.5.

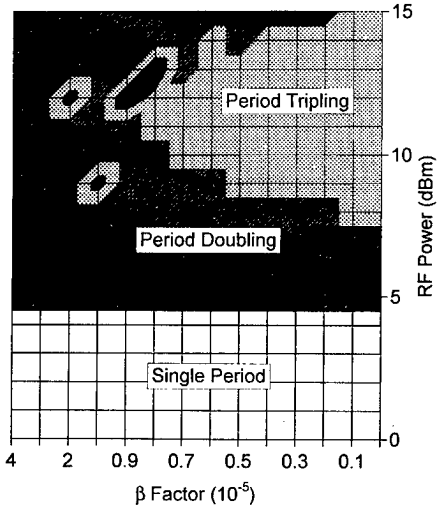


FIGURE 7 Simulated dependence of nonlinear behaviour on P_{RF} and β in the laser. Simulations were performed with a modulation frequency of 6 GHz and P_{out} set at 0.6 mW. The black areas in the diagram indicate regions of higher order bifurcations.

with power level which is known to be an important factor in characterising the nonlinear dynamics of laser diodes [9].

Liu *et al.* [2] suggest that the promotion of period tripling in their bulk DFB laser was due to the unusually low value of β measured in their laser. To investigate whether this was the case in our MQW DFB laser we analysed the effect of variations in β

on our simulation results. Figure 7 shows the regions of nonlinear behaviour as β and P_{RF} are varied. It is clear that as β is reduced the period tripling behaviour becomes more prominent. It would therefore seem likely that the observed period tripling in our DFB laser can be attributed to the low value of β measured in this laser compared to that measured in typical Fabry Perot lasers.

4. CONCLUSION

In summary, we have performed a detailed theoretical (using rate equations) and experimental analysis of the nonlinear dynamics of a MQW DFB laser diode under direct modulation. We observed period doubling and period tripling in both measurements and simulations. Period doubling was evident over a wide range of modulation frequencies. Computational results using rate equations show good agreement with the experimental results.

Acknowledgements

The authors would like to acknowledge Dr. L. D. Westbrook and Dr. I. F. Lealman of British Telecom Research Laboratories for informative discussions and the supply of sample lasers.

References

[1] Hemery, E., Chusseau, L. and Lourtioz, J. M. "Dynamic behaviors of semiconductor lasers under strong sinusoidal current modulation: Modeling and experiments at 1.3 μm", *IEEE J. Quantum Electron.*, **26**, 633–641, April 1990.

[2] Liu, H. F. and Ngai, W. F. "Nonlinear dynamics of a directly modulated 1.55 μm InGaAsP distributed feedback semiconductor laser", *IEEE J. Quantum Electron.*, **29**, 1668–1675, June 1993.

[3] Lealman, I. F., Harlow, W. F. and Perrin, S. D. "Effects of Zn doping on modulation bandwidth of 1.55 μm InGaAs/InGaAsP multiquantum well DFB lasers", *Electron. Lett.*, **29**, 1197–1198, June 1993.

[4] Nagarajan, R., Ishikawa, M., Fukushima, T., Geels, R. S. and Bowers, J. E. "High speed quantum-well lasers and carrier transport effects", *IEEE J. Quantum Electron.*, **28**, 1990–2007, Oct. 1992.

[5] Seltzer, C. P., Westbrook, L. D. and Wickes, H. J. (1995). "The 'Gain-Lever' effect in InGaAsP/InP multiple quantum well lasers", *IEEE J. Lightwave Technol.*, **13**(2), 283–289.

[6] McIlroy, P. W. A., Kurobe, A. and Uematsu. "Analysis and application of theoretical gain curves to the design of multi-quantum-well lasers", *IEEE J. Quantum Electron.*, **21**, 1958–1963, Dec. 1985.

[7] Lee, C. H., Yoon, T. H. and Shin, S. Y. "Period doubling and chaos in a directly modulated laser diode", *Appl. Phys. Lett.*, **46**, 95–97, Jan. 1985.

[8] Hori, Y., Serizawa, H. and Sato, H. "Chaos in a directly modulated semiconductor laser", *J. Opt. Soc. Amer. B.*, **5**, 1128–1133, May 1988.

[9] Yoon, T. H., Lee, C. H. and Shin, S. Y. "Perturbation analysis of bistability and period doubling bifurcations in directly-modulated laser diodes", *IEEE J. Quantum Electron.*, **25**, 1993–2000, Sept. 1989.

Authors' Biographies

Stephen Bennett was born in Blackpool, England, in 1971. He received the B.Eng. (Honours) degree in electronic and electrical engineering from the University of Leeds in 1993. After graduating he joined the Microwave and Terahertz Technology Group at the University of Leeds where he has been working towards the Ph.D. degree on the high frequency nonlinear dynamics of semiconductor lasers. Since 1996 he has worked as a Research Fellow within the Microwave Optoelectronics group at University College London. His research interests include semiconductor laser dynamics, soliton control, optical comb generation and MQW saturable absorption.

Christopher M. Snowden received the B.Sc., M.Sc. and Ph.D. degrees from the University of Leeds. After graduating in 1977 he worked as an Applications Engineer for Mullard, Mitcham. His Ph.D. studies were conducted in association with Racal-MESL and were concerned with the large-signal characterisation of MESFET microwave oscillators. In 1982 he was appointed Lecturer in the Department of Electronics at the University of York. He joined the Microwave Solid State Group in the Department of Electrical and Electronic Engineering at the University of Leeds in 1983. He now holds the Chair of Microwave Engineering in the Microwave and Terahertz Technology Research Group and is also currently Head of the

Department of Electronic and Electrical Engineering. During 1987 he was a Visiting Research Associate at the California Institute of Technology. He has been a Consultant to M/A-COM Inc., Corporate Research and Development since 1989, where he was on sabbatical leave during the period 1990-91. During this year he represented M/A-COM as Senior Staff Scientist. He was Chairman of the 1995 international Microwaves and RF Conference. He is a Member of the MIT Electromagnetics Academy. He is also a Top Scientist at the International Research Centre for Telecommunications-Transmission and Radar, Delft University of Technology, Netherlands.

Professor Snowden is a Fellow of the IEEE and a Fellow of the IEE. He is a Distinguished Lecturer (1996/7) for the IEEE (Electron Devices Society). He is co-Chairman of the MTT-1 Committee and a Member of the 1997 IEEE MTT-S Technical Program Committee. His main research interests include compound semiconductor device modelling, microwave, terahertz and optical nonlinear subsystem design and advanced semiconductor devices. He has written 7 books and over 190 papers.

Stavros Iezekiel was born in Coventry, England, in 1966. He received the B.Eng. and Ph.D. degrees in electronic and electrical engineering from the University of Leeds in 1987 and 1991, respectively. From 1991 to 1993, he worked as a Research Fellow developing hybrid optoelectronic packaging systems for M/A-Com Corporate Research and Development Centre. In 1993, he was appointed as a Lecturer in high-frequency analogue electronics at the University of Leeds. His research interests include high-speed semiconductor laser diode modelling, optoelectronic packaging, nonlinear systems, and microwave-optoelectronic subsystem design.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 361–365
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Temperature Dependence of the Electron and Hole Scattering Mechanisms in Silicon Analyzed through a Full-Band, Spherical-Harmonics Solution of the BTE

SUSANNA REGGIANI, MARIA CRISTINA VECCHI and MASSIMO RUDAN*

Dipartimento di Elettronica, Informatica e Sistemistica, Università di Bologna, Viale Risorgimento 2, 40136 Bologna, Italy

By adopting the solution method for the BTE based on the spherical-harmonics expansion (SHE) [1], and using the full-band structure for both the electron and valence band of silicon [2], the temperature dependence of a number of scattering mechanisms has been modeled and implemented into the code HARM performing the SHE solution. Comparisons with the experimental mobility data show agreement over a wide range of temperatures. The analysis points out a number of factors from which the difficulties encountered in earlier investigations seemingly originate, particularly in the case of hole mobility.

Keywords: Carrier mobility, scattering mechanisms, Boltzmann equation, spherical-harmonics expansion

1 INTRODUCTION

We present here an investigation of the temperature dependence of the scattering mechanisms described at microscopic level through the solution of the BTE based on an expansion in spherical harmonics. Such investigation is also relevant from the application viewpoint because it allows for a detailed analysis of the temperature and impurity dependence of mobility, hence of the performance of the submicron and power semiconductor devices. The mechanisms considered here are acoustical- and optical-phonon scattering, ionized-impurity scattering, and impact ionization. In particular, phonon and impurity scattering have a strong influence on mobility also at low energies. In addition, the effect of incomplete ionization of dopant has been introduced as well, to account for the behavior of carrier mobility at high impurity concentration and low temperatures.

By this approach the experimental mobility curves of electrons and holes are well reproduced in a wider range of temperatures and impurity concentrations than in earlier investigations.

* Corresponding author. Tel.: +39 (51) 64430-16, Fax: 64430-73. e-mail: mrudan@deis.unibo.it.

2 PHYSICAL MODELS

The mathematical model resulting from the SHE of the BTE is briefly summarized here, in the homogeneous case, with reference to both the conduction and valence band:

$$-q^2F^2\frac{\partial}{\partial E}\left[\tau g(E)u_g^2(E)\frac{\partial f_0}{\partial E}\right] = \quad (1)$$

$$+3c_{op}g(E)\mathcal{M}_{op} - 3c_{ii}\,g^2(E)f_0(E)$$

$$+\,3g(E)\int A(E',E)f_0(E')g(E')dE',$$

where $\mathcal{M}_{op}=(N_{op}^+f_0^+-N_{op}f_0)g^+-(N_{op}^+f_0-N_{op}f_0^-)g^-$. The symbols have the following meaning: $g(E)$ is the density of states, $u_g(E)$ the modulus of the group velocity, τ the total scattering rate, F the electric field, c_{op} a constant proportional to the optical-phonon coupling constant, N_{op} the optical-phonon occupation number, $N_{op}^+=N_{op}+1$, $g^{\pm}(E)=g(E\pm\hbar\omega_{op})$, where $\hbar\omega_{op}$ is the optical-phonon energy, and similarly for $f_0{}^{\pm}(E)$. Impact ionization is also considered: $c_{ii}\,g(E)$ is the total impact-ionization scattering rate and $A(E',E)$ is a suitable kernel [3]. All the scattering mechanisms considered here exhibit a temperature dependence:

- acoustical-phonon scattering shows an explicit dependence on temperature [4];
- optical-phonon scattering depend on temperature through the occupation number N_{op} [4];
- ionized-impurity scattering depend on temperature through the partial ionization effect;
- impact ionization depend on temperature through the energy gap.

In the following we will focus on the effect of temperature on the carrier mobility as a function of impurity concentration, therefore we will describe with more details the scattering term strongly related to this kind of collision.

The ionized-impurity scattering is described according to the Brooks-Herring (BH) model:

$$S_{BH}=\frac{N_ie^2I^2}{(2\pi)^2\hbar}\left[\frac{Ze}{\varepsilon_s(2k^2(1-\cos\xi)+\beta^2)}\right]^2 \times\delta(E'-E) \quad (2)$$

where N_i is the ionized-impurity concentration, I is the overlap factor, Ze the impurity charge, ε_s the semiconductor dielectric constant, k the wave vector magnitude, ξ the scattering angle, β the inverse screening length.

The overlap factor accounts for the interaction between the initial and final states in a collision event and depends on the shape of the band. According to [4], it is considered constant for the conduction band, incorporating its value in the coupling constant of each of the scattering mechanisms. For the valence band the expression obtained by Wiley [4] has been adopted:

$$I^2(k,k')=\tfrac{1}{4}(1+3\cos^2\xi),$$

which produces an additional angular dependence in the spherical harmonics expansion of the impurity scattering term, while, for the collisions with phonons the expansion of the scattering terms turns out to be unmodified, apart from a constant factor.

The inverse screening length is usually given by the Debye expression which is obtained by solving Poisson's equation considering an isotropic parabolic band and Boltzmann statistics. In this work we consider a more complicated description of the band structure, based on pseudo-potential calculation, and therefore we compute the inverse screening length following the general definition [5]:

$$\beta^2=\frac{e^2}{\varepsilon_sK_BT_0}\int g(E)f_0(E)(1-f_0(E))dE.$$

where e electron charge, K_B Boltzmann constant, T_0 lattice temperature, $g(E)$ density of states, $f_0(E)$ first term of SHE of distribution function. It is necessary to include additional corrections to the Brooks-Herring (2) scattering term to take into account the effects of the Born approximation of order higher than 1 for incoherent collision with a

single impurity and for coherent collisions with impurity-pairs. The theory of multiple collisions of [6], in agreement with the Brooks-Herring formulation, considers an electron gas interacting with N_i impurities in the crystal. The main difference between the results of [6] and that of Brooks-Herring is that the latter considers the potential of a single impurity, resulting in a total collision probability equal to N_i times the collision probability of a single impurity. The result of [6], instead, keeps the information on the interaction potential between electrons and impurities, which is expressed by $\sum_{j=1}^{N} V(\mathbf{r} - \mathbf{R}_j)$, with $\mathbf{R}_j$ position vector for the j-th ion, and modifies the scattering matrix in such a way that it becomes:

$$S(E, E') = S_{BH}(E, E')[1 + \delta_M + \delta_D]$$

where δ_M, δ_D are multiple collisions and impurity dressing constants.

The effect of impurity clustering, relevant for high doping concentrations, has also been taken into account. At high doping densities it is described as if the carrier scatters with a cluster of Z ions, with Z number of charge. The partial ionization, relevant for both high doping concentration and low temperature, has been modeled redefining the donor and acceptor concentrations according to the following expressions:

$$N_D^+ = \frac{N_D}{1 + 2\exp[0.045/(K_B T)]}$$

for electrons; and

$$N_A^- = \frac{N_A}{1 + 4\exp[0.0438/(K_B T)]}$$

for holes, along with the description of the equilibrium carrier concentration in terms of Fermi statistics.

3 RESULTS

The model described in the previous section has been applied to the investigation of electron and hole mobility as a function of impurity concentration and temperature.

The experiments show that the dependence of the electron mobility on temperature in intrinsic silicon differs from that of holes. In particular, at low temperatures ($T < 100$ K) it is $\mu_n \approx T^{-1.5}$, $\mu_p \approx T^{-2}$. On the other hand, at higher temperatures (up to 500 K) several investigators agree on that the mobility of both species of carriers is well described by a relation of the form $\mu_{n,p} = \mu_{max}(T/300)^{-\gamma}$, with $2.2 < \gamma < 2.3$. The relative departure of the mobility laws in terms of the lattice temperature was initially ascribed to the different structure of the conduction and valence bands. However, after implementing the full-band structure in HARM, a quantitative agreement was found in a wide range of temperatures for the electron mobility only; on the contrary, the agreement of the hole mobility was acceptable at 300 K, whereas a difference of about 20% was found at 200 and 400 K. This behavior can be observed in Figures 1 and 2, where the intrinsic

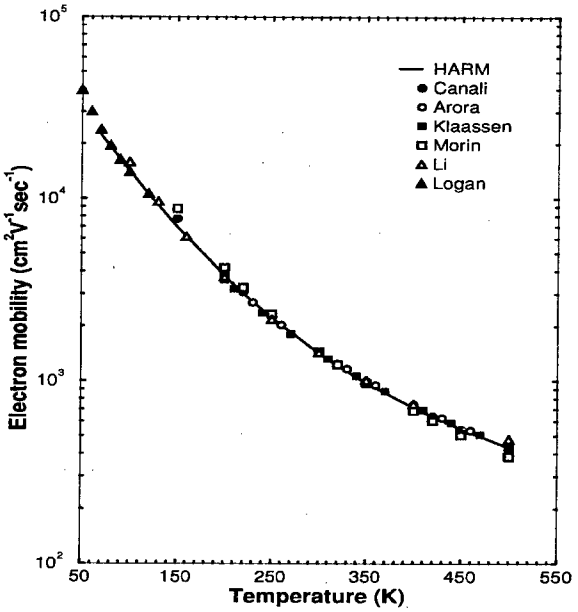


FIGURE 1 Experimental intrinsic electron mobility as a function of temperature compared with the results obtained with the HARM code.

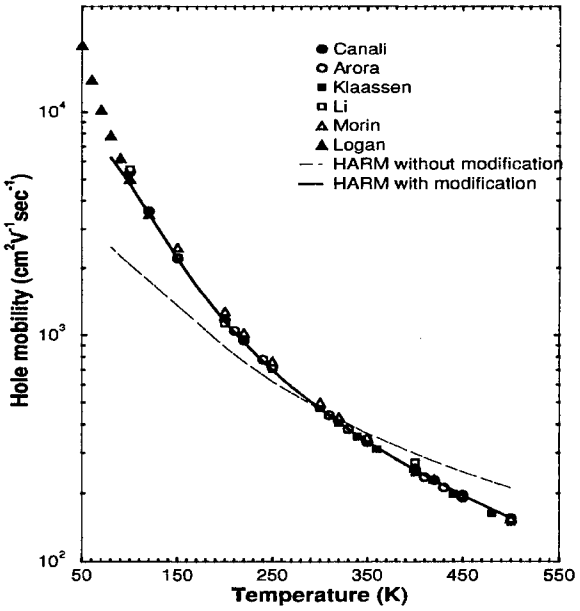


FIGURE 2 Experimental intrinsic hole mobility as a function of temperature compared with the results obtained with the HARM code. For the latter two different results are provided: the one labeled as "HARM (without modification)" refers to an acoustical-phonon scattering model with no dependence of the coupling constants on temperature; this dependence has been added in the results labeled as "HARM (with modification)".

mobility for electrons and holes as a function of temperature is reported. The curve for holes which corresponds to the physical model illustrated so far is labeled "HARM (without modification)".

Another effect whose temperature dependence proved relevant is the scattering with ionized impurities. To correctly describe such mechanism it is necessary to take the incomplete ionization into account, which becomes relevant when the total impurity concentration exceeds 10^{17} cm^{-3} and produces a significant lowering in the majority-carrier concentration with respect to that of the impurities. As already mentioned, incomplete ionization has been implemented into HARM, along with the description of the equilibrium carrier concentration in terms of Fermi statistics.

A final investigation is related to the phonon scattering, whose effect is modeled here following [4]. The values of the acoustical- and optical-phonon coupling constants were preliminarily fitted on the experimental data at 300 K. Then, a model for the temperature dependence of the

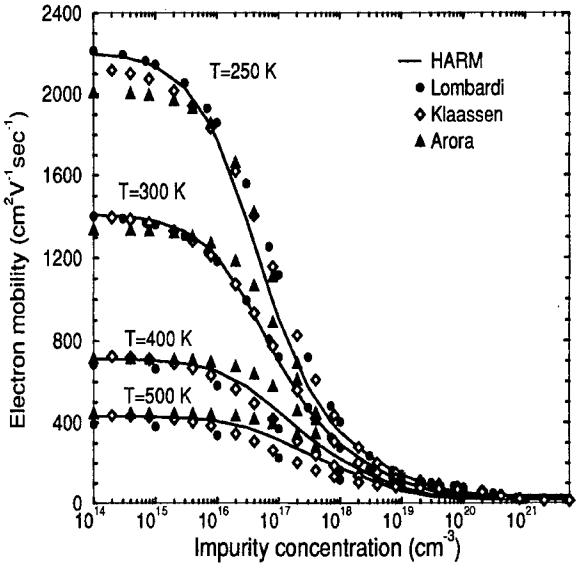


FIGURE 3 Experimental electron mobility as a function of impurity concentration at different temperatures compared with the results obtained with the HARM code.

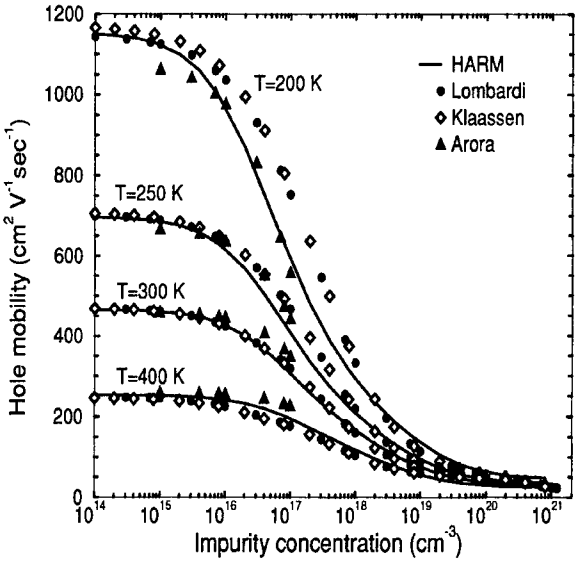


FIGURE 4 Experimental hole mobility as a function of impurity concentration at different temperatures compared with the results obtained with the HARM code.

acoustical-phonon coupling constant has been implemented into HARM and the values of its parameters have been derived from the experiments. The result obtained with this additional dependence on temperature of the acoustical-phonon coupling constant is indicated in Figure 2 with the label "HARM (with modification)".

A number of runs have been carried out to test the model described above in a typical range of temperatures as a function of impurity concentration. The comparison with experimentally-fitted curves of electron and hole mobility reported by different investigators is shown in Figures 3 and 4: in both cases a satisfactory agreement with experimental data has been obtained.

References

[1] Gnudi, A., Ventura, D., Baccarani, G. and Odeh, F. (1993). Two-dimensional MOSFET Simulation by Means of a Multidimensional Spherical Harmonics Expansion of the Boltzmann Transport Equation. *Solid-State Electronics*, **36**(4), 575–581.

[2] Vecchi, M. C., Ventura, D., Gnudi, A. and Baccarani, G. (1994). Incorporating Full Band-Structure Effects in the Spherical-Harmonics Expansion of the Boltzmann Transport Equation. In H. S. Bennet and M. E. Law, editors, *Proc. of the NUPAD V Conference*, pp. 55–58, Honolulu, IEEE.

[3] Gnudi, A., Ventura, D. and Baccarani, G. Modeling Impact Ionization in a BJT by Means of a Spherical Harmonics Expansion of the Boltzmann Transport Equation. *IEEE Trans. on CAD of ICAS*, **CAD-12**(11), 1706–1713, November 1993.

[4] Jacoboni, C. and Reggiani, L. (1983). The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials. *Reviews of Modern Physics*, **55**, 645.

[5] Haug, H. and Koch, S. W. (1993). *Quantum Theory of the Optical and Electronic Properties of Semiconductors*. World Scientific, Singapore.

[6] Moore, E. J. (1967). Quantum-transport theories and multiple scattering in doped semiconductors. I. Formal theory. *Physical Review*, **160**(3), 607–617, August.

Authors' Biographies

Susanna Reggiani received the degree in Electrical Engineering from the University of Bologna, Italy, in 1997. Since April 1997 she has been working at the Department of Electronics (DEIS) of the same University in the field of the numerical simulation of semiconductor devices.

Maria Cristina Vecchi received the degree and the Ph.D. in Electrical Engineering from the University of Bologna, Italy, in 1990 and 1995, respectively. Since 1990 she has been working at the Department of Electronics (DEIS) of the same University in the field of the numerical simulation of semiconductor devices. In 1992–93 she visited the IBM T. J. Watson Research Center of Yorktown Heights, NY, working on higher-order transport models in semiconductors. In 1994 she visited the Integrated System Laboratories of the ETH, Zürich, working on the deterministic solution of the Boltzmann transport equation.

Massimo Rudan received a degree in Electrical Engineering in 1973 and a degree in Physics in 1976, both from the University of Bologna, Italy. He joined the Department of Electronics (DEIS) of the University of Bologna in 1975. From 1978 he has been teaching a course of Quantum Electronics in the Faculty of Engineering of the same University. Since 1983 he has been working in a group involved in numerical analysis of semiconductor devices. In 1986 he has been a visiting scientist at the IBM Thomas J. Watson Research Center at Yorktown Heights, NY, studying the discretization techniques for the higher-order moments of the Boltzmann Equation. In 1990, he was appointed Full Professor of Microelectronics at the University of Bologna.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 367–373
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Monte Carlo Simulations of Intersubband Hole Relaxation in a GaAs/AlAs Quantum Well

R. W. KELSALL

Institute of Microwaves and Photonics, School of Electronic and Electrical Engineering, University of Leeds, Leeds LS2 9JT, UK

An ensemble Monte Carlo code has been developed for the simulation of hole relaxation processes in a GaAs/AlAs quantum well. The code includes a realistic **k.p** model of the valence subbands and corresponding wavefunctions. Intra- and inter-subband phonon scattering rates are calculated for polar and non-polar interactions via both optical and acoustic modes. The code is used to simulate the cooling of non-equilibrium photogenerated hole populations. A lifetime of 90 fs is extracted for optical phonon mediated depopulation of the 4th subband at 77 K. By contrast, the 2nd subband exhibits fast re-population, but slow de-population, with extracted lifetimes of up to 160 ps. The slow depopulation is attributed to the small energy separation of the 1st and 2nd subbands (less than the optical phonon energy) and the large density of states in the 2nd subband off-zone-centre minimum.

Keywords: Monte Carlo, intersubband, quantum well

1. INTRODUCTION

Interest in intersubband transitions has been given new impetus by recent developments in "Quantum Cascade" or intersubband lasers for Far-Infra-Red (FIR) operation [1]. Whilst the Quantum Cascade devices depend on electron intersubband transitions, *p*-type quantum well systems offer several advantages – high density of states, wide tunability and the possibility of normal incidence operation – and FIR emission has already been demonstrated in *p*-type germanium devices [2].

Design of FIR emitting devices requires detailed knowledge of the electronic bandstructure and transient carrier dynamics. The Monte Carlo method is an ideal simulation tool for these structures, since the time-dependent non-equilibrium carrier distribution function is intrinsically determined without recourse to any approximation, enabling accurate determination of intersubband lifetimes.

In this paper, an ensemble Monte Carlo code is used to simulate transient hole relaxation and energy loss phenomena in a GaAs/AlAs quantum

well. The quantum confined valence bandstructure is calculated using an 8-band **k.p** scheme, to account for heavy-light hole mixing effects. The **k.p** wavefunctions are used to calculate hole-phonon scattering rates for the acoustic deformation potential, polar-optical, non-polar optical and piezoelectric interactions, via Fermi's Golden Rule. A total of 192 distinct intra- and inter-subband scattering processes are included in the simulation. Degeneracy (Pauli exclusion) effects are also included. The **k**-space trajectories of typically 100,000 particles are simulated, and cooling from a range of non-equilibrium carrier distributions is investigated.

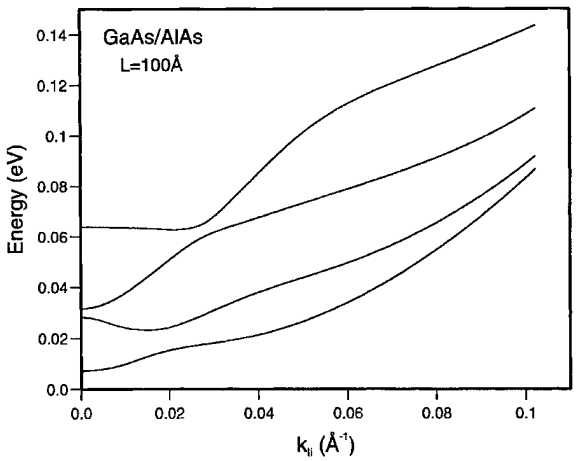


FIGURE 1 Energy vs. in-plane wavevector for the first four valence subbands in a 100 Å GaAs/AlAs quantum well.

2. BANDSTRUCTURE AND SCATTERING RATES

Figure 1 shows the in-plane hole energy dispersion for the lowest four subbands of a 100 Å GaAs quantum well. The bandstructure is calculated in the axial approximation, in which **k**-space anisotropy is not resolved [3]. In practice, anisotropy in quantum wells is not as significant as that in the bulk GaAs valence bands. Figure 1 clearly shows the familiar distortion of the dispersion curves due to heavy-light hole mixing, including the presence of negative effective mass regions in the second and fourth subbands, and corresponding off-zone-centre minima. These minima give rise to Van-Hove singularities in the density of states, which result in strongly localised peaks in the scattering rates. When calculating the rates, we include a Lorentzian broadening function to model lifetime broadening effects which will 'smear out' these singularities [4]. The **k.p** calculation yields pairs of degnenerate eigenstates which have mixed spin parities. In this simulation, the spin states of individual carriers are not recorded, and the total rate for a given transition is obtained by summing the scattering rates into the two spin-degenerate states. The phonon scattering rates were calculated via Fermi's Golden Rule, assuming bulk-like phonon modes throughout. Figures 2 and 3 show

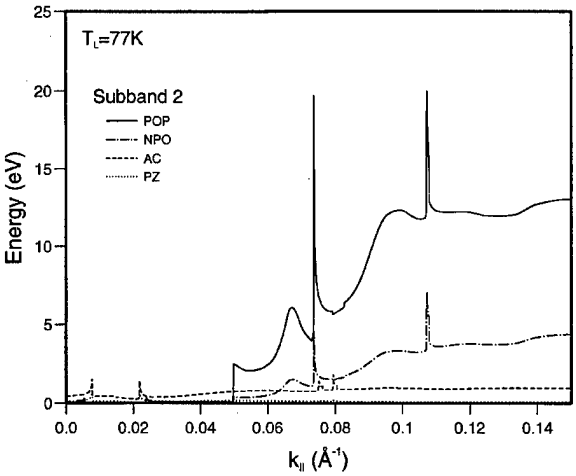


FIGURE 2 Phonon scattering rates for a hole in the 2nd subband, as a function of in-plane wavevector, for a lattice temperature of $T_L = 77$ K, and a total hole density of $p = 10^{11}$ cm^{-2}. The graph shows contributions due to scattering via polar optical (POP), non-polar optical (NPO), acoustic deformation potential (AC) and piezoelectric (PZ) processes.

the total scattering rates for a hole resident in the 2nd and 4th subband, respectively, for a lattice temperature of 77 K. In both cases, the dominant scattering processes are those due to polar optical phonons (Frölich interaction). However, for a carrier in the 2nd subband, optical phonon emission is only possible above a certain energy (wavevector) threshold, whereas in the 4th sub-

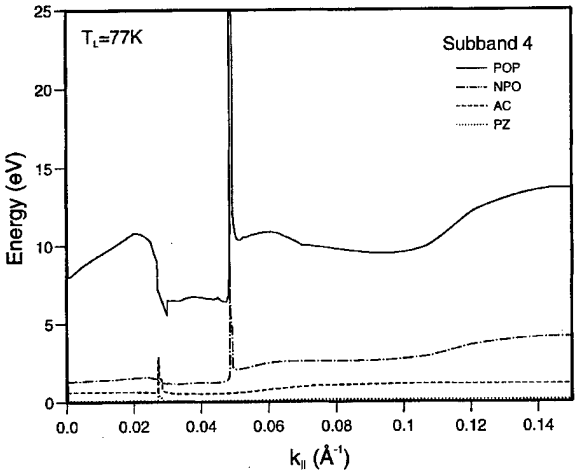


FIGURE 3 Phonon scattering rates for a hole in the 4th subband: details as for Figure 2.

band, the 4–1 and 4–2 optical phonon emission processes are allowed at zero wavevector. Note that, in the valence bands, the deformation potential (non-polar) optical phonon interaction is not forbidden (as it is the conduction bands at the Γ point), although the scattering rates are smaller than those for the polar interaction. Acoustic phonon scattering (modelled here by a single rate to represent the combined effects of the LA and TA modes) is only important at low carrier energies, and the piezoelectric interaction is weaker still, even at 77 K. However, in our zero-field carrier cooling simulations, it is essential to include the acoustic phonon interactions, since they represent the principal energy relaxation processes for low energy carriers.

3. HOLE RELAXATION SIMULATIONS

The Monte Carlo model includes the four lowest energy subbands in the valence band quantum well of GaAs/AlAs. The subband energy dispersions, hole group velocities, and densities of states were tabulated as functions of in-plane wavevector. Similarly, the matrix elements for all permutations of intra- and inter-subband scattering via the relevant phonon modes were precalculated and supplied to the simulation as data. For the polar optical and piezoelectric processes, the matrix elements vary with the in-plane phonon wavevector ($q_\parallel$), and hence with scattering angle. For polar optical scattering, we approximated this dependence by a $1/q_\parallel^2$ variation for all cases. Similar analytic functions were derived to describe the angular dependence of piezoelectric scattering [5]. When calculating the new carrier states after an acoustic or piezoeletric phonon scattering event, the phonon energy E_{AC} was calculated using a linear phonon dispersion curve. For piezoelectric processes, a self consistent scheme was implemented to ensure that the choice of $q_\parallel$ and the corresponding E_{AC} satisfied wavevector and energy conservation respectively, as well as following the appropriate probability distribution for $q_\parallel$. The optical phonon energy was assumed to be independent of wavevector.

The well was taken to be nominally undoped: no impurity scattering was included in the simulation. The results shown here are for simulations of an ensemble of 100,000 particles, whose energies, in-plane wavevectors, and subband indices are monitored at each timestep. Degeneracy (Pauli exclusion) was included by compiling the hole distribution function in each subband at each timestep, and accounting for band-filling via a rejection technique. Since the initial energy distributions were highly non-equilibrium, a short timestep (1 fs) was chosen to ensure that band-filling was not underestimated due to undersampling of the distribution. The simulations were carried out under zero field conditions: for this special case, the probability function which describes the distribution of free flight times can be inverted directly, without recourse to the self-scattering technique. This greatly increases the efficiency of the simulation.

As yet, there is very little experimental data on hole relaxation. Therefore, the aim of this initial work was to obtain information on the hole subband lifetimes and the many-subband cooling process, rather than attempting to simulate photo-excitation experiments. Such a direct comparison

would also require inclusion of carrier–carrier interactions, which are not yet present in the simulation. In the following computer experiments, we assume that a highly non-equilibrium hole population has been photogenerated in an upper subband. The initial energy distribution is highly peaked at the assumed pump energy, with a narrow Gaussian distribution of carrier wavevectors to model Pauli exclusion within the subband. Two cases are presented below: firstly, photogeneration in the 4th subband, and secondly, into the 2nd subband; in both cases the pump energy corresponds to the subband energy minimum. The generated carrier density in both cases was 10^{11} cm^{-2}.

Figure 4 shows the time dependence of the subband populations following carrier photogeneration in subband 4. As expected, subband 4 depopulates within 1 ps, due to fast polar optical phonon scattering into subbands 1 and 2. Note that very few carriers are scattered into subband 3, since the energy separation between subbands 4 and 3 is less than the optical phonon energy. For $t < 1$ ps, subbands 1 and 2 are populated at approximately the same rate; however, for $t > 1$ ps, carriers begin to transfer from subband 2 to subband 1. This transfer takes over 100 ps to complete, since the subband edge separation here, too, is less than the optical phonon energy, and most of the transitions are mediated via acoustic phonon emission (although there is some optical phonon emission from carriers located higher up the 2nd subband). One of the striking features about the population dynamics in this case is that the 2nd subband population rises to approximately 5 times its equilibrium value: in other words, this subband acts as a metastable carrier reservoir, which populates quickly, but depopulates much more slowly. This behaviour is, in part, due to the relative energy separations between subbands 4 and 2, and 2 and 1, but also due to the large density of states at the 2nd subband (off-zone-centre) energy minimum.

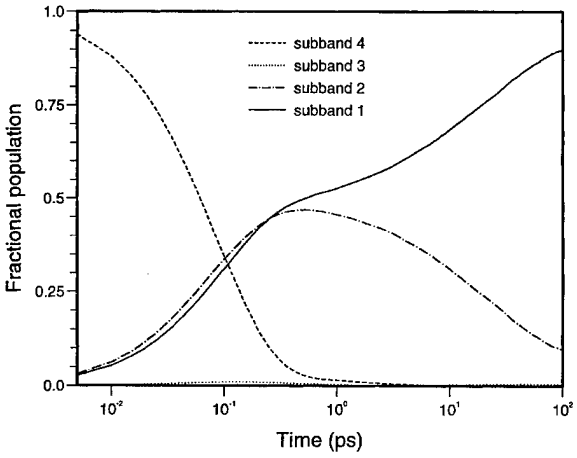


FIGURE 4 Simulated time dependence of the subband populations (normalised to unity) for an initial population photogenerated in subband 4: $T_L = 77$ K, $p = 10^{11}$ cm^{-2}.

We can use the simulation data to extract a value for the overall phonon scattering lifetime. Obviously, we already have information on lifetimes for individual phonon scattering processes, but in most photoluminescence experiments these processes cannot be separately resolved, and the extracted lifetime represents the combined effect of all participant scattering processs. Furthermore, the extracted result is modified by the exact non-equilibrium energy distribution of the carriers; a quantity which cannot readily be obtained by analytic methods, but which is inherently included in a Monte Carlo simulation. Not surprisingly, for such a complex system, we find that the lifetime approximation itself only holds within limited time ranges. For the initial cooling phase we estimate an overall depopulation lifetime for the 4th subband of approximately $\tau_4 = 90$ fs. In contrast, in the latter phase of the simulation where carrier transfer from the 2nd to 1st subband occurs, we estimate a an overall depopulation lifetime for the 2nd subband of $\tau_2 = 90$ ps.

Figure 5 shows the energy loss rate per carrier for the same simulation. The rate has also been resolved into components due to intra- and inter-subband optical and acoustic phonon scattering, respectively. As expected, the initial rapid cooling is primarily due to optical phonon emission, but a crossover occurs after approximately 2 ps, where-

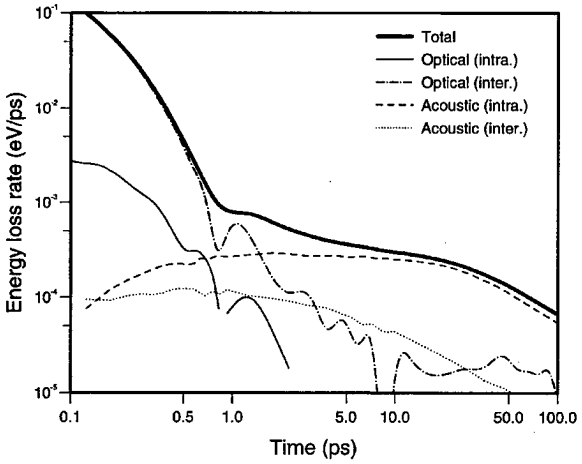


FIGURE 5 Total energy loss rate per hole, and energy loss rates due to intra- and inter subband optical and acoustic phonon processes: simulation conditions as for Figure 4.

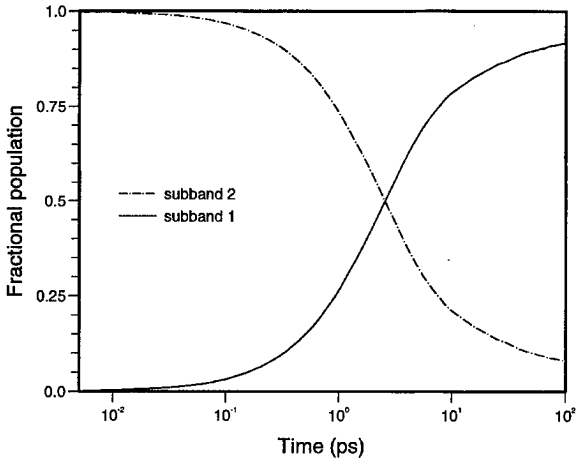


FIGURE 6 Simulated time dependence of the subband populations for an initial population photogenerated in subband 2: $T_L = 77\,\mathrm{K}$, $p = 10^{11}\,\mathrm{cm}^{-2}$.

upon acoustic scattering becomes the dominant energy loss mechanism.

Figure 6 shows the time dependence of the subband populations following carrier photogeneration in subband 2. In this case, only two subbands participate in cooling, and we see a slow transfer of carriers from subband 2 to 1. In the sub-picosecond phase, the transfer is characterised by a lifetime of approximately $\tau_2 = 4\,\mathrm{ps}$, whereas in the $t > 10\,\mathrm{ps}$ phase, the extracted lifetime is $\tau_2 = 160\,\mathrm{ps}$. Figure 7 shows the energy loss rates for this simulation. The total rate is actually negative for $t < 2\,\mathrm{ps}$. Examination of the simulation statistics reveal substantial (intrasubband) absorption of optical phonons by the 2nd subband carriers in the initial phase of the simulation, followed by intersubband optical phonon emission to achieve transfer into subband 1. The figure also shows some initial intersubband transfer by acoustic phonon emission, but this process is slower than the optical phonon absorption, and certainly much less effective in exchanging energy with the lattice. It is this intersubband optical phonon absorption which gives rise to the 4 ps depopulation lifetime – substantially longer than the lifetime in subband 4, but still much shorter than typical acoustic phonon scattering lifetimes. For $t > 2\,\mathrm{ps}$, a net energy loss is observed, and is

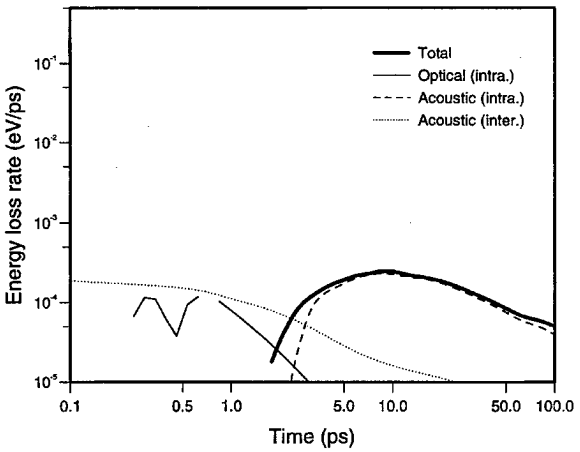


FIGURE 7 Total energy loss rate per hole, and energy loss rates due to intra- and inter subband optical and acoustic phonon processes: simulation conditions as for Figure 6.

dominated by intrasubband acoustic phonon emission, as carriers relax within subband 1. Comparing Figures 5 and 7 it can be seen that, for initial photogeneration in subband 4, some intersubband (2–1) optical phonon emission persists for $t > 10\,\mathrm{ps}$ whereas, for photogeneration in subband 2, this is not the case. This observation explains the difference in τ_2 for the two simulations in the $t > 10\,\mathrm{ps}$ range: for initial photogeneration in subband 4, the average carrier energy is

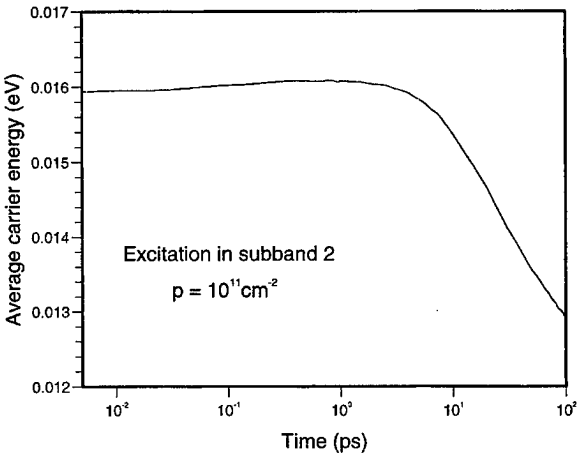


FIGURE 8 Average hole energy vs. time, for an initial population photogenerated in subband 2.

higher, and hence some carriers are still able to emit optical phonons even after more than 10 ps of cooling time; thus τ_2 is enhanced by optical phonon scattering. On the other hand, for photogeneration in subband 2, the average carrier energy is lower, and $\tau_2(t > 10\,\text{ps})$ is entirely acoustic-phonon limited. Figure 8 shows the time dependence of the overall average carrier energy for this simulation. A small increase in average energy is indeed observed throughout the range $t < 5$ ps, with a clear decrease thereafter. Again, this implies that the second subband acts as a bottleneck to the hole cooling process. Clearly, this conclusion may be affected significantly by carrier–carrier scattering, dependent on the extent to which intersubband inter–carrier processes will redistribute both energy and population. Calculations for electrons in quantum wells show that such processes occur over a range of timescales of the order 10^0 ps [6].

4. CONCLUSIONS

An ensemble Monte Carlo code has been developed which includes a realistic description of the band-structure, density of states, and phonon scattering rates for holes in a GaAs/AlAs quantum well. Simulations of intersubband hole relaxation indicate that the 2nd valence subband of the quantum well acts as a bottleneck for hole cooling, due to both the small 2–1 subband energy separation and the large density of states in the off-zone centre minimum. Consequently, a 2nd subband lifetime as long as 160 ps is predicted at 77 K, compared to a lifetime of 90 fs for the 4th subband. The large difference in the re-population and de-population rates of the 2nd subband can result in the accumulation of large excess carrier populations; which implies that Far-Infra-Red photon emission may be attainable between the 1st and 2nd subbands under suitable excitation conditions.

Acknowledgment

This work was initially begun by the author at the University of Durham, UK, with support from the UK Engineering and Physical Sciences Research Council. Advice and guidance from Prof. R. A. Abram at Durham is gratefully acknowledged.

References

[1] Faist, J., Capasso, F., Sivco, D. L., Sirtori, C., Hutchinson, A. L. and Cho, A. Y. (1994). "Quantum cascade laser", *Science*, **264**, 553.
[2] Heiss, W., Unterrainer, K., Gornik, E., Hansen, W. L. and Halter, E. E. (1994). "Influence of impurity absorption on Germanium hot-hole laser spectra", *Semicond. Sci. Technol.*, **5**, SS638.
[3] Eppenga, R., Schuurmans, M. F. H. and Colak, S. "New **k.p** theory for GaAs/AlGAAs type quantum wells", *Phys. Rev. B.*, **36**, 1554.
[4] Kelsall, R. W., Wood, A. C. G. and Abram, R. A. (1991). "Phonon scattering and mobility of holes in a GaAs/AlAs quantum well", *Semicond. Sci. Technol.*, **6**, 841.
[5] Kelsall, R. W. (1989). *Ph.D. Thesis*, University of Durham.
[6] Goodnick, S. M. and Lugli, P. (1988). "Effect of electron-electron scattering on non-equilibrium transport in quantum well system", *Phys. Rev. B.*, **37**, 2578.

Author Biography

Robert W. Kelsall was born in Rotherham, England in 1964. He received the B.Sc. and Ph.D. degrees from the University of Durham,

UK, in 1985 and 1989 respectively; the latter for research in Monte Carlo modelling of electronic transport in GaAs quantum wells.

From 1989 to 1993 he worked as a research assistant, developing self-consistent Monte Carlo simulations of HEMTs and multilayer MOSFETs. He was appointed to a lectureship at the University of Leeds, UK in 1993, and is a member of the newly formed Institute of Microwaves and Photonics at Leeds, where he is working on semiconductor device simulations for hot-carrier degradation studies in MOSFETs, quantum confinement effects in HEMTs, and far-infra-red intersubband lasers.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 375–379
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Bi-Dimensional Simulation of the Simplified Hydrodynamic and Energy-Transport Models for Heterojunction Semiconductors Devices Using Mixed Finite Elements

A. MARROCCO and PH. MONTARNAL*

INRIA, M3N Project, B.P. 105, 78153 Le Chesnay Cedex, France

We study the application of the mixed finite elements method (MFE) for the bi-dimensional simulation of the simplified hydrodynamic and energy-transport models. The two main points are the use of entropic variables which gives a symmetric positive definite problem and a coupled computation of the equations for electrons which requires a generalization of the MFE method for vector valued problems. We give numerical results on JFET and HEMT devices.

Keywords: Semiconductor device simulation, hydrodynamic and energy-transport models, entropic variables, mixed finite elements, GMRES algorithms, HEMT devices

1. INTRODUCTION

The Drift-Diffusion (DD) model, which is currently the most usual model in industrial simulation of electronic devices, is not sufficiently accurate for sub micrometer device modeling. The two main contenders are the Hydrodynamic (HD) [2, 6, 10] and the Energy Transport (ET) [11, 1] models which provides a better description of such effects as velocity overshoot and carrier heating. Both models can be derived from the Boltzmann Transport Equation (BTE) by the moments method. If we remove the convective term from the HD model, we obtain the Simplified Hydrodynamic (SHD) model. Indeed, this convective term can be neglected for some applications, otherwise numerical difficulties may occurs as in fluid dynamics occurs.

The Mixed Finite Elements (MFE) method, which was first developed in the study of structural mechanics, is well adapted to these equations because it gives a good conservation of the currents. For the DD equations this method was used by Brezzi *et al.* [3, 8] with the carrier concentration variables and by Marrocco *et al.* [5, 7] with the quasi-Fermi level variables. This last

*Corresponding author.

approach seems well adapted when considering heterojunctions. Indeed, the quasi-Fermi levels are continuous at the heterojunctions when densities are not.

We consider the bi-dimensional stationary case and take a simplified hydrodynamic model or an energy transport model for the electrons coupled with a drift-diffusion model for the holes through the Poisson equation. We use a Boltzmann statistic. We express the problem in a conservative form. In order to obtain a symmetric positive definite system of elliptic equations, we introduce appropriate variables. We first use the electrostatic potential ϕ, the hole quasi-Fermi level φ_p and for the electrons we introduce entropic variables v_1 and v_2 (functions of the electron quasi-Fermi levels φ_n and the electron temperature T_n, see (9)) which lead to a symmetric formulation. We consider the usual "flow" variables like the electric displacement D, the hole current density J_p, the electron current density J_n and a "flow" function J_w which depends on J_n and on the thermal energy flow density S_n (see (10)). The boundary conditions are non-homogeneous Dirichlet conditions on the contacts and homogeneous Neumann conditions on the insulated boundaries.

We extend the numerical methods developed for the DD model in [5, 7] and for the uni-dimensional ET model in [9]. The solution of the problem is obtained as the limit, when t goes to infinity, of an artificial transient problem discretized with local time steps. We use block relaxation techniques as we consider successively the Poisson equation, the equation for the holes and the two equations for the electrons. The two first equations are solved like in [5, 7]. On the other hand, in order to treat the electronic conservation laws, we extend the MFE method for a vector valued problem. For each case "implicit type" discretization for time and mixed finite elements for space are used. The non-linear problems are solved by Newton-Raphson algorithms.

An outline of the paper is as follows. In section 2 the physical model is presented in his symmetric form. We present the general method of solution in section 3. The extension of the MFE method for vector valued problems is developed in section 4. And finally we present numerical results on JFET and HEMT in section 5.

2. PHYSICAL MODELS

We consider the following problem on $\Omega \subset \mathbb{R}^2$

$$-\operatorname{div} D + q(N - P - dop) = 0, \tag{1}$$

$$-\operatorname{div} J_p - qU = 0, \tag{2}$$

$$-\operatorname{div} J_n + qU = 0, \tag{3}$$

$$-\operatorname{div} J_w + (w_n - q(\phi - \chi))U - N\frac{w_0 - w_n}{\tau_w} = 0, \tag{4}$$

where D, J_p, J_n and J_w are defined by

$$D = \varepsilon \nabla \phi, \tag{5}$$

$$J_p = q\mu_p P \nabla \varphi_p, \tag{6}$$

$$J_n = A_{1,1}\nabla\left(\frac{\varphi_n}{T_n}\right) + A_{1,2}\nabla\left(\frac{-1}{T_n}\right), \tag{7}$$

$$J_w = A_{2,1}\nabla\left(\frac{\varphi_n}{T_n}\right) + A_{2,2}\nabla\left(\frac{-1}{T_n}\right), \tag{8}$$

and the matrix A is defined by

$$A_{1,1} = N\mu_n qT_n,$$

$$A_{1,2} = N\mu_n qT_n\left(\beta\frac{k_B T_n}{q} - \phi + \chi\right),$$

$$A_{2,1} = A_{1,2},$$

$$A_{2,2} = N\mu_n qT_n\left(\left(\beta\frac{k_B T_n}{q} - \phi + \chi\right)^2 + (\beta - c)\left(\frac{k_B T_n}{q}\right)^2\right),$$

with *dop* the doping profile, U the net recombination rate per unit volume, w_n the electron mean energy given by (3/2) $k_B T_n$, w_0 the electron mean energy at equilibrium, τ_w the energy relaxation time (constant), T_L the lattice temperature, ε the

permittivity, μ_p the hole mobility and μ_n the electron mobility given by $\mu_n^0(T_L/T_n)^\alpha$.

Let us remark that the matrix A is symmetric positive definite. This proves the interest of the variables

$$v_1 = \frac{\varphi_n}{T_n} \quad \text{and} \quad v_2 = \frac{-1}{T_n}. \tag{9}$$

The link between J_w and J_n, S_n is the following:

$$J_w = -S_n - (\phi - \chi) J_n. \tag{10}$$

The boundary conditions are non-homogeneous Dirichlet conditions on the ohmic contacts and homogeneous Neumann conditions on the insulated boundaries:

$$\phi = g_\phi, \quad P = g_P, \quad N = g_N, \quad T_n = g_{T_n} \quad \text{on } \Gamma_D,$$

$$\frac{\partial \phi}{\partial n} = \frac{\partial P}{\partial n} = \frac{\partial N}{\partial n} = \frac{\partial T_n}{\partial n} = 0 \quad \text{on } \Gamma_N = \Omega - \Gamma_D.$$

The difference between the models comes from the definition of α, β and c : $\alpha = 1$, $\beta = 5/2$ and $0 \leq c \leq 5/2$ for the SHD model, $\alpha = 1$, $\beta = 3/2$ and $c = 0$ for the Stratton ET model and $\alpha = 1/2$, $\beta = 2$ and $c = 0$ for the Degond *et al.* ET model.

3. NUMERICAL METHODS

The solution of the problem (1)–(8) is obtained as the limit, when t goes to infinity, of the following artificial transient problem

$$s^\phi \partial_t \phi - \operatorname{div} D + q(N - P - dop) = 0, \tag{11}$$

$$s^{\varphi_p} \partial_t \varphi_p - \operatorname{div} J_p - qU = 0, \tag{12}$$

$$S\partial_t \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} - \operatorname{div} \begin{pmatrix} {}^t J_n \\ {}^t J_w \end{pmatrix} + \begin{pmatrix} qU \\ (w_n - q(\phi - \chi))U - n\frac{w_0 - w_n}{\tau_w} \end{pmatrix} = 0, \tag{13}$$

where D, J_p, J_n, J_w are respectively given by Eqs. (5–8).

Our choice of functions s^ϕ, s^{φ_p} and $S_{i,j}$ is motivated by the following considerations: we must keep the homogeneity of the equations, S must be a positive definite matrix and functions s^ϕ, s^{φ_p} must be strictly positive and we want to optimize the convergence. In our computations, we take as in [5, 7]

$$s^\phi = \frac{1}{(\Delta x)^2}\varepsilon, \quad s^{\varphi_p} = \frac{1}{(\Delta x)^2} q\mu_p P,$$

and by natural extension

$$S = \frac{1}{(\Delta x)^2} A.$$

We use block relaxation techniques as we consider successively the Poisson Eq. (11), the equation for the holes (12) and the system of equations for the electrons (13). So we have to solve scalar and vector valued non linear parabolic equations. We use an implicit type discretization for time and mixed finite elements for space. The non-linear problems are solved by Newton-Raphson algorithms.

The two first problems (11, 12) are solved like in [5, 7]. On the other hand, for the computation of the system (13), we extend the MFE method for a vector valued problem. We develop this in the next section.

For the linear systems, direct solvers (Cholesky or LU factorization) are used. The non-symmetric linear system associated with the coupled Eq. (13) can efficiently (memory/time) be solved via GMRES algorithms with block-diagonal pre-conditioning.

4. SOLUTION OF THE TWO ELECTRON EQUATIONS: EXTENSION OF THE MFE METHOD FOR A VECTOR VALUED PROBLEM

At each time step k, after solving the Poisson equation and the hole conservation equation, we

solve the following problem: find v_1^{k+1}, v_2^{k+1}, J_n^{k+1}, J_w^{k+1} solution of the following system:

$$\frac{S(x)}{\Delta t}\begin{pmatrix} v_1{}^{k+1} - v_1{}^{k} \\ v_2{}^{k+1} - v_2{}^{k} \end{pmatrix} - \operatorname{div}\begin{pmatrix} {}^{t}J_n{}^{k+1} \\ {}^{t}J_w{}^{k+1} \end{pmatrix} + G(x, v_1{}^{k+1}, v_2{}^{k+1}) = 0, \tag{14}$$

$$\begin{pmatrix} {}^{t}J_n{}^{k+1} \\ {}^{t}J_w{}^{k+1} \end{pmatrix} = A(x, v_1{}^{k+1}, v_2{}^{k+1})\nabla\begin{pmatrix} v_1{}^{k+1} \\ v_2{}^{k+1} \end{pmatrix}, \tag{15}$$

with boundary conditions

$$\begin{pmatrix} v_1{}^{k+1} \\ v_2{}^{k+1} \end{pmatrix} = \begin{pmatrix} g_{v_1} = \frac{g_{\varphi_n}}{g_{T_n}} \\ g_{v_2} = \frac{-1}{g_{T_n}} \end{pmatrix} \quad \text{on } \Gamma_D$$

$$\text{and} \quad \frac{\partial}{\partial n}\begin{pmatrix} v_1{}^{k+1} \\ v_2{}^{k+1} \end{pmatrix} = 0 \text{ on } \Gamma_N.$$

Let us introduce the following Sobolev spaces

$$H(\operatorname{div}) = \{\omega | \omega \in (L^2(\Omega))^p, \quad \operatorname{div}(\omega) \in L^2(\Omega)\},$$
$$V_0 = \{\omega | \omega \in H(\operatorname{div}), \quad \omega.n = 0 \text{ on } \Gamma_N\}.$$

The variational mixed dual formulation is the following (with $B = A^{-1}$) : find $(v_1^{k+1}, v_2^{k+1}) \in (L^2(\Omega))^2$ and $(J_n^{k+1}, J_w^{k+1}) \in (V_0)^2$ such that

$$\int_\Omega u \cdot \frac{S}{\Delta t}\begin{pmatrix} v_1^{k+1} - v_1^{k} \\ v_2^{k+1} - v_2^{k} \end{pmatrix} dx - \int_\Omega u \cdot \operatorname{div}\begin{pmatrix} {}^{t}J_n^{k+1} \\ {}^{t}J_w^{k+1} \end{pmatrix} dx + \int_\Omega u \cdot G dx = 0 \tag{16}$$
$$\forall u \in (L^2([0, L]))^2,$$

$$\int_\Omega \omega \cdot B\begin{pmatrix} {}^{t}J_n^{k+1} \\ {}^{t}J_w^{k+1} \end{pmatrix} dx + \int_\Omega \operatorname{div}(\omega) \cdot \begin{pmatrix} v_1^{k+1} \\ v_2^{k+1} \end{pmatrix} dx - \int_{\Gamma_D} \begin{pmatrix} g_{v_1} \\ g_{v_2} \end{pmatrix} \omega \cdot n d\Gamma = 0 \quad \forall \omega = \begin{pmatrix} {}^{t}\omega_1 \\ {}^{t}\omega_2 \end{pmatrix} \tag{17}$$
with $\omega_i \in V_0$.

The discrete form of the problem can be directly derived from (16, 17) by using Raviart-Thomas spaces of lowest degree. Thus, we obtain a non-linear problem. We apply a Newton-Raphson iterative technique wich gives at each step l: a linear system for the only variables $J_{n_h}^{l+1}, J_{w_h}^{l+1}$ and a explicit formula for $v_{1_h}^{l+1}, v_{2_h}^{l+1}$ (on each element ε of the mesh):

$$\begin{pmatrix} v_{1\varepsilon}^{l+1} - v_{1\varepsilon}^{l} \\ v_{2\varepsilon}^{l+1} - v_{2\varepsilon}^{l} \end{pmatrix} = -\Lambda_\varepsilon \left[\frac{S_\varepsilon}{\Delta t}\begin{pmatrix} v_{1\varepsilon}^{l} - v_{1\varepsilon}^{k} \\ v_{2\varepsilon}^{l} - v_{2\varepsilon}^{k} \end{pmatrix} + G_\varepsilon(v_{1\varepsilon}^{l}, v_{1\varepsilon}^{l}) - \operatorname{div}\begin{pmatrix} {}^{t}J_{n_h}{}^{l+1} \\ {}^{t}J_{w_h}{}^{l+1} \end{pmatrix} \right],$$

where Λ_ε depends on v_1^l, v_2^l.

5. NUMERICAL RESULTS

We have validated our method either with uni-dimensional devices such that balistic diodes (with or without heterojunctions) or $p-n$ diodes and bi-

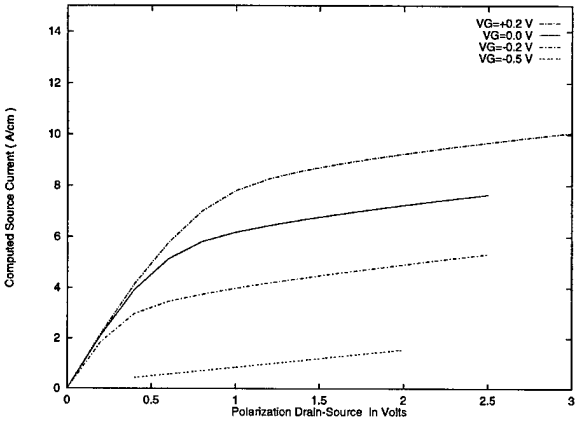


FIGURE 1 Current-Voltage characteristic.

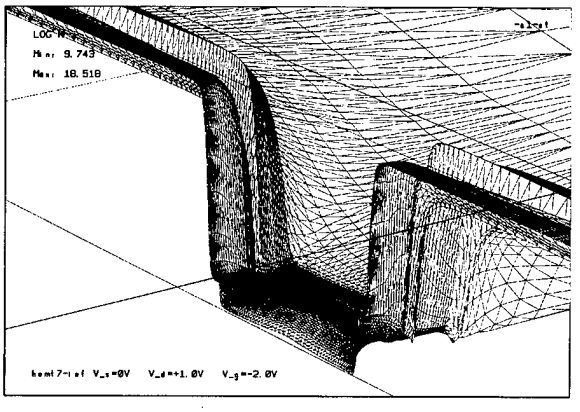


FIGURE 2 Log of the electron concentration.

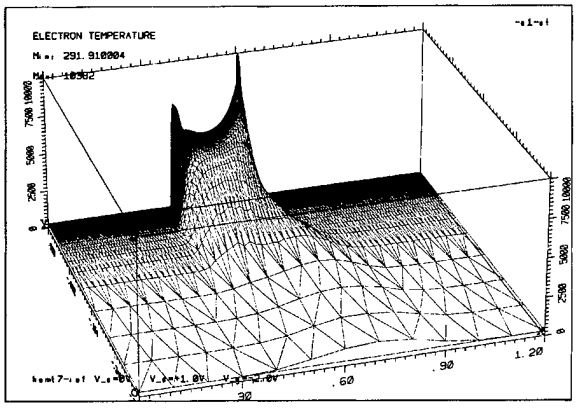


FIGURE 3 Electron temperature.

dimensional devices such that JFET or HEMT with 0.2 μm gates length.

We give here the charateristic I_S (V_{DS}) curves and the electron concentration and temperature distribution for the HEMT device for applied potential $V_S = 0\text{V}$, $V_d = +1\text{V}$, $V_G = -2\text{V}$. We use a mesh with 20 000 elements which gives 30 000 degrees of freedom for each flow variable.

Acknowledgements

The authors are grateful to FRANCE-TELECOM (CNET-Bagneux) for scientific discussions and financial support and to IEF for scientific collaboration

References

[1] Ben Abdallah, N., Degond, P. and Genieys, S. (1996). An energy-transport model for semiconductors derived from the Boltzmann equation. *Journal of Statistical Physics*, **84**(1–2), 205–231.

[2] Blotekjaer, K. (1970). Transport equations for electrons in two-valley semiconductors. *IEEE Trans. Electron. Device*, ED-17, 38–47.

[3] Brezzi, F., Marini, L. D. and Pietra, P. (1989). Numerical simulation of semiconductor devices. *Computer Methods in Applied Mechanics and Engineering*, **75**, 493–514.

[4] Chen, D., Kan, E. C., Ravaioli, U., Shu, C. W. and Dutton, R. W., An improved energy transport model including nonparabolicity and non-Maxwellian distribution effects. *IEEE electron. devices letters*, **13**(1), 26–28, January 1992.

[5] El Boukili, A. and Marrocco, A. (1994). Mixed finite element approach and nonlinear implicit schemes for drift-diffusion equation solution of 2d heterojunction semiconductor devices. *COMPEL*, **13**(4), 609–640.

[6] Gnudi, A., Odeh, F. and Rudan, M. (1990). Investigation of non-local transport phenomena in small semiconductor devices. *European trans. on telecommunications and related technologies*, **1**(3), 77–82.

[7] Hecht, F. and Marrocco, A. (1994). Mixed finite elements simulation of heterojunction structures including a boundary layer model for the quasi-fermi levels. *COMPEL*, **13**(4), 757–770.

[8] Marini, L. D. and Pietra, P. (1990). New mixed finite element schemes for current continuity equations. *Int. Jour. Comp. Math. Electre. Engi.*, **9**(4), 268–275.

[9] Marrocco, A. and Montarnal, Ph. (1996). Simulation of the energy transport models via mixed finite elements. *C. R. Acad. Sci. Paris, t. 323, Série I*, 535–541.

[10] Montarnal, Ph., Derivation of a hydrodynamic model from a two collision scales Boltzmann equation to appears.

[11] Stratton, R. (1962). Diffusion of hot and cold electrons in semiconductor barriers. *Phys. Rev.*, **126**(6), 2002–2014.

Authors' Biographies

Americo Marrocco received his Ph.D. (spec. numerical analysis) form the University of Paris in 1970. From this time, he is working at INRIA, within a research group involved in the elaboration and development of numerical methods in engineering sciences. Firstly he was interested with the numerical simulation of rotating electrical machines and from 1980, within a sustained collaboration with FRANCE-TELECOM-CNET, the main research topics were connected with numerical simulation in semiconductor area, first on process modelling and then on device modelling.

Philippe Montarnal received an Ingenior Degree in Applied Mathematics of the University of Grenoble in 1993 together with a DEA (predoctoral degree) from the same University. He spent one year, as a scientific trainee at the CEA (research Center for the Atomic Energy) working on asymptotic analysis in electromagnetics (study of different Ansatz, implementation of algorithms using Maple). Since then, he has been preparing his Ph.D. from Paris VI University with Prof. Perthame on the semiconductor energy transport and hydrodynamic equations (mathematical and numerical point of view) supported by INRIA. He is a member of SMAI.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 381–385
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Semiconductor Device Noise Computation Based on the Deterministic Solution of the Poisson and Boltzmann Transport Equations

ALFREDO J. PIAZZA* and CAN E. KORMAN

Department of Electrical Engineering and Computer Science, The George Washington University, Washington, DC 20052

Numerical simulation results of noise due to current fluctuations along an n^+-n-n^+ submicron structure are presented. The mathematical framework is based on the interpretation of the equations describing electron transport in the semiclassical transport model as stochastic differential equations (SDE). According to this formalism the key computations for the spectral density describing the noise process are reduced to a special initial value problem for the Boltzmann transport equation (BTE). The algorithm employed in the computation of the space dependent noise autocovariance function involves two main processes: the stationary self-consistent solution of the Boltzmann and Poisson equations, and a transient solution of the BTE with special initial conditions. The solution method for the BTE is based on the Legendre polynomial method. Noise due to acoustic and optical scattering and the effects of non-parabolicity are considered in the physical model.

Keywords: Stochastic differential equations, Boltzmann transport equation, noise, Legendre polynomials

1. INTRODUCTION

Current noise in semiconductors is due to the inherent randomness of the scattering mechanisms that govern electronic transport. These current fluctuations around a stationary value are generally characterized by the associated autocovariance function or equivalently, by their spectral density. Employing the machinery of SDE, a new noise model [1] shows that the key computations for the noise autocovariance function are reduced to the transient solution of the BTE with special initial conditions. This novel approach was previously utilized [2] for the computation of the noise spectral density in bulk silicon. The BTE was deterministically solved using the Legendre polynomials method and the results were in excellent agreement with those obtained using the Monte

* Corresponding author.

Carlo technique. In this paper we study the impact that the spatial variation of the doping has on the current noise autocovariance function. The subject of our study is a one-dimensional n^+-n-n^+ submicron structure in the stationary regime. The outline of our paper is as follows. Sections II and III concentrate on the noise and physical models respectively. Section VI describes the algorithm and Section V presents the numerical results. The last section is devoted to the conclusions.

2. THE NOISE MODEL

According to semiclassical transport theory, the motion of an electron in a semiconductor is described by the following stochastic differential equations:

$$\frac{d\vec{x}}{dt} = \vec{v}(\vec{k}) = \frac{1}{\hbar}\nabla_{\vec{k}}\varepsilon(\vec{k}), \quad d\frac{\vec{p}}{dt} = \hbar\frac{d\vec{k}}{dt} = -q\vec{E} + \vec{F}_r,$$
$$\text{and} \quad \vec{F}_r = \sum_i \hbar\vec{u}_i\delta(t - t_i), \tag{1}$$

where $\vec{x}, \vec{v}, \vec{p}$ and $\vec{k}$ are the electron position, velocity, momentum and wave vector, respectively, $\vec{E}$ is the electric field, $\varepsilon(\vec{k})$ is the energy-wave vector relationship in the given energy band and $\vec{F}_r$ is the random impulse force on the electron due to scattering. The random force is characterized by the *transition rate* $W(\vec{k},\vec{k}')$. Accordingly, the probability of scattering is given by

$$\Pr\{t_i - t_{i-1} > \tau\} = \exp\left\{-\int_{t_{i-1}}^{t_{i-1}+\tau} \lambda(\vec{k}(t'))dt'\right\},$$
$$\lambda(\vec{k}) = \int W(\vec{k},\vec{k}')d\vec{k}' \tag{2}$$

where $\lambda(\vec{k})$ is the *scattering rate*. Therefore, given the electron wave vector $\vec{k}, \lambda(\vec{k})\Delta t$ is the probability that a jump in momentum will occur in a small time interval Δt. Assuming that a scattering event has occurred at some time t_i, the probability density function for the amplitude of the jump is given by,

$$\chi_{\vec{k}_i}(\vec{u}_i) = \frac{W(\vec{k}_i, \vec{k}_i + \vec{u}_i)}{\lambda(\vec{k}_i)} \tag{3}$$

where $\vec{k}_i = \vec{k}(t_i^-)$ and $\vec{k}_i + \vec{u}_i = \vec{k}(t_i^+)$.

These same equations describe the electron motion in Monte Carlo simulations.

In the context of the SDE theory, these equations correspond to a Markov process, which is characterized by a transition probability function satisfying the Kolmogorov-Feller equation. In the case of semiclassical transport the latter equation is identical to the linear (non-degenerate) BTE (eq. 4)

$$\frac{\partial\rho}{\partial\tau}(\vec{x}',\vec{k}',0;\vec{x},\vec{k},\tau) + \vec{v}(\vec{k})\cdot\nabla_{\vec{x}}\rho - \frac{q}{\hbar}\vec{E}(\vec{x},\tau)\cdot\nabla_{\vec{k}}\rho =$$
$$\int \rho(\vec{x}',\vec{k}',0;\vec{x},\vec{k}'',\tau)W(\vec{k}'',\vec{k})d\vec{k}'' - \lambda(\vec{k})\rho, \tag{4}$$

subject to the following initial condition:

$$\rho(\vec{x},\vec{k}',0;\vec{x},\vec{k},\tau)|_{\tau=0} = \delta(\vec{k}-\vec{k}')\cdot\delta(\vec{x}-\vec{x}'), \tag{5}$$

where $\delta(\cdot)$ is the Dirac delta function and the following notation has been adopted:

$$\rho(\vec{x},\vec{k},\tau|\vec{x}',\vec{k}',0) \equiv \rho(\vec{x}',\vec{k}',0;\vec{x},\vec{k},\tau) \tag{6}$$

Generally, noise in semiconductors is characterized by the spectral density of current fluctuations. The spectral density is defined as the Fourier transform of the autocovariance function. The autocovariance of any random process can be found from the transition probability density function of such random process. Since the transition probability function satisfies the Kolmogorov-Feller equation, and the BTE is identical to it, the transition probability density function can be obtained from the solution of the BTE subject to appropriate initial conditions.

In [1] it was shown that the current longitudinal noise autocovariance function can be obtained

from the transient solution of the BTE subject to the following "special" initial condition:

$$\rho(\vec{x},\vec{k},\tau)|_{\tau=0} = \left(v(\vec{k}) - \langle\vec{v}\rangle_{\vec{x}}\right) f(\vec{x},\vec{k}). \quad (7)$$

where

$$\langle\vec{v}\rangle_{\vec{x}} \equiv \int v(\vec{k}) f(\vec{x},\vec{k})\, d\vec{k}. \quad (8)$$

Here, in (7) $f(\vec{x},\vec{k})$ represents the steady state solution of the BTE. The current autocovariance function is computed

$$K_{\vec{J}}(\vec{x},\tau) = q^2 \int v(\vec{k})\rho(\vec{x},\vec{k},\tau)d\vec{k}, \quad \tau \geq 0. \quad (9)$$

It is very important to note that this approach for the current noise autocorrelation computation is strictly within the framework of semiclassical transport. This approach directly connects the physics of scattering with the current noise characteristics and makes no additional assumptions regarding the nature of the noise.

3. THE DEVICE MODEL

A device under stationary conditions is appropriately characterized by the Poisson Eq. (10) and the space-dependent BTE (11). The self-consistent solution of these equations provides $f(\vec{x},\vec{k})$, the stationary probability density function and $\phi(\vec{x})$, the electrostatic potential throughout the device. In mathematical form these equations are given by:

$$\nabla^2_{\vec{x}}\phi(\vec{x}) = \frac{q}{\varepsilon_s}[n(\vec{x}) - N(\vec{x})]. \quad (10)$$

where $N(\vec{x})$ is the doping concentration and $n(\vec{x}) = \int f(\vec{x},\vec{k})d\vec{k}$ is the free charge concentration and

$$\vec{v}(\vec{k})\cdot\nabla_{\vec{x}} f(\vec{x},\vec{k}) - \frac{q}{\hbar}\vec{E}(\vec{x})\cdot\nabla_{\vec{k}} f = \int f(\vec{x},\vec{k}'')W(\vec{k}'',\vec{k})d\vec{k}'' - \lambda(\vec{k})f \quad (11)$$

Since in our example we are using a 1-D $n^+ - n - n^+$ structure, $\vec{E}$ and $\vec{x}$ are parallel and can be replaced by the scalars E and x. It is also assumed that $\vec{E}$ is in the symmetric [1 1 1] crystallographic direction and therefore a single band distribution function accurately represents the state of the momentum space.

4. THE ALGORITHM

The space-dependent noise auto-correlation function is computed according to the following straight-forward procedure:

1. Obtain $f(x,\vec{k})$ and $\phi(x)$ from the self consistent solution of the Poisson Eq. (10) and space-dependent BTE (11). The self-consistent solution to the Poisson and Boltzmann equations is achieved by concurrently solving the corresponding discretized equations in a Gummel-type iteration. This process comprises the following steps:
   (a) Generate initial values for $\phi(x)$ and $n(x)$ using the Drift-Diffusion model.
   (b) Generate initial values for $f(x,\vec{k})$ consistent with $\phi(x)$ and $n(x)$ calculated in step (a), solving the homogeneous BTE.
   (c) Obtain $f^{i+1}(x,\vec{k})$ from $f^i(x,\vec{k})$ solving the space-dependent BTE until the current along the device converges. The BTE is solved using the electric field that corresponds to $\phi^i(x)$.
   (d) Solve Poisson Eq. (10) for $\phi^{i+1}(x)$ given $n^{i+1}(x)$.
   (e) Steps (c) and (d) are repeated until the whole system converges.
2. Compute special initial condition $\vec{g}(x,\vec{k},\tau)|_{\tau=0}$ as prescribed by (7).
3. Obtain $\vec{g}(x,\vec{k},\tau)$ solving the transient BTE (4) with the special initial condition computed in step 2, and the electric field $E(x)$ corresponding to $\phi(x)$ obtained in step 1. The transient solution to the BTE is based on implicitly approximating the time derivative in Eq. (4).

Therefore $g(x,\vec{k},\tau^{k+1})$ can be computed from the solution of Eq. (4) and the knowledge of $g(x,\vec{k},\tau^{k})$, for any k. This process is comprised of the following steps:

(a) Generate initial guess values for $g(x,\vec{k},\tau^{k+1}) = g(x,\vec{k},\tau^{k})$
(b) Obtain $g^{i+1}(x,\vec{k},\tau^{k+1})$ from $g^{i}(x,\vec{k},\tau^{k+1})$ solving the transient BTE, Eq. (4), until total current converges.
(c) Advance the time index and go to step (a).

4. Compute the autocovariance function according to Eq. (9).

5. SOLUTION OF THE BTE EMPLOYING LEGENDRE POLYNOMIALS

In order to map ellipsoidal energy surfaces into spherical ones, the Herring-Vogt transformation is employed. In this transformation the original coordinates $\vec{x},\vec{E}$ and $\vec{k}$ are mapped into $\vec{x}^*,\vec{E}^*$ and $\vec{k}^*$, respectively and the dispersion relationship becomes spherical $\gamma(\varepsilon) = \varepsilon + \beta\varepsilon^2 = \hbar^2 k^{*2}/2m_0$ in the new domain. The BTE remains unchanged in the stared variables and the direction of $\vec{E}^*$ defines a symmetry axis ($\hat{z}$). The density function dependence on momentum can be expressed in terms of only two independent variables: ε and θ. We expand $f(x^*,\vec{k}^*,t)$ in Legendre polynomials according to:

$$f(x^*,\vec{k}^*,t) = f_0(x^*,\varepsilon,t) + k^* g(x^*,\varepsilon,t)\cos\theta + k^{*2} h(x^*,\varepsilon,t)(3\cos 2\theta - 1). \quad (12)$$

This representation for $f(x^*,\vec{k}^*,t)$, is replaced in Eq. (4). Recalling that the Legendre polynomials are orthogonal, the resulting equation is solved for the corresponding coefficients independently. This results in a system of three coupled differential-difference equations for the functions f_0, g and h, in terms of t, x^* and ε. The system is solved for $f_0(x^*, \varepsilon, t)$ in the energy domain using standard finite difference method. The functions $g(x^*, \varepsilon, t)$ and $h(x^*, \varepsilon, t)$ are then obtained from $f_0(x^*, \varepsilon, t)$.

6. NUMERICAL RESULTS

Figure 1 shows the parameters employed in the simulation: A Gaussian doping profile ranging from 10^{16} to 10^{17} cm^{-3} with the highest doping at the device boundaries and minimum in the middle, 2 Volts of applied voltage, a device length of 0.5 μm, etc. Figure 2 shows the space-dependent current noise autocovariance function computed for this device. Several comments are in order: a) The maximum value of σ^2, the noise power, occurs at the 0.15 μm point in which the electrostatic field reaches its maximum value. This result had already been observed for Bulk silicon in [2], b) The auto-covariance function remains positive for points in the middle of the device for which the

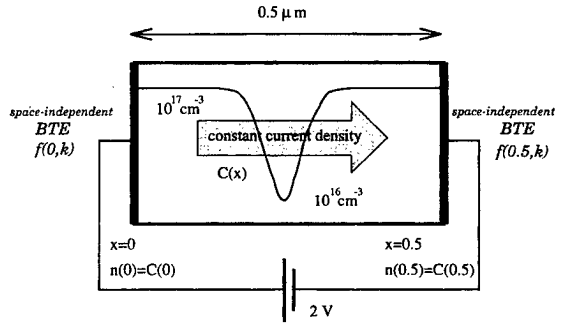


FIGURE 1 shows the device considered in our simulations. The device is 0.5 μm long, the applied potential is equal to 2 volts, the doping profile is Gaussian and symmetrical with its lowest value in the middle of the device.

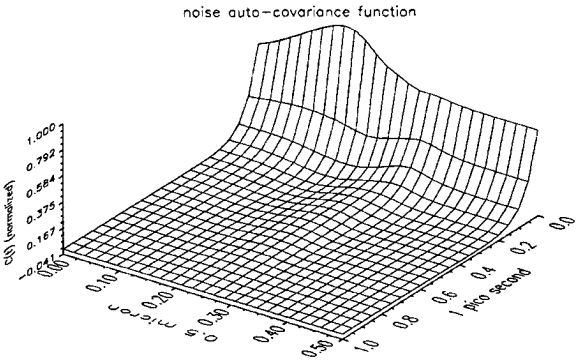


FIGURE 2 shows the current longitudinal noise autocorrelation function corresponding to our simulations. The value of σ^2 reaches it maximun value at 0.15 μm, point for which the electrostatic field has reached its maximum value.

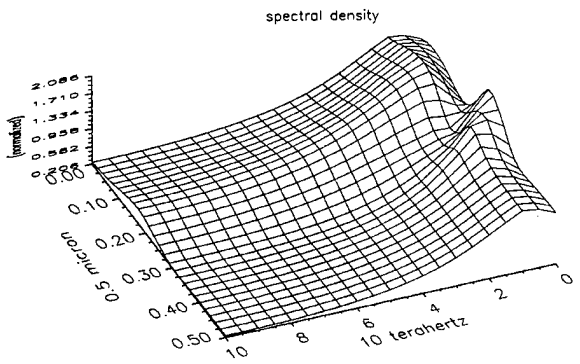


FIGURE 3 shows the spectral density function corresponding to Figure 2. This figure shows that the noise at 0.15 μm, point for which the electrostatic field has reached its maximum value, has important high frequency components.

doping reaches its minimum and c) the auto-covariance functions at both ends are identical, a result that is in complete agreement with the assumed space-independence condition for the boundaries of the device. Figure 3 shows the corresponding spectral density function to Figure 2. The most important feature of this figure is that higher frequency components are observed again for the points in the area of large electrotatic field.

7. CONCLUSIONS

Numerical results for the space-dependent auto-covariance and spectral density function of noise due to acoustic and optical phonon scattering were presented. A simple algorithm to compute the noise auto-correlation function that accounts for the effect of the spatial variation of the doping was described. The algorithm was succesfully employed for noise calculations for a sub-micron stucture in stationary regime. Hence the SDE noise model is demonstrated to be a viable approach for space-dependent noise computations .

References

[1] Korman, C. E. and Mayergoyz, I. D. (1996). "Semiconductor Noise in the Framework of Semiclassical Transport", *Phys. Rev. B*, **54**, 24.

[2] Piazza, Alfredo J. and Korman, C. E., "Computation of the Spectral Density of Current Fluctuations in Bulk Silicon Based on the Solution of the Boltzmann Transport Equation", accepted for publication.

[3] Hongchin, Lin, Neil, Goldsman and Mayergoyz, I. D. (1992). "Device Modeling by Deterministic Self-Consistent Solution of Poisson and Boltzmann Transport Equations", *Solid-St. Electron*, **35**(6), pp. 769–778.

Authors' Biographies

Can E. Korman was born in Silver Spring, Maryland. He received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from the University of Maryland in 1985, 1987 and 1990, respectively. In 1991, he joined the Department of Electrical Engineering and Computer Science at the George Washington University, Washington, DC where he is currently an Assistant Professor.

His current research is in the area of numerical simulation of semiconductor devices, semiconductor noise characterization, parallel algorithms, stochastic differential equations, VLSI and magnetic viscosity. He is the recipient of the National Science Foundation Research Initiation Award. He has also received the following teaching awards at the George Washington University: the Eta Kappa Nu Association Teacher of the Year Award for 1993–94 and the Engineer's Council Faculty of the Year Award for 1994–95. He is a member of the IEEE Electron Devices Society and the Circuits and Systems Society.

