Annual Report on Grant AFOSR-91-0211

Quantum Transport

Craig S. Lent, Wolfgang Porod, Supriyo Bandyopadhyay, and
Gary H. Bernstein
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

Department of
ELECTRICAL ENGINEERING

UNIVERSITY OF NOTRE DAME, NOTRE DAME, INDIANA

Annual Report on Grant AFOSR-91-0211

Quantum Transport

Craig S. Lent, Wolfgang Porod, Supriyo Bandyopadhyay, and
Gary H. Bernstein
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

14 May, 1993

Annual Report for Period 15 March 1992 — 14 March 1993

Prepared for
U. S. Air Force Office of Scientific Research
Bolling Air Force Base
Washington DC 20332-6448

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: University of Notre Dame, Electrical Engineering Dept. / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: Same as 6a
6c. ADDRESS (City, State and ZIP Code): Notre Dame, IN 46556	7b. ADDRESS (City, State and ZIP Code): Same as 6c
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: U.S. Air Force, Office of Scientific Research / 8b. OFFICE SYMBOL (If applicable): NE	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-91-0211
8c. ADDRESS (City, State and ZIP Code): Bolling Air Force Base, Washington, D.C. 20332-6448	10. SOURCE OF FUNDING NOS. — PROGRAM ELEMENT NO.: 61102F / PROJECT NO.: 2305 / TASK NO.: ES / WORK UNIT NO.
11. TITLE (Include Security Classification): Quantum Transport	
12. PERSONAL AUTHOR(S): Lent, Porod, Bandyopadhyay & Bernstein	
13a. TYPE OF REPORT: Annual / 13b. TIME COVERED: FROM 92-3-15 TO 93-3-14	14. DATE OF REPORT (Yr., Mo., Day): 93-5-14 / 15. PAGE COUNT
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES — FIELD / GROUP / SUB. GR	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

During the period April 15, 1992 to April 14, 1993, research carried out by the Nanostructures Group in the Department of Electrical Engineering at Notre Dame was concerned with a variety of quantum transport in mesoscopic structures. This research was funded by the Air Force Office of Scientific Research under Grant No. AFOSR-91-0211. The major issues examined included quantum transport in high magnetic fields and modulated channels. Coulomb-coupled quantum dot systems, transmission resonances and zeroes in resonant transport, self-consistent Hartree calculations of transport, lateral quantum wires an *pn*-junction formation, quantum magnetotransport in disordered systems, magnetoelectric states in quantum wires, anomalous magnetoresistance, electromigration, collision retardation and phonon effects in hot-electron transport, spin-polarized single electronics, single-particle lifetimes in quasi-1D structures, quantum transport experiments in metals, the mesoscopic photovoltaic effect, and new techniques for fabricating quantum structures in semiconductors.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ OTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code): 202 767-4931 / 22c. OFFICE SYMBOL: NE

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

SECURITY CLASSIFICATION OF THIS PAGE

Table of Contents

I. Introduction

During the period April 15, 1992 to April 14, 1993, research carried out by the Nanostructures Group in the Department of Electrical Engineering at Notre Dame was concerned with a variety of quantum transport in mesoscopic structures. This research was funded by the Air Force Office of Scientific Research under Grant No. AFOSR-91-0211. The major issues examined included quantum transport in high magnetic fields and modulated channels, Coulomb-coupled quantum dot systems, transmission resonances and zeroes in resonant transport, self-consistent Hartree calculations of transport, lateral quantum wires an *pn*-junction formation, quantum magnetotransport in disordered systems, magnetoelectric states in quantum wires, anomalous magnetoresistance, electromigration, collision retardation and phonon effects in hot-electron transport, spin-polarized single electronics, single-particle lifetimes in quasi-1D structures, quantum transport experiments in metals, the mesoscopic photovoltaic effect, and new techniques for fabricating quantum structures in semiconductors.

The research supported by this grant resulted in 33 journal papers, 12 conference presentations, and 4 research seminars. The research contributed to the granting of 2 M.S. degrees,1 Ph.D. degree, and supports continuing M.S. and Ph.D. thesis research.

II. Research Description

Quantum Transport in Magnetic Fields

Investigators: Craig S. Lent

We have been investigating ballistic electron transport in quantum channels with an applied magnetic field. The magnetic field is in the direction perpendicular to the plane of the two-dimensional electron gas from which the channels are formed. Electrons in such geometries form magnetic edge states, which carry current, and circulating Landau levels which do not. We solve the Schrödinger equation with the magnetic field included through a vector potential. The precise eigenstates of the system, and the current density distribution can the be calculated. We have extended the Quantum Transmitting Boundary Method, an adaptation of the Finite Element Method for current-carrying quantum states, to include the effect of the applied magnetic field.

We have used this technique to examine transport in quantum channels with a periodic modulation in the channel width. This investigation yielded two surprising results: 1) that, in the high-field regime, the electron in an infinite modulated channel propagates as though it were completely free, but with a *renormalized effective mass*, and 2) that the quantization of conductance characteristic of short ballistic constrictions is recovered for longer (but finite) periodically-modulated channels. This latter effect has lead us to make new predictions concerning the quenching of integer quantum Hall effect plateaus in such structures.

Resulting Publications and Presentations:

Manhua Leng and Craig S. Lent, "Magnetic Edge States in a Quantum Channel with a Periodic Array of Antidots", *Superlattices and Microstructures* **11**, 351 (1992).

Manhua Leng and Craig S. Lent, "Recovery of Quantized Ballistic Conductance in a Periodically Modulated Channel," submitted to *Physical Review Letters*.

Manhua Leng and Craig S. Lent, "The Quantum Transmitting Boundary Method in an Applied Magnetic Field," to be submitted to *Journal of Applied Physics*.

Coulomb-coupled Quantum Dot Systems

Investigators: Craig S. Lent, Wolfgang Porod, and Gary Bernstein

We analyze the interaction between quantum dot cells which may be potential building blocks for quantum cellular automata architectures. Each cell holds a few electrons and interacts Coulombically with nearby cells. In particularly interesting cell geometries, the charge density tends to align along one of two cell axes. Thus, we can define a cell "polarization" which might be used to encode binary information. The polarization of a cell is affected in a very nonlinear manner by the polarization of its neighbors. We quantify this interaction by calculating a cell-cell response function which makes clear the

non-linear bistability in the charge density of the coupled cell system. Effects of non-zero temperature on the response of a model cell have been investigated by taking quantum averages over the canonical ensemble. We have also examined the effects of multiple neighbors on a cell and developed programmable logic gate structures based on these ideas.

The underlying formalism we have developed involves expressing the electron-electron interaction in the form of a second-quantized Hamiltonian, and solving the resulting Schrödinger equation directly in the basis of many-electron states. This approach, which is numerically tractable for few-electron problems, is one we hope to generalize to related transport problems. It seems ideal for attacking certain problems in the quantum theory of dissipation.

Resulting Publications and Presentations:

C. S. Lent, P. Douglas Tougaw, and Wolfgang Porod, "Bistable Saturation in Coupled Quantum Dots for Quantum Cellular Automata," *Appl. Phys. Lett.*, **62**, 714 (1993).

C. S. Lent, P. Douglas Tougaw, Wolfgang Porod and Gary H. Bernstein, "Quantum Cellular Automata," to appear in *Nanotechnology* **4**, (1993).

P. Douglas Tougaw and C.S. Lent, "Lines of Interacting Quantum-dot Cells: a Binary Wire," submitted to Journal of Applied Physics.

C. S. Lent, P. Douglas Tougaw, and Wolfgang Porod, "Bistable Saturation in Coupled Quantum-dot Cells," submitted to Journal of Applied Physics.

C. S. Lent, P. Douglas Tougaw, and Wolfgang Porod, "A Bistable Quantum Cell for Cellular Automata," Proceedings of the International Workshop on Computational Electronics, University of Illinois at Urbana-Champaign, May 1992.

C.S. Lent, "Quantum Cellular Automata," seminar at the Department of Electrical Engineering, University of Illinois at Urbana, October 6, 1992.

C. S. Lent, P. D. Tougaw, W. Porod, and G. H. Bernstein, "Quantum Cellular Automata," presented at the "International Symposium on New Phenomena in Mesoscopic Structures," in Hawaii, December, 1992.

W. Porod, "Dissipation in Computation," seminar at the School of Electrical Engineering, Purdue University, October 1992.

C.S. Lent, "Quantum Cellular Automata," seminar at the Department of Electrical Engineering, Purdue University, April 14, 1993.

Transmission Resonances and Zeros in Quantum Waveguides with Resonantly-Coupled Cavities

Investigators: Wolfgang Porod and Craig S. Lent

We have studied transmission in quantum waveguides in the presence of resonant cavities. This work was inspired by our previous modeling of the conductance of resonantly-coupled quantum wire systems. We expected to find qualitatively the same phenomena as in the much studied case of double-barrier resonant tunneling, which can be viewed as transmission in a waveguide which contains both barriers. Our investigation showed surprising new behavior of the transmission amplitude for the case of the waveguides with the attached resonators: (i) The transmission probability possesses zeros, which are a consequence of unitarity. (ii) Each quasi-bound resonator state is represented by a zero-pole pair in the complex-energy plane. (iii) The close proximity of a zero and a pole results in very sharp resonance/antiresonance features of the transmission probability. (iv) The location of the transmission peaks does *not* give the location of the quasi-bound states, like for double-barrier resonant tunneling.

In current work, we study techniques to infer from the transmission probability on the real-energy axis the positions of zero-pole pairs in the complex-energy plane. This will allow us to obtain the lifetime of the corresponding quasi-bound states. For quantum waveguides, such a theory will be the analogue of the Breit-Wigner formulation of transmission resonances for double-barrier resonant tunneling. We also refine techniques to obtain the exact locations of zeros and poles from eigenvalue problems.

Resulting Publications and Presentations:

Wolfgang Porod, Zhi-an Shao, and Craig S. Lent, "Transmission Resonances and Zeros in Quantum Waveguides with Resonantly-Coupled Cavities," Applied Physics Letters 61, 1350 - 1352 (1992).

Wolfgang Porod, Zhi-an Shao, and Craig S. Lent, "Transmission Resonances and Zeros in Quantum Waveguides with Resonantly-Coupled Cavities," seminar at the Beckman Institute, University of Illinois, Urbana, Illinois, April 1992.

Zhi-an Shao, Wolfgang Porod, and Craig S. Lent, "A Numerical Study of Transmission Resonances and Zeros in Quantum Waveguide Structures," Proceedings of the International Workshop on Computational Electronics, pp. 253 -256; presented at the International Workshop on Computational Electronics, Urbana, Illinois, May 1992.

Zhi-an Shao, Wolfgang Porod, and Craig S. Lent, "Transmission Resonances and Zeros in Quantum Waveguide Systems with Attached Resonators," submitted to Physical Review B.

Lateral *P-N* Junctions and Quantum Wires Formed by Quasi Two-dimensional Electron and Hole Systems at Corrugated GaAs/AlGaAs Interfaces

Investigators: Wolfgang Porod

We have proposed a novel system for the design of quantum wires. Using the surface orientation dependent amphoteric nature of Si-doping, we have demonstrated in our numerical models the existence of lateral p-n junctions and p-n-p quantum wire structures

at corrugated *GaAs/AlGaAs* interfaces. Among the novel properties of such structures are the coexistence of quasi two-dimensional electron and hole states next to a quasi one-dimensional electron or hole system.

Si acts as an amphoteric dopant depending on substrate orientation during MBE growth of *GaAs* and *AlGaAs*, which may be utilized to achieve different doping behavior on different facets of a corrugated *GaAs/AlGaAs* interface. In particular, growth on {100} surfaces invariably leads to n-type doping, while p-type behavior has been reported for Ga-terminated {111}*A* surfaces. For a model geometry, we determine the potential landscape and the electron and hole charge densities within a semi-classical Thomas-Fermi screening model, and then solve the two dimensional Schrödinger equation using finite elements for the quantized electron and hole states at the heterointerfaces. At V-grooves etched into a semi-insulating GaAs substrate, a p-n junction forms at the intersection of (100) and (111) planes if the doping in the overlayer exceeds 5×10^{17} cm^{-3}. We have demonstrated the existence of a quantum wire which may form at the bottom of the V-groove. We find that a quantum wire is to be expected in this *p-n-p* structure for a length of the n-type region on the order of 100 *nm* and a Si-doping density of 1×10^{18} cm^{-3}.

Resulting Publications and Presentations:

Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "Lateral p-n junctions between quasi two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces," Journal of Vacuum Science and Technology B 10, 2051 -- 2055 (1992).

Wolfgang Porod, Henry K. Harbury, and Stephen M. Goodnick, "Lateral p-n junctions and quantum wires formed by quasi two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces," Applied Physics Letters 61, 1823 -- 1825 (1992).

Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "A Novel Quantum Wire Formed by Lateral p-n-p Junctions Between Quasi-Two-Dimensional Electron and Hole Systems at Corrugated GaAs/AlGaAs Interfaces," Journal of Applied Physics 73, 1509 -- 1520 (1993).

Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "A Numerical Study of Lateral p-n Junctions between Quasi Two-Dimensional Electron and Hole Systems at Corrugated GaAs/AlGaAs Interfaces," Proceedings of the International Workshop on Computational Electronics, pp. 249 -- 251; presented at the International Workshop on Computational Electronics, Urbana, Illinois, May 1992.

Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "Novel Quantum Wire Formed by Lateral p-n-p Junctions Between Quasi-Two-Dimensional Electron and Hole Systems at Corrugated GaAs/AlGaAs Interfaces," presented at the International Workshop on Quantum Structures, Santa Barbara, California, March 1993.

Carrier Dynamics in Quantum Wires

Investigators: Wolfgang Porod

Using the Monte Carlo technique, we have investigated the thermalization of carriers following a laser pulse excitation. Realistic quantum wire structures are investigated with multiple electronic subbands (on the order of 20). Our simulations include standard polar-optical phonon rates, and an improved technique to account for carrier-carrier scattering. We find that the reduced phase space in these quasi-one-dimensional systems leads to reduced carrier cooling when compared to bulk samples. These findings are in agreement with recent experiments.

Resulting Publications and Presentations:

L. Rota, F. Rossi, S. M. Goodnick, P. Lugli, E. Molinari, and W. Porod, "Reduced carrier cooling and thermalization in semiconductor quantum wires," Physical Review B 47, 1632 -- 1635 (1993).

L. Rota, F. Rossi, P. Lugli, E. Molinari, S. M. Goodnick, and W. Porod, "Monte Carlo Simulation of a 'True' Quantum Wire," presented at the 1992 Symposium on Compound Semiconductor Physics and Devices, of SPIE -- The International Society of Optical Engineering, Somerset, New Jersey, March 1992.

Local Field Effects and Electromigration in Mesoscopic Structures

Investigators: Wolfgang Porod and Craig S. Lent

In this ongoing effort, we study the effect of space charges on the current - voltage characteristics of two-dimensional mesoscopic structures. In particular, we focus on the microscopic potential and current variations in the vicinity of scattering centers. The resulting non-uniform carrier distribution gives rise to local-field effects and the so-called residual resistivity dipole. These phenomena are thought to be responsible for the driving force in electromigration, and we believe that the study of such mesoscopic systems may shed light on an understanding of this important effect.

The (non-uniform) electronic density along the channel influences the electronic wavefunctions through its space charge potential. The Schrödinger and Poisson equations are solved self-consistently within the Hartree approximation. Like in our past studies, the finite element method is used for the discretization of the Schrödinger and Poisson equations. The Hartree term is obtained by numerically integrating the contributions from the electronic charge density along the channel. An iterative procedure is then employed until self consistency in the wavefunctions and in the space charge potential is achieved.

Resulting Publications and Presentations:

Henry K. Harbury, Wolfgang Porod, and Craig S. Lent, "Field Effects in Self-consistent Transport Calculations for Narrow Split-gate Structures", *Superlattices and Microstructures* **11**, 189 - 193 (1992).

Spatial Distributions of Transport Variables in Quantum Transport

Investigators: Supriyo Bandyopadhyay

We have calculated the spatial distributions of current, carrier concentration, electrostatic and chemical potential, residual resistivity dipole fields, etc. around localized elastic scatterers in phase coherent quantum transport. We have observed many interesting features such as the collapse of all current carrying states into "edge states" in the quantum Hall regime, the formation of magnetic bound states pinned by an impurity and the associated current patterns around the impurity, the destruction of these states by a magnetic field, the destruction of quantum Hall effect by magnetic bound states, the vanishing of the longitudinal resistance and precise quantization of the Hall resistance in the quantum Hall regime, current patterns around attractive and repulsive scatterers and their dependence on a magnetic field, the difference between majority and minority carrier transport and mobility in the quantum regime, current vortices forming as result of quantum interference between scatterers, dependence of the sign of magnetoresistance in a disordered structure on the impurity configuration, etc. All of these have revealed new physics and have helped to visualize many other physical phenomena. Our most important contribution is that we have demonstrated the Quantum Hall effect from a microscopic approach (directly from the Schrödinger equation) in the phase coherent regime, for the first time. We have also calculated the spatial distributions of transport variables in the coherent regime associated with the Quantum Hall Effect for the first time.

Resulting Publications and Presentations:

S. Chaudhuri and S. Bandyopadhyay, ``Spatial Distribution of the Current and Fermi Carriers Around Localized Elastic Scatterers in Quantum Transport" Phys. Rev. B, 45, 11126 - 11135 (1992).

S. Chaudhuri, S. Bandyopadhyay, and M. Cahay, ``Spatial distribution of the current, Fermi carrier density, potential and electric field in a disordered quantum wire in a magnetic field" Phys. Rev. B (to appear in May, 1993).

Magnetoelectric States in Electron Waveguides

Investigators: Supriyo Bandyopadhyay

We have calculated the wavefunctions and energy dispersion relations of hybrid magnetoelectric states in a quasi one-dimensional structure numerically from the Schrödinger equation. From these, we have also obtained the density of one dimensional magnetoelectric states. These results have been compared with semiclassical results based on the Bohr-Sommerfield quantization rule. The results are very important in many areas of quantum magnetotransport.

Resulting Publications and Presentations:

S. Chaudhuri and S. Bandyopadhyay, ``Numerical Calculation of Hybrid Magnetoelectric States in an Electron Waveguide", J. Appl. Phys., 71, 3027 - 3029 (1992).

S. Chaudhuri and S. Bandyopadhyay, ``Quantum Transport in a Disordered Quantum Wire in the Presence of a Magnetic Field", Superlattices and Microstructures, 11, 241 - 244 (1992).

S. Chaudhuri, S. Bandyopadhyay and M. Cahay, ``Numerical study of quantum magnetotransport in disordered non-adiabatic constrictions", Proceedings of the International Conference on Computational Electronics, 305 - 308, (1992).

Anomalous Magnetoresistance

Investigators: Supriyo Bandyopadhyay and Gary H. Bernstein

We have experimentally observed anomalous magnetoresistance in a two-dimensional structure at the liquid helium temperature of 4.2 K. This has been explained by invoking subtle correlations between impurity scattering events in the structure. A Monte Carlo simulation was performed and supports this picture.

Resulting Publications and Presentations:

N. Telang and S. Bandyopadhyay, ``A Monte Carlo Study of Correlations Between Impurity Scattering Events in a Two dimensional Electron Gas Causing Inhomogeneous Magnetoresistance, Superlattice and Microstructures 11, 99 - 102 (1992).

S. Subramaniam, S. Bandyopadhyay, B. Das, G. H. Bernstein, and P. A. Sekula-Moise, ``Anomalous Magnetoresistance Due to Correlations Between Disorder and Boundary Scattering in a Two-Dimensional Electron Gas Channel" Phys. Rev. B, 45, 3826 - 3829 (1992).

Electromigration

Investigators: Supriyo Bandyopadhyay

We have calculated the weak electromigration forces causing 1/f noise and impurity migration in quasi one-dimensional structures in linear response transport. We have shown that the wind force and direct force can be quenched or reoriented by an external magnetic field. We have also performed these calculations in the hot electron regime using a Monte Carlo simulation.

Resulting Publications and Presentations:

S. Bandyopadhyay, ``Coupling and Crosstalk Between High Speed Interconnects in Ultra Large Scale Integrated Circuits", IEEE J. Quant. Elec., 28, 1554 - 1561 (1992).

S. Bandyopadhyay, S. Chaudhuri, B. Das and M. Cahay, ``Features of quantum magnetotransport and electromigration in mesoscopic systems" Superlattices and Microstructures, 12, 123 - 132 (1992). M. Cahay and S. Bandyopadhyay,

S. Bandyopadhyay, S. Chaudhuri, B. Das and M. Cahay, ``Magnetotransport and electromigration in mesoscopic systems'', Sixth International Conference on Superlattices and Microstructures, Xian, People's Republic of China, August 1992.

Collision Retardation Effects in Hot Electron Transport in Two Dimensional Systems

Investigators: Supriyo Bandyopadhyay

We have performed a Monte Carlo simulation of hot electron transport in a quasi two dimensional structure including the effects of finite collision duration (collision retardation). We have found that collision retardation tends to make the electrons hotter by suppressing energy relaxation collisions. It also increases velocity and energy overshoot.

Resulting Publications and Presentations:

N. Telang and S. Bandyopadhyay, ``The effect of collision retardation on hot electron transport in a quantum well'', Phys. Rev. B (to appear in April, 1993.)

The Effects of a Magnetic Field on Phonon Scattering in Quasi One Dimensional Systems:

Investigators: Supriyo Bandyopadhyay

We have rigorously calculated electron-phonon scattering rates in quasi one dimensional systems in the presence of an external magnetic field. We have found that a magnetic field dramatically reduces both polar and non-polar acoustic phonon scattering by orders of magnitude which has important implications for the quantum Hall effect. The reduction is caused by a decrease in the overlap between initial and final state electron wavefunctions. The scattering rates due to longitudinal polar and non-polar optical phonons and surface phonons increases in a magnetic field (though the change is not as dramatic as in the case of acoustic phonons). This is attributed to the opening of many new scattering channels associated with the breaking of orthogonality between electron wavefunctions and confined phonon modes in the presence of a magnetic field.