Alfredo Piazza was born in Lima, Peru, in 1963. He received the B.S. degree in industrial engineering from the Pontificia Universidad Catolica del Peru in 1988. In 1995 he obtained the M.S. degree in electrical engineering with concentration in electrophysics from the George Washington University. He continues toward the Ph.D. in electrical engineering with concentration in systems science, networks and controls. His research concentrates in the numerical simulation of semiconductor devices. His current research is on the topic of noise characterization of semiconductor materials by means of the direct solution of the Boltzmann Transport Equation.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 387–391
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Consistent Hydrodynamic and Monte-Carlo Simulation of SiGe HBTs Based on Table Models for the Relaxation Times

B. NEINHÜS*, S. DECKER, P. GRAF, F. M. BUFLER and B. MEINERZHAGEN

Institut für Theoretische Elektrotechnik und Mikroelektronik, Universität Bremen, Kufsteiner Strasse, Postfach 33 04 40, 28334 Bremen, Germany

Good agreement between a hydrodynamic and a Monte-Carlo device model is demonstrated in this paper for an advanced SiGe Heterojunction Bipolar Transistor. This result is based on two principles: 1) Extraction (from the Monte-Carlo bulk model under homogeneous conditions) of the relaxation times τ at discrete points of the parameter space spanned by the Ge-content x, doping density N, carrier temperature T_C and lattice temperature T_L. 2) Modeling of the relaxation times $\tau(x, N, T_C, T_L)$ by splines.

Keywords: Silicon-Germanium (SiGe), Hetereo Bipolar Transistor (HBT), hydrodynamics, transport parameter, spline model, GHDM, Monte-Carlo

1. INTRODUCTION

Among all SiGe device concepts, SiGe Heterojunction Bipolar Transistors (HBTs) currently have the highest potential for commercial applications. In order to support the design of SiGe HBTs, accurate and efficient device simulation tools are necessary. However, though even some of the commercially available simulators offer the capability of simulating heterojunction devices, reliable transport parameters for these devices are not available for most design tasks. The underlying reason for this dilemma is that the Ge content x is variable in SiGe devices, which has added an additional dimension to the problem of determining transport parameters.

For example, the relaxation times of a hydrodynamic (HD) model for SiGe HBTs $\tau(N, x, T_C, T_L)$ depend on four independent quantities instead of three, namely the total doping density N, the Ge content x, the carrier temperature T_C and the lattice temperature T_L. Therefore the traditional approach of extracting transport parameters predominantly from experimental data, which worked

* Corresponding author.

well for silicon for a long time, is no longer feasible for SiGe devices because reliable experimental data, especially for strained SiGe, are hardly available.

2. TRANSPORT PARAMETERS

In order to overcome this problem the comprehensive and experimentally verified Monte-Carlo bulk transport model described in [1] was applied to generate transport parameters for the drift-diffusion (DD) or HD device simulations at discrete mesh points of the 4D space spanned by N, x, T_C, T_L.

To generate the smooth functions $\tau(N, x, T_C, T_L)$, that are needed for example for HD device simulations from the resulting table of transport data, a flexible monotonicity preserving spline interpolation scheme has been developed.

3. THE SPLINE MODEL

Compared to the traditional approach of using closed form analytic expressions with only few model parameters for $\tau(N, x, T_C, T_L)$ our spline approximation scheme has several advantages: For example, it adapts itself automatically to model extensions like an extended range of N, x, T_C or T_L. Moreover, it is easy to control the accuracy of the spline interpolation by just generating a denser table of transport data for the range of N, x, T_L, T_C, that is of highest interest. In addition, our method is also capable of working on a non-rectangular grid in the N-x-T_L-T_C space.

The underlying algorithm considers the functional dependencies of the transport parameters in two stages. First, the dependence on variables not being influenced by device simulation like N, x_{Ge} and sometimes T_L are considered by multilinear interpolation. In order to process data on a non-rectangular grid the algorithm performs a multi-dimensional search of the nearest neighbour data points suitable for the interpolation.

The second stage deals with dependencies on variables of the hydrodynamic equations itself, like the carrier temperature T_C. This is considered by using a variant of cubic splines called AKIMA Subsplines [2]. The spline curves generated by this method do not contain artificial oscillations which can cause artificial modeling results or convergence problems for the solution algorithm.

The final result that influences the memory requirements of the device simulator is a set of spline coefficients for each grid point of the device. For medium grid sizes (3000 points) roughly 20 MBytes of memory are necessary to hold the spline coefficients.

4. SIMULATION RESULTS

In order to verify the validity of our modeling approach we have simulated the two-dimensional SiGe HBT structure shown in Figure 1. The structure is very demanding for numerical device models because of its narrow base and piecewise constant profiles for N and x that give rise to abrupt junctions and steps in the valence and conduction band edges. Consequently this device is well suited for testing the modeling accuracy of

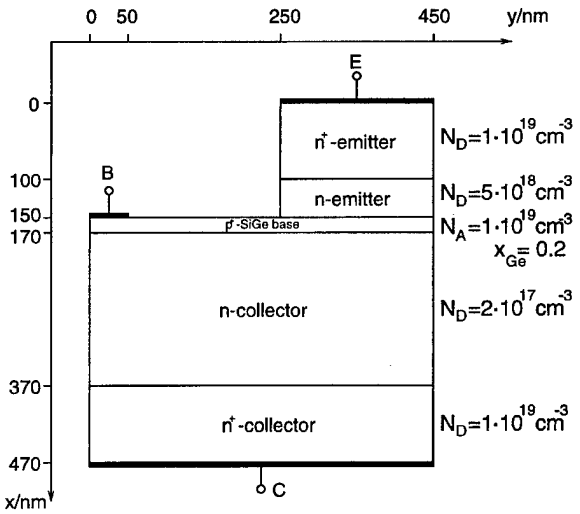


FIGURE 1 The 2D SiGe-HBT test structure.

classical device simulation in comparison to Monte-Carlo device simulation. The structure has been simulated by three different models. The first model is the DD model. The second model is an extension of the Generalized HD Model reported in [3] for devices with position-dependent band structure. Moreover, the heat flux reduction that has been proven to be beneficial for ultra short MOSFETs [4] has been adopted. The third model is a newly developed Monte-Carlo device model for SiGe heterojunction devices. Details of this MC model will be published elsewhere. Since all transport parameters of the DD and HD simulators have been derived from the SiGe MC bulk model and because the MC device model used exactly the same band structure and scattering models as the MC bulk simulator, all three models are fully consistent under homogeneous material and field conditions. Moreover, all three device models use exactly the same offsets for the valence and conduction band edges.

The results of the 2D simulations for $V_{BE} = 0.75$ V, $V_{CE} = 1$ V and 300 K are summarized in Figures 2–5 for the electric potential, the dynamic temperature, the drift velocity and the electron density. In all cases the results of the three models are shown along the vertical line at $y = 450$ nm. It can be seen that at the base collector junction even in the MC model the drift velocity is more than a factor of two higher than the maximum drift velocity under homogeneous field conditions. Despite this overshoot, which would be extremely large for Si-based devices, it can be clearly stated that the hydrodynamic results are in good agreement with the MC-results. Especially the electron density profile in the base and the space charge region, which is important for the transient behavior of the HBT, agrees well for the HD

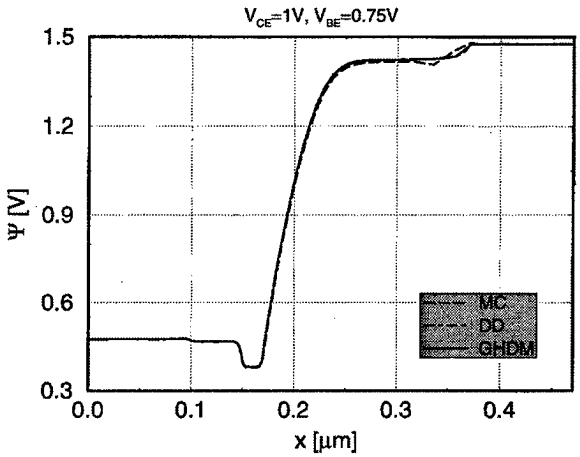


FIGURE 2 Comparison of the electrostatic potential profiles resulting from the DD, HD and MC models. All models are in good agreement.

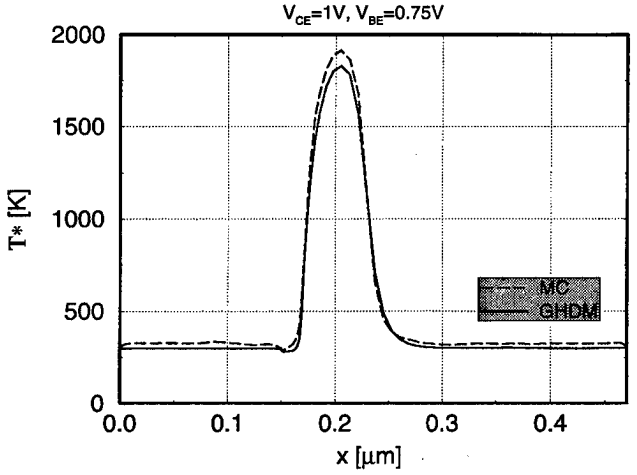


FIGURE 3 Comparison of the dynamic temperature profiles resulting from the HD and MC models. Both model are in good agreement.

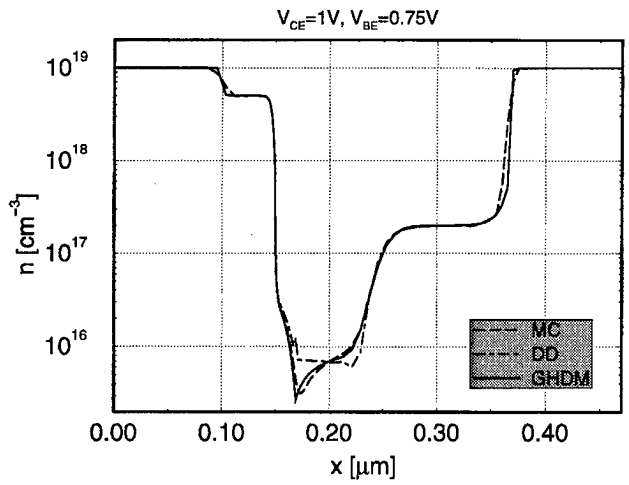


FIGURE 4 Comparison of the electron densities resulting from the DD, HD and MC models. The HD and MC density distributions are in good agreement. The DD model deviates substantially.

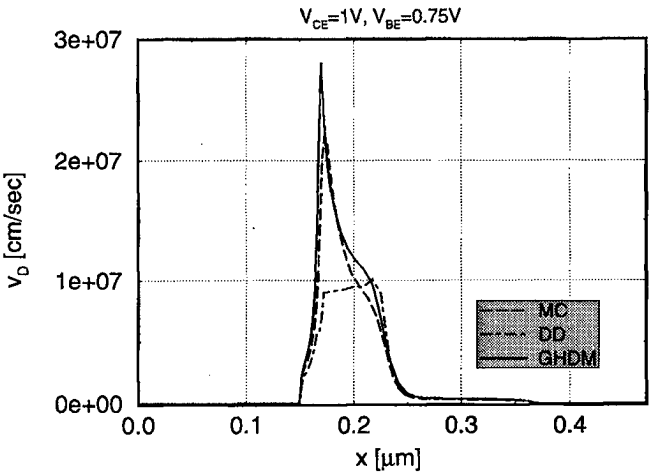


FIGURE 5 Comparison of the drift velocity profiles resulting from the DD, HD and MC models. Please note the velocity overshoot which is extreme for Si-based devices.

and the MC simulation. On the other hand, the DD simulation deviates much more from the MC-reference, which may lead to intolerable errors of the DD-model for agressively scaled SiGe HBTs in the near future [5].

To the best of our knowledge the results reported in this paper represent the first 2D simulations of a SiGe HBT with fully consistent DD, HD and MC device models.

Acknowledgements

This work was supported in part by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie under Contract No. 01M2416A and by Siemens AG in Munich. The authors like to thank H. Förster, D. Nuernbergk and F. Schwierz from the Technical University of Ilmenau for proposing the SiGe HBT in Figure 1 and several fruitful discussions.

References

[1] Bufler, F. M., Graf, P., Meinerzhagen, B., Adeline, B., Rieger, M. M., Kibbel, H. and Fischer, G. (1997). "Analysis of electron transport properties in unstrained and strained $Si_{1-x}Ge_x$ alloys", *IEEE Trans. Semicond. Technol. Modeling and Simulation*, **1**.

[2] Müllges, G. E. and Reutter, F. (1993). *Numerik-Algorithmen mit ANSI-C Programmen*, BI-Wissenschaftsverlag, Mannheim, 2nd edition.

[3] Thoma, R., Emunds, A., Meinerzhagen, B., Peifer, H. J. and Engl, W. L. (1991). "Hydrodynamic equations for semiconductors with nonparabolic bandstructures", *IEEE Trans. Electron Devices*, **38**, 1343–1352.

[4] Bork, I., Jungemann, C., Meinerzhagen, B. and Engl, W. L. (1994). "Influence of heat flux on the accuarcy of hydrodynamic models for ultrashort Si MOSFETs", in *NUPAD Tech. Dig.*, Honolulu, **5**.

[5] To be published at ESSDERC 97.

Authors' Biographies

Burkhard Neinhüs received the Diploma in physics from the RWTH Aachen, Germany, in 1995. In 1996 he joined the Insitut für Theoretische Elektrotechnik und Mikroelektronik at the University of Bremen and is working on hydrodynamic device simulations.

Stefan Decker received the Diploma in electrical engineering from the RWTH Aachen, Germany, in 1993. From 1993 to 1995 he worked at the Institut für Theoretische Elektrotechnik, RWTH Aachen, on hydrodynamic and Monte-Carlo simulation methods. Since 1995 he was working at the Insitut für Theoretische Elektrotechnik und Mikroelektronik at the University of Bremen.

Peter Graf received the Diploma in physics from the RWTH Aachen (Technical University of Aachen), Germany, in 1991. At the Institut für Theoretische Elektrotechnik, RWTH Aachen, he then worked on the Monte-Carlo simulation of homogeneous systems and its benefits for hydrodynamic charge transport parameters. After joining the Institut für Theoretische Elektrotechnik und Mikroelektronik, Universität Bremen, he dealt with algorithms for Monte-Carlo device simulation and their implementation into the institute's Monte-Carlo heterojunction device simulator.

Fabian M. Bufler studied physics at the TU Braunschweig and RWTH Aachen (both Germany) including an academic year at the Université de Grenoble I (France) with a scholarship of the Studienstiftung des deutschen Volkes and received the Dipl.-Phys. degree in 1992. Since then he was working as a research assistant at the Institut für Theoretische Elektrotechnik, RWTH Aachen, and moved in 1995 together with the group of Prof. Bernd Meinerzhagen to the Institut für Theoretische Elektrotechnik und Mikroelektronik, Universität Bremen (Germany). His main research interests are concerned with transport theory and full band Monte Carlo simulation for SiGe based devices.

Bernd Meinerzhagen received the Dipl.-Ing. degree in electrical engineering in 1977, the Dipl.-Math. degree in mathematics in 1981, the Dr.-Ing. degree in electrical engineering in 1985 and the "venia legendi" in 1995 all from the RWTH Aachen (Germany). From 1978 to 1986 as a Research and Teaching Assistant at the RWTH Aachen he worked mainly on the development of numerical device modeling codes. In 1986 he joined AT&T Bell Laboratories in Allentown, PA, as a Member of Technical Staff, where he developed advanced numerical models for MOS substrate and gate currents. He went back to the RWTH Aachen in 1987 and became head of the research and development group for Silicon technology modeling and simulation (TCAD). In 1995 he was appointed Professor at the University of Bremen, where his current research interests include TCAD and the theory of electromagnetic fields and networks.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 393–399
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Additive Decomposition Applied to the Semiconductor Drift-Diffusion Model

ELIZABETH J. BRAUER [a,*], MAREK TUROWSKI [b] and JAMES M. McDONOUGH [c]

[a] *P.O. Box 15600, Department of Electrical Engineering, Northern Arizona University, Flagstaff, AZ 86011-5600;* [b] *Department of Electrical Engineering, University of Kentucky, 453 Anderson Hall, Lexington, KY 40506-0046, USA;* [c] *Department of Mechanical Engineering, University of Kentucky, 514h Ctr for Robotics and Mfg Systems, Lexington, KY 40506-0108, USA*

A new numerical method for semiconductor device simulation is presented. The additive decomposition method has been successfully applied to Burgers' and Navier-Stokes equations governing turbulent fluid flow by decomposing the equations into large-scale and small-scale parts without averaging. The additive decomposition (AD) technique is well suited to problems with a large range of time and/or space scales, for example, thermal-electrical simulation of power semiconductor devices with large physical size. Furthermore, AD adds a level of parallelization for improved computational efficiency. The new numerical technique has been tested on the 1-D drift-diffusion model of a *p-i-n* diode for reverse and forward biases. Distributions of ϕ, n and p have been calculated using the AD method on a coarse large-scale grid and then in parallel small-scale grid sections. The AD results agreed well with the results obtained with a traditional one-grid approach, while potentially reducing memory requirements with the new method.

Keywords: Numerical methods, semiconductor, simulation, drift-diffusion, decomposition

INTRODUCTION

The numerical method, additive decomposition, has been successfully applied in mechanical and chemical engineering to Burgers' equation and the Navier-Stokes equations governing turbulent fluid flow by decomposing governing equations into large-scale and small-scale parts without averaging, e.g., [1–3]. The additive decomposition (AD) technique is well suited to problems with a large range of time and/or space scales, for example, thermal-electrical simulation of power semiconductor devices with large physical size. Thermal-electrical effects have a large range of

* Corresponding author.

time scales since the electrical time constants are much faster than the thermal time constants. Furthermore, additive decomposition adds a level of parallelization for improved computational efficiency. Thus, semiconductor device simulation is a natural application of the additive decomposition numerical technique. Initially, we decompose the simplest device equations, the drift-diffusion model, to test the method. After successful implementation for the drift-diffusion equations, we plan to apply the new approach to the hydrodynamic semiconductor model.

ADDITIVE DECOMPOSITION OF THE DRIFT-DIFFUSION EQUATIONS

The standard drift-diffusion model of semiconductors consists of the following equations. Poisson's equation is

$$\nabla^2\phi = -\frac{q}{\varepsilon_s}(p - n + N_d - N_a) \tag{1}$$

and the continuity equations for electrons and holes are

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla[D_n \cdot \nabla n - \mu_n \cdot n \cdot \nabla\phi] - R \tag{2}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla[-D_p \cdot \nabla p - \mu_p \cdot p \cdot \nabla\phi] - R \tag{3}$$

where the symbols have their standard meaning [4].

Additive decomposition of the drift-diffusion semiconductor model proceeds as follows. First, the basic variables, electrostatic potential, electron density and hole density, are divided into large-scale and small-scale components:

$$\phi = \bar{\phi} + \phi^*, \quad n = \bar{n} + n^*, \quad p = \bar{p} + p^*. \tag{4}$$

here $\bar{\phi}$ is the large-scale component and ϕ^* is the small-scale component of the potential, and similarly for n and p. After substituting Eq. (4) into Eq. (1) through (3), the equations are then decomposed additively into large-scale and small-scale equations:

$$\nabla^2\bar{\phi} = \frac{q}{\varepsilon_s}(\bar{n} - \bar{p} - \overline{N_{da}}) \tag{5a}$$

$$\nabla^2\phi^* = \frac{q}{\varepsilon_s}(n^* - p^* - N^*_{da}) \tag{5b}$$

where $N_{da} = N_d - N_a$, and

$$\frac{\partial\bar{n}}{\partial t} = \nabla\Big[D_n\nabla\bar{n} - \mu_n\bar{n}\nabla\bar{\phi} - \mu_n \cdot (1-\beta) \cdot \left(\bar{n}\nabla\phi^* + n^*\nabla\bar{\phi}\right)\Big] - \bar{R} \tag{6a}$$

$$\frac{\partial n^*}{\partial t} = \nabla\Big[D_n\nabla n^* - \mu_n n^*\nabla\phi^* - \mu_n \cdot \beta \cdot \left(\bar{n}\nabla\phi^* + n^*\nabla\bar{\phi}\right)\Big] - R^* \tag{6b}$$

Equation (5a) is the large-scale Poisson's equation, (5b) is the small-scale Poisson's equation, (6a) is the large-scale current continuity equation for electrons, and (6b) is the small-scale current continuity equation for electrons. The decomposition of the current continuity equation for holes is similar to (6). The additive decomposition is done in such a manner, that Eqs. (5a) and (5b) added together are equivalent to Eq. (1), and in the same way, Eqs. (6a) and (6b) added together result in the original Eq. (2).

The decomposition parameter β is a coupling coefficient between large-scale and small-scale solutions [1, 2]. The parameter β controls the large-scale/small-scale interaction. For example, in the limit $\beta = 1$, the effect of large scale changes is transferred to the small scale, but not vice versa. In principle, β can be assigned any value between 0 and 1 implying nonuniqueness of the decomposition, but Brown *et al.* [5] have recently demonstrated optimal values for the incompressible Navier-Stokes equations. In the present study, $\beta = 1$ has been used.

Decomposition of the variables in the recombination term R requires a special treatment. In the Shockley-Read-Hall expression for recombination-generation, n and p appear in the numerator

as well as in the denominator. This makes a straight implementation of the additive decomposition impossible. One possible solution to this problem is to calculate the value of the denominator with the n and p values from the previous iteration. Then, treat the denominator as a constant and decompose only the numerator, using the decomposition parameter β, which will generate both R^* and $\bar{R}$. Another possible method is to linearize R by expanding in a Fréchet-Taylor series, and then decompose the basic variables n and p of the linearization in the usual way. In the preliminary tests of the AD method presented here, the authors used the first method to decompose the recombination term. It is important to note that the decomposition has been designed so that sums of the large and small-scale equations are the original Eqs. (1–3). For linear problems, this represents a straightforward approach, the validity of which is trivial to demonstrate. For problems of the type considered here, the proof is more involved, and we refer the reader to Brown *et al.* [5] for an indication of what is needed.

For a 1-D simulation, each large-scale equation is solved on the large-scale grid with N points across the device (N is quite a small number), and the small-scale equations are solved on the small-scale grid over a section of the device centered around a large-scale grid point; see Figure 1. There is a separate set of small-scale equations (for small-scale space or time points) for *each* large-scale grid point. Thus, the small-scale equations can be solved independently and in parallel in N sections. The large-scale solution, which depends on the small-scale solution, combines these to produce total results over the whole simulation domain.

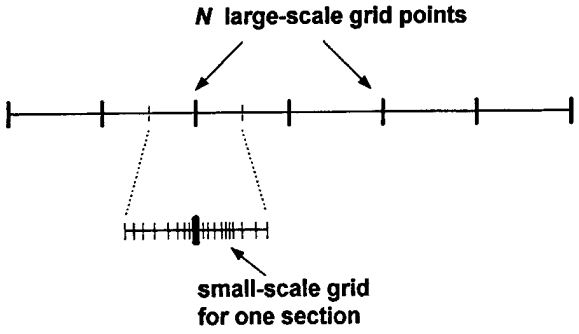


FIGURE 1 Illustration of large-scale and small-scale discretization grid.

The large-scale function may be defined in various ways between the large-scale grid points. In our first calculations, the large-scale values were set constant within one section, equal to the values obtained from the large-scale solution for the large-scale point of the section; i.e., the large-scale solution representation is piecewise constant when used in the small-scale equations. The small-scale values are variations of the exact solution around the (local) constant large-scale value within one section. The total small-scale function along the entire device length is, of course, not continuous, although the derivative of the small-scale function is the same as the derivative of the total solution where both are well defined. We also tried the approach which is a linear interpolation of the large-scale solution between the large-scale grid points. In this case, the small-scale function is continuous but has discontinuous slope between sections. Nevertheless the convergence of the final solution was much worse with this method than with the first one.

Boundary conditions between the separate sections pose a significant problem, particularly if the sections are computed in parallel. Two types of boundary conditions have been tested, viz., the Neumann type and the Dirichlet type, and it has appeared that properly calculated Dirichlet conditions result in more stable solutions.

NUMERICAL EXPERIMENTS

As a first test, the new AD technique has been applied to the 1-D solution of the drift-diffusion transport equations of the *p-i-n* diode for reverse-bias and forward-bias conditions. The test *p-i-n* diode structure is presented in Figure 2. As a computational platform, the PASSC semiconduc-

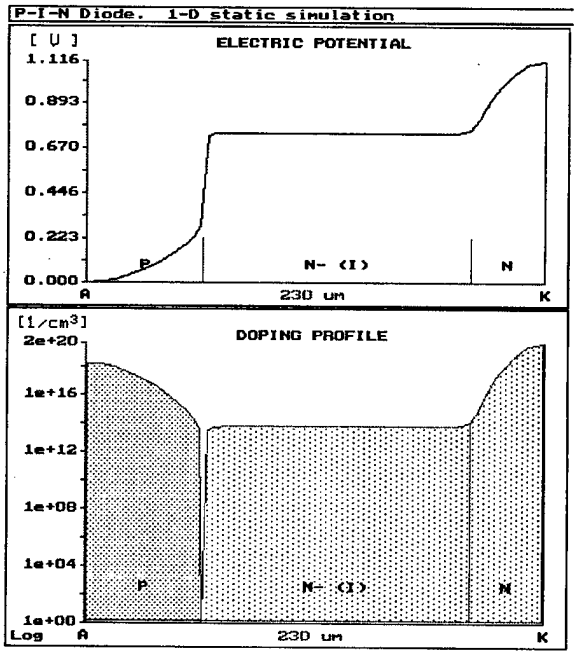


FIGURE 2 Potential distribution for zero bias and doping profile of the best *p-i-n* diode structure.

tor simulation program [6] has been used, with extensive physical models included [7, 8]. In the present calculations, the decomposition parameter β has been set to 1. The distributions of electrostatic potential ϕ, electron concentration n and hole concentration p have been calculated as basic results. As a measure of quality of computed results, the space charge formation has been used in the case of zero bias [9] and reverse bias, and the distributions of ϕ, n, and p as well as current density in the case of forward bias.

First, as a reference, the solution with the traditional method has been obtained, using a non-uniform grid with 40 points in one dimension. Then, a large-scale solution has been computed on N grid points uniformly distributed along the device. In the present study, we have tried $N = 9$ and $N = 5$. Small-scale results have been calculated in separate sections surrounding each large-scale point, except the border ones, which resulted in 7 or 3 separate sections along the device, respectively. With the boundary points of neighboring sections overlapping, the sections included from 5 to 12 small-scale grid points for 7 sections, and 15, 9 and 18 small-scale grid points per section in the case of 3 sections. After the large-scale solution has been calculated, the small-scale initial guess values are updated by interpolation using the large-scale solution. Then the accurate small-scale solutions are computed in separate (parallel) sections around each large-scale point. The final results for the forward bias $V_{AK} = 1.0\,\mathrm{V}$, calculated with the large grid of $N = 5$, are shown in Figure 3. These results are very close to the reference results obtained with the standard one-grid method. The biggest problem in this approach is setting up appropriate boundary conditions between sections, because the final results are very sensitive to the slightest change in boundary conditions of each section. The current density

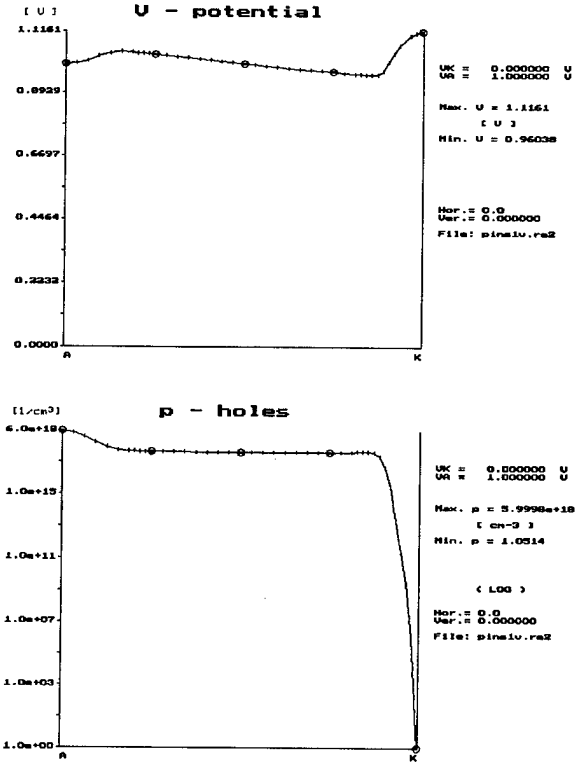


FIGURE 3 Final additive decomposition solution for the forward bias $V_{AK} = 1.0\,\mathrm{V}$ the thicker grid lines (circled) correspond to the large-scale grid, $N = 5$.

from the reference results for $V_{AK} = +1.0\,\mathrm{V}$ was $134.7\,\mathrm{A/cm^2}$. The current density calculated using only the large-scale solution was $168.4\,\mathrm{A/cm^2}$ for $N = 5$ (error 25%), and $135.8\,\mathrm{A/cm^2}$ for $N = 9$ (error 0.8%), and the values were uniform along the entire device length. After solving the small-scale equations in each section, the final current density values obtained with the AD method were slightly different in different sections (within 10%), which was due to inaccurate small-scale boundary conditions for separate sections. Iterating the small-scale results to recalculate small-scale boundary conditions and re-solving small-scale equations improved the uniformity of the resulting current density, although the rate of improvement was slow.

In a similar manner, results for the reverse bias were calculated. As mentioned already, in this case, the space charge formation has been used as a measure of accuracy of results. Figure 4 shows the net charge ($Q = p - n + N_d - N_a$) calculated on a non-uniform single grid (40 points in 1-D) with the traditional method for the reverse bias $V_{AK} = -1.0\,\mathrm{V}$. The net charge distribution calculated only with the coarse large-scale grid ($N = 9$) is presented in Figure 5a. After solving the small-scale equations in separate sections and adding the large and small components according to Eq. (4), the resulting space charge distribution (Fig. 5b)

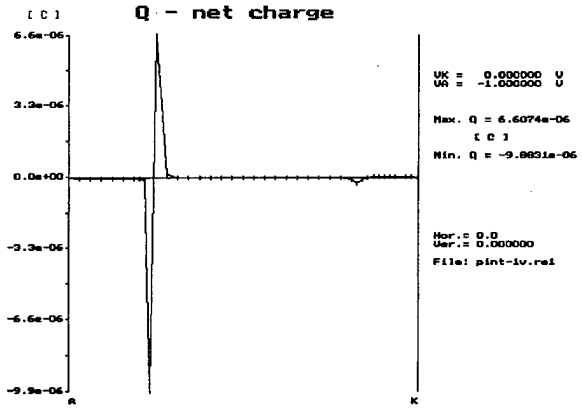


FIGURE 4 Net charge ($Q = p - n + N_d - N_a$) for the reverse bias $V_{AK} = -1.0\,\mathrm{V}$, calculated with the traditional method on one non uniform discretization grid.

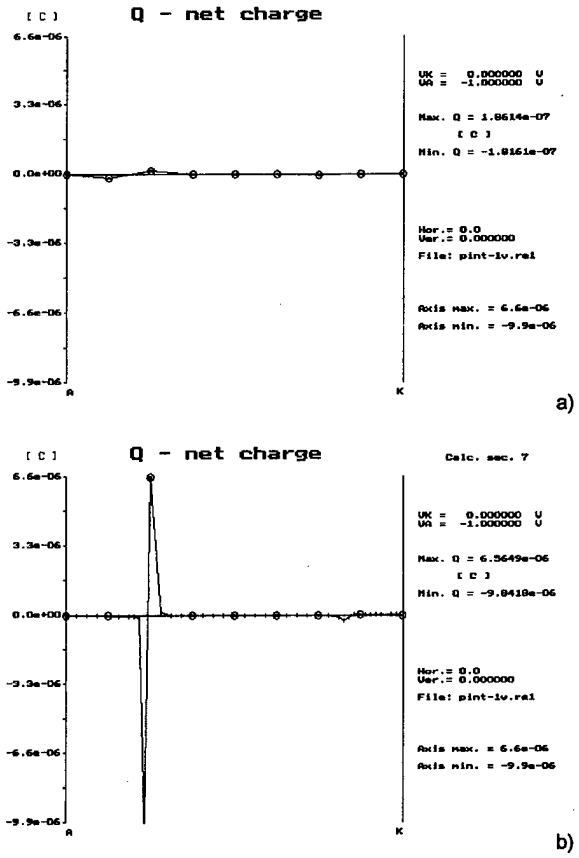


FIGURE 5 Net charge distributions for the reverse bias $-1.0\,\mathrm{V}$: a) results calculated using only the large-scale grid with 9 points uniformly distributed, b) results obtained using AD method after adding the large-scale and small-scale components (the thicker, circled grid lines correspond to the large-scale grid); note the good agreement with the reference results in Figure 4.

was almost identical to the reference space charge distribution obtained with the traditional method (Fig. 4). While the current density for the reverse bias calculated using only the coarse large-scale grid is the same order as the reference solution (error below 36%), the local charge distribution is absolutely missing from large-scale solution (Fig. 5a), and is updated only by adding the small-scale components (Fig. 5b).

Here, we discuss if there is any physical meaning of the basically arithmetic operations leading to the additive decomposition of the semiconductor model into large-scale and small-scale compon-

ents. The method was originally conceived for a more accurate description of turbulent fluid flow, where the two types of phenomena could be observed: large-scale fluid flow in the main stream, and turbulences which might be perceived as small-scale fluctuations over the large-scale flow. Is there anything similar in semiconductors? The results shown above suggest that the current density may be treated as a "large-scale value" as it can be calculated with a reasonable accuracy even on very coarse grids. We have checked various forward bias values; the error of large-scale current density ranged from 0.8% for $V_{AK} = +1.0\,\mathrm{V}$ to 20% for $V_{AK} = +2.0\,\mathrm{V}$. In the 1-D case, the value of the total current density should be uniform and equal over the entire length of the device, which again shows its large-scale character. On the other hand, results in Figure 5 indicate that the space charge shows a "small-scale character", as it can be only calculated using a fine small-scale grid. These observations actually agree with the more rigorous analysis of the basic semiconductor equations using a singular perturbation approach [10]. There, it is concluded that if the Scharfetter-Gummel discretizetion [11] of the continuity equations is used (which is used also in our calculations), the current density can be calculated accurately even on a coarse discretization grid, virtually ignoring "layers", i.e., thin regions of rapid variations of the potential and the carrier densities, which physically correspond to junctions between differently doped regions of the device. However, to solve the continuity equations and Poisson's equation accurately in the neighborhood of junctions, a much finer grid is required. Therefore, the distributions of the potential and carrier densities (hence, also the net charge) indicate a "small-scale character" near the junctions. The presented additive decomposition approach revealed also this feature of the semiconductor drift-diffusion equations.

In both cases, i.e., forward bias and reverse bias of the *p-i-n* diode, the additive decomposition results agreed quite well with the results obtained with traditional one-grid approach, while the AD method allows for potentially significant reduction of matrix computation and computer memory requirements. Although in this case the sections with small-scale grid were calculated sequentially, the code has been written as if the sections were processed in parallel. Hence, the possible parallel implementation of AD has been verified. Extensive parallelization studies have been conducted for 1-D, 2-D and 3-D Navier-Stokes equations solved with this general approach by McDonough and co-workers [12–14].

CONCLUSION

The main advantage of the new method is that the matrix equations for sections with small-scale points can be solved independently, which allows for parallel computations and smaller memory requirements. Furthermore, computations for few large-scale points are very fast. After the large-scale solution is computed an initial guess for the small-scale solution may be updated using interpolation of large-scale solution, which also leads to more efficient computation.

After successful implementation and testing of the AD method for the drift-diffusion equations, the authors intend to apply the new approach to the hydrodynamic semiconductor model, for which the AD technique should be also suitable.

References

[1] McDonough, J. M. and Bywater, R. J., "Large-Scale Effects on Local Small-Scale Chaotic Solutions to Burgers' Equation", *AIAA Journal*, **24**(12), 1924–1930, December 1986.

[2] McDonough, J. M. and Bywater, R. J., "Turbulent Solutions from an Unaveraged, Additive Decomposition of Burgers' Equation", *The Third Joint ASCE/ASME Mechanics Conference*, University of California, San Diego, La Jolla, California, 7–11, July 9–12, 1989.

[3] McDonough, J. M. and Saito, K. (1994). "Local Small-Scale Interaction of Turbulence with Chemical Reactions in H_2–O_2 Combustion", *Fire Science and Technology*, **14**(1–2), 1–18.

[4] Selberherr, S. (1984). *Analysis and Simulation of Semiconductor Devices*, Vienna – New York: Springer-Verlag.

[5] Brown, R., Perry, P. and Shen, Z., "The Additive Decomposition for the Two-Dimensional Incompressible Navier-Stokes Equations: Convergence Theorems and

error Estimates," accepted by *SIAM Journal on Mathematical Analysis*.
[6] Turowski, M., Grecki, M. and Napieralski, A., "PASSC – Package for 2-D Analysis of Semiconductor Structures in Circuits", *XIII Symposium "Electromagnetic Phenomena in Nonlinear Circuits"*, Poznan, Poland, pp. 189–194, May 25–28, 1994.
[7] Turowski, M. and Napieralski, A., "Two-dimensional analysis of GTO switching under the influence of external circuit", *IEE Proc.-Circuits, Devices and Systems*, **141**(6), 483–488, December 1994.
[8] Turowski, M. (1995). "Educational Software for the Analysis of the Behaviour of Power Semiconductor Devices", Chapter 10 in monographic book: *Computers in Electrical Engineering Education – Research, Development and Application*, pp. 153–165, Monash Engineering Education Series, EEERG-USICEE, Melbourne, Australia.
[9] Brauer, E. J., Turowski, M. and McDonough, J. M., "Additive Decomposition of the Drift-Diffusion Model", *Fifth International Workshop on Computational Electronics, IWCE-5*, Notre Dame, Indiana, USA, May 28–30, 1997.
[10] Markowich, P. A., Ringhofer, C. A., Selberherr, S. and Lentini, M. (1983). "A Singular Perturbation Approach for the Analysis of the Fundamental Semiconductor Equations", *IEEE Trans. Electron Devices*, **ED-30**(9), 1165–1180.
[11] Scharfetter, D. L. and Gummel, H. K. (1969). "Large scale analysis of a silicon Read diode oscillator", *IEEE Trans. Electron Devices*, **ED-16**(1), 64–77.
[12] McDonough, J. M. and Wang, D. (1995). "Additive Turbulent Decomposition: A Highly Parallelizable Turbulence Simulation Technique", in *Parallel Computational Fluid Dynamics*, Satofuka *et al.* (eds.), Elsevier Science B.V., pp. 129–136.
[13] Mukerji, S. and McDonough, J. M. (1996). "Parallel Computation of 3-D Small-Scale Turbulence Via Additive Turbulent Decomposition", in *Parallel Computational Fluid Dynamics*, Ecer *et al.* (eds.), Elsevier Science B.V., pp. 465–472.
[14] McDonough, J. M. (1997). "Parallelization of ATD/Chaotic Map Turbulence Models", in *Parallel Computational Fluid Dynamics*, Schiano *et al.* (eds.), Elsevier Science B.V., pp. 92–99.

Authors' Biographies

Elizabeth J. Brauer received the Ph.D. degree from University of Illinois at Urbana-Champaign. She worked at Motorola, Inc., National Semiconductor, Inc. and the University of Kentucky. Presently, she is the Director of the Advanced Microelectronics Laboratory in the Department of Electrical Engineering at Northern Arizona University in Flagstaff, AZ. Her research interests involve CAD, verification, and testing of integrated circuits, modeling of semiconductor devices, and neural networks.

Marek Turowski received the M.Sc. and Ph.D. degrees from the Technical University of Lodz, Poland, in 1983 and 1992, respectively. Since 1983 he has been with the Technical University of Lodz. Currently he is a post-doc scholar at the University of Kentucky in Lexington, USA. In 1992/93 he was for one year at the University of Sydney, Australia. His research interests include numerical methods for 2-D and 3-D modeling and simulation of semiconductor devices, electromagnetic fields, and thermal phenomena in electronic devices. He is author or co-author of over 50 papers in international journals and conference proceedings.

James M. McDonough is an Associate Professor in the Department of Mechanical Engineering at the University of Kentucky. He received a BS in Aero-Astro Engineering from the Ohio State University (1968) and MA, Applied Mathematics (1975) and Ph.D., Engineering (1980) degrees from the University of California, Los Angeles. He has nearly 15 years of industrial experience, primarily in the aerospace industry, in addition to his academic experience. His research interests span a broad spectrum of topics in numerical computation, CFD and nonlinear chaotic dynamical systems. He has published more than 60 papers in archival journals and conference proceedings.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 401–405
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Monte Carlo Simulations of High Field Transport in Electroluminescent Devices

MANFRED DÜR[a,*], STEPHEN M. GOODNICK[a], MARTIN REIGROTZKI[b], and RONALD REDMER[b]

[a] *Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287-5706, USA;*
[b] *Universität Rostock, Fachbereich Physik, Universitätsplatz 3, D-18051 Rostock, Germany*

High field transport in phosphor materials is an essential element of thin film electroluminescent device performance. Due to the high accelerating fields in these structures (1–3 MV/cm), a complete description of transport under high field conditions utilizing information on the full band structure of the material is critical to understand the light emission process due to impact excitation of luminescent impurities. Here we investigate the role of band structure for ZnS, GaN, and SrS based on empirical pseudopotential calculations to study its effect on the high field energy distribution of conduction band electrons.

Keywords: High field transport, Monte Carlo method, ZnS, GaN, SrS

INTRODUCTION

Over the past ten years, thin film electroluminescent (TFEL) devices have become an important technology for flat-panel display applications [1]. A TFEL device essentially consists of a thin layer of a phosphor material sandwiched between two insulating layers which are contacted by electrodes. Wide band gap semiconductors such as ZnS and SrS heavily doped with a luminescent impurity are currently the most important commercial phosphor materials. Above a critical potential applied to the electrodes of a ZnS based TFEL device, electrons are injected into the phosphor layer from deep trap states at the phosphor-insulator interface. For sufficiently high phosphor fields, the electrons gain kinetic energy large enough to initiate impact excitation of the luminescent impurities such as Mn^{2+} to excited levels. Visible light is emitted when the impurities return back to the ground state. SrS doped with Ce^{3+} has recently demonstrated good efficiency as a blue phosphor material in TFEL applications, which is critical to realizing full color (RGB) displays. Typically, the turn on fields are higher than those of ZnS devices, which may be

* Corresponding author.

associated with the high field transport properties of this material. GaN is another wide band gap material with a similar band structure to ZnS, which is currently of interest for high temperature electronics and visible optical sources. However, the efficiency of GaN phosphors in TFEL devices is poor, which is not understood at this time.

The purpose of the present work is to investigate the relative high field transport properties of zincblende-type ZnS and GaN as well as SrS based on band structure features by Monte Carlo methods. In order to better understand the operation of electroluminescent devices, it is critical to know to which extent the TFEL performance is determined by the transport in the phosphor material.

MONTE CARLO MODEL

The ensemble Monte Carlo code presented in this paper includes the full band structure information of the phosphor material and all the pertinent scattering mechanisms to solve the semi-classical Boltzmann transport equation under high field conditions. Our approach basically follows the model developed by the Illinois group, which is described in detail elsewhere [2].

The standard empirical pseudopotential method (EPM) is employed to obtain the band structure for ZnS, GaN, and SrS. In the case of ZnS, we have extended the local EPM calculation reported by Walter and Cohen [3] and have included also nonlocal corrections in the pseudopotential [4]. The new nonlocal parameter set was determined by comparing to optical data for this material [5]. For GaN, the local EPM parameters published by Fan *et al.* [6] have been used in the calculation. Their form factors were obtained by adjusting the band structure results along directions of high symmetry to results based on *ab initio* techniques and to experimental data for band gaps and effective masses. In the case of SrS, we have derived a nonlocal EPM band structure, since a reliable EPM calculation is not yet available in the literature. As a starting point, we have utilized the atomic pseudopotentials for Sr and S, and combined them using the method suggested by Harrison [7] to give a first estimate of the band structure. The resulting energy dispersion gives a direct gap at the Γ point. Relying on various first principles calculations [8,9], SrS is believed to be an indirect semiconductor with an indirect gap close to 4 eV occuring at the X point. We have therefore added nonlocal corrections to the local form factors to accomplish the transition to an indirect band gap material. Other than the energy band gaps at the X and Γ point [10], we are aware of no experimental band structure data for this material. Thus, the accuracy of the calculated energy dispersion is at present unknown. Furthermore, the number of 181 plane wave states employed in the expansion of the crystal wave functions might not be large enough in the case of SrS to ensure convergence of the energy levels due to its pronounced ionic character [11]. Figure 1 shows a comparison of the DOS for ZnS, GaN, and SrS. As can be seen, the DOS for SrS is dramatically larger than the DOS for ZnS and GaN.

The scattering rates for various mechanisms in the semiconductor material are calculated based on first-order time dependent perturbation theory. For energies below higher valley minima, we use a

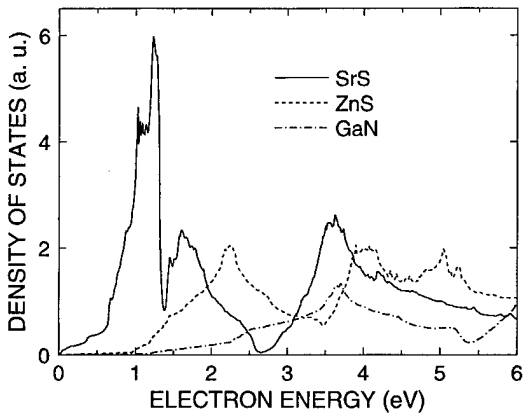


FIGURE 1 Comparison of the density of states for ZnS, GaN and SrS.

nonparabolic band model to compute the scattering rates in the lowest lying valley [12]. The effective mass and nonparabolicity parameter of the valley entering into the scattering rates were extracted directly from the EPM calculation of the band structure. The scattering mechanisms taken into account include scattering due to polar optical phonons, acoustic and optical phonons via the deformation potential interaction and ionized impurities. At energies above higher valley thresholds, electron scattering is dominated by the optical deformation potential mechanism. To lowest order, this mechanism is isotropic in nature and depends only on the density of final states. Therefore, the DOS shown in Figure 1 is used directly to determine the scattering rate at high energy. Reported measurements of the optical deformation potentials for inter- and intravalley scattering in ZnS, GaN, or SrS do not exist to our knowledge. For this reason, we treat the deformation potentials as adjustable parameters by fitting to carrier multiplication data, when available.

An inevitable scattering mechanism for the stabilization of the electron distribution function at high electric fields is band-to-band impact ionization of electron-hole pairs by energetic electrons. Here, we use an energy dependent impact ionization rate averaged over all directions in wave vector space, calculated numerically from the full band structure of the wide band gap semiconductor and well represented by a power law [5], [13]. In our Monte Carlo simulations, this fit formula is incorporated into the collision term of the Boltzmann equation.

RESULTS AND DISCUSSION

In the ensemble Monte Carlo simulations presented here, we have included the first four conduction bands of each semiconductor studied. The full band structure is represented using 916 points in the irreducible wedge of the first Brillouin zone. Typically, an ensemble of 96,000 particles is followed in order to achieve sufficient accuracy in simulating rare events such as impact ionization and impact excitation.

Although systematic experimental studies of high field transport in wide band gap semiconductors such as ZnS, GaN, or SrS are currently not available, information regarding the high field distribution may be indirectly deduced from measurement of carrier multiplication due to band-to-band impact ionization. Such measurements were performed for ZnS by Thompson and Allen [14]. We have used a set of two optical deformation potentials, one characterizing the lowest conduction band, and the other characterizing the upper bands, to bring our simulated results for the impact ionization coefficient versus field into agreement with Thompson and Allen's data. The details of this choice of proper deformation potentials have been discussed in Reference [5]. We are not aware of any carrier multiplication measurement for GaN and SrS. Owing to this lack of information, values of deformation potentials identical to those in ZnS have been assumed in the Monte Carlo simulations. By assuming the same deformation potentials for each material, we are able to make a comparison of the high field transport properties of these three materials based on the band structure itself.

Figure 2 shows the simulated steady-state particle distribution (i.e., the number of particles in a given energy range) in ZnS, GaN, and SrS for an electric field of 2 MV/cm at 300 K. For Mn^{2+} impurities in ZnS with an excitation threshold energy of approximately 2.1 eV, the comparison of ZnS to GaN indicates that there are more electrons above the impact excitation threshold due to the lower DOS for GaN, and significantly fewer electrons in the case of SrS due to its high DOS. The secondary peak in the distribution for ZnS at about 4 eV is directly related to the minimum in the DOS shown in Figure 1. Since a lower DOS translates to a reduced optical deformation potential scattering rate, the trends in the DOS are clearly reflected in the simulated high field distributions, provided that the assumption of identical deformation potentials is not too

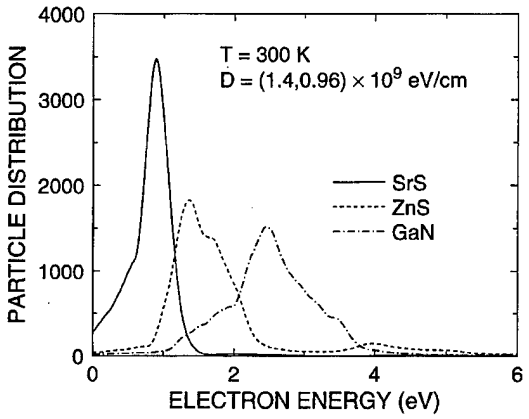


FIGURE 2 Simulated particle distribution in ZnS, GaN and SrS for an electric field of 2 MV/cm at a lattice temperature of 300 K assuming identical deformation potentials.

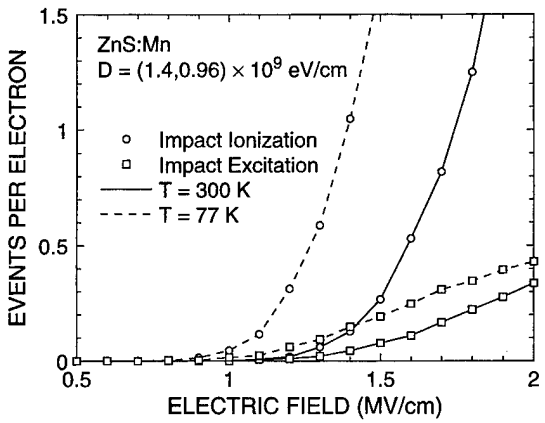


FIGURE 3 Simulated impact ionization and impact excitation yields for various electric fields in a ZnS layer of 0.5 μm thickness doped with Mn. Shown are the results for two different lattice temperatures.

unrealistic. Our results seem to suggest that the high field properties of GaN would be superior to those of ZnS in TFEL device operation, and that SrS has less favorable transport properties. High field transport in SrS might improve at higher fields because of the tendency towards reduced DOS with increasing kinetic energy, as can be seen from Figure 1. The uncertainty with regard to the operation of a SrS device is how much of a role is played by space charge and how large is the cathode field under device operation. There is experimental evidence that the cathode field may be significantly larger than the measured, average field due to large amounts of space charge [15]. These effects have been ignored in the present work.

Figure 3 shows the impact excitation yield and impact ionization yield for electrons traversing a ZnS layer of 0.5 μm thickness for various electric fields. The impact excitation rate of Mn luminescent impurities in ZnS used in the Monte Carlo simulations has been calculated from the transition rate derived by Bringuier [16] based on the exchange scattering process. For Mn impurities embedded in ZnS, exchange scattering is known to be the dominant excitation path. We have performed the high field simulations for two different lattice temperatures: $T = 300$ K and $T = 77$ K. For both temperatures, the impact ionization yield is larger than the corresponding excitation yield for a density of 10^{20} cm^3 Mn atoms. The comparison reveals that impact excitation as well as impact ionization increases with decreasing lattice temperature since the scattering of electrons due to phonons is reduced. At the lower temperature, the electron system reaches a higher average energy in the steady state so that more electrons are capable to impact excite luminescent impurities or to generate electron-hole pairs by band-to-band impact ionization.

CONCLUSIONS

We have simulated high field transport in ZnS, GaN, and SrS including the full band structure of these materials in a Monte Carlo simulation. The simulated steady-state distributions of electrons at typical phosphor fields reveal good transport properties of ZnS and GaN. High field transport in SrS appears to be less favorable, although space charge effects might be very important. We find good agreement in the expected trends of the impact excitation yield with temperature in ZnS

based TFEL devices. The optical deformation potential is the greatest uncertainty in the present Monte Carlo model, which suggests further investigation of the electron-phonon coupling in phosphor materials is needed in order to understand their high field behavior.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft under Grants No. Rė 882/6-1 and No. Scha 360/8-1, by the U.S. Army Research Office under Contract No. DAAH04-94-G-0324, and by the Advanced Research Projects Agency under the Phosphor Technology Center of Excellence, Grant No. MDA 972-93-1-0030.

References

[1] Ono, Y. A. (1995). *Electroluminescent Displays* (World Scientific, Singapore).
[2] Hess, K., Ed. (1991). *Monte Carlo Device Simulation: Full Band and Beyond* (Kluwer Academic Publishers, Boston).
[3] Walter, J. P. and Cohen, M. L. (1969). "Calculation of the Reflectivity, Modulated Reflectivity, and Band Structure of GaAs, GaP, ZnSe and ZnS", *Phys. Rev.*, **183**, 763.
[4] Chelikowsky, J. R. and Cohen, M. L. (1976). "Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors", *Phys. Rev. B*, **14**(2), 556.
[5] Reigrotzki, M., Redmer, R., Lee, I., Pennathur, S. S., Dür, M., Wager, J. F., Goodnick, S. M., Vogl, P., Eckstein, H. and Schattke, W. (1996). "Impact ionization rate and high-field transport in ZnS with nonlocal band structure", *J. Appl. Phys.*, **80**(9), 5054.
[6] Fan, W. J., Li, M. F., Chong, T. C., and Xia, J. B. (1996). "Electronic properties of zinc-blende GaN, AlN, and their alloys $Ga_{1-x}Al_xN$", *J. Appl. Phys.*, **79**(1), 188.
[7] Harrison, W. A. (1996). *Pseudopotentials in the Theory of Metals* (Benjamin, New York).
[8] Hasegawa, A. and Yanase, A. (1980). "Electronic structure of Sr monochalcogenides", *J. Phys. C: Solid St. Phys.*, **13**, 1995.
[9] Stepanyuk, V. S., Szasz, A., Farberovich, O. V., Grigorenko, A. A., Kozlov, A. V. and Mikhailin, V. V. (1989). "An Electronic Band Structure Calculation and the Optical Properties of Alkaline-Earth Sulphides", *Phys. Stat. Sol. B*, **155**, 215.
[10] Kaneko, Y. and Koda, T. (1988). "New developments in IIa–VIb (alkaline-earth chalcogenide) binary semiconductors," *J. Cryst. Growth*, **86**, 72.
[11] Ferhat, M., Zaoui, A., Certier, M. and Khelifa, B. (1996.) "Applicability of the empirical pseudopotential method to semiconductors with d valence electrons", *Phys. Lett. A*, **216**, 187.
[12] Bhattacharyya, K., Goodnick, S. M. and Wager, J. F. (1993). "Monte Carlo simulation of electron transport in alternating-current thin-film electroluminescent devices", *J. Appl. Phys.*, **73**(7), 3390.
[13] Reigrotzki, M., Stobbe, M., Redmer, R. and Schattke, W. (1995). "Impact ionization rate in ZnS", *Phys. Rev. B*, **52**(3), 1456.
[14] Thompson, T. D. and Allen, J. W. (1987). "Band-to-band impact ionisation rates in ZnS", *J. Phys. C: Solid State Phys.*, **20**, L499.
[15] Wager, J. F., private communication.
[16] Bringuier, E. and Bhattacharyya, K. (1995). "Hot-electron impact cross sections of impurities in semiconductors", *Semicond. Sci. Technol.*, **10**, 1065.

Authors' Biographies

Manfred Dür received the M.S. and Ph.D. degrees in physics from the University of Innsbruck, Austria, in 1990 and 1995, respectively. He is currently working as a Faculty Research Associate at the Department of Electrical Engineering at Arizona State University. His research interests include high field transport in electroluminescent materials and relaxation processes in quantum confined systems.

Stephen M. Goodnick received the Ph.D. degree in electrical engineering from Colorado State University in 1983. Currently, he is Professor and Chair in the Electrical Engineering Department at Arizona State University. His research interests include semiconductor transport, quantum and nanostructure devices and device technology, and high frequency devices.

Martin Reigrotzki studied physics at the Swiss Federal Institute of Technology in Zürich where he received his diploma in theoretical physics in 1994. Since 1994, he is working as a Ph.D. student with Professor Redmer at the University of Rostock on high field electron transport in semiconductors.

Ronald Redmer received his Ph.D. in theoretical physics in 1986 and was awarded the habilitation grade in 1991, both from the University of Rostock. He is now Professor for theoretical physics at the University of Rostock. His primary interests concern the high field electron transport in semiconductors and the properties of dense plasmas and expanded fluids.

VLSI DESIGN
1998, Vol. 8, Nos. (1-4), pp. 407-412
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Modeling of Radiation Fields in a Sub-Picosecond Photo-Conducting System

K. A. REMLEY[a], A. WEISSHAAR[a], V. K. TRIPATHI[a] and S. M. GOODNICK[b,*]

[a] *Department of Electrical and Computer Engineering, Oregon State University, Corvallis, Oregon, USA 97331;*
[b] *Department of Electrical Engineering, Arizona State University, Tempe, Arizona, USA 85287*

FDTD and Monte Carlo methods are combined to simulate the terahertz radiation from a coplanar photoconducting structure. The simulation tool under consideration allows calculation of potentials, particle distributions, current densities, and the near field electromagnetic fields anywhere in the computational domain. To model the far field radiation, it is not efficient nor, in many cases, physically possible at present to use the FDTD technique directly because of the excessive computational burden. Techniques are discussed for modeling both the near field and the far field radiation. Computational results showing the far field radiation are in qualitative agreement with published experimental results.

Keywords: Terahertz radiation, subpicosecond systems, photo-conducting systems, FDTD, time domain analysis, computational modeling, electromagnetic radiation, far-field radiation, near-to-far-field transformation, Hertzian dipole

INTRODUCTION

Characterization of the electromagnetic radiation arising from ultrafast electronic structures whose dimensions approach the wavelength of operation is a subject of continued research. The design of terahertz radiators [1, 2] as well as applications such as electrooptic sampling [3], switching [4], and terahertz spectroscopy [5] all require an accurate model of the radiated electromagnetic field. In some simple cases the radiated field may be calculated directly or approximated from a known excitation. Alternatively, numerical techniques such as the finite-difference time domain (FDTD) method allow computation of the fields radiated from structures with arbitrary physical geometry or material composition. For an accurate portrait of the radiated fields, the open region simulation space used in the FDTD should be terminated with absorbing boundary conditions (ABCs).

* Corresponding author.

ABCs eliminate reflections at the edges of the computational domain, allowing simulation of propagation in infinite free space.

The computational burden of the FDTD method increases with the dimensions of the physical simulation space, and thus becomes less efficient for calculations far from the source. The FDTD also suffers from numerical grid dispersion for large computational domains. Additionally, the larger the simulation space, the larger the computational burden involved in the ABC calculations. It is, therefore, desirable to use alternate techniques for determination of the far field radiation.

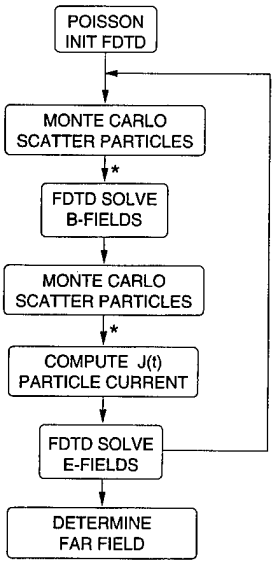


FIGURE 1 Computer algorithm used in the present work.

FORMULATION

Simulation Tool

The present work utilizes a simulation tool which has been developed at Oregon State University [6]. The tool couples a Monte Carlo particle simulator to either a Poisson solver or an FDTD electromagnetic solver to determine device characteristics such as potential, particle distribution, and current density anywhere in the computational domain. In the present case, the FDTD solver is used, enabling determination of the electromagnetic (EM) fields throughout the simulation space. The Monte Carlo simulator and the FDTD solver are coupled together in a leap frog manner, allowing continual updates of both the particle distribution and the EM fields, see Figure 1.

The Monte Carlo particle simulator allows stochastic solution to the Boltzmann transport equation for particle motion. Particle motion is modeled as a series of free flights subject to magnetic and electric forces. Motion is terminated by instantaneous, random scattering events. Random scattering times are generated using a random number generator and the calculated quantum mechanical scattering cross-section. Important scattering mechanisms for GaAs are included in a three band model.

The FDTD scheme used in the present work allows modeling of complex, three dimensional structures. Centered differences are used in time and space, utilizing the Yee cell formulation [7]. This technique allows broadband response predictions and access to field components throughout the computational domain. The absorbing boundary conditions used in the simulations space are Berenger's Perfectly Matched Layer (PML) [8]. Impedance matching to a highly lossy medium is accomplished by utilizing electric and magnetic loss terms to attenuate only the outgoing portion of the wave. In the PML, absorption is theoretically independent of frequency and angle of incidence [8].

Photo-Conducting Experiment

The radiation arising from a photo-conducting system is modeled using the simulation tool described above. In this system, a sub-picosecond laser pulse incident on a GaAs substrate creates electron-hole pairs, as shown in Figure 2. Biased electrodes on either side of the pulsed area cause the electrons and holes to migrate away from each other, creating a dipole moment. A time-varying current density is induced in the GaAs substrate.

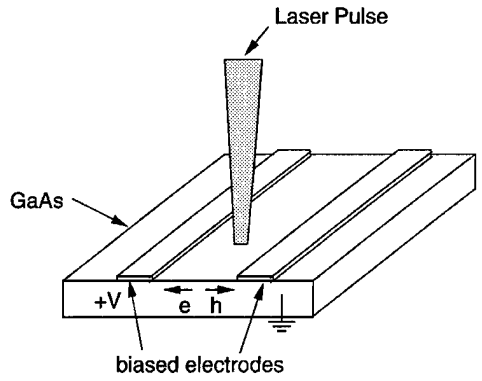


FIGURE 2 Typical structure for an electrooptic sampling experiment.

The rise time of the current pulse is extremely fast, on the order of a few hundred femtoseconds, as illustrated in Figure 3a. This current density is the source of the EM radiation in the photo-conducting experiment. Once the excitation is removed, the biased electrodes continue to attract the electrons and holes, therefore recombination occurs over a timescale much longer than the simulation time.

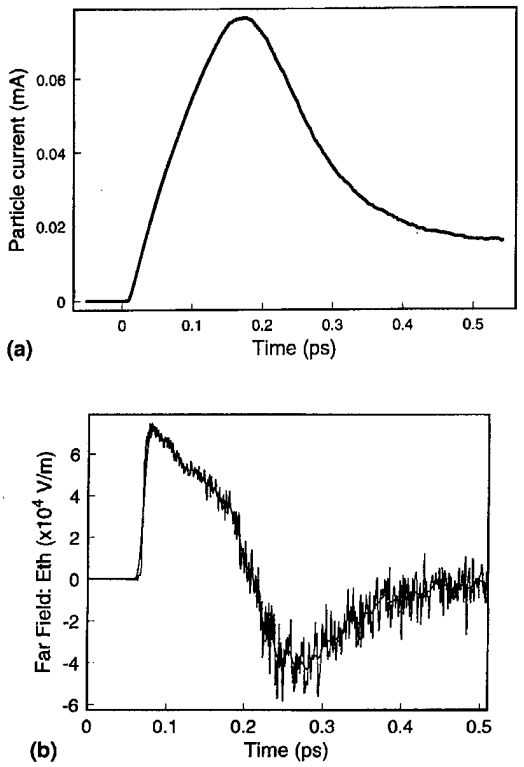


FIGURE 3 (a) Current density in the GaAs substrate for an injection rate of 1e15/cm^3 and bias of 40 volts. (b) Numerically smoothed time derivative of the current pulse shown in (a).

RESULTS

Figures 4a and 5a show the near field radiation along a surface above and parallel to the GaAs substrate of the structure shown in Figure 2. These fields were calculated using the simulation tool described above, with a bias of 40 volts on the metal electrodes, and an injection rate of 1e15/cm^3. In this simulation it was assumed that the laser pulse had a Gaussian wavefront with beam radius standard deviation of 2 μm, a Gaussian-distributed energy variation between 1.42 eV and 1.62 eV (larger than the bandgap of GaAs), and a pulse duration of approximately 20 fs.

If the source is modeled as a Hertzian dipole, i.e., a dipole with negligible length, the radiation may be found using the well-known equations for an infinitesimal dipole, repeated, for convenience, below. See e.g. [9], for the derivation.

$$\vec{E}(R,t) = \hat{a}_R\left[\frac{2\cos\theta}{4\pi\varepsilon_0}\left(\int_t \frac{\int_v J\left(\tau-\frac{R}{c}\right)dv'}{R^3}d\tau + \frac{\int_v J\left(t-\frac{R}{c}\right)dv'}{cR^2}\right)\right] - \hat{a}_\theta\left[\frac{\sin\theta}{4\pi\varepsilon_0}\left(\int_t \frac{\int_v J\left(\tau-\frac{R}{c}\right)dv'}{R^3}d\tau - \frac{\int_v J\left(t-\frac{R}{c}\right)dv'}{cR^2} + \frac{\frac{\partial}{\partial t}\int_v J\left(t-\frac{R}{c}\right)dv'}{c^2R}\right)\right] \quad (1)$$

Here R is the distance from the dipole to the far field observation point, J is the volume current density induced by the dipole, and θ is measured off the dipole axis. Figures 4b and 5b show the fields from an ideal dipole for comparison to that of the near field radiation found using the FDTD

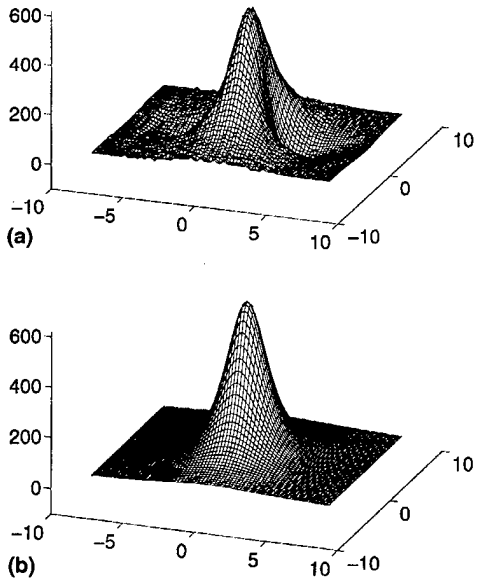


FIGURE 4 E_θ: (a) Near field radiation from the photo-conducting structure found using FDTD. (b) Near field radiation from an ideal Hertzian dipole.

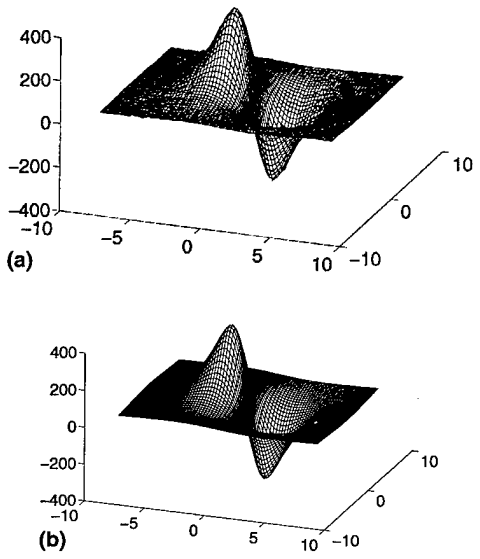


FIGURE 5 E_R: (a) Near field radiation from the photo-conducting structure found using FDTD. (b) Near field radiation from an ideal Hertzian dipole.

simulation in Figures 4a and 5a. The effects of the various material parameters in the simulation space may be seen in Figure 4a. Specifically, the GaAs/air interface causes some variation in the field distribution near the center of the surface.

To model the far field radiation, only terms which decrease as $1/R$ are retained from Eq. (1). In this case, the electric field is given by:

$$E_\theta(R,t) = \frac{\sin\theta}{4\pi\varepsilon_0 Rc^2}\frac{\partial}{\partial t}\int\int\int_V J_V\, dV' \qquad (2)$$

The far field radiation is proportional to the time derivative of the excitation. Eq. (2) is often used to approximate the far field radiation arising from photoconducting antennas and arrays [1]. However, (2) assumes a dipole source of negligible length and free space propagation. The use of (2) in the photo-conducting experiment also neglects the effects of the GaAs substrate and metal electrodes. In addition, numerical implementation of the time differentiation emphasizes the high frequency noise components. Figures 3a and 3b show a typical current pulse and two numerically smoothed versions of its derivative with respect to time. Smoothing is accomplished using a moving average filter. Figure 3b compares to measured results such as [10].

It is expected that a more accurate representation of the far field radiation would be given with a near-to-far field transformation. This technique, which has been utilized by microwave and RF engineers for the analysis of radiation from, for example, horn antennas [11], utilizes an equivalent source methodology. The fields found on a closed virtual surface surrounding the source are used as new sources of radiation, replacing the original source, as illustrated in Figure 6. In this technique, the near field radiation arising from a complex structure may be calculated using the FDTD method, and the far field radiation may be found using the equivalent source in a homogeneous external problem space. Several formulations for an equivalent source representation exist, including vector forms such as the Schelkunoff equivalence theorem [12], the Stratton-Chu formulation [13], the Franz formulation [14], and the scalar Kirchhoff Surface Integral Representation, de-

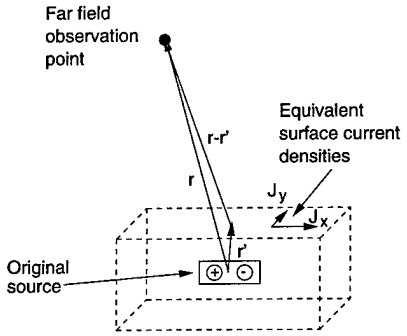


FIGURE 6 Equivalent current source model.

scribed, for example, in [15]. All provide a relationship between the electromagnetic fields on a surface surrounding a source and the field at a point outside the surface. The field produced by the equivalent source at the observation point will be the same as that produced by the original source. The Kirchhoff formulation would be the preferred technique in modeling the photo-conducting structure because the technique requires only the values of the electric fields on the equivalent surface. This is an advantage when used with the FDTD since the electric and magnetic fields are offset spatially by one half cell [16, 17]. Additionally, since this is a scalar formulation, each field component may be calculated independently, allowing placement of the virtual surfaces to coincide with each field component's placement in the Yee cell. Application of this technique to the photoconducting structure is currently being implemented by the present authors.

CONCLUSION

Techniques for describing and modeling the terahertz radiation from a photo-conducting structure have been discussed. The physical source of radiation for laser excitation on a GaAs substrate was described, and an approximation of the source in terms of an ideal Hertzian dipole was presented. A simulation tool which couples a Monte Carlo particle simulator with an FDTD electromagnetic solver was used to numerically determine the near field radiation. Computational results showing near field radiation patterns were shown. Techniques for approximating the far field radiation were discussed. Simulation results for a technique based on calculation of the derivative of the excitation were presented.

Acknowledgement

This work is supported by grant #ECS-9312240 from the National Science Foundation. K. A. Remley is also supported under a National Science Foundation Graduate Research Fellowship.

References

[1] Froberg, N., Hu, B., Zhang, X. and Auston, D. H., "Terahertz radiation from a photoconducting antenna array", *IEEE J. Quantum Electron.*, **28**, 2291–2301, Oct. 1992.

[2] Smith, P. R., Auston, D. H. and Nuss, M., "Subpicosecond photoconducting dipole antennas", *IEEE J. Quantum Electron.*, **24**, 255–260, Feb. 1988.

[3] Valdmanis, J. and Mourou, G., "Subpicosecond electro-optic sampling: principles and applications", *IEEE J. Quantum Electron.*, QE-**22**, 69–78, Jan. 1986.

[4] El-Ghazaly, S. M., Joshi, R. P. and Grondin, R. O., "Electromagnetic and transport considerations in subpicosecond photoconductive switch model", *IEEE Trans. Microwave Theory Tech.*, **38**, 629–637, May 1990.

[5] Greene, B., Federici, J., Dykaar, D., Levi, A. and Pfeiffer, L., "Picosecond pump and probe spectroscopy utilizing freely propagating terahertz radiation", *Opt. Lett.*, **16**, 48–49, Jan. 1991.

[6] Goodnick, S., Pennathur, S., Ranawake, U., Lenders, P. and Tripathi, V. (1995). "Parallel implementation of a Monte Carlo particle simulation coupled to Maxwell's equations", *Intl. J. of Num. Modelling*, **8**, pp. 205–219.

[7] Yee, K., "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media", *IEEE Trans. Antennas Propagat.*, AP-**14**, 302–307, Dec. 1966.

[8] Berenger, J. P. (1994). "A perfectly matched layer for the absorption of electromagnetic waves",*J. Computational Physics*, **114**, 185–200.

[9] Franceschetti, G. and Papas, C. H., "Pulsed Antennas", *IEEE Trans. Antennas Propagat.*, AP-**22**, 651–661, Sept. 1974.

[10] Son, J., Sha, W., Norris, T., Whitaker, J. and Mourou, G., "Transient velocity overshoot dynamics in GaAs for electric fields >200 kV/cm", *Appl. Phys. Lett.*, **63**, 923–925, Aug. 1993.

[11] Balanis, C. (1989). *Advanced Engineering Electromagnetics*. New York: Wiley.
[12] Schelkunoff, S. A. (1936). "Some equivalence theorems of electromagnetics and their application to radiation problems", *Bell System Tech. J.*, **15**, 92–112.
[13] Stratton, J. A., and Chu, L. J. (1939). "Diffraction Theory of Electromagnetic Waves", *Phys. Rev.*, **56**, 99–107.
[14] Franz, V. W. (1948). "Zur Formulierung des Huygensschen Prinzips", *Z. Naturforsch. A*, **3a**, 500–506.
[15] Jackson, J. D. (1962). *Classical Electrodynamics*. New York: Wiley.
[16] De Moerloose, J. and Dezutter, D. "Surface Integral Representation Radiation Boundary Condition for the FDTD Method", *IEEE Trans. Antennas Propagat.*, **41**, 890–896, July 1993.
[17] Ramahi, O. "Near- and Far-Field Calculation in FDTD Simulations Using Kirchhoff Surface Integral Representation", *IEEE Trans. Antennas Propagat.*, **45**, 753–759, May 1997.

Authors' Biographies

Kate Remley was born in Ann Arbor, Michigan in December, 1959. She worked as a broadcast engineer in Eugene, Oregon between 1983 and 1992, and was Chief Engineer of an AM/FM broadcast station from 1989–1992. She received the B.S. degree *magna cum laude* from Oregon State University in 1993 and the M.S. degree in 1995. She is currently working toward the Ph.D. degree at Oregon State University as a National Science Foundation Graduate Research Fellow. Her research interests include computational modeling and measurement of guided-wave optical, microwave, and millimeter-wave devices, modeling of quantum waveguide structures, and the theory of semiconductor device fabrication technology.

Andreas Weisshaar received the Diplom-Ingenieur (Dipl.-Ing.) degree in electrical engineering from the University of Stuttgart, Germany, in 1987, and the M.S. and Ph.D. degrees in electrical and computer engineering from Oregon State University in 1986 and 1991, respectively. Since 1992 he has been an Assistant Professor at the Department of Electrical and Computer Engineering at Oregon State University. His current research interests are in the areas of CAD of microwave circuits and components, interconnects and packaging, fiber optic sensors, and wireless communications.

Vijai K. Tripathi received the B.Sc degree from Agra University, India, M.Sc Tech. degree in electronics and radio engineering from Allahabad University, India and the M.S.E.E. and Ph.D. degrees in electrical engineering from the University of Michigan, Ann Arbor, in 1958, 1961, 1964 and 1968, respectively. He is a Professor and Head of Electrical and Computer Engineering at Oregon State University in Corvallis, Oregon. Prior to joining Oregon State in 1974, he had been with the Indian Institute of Technology in Bombay, the University of Michigan in Ann Arbor and the University of Oklahoma in Norman, Oklahoma. Over the years he has been a consultant to many industrial organizations including AVANTEK, EEsof Inc., Teledyne MMIC and Tektronix. His research activities are in the general areas of microwave and distributed parameter circuits and systems, computations electromagnetics, electronic packaging and interconnects.

Stephen M. Goodnick received his B.S. degree in Engineering Science from Trinity University in 1977, and his M.S. and Ph.D. in Electrical Engineering from Colorado State University in 1979 and 1983 respectively. He was an Alexander von Humboldt Fellow at the Technical University of Munich, Germany and the University of Modena, Italy in 1985 and 1986. He was the Melchor visiting chair at the University of Notre Dame in 1991. He was a faculty member from 1986 to 1997 in the Department of Electrical and Computer Engineering at Oregon State University, Corvallis, Oregon, where he held the rank of Professor. He is presently Chair and Professor of Electrical Engineering at Arizona State University.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 413–417
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

New "Irreducible Wedge" for Scattering Rate Calculations in Full-Zone Monte Carlo Simulations

JOHN STANLEY* and NEIL GOLDSMAN

Department of Electrical Engineering, University of Maryland, College Park, MD, 20742

To facilitate the implementation of accurate Brillouin zone integration codes for diamond-type semiconductors such as Si we have constructed a new "irreducible wedge" to be used in conjunction with tetrahedral k-space interpolations. This new wedge, defined by the vertices (000), (100), (110) and $(\frac{1}{2}\frac{1}{2}\frac{1}{2})$ (normalized to $|X| = 1$), spans portions of the first three Brillouin zones and contains the equivalent of 1/48 of a Brillouin zone volume. Simple symmetry arguments may be used to establish the equivalency of this new wedge with the standard irreducible wedge.

Keywords: Monte Carlo, Brillouin zone integration, irreducible wedge

1. INTRODUCTION

Carrier transport investigations based on full-zone band structure require accurate and efficient methods of computing singular integrals over the Brillouin zone. For example, in Monte Carlo simulations, net (integrated) scattering rates and scattering rate densities (in k-space) are needed to determine scattering event occurrences and post-scattering states for carriers [1]. Such integrals arise from Fermi's Golden rule and are of the general form

$$g(\varepsilon(\boldsymbol{k}')) = \int d^3k M(\boldsymbol{k},\boldsymbol{k}')\delta\{\varepsilon(\boldsymbol{k}) - [\varepsilon(\boldsymbol{k}') + \Delta\varepsilon(\boldsymbol{k},\boldsymbol{k}')]\}.$$

In particular, if $\Delta\varepsilon\ (\boldsymbol{k}, \boldsymbol{k}') = 0$, $M(\boldsymbol{k}, \boldsymbol{k}') = 1/4\pi^3$ and $\varepsilon' \equiv \varepsilon\ (\boldsymbol{k}')$, $g(\varepsilon')$ gives density of states [2]. Integrals of this form usually can only be evaluated numerically, particularly if one is using a full-zone band structure computation for the dispersion relation $\varepsilon(\boldsymbol{k})$. In such cases one usually replaces $\varepsilon(\boldsymbol{k})$, $M(\boldsymbol{k}, \boldsymbol{k}')$, and $\Delta\varepsilon(\boldsymbol{k}, \boldsymbol{k}')$ with appropriate piecewise linear (for $M(\boldsymbol{k}, \boldsymbol{k}')$ and $\Delta\varepsilon(\boldsymbol{k}, \boldsymbol{k}')$, often piecewise constant) interpolants defined on a regular mesh in k-space. The underlying k-space mesh, $\{\boldsymbol{k}_n\}_n$, is usually cubic; interpolatory elements are then chosen to be either cubes [1], or tetrahedra (each cube being subdivided into five (equi-edge-length) or six (equi-volume) tetrahedra [3]. Band structure, $\varepsilon(\boldsymbol{k}_n)$, gradients, $\nabla_k\varepsilon(\boldsymbol{k}_n)$, and scattering rates, $S(\boldsymbol{k}_n)$,

*Corresponding author

are then pre-computed and stored for interpolation during Monte Carlo execution. Since the $\varepsilon(\boldsymbol{k})$ relation possesses the (translation and point group) symmetry of the crystal, rather than store data at mesh points throughout the entire first Brillouin zone, it is often expedient to use a much smaller region of k space known as an "irreducible wedge". For Si, the irreducible wedge is 1/48 of the Brillouin zone volume and $\varepsilon(\boldsymbol{k})$ for any $\boldsymbol{k}$ may be computed as $\varepsilon(\boldsymbol{k}_I)$, where $\boldsymbol{k}_I$ is a symmetry-equivalent wavevector in the irreducible wedge.

Density of states and scattering rate integrals are thus reduced to sums of element integrals over the irreducible wedge, each element integral being relatively simple to evaluate. However, for the standard irreducible wedge in common use (described below), element integrals on some boundary planes and edges are cumbersome to evaluate. This occurs because the standard irreducible wedge cannot be precisely filled using a regular[1] cubic or tetrahedral partition; along some boundaries there inevitably will be "overhang" of elements outside the irreducible wedge. In diamond-type semiconductors, elements intersecting [111] faces of the Brillouin zone fall into this category so that special care must be exercised to ensure accurate treatment of these contributions to Brillouin zone integrals.

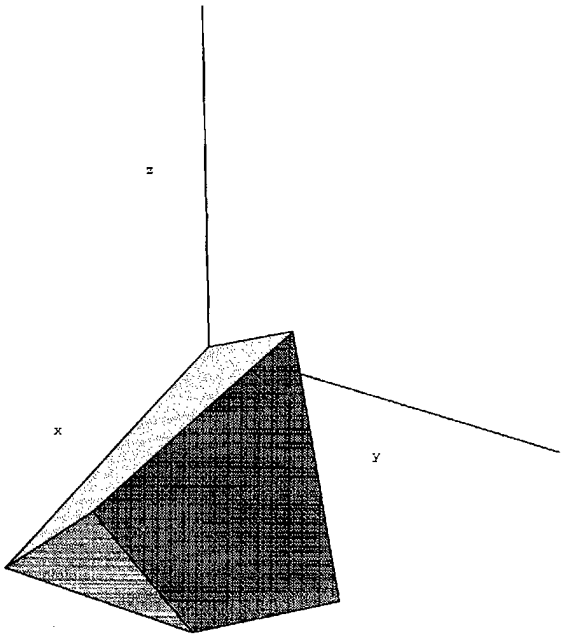


FIGURE 1 Standard irreducible wedge for Si.

2. NEW IRREDUCIBLE WEDGE

We have constructed a new irreducible wedge for use in diamond-type materials, which, compared with the standard irreducible wedge in common use (Fig. 1) considerably simplifies the handling of boundary element contributions to Brillouin zone integrals. This is especially true when the $\varepsilon(\boldsymbol{k})$ argument of the delta function is treated to quadratic order. Our irreducible wedge is depicted in Figure 2. Notice that the standard irreducible wedge lies entirely in the first Brillouin but has a

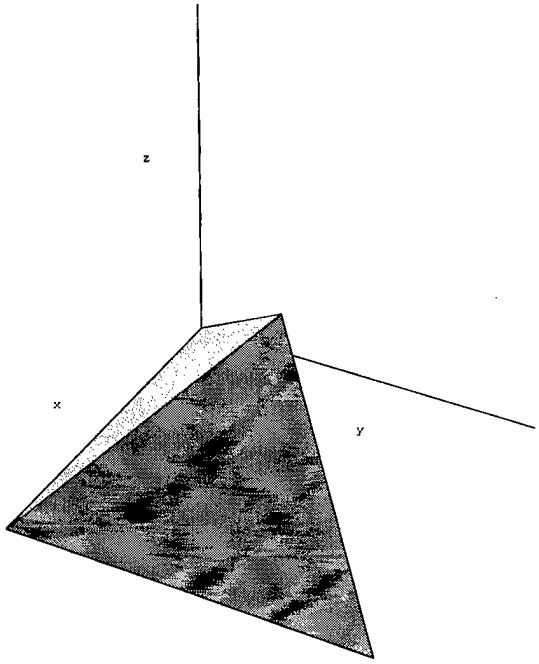


FIGURE 2 New irreducible wedge.

[1]By "regular" we mean identical up to translation and possibly a rotation.

somewhat awkward geometry, while the irreducible wedge shown in Figure 2 (a tetrahedron of the same volume as the standard wedge) has simpler geometry but contains contributions from the first three Brillouin zones.

In contrast to the standard irreducible wedge, this new wedge allows a regular tetrahedral discretization in which there is essentially no distinction between boundary element and interior element contributions to integrals. That is, the new wedge may be precisely filled with a regular array of tetrahedral elements, while for the standard wedge, one inevitably finds elements intersecting the LUWK boundary plane to have portions extending outside the wedge. The presence of such "overhang", while posing no significant difficulties for element interpolations, can have a troublesome impact on the evaluation of integrals. The reason for this is as follows. Assuming a tetrahedral interpolation scheme, first note that evaluating interior element contributions to Brillouin zone integrals is relatively straight-forward (at least for the case where energy is interpolated only to linear order); tetrahedral element integrals reduce to simple surface integrals over planar triangles contained within the element. The difficulty for boundary elements not conforming to the wedge boundary arises from the need to include only that portion of the (constant energy or post-scattering state) surface which lies within the tetrahedron and also within the wedge. Clearly, no such difficulties occur for interior elements or for boundary elements conforming to the wedge boundary.

To illustrate the construction of this new irreducible wedge, refer to Figures 3–6. In Figure 3 we show an octahedron, which, if the vertices are sliced off by [100] planes, gives the first Brillouin zone (Fig. 4). The surface of the second Brillouin zone is shown in Figure 5. Note that both the first and second Brillouin zones are defined entirely by [100] and $[\frac{1}{2}\frac{1}{2}\frac{1}{2}]$ planes.

An examination of the second Brillouin zone reveals, by translation symmetry, that the portions of the octahedron vertices extending beyond the ⟨100⟩ points may equivalently be subdivided into

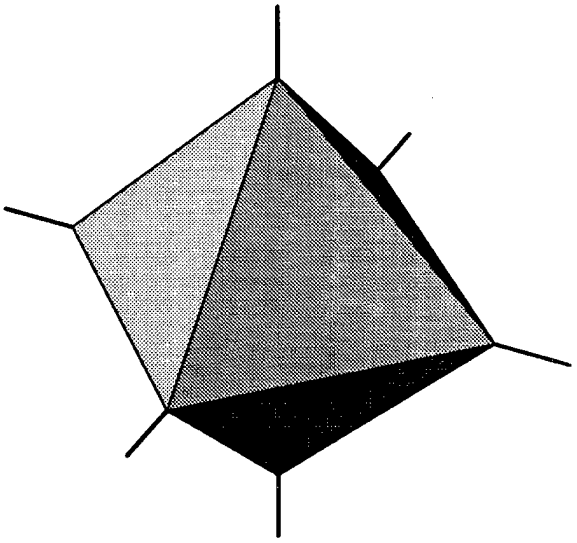


FIGURE 3 An octahedron.

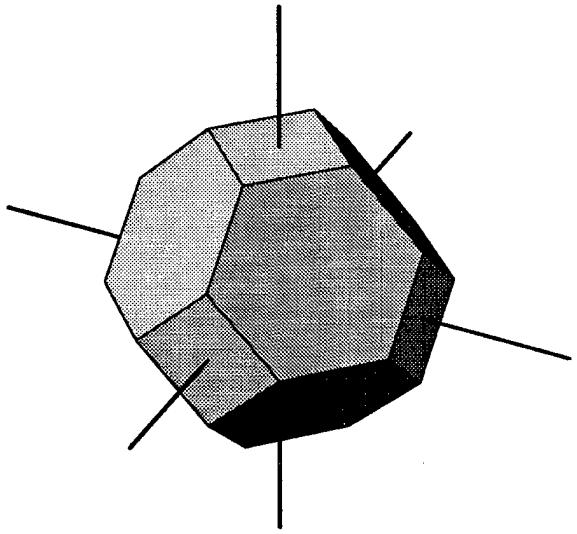


FIGURE 4 1st Brillouin zone of fcc lattice.

four identical tetrahedrons and moved pairwise from the ⟨100⟩ axes to fill the "notches" present at the surface along ⟨110⟩ directions. In carrying out this manipulation one converts the second Brillouin zone into a cube without changing the volume enclosed. The legitimacy of this transformation is readily established by appealing to the underlying space group symmetry. If we now

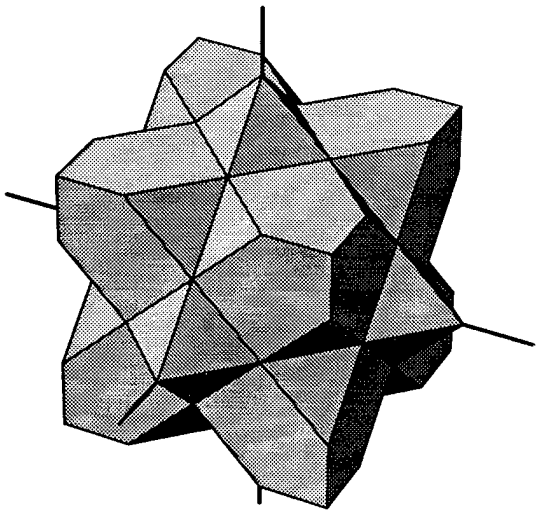


FIGURE 5 Surface of 2nd Brillouin zone of fcc lattice.

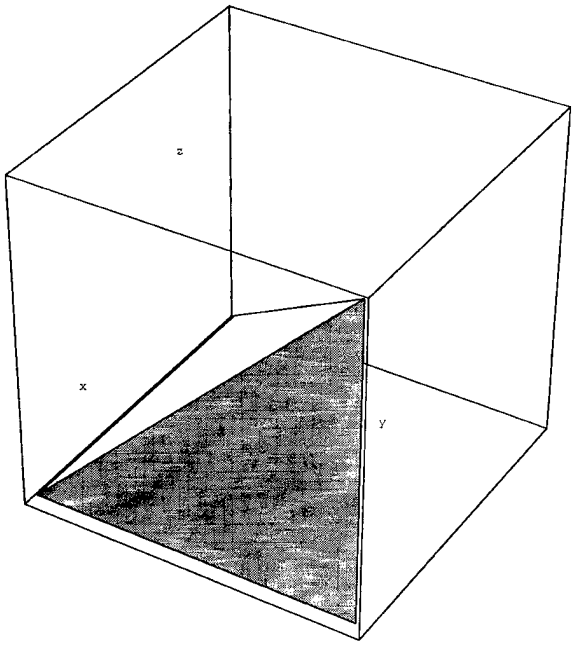


FIGURE 6 Wedge of modified 2nd Brillouin zone.

extend the standard irreducible wedge to pierce the faces of this cube, we obtain the right-tetrahedron shown in Figure 6. This irreducible wedge is exactly twice the volume of the standard irreducible wedge, containing 1/48 of a zone in the first Brillouin zone and (the equivalent of) 1/48'th of a zone in the second and third Brillouin zones.

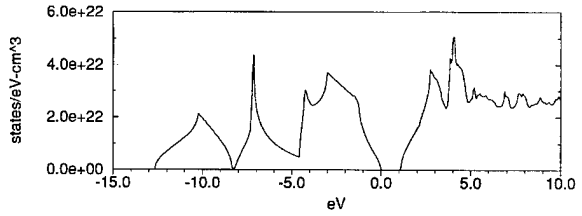


FIGURE 7 Density of states in Si computed using new irreducible wedge.

To reduce this tetrahedron to an equivalent irreducible wedge, we again appeal to symmetry. The tetrahedron shown in Figure 6 may be halved by cleaving along a $[\frac{1}{2}\frac{1}{2}\frac{1}{2}]$ plane; the two halves are identical by virtue of (diamond-type) reflection symmetry in [111] planes. We thus arrive at the new irreducible wedge shown in Figure 2 .

3. DENSITY OF STATES CALCULATION

To illustrate the quality of integration provided by the new wedge, Figure 7 shows the results of a density of states calculation for the lowest ten (4 valence plus 6 conduction) bands in Si using a local, 113 plane-wave pseudopotential calculation [4] (without spin-orbit coupling included). The mesh used consists of 4000 tetrahedrons (946 distinct vertices) filling the new irreducible wedge. With a denoting lattice constant, this corresponds to $0.05 \times (2\pi/a)$ increments for mesh points along $\langle 100 \rangle$ directions. By using a regular tetrahedral discretization of the new wedge, one obtains remarkably smooth results with relatively little coding effort. For the present calculation the $\varepsilon(\boldsymbol{k})$ argument of the delta function is treated to linear order leading to "wiggles" near van Hove singularities [5].

4. CONCLUSIONS

We have defined a new 1/48 Brillouin zone irreducible wedge applicable to diamond-type

semiconductors. This new irreducible wedge facilitates accurate numerical evaluation of Brillouin zone integrals such as density of states and scattering rates by simplifying considerably the handling of boundary element contributions to such integrals.

References

[1] Fischetti, M. V. and Laux, S. E. (1988). " Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects", *Phys. Rev. B*, **38**, 9721–9745.

[2] Callaway, J. (1991). *Quantum Theory of the Solid State*, Academic Press: Boston, Second Edn.

[3] Kunikiyo, T. *et al.* (1991). " A Monte Carlo simulation of anisotropic electron transport in silicon including full band structure and anisotropic impact-ionization model", *J. Phys.: Condens. Mat.*, **3**, 6721–6743.

[4] Chelikowsky, J. R. and Cohen, M. L. (1976). "Nonlocal Pseudopotential Calculations for the Electronic Structure of Seven Diamond and Zinc-blende Semiconductors", *Phys. Rev. B*, **14**, 556–582.

[5] Wiesenekker, G. and Baerends, E. J. (1991). "Quadratic integration over the three-dimensional Brillouin zone", *J. Phys.: Condens. Mat.*, **3**, 6721–6743.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 419–422
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Formulation of a Self-Consistent Model for Quantum Well *pin* Solar Cells: Dark Behavior

S. RAMEY and R. KHOIE*

Department of Electrical and Computer Engineering, University of Nevada, Las Vegas, 89154

(Received 28 May 1997; In final form 15 July 1997)

A self-consistent numerical simulation model for a *pin* single-cell solar cell is formulated. The solar cell device consists of a p–AlGaAs region, an intrinsic i–AlGaAs/GaAs region with several quantum wells, and a n–AlGaAs region. Our simulator solves a field-dependent Schrödinger equation self-consistently with Poisson and drift-diffusion equations. The field-dependent Schrödinger equation is solved using the transfer matrix method. The eigenfunctions and eigenenergies obtained are used to calculate the escape rate of carriers from the quantum wells, the capture rates of carriers by the wells, the absorption spectra in the wells, and the non-radiative recombination rates of carriers in the quantum wells. These rates are then used in a self-consistent finite-difference numerical Poisson-drift-diffusion solver. We believe this is the first such comprehensive model ever reported.

Keywords: Quantum well, solar cell, Schrödinger, escape, capture, recombination

1. INTRODUCTION

The conversion efficiency of a single cell *pin* solar cell can be enhanced by incorporating quantum wells in the intrinsic region of the device. [1] The incorporation of the quantum wells has two counteracting effects: the short-circuit current is increased because of the additional absorption of the low-energy photons in the lower bandgap quantum well and the open-circuit voltage is decreased because of the increase in the recombination of the photoexcited carriers trapped in the quantum well. Experimental results have shown, nevertheless, that the additional photocurrent resulting from the extension of the absorption spectrum to lower energies can outweigh the accompanying drop in the open-circuit voltage [2–3].

Along with these experimental studies, a number of theoretical investigations have been performed. Corkish and Green [4] studied the effects of recombination of carriers in the quantum well and concluded that although the increased recombination reduces the open-circuit voltage, limited

* Corresponding author.

enhancement in the conversion efficiency can be obtained with incorporation of the quantum well, albeit not as much as previously reported by Barnham and Duggan [1]. Araujo *et al.* [5] used detailed balance theory and predicted that the conversion efficiency of the quantum well cell would not exceed that of the base-line device. The results of photoresponse calculations by Renaud *et al.* [6] revealed that introducing the quantum wells can lead to improved photocurrent without much degradation of the open-circuit voltage. Most recently, Anderson [7] presented an ideal model for the quantum well solar cell device, incorporating the recombination and generation in the quantum wells. Anderson concluded that the improvement in efficiency is achieved only when the depth of the quantum well is less than about 200 meV.

The need for a comprehensive model is rather obvious, now that there seems to be an unsolved debate as to the ultimate advantage of incorporating quantum wells in the intrinsic region of a *pin* solar cell. In this paper we present formulation of one such model in which we self-consistently include the effects of:

1) Capture of electrons by the wells,
2) Escape of electrons from the wells,
3) Absorption of light in the wells, and
4) Recombination of carriers in the wells.

The standard drift-diffusion equations are modified to account for generation and recombination in the quantum wells and the transfer of electrons and holes between the bulk and quantum well systems.

2. SELF-CONSISTENT MODEL

The steady-state transport of carriers in the *pin* structure is described by current continuity equation written for the bulk regions as:

$$\frac{\partial n_b}{\partial t} = G_b - U_b + \frac{n_w}{\tau_e^n} - \frac{n_b}{\tau_c^n} + \frac{1}{q}\frac{dJ_n}{dx} = 0, \quad (1)$$

where n_b and n_w are electron densities in the bulk and quantum wells, respectively. J_n is electron current density. The two additional terms to the standard bulk current continuity equation incorporate the effect of carrier transfer into and out of the quantum wells. The terms n_w/τ_e^n and n_b/τ_c^n are the electrons escape and capture rates, respectively. These rates are also used in the continuity equations for the quantum wells, as given by:

$$\frac{\partial n_w}{\partial t} = \frac{n_b}{\tau_c^n} - \frac{n_w}{\tau_e^n} + G_w - U_w = 0. \quad (2)$$

In the above equations τ_e^n and τ_c^n are the electrons escape and capture times, respectively. The escape times of carriers are calculated using the model reported by Moss *et al.* [8] and capture times are extrapolated from theoretical and experimental data reported by Blom *et al.* [9]. Similar equations are written for holes. The boundary conditions for continuity equations are derived from surface recombination velocity model.

In the above continuity equations, the recombination in the bulk is modeled with radiative and non-radiative mechanisms. The term G_b is bulk generation rate and is given by:

$$G_b = \int_0^{\lambda_c} \alpha(\lambda) \cdot N_{ph} \cdot \left[\exp\left(-\int_0^x \alpha(\lambda)dx\right)\right] d\lambda \quad (3)$$

where λ_c is set to correspond to the bandgap of the material. The recombination in the quantum well U_w is a modified Shockley-Read-Hall recombination rate given by:

$$U_w = \frac{\sigma_n \sigma_p v_{th} N_t [pn - p_0 n_0]}{\sigma_n [n + n_t] + \sigma_p [p + p_t]} \quad (4)$$

where the trap density, N_t, is derived from the density of the interface states. The generation term G_w is calculated from Eq. (3) with the bulk absorption coefficient replaced with that of the quantum well. The absorption coefficients of the quantum wells are calculated by a model reported by Stevens *et al.* [10].

Calculation of the absorption spectra in the quantum wells as well as the escape and capture times of carriers require the eigenfunctions and eigenenergies of the carriers in the quantum wells, which are obtained from a field-dependent Schrödinger equation given by:

$$\left[\frac{\hbar^2}{2}\frac{d}{dx}\frac{1}{m^*(x)}\frac{d}{dx} + V(x)\right]\psi(x) = E_i\psi(x) \qquad (5)$$

where $\psi(x)$ is the envelope function, E_i are the eigenenergies and $V(x)$ is the potential profile. Non-constant effective mass $m^*(x)$ is assumed. The Schrödinger equation is solved using the transfer matrix method [11]. The above equations, together with Poisson and drift-diffusion equations are solved using a finite difference scheme.

3. RESULTS

The energy band diagram of the *pin* solar cell device with four quantum wells in the intrinsic region is shown in Figure 1. We simulated five *pin* devices all with a acceptor doping level of 10^{18}/cc

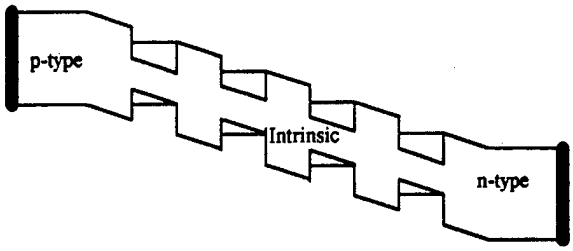


FIGURE 1 Energy band diagram of the solar cell with four quantum wells.

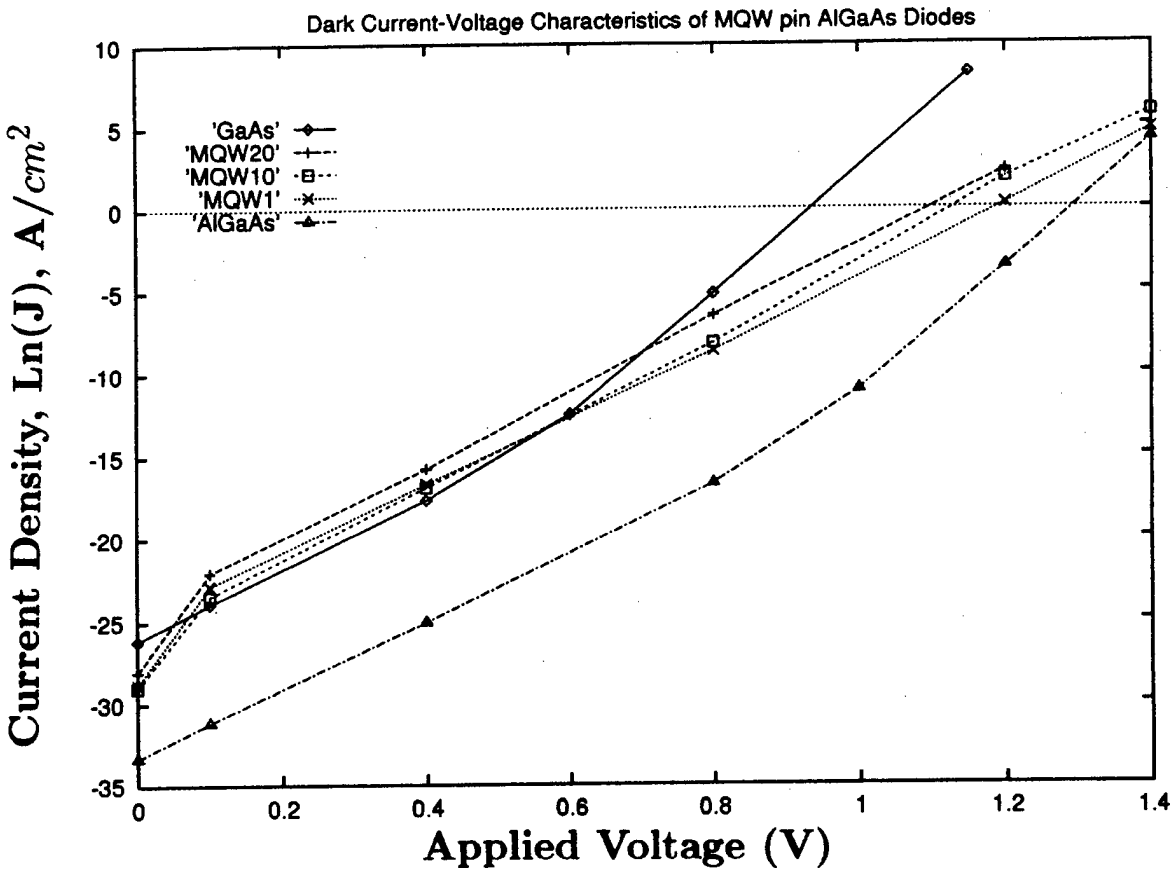


FIGURE 2 Dark characteristics of MQW *pin* solar cells.

and donor doping level of 4×10^{17}/cc in the 0.2 μm *p*-region and 0.3 μm *n*-region, respectively. The intrinsic region of all five devices is 0.5 μm. The first device is a base-line GaAs, the second device is a base-line AlGaAs device and the remaining three devices are AlGaAs devices with 1, 10 and 20 quantum wells, respectively. The mole fraction of Al is 0.3 in all AlGaAs devices and all quantum wells are 100 Å wide.

We simulated the dark characteristics of all five cells and the results are shown in Figure 2. Because the model incorporates both radiative and non-radiative recombination for the bulk material, the slope of the dark current is qv/nkT, with the ideality factor n ranging from one for exclusively radiative recombination to two for exclusively non-radiative. In the *pin* devices without quantum wells, n approaches one at high forward bias where radiative recombination becomes more important. At low biases the non-radiative recombination dominates and n is 1.9 for the AlGaAs device at 0.1 volts applied.

The recombination in the quantum wells is modeled as non-radiative interface recombination using a modified Shockley-Read-Hall expression. With the introduction of quantum wells into the *pin* device, this non-radiative recombination dominates at all biases and the ideality factor equals about 1.7 for all the quantum well devices. The observed crossover of the $i-v$ curves of the quantum well devices and the GaAs device has also been reported by Ragay *et al.* [3].

References

[1] Barnham, K. and Duggan, J. (1990). "A new approach to high-efficiency multi-band-gap solar cells", *Journal of Applied Physics*, **67**, 3490–3493.

[2] Barnham, K., Braun, B., Nelson, J. and Paxman, M. (1991). "Short-circuit current and energy efficiency enhancement in a low-dimensional structure photovoltaic device", *Applied Physics Letters*, **59**, 135–137.

[3] Ragay, F., Wolter, J., Marti, A. and Araujo, G. (1994). "Experimental analysis of the efficiency of MQW solar cells", *Proceedings of the 12th European Community Photovoltaic Solar Energy Conference*, pp. 1429–1433.

[4] Corkish, R. and Green, M. (1993). "Recombination of carriers in quantum well solar cells", *Proceedings of the 23rd Photovoltaics Specialist Conference*, pp. 675–680.

[5] Araujo, G., Marti, A., Ragay, F. and Wolter, J. (1994). "Efficiency of multiple quantum well solar cells", *Proceedings of the 12th European Community Photovoltaic Solar Energy Conference*, pp. 1435–1439.

[6] Renaud, P., Vilela, M., Freundlich, A., Bensaoula, A. and Medelci, N. (1995). "Modeling *p-i* (MQW)-*n* solar cells: A contribution for a near optimum design", *Proceedings of the 1994 IEEE First World Conference on Photovoltaic Energy Conversion Hawaii*, IEEE, NY, pp. 1787–1790.

[7] Anderson, N. (1995). "Ideal Theory of Quantum Well Solar Cells", *Journal of Applied Physics*, **78**, 1850–1861.

[8] Moss, D., Ido, T. and Sano, H. (1994). "Calculation of photo-generated carrier escape rates from GaAs/AlGaAs quantum wells", *IEEE Journal of Quantum Electronics*, **30**, 1015–1026.

[9] Blom, P., Smit, C., Haverkort, J. and Wolter, J. (1993). "Carrier capture in a semiconductor quantum well", *Physical Review B*, **47**, 2072–2081.

[10] Stevens, P., Whitehead, M., Parry, G. and Woodbridge, K. (1988). "Computer modeling of the electric field dependent absorption spectrum of multiple quantum well material", *IEEE Journal Quantum Electronics*, **24**, 2007–2015.

[11] Jonsson, B. and Eng, S. (1990). "Solving the Schrodinger equation in arbitrary quantum-well potential profiles using the transfer matrix method", *IEEE Journal of Quantum Electronics*, **26**, 2025–2035.

Authors' Biographies

Steve Ramey received his B.S. in Physics from Carnegie Mellon University in 1992. He is currently completing his M.S. degree in Electrical Engineering at University of Nevada, Las Vegas. His research interests include the optoelectronic properties of quantum well structures, numerical modeling of semiconductor devices and applications of parallel supercomputers. He has accepted a research position with Solid State Measurements in Pittsburgh, Pennsylvania exploring the physics of metal-semiconductor contacts.

Rahim Khoie received his Ph.D. degree in Electrical Engineering from University of Pittsburgh, Pennsylvania, in 1986. He is presently an Associate Professor of Electrical and Computer Engineering at University of Nevada, Las Vegas. He has published numerous papers in journals and proceedings and has received several research grants. His research interest include high speed electron devices, heterostructures, quantum well devices, photovoltaics and optoelectronics. Dr. Khoie has won a number of local and national teaching awards.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 423–428
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Hydrodynamic (HD) Simulation of *N*-Channel MOSFET's with a Computationally Efficient Inversion Layer Quantization Model

HAIHONG WANG*, WEI-KAI SHIH, SUSAN GREEN, SCOTT HARELAND, CHRISTINE M. MAZIAR and AL. F. TASCH JR.

Microelectronics Research Center, The University of Texas at Austin, Austin, TX 78712, USA

A quantum mechanical treatment of electron inversion layers is incorporated in the hydrodynamic (HD) transport model used in UT-MiniMOS. A physically based, yet computationally efficient, three-subband model is implemented in the HD simulation tool. The three-subband model, which is based upon solutions to Schrodinger's equation, has the important advantage of more accurately predicting the distribution of electrons in the inversion layers than does more conventional classical models. A more simplified quantum mechanical model with carrier heating effects included has also been developed. Terminal currents are calculated using these quantum mechanical models and the comparison with results from classical calculations indicates the importance of quantum mechanical effects in the deep submicron device simulations.

Keywords: MOS, inversion layer, quantization, hydrodynamic simulation, carrier confinement

INTRODUCTION

The quantum mechanical treatment of carrier confinement in silicon MOSFET inversion layers is increasingly essential for accurate simulation of devices scaled into the deep submicron region. It is well known that a high transverse electric field at the Si/SiO_2 interface leads to a large energy band bending, which in turn can cause quantization of the inversion layer carriers in the direction perpendicular to the interface [1]. Self-consistent solutions of the Schrödinger's and Poisson equations have been adopted by a number of researchers to investigate the two-dimensional nature of electrons in inversion layers [2, 3]. However, from the point of view of the device design engineer, less computationally expensive approaches that may be routinely used with existing device simulators are necessary. Van Dort's simple quantum model [4] and three subband quantization model (3SB) [5]

*Corresponding author.

have been implemented in drift-diffusion simulation tools and have facilitated the prediction of device electrical behavior (such as threshold voltage and capacitance).

The three subband model introduces an additional level of physical sophistication by solving for the energy-level and wave-function for the first three subbands, using approximate solutions to Schrödinger's equation based upon physically reasonable estimates of the form of the potential well. This approach has the advantage of correctly predicting the electron distribution in the inversion layers due to quantum mechanical effects and, in turn, leads to better predictive capabilities for C–V simulation and physical oxide thickness estimation. In the hydrodynamic (HD) 3-subband simulation, carrier temperature instead of lattice temperature is used to distribute electrons among each subband and between the classical (three dimensional electron gas) and the quantum (two dimensional electron gas) domains. As a result, under high drain bias condition, a large portion of electrons near the drain edge will be in the classical domain due to the carrier heating effect. In this paper, we present, for the first time, HD simulation results with a 3SB model included using UT-MiniMOS [6].

CARRIER CONFINEMENT AND QUANTIZATION MODEL IMPLEMENTATION

When carriers are quantized in the potential well, the density of states must be represented by a two-dimensional (2-D) density of states instead of the three-dimensional (3-D) density of states that is assumed in classical calculations. Knowing the eigenvalue and eigenfunction for each subband, we can represent the distribution of charge in a 2-D system by the following expression:

$$n_{\mathrm{QM}}(z) = \frac{kT_{\mathrm{carrier}}}{\pi\hbar^2}\sum_i n_{vi}m_{zi}$$

$$\sum_j \ln\frac{\left\{1+\exp\left(\frac{(E_F-E_{ij})}{kT_{\mathrm{carrier}}}\right)\right\}}{\left\{1+\exp\left(\frac{(E_F-E_{\mathrm{cl}})}{kT_{\mathrm{carrier}}}\right)\right\}}|\zeta_{ij}(z)|^2 \qquad (1)$$

where n_{vi} denotes the degeneracy of the i^{th} subband, E_{ij} the subband energy level of the j^{th} valley in the i^{th} subband, E_{cl} the energy level of the classical domain, and $\zeta_{ij}(z)$ the wavefunction describing the distribution of carriers in the subband. In the three subband model, E_{ij} and ζ_{ij} are calculated via self-consistent solutions of Schrödinger's equation and Poisson's equation. However, *a priori* assumptions about the shape of the potential well are used to solve (1) based upon analytical expressions for E_{ij} and ζ_{ij}. A triangular potential well is used in the three subband model. For this form of the potential, the solution to Schrödinger's equation is the Airy function where the eigenenergies for the system can be calculated exactly as a function of the surface field [2, 5]. To account for the perturbation of the potential due to the inversion layer charges, a variational wavefunction originally proposed by Fang and Howard [7] was used for the lowest subband. This bypasses the computational expense of solving Schrödinger's equation explicitly. In order to maintain additional computational efficiency, only the first three subbands are used in the calculation of (1). Above the energy level E_{cl}, the carriers are all treated as classical carriers with a 3-D density of states. This approach ensures agreement with classical calculations in the limit of low channel doping in addition to reducing the total computation time.

Since the electron temperature (or average energy) is used as an additional variable in the HD simulation, the temperature effect must be included in the quasi-Fermi level calculation in order to obtain a correct carrier distribution among subbands. Also, due to the electron heating, the classical carrier component may play an important role in the carrier density calculation for the region near the drain edge. The quasi-

Fermi level and the classical carrier component need to be calculated correctly in the 3SB model to ensure that the HD simulation converges and that the simulation results are physically correct. In our approach, the fractional electron population in the quantum domain as a function of the electron energy is modeled using our quantum-window enhanced uniform-field Monte Carlo (MC) simulator [8], as shown in Figure 1. The fractional population-energy relationship in our HD simulation is obtained from Monte Carlo simulation under the low transverse effective field condition ($E_{\text{trans}} = 1.9 \times 10^5$ V/cm). Very little impact on the terminal current, even under the high drain bias condition, is observed by choosing a different fractional population-energy relationship from Monte Carlo simulations with different transverse effective fields. The reasons for the insensibility of terminal current on the fractional population-energy relationship are being under investigations. The fractional population, so modeled, can then be used to calculate the sheet quantum carrier concentration (N_{qm}) and classical carrier concentration (N_{cl}) for a given total sheet carrier concentration ($N_{\text{qm}} + N_{\text{cl}}$) after the drift-diffusion solution is obtained. By knowing N_{qm} and N_{cl}, we can calculate the quasi-Fermi level and classical

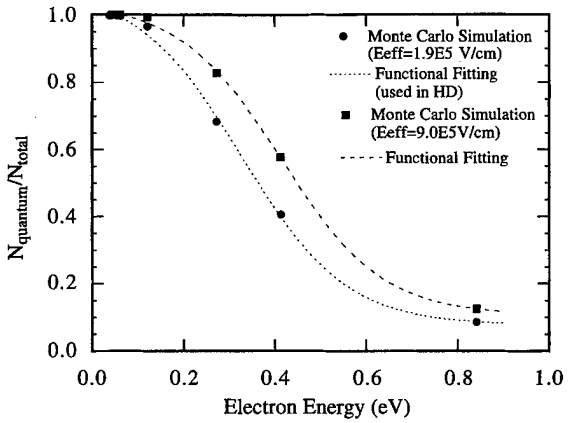


FIGURE 1 Fractional population of electrons in the quantum domain (first three subbands) as a function of the electron average energy from our uniform-field Monte Carlo simulation. A simple analytic fit used in the hydrodynamic simulation is also plotted.

effective density of states. The carrier distribution along the depth direction is subsequently calculated in the 3SB model to obtain the effective intrinsic carrier density, following the similar approach of Ref. [5].

In addition to the more physically accurate three-subband model, more simplified models are of interest to the modeling community. While these models do not attempt to predict the quantum mechanically consistent charge distribution, they are able to predict the effect of quantization on the terminal characteristics when this is the only information required. This is accomplished by altering the bandgap of the silicon near the interface via the intrinsic carrier concentration [4]. The change in the bandgap is modeled as follows:

$$\Delta E_G = \beta \left(\frac{\varepsilon_{\text{si}}}{4qkT_{\text{carrier}}} \right) E_s^{2/3} \tag{2}$$

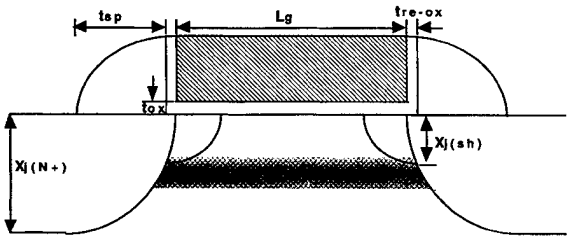
where E_s denotes the transverse electric field and β the fitting parameter used to calibrate the model to experimental data and/or more physically based quantization models. This modified bandgap is then used to make corrections to the charge density near the interface. The carrier temperature is used when the HD models are turned on, otherwise the lattice temperature is used. This modification permits the model to approach the classical results when the energy distribution of the carriers is highly energetic, and hence more classical-like.

SIMULATION RESULTS AND DISCUSSION

In order to explore the importance of including carrier confinement models, as well as carrier heating models, in device simulation tools, we have examined the behavior of the drift-diffusion and hydrodynamic transport models both with quantum mechanical carrier confinement models and without. We used as our prototypical device structure the simple 0.18 micron nMOS device

presented in Figure 2. The doping profiles, junctions depths and oxide thickness were selected to be representative of devices of this technology generation.

An examination of the electron concentration near drain edge at low drain bias (V_d= 0.1 V) illustrates that the electron distribution obtained from the HD 3SB calculation is almost the same as that from a DD-3SB calculation (Fig. 3(a)). This result can be understood by the fact that there is almost no carrier heating for such a low drain bias. At high drain bias (V_d= 1.8 V), on the other hand, the distribution near the drain edge obtained from the HD- 3SB calculation differs significantly from that obtained with DD 3SB (Fig. 3(b)). In the drift-diffusion approach, regardless of lateral field conditions, the carriers are assumed to be at the lattice temperature, so the carriers are assumed to occupy the lowest subbands (i.e., quantum mechanical regime) more readily, and hence have a smaller classical contribution (i.e., the unbound states). However, in the hydrodynamic approach, under high lateral electric field conditions, significant carrier heating can occur near the drain edge. Such energetic carriers tend to occupy bound states in the higher subbands and the unbound states. Hence, for high drain bias, the electron distribution near the drain edge calculated from HD 3-subband is more classical-like.



L_{gate}=0.18 μm
W_{gate}=10.0μm
t_{ox}=4.0nm
t_{reox}(poly)=15nm
Substrate Doping 1x10^{16}cm^{-3}
Channel Profiles: APT + V_T adjust

Source/Drain	Shallow Junction	Deep Junction
x_j(nm)	50	150
Profile	Gaussian	Gaussian
Peak Doping(cm^{-3})	4x10^{19}	1x10^{20}
Lateral Extension(%)	60	60

FIGURE 2 Device structure for prototypical 0.18 micron nMOS transistor. This structure is used throughout the results section of the paper.

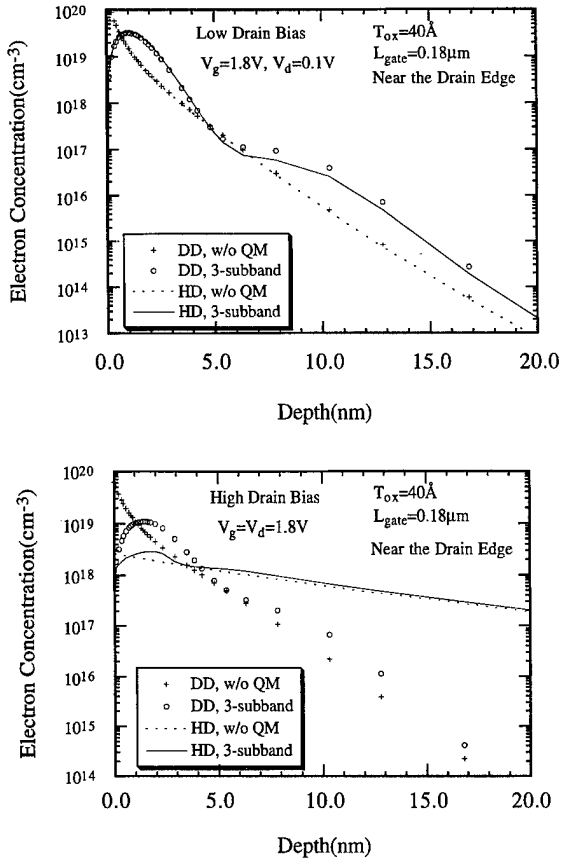


FIGURE 3 (a) Electron concentration along the depth near the drain edge at the low drain bias (V_d=0.1 V) for both DD and HD simulations (classical and 3-subband) (b) Electron concentration along the depth near the drain edge at the high drain bias (V_d=1.8 V) for both DD and HD simulations (classical and 3-subband).

The I_d–V_d characteristics from the HD simulations are presented in Figure 4. Either a temperature corrected van Dort's model or the 3-subband quantum mechanical model is used in the HD quantum mechanical calculations. The results from classical HD simulation are also presented for comparison. It should be noted that the drain currents from the temperature corrected van Dort's model agree very well with those from the three subband HD simulations. Also, the observable difference in saturation currents between

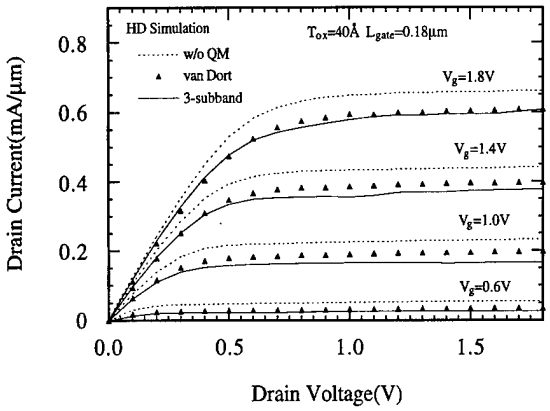


FIGURE 4 Drain currents versus drain voltage from the hydrodynamic simulation, using classical, temperature corrected van Dort's model and 3-subband model, respectively.

classical and quantum mechanical calculation suggests that the carrier confinement effects should be taken into account for the deep submicron device simulation. The I_d-V_g characteristics from HD simulation with classical, temperature corrected van Dort model, and 3-subband model, respectively (Fig. 5), indicate that the two quantum models predict almost the same threshold voltage, while the threshold voltage from classical HD calculations is smaller than that from quantum mechanical HD calculations. The difference in the threshold voltage between the classical and quantum mechanical calculations is believed to contribute to the difference in the saturation drain current between HD classical and HD quantum mechanical calculations, as shown in Figure 4.

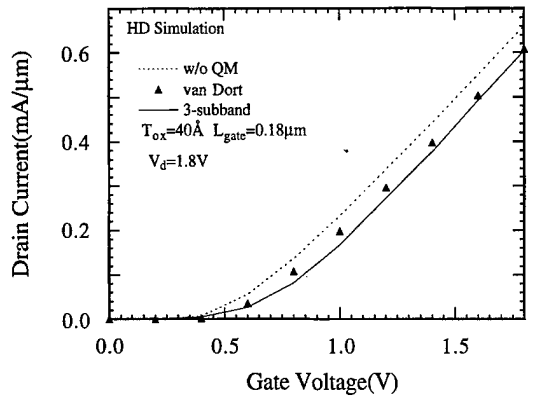


FIGURE 5 Drain current versus gate voltage from the hydrodynamic simulation, using classical, temperature corrected van Dort's model and the 3-subband model, respectively.

CONCLUSION

Carrier confinement and quantization models have been successfully implemented in the hydrodynamic (HD) simulation tool UT-MiniMOS. Under high drain bias conditions, the electron concentration near the drain edge calculated from HD simulation differs significantly from that of DD simulation due to the carrier heating effect, both with and without quantum effects included. Under low dran bias conditions, quantum confinement effects are significant while carrier heating effects are negligible. The I_d-V_d and I_d-V_g characteristics suggest the need for both the HD treatment and the QM picture either through the use of the 3SB model or a temperature corrected van Dort model.

Acknowledgements

This work was supported, in part by the Semiconductor Research Corporation (SRC), the Joint Services Electronics Program (F49620-95-C-0045), the Texas Advanced Technology Program (TATP) and Motorola. The United States Government is authorized to distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

References

[1] Schriffer, J. R. (1957). "Mobility in Inversion Layers: Theory and Experiment", in *Semiconductor Surface Physics*, p. 55.

[2] Stern, F. (1972). "Self-Consistent Results for *n*-type Si Inversion Layers", *Phys. Rev. B*, **5**(12), 4891.

[3] Ando, T., Fowler, A. B. and Stern, F. (1982). "Electronic Properties of Two-Dimensional System", *Rev. Mod. Phys.*, **54**(2), 437.

[4] van Dort, M. J., Woerlee, P. H. and Walker, A. J. (1994). "A Simple Model for Quantization Effects in Heavily-Doped Silicon MOSFETs at Inversion Conditions", *Solid-State Electron.*, **37**(3), 411.

[5] Hareland, S. A., Krishnamurthy, S., Jallepalli, S., Yeap, C.-F., Hasnat, K., Tasch, A. and Maziar, C. (1996). "Computationally Efficient Models for Inversion Layer Quantization Effects in Deep Submicron *N*-Channel MOSFETs", *IEEE Trans. Electron Devices*, **43**(1), 90.
[6] *UT-MiniMOS 5.2-3.2 User's Guide*, Microelectronics Research Center, The University of Texas at Austin, 1997.
[7] Fang, F. F. and Howard, W. E. (1966). "Negative Field-Effect Mobility on (100) Si Surface", *Phys. Rev. Lett.*, **16**(8), p. 797.
[8] Wei-Kai Shih, Scott Hareland, Srinivas Jallepalli, Maziar, Christine M. and Tasch, Al. F. (1996). "An Accurate Preprocessor for Monte Carlo Study of Electron Transport in Inversion Layers of Silicon nMOSFETs", *54th Annual Device Research Conference Digest*, p. 28.

Authors' Biographies

Haihong Wang received the B.S. degree in physics from Beijing University, Beijing, P.R.C., in 1988, and the M.S. degree in electrical engineering from the University of Texas in Austin in 1995. He is currently pursuing the Ph.D. degree in electrical engineering at the University of Texas at Austin. His research interests include advanced transport models and quantum-mechanical effects in deep submicron MOSFET's, low-power device design and analysis, and improved structures for MOSFET scaling.

Wei-Kai Shih received the B.S. degree in physics from National Taiwan University, Taipei, Taiwan, R.O.C., in 1989, and the M.A. degree in physics from the University of Texas at Austin in 1994. He is currently pursuing the Ph.D. degree in electrical engineering at the University of Texas at Austin. His research interests include hot-carrier effects, non-local phenomena, and quantum-mechanical effects in deep submicron MOSFET's.

Susan Green received the B.S.E.E. degree from the University of Arizona in 1994 and the M.S.E. from the University of Texas at Austin in 1996. She is currently with Intel in Chandler, Arizona.

Scott Hareland received the B.S.E.E. degree from Rice University in 1992 and the M.S.E. and Ph.D. degrees from the University of Texas at Austin in 1994 and 1997. He is currently with Intel in Portland, Oregon. His research interests include device, process and circuit modeling, device physics, and improved structures for MOSFET scaling. Dr. Hareland was awarded an Intel Foundation Robert Noyce Memorial Fellowship in 1994 and is a member of Phi Beta Kappa, Tau Beta Pi and Eta Kappa Nu.

Christine Maziar received the B.S.E.E. (with Highest Distinction), the M.S.E.E. and the Ph.D. degrees from Purdue University in 1981, 1984, and 1986. In January of 1987, she joined the faculty of the University of Texas at Austin, where she is a Professor of Electrical and Computer Engineering and Vice Provost. She holds the Archie W. Straition Endowed Teaching Fellowship in Electrical and Computer Engineering. Her research interests include modeling of charge transport in high performance device structures, semiconductor device physics, and device simulator enhancement. She has authored or co-authored over 100 journal or conference publications. She is a member of IEEE, ASEE, APS, Eta Kappa Nu, Tau Beta Pi and Sigma Pi Sigma.

Al. F. Tasch received his Ph.D. degree in Physics in 1969 form the University of Illinois, Urbana-Champaign. He worked in industry for 17 years at Texas Instruments and Motorola, and joined the faculty of the Department of Electrical and Computer Engineering at the University of Texas at Austin in 1986 as a Chair Professor. He has been awarded 38 U.S. patents and was elected to the National Academy of Engineering in 1989.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 429–435
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Study of Electron Velocity Overshoot in NMOS Inversion Layers

WEI-KAI SHIH, SRINIVAS JALLEPALLI, MAHBUB RASHED, CHRISTINE M. MAZIAR* and AL. F. TASCH JR.

Microelectronics Research Center, The University of Texas at Austin, Austin, TX 78712, USA

Non-local electron transport in *n*MOSFET inversion layers has been studied by Monte Carlo (MC) simulations. Inversion layer quantization has been explicitly included in the calculation of density of states and scattering rate for low-energy electrons while bulk band structure is used to describe the transport of more energetic electrons. For uniform, high-lateral field conditions, the effects of quantization are less pronounced due to the depopulation of electrons in the lower-lying subbands. On the other hand, Monte Carlo results for carrier transport in spatially varying lateral fields (such as those in the inversion layer of MOSFETs) clearly indicate that depopulation of the low-lying subbands is less evident in the non-local transport regime. Quasi-2D simulations have shown that, at high transverse effective field, the inclusion of a quantization domain does have an impact on the calculated spatial velocity transient .

Keywords: *n*MOS, inversion layer, velocity overshoot, quantization, Monte Carlo simulation

INTRODUCTION

With the continued scaling of the feature size of MOS devices, carrier transport in the MOS inversion layers has entered the regime where non-local effects are no longer negligible. In a previous study, velocity overshoot was found to account for approximately 20% of the disagreement in drain current between measurement and drift-diffusion simulation in a 0.12 μm SOI MOSFET [1]. Despite a few pioneering studies on probing the electron spatial transient velocity in compound semiconductor devices [2], a mature experimental technique that allows the direct measurement of carrier velocity in silicon MOSFETs has not been available. In order to accurately express these non-local effects in physically based models, the Monte Carlo (MC) technique has become increasingly indispensable for device scientists and engineers. In conventional MC tools, effects due to channel quantization resulting from the steep bulk band bending are routinely ignored. With phenomenological surface-roughness scattering models, some conven-

* Corresponding author.

tional MC tools [3,4] using bulk silicon band structure have demonstrated the ability to reproduce the experimental inversion-layer velocity-field characteristics in the local transport regime within a limited range of temperatures. However, the validity and accuracy of this type of approach remains questionable in the non-local transport regime.

In favor of the semi-classical approach, one might argue that non-local transport is most generally observed under conditions of high electric field in the transport direction. A condition in which the inversion layer electrons populate higher energy states and the effects of quantization are less pronounced. However, the issue regarding how fast the 2DEG can respond to an abruptly changing field, hence become more classical-like, has not been investigated. The purpose of this paper is to qualitatively address these issues with simulations performed on quasi-2D test structures. As will be shown, inversion-layer quantization does have an impact on the carrier average velocity in the non-local regime, despite the presence of high lateral electric field.

THE MONTE CARLO SIMULATOR

The MC program used in this work is an integrated tool consisting of two simulation domains that partition the entire inversion-layer electron population into quantum (2D) and classical (3D) components. The 2D domain is used to simulate transport of low-energy electrons by directly solving the multi-subband 2D Boltzmann transport equation [5]. The 3D domain, employing three fitted conduction bands, is used for treating more energetic electrons that have motions that are more classical-like [6]. The design of the integrated tool allows the 3D domain to execute without invoking the 2D domain. In its stand-alone mode, the 3D domain takes into account phonon scattering and impact ionization and has been calibrated to reproduce uniform transport characteristics in bulk silicon. It also employs a phenomenological model, which was previously validated, to take into account surface roughness scattering [4]. In the 2D domain, the effective-mass approximation (EMA) including bulk nonparabolicity is used and effects of scattering due to bulk phonon and surface roughness have been accounted for. Coulomb scattering is greatly suppressed due to carrier screening in the strong-inversion regime and has been ignored in the present treatment. The envelope functions obtained within the EMA are used to calculate the electron-phonon form factors needed for computing the phonon scattering rate. The parameters used in phonon and surface roughness scattering models have been adjusted to reproduce the experimental mobility values. As shown in Figure 1, the inversion layer ohmic mobilities obtained with the 2D domain agree well with those obtained in previous experimental studies (at 413 K and 77 K) and the extensively validated UT-mobility model (at 300 K) [7,8]. It should be emphasized that all the MC mobility values are obtained with a single set of scattering model parameters.

Physical boundary conditions between the 2D and 3D domains are employed. The implementation is similar to that proposed by Fischetti *et al.* [9] except for the use of a 2D/3D boundary (in the energy space) more consistent with the EMA and our previous pseudo-potential calculation [10]. Within the EMA, the envelope function of a given

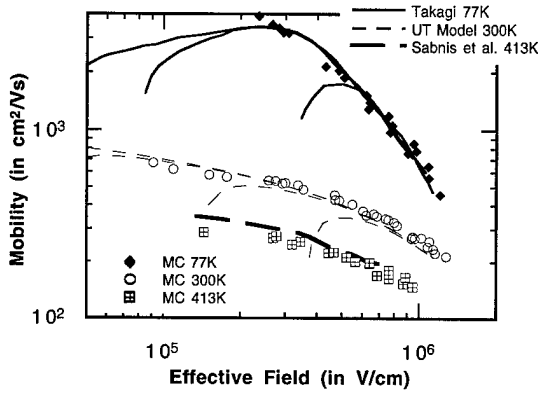


FIGURE 1 low-field mobility obtained with the single particle MC compared with experimental and UT-mobility model values.

subband is independent of the parallel momentum. Such an insensitivity, confirmed in [10], indicates that an electron with a large parallel momentum is as well quantized as that with a small parallel momentum in the same subband, despite the difference between their total energies. Instead of defining the 2D/3D boundary in terms of total energy, as suggested in [9], we use a criterion based on the electron total energy in the quantization (z-) direction. The concept of "buffer subbands" suggested in [9] has been adopted to allow 2D electrons to enter the 3D domain via scattering. The six lowest-lying subbands and 64 buffer subbands have been used. Applying the above criteria, as illustrated in Figure 2, a 3D electron is converted to the quantum domain when the following two conditions are met: (1) The electron is located in the lowest bulk valley. (2) The total energy of the 3D electron in the z-direction (E_z) is less than $E_z^{(th)}$, where $E_z^{(th)}$, is defined to be the edge of the lowest buffer subband and E_z is evaluated within the parabolic approximation. On the other hand, a 2D electron is converted to the classical domain when it enters one of the buffer subbands via intersubband scattering. Conservation of the electron in-plane momentum is imposed in the conversion process. Since the energy spectrum in the 3D domain is a continuum, energy conservation is enforced for the 2D to 3D transition. However, energy cannot be conserved during the 3D to 2D transition because only discrete subband levels are available in the 2D domain. In this case, the subband that minimizes the energy mismatch is chosen. If the energy conservation law were strictly enforced during the simulation, the 2D domain would be depleted of electrons at steady state under uniform, high-lateral-field condition, which is clearly unphysical. To circumvent this problem without violating the energy conservation, one needs to directly calculate the scattering rate between the 3D and the 2D states. This alternative approach is expected to be comparable to our current implementation since the 2D density of states obtained with the quantum treatment is expected to be comparable to that obtained with the classical treatment at sufficiently high energy (the correspondence principle).

With the integrated MC tool, the velocity-versus-field relations obtained at various transverse E_{eff} are observed to agree well with the UT-mobility model (see Fig. 3). As expected, the calculated saturation velocity is the same as that in bulk silicon since at high field, the 3D population dominates and surface roughness scattering is less effective. This is further evidenced in Figure 4, where the populations of electrons in the quantum and classical domains are separately plotted with respect to the lateral field. At low lateral field, the 3D electrons have a larger population for the less inverted channel due to the weaker quantum confinement at lower E_{eff}. For the same reason, the 3D component takes over the 2D component at a lower lateral field in the less inverted channel.

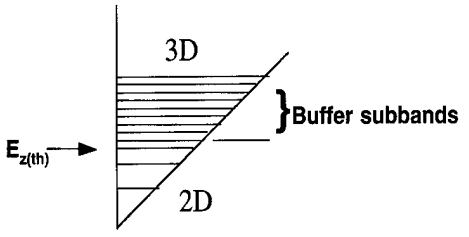


FIGURE 2 Illustration of the partition of simulation domain in the inversion-layer quantum well.

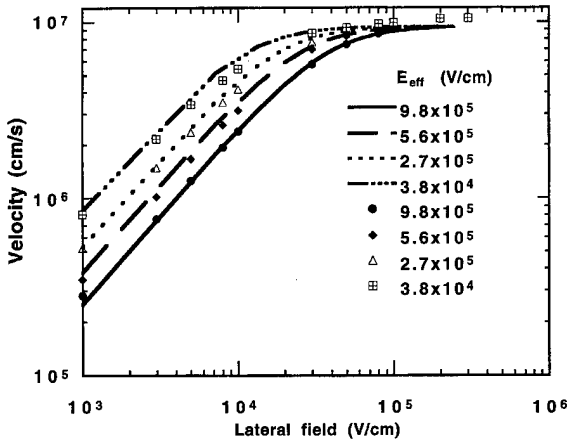


FIGURE 3 Velocity versus field obtained from the MC tool including both 2D and 3D domains (symbols) compared to that obtained with the UT mobility model (lines).

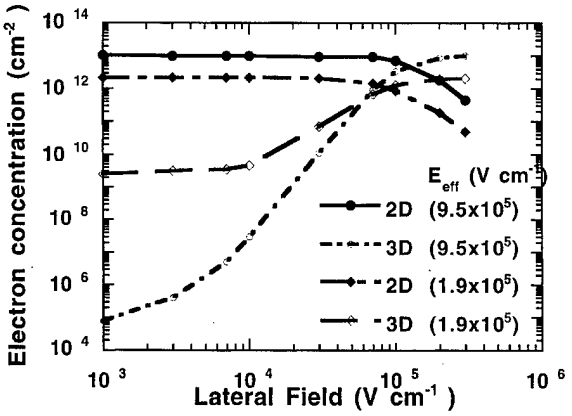


FIGURE 4 Populations of electrons in the quantum and classical domains.

SIMULATION PROCEDURE AND RESULT

Quasi-2D structures with uniform MOS inversion layers subjected to artificial lateral field profiles that vary abruptly along the channel (x-) direction are studied. A 1D Schrödinger-Poisson iteration following classical drift-diffusion simulation is used to obtain the self-consistent subband dispersion and scattering rate for given substrate doping levels (N_{sub}) and gate biases (V_G). Either a step or a ramp profile is chosen for the lateral field across a 0.9 μm channel. The low-field region is sufficiently long for transport to reach the steady state corresponding to the local field. A periodic boundary condition is applied to both ends of the channel with carrier injection conserving both momentum and subband index. To directly assess velocity overshoot due solely to the applied field, Poisson feedback resulting from the carrier redistribution along the channel has been ignored.

To appreciate the importance of the quantum domain in the non-local transport regime, the fractional electron population in the 2D domain under step-like and ramp-shaped lateral field profiles is examined. The ramp field profile corresponds to a 2V voltage drop across the 0.2 μm channel and roughly conforms with the power-supply scaling trend associated with the scaling of device dimensions. The step profile represents a worse case scenario where a power supply of 4V is applied across the channel. In Figure 5, it is observed that the 2D electron population in the high lateral-field region remains significant (>40%) in the overshoot regions (around $x=0.5\,\mu m$ for the ramp profile and $x=0.35\,\mu m$ for the step profile), suggesting that quantization might be of importance in the non-local transport regime in spite of the presence of the high lateral fields. It should be noted that, had there been no non-local effects, inspection of Figure 4 would have lead to the estimated fractional 2D populations at these locations of less than 20% even without taking into account

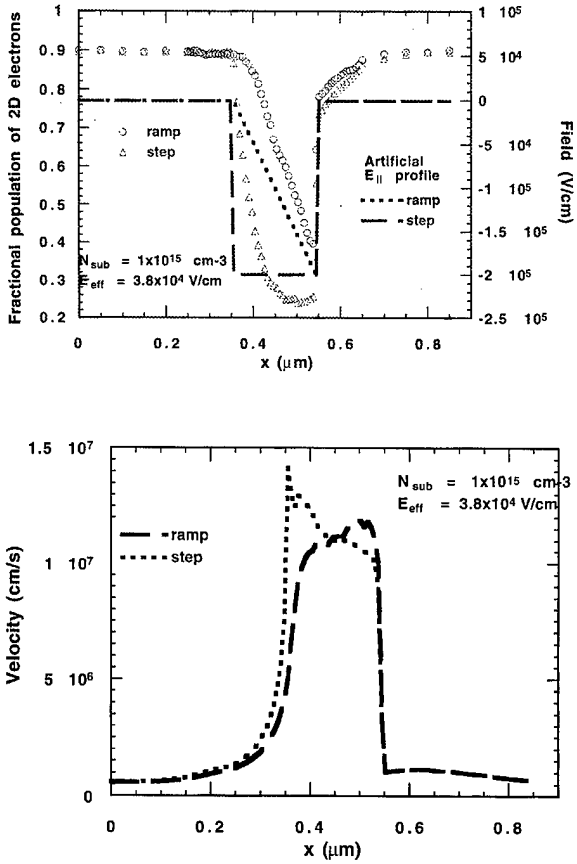


FIGURE 5 (a) Fractional population of electrons in the 2D simulation domain (symbols). Applied lateral field profiles are also shown (lines). (b) Velocity profiles along the lateral direction obtained with the lateral field profiles in (a). Overshoot occurs near $x=0.5\,\mu m$ and $x=0.35\,\mu m$ for the step profile and ramped profile, respectively.

the fact that the E_{eff} used in Figure 5 is lower than those used in Figure 4.

With the same ramped lateral field profile shown in Figure 5(a), Figure 6 shows that the transient velocity obtained at different E_{eff}'s and N_{sub}'s. In general, the transient velocity seems to have a stronger dependence on E_{eff} than on N_{sub}. In the low-field region, this is expected since the quantum domain explains universal mobility quite well. The universality of velocity (and mobility) dependence on E_{eff} seems to be retained along the rising edge of the transient velocity until the latter becomes sufficiently high, where N_{sub} starts to have some impact. At higher E_{eff}, the lower velocity observed on the rising edge can be qualitatively explained by the lower mobility of the ground subband due to stronger surface roughness scattering. Near the velocity peak, on the other hand, a simple explanation for the different velocities is not available since all subbands in the quantum domain and the classical domain are equally significant.

To see the quantization effects on the MC-simulated velocity profile, Figures 7(a) and 7(b) compare the velocity profiles obtained with conventional MC with those obtained with the integrated MC tool at $E_{eff} = 5.6 \times 10^5$ V/cm and 9.8×10^5 V/cm, respectively. In both cases, lower

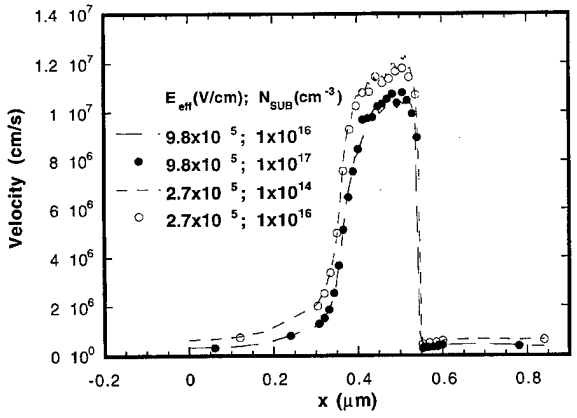


FIGURE 6 Velocity along the channel with ramped field profile at two different E_{eff}. More pronounced overshoot is seen at the lower E_{eff}. For each E_{eff}, two different substrate doping levels are used.

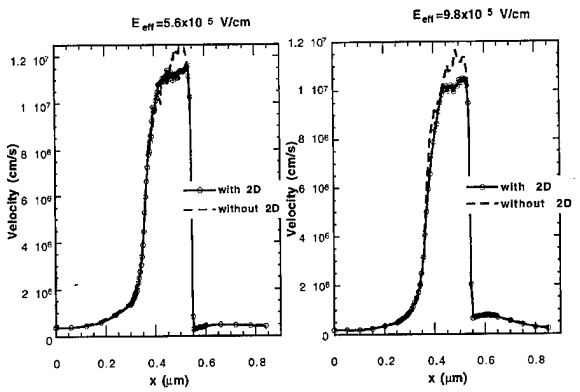


FIGURE 7 Velocity profiles obtained from conventional MC simulations compared to those obtained from the integrated tool. Different E_{eff}'s and same substrate doping level (1×10^{16} cm^{-3}) are used in (a) and (b).

peak velocities are predicted by the integrated MC. Although the difference between two MC results slightly increases as quantization becomes stronger, we are reluctant make the intuitively appealing statement that the difference is indeed enhanced as channel becomes more strongly inverted. If such a trend does exist, the impact of including channel quantization on the predicted velocity overshoot in conventional deep-submicron MOSFETs might be limited as the quantum confinement is weaker near the drain. However, as aggressive device design with asymmetric channel doping profile has been carried out [11] to take advantage of velocity overshoot near the source, where the silicon surface is strongly inverted, quantization effects on the spatial velocity transient is likely to be more significant in these devices.

CONCLUSION

Quasi-2D Monte Carlo simulations that take into account the effects due to size quantization have been used to study carrier velocity overshoot in *n*MOS inversion layers. Non-local effects are seen not only in carrier velocity but also in the carrier population. With a given ramp-field profile, the spatial transient velocity is observed to depend on the transverse effective field more strongly than on

the substrate doping. In this work, more pronounced velocity overshoot is observed at lower E_{eff}.

The impact of quantization on simulated velocity profiles has been observed by comparing results obtained with the conventional MC to those obtained with the integrated MC tool. While quantization seems to have a slightly larger impact at higher E_{eff}, more thorough study is needed to support such a viewpoint, given the uncertainties in models used in both MC tools. Since the comparison is based on two different simulation models, each of which contains certain number of parameters calibrated such that uniform-condition transport characteristics can be reproduced, we cannot exclude the possibility that there exists a different region in the parameter space where one can draw a conclusion different from ours. In fact, even among the currently existing well-calibrated classical MC codes, disagreement in spatial transient velocity is very likely to exist. The same statement can also apply to the quantum MC codes. More studies are needed to clarify the issues related to the impact of quantization on velocity overshoot.

Acknowledgements

This work was supported, in part by the Semiconductor Research Corporation (SRC), the Joint services Electronics Program (F49620-95-C-0045), the Texas Advanced Technology Program (TATP) and Motorola. The United States Government is authorized to distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

References

[1] Assaderaghi, F., Sinitsky, D., Gaw, H., Bokor, J., Ko, P. K. and Hu, C. (1994). "Saturation velocity and velocity overshoot of inversion layer electron and holes", *IEDM Tech. Dig.*, pp. 479–482.

[2] Grann, E. D., Tsen, K. T., Sankey, O. F., Ferry, D. K., Salvador, A., Botcharev, A. and Morkoc, H. (1995). "Electron velocity overshoot in a GaAs-based p-i-n nanostructure semiconductor observed by transient subpicosecond Raman spectroscopy", *Appl. Phys. Lett.*, **67**, 1760–1762.

[3] Sangiorgi, E. and Pinto, M. R. (1992). "A semi-empirical model of surface scattering for Monte Carlo simulation of silicon *n*-MOSFETs", *IEEE Trans. Electron Devices*, **39**, 356–361.

[4] Jallepalli, S. (1992). "Surface scattering in a self-consistent Monte Carlo analysis of electron transport in *n*-MOSFETs", Master Thesis, The University of Texas at Austin.

[5] Shih, W.-K., Jallepalli, S., Yeap, C.-F., Rashed, M., Maziar, C. M. and Tasch, A. F. (1995). "A Monte Carlo study of electron transport in Silicon nMOSFET inversion layers", presented at International Workshop on Computational Electronics, Phoenix, Arizona, USA.

[6] Wang, X., Chandramouli, V., Maziar, C. M. and Tasch, A. F. (1993). "Simulation program suitable for hot carrier studies: an efficient multiband Monte Carlo model using both full and analytical band structure for silicon", *J. Appl. Phys.*, **71**, 3339–3347.

[7] Shin, H., Yeric, C. M., Maziar, C. M. and Tasch, A. F. (1991). "Physically-based models for effective mobility and local-field mobility of electrons in MOS inversion layers", *Solid-State Electronics*, **34**, 545.

[8] Khan, S. A., Hasnat, K., Tasch, A. F. and Maziar, C. M. (1995). "Detailed evaluation of different inversion layer electron and hole mobility models", presented at *Proceeding of the Eleventh Biennial University Government Industry Microelectronics Meeting*, Austin, Texas.

[9] Fischetti, M. V. and Laux, S. E. (1993). "Monte Carlo study of electron transport in silicon inversion layers", *Phys. Rev. B*, **48**, 2244.

[10] Jallepalli, S., Bude, J., Shih, W.-K., Pinto, M. R. and Maziar, C. M., "Quantization of electrons and holes in silicon inversion layers beyond the effective mass approximation", Submitted to *Phys. Rev. B*.

[11] Hiroki, A., Odanaka, S. and Hori, A. (1995). "A high performance 0.1 μm MOSFET with asymmetric channel profile", *IEDM Tech. Dig.*, pp. 439–442.

Authors' Biographies

Wei-Kai Shih received the B.S. degree in physics from National Taiwan University in 1989 and M.A. degree in physics from the University of Texas at Austin in 1994. He is currently working towards a Ph.D. degree in electrical engineering at the University of Texas at Austin. His research interests include theoretical and computational study of hot-carrier effects, non-local phenomena, effects due to inversion-layer quantization on carrier transport and development of physical models for hydrodynamic and drift-diffusion simulation of deep submicron silicon MOSFETs.

Srinivas Jallepalli was born in Srikakulam, India. He received the B. Tech degree from the Indian Institute of Technology, Madras, in 1991 and the M.S.E. and Ph.D. degrees from the

University of Texas at Austin, in 1993 and 1996, respectively. He is currently with the Advanced Device Development group at STL, Motorola, Austin, TX. His research interest include device physics and modeling with special emphasis on the use of the Monte Carlo tool for advanced transport analysis and for studying the physics and effects of carrier transport in deep submicron MOS structures.

Mahbub Rashed received his Bachelor of Science in Electrical Engineering from Bangladesh University of Engineering and Technology in May, 1991. He joined in the Electrical and Computer Engineering department of the University of Texas at Austin in January 1992. He obtained his Masters degree from this department in the summer of 1993. He continued his education at the same university, pursuing a Ph.D. degree in Electrical Engineering. He focused on the applications of Monte Carlo tools for studying carrier transport in silicon and strained-Si/$Si_{1-x}Ge_x$ based devices. In December, 1996 he was awarded the Ph.D. degree. He is currently with the Semiconductor Technologies Laboratory, Motorola, Austin.

Christine Maziar received the B.S.E.E. (with Highest Distinction), the M.S.E.E. and the Ph.D. degrees from Purdue University in 1981, 1984 and 1986. In January of 1987, she joined the faculty of the University of Texas at Austin, where she is a Professor of Electrical and Computer Engineering and Vice Provost. She holds the Archie W. Straition Endowed Teaching Fellowship in Electrical and Computer Engineering. Her research interests include modeling of charge transport in high performance device structures, semiconductor device physics and device simulator enhancement. She has authored or coauthored over 100 journal or conference publications. She is a member of IEEE, ASEE, APS, Eta Kappa Nu, Tau Beta Pi and Sigma Pi Sigma.

Al. F. Tasch received his Ph.D. degree in Physics in 1969 form the University of Illinois, Urbana-Champaign. He worked in industry for 17 years at Texas Instruments and Motorola and joined the faculty of the Department of Electrical and Computer Engineering at the University of Texas at Austin in 1986 as a Chair Professor. He has been awarded 38 U.S. patents and was elected to the National Academy of Engineering in 1989.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 437–442
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Study on Possible Double Peaks in Cutoff Frequency Characteristics of AlGaAs/GaAs HBTs by Energy Transport Simulation

T. OKADA and K. HORIO*

Department of Electronic Information Systems, Faculty of Systems Engineering, Shibaura Institute of Technology, 307 Fukasaku, Omiya 330, Japan

By using an energy transport model, we simulate cutoff frequency f_T versus collector current density I_C characteristics of npn^-n AlGaAs/GaAs heterojunction bipolar transistors (HBTs) with various n^--collector thickness and n^--doping densities. It is found that the calculated f_T characteristics show double peak behavior when the n^--layer is thick enough and the n^--doping is high enough to allow existence of neutral n^--region. The mechanism of the double peak behavior is discussed by studying energy band diagrams, electron-energy profiles and electron-velocity profiles. Particularly, we discuss the origin of the second peak (at higher I_C) which is not usually reported experimentally.

Keywords: GaAs, heterojunction bipolar transistor, energy transport model, cutoff frequency, double peaks, velocity overshoot

1. INTRODUCTION

Recently, AlGaAs/GaAs heterojunction bipolar transistors (HBTs) have received great interest for application to high-speed and high-frequency devices. Since non equilibrium carrier transport becomes important in the HBTs, carrier energy should be considered in the modeling of them. For this purpose, the Monte Carlo simulation [1, 2] and so-called an energy transport model [3–7] have been applied to analyze the characteristics of AlGaAs/GaAs HBTs.

Cutoff frequency f_T is one of the figure-of-merits of HBTs' high-frequency performance. Usually, the cutoff frequency of AlGaAs/GaAs HBTs increases with the collector current density I_C and begins to decrease at a certain I_C, showing a single peak in the experimental $f_T - I_C$ characteristics. However, according to the simulation using a drift-diffusion model, the $f_T - I_C$ characteristics show a steep second peak in some cases [5, 8]. This is attributed to the fact that in the drift-diffusion approximation, electron mobility is given as a function of *local* electric field and the electron

*Corresponding author: Fax: +81-48-687-5198, e-mail: horio@sic.shibaura-it.ac.jp.

velocity versus electric field curve of GaAs shows a peak behavior. By the other simulation methods, the second peak has not been reported yet.

In this work, we have systematically and carefully analyzed the f_T characteristics of AlGaAs/GaAs HBTs by using an energy transport model [9] in which electron mobility is determined by an electron energy (not by local electric field) and velocity overshoot can be treated. As a result, we have found that the second peak can arise (at rather high current levels) also when using this model. Therefore, we discuss here the physical reason why the double peak behavior in the f_T characteristics arises.

2. PHYSICAL MODEL

2.1. Device Structure and Basic Equations

Figure 1 shows an npn^-n AlGaAs/GaAs HBT structure analyzed here. Al composition changes from 0.3 to 0 in the emitter and base regions, and so this is a graded band-gap base HBT. Doping densities are different in respective regions, and hence in general, transport parameters should be given as functions of Al composition, doping density and carrier energy. We have simulated the device characteristics as parameters of n^--collector thickness L_{C1} and its doping density N_{C1}.

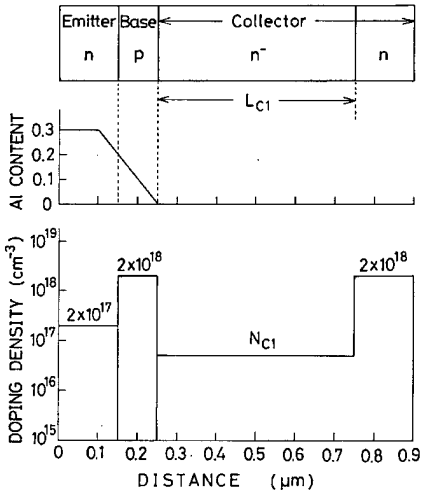


FIGURE 1 AlGaAs/GaAs HBT structure simulated in this study.

Next we describe the derivation of electron transport equations. Assuming parabolic energy bands, three conservation equations, that is, a particle conservation equation, a momentum conservation equation and an energy conservation equation are obtained by taking moments of the Boltzmann equation [10]. To make these equations tractable and applicable to the HBTs, the following assumptions are made [5]. Firstly, we adopt relaxation time approximation for collision terms. Secondly, we adopt an equivalent one-valley model. If we assume that electron drift energy is negligible small as compared to its thermal energy, the average electron energy w_n can be written as

$$w_n = \frac{3}{2} k T_n + F_U \Delta_{LU} \tag{1}$$

where T_n is electron temperature, F_U is upper-valley fraction and Δ_{LU} is energy difference between upper and lower valleys. To treat HBTs, some factors are also considered. Carrier recombination is treated by SRH statistics, and an effective field acting on electrons which arises due to the position dependence of band structure is included. Then the electron transport equations are simplified as follows.

$$\frac{dJ_n}{dz} = q U \tag{2}$$

$$J_n = -qnv_n = q\mu_n(w_n)\left\{n E_n + \frac{d}{dz}\left(n\frac{k T_n}{q}\right)\right\} \tag{3}$$

$$-\frac{d}{dz}\left\{\frac{J_n}{q}(w_n + kT_n)\right\} = J_n E_n - w_n U - n\frac{w_n - w_0}{\tau_w} \tag{4}$$

where U is the recombination rate, v_n is the average electron velocity, μ_n is the electron mobility, E_n is the effective field acting on electrons, τ_w is the energy relaxation time and w_0 is the equilibrium value of w_n. In the electron transport equations, parameters that should be given are

the electron mobility μ_n, the energy relaxation time τ_w and the upper-valley fraction F_U.

For holes, we use drift-diffusion type equations. In addition to the transport equations, we include Poisson's equation, which completes the basic equations for device simulation.

2.2. Transport Parameters

Here we describe methods of giving transport parameters such as μ_n, τ_w and F_U. To do this, usually, homogeneous bulk is assumed ($d/dz = 0$). Then, the following two equations are obtained from the previous electron transport equations.

$$v_n = -\mu_n E \tag{5}$$

$$w_n = w_0 - q\tau_w v_n E \tag{6}$$

where E is electric field. By using the Monte Carlo method, v_n, w_n and F_U are obtained as a function of electric field E. Thus, by using Eqs. (5) and (6), the energy-dependent μ_n, τ_w and F_U can be obtained.

In this work, we must give the transport parameters as functions of Al composition x, electron energy w_n and doping density N. To do this, we first evaluate μ_n, τ_w and F_U for $x = 0$, 0.05, 0.1, 0.15, 0.2, 0.25 and 0.3 by a Monte Carlo method. Doping densities are also varied. Once these parameters are available as fit curves or tables for a given doping density, parameters for any x between 0 and 0.3 can be obtained (as a function of w_n) by a linear extraporation method. That is, if x lies between x_1 and x_2, the parameter f is given by the following equation.

$$f(w_n) = f_1 + \frac{x - x_1}{x_2 - x_1}(f_2 - f_1) \tag{7}$$

where f_1 and f_2 are corresponding values of f for $x = x_1$ and x_2, respectively.

Next we show some examples of transport parameters estimated by a Monte Carlo method. Figure 2 shows electron mobility μ_n, energy relaxation time τ_w and upper valley fraction F_U as a parameter of Al composition x. The doping density is 2×10^{18} cm^{-3}. Here, we consider L-valley as upper valley and use L-valley's parameters for Δ_{LU} and upper-valley effective mass. We also treat a case somewhat considering X-valley's contribution, but x and energy dependences of estimated parameters are essentially similar to those shown in Figure 2 [9]. As described before, once the figures like Figure 2 are obtained for a given doping density, the transport parameters for any x and for any electron energy are obtained by the linear extraporation method. Of course, when the doping density becomes different, another figure is required. Furthermore, the parameters must be evaluated in every mesh point. However, this approach is simple in itself and can be easily implemented.

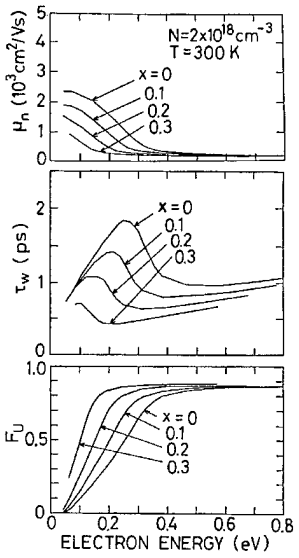


FIGURE 2 Electron mobility μ_n, energy relaxation time τ_ω and upper-valley fraction F_U versus electron energy ω_n curves as a parameter of x in $Al_xGa_{1-x}As$, calculated by a Monte Carlo method. The energy difference between upper and lower valleys Δ_{LU} and upper-valley effective mass m_U are set to $0.284 - 0.605x$ (eV) and $0.23m_0$, respectively.

3. SIMULATED $f_T - I_C$ CHARACTERISTICS

We calculate $f_T - I_C$ characteristics of the AlGaAs/GaAs HBT as parameters of the n^- collector

thickness L_{C1} and its doping density N_{C1}. Here f_T is calculated from the following equation.

$$f_T = (1/2\pi)(\partial I_C/\partial Q_n)_{V_{CE}} \quad (8)$$

where Q_n is electron charges in the device and V_{CE} is the collector-emitter voltage. V_{CE} is set to 1.5 V in this study.

Figure 3 shows calculated $f_T - I_C$ characteristics of the HBT as a parameterof n^--layer doping density N_{C1}, where L_{C1} is set to 0.5 μm. Figure 4 shows calculated $f_T - I_C$ characteristics of the HBT as a parameter of n^--layer thickness L_{C1}, where N_{C1} is set to 5×10^{16} cm^{-3}. As I_C increases, f_T increases because the emitter charging time and the collector charging time are reduced. From these figures, we see that for lower N_{C1} (10^{16} cm^{-3} in Fig. 3) and for thinner L_{C1} (0.1 μm and 0.2 μm in Fig. 4), f_T characteristics show a single peak. These cases correspond to the situation that n^--collector layer is almost or fully depleted already when the base-emitter voltage (V_{BE}) is 0 V. In the other cases,two peaks are clearly seen in the f_T characteristics. In these cases, neutral n^--region exists at $V_{BE} = 0$ V.

Up to the first peak, f_T is higher for higher N_{C1}, as seen from Figure 3. Also, the value of I_C where f_T begins to decrease is higher for higher N_{C1}. These are because around the peak region, the transit time through n^--collector depletion layer (which is thinner for higher N_{C1}) is dominant [5] and f_T begins to decrease due to a high injection effect which leads to expanding the depletion layer and increasing the collector transit time. As seen in Figure 4, the f_T characteristics are essentially similar between the two cases with $L_{C1} = 0.5$ μm and 0.7 μm. This is because the thickness of n^--collector depletion layer is determined by N_{C1}. As shown in Figure 3, when N_{C1} is higher, the value of I_C where f_T takes the second peak is higher. We will discuss below why the f_T characteristics show double peak behavior.

Figure 5 shows energy band diagrams as a parameter of I_C for the HBT where $N_{C1} = 5 \times 10^{16}$ cm^{-3} and $L_{C1} = 0.5$ μm. Figures 6 and 7 show the corresponding electron-energy profiles and electron-velocity profiles, respectively. In these figures $I_C = 10^4$ A/cm^2, 3×10^4 A/cm^2, 8×10^4 A/cm^2, 1.2×10^5 A/cm^2 and 1.5×10^5 A/cm^2 correspond to the regions before the first peak, around the first peak,

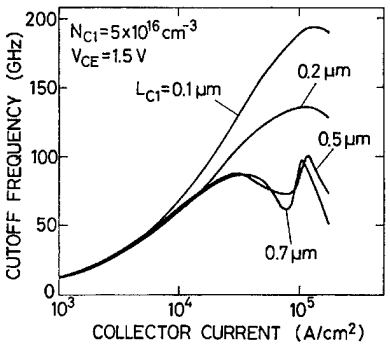


FIGURE 4 Calculated $f_T - I_C$ curves of AlGaAs/GaAs HBTs as a parameter of n^--collector thickness L_{C1}. The n^--collector doping density N_{C1} is 5×10^{16} cm^{-3}.

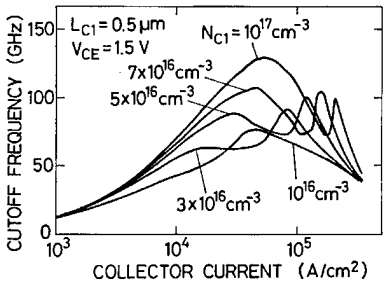


FIGURE 3 Calculated $f_T - I_C$ curves of AlGaAs/GaAs HBTs as a parameter of n^--collector doping density N_{C1}. The n^--collector thickness L_{C1} is 0.5 μm.

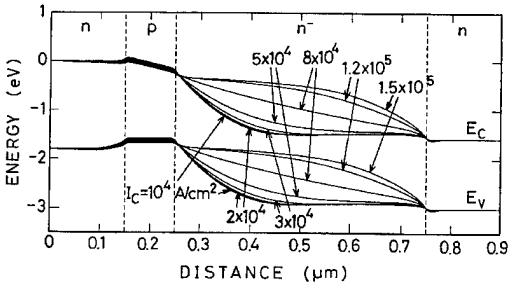


FIGURE 5 Energy band diagrams of an AlGaAs/GaAs HBT with $L_{C1} = 0.5$ μm and $N_{C1} = 5 \times 10^{16}$ cm^{-3} as a parameter of I_C. $V_{CE} = 1.5$ V.

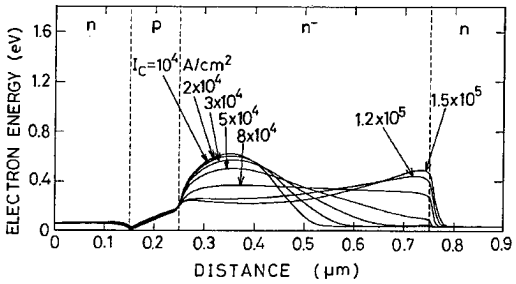


FIGURE 6 Electron-energy profiles of an AlGaAs/GaAs HBT, corresponding to Figure 5. $L_{C1}=0.5$ μm and $N_{C1}=5\times10^{16}$ cm^{-3}.

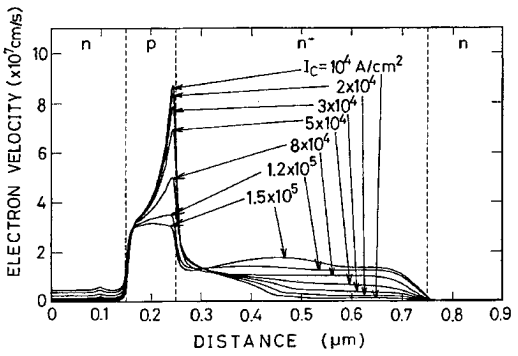


FIGURE 7 Electron-energy profiles of an AlGaAs/GaAs HBT, corresponding to Figure 5. L_{C1} = 0.5 μm and N_{C1} = 5 × 10^{16} cm^{-3}.

at the local minimum, at the second peak, and after the second peak respectively, in the f_T-I_C characteristics. From these, we interpret the double peak behavior in the following way.

As is understood from Figure 5, the first peak arises because the depletion layer in the n^--collector layer expands due to an high injection effect, and hence the collector transit time increases. The fall of I_C should last until the n^--collector layer becomes entirely depleted. Around $I_C=8\times10^4$ A/cm^2, the n^--layer is entirely depleted as seen from Figure 5 and at this current, f_T characteristics show the local minimum. As can be seen from the band diagrams, the electric field at the n^--collector layer near the base becomes weaker when I_C increases further. Hence, as shown in Figure 6, the electron energy in this region becomes lower. Consequently, the electron mobility becomes high, leading to the higher electron velocity than the saturation velocity, for example, at $I_C=1.2\times10^5$ A/cm^2 as shown in Figure 7. Therefore, the collector transit time becomes shorter temporarily and hence f_T begins to increase again. In the end, however, f_T falls because of the base-push-out effect (Kirk effect) which leads to a lower electron velocity around the base-collector interface and higher collector capacitance due to the injected electrons whose densities become higher than N_{C1}. Thus the second peak arises. As is evident from the above discussion, the value of I_C where f_T shows a second peak becomes higher for higher N_{C1}.