Resulting Publications and Presentations:

N. Telang and S. Bandyopadhyay, ``Quenching of Acoustic Phonon Scattering of Electrons in Semiconductor Quantum Wires Induced by a Magnetic Field'' Appl. Phys. Lett. (in press).

N. Telang and S. Bandyopadhyay, ``Electron phonon scattering in quantum wires subjected to high magnetic field'', Proceedings of the International Conference on Computational Electronics, 237 - 240, (1992).

N. Telang and S. Bandyopadhyay, ``The Effects of a Magnetic Field on Electron Phonon Scattering in Quantum Wires by a Magnetic Field'', submitted to Phys. Rev. B.

Supercomputing with Spin Polarized Single Electrons

Investigators: Supriyo Bandyopadhyay

We have proposed a novel quantum technology for ultra-fast, ultra-dense and ultra-low power supercomputing. The technology utilizes single electrons as binary logic devices in which the *spin* of the electron encodes the bit information. The architecture mimics two dimensional cellular automata. It is realized by laying out on a wafer regimented arrays of nanophase particles each hosting an electron. Various types of logic gates, combinational circuits for arithmetic logic units, and sequential circuits for memory can be realized.

The technology has many advantages such as (1) the absence of physical interconnects between devices (inter-device interaction is provided by quantum mechanical coupling between adjacent electrons), (2) ultrafast switching times of 1 picosecond for individual devices, (3) extremely high bit density approaching 10 Terabits/sq-cm, (4) non-volatile memory, (5) robustness and possible *room temperature* operation with very high noise margin and reliability, (6) a very low power delay product for switching a single bit 10^{-20} Joules), and (7) a very small power dissipation of a few tens of nanowatts per bit. We have also proposed a new non-lithographic fabrication technology for realizing these chips.

Resulting Publications and Presentations:

S. Bandyopadhyay, B. Das and A.E. Miller, "Supercomputing with Spin Polarized Single Electrons in a Quantum-coupled Architecture," submitted to J. Appl. Physics.

S. Bandyopadhyay, Single Electronics with Nanophase Materials, invited lecture presented at Argonne National Laboratory, Argonne, Illinois, June 1992.

S. Bandyopadhyay, Supercomputing with spin polarized single electrons, invited seminar presented at School of Electrical Engineering, Purdue University, March 1993.

Single Particle Lifetimes in Quasi-one-dimensional Structures

Investigators: Supriyo Bandyopadhyay

We have experimentally measured, for the first time, the relaxation time for single particle excitations in quantum wires as a function of both wire width and carrier concentration. The measurements were made in a back-gated AlGaAs/GaAs heterostructure at 4.2 K. The experimental data show excellent agreement with a theoretical model which uses only one fitting parameter (to account for screening).

Resulting Publications and Presentations:

M. Cahay, T. Singh and S. Bandyopadhyay, ``Electron emission from a quantum well as a result of exchange and Coulomb interactions", Proceedings of the International Conference on Computational Electronics, 147 - 150, (1992).

Quantum Transport in Nanofabricated Structures: Program Overview

Investigators: Gary H. Bernstein

The experimental aspect of the second year of the grant was very successful. We made enormous progress in the area of fabrication of advanced nanometer-scale and quantum structures, and have bridged the gap from pushing the limits of nanofabrication to fabricating devices in metal, silicon and III-V materials for testing both at Notre Dame and other universities.

Electron Beam Lithography Technology for Advanced Quantum Device Structures

Investigators: Gary H. Bernstein

In the area of nanostructure fabrication, work progressed on several fronts. First, the very-low-noise electron beam lithography (EBL) system was completed. The true capabilities of the system were demonstrated through the achievement of high resolution over very large scan fields. Asdiscussed in the resulting publication (Bazán and Bernstein), poor noise performance of the associated EBL electronics can usually be compensated for by performing exposures at small scan fields. This reduces the effects of the noise, but also limits the flexibility of the system. By making major improvements in noise and system performance, we accomplished sub-100-nm lithography in scan fields up to 1.4 μm (1400 μm) in size. Figure 1 shows a 55-nm metal line fabricated at a distance of about 300 μm from the center of a 1 mm scan field exposed in our improved EBL system. The scan fields discussed here are up to 25 times larger in area than those typically reported in the literature. Our publication (presented at the American Vacuum Society Meeting in Chicago, November, 1992) demonstrated that our results are quite useful over the full scan field.

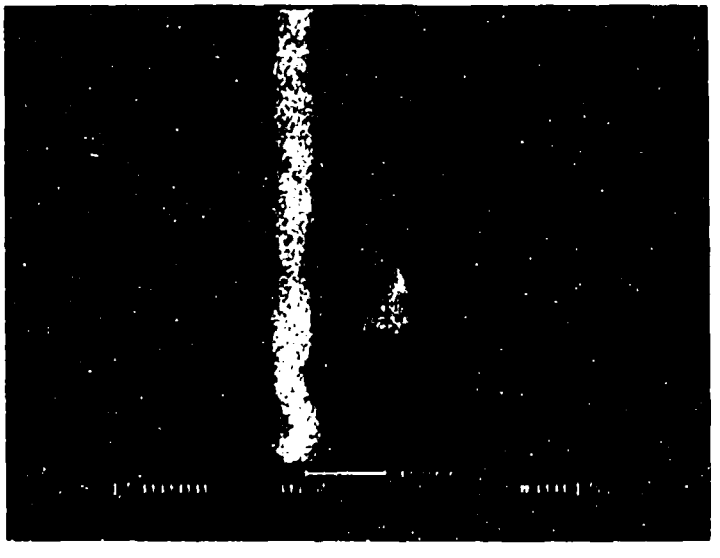


FIGURE 1. 55-nm gold line fabricated 300 μm from the center of a 1 mm scan field demonstrating noise properties of improved EBL system.

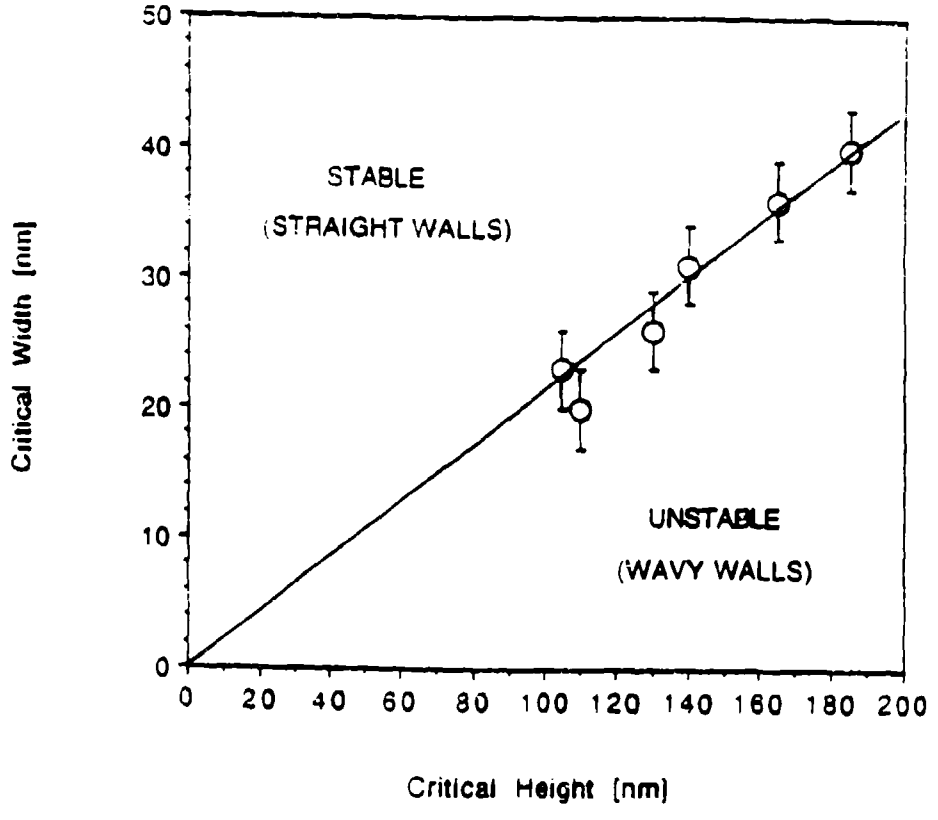


FIGURE 2. Minimum achievable remaining wall thickness possible for successful lift-off versus starting resist thickness.

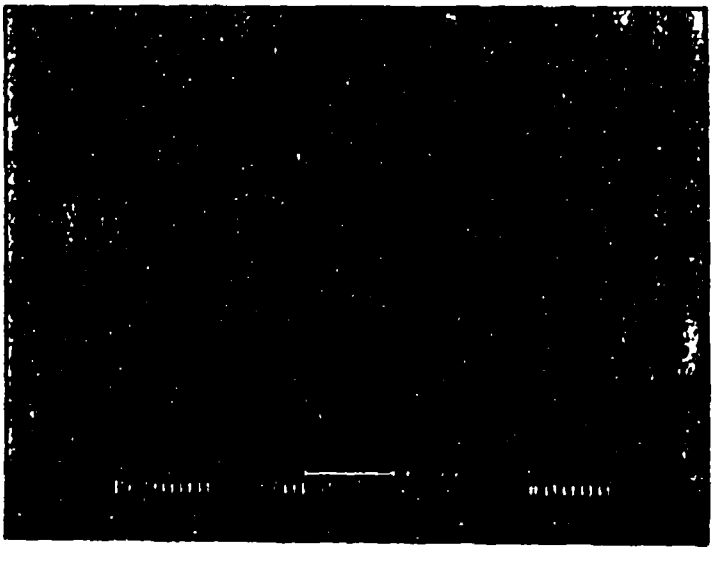


FIGURE 3. Our best gold line density achieved on silicon to date- approximately 38 nm pitch.

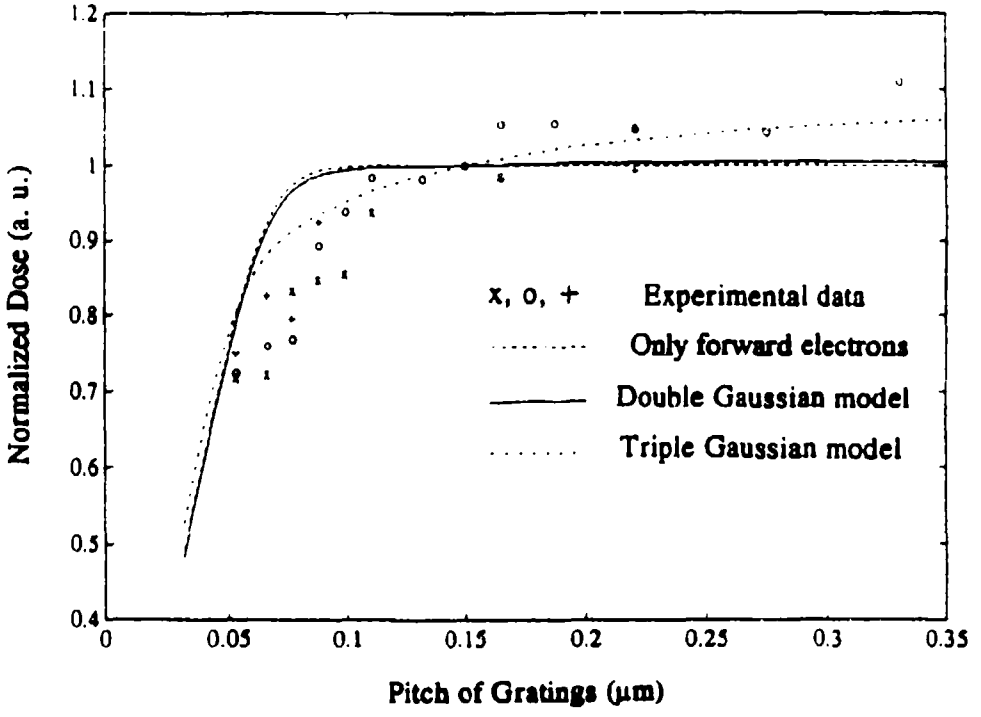


FIGURE 4. Normalized area dose necessary to effect successful lift-off of well-defined patterns.

Secondly, new developers for poly (methyl methacrylate) (PMMA), as discussed in the report for year one, bore fruit. As predicted, the new developers were instrumental in achieving very high metal pattern densities. We fabricated dots and lines with pitches well below 40 nm. Figure 2 shows the best line density we have achieved on gratings (on silicon) to date- approximately 38 nm pitch. Some new properties of PMMA were discovered, leading to a conference presentation (Bernstein *et al.*, Chicago, 1992) and several publications

It was found that the physical strength of PMMA limits the achievable metal thickness obtained with the lift-off process. Figure 3 shows the minimum achievable remaining wall thickness (necessary for successful lift-off) versus starting resist thickness. We found that due to interwall forces and swelling of the PMMA at these dimensions, a maximum ratio of 4.7 of resist thickness to wall width is possible before failure of the lines occurs.Also at these dimensions, proximity effects increase dramatically. Figure 4 shows the normalized area dose (normalized to the area dose necessary to create gratings with 40-nm- wide lines on a 150-nm-pitch) necessary to effect successful lift-off of well-defined patterns. We used a triple Gaussian model to fit the data, and found evidence that fast secondary electrons, predicted by D. Joy[1], were responsible for a sharp decrease in exposure dose as line pitch decreased below about 100 nm. Figure 5 shows that as the number of lines in the grating

1. D. C. Joy, *Microelectronic Engineering*, Vol. 1, p. 103 (1983).

increases, the required dose drops very quickly due to these fast secondary electrons, leveling out only for gratings with more than about 15 lines. To our knowledge, ours is the first detailed study of proximity effects on this very small size scale reported in the literature

X. Huang, G. H. Bernstein, G. Bazán, and D. A. Hill, "Spatial Density of Lines in PMMA by Electron Beam Lithography," *J. Vac. Sci. Technol. B.*, in press.

G. Bazán and G. H. Bernstein, "Electron Beam Lithography Over Very Large Scan Fields," *J. Vac. Sci. Technol. B*, in press.

D. A. Hill, X. Huang, G. Bazán, and G. H. Bernstein, "Swelling and Surface Forces-Induced Instabilities in Nanoscopic Polymeric Structures," *J. Appl. Phys.*, 72(9), pp. 4088-4094 (1992).

X. Huang, G. Bazán, G. H. Bernstein, and D. A. Hill, "Stability of Thin Resist Walls," *J. Electrochem. Soc.*, **139**(10), pp. 2952-2956 (1992).

G. H. Bernstein and D. A. Hill, "On the Attainment of Optimum Development Parameters for PMMA Resist," *Superlat. and Microstruct.*, **11**, No. 2, PP. 237-240 (1992).

G. H. Bernstein, D. A. Hill and W. P. Liu, "New High-Contrast Developers for PMMA Resist," *J. Appl. Phys.*, **71** (8) pp. 4066-4075 (1992).

G. H. Bernstein, D. A. Hill, X. Huang, and G. Bazán, "Failure Mechanisms of Very Narrow PMMA Walls," presented at the 39th National Symposium of the American Vacuum Society Chicago, IL, November 1992,.

G. Bazán and G. H. Bernstein, "Electron Beam Lithography Over Large Scan Fields," presented at the 39th National Symposium of the American Vacuum Society, Chicago, IL, November 1992

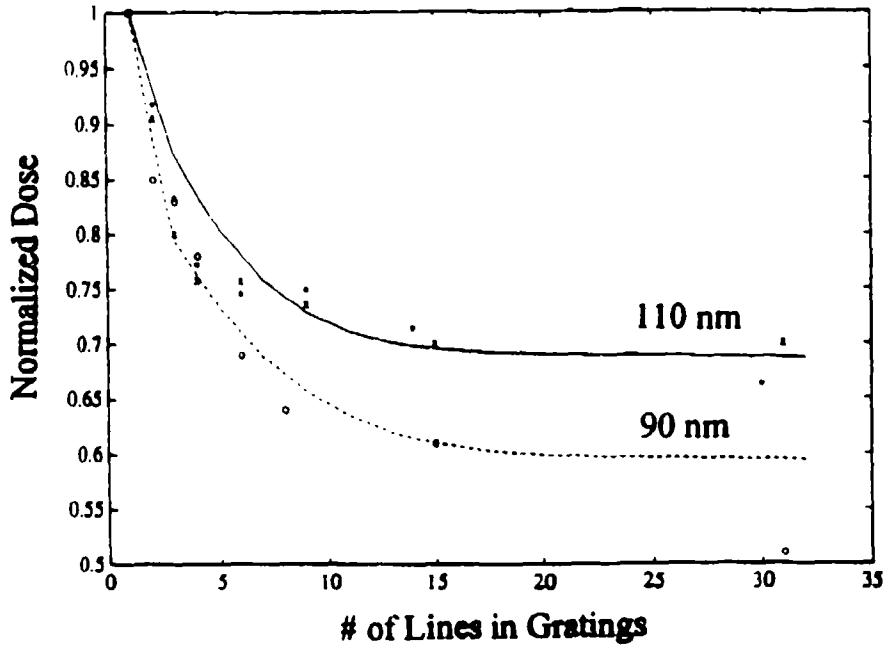


FIGURE 5. Normalized dose necessary to effect successful lift- off versus total number of lines per grating.

Quantum Transport in Metals

Investigators: Gary H. Bernstein

We have completed the first phase of building our low-temperature measurement capability. Proof of this is demonstrated by the replication of an experiment performed previously by Benoit et al.[2] and Skocpol et al.[3] which demonstrates the dependence of the magnitude of conductance fluctuations on the length of metal wires. Figure 6 shows details of our data indicating UCF in gold lines, 70 nm wide and 30 nm thick, at 1.7 K. The solid and dashed lines represent scans up and down in magnetic field, respectively. (The lines are offset slightly as a result of mismatch between time constants of the lock-in

2. A. Benoit, C. P. Umbach, R. B. Laibowitz, and R. A. Webb, *Phys. Rev. Lett.*, Vol. 58, No. 22, p. 2343 (1987).

3. W. J. Skocpol, in Physics of Quantum Electron Devices, ed. by Federico Capasso, p. 367, Springer- Verlag, New York (1990).

amplifier as compared with magnetic field scan rates.) Figure 7 shows the RMS magnitude of the conductance fluctuations as a function of normalized line length (where L_{ϕ}, the phase breaking length, is taken as 1 μm based on values found by Benoit et al.). The data agrees very well with that of previous researchers with an exponential dependence with a power of -3/2 to 2. The agreement of our data with that of previous researchers shows that we are now capable of performing meaningful low-temperature experiments on a variety of systems.

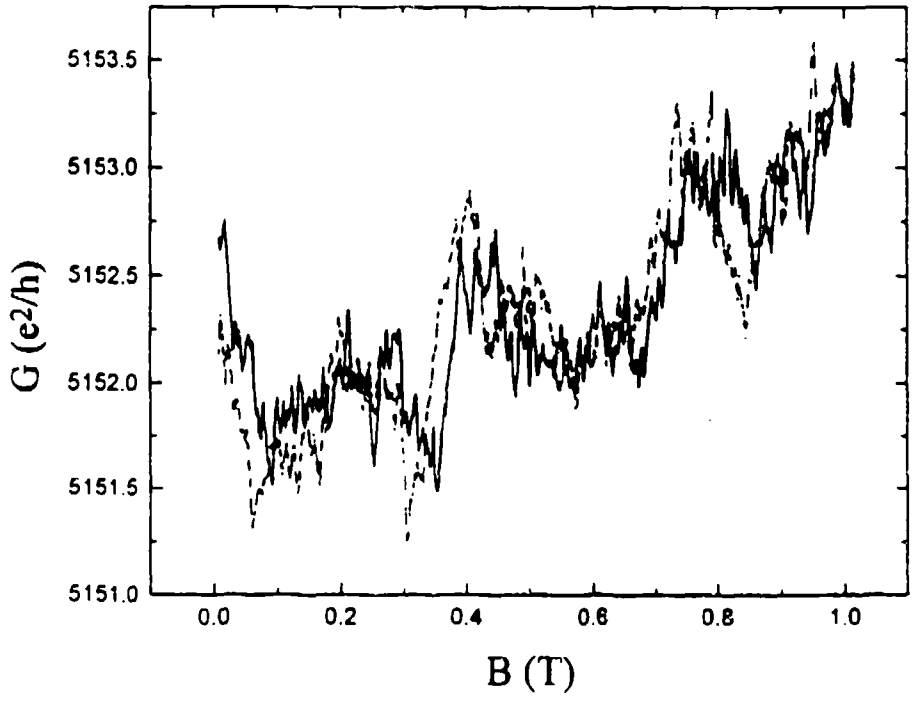


FIGURE 6. Data indicating UCF in gold lines at 1.7 K.

We have succeeded in fabricating pairs of metal lines with widths of about 20 nm spaced by about 16 nm, as shown in Figure 8. This is an improvement of a factor of about three in device size as compared with a year ago. It was hoped that virtual phonon coupling would be evident on these size scales at 1.8 K, but so far no effects of quantum coupling has been demonstrated. We are continuing to decrease noise and capacitive coupling in our measurement system and to decrease the spacing even further. In addition, a new low temperature measurement system with capability of 300 mK and 11 Tesla will be operational in the near future.