As described above, we have shown theoretically that double peak behavior can be seen in the f_T-I_C characteristics of AlGaAs/GaAs HBTs. Physical mechanism of this behavior has been explained.

4. CONCLUSION

By using an energy transport model, we have simulated f_T-I_C characteristics of AlGaAs/GaAs HBTs with various n^--collector thickness and n^--doping densities. It is found that the calculated f_T characteristics show the double peak feature when the n^--layer is thick enough and the n^--doping is high enough to allow existance of the neutral n^--region. It is interpreted that the first peak arises because the depletion region in the n^--layer begins to expand due to a high injection effect and the collector transit time increases. The fall of f_T lasts until the n^--layer becomes entirely depleted. When the base voltage is raised further, the electric field in the n^--layer near the base becomes lower, leading to the lower electron energy there. Then, the electron velocity in the n^--layer becomes higher, resulting in shorter collector transit time. Therefore, f_T begins to increase again. Finally, f_T decreases due to the base-push-out effect (Kirk effect), resulting in the second peak. We can say that the double peak behavior can be seen in the f_T

characteristics of real AlGaAs/GaAs HBTs if the collector current is raised rather high.

References

[1] Tomizawa, K., Awano, Y. and Hashizume, N. (1984). "Monte Carlo simulation of AlGaAs/GaAs heterojunction bipolar transistors", *IEEE Electron Device Lett.*, **EDL-5**, 362–364.

[2] Katoh, R., Kurata, M. and Yoshida, J. (1989). "Self-consistent particle simulation for (AlGa)As/GaAs HBT's with improved base-collector structures", *IEEE Trans. Electron Devices*, **36**, 846–853.

[3] Horio, K. and Yanai, H. (1987). "Numerical modeling of energy transport effects in AlGaAs/GaAs heterojunction bipolar transistors", *Proceedings of NASECODE V*, pp. 231–236.

[4] Azoff, E. M. (1989). "Energy transport numerical simulation of graded AlGaAs/GaAs heterojunction bipolar transistors", *IEEE Trans. Electron Devices*, **36**, 609–616.

[5] Horio, K., Iwatsu, Y. and Yanai, H. (1989). "Numerical simulation of AlGaAs/GaAs heterojunction bipolar transistors with various collector parameters", *IEEE Trans. Electron Devices*, **36**, 617–624.

[6] Tomizawa, K. and Pavlidis, D. (1990). "Transport equation approach for heterojunction bipolar transistors", *IEEE Trans. Electron Devices*, **37**, 519–529.

[7] Teeter, D. A., East, J. R., Mains, R. K. and Haddad, G. I. (1993). "Large-signal numerical and analytical HBT models", *IEEE Trans. Electron Devices*, **40**, 837–845.

[8] Chen, J., Gao, G. B., Unlu, M. S. and Morkoc, H. (1991). "High-frequency output characteristics of AlGaAs/GaAs heterojunction bipolar transistor for large-signal applications", *Solid-State Electron.*, **34**, 1263–1273.

[9] Nakatani, A. and Horio, K. (1995). "Energy transport modeling of HBTs considering composition-, doping- and energy-dependence of transport parameters", *Inst. Phys. Conf. Ser.*, **141**, 633–638.

[10] Blotekjaer, K. (1970). "Transport equations for electrons in two-valley semiconductors", *IEEE Trans. Electron Devices*, **ED-17**, 38–47.

Authors' Biographies

Tadayuki Okada received the M.E. degree in electrical engineering from Shibaura Institute of Technology, Japan in 1997. His research was concerned with energy transport simulation of AlGaAs/GaAs HBTs. He now works at Toshiba Information Systems Corporation, Japan. He is a member of the Japan Society of Applied Physics.

Kazushige Horio received the Ph.D. degree in electronic engineering from University of Tokyo, Japan in 1982. Now, he is a Professor at Department of Electronic Information Systems in Shibaura Institute of Technology, Japan. His research interests include modeling and simulation of high-speed and high-frequency devices such as GaAs MESFETs, HBTs and HEMTs, and simulation methods for heterojunctions and energy transport effects. He has authored more than 50 international papers in these fields. He is a senior member of IEEE and a member of the Japan Society of Applied Physics. He is listed in *Who's Who in the World.*

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 443–447
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Shell-Filling Effects in Circular Quantum Dots

M. MACUCCI [a,*] and KARL HESS [b]

[a] *Dipartimento di Ingegneria dell' Informazione, Via Diotisalvi, 2 I-56126 Pisa, Italy;*
[b] *Beckman Institute, 405 N Mathews, Urbana, IL 61801*

We investigate the dependence of the capacitive energy associated with a circular quantum dot on the number of electrons in the dot and on the parameters defining the potential confinement. Our results reproduce the shell-filling behavior that has recently been experimentally observed and allow us to estimate the shape of the confinement potential and the dot size of the experiments.

Keywords: Quantum dot, shell-filling, charging energy, density functional

1. INTRODUCTION

Recent advances in nanofabrication techniques have allowed a more detailed investigation of the electronic structure of quantum dots. In particular, it has become possible to measure, with greater precision and reliability, the values for the charging energies in a quantum dot. Tarucha *et al.* [1] have fabricated a circular quantum dot in which vertical confinement has been achieved by means of AlGaAs barriers 12 nm apart, while lateral confinement has been obtained by etching and by means of the depletion due to a lateral Schottky contact. Tarucha *et al.* have evaluated the charging energies, which they improperly name "addition energies", by measuring the $I-V$ characteristic for several different values of the gate voltage: in this way it is possible to construct the "diamond diagram", typical of Coulomb blockade studies, which yields directly the values for the charging energies.

For the first time, the results in Ref. [1] show atomic-like shell-filling effects in a quantum dot, as predicted in Refs. [2], [4]. In order to understand the meaning of these results, we must first define rigorously what the charging energy is. We can start from the classical definition of the capacitive energy, i.e., the variation in the potential energy of a capacitor as a consequence of the addition of charge. For a quantum dot the charging energy can be defined as the variation of the chemical potential as a consequence of the addition of one electron: $E_C=\mu(N+1)-\mu(N)$. If the dot were a classical capacitor, this would be a constant

* Corresponding author.

quantity; in the quantum case we can define a quantum capacitance [2], depending on the number N of electrons contained in the dot, as $C(N)=e^2/E_C$, where e is the charge of the electron.

As long as we add electrons to the same orbital, the behavior of the quantum dot is quite similar to that of a classical capacitor: the charging energy is substantially constant and depends mostly on the geometrical dimensions. However, when we start populating a new orbital, the charging energy exhibits a peak, due to the variation in the quantum confinement energy. Thus, the charging energy is about constant, with a peak every time that a new shell opens. In a rectangular dot, for example, each orbital accomodates only two electrons, so that the charging energy has an oscillating behavior, as a function of the electron number. A circular dot, instead, has several degeneracies, which lead to a much smaller number of peaks, as will be discussed in the following.

2. MODEL

We have considered a 2D model of a quantum dot characterized by parabolic confinement and hard walls located at a distance from the center corresponding to the geometrical radius. For purposes of easy comparison with the experimental results, we assume the parabolic potential given by $V=1/2\, m^*\omega^2 r^2$, where r is the radial coordinate and m^* is the electron effective mass, and express the slope of such potential by means of the product $\hbar\omega$, $\hbar$ being the reduced Planck constant. The 2D approximation is allowed by the fact that the vertical dimension of the dot (12 nm) is much smaller than the lateral radius, therefore only the lowest state along the vertical direction is occupied, and this represents just a constant shift in the energy, which does not affect the results for the charging energy. The 2D Schrödinger equation can be separated [3] into a straightforward angular equation and into a radial equation, which is solved by means of a standard numerical finite-difference procedure. The two equations read:

$$\frac{d^2\Phi(\phi)}{d\phi^2}+\nu^2\Phi(\phi)=0 \tag{1}$$

$$\frac{d^2P(\rho)}{d\rho^2}+\frac{1}{\rho}\frac{dP(\rho)}{d\rho^2}-\frac{2m^*}{\hbar^2}V(\rho)P(\rho)+\left(k^2-\frac{\nu^2}{\rho^2}\right)P(\rho)=0, \tag{2}$$

where ν^2 is the constant for the separation of variables and $k^2=2m^*E/\hbar^2$, E being the energy eigenvalue. If we enforce periodic boundary conditions, the angular equation has the simple solution $\Phi=c\exp(\pm i\phi)$. When $\nu^2=0$, there is just one solution, while for $\nu^2\neq 0$ two different degenerate solutions exist, corresponding to the up and down orientations of the angular momentum.

The potential $V(\rho)$ is computed self-consistently at each iteration from the electron density obtained at the previous iteration. It includes also a term accounting for the exchange and correlation interactions, in the framework of a local density functional approach [3]. The iterative procedure is stopped when the electron density differs from that of the preceding iteration less than a given amount. Obtaining convergence is simple as long as the Coulomb interaction among the electrons represents just a perturbation of the quantum confinement energy. This condition is satisfied for very small dots, since the electrostatic interaction energy scales with the inverse of the distance, while the quantum confinement energy scales with the inverse of the square of the dot size. Thus, for dots larger than about 100 nm, particular care has to be taken to ensure convergence, with the usage of underrelaxation techniques, which help controlling the instabilities typical of fixed-point iteration schemes.

The chemical potential is obtained applying Slater's transition rule [3], which allows better numerical precision than that obtainable from a differentiation of the total energy.

3. NUMERICAL RESULTS AND DISCUSSION

The reason why we have chosen the particular potential landscape described in the previous section is that it qualitatively explains a feature observed in the experimental results: the charging energy tends to saturate to a limiting value as if the size of the dot were growing while the number of electrons is increased, until any further growth is limited by the presence of hard walls.

We have first considered a quantum dot with a geometrical radius of 90 nm and we have computed the capacitive energy for three different slopes of the parabolic confinement potential. The results are reported in Figure 1, where the charging energy is plotted versus the number of electrons in the dot for $\hbar\omega = 4$ meV (thick line), 3 meV (thin line) and 2.5 meV (dashed line). In all three cases the capacitive energy tends to decrease and to saturate down to a limiting value. This is the consequence of the variation of the effective dot size (defined as the area over which the electron density is nonnegligible) as a function of the number of electrons: when there are just a few electrons, they tend to concentrate in the deepest region of the potential landscape, near the dot center, while for larger numbers of electrons such minimum is effectively screened and the charge is distributed over a larger area, up to the maximum size allowed by the hard walls. The classical charging energy E_{cc} for a conducting disk with a radius of 90 nm is given [5] by $E_{cc} = e^2/(8R\varepsilon_0\varepsilon_r) = 1.94$ meV, where R is the disk radius, ε_0 is the vacuum permittivity and ε_r is the relative permittivity of the medium. This is in good agreement with the limiting value of approximately 2 meV obtained from the simulation. Another major feature of the computed charging energy consists in the presence of peaks for $N = 2, 6, 12, 16, 20, 24$, corresponding to those found in the experimental results. Each of them is associated with the filling of a shell (orbital) and its height depends on the difference between the energy eigenvalue of the new shell and that of the previous one. If the confinement potential were perfectly parabolic, without any contribution from electron-electron interaction, there would be peaks only for $N = 2, 6, 12, 20, 30$, because more degeneracies are present with a parabolic potential than with a generic potential characterized by circular symmetry. Circular symmetry only implies that when $\nu = 0$ there is a double degeneracy (the spin degeneracy) and when $\nu \neq 0$ the degeneracy is fourfold (spin degeneracy plus angular momentum degeneracy). We do not observe any peak for $N = 10$, because the energy splitting between the two related orbitals is extremely small (see Ref. [6] for more details).

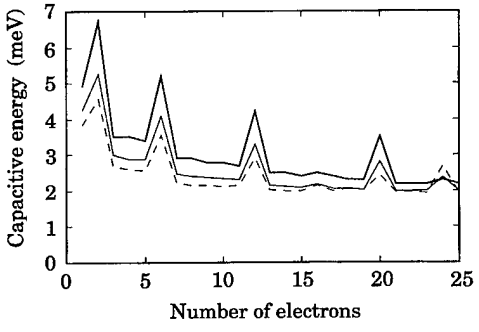


FIGURE 1 Charging energy for a quantum dot with a radius of 90 nm versus the number of electrons. The bare confinement potential is parabolic with $\hbar\omega = 4$ meV (thick line), 3 meV (thin line), 2.5 meV (dashed line).

The saturation of the charging energy towards a limiting value determined by the maximum dot size is apparent in Figure 2, where we report the capacitive energies for $\hbar\omega = 3$ meV and a dot radius of 75 nm (thick line), 90 nm (thin line) and 120 nm (dashed line). When the number of electrons in the dot is small, the effective size is independent of the geometrical radius and is determined only by the shape of the parabolic potential, therefore for small N the three curves overlap. For large numbers of electrons, instead, the three curves tend to saturate to the corresponding classical capacitive energies (the curve for $R = 120$ nm does not reach the limiting value for the maximum number of electrons included in our plot).

The experimental results by Tarucha *et al.* show a saturation value of approximately 1.25 meV, which would correspond to a classical radius of

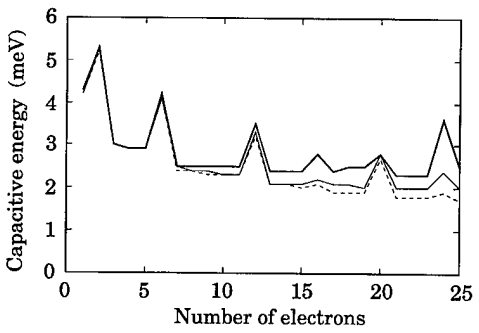


FIGURE 2 Charging energy for a quantum dot with a parabolic confinement potential defined by $\hbar\omega = 3$ meV and with a geometrical radius of 75 nm (thick line), 90 nm (thin line) and 120 nm (dashed line).

140 nm. We have therefore performed a calculation for a dot with a geometrical radius of 140 nm, choosing a confinement potential with $\hbar\omega = 2.48$ meV, on the basis of the best possible agreement between the results of the simulation and the experimental data. The computed capacitive energy as a function of the electron number is reported in Figure 3: the only significant discrepancy with the experimental results is represented by the absence of the peak for $N = 4$, which cannot be explained if we assume rigorous circular symmetry. It is however possible that some irregularity exists in the confinement potential of the experimental dot and that it produces measurable effects only for small numbers of electrons, while it is screened out for larger values of N.

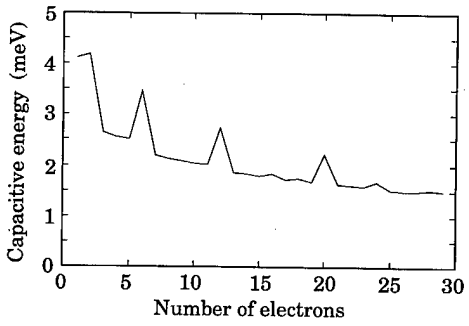


FIGURE 3 Charging energy for a quantum dot with a parabolic confinement potential defined by $\hbar\omega = 2.48$ meV and with a geometrical radius of 140 nm.

4. CONCLUSIONS

The charging energy for a circular model quantum dot has been evaluated with the inclusion of exchange and correlation effects. The numerical results are in quantitative agreement with experimental data that have recently been obtained and reproduce the position of the observed peaks, corresponding to shell-filling events.

Acknowledgements

This work has received partial support from the NATO Collaborative Research Grant n. 950753. One of the authors (M. M.) acknowledges support from the Italian Ministry of the University and Scientific Research. K. H. acknowledges the support of the US Army Research Office.

References

[1] Tarucha, S., Austing, D. G., Honda, T., van der Hage, R. J. and Kouwenhoven, L. P. (1996). "Shell Filling and Spin Effects in a Few Electron Quantum Dot", *Phys. Rev. Lett.*, **77**, 3613.

[2] Macucci, M., Hess, K. and Iafrate, J. G. (1993). "Electronic Energy Spectrum and the Concept of Capacitance in Quantum Dots", *Phys. Rev. B*, **48**, 17354.

[3] Macucci, M.; Hess, K. and Iafrate, G. J. (1995). "Simulation of electronic properties and capacitance of quantum dots", *J. Appl. Phys.*, **77**, 3267.

[4] Iafrate, G. J., Hess, K., Krieger, J. B. and Macucci, M. (1995). "Capacitive nature of atomic-sized structure", *Phys. Rev. B*, **52**, 10737.

[5] van Houten, H., Beenakker, C. W. J. and Staring, A. A. M. (1992). In "Single Charge Tunneling", edited by Grabert, H. and Devoret, M. H. (Plenum Press, New York).

[6] Macucci, M., Hess, K. and Iafrate, G. J. (1997). "Numerical simulation of shell-filling effects in circular quantum dots", *Phys. Rev. B*, **55**, R4879.

Authors' Biographies

Massimo Macucci is serving on the faculty of the Electrical Engineering Department at the University of Pisa, Italy. His research interests include quantum-interference and single-electron devices, Coulomb Blockade phenomena, modeling and measurements of noise in electron devices.

Karl Hess holds the Swanlund Chair in Electrical and Computer Engineering at the University of Illinois in Urbana-Champaign. His main interests are in computational electronics, optoelectronics and most recently in the physics of semiconductor nanostructures.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 449–453
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Modeling of Shot Noise in Resonant Tunneling Structures

G. IANNACCONE*, M. MACUCCI and B. PELLEGRINI

Dipartimento di Ingegneria dell' Informazione: Università degli studi di Pisa, Via Diotisalvi 2, I-56126 Pisa, Italy

In this paper, we present insights into the transport properties and the geometrical structure of resonant tunneling devices that can be obtained by the study of their noise properties. We stress the importance of including noise behavior among the objectives of device simulations. The reason is twofold: on one hand, as the number of carriers involved in device operation decreases, fluctuations become more relevant; on the other hand, in devices whose functionality is based on quantum effects, noise properties strongly depend on the details of device geometry.

Keywords: Resonant tunneling, nanoelectronics, device modeling, noise modeling, shot noise

1. INTRODUCTION

In recent years, noise characterization has emerged as a powerful tool for obtaining information about the structure and the transport properties of nanoscale devices complementary to those given by the DC characteristics and the small signal AC response. In fact, since the number of charge carriers involved in device operation is decreased with respect to semiclassical devices, it is apparent that fluctuations, and in particular those due to the granularity of charge (the so-called "shot noise"), acquire an increasing importance. Furthermore, noise in such structures exhibits a behavior which is strongly dependent upon the details of device geometry.

Therefore, it is important to extend the domain of device simulations to noise properties. Here, we focus on resonant tunneling structures, for which a greater number of experimental data is available.

Since the pioneering work of Lesovik [1] and the first experimental results of Li and coworkers [2], many theoretical studies [3–6] and experimental results [5, 7, 8] have appeared in the literature, assessing that the power spectral density S of the noise current in such devices may be suppressed down to half the "full" shot noise value $S_{\text{full}} = 2q\langle i\rangle$, due to correlation in the motion of individual electrons introduced by electrostatic force and/or Pauli exclusion.

In this paper, we presents insights into the transport properties and the geometrical structure

*Corresponding author.

of resonant tunneling devices that can be obtained by the study of their noise properties. Our calculations are based on a model for transport and noise in generic resonant tunneling structures which has been presented elsewhere [6, 9], so that only a brief description will be given in Section 2, while the interested reader can find analytical details in the cited papers. Numerical results and a comparison with available experimental results will be shown in Section 3, while a discussion of the results obtained and of the future developments ends the paper.

2. MODEL

Let us consider the one-dimensional structure sketched in Figure 1: it consists of three regions Ω_l, Ω_w and Ω_r, i.e., the left reservoir, the well region and the right reservoir, respectively, that are only weakly coupled through the two tunneling barriers 1 and 2. Moreover, we suppose that electron transport is well described in terms of sequential tunneling: an electron in Ω_l traverses barrier 1, loses phase coherence and relaxes to a quasi-equilibrium energy distribution in the well

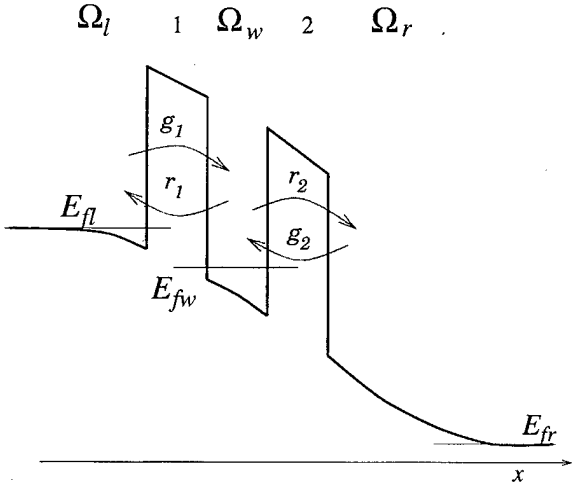


FIGURE 1 A generic resonant tunneling structure consists of three isolated regions Ω_l, Ω_w, Ω_l weakly coupled by tunneling barriers, indicated here with 1 and 2. Coupling between different regions has to be small enough to be treated with first order perturbation theory.

region Ω_w, then traverses barrier 2 and leaves through Ω_r. Such hypothesis is very reasonable, except at millikelvin temperatures, where inelastic processes are strongly suppressed and no more effective in thermalizing electrons in the well.

The typical resonant current peaks in the I–V curve are due to the shape of the density of longitudinal states in Ω_w, which is strongly affected by confinement: for the material parameters considered here, it has a single narrow peak in correspondence of the allowed longitudinal energy level of Ω_w; the rate of inelastic scattering processes affects the width of such peak and in our model this effect is taken into account through a phenomenological parameter, the mean free path l, which plays the role of a relaxation length for phase and energy. The density of states in the well is calculated using a compact formula derived in Ref. [11].

A state in $\Omega_s (s = l, r, w)$ is characterized by its longitudinal energy E, its transverse wave vector $\boldsymbol{k}_T$ and its spin σ. Tunneling is treated as a transition between levels in different regions [10] in which E, $\boldsymbol{k}_T$ and σ are conserved.

Following Davies *et al.* [4], we introduce the "generation" rates g_1 and g_2, i.e., the transition rates from Ω_l to Ω_w and from Ω_r to Ω_w, obtained by summing the transition rates given by the Fermi golden rule over all pairs of occupied states in the initial region and available states in the final region. Analogously, we define the "recombination" rates, i.e., the transition rates r_1, from Ω_w and Ω_l, and r_2, from Ω_w to Ω_r, the total generation rate $g = g_1 + g_2$ and the total recombination rate $r = r_1 + r_2$. We refer the reader to Ref. [6] for detailed analytical expression.

The occupation factor in the well, which, under the assumption of complete relaxation, depends only on the value of the quasi-fermi level E_{fw} in the well, has to be calculated in the steady state condition, i.e., by imposing $\langle g \rangle = \langle r \rangle$, where we denote the steady state value of a generic quantity a as $\langle a \rangle$. The average current is

$$\langle i \rangle = q\langle g_1 - r_1 \rangle = q\langle r_2 - g_2 \rangle. \qquad (1)$$

The problem of transport is solved self-consistently, since the charge accumulated in the well affects the conduction band profile of the structure. We also need to obtain the complete curve of g and r as a function of the number of electrons in the well N (that is to say, of the quasi-Fermi level in the well). Then, we expand g and r to first order in N, around the steady state value $\tilde{N}$ and introduce the following characteristic times:

$$\frac{1}{\tau_{g1}} \equiv -\left.\frac{dg_1}{dN}\right|_{N=\tilde{N}}, \qquad \frac{1}{\tau_{g2}} \equiv -\left.\frac{dg_2}{dN}\right|_{N=\tilde{N}}, \tag{2}$$

$$\frac{1}{\tau_{r1}} \equiv \left.\frac{dr_1}{dN}\right|_{N=\tilde{N}}, \qquad \frac{1}{\tau_{r2}} \equiv \left.\frac{dr_2}{dN}\right|_{N=\tilde{N}}; \tag{3}$$

from which we can define

$$\tau_1^{-1} \equiv \tau_{g1}^{-1} + \tau_{r1}^{-1}, \quad \tau_2^{-1} \equiv \tau_{g2}^{-1} + \tau_{r2}^{-1} \tag{4}$$

and

$$\tau^{-1} = \tau_1^{-1} + \tau_2^{-1}. \tag{5}$$

The power spectral density $S(\omega)$ of the noise current at low frequency ($\omega\tau \ll 1$) can be written as

$$S(\omega) = 2q^2\left(\frac{\tau^2\langle g_1 + r_1\rangle}{\tau_2^2} + \frac{\tau^2\langle g_2 + r_2\rangle}{\tau_1^2}\right); \tag{6}$$

a detailed derivation of this result can be found in Ref. [6]; it suffices here to say that no additional hypothesis is required to arrive at (6).

An important parameter is the noise suppression factor γ, also called "Fano factor", i.e., the ratio between $S(\omega)$ and the "full shot" noise value $S_{\text{full}} = 2q\langle i\rangle$. From (1), (4), (5) and (6), it is apparent that it can reach a minimum of $\gamma = 0.5$ if $\tau_1 = \tau_2$, $\langle g_2\rangle \ll \langle r_2\rangle$ and $\langle r_1\rangle \ll \langle g_1\rangle$.

3. RESULTS

We consider a device with the following layer structure: a Si-doped ($N_d = 1.4 \times 10^{18}\,\text{cm}^{-3}$) 500 nm-thick GaAs buffer layer, an undoped 20 nm-thick GaAs spacer layer to prevent silicon diffusion into the barrier, an undoped 11.5 nm-thick AlGaAs barrier (barrier 2), an undoped 5 nm-thick GaAs quantum well, an undoped 10 nm-thick AlGaAs barrier (barrier 1), an undoped 10 nm-thick GaAs spacer layer and a Si-doped 500 nm-thick cap layer. The aluminum mole fraction in both barriers is *0.36*. Noise measurements on such structure as a function of current and temperature have been presented in Ref. [8].

In Figure 2 the experimental forward $I-V$ characteristic at 77 K is compared with the simulation result. The mean free path is chosen as a fitting parameter, and is equal to 15 nm. We have found that the so called "peak to valley ratio", i.e., the ratio between the peak current and the valley current, is almost linearly dependent on the mean free path, while the other parameters are practically independent. For the structure being investigated, a mean free path close to 15 nm seems to be the best fit at all considered temperatures (since the mean free path accounts for all randomizing effects, it seems that some temperature independent cause, e.g., interface roughness, is predominant); therefore, the decrease of the peak to valley ratio with increasing temperature seems to be due only to the Fermi distribution spreading.

The noise current power spectral density S as a function of current is plotted in Figure 3 for three different temperatures, and for very low bias. As expected, at equilibrium S tends to the thermal

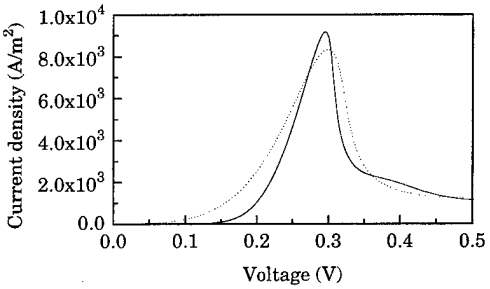


FIGURE 2 Forward $I-V$ curves at 77 K: comparison between experiment (solid) and simulation with mean free path 15 nm (dotted).

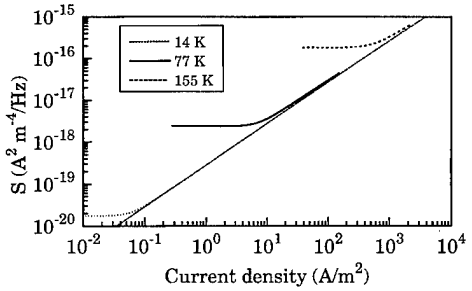


FIGURE 3 Noise current power spectral density as a function of current density for very low biases at three different temperatures: 14 K (dotted), 77 K (solid), 155 K (dashed). The straight line is $S_{\text{full}} = 2q\langle i\rangle$.

noise value $S = 4Gk_BT$, while, as the bias is increased, it approaches the $2q\langle i\rangle$ curve. This is an important check for the validity of our noise model.

In Figure 4, the noise suppression factor γ as a function of current in the resonant region of the $I-V$ curve is plotted for different temperatures and compared with the experimental results. As

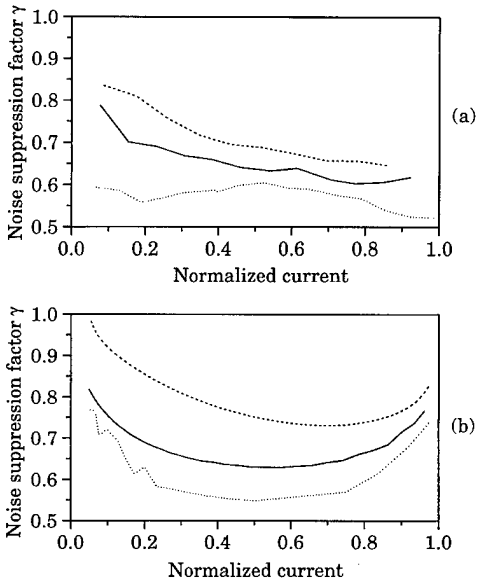


FIGURE 4 Noise suppression factor as a function of normalized current at the temperatures of 14 K (dotted), 77 K (solid), 155 K (dashed): comparison between experiments (a) and simulations (b). Biases lower than that of the resonant peak are considered and the current at the resonant peak is taken as unity.

can be seen, γ increases with increasing temperature. This behavior is described in [6].

For the same structure, in Figure 5, γ is plotted as a function of the applied voltage at the temperature of 77 K. For bias voltages in the negative differential resistance (NDR) region, enhanced shot noise is to be expected, while for voltages greater than that corresponding to the valley current, full shot noise is obtained (Pauli exclusion and Coulomb repulsion are no more effective in correlating current pulses). For the latter result, an experimental evidence can be found in [8], while, for the former, measurements are in progress.

The enhanced shot noise in the NDR region depends on the fact that the characteristic time τ_{g1} is negative, i.e., the transition rate g_1 increase with increasing N. The reason is that the peak in the density of states is below the conduction band bottom of the cathode: when an electron enters the well, the conduction band bottom of the well is raised, and more states are available for tunneling from the left electrode, so that electron crossings through the whole structure are positively correlated.

4. DISCUSSION

We have shown that numerical modeling of noise in resonant tunneling devices provides new insights into device physics and structure, allows us to recover the results of experiments with a reason-

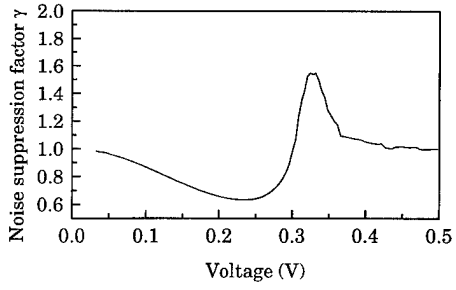


FIGURE 5 Theoretical noise suppression factor at 77 K as a function of voltage.

able accuracy and to predict interesting new results. The results obtained give us confidence in the model for transport and noise in resonant tunneling structures described in [6, 9, 11].

Further developments will include a more refined model to predict experimental data with better accuracy; in particular, greater attention will be devoted to the connection between regions in which transport can be described as semiclassical and regions in which the quantum nature of transport must be taken into full account.

Acknowledgements

This work has been supported by the Ministry for the University and Scientific and Technological Research of Italy and by the Italian National Research Council (CNR).

References

[1] Lesovik, G. B. (1989). "Excess Quantum Noise in 2D Ballistic Point Contacts", *JETP Lett.*, **49**, 592–594 (*Pis'ma Zh. Eksp. Teor. Fiz.*, **49**, 513).

[2] Li, Y. P., Zaslavsky, A., Tsui, D. C., Santos, M. and Shayegan, M. (1990). "Noise Characteristic of Double-Barrier Resonant-Tunneling Structures Below 10 KHz", *Phys. Rev. B*, **41**, 8388–8391.

[3] Büttiker, M. (1992). "Scattering Theory of Current and Intensity Noise Correlations in Conductors and Wave Guides", *Phys. Rev. B*, **46**, 12485–12507.

[4] Davies, J. H., Hyldgaard, P., Hershfield, S. and Wilkins, J. W. (1992). "Classical Theory of Shot Noise in Resonant Tunneling", *Phys. Rev. B*, **46**, 9620–9633.

[5] Brown, E. R. (1992). "Analytic Model of Shot Noise in Double-Barrier Resonant Tunneling Diodes", *IEEE Trans. Electron Devices*, **39**, 2686–2693.

[6] Iannaccone, G., Macucci, M. and Pellegrini, B. (1997). "Shot Noise in Resonant Tunneling Structures", *Phys. Rev. B*, **55**, 4539–4550.

[7] Liu, H. C., Li, J., Aers, G. C., Leavens, C. R. and Buchanan, M. (1995). "Shot-Noise Suppression in Resonant Tunneling", *Phys. Rev. B*, **51**, 5116–5120.

[8] Ciambrone, P., Macucci, M., Iannaccone, G., Pellegrini, B., Sorba, L., Lazzarino, M. and Beltram, F. (1995). "Noise Measurements in Resonant Tunneling Structures as a Function of Current and Temperature", *Electronics Lett.*, **31**, 503–504.

[9] Iannaccone, G. and Pellegrini, B. (1995). "Unified Model to Electron Transport in Double Barrier Structures", *Phys. Rev. B*, **52**, 17406–17412.

[10] Bardeen, J. (1961). "Tunneling from a Many-Particle Point of View", *Phys. Rev. Lett.*, **6**, 57–59.

[11] Iannaccone, G. and Pellegrini, B. (1996). "Compact Formula for the Density of States in a Quantum Well", *Phys. Rev. B*, **53**, 2020–2025.

Authors' Biographies

G. Iannaccone is serving on the faculty of the Electrical Engineering Department (Dipartimento di Ingegneria dell' Informazione) at the University of Pisa, Italy. His research interests include transport and noise modeling in mesoscopic and heterostructure devices, architectures and devices for nanoscale integrated circuits.

M. Macucci is serving on the faculty of the Electrical Engineering Department (Dipartimento di Ingegneria dell' Informazione) at the University of Pisa, Italy. His research interests include quantum-interference and single-electron devices, Coulomb Blockade phenomena, modeling and measurements of noise in electron devices.

B. Pellegrini is Professor of Electronics at the Electrical Engineering Department (Dipartimento di Ingegneria dell' Informazione) of the University of Pisa, Italy. His recent research interest include modeling and characterization of transport and noise in electron devices.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 455–461
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Impact Ionization and Hot-Electron Injection Derived Consistently from Boltzmann Transport

PAUL HASLER*, ANDREAS G. ANDREOU, CHRIS DIORIO,
BRADLEY A. MINCH and CARVER A. MEAD

California Institute of Technology, Pasadena, CA 91125

(Received 28 May 1997)

We develop a quantitative model of the impact-ionizationand hot-electron–injection processes in MOS devices from first principles. We begin by modeling hot-electron transport in the drain-to-channel depletion region using the spatially varying Boltzmann transport equation, and we analytically find a self consistent distribution function in a two step process. From the electron distribution function, we calculate the probabilities of impact ionization and hot-electron injection as functions of channel current, drain voltage, and floating-gate voltage. We compare our analytical model results to measurements in long-channel devices. The model simultaneously fits both the hot-electron-injection and impact-ionization data. These analytical results yield an energy-dependent impact-ionization collision rate that is consistent with numerically calculated collision rates reported in the literature.

Keywords: Impact ionization, hot electron injection, floating gate devices, silicon electron transport, MOSFET modeling

We develop a quantitative analytical model of the impact-ionization and hot-electron–injection processes in MOS devices that is derived consistently from a single spatially varying hot-electron distribution function. This approach not only provides a useful circuit model, but also complements and validates numerical results from Monte Carlo simulations.

We measure hot-electron–injection (gate) and impact-ionization (substrate) currents using an *n*-type MOSFET built with a high substrate doping ($1\times10^{17}\text{cm}^{-3}$) operating with subthreshold currents. Figure 1 illustrates the cross section of the device. In subthreshold the channel current of a MOSFET is sufficiently small so that the mobile charge does not affect the surrounding electrostatics, resulting in a constant surface potential. Consequently, by operating the MOSFET in subthreshold, we obtain a high field region whose properties are independent of the channel current. This higher substrate doping is consistent with a 0.3 μm channel length CMOS process; thus, these

*Corresponding author: E-mail phasler@ee.gatech.edu.

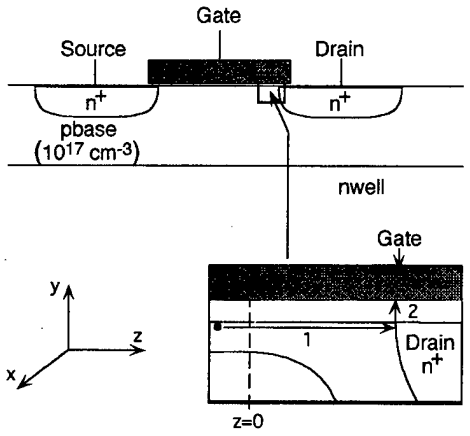


FIGURE 1 Cross section of the MOSFET device we used to measure the hot-electron effects. It uses a highly doped ($1 \times 10^{17}\text{cm}^{-3}$) substrate to achieve a high threshold voltage which allows hot-electron injection for bias current levels in subthreshold. The *n* well isolates the highly doped substrate region from the surrounding substrate, and allows measurement of substrate current. Holes resulting from impact ionization are measured at the *p* base contact. The hot-electron injection process is identical for the FET with or without the isolating *n* well. Inset: the electron is accelerated through the drain depletion (path 1), and when it gains energy greater than the Si—SiO_2 barrier, the electron is injected over the Si—SiO_2 barrier to the floating-gate (path 2).

effects are directly applicable to modern processes.

For an electron to reach the floating gate, it must have energy greater than the oxide barrier height and must be directed towards the SiO_2 when the electron reaches that energy. The high electric fields in the drain-to-channel depletion region accelerate channel electrons to high energies (path 1). The high substrate doping increases the threshold voltage (≈ 6 V) and the drain-to-channel electric field, which generates high-energy electrons at subthreshold currents for positive gate-to-drain voltages; therefore, an electron surmounting the Si—SiO_2 barrier will be transported to the gate by the resulting oxide field (path 2).

As an electron gains energy due to the electric field in the z direction, the electron is confined by the electric field and the silicon—silicon-dioxide interface in the y direction. The resulting electron distribution in y and k_y is nearly independent of the electron distribution in the other coordinates; therefore, some electrons at $y=0$ are directed toward the SiO_2, and these electrons will enter the SiO_2 if they have gained sufficient energy.

1. ELECTRON TRANSPORT IN THE DRAIN TO CHANNEL DEPLETION REGION

We begin by modeling hot-electron transport in the drain-to-channel depletion region using the spatially varying Boltzmann transport equation. We can simplify the general Boltzmann equation to a 1-D problem along the channel (z) axis [1]; Figure 2 shows the conduction band as a function of position through the MOSFET's channel region. Following a similar procedure to Baraff [2], we get

$$\frac{\partial f}{\partial z} + q\mathcal{E}\frac{\partial f}{\partial E} + q\mathcal{E}\frac{1-\zeta^2}{\zeta E}\frac{\partial f}{\partial \zeta} = \frac{m^*(c)}{\zeta c}S(f), \qquad (1)$$

where $f(z, c\zeta)$ is the distribution function, $\mathcal{E}(z)$ is the component of the electric field in the z direction, c is the magnitude of the average momentum vector, ζ is the cosine of the angle of momentum vector and the z axis, and $S(f)$ is the collision operator. $E=c^2/m^*(c)$ is the electron energy, where $m^*(c)$ is the effective mass of the electron that depends upon the silicon band structure.

Starting from Conwell's optical-phonon collision operator [3], we derive the following approximate optical-phonon collision operator for $E \gg E_R$ [1]:

$$S_{op}(f) \approx \left(\frac{c(E)}{m^*(E)}\right)\left(\frac{E_R^2}{2\lambda}\frac{\partial^2 f}{\partial E^2} + \frac{E_R}{\lambda}\frac{\partial f}{\partial E}\right), \qquad (2)$$

where E_R is the energy of an optical phonon ($E_R = 63$ meV in Si). A similar expansion and simplification has been done for polar optical phonons [4]. The mean free length for phonon collisions (λ) is known to be approximately constant for high energies. We can remove the bandstructure effects in [1] by developing our collision models only in terms of a mean free

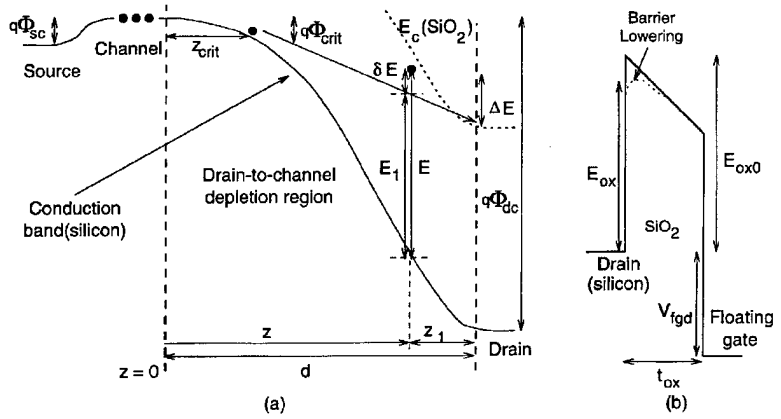


FIGURE 2 Band diagram illustrating hot-electron injection in a MOSFET biased in subthreshold. The appropriate variables in the Boltzmann transport equation and its variable transformations are shown on the graphs. (a) Band diagram along the surface of the Si—SiO_2 barrier. This region is the lowest local potential in either material; therefore the electrons are most likely to travel along this path. This region corresponds to path 1 in the inset in Figure 1. (b) Band Diagram of at the drain edge. This region corresponds to path 2 in the inset in Figure 1.

length and terms of $c(E)/m^*(E)$.[1] Phonons have momentum, and the total momentum involved for a phonon absorption or emission must be conserved. To precisely model this effect, one would need to know the distribution function of momentum for the phonons in the drain-to-channel depletion region. Elsewhere we show that the scattering of the momentum distribution has a small effect on our zeroth order expressions [1].

Most proposed impact-ionization collision operators can be formulated in general as

$$S(f)_{\text{ion}} = -\frac{f}{\tau_{\text{ion}}(E)} = -\frac{c(E)}{m^*(E)}\frac{1}{L(E)}f \quad (3)$$

where τ_{ion} is the mean free time for an impact ionization collision, and $L(E)$ is the mean free path, which is a function of the electron energy. We propose the following model for the energy dependence for the impact-ionization mean-free length

$$L(E) = (0.181\text{Å})\exp\left(\sqrt{\frac{119\,\text{eV}}{E - 0.95\,\text{eV}}}\right), \quad (4)$$

which is based on our experimental measurements of the impact-ionization mean free length, and corresponds to previous numerical calculations [5–7]. Figure 3 shows our functional form with these three numerically calculated models. We have assumed a constant velocity of 8.1×10^6 cm/s in converting from $L(E)$ to impact-ionization scattering rate, since our measured data is directly related to $L(E)$. This functional form is a curve fit to experimental data of $L(E)$ derived from our experimental measurements of hot-electron-injection and impact-ionization currents in Section III.

2. SOLUTION OF THE TRANSPORT EQUATION

We analytically solve the resulting Boltzmann transport equation,

$$\frac{\partial f}{\partial z} + \left(q\mathcal{E} - \frac{E_R}{\lambda\zeta}\right)\frac{\partial f}{\partial E} + q\mathcal{E}\frac{1-\zeta^2}{\zeta E}\frac{\partial f}{\partial \zeta} = \frac{E_R^2}{2\lambda\zeta}\frac{\partial^2 f}{\partial E^2} - \frac{f}{L(E)}, \quad (5)$$

[1]Canceling out the effects of the bandstructure may limit the predictive power of this model. This insight by Karl Hess is appreciated.

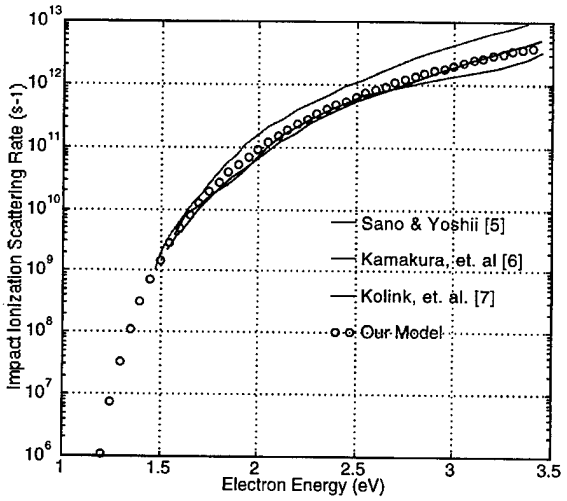


FIGURE 3 Plot of previous calculations of impact-ionization rate versus electron energy in silicon and our derived impact-ionization rate from our measured impact-ionization and hot-electron injection data. We have assumed a constant velocity, since our model measures the impact-ionization mean-free length. Our measured data is directly related to $L(E)$ and not impact-ionization scattering rate.

for a self-consistent distribution function using a two-step process. Elsewhere, we show that the transport along $\zeta = 1$ for hot-electron injection and impact ionization closely approximates the exact solution [1] for clarity, we will only consider the $\zeta = 1$ case here. In the first step, we solve for the average hot-electron trajectory in energy and direction as a function of position through the depletion region. The average hot-electron trajectory is the flow line for the hyperbolic P.D.E. operator, and is related to the numerical method that Budd presented previously [9]. In this model, the average electron starts gaining energy at the position (z_{crit} in Fig. 2) where the phonon restoring force is equal to the energy increase due to the local electric field ($\mathcal{E}(z)$). This breakaway field – the minimum electric field at which the electron gains energy at the same rate as it loses energy to phonon collisions – is expressed as $E_{R/q\lambda}$, which for our parameters is 9.7 V/μm. The average energy, $E_1(z)$, that the electron gains after reaching z_{crit} is

$$E_1(z) = qV(z) - qV(z_{crit}) - E_R \frac{z - z_{crit}}{\lambda}, \qquad (6)$$

or the difference between the potential from z_{crit} to the position z in the drain-to-channel depletion region, and the number of phonon collisions in this region.We show the electron in Figure 2 taking a linear path because of the functional form of Eq. (6).

In the second step, we solve for the electron distribution function around this average electron trajectory. In this coordinate system, phonon collisions diffuse the electron distribution spatially, and impact-ionization collisions remove high-energy electrons. To simplify the analysis, we assume that the electron leaving at z_{crit} dominates the behavior of hot-electron injection and impact-ionization for a wide range of drain voltages; the limitations of this approximation are illustrated in Figure 4. Using a more complicated initial and boundary conditions, $f(z, E)$ nearly follows an effective temperature solution for high electron energies, and $f(z, E)$ is the convolution of several Gaussians at low energies. From this analysis, the solution for the distribution function, $f(z, E)$, is

$$f(z,E) = \exp\left(-\frac{\lambda}{z - z_{crit}}\left(\frac{E - E_1(z)}{2E_R}\right)^2\right) a(z,E), \qquad (7)$$

where $a(z, E)$ models the electrons lost to impact ionization, and is approximated by

$$a(z,E) = \exp\left(-\frac{1}{(q\mathcal{E}(z)\lambda - E_R)}\int_{E=0}^{E}\frac{\lambda}{L(E)}dE\right). \qquad (8)$$

This solution shows that the assumption of a constant electron temperature is not valid at energies at which impact ionization and hot-electron injection occur.

3. COMPARING THEORY WITH EXPERIMENT

From the electron distribution function in (7), we can calculate the probabilities of impact ionization

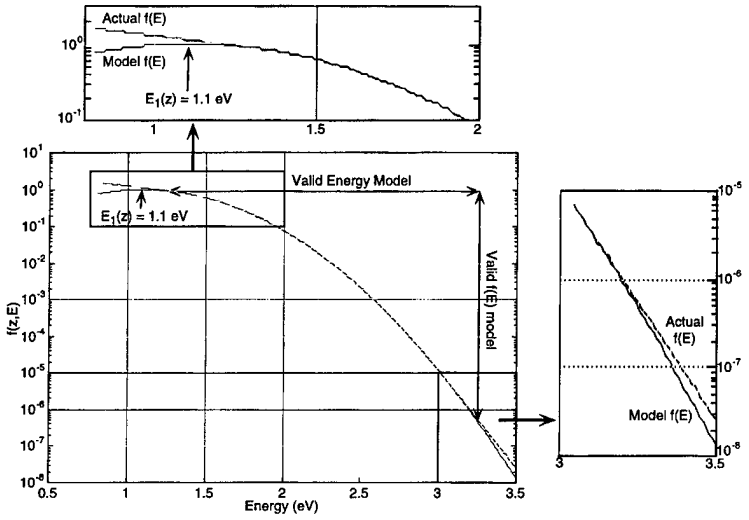


FIGURE 4 Picture of the distribution function for an electron in the drain-to-channel beyond $z = z_{crit}$. This figure compares our approximate model to the solution using the exact conditions around z_{crit}. For energies at or below the average electron energy, the distribution function shows the cumulative effect of electrons leaving the conduction band after z_{crit}. For large positive energies, the distribution function does not change as fast as the Gaussian, but rather at a slope due to an effective temperature.

and hot-electron injection as functions of channel current and drain-to-channel voltage (Φ_{dc}). We use two free parameters, λ and E_{ox}, as well as our functional form for $L(E)$. The hot-electron–injection efficiency–the ratio of the injection current (I_{inj}) and the source current (I_s)–is approximately given by

$$\frac{I_{inj}}{I_s} = B_2 \exp\left(-\frac{\lambda}{d - z_{crit}}\left(\frac{E_{ox} - E_1(d)}{2E_R}\right)^2 - \frac{7.102}{\sqrt{E_1(d)}}\right), \tag{9}$$

where λ is equal to 6.5 nm, $E_{ox} \approx 2.8$ eV is the Si—SiO_2 barrier height at the drain, $B_2 = 4.55 \times 10^{-3}$, $d(\Phi_{dc})$ is the width of the drain-to-channel depletion region, $\sqrt{E_1(d)} = \sqrt{\Phi_{dc}} - \sqrt{\Phi_{crit}}$, and Φ_{crit} is the potential drop from $z = 0$ to $z = z_{crit}$. Our experimental data on the Early voltage versus Φ_{dc} show that the channel doping profile is approximately a step junction for a fixed gate voltage [1]. Figure 5 shows measured data of hot-electron–injection efficiencies as a function of drain-to-channel voltage for two channel currents; the hot-electron–injection efficiency is independent of source current. Figure 5 shows (9) fitted to the injection efficiency data. The curve fit shows close agreement to (9) except at large Φ_{dc} (> 5.0 V), due to average-electron energy being near the energy of the silicon—silicon-dioxide barrier, and at small Φ_{dc}, probably due to the simplified modeling of the band-structure effects in the collision operators.

The impact-ionization efficiency–the ratio of the substrate current (I_{sub}) and the source current (I_s)– is

$$\frac{I_{sub}}{I_s} = \frac{\int_0^\infty (1 - a(d,E)) e^{-\frac{\lambda}{d - z_{crit}}\left(\frac{E - E_1(d)}{2E_R}\right)^2} dE}{\int_0^\infty e^{-\frac{\lambda}{d - z_{crit}}\left(\frac{E - E_1(d)}{2E_R}\right)^2} dE.} \tag{10}$$

We get an approximate solution by substituting (4), (8) and expanding the function in the exponent around the function's maximum value in E. We show the general solution elsewhere [1]; the solution for substrate doping of $N_a = 10^{17}\text{cm}^{-3}$ is

$$\alpha = \frac{I_{sub}}{I_s} = e^{-\sqrt{\frac{179.5\text{ V}}{\Phi_{dc}}}} \exp\left(\frac{\sqrt{28\text{ V}}\left(1 - \frac{E_1(d)}{3.36\text{ eV}}\right)^2}{\sqrt{\Phi_{dc}} - \sqrt{\Phi_{crit}}}\right) \tag{11}$$

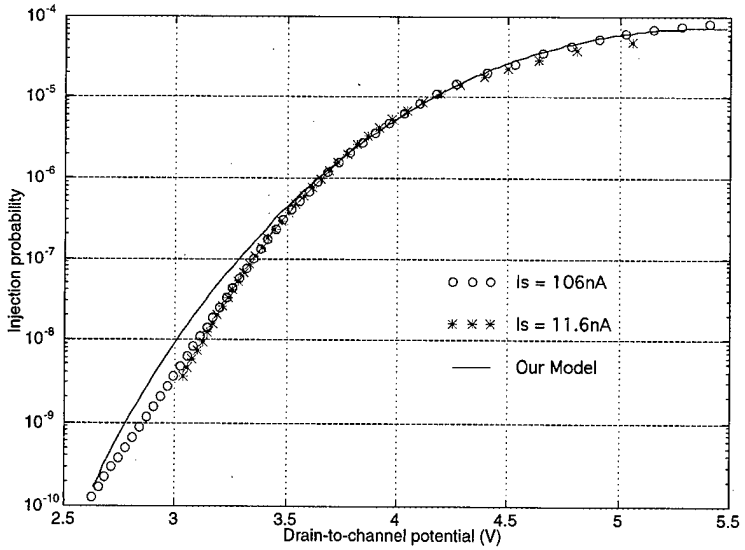


FIGURE 5 Measurements of hot-electron–injection efficiency verses drain-to-channel voltage for two values of source current. The drain-to-channel voltage is computed from the source current and the drain-to-source voltage. For each sweep, we used a constant gate voltage to chose a particular channel current; the actual oxide barrier height changes slightly due to image force lowering, because the floating-gate-to-drain voltage is not constant.

Figure 6 shows experimental measurements of α versus drain-to-channel potential.The solid line is the curve fit of (11) to the experimental data; the fit closely agrees with the measured data. From measured values of α versus Φ_{dc}, our analytical model allows us to measure the energy-dependent impact-ionization collision rate from experimental data; (4) is a curve fit to these data.

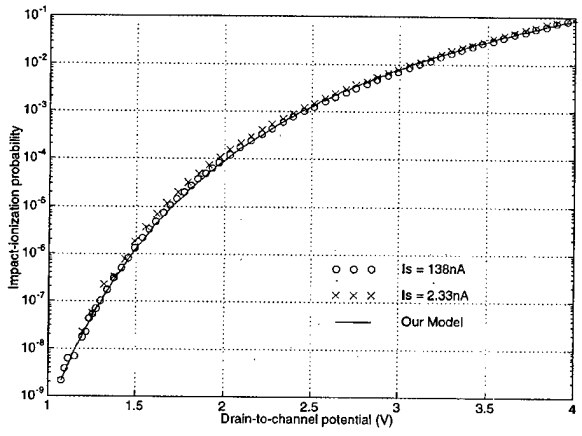


FIGURE 6 Measurements of impact-ionization efficiency vs. drain to channel voltage for two source currents (gate voltages). We plot a curve fit to the analytic model in (11); the model closely agrees with the experimental data.

Acknowledgements

We thank K. Hess for several helpful comments and suggestions, and L. Dupre for editing this manuscript.

References

[1] Hasler, P., "Impact Ionization and Hot-Electron Injection in MOSFETs Derived Consistently from Boltzmann Transport", in *Foundations of Learning in Analog*, VLSI, Ph.D. Thesis, Computation and Neural Systems, California Institute of Technology, February 1997. *Also at www.pcmp.caltech.edu/anaprose/paul.*

[2] Baraff, G. A. (1964). "Maximum Anisotropy Approximation for Calculating Electron Distributions; Application to High Field Transport in Semiconductors", *Physical Review*, **133**, A26–A33.

[3] Conwell, E. M. (1967). *High Field Transport in Semiconductors*, Academic Press, New York.

[4] LeBurton, J. P. and Hess, K., "Energy-diffusion equation for an electron gas interacting with polar optical phonons", *Physical Review B*, **26**(10), 5623–5633.

[5] Sano, N. and Yoshii, A. (1994). "Impact ionization rate near thresholds in Si", *Journal of Applied Physics*, **75**, 5102–5105.

[6] Kamakara, Y., Mizuno, H., Yamaji, M., Morifuji, M., Taniguchi, K., Hamaguchi, C., Kunikiyo, T. and Takenaka, M. (1994). "Impact ionization model for full band Monte Carlo simulation", *Journal of Applied Physics*, **75**, 3500–3506.
[7] Kolnik, J., Wang, Y., Oguzman, I. H. and Brennan, K. F. (1994). "Theoretical investigation of wave-vector-dependent analytical and numerical formulationsof the interband impact-ionization transition rate for electrons in bulk silicon and GaAs", *Journal of Applied Physics*, **76**, 3542–3551.
[8] Lee, C. H., Ravaioli, U., Hess, K., Mead, C. and Hasler, P. (1995). "Simulation of a long term memory device with a full bandstructure Monte Carlo Approach", *IEEE Electron Device Letters*, **16**(8), 360–362.
[9] Budd, H. (1967). "Path variable formulation of the hot carrier problem", *Physical Review*, **158**(3), 798–804.

Authors' Biographies

Paul Hasler received his B.S.E. and M.S. degrees in electrical engineering from Arizona State University in August 1991, and a Ph.D. in computation and neural systems from the California Institute of Technology in 1997. He is currently an Assistant Professor in electrical engineering at Georgia Institute of Technology. His research interests include using floating-gate MOS transistors to build adaptive systems in silicon, investigating the solid-state physics of floating-gate devices, and modeling high-field carrier transport in Si and SiO_2.

Andreas Andreou received his M.S.E. and Ph.D. degrees in electrical engineering and computer science from Johns Hopkins University in 1983 and 1986, respectively. From 1987 and 1989 he was a postdoctoral fellow and associate research scientist at Johns Hopkins University, where he became assistant professor in 1989, associate professor in 1993 and full professor in 1997. His research interests are in the areas of device physics, integrated circuits/systems and neural computation. During 1995–1996 he was a visiting associate professor of computation and neural systems at Caltech.

Chris Diorio received a B.A. in physics from Occidental Collegein 1983, an M.S. in electrical engineering from the California Institute of Technology in 1984, and a Ph.D. in electrical engineering from the California Institute of Technology in 1997. He is currently an Assistant Professor in computer science at the University of Washington. His interests include analog integrated circuit design, ultra-high-speed digital circuit design, and semiconductor physics. His current research involves using floating-gate MOS transistors for silicon learning applications. He is currently employed as a Staff Engineer at TRW Inc. in Redondo Beach, CA and has worked as a Senior Staff Scientist at American Systems Corporation in Chantilly, VA and as a Technical Consultant at The Analytic Sciences Corporation in Reston, VA.

Bradley A. Minch received a B.S. in electrical engineering with distinction from Cornell University in 1991, and a Ph.D. in computation and neural systems from the California Institute of Technology in 1997. He is currently an Assistant Professor in electrical engineering at Cornell University. His research interests include current-mode circuits and signal processing, the use of floating-gate MOS transistors to build adaptive systems in silicon and silicon models of dendritic computation.

Carver A. Mead Gordon and Betty Moore Professor of Engineering and Applied Science, has taught at the California Institute of Technology for more than 30 years. He has contributed in the fields of solid-state electronics and the management of complexity in the design of very large scale integrated circuits and has been active in the development of innovative design methodologies for VLSI. He has written, with Lynn Conway, the standard text for VLSI design, *Introduction to VLSI Systems*. His recent work is concerned with modeling neuronal structures, such as the retina and the cochlea using analog VLSI systems. His newest book on this topic, *Analog VLSI and Neural Systems*, was published in 1989 by Addition-Wesley. Professor Mead is a member of the National Academy of Sciences, the National Academy of Engineering, the American Academy of Arts and Sciences, a foreign member of the Royal Swedish Academy of Engineering Sciences, a Fellow of the American Physical Society, a Fellow of the IEEE and a Life Fellow of the Franklin Institute. He is also the recipient of a number of awards, including the Centennial Medal of the IEEE.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 463–468
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Inclusion of Bandstructure and Many-Body Effects in a Quantum Well Laser Simulator

F. OYAFUSO*, P. VON ALLMEN, M. GRUPEN and K. HESS

Beckman Institute, University of Illinois, Urbana, IL 61801

A self-consistent eight band **k·p** calculation, which takes into account strain and includes Hartree, exchange, and correlation terms (determined from a local density approximation) is incorporated into a QW laser simulator (MINILASE-II). The computation is performed within the envelope function approximation for a superlattice, in which all spatially varying terms of the **k·p** Hamiltonian, including the exchange and correlation energies are expanded in plane waves. The **k·p** eigenvalue equation, and Poisson's equation are solved iteratively until self-consistency is attained. Results from the **k·p** calculation are exported to MINILASE-II via a density of states and an energy dependent optical matrix element factor, renormalized by a Coulomb enhancement factor to account for electron-hole attraction. Results are presented for the gain spectrum and modulation response for a $Ga_{0.8}In_{0.2}As/Al_{0.1}Ga_{0.9}As$ quantum well laser with and without the inclusion of the Coulomb enhancement factor.

Keywords: Gain, **k·p**, laser, quantum well, modulation response, bandstructure

MINILASE-II [1] is a semiconductor laser simulator that self-consistently solves Poisson's equation, the carrier transport equations (both drift diffusion and thermionic emission across heterostructures), quantum well (QW) capture, and the photon rate equations for arbitrary two dimensional geometries. Many important effects arise from the bandstructure near the quantum well. However, it is computationally impractical to perform an accurate self-consistent bandstructure calculation within the laser simulator at each iteration. The purpose of this paper is to explain how we connect a separate eight band **k·p** superlattice calculation, including many-body effects, with MINILASE-II and to present results for the gain and modulation response for a strained-layer $In_{0.2}Ga_{0.8}As/Al_{0.1}Ga_{0.9}As$ quantum well laser.

The **k·p** calculation involves diagonalizing a Hamiltonian which at the center of the Brillouin zone can be expressed as

$$H_{n\ell}^{(kp)}(\mathbf{k}_{\|}, k_z) = \left(E_n(0) + \frac{\hbar^2(\mathbf{k}_{\|}^2 + k_z^2)}{2m}\right)\delta_{n\ell} + (\mathbf{k}_{\|} + k_z\hat{\mathbf{e}}_z)\cdot\mathbf{P}_{n\ell} + \sum_{\alpha\beta}\Gamma_{n\ell}^{\alpha\beta}k_\alpha k_\beta + H^{(\text{strain})} \quad (1)$$

* Corresponding author.

where the indices run over the bands in our basis set B, $\mathbf{P}_{n\ell}$ are the momentum matrix elements, and $\Gamma_{n\ell}^{\alpha\beta}$ are renormalization constants describing the contribution from bands not contained in B. To characterize the bandstructure of crystals with zincblende symmetry near Γ, it is sufficient to include the heavy hole (HH) and light hole bands (LH) (Γ_8^v), the split-off bands (SO) (Γ_7^v) and the lowest conduction bands (Γ_6^c). $H^{(\text{strain})}$, a k-independent term, describes the strain, which is assumed to be confined to the well region. The resulting 8×8 Hamiltonian has been described in the literature [2, 3].

For the superlattice calculation, z is taken to be the growth direction and the usual substitution $k_z \rightarrow \partial_z/i$ is made to obtain the effective mass Eq. [4],

$$\sum_{\ell \in \mathrm{B}} H_{m\ell}(\mathbf{k}_\parallel, \partial_z/i)\phi_\ell^{(n\mathbf{k}_\parallel k_q)}(z) = E^{(n)}\phi_m^{(n\mathbf{k}_\parallel k_q)}(z), \quad (2)$$

where k_q is the wave number for the 1D Brillouin zone of the superlattice, and $\phi_m^{(n\mathbf{k}_\parallel k_q)}(z)$, the superlattice envelope functions, are related to the wave functions to second order by

$$\psi_{n\mathbf{k}_\parallel} = \exp(i\mathbf{k}_\parallel \cdot \mathbf{r}) \sum_{m \in B} \left(u_{m0} + \sum_{\ell \notin B} \frac{\mathbf{k}_\parallel \cdot \mathbf{P}_{\ell m}}{E_\ell(0) - E_m(0)} u_{\ell 0} \right) \phi_m^{(n\mathbf{k}_\parallel k_q)}(z). \quad (3)$$

The band parameters which enter the Hamiltonian are allowed to have different values in the well and the barrier regions [5], and the resulting operator is Fourier transformed. For a superlattice period of 500 Å, typically about thirty Fourier components are required to ensure convergence of the ground state eigenenergies to within 1 μeV. The figures throughout this paper were generated for a superlattice consisting of $Al_{0.1}Ga_{0.9}As$ barriers of 400 Å, and $In_{0.2}Ga_{0.8}As$ wells of 80 Å. The barriers are wide enough to decouple adjacent wells so that the superlattice is in practice a collection of independent quantum wells. Most band parameters were determined from [6] except for the conduction band and valence band deformation potentials used to determine band edge shifts due to strain and the fraction conduction band discontinuity, which are not well known. We assume a fraction conduction band discontinuity of 0.7 and 11 eV for the difference in conduction band and valence band deformation potentials.

To account for carrier–carrier interactions additional terms which depend on electron and hole densities are added as diagonal terms in the **k·p** Hamiltonian. The direct Coulomb interaction gives the Hartree potential V_H which is determined by solving the 1D Poisson equation:

$$\partial_z(\varepsilon \partial_z V_H) = -\frac{4\pi}{\varepsilon}\rho_{3D}(z) \quad (4)$$

$$\rho_{3D}(z) = \frac{\rho_{2D}}{L}\int_{BZ}\frac{d^2k}{(2\pi)^2}\sum_j |\phi_j(\mathbf{k}, z)|^2 f(E_j(k), \mu(n_{2D})) \quad (5)$$

where the envelope functions ϕ_j have been normalized to the superlattice period L. The Fermi distribution $f(E, \mu)$ depends on the energy dispersions found from the **k·p** calculation and on the quasi-chemical potential μ, determined from the input parameter n_{2D}, the carrier density per unit surface in one period of the superlattice.

$$n_{2D} = \sum_j \int_{BZ}\frac{d^2k}{(2\pi)^2} f(E_j(k), \mu) \quad (6)$$

The interaction of a carrier with its exchange-correlation hole lowers its energy and results in a narrowing of the bandgap. This effect is taken into account in the local density approximation. We use an interpolated expression derived by Hedin and Lundquist [7] for three dimensional systems, and treat electrons and holes as separate plasmas. The bandgap renormalization (BGR) obtained is in qualitative agreement with experiment and can be improved by using an expression more appropriate for 2D systems.

Because these additional terms depend on the eigenfunctions determined from the **k·p** calcula-

tion, an iterative approach must be used. First the **k·p** Hamiltonian is diagonalized for a suitable initial guess for the Hartree, exchange and correlation energies. For a given 2D carrier density, the chemical potentials are then computed assuming a Fermi distribution function. The z dependence of the 3D carrier density can then be determined and used to compute a new estimate of the Hartree and exchange and correlation potentials. To assure convergence, an underrelaxation scheme is used for the exchange and correlation potentials. The approach is illustrated in the flow chart in Figure 1.

Once convergence has been attained, optical matrix elements (OME) given by $\mathbf{M}_{nn'} = \langle \psi_{nk\|} | \mathbf{p} | \psi_{n'k\|} \rangle$ are computed for later use in MINILASE II. The square of the OME gives the strength of the electron-hole coupling through the photon interaction and therefore enters the expressions for spontaneous and stimulated emission. Figure 2a shows the OME dispersion for TE polarization and two directions of $\mathbf{k}_{\|}$, parallel to and perpendicular to $\hat{\mathbf{e}}$. At Γ, CB1 couples only to HH1 and LH1 because of the orthogonality relation between the envelope functions. However, away from Γ, band mixing allows formerly forbidden transitions to take place. As expected for this polarization, the coupling at Γ between CB1 and HH1 is about three times stronger than between CB1 and LH1 [8].

Due to the Coulomb attraction between electrons and holes, spontaneous and stimulated emission are enhanced. This Coulomb enhancement, which may be thought of a renormalization of the optical matrix elements, is computed external to MINILASE II for various carrier densities and lasing energies. The expression used for the Coulomb enhancement factor has been

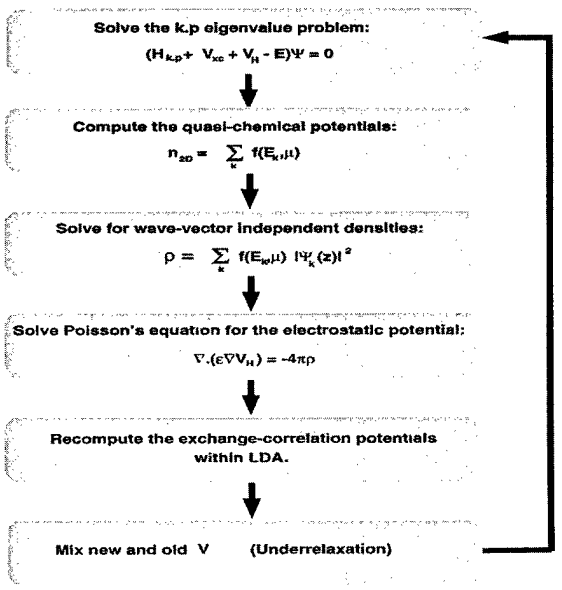


FIGURE 1 Flowchart for self-consistent **k·p** calculation.

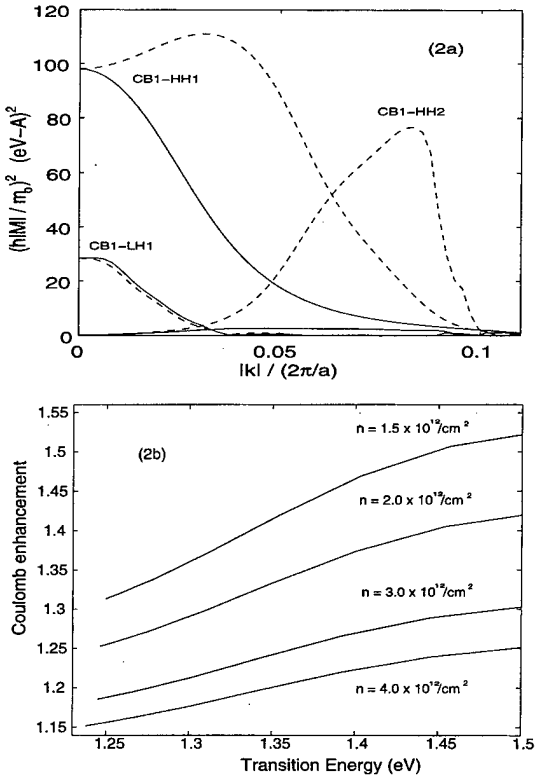


FIGURE 2 (a) Optical matrix elements for $\mathbf{k}_{\|}$ parallel to (solid) and normal to (dashed) the polarization. Transitions between the lowest conduction subband CB1 and the valence subbands HH1, HH2 and LH1 are shown. (b) Density dependence of the Coulomb enhancement factor.

derived in the literature [9] and is given by

$$\frac{1}{1-q(k)}$$

where

$$q(k)=\frac{1}{M(k)}\int dk'\, V^{(s)}(k-k')\,\frac{iM(k')}{\hbar}\,\frac{f_n(k')+f_p(k')-1}{i(E_c(k')-E_v(k')-\nu)+1/\tau} \tag{7}$$

Here f_n and f_p are Fermi occupation factors, τ is the carrier-carrier scattering time, assumed to be 100 fs, and $V^{(s)}$ is the 2D screened Coulomb potential which in the static long wavelength limit of RPA is given by

$$V^{(s)}(q)=\frac{2\pi e^2}{\varepsilon_0 q}\frac{1}{1+\kappa/q}$$

where

$$\kappa=\frac{2\pi e^2}{\varepsilon_0}\left(\frac{\partial n}{\partial \mu_n}+\frac{\partial p}{\partial \mu_p}\right)$$

is the inverse screening length. Figure 2b shows the renormalization factor for the CB1-HH1 transition between the lowest conduction subband and lowest valence subband for several carrier densities above threshold. In this calculation, an axial average of the matrix elements is performed, so that the OME may be treated strictly as a function of energy rather than in-plane wave vector $\mathbf{k}_\parallel$. We see that the modest enhancement of the OME decreases as the carrier density increases. This result stems from the fact that screening is enhanced at higher densities.

The results of these calculations are exported to MINILASE II in three forms: the spatial profile of the envelope functions, a density of states and an optical matrix element factor $P_n(\Omega)$ (associated with each conduction band energy grid point n) given by

$$P_n(\Omega)=\sum_{ij}\int\frac{d^2k_\parallel}{(2\pi)^2}|\hat{\mathbf{e}}\cdot\mathbf{M}_{ij}(\mathbf{k}_\parallel)|^2\eta_n(E_c^i(\mathbf{k}_\parallel))\,\delta(E_c^i(\mathbf{k}_\parallel)-E_v^j(\mathbf{k}_\parallel)-\Omega) \tag{8}$$

The direction of the light polarization is $\hat{\mathbf{e}}$, and η_n is a hat function with support on the energy range defined by grid points $n-1$ and $n+1$ and whose area is normalized to unity. The sum extends over all conduction subbands i and valence subbands j. The gain coefficient is then given by

$$g^{(0)}(\Omega)\sim\frac{1}{\hbar\Omega}\sum_n(\tilde{f}_e(\varepsilon_n)+\tilde{f}_h(\varepsilon_n-\Omega)-1)\,P_n(\Omega)\mathcal{L}(\Omega,\tau) \tag{9}$$

where $\mathcal{L}$ is a Lorentzian broadening function and $\tilde{f}_e$ and $\tilde{f}_h$ are the non-equilibrium electron and hole distribution functions, respectively. Minilase II incorporates the $\mathbf{k}\cdot\mathbf{p}$ data in an iterative manner. That is, the simulator first solves the carrier transport equations, Poisson's equation and the photon rate equations through Newton's method. The resulting carrier densities are used to recompute the band edges from an interpolated expression for the bandgap renormalization. Then, files appropriate for the computed density are read from disk. Since the exchange-correlation terms are added as k-independent diagonal elements, their inclusion does not significantly alter the effective masses and therefore the densities of state, except for shifting the band-edges.

We now briefly describe a few results obtained from MINILASE-II for an operating regime beyond the lasing threshold. Figure 3(a) shows the optical gain spectra with and without the Coulomb enhancement factor for the same applied current. In steady state, the maximum height of the gain is pinned by the losses in the laser which is the same in each case. There are two salient features that distinguish the gain curve without the Coulomb enhancement from the other curve. These are the reduced threshold for the onset of gain and the larger transparency point. Both features stem from the smaller effective matrix element in the absence of Coulomb enhancement

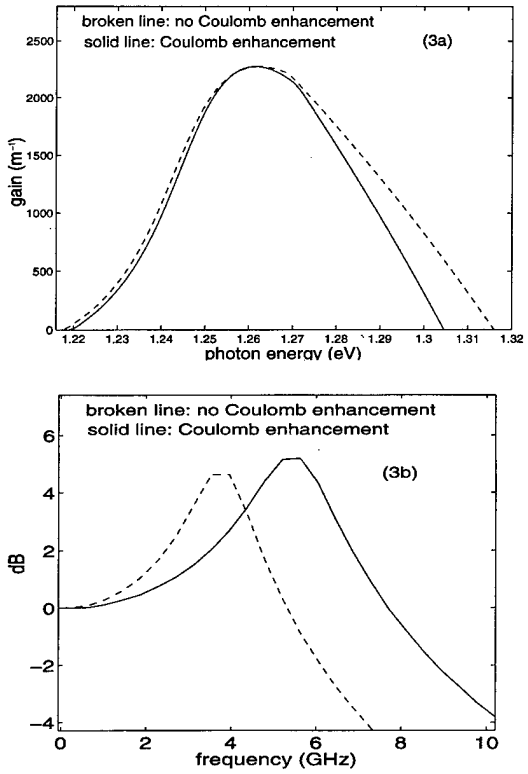


FIGURE 3 Gain with and without Coulomb enhancement.

that requires greater pumping of the laser to achieve a fixed gain. Thus, more carriers are required which results in a reduction in the band gap and implies a greater separation of the electron and hole quasi-fermilevels and thus a larger transparency point.

Figure 3(b) displays the laser's response to a small square pulse in the applied voltage. We note that the frequency peak is blue shifted when a larger effective matrix element is used. This may be explained as follows. The photon rate equation has the form

$$\dot{S}_\nu = (G_\nu - 1/\tau_\gamma)S_\nu + R$$

where ν is a mode index, S_ν is the photon population, G_ν is the modal gain, τ_γ is the photon lifetime and R is the spontaneous emission rate, which is much smaller than the other terms of the right hand side under lasing conditions. Since the losses $1/\tau$ are fixed, the gain is roughly proportional to the time derivative of the photon density in the cavity which implies a more swiftly responding laser.

In conclusion, we described how an accurate **k·p** calculation was connected to a sophisticated laser simulator and demonstrated how the larger effective matrix element that stems from the Coulomb enhancement affects the gain and resonance frequency in the modulation response of a laser.

Acknowledgement

We acknowledge the support of NSF through NCCE and of the Office of Naval Research.

References

[1] Grupen, M., Hess, K. and Rota, L. "Simulating Spectral Hole Burning and the Modulation Response of Quantum Well Laser Diodes", *Proc. of the SPIE Conf.*, 6–9 February 1995, San Jose, California.
[2] Cardona, M., Christensen, N. E. and Fasol, G. "Relativistic Band Structure and Spin Orbit Splitting of Zinc-blende-type Semiconductors", *Phys. Rev. B*, **38**, 1806, July 15, 1988.
[3] O'Reilly, E. P. (1989). "Valence Band Engineering in Strained-Layer Structures", *Semicond. Sci. Technol.*, **4**, 121.
[4] von Allmen, P. (1992). "Derivation of the Effective Mass Equation for a Superlattice: a Perturbational Approach", *Phys. Rev. B*, **46**, 15377, December 15.
[5] von Allmen, P. "Conduction Subbands in a GaAs/$Al_xGa_{1-x}As$ Quantum Well: Comparing Different **k·p** models", *Phys. Rev. B*, **46**, 15382, December 15, 1992.
[6] Landoldt-Bornstein, L. (1982). "Numerical Data and Functions in Science and Technology", edited by O. Madelung. Berlin: Springer, **17a** and **22a**.
[7] Hedin, L. and Lundquist, B. I. (1971). "Explicit local exchange correlation potentials", *J. Phys. C*, **4**, 2064.
[8] Chuang, S. L. (1995). *Physics of Optoelectronic Devices*, John Wiley and Sons.
[9] Weng Chow, Stephen Koch and Murray Sargent. (1994). *Semiconductor Laser Physics*, Springer-Verlag.

Authors' Biographies

Fabiano Oyafuso received B.A. degrees in mathematics (with honors) and physics from UC Berkeley in 1992. He is currently pursuing the Ph.D. degree in physics at the University of Illinois

at Urbana-Champaign. His thesis involves the study of the temperature dependence of threshold current densities in quantum well lasers.

Paul von Allmen obtained his B.S. and Ph.D. in physics from the Swiss Federal Institute of Technology in Lausanne. He joined the Zurich IBM research laboratory as a postdoctoral research associate in 1990. Since 1992 he has been an invited scholar at the Beckman Institute for Advanced Science and Technology at the University of Illinois in Urbana-Champaign. His interests range from subband structures and many-body effects in confined electron systems to the dynamical properties of nanostructures and surfaces. His main field of interest is presently the study of the desorption mechanism of hydrogen and deuterium from a silicon surface and the related isotope effect which has important consequences for the resistance of MOS-transistors against hot carrier degradation.

Matt Grupen received his B.S. from Penn State University in Engineering Science n 1985. He then attended UCLA where he received an M.S. in Electron Device Physics in 1989. In 1994, he received a Ph.D. in Computational Electronics at the University of Illinois, where he has held a postdoctoral position.

Karl Hess has dedicated the major portion of his research career to the understanding of electronic current flow in semiconductors and semiconductor devices with particular emphasis on effects pertinent to device miniaturization. His theories and use of large computer resources are aimed at complex problems with clear applications and relevance to miniaturization of electronics. He is currently the Swanlund Professor of Electrical and Computer Engineering, Professor of Physics, Adjunct Professor for Supercomputing Applications and a Research Professor in the Beckman Institute working on topics related to Molecular and Electronic Nanostructures. He has received numerous awards, for example the IEEE David Sarnoff Field Award for electronics in 1995.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 469–473
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Optical and Electronic Properties of Semiconductor 2D Nanosystems: Self-consistent Tight-binding Calculations

ANDREA REALE, ALDO DI CARLO*, SARA PESCETELLI, MARCO PACIOTTI and PAOLO LUGLI

INFM-Dipartimento di Ingegneria Elettronica, Università di Roma "Tor Vergata", 00133 Roma, Italy

A tight-binding models which account for band mixing, strain and external applied potentials in a self-consistent fashion has been developed. This allows us to describe electronic and optical properties of nanostructured devices beyond the usual envelope function approximation. This model can be applied to direct and indirect gap semiconductors thus allowing for instance the self-consistent calculation of band profile and carrier control in pseudomorphic InGaAs/GaAs HEMTs and SiGe/Si MODFETs.

Keywords: Semiconductors, nanostructures, electronic band structure, tight-binding, optical properties

1. INTRODUCTION

Electronic and optical properties of semiconductor nanostructures based on homo- and heterojunctions have been investigated theoretically by means of a variety of tools. These range from *ab-initio* approaches [1], which are very precise but require a large computational effort and consequently, are limited only to very small nanostructures, to approximate but easy-to-handle and fast methods such as for example those based on the envelope function approximation (EFA) [2]. In its simplest form, the EFA leads to the evaluation of the energy levels of nanosystems by simply solving a one-electron Schrödinger equation where each semiconductor is described in terms of effective masses and band edges. Such an approach and its generalizations, have been very successful [2–4], even though several problems still cannot be easily handled within the EFA context. The empirical tight binding method (TB) [5–8] has been shown to be a valid alternative to EFA, since it improves the physical content in the description of the nanostructure with respect to EFA without requiring a much higher computational effort. In particular, it allows us to treat indirect-gap semiconductors, heterostructures made by indirect/direct materials and to describe very thin nanostructures [5–7, 9]. Other advantages of TB with respect to EFA consist in the realistic description of the band structure of the whole Brillouin zone and in the possibility to use

* Corresponding author.

different Bloch basis functions for each component material of an heterosystem. TB has been mainly used in the calculation of the electronic properties of nanostructures without taking into account self-consistent charge redistribution, which is an important requirement when we deal with real systems. However, very recently, we have shown [10] that a self consistent tight-binding procedure can be defined. In this paper we will apply the self-consistent tight-binding approach to realistic nanostructures such as pseudomorphic HEMT and we will show how the method can be easily extended to handle the case of indirect band gap material such as in SiGe MODFETs.