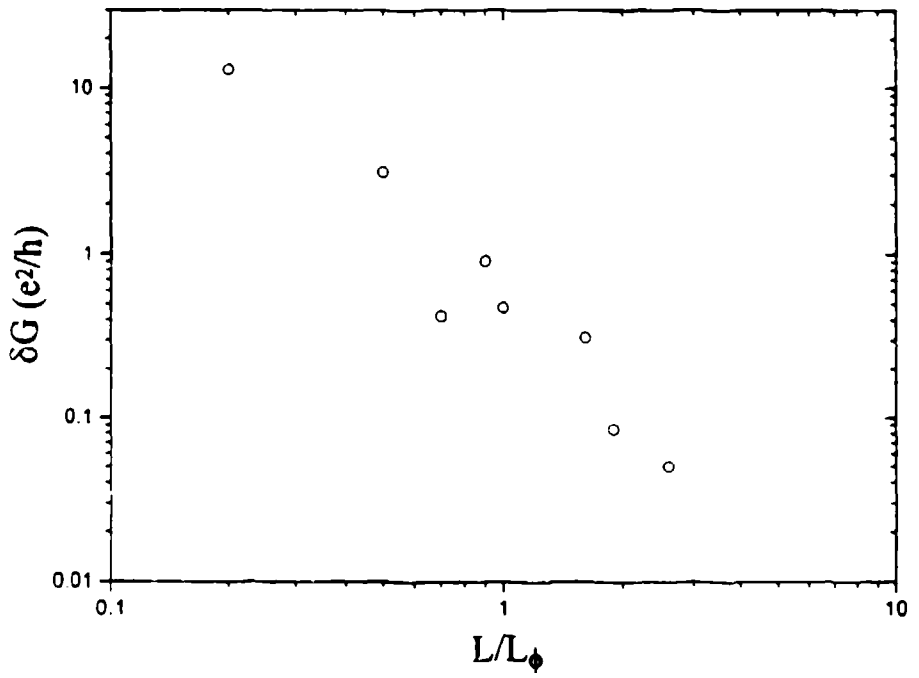


FIGURE 7. RMS magnitude of the conductance fluctuations as a function of normalized line length (where L_ϕ is the phase breaking length taken as 1 μ m).

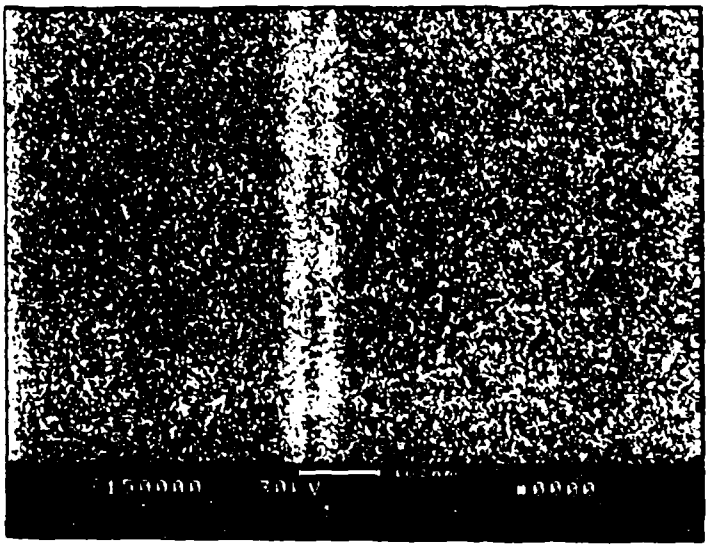


FIGURE 8. Pair of metal lines with widths of about 20 nm spaced by about 16 nm for use in phonon coupling experiments.

Mesoscopic Photovoltaic Effect

Investigators: Gary H. Bernstein

We maintain a close collaboration with Dr. N. Giordano and graduate student Mr. R. Bartolo at Purdue University. Their cryogenic measurement system is capable of in situ microwave irradiation. Using this system, they have observed d.c. voltages in mesoscopic systems which arise as a direct result of the microwave energy[4]. This photovoltaic (PV) effect has been observed in different systems[5], but we are attempting to directly locate the source of the rectification of the electron motion by separating the mesoscopic structure

4. J. Liu, M. A. Pennington, and N. Giordano, *Phys. Rev. B*, Vol. 45, No. 3, p. 1267 (1992).
5. A.A. Bykov, G. M. Gusev, Z. D. Kvon, and B. I. Fomin, *Sov. Phys. JETP* Vol. 70, No. 1, p. 140 (1990).

into segments and isolating the individual scatterers. Figure 9 shows a sample built at Notre Dame used in this experiment. These experiments are in progress.

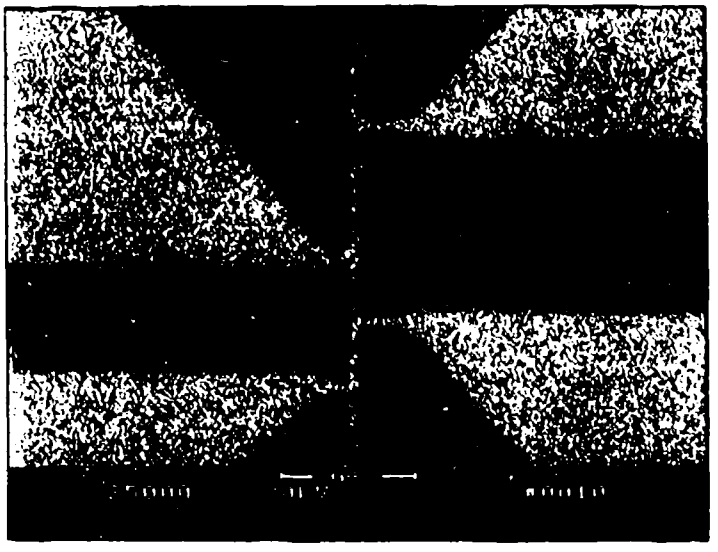


FIGURE 9. Metal line fabricated on glass slide at Notre Dame showing voltage probes and interconnect pads. This pattern is currently employed to study the origin of the photovoltaic effect in mesoscopic structures.

Additionally, we are trying to identify interference phenomena which might play a role in the PV effect. Figure 10 shows a gold Aharonov-Bohm ring on a glass substrate fabricated at Notre Dame. Figure 11 shows PV data obtained at Purdue on the ring sample. The

Fourier transform of the data indicates an h/e component of the PV effect which could prove that the effect is strongly related to interference phenomena.

Resulting Publications and Presentations:

R. E. Bartolo, N. Giordano, X. Huang, and G. H. Bernstein, "Giant Oscillations in Mesoscopic Photovoltaic Effect," poster presented at the 40th Midwest Solid State Conference, Urbana-Champaign, Oct. 1992. (Judged Second-Best Poster at Conference.)

FIGURE 10. Gold Aharonov-Bohm ring on a glass substrate fabricated at Notre Dame used to determine the role of quantum interference in the photovoltaic effect.

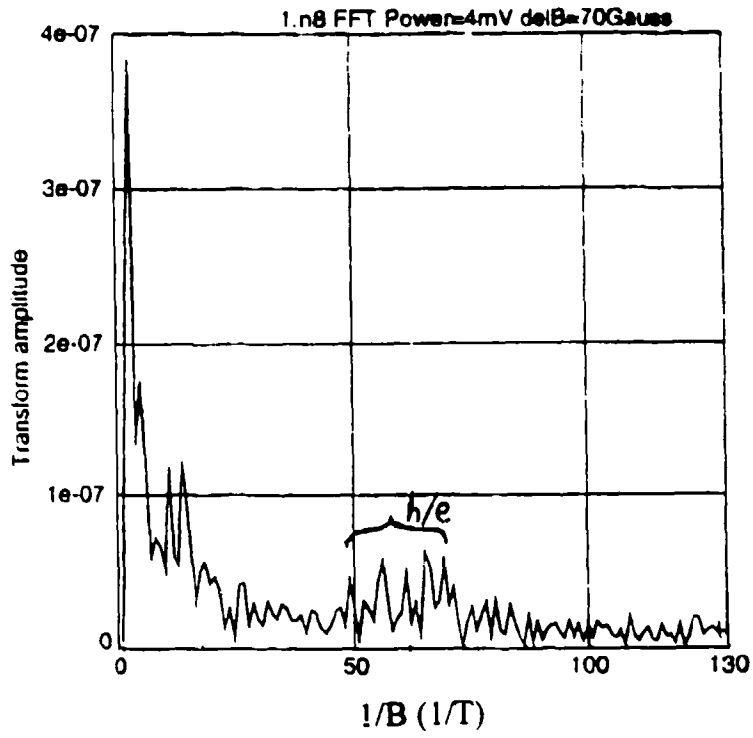
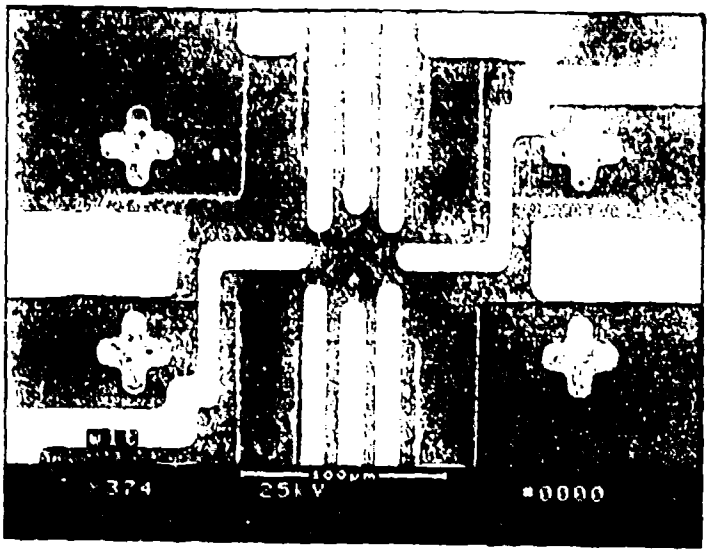


FIGURE 11. Photovoltaic effect data obtained at Purdue University showing possible h/e oscillations indicative of Aharonov Bohm effect. If confirmed, this data would show that phase coherence plays a role in the PV effect.

Quantum Transport in III-V Materials

Investigators: Gary H. Bernstein

In collaboration with Dr. S. Washburn at the University of North Carolina, Chapel Hill, we have fabricated prototype structures with up to 8 control gates. Figure 12(a) shows one prototype device requiring 14 leads consisting of both ohmic and Schottky contacts. Evident is the mesa pattern with ohmic contacts and leads for Schottky contacts. Figure 12(b) shows the same device with a close-up of the Schottky gates on the mesa. Spaces between Schottky contacts are as narrow as 40 nm. This structure will allow a large variety of effects to be observed in one device ranging from single electron tunneling to split gate behavior with a series of gates. We are in the process of transferring the fabrication technology to device-grade material for testing at UNC. Similar devices will be tested at Notre Dame when the milliKelvin cryostat is operational.



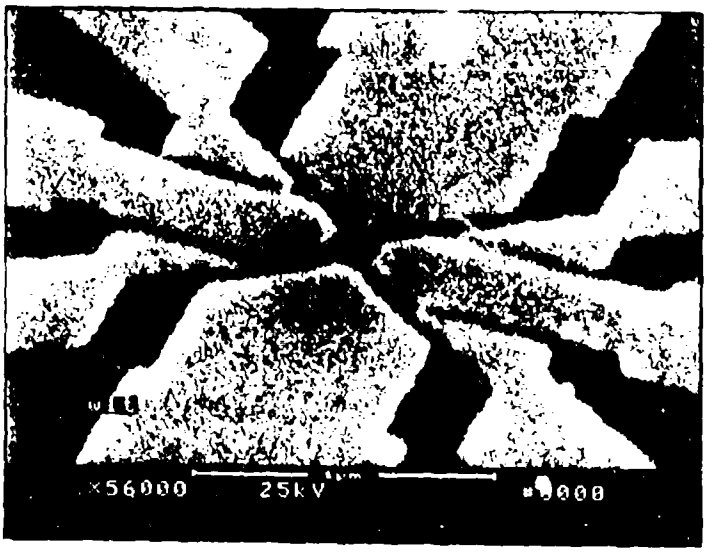


FIGURE 12. a) One prototype device requiring 14 leads consisting of both ohmic and Schottky contacts. Evident is the mesa pattern with ohmic contacts and leads for Schottky contacts. b) Same device with a close-up of the Schottky gates on the mesa. Spaces between Schottky contacts are as narrow as 40 nm.

Quantum Transport in Silicon

Investigators: Gary H. Bernstein

We have undertaken an exciting new area of fabrication for silicon nanostructures. Initially, we demonstrated that positive charges resulting from electron beam damage to SiO_2 remain fixed in space (Bernstein et al.). These positive charges cause a shift in the MOSFET threshold voltage by up to -15 volts, so that at a gate voltage where an n-channel MOSFET is still off, the area that has been irradiated by the beam has a significant inversion layer in place. We conceived of the possibility of exposing narrow regions between the source and drain of an n-channel MOSFET which will remain conducting while the rest of the device is nonconducting. We hope to demonstrate 1-D and quantum transport in MOSFET's which have been exposed in this way over narrow regions between the source and drain. It should be emphasized that the exposure is a straightforward application of EBL patterns performed through the existing gate metal and oxide of a simple MOSFET.

So far, we have used our silicon IC processing laboratory to make test MOSFET's used in these experiments. We have recently exposed MOSFET's as described above and have observed some very compelling evidence of confinement due to the narrow potential well created by the positive charge left after irradiation by the electron beam. Modulation of the gate voltage is expected to narrow or widen the potential well thus controlling its

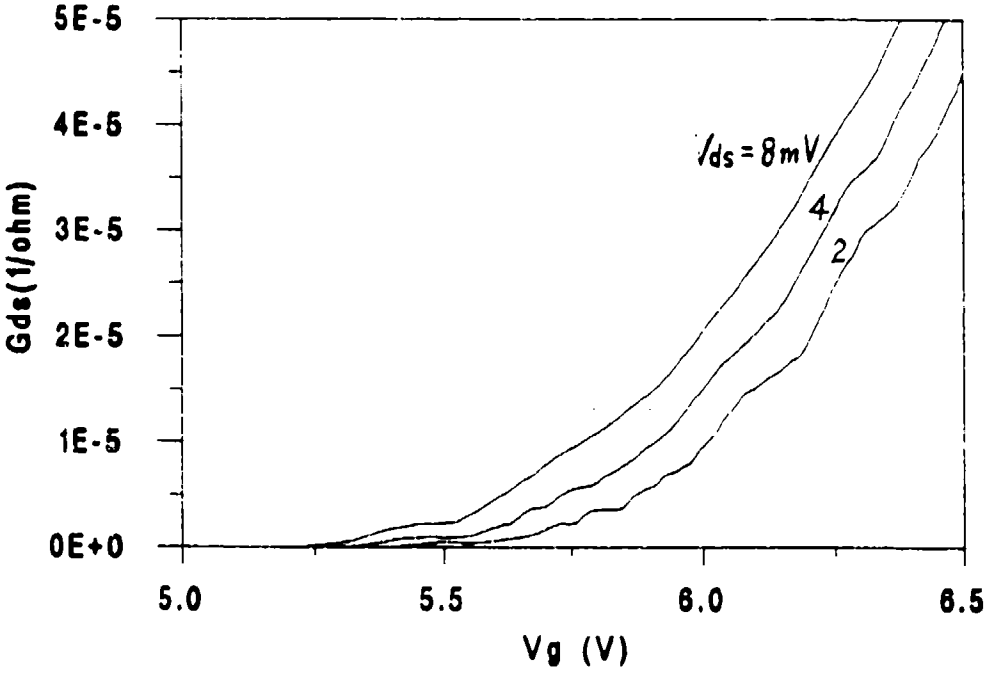


FIGURE 13. Structure in the drain conductance, g_d, between source and drain of an 8 μm long MOSFET at 1.83 K as a function of gate voltage, V_g. The 1-D channel was created by EBL exposure and resulting positive oxide charge creating a potential well at the interface.

properties. Figure 13 shows structure in the drain conductance, g_d, between source and drain of an 8 μm long MOSFET at 1.83 K as a function of gate voltage, V_g. As expected for confinement effects, the structure in the g_d vs. V_{ds} data disappears at higher temperatures and at larger V_{ds}. Also, there is no evidence of the structure for unexposed MOSFET's fabricated in our laboratory, or for commercially obtained, prepackaged MOSFET's.

We have much work to do in order to verify that the structure is due to confinement of the conducting channel brought on by the positive charge generation. We hypothesize so far that the structure is due either to UCF or single electron tunneling phenomena. Many more experiments are planned in order to determine the precise nature of the effect. The pending confirmation of confinement effects implies that a whole range of quantum devices ranging from lateral surface superlattices to turnstile devices might be possible by simple processing of existing silicon devices.

Resulting Publications and Presentations:

G. H. Bernstein, S. W. Polchlopek, R. Kamath, and W. Porod, "Determination of Electron-Beam-Induced Positive Oxide Charge," *Scanning*, Vol. 14, pp. 345-349 (1992).

III. Research Personnel Supported

Craig S. Lent	Assoc. Professor	1 Summer month
Wolfgang Porod	Professor	1 Summer month
Supriyo Bandyopadhyay	Assoc. Professor	1 Summer month
Gary H. Bernstein	Assoc. Professor	1 Summer month
Zhi-an Shao	Graduate student	6 months
Manhua Leng	Graduate student	6 months
Suresh Subramaniam	Graduate student	6 months
Xiaokang Huang	Graduate student	6 months

IV. Theses supervised

Santanu Chaudhuri "Quantum Transport in Disordered Systems," MSEE May 1992 (advisor: S. Bandyopadhyay)

Renu Kamath, "Breakdown in Thin SiO_2 Films", MSEE August 1992 (advisors: G.H. Bernstein and W. Porod)

Suresh Subramaniam, "A Study of Electron Transport in Backgated AlGaAs/GaAs Modulation Doped Heterostructures for Novel Device Applications", May 1993.

V. Publication Reprints and Preprints

Journal Articles

Manhua Leng and Craig S. Lent, "Magnetic Edge States in a Quantum Channel with a Periodic Array of Antidots", *Superlattices and Microstructures* **11**, 351 (1992).

Manhua Leng and Craig S. Lent, "Recovery of Quantized Ballistic Conductance in a Periodically Modulated Channel," submitted to *Physical Review Letters*.

Manhua Leng and Craig S. Lent, "The Quantum Transmitting Boundary Method in an Applied Magnetic Field," to be submitted to *Journal of Applied Physics*.

C. S. Lent, P. Douglas Tougaw, and Wolfgang Porod, "Bistable Saturation in Coupled Quantum Dots for Quantum Cellular Automata," *Appl. Phys. Lett.*, **62**, 714 (1993).

C. S. Lent, P. Douglas Tougaw, Wolfgang Porod and Gary H. Bernstein, "Quantum Cellular Automata," to appear in *Nanotechnology* **4**, (1993).

P. Douglas Tougaw and C.S. Lent, "Lines of Interacting Quantum-dot Cells: a Binary Wire," submitted to Journal of Applied Physics.

C. S. Lent, P. Douglas Tougaw, and Wolfgang Porod, "Bistable Saturation in Coupled Quantum-dot Cells," submitted to Journal of Applied Physics.

Wolfgang Porod, Zhi-an Shao, and Craig S. Lent, "Transmission Resonances and Zeros in Quantum Waveguides with Resonantly-Coupled Cavities," Applied Physics Letters 61, 1350 - 1352 (1992).

Zhi-an Shao, Wolfgang Porod, and Craig S. Lent, "Transmission Resonances and Zeros in Quantum Waveguide Systems with Attached Resonators," submitted to Physical Review B.

Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "Lateral p-n junctions between quasi two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces," Journal of Vacuum Science and Technology B 10, 2051 -- 2055 (1992).

Wolfgang Porod, Henry K. Harbury, and Stephen M. Goodnick, "Lateral p-n junctions and quantum wires formed by quasi two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces," Applied Physics Letters 61, 1823 -- 1825 (1992).

Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "A Novel Quantum Wire Formed by Lateral p-n-p Junctions Between Quasi-Two-Dimensional Electron and Hole Systems at Corrugated GaAs/AlGaAs Interfaces," Journal of Applied Physics 73, 1509 -- 1520 (1993).

L. Rota, F. Rossi, S. M. Goodnick, P. Lugli, E. Molinari, and W. Porod, "Reduced carrier cooling and thermalization in semiconductor quantum wires," Physical Review B 47, 1632 -- 1635 (1993).

Henry K. Harbury, Wolfgang Porod, and Craig S. Lent, "Field Effects in Self-consistent Transport Calculations for Narrow Split-gate Structures", *Superlattices and Microstructures* **11**, 189 - 193 (1992).

S. Chaudhuri and S. Bandyopadhyay, ``Spatial Distribution of the Current and Fermi Carriers Around Localized Elastic Scatterers in Quantum Transport" Phys. Rev. B, 45, 11126 - 11135 (1992).

S. Chaudhuri, S. Bandyopadhyay, and M. Cahay, ``Spatial distribution of the current, Fermi carrier density, potential and electric field in a disordered quantum wire in a magnetic field" Phys. Rev. B (to appear in May, 1993).

S. Chaudhuri and S. Bandyopadhyay, ``Numerical Calculation of Hybrid Magnetoelectric States in an Electron Waveguide", J. Appl. Phys., 71, 3027 - 3029 (1992).

S. Chaudhuri and S. Bandyopadhyay, ``Quantum Transport in a Disordered Quantum Wire in the Presence of a Magnetic Field", Superlattices and Microstructures, 11, 241 - 244 (1992).

N. Telang and S. Bandyopadhyay, ``A Monte Carlo Study of Correlations Between Impurity Scattering Events in a Two dimensional Electron Gas Causing Inhomogeneous Magnetoresistance, Superlattice and Microstructures 11, 99 - 102 (1992).

S. Subramaniam, S. Bandyopadhyay, B. Das, G. H. Bernstein, and P. A. Sekula-Moise, ``Anomalous Magnetoresistance Due to Correlations Between Disorder and Boundary Scattering in a Two-Dimensional Electron Gas Channel" Phys. Rev. B, 45, 3826 - 3829 (1992).

S. Bandyopadhyay, ``Coupling and Crosstalk Between High Speed Interconnects in Ultra Large Scale Integrated Circuits", IEEE J. Quant. Elec., 28, 1554 - 1561 (1992).

S. Bandyopadhyay, S. Chaudhuri, B. Das and M. Cahay, ``Features of quantum magnetotransport and electromigration in mesoscopic systems" Superlattices and Microstructures, 12, 123 - 132 (1992). M. Cahay and S. Bandyopadhyay,

N. Telang and S. Bandyopadhyay, ``Quenching of Acoustic Phonon Scattering of Electrons in Semiconductor Quantum Wires Induced by a Magnetic Field" Appl. Phys. Lett. (in press).

N. Telang and S. Bandyopadhyay, ``The Effects of a Magnetic Field on Electron Phonon Scattering in Quantum Wires by a Magnetic Field", submitted to Phys. Rev. B.

S. Bandyopadhyay, B. Das and A.E. Miller, "Supercomputing with Spin Polarized Single Electrons in a Quantum-coupled Architecture," submitted to J. Appl. Physics.

X. Huang, G. H. Bernstein, G. Bazán, and D. A. Hill, "Spatial Density of Lines in PMMA by Electron Beam Lithography," *J. Vac. Sci. Technol. B.*, in press.

G. Bazán and G. H. Bernstein, "Electron Beam Lithographv Over Very Large Scan Fields," *J. Vac. Sci. Technol. B*, in press.

D. A. Hill, X. Huang, G. Bazán, and G. H. Bernstein, "Swelling and Surface Forces-Induced Instabilities in Nanoscopic Polymeric Structures," *J. Appl. Phys.*, **72**(9), pp. 4088-4094 (1992).

X. Huang, G. Bazán, G. H. Bernstein, and D. A. Hill, "Stability of Thin Resist Walls," *J. Electrochem. Soc.*, **139**(10), pp. 2952-2956 (1992).

G. H. Bernstein and D. A. Hill, "On the Attainment of Optinum Development Parameters for PMMA Resist," *Superlat. and Microstruct.*, **11**, No. 2, PP. 237-240 (1992).

G. H. Bernstein, D. A. Hill and W. P. Liu, "New High-Contrast Developers for PMMA Resist," *J. Appl. Phys.*, **71** (8) pp. 4066-4075 (1992).

G. H. Bernstein, S. W. Polchlopek, R. Kamath, and W. Porod, "Determination of Electron-Beam-Induced Positive Oxide Charge," *Scanning*, Vol. 14, pp. 345-349 (1992).

Reprints and preprints of journal article follow. Articles appear in the order in which they are referenced in Section II.

MAGNETIC EDGE STATES IN A 1D CHANNEL WITH A PERIODIC ARRAY OF ANTIDOTS

Manhua Leng and Craig S. Lent
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556
(Received 19 May 1991)

Recent experiments in transport through a ballistic constriction with a periodical corrugation on one wall to modulate the width [1] have revealed miniband structure in the presence of a magnetic field. We have previously explored the properties of edge states in a similar structure[2]. Here we present our calculation of bandstructure and edge states on a 1D channel having an array of antidots placed along its central axis. Our interest is to examine the allowed states in such a channel and, by direct calculating the particle current density, to visualize the current flow via magnetic edge states as well as localized Landau states.

1. Introduction

Quantum transport of 2DEG systems in the presence of a perpendicular magnetic field has the remarble result of a quantized Hall resistance. A transport theory developed by Büttiker [3] has been employed to explain the integer quantum Hall effect of ballistic channel. In this formalism, magnetic edge states play an important role in the sense that the Hall resistance can be expressed in terms of the transmission properties of these edge channels. Recently Kouwenhoven *et al.*, studied the conductance of a periodically modulated channel and observed miniband formation in zero magnetic field and quantized conductance in a 2T field. We have investigated a similar but infinitely long corrugated channel .The bandstructure and current distribution of edge states in the channel have been discussed.

Here we extend this work to examine transport of an infinitely long 1D channel with an array of antidots sitting in the channel [4]. The geometry is shown in Figure 1. The bandstructure and the current distribution in a perpendicular magnetic field have been calculated.

2. Model And Method

The structure is shown in Figure 1 with dimensions indicated. The width of the channel is d, and an array of antidots with diameter D sitting at the channel center results in an effective periodic potential profile in $\hat{x}$ direction with period of a. In the results shown in the following section, $d=a$ and $D/a=3/10$. For simplicity, the potential outside the channel and inside the antidots is taken to be infinite. The potential in the conducting region of the channel is zero.

The single-band, effective-mass model has been employed in calculating the bandstructure by solving Schrödinger equation for Bloch-state wavefunctions,

$$\Psi(x,y) = e^{ikx}u_{n,k}(x,y). \tag{1}$$

with eigen-energy $E_n(k)$ and $m^* = 0.05m_0$. A magnetic field is applied in $\hat{z}$ direction and the Landau gauge is chosen so that the vector potential $\mathbf{A} = -yB\hat{x}$.

The resulting time-independent Schrödinger equation for $u_{n,k}(x,y)$ and $E_n(k)$ is,

$$\left(-\nabla^2 - 2i(ka - \beta y)\frac{\partial}{\partial x} + (ka - \beta y)^2 + V\right) u_{n,k}(x,y) = E_n(k)u_{n,k}(x,y). \tag{2}$$

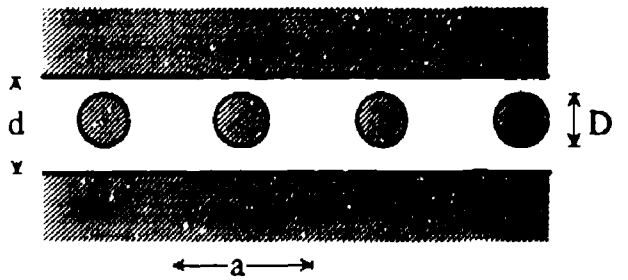


FIGURE 1. Geometry. Unit cell is marked with dashed lines.

0749-6036/92/030351+05 $02.00/0

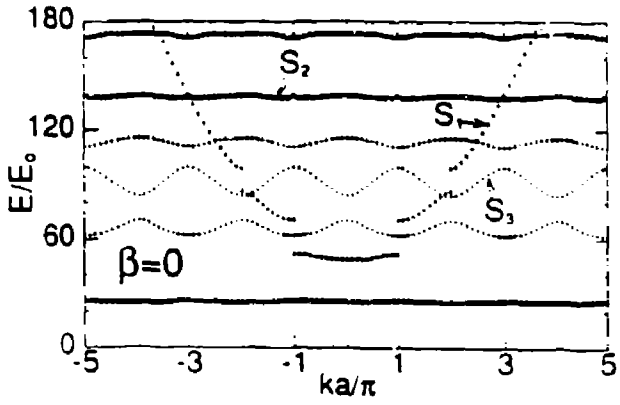


FIGURE 2. Shown is the bandstructure when $\beta = 0$ in extended zone scheme.
"••" - traveling states; "oo" - localized states; "++" - mixed states.

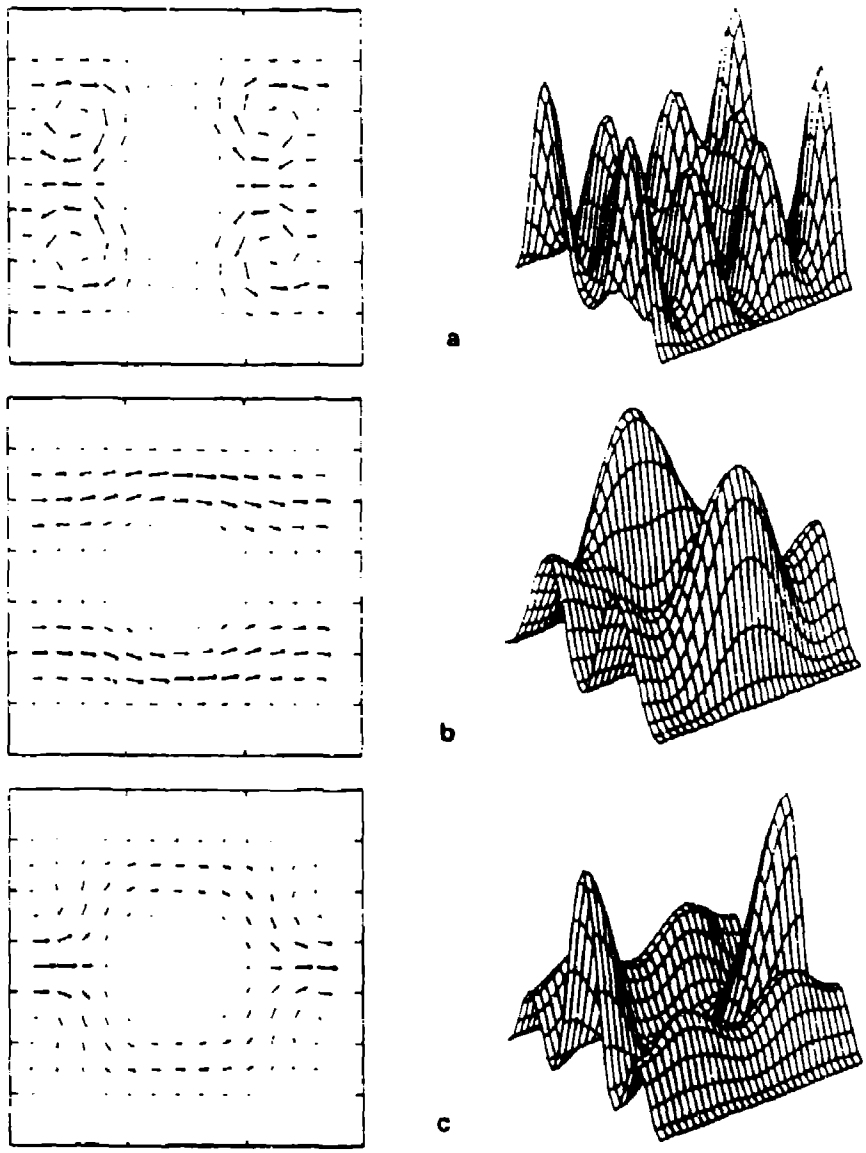


FIGURE 3. Shown are the current density and probability density distributions in one unit cellfor zero magnetic field. (a) localized state S1; (b) traveling state S2; (c) mixed state S3.

Here all the variables are expressed in a dimensionless fashion. Lengths are expressed in units of a, energy and potential are expressed in units of $E_0 \equiv \hbar^2/(2m^*a^2)$, and magnetic field is specified by $\beta \equiv e\beta a^2/\hbar$.

For a unit cell of a = 1000Å, a value of 30 for β corresponds to a 2 Tesla magnetic field.

We used the finite element method to discretize Equation (1). Numerical solutions for the first 10 eigen-energies and eigen-functions were obtained using a subspace iteration technique in first Brillouin zone. The current densities therefore were directly calculated from the wavefunctions.

3. Results

A. Bandstructure

Figure 2 shows the energy bandstructure for zero magnetic field. Minigaps are formed because the antidots placed in the channel act as a periodic potential. The remnant of a parabolic band("••" lines) represents the states traveling via the region between the antidots and the channel edge. Some bands ("oo" lines) are fairly flat through the whole Brillouin zone and represent those states which are confined in the wider region by two neighboring antidots. This confinement is effective up to very high energy. The left bands("++" lines) are mixtures of the above two. The particle current densities and probabilities in a unit cell for three typical states S_1, S_2, S_3 are shown in Figure 3.

We have calculated the bandstructures for β = 20, 30, and 40 but only present the result for β = 40 here. We find the energy bands change by filling in some of the minigaps as the magnetic field increases. At high field, as displayed in Figure 4, energy levels are resolved into quasi-parabolic band and a series of flat bands. The former

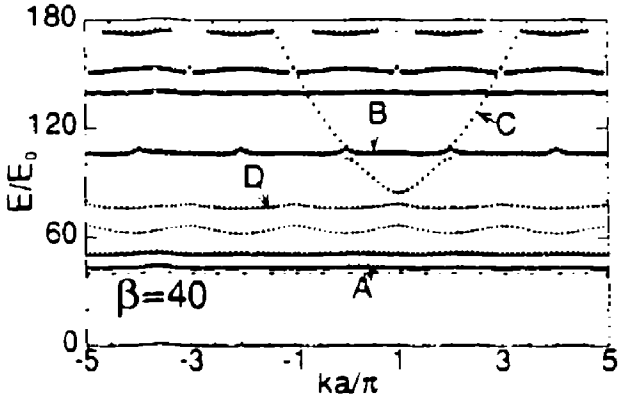


FIGURE 4. Shown is the bandstructure when β = 40 in extended zone scheme.
"••" - traveling edge states on the walls; "oo" - circulating edge states; "ΔΔ"–localized states between antidots; "++" - mixed states; dashed line is 2D bulk Landau Level.

emerged because the applied magnetic field pushes electron to the wall and suppresses back-scattering on the antidots. The flat bands consists of two major kinds of states. One kind originates from Landau orbits which are mainly localized in the wider region between two antidots(bulk Landau level energies are at $E_n/E_0 = (2n+1)\beta$, i.e., 40 and 120 for the first two with β = 40). Another kind of flat band belongs to the family of edge states: they are pushed not against either wall of the channel but against the antidot therefore circulate around the periphery of the antidot.

B. Magnetic Edge States

Edge states have been discussed in quantum transport but often indicated schematically. With the wavefunctions in hands from our numerical solution, we are readily able to calculate the current density in the structure. The particle current density, $j_{n,k}$, is calculated from the wavefunction as,

$$j_{n,k} = \frac{\hbar}{m^*}\left\{\left(k - \beta\frac{y}{a^2}\right)|u_{n,k}|^2 \hat{x} + \frac{i}{2}\left(u_{n,k}\nabla u^*_{n,k} - u^*_{n,k}\nabla u_{n,k}\right)\right\}. \quad (3)$$

Several typical states taken from Figure 4 are shown in Figure 5. State A corresponds to the lowest Landau level and is approaching its bulk value from above at this high field. State B is the magnetic edge state surrounding the antidot. This kind of edge states does not carry overall net current. State C is the magnetic edge state pushed to the wall of the channel by magnetic field. Current is mostly carried by this kind of state. State D is a mixed state of all kinds of states described above but is mainly localized.

4. Summary

We have calculated the bandstructure of a 1D channel with periodic potential barriers represented by an array of antidots at center along the channel. At zero magnetic field, minigaps form due to the periodic potential. Flat bands appear because of the constriction between the walls and the antidots. At high magnetic fields, some of the minigaps close because of the suppression of back-scattering by the magnetic field. There are two kinds of magnetic edge states present in the structure: circulating states which move around the antidot's periphery , and traveling edge states which move along the the channel walls. The circulating states carry no net current down the channel. Localized Landau orbits are also present.

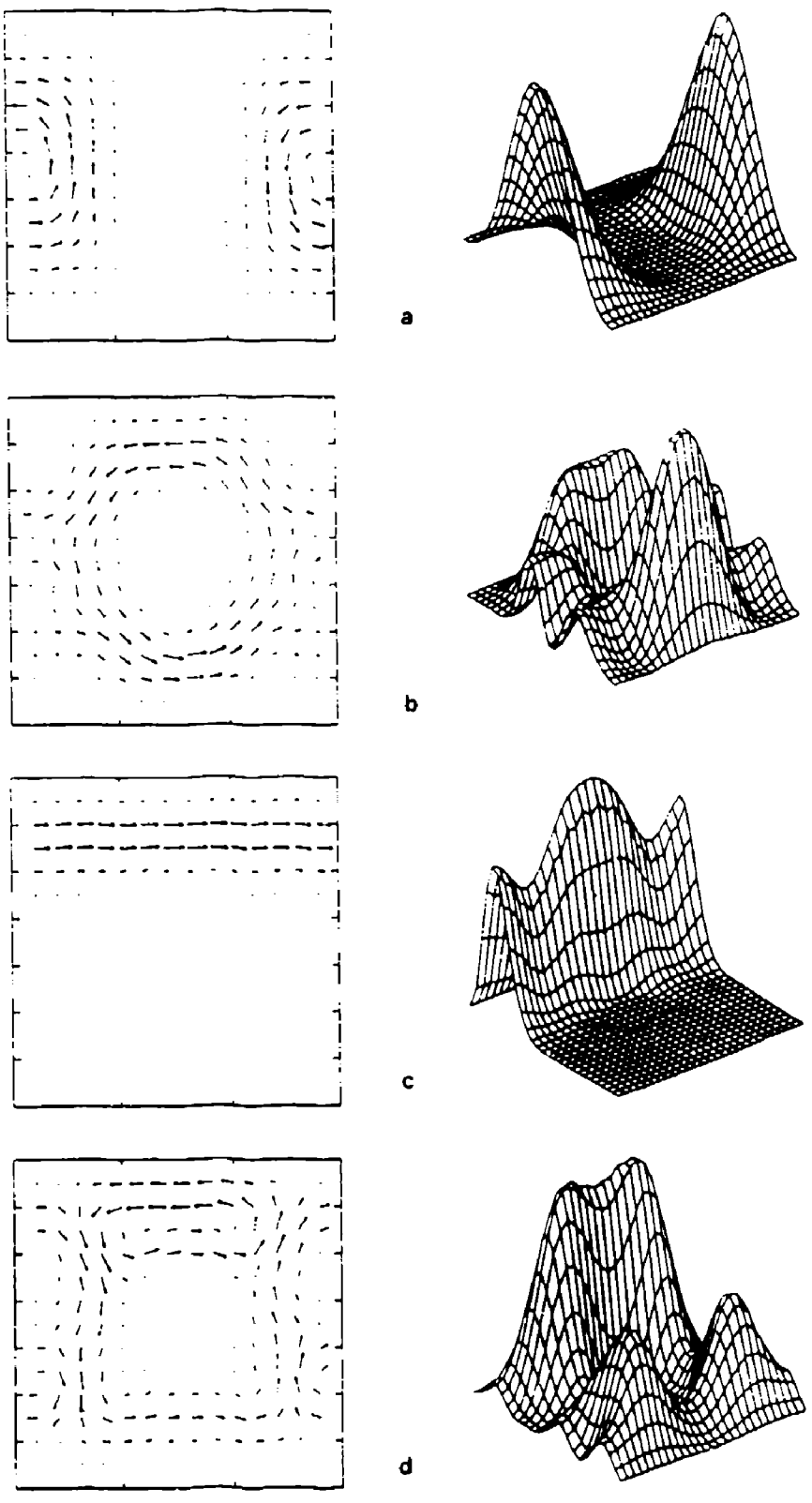


FIGURE 5. Shown are the current density and probability density distributions in one unit cellfor $\beta = 40$.
(a) first Landua orbit A; (b) circulating edge state B; (c) traveling edge state C; (d) mixed state D.

Acknowledgement - This work was supported by the Air Force Office of Scientific Research, by Texas Instruments, and by the National Science Foundation under grant number ECS890025 through the National Center for Computational Electronics, and utilized the Cray -2 at he National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign.

References

[1] L. P. Kouwenhoven, F. W. J. Hekking, B. J. von Wees, C. J. P. M. Harmans, C. E. Timmering, and C. T. Foxon, Phys. Rev. Lett. 65, 361 (1990).

[2] Craig S. Lent and Manhua Leng, Appl. Phys. Lett. 58, 1650 (1991).

[3] M. Büttiker, Phys. Rev. B 38, 9375 (1988).

[4] K. Kern, D. Heitmann, P. Grambow, Y-H. Zhang and K. Ploog, Phys. Rev. Lett. 66, 1618 (1991).

Recovery of quantized ballistic conductance in a periodically modulated channel

Manhua Leng and Craig S. Lent
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

(submitted to PRL April 6, 1993)

Abstract

We examine the ballistic conductance of a long, periodically-modulated electron channel. In general, the conductance of a modulated channel is a complicated function of the energy. For a long periodically-modulated channel, however, the quantized conductance characteristic of a simple quantum point contact is recovered. The value of the quantized conductance is no longer a monotonically increasing function of energy, however. Conductance can step down as well as up with increasing energy. We explain these results by comparing the calculated conductance with the bandstructure of the corresponding infinite periodically-modulated channel. The comparison shows a direct correspondence between the index of the conductance plateau and the number of positive-velocity bands at a given energy. The results persist in the presence of an applied magnetic field. In the high-field regime they can be interpreted as selective resonant back-scattering of edge states. We predict non-monotonic steps in the integer quantum Hall resistance in such structures.

PACS: 73.20.Dx, 72.10.-d, 73.50.Jt

Quantization of the conductance of a quantum point contact (QPC), such as those fabricated by lateral confinement of a two-dimensional electron gas in a semiconductor, is now well known and well understood [1-3]. The quantization is due to the creation of lateral sub-band modes, analogous to waveguide modes, and the fact that each mode carries the same amount of current. The linear-response conductance for a QPC is simply $G(E)=N(E)(2e^2/h)$, where $N(E)$ is an integer function of the energy which counts the number of travelling modes with energy below E. The conductance increases monotonically in steps as additional modes become available. If the ballistic channel is patterned with further features, bends, constrictions, or other obstructions, conductance quantization is lost and a complicated structure for $G(E)$ emerges due to the details of quantum interference and backscattering in the channel [4-5].

In this letter we show that if a ballistic channel has a periodically modulated structure, the quantization in $G(E)$ is recovered but it is no longer a monotonic function of energy. We numerically calculate transmission coefficients and the linear-response conductance of such a channel. We also calculate the energy bandstructure for the corresponding infinite modulated channel and show that simple features of the bandstructure explain the conductance behavior [6].