2. SELF-CONSISTENT TIGHT BINDING

In this section we discuss the self-consistent tight-binding model for a system where symmetry is broken in one direction, for example the growth axis (z). The wave function $|E, \boldsymbol{k}_\parallel\rangle$ can be written as linear combination of planar Bloch sums, $|\alpha, m\rangle$ [11]

$$|E, \boldsymbol{k}_\parallel\rangle = \sum_{\alpha, m} C_{\alpha,m}(\boldsymbol{k}_\parallel, E)|\alpha, m\rangle \qquad (1)$$

with

$$|\alpha, m\rangle = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{R}_\parallel^m} e^{i\boldsymbol{k}_\parallel \cdot \boldsymbol{R}_{\alpha\parallel}^m} |\alpha, \boldsymbol{R}_\alpha\rangle \ , \qquad (2)$$

where $|\alpha \boldsymbol{R}_\alpha\rangle$ is a localized orbital, $\boldsymbol{k}_\parallel$ is the in-plane wave vector and N is the number of unit cells in the atomic plane. The subindex α refers both to the basis atom index and to the atomic orbital index. The lattice vector, $\boldsymbol{R}_\alpha = \boldsymbol{R} + v_\alpha$, (where $\boldsymbol{v}_\alpha$ is the basis atom displacement), can be written as $\boldsymbol{R}_\alpha = m\boldsymbol{d} + \boldsymbol{R}_{\alpha\parallel}^m$ where m is an integer, $\boldsymbol{d}$ a vector parallel to the growth direction with modulus equal to the distance between two atomic planes and $\boldsymbol{R}_{\alpha\parallel}^m$ is a vector on the m-th atomic plane. For a given $\boldsymbol{k}\parallel$, the eigenstates E are calculated by solving the secular equation $(H + V_H)|E, \boldsymbol{k}_\parallel\rangle = E|E, \boldsymbol{k}_\parallel\rangle$ where H is the system tight-binding hamiltonian and V_H is the Hartree potential. In order to calculate electronic and optical properties for real heterosystems, the presence and possible rearrangement of free charges has to be taken into account. The influence of the electronic charge rearrangement can be included at a Hartree level by solving the Poisson equation for the Hartree potential, $d^2V_H/dz^2 = -\rho(z)/\ \varepsilon$, where ε is the static dielectric constant. The charge density $\rho(z)$ is defined by the square of the wave function averaged over a layer z and weighted by the Fermi function $f(E)$:

$$\rho(z) = \frac{e}{(2\pi)^2} \int d\boldsymbol{k}_\parallel \sum_n |\Psi_{n,\boldsymbol{k}_\parallel}|^2 f(E_n), \qquad (3)$$

where e is the electron charge and n labels the energy levels for a given $\boldsymbol{k}_\parallel$. A full $\boldsymbol{k}_\parallel$ integration is performed in the 2D Brillouin zone by using the special k-points technique in the irreducible wedge [12]. The convergence of this integration has been obtained by using 5 special points with $|\boldsymbol{k}_\parallel| \leq 0.06\ 2\pi/a$ for direct band gap material. To achieve self-consistency in indirect band gap material we use 8 points with $|\boldsymbol{k}_\parallel| \leq 0.2\ 2\pi/a$.

The Poisson and Schrödinger equations in the tight-binding representation are iteratively solved until convergence is reached. An open-chain (infinite well) boundary condition is used for Schrödinger's equation. In order to avoid influences on calculated electronic levels, boundaries are chosen far away from the nanostructure region where the density of conduction electrons or valence holes is high. Although a better choice for the boundary condition is provided by complex band structure states as explained in Ref. [11], for all the situations discussed here the open chain condition represents a valid and simple choice. In order to speed up the self-consistent algorithm we have introduced a hybrid method to diagonalize the tight-binding hamiltonian which uses a standard (LAPACK [13]) routine to calculate eigenvalues and an inverse iteration scheme to calculate eigenvectors. The initial conditions for Schrödinger and Poisson equations are obtained self consistently in the effective mass approximation.

When optical properties are of interest, one can make use of the Kubo formula to define the susceptibility tensor which is related to the current-current response function of the electromagnetic perturbation [10, 14].

3. RESULTS

In order to show some results obtained with TB coupled with carrier redistribution, we consider several HEMT-like structures.

If we consider a typical device with a 430 Å n-doped ($n = 10^{18}$ cm^{-3}) cap layer, 20 Å undoped spacer in $Al_{0.3}Ga_{0.7}As$, 100 Å GaAs 2D channel and $Al_{0.3}Ga_{0.7}As$ buffer, TB shows small improvements in band profile, charge density and energy levels with respect to EFA. However, a more critical situation is obtained when higher Al concentrations are considered in the cap layer, which bring the barrier region to be almost indirect gap (e.g., $E_\Gamma = E_X$). Thus, we consider in Figure 1 a HEMT similar to the previous one, but with $Al_{0.45}Ga_{0.55}As$ as cap and spacer layer material. It becomes evident (see Fig. 2) that EFA completely fails in describing the charge redistribution between the 2D channel and the parasitic channel which forms due to the large X-valley contribution to the charge density.

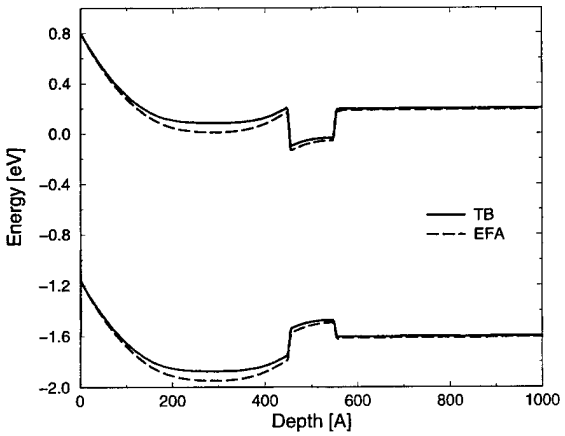


FIGURE 1 Self consistent band profile of a $Al_{0.45}Ga_{0.55}As$/GaAs/$Al_{0.3}Ga_{0.7}As$ HEMT structure

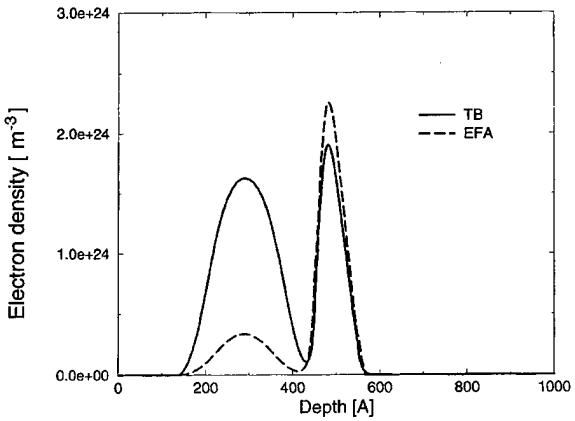


FIGURE 2 Self consistent charge density of the HEMT of Figure 1

The second example refers to a pseudomorphic (PM) structure consisting of a 360 Å n-doped ($n = 10^{18}$ cm^{-3}) $Al_{0.2}Ga_{0.8}As$, 20 Å undoped $Al_{0.2}Ga_{0.8}As$, a 265 Å strained well of $In_{0.15}Ga_{0.85}As$ and a GaAs substrate. Such a modulation-doped structure is typical for PM HEMT's, a device of great importance in the microwave field [15]. In this calculation the background doping charge is simply added to the free charge as given by Eq. (3). The self-consistent potential profile of the structure for two applied gate voltages of −0.6 and 0.0 V respectively is shown in Figures 3a and 3b, respectively. Strain effects in the InGaAs region, which are taken into account by scaling the hamiltonian matrix elements [16], split the *HH* and *LH* valence bands. Here the labels *HH* and *LH* refer to the character of the valence band in the growth direction. In the parallel direction the uppermost valence band has a light hole character

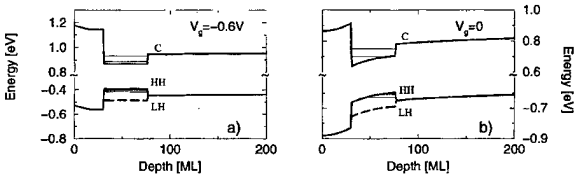


FIGURE 3 Self consistent band edge profile of the $Al_{0.2}Ga_{0.8}As/In_{0.15}Ga_{0.85}As$/GaAs pseudo-morphic structure at applied gate voltages of −0.6 V (a) and 0.0 V (b). The depth is measured in monolayers [ML].

while the other a heavy hole one. Strain also induces large changes on the effective masses of the valence bands. For the unstrained situation we have $m_{HH} = 0.4475\ m_0$ and $m_{LH} = 0.06481\ m_0$, while for bulk $In_{0.15}Ga_{0.85}As$ strained on GaAs the effective masses are $m_{HH}[001] = 0.4328\ m_0$, $m_{LH}[001] = 0.09596\ m_0$, $m_{LH}[100] = 0.08599\ m_0$ and $m_{HH}[100] = 0.1391\ m_0$. In the confined system, the hole energy levels present a different in-plane mass which depends on the energy. For an applied potential of -0.6 V the first three hole levels effective masses are $m_{h1}[100] = 0.10092\ m_0$, $m_{h2}[100] = 0.14799\ m_0$, $m_{h3}[100] = 0.2421\ m_0$, respectively. We observe that carrier confinement induces an enhancement of the effective mass compared to strained bulk material, a consequence of band non-parabolicity.

The calculated photoluminescence spectra for several applied potentials are shown in Figure 4. Since the *HH* quantized levels are now above the *LH* states, the luminescence transitions occur between conduction levels and the heavy hole levels. For an applied voltage of -0.6 V, which corresponds to a nearly flat InGaAs band, the main luminescence peaks are related to the $E_1 \rightarrow HH_1$ and $E_2 \rightarrow HH_2$ transitions. This is in agreement with the selection rules $\Delta_{nm} = 0$, where n labels the conduction states and m the valence states [2]. The emitted light is mainly polarized in the in-plane direction since the levels have essentially a heavy-hole character. When the applied potential is reduced, two other peaks form which correspond to the $E_1 \rightarrow HH_2$ and $E_2 \rightarrow HH_1$ transitions. Indeed, by decreasing the applied potential, the increasing channel electron density induces a sizable band bending which is responsible for the loss of symmetry (see Fig. 3b) and the transitions $E_1 \rightarrow HH_2$ and $E_2 \rightarrow HH_1$ are no longer forbidden. We also notice the presence of a red shift (quantum Stark shift) of the $E_1 \rightarrow HH_1$ transition due to the presence of an electric field as the applied potential reduces from -0.6 to 0.0 V.

As final example we show in Figure 5 the calculated band profile (a) and charge density (b) of a *p*-type SiGe/Si MODFET. Strain is responsible for the splitting of heavy and light hole bands

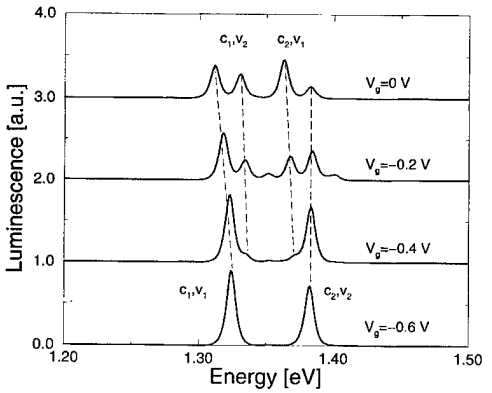
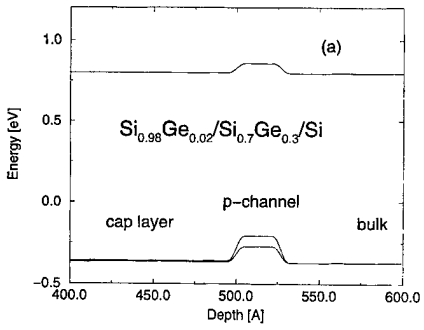


FIGURE 4 Photoluminescence spectra of the pseudomorphic structure for several applied voltages. The dashed lines follow the different luminescence peaks.



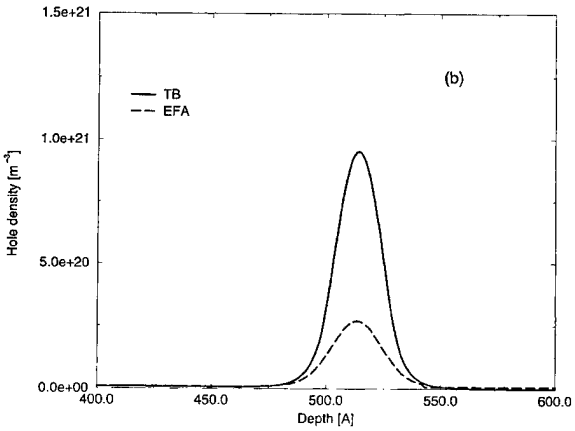


FIGURE 5 Self consistent band profile (a) and charge density (b) of a $Si_{0.98}Ge_{0.02}/Si_{0.7}Ge_{0.3}/Si$ *p*-channel MODFET structure.

as in the previous case. The charge density shows a strong difference between TB and EFA. This is is mainly due to the high non-parabolicity and anisotropy of the valence band.

In conclusion, we have shown that a self-consistent tight-binding approach can be used to evaluate the electronic structure and optical properties of semiconductor nanostructures. This represents a further step, with respect to the envelope function model, towards an *ab initio* calculation of such properties.

References

[1] Wood, D. M., Wei, S.-H. and Zunger, Alex. (1988). *Phys. Rev.*, **B37**, 1342; Park, C. H. and Chang, K. J. (1993). *Phys. Rev.*, **B47**, 12709.

[2] Bastard, G. (1988). *Wave mechanics applied to semiconductor heterostructure*: Les Edition de Physique, Les Ulis Cedex.

[3] White, S. R. and Sham, L. J. (1981). *Phys. Rev. Lett.*, **47**, 879; Altarelli, M. (1983). *Phys. Rev. B.*, **28**, 842.

[4] Los, J. (1996). *Phys. Rev.*, **B53**, 4630.

[5] Boykin, T. B., van der Wagt, J. P. A. and Harris, J. S. (1991). *Phys. Rev.*, **B43**, 4777.

[6] Di Carlo, A. and Lugli, P. (1995). *Semicon. Sci. Technol.*, **10**, 1673.

[7] Schulman, J. N. and Chang, Y. C. (1985). *Phys. Rev.*, **B31**, 2056.

[8] Slater, J. C. and Koster, G. F. (1954). *Phys. Rev.*, **94**, 1498; Bullet, D. W. (1980). *Solid State Physics*, **35**, 129; Majewski, J. A. and Vogl, P. (1989). *The structure of binary compounds*, Amsterdam: Elsevier, 1989. Edited by F. R. de Boer and D. G. Pettifor.

[9] Zunger, A., Yeh, C.-Y., Wang, L.-W. and Zang, S. B. (1994). *Proceedings ICPS-22*, 1763.

[10] Aldo Di, Carlo, Pescetelli, S., Paciotti, M., Lugli, P. and Graf, M. (1996). Solid State Communications **98**, 803.

[11] Di Carlo, A., Vogl, P. and Pötz, W. (1994). *Phys. Rev.*, **B50**, 8358.

[12] Froyen, S. (1989). *Phys. Rev.*, **B39**, 3168.

[13] Anderson, E., Bai, Z., Bischof, C., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A., Ostrouchov, S. and Sorensen, D. (1992). *LAPACK User's Guide*, Philadelphia: SIAM.

[14] Graf, M. and Vogl, P. (1995). *Phys. Rev.*, **B51**, 4940.

[15] Morkoç H., Unlu, H. and Ji, G. (1991). *Principles and Technology of MODFETS*, Chichester: John Wiley & Sons.

[16] Priester, C., Allan, G. and Lannoo, M. (1988). *Phys. Rev.*, **B37**, 8519.

Authors' Biographies

Andrea Reale was born in Italy in 1971 and received the degree in Electronic Engineering from the University of Rome "Tor Vergata" in 1997. He is currently working in the Optoelectronics Laboratory of the Electronic Engineering Departement of the same University, supported by a grant from CSELT-INFM collaboration. His interests include MQW-Semiconductor Optical Amplifiers, optical switches and self consistent band structure characterization of semiconductor devices.

Aldo Di Carlo graduated in Physics at the University of Roma, Italy, in 1991, obtained the Ph.D. degree in Physics in 1995 at the Walter Schottky Institute of the Technical University of Munich (Germany). He is currently researcher at the El. Eng. Dept. of the University of Rome "Tor Vergata", where his work is mainly focused on the theoretical study of optical and transport processes in semiconductor nanostructures, devices and polymers.

Paolo Lugli born in Carpi (Italy) in 1956, obtained his Laurea in Physics at the University of Modena in 1979 and his Ph.D. in El. Eng. at Colorado State University in 1985. He is currently Full Professor of Optoelectronics at the University of Rome "Tor Vergata". His main field of activity is the theoretical study and numerical simulation of semiconductor nanostructures and devices.

Marco Paciotti born in Rome in 1966, received the degree in Electronic Engineering in 1995 at the University of Rome "Tor Vergata". His main interests concern the modeling and the numerical simulation of semiconductor devices, with a particular knowledge in self consistent calculation of energy structure and charge redistribution. Currently at "Telecom Italia s.p.a.".

Sara Pescetelli born in Rome (1964), graduated in Physics in 1995 at the University of Rome "Tor Vergata". Her main field of activity concerns tight binding calculations of electronic and optical properties of semiconductors. Currently she is working into the El. Eng. Dept. of the University of Rome "Tor Vergata".

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 475–480
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Transient Phenomena in High Speed Bipolar Devices

MICHAEL S. OBRECHT*, EDWIN L. HEASELL, J. VLACH
and MOHAMED I. ELMASRY

Department of Electrical and Computer Engineering, University of Waterloo, Ontario N2L 3G1 Canada

A new numerical method is applied to the analysis of the charge partitioning in the quasi-neutral base of a BJT. The results show that the conventional, 1:2 collector/emitter partitioning is not valid in general. High level injection increases the collector fraction, whilst fast switching decreases it.

Keywords: BJT capacitance, transient semiconductor device simulation

1. INTRODUCTION

As shown in [1], the charge partitioning may significantly affect device delay times. Non-Quasi-Static (NQS) BJT modeling [2] also demonstrates the practical importance of such effects in modern devices. The ever increasing speed of microelectronic devices [3, 4, 7] requires the development of more accurate models to describe transient behavior.

Papers [1, 2] are based on "first order" NQS corrections. Although the latter approaches are significantly different, they are similar in their use of the quasi-static approximation for the charge partitioning factor in the quasi-neutral base (QNB). The accuracy of the NQS method [2] is higher since it takes into account a delay in the QNB charge, not only in the collector current.

In the present paper we use an exact, 2D numerical technique, for the evaluation of the charge partitioning in semiconductor devices [5]. It is applied here to study the partitioning of the QNB charge (related to the diffusion capacitances of a BJT). The numerical method allows the separation of the charges into quasi-neutral and space charge components, and the extraction of the transient components of terminal currents, responsible for changes in the device internal charge.

Previous results, obtained for a 1D *pn*-diode [5], demonstrated a dramatic change of the charge partitioning of the charge injected into the quasi-neutral region depending on the ramp speed. It was shown that the injection level strongly influenced the charge partitioning which could exceed 0.6 of the net base charge in very high level

* Also with Siborg Systems Inc, Waterloo, Ontario N2L 5B1 Canada, Phone (519) 888-4567 ext. 2082, Fax (519) 746-5195, E-mail: obrecht@siborg.ca, Web www.siborg.ca.

injection. In contrast, in low level injection, a partitioning of 0.33 is obtained which agrees well with the conventional diffusion theory. A similar trend was noticed in [2] where small-signal 1D analysis of a BJT was performed (α_m parameter in Fig. 4, [2]). Charge-based models e.g., those stemming from [8, 9] are preferred to describe transient device behavior and there has been renewed interest in charge partitioning [1, 10–21].

The majority of authors [1, 10–21] have addressed the partitioning problem using simplified analytic methods. Analytic models, such as those proposed for MOSFETs [10, 11] and BJTs [12, 13, 15, 16, 18, 19] provide charge conserving, partitioned charge descriptions. Analytic models, based on first or higher order, non-quasi-static, solutions of the continuity and transport equations have also been proposed [12, 17, 20, 21]. To obtain closed form expressions, useful for device models, such analysis requires a significant number of simplifying assumptions.

A numerical techniques needed to extract the various charge/current components, necessary to accurately characterize the charge partitioning, have been developed recently [5]. These techniques were implemented in TRASIM [22] which provides an accurate, stable, and rapidly convergent, time-dependent, numerical solution of the two-dimensional Poisson and continuity equations, for specified, time dependent boundary conditions. The numerical results show that there are significant differences between the exact charge partitioning and the predictions of the analytic models.

Both the present and previous analytic models are based on the conventional, drift-diffusion model for carrier transport. The numerical analysis confirms the validity of this approach in BJTs, since the electric fields (or Fermi-level gradients) remain modest, at even the highest switching speeds.

Non-the-less, the Ohmic electric fields, associated with the removal of the neutralizing, majority carriers, are large enough to cause major changes in the minority carrier partitioning. Such effects are completely omitted in analytic models.

2. DEPLETION AND DIFFUSION CHARGES

Conventional device models associate the motion of the space-charge region boundaries with charge storage in a "depletion capacitance", and the storage of minority carriers in quasi-neutral regions with a "diffusion" capacitance. A conventional numerical model provides no such distinction. The total charge for each carrier type must be split into a) a space-charge part, corresponding to a depletion capacitance and b) a quasi-neutral part, corresponding to a diffusion capacitance.

The extraction of the space charge and quasi-neutral charge have been described in detail earlier [5]. Following the methods described in [5] the depletion and diffusion charges are found from

$$Q_{n,p}^{\text{dep}} = \int_{\Omega} \rho_{n,p}^{sc}\, dxdy, \quad Q_{n,p}^{\text{dif}} = \int_{\Omega} \rho_{n,p}^{qn}\, dxdy \quad (1)$$

where Ω is the computational domain. The charges defined in Eq. (1) represent only the changes in the device charges/carrier concentrations that occur as a result of the applied switching ramp.

3. THE PASS-THROUGH CURRENT

It is necessary to compute the "pass-through" current [5], defined as the current that would flow across the device if the instantaneous bias and the instantaneous, transient internal potential distribution were held constant (c.f. "convective current" in [23]). The time dependent pass-through current must be subtracted from the terminal currents since it does not contribute to changes in the stored charges.

The "pass-through" current J_n^{pt} and J_p^{pt} are computed in a separate series of iterations, solving only the steady state, non-linear continuity equations

$$\nabla J_n^{pt} = qR; \quad \nabla J_p^{pt} = -qR. \quad (2)$$

During a fast transient, or as a result of high level injection, the electrostatic and majority carrier

quasi-Fermi potential distributions ψ (x, y) and $\varphi_p(x, y)$ change during the transition. It is insufficient to merely interrupt the changes in terminal voltages and to then compute the conventional steady-state current.

For even a very slow transient errors arise: Although the change in the potential distribution will be smaller, the integration time is increasing and still leads to an erroneous calculation of the partitioning. Unfortunately the effects of internal potential variation, imply also that the conventional "steady-state" part of a circuit model is no longer adequate to predict the corresponding terminal currents.

Integration of the solutions of Eq. (2) over the device contacts, gives $\sum_{k=1}^{N_c} J_k^{pt} = 0$. The total charge, arriving at the k-th terminal is defined as $Q_k(t) = \int_0^t \left(J_{c,k}(t') - J_{c,k}^{pt}(t')\right) dt'$.

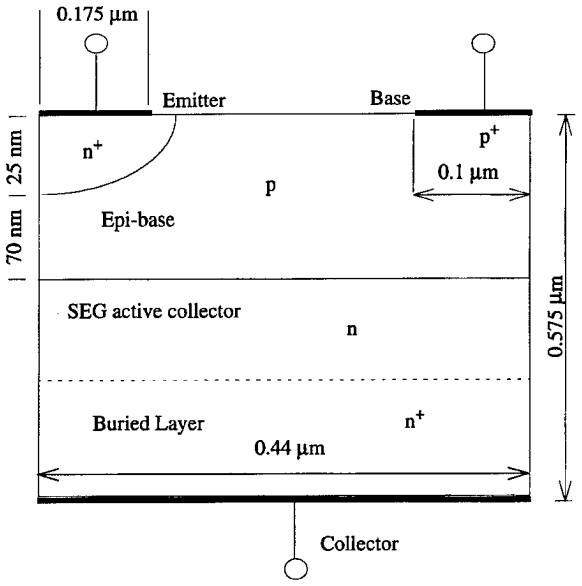


FIGURE 1 Simulated BJT structure.

4. SIMULATION RESULTS

The modifications described above were implemented in TRASIM [22], and the modified version of TRASIM used for transient 2D numerical simulation of a CML BJT [4]. Depending on the functional purpose of the transistor and the desire to increase speed of the ECL circuit, the transistor may operate in a high current mode, causing high level injection. Rather few papers have been published on the high injection behavior of the BJT (e.g. [6]). We believe that only thorough numerical investigation will shed light on the highly nonlinear, transient problem.

4.1. Device Structure

The transistor structure is shown in Figure 1 (we follow the data and SEM photograph from [4]). Only one half of the transistor is simulated to reduce the computational burden. The depth of the buried layer was 0.575 μm, having Gaussian impurity distribution with exponent of 0.1 μm and peak concentration of $10^{19}\,\mathrm{cm}^{-3}$. The selectively epitaxially grown active collector has impurity concentration of $1.8 \cdot 10^{17}\,\mathrm{cm}^{-3}$. The epitaxial base layer has a concentration of $2.5 \cdot 10^{18}\,\mathrm{cm}^{-3}$ and width of 70 nm. The polysilicon emitter was modeled by a surface recombination 10^4 cm/s, effective p-n junction depth of 25 nm and a half emitter with and area of 0.175 μm and 0.35×2.6 $\mu\mathrm{m}^2$ respectively. SRH, Auger and high doping effects were taken into account. This device shows β about 95 and I_C of 1 mA at $V_{EB}^0 = -0.92$ and $V_{CB} = 2\,\mathrm{V}$ which matches the experimental data [4].

4.2. Charge Partitioning for Different Injection Levels

The transients were simulated at $V_{CB} = 2\,\mathrm{V}$ with V_{EB} switched from V_{EB}^0 to 0 V with a linear ramp of duration τ_r seconds. The collector charge fraction P_C was calculated from $P_C = Q_C(t)/(Q_C(t) + Q_E(t))$, where $Q_E(t)$ and $Q_C(t)$ are charges recaptured at the emitter and collector terminals respectively after time t. Figure 2 shows the time evolution of the dynamical charge partitioning of the QNB charge. As we see, increasing the injection level leads to a bigger

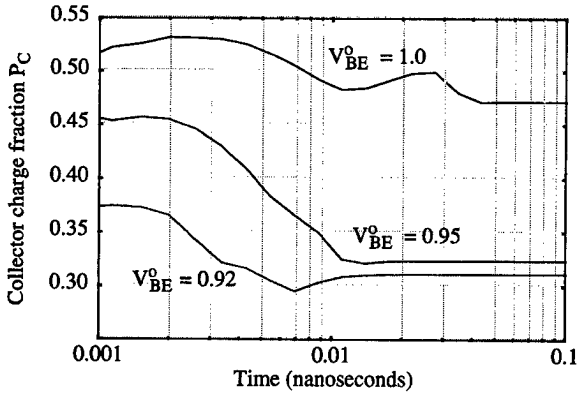


FIGURE 2 Charge partitioning for different injection levels $\tau_r = 0.01$ ns.

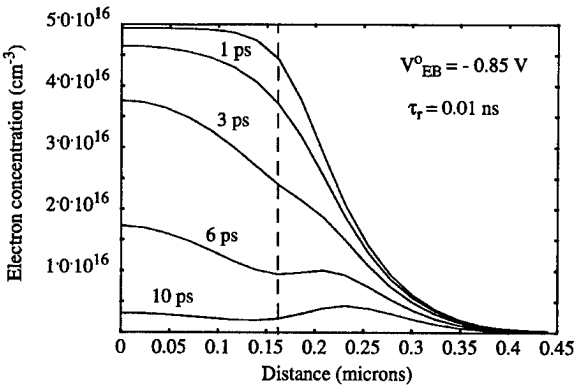


FIGURE 3 Electron density profile across the current flow in the QNB.

collector charge fraction. This effect is due to the built-in electric field in the QNB associated with the high injection level and the extraction of the compensating majority carriers.

Figure 3 shows 2D effects present when discharging the base of the transistor, in low injection. The vertical dashed line in the Figure 3 shows the "current channel boundary". Electrons injected outside this channel are diverted to the collector rather than to the emitter. After switching-off the transistor they are collected by the collector effectively increasing the collector fraction.

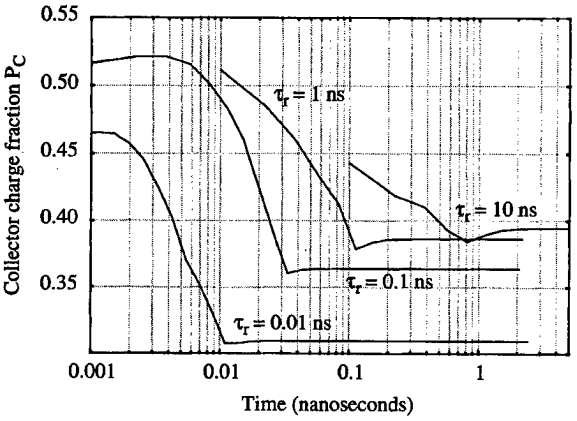


FIGURE 4 Charge partitioning for different ramps, $V^0_{EB} = -0.95$ V.

4.3. Charge Partitioning for Different Ramp Speeds

Figure 4 shows the dynamical charge partitioning of the QNB charge for different ramp speeds at $V^0_{EB} = -0.95$ V. Increasing ramp speed leads to smaller collector charge fraction which is consistent with the earlier results for a *p-n* diode [5]. This effect is due to the transient electric field in the QNB caused by the hole current.

Figure 5 shows the vertical component of the electron current density during transient for a slow and fast ramps under high level injection. The lateral electric field, caused by the fast base discharge, pinches the electron current flowing from the emitter to the collector, leading to a non-

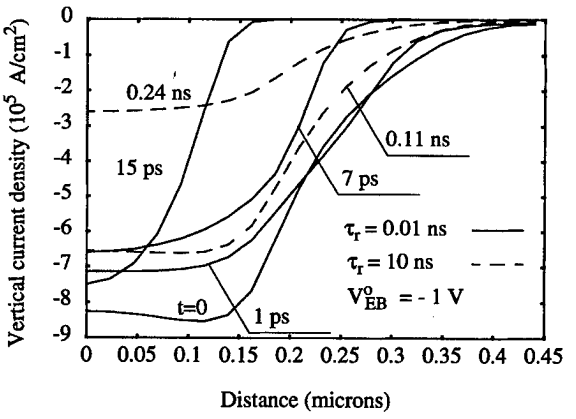


FIGURE 5 Vertical current density component for a slow and fast transient.

monotonic dependence of the electron current density underneath the center of the emitter. For the slow transients the current density profiles retain their shape during the transient.

CONCLUSIONS

A number of different BJT operating conditions were simulated. Some results depend on the device structure, nevertheless a few general conclusions seem to be possible at this point:

i) At low level injection charge partitioning deviates from the theoretical 1:2 collector:base partition due to the 2D effects.
ii) In a high current mode, when high injection occurs, the resulting electric built-in field the QNB leads to a collector charge partition that grows with V_{BE}.
iii) The ramp speed affects charge partitioning in both high and low injection level by effectively reducing the collector charge fraction at faster ramps.

It is not possible draw comparisons with [6] which models a device of base width 200 microns and zero recombination.

References

[1] Fossum, J. G. and Veeraraghavan, S. "Partitioned-Charge-Based Modeling of Bipolar Transistors for Non-Quasi-Static Circuit Simulation", *IEEE. Electron Device Letters*, **EDL-7**, 652–654, December 1986.

[2] Schroter, M. and Rein, H.-M. "Investigation of Very Fast and High-Current Transients in Digital IC's Using Both a New Compact Model and a Device Simulator", *IEEE Journ. Sol. State Circuits*, **30**(5), 551–562, May 1995.

[3] Meister, T. F., Stengl, R., Meul, H. W., Weyl, R., Packan, P., Felder, A., Klose, H., Schreiter, R., Popp, J., Rein, H. M. and Treitinger, L. (1992). "Sub-20ps Silicon Bipolar Technology Using Selective Epitaxial Growth", IEDM'92, *Technical Digest Papers*, pp. 401–404.

[4] Felder, A., Stengl, R., Hauenschild, J., Rein, H. M. and Meister, T. F. (1993). "25 to 40 Gb/s Si IC's in Selective Epitaxial Bipolar Technology", ISSCC'93, *Digest Technical Papers*, pp. 156–157.

[5] Obrecht, M. S. and Heasell, E. L. "A numerical analysis of transient charge partitioning", *IEEE Trans. Electron Devices*, **43**(3), 424–430, March 1996.

[6] Seitchik, J. A. and Hamel, J. S. "Transient Base Dynamics of Bipolar Transistors in High Injection", *IEEE Trans. ED*, **41**, 2385–2390, Dec. 1994.

[7] Yamaguchi, T., Uppli, S., Lee, J. S., Kawamoto, G. H., Dosluoglu, T. and Simpkins, S. (1993). "Process and Device Characterization for a 30 GHz submicrometer, Double Poly-Si Bipolar Transistor Technology using Implanted Base with Rapid Thermal Annealing Process," *IEEE Trans. Electron Devices*, **40**, 1484–1495.

[8] Ebers, J. J. and Moll, J. L. (1954). "Large-Signal Behavior of Junction Transistors", *Proc. IRE*, **42**, 1761–1772.

[9] Gummel, H. K. and Poon, H. C. (1970). "An Integral Charge Control Model of Bipolar Transistors", *Bell System Technical Journal*, **49**, 827–852.

[10] Ward, D. E. and Dutton, R. W. (1978). "A charge-oriented model for MOS transient capacitances", *IEEE J. Solid-State Circuits*, **13**, 703–707.

[11] Yang, P. (1986). "Capacitance Modelling for MOSFETs", *Advances in CAD for VLSI*, **3**, part 1, Editor A. E. Ruehli, North Holland.

[12] Hamel, J. S. (1993). "Integral relation for determing non-quasi-static charge partitioning in bipolar devices from static charge distributions", *IEEE Trans. Electron Devices*, **40**, 1713–1716.

[13] Lu, T. C. and Kuo, J. B. (1993). "A closed form analytical forward transit time model considering specific models for bandgap-narrowing effects and concentration-dependent diffusion coefficients for BJT devices operating at 77 K", *IEEE Trans. Electron Devices*, **40**, 766–772.

[14] Suzuki, K., Sato, S. and Nakayama, N. (1992). "Transient analysis of stored charge in neutral base region", *IEEE Trans. Electron Devices*, **39**, 1164–1169.

[15] Hamel, J. S. and Selvakumar, C. R. (1991). "The general transient charge control relation: a new charge control relation for semiconductor devices", *IEEE. Trans. Electron Devices*, **38**, 1467–1476.

[16] Wu, B. S. and Lindholm, F. A. (1990). "One-dimensional all injection non-quasi-static models for arbitrarily doped quasi-neutral layers in bipolar junction transistors including plasma-induces energy-gap narrowing", *IEEE. Trans. Electron Devices*, **37**, 250–261.

[17] Sipila, M., Parry, V. and Valtonen, M. (1989). "Improved description of base dynamics in the modelling of bipolar transistors", *International Jour. Circuit Thoery and Applications*, **17**, 465–482.

[18] Wu, B. S. and Lindholm, F. A. (1989). "One-dimensional non-quasi-static models for arbitrarily and heavily doped quasi-neutral layers in bipolar transistors", *IEEE Trans. Electron Devices*, **36**, 727–737.

[19] Klose, H. and Wieder, A. W. (1987). "The transient integral charge control relation – a novel formulation of the currents in a bipolar transistor", *IEEE. Trans. Electron Devices*, **34**, 1090–1099.

[20] Laux, S. E. (1985). "Techniques for small-signal analysis of semiconductor devices", *IEEE Trans. Computer-Aided Design*, **4**, 472–481.

[21] Hurx, G. A. M. (1988). "A New Approach to the AC Characterization of Bipolar Transistors", *Solid State Electr.*, **31**, 1269–1275.

[22] Obrecht, M. S., Elmasry, M. I. and Heasell, E. L. "TRASIM: compact and efficient two-dimensional transient simulator for arbitrary planar semiconductor devices", *IEEE Trans. CAD*, **14**(4), 447–458, April 1995.

[23] Cherry, E. M. (1971). "Active-Device Capacitances", *IEEE Trans. Electron Devices*, **18**, 1166–1168.

Authors' Biographies

Michael S. Obrecht received his M.Sc. (Physics and Applied Mathematics, 1975) and Ph.D. (Theoretical Physics, 1983) degrees from the Novosibirsk State University, Russia. He has worked in the area of semiconductor device numerical modeling for the last 14 years. In 1991 he joined Electrical and Computer Engineering Department of the University of Waterloo as a Research Associate Professor. Since 1994 he is also with Siborg Systems Inc. He has authored and coauthored over 20 papers and developed software tools for two-dimensional process-simulation which are used worldwide. Recently Dr. Obrecht is developing numerical algorithms and software tools for steady-state and transient, 2D/3D semiconductor device simulation.

Professor Heasell graduated in Physics (BSc.) and Electrical Engineering (Ph.D.), from Imperial College, London. As a research associate at AEI, Rugby he studied high current silicon rectifiers, before returning the EE.Dept. at Imperial College. His early research was directed to the study of 3–5 materials and especially InSb. On joining Waterloo the latter interest gave way to silicon device processing and modeling. His interests remain the analytical and numerical study of silicon devices.

Jiri Vlach (SM '67 – F' 82) was born in Prague, Czechoslovakia, where he received the equivalents of M.Sc. and Ph.D. degrees in Electrical Engineering from the Czech Technical University. Since 1969 he has been Professor in the Department of Electrical and Computer Engineering, Uinversity of Waterloo, Ontario, Canada. His research interests include computer-aided design, analog, digital, switched networks and numerical methods. In 1982 he was elected IEEE Fellow for "Contributions to computer-aided analysis and design of electrical networks". Presently J. Vlach is the Associate Editor of the IEEE Transactions on Circuits and Systems.

Mohamed I. Elmasry was born in Cairo, Egypt and received the BSc. degree from Cairo University, Cairo, Egypt, and the MASc. and Ph.D. degrees from the University of Ottawa, Ottawa, Ontario, Canada, all in Electrical Engineering in 1965, 1970 and 1974 respectively. He has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario Canada, since 1974, where he is a Professor and founding Director of the VLSI Research Group. He has served as a consultant to research laboratories in Canada, United States and Japan. Dr. Elmasry is a member of the Association of Professional Engineers of Ontario and is a Fellow of the IEEE for his contributions to "digital integrated circuits".

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 481–487
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Transverse Patterns in the Bistable Resonant Tunneling Systems Under Ballistic Lateral Transport

V. A. KOCHELAP[a,b], B. A. GLAVIN[a,b] and V. V. MITIN[b,*]

[a] *Institute of Semiconductor Physics, Ukrainian Academy of Sciences, Pr. Nauki 45, Kiev 252028, Ukraine;* [b] *Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202, USA*

We report the theoretical investigation of the phenomenon of the formation of patterns transverse to the tunneling current in resonant tunneling double-barrier heterostructures in the case of wide range of bistable voltages. In contrast to the case of the patterns in the structures with small region of bistability, for pronounced bistability electron lateral transport is strongly nonlocal. We performed numerical simulations of the stationary and mobile patterns using special variational procedure. Our results revealed that though the possible types of patterns remains the same as for the structures with small bistability region, their characteristics are modified considerably.

Keywords: Resonant tunneling, bistability, patterns, ballistic transport

1. INTRODUCTION

It is well established that the resonant tunneling in the double barrier heterostructures (DBH) is supplemented by the dynamic charge accumulation in the quantum well. This charge accumulation is particularly substantial in the case of the structure with asymmetrical barriers. The important effect, induced by the built-up charge is intrinsic bistability of the system under consideration. For some range of biases two stable states exist at the same bias. One state is characterized by a large built-up charge, resonant tunneling conditions and a large current, the other one corresponds to resonance breaking, lowering of the quasi-bound state below the bottom of the emitter band and a low charge built-up and current. The effect of bistability transforms the shape of the current-voltage characteristic of the resonant tunneling diode from N-type to Z-type. The effect of bistability was experimentally observed in [1–3]. Theoretical investigations of the bistability were performed in [4–6]. In these papers, the tunneling was considered as one-dimensional and the transport through the DBH was supposed to be dependent on only one coordinate, perpendicular to the barriers. Actually, most of the DBH are layered ones and a tunneling electron moves

* Corresponding author.

not only across the layers (vertical transport), but also along these layers (horizontal, or lateral, transport). In previous works [7, 8] we have shown that under the bistable conditions not only laterally uniform, but also nonuniform configurations of built-up charge and tunneling current (patterns) can exist. Results of [7, 8] were applicable mainly to DBHs with a small range of bistable voltages, where local approach for electron lateral transport was applied. In this work we present results of numerical simulations of stationary and mobile patterns for the DBHs with wide voltage range of bistability, where electron lateral transport is ballistic and strongly nonlocal.

2. MODEL AND BASIC EQUATIONS

Since the problem of the transverse patterns requires at least a two dimensional spatial analysis, we use a simple model, showing the main features of the bistability and the patterns. We deal with the model of a resonant tunneling heterostructure, schematically shown in Figure 1. The structure is treated as a system of three parts, weakly coupled by tunneling: emitter (E), quantum well (QW) and collector (C). The electrodes E and C are usually heavy doped semiconductors and are supposed to

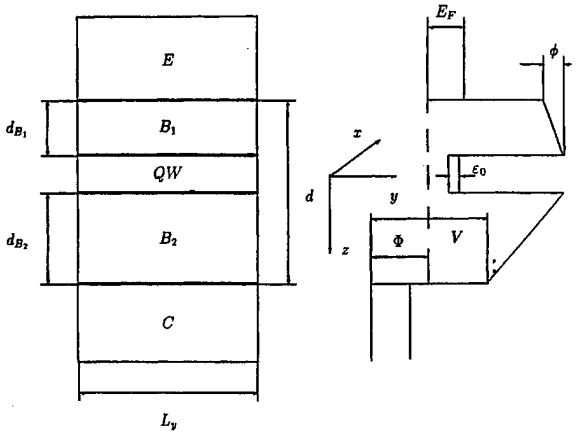


FIGURE 1 The scheme of the resonant tunneling structure and its energy band diagram under bias.

be ideal conductors. The energy height of the barriers B_1, B_2 is V and their thicknesses are d_{B_1}, d_{B_2}, respectively. Charge accumulation in the well causes a change of the potential profile in the whole structure. It alters the position of the quasi-bound state with respect to the bottom of the energy band of the emitter and, in general, with respect to the bottom of the quantum well. We disregard the latter effect and consider, that the built-up charge shifts equally the well bottom and the quasi-bound level. Such a case corresponds to the very thin quantum well, where the built-up charge can be accounted for as an infinitely thin sheet.

The horizontal electron transfer is the main process, determining the transverse patterns. This transfer can be thought as follows. The electron is injected from the emitter to the well in general with a finite horizontal component of the momentum p or velocity $v=p/m^*$ (m^* is effective mass). The velocity depends on the position of the quasi-bound state energy with respect to the Fermi energy E_F in the emitter: when the energy quasi-bound level, ε_0, moves from E_F through the bottom of the emitter band E_0, the velocity changes from zero to the Fermi velocity $v_F = \sqrt{2E_F/m^*}$. For estimates, one can account that v and v_F have the same order of magnitude. We can introduce the characteristic time for horizontal transfer: time of tunneling escape from the well τ_{es}. As we show below, the existence of the developed intrinsic bistability is strongly related with the character of electron lateral transport. Namely, the time of electron tunneling escape from the QW should be smaller than the time of scattering on phonons, impurities, etc. It means, that between the tunneling events electron moves in the QW *ballistically*. As a result, the horizontal characteristic distance is $L_{ch} = v\tau_{es}$. One can expect that the scale of the patterns in question is of the order of L_{ch}.

For narrow resonant level from the uncertainty relation we can write $\varepsilon_0\tau_{es} \gg \hbar$. Combining this inequality with the fact, that E_F, ε_0 and the kinetic energy of the horizontal motion $m^*v^2/2$ are of the same order of magnitude, we find for the in-plane

wave vector k:

$$kL_{ch} = \frac{p}{\hbar} L_{ch} = \frac{m^* v^2}{\hbar} \tau_{es} \sim \varepsilon_0 \tau_{es} \gg 1. \quad (1)$$

The latter estimate shows, that horizontal transfer can be considered as *classical*, while the vertical transport should be treated as *quantum*. As a result, we can introduce the distribution function of the resonant electrons in the QW. Assuming weak tunneling coupling between E, QW and C, we derive the kinetic equation for the distribution function f:

$$\frac{\partial f}{\partial t} + \frac{\vec{p}}{m^*} \frac{\partial f}{\partial \vec{r}} - \frac{\partial \phi}{\partial \vec{r}} \frac{\partial f}{\partial \vec{p}} = G(\phi(\vec{r}, t), \vec{p}) - \frac{f}{\tau_{es}} + I\{f\} \quad (2)$$

where $\phi(\vec{r})$ is the electrostatic potential energy in the well, $G(\phi(\vec{r}, t), \vec{p})$ is the local rate of tunneling from the emitter to the well, τ_{es} is the tunneling escape time, $I\{f\}$ is the collision integral for the electrons inside the well, G and τ_{es} are functions of ϕ at fixed $\vec{r}$. They are expressed through the tunneling probabilities and the Fermi distribution of electrons in the emitter.

From the uncertainty condition (1) we can deduce, that the characteristical scale L_{ch} greatly exceeds the well width. We assume L_{ch} is much larger than the thickness of the DBH: $L_{ch} \gg d$. In this case one can use the quasi-local relation between ϕ and electron density $n \equiv \Sigma_{\vec{p}} f(\vec{r}, \vec{p}, t)$:

$$\left(\phi(\vec{r}) + \frac{d_{B_1}}{d} \Phi\right) \frac{\kappa}{4\pi e^2} \frac{d}{d_{B_1} d_{B_2}} = n(\vec{r}), \quad (3)$$

which can be derived from the Poisson equation. Here κ is dielectric permittivity, Φ is the external bias in the energetical units. Eqs. (2), (3) compose the system of coupled nonlinear equations which describe the system under consideration.

3. BISTABILITY IN THE UNIFORM STRUCTURE

Let us show, that the model formulated above allows the bistable vertical transport regimes with uniform tunneling in the x, y plane. In such a case the $\vec{r}, t$ dependences are absent and from kinetic Eq. (2) one can find the areal electron concentration

$$n(\phi) = n_0(\phi) \equiv \tau_{es}(\phi) \sum_{\vec{p}} G(\phi(\vec{r}, t), \vec{p}). \quad (4)$$

Since the left-hand side of (4) is a function of ϕ, we get two algebraic Eqs. (3) and (4) for two variables n, ϕ. It is convenient to rewrite this system as

$$L(\phi) \equiv n_0(\phi) = \frac{\kappa d}{4\pi e^2 d_{B_1} d_{B_2}} \left(\phi + \frac{d_{B_1}}{d} \Phi\right) \equiv R(\phi). \quad (5)$$

For the particular heterostructure the latter equation has one controlling parameter, external bias Φ.

We have calculated the functions $\tau_{es}(\phi)$ and $g(\phi)$ for the heterostructure with parameters: $V = 1$ eV, $m^* = 0.067\, m_0$, $d = 5.7$ nm, $d_{B_1} = 2$ nm, $\varepsilon_0 = 0.1$ eV, $\kappa = 11.5$. In Figure 2 the left and right-hand sides are shown for $E_F = 56$ meV and zero temperature. The dependence of L on Φ is weak and only one curve $L(\phi)$ is shown. Qualitatively the dependence $L(\phi)$ corresponds to the switching on of the resonant tunneling when the resonant level crosses the bottom of emitter conduction band and further gradual decrease of the number of resonant electrons in the emitter while the resonant level is shifted toward the Fermi level of the emitter. $R(\phi)$ is just a straight line whose vertical shift is determined by the bias. From Figure 2 one can see that our system possess the property of bistability in the particular range of biases $\Phi_l < \Phi < \Phi_h$. Corresponding current-voltage characteristic is shown in the insert of Figure 2.

The approach, used in Eq. (2) allows to include into the tunneling generation rate G the effect of scattering of the resonant electrons in the QW (see [9]). The analysis of this effect has shown that the wide range of the bistability can take place if $\tau_{ew} \ll \tau_{sc}\tau_{cw}/(\tau_{cw}\tau_{sc})$. Here τ_{ew}, τ_{cw} are the char-

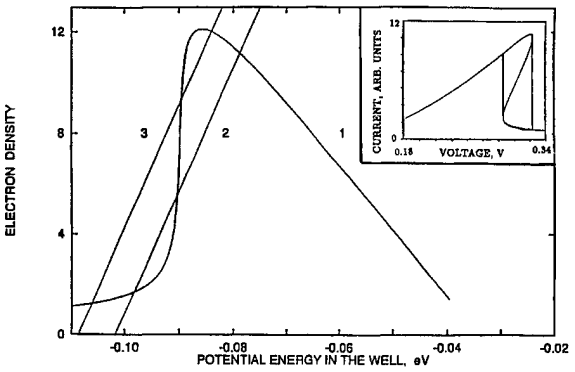


FIGURE 2 Illustration of the self-consistent solutions of the bistability problem under uniform tunneling for the structure described in the text. Curve 1 is almost independent on the external bias $L(\phi)$ and other curves are $R(\phi)$ for two biases: 0.29 eV (curve 2) and 0.31 eV (curve 3). The insert shows the current-voltage characteristic. L and R are in units $10^{11}\,\mathrm{cm}^{-2}$.

acteristical times of tunneling from the emitter to the QW and from the QW to the collector, calculated for the electrons with the energy of vertical motion equal to the Fermi energy in the emitter, and τ_{sc} is the characteristical time of scattering. It means, that the bistable range of voltages is substantial for the structure with high asymmetry in the barriers transmission coefficients (the collector barrier should be less transparent) and with *ballistic* or *quasi-ballistic* character of lateral electron transport. Furthermore, this fact explains dependence of the bistability on the temperature, which is usually observed in experiments: at higher temperatures τ_{sc} decreases and this washes out the asymmetry of tunneling in the system and the effect of bistability.

4. PATTERNS UNDER BALLISTIC REGIME OF HORIZONTAL TRANSFER

We consider the theory of the *one-dimensional* patterns, in which all physical values (built-up charge, tunneling currents, etc.) depend on only one coordinate, namely y. Then, we assume completely ballistic lateral electron transport. In this case (2) becomes

$$\frac{\partial f}{\partial t}+\frac{p}{m^*}\frac{\partial f}{\partial y}-\frac{\partial \phi}{\partial y}\frac{\partial f}{\partial p}=G-\frac{f}{\tau_{es}}. \tag{6}$$

here p labels the y-component of momentum $\vec{p}$. One can solve (6) in terms of the characteristic curves

$$p=\pm\sqrt{p_0^2+2m^*(\phi(y_0)-\phi(y))}\equiv\mathcal{P}(p_0,y_0,y), \tag{7}$$

where p_0 is the momentum of the electron, injected into the well at the point $y=y_0$. The general solution of the kinetic equation has the form

$$f(y,\vec{p})=\int m^*\frac{dy'}{\mathcal{P}(p,y,y')}M(p,y,y')\,G(\mathcal{P}(p,y,y'),y'), \tag{8}$$

where the kernel $M(p,y,y')$ depends on the particular shape of the potential $\phi(y)$ and can be easily calculated in the explicit form.

Unfortunately, iteration methods of solution fail in the solution of (3), (8) due to stability problems. Because of that we have applied the following variational procedure for the self-consistent solution of (3), (8). We introduce the functional

$$J\{\phi\}=\int dy\,(\phi-\mathcal{L}\{\phi\})^2, \tag{9}$$

where

$$\mathcal{L}\{\phi\}=-\frac{d_{B_1}}{d}\Phi+\frac{4\pi e^2 d_{B_1}d_{B_2}}{\kappa d}m^*\sum_{\vec{p}}\int\frac{dy'}{\mathcal{P}}MG. \tag{10}$$

Functional J equals zero for the exact solution of (3), (8). For a particular solution we can choose some probe functions $\phi_{pr}(y,c_i)$, where c_i are variational parameters. These parameters are determined by the condition of minimization of $J(c_i)$. It can be shown that our variational formulation of the problem of patterns is equivalent to the initial system of Eqs. (3), (6).

Using this method we have analyzed the possible types of stationary patterns. The possible types of patterns are the same as for the case of the structures with small range of bistable biases. The bistable range of biases can be split into the two regions $\Phi_l < \Phi < \Phi_c$ and $\Phi_c < \Phi < \Phi_h$ with the different types of patterns. In the first region the soliton-like patterns can exist. In these patterns at large $|y|$ the system is in the low-current state, while in the certain spatial region the local increase of the built-up charge and tunneling current takes place. At $\Phi_c < \Phi < \Phi_h$ the possible patterns are anti-soliton-like with the opposite characteristics: at high $|y|$ the system is in the high current state and local decrease of the built-up charge and the tunneling current in the finite spatial region. The $\Phi = \Phi_c$ is the critical bias: in this case the special kink-like pattern occurs. In this pattern high and low current states coexist in the different spatial regions.

Our calculations revealed that the anti-soliton patterns are substantially wider than soliton patterns. This is because for soliton pattern in the region of nonuniformity tunneling via both barriers is possible, while for anti-soliton only tunneling via collector barrier takes place. This influence the characteristical length which an electron can pass in the quantum well, and consequently the width of the pattern. The degree of the asymmetry in the spatial scales of the soliton anti-soliton patterns strongly depends on the degree of asymmetry of the transmission coefficients of emitter and collector barriers. To avoid numerical difficulties, we performed numerical simulations of the stationary patterns for a structure with thinner collector barrier with respect to the structure described in Section 3, namely with $d_{B_2} = 3.4\,\text{nm}$. The dependence of the L_s for the case of soliton-like pattern and L_a for the case of anti-soliton pattern on the dimensionless parameter $q \equiv (\Phi - \Phi_l)/(\Phi_h - \Phi_l)$ is presented in Figure 3.

The important property of patterns in the case of pronounced bistability is substantial difference between the spatial regions of localization of nonuniformities of the emitter-QW and the QW-collector tunneling currents. This is due to i) a strong dependence of the tunneling injection rate from the emitter to the QW on the position of the resonant level with respect to the bottom of the emitter conduction band and ii) a ballistic leakage of the injected electrons over the QW before the tunneling escape to the contacts. This is illustrated in Figure 4. In the upper part of the figure the current field in the structure is shown (spacing between the current lines is proportional to the value of the current density). In the lower part of the figure spatial dependences of the emitter-QW,

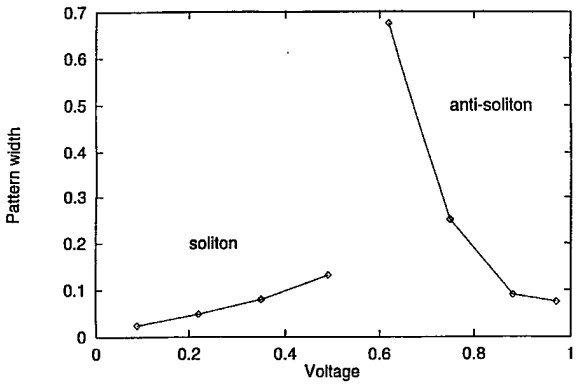


FIGURE 3 Dependence of the width of the soliton pattern, L_s, and anti-soliton pattern, L_a, on the dimensionless voltage parameter $q \equiv (\Phi - \Phi_l)/(\Phi_h - \Phi_l)$.

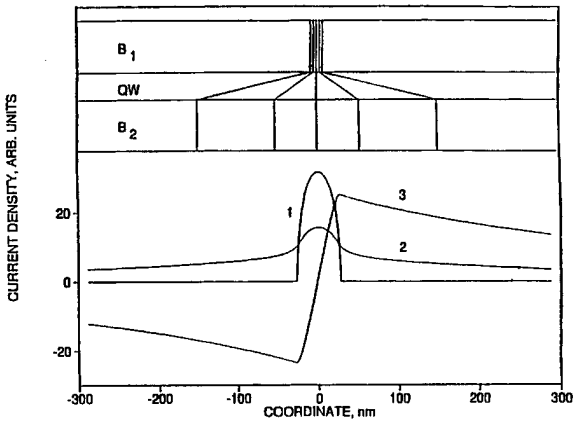


FIGURE 4 The spatial dependence of the current for the soliton-like pattern at $q = 0.3$. In the upper part the current field is depicted. In the lower part the emitter (curve 1), collector (curve 2, multiplied by factor 5) and two-dimensional lateral (curve 3) currents are presented.

QW-collector and two-dimensional lateral currents are depicted. The bias at which calculations were performed corresponds to $q=0.3$.

In addition to the stationary, we have investigated mobile autowave patterns. In fact, they are the moving kink-like patterns, describing switching of the DBH from one uniform state to the other. For the particular value of the bias the unique value of the switching wave velocity and the type of the switching exist. Namely, for $\Phi_l<\Phi<\Phi_c$ the switching from the high to the low current state is possible, while for $\Phi_c<\Phi<\Phi_h$ the switching from the high to the low current state can occur. Due to the ballistic character of the horizontal transport the characteristical velocity of the switching waves is determined by the Fermi velocity of electrons in the emitter v_F. In Figure 5 the dependence of the switching wave velocity (in units v_F) on q is presented for the structure described in Section 3. The insert shows the spatial dependence of the built-up charge in the stationary kink at $\Phi=\Phi_c$. The positive sign of velocity, in accordance with mentioned above, corresponds to the switching from the high to the low current state.

Note, that (3) is obtained for the infinitely conducting emitter and collector. For autowaves, moving with the velocity of the order of v_F and spatial scale of the kink transition region $v_F\tau_{es}$, this assumption is valid for the structures, in which τ_{es} is substantially greater than the characteristical time of relaxation in electrodes. This is true for the heterostructures with thick enough barriers and electrodes with high conductivity. Otherwise, the characteristics of autowaves (their velocity, for example) can be modified by the relaxation processes in the electrodes.

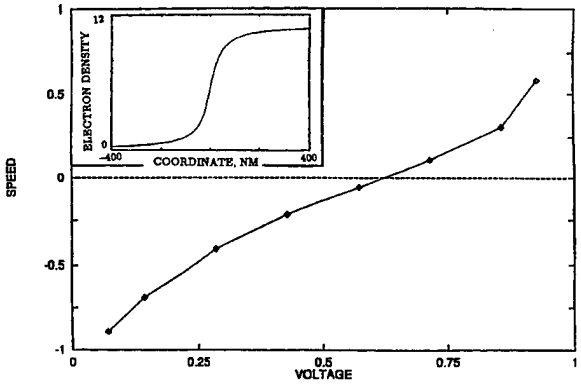


FIGURE 5 The dependences of the switching wave velocity (in units v_F) on dimensionless voltage parameter $(\Phi-\Phi_l)/(\Phi_h-\Phi_l)$. The spatial dependence of the built-up charge (in units $10^{11}\,cm^{-2}$) in the stationary kink is shown in the insert.

5. SUMMARY

In a conclusion, we have investigated the phenomenon of the transverse patterns formation in the resonant tunneling double barrier diode with the wide voltage range of intrinsic bistability of the current-voltage characteristic. These patterns are stationary or mobile nonuniform distributions of the built-up charge and tunneling current. For the pronounced bistability the characteristics of the patterns are determined by the **ballistic** and **nonlocal** character of lateral transport of the resonant electrons in the QW. This fact gives rise to specific features of the patterns with respect to the similar phenomena in DBHs with a small range of bistable voltages. For numerical simulations a special variational procedure was developed.

Acknowledgements

The authors would like to thank Drs. J. Schulman, F. Vasko and V. Sheka for discussions.

This work was supported by US ARO and by the Ukrainian State Committee for Science and Technology (grant No. 2.2/49).

References

[1] Goldman, V. J., Tsui, D. C. and Cunningham, J. E. (1987). "Observation of Intrinsic Bistability in Resonant Tunneling Structures", *Phys. Rev. Lett.*, **58**, 1256.

[2] Leadbeater, M. L., Alves, E. S., Eaves, L., Henini, M., Hughes, O. H., Sheard, F. W. and Toombs, G. A. (1988).

"Charge Build-up and Intrinsic Bistability in Resonant Tunneling Structures", *Semicond. Sci. Technol.*, **3**, 1060.
[3] Zaslavski, A., Goldman, V. J., Tsui, D. C. and Gunningham, J. E. (1988). "Resonant Tunneling and Intrinsic Bistability in Asymmetric Double Barrier Heterostructures", *Appl. Phys. Lett.*, **53**, 1408.
[4] Sheard, F. W. and Toombs, G. A. (1988). "Space Charge Buildup and Bistability in Resonant Tunneling Double Barrier Structures", *Appl. Phys. Lett.*, **52**, 1228.
[5] Kluksdhal, N. C., Kriman, A. M., Ferry, D. K. and Ringhofer, C. (1989). "Selfconsistent Study of the Resonant Tunneling Diode", *Phys. Rev. B*, **39**, 7720.
[6] Jensen, K. and Buot, F. (1991). "Numerical Simulations of Intrinsic Bistability and High Frequency Current Oscillations in Resonant Tunneling Structures", *Phys. Rev. Lett.*, **66**, 1078.
[7] Kochelap, V. A., Glavin, B. A. and Mitin, V. V. (1996). "Patterns in Bistable resonant Tunneling Systems", in *Hot Carriers in Semiconductors*, ed. by K. Hess, J.-P. Leburton and U. Ravaioli, New York: Plenum Press, p. 551.
[8] Glavin, B. A., Kochelap, V. A. and Mitin, V. V. (1996). "Patterns in Bistable resonant Tunneling Diode: Possibility of Novel Electron Device", *Proc. of the International Conference on Quantum Device and Circuits*, Imperial College Press, p. 170.
[9] Iannaccone, G. and Pellegrini, B. (1995). "Density of States in Double Barrier Resonant Tunneling Systems", *Phys. Rev. B*, **53**, 2020.

Authors' Biographies

V. A. Kochelap Viacheslav Kochelap received M.S. degree from Kiev State University (Ukraine) and Ph.D. degree from Institute of Semiconductors (Kiev, Ukraine). At present he is chairman of the Department of Theoretical Physics of the Institute of Semiconductors (Kiev, Ukraine).

His research interests are in the field of theory of semiconductors and semiconductor devices, electron-phonon interaction, electron transport and noise in low-dimensional structures, quantum electronics, nonlinear optics.

B. A. Glavin Boris Glavin received M.S. degree from Kiev State University. Currently he is working on his Ph.D. dissertation in Electrical Engineering. He works in the field of nonlinear effects in resonant tunneling semiconductor heterostructures and theory of phonons in low-dimensional semiconductor structures.

V. V. Mitin Vladimir Mitin received M.S. degree from Rostov State University (Russia) and Ph.D. degree from Institute of Semiconductors (Kiev, Ukraine). Currently he is professor of Electrical and Computer Engineering at Wayne State University, Detroit, USA.

His scientific activities are in the sphere of electronic and optoelectronic devices and materials, simulations of electron transport and noise in low dimensional semiconductor structures, growth of semiconductors, heat removal in semiconductor devices.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 489–493
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

SPIN – A Schrödinger-Poisson Solver Including Nonparabolic Bands

H. KOSINA* and C. TROGER

Institute for Microelectronics, TU Vienna, Gusshausstrasse 27-29, A-1040 Vienna, Austria

Nonparabolicity effects in two-dimensional electron systems are quantitatively analyzed. A formalism has been developed which allows to incorporate a nonparabolic bulk dispersion relation into the Schrödinger equation. As a consequence of nonparabolicity the wave functions depend on the in-plane momentum. Each subband is parametrized by its energy, effective mass and a subband nonparabolicity coefficient. The formalism is implemented in a one-dimensional Schrödinger-Poisson solver which is applicable both to silicon inversion layers and heterostructures.

Keywords: Two-dimensional electron gas, heterostructures, MOS capacitor, inversion layer, Schroedinger-Poisson solver, self-consistent

1. INTRODUCTION

To accurately model the high-field transport in silicon inversion layers, several authors [3, 4] have introduced a nonparabolicity correction in the subband dispersions. In this work we quantitatively analyze nonparabolicity effects in various two-dimensional electron gases. For this purpose a self-consistent Schrödinger-Poisson solver has been developed, capable of dealing with silicon inversion layers and heterostructures. For heterostructures, position-dependent material parameters are taken into account. As a result each subband is characterized by three parameters, E_n, m_n, α_n, which denote the subband energy, mass and nonparabolicity coefficient, respectively. This set of parameters is intended to serve as input for high-field transport calculations.

Our approach relies on the effective-mass approximation which is applicable if the confining potential, $V(z)$, satisfies two conditions [3]:

1. $V(z)$ is slowly varying over a unit cell,
2. matrix elements of $V(z)$ between Bloch functions of different bands are negligible.

2. SILICON INVERSION LAYERS

Within the effective-mass approximation a one-dimensional Schrödinger equation can be derived

* Corresponding author: Tel.: +43/1/58801-3719, Fax: +43/1/5059224, e-mail: kosina@iue.tuwien.ac.at.

from the three-dimensional one provided that the potential only varies in one dimension.

$$\left(\varepsilon\left(-i\frac{\partial}{\partial z};K\right)+V(z)\right)\zeta_n(z;K)=E_n(K)\zeta_n(z;K) \tag{1}$$

In this equation, ζ_n denotes the envelope function, K is the in-plane wave number, and $E(K)$ represents the in-plane dispersion relation. The bulk dispersion relation, $\varepsilon(\boldsymbol{k})$, is assumed to have ellipsoids as equi-energy surfaces, $\varepsilon(\boldsymbol{k})=\varepsilon(K^2/m_{xy}+k_z^2/m_z)$. In principle, (1) can be solved numerically for different values of K so as to obtain a point-wise representation of $E(K)$. However, since the bulk dispersion is given by an analytic function one usually is interested in obtaining analytic subband dispersions as well. The latter can be found by applying perturbation theory at $K=0$. The kinetic energy operator is expanded into a Taylor series, and terms up to K^4 are retained in order to get information on the subband nonparabolicity.

$$\varepsilon\left(\frac{K^2}{m_{xy}}-\frac{1}{m_z}\frac{\partial^2}{\partial z^2}\right)\approx \boldsymbol{T}_0+\boldsymbol{T}_1K^2+\boldsymbol{T}_2K^4 \tag{2}$$

The unperturbed problem is defined by $\boldsymbol{T}_0$, and the terms containing the in-plane wave number are considered as perturbation. The operators $\boldsymbol{T}_i$ are given by:

$$\begin{aligned}\boldsymbol{T}_0 &= \varepsilon\left(-\frac{1}{m_z}\frac{\partial^2}{\partial z^2}\right)\\ \boldsymbol{T}_1 &= \frac{1}{m_{xy}}\cdot\varepsilon'\left(-\frac{1}{m_z}\frac{\partial^2}{\partial z^2}\right)\\ \boldsymbol{T}_2 &= \frac{1}{2m_{xy}^2}\cdot\varepsilon''\left(-\frac{1}{m_z}\frac{\partial^2}{\partial z^2}\right)\end{aligned}$$

3. HETEROSTRUCTURES

For heterostructures we assume nonparabolic dispersion relations for the different semiconductors. The usual implicit definition of the kinetic energy can be generalized to a kinetic energy operator, even when the parameters are position-dependent.

$$\boldsymbol{T}+\boldsymbol{T}\alpha(z)\boldsymbol{T}=\frac{\hbar^2}{2}\left(\frac{K^2}{m_{xy}(z)}-\frac{\partial}{\partial z}\frac{1}{m_z(z)}\frac{\partial}{\partial z}\right) \tag{3}$$

This equation, which is self-adjoint, can be solved for the kinetic energy operator.

$$\begin{aligned}\boldsymbol{T} &= \alpha^{-1/2}h(\boldsymbol{G}_0+\boldsymbol{G}_1K^2)\alpha^{-1/2} \qquad (4)\\ \boldsymbol{G}_0 &= -\alpha^{1/2}\frac{\partial}{\partial z}\frac{\hbar^2}{2m_z}\frac{\partial}{\partial z}\alpha^{1/2}\\ \boldsymbol{G}_1 &= \frac{\hbar^2}{2}\cdot\frac{\alpha(z)}{m_{xy}(z)}\\ h(x) &= \frac{2x}{1+\sqrt{1+4x}}\end{aligned}$$

In analogy with (2) the kinetic energy operator has to be decomposed as follows,

$$\boldsymbol{T}(\boldsymbol{G}_0+\boldsymbol{G}_1K^2)\approx \boldsymbol{T}_0+\boldsymbol{T}_1K^2+\boldsymbol{T}_2K^4, \tag{5}$$

where the determination of the operators $\boldsymbol{T}_i$ for heterostructures is more complicated than for uniform material parameters since the operators $\boldsymbol{G}_0$ and $\boldsymbol{G}_1$ no longer commute.

4. SUBBAND DISPERSION RELATION

The eigenvalues of (1) at $K=0$ are denoted by E_n^0 and the eigenfunctions by $\zeta_n^0(z)$ $(n=0,1,2\ldots)$. The matrix elements of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ in the $\{\zeta_n^0\}$ basis are $T_{1,mn}$ and $T_{2,mn}$, respectively. Perturbation theory yields a polynomial representation for $E_n(K)$:

$$E_n(K)=E_n^0+T_{1,nn}K^2+\left(\sum_{m\neq n}\frac{|T_{1,mn}|^2}{E_n^0-E_m^0}+T_{2,nn}\right)K^4 \tag{6}$$

This expression allows to characterize each subband by an effective mass, m_n, and a nonparabo-

licity coefficient, α_n defined by

$$m_n = \frac{\hbar^2}{2T_{1,nn}} \tag{7}$$

$$\alpha_n = -\frac{1}{T_{1,nn}^2}\left(\sum_{m\neq n}\frac{|T_{1,mn}|^2}{E_n^0 - E_m^0} + T_{2,nn}\right) \tag{8}$$

Although (6) suggests a representation of $E_n(K)$ as a polynomial in K it appears favorable to assume a second-order polynomial in the energy instead.

$$(E - E_n^0)(1 + \alpha_n(E - E_n^0)) = \frac{\hbar^2 K^2}{2m_n} \tag{9}$$

By comparison with $E_n(K)$ obtained from a solution of (1) at discrete K-points it was found that (9) gives an excellent approximation for $E_n(K)$ in a wide K-range, whereas the polynomial (6) is applicable only for sufficiently small K.

5. ELECTRON DENSITY

The electron density is given as the sum of all position probabilities of the states $|n, K\rangle$ weighted by their occupation probabilities.

$$n(z) = g_v \sum_n \sum_K |\zeta_n(z; K)|^2 f(E_n(K)) \tag{10}$$

In (10) f denotes the Fermi-Dirac distribution function and g_v the valley degeneracy factor. The summation over K is usually converted to an integral by employing the two-dimensional density of states, $\rho_{2D,n}$.

$$\rho_{2D,n(E)} = g_v \frac{m_n}{\pi\hbar^2}(1 + 2\alpha_n(E - E_n^0)) \tag{11}$$

$$n(z) = \sum_n \int_{E_n^0}^{\infty} |\zeta_n(z; K)|^2 f(E)\rho_{2D,n(E)} \mathrm{d}E \tag{12}$$

The hole density is calculated using Boltzmann statistics. Once the carrier densities are known the electrostatic potential is obtained by solving the one-dimensional Poisson equation. During each self-consistent iteration step the Schrödinger equation and the linearized Poisson equation are solved.

6. DISCRETIZATION

The wave functions, which are assumed to vanish at the boundaries, are represented by a Fourier series. After truncation of the series the Schrödinger equation (1) is converted into an algebraic eigenvalue equation [1].

The linearized Poisson equation is discretized in real space using the finite difference method. This yields a tridiagonal equation system which can be efficiently solved. During each iteration step the Fast Fourier Transform is applied to transfer the required quantities from real-space representation to momentum representation and vice versa.

7. RESULTS AND DISCUSSION

7.1. Silicon Inversion Layer

A MOS capacitor has been simulated. The parameters were chosen as follows: $m_l = 0.98$, $m_t = 0.19$, $\alpha = 0.7\,\mathrm{eV}^{-1}$, $t_{ox} = 7\,\mathrm{nm}$, $N_A = 5 \cdot 10^{16}$ cm^{-3}, $V_{GB} = 2.5\mathrm{V}$, $T = 300\,\mathrm{K}$. A nonparabolic bulk dispersion relation was assumed:

$$\varepsilon = \frac{2\gamma}{1 + \sqrt{1 + 4\alpha\gamma}}, \qquad \gamma = \frac{\hbar^2}{2}\left(\frac{K^2}{m_{xy}} + \frac{k_z^2}{m_z}\right). \tag{13}$$

In Figure 1 the Fourier coefficients of the first two wave functions are plotted. This corresponds to a representation of the wave functions in momentum space. The number of harmonics equals $N = 64$. Due to nonparabolicity the motion of the carriers normal to the interface is no longer decoupled from the motion parallel to the interface, as is the case for parabolic bands. For nonvanishing K narrower wave functions are observed than for $K = 0$.

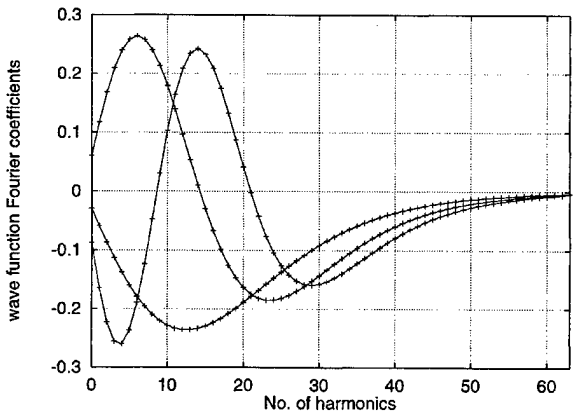


FIGURE 1 Spectrum of the first three wave functions in a silicon inversion layer.

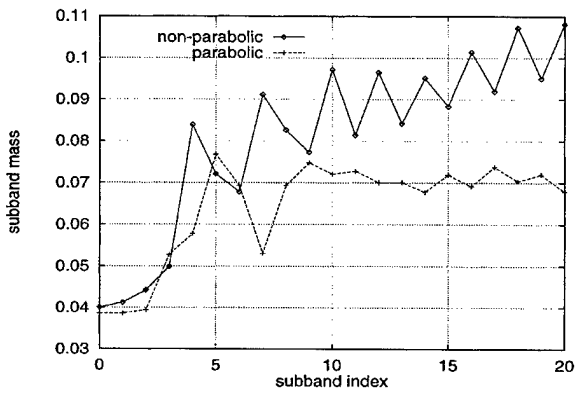


FIGURE 2 The subband masses for the InP heterostructure for both the nonparabolic and parabolic case.

7.2. InP-Based Pseudomorphic HEMT

A heterostructure after [2] has been simulated (Tab. I). In the simulation both nonparabolic and parabolic bulk dispersions have been considered. When moving across this heterostructure an electron will experience considerable variations of the bandstructure parameters.

The subband parameters m_n and α_n are plotted in Figures 2 and 3 for 20 subbands. Strong variations of these parameters can be observed. In the nonparabolic case, the mass increases significantly when going from subband 3 to 4, while in turn the nonparabolicity coefficient decreases. This means that the changes of m_n and α_n are correlated. This behavior can be understood when considering the wave functions (Fig. 4). While an electron in subband 3 resides preferably in the InGaAs channel (low mass, high α), an electron in subband 4 feels predominantly the material properties of the AlInAs barrier.

TABLE I Parameters of the InP heterostructure. The donor layer is doped with $N_D = 2.10^{18}$ cm^{-3}

Layer	Material	t (nm)	mass (m_0)	α (eV^{-1})
barrier	$Al_{0.48}In_{0.52}As$	20	0.082	0.84
donor	$Al_{0.48}In_{0.52}As$	12.5	0.082	0.84
spacer	$Al_{0.48}In_{0.52}As$	2	0.082	0.84
channel	$In_{0.53}Ga_{0.47}As$	20	0.038	1.02
substrate	InP	20	0.077	0.83

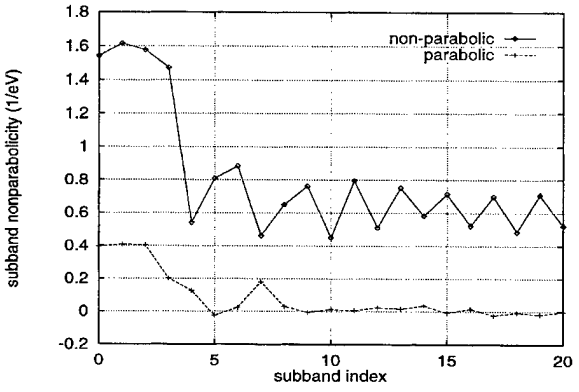


FIGURE 3 The subband nonparabolicity coefficients for the InP heterostructure. Although for the lower curve parabolic bulk dispersions are assumed, the subbands are nonparabolic due to the material inhomogeneity.

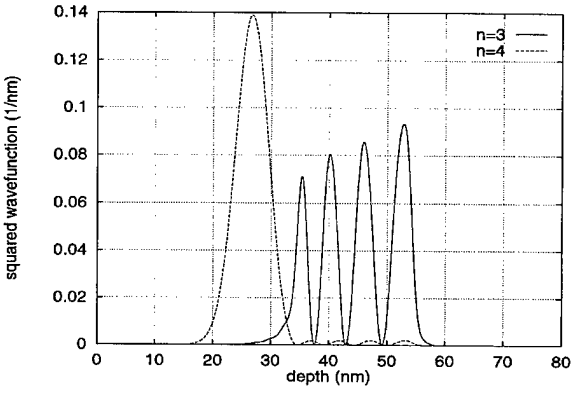


FIGURE 4 Wavefunctions 3 and 4 in the InP heterostructure.

Acknowledgment

This work has been supported by the "*Fonds zur Förderung der wissenschaftlichen Forschung*", Project No. P10642 PHY.

References

[1] Fischetti, M. and Laux, S. (1993). "Monte Carlo Study of Electron Transport in Silicon Inversion Layers", *Physical Review B*, **48**(4), 2244–2274.

[2] Jungemann, C., Emunds, A. and Engl, W. (1993). "Simulation of Linear and Nonlinear Electron Transport in Homogeneous Silicon Inversion Layers", *Solid-State Electron.*, **36**(11), 1529–1540.

[3] Abou-Elnour, A. and Schuenemann, K. (1994). "An Efficient and Accurate Self-Consistent Calculation of Electron States in Modulation Doped Heterostructures", *Solid-State Electron.*, **37**(1), 27–30.

[4] Ali, F. and Gupta, A. (1991). *HEMT's and HBT's: Devices, Fabrication and Circuits*. Artech House.

Authors' Biographies

Hans Kosina received the 'Diplomingenieur' degree in electrical engineering and the Ph.D. degree from the Vienna University of Technology in 1987 and 1992, respectively. He was for one year with the 'Institut für flexible Automation' and joined then the 'Institut für Mikroelektronik', where he is currently employed as an Assistant Professor. In summer 1993 he was Visiting Scientist at the Advanced Product Research and Development Laboratory at Motorola Inc., Austin, Texas. His current interests include modeling of hot carrier phenomena and quantum effects in semiconductor devices, and computer aided engineering in VLSI technology.

Christian Troger was born in Vienna, Austria, in 1969. He studied electrical engineering at the Vienna University of Technology, where he received the degree of 'Diplomingenieur' in 1995. He joined the 'Institut für Mikroelektronik' in September 1995, where he is currently working for his doctoral degree. His work is focused on the simulation of quantum mechanical devices.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 495–500
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Advantages of Semiconductor Device Simulator Combining Electromagnetic and Electron Transport Models*

S. M. SOHEL IMTIAZ, SAMIR M. EL-GHAZALY† and ROBERT O. GRONDIN

Department of Electrical Engineering, Telecommunications Research Center, Arizona State University, Tempe, AZ 85287-7206

(Received 18 May 1997; In final form 10 July 1997)

Physical simulation of semiconductor devices at high frequencies involves not only semiconductor transport issues but also electromagnetic wave propagation issues. In order to obtain the nonlinear and the large-signal characteristics of the semiconductor devices, an electromagnetic model should replace the traditional quasi-static model in the device simulator. In this paper, the advantages of a semiconductor device simulator combining an electromagnetic and an electron transport models are presented. This study is based on a semiconductor device simulator that couples a semiconductor model to the 3D time-domain solution of Maxwell's equations. The electromagnetic wave propagation effects on the millimeter-wave FETs are thoroughly analyzed. The use of the electromagnetic model over the conventional quasi-static model provides the actual device response at high frequencies. It also shows the nonlinear energy build-up along the device width whereas the quasi-static model provides a linear increase of energy. The combined model is capable of predicting the device nonlinearity and harmonic distortion of amplifier circuits at large signal.

Keywords: Device simulation, hydrodynamic models, FDTD, full-wave simulators

1. INTRODUCTION

With the advancement of semiconductor technology, the techniques required to analyze, design, and optimize the semiconductor devices are becoming increasingly sophisticated. The computer simulation programs are now essential tools for device engineers. These numerical simulations based on physical modeling can be used to predict and provide better understanding of the device behavior. However, the down-sizing of the active device dimensions has presented new challenges to the device and circuit designer. In submicron semiconductor devices, several new transport

* This work is supported by the Army Research Office under contract # DAAH04-95-1-0252.
†Corresponding author. Phone: 602-965-5322, Fax: 602-965-8325, Email: sme@asu.edu.

phenomena develop and, consequently, have to be considered in device modeling. The electrons do not reach equilibrium transport conditions while traveling along the conducting channel. To incorporate the effects of nonstationary dynamics in semiconductor devices, the hydrodynamic model based on moments of Boltzmann's transport equation is used [1–3]. The transport parameters are taken as functions of average electron energy rather than the local electric field.

The device modeling at high frequencies requires special attention. At high frequencies, the coupling between the electrons and the propagating electromagnetic waves can not be neglected in submicron devices. The short period of the propagating EM wave approaches the electron relaxation time and as the electrons need a finite time to adjust their velocities to the changes in field, electron transport is directly affected by the propagating wave. In such cases, the quasi-static semiconductor device models fail to represent accurately the exact device response. In addition, the electrodes extending along the device width behave like transmission lines with nonlinear characteristics. These facts call for the necessity of incorporating wave effects in a three-dimensional model. This goal can be achieved by taking full account of the varying fields inside the device. The acceptable method for representing these various forces is to combine an electromagnetic model with a semiconductor device model which leads to the Combined Electromagnetic and Solid-State (CESS) simulator [4].

On the other hand, the increasing demand of processing and transmitting more informations at a faster rate, drives the analog and digital electronic systems to operate at higher clock speeds. At the same time, to curtail the production cost, the manufacturers are more inclined towards heavily densed integrated circuits. In these high density integrated circuits, there are many closely spaced active and passive devices. As a result, there are some detrimental effects on the circuit performance at high frequencies due to crosstalk caused by coupling, surface waves and radiation effects. In such cases, the circuit modeling issue becomes more intensive. The circuit design should be based on advanced global model which takes the electromagnetic wave effects into consideration.

The issues like device-wave interaction, electromagnetic coupling, discontinuity problem, linear and nonlinear behavior of passive and active devices, and EM radiation effects are addressed in the global modeling. The computer memory requirement as well as the simulation time is reduced by using a hybridization approach in global modeling. The amplifier is divided into three regions, preserving the physical characteristics of the amplifier circuit by taking the reflections at the breaking points into consideration. The full-wave analysis of each region is performed individually and coupled to the next stage properly with all the required informations from the preceding stage. This technique enables one to use large space step, and hence, large time step in matching networks.

2. NUMERICAL MODEL

The CESS simulator is a physically based model which takes care of nonisothermal transport and nonstationary electron dynamics as well as electromagnetic wave propagation effects. This model couples the hydrodynamic model to a 3D time-domain solution of Maxwell's equations. The hydrodynamic model is based on the moments of the Boltzmann's transport equation obtained by integration over the momentum space. The electromagnetic wave propagation effects can be completely characterized by solving Maxwell's equations. These equations are first-order linearly coupled differential equations relating the field vectors, current densities and charge densities at any point in space at any time. The coupling between the two models is established by using the fields obtained from the solution of Maxwell's equation in the semiconductor model to calculate the current densities inside the device. These current densities are used to update the electric and the magnetic fields using Maxwell's equations

with an applied high frequency sinusoidal excitation. The initialization is provided by solving the hydrodynamic model for the dc charges and currents in response to a specified dc operating point. In this manner, the coupling between the two models results in the overall high frequency characteristics of the semiconductor devices. The details of the mathematical representation and the coupling procedure can be found in [4]. The finite-difference time-domain scheme is used in semiconductor device discretization.