We examine the structure shown schematically in Fig. 1. A channel of width d is periodically narrowed to a width of d-h. The period of the modulation is a, the number of narrow regions is N, and the length of each narrow region is w. We present results for the particular case where $d/a = 2.0$, $h/a = 0.6$, and $w/a = 0.4$. A magnetic field of magnitude B is applied perpendicular to the plane in the $\hat{z}$ direction. The Landau gauge is chosen for the vector potential so $\vec{A} = -yB\hat{x}$. We adopt a single-band effective-mass model with an effective mass appropriate for GaAs, $m^*/m_0 = 0.067$, and ignore spin throughout. Hard wall potentials are assumed to define the channel modulation while the potential inside the channel is taken to be zero.

We solve the two-dimensional time-independent Schrödinger equation.

$$\left(\frac{-\hbar^2}{2m^*}\nabla^2 - \frac{ieh\hat{B}v}{m^*}\frac{\partial}{\partial x} + \left(\frac{e^2\hat{B}^2y^2}{2m^*}\right) + V_0(x,y)\right)\psi(x,y) = E\psi(x,y) \quad . \qquad (1)$$

To calculate the conductance of a channel of finite length and N periods, we solve Eq. (1) to obtain the complex energy-dependent transmission and reflection amplitudes for each transverse mode (defined in the wide regions). We solve directly for the wavefunction and transmission amplitudes in one unit cell (marked by dashed lines in Fig. 1), then use a scattering matrix cascading method [7] to obtain transmission and reflection amplitudes for the whole structure. Transmission and reflection into evanescent modes must be included in the cascading process. The conductance in the linear response regime is then obtained using the Landauer equation [8],

$$G = \frac{2e^2}{h}Tr(tt^{\dagger}) \quad , \qquad (2)$$

where $t_{i,j}$ is the transmission amplitude from mode j into mode i for the whole structure.

The numerical solution of Eq. (1) for the unit cell with open boundaries is accomplished using the Quantum Transmitting Boundary Method [9], a numerical algorithm we have developed based on the Finite Element Method for solving the two-dimensional Schrödinger equation for current-carrying states. We employ a recent extension of the method to include the case of an applied magnetic field [10].

We compare the conductance for the finite system with N periods of modulation with the bandstructure of the infinite periodic system. For the infinite system, we can use the Bloch theorem and look for a solution of the form $\psi_{n,k}(x,y) = e^{ikx}u_{n,k}(x,y)$ where $u_{n,k}(x,y)$ is the periodic part of the Bloch function. We solve Eq. (1) with this substitution as an eigenvalue problem for $E_n(k)$ and $u_{n,k}(x,y)$. We use the finite element method to achieve

numerical discretization over the unit cell. Meshes of up to 5151 nodes were used to achieve convergence at high values of magnetic field.

Figure 2 shows the calculated conductance and transmission coefficients $T_j = \sum_i |t_{i,j}|^2$ for a finite channel and the bandstructure for an infinite channel when no magnetic field is applied. Energy is expressed in units of E_1, the energy of the first transverse mode (subband) in the wide regions. Figure 2(a) shows the transmission of a channel with three unit cells (three constrictions). The total transmission for various incoming modes are shown separately. The energy bandstructure for the infinite modulated channel is shown for the first Brillouin zone in Figure 2(b). Energy is plotted on the horizontal axis so that the relationship to the conductance results below is clear. Figure 2(c) shows the conductance calculated from Eq. (2) for a long channel with N = 85 periods. Figure 2(d) shows the transmission for individual modes in the long modulated structure [11].

For only three constrictions in the channel the total transmission, shown in the solid curve of Figure 2(a), which is proportional to the conductance, is not quantized. This is to be expected because of mode mixing due to the abrupt narrowing of the channel [2,3]. The conductance of the long modulated channel, shown in Figure 2(c), is striking in that the conductance is essentially quantized in units of $2e^2/h$ [12]. Unlike the usual quantization of QPC ballistic conductance, however, the conductance does not increase monotonically but rather steps up *and down* between quantized levels, sometimes going to zero[13]. The ballistic conductance of the very long channel can be written as $G(E)=N_c(E)(2e^2/h)$, where $N_c(E)$ is the integer index corresponding to the quantized conductance plateau for energy E.

The conductance quantization in Figure 2(c) can be understood by examining the bandstructure shown in Figure 2(b). For each value of the energy, define an integer $N_b^+(E)$ to be the number of energy bands (distinct $E_n(k)$ curves) with positive group velocity (slope). $N_b^+(E)$ is also a non-monotonic function; it is zero in energy gaps and steps up

and down as a function of energy. By comparing Figures 2(b) and 2(c) we see that, in fact, $N_c(E)=N_b^+(E)$. The number of positive velocity bands for the infinite system yields the quantization of the conductance in the periodically-modulated finite system. The colored regions of the figures illustrate this correspondence. Each value of $N_c(E)=N_b^+(E)$ is represented by a different color (white=0, red=1, yellow=2, green=3, blue=4).

For a straight channel or a short ballistic point contact, $N_b^+(E)$ is simply the sub-band number and increases monotonically. The well known cancellation of velocity and density-of-states factors leads to identical current being carried in quasi-one-dimensional sub-band. Thus $N_c(E)$ is quantized and monotonic in the same manner.

The non-monotonic behavior of $N_b^+(E)$ in the periodically modulated channel is due to the significant amount of reflection and mode mixing caused by even a single constriction in the channel. Transport is by no means adiabatic. This mode mixing in the finite channel results in band mixing in the bandstructure of the infinite channel. The band mixing results in the appearance of forbidden gaps and allowed energy regions with differing numbers of energy bands. Just as sub-bands in a straight channel each carry the same amount of current, Bloch bands in the periodic structure each carry the same current.

These results have no strictly one-dimensional analogue. The bandstructure in a one-dimensional periodic system has the same number of bands (two, one with positive velocity and one with negative velocity) in each allowed region. $N_b^+(E)$ in one dimension only takes the values of 1 in a band or 0 in a gap. The integer values, (0,1,2,3, ...) of $N_b^+(E)$ for the modulated channel manifest the two-dimensional character of the channel mixing with the quasi-one-dimensional character of the current flow.

It is important to note that the transmission of individual modes is *not quantized* but the total transmission is. Figure 2(d) shows that for no individual mode is the conductance

quantized. The quantization occurs as the various modes are mixed by the periodic scattering.

The recovery of ballistic conductance quantization persists in the presence of a perpendicular magnetic field. Figure 3 shows the conductance and bandstructure calculated for the case of a moderately high magnetic field $\beta \equiv (\frac{eB}{\hbar})ad = 24.3$ (for a unit cell with a=200A, this corresponds to B= 20 T) [14]. Here conduction can be described in terms of edge states and the energy is naturally expressed in units of the first bulk Landau level $E_L = \hbar(\frac{eB}{2m^*})$.

In the limit of extremely high magnetic fields, all transport would be through edge states and the suppression of back-scattering between edge-states on opposite sides of the channel would guarantee monotonically increasing and quantized conductance [15]. The interesting result here is that for intermediate field strengths (or for higher edge-state indices), where backscattering can still occur between edge states, the conductance is nevertheless quantized. Again, however, the quantization is non-monotonic and related directly to the number of positive-velocity bands in the bandstructure for the corresponding infinite system. A step *down* in conductance as energy is increased means that an edge state which was contributing to the conduction has now been turned off at the higher energy because it resonantly back-scatters from the constrictions.

Notice in Figure 3(c) that the conductance is lowered by one unit for energies just above 6 E_L. Examination of the individual mode transmission coefficients in Figure 3(d) shows that it is the *first* edge state (with a slight admixture of the third) which has been resonantly reflected. For this energy range the second and third edge states are almost entirely transmitted but the first and outermost edge state is reflected. This selective reflection of edge states is similar to the experimental results of Müller *et al.* [16], who used an applied potential from a metal gate to reflect individual edge states. The consequence of this reflection was a deviation from the usual integer quantum Hall effect (IQHE) plateaus, a

deviation understandable in the edge-state picture of the IQHE [15]. The reflection of selective edge states seen in our calculation for a modulated channel suggests that a similar IQHE deviations, steps up and down between Hall resistance plateaus, should be observable in these geometries.

In conclusion, we have studied the ballistic transport properties of a periodically modulated channel. Our results show that a long modulated channel has a ballistic conductance which is quantized, but a non-monotonic function of energy. The index of a quantized conductance plateau has a one-to-one correspondence to the number of positive-velocity states in the energy bandstructure for the corresponding infinite modulated channel. This phenomenon persists at high magnetic field, where it can be interpreted as resonant reflection of particular edge states and should produce anomalous IQHE behavior.

ACKNOWLEDGMENTS —This work was supported by the Air Force Office of Science Research, and by the National Science Foundation. Numerical calculations were carried out on the Cray-2 at the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign. The authors would like to thank Dr. W. Porod for helpful discussions.

REFERENCES

[1] D.A. Wharam, T.J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J.E.F. Frost, D.G. Hasko, D.C. Peacock, D.A. Ritchie, and G.A.C. Jones, J. Phys. C **21**, L209 (1988); B.J. van Wees, H. van Houten, C.W.J. beenakker, J.G. Williamson, L.P. Kouwenhoven, D. van der Marel, and C.T. Foxon, Phys. Rev. Lett. **60**, 848 (1988).

[2] See discussion by G. Timp, R. Behringer, S. Sampere, J.E. Cunningham, and R.E. Howard, in *Nanostructure Physics and Fabrication* , Mark A. Reed and Wiley P. Kirk, eds., p 331 (Academic Press, Boston, 1989).

[3] E. Castano and G. Kirczenow, Phys. Rev. B **45**, 1514 (1992).

[4] A. Weisshaar, J. Lary, S.M. Goodnick and V.K. Tripathi, SPIE Proceedings, Vol. 1284, p. 45 (1990); S.E. Ulloa, E. Castano, and G. Kirczenow, SPIE Proceedings, Vol 1284, p. 57 (1990); Hua Wu and D. W. Sprung, Phys. Rev. B 47, 1500 (1993).

[5] Craig S. Lent, Srinivas Sivaprakasam and D.J. Kirkner, Solid State Electronics **32**, 1137 (1989); Craig S. Lent, Appl. Phys. Lett. **56**, 2554 (1990); Craig S. Lent, Appl. Phys. Lett. **57**, 1678 (1990); Henry Harbury, Wolfgang Porod, and Craig S. Lent, Superlattices and Microstructures **11**, 189 (1992).

[6] An experimental investigation of the conductance of a short modulated channel (16 periods) of this type has been performed. See L.P. Kouwenhoven, F. W. J. Hekking, B. J. von Wees, C. J. P. M. Harmans, C. E. Timmering, and C. T. Foxon, Phys. Rev. Lett. **65** , 361 (1990). Note that their results are for the energy region *below* the second conductance plateau and so probe only the purely one-dimensional aspect of this problem.

[7] S. Datta, *Quantum Phenomena*, p. 21, (Addison-Wesley, Reading, 1989).

[8] A. D. Stone and A. Szafer, IBM J. Res. Develop. 32, 384 (1988).

[9] Craig S. Lent and D.J. Kirkner, J. Appl. Phys. 67, 6353 (1990).

[10] Manhua Leng and Craig S. Lent (unpublished).

[11] Although the basic features of the quantization emerge with comparitively few periods (say N=20), the vary narrow plateaus visible in Figures 2 and 3 only reveal themselves for a rather long channel.

[12] Variations from quantized values are due to the still-finite length of the modulated region. Additionally an extremely fine energy grid would be necessary to completely resolve the rapid oscillations in conductance due to the finite lenth of the region.

[13] The appearance of conductance zeros corresponding to the gaps in the bandstructure has been noted by others. See reference 4.

[14] For a full examination of the bandstructure of this structure in a magnetic field see Craig S. Lent and Manhua Leng, Appl. Phys. Lett. 58, 1650 (1991), and Craig S. Lent and Manhua Leng, J.Appl. Phys. 70 , 3157 (1991).

[15] M. Büttiker, Phys. Rev. B 38 , 9375 (1988).

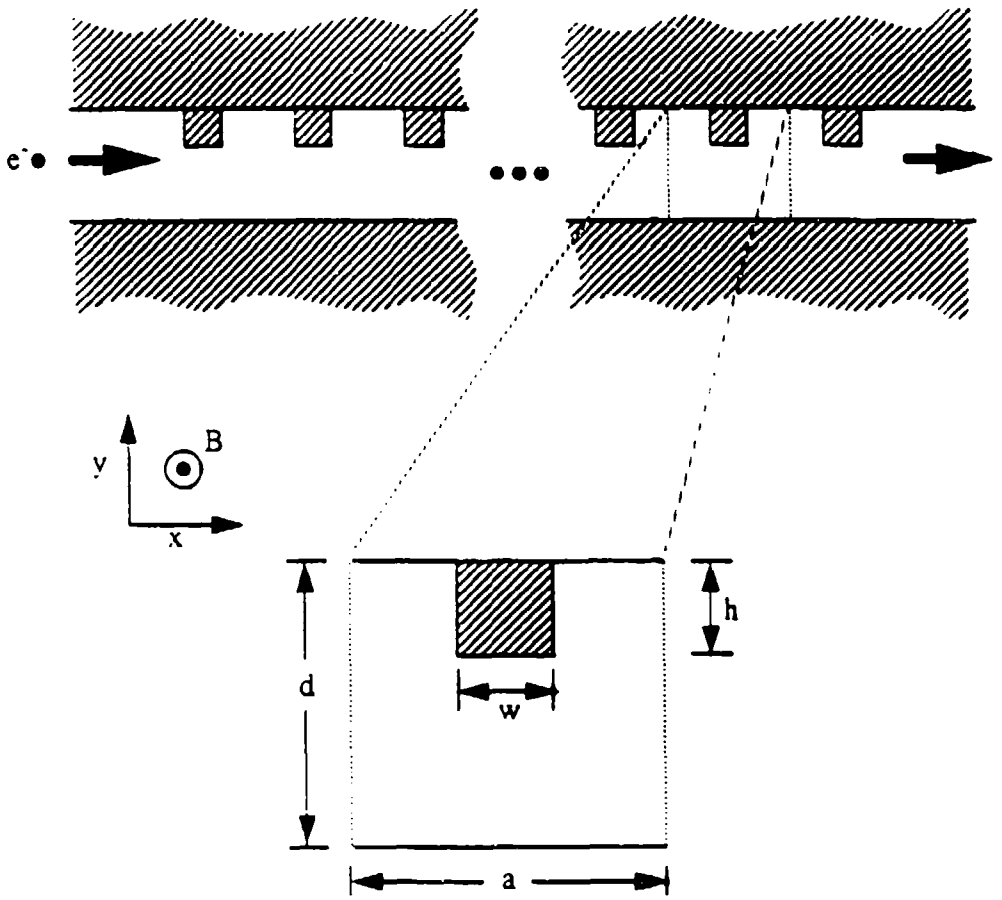
[16] G. Müller, D. Weiss, A.V. Khaetskii, K. von Klitzing, S. Koch, H. Nickel, W. Schlapp, and R. Lösch, Physical Review B 45, 3932 (1992).

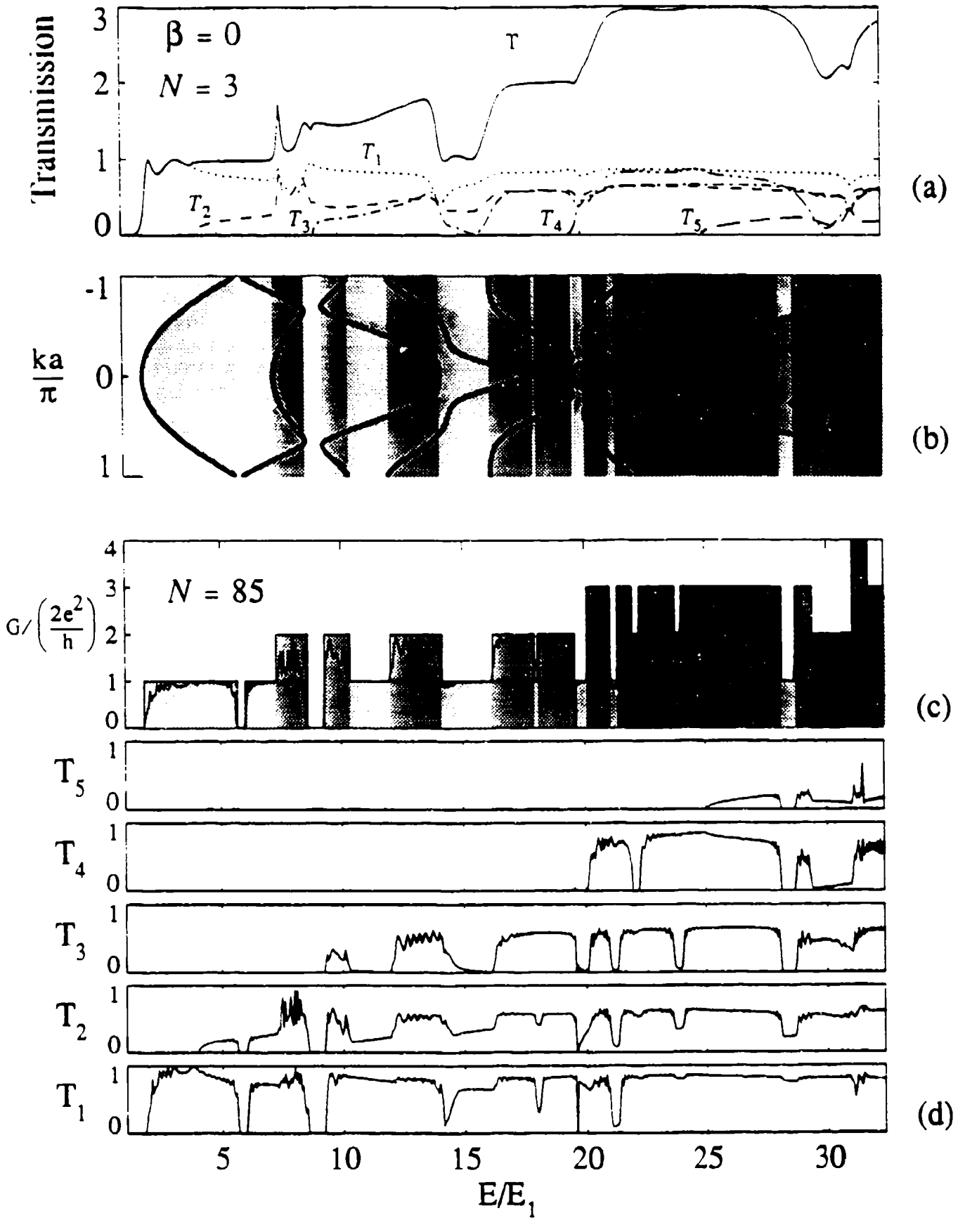
FIGURE CAPTIONS

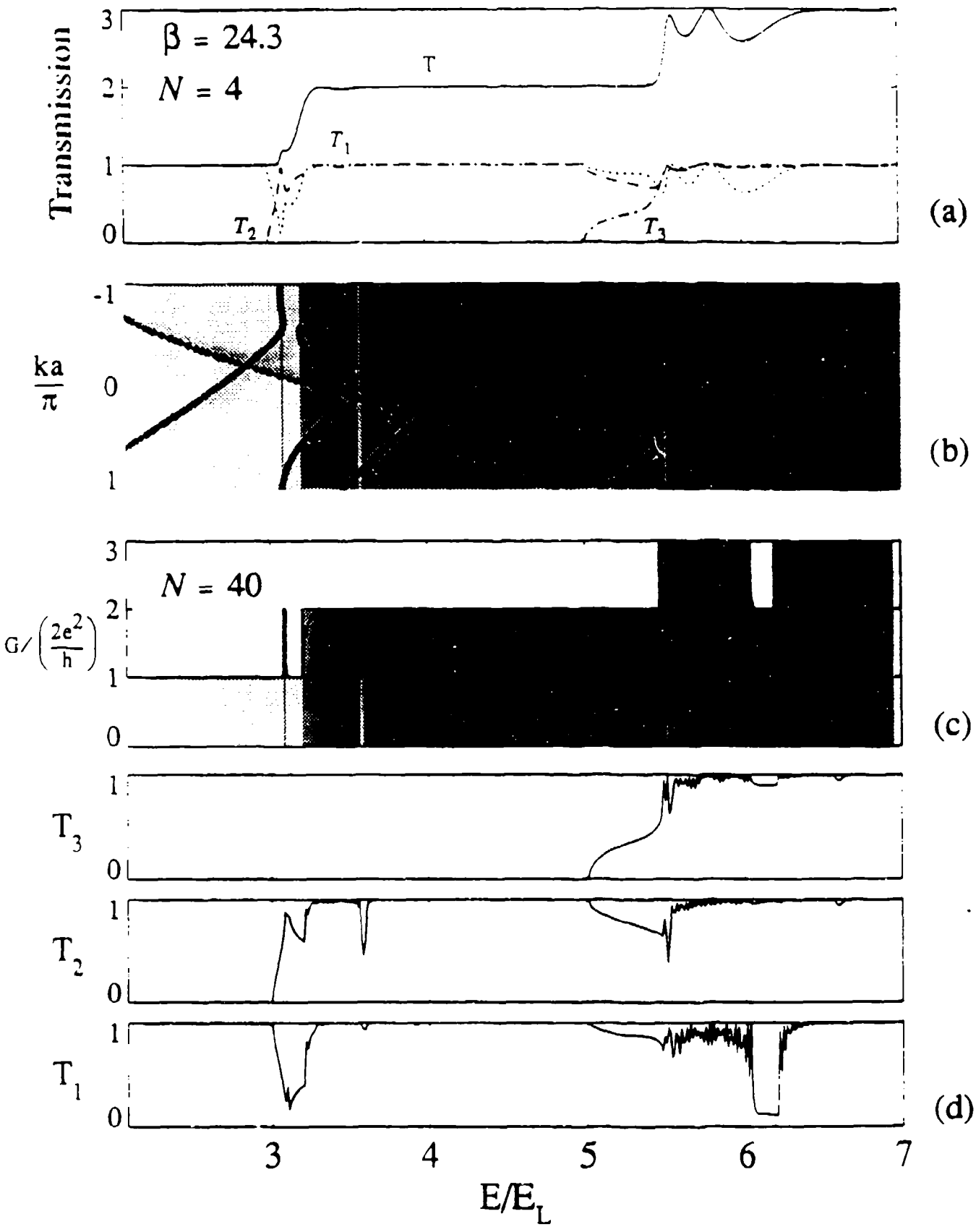
FIGURE 1. The geometry of the periodically modulated quantum channel.