3. RESULTS AND DISCUSSION

The MESFET (Fig. 1) used in this work have the following parameters. Gate-source spacing = 0.4 μm, gate-drain spacing = 0.56 μm, gate length = 0.24 μm, undoped GaAs thickness = 0.3 μm, active layer thickness = 0.1 μm, gate width = 250 μm. In order to validate the CESS simulator, a MODFET structure similar to Shawki *et al.* [5] is simulated to compare the performances. The transconductances are compared with/without taking the traps into account. They exhibit reasonable agreement with each other in Figure 2.

To demonstrate the electromagnetic-wave propagation effects for MESFET, a sinusoidal excitation of peak 0.1 V and frequency 80 GHz is applied between the gate and the source electrodes. The excitation is applied as a plane source at $z = 0$, as shown in Figure 1. The CESS model is then solved for a few rf cycles to avoid the effects of the

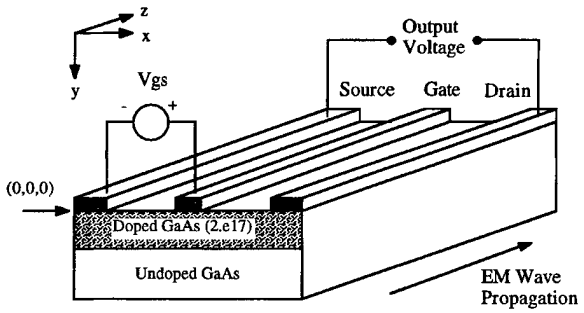


FIGURE 1 The simulated MESFET structure.

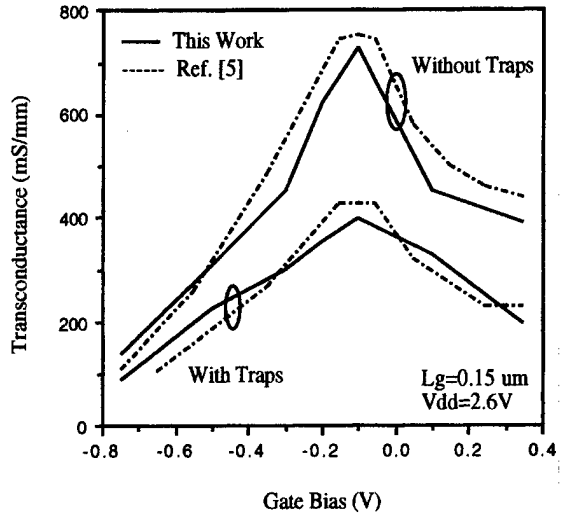


FIGURE 2 Comparison of transconductance of MODFET with Ref. [5] with or without including the traps.

transients on the ac solution. The output is obtained across the drain and the source at several points along the device width in the z-direction. The output voltage wave, as shown in Figure 3, first decreases and then increases along the device width. Early in the simulation, the electronic effect is not present and the wave amplitude decreases along the device width. Later, as more and more electromagnetic energy is propagated along the device width, the wave energy builds up, and the wave amplitude increases. This figure clearly

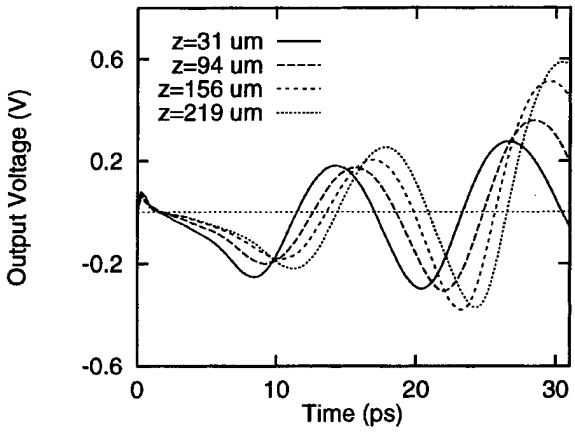


FIGURE 3 The electromagnetic wave propagation effects on output voltages of MESFET for different device widths.

demonstrates the direct relationship between the device gain characteristics and the electromagnetic wave propagations.

The advantage of using the electromagnetic model in device simulation is demonstrated in Figure 4. MESFET is simulated using the quasi-static model as well as the CESS simulator. In quasi-static model, Poisson's equation is solved to get the electric fields. In electromagnetic model, Maxwell's equations are solved to obtain the electric and the magnetic fields. In Figure 4, the output voltage wave monotonously increases along the device width in quasi-static model. On the other hand, in electromagnetic model, the output voltage wave nonlinearly increases with the device width. This phenomenon is expected due to the device-EM wave interaction. The exchange of energy between the electrons and the electromagnetic wave takes place along the device width. This behavior is absent when the output is obtained from the quasi-static analysis. This figure strongly supports the use of the electromagnetic model for device simulation at high frequencies.

To demonstrate the large signal potential of the CESS simulator, the gain of a 0.5 μm×1000 μm MESFET is calculated for a small signal of 0.1 V and a large signal of 0.3 V. As shown in Figure 5, the gain becomes lower as the amplitude increases, which is expected. The strength of this approach is not in simply confirming that larger amplitudes reduce the gain, but in estimating the reduction directly, using the physical model.

The potential of the CESS model is further investigated by simulating an amplifier circuit with input and output matching networks as shown in Figure 6. The physical simulation of the entire amplifier is performed using a global modeling technique [6]. The optimized transistor parameters (in terms of maximum gain) used in the amplifier

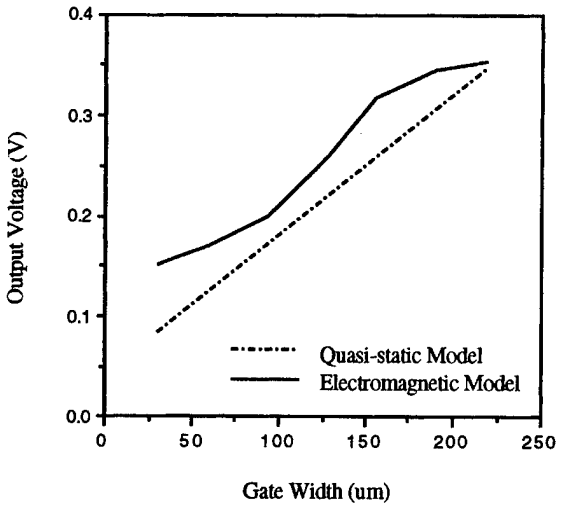


FIGURE 4 The comparison of output voltage variations with device-width obtained from the electromagnetic model and the quasi-static model for MESFET.

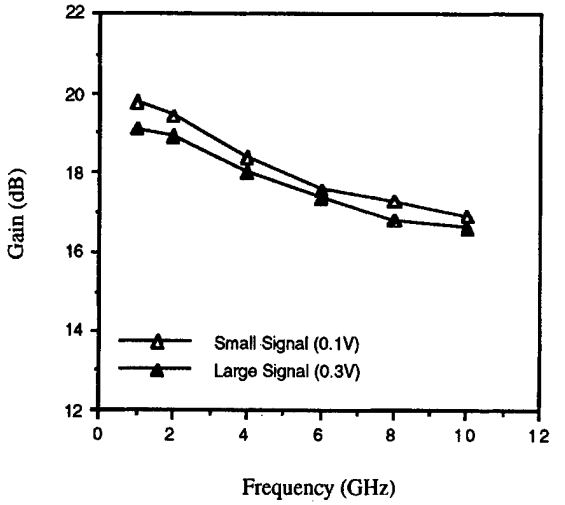


FIGURE 5 The comparison of gain characteristics of MESFET at small and large signal operations.

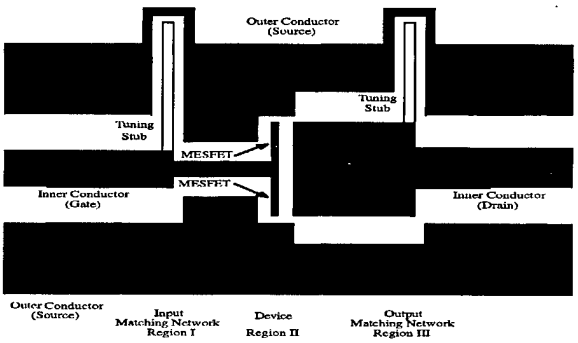


FIGURE 6 GaAs transistor amplifier with input and output matching networks.

are: gate length = 0.22 μm, aspect ratio = 2.5 and the active layer doping = $2.2\times10^{17}/cm^3$. The amplifier is designed for 40 GHz frequency. The entire amplifier is simulated by applying a small signal of amplitude 0.1 V and large signal of amplitudes 0.3 and 1.0 V at different frequencies. In Figure 7, S_{21} is presented for small and large signals at different frequencies. At the design frequency of 40 GHz, the gain is 7.38 dB for small signal and 6.44 and 5.15 dB for large signals of 0.3 and 1.0 V, respectively. Thus it is observed that for the same design the gain drops at large signals. In small signal, the gain drops slightly as the frequency is shifted away from the 40 GHz point. On the other hand, in large signal, the gain reduction at frequencies other than design frequency is higher. Nonlinearity in the device behavior is evident from this figure.

Once the large signal response of the amplifier is obtained, it is interesting to study its frequency content and to identify the harmonics. The output wave contains a significant amount of third and fifth harmonic components as shown in Figure 8. For the large signal input of 0.3 V, the output power contains 4.8%, 11.6% and 13% of the fundamental at 30, 40 and 50 GHz, respectively, at the third harmonic. The fifth harmonic contains 6%, 0.7% and 0.3% of the fundamental at 30, 40 and 50 GHz, respectively. This shows that a considerable amount of power is transferred at the harmonic components. Thus the global model is able to predict the nonlinearity of the device behavior and show the different harmonic components generated at the amplifier output due to nonlinearity.

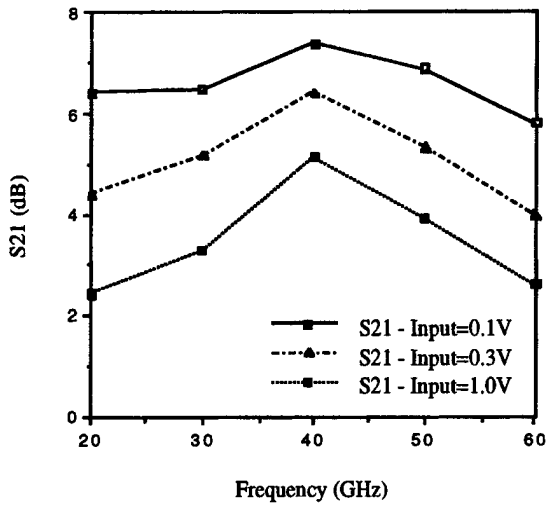


FIGURE 7 The dependence of scattering parameter S_{21} on frequency at small and large signals for the amplifier with the optimized transistor.

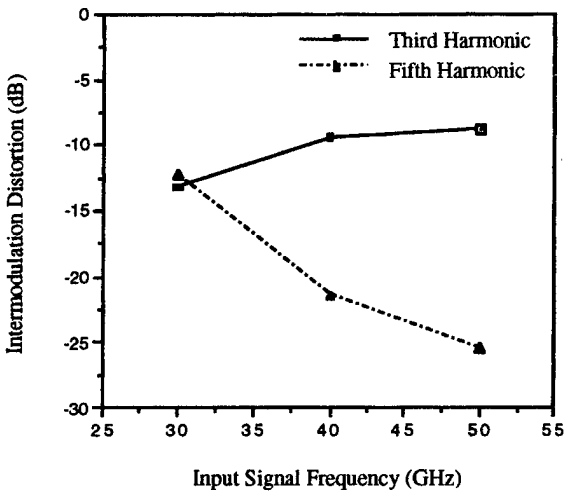


FIGURE 8 The intermodulation distortion at the third and the fifth harmonic components for the large signal of amplitude 0.3 at 30, 40 and 50 GHz.

4. CONCLUSIONS

Advantages of semiconductor device and circuit simulator combining an electromagnetic and an electron transport models are presented. The simulation confirms that a significant device-wave interaction takes place in semiconductor devices at high frequencies. The use of the electromagnetic model over the conventional quasi-static model provides the actual device response at high frequencies. It also shows the nonlinear energy build-up along the device width whereas the quasi-static model provides a linear increase of energy. An approach towards global modeling of milli-

meter-wave circuits is also presented in this paper. The global model is able to characterize the electromagnetic coupling, device-EM wave interaction and the EM radiation effects of the very closely spaced integrated circuit amplifier. The global modeling technique is capable of representing the nonlinearity and the harmonic distortion of the amplifier circuit. More effects will be added in the future including thermal and packaging effects. By incorporating all these effects in the circuit simulation, a milestone will be reached towards the comprehensive global modeling.

References

[1] Carnez, B., Cappy, A., Kaszynski, A., Constant, E. and Salmer, G., "Modeling of a sub-micrometer gate Field-effect transistors including effects of nonstationary electron dynamics", *J. Appl. Phys.*, **51**(1), 784–790, Jan. 1980.

[2] Snowden, C. M. and Loret, D., "Two-dimensional hot-electron models for short-gate-length GaAs MESFET's", *IEEE Trans. Electron Devices*, **ED-34**(2), 212–223, Feb. 1987.

[3] El-Ghazaly, S. M. and Itoh, T., "Two-dimensional numerical simulation of short-gate-length GaAs MESFETs and application to the traveling Gunn domain phenomenon", *Int. J. Numerical Modeling*, **1**, 19–30, Jan. 1988.

[4] Alsunaidi, M. A., Sohel Imtiaz, S. M. and El-Ghazaly, S. M. "Electromagnetic wave effects on microwave transistors using a full-wave time domain model", *IEEE Trans. Microwave Theory Tech.*, **44**, 799–808, June 1996.

[5] Shawki, T., Salmer, G. and El-Sayed, O., "MODFET 2-D hydrodynamic energy modeling: optimization of subquarter-micron gate structure", *IEEE Trans. Electron Devices*, **37**, 21–30, Jan. 1990.

[6] Sohel Imtiaz, S. M., "Physical simulation of high frequency semiconductor devices and amplifier circuits", Ph.D. Dissertation, May 1997.

Authors' Biographies

S. M. Sohel Imtiaz was born in Dhaka, Bangladesh, in 1966. He received the B.S. and the M.S. degree in electrical engineering in 1988 and 1990, respectively, from Bangladesh University of Engineering and Technology (BUET), Dhaka. He received his Ph.D. degree in electrical engineering from Arizona State University in 1997. Dr. Imtiaz worked as a Lecturer in the department of electrical engineering in BUET from 1989 to 1991. His research interests include modeling, simulation and characterization of microwave semiconductor devices and circuits, device-wave interactions, numerical techniques, and the simulation of microwave amplifiers. He joined Micro Linear Corporation as a Sr. Device Engineer in June 1997.

Samir M. El-Ghazaly received the Ph.D. degree, in Electrical Engineering, from the University of Texas at Austin, Texas, in 1988. He joined Arizona State University as Assistant Professor in August 1988, and became Associate Professor in 1993. Dr. El-Ghazaly is a senior member of IEEE, an elected member of Commissions A and D of URSI, a member of Tau Beta Pi, Sigma Xi and Eta Kappa Nu. He is the secretary of US National Committee of URSI, Commission A. He is also the Chairman of the Chapter Activities Committee of the IEEE Microwave Theory and Techniques Society.

Robert Grondin was born and raised in Michigan. He attended the University of Michigan obtaining the BS, MS and Ph.D. degrees. From 1981 to 1983 he was a post-doctoral research associate at Colorado State university. In 1983 he joined the faculty of the Department of Electrical Engineering at Arizona State University where he is presently an associate professor. He is a senior member of the IEEE and a member of the AAAS.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 501–505
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Quantum Transport and Thermoelectric Properties of InAs/GaSb Superlattices

J.-F. LIN and D. Z.-Y. TING*

Department of Physics, National Tsing Hua University, Hsinchu, Taiwan 300, ROC

In recent years, artificially layered microstructure have been considered as candidates for better thermoelectrics. In this work we examine transport properties of the type-II broken-gap InAs/GaSb superlattice. We use the effective bond orbital model for an accurate description of the band structures. Theoretical results of thermoelectric transport coefficients and the dimensionless figure of merit for an (8, 8)-InAs/GaSb type-II superlattice are presented.

Keywords: Thermoelectric, type-II superlattice, superlattice transport, InAs, GaSb

1. INTRODUCTION

From about 1940 to 1965, many scientists made tremendous effort looking for superior thermoelectric materials. The advantage of using these solid state devices are compactness, quietness (no moving parts), freedom from corrosion, localized heating or cooling, and reliability. The effort dwindled eventually because performance of most thermoelectric materials found at that time were too poor to be used for practical commercial applications. Recently, there is a renewed interest in this area [1, 2], driven by the following reasons: First, thermoelectric technology is environmentally cleaner than traditional compressor-based refrigeration technology since it does not use chlorofluorocarbons. Second, several new materials were identified as potential candidates for better thermoelectrics, including the filled skutterudite antimonides [3] and $PbTe/Pb_{1-x}Eu_xTe$ multiple-quantum-well structures [4]. Among the new materials, superlattices attracted many scientists' attention [5, 6, 7, 8, 9]. The interest can be traced back to the quantitative results first obtained by Hicks and Dresselhaus [5] where huge enhancement of thermoelectric properties was predicted for superlattice structures.

In the present work, we focus our interest on the type-II broken-gap InAs/GaSb superlattices. Superlattices consisting of combinations of III–V semiconductors with type-II band alignments are of interest for infrared applications, including IR detectors [10] and laser diodes [11, 12]. This is because their energy gaps can be made smaller

* Corresponding author.

than their constituents. However, for most of these applications, cooled operation are desirable for good performance. If reasonable thermoelectric properties of such superlattice structure could be obtained, directly integration of thermoelectric cooling devices with a set of IR detectors or laser diodes may offer some advantage. Our aim here is to present the early results of our ongoing theoretical effort to identify and characterize the thermoelectric properties of the type-II broken-gap InAs/GaSb superlattices. In Sec. II, the theoretical framework for calculating electronic contributions to the thermoelectric properties of superlattices using realistic band structure models is presented. Our results for an (8, 8)-InAs/GaSb superlattice is presented in Sec. III.

2. MODEL

The dimensionless figure of merit for a thermoelectric material is given by [14, 15]

$$ZT = \frac{S^2 \sigma T}{\kappa} \tag{1}$$

where S is the thermopower (thermoelectric power or Seebeck coefficient), σ is the electrical conductivity, κ is the total thermal conductivity ($\kappa = \kappa_l + \kappa_e$, the lattice and electronic contributions, respectively), and T is the temperature. The maximum Coefficient Of Performance (COP) is directly related to the dimensionless figure of merit of the material by [14, 15]

$$COP = \frac{T_c}{T_h - T_c} \frac{\sqrt{1+ZT} - T_h/T_c}{\sqrt{1+ZT}+1}, \tag{2}$$

where T_c, T_h are the temperatures at the cold and hot junctions, respectively, and T, which is equal to $(T_h + T_c)/2$, is the mean absolute temperature, for a single-stage thermoelectric cooler. As ZT goes to infinity, the COP goes to the thermodynamic limit of Carnot efficiency.

The low-field transport coefficients for thermoelectric materials are defined by

$$\mathbf{J} = \sigma \mathbf{E} - \sigma S \nabla T \tag{3}$$

$$\mathbf{J_Q} = T\sigma S \mathbf{E} - \kappa_0 \nabla T \tag{4}$$

where $\mathbf{J}$ is the electric current density, $\mathbf{J_Q}$ the thermal current density, $\mathbf{E}$ the electric field, T the temperature, σ the electrical conductivity, S the thermopower, and κ_0 the thermal conductivity at zero electric field. In general, σ, S, and κ_0 are 3×3 matrices in the Cartesian coordinates. Note that the value of κ needed for computing Z is the total heat flow for zero electrical current, so it is given by

$$\kappa = \kappa_0 - T\sigma S^2 + \kappa_l \tag{5}$$

where κ_l is the lattice contribution to the thermal conductivity.

Using the relaxation time approximation to the Boltzmann equation, the transport coefficients are given by [16]

$$\sigma = e^2 \int_{-\infty}^{\infty} d\varepsilon \left(-\frac{\partial f_0}{\partial \varepsilon}\right) \boldsymbol{\Sigma}(\varepsilon), \tag{6}$$

$$T\sigma S = -e \int_{-\infty}^{\infty} d\varepsilon \left(-\frac{\partial f_0}{\partial \varepsilon}\right) \boldsymbol{\Sigma}(\varepsilon)(\varepsilon - \mu), \tag{7}$$

$$T\kappa_0 = \int_{-\infty}^{\infty} d\varepsilon \left(-\frac{\partial f_0}{\partial \varepsilon}\right) \boldsymbol{\Sigma}(\varepsilon)(\varepsilon - \mu)^2, \tag{8}$$

where μ is the chemical potential, e the electron charge, and f_0 the Fermi-Dirac distribution function. Here $\boldsymbol{\Sigma}(\varepsilon)$ is given by:

$$\boldsymbol{\Sigma}(\varepsilon) = \tau(\varepsilon) \sum_k \mathbf{v}(k)\mathbf{v}(k)\delta(\varepsilon - \varepsilon(k)) \tag{9}$$

where $\mathbf{v}(k) = \nabla_k \varepsilon(k)/\hbar$ is the semi-classical carrier velocity, $\varepsilon(k)$ is the dispersion relation for the carriers, and $\tau(\varepsilon)$ is the energy-dependent relaxation time tensor, taken to be k-independent. The summation is over the first Brillouin zone. Equations (6), (7), and (8) are the basic results of the theory of electronic contributions to the

thermoelectric effects. It should be noted that phonon contributions are ignored here.

The theoretical scheme we used for calculating of the electronic structure of InAs/GaSb superlattices is the Effective Bond-Orbital Model (EBOM) [13]. This method combines the virtues of the $\mathbf{k}\cdot\mathbf{p}$ and the tight-binding methods. The basic idea is to use a minimum number of bond orbitals to describe, as accurately as possible, the most relevant portion of the bulk band structures, and then use them in a supercell calculation to obtain superlattice band structures. In our case, eight bond orbitals per unit cell, including the s-like bond orbitals with spin up and spin down, four bond orbitals each with total angular momentum $J = 3/2$ (made of p-like states coupled with spin) and two additional bond orbitals each with total angular momentum $J = 1/2$ (also made of p-like states coupled with spin), are used. This is because the superlattice states of interest contain admixture of both valence-band and conduction-band characteristics. We assume that all the bond orbitals are sufficiently localized so that the interaction between orbitals separated farther than the nearest neighbor distance can be ignored. All nonvanishing interaction parameters can then be directly related to the effective masses or other band parameters of the $\mathbf{k}\cdot\mathbf{p}$ perturbation theory. Based on this model, an accurate band structure could be obtained for values of $\mathbf{k}$ near the zone center.

Our strategy for computing thermoelectric properties is as follows. First, we calculate the superlattice band structures by using SOBO model. Then, we perform full Brillouin zone integration to obtain $\boldsymbol{\Sigma}(\varepsilon)$. Finally, we use Equations (6), (7) and (8) to compute the transport coefficients.

3. RESULTS AND DISCUSSION

Figure 1 shows our calculated band structure for an InAs/GaSb superlattices eight monolayers of InAs and eight monolayeres of GaSb per period.

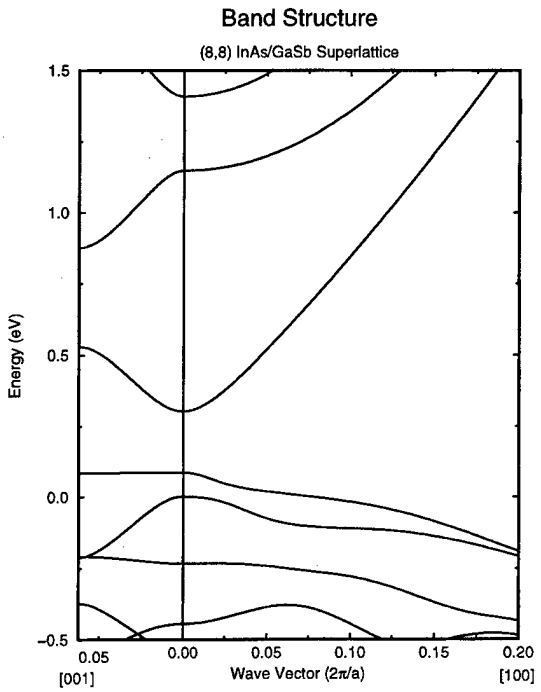


FIGURE 1 Illustration of band structure of (4, 4) InAs/GaSb superlattices. Note that in our calculation the strain effect is included.

The conduction band minimum of bulk InAs is taken to be at 0 eV. Experimental band offset value [17] between InAs and GaSb then puts the valence band top of GaSb at approximately 200 meV above the conduction band minimum of the InAs conduction band edge. In the figure, superlattice band structure perpendicular and parallel to growth directions are shown. The valence subband maximum is found at 0.086 eV, while the conduction subband minimum at 0.303 eV, yielding a superlattice bandgap of 0.217 eV. Note that the conduction subband structure along the superlattice growth direction is still very dispersive. This is due the broken-gap band alignment, the small InAs conduction band effective mass, and the relatively short period.

To compute thermal electric properties, we need to know relaxation times. Lacking specific knowledge of relaxation times in superlattice structures, we made an estimate from the mobility data for bulk InAs and GaSb, and assume a value of

$\tau(\varepsilon) \approx 10^{-13}$ sec. All of our calculations assume a temperature of 300 K. In Figure 2, we plot the thermoelectric transport coefficients as functions of the chemical potential. Due to the fact that transport along the growth (transverse) direction is impeded by scattering from the superlattice interfaces, the transverse components (zz) of electrical conductivity (σ) and electronic contribution to thermal conductivity (κ_e) are always smaller than the parallel (xx) components. The transverse components of all three transport coefficients show strong oscillations as functions of chemical potential. This can be understood by examining the superlattice minibands along the transverse ([001]) direction. We can easily see that the oscillations track the miniband edges.

In order to calculate the thermoelectric figure of merit for this structure we need the lattice contribution to the thermal conductivity. From the experimental data of bulk InAs and GaSb, we take the lattice thermal conductivity of superlattice as the average of two bulk values, which is approximately 35 W/K m. This procedure probably overestimates the values of lattice conductivity since we would expect interface scattering to lower thermal conductivities. In Figure 3, we show the dimensionless figure of merit ZT vs. the chemical potential. For parallel transport, the optimum value occurs when chemical potential is about 0 eV. Its value is about 0.033 which is not better than the bulk value. This result is opposite to that of Hicks' calculation [5] where ZT for a type-I superlattices better than bulk materials. The difference is probably due to the much stronger inter-well coupling in the type-II broken gap superlattices.

While our calculations have carefully treated band structure effects, the models we used to estimate relaxation times and lattice thermal conductivities have been rather crude. In particular, it is likely that we considerably over-estimated the superlattice thermal conductivity [2]. These issues must be addressed before we could obtain a more realistic description of superlattice thermoelectric properties. In summary, we have calculated the thermoelectric properties, including conductivity, electronic contribution to thermal conductivity, and thermal power, for a type-II

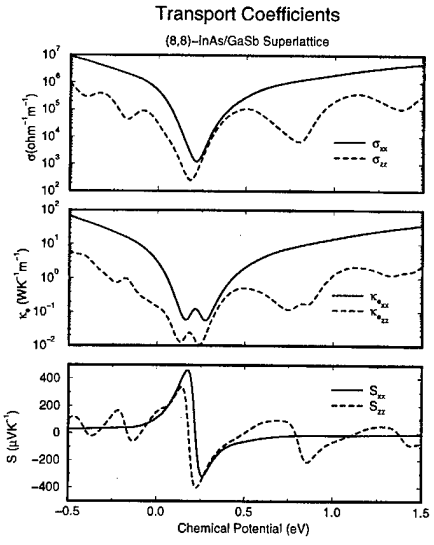


FIGURE 2 Shown are the conductivity, electronic contribution to the thermal conductivity, and thermopower vs. the chemical potential at $T = 300$ K.

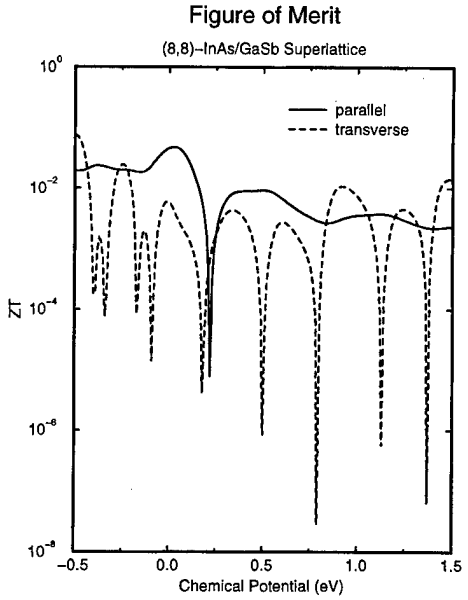


FIGURE 3 ZT vs. chemical potential for the (8, 8)-InAs/GaSb superlattices.

broken-gap InAs/GaSb superlattice using realistic band structure models. The computational tools developed form a basis for further explorations of superlattice thermoelectric properties.

Acknowledgements

The authors would like to thank C.C. Chi for helpful discussions. This work was supported by the ROC National Science Council under Grant No. NSC 86-2112-M-007-001. The use of ROC National Center for High-Performance Computing facilities is acknowledged.

References

[1] Tritt, T. M. (1996). "Thermoelectrics run hot and cold", *Science*, **272**, 1276–1277.
[2] Mahan, G., Sales, B. and Sharp, J. "Thermoelectric materials: new approaches to an old problem", *Physics Today*, March 1997, 42–47.
[3] Sales, B. C., Mandrus, D. and Williams, R. K. (1996). "Filled skutterudite antimonides: a new class of thermoelectric materials", *Science*, **272**, 1325–1328.
[4] Hicks, L. D., Harman, T. C., Sun, X. and Dresselhaus, M. S. (1996). "Experimental study of the effect of quantum-well structures on the thermoelectric figure of merit", *Phys. Rev.*, **B53**, 10493–10496.
[5] Hicks, L. D. and Dresselhaus, M. S. (1993). "Effect of quantum-well structures on the thermoelectric figure of merit", *Phys. Rev.*, **B47**, 12727–12731.
[6] Sofo, J. O. and Mahan, G. D. (1994). "Thermoelectric figure of merit of superlattices", *Appl. Phys. Lett.*, **65**(21), 2690–2692.
[7] Mahan, G. D. and Lyon, H. B. (1994). "Thermoelectric devices using semiconductor quantum wells", *J. Appl. Phys.*, **76**(3), 1899–1901.
[8] Whitlow, L. W. and Hirano, T. (1995). "Superlattice applications to thermoelectricity", *J. Appl. Phys.*, **78**(9), 5460–5466.
[9] Lin-Chung, P. J. and Reinecke, T. L. (1995). "Thermoelectric figure of merit of composite superlattices systems", *Phys. Rev.*, **B51**, 13244–13248.
[10] Chow, D. H., Miles, R. H., Schulman, J. N., Collins, D. A. and Mcgill, T. C. "Type-II superlattices for infrared detectors and devices", *Semi. Sci. Tech.*, **v6**(12C), 47–51 (1991 Dec.).
[11] Chow, D. H., Miles, R. H., Hasenberg, T. C., Kost, A. R., Zhang, Y. H., Dunlap, H. L. and West, L. "Mid-wave infrared diode-lasers based on GaInSb/InAs and InAs/AlSb superlattices", *Appl. Phys. Let.*, **67**(25), 3700–3702 (1995 Dec. 18).
[12] Miles, R. H., Chow, D. H., Zhang, Y. H., Brewer, P. D. and Wilson, R. G. "Midwave infrared stimulated-emission from a GaInSb/InAs superlattices", *Appl. Phys. Let.*, **66**(15), 1921–1923 (1995 April. 10).
[13] Chang, Y.-C. (1988). "Bond-orbital models for superlattices", *Phys. Rev.*, **B37**, 8215–8222.
[14] Ioffe, A. F. *Semiconductor Thermoelements and Thermoelectric Cooling* (Infosearch Ltd., 1957).
[15] Goldsmid, H. J. *Applications of Thermoelectricity* (London: Methuen & Co Ltd. New York: John Wiley & Sons Inc., 1960).
[16] Ashcroft, N. W. and Mermin, N. D. *Solid State Physics*, Chapter 13 (Harcourt Brace College Publishers, International Edition).
[17] Sai-Halasz, G. A., Chang, L. L., Walter, J. M., Chang, C. A. and Esaki, L. (1978). *Solid State. Commun.*, **27**, 935.

Authors' Biographies

Jie-Feng Lin received a B.S. degree in Electrophysics from the National Chiao Tung University, Taiwan in 1995, and an M.S. Degree in Physics from the National Tsing Hua University, Taiwan in 1997. His involves theoretical studies of semiconductor heterostructures.

David Z.-Y. Ting is an Associate Professor of Physics at the National Tsing Hua University in Hsinchu, Taiwan, ROC, and a Visiting Research Associate at the California Institute of Technology. His research activities include theoretical studies of electronic and optical properties of semiconductor alloys and heterostructures, quantum transport in nanostructures and tunnel devices, and optical simulations.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 507–513
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Multiband Quantum Transmitting Boundary Method for Non-orthogonal Basis

G.-C. LIANG[a], Y. A. LIN[a], D. Z.-Y. TING[a,*] and Y.-C. CHANG[b]

[a] *Department of Physics, National Tsing Hua University, Hsinchu, Taiwan 300, ROC;*
[b] *Department of Physics, University of Illinois at Urbana-Champaign 1110 W. Green St., Urbana, Illinois 61801*

We generalize the Multiband Quantum Transmitting Boundary Method (MQTBM) for computing transmission coefficients in heterostructure to tight-binding-like band structure models with non-orthogonal basis and multiple neighbor interactions. We implement this method based on the newly developed planar-basis pseudopotential method which uses the generalized planar Wannier function basis. We demonstrate the method by computing transmission coefficients for a GaAs/AlAs double barrier structure.

Keywords: Resonant tunneling, tight binding, non-orthogonal basis, transmission coefficient

1. INTRODUCTION

Since Chang, Esaki and Tsu first observed resonant tunneling effect in heterostructures in 1974 [1], many theoretical methods have been developed to model the physics of resonant tunneling phenomenon in semiconductor heterostructures. The most commonly used formalism for computing tunneling probabilities in these structures is the transfer-matrix method. Unfortunately, this method is numerically unstable, and especially so for multiband models. One solution which circumvents the instability problem is the Multiband Quantum Transmitting Boundary Method (MQTBM) [2]. The MQTBM is a multiband realization of the Quantum Transmitting Boundary Method originally developed by Lent and Kirkner [3] for treating quantum transport in nanostructures. Thus far, MQTBM has been implemented for several realistic, empirical, multiband models, including the tight-binding model, the effective bond orbital model and the $\mathbf{k} \cdot \mathbf{p}$ method [4].

Recently, Chang has developed the planar-basis pseudopotential band structure model for treating electronic properties of surfaces and heterostructures [5, 6]. For treating structures with layer-like local geometry, this model is considerably more efficient than plane-wave based pseudopotential

*Corresponding author. Tel.: +886.3.574.2518, Fax: +886.3.572.3052. E-mail: dzt@phys.nthu.edu.tw.

methods, and leads to better analysis of surface and interface properties. Furthermore, it can be constructed directly from first principles and thus offers the possibility for more physically accurate descriptions than could be obtained with empirical band structure models. The method has been successfully applied to the studies of electronic and optical properties of surface and interfaces. We implement MQTBM for this new band structure model so that it can be used to treat quantum transport problems as well. The present work differs from MQTBM implementations in the past in that the basis functions used are non-orthogonal, and thus introduces some new complications.

2. METHOD

A. Planar-Basis Pseudopotential Method and Planar Wannier Functions

We briefly introduce the planar-basis pseudopotential method and planar Wannier functions, and refer the readers to the work by Chang [5, 6] for more details. In the planar-basis pseudopotential method, planar-basis functions consisting of 2D plane waves in the x-y plane and 1D gaussian functions in the z direction are used in heterostructure electronic calculations [5, 6]. The method can be made formally equivalent to either empirical or self-consistent first principles pseudopotential methods. Using these basis functions, we can find Bloch states associated with band n and wave vector $\boldsymbol{k}$, $\psi_{n,k}(\boldsymbol{r})$ and use them in turn to construct the planar Wannier functions. Assuming the heterostructure growth direction is along the z axis, a planar Wannier function is a translationally invariant function in the x-y plane characterized by an in-plane wavevector $\boldsymbol{k}_\parallel$, and centered along the growth direction at R_z :

$$\langle \boldsymbol{r}|W_{n,k_\parallel}(R_z)\rangle = \int dk_z e^{-ik_z R_z}\psi_{n,k}(\boldsymbol{r}). \qquad (1)$$

For a fixed value of $\boldsymbol{k}_\parallel$, the set of Wannier functions with same band index n and centered at different R_z may be thought of as a 1-band tight-binding basis which can be used to calculated the bulk band structure for the nth band. However, the interaction among those functions will in principle extend to large distances, which makes computations inefficient. The Generalized Planar Wannier Functions (GPWF) can be used to solve this problem. The generalized planar Wannier functions are proper linear combinations of $W_{n,k\parallel}(R_z)$ for the interested bands such that they are more localized along the z direction. In this new sub-basis, the interaction parameters can be cut off at shorter distances, but at the cost of leaving the Hamiltonian in the GPWF sub-basis non-diagonal in band index. In addition, depending on the number of neighboring interactions kept, the GPWF sub-basis could be either orthogonal or non-orthogonal.

For the generalized planar Wannier functions, the elements of the Hamiltonian matrix are given by,

$$H_{m,m'}(\mathbf{k}) = \sum_{R_z} e^{ik_z R_z}\langle W'_{m,\mathbf{k}_\parallel}(0)|H|W'_{m',\mathbf{k}_\parallel}(R_z)\rangle \quad (2)$$

where $W'_{m',k_\parallel}$ is a generalized planar Wannier function. By storing the interaction terms $\langle W'_{m,k_\parallel}(0)|H|W'_{m',k_\parallel}(R_z)\rangle$ up to the desired number of neighbors (two for non-orthogonal, and four for orthogonal basis) for each fixed value of $\boldsymbol{k}_\parallel$, we can converted the problem into a 1D multiband tight-binding model.

Let d be monolayer separation along the growth direction. We can then write $R_z = \sigma$d, where σ is a integer layer label. We can denote the generalized Wannier planar functions with the following tight-binding-like simplified notation :

$$|\sigma, \mathbf{k}_\parallel\rangle \equiv |W'_{m,\mathbf{k}_\parallel}(R_z = \sigma d)\rangle \qquad (3)$$

where $m = 1, 2, \ldots, M$ labels the basis functions within a monolayer ($M = 16$ in our case). We then appeal to MQTBM for transmission coefficient calculations.

B. Transmission Coefficient Calculations in Non-Orthogonal Basis

The multiband quantum transmitting boundary method is a numerically stable and efficient formulation for calculating heterostructure transmission coefficients using realistic multiband tight-binding band structure model. As seen in the previous section, the Generalized Planar Wannier Functions (GPWF) may be viewed as a non-orthogonal tight-binding basis with second-neighbor interactions. Previously, MQTBM has been implemented for several nearest-neighbor tight-binding models with orthogonal bases. The generalization to orthogonal basis with multiple neighbor interactions is straight-forward. However, a somewhat different formulation is required for non-orthogonal bases. In this work, we will introduce a more general version of MQTBM capable of handling multiple-neighbor interactions in a non-orthogonal basis. We consider a heterostructure as consisting of an active region sandwiched between two semi-infinite flat-band electrode regions. The computational domain consists of a sequence of N monolayers parallel to heterointerfaces. It includes all of the active region, and extends out from both sides of the active region to include $2N_I$ layers of each of the flat-band electrodes; N_I is the number of neighboring interactions for the particular tight-binding model used. As will be been, these inclusions are sufficient to allow us to fully incorporate the scattering boundary conditions within the finite computational domain.

We label the layers in the computational domain by $\sigma = 1, 2, \ldots, N$. For a second-neighbor model, for example, $\sigma = 1, 2, 3, 4$ will be in the left electrode, $\sigma = N-3, N-2, N-1, N$ the right and the rest are the active layer. From this point on we will specifically examine the case of a second-neighbor model for simplicity. However, all derivations generalize trivially to models with larger interaction range.

A wave function in the tight-binding basis may be written as

$$|\Psi\rangle = \sum_{\sigma m} C_{\sigma m} |\sigma m, \boldsymbol{k}_\parallel\rangle, \tag{4}$$

where, again, $\boldsymbol{k}_\parallel$ is the planar orbital in-plane wave vector, σ the layer label and m the orbital label.

The Schrödinger equation $(H - E)|\Psi\rangle = 0$ in this basis may be written as

$$\begin{aligned}\bar{\boldsymbol{H}}_{\sigma,\sigma-2}\boldsymbol{C}_{\sigma-2} + \bar{\boldsymbol{H}}_{\sigma,\sigma-1}\boldsymbol{C}_{\sigma-1} + \bar{\boldsymbol{H}}_{\sigma,\sigma}\boldsymbol{C}_{\sigma} \\ + \bar{\boldsymbol{H}}_{\sigma,\sigma+1}\boldsymbol{C}_{\sigma+1} + \bar{\boldsymbol{H}}_{\sigma,\sigma+2}\boldsymbol{C}_{\sigma+2} = 0,\end{aligned} \tag{5}$$

where C_σ is a vector of length M,

$$\boldsymbol{C}_\sigma = \begin{pmatrix} C_{\sigma 1} \\ C_{\sigma 2} \\ \vdots \\ C_{\sigma M} \end{pmatrix} \tag{6}$$

and $\bar{\boldsymbol{H}}_{\sigma,\sigma'}$ are $M \times M$ matrices,

$$\bar{\boldsymbol{H}}_{\sigma,\sigma'} = \boldsymbol{H}_{\sigma,\sigma'} - E\boldsymbol{S}_{\sigma,\sigma'}. \tag{7}$$

The elements of the Hamiltonian matrix $\boldsymbol{H}$ and overlap matrix $\boldsymbol{S}$ are given, respectively, by :

$$(\boldsymbol{H}_{\sigma,\sigma'})_{m,m'} = \langle \sigma m, \boldsymbol{k}_\parallel | H | \sigma' m', \boldsymbol{k}_\parallel \rangle, \tag{8}$$

and

$$(\boldsymbol{S}_{\sigma,\sigma'})_{m,m'} = \langle \sigma m, \boldsymbol{k}_\parallel | \sigma' m', \boldsymbol{k}_\parallel \rangle. \tag{9}$$

For convenience, we do not explicitly show the dependence on $\boldsymbol{k}_\parallel$ and E in our notation.

The boundary conditions in scattering problems are usually described in terms of Bloch states, i.e., as amplitudes of incoming and outgoing plane waves. To translate them into the tight-binding basis, we make use of the transfer matrix and Bloch's theorem. Together, they provide a connection between the plane-wave and the local-orbital descriptions through the concept of complex band structures [7].

We can write Eq. 5 in the transfer-matrix form,

$$\begin{pmatrix} -\bar{H}^{-1}_{\sigma,\sigma-2}\bar{H}_{\sigma,\sigma-1} & -\bar{H}^{-1}_{\sigma,\sigma-2}\bar{H}_{\sigma,\sigma} & -\bar{H}^{-1}_{\sigma,\sigma-2}\bar{H}_{\sigma,\sigma+1} & -\bar{H}^{-1}_{\sigma,\sigma-2}\bar{H}_{\sigma,\sigma+2} \\ I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \end{pmatrix} \begin{pmatrix} C_{\sigma-1} \\ C_{\sigma} \\ C_{\sigma+1} \\ C_{\sigma+2} \end{pmatrix} = \begin{pmatrix} C_{\sigma-2} \\ C_{\sigma-1} \\ C_{\sigma} \\ C_{\sigma+1} \end{pmatrix} \tag{10}$$

Now, consider, plane-wave state in either of the flat-band electrode regions,

$$|\psi_I(\boldsymbol{k}\rangle = |\psi_I(\boldsymbol{k}_{\|}, \boldsymbol{k}_{\perp})\rangle = \sum_{\sigma m} B_{\sigma m}|\sigma m, \boldsymbol{k}_{\|}\rangle. \tag{11}$$

Since $|\psi_I(\boldsymbol{k}\rangle$ is a Bloch state, by Bloch's theorem, the tight-binding coefficients must obey the relation

$$\boldsymbol{B}_{\sigma} = e^{ik_{\perp}d}\boldsymbol{B}_{\sigma-1}, \tag{12}$$

where d is the distance between monolayers, and $k_{\perp}$is the component of the crystal momentum along the growth direction. Then, together with Eq. (10), we can formulate the following eigenvalue problem:

$$\begin{pmatrix} -\bar{H}^{-1}_{\sigma,\sigma-2}\bar{H}_{\sigma,\sigma-1} & -\bar{H}^{-1}_{\sigma,\sigma-2}\bar{H}_{\sigma,\sigma} & -\bar{H}^{-1}_{\sigma,\sigma-2}\bar{H}_{\sigma,\sigma+1} & -\bar{H}^{-1}_{\sigma,\sigma-2}\bar{H}_{\sigma,\sigma+2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} B_{\sigma-1} \\ B_{\sigma} \\ B_{\sigma+1} \\ B_{\sigma+2} \end{pmatrix} = e^{ik_{\perp}d} \begin{pmatrix} B_{\sigma-1} \\ B_{\sigma} \\ B_{\sigma+1} \\ B_{\sigma+2} \end{pmatrix}. \tag{13}$$

Recall the implicit dependence of $\bar{H}_{\sigma,\sigma'}$ on E. Solving the above yields a set of eigenvalues $\{e^{ik_{\perp}d}\}$, from which a corresponding set of complex $\{k_{\perp}\}$ values are readily calculated. $E(k_{\perp})$ is the bulk complex band structure [7]. As an example, the complex band structure for bulk GaAs calculated using the planar-basis pseudopotential band structure model is shown in Figure 1.

Each of the $4M$ eigenvectors contain the coefficients of tight-binding orbitals on the one hand, and is associated with a particular complex wave vector $(\boldsymbol{k}\|, \boldsymbol{k}\perp)$ on the other. In this way, they provide a connection between the local tight-binding basis and Bloch (plane-wave) states [4]. We sort the eigenvectors such that the first $2M$ vectors represent state which propagate or decay to the right, and the remainder $2M$ to the left. The sorted eigenvectors are then grouped in a $4\times4M$ matrix $\boldsymbol{D}$.

Assume that we have a known coefficient set of incoming plane wave, $\tilde{I}$, and unknown coefficient sets of reflected and transmitted parts, $\tilde{r}$ and $\tilde{t}$, where $\tilde{I}$, $\tilde{r}$ and $\tilde{t}$ are the column vectors of the length $2M$. Then we can find the relations:

$$\begin{pmatrix} \tilde{C}_1 \\ \tilde{C}_2 \end{pmatrix} = \boldsymbol{D}^L \begin{pmatrix} \tilde{I} \\ \tilde{r} \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}^L_{11} & \boldsymbol{D}^L_{12} \\ \boldsymbol{D}^L_{21} & \boldsymbol{D}^L_{22} \end{pmatrix} \begin{pmatrix} \tilde{I} \\ \tilde{r} \end{pmatrix}, \tag{14}$$

and

$$\begin{pmatrix} \tilde{C}_{(N/2)-1} \\ \tilde{C}_{N/2} \end{pmatrix} = \boldsymbol{D}^R \begin{pmatrix} \tilde{t} \\ 0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}^R_{11} & \boldsymbol{D}^R_{12} \\ \boldsymbol{D}^R_{21} & \boldsymbol{D}^R_{22} \end{pmatrix} \begin{pmatrix} \tilde{t} \\ 0 \end{pmatrix}, \tag{15}$$

where $\boldsymbol{D}^L$ and $\boldsymbol{D}^R$ are $4M \times 4M$ matrices formed by the eigenvectors obtained by solving Eq. (13), for the bulk materials making up the left and right electrodes, respectively, and the $\tilde{C}_j$ is a vector of length $2M$ constructed by joining two consecutive $\boldsymbol{C}$ vectors : $\tilde{C}_1$ contain $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$, $\tilde{C}_2$ contain $\boldsymbol{C}_3$ and $\boldsymbol{C}_4$, $\tilde{C}_{(N/2)-1}$ contain $\boldsymbol{C}_{N-3}$ and $\boldsymbol{C}_{N-2}$ and $\tilde{C}_{N/2}$

contain C_{N-1} and C_N instead. Note that we have divided each D_{ij} matrix into $2M \times 2M$ sub-matrices for convenience.

Eliminating $\tilde{r}$ and $\tilde{t}$ from Eqs. (14) and (15), we obtain

$$\tilde{C}_1 - D_{12}^L D_{22}^{L\,-1} \tilde{C}_2 = D_{11}^L \tilde{I} - D_{11}^L D_{22}^{L\,-1} D_{21}^L \tilde{I}, \quad (16)$$

and

$$-D_{21}^R D_{11}^{R\,-1} \tilde{C}_{(N/2)-1} + \tilde{C}_{N/2} = 0 \quad (17)$$

Combining the above equations and Eq. (5), we obtain MN linear equations written in the matrix form as

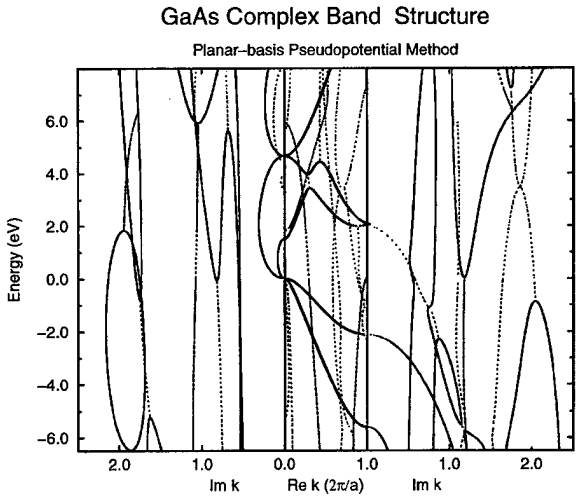


FIGURE 1 Complex bands of GaAs calculated using the planar-basis pseudopotential method.

$$\begin{pmatrix} I_M & 0 & [-D_{12}^L D_{22}^{L\,-1}] & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & I_M & & 0 & \cdots & \cdots & \cdots & 0 \\ \bar{H}_{31} & \bar{H}_{32} & \bar{H}_{33} & \bar{H}_{34} & \bar{H}_{35} & 0 & \cdots & 0 \\ 0 & \bar{H}_{42} & \bar{H}_{43} & \bar{H}_{44} & \bar{H}_{45} & \bar{H}_{46} & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & [-D_{21}^R D_{11}^{R\,-1}] & I_M & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & & 0 & I_M \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ \vdots \\ C_{N-1} \\ C_N \end{pmatrix}$$

$$= \begin{pmatrix} [D_{11}^L \tilde{I} - D_{12}^L D_{22}^{L\,-1} D_{21}^L \tilde{I}] \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad (18)$$

where I_M is an $M \times M$ identity matrix and the bracketed terms are $2M \times 2M$ sub-matrices. Solving the matrix equation yields a set of coefficient, $\{C_1, C_2, \ldots, C_N\}$. By Eq. (15), the coefficients of the transmitted plane-wave states can be computed by

$$\tilde{t} = D_{21}^{R\,-1} C_{N/2} \quad (19)$$

We can use $\tilde{t}$ to compute the transmission coefficient :where $v_I(E, k_\parallel; L)$ and $v_j(E, k_\parallel; R)$ are the velocities of the incident and the transmitted bulk plane-waves, respectively.

3. RESULTS AND DISCUSSION

We have successfully applied the multiband quantum transmitting boundary method to tight-binding band structure models with nonorthogonal basis and multiple-neighbor interactions.

Figure 2 shows transmission spectra of two different GaAs/AlAs double barrier structures calculated for using the Generalized Planar Wannier Functions (GPWF) band-structure model with second-neighbor interactions. Both Γ- point and the *X*-point resonant tunneling are seen in Figure 2. The Γ- and *X*-point resonances are associated with quasibound states localized in the GaAs and the AlAs layers respectively. The two structures used have the same GaAs layer widths (16 monolayers) but different AlAs layer widths (4 and 8 monolayers). Consequently, the positions of the Γ-point resonances remain the same for the two structures, while the *X*-point resonances showed a definite position shift.

Acknowledgment

This work was supported by the ROC National Science Council under Grant No. NSC 86-2112-M-007-001, and by the US Office of Naval Research under contract No. ONR-N00014-90-J-1267. The use of ROC National Center for High-Performance Computing facilities is acknowledged.

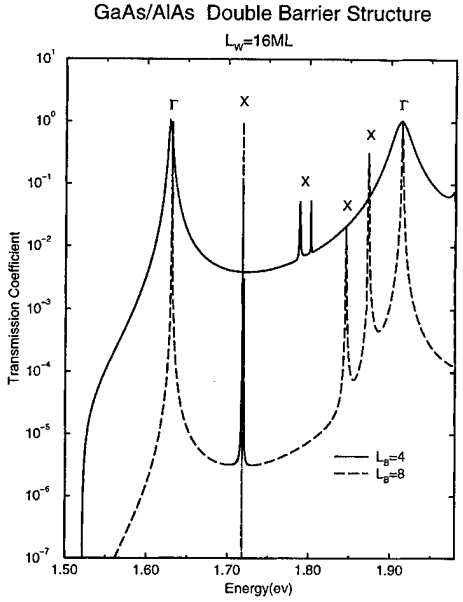


FIGURE 2 Transmission coefficients of the GaAs/AlAs double barrier structure.

References

[1] Chang, L. L., Esaki, L. and Tsu, R. (1974). "Resonant tunneling in semiconductor double barrier", *Appl. Phys. Letts.*, **24**(12), 593.
[2] Ting, D. Z.-Y., Yu, E. T. and McGill, T. C. (1992). "Multiband treatment of quantum transport in interband tunnel devices", *Phys. Rev. B.*, **45**(7), 3583.
[3] Lent, C. S. and Kirkner., D. J. (1990). "The quantum transmitting boundary method", *J. Appl. Phys.*, **67**, 6353.
[4] Ting, D. Z.-Y. and McGill, T. C. "Multiband and Multidimensional Analysis of Quantum Transport in Ultrasubmicron Device", published in *Quantum Transport Ultrasmall in Device*, edit by D. K. Ferry, H. L. Grubin, C. Jacoboni and A.-P. Jauho, NATO ASI Series, Series B : Physic; **342**, (Plenum Press, New York, 1995), pp. 417–436.
[5] Chang, Y.-C. (1995). "Device modeling from first principles at the atomic level", *Proceedings of the 2nd international symposium on compound semiconductors*, Cheju Island, Korea.
[6] Chang, Y.-C. and Li, G. (1996). "Planar-basis pseudopotential method and planar Wannier functions for surfaces and heterostructures", *Comp. Phys. Commun.*, **95**(2–3), 158.
[7] Chang, Y.-C. and Schulman, J. N. (1983). "Complex band structures of crystalline solids", *Phys. Rev. B.*, **27**, 2346.

Authors' Biographies

Geng-Chiau Liang received a B.S. degree (1995) and an M.S. degree (1997) in Physics from the National Tsing Hua University, Taiwan. His research involves theoretical studies of transport in semiconductor heterostructures.

Yiping A. Lin received a B.S. degree (1995) and an M.S. degree (1997) in Physics from the National Tsing Hua University, Taiwan. Her research interests include modeling of scanning probe microscopy and semiconductor physics.

David Z.-Y. Ting is an Associate Professor of Physics at the the National Tsing Hua University in Hsinchu, Taiwan, ROC, and a Visiting Research Associate at the California Institute of Technology. His research activities include theoretical studies of electronic and optical properties of semiconductor alloys and heterostructures, quantum transport in nanostructures and tunnel devices, and optical simulations.

Yia-Chung Chang is a Professor of Physics at the University of Illinois at Urbana-Champaign. His research interests include shallow impurities and excitons bound to shallow impurities in semicon-

ductors, shallow impurities in semiconductor quantum wells, many body effects in semiconductors with high carrier densities, exciton dynamics, electronic and optical properties of superlattices, magnetotransport and exchange interactions in magnetic multilayers, optical properties of disordered semiconductors, electronic properties of semiconductor surfaces, nonlinear optical properties of semiconductors, impact ionization in semiconductors, and physics of heterostructure resonant tunneling devices, nanostructure lasers, and infrared detectors.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 515–520
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Calibration of a One Dimensional Hydrodynamic Simulator with Monte Carlo Data

O. MUSCATO[a,*], S. RINAUDO[b] and P. FALSAPERLA[a]

[a] *Dipartimento di Matematica, Viale Andrea Doria, 95125 Catania (Italy);*
[b] *SGS-THOMSON Microelectronics, Stradale Primo Sole 50, 95121 Catania (Italy)*

In this paper we use the code **Exemplar** for matching a hydrodynamic 1D, time-dependent simulator and the transport coefficients obtained by the Monte Carlo simulator **Damocles**. This code is based on the Least Square method and it does not require any a priori knowledge about the simulator (analytical form of the equations etc.). The stationary electron flow in a one dimensional n^+-n-n^+ submicron silicon diode is simulated.

Keywords: TCAD, VLSI, BTE, transport theory, fluid mechanics, electronic devices

1. INTRODUCTION

Electronic transport in semiconductors can be described by hydrodynamic models (hereafter HM), obtained by taking the moments of the Boltzmann transport equation (hereafter BTE): the resulting mathematical model consists of an infinite hierarchy of Partial Differential Equations expressing balance laws for the particle number n, velocity $\vec{u}$, total energy E, deviatoric stress tensor $\hat{\theta}_{\langle ij\rangle}$, energy flux $\vec{S}$ (or heat flux $\vec{h}$) and so on, coupled with the Poisson equation.

Recently Anile and Muscato [1] presented an Extended Hydrodynamic model where the closure of the moment hierarchy is obtained by exploiting the entropy principle: this system, which is hyperbolic, consists of 13 scalar equations in the 13 unknowns $(n, \vec{u}, E, \hat{\theta}_{\langle ij\rangle}, \vec{h})$ and completely determines the description of the stress and of the heat flux, at variance with the other models. Such a model has been tested successfully with Monte Carlo (hereafter MC) simulations in Silicon [2]. The Left H and Side of the balance equations, called *production terms*, represent the average rate of change of carrier total energy Q_w, momentum $\vec{Q}_p$, energy-flux $\vec{Q}_s$ and stress $Q_{\langle ij\rangle}$ due to the scattering of carriers with the lattice. Usually they are approximated with *ad hoc* empirical formula [3] or as relaxation terms [4]: this last approximation leads to a serious inconsistency with one of the fundamental principles of Linear Irreversible Thermodynamics, the Onsager Reciprocity Princi-

*Corresponding author.

ple. In order to tackle this problem, we expanded the electron distribution function in Hermite polynomials around the state of global thermal equilibrium (limiting ourselves to the first two polynomials), where the electrons lie close to the bottom of the conduction band. By using a well known procedure due to Grad [5] we obtained the following equations:

$$Q_w = -\frac{E - E_0}{\tau_w} \tag{1}$$

$$\vec{Q}_p = \left(a_0^p + a_1^p \frac{T}{T_0} + a_2^p \frac{n}{n_0}\right) n\vec{u} + \left(b_0^p + b_1^p \frac{T}{T_0} + b_2^p \frac{n}{n_0}\right) \frac{n\vec{S}}{k_B T_0} \tag{2}$$

$$\vec{Q}_S = \left(a_0^S + a_1^S \frac{T}{T_0} + a_2^S \frac{n}{n_0}\right) k_B T_0 n\vec{u} + \left(b_0^S + b_1^S \frac{T}{T_0} + b_2^S \frac{n}{n_0}\right) n\vec{S} \tag{3}$$

where T_0 is the room temperature, E_0 the lattice total energy, n_0 a reference impurity concentration ($10^{18}\,\mathrm{cm}^{-3}$). Since in this approximation the distribution function is quasi-isotropic [2] the deviatoric stress tensor $\hat{\theta}_{ij}$ is negligible and by this method we cannot extract the corresponding production term (unless we consider higher order Hermite polynomials).

For the sake of simplicity we modeled the stress production as a relaxation term:

$$Q_{\langle ij \rangle} = -\frac{\hat{\theta}_{\langle ij \rangle}}{\tau_\sigma} \tag{4}$$

We should emphasize that:

- the coefficients appearing in Eqs. (1–4) are not fitting parameters but rather will be extracted by MC data obtained by the **Damocles** code [6], in the case of parabolic spherical band approximation, in order to be consistent with the Anile and Muscato model, obtained under these restrictions;
- these coefficients are not functions of the positions in the device as some authors claim.

With such coefficients we obtained a closed set of hydro equations which has been solved by an adequate numerical scheme. Since the Monte Carlo method gives a stochastic solution of the BTE, the results are noisy: how does the hydro solution change for small variations of the parameters τ_w, τ_σ, a_0, a_1, a_2, b_0, b_1, b_2? In order to answer this question the simulator **Exemplar**, developed by one of us [7], has been used.

The plan of the paper is the following: in section 2 we discuss the optimization problem; in section 3 we simulate the usual $n^+ - n - n^+$ diode with the **Damocles** code and with our Extended hydrodynamic model. We discuss the range of validity of the optimization procedure for various biases and conclusion are drawn.

2. THE OPTIMIZATION PROBLEM

The optimization problem is well known: defining the residual function

$$R(\vec{a}) = \sum_{i=1}^{N} w_i (f(x_i, \vec{a}) - y_i)^2 \tag{5}$$

where $\vec{a}$ is the vector of the parameters, $f(x_i, \vec{a})$ is the analytical description of the event computed in x_i (modeled by the vector parameter $\vec{a}$), y_i the real event value obtained in x_i and w_i an appropriate weight factor. In a mathematical language the problem of the least squares method could be expressed in the form:

$$\min\{R(\vec{a}) / \vec{a} \in R^n\} \tag{6}$$

where n is the dimension of the vector $\vec{a}$ and $\vec{a}$ the solution of the system, generally, non-linear

$$\nabla R(\vec{a}) = \vec{0} \tag{7}$$

and R^n the n dimensional space of the real R.

If we know the analytical equation of $f(x, \vec{a})$, we could evaluate and resolve the system $\nabla R(\vec{a}) = \vec{0}$,

using one of many methods proposed (e.g. Marquardt). However, if we do not know the analytical form of $f(x, \vec{a})$, it is not possible to find the solution of the problem (6) finding the solution of the system (7). This happens when we want to determine a set of parameters to model a general device with any kind of simulator: we run the simulator obtaining the macroscopic quantities but not the derivatives of the Residual function. In this case we can resort numerical methods as the simplex method and Powell's method, which minimize the Residual function, without having to know its derivatives.

The **Exemplar** code runs the simulator to calculate the function $f(x_i, \vec{a})$ $(i=1,\ldots,N)$ then compute and optimize the Residual function with the Simplex or Powell method (according the user choice) and the procedure restarts until a tolerance of the Residual function is obtained.

The proposed algorithm can be used with any simulators which can read their input from files and store their output in column-formatted ASCII files. It was implemented in a software program with a friendly user interface based on X-Window, which resolves the problem of the optimization well, but, since during the optimization it runs the simulator, it needs a large computational time.

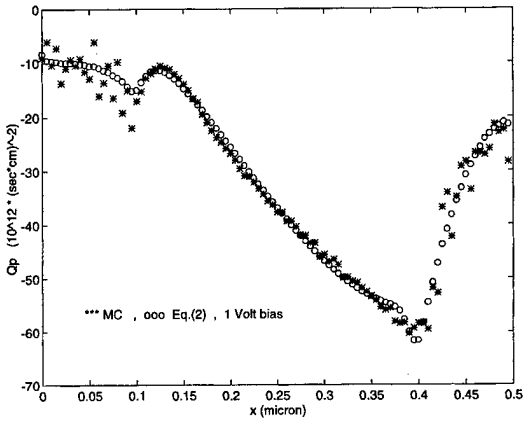


FIGURE 1 Average rate of change of carrier momentum Q_p vs. distance: Monte Carlo data (with ***) and the fitting formula Eq. (2) (with ooo).

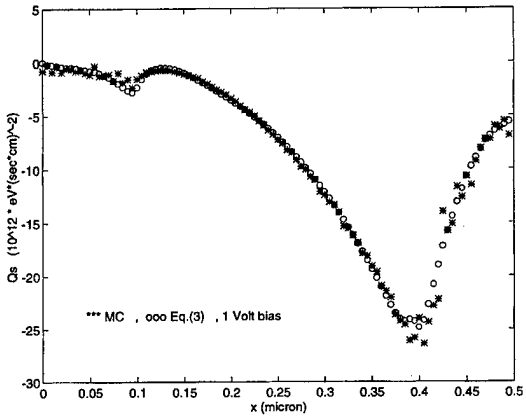


FIGURE 2 Average rate of change of carrier energy flux Q_S vs. distance: Monte Carlo data (with ***) and the fitting formula Eq. (3) (with ooo).

3. SIMULATION OF A ONE DIMENSIONAL n^+-n-n^+ SILICON DIODE

The n^+-n-n^+ diode consists of two n^+ regions 0.1 μm−long doped to a density of $N = 10^{18}$ cm^{-3}, while the central n region is 0.3 μm wide, with a doping density of $N = 10^{16}$ cm^{-3}. We consider Ohmic boundary conditions, $T_0 = 300$ K, and 1 V of applied bias; simulations refer to the stationary regime.

In Figures 1 and 2 we plot the MC data for Q_p and Q_S and the fitting with eqs. (2) and (3) (with ooo): we see that our functional form fits well the data. We observe that the MC data are noisy especially near the contacts (this run needed 1 month of CPU in a IBM Risc 6000 590) and consequently the fitting coefficients are not unique.

By using these coefficients the Extended Hydrodynamic model is solved by using a simulator based on the splitting method between relaxation and convection (Tadmor scheme) [8]. Then we run the **Exemplar** code in order to find the best coefficients which fit well the MC data. In Figures 3, 4 and 5 we compare the MC data (with ***), the hydro data without optimization (with ooo) , the hydro data optimized with **Exemplar** (with xxx).

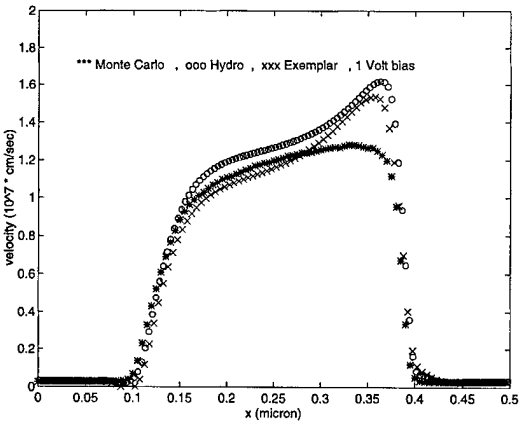


FIGURE 3 Average drift velocity vs. distance computed by using the hydrodynamic simulator without optimization (ooo), the hydro data optimized with the **Exemplar** code (xxx) and Monte Carlo data (***), with 1 Volt bias.

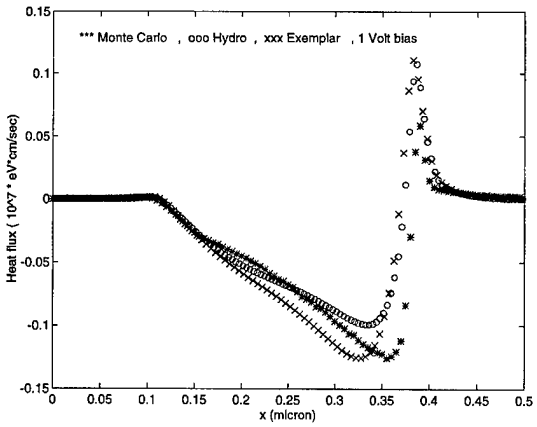


FIGURE 5 Heat-flux vs. distance computed by using the hydrodynamic simulator without optimization (ooo), the hydro data optimized with the **Exemplar** code (xxx) and Monte Carlo data (***), with 1 Volt bias.

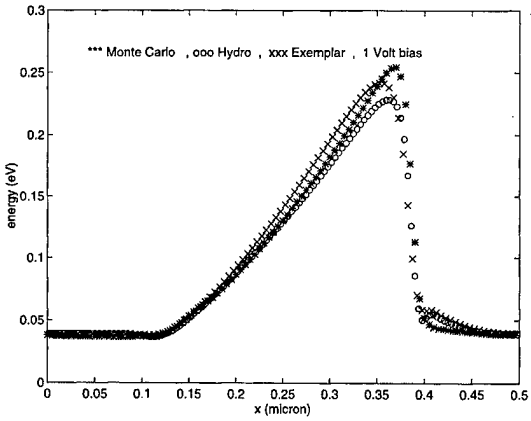


FIGURE 4 Average electron energy vs. distance computed by using the hydrodynamic simulator without optimization (ooo), the hydro data optimized with the **Exemplar** code (xxx) and Monte Carlo data (***), with 1 Volt bias.

The optimized coefficients (shown in Tab. I) differ from the previous ones at maximum by 40%. We notice that the well known 'spurious' peak in the velocity curve Figure 3 is lowered and the energy and heat flux curves Figures 4 and 5 are closer to the MC data. However this 'spurious' peak cannot be eliminated by any small change in the parameters and therefore its persistence calls for a more radical examination of the basic assumptions of the model.

A crucial question to be addressed is what is the range of validity of the optimization procedure? In order to answer this question we simulate the device with the coefficients of Table I (obtained with 1 Volt) for various biases. The result is that in the range $0.8 \div 1.2$ Volt the hydro simulations fit well the respective MC data (see Figs. 6, 7 and 8 in case of 1.2 V): for higher biases the hydro simulations diverge considerably with respect to the MC data.

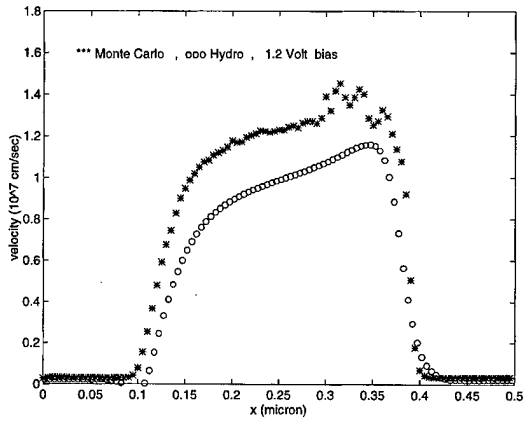


FIGURE 6 Average drift velocity vs. distance computed by using the hydrodynamic simulator with transport coefficients of Table I (obtained with 1 Volt) (ooo) and Monte Carlo data (***), with 1.2 Volt bias.

TABLE I Transport coefficients (in $psec^{-1}$), channel 0.3 μm, 1 Volt bias

	a_0	a_1	a_2	b_0	b_1	b_2
Q_p	−0.1376	−2.3042	1.3978	−1.6078	0.1930	−0.00769
Q_s	5.7677	27.7312	24.1639	−21.7514	−2.6697	−3.5269
τ_w			0.4094 (in psec.)			
τ_σ			0.0712 (in psec.)			

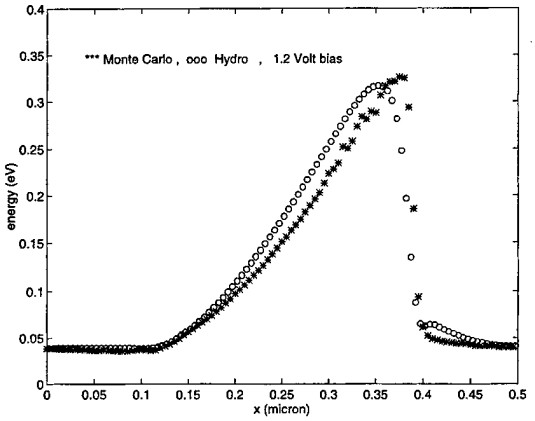


FIGURE 7 Average electron energy vs. distance computed by using the hydrodynamic simulator with transport coefficients of Table I (obtained with 1 Volt) (ooo) and Monte Carlo data (***), with 1.2 Volt bias.

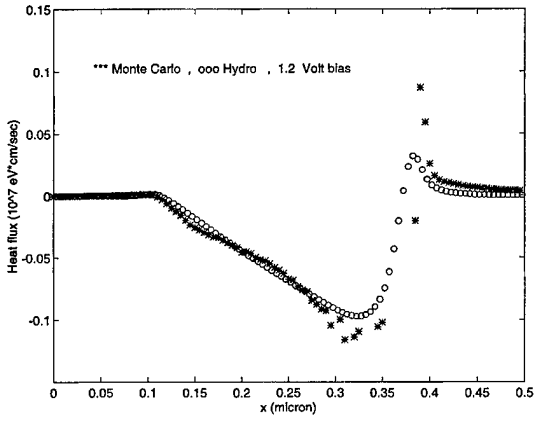


FIGURE 8 Heat flux vs. distance computed by using the hydrodynamic simulator with transport coefficients of Table I (obtained with 1 Volt) (ooo) and Monte Carlo data (***), with 1.2 Volt bias.

Finally we say that the modeling of the production terms and the closure of the hydro model are crucial points. They should be obtained by using the physics (e.g., an entropy principle, expansions of the distribution function). The transport coefficients can be obtained either by experiments or by MC simulations, and their validity is restricted to a neighborhood.

We are trying to improve our model and to simulate 2D devices: work along this line is in progress and will presented elsewhere.

Acknowledgements

This work has been supported by the C.N.R. Progetto Speciale *Modelli Matematici per i semiconduttori* 1996, the MURST project 40% and 60% 1996.

References

[1] Anile, A. M. and Muscato, O. (1995). *Phys. Rev. B.*, **51**, 16728.
[2] Anile, A. M. and Muscato, O. (1996). *Cont. Mech Therm.*, **1**, 1.
[3] Tang, T. W., Ramaswamy, S. and Nam, J. (1993). *IEEE Trans. Elec. Dev.*, **40**, 1469.
[4] Baccarani, G. and Wordemann, M. E. (1982). *Solid State Elec.*, **29**, 970.
[5] Grad, H. (1958). In *Thermodynamics of gases* edited by S. Flugee, Handerbruch der Physik **12**, Springer-Verlag, Berlin.
[6] Laux, S. E., Fischetti, M. V. and Frank, D. J. (1990). *IBM J. Res. Develop.*, **34**.
[7] Rinaudo, S. (1995). *A General algorithm to calibrate any kind of simultators without any knowledge about analytical form of the implemented models*, SGS-THOMSON Microelectronics, internal report.
[8] Falsaperla, P. and Trovato, M. (1997). Preprint University of Catania.

Authors' Biographies

Orazio Muscato is Assistant Professor of Theoretical Mechanics at Catania University. His research interests include mathematical models for semiconductors, Monte Carlo simulations, wave propagation.

Salvatore Rinaudo is research staff member at ST. His research interests include process and devices numerical simulations and technology CAD.

Paolo Falsaperla is Graduade student at Catania University. His research interests include numerical simulations and mathematical models for semiconductors.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 521–525
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Hyperbolic Hydrodynamical Model of Carrier Transport in Semiconductors

ANGELO MARCELLO ANILE[a], VITTORIO ROMANO[b,*] and GIOVANNI RUSSO[c]

[a] *Dipartimento di Matematica, Universitá di Catania, viale A. Doria, 6-95125 Catania, Italy;*
[b] *Politecnico di Bari, sede di Taranto, Viale del Turismo 8-74100 Taranto, Italy;*
[c] *Università dell'Aquila, Via Vetoio, loc. Coppito - 67100 L'Aquila*

A hydrodynamical model for semiconductors based on Extended Thermodynamics is presented and a suitable numerical scheme is proposed.

Keywords: TCAD, VLSI, BTE, transport theory, fluid mechanics, electronic devices

1 INTRODUCTION

Hydrodynamical models for carrier transport in semiconductors have been widely investigated by several authors [1, 2, 3, 4, 5, 6 and al.]. Most hydrodynamical models are based on phenomenological assumptions and not derived from a consistent physical model. Here we present a hydrodynamical model based on Extended Thermodynamics which is the appropriate thermodynamic theory of irreversible processes far from thermodynamic equilibrium. The model is obtained from the set of moment equations of the Boltzmann transport equation for electrons at the level of 13 moments. The closure relations are obtained in the framework of Extended Thermodynamics [7, 8] by employing the entropy principle. The balance equations for the macroscopic variables form a hyperbolic system for values of heat flux and stress tensor which are routinely encountered in the simulation of semiconductor devices of interest for industrial purposes. In the present paper we model the productions as relaxation terms, with relaxation times functions of the moments.

A suitable numerical scheme, developed in [10], has been successfully applied to the present model. It is based on the splitting method between relaxation and convection and allows an easy implementation in numerical codes. For the convective steps we used the Nessyahu-Tadmor [11] scheme for quasilinear hyperbolic systems in divergence form. It is a second order shock capturing method and does not require the analytical expressions of the eigenvalues and eigenvectors of the Jacobian matrix of fluxes.

*Corresponding author.

The latter property is crucial because explicit expressions for the eigenvalues and eigenvectors are not available for the balance equations of Extended Thermodynamics. For the relaxation part a suitable combinations of implicit Euler steps assures a global second order accuracy.

In particular we present the simulation of a $n^+ - n - n^+$ silicon diode.

The numerical results shows a uniform accuracy in all the fields (energy density, velocity, particle number, heat flux) within a reasonable degree of tolerance and could be employed for a systematic use as a CAD tool.

2 MATHEMATICAL MODEL

In the one dimensional case the evolution equations become (see [9] for the complete derivation)

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x}(nv) = 0, \tag{1}$$

$$\frac{\partial}{\partial t}(nv) + \frac{\partial}{\partial x}\left(nv^2 + \frac{p}{m^*} + \frac{\sigma}{m^*}\right) = -\frac{nv}{\tau_p} - \frac{neE}{m^*}, \tag{2}$$

$$\frac{\partial}{\partial t}\left(nv^2 + \frac{3p}{m^*}\right) + \frac{\partial}{\partial x}\left(nv^3 + \frac{5vp}{m^*} + \frac{2\sigma v}{m^*} + \frac{2q}{m^*}\right) = -2\frac{W - W_0}{m^*\tau_\omega} - \frac{2nevE}{m^*}, \tag{3}$$

$$\frac{\partial}{\partial t}\left(\frac{2}{3}nv^2 + \frac{\sigma}{m^*}\right) + \frac{\partial}{\partial x}\left(\frac{2}{3}nv^3 + \frac{4}{3}\frac{vp}{m^*} + \frac{7}{3}\frac{v\sigma}{m^*} + \frac{8}{15}\frac{q}{m^*}\right) = -\frac{1}{m^*\tau_s}\left(nv^2 + \frac{\sigma}{m^*}\right) - \frac{4nevE}{3m^*} \tag{4}$$

$$\frac{\partial}{\partial t}\left(nv^3 + \frac{5vp}{m^*} + \frac{2\sigma v}{m^*} + \frac{2q}{m^*}\right) + \frac{\partial}{\partial x}\left[nv^4 + 5\frac{p^2}{n(m^*)^2} + 7\frac{\sigma p}{n(m^*)^2} + \frac{32}{5}\frac{qv}{m^*} + v^2\left(8\frac{p}{m^*} + 5\frac{\sigma}{m^*}\right) + \frac{148}{25}\frac{q^2}{m^* p}\right]$$

$$= -\frac{1}{\tau_q}\left(nv^3 + \frac{5vp}{m^*} + \frac{2\sigma v}{m^*} + \frac{2q}{m^*}\right) - \frac{eE}{m^*}\left(3nv^2 + \frac{5p}{m^*} + \frac{2\sigma}{m^*}\right). \tag{5}$$

In the previous equations, n is the electron number density, v the mean velocity, $p = nk_B T$ the hydrostatic pressure, q the x-component of the heat flux, σ the xx-component of the viscosity tensor, m^* the effective electron mass in the parabolic band approximation and E the x-component of the electric field

$$E = \frac{\partial \phi}{\partial x},$$

$$\varepsilon \frac{\partial^2 \phi}{\partial x^2} = -e(N_D - N_A - n),$$

where ϕ is the electric potential, N_D and N_A are respectively the donor and acceptor density, e the elemetary charge and ε the dielectric constant.

W is the electron energy density $W = (1/2)nv^2 + (3/2)nk_B T$ and $W_0 = (3/2)nk_B T$. For the energy relaxation time, τ_w, and viscous relaxation timei, τ_σ, we take the constant values $\tau_w = 0.47$ ps and $\tau_\sigma = 0.02$ ps. The relaxation times of momentum and energy flux have been obtained by the homogeneous Monte Carlo [12] data by using the following fitting formulas (see [9] for the value of the constants c_i and $\tilde{c}_i$)

$$\tau_p = c_1 + c_2\frac{S}{J} + c_3\frac{S^2}{J^2}, \tag{6}$$

$$\tau_q = \tilde{c}_1 + \tilde{c}_2\frac{S}{J} + \tilde{c}_3\frac{S^2}{J^2}. \tag{7}$$

System (1–5) has the structure of a hyperbolic system of balance laws. Hyperbolicity, however, is restricted to a suitable neighborhood of thermodynamical equilibrium (see [9]).

Numerical solutions of the system are obtained by a splitting scheme. Let us consider a system of the form

$$\frac{\partial v}{\partial t} + \frac{\partial F(v)}{\partial x} = G(v), \tag{8}$$

with $v \in \mathbf{R}^m$ and $F: \mathbf{R}^m \to \boldsymbol{R}^m$.

The splitting scheme is based on the solution of the two steps:

Convection step

$$\frac{\partial v}{\partial t}+\frac{\partial F}{\partial v}=0$$

Relaxation step

$$\frac{\partial v}{\partial t}=G(v)$$

Each convection step is a predictor–corrector scheme of the form

$$v_{j+1/2}^{n+1/2}=\frac{1}{2}(v_j{}^{n+1}+v_{j+1}{}^{n})+\frac{1}{8}(v'_j-v'_{j+1}) - \lambda\left[F(v_{j+1}^{n+1/2})-F(v_j^{n+1/2})\right], \tag{9}$$

$$v_j^{n+1/2}=v_j^n-\frac{\lambda}{2}F'_j. \tag{10}$$

where $\lambda = dt/dx$. The time step Δt must satisfy a stability condition

$$\lambda\cdot\max\rho(A(v(x,t)))<\frac{1}{2}. \tag{11}$$

where $\rho\,(A(v(x,t)))$ is the spectral radius of the Jacobian matrix,

$$A=\frac{\partial F}{\partial v}.$$

This condition will ensure that the generalized Riemann problems with piecewise smooth data at time t_n will not interfere during the time step Δt.

The values of $v'_j/\Delta x$ and $F'_j/\Delta x$ are a first order approximation of the space derivatives of the field and of the flux, computed from cell averages by using Uniform Non Oscillatory reconstruction (UNO),

$$v'_j=\mathrm{MM}\left(d_{j-\frac{1}{2}}v+\frac{1}{2}\mathrm{MM}(D_{j-1},D_j),\ d_{j+\frac{1}{2}}v-\frac{1}{2}\mathrm{MM}(D_j,D_{j+1})\right) \tag{12}$$

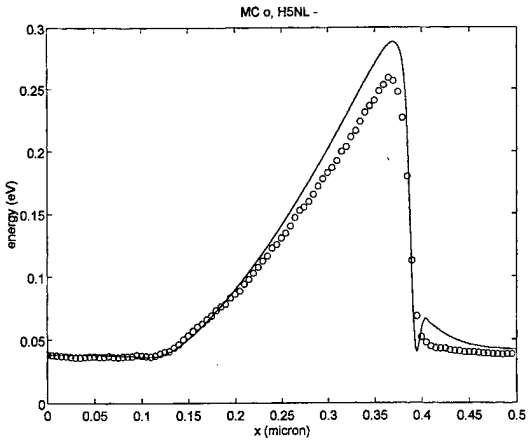


FIGURE 1 Average electron energy vs. distance obtained with the hydrodynamical model (continuous line) compared to Monte Carlo results (circles).