FIGURE 2. The modulated channel in no applied magnetic field. (a) Transmission coefficients for a short channel with three unit cells (N=3). (b) The energy bandstructure for the infinite periodic channel. The colors indicate the number of individual energy bands with positive group-velocity in each energy region (white=0, red=1, yellow=2, green=3, blue=4). (c) Conductance for a long channel with 85 unit cells (N=85). Colors indicate the index of the quantization plateau (white=0, red=1, yellow=2, green=3, blue=4). The correspondence of the color schemes in (b) and (c) illustrates that the conductance of the finite channel is related to the number of positive-velocity bands in the bandstructure of the infinite channel. (d) Transmission coefficients of individual modes for the long modulated channel.

FIGURE 3. The modulated channel in an applied magnetic field. (a) Transmission coefficients for a short channel with four unit cells (N=4). (b) The energy bandstructure for the infinite periodic channel. The colors indicate the number of individual energy bands with positive group-velocity in each energy region (white=0, red=1, yellow=2, green=3, blue=4). (c) Conductance for a long channel with 40 unit cells (N=40). Colors indicate the index of the quantization plateau (white=0, red=1, yellow=2, green=3, blue=4). The correspondence of the color schemes in (b) and (c) illustrates that the conductance of the finite channel is related to the number of positive-velocity bands in the bandstructure of the infinite channel. (d) Transmission coefficients of individual modes (edge states) for the long modulated channel.


e⁻
y
B
x
h
d
w
a


Transmission
β = 0
N = 3
T
T1
T2
T3
T4
T5
(a)
ka/π
(b)
N = 85
G/(2e²/h)
(c)
T5
T4
T3
T2
T1
(d)
E/E1


Transmission
β = 24.3
N = 4
T
T_1
T_2
T_3
(a)
ka/π
(b)
N = 40
G/(2e²/h)
(c)
T_3
T_2
T_1
(d)
E/E_L

Bistable saturation in coupled quantum dots for quantum cellular automata

Craig S. Lent, P. Douglas Tougaw, and Wolfgang Porod
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 5 May 1992; accepted for publication 23 November 1992)

A simple model quantum dot cell containing two electrons is analyzed as a candidate for quantum cellular automata implementations. The cell has eigenstates whose charge density is strongly aligned along one of two directions. In the presence of the electrostatic perturbation due to a neighboring cell, the ground state is nearly completely aligned (polarized) in one direction only. The polarization is a highly nonlinear function of the perturbing electrostatic fields and shows the strong bistable saturation important for cellular automation function.

Nanoscale quantum structures with potential device applications have been an active area of exploration for several years. A frequent criticism[1] of many of these structures is the absence of the saturating behavior which forces conventional transistor elements into one of two stable states, "on" or "off." Such bistable saturation is important to keep device performance robust in the presence of physical inhomogeneities and noise.[2]

The possibility of realizing cellular automata (CA) with regular arrays of quantum dots has been suggested by Bate and others.[3] In one example, the necessary nonlinear response of each dot is the result of resonant tunneling through the dot.[4] We focus on a different paradigm in which each cell of the CA is composed of groups of coupled quantum dots. The confining potentials are such that electrons can tunnel between dots in the same cell but not between different cells. Quantum mechanics and the Coulomb interaction in each cell determine the possible cell states. The Coulomb interaction between electrons in different cells provides a local intercellular coupling mechanism. The nonlinear response of the cell to its electrostatic environment must be a feature of the internal cell dynamics. Recent success in fabricating arrays of very small quantum dots with one or two electrons per dot[5] prompts us to investigate possible few-electron coupled-dot cell geometries which provides the sort of bistable saturation so desirable. In this letter, we analyze a possible cell geometry with two electrons in the cell. We show that quantum confinement and the intracellular Coulomb interaction together yield the nonlinear saturation behavior which is essential.

We examine a simple nanostructure model cell containing five coupled quantum dots. The model cell is shown schematically in Fig. 1. It consists of a central site and four neighboring sites. Tunneling is possible both between the outer sites and the central site, and between adjacent outer sites. We first consider such a cell holding two electrons (the contrasting case of single and triple cell occupancy is discussed below). We show below that the Coulomb repulsion between the two electrons causes the ground state of the system to be one in which the electrons occupy antipodal sites.

We model the cell using a Hubbard-type Hamiltonian with Coulomb repulsion. The Hamiltonian for a single isolated cell can be written,

$$H_0^{\text{cell}} = \sum_{i,\sigma} E_0 n_{i,\sigma} + \sum_{i,j,\sigma} t_{i,j}(a_{i,\sigma}^{\dagger} a_{j,\sigma} + a_{i,\sigma}^{\dagger} a_{i,\sigma}) + \sum_{i} E_Q n_{i,\uparrow} n_{i,\downarrow} + \sum_{i>j,\sigma,\sigma'} V_Q \frac{n_{i,\sigma} n_{j,\sigma'}}{|r_i - r_j|}, \quad (1)$$

where the number operator $n_{i,\sigma} = a_{i,\sigma}^{\dagger} a_{i,\sigma}$ and the operator $a_{i,\sigma}^{\dagger}$ creates an electron at site i with spin σ. The cell parameters which define the Hamiltonian are then the on-site energy, E_0, the tunneling energies, $t_{i,j}$, and the on-site Coulomb charging energy, E_Q. The parameter V_Q is determined by fundamental constants and the dielectric constant of the material in which the dots are formed. A fixed positive charge $\bar{\rho}$ is assumed at each site sufficient to maintain overall cell charge neutrality. For an isolated cell, this only renormalizes E_0, but it is important in calculating the interaction between cells as is done below.

For the numerical results we discuss here we choose parameters based on a simple, experimentally accessible model. We consider a cell in a semiconductor with $m^* = 0.067 m_0$, which is composed of circular quantum dots of diameter $D = 10$ nm. The near-neighbor distance between the cells is 20 nm. The dielectric constant for the semiconductor is 10. We take $t = 0.3$ meV for coupling to the center site and $t = 0.03$ meV for coupling between outer dots. These tunneling energies can be varied greatly by adjusting the potential barriers between dots. We take $E_Q = V_Q/(D/3)$. We will assume here that the two electrons in the cell have antiparallel spins. The parallel spin case yields results which are qualitatively very similar.

The eigenstates of the Hamiltonian [Eq. (1)] can now be calculated for this specific choice of cell parameters. The Hamiltonian is diagonalized directly in the basis of few-electron states. From the two-electron wave function we calculate the single particle density at each site, ρ_i by find-

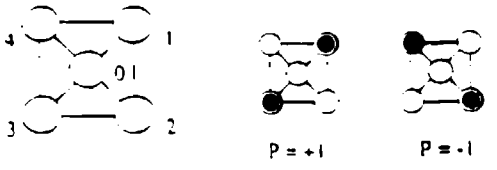


FIG. 1. The quantum cell consisting of five quantum dots which are occupied by two electrons. The mutual Coulomb repulsion between the electrons results in bistability between the $P = +1$ and $P = -1$ states.

 0003-6951/93/070714-03$06.00

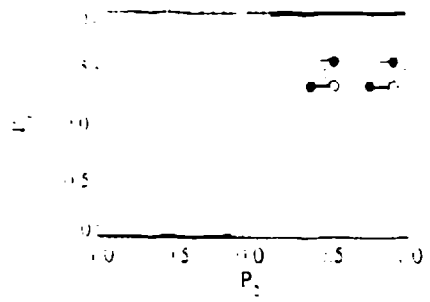


FIG. 2. The cell-cell response function. The induced polarization of cell 1 is shown as a function of the polarization in a neighboring cell 2 (inset). The solid line represents the polarization of the ground state and the dashed line represents the polarization of the first excited state.

ing the expectation value of the total number operator, $n_i = n_{i\uparrow} + n_{i\downarrow}$, at each site.

It is helpful to define a scalar quantity which represents the degree to which the electron density is aligned either along the line through sites 1 and 3, or along the line through sites 2 and 4. To this end we define the polarization of the cell as

$$P = \frac{(\rho_1 + \rho_3) - (\rho_2 + \rho_4)}{\rho_0 + \rho_1 + \rho_2 + \rho_3 + \rho_4}. \tag{2}$$

If sites 2 and 4 are vacant, the cell is completely in the $P = +1$ polarized state as shown in Fig. 1. If sites 1 and 3 are vacant, the cell is completely in the $P = -1$ polarized state. Clearly if the on-site energies are the same for all sites, the ground state is degenerate, comprising a combination of both polarizations, with no polarization preferred.

We examine the polarization of the low-lying eigenstates of the cell when perturbed by the presence of a nearby cell. We denote the target cell as cell 1 and the perturbing cell as cell 2. The potential at each site i of cell 1 is altered by the Coulomb interaction with the charge $\rho_{2,j}$ at site j of cell 2. The Hamiltonian for cell 1 can be written as the sum of the isolated cell Hamiltonian and a perturbation due to cell 2.

$$H^{\text{cell}} = H_0^{\text{cell}} + H_{12}^{\text{cell}}, \tag{3}$$

where

$$H_{12}^{\text{cell}} = \sum_{i,j,\sigma} V_Q \frac{\rho_{2,j} - \bar{\rho}}{|\mathbf{R}_{2,j} - \mathbf{R}_{1,i}|} n_{i,\sigma}. \tag{4}$$

Here $R_{m,i}$ denotes the position of site i in cell m. We solve for the eigenstates of the Hamiltonian [Eq. (3)] as the polarization of cell 2 is varied in the range $P_2 = [-1, 1]$. The occupancy of the central site in cell 2 is assumed to be zero, so that the charge densities, $\rho_{2,j}$ are simple functions of the polarization P_2. The distance between cell centers is three times the near-neighbor distance in a cell. For each value of P_2, we find the eigenstates and the associated charge densities and polarizations (Eq. 2). The result is the cell-cell response function—the polarization of cell 1 induced by a polarization of cell 2.

Figure 2 shows the polarization P_1 of the lowest two cell eigenstates as a function of the perturbing cell polarization P_2. This figure displays the central result of this letter—that the cell-cell response function is highly nonlinear and bistable. Even a very slight polarization of a cell induces nearly complete polarization of a neighbor.

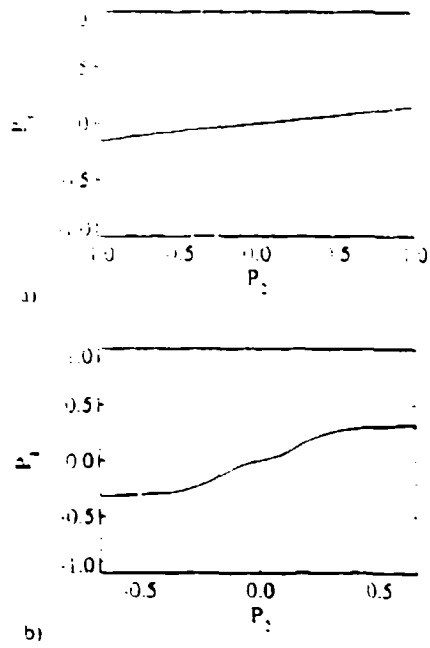


FIG. 3. The cell-cell response function for other cell occupancies. (a) A single electron in the cell. (b) Three electrons (two spin up and one spin down) in the cell.

The strongly nonlinear saturation of the polarization does not occur if only a single electron is in the cell.[7] Figure 3(a) shows the cell-cell response function for the ground state of a cell occupied by a single electron. The polarization is a very weak and nearly linear function of the perturbation.[8] The response function for a triply occupied cells (two spin up, one spin down) is shown in Fig. 3(b). Although the response is clearly not linear, it is not nearly as strong as the two-electron case. The bistable saturation present in the doubly occupied cells is a result of the distinct splitting of the degenerate ground state by the perturbation of a neighboring cell.

The rapid saturation of the polarization is the essential nonlinear effect which suggests this type of cell could provide the basis for a CA-type array. In such an array of cells, the differing polarizations of neighboring cells would provide the electrostatic perturbation which leads to a definite polarization of the ground state of the cell. It is possible to extract a CA rule set by finding the cell polarization induced from the various combinations of neighboring polarizations. This process, and the behavior of arrays of quantum cells will be discussed at greater length elsewhere.[9]

It must be noted that to date, quantum dot fabrication techniques have produced dots which tend to be rather far apart, thus only rather weak Coulombic coupling exists.[4] Our analysis presumes that fabrication techniques will shortly overcome these difficulties, possibly through macromolecular rather than semiconductor implementations.[10]

In conclusion, we have shown that strongly nonlinear saturation effects occur in a model two-electron nanoscale system. The charge density "snaps" into one of two positions, depending sensitively on asymmetries in the surrounding charge. This type of very desirable bistable saturation suggests the possibility of quantum cellular automata arrays based on this type of cell.

This work was supported in part by the Air Force Office of Scientific Research, the Office of Naval Research, and by the National Science Foundation under Grant Number ECS8900025 through the National Center for Computational Electronics, and utilized the Cray-2 at the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign. This material is based upon work supported under a National Science Foundation Graduate Fellowship.

R. Landauer, Physics Today 42, 119 (1989).
K. Obermayer, G. Mahler, and H. Haken, Phys. Rev. Lett. 58, 1792 (1987).
R. T. Bate, Bull. Am. Phys. Soc. 22, 407 (1977); J. N. R. Reed, and G. A. Frazier, J. Vac. Sci. Technol. B 7, 1398
Ferry, L. A. Akers, and E. W. Greeneich, *Ultra Large S tected Microelectronics* (Prentice Hall, Englewood Cliffs
J. N. Randall, A. C. Seabaugh, Y.-C. Kao, J. H. Luscon Newell, J. Vac. Sci. Technol. B 9, 2893 (1991).
B. Meurer, D. Heitmann, and K. Plong, Phys. Rev. L 1992).
A polarization with magnitude near 1 can be achieved o negligible charge density on the central site. That such str tion is likely is justified by the following results which strong polarization induced by a neighboring cell.
C. S. Lent, *Nanostructures and Mesoscopic Systems*, edit Kirk and M. A. Reed (Academic, Boston, 1992), p. 183.
Single electron cells have been discussed by P. Bakshi, D and K. Kempa, J. Appl. Phys. 70, 5150 (1991). The fact damental bistability in the response of a singly occupied cel approach less promising.
C. S. Lent, P. D. Tougaw, and W. Porod (unpublished).
For a survey see *Molecular Electronics: Materials and Me* by P. I. Lazarev (Kluwer, Boston, 1991).

QUANTUM CELLULAR AUTOMATA

To appear in *Nanotechnology*

Craig S. Lent, P. Douglas Tougaw, Wolfgang Porod, and Gary H. Bernstein

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

Abstract

We formulate a new paradigm for computing with cellular automata (CA) composed of arrays of quantum devices — quantum cellular automata. Computing in such a paradigm is *edge-driven*. Input, output, and power are delivered at the edge of the CA array only; no direct flow of information or energy to internal cells is required. Computing in this paradigm is also *computing with the ground-state*. The architecture is so designed that the ground state configuration of the array, subject to boundary conditions determined by the input, yields the computational result. We propose a *specific realization* of these ideas using two-electron cells composed of quantum dots, which is within the reach of current fabrication technology. The charge density in the cell is very highly polarized (aligned) along one of the two cell axes, suggestive of a two-state CA. The polarization of one cell induces a polarization in a neighboring cell through the Coulomb interaction in a very nonlinear fashion. Quantum cellular automata can perform useful computing. We show that AND gates, OR gates, and inverters can be constructed and interconnected.

1. Introduction

The continual down-scaling of device dimensions in microelectronics technology has led to faster devices and denser circuit arrays with obvious benefits to chip performance. Dramatic as they have been, these changes have been evolutionary in nature in that even the most advanced chips use the same paradigms for computing as their more primitive ancestors. There is now much expectation that the availability of very dense device arrays might lead to new paradigms for information processing based on locally-interconnected architectures such as cellular automata (CA) and cellular neural networks [1].

There has also been considerable interest in quantum mesoscopic structures for their possible application as devices[2]. Much has been learned about the behavior of electrons flowing through very small structures in semiconductors. Various investigators have pointed out the natural link between mesoscopic quantum systems and cellular automata architectures [3,4,5]. Because quantum structures are necessarily so small, it is difficult to conceive of a regime in which a single quantum device could drive many other devices in subsequent stages [6]. Furthermore, the capacitance of ultra-small wires forming the connections to each device would tend to dominate the behavior of an assembly of quantum devices. For these reasons locally interconnected structures such as cellular neural networks and cellular automata may provide the natural architecture for quantum devices.

We focus here on the idea of employing cellular automata (CA) architectures which are compatible with nanometer-scale quantum devices — thus, *quantum cellular automata* (QCA). A quantum cellular automaton would consist of an array of quantum device cells in a locally-interconnected architecture. The cell state becomes identified with the quantum state of the mesoscopic device. Two-state CA's are attractive because they naturally admit to encoding binary information. For a two-state QCA, each cell should have two stable quantum states. The state of a given cell should influence the state of the

neighboring cells. Two ingredients are essential then: 1) the bistability of the cell, and 2) coupling to neighboring cells.

We propose a cell which is composed of coupled quantum dots occupied by two electrons [7]. The requisite bistability is accomplished through the interaction of quantum confinement effects, the Coulomb interaction between the two electrons, and the quantization of charge [8]. The intercellular interaction is provided by the Coulomb repulsion between electrons in different cells. We analyze this cell and the interactions between neighboring cells in Section 2.

In Section 3 we propose a new paradigm for how computation could be done with an array of quantum devices. Because no direct connections can be made to interior cells, information or *energy* can enter the array only from the edges. *Edge-driven computation* imposes further constraints on the nature of the computing process [9]. The lack of direct connections to the interior cells also means that no mechanism exists for keeping the array away from its equilibrium ground-state configuration. We are therefore led to use the ground-state of the array to do the computation. *Computing with the ground state* means that the physics of the array must perform the computation by dissipating energy as it relaxes to the ground state. This has the distinct advantages that the computing process is independent of the details of the energy relaxation mechanisms and that the unavoidable energy dissipation is useful to the computing process.

Section 4 demonstrates that QCA's can perform useful functions. We show how logical gates and inverters can be constructed with arrays of the two-electron bistable quantum cell we propose. Section 5 discusses some key issues in realizing QCA's as a viable technology and Section 6 identifies technological advantages that a successful QCA implementation would enjoy.

2. Few-electron Quantum Cells

The specific cell we consider here is shown in Figure 1. Four quantum dots are coupled to a central dot by tunnel barriers. The two electrons tend to occupy antipodal sites in one of two configurations, shown in the figure as the $P=+1$ and $P=-1$ configurations. Our analysis below will show that the cell is indeed in one of these two stable states, and that an electrostatic perturbation, perhaps caused by neighboring cells, switches the cell between these two states in a very abrupt and nonlinear way. This permits the encoding of bit information in the cell.

The essential ingredients that produce the bistable saturation behavior [10] so desirable are 1) quantum confinement, 2) Coulomb interaction between electrons, 3) few-electron quantum mechanics, and 4) the discreteness of electronic charge.

A model for the quantum cell

We model the cell shown in Figure 1 using a Hubbard-type Hamiltonian. For the isolated cell, the Hamiltonian can be written,

$$H_0^{cell} = \sum_{i,\sigma} E_{0,i} n_{i,\sigma} + \sum_{i,\sigma} t\,(a_{i,\sigma}^{\dagger} a_{0,\sigma} + a_{0,\sigma}^{\dagger} a_{i,\sigma}) + \sum_i E_Q n_{i,\uparrow} n_{i,\downarrow} + \sum_{i>j,\sigma,\sigma'} V_Q \frac{n_{i,\sigma} n_{i,\sigma'}}{|\vec{R}_i - \vec{R}_j|} \quad (1)$$

Here $a_{i,\sigma}$ is the annihilation operator which destroys a particle at site i (i=0,1,2,3,4) with spin σ. The number operator for site i and spin σ is represented by $n_{i,\sigma}$. The on-site energy

FIGURE 1. The quantum cell consisting of five quantum dots which are occupied by two electrons. The mutual Coulombic repulsion between the electrons results in bistability between the P=+1 and P=-1 states.

for the i^{th} dot is $E_{0,i}$; the coupling to the central dot is t; the charging energy for a single dot is E_Q. The last term represents the Coulombic potential energy for two electrons located at sites i and j at positions R_i and R_j. Unless otherwise noted, we will consider the case where all the on-site energies are equal, $E_{0,i}=E_0$.

For our standard model cell, on which the numerical results reported here are based, we obtain the values of the parameters in the Hamiltonian from a simple, experimentally reasonable model. We take each site to be a circular quantum dot with diameter $D=10$ *nm*, and take E_0 to be the ground state energy of such a dot holding an electron with effective mass $m^* = 0.067\ m_0$. The near-neighbor distance between dot centers, a, is taken to be *20 nm*. The Coulomb coupling strength, V_Q, is calculated for a material with a dielectric constant of *10*. We take $E_Q=V_Q/(D/3)$ and $t=0.3$ *meV*.

It is useful to define a quantity which represents the degree to which the charge density for a given eigenstate of the system is aligned linearly. This alignment could be either along the line through sites *1* and *3* or along the line through sites *2* and *4*. For each site, we calculate the single particle density ρ_i, which is simply the expectation value of the total number operator for the two-electron eigenstate. The polarization, P, is defined as

$$P = \frac{(\rho_1 + \rho_3) - (\rho_2 + \rho_4)}{\rho_0 + \rho_1 + \rho_2 + \rho_3 + \rho_4}. \tag{2}$$

For an isolated cell with all on-site energies equal, no polarization is preferred. We will see below that perturbations due to charges in neighboring cells can result in a strongly polarized ground state. The polarization thus defined is not to be confused with the usual dipole polarization of a continuous medium. It simply represents the degree to which the electrons in the cell are aligned and in which of the two possible directions the alignment occurs. For the states of interest here, the cell has no dipole moment.

The interaction of the cell with the surrounding environment, including other neighboring cells, is contained in a second term in the Hamiltonian which we write as H_{inter}^{cell}. We solve

the time independent Schrödinger equation for the state of the cell, $|\Psi_n\rangle$, under the influence of the neighboring cells:

$$(H_0^{cell} + H_{inter}^{cell})\,|\Psi_n\rangle = E_n\,|\Psi_n\rangle\,. \tag{3}$$

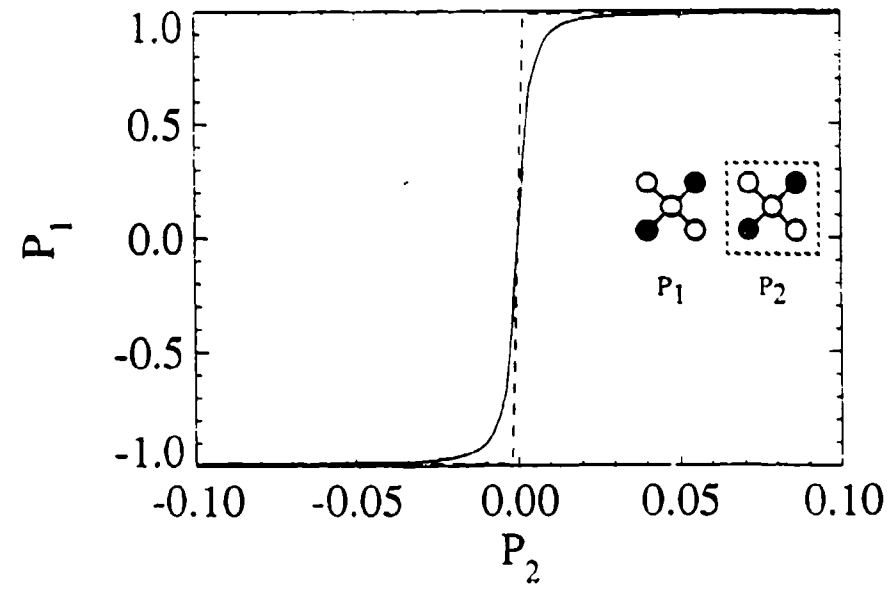
The spins of the two electrons can be either aligned or anti-aligned, with corresponding changes in the spatial part of the wavefunction due to the Pauli principle. We will restrict our attention to the case of anti-aligned spins here because that is the ground-state configuration; the spin-aligned case exhibits nearly identical behavior. The Hamiltonian is diagonalized directly in the basis of few-electron states. We calculate single particle densities, ρ_i, from the two-particle ground-state wavefunction $|\Psi_0\rangle$,

$$\rho_i = \sum_\sigma \langle\Psi_0|\, n_{i,\sigma}\,|\Psi_0\rangle, \tag{4}$$

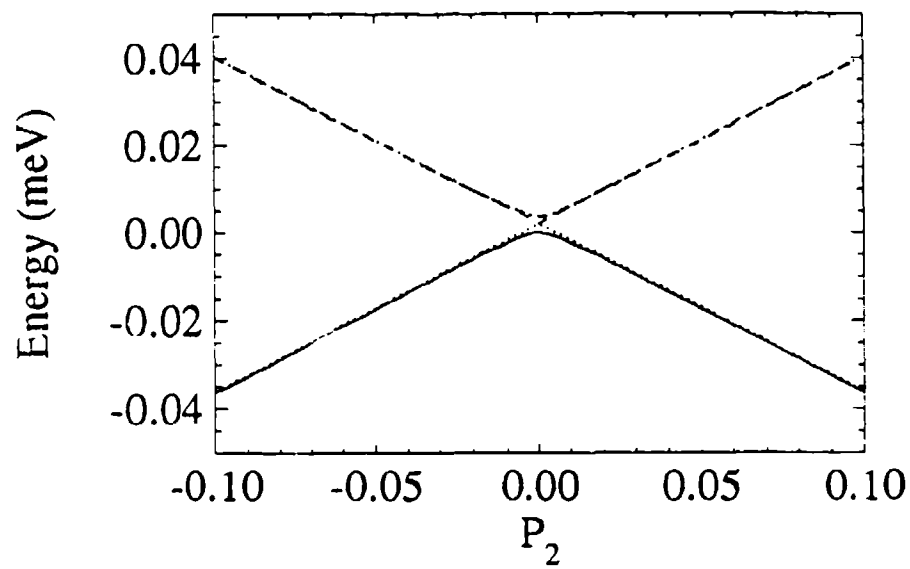
and from the densities calculate the resultant polarization P (EQ. 2). To maintain charge neutrality, a fixed positive charge, $\tilde{\rho}$, with magnitude $(2/5)e$ is assumed at each site. For the isolated cell, this has no effect and is included in the on-site energies. For several cells in close proximity, as will be considered below, the maintenance of overall cell charge neutrality means that the intercellular interaction is due to dipole, quadrupole, and higher moments of the cell charge distribution. If cells had a net total charge then electrons in cells at the periphery of a group of cells would tend to respond mostly to the net charge of the other cells.

The cell-cell response function

To be of use in a CA architecture, the polarization of one cell must be strongly coupled to the polarization of neighboring cells. Consider the case of two nearby cells shown in the inset to Figure 2. Suppose we fix the charge distribution in the right cell, labeled cell *2*. We assume cell *2* has polarization P_2, and that the charge density on site *0* is negligible (this means the charge density is completely determined by the polarization). For a given polarization of cell *2*, we can compute the electrostatic potential at each site in cell *1*. This



a)



b)

FIGURE 2. The cell-cell response function. The polarization of the right cell is fixed and the induced polarization in the left cell is calculated. The top figure shows the calculated polarization of cell 1 as a function of the polarization of cell 2. Note that the range of P_2 shown is only from -0.1 to +0.1. This is because the transition in the induced polarization is so abrupt. The lower figure shows the first four eigen-energies of cell 1. The polarization of the lowest two are shown in the first figure.

additional potential energy is then included in the total cell Hamiltonian. Thus the perturbing Hamiltonian is

$$H_{inter}^{cell} = H_1^{cell} = \sum_{i \in cell\ 1, \sigma} V_i^1 n_{i,\sigma}, \tag{5}$$

where

$$V_i^m = \sum_{k \neq m, j} V_Q \frac{(\rho_j - \bar{\rho})}{|\vec{R}_{k,j} - \vec{R}_{m,i}|} \tag{6}$$

is the potential at site i in cell m due to the charges in all other cells k. We denote the position of site j in cell k as $\vec{R}_{k,j}$. The total Hamiltonian for cell *1* is then

$$H^{cell} = H_0^{cell} + H_1^{cell}. \tag{7}$$

The two-electron Schrödinger equation is solved using this Hamiltonian for various values of P_2. The ground state polarization of cell *1*, P_1, is then computed as described in the previous section.

Figure 2b shows the lowest four eigen-energies of cell 1 as a function of P_2. The perturbation rapidly separates states of opposite polarization.The excitation energy for a completely polarized cell to an excited state of opposite polarization is about *0.8* meV for our standard cell. This corresponds to a temperature of about 9 K. Figure 2a shows P_1 as a function of P_2 — the cell-cell response function. A very small polarization in cell *2* causes cell *1* to be very strongly polarized. This nonlinear response is the basis of the CA's we describe here. As the figure shows, the polarization *saturates* very quickly. This observation yields two important results:

1. ***The bipolar saturation means that we can encode bit information using the cell polarization.*** A cell is almost always in a highly polarized state with $P \approx \pm 1$. We define the P=+1 state as a bit value of 1 and the P=-1 state as a bit value of 0. Only if the electrostatic environment due to other cells is nearly perfectly symmetric will there be no polarization.

2. ***The polarization of one cell induces a polarization in its neighbor.*** Figure 2 shows that even a very slight polarization will induce nearly complete polarization of a neighboring cell. This cell-cell Coulomb coupling provides the mechanism for CA-like behavior. The rapid saturation of the cell-cell response function is analogous to the gain necessary to preserve digital logic levels from stage to stage.

The abruptness of the cell-cell response function depends on the ratio of the dot-to-dot coupling energy, t in Eq. (1), to the Coulomb energy for electrons on different sites. The magnitude of the coupling depends exponentially on both the distance between the dots and the height of the potential barrier between them [11], each of which can be adjusted as engineering parameters. Figure 3 shows how the cell-cell response function varies with t.

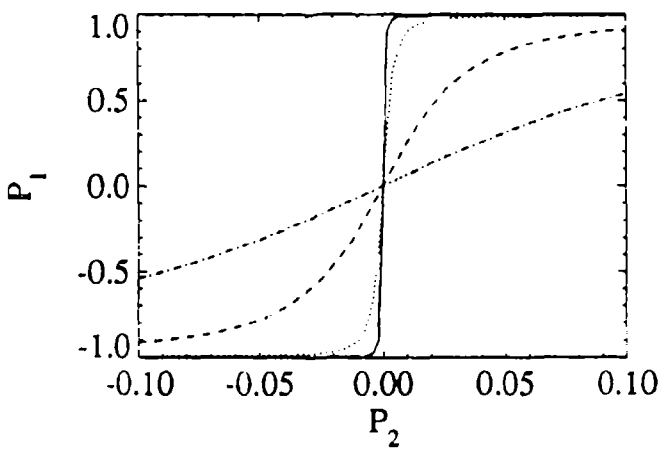


FIGURE 3. The cell-cell response function for various values of the dot-to-dot coupling energy (t in Eq.1). The induced cell polarization P_1 is plotted as a function of the neighboring cell polarization P_2. The results are shown for values of the coupling energy, t=-0.2 (solid),-0.3 (dotted),-0.5 (dashed), and -0.7 (dot-dashed) meV. Note that the response is shown only for P_2 in the range [-0.1, +0.1].

Self-consistent analysis of several quantum cells

In the analysis of the previous two sections, the two-electron eigenstates were calculated for a single cell. It is important to note that for the Hamiltonian employed, these are exact two-particle eigenstates. Exchange and correlation effects have been included exactly. This was possible because we could explicitly enumerate all possible two-electron states and diagonalize the Hamiltonian in this basis set. We want to analyze clusters and arrays of cells to investigate possible device architectures. To do so we need to calculate the ground-state wavefunction of a group of cells. Exact diagonalization methods are then no longer tractable because the number of possible many-electron states increases so rapidly as the number of electrons increases. We must therefore turn to an approximate technique.

The potential at each site of a given cell depends on the charge density at each site of all other cells. We will treat the charge in all other cells as the generator of a Hartree-type potential and solve iteratively for the self-consistent solution in all cells. This approximation, which we call the Intercellular Hartree Approximation (ICHA), can be stated formally as follows. Let Ψ_0^k be the two-electron ground-state wavefunction for cell k, and ρ_j^m be the single particle density at site j in cell m. We begin with an initial guess for the densities. Then, for each cell we calculate the potential due to charges in all other cells.

$$V_i^k = \sum_{m \neq k, j} V_Q \frac{(\rho_j^m - \bar{\rho})}{|\vec{R}_{m,j} - \vec{R}_{k,i}|}. \tag{8}$$

Although the neighboring cells will normally dominate this sum, we do not examine only near-neighbors but include the effect of all other cells. For each cell k, this results in a perturbation of the basic cell Hamiltonian of equation (1):

$$H_k^{cell} = \sum_{i \in cell\ k, \sigma} V_i^k n_{i,\sigma}. \tag{9}$$

The Schrödinger equation for each cell is now solved for the two-electron ground state eigenfunction:

$$(H_0^{cell} + H_k^{cell}) \, |\Psi_0^k\rangle = E_0^k \, |\Psi_0^k\rangle. \tag{10}$$

From the ground state eigenfunctions we calculate the improved single particle densities.

$$\rho_j^k = \sum_\sigma \langle \Psi_0^k | \, n_{j,\sigma} \, |\Psi_0^k\rangle \tag{11}$$

The improved densities are then used in Equation (8) and the system is iterated until convergence is achieved. Once the system converges, the many-electron energy, E_{total}, is computed from the sum of the cell eigen-energies using the usual Hartree correction term to account for over-counting of the Coulomb interaction energy between cells:

$$E_{total} = \sum_k E_k^0 - \sum_{k>q,\, i,\, j} V_Q \frac{\rho_i^k \rho_j^q}{|\vec{R}_{k,i} - \vec{R}_{q,j}|}. \tag{12}$$

It should be stressed that the ICHA still treats Coulombic, exchange, and correlation effects between electrons in the *same* cell exactly. The Hartree mean field approach is used to treat self-consistently the interaction between electrons in *different* cells. Since electrons in different cells are physically distinguishable (there being no wavefunction overlap), the exchange coupling between them is zero. The Hartree and Hartree-Fock approximations are therefore equivalent in this case.

The converged ICHA solution will be an (approximate) eigenstate of the entire system. In general, however, it need not be the ground state. As with the usual Hartree approximation, which of the eigenstates the scheme converges to is determined by the choice of the initial guess. To find the ground state we must try many initial state guesses and determine which converged solution has the lowest energy. Typically, this does not present a serious problem for the type of cellular arrays considered here because the set of likely ground states is easily discerned. In general, a systematic search may be required.

The procedure described above uses, at each stage of the iteration, only the ground-state wavefunction of each cell. If all the excited states of the entire system were desired, we would have to include states composed of excited cell states as well. Since our interest is

in the ground-state, this is not necessary. It is relevant to point out however, that because each cell is in a "local" ground state, we do not require coherence of the many-electron wavefunction across the whole array of cells. All that is required to support this analysis is that the wavefunction is *coherent across a single cell.* No information about the phase of the wavefunction in other cells is relevant to the wavefunction in a given cell — only the charge densities in other cells need be known.

3. Computing with Quantum Cellular Automata

We present a new paradigm for computing with quantum cellular automata. This represents a complete picture of how quantum devices could be coupled in a CA architecture to perform useful functions. The paradigm we propose is shown schematically in Figure 4. We will focus on the zero temperature case; temperature effects will be considered below. As shown in the figure, the inputs are along an edge of the array. Specifying the inputs consists of electrostatically fixing the polarization of the input cells. This could be accomplished by simply applying voltages to conducting "set" lines which come in close proximity to the cells, but any method that fixes the cell polarization state would do. The output cells are not fixed; their polarization state is sensed, perhaps by electrostatic coupling to "sense" lines. There could also be several input and output edges. Computation proceeds in the following steps:

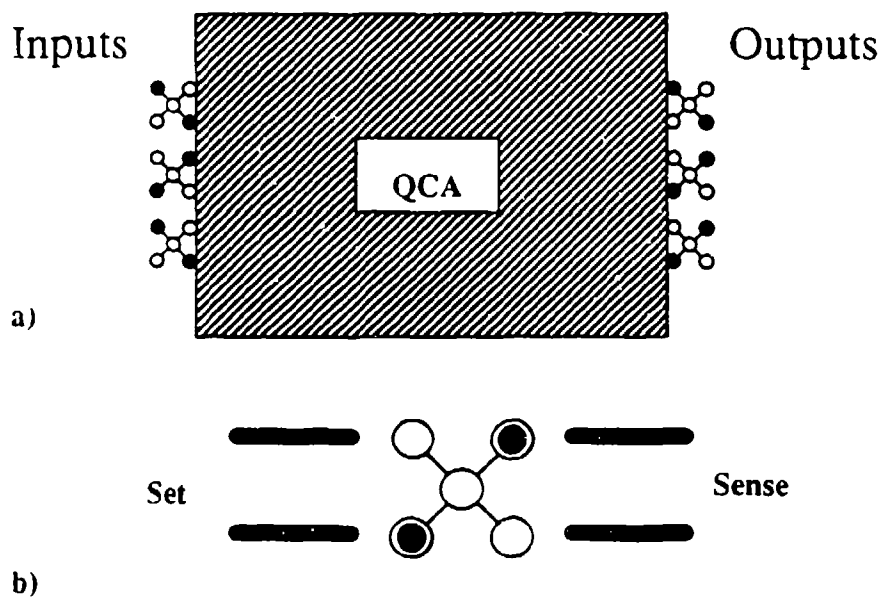


FIGURE 4. The new paradigm for computing with quantum cellular automata. The input to the QCA is provided at an edge by setting the polarization state of the edge cells (*edge-driven computation*). The QCA is allowed to dissipatively move to its new ground-state configuration and the output is sensed at the other edge (*computing with the ground state*). The "set" and "sense" lines are shown schematically.

1. Write the input bits by fixing the polarization state of cells along the input edge (edge-driven computation).

2. Allow the array to relax to its ground state with these inputs (computing with the ground state).

3. Read the results of the computation by sensing the polarization state of cells at the output edge.

The essential elements that define this computing paradigm are *computing with the ground state* and *edge-driven computation*, which we discuss below.

Computing with the Ground State

The advantage of computing with the ground state is that it leaves the computing process insensitive to the details of the dissipative processes which couple the electrons in the array to the environment. Consider a QCA at zero temperature for which all the input cells have been held in a fixed state. Dissipative processes have brought the array to its ground state configuration for these boundary conditions. Suppose at time $t=0$ the input cell states are set to their new input values completely abruptly. Just after the inputs are applied at the edge of the QCA, the array is no longer in the ground state but is now in an excited non-stationary state for the new boundary conditions. In the time between 0 and t_r, a characteristic relaxation time, various dissipative processes will bring the array to its new ground-state configuration. After that, the array will be stable until the boundary conditions are changed again. During the relaxation time the temporal evolution of the system is *very* complicated. Even without dissipation, the system will undergo quantum oscillations due to interference between the various eigenstates which compose the $t=0^+$ state. The dissipative processes, like phonon emission, introduce extraordinary complication in the temporal evolution. The exact state of the system at a particular time $t<t_r$ depends not just on phonon emission rates, but on the *particular* phonons emitted by

these *particular* electrons. In short, the temporal evolution before $t=t_r$ depends on the precise microscopic details of the dissipative dynamics. By contrast, *the ground state configuration to which the system relaxes is completely independent of the dissipation mechanisms*. Hence we choose to do computing with the ground state only.

Edge-driven Computation

In the QCA computing paradigm we are proposing, the input data is represented by edge cells whose polarization is fixed. Computing then proceeds by allowing the physics interior to the QCA to "solve" the dissipative many-electron problem for this new set of boundary conditions. The array is designed so the part of the ground-state "solution" of the many-body problem which appears at the output edge corresponds to the solution of the computing problem posed by the input data.

The advantage of writing input and reading output only at the edges of the array is that no separate connections to the array interior need be made. Because quantum devices are of necessity extremely small, the problem of making contacts to each element or device becomes severe. If a single array contains thousands of individual cells, the "wiring" problem is overwhelming.

Edge-driven computation is, in fact, the practical requirement which makes *computing with the ground state* necessary. If no connections can be made to the interior of the array, there is no controlled mechanism for keeping the system away from the ground state. Neither clocking nor refresh mechanisms are available. With a change in input, the system will dissipate energy and find a new equilibrium ground state. The only choice is whether to try to do computation with the system's transient response, or with its ground state. For the reasons discussed above, the ground-state approach is preferable.

Conventional computing, by contrast, is done using very highly-excited, non-equilibrium states. Because each element (device) can be separately contacted, energy can be fed into

the system at each point. The entire system can thereby be maintained in non-equilibrium states. The advantage of this is that the energy difference between the states used for computing can be very much larger than $k_B T$. The requirement that each element be driven far from equilibrium ultimately contributes to the difficulty of reducing the scale of conventional technology to the nanometer level. The breakdown of the operating device physics at small scales also plays a crucial role in the scale-down problem.

Ultimately, temperature effects are the principal problem to be overcome in physically realizing the QCA computing paradigm. The critical energy is the energy difference between the ground state and the first excited state of the array. If this is sufficiently large compared with $k_B T$, the system will be reliably in the ground state after time t_r. Fortunately, this energy difference increases quadratically as the cell dimensions shrink. If the cell size could be made a few Angstroms, the energy differences would be comparable to atomic energy levels — several electron volts! This is, of course, not feasible with semiconductor implementations, but may ultimately be attainable in molecular electronics. It may, however, be possible to fabricate cells in semiconductors small enough to work reliably at reasonable cryogenic temperatures.

Relation to synchronous CA rules

The relationship between the Quantum Cellular Automata described here and traditional rule-base CA's is not direct. Cellular automata are usually described by a set of CA rules which govern the temporal evolution of the array [12]. Time proceeds in discrete increments called generations. The rules determine the state of the array based on its configuration in the previous generation. Clearly, for the QCA described here, the temporal evolution proceeds not through discrete generations but through continuous physical time. Moreover, as argued above, we are not particularly interested in the temporal evolution of the QCA in order to do computing. We are only concerned with the final ground state configuration associated with a particular input state. Like the rule-based

synchronous CA, the QCA is an array of interacting multi-state cells and the behavior is dominated by near-neighbor interactions between cells. Thus, the QCA is chiefly related to traditional CA's by *analogy*.

Nevertheless, it is possible to construct a rule-based CA from the QCA interacting cell Hamiltonian (Eq. 10). The CA so constructed may be useful, perhaps not in describing the transient state of the QCA, but rather in calculating the ground state configuration, which is our primary concern anyway.