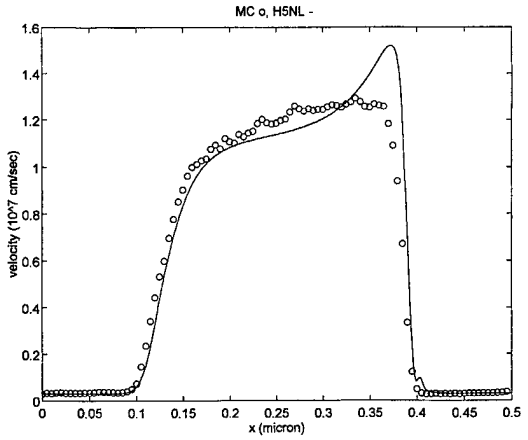


FIGURE 2 As in Figure 1 for the average drift velocity vs. distance.

where

$$D_j=v_{j+1}-2v_j+v_{j-1},$$
$$d_{j+1/2}=v_{j+1}-v_j$$

and

$$\mathrm{MM}(x,y)=\begin{cases}\mathrm{sgn}(x)\cdot\min(|x|,|y|) & \text{if } \mathrm{sgn}(x)=\mathrm{sgn}(y)\\ 0 & \text{otherwise.}\end{cases}$$

A similar procedure is used for computing F'_j.

For the relaxation step an unconditionally stable second order scheme can be obtained by analytical integration of the linearized relaxation equation, where linearization is obtained by *freezing* the coefficients at time t_n. The electric potential is computed by a standard procedure. In order to obtain second order accuracy in time we combinine the two steps according to the following scheme

$$\begin{aligned} U_1 &= U^n - R(U_1, E^n, \Delta t), \\ U_2 &= \frac{3}{2}U^n - \frac{1}{2}U_1, \\ U_3 &= U_2 - R(U_3, E^n, \Delta t), \\ U_5 &= \mathcal{C}_{\Delta t} U_3, \\ E^{n+1} &= \mathcal{P}(U_4), \\ U^{n+1} &= U_4 - R(U^{n+1}, E^{n+1}, \Delta t/2) \end{aligned}$$

where R represents the discrete operator correspoding to the relaxation step, $\mathcal{C}_{\Delta t}$ is the discrete operator corresponding to NT scheme, and $\mathcal{P}(U)$gives the solution to Poisson's equation.

3 NUMERICAL RESULTS

As test problem we consider a ballistic diode $n^+ - n\ n^+$, which models a MOSFET channel. The diode is made of silicon, and the bulk temperature is supposed to be 300 K. The n^+regions are 0.1 μm long and the channel is 0.3 μm long. The doping density is $N_D^+ = 0.5 \times 10^{17}\,\mathrm{cm}^{-3}$ in the n^+regions and $N_D = 0.2 \times 10^{16}\,\mathrm{cm}^{-3}$ in the channel.

For the electron effective mass in the approximation of parabolic band we use $m^* = 0.32\ m_e$ when m_e is the electron mass. The silicon dielectric constant is given by $\varepsilon = \varepsilon_r \varepsilon_0$, where $\varepsilon_r = 11.7$ is the relative dielecric constant and $\varepsilon_0 = 8.85 \times 10^{-18}$ C/Vμm is the dielectric constant of vacuum.

The initial electron temperature is the lattice temperature $T_0 = 300$ K and the charges are at rest. A bias voltage of 1 Volt is applied, and this determines a charge flux in the semiconductor.

The initial conditions for the system are;

$$n(x,0) = n_0(x), \quad T(x,0) = 300\,\mathrm{K}, \quad v(x,0) = 0, \quad q(x,0) = 0, \quad \sigma(x,0) = 0.$$

The doping profile is regularized according to the function

$$n_0(x) = n_0 - d_0\left(\tanh\frac{x - x_1}{s} - \tanh\frac{x - x_2}{s}\right),$$

where $n_0 = n_0(0)$, $d_0 = n_0\ (1 - ND/\ N_D^+)/2$, $x_1 = 0.1\mu\mathrm{m}$, and $x_2 = 0.4\mu\mathrm{m}$.

For the boundary conditions see [9]. The stationary solution is reached within a few picosends. The numerical results obtained for the stationary case have been compared with the Monte Carlo results, obtained by the DAMOCLES code [13].

Comparison with MC simulations shows a good overall agreement. In particular the agreement is very good everywhere in the device but in the second junction, where the discrepancy with MC simulations can reach up to 10%.

References

[1] Haensh, W. (1991). The drift-diffusion equation and its application in MOSFET modeling, Springer-Verlag, Wien.

[2] Markowich, A., Ringhofer, C. A. and Schmeiser, C. (1990). Semiconductor Equations, Springer-Verlag, Wien.

[3] Blotekjaer, K. (1970). *IEEE Trans. on Electron Devices*, **ED-17**, 38.

[4] Baccarani, G. and Wordeman, M. R. (1982). *Solid-state Electronics*, **29**, 970.

[5] Stettler, M. A., Alam, M. A. and Lundstrom, M. S. (1993). *IEEE Trans. on Electron devices*, **40**(4), 733.

[6] Lee, S.-C. and Tang, T.-W. (1992). *Solid-State Electronics*, **35**(4), 561.

[7] Müller, I. and Ruggeri, T. (1993). Extended Thermodynamics, Springer-Verlag, Berlin.

[8] Jou, D., Casas-Vazquez e, J. and Lebon, G. (1993). Extended Irreversible Thermodynamics, Springer-Verlag, Berlin.

[9] Anile, A., Romano, V. and Russo, G. Extended Hydrodynamical Models of Carrier Transport in Semiconductors, preprint.

[10] Romano, V. and Russo, G. (1996). Numerical solution for hydrodynamical models of semiconductors, submitted to *IEEE trans. on CAD*.

[11] Nessyahu e, H. and Tadmor, E. (1990). *J. Comput. Phys.*, **87**(2), 408.

[12] Anile, A. M. and Muscato, O. Improved hydrodynamical model for carrier transport in semiconductors, *Phys. Rev. B.*, **51**, 16728.
[13] Fischetti, M. V. and Laux, S. (1993). *Phys. Rev. B.*, **48**, 2244.

Authors' Biographies

Angelo Marcello Anile is full professor of Mathematical Physics at Catania University (Italy). His research interests include mathematical models of semiconductors, nonlinear wave propagation, fluid dynamics, plasma physics, general relativity.

Vittorio Romano is assistent professor at Polytechnic of Bari (Italy). His research interests include mathematical models for semiconductors, asymptotic wave, relativistic fluids, radiation hydrodynamics.

Giovanni Russo is associate professor at the University of L'Aquila (Italy). His research interests include numerical analysis, mathematical models for semiconductors, wave propagation.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 527–532
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

A Hydrodynamic Model for Transport in Semiconductors without Free Parameters

P. FALSAPERLA and M. TROVATO*

Dipartimento di Fisica, Università di Catania, Corso Italia 57, I-95129, Catania, Italy;
Dipartimento di Matematica, Università di Catania, Viale A.Doria 6, I-95125, Catania, Italy

We derive, using the Entropy Maximum Principle, an expression for the distribution function of carriers as a function of a set of macroscopic quantities (density, velocity, energy, deviatoric stress, energy flux). Given the distribution function, we obtain, for these macroscopic quantities, a hydrodynamic model in which all the constitutive functions (fluxes and collisional productions) are explicitely computed starting from their kinetic expressions. We have applied our model to the simulation of some one-dimensional submicron devices in a temperature range of 77–300 K, obtaining results comparable with Monte Carlo. Computation times are of order of few seconds for a picosecond of simulation.

Keywords: Entropy maximum principle, electron transport, submicron devices

1. INTRODUCTION

Modeling of modern semiconductor devices is currently performed by means of two distinct approaches, kinetic models and Fluid Dynamic (FD) models. The most accurate kinetic description is given by Monte Carlo methods, which can take into account explicitely both the band structure and the various scattering phenomena [1, 2]. Other kinetic approaches are based on the choice of particular forms of the non-equilibrium distribution function of carriers. Common examples are the simple shifted Maxwellian [3] or an expansion of the distribution in spherical harmonics [4]. The cylindrical symmetry constraint in momentum space and the reduced number of terms of the expansion that can be practically used do not permit, however, to describe transport properties of carriers in conditions very far from equilibrium [5]. The FD models are obtained considering a set of moments of the Boltzmann Transport Equation (BTE). These models need the knowledge of *constitutive functions* (fluxes and collisional productions) present in the hierarchy of equations, that are usually fixed on a phenomenological basis, introducing parameters, such as e.g. relaxation times and transport coefficients, which have an unknown dependence on the geometry and working conditions of the simulated devices. The presence of these *free parameters* [6] has always been

*Corresponding author: e-mail:falsaperla@ct.infn.it-trovato@ct.infn.it.

a limit to a practical use of FD models, because, in general, they are determined in each case on the basis of MC simulations or experimental data.

We have developed a HydroDynamic (HD) model for the simulation of transport phenomena in semiconductors, based on the Entropy Maximum Principle (EMP). Following this principle, we find the distribution that makes best use of the 'information' deriving from the knowledge of a given set of moments. This distribution turns out to be a strongly non-linear function of the moments. Given the distribution, we determine the unknown constitutive functions appearing in the hierarchy of the equations that describes the time evolution of the moments. We point out that the computation of collisional productions is then based on the sole knowledge of the scattering kernels and the physical quantities they contain. Our HD model is then fully closed, and, contrarily to other HD models, does not contain any *free parameter*. On the other hand, the distribution function we obtain has no particular symmetry restrictions and is fully suitable for three-dimensional models. Its strong non-linearity is also capable of describing transport phenomena even in conditions far from thermodynamic equilibrium, as those present in submicron devices with very high electric fields and field gradients ($E \simeq 10^5$ V/cm, $E/(dE/dx) \simeq 100$Å).

2. PHYSICAL CHARACTERISTICS OF THE MODEL

We consider here a HD model for transport phenomena in silicon. Our main purpose in the development of this model, has been to test how accurately our distribution function describes strong non-equilibrium conditions. Therefore we have used a simplified band structure. As is well known, electrons contributing to transport are mainly those belonging to the six equivalent X valleys which, up to an energy of about 0.5 eV, can be considered approximately parabolic. Electrons can then be described by a density of states effective mass $m^* = 0.32\,m_e$. In the same energy range, the main scattering phenomena are due to electron-phonon interactions, and we will consider intervalley transitions caused both by f type and g type phonons; for intravalley transitions we will consider scattering with acoustic phonons, which will be regarded as approximately elastic. For the evaluation of the scattering terms we have used the parameters reported in [2]. We will show (by a comparison with MC simulations performed under the same physical approximations) that in this way it is possible to describe accurately some simple Si devices even in conditions very far from thermodynamic equilibrium. An extension of the model toward a non parabolic band structure (for Si or other semiconductors) and the inclusion of further scattering terms (ionized impurities, polar optical phonons, ...) do not present, however, conceptual difficulties, and they are at present in development.

3. MAXIMIZATION OF ENTROPY

Let f be the the distribution function of electrons in phase space. The entropy density is defined by

$$h = -C \int f \log(f)\, d\vec{k}, \tag{1}$$

where C is a constant. Given a set of kinetic quantities $\psi_A(\vec{k})$, we can calculate the corresponding moments F_A of the f by means of

$$F_A = \int \psi_A(\vec{k}) f\, d\vec{k}. \tag{2}$$

We search the distribution f that maximizes the entropy density (1), under the constraints given by the relations (2). For this problem we have considered the set of kinetic quantities [7]:

$$\psi_A(\vec{k}) = \left\{1, \frac{\hbar}{m^*}k_i, \frac{\hbar^2}{2m^*}k^2, \frac{\hbar^2}{m^*}k_{\langle i}k_{j\rangle}, \frac{\hbar^3}{2(m^*)^2}k^2 k_i\right\}$$

and the corresponding moments: density of carriers (n), flux of carriers (nv_i), total energy density (W), traceless part of momentum flux density ($\Sigma_{\langle ij\rangle}$), energy flux density ($S_i$), being

$$W = \tfrac{3}{2}p + \tfrac{1}{2}nm^*v^2,$$
$$\Sigma_{\langle ij\rangle} = \sigma_{\langle ij\rangle} + nm^*v_{\langle i}v_{j\rangle},$$
$$S_i = q_i + \sigma_{\langle ij\rangle}v_j + \tfrac{5}{2}pv_i + \tfrac{1}{2}nm^*v^2v_i$$

where v_i is the mean velocity, $p = nK_BT$ is the pressure, $\sigma_{\langle ij\rangle}$ is the non convective part of tensor $\Sigma_{\langle ij\rangle}$ and q_i is the heat flux. Application of the EMP requires then the maximization of the functional

$$h' = h - \sum_{A=1}^{13}\lambda_A\left[\int \psi_A(\vec{k})f\,d\vec{k} - F_A\right],$$

where λ_A's are Lagrange multipliers. The resulting distribution will have the functional form:

$$f = \exp(-\Pi), \qquad \Pi = \sum_A \psi_A(\vec{k})\lambda_A. \qquad (3)$$

From (2) we have that $F_A = F_A(\lambda_B)$, and so, to determine f, we must invert (2) obtaining $\lambda_A = \lambda_A(F_B)$. This inversion is extremely difficult and can be obtained only by numerical integration or by a series expansion of f [8]. We have followed the latter approach, expanding f to third order in $\tilde{F}_A = \{v, \sigma, q\}$ around an equilibrium configuration defined by a local Maxwellian f_M. Introducing this expression into (2) all the quadratures can be done analitically, and the resulting relations can be inverted. In this way we express f as

$$f(\vec{r}, t, \vec{k}) \equiv f(\tilde{F}_A(\vec{r}, t), \vec{k}),$$

which is a strongly non-linear function of $\tilde{F}_A$.

4. THE HYDRODYNAMIC MODEL

The model resulting from the procedure described in the previous section is fully three-dimensional, but we show here one-dimensional results. The only independent variables are in this case $F_A = \{n, nv_x = nv, W, \Sigma_{\langle xx\rangle} = \Sigma, S_x = S\}$, satisfying the following balance equations

$$\frac{\partial n}{\partial t} + \frac{\partial nv}{\partial x} = 0,$$

$$\frac{\partial nv}{\partial t} + \frac{1}{m^*}\frac{\partial(\Sigma + (2/3)W)}{\partial x} = -\frac{neE}{m^*} + P_{nv},$$

$$\frac{\partial W}{\partial t} + \frac{\partial S}{\partial x} = -nveE + P_w,$$

$$\frac{\partial \Sigma}{\partial t} + \frac{\partial\left(\frac{2}{3}nm^*v^3 + \frac{4}{3}pv + \frac{7}{3}\sigma v + \frac{8}{15}q + G_\Sigma\right)}{\partial x} = -\tfrac{4}{3}nveE + P_\Sigma,$$

$$\frac{\partial S}{\partial t} + \frac{\partial\left(\frac{1}{2}nm^*v^4 + \frac{16}{5}qv + 4pv^2 + \frac{5}{2}\sigma v^2 + G_s\right)}{\partial x} = -eE\left(\tfrac{3}{2}nv^2 + \tfrac{5}{2}\frac{p}{m^*} + \frac{\sigma}{m^*}\right) + P_s,$$

where $E_x = E$ is the electric field, $G_A = \{G_\Sigma, G_S\}$ are terms of the fluxes dependendig on higher order moments and $P_A = \{P_{nv}, P_W, P_\Sigma, P_S\}$ are the collisional productions. Such unknown constitutive functions have been determined by integration of their kinetic expression using the distribution function given by the EMP. The fluxes and the collisional productions are very complex functions of $\tilde{F}_A$ which we do not show here.

5. DEVICE SIMULATIONS

As test case, we have considered four unipolar and one-dimensional devices n^+nn^+ (labelled A, B, C, D) at different temperatures, doping and applied biases, as summarized in Table I. All the results are compared with a MC model using the same physical approximations.

In Figure (1) we report the results of a series of simulations for device A. We see that for devices with such high fields and field gradients a strongly

TABLE I Devices parameters

Dev.	Temp. (K)	Bias (V)	$N^+(cm^{-3})$	$N\ (cm^{-3})$	Chan. (μm)
A	300	2	10^{19}	10^{17}	0.25
B	150	1	10^{18}	10^{16}	0.25
C	77	1	10^{17}	10^{15}	0.45
D	300	1	10^{18}	10^{16}	0.3

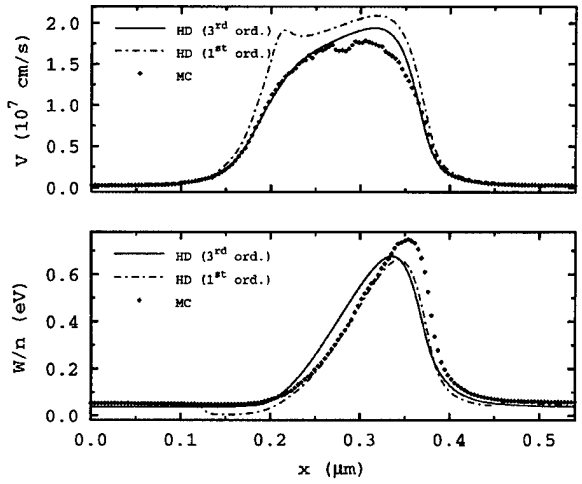


FIGURE 1 Velocity and total energy as a function of position for device A. We show a comparison with MC data and results for the HD model with production terms computed at different orders of the expansion. Differences in the values of energy in the highly doped regions are due to the correlation energy of electrons, included in MC model.

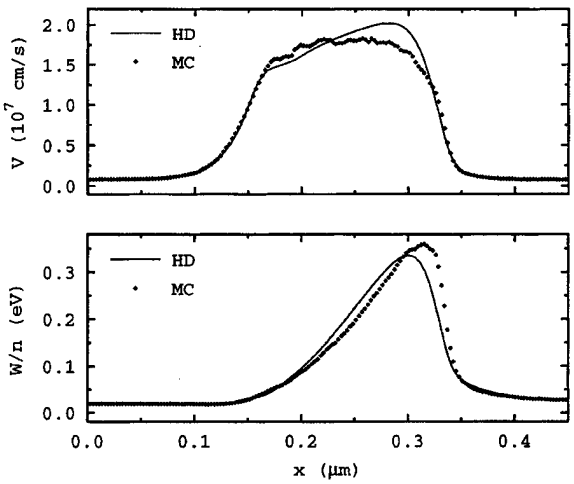


FIGURE 2 Velocity and total energy for device B. Results from a MC simulation and the HD model.

non linear description is necessary to obtain a good agreement with MC simulations. In Figures (2), (3) we report two examples of computations at lower temperatures (with production term evaluated at third order). Note that MC data for the 77 K simulation show evident oscillations on both velocity and energy. This phenomenon is probably due to single-phonon interactions, and is partially reproduced by our model. Note also the ballist pick, characteristic of low-temperature devices. Figure (4) shows velocity, energy, traceless stress and heat flux for device D. Current densities for devices A, B, C and electric fields for devices A, B, D are shown in Figures (5), (6). Computation times for the present results have been in a range of 1÷10 sec for a ps of simulation.

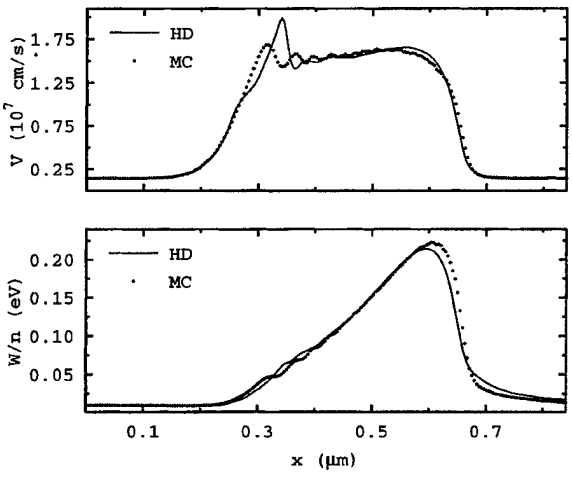


FIGURE 3 Velocity and total energy for device C. Results from a MC simulation and the HD model.

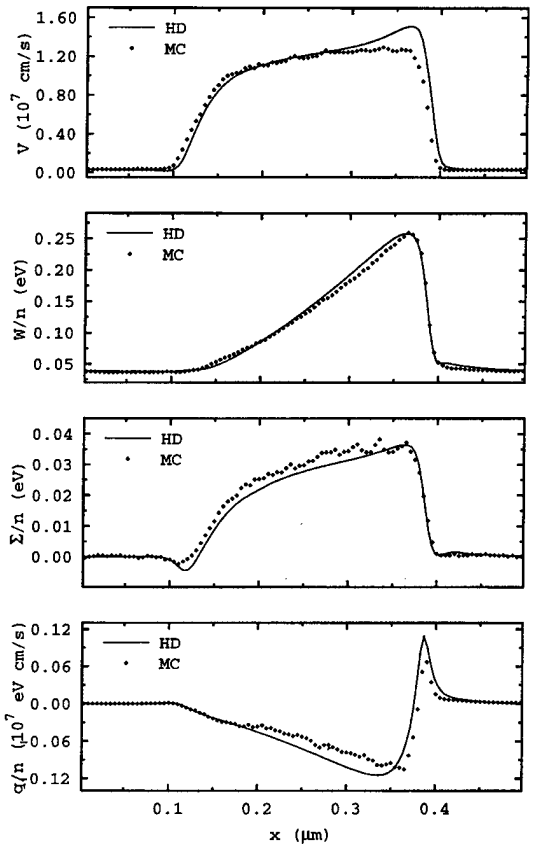


FIGURE 4 Velocity, total energy, traceless stress and heat flux for device D from a MC simulation and the HD model.

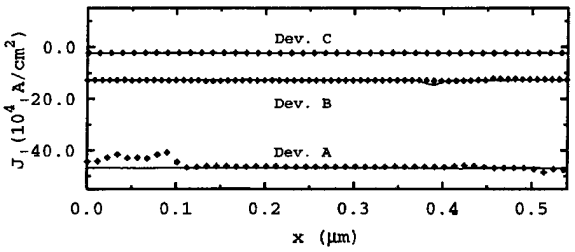


FIGURE 5 Current densities J for devices A, B and C, from MC (points) and HD (lines).

6. CONCLUSIONS

We have shown that the EMP allows to create a closed HD model to describe transport phenomena in Si in strong non-equilibrium conditions. We point out that an effective use of the EMP depends both on the choice of a set of 'constraints' (moments of the distribution function) and on the determination of a set of evolution equations for these 'constraints' that takes explicitely into account the underlying physical processes (the various scattering phenomena, in this case).

A better model of transport in semiconductors will require a more accurate description of band structure and the inclusion of further scattering terms. These extensions are now in development.

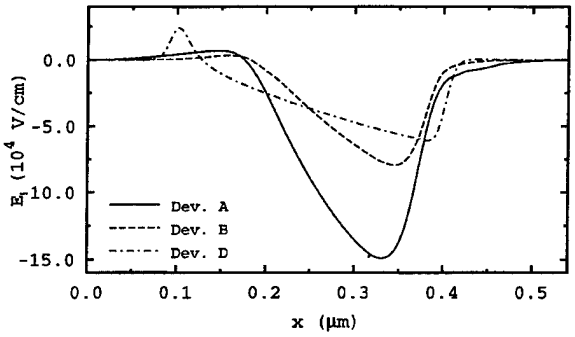


FIGURE 6 Electric fields for devices A, B and D, from the HD model.

Acknowledgments

The authors thank M.V. Fischetti for helpful discussions on transport physics and for providing most of MC data, and O.Muscato for providing MC data of device D.

References

[1] Fischetti, M. and Laux, S. (1993). *Phys. Rev. B.*, **48**, 2244.
[2] Jacoboni, C. and Reggiani, L. (1983). *Rev. Mod. Phys.*, **55**, 645.
[3] Stratton, R. (1962). *Phys. Rev.*, **126**, 2002.
[4] Gnudi, A., Ventura, D. and Baccarani, G. (1993). *IEEE Trans. on Comp.-Aided Design*, **12**, 1706.
[5] Laux, S. E. and Fischetti, M. V. *Proceedings Bipolar-Bicmos Tech. Meeting, Minneapolis*, Oct. 1–3, 1995.
[6] Tang, T., Ramaswamy, S. and Nam, J. (1993). *IEEE Trans. on Elec. Devices*, **40**, 1469.
[7] Here and in the following symbol $A_{\langle ij \rangle}$ denotes the traceless part of a symmetric tensor A, i.e. $A_{\langle ij \rangle} = A_{ij} - (1/3)\delta_{ij}\mathrm{Tr}(A)$.
[8] Anile, A. M. and Trovato, M. (1997). *Phys. Lett. A*, **230**, 387.

Authors' Biographies

Paolo Falsaperla graduated in Physics in 1992 at University of Catania, Italy. He has worked on mathematical methods for atomic structures in high magnetic fields. His current interests include Monte Carlo simulation of noise and hydrodynamic models for semiconductor devices.

Massimo Trovato graduated in Physics in 1993 at University of Catania, Italy. He is currently working toward the Ph.D. degree in Physics at University of Catania. He has worked on the application of the Entropy Maximum Principle to transport problems. His current research interest is in the modeling and simulation of hot-carrier effects in submicrometer devices.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 533–537
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Modeling of Poly-Silicon Carrier Transport with Explicit Treatment of Grains and Grain Boundaries

EDWIN C. KAN[a] and ROBERT W. DUTTON[b]

[a] *404 Phillips Hall, Cornell University, Ithaca, NY 14853;*
[b] *CIS-X 333, Stanford University, Stanford, CA 94305-4075*

Explicit treatment of grains and grain boundaries is necessary to model the carrier transport in poly-silicon devices whose feature size is comparable to the grain size. The grain boundaries were modeled by interface traps, and comparison was made between thermionic and diffusion transport across the grain boundaries. It was found that the numerical model for diffusion transport with total trap conservation in grain boundary areas is not physically convergent and shows a strong grid sensitivity. Effects of the critical doping level and the lattice temperature are demonstrated on poly-silicon resistors with 1-D bamboo-type and 2-D realistic microstructures.

Keywords: Microstructure, poly-crystalline silicon, carrier transport, device simulation, interface trap

Poly-silicon devices such as TFT for SRAM, LCD, and image sensor, and high-value resistance in analog circuit, have recently received much attention in consideration with its good integration capability with standard CMOS technology [1]. The carrier transport in poly-Si is conventionally modeled like a uniform Si material with distributed trap density in the band gap [2] related to the average grain size, since more than tens of grains usually exist within the device area of interest [1]. However, with the aggressive scaling of VLSI technology and recent use of laser annealing [3], TFT for SRAM and high resistance poly-Si resistors may have only a few grains within the device, and the device behavior will be heavily dependent on the process conditions owing to the microstructure effects [4]. Nonuniform and anisotropic phenomena from the limited number of grains within the active device area may be important and explicit treatment of carrier transport within grains and across grain boundaries will be necessary. Physical position and size of grains in 2-D can be obtained by AFM (atomic force microscopy) and EBIC (electron-beam induced current) [5] instrumentation.

If the grain size in poly-Si is still larger than 100 Å, deviation from single-crystal Si transport (based on the periodic potential approximation) within grains may be negligible. Bandtail transport [4] within grains can be used as the next level of approximation. The grain boundaries in silicon are usually an interface between different crystal orientations and are therefore atomically thin. It is impractical and unphysical to assign nontrivial regions on the order of 50 Å [3] to model the carrier transport across grain boundaries, since the drift-diffusion model is doubtful for that geometrical scale, and gridding for multi-dimension cases

requires a large number of grid points. Alternatively, the grain boundary can be modeled by interface traps (no physical size assigned to grain boundaries):

$$\nabla(\varepsilon\nabla\psi) = q\,(n - p - N_D^+ + N_A^- - N_{DD}^+ + N_{AA}^-) \tag{1}$$

$$\frac{\partial N_{DD}^+}{\partial t} = (C_{pn}p + e_{nn})(N_{DD} - N_{DD}^+) - (C_{nn}n + e_{pn})\,N_{DD}^+ \tag{2}$$

$$\frac{\partial N_{AA}^-}{\partial t} = (C_{np}n + e_{pp})(N_{AA} - N_{AA}^-) - (C_{pp}p + e_{np})\,N_{AA}^- \tag{3}$$

where ε is the dielectric constant, ψ is the potential, n and p are the electron and hole concentrations, N_D^+ and N_A^- are non-trapping ionized dopants, and N_{DD}^+ and N_{AA}^- are donor- and acceptor-state traps. The effect of traps are calculated in rate equations (2) and (3), and then fed back to the Poisson equation (1). The transport equations for n and p follow the conventional forms. The interface trap density is mapped into the discretization scheme of (1) by multiplying the half mesh length of the closest cell to convert into volume trap density. Carrier transport across the grain boundary can be modeled by either *thermionic emission* or *diffusion*. The former approach has been taken in [7], although a Monte Carlo simulations is used. The drift-diffusion model is much more computationally efficient and numerically stable, and thus we have chosen to implement the grain boundary treatment under the drift-diffusion model. The thermionic emission model assumes two distinct distribution function at either side of the barrier, and the carrier flux is estimated by the population which has energy over the barrier height. The barrier height is corrected to account for image force and field emission (tunneling), but no scattering event is considered at the barrier. The diffusion transport across the grain boundaries is similar to the case of delta-doped traps, i.e., transport is affected by the barrier height, but scattering mechanisms are homogeneously treated. Detailed model description can be found in classical textbook such as [8].

A poly-Si resistor with 1-D bamboo-like microstructure will be used as the first example for explicit treatment of grain boundaries. Three one-micron-long grains and two grain boundaries are used. The grain boundary is modeled by surface trap concentration N_S of $10^{12}\,\mathrm{cm}^{-2}$. The critical doping level N_T [1] will be $N_S/\lambda = 10^{16}\,\mathrm{cm}^{-3}$, where λ is the average grain size. When the grain is doped with less than N_T, the grain will be fully depleted and the conduction band will have no flat region. The barrier height will reach a maximum when the doping level is close to N_T. When the doping level exceeds N_T, most of the traps are filled, the barrier height starts decreasing, and charge neutral regions (flat conduction band) grow from the center of grain toward the grain boundary. This can be clearly observed in both Figures 1 and 2, where thermionic (1a) and diffusion (1b) transport have been demonstrated respectively. However, when diffusion transport is used and the total number of traps is conserved in the delta-doped representation, the barrier height is strongly dependent on grid spacing since the depletion width does not vary with N_D^{-1}, but with $N_D^{-0.5}$ for abrupt junctions. This indicates serious deficiency of using the combination of interface traps and diffusion model, since their physical concepts of interface condition conflict each other. On the other hand, when thermionic transport is used, the barrier height shows no grid dependence for 10 Å to 100 Å minimal grid spacing, (Fig. 2).

As shwon in Figure 3(a), the lattice temperature effect can also be well modeled by the thermionic transport as compared to the experimental results in [1]. When the barrier height is large, the resistivity is much more sensitive to the lattice temperature T_L, since the thermionic current is proportional to $T_L^2 \exp(-E_B/k\,T_L)$, which is the limiting factor in the resistor studied here. In Figures 3(b–d) where the minimal grid spacings are 100 Å, 50 Å and 10 Å, the diffusion transport

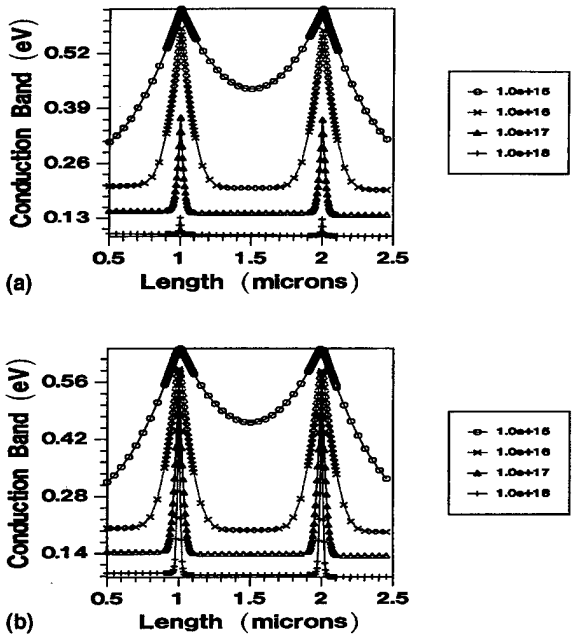


FIGURE 1 Conduction band for various doping levels. (a) by thermionic emission and (b) by diffusion. The critical doping level is around $10^{16}\,cm^{-3}$.

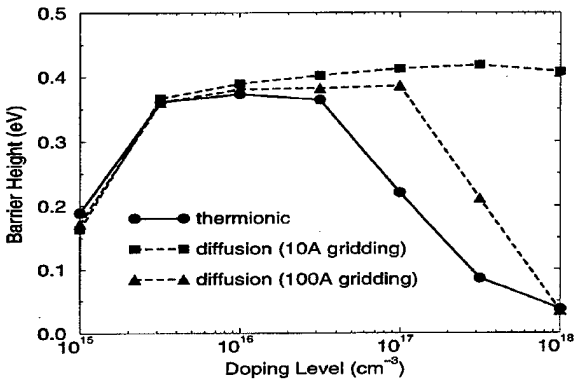


FIGURE 2 Grid dependence of barrier height when diffusion transport across grain boundaries is used. The thermionic model shows no grid dependence for minimal gridding from 100 Å to 10 Å.

shows not only strong grid dependence, but also inaccurate resistivity transition when the doping level is in the vicinity of N_T. It can be observed that the resistivity has less doping-level dependence owing to the estimation of barrier height, and the slope of resistivity with respect to $1/T_L$ does not show a gradual change for doping level across N_T.

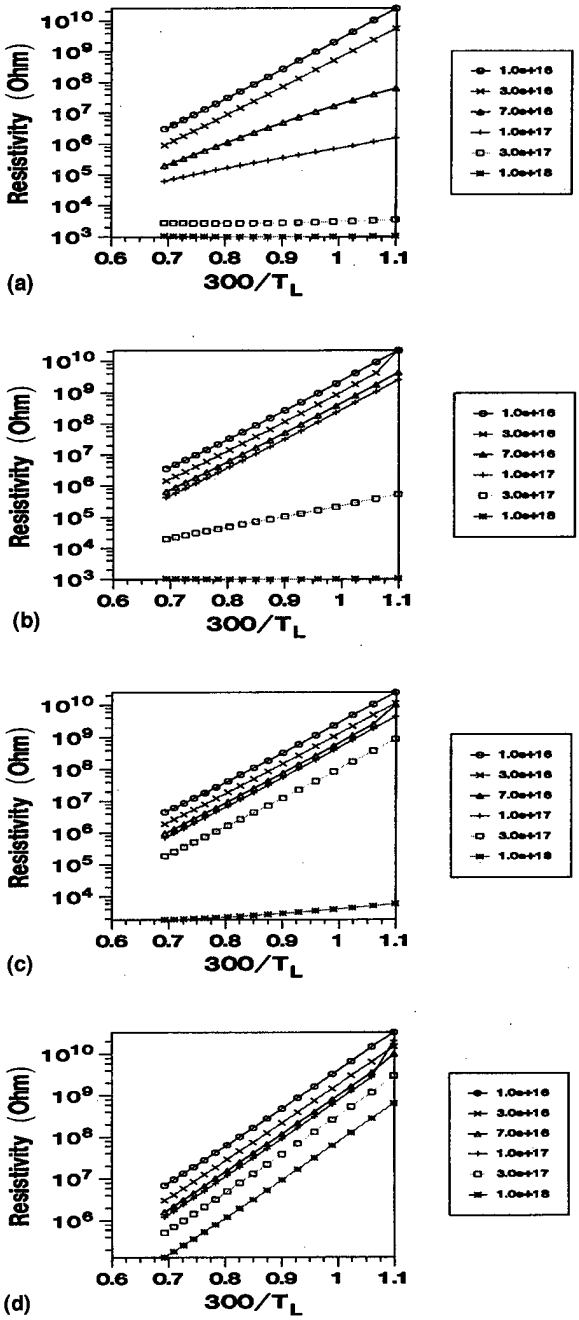


FIGURE 3 The lattice temperature effects on poly-Si resistivity. (a) uses thermionic emission, (b), (c) and (d) use diffusion transport with minimal grid spacing of 100 Å, 50 Å and 10 Å respectively.

A poly-Si resistor with microstructure similar to that obtained from EBIC measurement [5] is shown in Figure 4. Thermionic emission transport

has been used since grid spacing is much harder to control for arbitrary 2-D cases. It can be seen that the current flow pattern is strongly affected by the grain boundaries when the doping level is smaller than N_T, and less affected when the doping level is larger than N_T owing to the barrier height. Stronger depletion can be observed at the grain boundary corners. The lattice temperature effect also shows more complicated behavior in Figure 5, which apparently can not be predicted by simple analytical models [1]. This can be explained by the dependence of 2-D current flow pattern on lattice temperature and barrier height. Detailed analyses are limited by the length of this paper, and will be included in future publications.

We have reported results for carrier transport analyses with explicit treatment of grains and grain boundaries. As the number and orientation of grains become influential for device electrical behaviors, our approach shows good promises for predictive simulation.

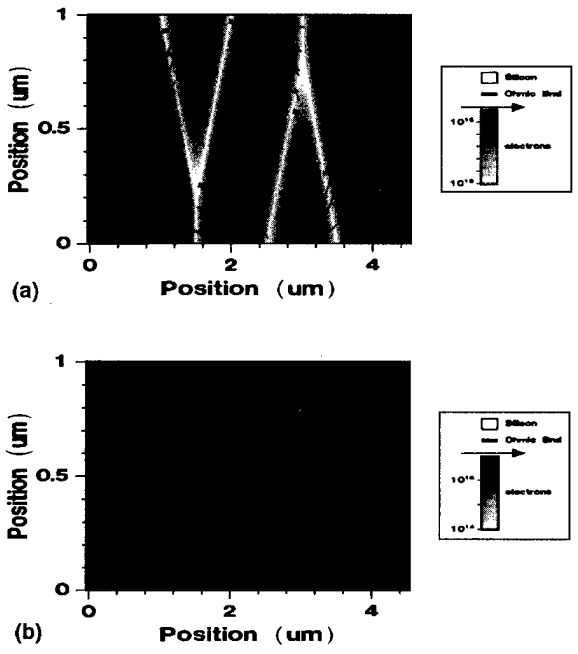


FIGURE 4 Electron concentration and current flow pattern of a 2-D poly-Si resistor. The grain boundaries can be clearly observed at the depletion region. (a) is doped with $10^{16}\,cm^{-3}$ and (b) with $10^{17}\,cm^{-3}$.

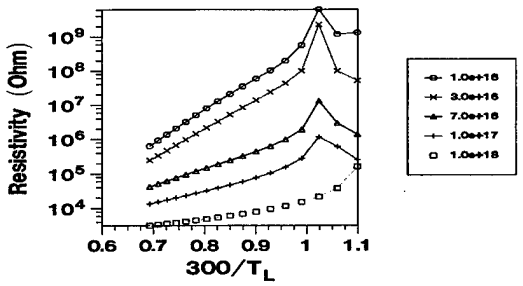


FIGURE 5 The lattice temperature effects on poly-Si resistivity with a 2-D microstructure.

Acknowledgements

This work was supported in part by DARPA under contracts DABT63-93-C-0053 and F30602-96-2-0308.

References

[1] Kamins, T. (1988). Polycrystalline Silicon for Integrated Circuit Applications, Kluwer Academic.
[2] Hack, M. and Shaw, J. G. "Numerical simulations of amorphous and polycrystalline silicon thin-fim transistor", *Proc. 22nd Int. Conf. Solid-State Devices and Materials* sendai, Japan, 1990, pp. 999–1003.
[3] Armstrong, G. A., Uppal, S., Brotherton, S. D. and Ayres, J. R. "Modeling of laser-annealed polysilicon TFT characteristics", *IEEE Electron Dev. Lett.*, **18**(7), 315–318, July 1997.
[4] Pennell, R. and Foerstner, J. "Resistivity control of boron-doped polysilicon resistors", *J. Electrochem. Soc.*, **138**(3), 860, March 1991.
[5] Wu, K., Dutton, R. W. and Johnson, N. M. "EBIC measurement and grain boundary recombination in SOI polycrystalline silicon", *IEEE Trans. Electron Devices*, ED-**33**(7), 1020, July 1986.
[6] Serikawa, T. *et al. Proc. of Intl. Conf. Solid State Dev. and Mat.*, Osaka, 1995, p. 662.
[7] Shimatani, T. and Koyanagi, M. (1995). "New Monte Carlo simulation for polycrystalline silicon thin-film transistor", *IEDM Tech. Dig.*, p. 297.
[8] Sze, S. M. Physics of Semiconductor Devices, 2nd ed., Wiley, 1981.

Author's Biographies

Edwin C. Kan received his B.S. degree from National Taiwan University in 1984 and M.S. and Ph.D. degrees from University of Illinois at Urbana-Champaign in 1988 and 1992, respectively, all in electrical engineering. From 1984 to 1986, he

served as a second lieutenant in Air Force, Taiwan, Republic of China. From January 1992, he had been with Dawn Technologies as the principal developer of advanced electronic and optical device simulators and technology CAD framework. He has been with Stanford University as a research associate since October 1993, leading projects such as TCAD 1-2-3D tool development, software architecture definition, model hierarchy and MEMS modeling. He joined Cornell University in July 1997 as an assistant professor of School of Electrical Engineering. His main research areas include VLSI technology, semiconductor device physics, composite CAD development, and numerical methods for PDE and ODE.

Robert W. Dutton is Professor of Electrical Engineering at Stanford University and Director of Research in the Center for Intergrated Systems. He received the B.S., M.S., and Ph.D. degrees from the University of California, Berkeley, in 1966, 1967 and 1970, respectively. He has held summer staff positions at Fairchild, Bell Telephone Laboratories, Hewlett-Packard, IBM Research, and Matsushita during 1967, 1973, 1975, 1977 and 1988 respectively. His research interests focus on Integrated Circuit process, device, and circuit technologies-especially the use of Computer-Aided Design (CAD) and parallel computational methods. Dr. Dutton has published more than 200 journal articles and graduated more than four dozen doctorate students. He was Editor of the IEEE CAD Journal (1984–1986), winner of the 1987 IEEE J. J. Ebers Award, 1988 Guggenheim Fellowship to study in Japan and was elected to the National Academy of Engineering in 1991. In December 1996, Dr. Dutton received the Jack A. Morton Award for his "seminal contributions to semiconductor process and device modeling".

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 539–544
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Formulation of the Boltzmann Equation as a Multi-Mode Drift-Diffusion Equation

K. BANOO, F. ASSAD and M. S. LUNDSTROM*

1285 School of Electrical Engineering, Purdue University, West Lafayette, IN 47907

(Received 2 June 1997; In final form 10 July 1997)

We present a multi-mode drift-diffusion equation as reformulation of the Boltzmann equation in the discrete momentum space. This is shown to be similar to the conventional drift-diffusion equation except that it is a more rigorous solution to the Boltzmann equation because the current and carrier densities are resolved into M×1 vectors, where M is the number of modes in the discrete momentum space. The mobility and diffusion coefficient become M×M matrices which connect the M momentum space modes. This approach is demonstrated by simulating electron transport in bulk silicon.

Keywords: Semi-classical Boltzmann equation, non-equilibrium transport, multi-mode drift-diffusion

1 INTRODUCTION

With the continued down-scaling of semiconductor devices, there is a need to develop device simulators that can treat carrier transport taking into account off-equilibrium carrier distributions by solving the Boltzmann Transport Equation as accurately as possible. Several techniques have been developed to do so – such as the Monte Carlo [1], hydrodynamic [2], spherical harmonic [3], and cellular automata methods [4], and the Scattering Matrix Approach [5, 6, 7]. Each method has its own limitations – for example, computational burden, calibration of parameters, low order approximation for the distribution function, "artificial diffusion" of carriers and restriction to fixed spatial square grids. Ideally, a simulation method should provide all the capabilities of drift-diffusion simulators (i.e., simulations from equilibrium to high bias with smooth results at low computational burden) while also resolving carrier distribution and treating scattering processes rigorously. Our objective in this paper is to take a step in this direction.

Therefore, here we will describe a re-formulation of the scattering matrix equations which expresses the 1-D spatial Boltzmann equation as a 1-D spatial drift-diffusion equation in a discretised 3-D momentum space. The current and carrier densities generalise to M×1 vectors, where

*Corresponding author. Tel.: (765) 494-3515. Fax: (765) 494-6441. E-mail: lundstro@ecn.purdue.edu.

M is the number of modes in the discrete 3-D momentum space. The mobility and diffusion coefficient become M×M matrices which connect the M momentum space modes. Solving the Boltzmann equation, then, reduces to solving a set of M coupled drift-diffusion equations which might be done by a generalisation of the standard techniques for solving the conventional drift-diffusion Equation [8].

2 FORMULATION OF THE MULTI-MODE DD EQUATION

The one-mode method of McKelvey [9, 10] and similar observations by Shockley [11] can be generalised to M-modes in 3-D momentum space and this gives us an expression for the differential flux equations in 1-D real space for the M-modes in momentum space as follows [7]

$$\begin{pmatrix} \frac{d}{dx}\overrightarrow{J^+}(x) \\ \frac{d}{dx}\overrightarrow{J^-}(x) \end{pmatrix} = \begin{bmatrix} -[\xi_{11}(x)] & [\xi_{12}(x)] \\ -[\xi_{21}(x)] & [\xi_{22}(x)] \end{bmatrix} \begin{pmatrix} \overrightarrow{J^+}(x) \\ \overrightarrow{J^-}(x) \end{pmatrix}, \tag{1}$$

where the M×1 vectors $\overrightarrow{J^+}(x)$ and $\overrightarrow{J^-}(x)$ are the fluxes discretised in positive and negative directions of momentum p_x respectively, at a position x.

The elements of the (M×M) differential matrices $[\xi_{ij}(x)]$ can be interpreted as the inverse of an *inter-mode* mean free path in presence of all possible scattering mechanisms. In general, these terms can be difficult to calculate analytically and therefore we use an indirect procedure. The space-independent differential $[\xi_{ij}]$ matrices are calculated from the matrix logarithm of the transmission matrix of a semiconductor slab divided by the slab thickness. The transmission matrix itself is obtained from the scattering matrix calculated by Monte Carlo techniques [7]. All possible information about the underlying physics of scattering (band structure, phonons, ionised impurities and electric field) that is included in the Monte Carlo simulation is automatically embedded in the scattering matrix and hence in the $[\xi_{ij}]$ matrices.

Now returning to Eq. (1), we find its symmetric and anti-symmetric components and relate each flux $J_i(x)$ to its velocity v_i and its population density $n_i(x)$. Thus, we obtain a multi-mode drift-diffusion equation and its associated continuity equation:

$$\vec{J}(x) = e[\mu(x)]\vec{n}(x)\mathcal{E}_x + e[D(x)]\frac{d}{dx}\vec{n}(x), \tag{2}$$

$$\frac{d}{dx}\vec{J}(x) = [\alpha(x)]\vec{J}(x) + [\beta(x)]\vec{n}(x), \tag{3}$$

where the diffusion, inverse Einstein and mobility matrices are defined as follows:

$$[D(x)] = 2\{[\xi_{11}(x)] + [\xi_{12}(x)] + [\xi_{21}(x)] + [\xi_{22}(x)]\}^{-1}[\mathcal{V}], \tag{4}$$

$$[E]^{-1} = \frac{1}{2\mathcal{E}_x}\{[\xi_{11}(x)] - [\xi_{12}(x)] + [\xi_{21}(x)] - [\xi_{22}(x)]\}, \tag{5}$$

$$[\mu(x)] = [D(x)][E]^{-1}, \tag{6}$$

and $[\mathcal{V}]$ is a diagonal matrix whose elements are the mode velocities.

The coefficient matrices for the continuity equation are

$$[\alpha(x)] = \frac{e}{2}\{[\xi_{11}(x)] + [\xi_{12}(x)] - [\xi_{21}(x)] - [\xi_{22}(x)]\}, \tag{7}$$

$$[\beta(x)] = \frac{e}{2}\{[\xi_{11}(x)] - [\xi_{12}(x)] - [\xi_{21}(x)] + [\xi_{22}(x)]\}[\mathcal{V}]. \tag{8}$$

Equations (2) and (3) are the key results of the paper. Note that they are very similar to the conventional drift-diffusion form. This multi-mode drift-diffusion equation has associated M×M mobility and diffusion matrices that depend only on the scattering mechanisms and momentum space discretisation.

It is important to note that the field dependence of the relevant matrices here. The discrete form of the Boltzmann Transport Equation [7] indicates

that each of the differential $[\xi_{ij}(x)]$ submatrices are *linearly* dependent on the field, for any given orientation of the field. This relation makes the $[\mu(x)]$, $[D(x)]$, $[\alpha(x)]$ and $[\beta(x)]$ matrices straightforward to calculate once the invariant zero-field $[\xi_{ij}(x)]$ matrices and the field-coefficient matrices are known.

3 RESULTS

In this section, we present some preliminary results for $\langle 111 \rangle$ electric fields in bulk Si. A simple spherical, non-parabolic energy band structure was assumed for calculating scattering rates and the 3-D momentum space was discretised into 400 modes – 20 modes in transverse $|p_t|$ and 20 modes in longitudinal p_x (this implies 20 $\overrightarrow{J^+}$ fluxes for $+p_x$ and 20 $\overrightarrow{J^-}$ fluxes for $-p_x$). Scattering matrices were computed by Monte Carlo simulation [5–7] and the $[\xi_{ij}(x)]$ matrices were extracted using the procedure described in Section 2. Higher levels of accuracy with respect to the underlying band-structure and scattering can be attained just by using more sophisticated Monte Carlo techniques. The distribution of the carriers was assumed to be *uniform* across the modes, which is adequate to demonstrate proof-of-concept here.

Having obtained the $[\xi_{ij}(x)]$ matrices, we then calculated the mobility and diffusion matrices and examined their structure. These are shown in Figure 1. Calculating these matrices involves taking matrix logarithms and inverses which produces a large number of elements that are numerically near zero. If we ignore these small elements, the matrices are very sparse ($\approx 1-2\%$ ignoring elements $<0.1\%$ of the largest element) and have a simple structure. The most significant elements of the mobility matrix are the diagonal elements and off-diagonal elements only where the ith mode is connected to the jth mode by field shift. The diffusion matrix is strongly *diagonal* and has significant but small off-diagonal elements only where the ith mode is connected to the jth mode by scattering.

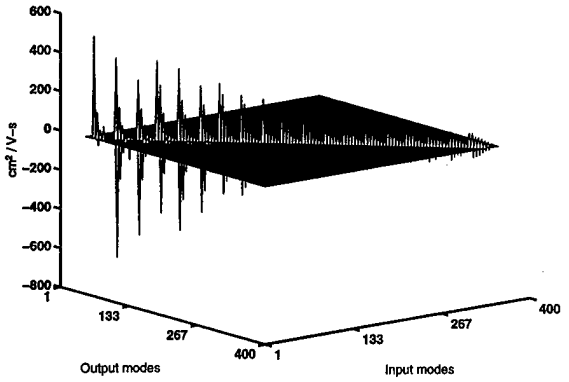


FIGURE 1a Mobility matrix for Si, non-parabolic spherical bands, 20 modes in longitudinal p_x (20 positive and 20 negative fluxes) and 20 modes in transverse $|p_t|$, $\langle 111 \rangle$ -1×10^4 V/cm.

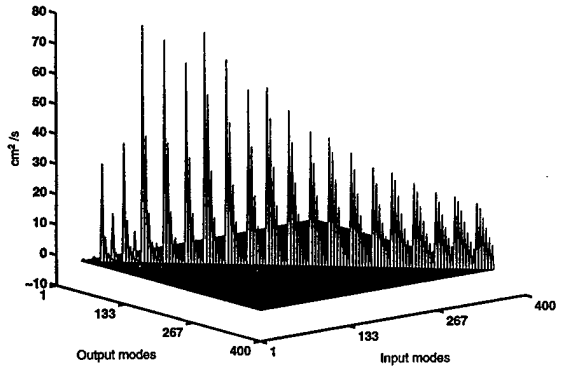


FIGURE 1b Diffusion matrix for Si, non-parabolic spherical bands, 20 modes in longitudinal p_x (20 positive and 20 negative fluxes) and 20 modes in transverse $|p_t|$, $\langle 111 \rangle$ -1×10^4 V/cm.

In order to test the formulation, we used the above matrices to simulate electron transport in bulk Si with $\langle 111 \rangle$ electric fields. The solution in bulk is simple because there are no spatial gradients in Eqs. (2) and (3). Substituting the expressions for $[\alpha]$, $[\beta]$ and $[\mu]$ from Eqs. (7), (8) and (6) respectively, we get

$$\begin{bmatrix} -[\xi_{11}] & [\xi_{12}] \\ -[\xi_{21}] & [\xi_{22}] \end{bmatrix} \begin{bmatrix} [\mathcal{V}] & [0] \\ [0] & [\mathcal{V}] \end{bmatrix} \begin{pmatrix} \overrightarrow{n^+} \\ \overrightarrow{n^-} \end{pmatrix} = \begin{pmatrix} \vec{0} \\ \vec{0} \end{pmatrix}, \quad (9)$$

which could also have been obtained directly from the differential flux Eq. (1).

The solution for the above Eq. (9) is a straightforward solution to a null-space problem (using svd factorisation in MATLAB) and it gives us the complete bulk carrier distribution function for any field $\mathcal{E}_x$. Some calculated bulk distributions for $\langle 111 \rangle$ electric fields are shown in Figure 2. Taking the average of the mobility and the diffusion matrices over the carrier distributions so obtained in the bulk,

$$\langle \mu \rangle = \frac{1}{\Sigma_i n_i} \sum_i \sum_j \mu_{ij} n_j, \tag{10}$$

$$\langle D \rangle = \frac{1}{\Sigma_i n_i} \sum_i \sum_j D_{ij} n_j, \tag{11}$$

we recover the values of macroscopic mobility and diffusion coefficients.

Plots of the results are shown in Figure (3), and we see that macroscopic mobility, diffusion coefficients and electron temperature so obtained have the expected behaviour with electric field. The low-field mobility is $\approx 30\%$ too low because of the assumed uniform intra-mode distribution. High-field results which do not suffer from this constraint are therefore more accurate – e.g. 9.915×10^6 cm/s at -1×10^5 V/cm. Note that the field dependence of $\langle \mu \rangle$ and $\langle D \rangle$ is a consequence of the field dependence of the distribution function in this approach.

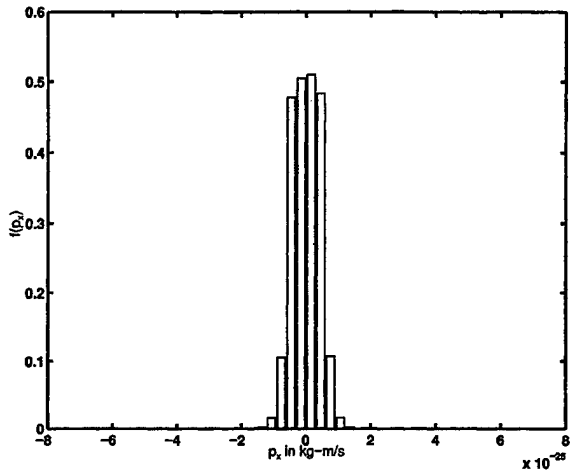


FIGURE 2a Bulk carrier distribution for $\langle 111 \rangle$ -1×10^2 V/cm in Si.

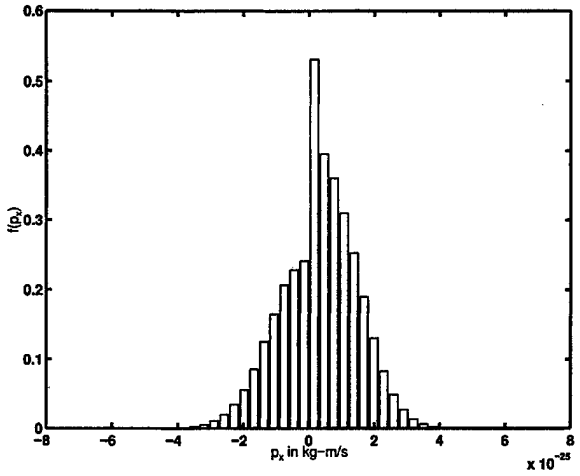


FIGURE 2b Bulk carrier distribution for $\langle 111 \rangle$ -1×10^5 V/cm in Si.

4 DISCUSSION

To illustrate how the equations would be solved under spatially varying conditions, we present a simple linear scheme using finite differences. By

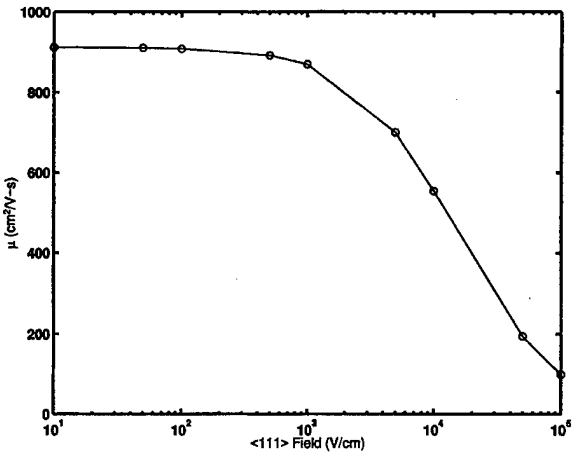


FIGURE 3a Macroscopic bulk mobility versus $\langle 111 \rangle$ electric field in Si.

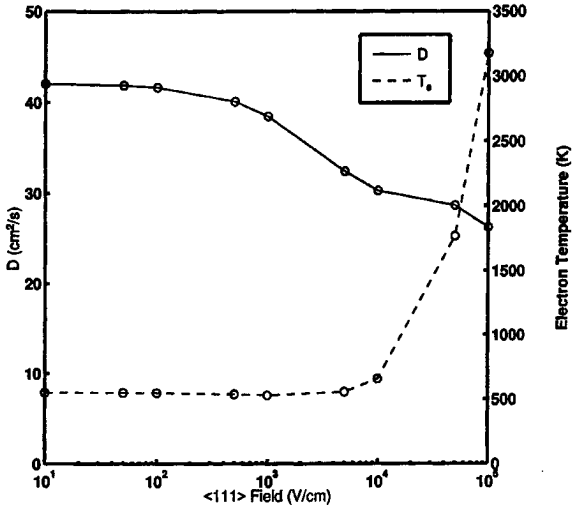


FIGURE 3b Macroscopic bulk diffusion coefficient and electron temperature (calculated from $D/\mu = k_B T_e/e$) versus ⟨111⟩ electric field in Si.

discretising $d\vec{J}(x)/dx$ on a uniform grid of size h, we obtain

$$\frac{1}{h}\left(\vec{J}\left(i+\frac{1}{2}\right)-\vec{J}\left(i-\frac{1}{2}\right)\right)=[\alpha(i)]\vec{J}(i)+[\beta(i)]\vec{n}(i), \tag{12}$$

In the simplest case, we could discretise the current equations using Eq. (2) and finite differences as follows

$$\vec{J}(i)=e[\mu(i)]\left(\frac{V(i+1)-V(i-1)}{2h}\right)\vec{n}(i) + e[D(i)]\frac{1}{2h}(\vec{n}(i+1)-\vec{n}(i-1)).$$

The result is a tridiagonal matrix form for the carrier distribution function across the device (i from 1 to N)

$$[\mathcal{L}(i-1)]\,\vec{n}(i-1)+[\mathcal{D}(i)]\,\vec{n}(i) + [\mathcal{U}(i+1)]\,\vec{n}(i+1)=\vec{0}, \tag{13}$$

where the elements of the tridiagonal matrix are now not scalars but M×M matrices and the variables are M×1 vectors. This is a large linear system whose solution is the position dependent carrier distribution function. However, the Scharfetter-Gummel method is normally the preferred scheme for discretising the conventional drift-diffusion equation because it is stable when the potential drop between adjacent nodes on a grid spacing h is greater than $2k_B T_e/e$. A corresponding result must be developed for the multi-mode semiconductor equations.

With regard to future applications, it is clear to see that finding the transitions rates for the fluxes in two (and three) spatial dimensions will give similar differential flux equations for two (and three) spatial dimensions. Therefore the multi-mode method will still hold in the most general case of transport. As a final note, we should point out that the multi-mode drift-diffusion/continuity equation formulation is formally equivalent to the differential flux equations, Eq. (1). These equations could, alternatively be integrated across the device in order to solve for carrier transport. The numerical advantages of one form over the other are not clear yet.

5 SUMMARY

To summarise, we have presented a 1-D spatial multi-mode drift-diffusion equation as reformulation of the 1-D spatial Boltzmann equation in a discrete 3-D momentum space. Although, the numerical aspects in this paper were not optimised for the best accuracy, the multi-mode drift-diffusion equation was solved in the bulk for the carrier distribution function and all the macroscopic properties that are incorporated as phenomenological models in conventional drift-diffusion were shown to fall out as a natural consequence of solving the multi-mode drift-diffusion.

The potential of this method lies in its close connection to conventional drift-diffusion – notably the equivalence of a one-mode case to conventional drift-diffusion, easy reduction to macroscopic quantities and similarity in solution

techniques. The key issue now is to formulate the problem for solution on a general 2-D spatial grid. Both deterministic and stochastic solution techniques for solving the resulting equations should be examined. From the results so far, the multi-mode drift-diffusion formulation of the Boltzmann equation promises to be a powerful approach and may overcome some of the limitations of the scattering matrix approach.

Acknowledgement

This work was supported by the Semiconductor Research Corporation.

References

[1] Fischetti, M. V. and Laux, S. E. (1988). "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects", *Phys. Rev. B*, **38**, 9721–9745.

[2] Meinerzhagen, B. and Engl, W. L. (1988). "The influence of the thermal equilibrium approximation on the accuracy of classical two-dimensional numerical modeling of silicon sub-micrometer MOS transistors", *IEEE Trans. Electron Devices*, **35**, 689–697.

[3] Liang, W., Goldsman, N., Mayergoyz, I. and Oldiges, P. J. (1997). "2-D MOSFET modeling including surface effects and impact ionization by self-consistent solution of the Boltzmann, Poisson and Hole-Continuity equations", *IEEE Trans. Electron Devices*, **44**, 257–267.

[4] Zandler, G., DiCarlo, A., Kometer, K., Lugli, P. and Gornik, E. (1993). "A comparison of Monte Carlo and Cellular Automata approaches for semiconductor device simulation", *IEEE Elec. Dev. Lett.*, **14**, 77–79.

[5] Das, A. and Lundstrom, M. S. (1990). "A scattering matrix approach to device simulation", *Solid-St. Electron.*, **33**, 1299–1307.

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

[7] Alam, M. A., Stettler, M. A. and Lundstrom, M. S. (1993). "Formulation of the Boltzmann equation in terms of scattering matrices", *Solid-St. Electron.*, **36**, 263–271.

[8] Snowden, C. H. (1986). *Introduction to Semiconductor Device Modelling*, World Scientific, Singapore.

[9] McKelvey, J. P., Longini, R. L. and Brody, T. P. (1961). "Alternative approach to the solution of added carrier transport problems in semiconductors", *Phys. Rev. B*, **123**, 51.

[10] McKelvey, J. P. and Balogh, J. C. (1964). "Flux methods for the analysis of transport problems in semiconductors in the presence of electric fields", *Phys. Rev.*, **137**, 1555.

[11] Shockley, W. (1962). "Diffusion and drift of minority carriers in semiconductors for comparable capture and scattering mean free paths", *Phys. Rev.*, **125**, 1570–1576.

Authors' Biographies

Kausar Banoo received her B.Tech. (Electrical Engineering) from the Indian Institute of Technology, Bombay in 1993 and her M.S. (Electrical Engineering) from Duke University in 1995. Currently she is enrolled in the Ph.D. program at Purdue University and is studying scattering matrix techniques for simulating electron transport in ultra-small MOSFETs.

Farzin Assad obtained his B.S.E.E and M.S.E.E, in 1989 and 1991 both from University of Minnesota. Since 1995 he has been involved in the scattering matrix simulation of silicon devices at Purdue University. His interest is in the area of high speed devices and circuits.

Mark Lundstrom received B.E.E. and M.S.E.E. degrees from the University of Minnesota in 1973 and 1974 and a Ph.D. from Purdue University in 1980. From 1974–1977 he was employed by Hewlett-Packard where he worked on the development of a second-generation NMOS integrated circuit process. He is currently Professor of Electrical and Computer Engineering at Purdue. Lundstrom's research interests center on the physics of semiconductor devices. He is an IEEE Fellow and the recipient of the ASEE Frederick Emmons Terman Award, and of two teaching awards from Purdue University.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 545–548
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

A 3D Nonlinear Poisson Solver

GYULA VESZELY

Department of Electromagnetic Theory, Budapest Technical University, 1521 Budapest, Hungary

The Poisson equation is solved in a rectangular prism of semiconductor with the boundary conditions commonly used in semiconductor device modeling. There is a planar heterojunction inside the prism. The finite difference formulation leading to a matrix of seven diagonals is used. The 3D version of the Stone's method is applied for the iterative solution of the matrix equation. The nonlinear dependence of the carrier concentration on the electrostatic potential is taken into account. The heterojunction is modeled by a potential jump. The advantages and limits of the method is presented.

Keywords: Poisson's equation, semiconductor device modeling, Stone's method, heterojunction

PROBLEM STATEMENT

The solution of the nonlinear Poisson's equation is a fundamental task in the modeling of semiconductor devices [1]. The problem to be examined was formulated in Ref. [2]. Figure 1 shows a possible realization in a rectangular geometry. The prism is combined from two different semiconductors contacting in a planar heterojunction. On the lateral faces and on the base zero Neumann boundary condition is supposed. On the upper surface on the rectangular spots (these are the exposed surfaces) zero Neumann boundary condition, on the metallized remaining part (this is the gate) Dirichlet condition is prescribed respectively. Chen and Porod [2] solved the problem by a coupled finite element/boundary element method. The drawback of this method is that it demands a matrix bandwidth reduction and a Newton iteration. The purpose of the present paper to solve the problem by the finite difference method combined with a 3D version of the Stone's [3] method.

NUMERICAL TREATMENT

A standard finite difference discretization is applied with variable grid distances. The boundary conditions can be taken into account in the usual manner. At the zero-Neumann boundaries mirror potentials were used, while at the i^{th} surface point where the potential is prescribed to be Φ_{given} the equation $\Phi_i = \Phi_{\text{given}}$ is valid, i.e., in the i^{th} row all of the matrix elements are zero except the main diagonal element, which equal to unity and at the right-hand side of the equation in the column vector of the excitation the i^{th} element is Φ_{given}. Continuously numbering the grid points a sparse matrix with seven diagonals is resulted.

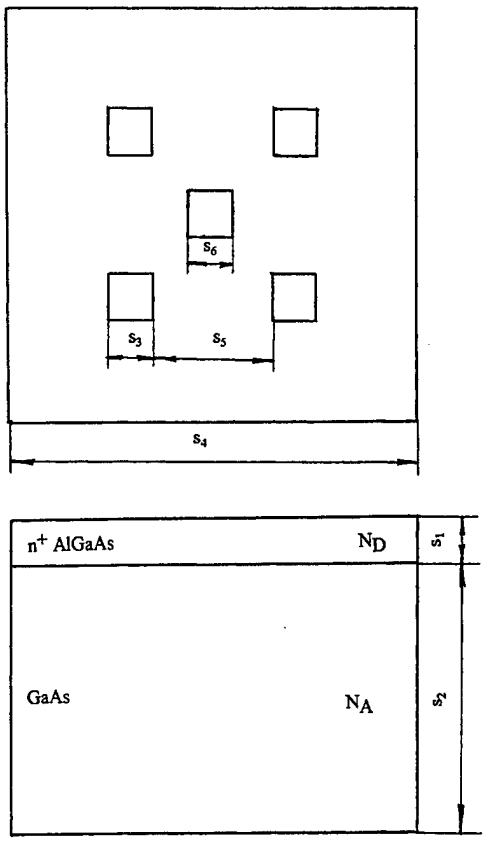


FIGURE 1 The structure to be analyzed. The upper part is a top view, while the lower part is a side-view. The small squares on the top view are the exposed surfaces.

The heterojunction is characterised by a potential jump [4]. This condition must be built into the sytem of equations so that no excess matrix elements to be appeared. We enforce the potential jump by a double layer. The connection between the moment of the doble layer and the given potential jump is

$$\sigma \Delta n = \varepsilon \Delta \Phi$$

where σ is the surface charge density, Δn is the distance between the positive and negative layer, ε is the permittivity and $\Delta\Phi$ is the potential jump characterizing the heterojunction (Fig. 2). Using the Gauss-theorem f.e. for the upper volume element in Figure 2 the field strengthes can be

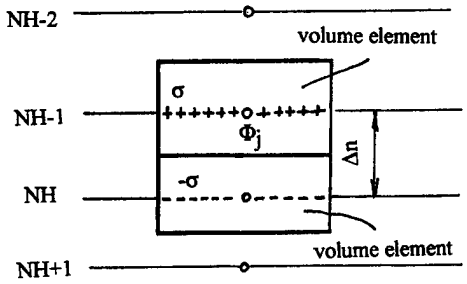


FIGURE 2 The potential jump is modeled by a double layer The Gauss-theorem is applied to the volume element.

calculated as the potential differences divided by the appropriate distances. So Φ_j and all of the six neighbouring potentials appear in the j^{th} equation as in an "ordinary" point (without surface charge). The given charge enclosed by the volume will be an element of the right-hand side column vector. The distance Δn is choosed $0{,}01^{th}$ part of the neighbouring (to the surface perpendicular) grid distances.

The treatment of the potential dependent carrier density is according to [5]. This is fitted itself naturally to the iterative procedure of Stone.

CONVERGENCE PROPERTIES

The numerical results are calculated under the following data (see Fig. 1): the depth of the heavily doped semiconductor $s_1 = 10\,\text{nm}$, the depth of the substrate $s_2 = 110\,\text{nm}$, the side length of the squares of the satellite excited surfaces $s_3 = 10\,\text{nm}$, the side length of the rectangular prism $s_4 = 120\,\text{nm}$, the distance between two satellite excited surface $s_5 = 40\,\text{nm}$, the side length of the square of the central excited surface $s_6 = 10\,\text{nm}$, the donor density of the heavily doped part $N_D = 10^{18}\,\text{cm}^{-3}$, the acceptor density of the substrate $N_A = 10^{15}\,\text{cm}^{-3}$, the voltage characterizing the Schottky contact is 0,8 V, the potential jump at the heterojunction is 0,3 V. The gate voltage is $-1\,\text{V}$.

When the number of unknowns exceeds a critical level the convergence speed degrades drastically (Fig. 3). 35 min CPU time is required for the 1700 iteration on a HP 9000 715/80 workstation. As it is known the α iteration parameter of the Stone's method is an important but theoretically not well established quantity. Most authors using the method have their own recipe for choosing the iteration parameter. We used the following formula proposed by Stone [3]

$$\alpha = 1 - b^e$$

where e varied cyclically with the values 0,6; 0,4; 0,2; 0,8; 0,2; 1. b generally equal to 10^{-3} but in the case of the curve a) in Figure 3 we had to change it to 10^{-2} at the iteration number 100 otherwise divergence occured.

RESULTS

The electron potential energy was calculated at the substrate side of the heterojunction Figure 4 shows a contour plot, while Figure 5 a surface plot of the energy. The diagrams presented are calculated under the grid dimension $59\times59\times59\approx205\,000$. In the vertical direction in the heavily doped part 50 grid points was applied uniformly and only 9 one in the substrate, the distances of the latter are increasing as a geometrical series. Using the symmetry only one quarter of the arrangement was calculated.

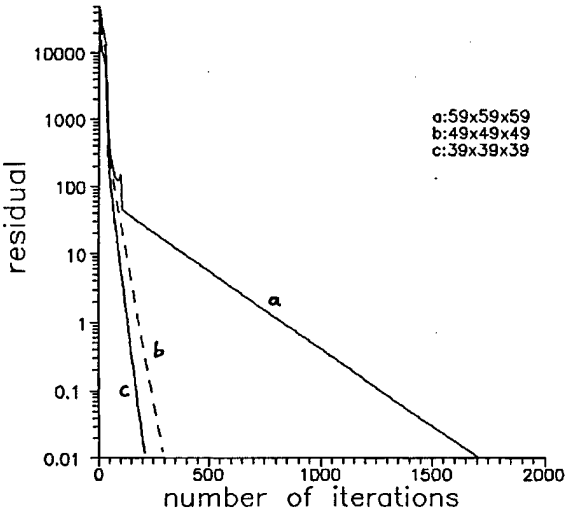


FIGURE 3 The residual versus the iteration number. In the inlet the numbers of grid point can be seen.

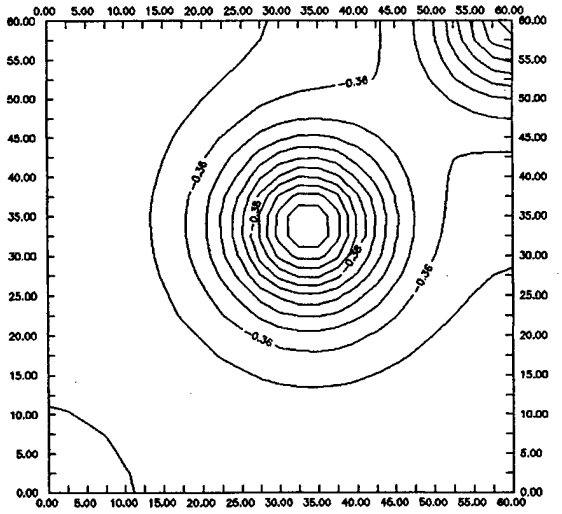


FIGURE 4 Contour plot of the electron potential energy. One satellite dot and a quarter of the central dot can be seen.

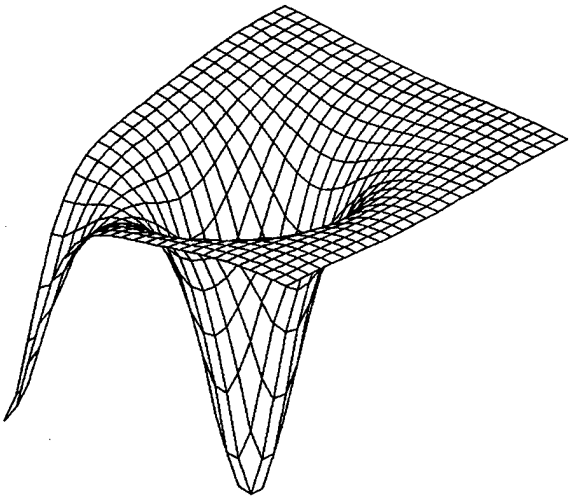


FIGURE 5 Surface plot of the electron potential energy . One satellite dot and a quarter of the central dot can be seen.

By the method the ratio of the depth of the central and satellite valleys and the absolute depth of the valleys can be calculated. These parameters, which are important in the charging and recharging process can be tailored by the proper choice of the geometry and the gate voltage.

CONCLUSIONS

The electrostatic problem of a semiconductor quantum device with exposed surface was solved by the finite difference/Stone's method.

The advantages of the method are: simple programing, low storage requirement, the iteration demanded by the nonlinearity can be naturally built into the iteration type solution.

The drawbacks are: after a limit the convergence becomes rather slow, there is no possibility to the local refinement of the grid.

The application of a more advanced iterative solver [6] is under development.

Acknowledgments

Valuable discussions with A. CSURGAY and W. POROD (University of Notre Dame) are gratefully acknowledged. This work has been supported by Hungarian National Science Foundation No. T-023337 and by a joint grant of NSF (National Science Foundation) and HAS (Hungarian Academy of Sciences) No. INT96-15497.

References

[1] Csurgay, A. and Simonyi, K. (1997). *Physics of Electronic Information Technologies*, (In Hungarian: *Az információtechnika fizikai alapjai*), BME Mérnöktovábbképzõ Intézet, Budapest.

[2] Chen, M. and Porod, W. "Design of Gate-Confined Quantum-Dot Structures in the Few-Electron Regime", *J. Appl. Phys.*, July 1995.

[3] Stone, H. L. "Iterative Solution of Implicit Approximations of Multidimensional Partial Differential Equations", *SIAM J. Numer. Anal.*, September 1968.

[4] Chen, M., Porod, W. and Kirkner, D. J. "Coupled Finite Element/Boundary Element Method for Semiconductor Quantum Devices with Exposed Surfaces", *J. Appl. Phys.*, March 1994.

[5] Gummel, H. K. "A Self-Consistent Iterative Scheme for One-Dimensional Steady-State Transistor Calculations", *IEEE Trans. on Electron Devices*, October 1964.

[6] van der Vorst, H. A. "Bi-CGSTAB: a Fast and Smoothly Converging Variant of Bi-CG for the Solution of Nonsymmetric Linear Systems", *SIAM J. Sci. Stat. Comput.*, March 1992.

Author Biography

Gyula Veszely was born on December 22, 1938 in Szombathely, Hungary. He obtained the master degree in electrical engineering in 1962 from the Budapest Technical University (BTU). In 1993 he obtained the Doctor of Science degree from the Hungarian Academy of Sciences. Since 1962 he has been working at the Department of Electrical Engineering and Informatics of BTU, since 1993 he has been a Professor and head of the Chair of Electromagnetic Theory. His main research interest is in the development and application of numerical methods for computing electromagnetic fields in complicated materials and geometries.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 549–553
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

An Alternative Geometry for Quantum Cellular Automata

PAUL G. KRAUSE, RACHEL M. MUELLER, P. DOUGLAS TOUGAW*
and JANELLE M. WEIDNER

Department of Electrical and Computer Engineering, Valparaiso University, Valparaiso, IN 46383 USA

We examine an alternative layout geometry for the quantum cellular automata (QCA) architecture. In the traditional QCA geometry, all of the cells are placed in a single plane, so that each cell interacts with a particular neighbor only along one of its edges. By rotating the cells out of the plane, we make it possible for neighbors to interact along all four edges at once. This increased interaction leads to a more bistable cell-cell response function and a 50% higher excitation energy. We also present a majority logic gate designed using three-dimensional cells.

Keywords: Quantum cellular automata, computer architecture, quantum-dot molecules

1. INTRODUCTION

The quantum cellular automata (QCA) architecture, which is composed of arrays of Coulombically coupled quantum dot molecules, is a revolutionary new paradigm for quantum computing [1]. QCA's offer new possibilities for computing at the nanometer scale, since they can be used to hierarchically implement any logical function [2–4].

In this paper, we present an alternative layout geometry for the QCA architecture by rotating the cells out of their plane. In the next section, we briefly review the QCA paradigm in two dimensions, while in Section 3 we examine the QCA architecture in three dimensions. By generalizing QCA cells to three dimensions and computing the resulting cell-cell response curve, we show that the three-dimensional cells are more bistable than the traditional two-dimensional cells.

In Section 4, we examine a three-dimensional majority logic gate. We show that two-dimensional and three-dimensional majority gates give identical functional results.

2. TWO-DIMENSIONAL QCA DEVICES

Figure 1 shows the basic building block in the QCA paradigm, the quantum dot molecule cell. Four quantum dots are spaced close enough for electrons to tunnel between neighboring pairs of dots. Two stable states are possible with this configuration. The first, corresponding to an

*Corresponding author: dtougaw@diamond.gem.valpo.edu.

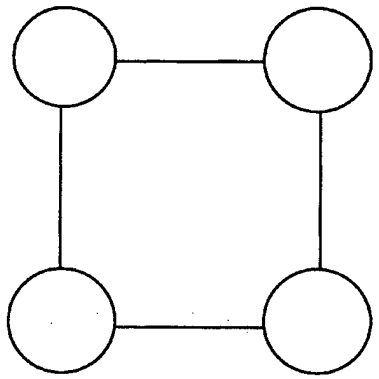


FIGURE 1 A schematic of the basic QCA cell. Electrons interact Coulombically and tunnel among the four sites.

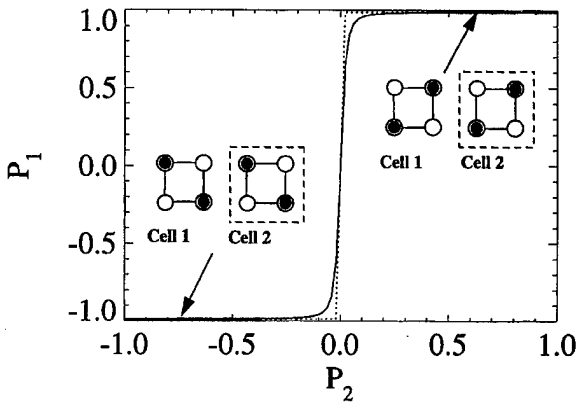


FIGURE 2 The cell-cell response function of two neighboring QCA cells. The bistable nature of the response function allows us to encode a binary bit in each cell.

electron occupying the upper right hand and lower left hand dots, represents *a* polarization of +1. The other, corresponding to an electron in the upper left hand and lower right hand dots, represents *a* polarization of −1. These two polarizations can represent binary values of "1" and "0", respectively.

The Hamiltonian used to describe each cell includes Coulombic interaction between each pair of sites, near-neighbor tunneling, next-near-neighbor tunneling, on-site energies and on-site charging costs. The on-site energy of each site will be affected by the polarization of its neighboring cells.

Thus, Coulombic forces cause the polarization of one cell to affect the polarization of neighboring cells. The cell-cell response curve shown in Figure 2 is computed by solving the two-particle, time-independent Schrödinger equation for the first cell in the presence of the second cell. The curve shows that the response of such cells is very bistable, which means that for even very small polarizations in one cell, the second cell is completely polarized. This feature is very attractive, since it allows these cells to encode a single binary bit in their quantum state. Furthermore, when propagating signals, the signal level is restored to the maximum polarization at every stage.

The most useful logic structure in the QCA paradigm is the three-input majority gate. A majority gate is one in which the output agrees with the majority of the inputs. This logic structure is essential because we can produce both AND and OR gates from it. A two-input OR gate can be produced by setting one of the three inputs equal to 1, while a two-input AND can be produced by setting one of the three inputs equal to 0.

Figure 3 shows a schematic for the coplanar QCA majority gate. In this figure, inputs A, B and C are used as inputs. The output is propagated down the line of cells acting as a binary wire. The ground state of this multiple-cell structure was

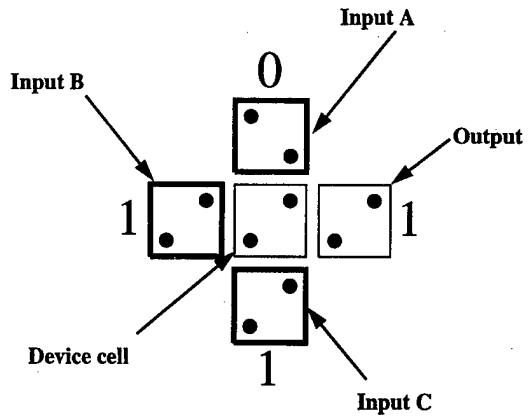


FIGURE 3 The coplanar majority voting gate. The device cell always matches a majority of the three fixed neighbors.

computed by a self-consistent solution of the Schrödinger equation.

3. AN ALTERNATIVE CELL GEOMETRY

The QCA paradigm presented thus far has included only cells lying in a two-dimensional plane. By generalizing the architecture to three dimensions, we can show that the cell-cell interaction is significantly stronger than in the two-dimensional case.

A schematic of our three-dimensional cell-cell layout is shown in Figure 4. Again, four dots are used, but in this case, the cells are facing each other so that all four of the dots are strongly interacting with the neighbor.

The cell-cell response curve of this layout is shown in Figure 5. The intercellular distance was computed such that the centers of the cells were at the same distance as the centers of the two-dimensional calculation. The code to solve the time-independent Schrödinger equation was generalized to allow for Coulombic interactions in all three dimensions.

It should be noted that three-dimensional neighbors tend to align opposite one another, so the cell-cell response function was horizontally

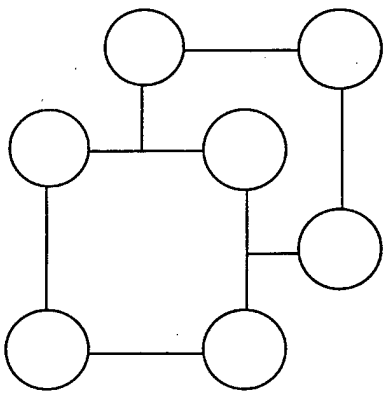


FIGURE 4 The cell layout for the three-dimensional cell-cell response function calculation.

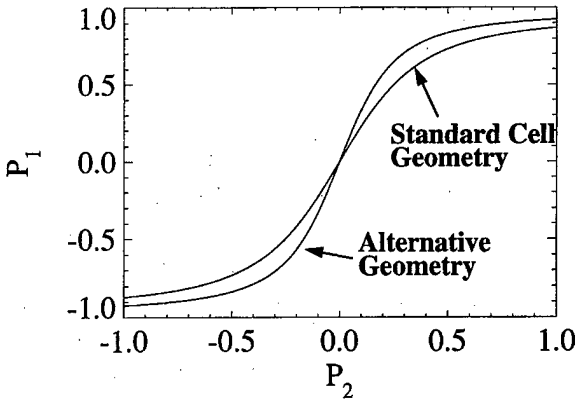


FIGURE 5 Cell-cell response function of the three-dimensional geometry. In addition to yielding a more bistable response function, use of the alternative geometry increases the excitation energy by 50%.

flipped to allow easier comparison of the two curves. As with the "inverter chain" of the two-dimensional QCA devices, this anti-alignment behavior can be overcome by using an even number of cells to transmit information. Of course, this will also make the implementation of three-dimensional inverters quite simple.

From the curves shown, it is easy to see that the three-dimensional curve is more bistable than the two-dimensional curve. In addition, the eigenvalue splitting between the ground state and the first excited state is 50% larger in the three-dimensional cells than it is in the two-dimensional cells. This is to be expected, since in the three-dimensional case, all four dots are Coulombically interacting, while in the two-dimensional case, only the near-neighbor dots strongly interact with each other.

Since the excitation energy between the ground state and the excited states is the quantity that controls the time-dependent evolution of the system, we can expect that this alternative geometry will propagate signals more quickly than the two-dimensional cells. In addition, the thermodynamic behavior, the adiabatic switching speed, and the maximum operating temperature are also dependent on the excitation energy, so similar improvements can be expected in those areas, as well.

4. A THREE-DIMENSIONAL MAJORITY GATE

As mentioned in the previous section, the most important logical device of the QCA paradigm is the majority logic gate. Figure 6 demonstrates the layout of a QCA majority gate in three dimensions. Cells 1 and 3 are coplanar to cell 4, so they tend to drive it to match their polarization. On the other hand, cell 2 drives cell 4 to an opposite polarity, since it interacts using the three-dimensional cell-cell response function. This can be cancelled by using an extra cell in that input, or it can be used to implicitly invert that input.

In order to ensure that the three inputs "vote" with equal weighting, the distances between cells were chosen such that when all four neighbors of the device cell are voting in opposite directions, the device cell exhibits no polarization.

To find this distance, the three coplanar cells (the device cell and two inputs) were placed a distance 60 nm apart from each other and the two coplanar drivers were set with $P = +1$. The other two neighbors were then set with $P = -1$ and moved away from the device cell until it exhibited zero polarization. At that point, we were certain that the horizontal and vertical neighbors were having the same impact on the polarization of the device cell. The distance between the non-coplanar cells at which this occurred was 1.7435 times larger than the distance between the coplanar cells, since the non-coplanar cells interact more strongly with each other.

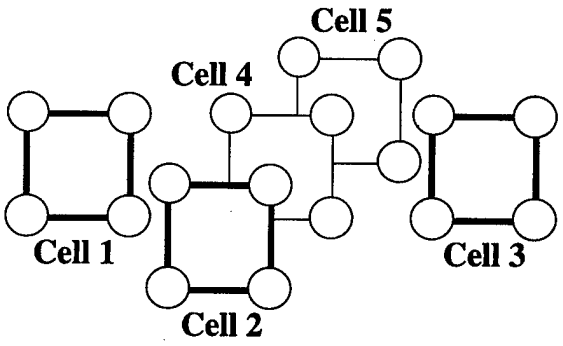


FIGURE 6 A schematic representation of the three-dimensional majority logic gate. Cells 1, 2 and 3 are drivers, while cell 4 is the device cell and cell 5 is the output. Cells 2 and 4 tend to anti-align, so the input at cell 2 is implicitly inverted.

We used the Intercellular Hartree Approximation to find the ground state of this multiple-cell system. The three input cells were fixed at a specific polarity (either $+1$ or -1), while the device cell and the output cell were allowed to change.

We tested each of the eight combinations of driver polarizations and found that in each case, the ground state corresponds to the correct result for a majority logic gate. The state of the output always matches the state of a majority of the three inputs, assuming the non-coplanar input has been implicitly inverted.

5. CONCLUSIONS

We have proposed an alternative geometry for the QCA computing architecture and we have shown that this geometry achieves a better cell-cell response than the two-dimensional implementation. We have also shown that an inverter and a majority logic gate can be constructed three-dimensionally, so any combinational function can be constructed using this alternative geometry.

Acknowledgements

The authors would like to acknowledge the support of the Valparaiso University College of Engineering, as well as many productive conversations with the University of Notre Dame Nanoelectronic Devices Group.

References

[1] Lent, C. S., Porod, W. and Tougaw, P. D. (1993). *Appl. Phys. Lett.*, **62**, 714.
[2] Lent, C. S., Tougaw, P. D., Porod, W. and Bernstein, G. (1993). *Nanotechnology*, **4**, 49.
[3] Tougaw, P. D., Lent, C. S. and Porod, W. (1993). *J. Appl. Phys.*, **74**, 3558.
[4] Lent, C. S. and Tougaw, P. D. (1993). *J. Appl. Phys.*, **74**, 6227.

Authors' Biographies

Paul G. Krause is a graduate student at the Uiversity of Minnesota, where he is pursuing an advanced degree in Biomedical Engineering. He earned his B.S.E.E. degree from Valparaiso University.

Rachel M. Mueller is currently employed by Motorola, where she is working as a software engineer in the iDEN group. She earned her B.S.Cp.E. degree from Valparaiso University.

P. Douglas Tougaw is an assistant professor at Valparaiso University. He received his B.S.E.E. degree from Rose-Hulman Institute of Technology and his Ph.D. degree from the University of Notre Dame.

Janelle Weidner is currently an undergraduate student of electrical engineering at Valparaiso University. She is participating in a cooperative education experience with the National Security Agency.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 555–558
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Electrostatic Formation of Coupled Si/SiO_2 Quantum Dot Systems

PER HYLDGAARD*, HENRY K. HARBURY and WOLFGANG POROD

Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556

We present three-dimensional numerical modeling results for gated Si/SiO_2 quantum dot systems in the few-electron regime. In our simulations, the electrostatic confining potential results from the Poisson equation assuming a self-consistent Thomas-Fermi charge model. We find that a very thin SiO_2 top insulating layer allows an effective control with single-electron confinement in quantum dots with radius less than 10 nm and investigate the detailed potential and resulting charge densities. Our three-dimensional finite-element modeling tool allows future investigations of the charge coupling in gated few-electron quantum-dot cellular automata.

Keywords: 3D simulation and modeling, finite element method, silicon/silicon-dioxide quantum dots, quantum-dot cellular automata

We present numerical simulations for electrostatically confined few-electron quantum dot systems in the technologically important Si/SiO_2 material system. Our emphasis is modeling a possible so-called *Quantum-dot cellular automata* (QCA) structure [1] in which a bi-stable occupation by two excess electrons in a small and strongly charge-coupled quantum-dot system defines logic 0/1.

The *bottom panel* of Figure 1 shows a schematic of the Si/SiO_2 material system: a thin silicon-dioxide layer serves as excellent insulation of the bottom silicon slab from the set of top gates. Applying finite biases at these gates allows the formation of electrostatically confined quantum dots just below the heterostructure interface (at $z = z_0$). Mesoscopic transport investigations in gate-induced quantum-dot arrays [2] documents the feasibility of fabricating few-electron quantum-dot systems in the Si/SiO_2 material system. This development has in turn resulted in a proposal for room-temperature single-electron Si/SiO_2 memory cells [3].

Previous (two-dimensional) modeling results [4, 5] of few-electron Si/SiO_2 quantum dot systems exploited an axial symmetry to investigate the electrostatic confinement within an individual dot. Encouraged by recent three-dimensional modeling of larger quantum-dot systems [6] we investigate

*Corresponding author.

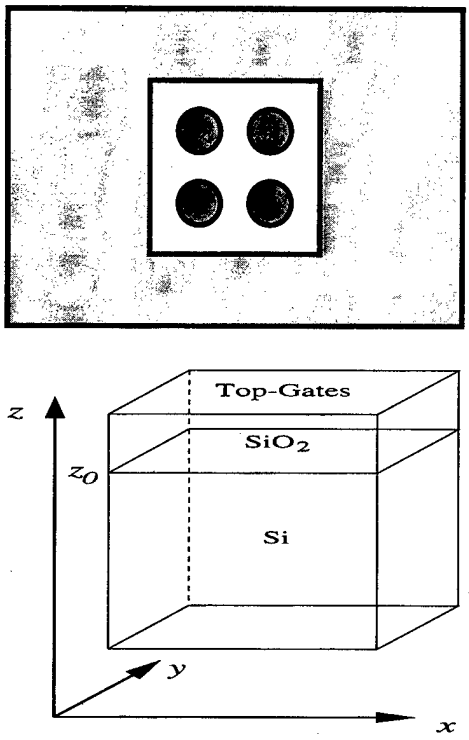


FIGURE 1 Schematics of possible gate-confined Si/SiO_2 quantum-dot cellular automata (QCA) The *lower panel* shows material composition: thin (5 nm) SiO_2 insulating layers grown on a Si block with an unintentional $p=10^{15}\,cm^{-3}$ p-doping. Top gates define quantum dot system just below the hetero-interface at $z=z_0$. The *upper panel* illustrates potential QCA cell formation: group of four positive small attractor gates (shaded circles) from charge-coupled system of close quantum dots. The entire QCA cell is surrounded by negative depletion gate (also shaded) to enhance confinement.

here a possible realization of a Si/SiO_2 quantum-dot cellular automata.

The *upper panel* of Figure 1 shows schematics of the small but experimentally accessible top-gate geometry assumed for our model Si/SiO_2 QCA system. A negative depletion gates (shaded) surrounds a group of four positive attractor gates (shaded circles) with a $r=10$ nm radius and mutual separation of 30 nm. The attractor gates are biased to about 1 V to ensure a single-electron equilibrium occupation in each of the four electrostatically confined quantum dots. For the top silicondioxide layer we assume a 5 nm thickness.

In our $T=100$ K simulations, the confining potential is obtained from the Poisson equation within a self-consistent Thomas-Fermi charge model. The silicon is assumed to have a small unintended but fully ionized doping, $p=10^{15}\,cm^{-3}$ and to ensure convergence we investigate a 1.5 nm thick bottom slab with in-plane extension of approximately 300 nm by 300 nm. The top metal depletion and attractor gates are described by Dirichlet boundary conditions. For the exposed SiO_2 surface we assume for simplicity a potential fixed at the mid-gap SiO_2 value, that is, again a Dirichlet boundary condition.

Our finite-element calculation uses a 129 by 129 nonuniform grid to allow a 1 nm resolution from the surface and well below the Si/SiO_2 interface, that is, around the quantum dot system. The *top panel* of Figure 2 shows how most nodal layers (at constant z) are connected in a mesh with alternating tetrahedron orientation to eliminate a geometrical bias. The *bottom panel* of Figure 2 illustrates the repeated thinning of our finite-element mesh undertaken deep below the interfaces where a high in-plane resolution is no longer needed. However, our numerical simulation still involves 6×10^5 nodal points for which we determine the electrostatic potential within the self-consistent Thomas Fermi screening model. Using the Newton-Ralphson procedure we solve in each iteration the resulting huge linear system using a quasi-minimal residual implementation [7].

Figure 3 shows our finite-element determination of the electron potential and charge distribution along the interface ($z=z_0$) for our possible QCA realization. The electron confinement potential (*top panel*) is calculated relative to the Fermi energy with negative values corresponding to a strongly enhanced electron concentration n_{el} (*bottom panel*). Note how the thin silicon-dioxide top layer allows a very well-defined set of quantum dots with a finite inter-dot potential barriers. The total equilibrium occupation is set to exactly four electrons and the strong Coulomb blockade effect will prevent multiple occupation of the individual quantum dots.

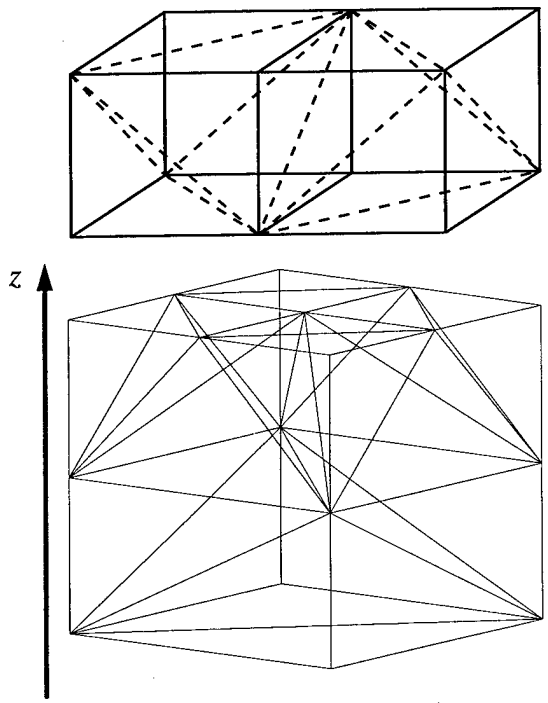


FIGURE 2 Three-dimensional finite-element modeling of gated Si/SiO_2 quantum-dot systems. The *upper panel* shows tetrahedron-based mesh with alternating cell orientation used in most in-plane layers (i.e., with constant z) to eliminate geometric bias. The *lower panel* illustrate repeated mesh thinning at layers deep below the interface ($z \ll z_0$) used to obtain a factor-of-two reduction in numerical complexity.

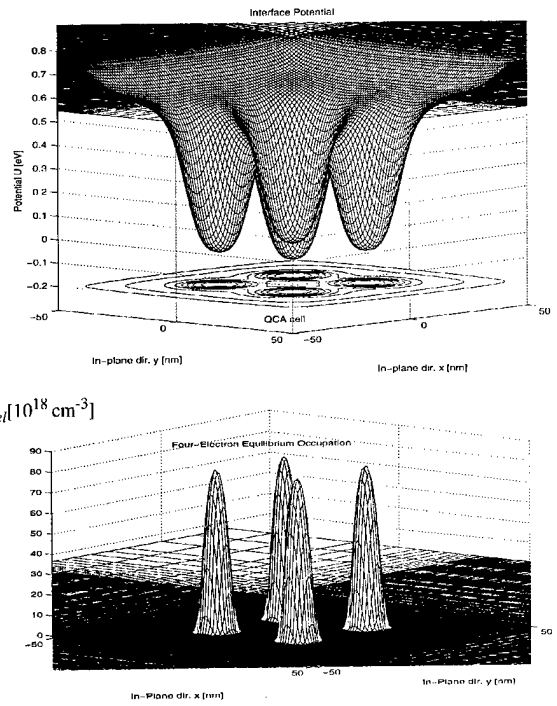


FIGURE 3 Electron potential and charge distribution of possible QCA cell at the heterostructure interface ($z = z_0$). Potential is calculated for an experimentally accessible gate geometry: four attractor gate of radius $r = 10$ nm with mutual 30 nm separation. The *upper panel* shows potential dips (measured relative to Fermi energy) with corresponding contour plot and illustrates the crisp gate control allowed by the thin SiO_2 top layer. The electron potential confines in equilibrium exactly four electrons with very well-defined charge distribution n_{el} (*lower panel*).

Figure 4 shows our modeling results for the simpler double-quantum-dot system in which we are preparing to investigate the *mutual* charge coupling between the quantum dots. We assume again $r = 10$ nm attractor gates with a mutual separation of 30 nm and adjust the positive bias to achieve a single-electron equilibrium occupation of each of the quantum dots. The *upper panel* shows the variation of the confinement potential both along the axes (x) connecting the two quantum dots and in the growth direction (z). The heterostructure-cut panel illustrates the excellent top-gate control of the electrostatic potential into the Si/SiO_2 slab well below the heterostructure interface, $z = z_0$.

The *lower panel* of Figure 4 shows the corresponding equilibrium charge distribution, n_{el}. Note that this electron distribution, i.e., the equilibrium quantum dot, is confined within 1 nm of the interface and about 5 nm of the attractor-gate center. Future modeling will investigate the charge coupling of such quantum-dot disks in the presence of the attractor and depletion gates.

In summary, we have presented three-dimensional finite-element calculations for gate-confined few-electron Si/SiO_2 quantum-dot systems. We have documented the feasibility of crisp electrostatic gate control for also few-electron quantum-dot systems and have investigated the detailed charge distribution and confinement potentials. Our modeling tool allows future investigations of charge coupling in few-electron quantum-dot cellular automata structures.

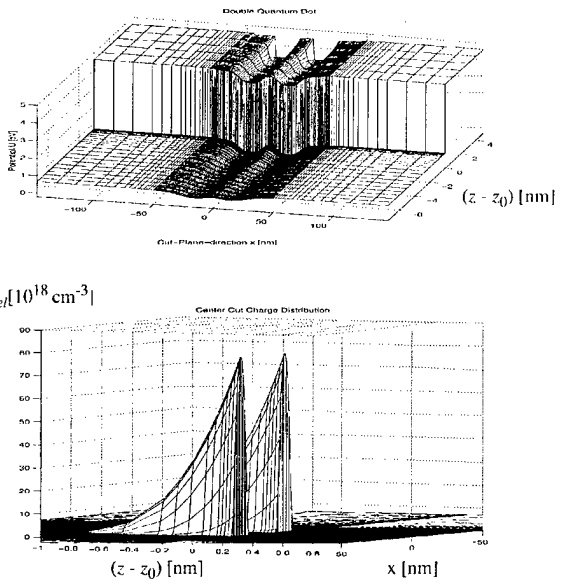


FIGURE 4 Growth direction (z) variation of confinement potential (*upper panel*) and corresponding (equilibrium) charge distribution (*lower panel*) in simpler double-dot model system. This system is defined by two $r = 10$ nm attractor dots also with a mutual separation of 30 nm. All z-positions are given relative to the interface $z = z_0$. Note that the equilibrium electron distribution n_{el} is confined within 1 nm of the interface and about 5 nm of the attractor-gate center.

Acknowledgements

This work has been supported by the Advanced Research Projects Agency through the Office of Naval Research.

References

[1] Lent, C. S., Tougaw, P. D., Porod, W. and Bernstein, G. H. (1993). "Quantum cellular automata", *Nanotechnology*, **4**, 49; Lent, C. S., Tougaw, P. D. and Porod, W. (1993). "Bistable saturation in coupled quantum dots for quantum cellular automata", *Appl. Phys. Lett.*, **62**, 714; Tougaw, P. D., Lent, C. S. and Porod, W. (1993). "Bistable saturation in coupled quantum-dot cells", *J. Appl. Phys.*, **74**, 3558.

[2] Matsuoka, H., Ichiguchi, T., Yoshimura, T. and Takeda, E. (1994). "Coulomb blockade in the inversion layer of a Si metal-oxide-semiconductor field-effect transistor with a dual-gate structure", *Appl. Phys. Lett.*, **64**, 586; and Matsuoka, H., Ichiguchi, T., Yoshimura, T. and Takeda, E. (1994). "Mesoscopic transport in Si metal-oxide-semiconductor field-effect transistor with a dual-gate structure", *J. Appl. Phys.*, **76**, 5561.

[3] Yano, K., Ishii, T., Hashimoto, T., Kobayashi, T., Murai, F. and Seki, K. (1994). "Room-Temperature Single-Electron Memory", *IEEE Trans. Electron Dev.*, **41**, 1628.

[4] Chen, M., Porod, W. and Kirkner, D. J. (1994). "Coupled finite element/boundary element method for semiconductor quantum devices with exposed surfaces", *J. Appl. Phys.*, **75**, 2545; Chen, M. and Porod, W. (1995). "Design of gate-confined quantum-dot structures in the few-electron regime", *J. Appl. Phys.*, **78**, 1050.

[5] Chen, M. and Porod, W., "Simulation of Quantum-Dot Structures in Si/SiO_2", *VLSI Design*, in press.

[6] Udipi, S., Vasileska, D. and Ferry, D. K. (1996). "Numerical Modeling of Silicon Quantum Dots", *Superlat. and Microstruct*, **20**, 343.

[7] Freund, R. W. and Nachtigal, N. M. (1991). "QMR: a Quasi-Minimal Residual Method for Non-Hermitian Linear System", *Numerische Mathematik*, **60**, 315; Lanczos, C. (1952). "Solution of Systems of Linear Equations by Minimized Iterations", *J. Research Nat. Bur. of Standards*, **49**, 33; see also C. Lanczos (1950). "An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators", *J. Research Nat. Bur. of Standards*, **45**, 255.

Authors' Biographies

Per Hyldgaard is a postdoctoral research associate in the Department of Electrical Engineering at the University of Notre Dame. His research interests include computational and theoretical modeling of Coulomb blockade systems. He is a member of the APS and of the Danish and European Physical Societies.

Henry K. Harbury is Director of the computation facilities of the College of Science at the University of Notre Dame. His interests are computational modeling of solid state electronic devices.

Wolfgang Porod is Professor of Engineering at the University of Notre Dame. His research interests include computational solid state electronics and the physics of computation. He is a member of the APS and a senior member of the IEEE.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 559–565
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Electron Transport In One-Dimensional Magnetic Superlattices

ZHEN-LI JI* and D. W. L. SPRUNG

Department of Physics and Astronomy, McMaster University, Hamilton, Ontario, Canada L8S 4M1

Electron transport properties of quantum wires in the presence of a periodically modulated magnetic field are investigated. For a short modulated wire, we find dips in conductance just below each mode threshold. The conductance dips are quite robust at low temperature. Increasing the number of periods of magnetic modulation can lead to the formation of minibands and gaps. The differences between the one dimensional (1D) *electric* superlattice and 1D *magnetic* superlattice are discussed. We also consider the spatial distributions of currents, which show dramatic differences between the magnetic superlattices and electric ones.

Keywords: Two-dimensional electron gas, magnetotransport, nanostructure, quantum wire, superlattice, miniband

Recently, much attention has been paid to electron transport in magnetic fields that are inhomogeneous on the nanometer scale. Experimentally, such magnetic modulation can be realized by laying metallic strips on the surface of the heterostructure, consisting of either superconducting or ferromagnetic material [1, 2, 3], or by varying the topography of a non-planar two-dimensional electron gas (2DEG) grown at a GaAs/AlGaAs heterojunction [4].

In this paper we point out that the experimental study of a new class of semiconductor nanostructures, the 1D *magnetic* superlattices, is now within reach and we present theoretical predictions of their transport properties.

We consider a model of a two-dimensional ballistic quantum wire shown schematically in Figure 1. This quantum wire has infinite length in the x direction and is confined laterally $-W/2 < y < W/2$ by hard wall potentials. Inside the channel the potential is taken to be zero. Our system consists of a finite magnetic superlattice and two straight leads. A uniform background field B_0 is applied to the entire sample, while a finite section of length $L = Na$ is subject to an additional periodically varying magnetic field with amplitude B_m perpendicular to the plane [5, 6].

For a quantum wire without modulation, it is well-known that conductance is quantized in units

*Corresponding author: Tel.: (905) 525-9140 Ext. 23181 Email: Zheji@physun.physics.mcmaster.ca Fax: (905) 546-1252.

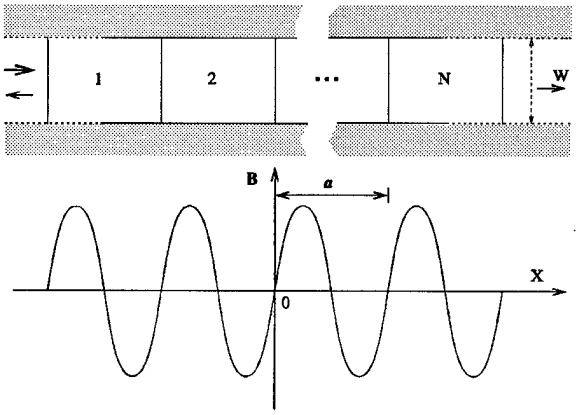


FIGURE 1 Schematic view of a quantum wire of width W with hard-wall confinement in the transverse direction. A finite section is subject to a periodically varying magnetic field, with profile shown in the lower graph.

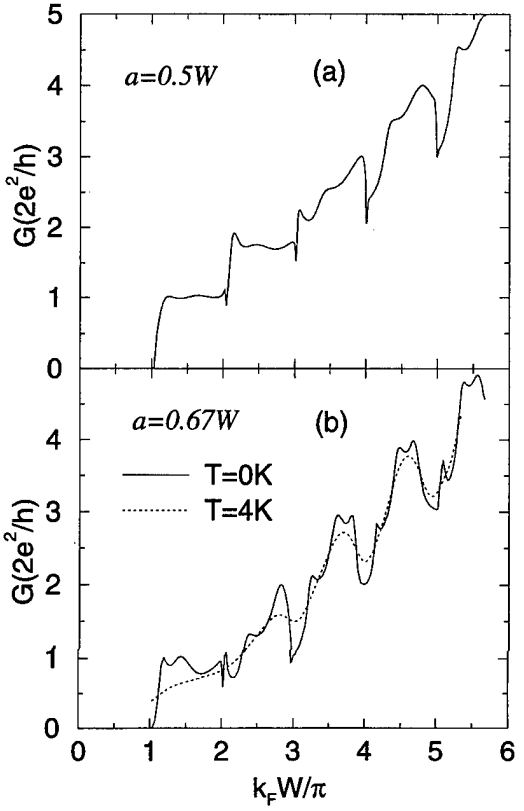


FIGURE 2 Conductance as a function of $k_F W/\pi$ for $N=3$, $B_m=1.2\,T$ and B_0 0.3 T. (a) and (b) correspond to $a=0.5W$ and $0.67W$, respectively.

of $2e^2/h$, due to propagation of a set of discrete modes through the wire. As the magnetic modulation is turned on, the transport will be affected by the oscillations of $\boldsymbol{B}(x)$, so some change in the character of the conductance is expected. We show in Figure 2 the conductance of the quantum wire as a function of $k_F W/\pi$ for $N=3$, $B_m=1.2\,T$ and $B_0=0.3\,T$, respectively. In the case of weak modulation, the conductance still shows a staircase structure with plateaus very close to the quantized values $2e^2n/h$ (n being the number of subbands below E_F) as shown in Figure 2(a). Notice that there are dips in conductance just below the threshold of each mode. Similar dips have also been found in the conductance of perturbed quantum wires subjected to local potential disorder, or a potential barrier, or coupled to a resonant cavity, rather than magnetic modulation. Tekman and Bagwell [7] and Berthod *et al.* [8] have argued that sub-threshold dips like these are Fano resonances; [9] that is, they are due to coupling between a discrete state and the continuum in which it is situated. They have shown how the nearby threshold can induce a discrete bound or anti-bound state and so give rise to the dip.

The coupling between localized states and propagating modes can be revealed by the local current distribution. In the non-magnetic case, it has been found that vortices in the current flow reverse their sense of rotation as the energy is raised through a resonance level [8, 10]. We looked at current flows at a number of energies around the first conductance dip in Figure 2(a). Away from the dip, the flow is laminar on either side. In a narrow energy window, close to the dip, a number of vortices appear, indicating the presence of a localized state. In our example, with three periods of oscillation, there was one large vortex in each period. These move a little as the energy is raised, but the sense of rotation did not change. Rather, the dip was associated with one vortex moving out of the scattering region, leaving a current pattern transverse to the channel along the end of the scattering region, accounting for the reduced transmission. Since the three main vor-

tices are counter-rotating, transmitted electrons follow a meandering path around them. The lack of reversal is understandable as these vortices are driven by the applied magnetic field.

When the background field B_0 is increased, the conductance dips shift to higher values of $k_F W/\pi$ due to the one-dimensional subbands in the leads, which gradually depopulate as an increasing, perpendicular magnetic field is applied (not shown here). In Figure 2(b) the length of the period a is $0.67W$. Lengthening the scattering region provides thicker magnetic barriers, which tends to reduce the transmission. We see that as a increases, the conductance dips become more pronounced. We also get some conductance oscillations between the dips as we increase a; these are due to multiple reflections within the scattering region.

The temperature dependence of the conductance dips is shown in Figure 2(b). We find that the conductance dips are quite robust at low temperatures. The dotted line in Figure 2(b) shows the conductance at $T=4$ Kelvin. The oscillations due to multiple reflections within the scattering region disappear, but the conductance dips are still apparent.

Figure 3 shows the conductance of a modulated quantum wire with $N=4$ and 8 unit cells, respectively. For a wire with $N=4$ unit cells the conductance shows an indication of miniband formation in a periodic structure. Although the basic features of the formation of minibands and gaps are already present for wires with a few unit cells, some narrow and fine minibands and gaps reveal themselves only for a rather long modulated wire, as seen in Figure 3 for the wire with $N=8$ periods.

In Figure 4 we show the calculated conductance of a wire with $N=16$ and the individual transmission coefficients for each incident mode. The conductance displays dips due to the formation of minigaps and rapid oscillations due to electron transmission through the coupled quasi-zero-dimensional states in the cavity regions between the magnetic barriers. Each group of conductance oscillations evolves into a continuous miniband in

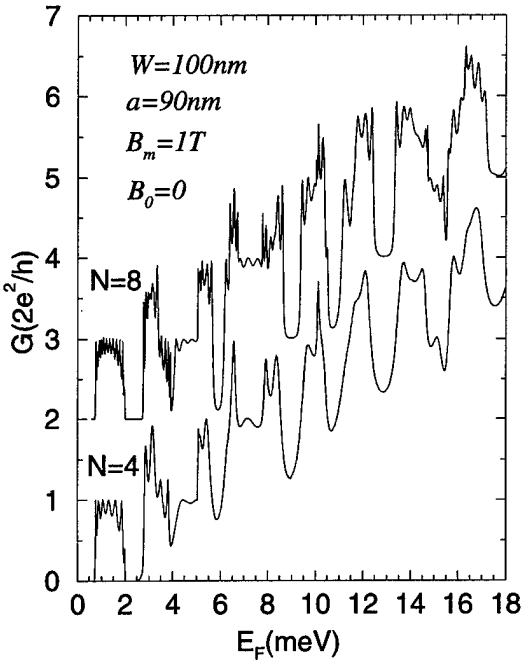


FIGURE 3 Conductance vs Fermi energy E_F for quantum wires with $N=4$ and 8, respectively. Curve for $N=8$ is vertically offset for clarity.

the limit of an infinitely long superlattice. The formation of minibands and minigaps in a finite lateral surface superlattice has been reported using split-gate structures, in which the 1D lateral surface superlattice is realized by a periodic modulation of constriction width instead of magnetic modulation [11].

In the electric modulated case, the number of oscillations in the first conduction plateau is equal to the number of cavity regions [12]. However, in our case the channel is modulated by the magnetic field. We can see that there is no simple direct correspondence between the number of unit cells and the number of oscillations. This result shows that the effective potential of the magnetic barriers for electron motion in the wire is complicated [13]. The oscillations in higher plateaus are more irregular because of the strong coupling between modes. The peak values of the conductance oscillations should be an integer multiple of $2e^2/h$. Some of the peaks fail to reach a quantized

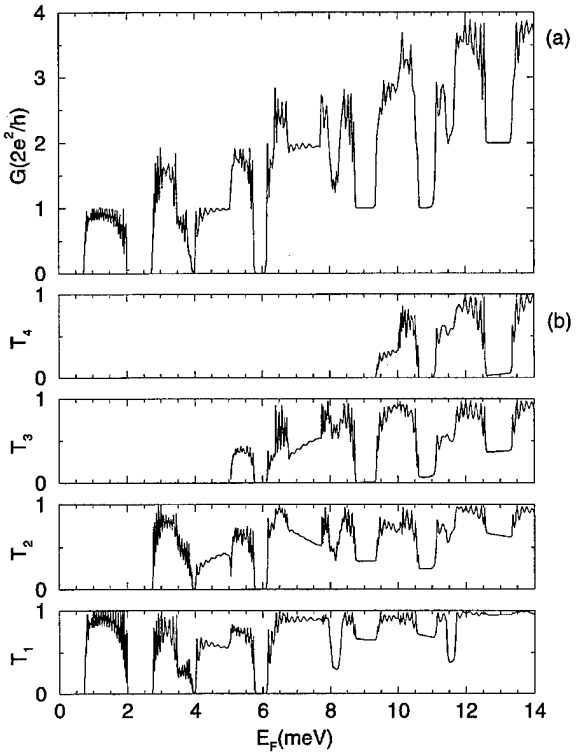


FIGURE 4 (a) Conductance as a function of Fermi energy E_F for a finite period magnetic superlattice with 16 unit cells. $a=90$ nm, $W=100$ nm, $B_0=0$ and $B_m=1.5$T. (b) Transmission coefficients of individual modes for the magnetic modulated quantum wire.

value due to the limited number of data points. The peaks are very narrow and therefore require more extensive computing. We find that the oscillations become more prominent as a becómes longer. It is thus suggested that the formation of an effective loop of edge states is essential for the conductance oscillations as discussed by Takagaki [14] and Yoshioka [15]. One is then dealing with an Aharonov-Bohm type interference.

The conductance, plotted in Figure 4(a), steps up and down between quantized levels. Conductance quantization is related to the band structure for the corresponding infinite system [12]. However, none of the individual modal transmission coefficients in Figure 4(b) shows quantization by itself. See for example the region near 9 meV. The quantization occurs as the various modes are mixed by the periodic modulation.

The calculated conductance as a function of the amplitude B_m, is displayed in Figure 5(a). The conductance steps down by two-units of $2e^2/h$ and up by one-unit, then down to zero as B_m is increased. Transmission coefficients T_i of indivi-

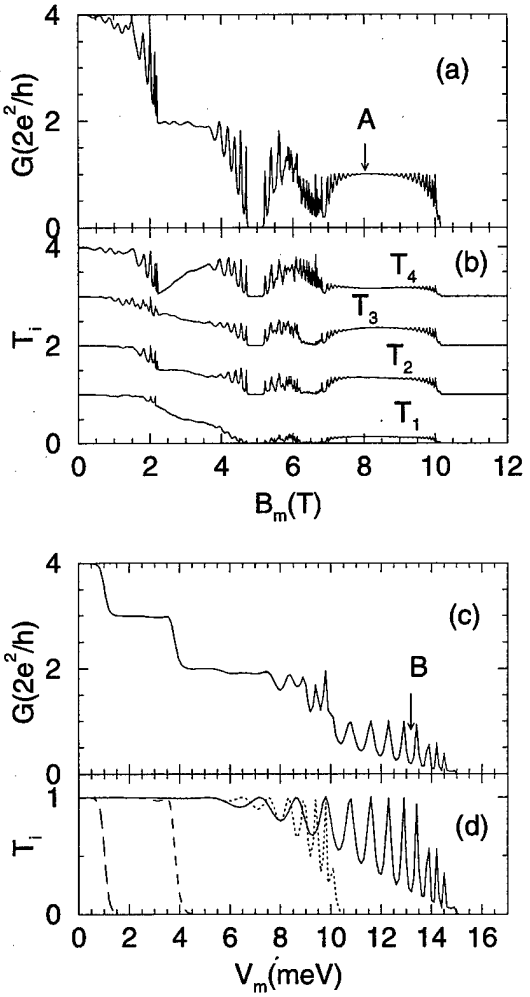
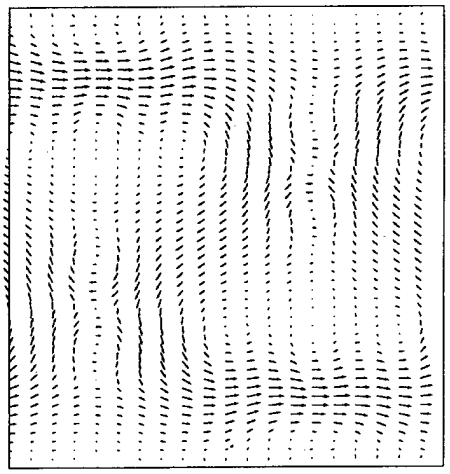


FIGURE 5 (a) The conductance as a function of the *magnetic* modulation amplitude B_m for a quantum wire with 16 unit cells. $a=90$ nm, $W=100$ nm, $B_0=0$ and $E_F=12$ meV. (b) Transmission coefficients T_i of individual modes for the *magnetic* superlattice. The curves are offset for clarity. (c) The conductance as a function of the *electric* modulation amplitude V_m for a finite period *electric* superlattice. The other parameters are the same as (a). (d) Transmission coefficients T_i of individual modes for the *electric* superlattice. The solid, dotted, dashed and long-dashed curves correspond to $i = 1, 2, 3$ and 4, respectively. The arrows A and B indicate the cases corresponding to the current distributions shown in Figures 6(a) and 6(b), respectively.

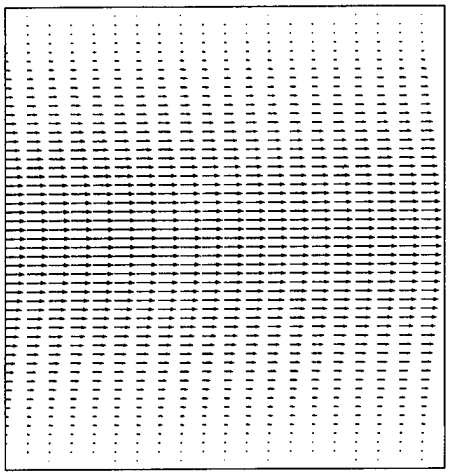
dual modes for the magnetic superlattice are shown in Figure 5(b); the curves are vertically offset for clarity. Although the total conductance, which is the sum of the modal transmissions, is essentially quantized, the individual transmissions, plotted in Figure 5(b), show complicated dependence on B_m.

We also calculated the conductance for an *electric* superlattice with amplitude V_m and sinusoidal barriers. The geometry of the structure is otherwise the same as that of Figure 5(a). The conductance for a 1D *electric* superlattice steps down monotonically as the propagating modes in the leads are blocked by the potential barriers with increasing electric modulation amplitude V_m, as shown in Figures 5(c) and 5(d). No two-unit drop is seen for increasing electric modulation amplitude V_m in Figure 5(c). Clearly, the unique features of the wave-vector-dependent effective potential for the magnetic barriers are critical for the two-unit drop in Figure 5(a). In our calculations, the two-unit drop in conductance with increasing B_m is predominant but not universal; occasionally we find cases with a one-unit drop. The reason for this is that the effective potential of the magnetic barriers for electron motion along the quantum wire depends on the mode of the incident electron in the leads [13].

Figures 6(a) and 6(b) show the current densities in one unit cell of the magnetic/electric superlattices for the cases labelled by A and B in Figures 5(a) and 5(c), respectively. Figure 6(a) corresponds to the current pattern for the case that there is one propagating mode in the sample. The current pattern clearly shows that the Lorentz force allows electrons to move in a meandering way. At sufficiently high magnetic fields, the current flow displays a complex pattern. Figure 6(b) shows the current distribution for an electric superlattice with sinusoidal barriers. Notice that in this case, the current flows straight across the cell. This illustrates that electron transport in this electric superlattice is equivalent to a one-dimensional system, in contrast to the magnetic one, in which it is not.



(a)



(b)

FIGURE 6 The current density distributions (a) and (b) corresponding to the cases labelled A and B indicated in Figures 5(a) and 5(c), respectively.

In conclusion, we have studied the transport properties of a quantum wire in spatially periodic magnetic and electrostatic fields. We found a conductance dip, an example of a Fano resonance, just below each mode threshold. These conductance dips shift to higher values of $k_F W/\pi$ when

the background field is increased. As the length a of the period increases, the conductance dips broaden and the dips for higher modes become more pronounced. The conductance dips are quite robust at low temperatures. We have shown that many features of electronic transport in 1D *magnetic* superlattices differ from those of the electrostatically modulated ones. We find that unlike the electric superlattice, the number of oscillations in the first conduction plateau has no simple direct correspondence to the number of unit cells N. The conductance of 1D magnetic superlattices does not decrease monotonically with increasing B_m but rather steps up and down between quantized levels, sometimes going to zero. The spatial distributions of current density show dramatic differences between the magnetic and electric superlattices.

This complex behaviour suggests that the experimental study of 1D *magnetic* superlattices should be very interesting. Several methods are available to produce such devices, including the laying down of patterned superconducting strips on top of a gated device.

Acknowledgements

Work supported by NSERC research grant SAPOO-3198.

References

[1] Carmona, H. A., Geim, A. K., Nogaret, A., Main, P. C., Foster, T. J., Henini, M., Beaumont, S. P. and Blamire, M. G. (1995). "Two-dimensional Electrons in a Lateral Magnetic Superlattice", *Phys. Rev. Lett.*, **74**, 3009–3012.

[2] Ye, P. D., Weiss, D., Gerhardts, R. R., Seeger, M., von Klitzing, K., Eberl, K. and Nickel, H. (1995). "Electron in a Periodic Magnetic Field Induced by a Regular Array of Micromagnets", *Phys. Rev. Lett.*, **74**, 3013–3016.

[3] Izawa, S., Katsumoto, S., Endo, A. and Iye, Y. (1995). "Two-dimensional Electrons in Lateral Magnetic Superlattice", *J. Phys. Soc. Jpn.*, **64**, 706–710.

[4] Leadbeater, L., Foden, C. L., Burke, T. M., Burroughes, J. H., Grimshaw, M. P., Ritchie, D. A., Wang, L. L. and Pepper, M. (1995). "Electron Transport in a Non-Uniform Magnetic Field", *J. Phys. Condens. Matter*, **7**, L307–L315.

[5] Ji, Zhen-Li and Sprung, D. W. L. (1996). "Electronic Transport in Quantum Wires in Nonuniform Magnetic Fields", *Phys. Rev. B*, **54**, 8044–8048.

[6] Ji, Zhen-Li and Sprung, D. W. L. (1997). "Transport Properties of Quantum Wires in Spatially Periodic Magnetic and Electrostatic Fields", *Semicond. Sci. Technol.*, **12**, 529–534.

[7] Tekman, E. and Bagwell, P. F. (1993). "Fano Resonances in Quasi-one-dimensional Electron Waveguides", *Phys. Rev. B*, **48**, 2553–2559.

[8] Berthod, C., Gagel, F. and Maschke, K. (1994). "DC Transport in perturbed multichannel quantum wires", *Phys. Rev. B*, **50**, 18299–18311.

[9] Fano, U. (1961). "Effects of Configuration Interaction on Intensities and Phase Shifts", *Phys. Rev.*, **124**, 1866–1870.

[10] Ji, Zhen-Li (1993). "Ballistic Transport through a Double Bend in an Electron Waveguide", *J. Appl. Phys.*, **73**, 4468–4472.

[11] Kouwenhoven, L. P., Hekking, F. W. J., van Wees, B. J., Harmans, C. J. P. M., Timmering, C. E. and Foxon, C. T. (1990). "Transport through a Finite One-dimensional Crystal", *Phys. Rev. Lett.*, **65**, 361–364.

[12] Leng, M. and Lent, C. S. (1993). "Recovery of Quantized Ballistic Conductance in a Periodically Modulated Channel", *Phys. Rev. Lett.* **71**, 137–140.

[13] Ibrahim, I. S. and Peeters, F. M. (1995). "Two-dimensional Electrons in Lateral Magnetic Superlattices", *Phys. Rev. B*, **52**, 17321–17334.

[14] Takagaki, Y. and Ploog, K. (1995). "Tunneling Transmission in Quantum Wires in the Presence of Magnetic Fields", *Phys. Rev. B*, **51**, 7017–7023.

[15] Yoshioka, H. and Nagaoka, Y. (1990). "Effect of Magnetic Field on the Conductance of Ballistic Quantum Wire", *J. Phys. Soc. Jpn.*, **59**, 2884–2892.

Authors' Biographies

Zhen-Li Ji received a Licentiate of Engineering degree in 1991 and a Ph.D. degree in 1993 from Linköping University. He remained at Linköping for two years of postdoctoral research experience. He is currently working in the Dept. of Physics and Astronomy, McMaster University as a postdoctoral fellow. His research interests include electronic transport properties in low-dimensional semiconductor nanostructures, such as ballistic transport, resonant tunnelling, conductance fluctuations, chaotic behaviour, impurities, geometrical effects, quantum Hall effect and nonlinear transport in semiconductor nanostructures.

D.W.L. Sprung received a Ph.D. in Mathematical Physics in 1961 from the University of Birmingham, with R. E. Peierls. Following a post-doctoral year with H. A. Bethe at Cornell he joined the faculty of McMaster University

where he has chaired the Departments of Physics and Astronomy and Applied Mathematics and has served as Dean of the Faculty of Science for two terms. His research interests have ranged over nuclear structure theory, few-body problems and quantum nanostructures.

VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 567–572
Reprints available directly from the publisher
Photocopying permitted by license only

© 1998 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in India.

Boundary Condition for the Modeling of Open-circuited Devices in Non-equilibrium

JOSEPH W. PARKS JR. and KEVIN F. BRENNAN*

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250

A boundary condition specifically designed to model open-circuited devices in a macroscopic device simulator is introduced. Other simulation techniques have relied on an external circuit model to regulate the current flow out of a contact thus allowing the potential to remain the controlled variable at the boundary. The limitations of these methods become apparent when modeling open-circuited devices with an exceptionally small or zero output current. In this case, using a standard ohmic-type Dirichlet boundary condition would not yield satisfactory results and attaching the device to an arbitrarily large load resistance is physically and numerically unacceptable. This proposed condition is a true current controlled boundary where the external current is the specified parameter rather than the potential. Using this model, the external current is disseminated into electron and hole components relative to their respective concentration densities at the contact. This model also allows for the inclusion of trapped interface charge and a finite surface recombination velocity at the contact.

An example of the use of this boundary condition is performed by modeling a silicon avalanche photodiode operating in the flux integrating mode for use in an imaging system. In this example, the device is biased in steady-state to just below the breakdown voltage and then open-circuited. The recovery of the isolated photodiode back to its equilibrium condition is then determined by the generation lifetime of the material, the quantity of signal and background radiation incident upon the device, and the impact ionization rates.

Keywords: Current boundary condition, macroscopic simulation, charge storage, avalanche photodiodes

1. INTRODUCTION

Many macroscopic device simulators provide for a connection to an external circuit primarily through either ohmic or Schottky contacts. In these cases, the system variables are constructed such that the electrostatic potential is the independent parameter describing the boundary condition, and

*Corresponding author. Email: kbrennan@ece.gatech.edu.

other values such as the current flowing through the contact are calculated in a post-processing step [1, 2]. These boundary conditions are adequate for situations where the device can be approximated as a voltage controlled current source; however, instances exist where this may not be the case. Most notably this occurs when the energy stored within the internal capacitance of a device becomes the dominant controlling factor for the prediction of the potential across the device. In these cases, the use of a current controlled boundary is more applicable.

A situation requiring a current controlled contact exists when simulating the recovery to equilibrium of an open circuited device. Such a condition occurs when examining the performance of an avalanche photodiode configured in the charge storage mode for imaging applications [3, 4]. In this configuration, the sensor is periodically reset by applying a large reverse bias to the device. Following the reset, the device is open-circuited and begins to equilibrate through carrier generation. If the device is used as a photodetector, light shining upon the structure generates excess charge which is then stored within the depletion region of the device. At the end of the integration cycle, the quantity of stored charge within the device is used as an indication of the total illumination upon the pixel [5].

The present work introduces a current controlled boundary condition which is capable of simulating an electrically isolated photodiode. With this model, a steady-state voltage is applied between the ohmic contacts using the traditional Dirichlet boundary condition. The device is then isolated from the external circuit. The voltage across the device is no longer fixed invalidating the use of a voltage controlled boundary condition. Alternatively, a current controlled boundary condition must be employed to describe the device which includes effects such as surface recombination, trapped surface charge, and leakage current from the read-out electronics. Application of this model to a generic APD is also presented.

2. BOUNDARY CONDITION

A current controlled boundary condition is required to accurately model the reverse recovery within a photodiode. When the diode is disconnected from the external circuit or driving bias, the voltage at the cathode floats. The metal contact can no longer be treated in its usual manner as an infinite source of carriers and the assumption that the electron and hole concentrations are dictated by their equilibrium values may not be valid. Moreover, the voltage across the device is determined solely by the distribution of charge and the internal capacitance of the diode and not by the external circuit. Simulation techniques which operate by allowing the potential to remain the independent variable, while acceptable for modeling switching characteristics, are inappropriate for cases with either very low or zero external current. Attempting to model the open-circuited device with the standard ohmic condition and a very large shunt load resistance is unacceptable on physical as well as numerical grounds.

The primary concern with developing the current controlled model lies in the apportioning of the electron and hole fluxes from the external current at the contact. It is assumed here that the total current is divided into the electron and hole currents relative to their respective densities at the contact. Thus, the relation between the currents is given by,

$$\frac{n}{n+p} \cdot J_p - \frac{p}{n+p} \cdot J_n = qR_{\text{surf}} \qquad (1)$$

where J_n and J_p are the electron and hole currents leaving the contact as indicated in the illustration of the discretized control volume as shown in Figure 1 [6]. The carrier concentrations are taken to be those of the neighboring control volume within the device. R_{surf} is the recombination rate owing to surface states at the contact [7] using the standard relation:

$$R_{\text{surf}} = \frac{S_n S_p (np - n_i^2)}{S_n(n + n_1) + S_p(p + p_1)} \cdot \delta(x). \qquad (2)$$

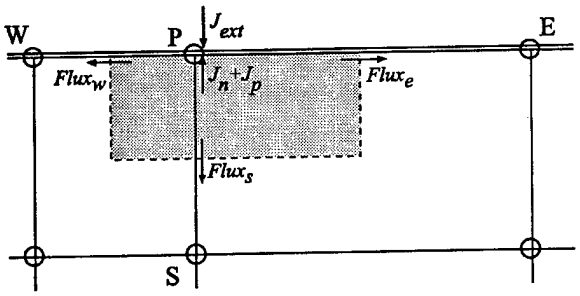


FIGURE 1 Illustration of the sample control volume at the surface of the simulation domain. Current continuity requires the conservation of flux within the control volume. Here, the current boundary is entered directly into the formulation where the sum of the electron and hole currents balance that of the external current.

where S_n and S_p are electron and hole recombination velocities. Additionally, Eqs. (3) and (4) below require that the carrier currents sum to the external current and that the band bending at the contact is proportional to the trapped interface charge.

$$J_n + J_p = J_{\text{ext}} \tag{3}$$

$$\varepsilon \frac{\partial \Psi}{\partial n} = Q_{\text{int}} \tag{4}$$

Combining Eqs. (1) and (3) yields expressions setting the values of the current into a contact.

$$J_p = qR_{\text{surf}} + \frac{p}{n+p} \cdot J_{\text{ext}} \tag{5}$$

$$J_p = -qR_{\text{surf}} + \frac{n}{n+p} \cdot J_{\text{ext}} \tag{6}$$

These currents are used directly within the control volume formulation of the discretized continuity equations [8]. Furthermore, partial derivatives of the current expressions with respect to the system variables are analytically obtained thus leading to increased convergence of the non-linear system.

A two-dimensional analysis of the photodiode is necessary to study the full carrier transport of effects such as premature edge breakdown and pixel crosstalk; however, this study is concerned solely with the examination of the charge storage capabilities of the APD. Therefore, a one-dimensional analysis is satisfactory. This greatly simplifies the model by allowing the external current density, J_{ext}, to be directly proportional to the cross sectional area of the device contact and the total current. Multi-dimensional extensions of this boundary can be produced by either imposing an additional constraint upon how the output currents sum over the range of the contact such as equipartitioning or by solving Gauss' law describing the current flow through the metal comprising the contact [9].

To study the charge storage of the photodiodes, the current boundary condition is incorporated into a drift-diffusion simulator which self-consistently solves Poisson's equation and the current continuity equations [10]. Since the smallest dimensions of the device structures are on the order of microns, it is believed that hot carrier transport can be neglected in these devices and that a full hydrodynamic simulation is unnecessary. Within this model, the generation-recombination rates include terms for SRH, Auger, and radiative recombination, impact ionization, and wavelength dependent photoillumination. Additionally, the standard field dependent mobility and impact ionization models for silicon are incorporated into the simulator [11]. By using a completely numerical model, many of the non-linear attributes of the carrier transport of a photodiode in charge storage mode can be included without the use of many of the simplifying approximations typically employed [3, 4].

3. EXAMINATION OF CHARGE STORAGE IN PHOTODIODES

To demonstrate the utility of the current controlled boundary condition, the reverse recovery of an avalanche photodiode designed for night-sky imaging is examined. Here the device, shown in Figure 2, is a reach-through avalanche photodiode 50 μm in length. During the initial reset period, a

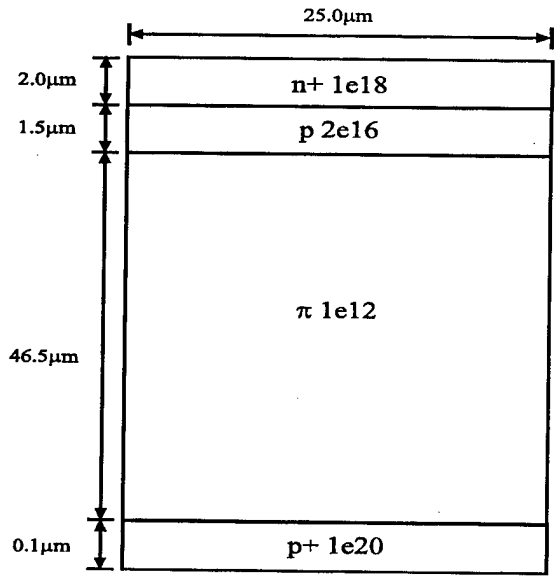


FIGURE 2 Doping profile and geometry of reach-through avalanche photodiode used to study the current boundary condition.

large electric field is established in the *p*-type multiplication region and a modest drift field exists in the intrinsic region. The illumination spectrum used in this study is that of a cloudless, moonless, night sky with a total illumination of 5.4×10^{16} photons per cm^2 over a range of wavelengths from 0.3 to 1.1 microns [12]. The total internal quantum efficiency for this spectrum is approximately 55%.

If the photodetector was continuously biased, the photo-generated carriers would be swept across the depletion region, collected at the device contacts, and interact directly with the external circuit. In the charge storage mode of operation, the connection to the external circuit is removed. In this case, the internally stored electric field still causes the carriers to be swept across the depletion region where they may impact ionize. However, the charge is stored near the edges of the depletion region as opposed to interacting with the external circuit. This accumulated charge counteracts the depleted charge from the reset and tends to drive the reverse recovery. Additional factors which promote the recovery of the diode include the leakage current inherent in the readout electronics, J_{ext}, the dark current due primarily to thermal generation of carriers from SRH centers, and surface recombination.

Figure 3 shows the dynamics of the charge storage within the APD for various photoillumination intensities. For this case, intensities ranging from 0.05 to 500 times the nominal night sky spectrum are considered (2.7×10^{15}– 2.7×10^{19} photons per cm^2 per second). The initial bias on the device is established to produce an initial steady-state gain of fifteen. The first part of the voltage recovery is dominated by the influence of holes filling the intrinsic region resulting in a linear decay of internal bias with only a very small degradation in gain. Once the drift region has been filled, the charge is stored at the edge of the high field multiplication region giving rise to the expected parabolic voltage recovery. A non-linear decrease in the internal gain is observed as the high field region recovers. Under extreme illumination conditions, the diode completely recovers to its equilibrium level and begins to become forward biased. Recombination processes prevent the diode from continued storage of charge once the diode becomes sufficiently forward biased.

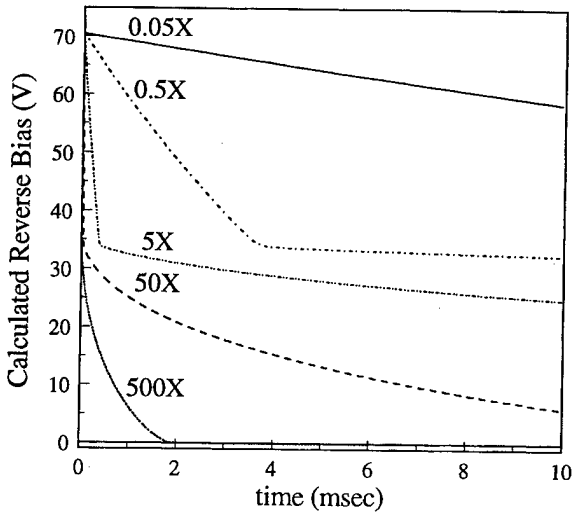


FIGURE 3 Calculated reverse bias recovery of the diode shown in Figure 2 as a function of integration time. The different curves represent separate values of photoillumination.

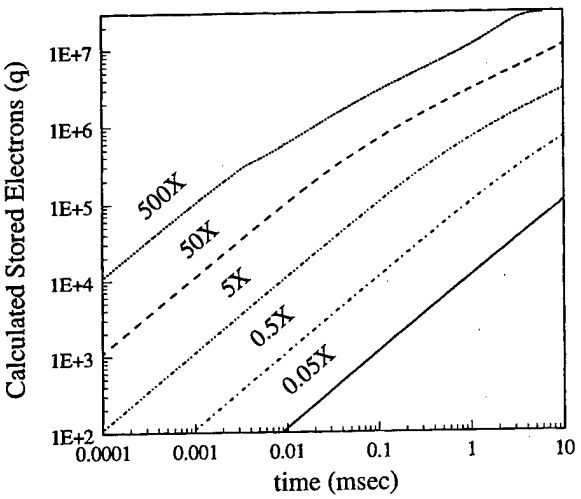


FIGURE 4 Calculated stored electron charge of the device shown in Figure 2 as a function of integration time for various illumination intensities.

Typically, the actual stored charge rather than the device bias is quantified at the end of the integration period to determine the level of photoillumination [5]. Figure 4 illustrates the total electron charge stored within the diode as a function of the integration time. As expected, the first part of the integration gives rise to a linear collection of charge with a constant gain. As the amount of charge increases, the gain begins to saturate with a square root dependence [10, 13]. Once the device has recovered to the point where internal gain becomes negligible, the linear recovery is again observed. As the diode becomes significantly forward biased, as in an open-circuited solar cell, the charge saturates.

4. CONCLUSIONS

A current controlled boundary condition applicable to macroscopic device simulators is presented. This boundary type establishes the systems variables such that the external current becomes the independent quantity at the contact, thus allowing the contact potential to be self-consistently set within the simulation. As an example of this model, a reach-through avalanche photodiode, operating in the charge storage mode, is examined. Here it is observed that the model allows one to capture the full transient recovery of the device from the initial reset point all the way through the forward bias condition in the case of extreme photoillumination.

References

[1] Iwamuro, N. and Tagami, S. (1993). "Stable Calculation for Power Device Simulation with Inductive Load Circuit: Application of an Integral Method", *Microelec. Jour.*, **24**, 139–145.

[2] Kausel, W., Nylander, J. O., Nanz, G., Selberherr, S. and Poetzl, H. (1990). "BAMBI- A Transient 2D-MESFET Model with General Boundary Conditions Including Schottky and Current Controlled Contacts", *Microelec. Jour.*, **21**(5), 5–21.

[3] Weckler, G. (1967). "Operation of *p-n* Junction Photodetectors in a Photon Flux Integrating Mode", *IEEE J. Solid State Cir.*, **SC-2**(3), 65–73.

[4] Komobuchi, H. and Ando, T. (1990). "A Novel High-Gain Image Sensor Cell Based on Si *p-n* APD in Charge Storage Mode Operation", *IEEE Trans. Elec. Dev.*, **37**(8), 1861–1868.

[5] Mendis, S., Kemeny, S. E. and Fossum, E. R. "CMOS Active Pixel Image Sensor", *IEEE Trans. Elec. Dev.*, **41**(3), 452–453.

[6] Basore, P. A. and Rover, D. T. (1985). "PC-1D Installation Manual and User's Guide", Iowa State University Research Foundation.

[7] Green, M. A. (1982). *Solar Cells*, Prentice-Hall, Englewood Cliffs, NJ, 1982.

[8] Pantankar, S. V. (1980). *Numerical Heat Transfer and Fluid Flow*, Hemisphere Publishing Corp., New York.

[9] Gough, P. A., Johnson, M. K., Higgins, S. A., Slatter, J. A. G. and Whight, K. R. (1987). "Two Dimensional Simulations of Power Devices with Circuit Boundary Conditions", NASCODE V, 213–8.

[10] Parks, J. W., Smith, A. W., Brennan, K. F. and Tarof, L. E., "Theoretical Study of Device Sensitivity and Gain Saturation of Separate Absorption, Grading, Charge, and Multiplication InP/InGaAs Avalanche Photodiodes", *IEEE Trans. Elec. Dev.*, **43**(12), 2113–2121, Dec. 1996.

[11] Selberherr, S. (1984). *Analysis and Simulation of Semiconductor Devices*, Springer-Verlag, Wein, New York.

[12] Engstrom, R. W. and Rodgers, R. L., "Camera Tubes for Night Vision", *Optical Spectra*, **26–30**, 438–442, Feb. 1971.

[13] Webb, P. P., McIntyre, R. J. and Conradi, J. (1974). "Properties of Avalanche Photodiodes", *RCA Review*, **35**, 234–278.

Authors' Biographies

Joseph W. Parks Jr. was born in Oak Ridge, TN on May 21, 1970. He received his B.S. degree in

electrical engineering from the University of Tennessee, Knoxville in 1992 and his M.S.E.E. from the Georgia Institute of Technology, Atlanta, GA, in 1993. He is presently working on his Ph.D. degree also in electrical engineering at the Georgia Institute of Technology. His research work involves the numerical modeling of semiconductor devices with emphasis in the drift-diffusion and hydrodynamic simulation of photodetectors and avalanche photodiodes.

Kevin F. Brennan received the B.S. degree in physics from the Massachusetts Institute of Technology, Cambridge, in 1978, and the M.S. degree in physics and Ph.D. degree in electrical engineering from the University of Illinois, Urbana-Champaign, in 1984.

He is currently Institute Fellow and Professor, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta. His current research interests include the physics and modeling of semiconductor devices. Of particular interest are the physics and modeling of avalanche photodiodes, confined state ionization devices, high field effects in semiconductors, photoconductors, and high speed transistors.

Dr. Brennan was the recipient of a Presidential Young Investigator Award through the National Science Foundation.

VLSI DESIGN

NOTES FOR CONTRIBUTORS

As Gordon and Breach moves into the delivery of journals in electronic format, it is essential that authors prepare their manuscripts according to established specifications. The effectiveness of the search capabilities offered by electronic delivery will depend upon the care used by authors in preparing their manuscripts. Therefore, contributors are strongly encouraged to follow these instructions carefully while preparing a manuscript for submission.

Gordon and Breach online "offices" enable our authors to submit abstracts and full-text articles in virtually any file format, from anywhere in the world, at any time. Visit http://www.gbhap.com/VLSI_Design/ to access a full range of internet author services.

The editors welcome theoretical and practice-oriented articles, technical notes (of 10–15 manuscript pages including artwork), and timely book reviews. Enterprising individuals are encouraged to submit proposals to the editor in chief to be guest editors for special issues on important and current VLSI topics. Those interested in serving as manuscript reviewers should contact the editor.

MANUSCRIPT SUBMISSION

Please submit three copies of the manuscript along with copies of camera-ready artwork (retain diskettes, permissions, and original artwork until post-review stage), to: **Dr. George W. Zobrist**, Editor in Chief, **VLSI DESIGN**, Department of Computer Science, University of Missouri at Rolla, 325 Math/Computer Science Building, 1870 Miner Circle Drive, Rolla, Missouri 65409, USA. Please include *complete* postal addresses, email addresses, and telephone numbers for all authors of an article. On acceptance of a paper, the author must forward to the editor both a diskette containing the manuscript and a hardcopy (in triplicate).

Submission of a paper to this journal will be taken to imply that it represents original work not previously published, that it is not being considered for publication elsewhere, and that, if accepted for publication, it will not be published elsewhere in the same form, in any language, without the consent of the Editor in Chief and publisher. It is a condition of acceptance by the Editor in Chief of a manuscript that the publisher acquires automatically the copyright of the manuscript throughout the world.

Papers, previously circulated as reports or as conference proceedings, but not available to a wide audience, will be considered. However, the author must disclose such previous history and obtain any necessary copyright clearances.

MANUSCRIPT PREPARATION

Please prepare the manuscript typed double-spaced, with one-inch (2.5 cm) margins left and right, top and bottom. Keep the manuscript layout as simple as possible. When preparing the manuscript on disk, please avoid the use of word-processing program-specific features that would need to be removed later by the typesetter. See the back of the journal for specific instructions on article submission on diskette.

The manuscript should consist of type and equations only, with all tables, line drawings, photographs, and examples of computer printouts grouped together as artwork at the end of the text.

In preparing any given manuscript page, please minimize the use of footnotes and avoid lengthy heading designations (for example: simply use 3, rather than "Smith, page 3," at the top of the page).

Permissions

If your paper includes material (quotations of about 60 or more words, tables, or figures) that has been taken exactly from, or has been adapted from, a published source, *you must:*

may be typewritten on a separate page, with their designated positions specified in the text. For manuscripts with extensive equations, T_EX is the preferred word processing program. To simplify typesetting of non-T_EX files, please use: (1) the "exp" form of complex exponential functions; (2) fractional exponents instead of root signs; and (3) the solidus (/) to simplify fractions—e.g., exp $x^{1/2}$.

Marking: The typesetter will set mathematical symbols in italics, except for obvious groups like sin and log. Any symbols which are left in roman (upright) type should be circled in pencil in the typescript and bold symbols should be underlined with a wavy line.

Units: Metric units should be used, with inch/pound equivalents in parantheses.

REFERENCES

Indicate references in the text by consecutive arabic numerals enclosed in square brackets, viz. De Pauw [1]. All sources cited in the text must be included in the reference list, and vice versa. The full list of references should appear on a separate page at the end of the paper. Below are examples of entries for a journal, a proceedings, and a book:

[1] B. Moszkowski, "A Temporal Logic for Multilevel Reasoning about Hardware," *Computer Magazine*, February 1985.

[2] M.C. McFarland, A.C. Parker and R. Camposano, "Tutorial on High-Level Synthesis," *Proc. of 25th Design Automation Conf.*, 1988.

[3] M. Born and E. Wolf, *Principles of Optics*, Oxford: Pergamon Press, Sixth Ed. 1986.

It is assumed that with the development of the World Wide Web (WWW), authors and/or the publisher will propose distribution of articles or parts of articles on the WWW. If the HTTP address of a referenced article is known, please include it at the end of the reference in the following style:

< www http://www.blouk.com/article.html >

where http://www.blouk.com/article.html is the HTTP address.

BIOGRAPHY

At the end of the manuscript please include a one-paragraph biographical sketch, **100 words maximum**, highlighting your education, experience, and other qualifications for writing this particular paper. There should be one biographical sketch for each author. Material submitted may be edited by the publisher.

The following information is needed (for each author): (1) Current research interests, or what is currently being worked on; (2) Pertinent extracurricular activities—for example, do you sit on any committees or have you chaired any symposia?; (3) Books or articles that you have written/are writing that are especially noteworthy; (4) Professional memberships or honors; (5) Current position; (6) Any electronic network connection (i.e., Bitnet).

SAMPLE

PRADIP K. SRIMANI has served on the faculty of the Indian Statistical Institute, Calcutta; Gesselschaft fuer Mathematik und Dalenverarbeitung, Bonn, Germany; the Indian Institute of Management, Calcutta; and Southern Illinois University, Carbondale. Currently, he is a Professor of Computer Science at Colorado State University in Ft. Collins. His research interests include parallel algorithms, distributed operating systems, software reliability, fault-tolerant computing, networks and graph theory applications. He is a senior member of the IEEE, and member of ACM, IEEE CS Press Editorial Board and the Editorial Board of the International Journal of Computer Simulation.

words capitalized. Underline the heading and start the text on the next line. For third-level headings, only the first letter should be a capital. Underline, then begin text after three spaces.

FIRST-LEVEL TEXT HEADINGS

Second – Level Text Headings

Third – level headings With text run on

FIGURES AND CAPTIONS

Please do not intersperse figures (line drawings, photos, computer printouts) throughout the text, but group them together at the end of the manuscript. Limit figures to one per page.

Number all figures consecutively with arabic numerals on the back of the camera-ready drawings, and include the author's name (Smith, Figure 1; Smith, Figure 2; etc.) and paper title (abbreviated). Please indicate "top" where this is not obvious. Be certain to refer to each figure in the text, e.g., (See Figure 2) as this is essential to efficient searching of articles on electronic delivery. In addition, please indicate where the typesetter should position artwork by marking and circling the first call-out to each figure in the manuscript margin.

Please include a separate sheet containing a list of figure captions which should be substantively one or two sentences. Captions will be set in type by the typesetter.

The original figures and one copy should be sent when the article is accepted.

Preparation

Figures submitted must be of a high enough standard for direct reproduction, with clear, sharp, black lines on a white background. Figures should be drawn on smooth white paper or computer generated and reproduced by a high-quality laser printer or plotter. Letters and numbers should be drawn professionally with a LeRoy Lettering Set and with clear black ink. No additional redrawing of artwork will be done. Note that the smallest correction will require reshooting. Therefore, check carefully for accuracy to avoid this costly and time-consuming procedure.

Photographs: Photographs intended for halftone reproduction must be good glossy original prints, of maximum contrast.

Computer Printouts: A computer printout of more than several lines should not be embedded in the manuscript, but treated as a figure. The printout should be labeled consecutively along with the line drawings (e.g., "Figure 4"), be referred to in the text, and be placed at the end of the manuscript with the other figures. Because the printout will be photographed as line art, it should be a clear, sharp printout produced on a high-quality laser printer.

Size: Illustrations should be prepared to fit either into a single column width (75 mm) or double column width (160 mm) with capital lettering 4 mm high, to accommodate reduction. Photographs for halftone reproduction should be about twice the desired size.

TABLES

Number tables consecutively with roman numerals, and give each a clear descriptive caption at the top. Be sure to refer to each table in the text; for example, (See Table I), and indicate in the margin where the typesetter should place tables. Limit tables to one per page. Avoid the use of vertical rules in the tables.

MATHEMATICAL EQUATIONS

Wherever possible, mathematical equations should be entered in the word processing program, with unusual or ambiguous symbols identified in the margin where they first occur. Alternatively, equations

1. write to the copyright owner (usually the publisher) and request permission to republish. The permission must be obtained in writing;
2. forward the original permission letter to the Editor in Chief, Dr. George W. Zobrist, immediately upon receipt and retain a copy for your own records;
3. credit the source (e.g., Reprinted from...: Adapted from...).

ANY TRADEMARKS THAT ARE USED MUST BE LISTED IN THE ACKNOWLEDGMENTS SECTION.

Paper Organization

Front Sheet: Title, Author(s), Affiliation(s), Address(es), Phone No.(s), Email Address(es), FAX No.(s).
Abstract/Key Words
Footnotes/Denote Principal Author for Contact
Text of Manuscript (Mark Figures/Tables/Computer Listing Placement)
Acknowledgement/Trademark
Figure Caption List
Table Caption List
Computer Listing List
Original Figures (1 to a Page)
Original Tables (1 to a Page)
Computer Listings
REMEMBER THE ORIGINALS OF THE FIGURES/TABLES/COMPUTER PRINTOUTS ARE *NOT* TO BE PLACED WITHIN THE MANUSCRIPT, BUT MUST BE PLACED AT THE END OF THE PAPER, ONE TO A PAGE. THERE ALSO MUST BE A FIGURE/TABLE/COMPUTER CAPTION LIST IN FRONT OF THE ORIGINALS.

Language: English only please.

Length: The text of a full-length manuscript should be between 15 and 40 double-spaced pages, plus artwork grouped at the end. The journal will also accept short technical notes of 10–15 manuscript pages, including artwork.

Abstract: On the first page of the manuscript, following the title and author byline (name, title, organization, city, state, country) please include a 100 to 150 word abstract summarizing your problem and findings.

The abstract should be as informative as possible, consisting of short, direct, and complete sentences. Length must be less than 150 words. The abstract should state the objectives of the work, summarize the results, and give the principal conclusions and recommendations. It should state clearly whether the focus is on theoretical developments or on practical questions, and whether subject matter or method is emphasized. The title need not be repeated. Work planned should not be described in the abstract. Because abstracts are extracted from a paper and used separately, one should not use the first person, display mathematics, nor use citation reference numbers. Avoid starting with the words "This paper...."

Keywords: On a line below the abstract please include six key words for indexing and database word searches.

Text Headings

Set first-level headings in the text to the left, typed all in capitals (upper case); begin the text on the following line. Second-level headings should be typed in small (lower case) letters but with all the main

FOOTNOTES

Authors are encouraged to minimize the use of footnotes. A footnote may include the designation of a corresponding author of the paper, current address information for an author (if different from that shown in the affiliation), and traditional footnote content. Information concerning grant support of research should appear in a separate Acknowledgments section at the end of the paper, not in a footnote. Acknowledgments of the assistance of colleagues or similar notes of appreciation also properly belong in an Acknowledgments section, not in footnotes.

Footnotes should be indicated in the text by the following symbols: * (asterisk or star), † (dagger), ‡ (double dagger), ¶ (paragraph mark), § (section mark), || (parallels), # (number sign). Do not use numerals for footnote call-outs, as they may be mistaken for bibliographical reference call-outs or exponents. Type each footnote at the bottom of the typescript page on which its text call-out appears.

Footnotes within a table should be indicated by the same symbols listed above. Reinitialize symbol sequence within tables.

SUBMISSION OF ARTICLES ON DISKETTE

The publisher encourages authors to submit accepted manuscripts on computer diskette. Authors must forward to the Editor in Chief a hard copy of the manuscript, in triplicate, along with the diskettes. The diskette files and hard copy must match exactly. Please refer to the back of the journal for instructions on submission of articles on diskette. These are also available from the journal Editor in Chief or the publisher.

SUBMISSION OF ANIMATION

Author-supplied animation related to articles accepted for publication will be included in the journal CD-ROM and published on the World Wide Web at no cost to authors. Animations are limited to a time duration of 30 seconds. Animation should be submitted to the journal Editor in Chief with the final manuscript, after it has completed the refereeing process.

Animations in the following forms (in order of preference) can be accepted from authors:

- Video tape
- AVI or QuickTime files
- A sequence of still images

The following formats can be accepted:

- all uncompressed formats widely used on PC, Mac and UNIX
- JPEG for colored and compressed images
- TIFF with a group IV compression for black and white compressed images
- FLI and FLC format from AutoDesk

Authors who submit animations are requested to provide the following information:

- Video tape - format used
- AVI or QuickTime files - version and system used for disk file creation
- Sequence of still images - format used, version and system used for disk file creation

Authors who are unable to supply video tape, AVI or QuickTime files may provide the publisher with a set of sequential still images. Note than an animated sequence will consist of 13 to 15 still images per second of animation; e.g., if an animated sequence is 10 seconds in duration, it is made up of 130 images. Authors who are unable to submit in any of the above-mentioned formats are advised to contact the publisher to discuss other options prior to submission.

Animation should be mentioned in the text. Indicate an approximate location for the animation call-out in the margin.

PROOFS

The first-named author will be airmailed proofs (including figures) for correction. These must be airmailed back to the typesetter within 48 hours of receipt. Please include a full postal address on the first page of the manuscript so proofs are not delayed. Author's alterations in excess of 10% of the original composition cost will be charged to the author.

REPRINTS

Twenty-five complimentary reprints will be sent to the first-named author of each paper. Additional reprints may be ordered by completing the appropriate form sent with proofs.

PAGE CHARGES

There are no page charges to individuals or to institutions. See the journal inside back cover for information on the publisher's negative page charge voucher and voluntary page charge programs.

A FINAL NOTE

Typescripts that do not conform to the required standard of preparation for submission outlined here will be returned to the authors for correction before review.

INSTRUCTIONS FOR AUTHORS

ARTICLE SUBMISSION ON DISK

The Publisher welcomes submissions on disk. The instructions that follow are intended for use by authors whose articles have been accepted for publication and are in final form. Your adherence to these guidelines will facilitate the processing of your disk by the typesetter. These instructions do not replace the journal Notes for Contributors; all information in Notes for Contributors remains in effect.

When typing your article, do not include design or formatting information. Type all text flush left, unjustified and without hyphenation. Do not use indents, tabs or multi-spacing. If an indent is required, please note it by a line space; also mark the position of the indent on the hard copy manuscript. Indicate the beginning of a new paragraph by typing a line space. Leave one space at the end of a sentence, after a comma or other punctuation mark, and before an opening parenthesis. Be sure not to confuse lower case letter "l" with numeral "1", or capital letter "O" with numeral "0". Distinguish opening quotes from close quotes. Do not use automatic page numbering or running heads.

Tables and displayed equations may have to be rekeyed by the typesetter from your hard copy manuscript. Refer to the journal Notes for Contributors for style for Greek characters, variables, vectors, etc.

Articles prepared on most word processors are acceptable. If you have imported equations and/or scientific symbols into your article from another program, please provide details of the program used and the procedures you followed. If you have used macros that you have created, please include them as well.

You may supply illustrations that are available in an electronic format on a separate disk. Please clearly indicate on the disk the file format and/or program used to produce them, and supply a high-quality hard copy of each illustration as well.

Submit your disk when you submit your final hard copy manuscript. The disk file and hard copy must match exactly.

If you are submitting more than one disk, please number each disk. Please mark each disk with the journal title, author name, abbreviated article title and file names.

Be sure to retain a back-up copy of each disk submitted. Pack your disk carefully to avoid damage in shipping, and submit it with your hard copy manuscript and complete Disk Specifications form (see reverse) to the person designated in the journal Notes for Contributors.

DISK SPECIFICATIONS

Journal name _____ **Date** _____

Paper title _____

_____ **Paper Reference Number** _____

Corresponding author _____

Address _____

_____

_____ **Postcode** _____

Telephone _____ **Fax** _____ **E-mail** _____

Disks Enclosed (file names and descriptions of contents)

Text

Disk 1 _____

Disk 2 _____

Disk 3 _____

Figures

Disk 1 _____

Disk 2 _____

Disk 3 _____

Computer make and model _____

Size/format of floppy disks

☐ 3.5" ☐ 5.25"

☐ Single sided ☐ Double sided

☐ Single density ☐ Double density ☐ High density

Operating system _____ **Version** _____

Word processor program _____ **Version** _____

Imported maths/science program _____ **Version** _____

Graphics program _____ **Version** _____

Files have been saved in the following format

Text: _____

Figure: _____

Maths: _____

PLEASE RETAIN A BACK-UP COPY OF ALL DISK FILES SUBMITTED.

(Continued from inside front cover)

World Wide Web Addresses
Additional information is also available through the Publisher's web home page site at http://www.gbhap.com. Full text on-line access and electronic author submissions may also be available.

Editorial enquiries by e-mail: <editlink@gbhap.com>.

Ordering Information

Four issues per volume. Subscriptions are renewed on an annual basis. 1998 Volume(s): 7–8

Orders may be placed with your usual supplier or at one of the addresses shown below. Journal subscriptions are sold on a per volume basis only; single issues of the current volume are not available separately. Claims for nonreceipt of issues will be honored, if made within three months of publication of the issue. **See Publication Schedule Information**. Subscriptions are available for microform editions; details will be furnished upon request.

All issues are dispatched by airmail throughout the world.

Subscription Rates Base list subscription price per volume: ECU 99.00.* This price is available only to individuals whose library subscribes to the journal OR who warrant that the journal is for their own use and provide a home address for mailing. Orders must be sent directly to the Publisher and payment must be made by personal check or credit card.

Separate rates apply to academic and corporate/government institutions. Postage and handling charges are extra.

*ECU (European Currency Unit) is the worldwide base list currency rate; payment can be made by draft drawn on ECU currency in the amount shown or in subscriber's local currency at the current conversion rate set by Publisher. Subscribers should contact their agents or the Publisher. All prices are subject to change without notice.

Publication Schedule Information To ensure your collection is up-to-date, please call the following numbers for information about the latest issue published: +44 (0) 118-956-0080 ext. 391; +1 973 643-7500 ext. 290; or web site: http://www.gbhap.com.reader.htm.

Note: If you have a rotary phone, please call our *Customer Service* at the numbers listed below.

Orders should be placed through one of the addresses below:

IPD Marketing Services
PO Box 310
Queen's House, Don Road
St. Helier, Jersey
Channel Islands JE4 0TH
Telephone: +44 (0) 118-956-0080
Fax: +44 (0) 118-956-8211

PO Box 32160
Newark, NJ 07102 USA
Telephone: +1 800 545-8398
Fax: +1 973 643-7676

Kent Ridge, PO Box 1180
Singapore 911106
Republic of Singapore
Telephone: +65 741-6933
Fax: +65 741-6922

Yohan Western Publications Distribution Agency
3-14-9, Okubo, Shinjuku-ku
Tokyo 169-0072, Japan
Telephone: +81 3 3208-0186
Fax: +81 3 3208-5308

Enquiries can also be sent by e-mail: <info@gbhap.com> and the world wide web: http://www.gbhap.com.

Rights and Permissions/Reprints of Individual Articles

Permission to reproduce and/or translate material contained in this journal must be obtained in writing from the Publisher.

This publication and each of the articles contained herein are protected by copyright. Except as allowed under national "fair use" laws, copying is not permitted by any means or for any purpose, such as for distribution to any third party (whether by sale, loan, gift, or otherwise); as agent (express or implied) of any third party; for purposes of advertising or promotion; or to create collective or derivative works. A photocopy license is available from the Publisher for institutional subscribers that need to make multiple copies of single articles for internal study or research purposes. Any unauthorized reproduction, transmission or storage may result in civil or criminal liability.

Copies of articles may be ordered through SCAN, the Publisher's own document delivery service. SCAN provides customers with the current contents and abstracts to all Gordon and Breach and Harwood Academic journals. Please contact one of the addresses listed above to receive SCAN, or view current contents and abstracts directly on the Web at http://www.gbhap.com, and for ordering.

The Publisher is also a member of the Copyright Clearance Center.

The Canada Institute for Scientific and Technical Information **(CISTI)** provides a comprehensive, worldwide document delivery service for Gordon and Breach/Harwood Academic journals. For more information, or to place an order for a copyright-cleared Gordon and Breach/Harwood Academic document, please contact Client Assistant, Document Delivery, CISTI, National Research Council Canada, Ottawa, Canada, KIA 0S2 (Tel: 613-993-9251; Fax: 613-993-7619; e-mail: <cisti.docdel@nrc.ca>).

Voluntary Page Charges The United States National Science Foundation has extended the allowance of page charge funds for payments to journals regardless of the Publisher's commercial status. Previously this was reserved for only nonprofit society-sponsored journals. Voluntary page charge payments are now accepted for this journal, and authors may elect to pay any amount up to a maximum of $25.00 per page. Payments will be refunded in the form of a voucher at 100% value of total payment. This voucher can be used by the author or the author's university library for any product or service offered by the Publisher, and thereby can be used to supplement library funding.

Article Vouchers The principal author of each article will receive a voucher for his contribution in the amount of ECU 15.00 (US$ 20.00, Yen 3,000), which can also be used to purchase the Publisher's products directly or through university libraries, thereby reducing costs of publications to those authors supporting the journal.

This journal is sold CIF with title passing to the purchaser at the point of shipment in accordance with the laws of The Netherlands. All claims should be made to your agent or the Publisher.

Printed in India.

July 1998