CA rules from the Schrödinger equation

The CA rule set is constructed as follows. For each cell, consider all possible polarization states ($P = \pm 1$) of the neighbors (neighbors out to any distance useful can be considered). For each configuration of the neighboring polarization, solve the Schrödinger equation (Eq. 10) and determine the target cell ground-state and its polarization. The map of neighbor polarizations to target cell polarization constitutes the CA rule set for that particular target cell. In general, a different rule set may apply to each cell. Typically, many cells will have similar environments and use the same rules.

The rule set obtained by this procedure can be recast in terms of a *weighted voting* procedure. In deciding the state of a particular cell, the neighboring cells vote according to their own state. The votes are weighted differently depending on the geometrical relationship between each neighbor and the target cell. The votes of closer cells are weighted more heavily than those of more distant cells. In addition, the weights can be negative, indicating that the energetics of the interaction between the neighbor and the target cell favor them having opposite polarizations. The CA rules generated by the solution of the Schrödinger equation for the target cell can then recast in the form of voting weights for the neighbors. Any set of voting weights which reproduces the CA rule set is equivalent.

Extended CA rules

This procedure so far has one problem which can be remedied by expanding the rules slightly. It is possible for the votes of the neighbors to result in a "tie". That is, the neighboring polarizations may be arranged so symmetrically that the ground-state polarization of the target cell is zero. It is desirable to break this tie by consulting the immediate history of the neighbors. The neighbors which flipped their polarization in the preceding generation are simply weighted more heavily than those which have not flipped. This introduces a notion of momentum which is otherwise absent in a two-state CA. With these *momentum rules*, ties are still possible but are now exceedingly rare events that can be handled by tie-breaking with a random number.

The CA rules corresponding to a particular QCA are thus derived from the Schrödinger equation and augmented by the momentum rule discussed above. The evolution of the synchronous CA is still not directly related to the temporal evolution of the physical QCA. The CA rules know nothing of the details of the dissipative dynamics, for example. However, in our experience, the synchronous CA with the momentum rules can be useful in determining the ground state of the QCA. If we start with a stable QCA state, and then flip the input cells to correspond to the new input condition, the synchronous CA will evolve to a stationary state which corresponds to the ground state of the physical QCA. That the final state is really the ground state can be checked by using the more rigorous self-consistent calculations described in the previous section.

4. Device Applications

Two types of QCA structures for computing can be envisioned. One type is a very large regular array of cells. We have work in progress exploring this type of array. It is widely appreciated that doing computing with large regular CA's is a significant challenge, particularly with a simple rule set. The solution to this difficult problem may have the greatest long-term potential, however, for exploiting the massive parallelism inherent in the QCA paradigm.

A second type of QCA structure involves a highly irregular array of cells. We show below that using simple irregular arrays one can produce structures analogous to wires, inverters, AND gates and OR gates. Since these can be connected together, more-complex devices such as adders and multipliers can be constructed. Because the individual devices are so small, this represents a potentially enormous increase in functional density in an architecture free of the usual interconnect problems. We examine below how these basic logical gates can be constructed from quantum cells.

The device configurations shown are the results of self-consistent calculations of the ground state using the ICHA described above. The figures show the calculated ground-state charge density on each site of the cellular array. In these figures the dot diameters reflect the relative electron density at each site (dot) in the cell.

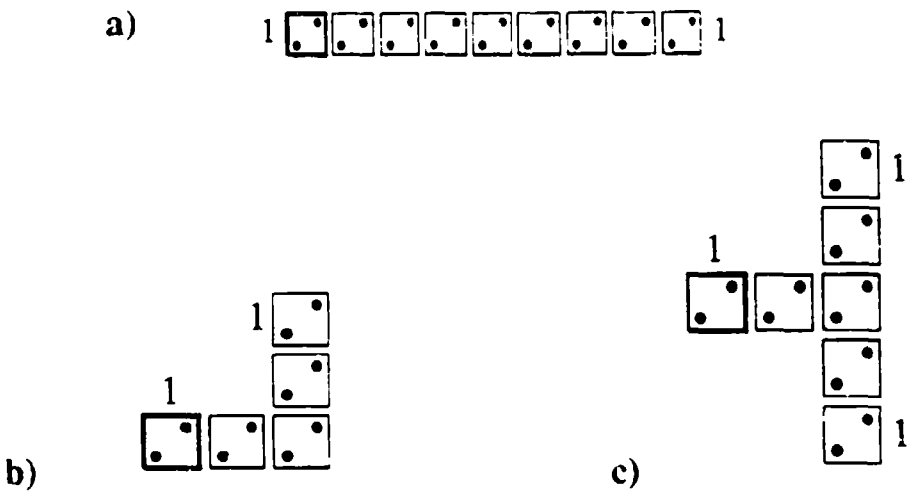


FIGURE 5. QCA wires. a) The basic wire. b) A corner in a wire. c) Fan-out of one signal into two channels. In each case the darker (left-most) cell has a fixed polarization which constitutes the input. Note that these figures are not simply schematic, but are a plot of the results of a self-consistent many-body calculation of the ground state for the cellular array. The diameter of each circle is proportional to the calculated charge density at each site.

Wires

A linear chain of cells oriented as shown in Figure 5a functions as a wire, transmitting a 0 or 1 (P=+1 or P=-1) from one end of the wire to the other. This is demonstrated by fixing the polarization of one end (the left), while letting the other end be unconstrained, and calculating the self-consistent ground state of the chain using the ICHA method. Figure 5a shows the results of that calculation. Not surprisingly, the ground state consists of all cells aligned with the same polarization as the end cell. The first excited state of the chain has a "kink" in it at the chain center, *i.e.*, half the cells polarized one way and half polarized the other. For our example, the energy of the first excited state is about *1* meV ($\Delta E/k_B T = 10\ K$) above the ground state energy. Wire bends and fan-out are also possible, as shown in Figures 5b and 5c respectively. Again, the left-most cell is fixed and the ground-state configuration calculated. This sort of fan-out is appropriate for the edge-driven paradigm discussed above.

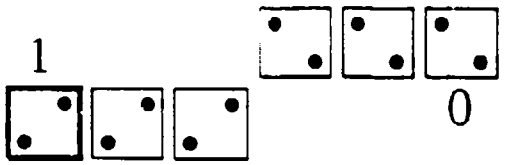


FIGURE 6. An inverter constructed from a quantum cell automaton.

Inverter

By offsetting one chain of cells from another, as shown in Figure 6, an inverter can be constructed. If the polarization of the one end is fixed, the polarization of the other end will be opposite.

AND and OR gates

AND and OR gates can be made from the intersection of two wires. Figure 7 shows an OR gate. The darker boxes are around the input cells. Their polarization is set to correspond to

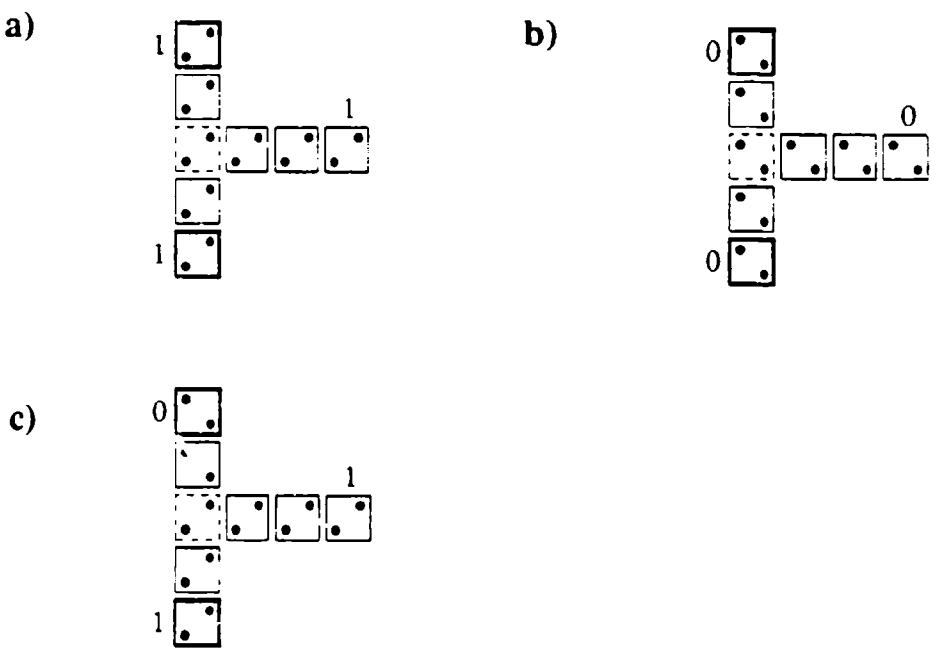


FIGURE 7. An OR gate. The cells in darker squares are fixed to the input states. The cell in the dashed square is biased slightly toward the "1" state.

the logical values shown. For the case when the inputs are *0* and *1* (Figure 7c), the central cell state would normally be indeterminate since a "tie vote" exists between the input cells. To resolve, this we *bias* the central cell by increasing the site energy on sites *2* and *4* slightly. This could be accomplished by making the quantum-dot diameter slightly smaller on these two sites. It is then slightly more energetically favorable for the cell to be in a *1* state, thus breaking the tie. The AND gate is constructed in exactly the same way except that the central cell is biased toward the 0 state. The AND gate is shown in Figure 8. Both these figures reflect the results of self-consistent solutions of the many-electron problem for the entire array shown.

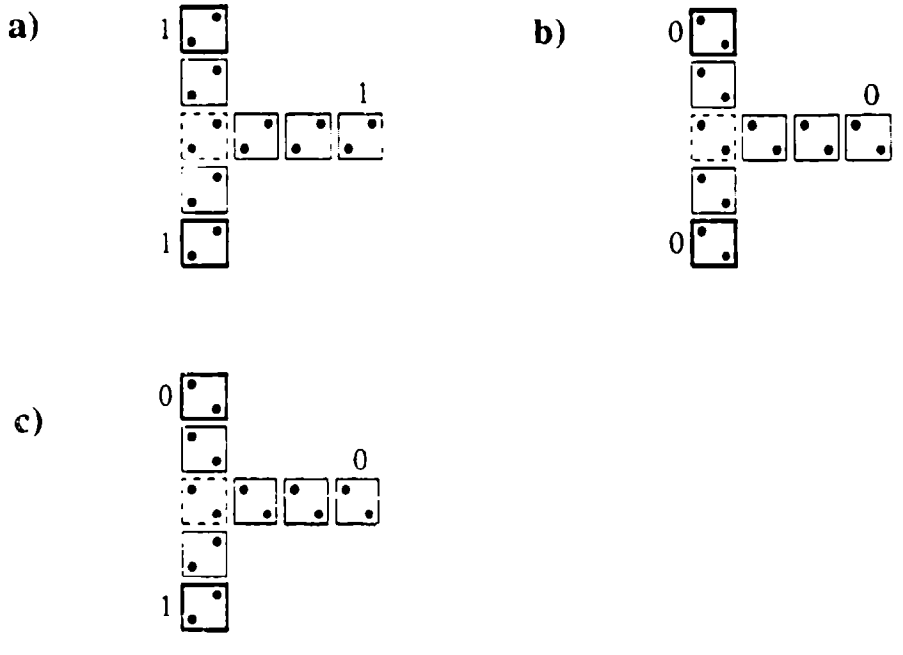


FIGURE 8. An AND gate. The cells in darker squares are fixed to the input states. The cell in the dashed square is biased slightly toward the "0" state.

Memory cell

A single quantum cell can act as a memory storage cell. Once prepared in an eigenstate with $P=+1$, for example, the cell will in principle remain in that configuration indefinitely. One problem is that slight variations in the potential environment may make it slip into the other eigenstates. To avoid this it may be desirable to use small or medium-size arrays of quantum cells to store each bit. This is shown schematically in Figure 9. One advantage of a regular rectangular array of cells is that it may be possible to use the interaction of many cells with the set and sense lines (the exact mechanism for setting and sensing is not critical here). The problem of making non-interfering address lines is certainly non-trivial.

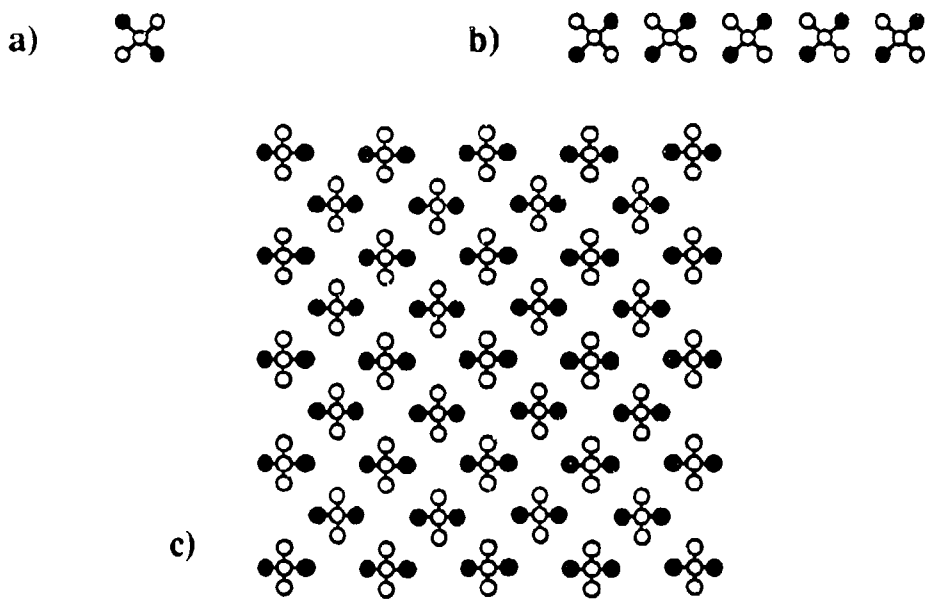


FIGURE 9. Quantum cellular arrays as memory storage cells. A single bit can be stored in a) a single cell, b) a line of cells, or c) an array of cells. Arrays of cells would make the storage more robust.

5. Issues for QCA as a Technology

Fabrication of QCA's in semiconductors appears to be within reach of current technology. The GaAs/AlGaAs system has proven fruitful as a means of fabricating quantum dot structures by imposing electrostatically a pattern on the two-dimensional electron gas formed at the heterojunction interface. Other materials systems, including molecular systems, are also candidates for realizing a QCA structure. Any implementation must deal with several issues key to the successful operation of the cell we have described.

Uniformity of cell occupancy

It is important for the operation of the QCA that each cell contain two electrons. The cell-cell response function degrades significantly if one or three electrons are in a cell. Fortunately, the physics of the cell acts to ensure that the occupancy will be very uniform. This is so because the Coulomb interaction causes significant energy-level splitting between the different cell charge states. The Coulomb energy cost to add the third electron is on the order of 10 meV for cells with a *30 nm* separation. Experiments by Obermeyer *et al.* [13] have shown that uniformity in the number of electrons/dot can be maintained in arrays of 10^8 dots.

Dot-size control

The size of the fabricated quantum dots must be fairly well controlled. Variations in the size of the dots translates into variations in the confinement energies on each dot. The cell bistability occurs because the Coulomb interaction is determinative in selecting a preferred polarization state. If the magnitude of the variation among the dots in the confinement energies is greater than the Coulomb energies involved, the cell will be pinned at a fixed polarization. Note that dot size variations are critical only within a single cell; variations between different cells are easily tolerated.

Temperature

The temperature of operation is a major factor. Our QCA quantum cell is expected to work at liquid helium temperatures for dot dimensions which are within the capability of current semiconductor fabrication technology. As technology advances to smaller and smaller dimensions on the few-nanometer scale, the temperature of operation will be allowed to increase. Perhaps our envisioned quantum cellular automaton will find its first room temperature implementation in molecular electronics.

6. Technological Benefits

If successful, quantum cellular automata would represent a revolutionary, rather than evolutionary, departure from conventional electronics. In this section we review some possible benefits a QCA technology might provide.

Quantum cellular automata solve the interconnection problem. It is widely acknowledged that the main challenge to further improvements in microelectronics is the interconnection and wiring problem. The QCA we discuss accommodate this challenge in a natural fashion. Interconnect lines are no longer necessary to provide the communication between cells; the Coulomb interaction provides the coupling mechanism. *Edge-driven computation* requires neither energy nor information to be transmitted directly to interior cells. *Computing with the ground state* makes both clocking and refresh signals unnecessary.

Quantum cellular automata make possible ultra-high density computing elements. The chief technological advantage of the proposed structures is the improved functional density of computing elements. With a *10* nm design rule, the cell dimensions would be about *50* nm *x 50* nm, which translates into an extremely high packing density of about 10^{10} cells/cm^2. Since, as shown above, a single cell can function as a logical gate, this represents an extremely high functional density.

Quantum cellular automata are extremely low in power dissipation. High packing density is usually accompanied by high power dissipation. However, in QCA structures, the information is stored in physical systems close to their ground state. The energy input to the array is the energy required to set each input bit — about *1* meV per *input* bit. This energy is dissipated in the time it takes for the QCA to relax to its new ground-state configuration, probably less than a few picoseconds (phonon scattering times). This represents a power dissipation of roughly 10^{-10} Watts per input bit, much less than conventional devices.

Quantum cellular automata offer the possibility of ultra-fast computing. As estimated above, the computation occurs in a QCA over the relaxation time for the electrons in the array, probably on the order of picoseconds. It is clear that this relaxation time is a function of the electron-phonon coupling and represents a fundamental speed limit for doing computation with electrons in a semiconductor.

Quantum cellular automata may facilitate fabrication of ultra-dense memory storage. The QCA cell encodes a bit of information. Writing and reading the bit involves very low power dissipation and is very fast. While problems of cell addressing, and cell volatility appear challenging, the possibility of solid-state electronic storage of information at these densities invites further investigation.

Summary

We have presented a specific model for using nanoelectronic devices in a cellular automata architecture and proposed a new paradigm for computing in this framework. Each cell consists of a central quantum dot and four neighboring dots occupied by two electrons. The Coulomb repulsion between the two electrons, quantum confinement effects, and the discreteness of the electronic charge, combine to produce strongly polarized (in the sense defined above) ground states. The response of this polarization to the electrostatic environment is highly nonlinear and exhibits the bistable saturation necessary for a two-state CA. The concept of *edge-driven computation* solves the interconnection problem. The concept of *computing with the ground state* in the QCA approach permits ultra-fast operation, eliminates problems of interconnect delays, resistive and capacitive effects, power dissipation, and limited densities associated with conventional architectures.

ACKNOWLEDGEMENTS

This work was supported in part by the Air Force Office of Scientific Research and the Office of Naval Research. This material is based upon work supported under a National Science Foundation Graduate Fellowship.

REFERENCES

1. D. K. Ferry, L. A. Akers, and E. W. Greeneich, *Ultra Large Scale Interconnected Microelectronics* (Prentice Hall, Englewood Cliffs N.J., 1988).

2. For a recent overview see *Nanostructures and Mesoscopic Systems*, Wiley P. Kirk and Mark A. Reed (eds.), (Academic Press, Boston, 1992).

3. R. T. Bate, *Bull. Am. Phys. Soc.* **22**, 407 (1977).

4. J.N. Randall, M.A. Reed, and G.A. Frazier, *J. Vac. Sci. Technol.* **B7**, 1398, (1989).

5. *Granular Nanoelectronics*, David K. Ferry, John R. Barker, and Carlo Jacoboni (eds.), (Plendum, New York, 1991).

6. R. Landauer, *Phys. Today* **42**, 119 (1989).

7. A proposal for cells with single electron occupancy has been made, but lacks the requisite bistable character. See P. Bakshi, D.A. Broido, and K. Kempa, *J. Appl. Phys.* **70**, 5150 (1991).

8. Craig S. Lent, *Nanostructures and Mesoscopic Systems*, Wiley P. Kirk and Mark A. Reed (eds.), 183 (Academic Press, Boston, 1992).

9. The edge-driven paradigm proposed here is to be distinguished from conventional systolic architectures. In systolic arrays *information* is input only at the edges, but *energy* must be separately fed to each computational element, typically through power lines to

each cell. In the paradigm discussed here, both energy and information are supplied only to the edge cells.

10. K. Obermayer, G. Mahler, and H. Haken, *Phys. Rev. Lett.* **58**, 1792 (1987).

11. A one dimensional treatment is given at length in M. Morrison, T. Estle, and N. Lane, *Quantum States of Atoms, Molecules, and Solids*, chapter 13 (Prentice-Hall, New Jersey, 1976).

12. T. Toffoli and N. Margolus, *Cellular Automata Machines*: A New Environment for Modeling (MIT Press, Cambridge, Massachusetts, 1987).

13. B. Meurer, D. Heitmann, and K. Ploog, *Phys. Rev. Lett.* **68**, 1371 (1992).

Lines of interacting quantum-dot cells: a binary wire

Craig S. Lent and P. Douglas Tougaw
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

submitted to Journal of Applied Physics Feb. 23, 1993

ABSTRACT

We examine the behavior of linear arrays of cells composed of quantum dots. Each cell holds two electrons and interacts Coulombically with neighboring cells. The electrons in the cell tend to align along one of two axes resulting in a cell "polarization" which can be used to encode binary information. The ground-state polarization of a cell is a highly nonlinear function of the polarization of its neighbors. The resulting bistable saturation can be used to transmit binary information along the line of cells, thus forming a binary wire.

PACS: 73.20.Dx, 85.90.+e, 71.90.+q

I. Introduction

Many investigators have noted the connection between quantum devices and locally interconnected architectures [1]. The small currents and charges inherent in quantum devices are poorly suited for driving large numbers of devices, particularly conventional devices. Requiring that a quantum device interact only with its neighbors is much more promising. Despite the appeal of this synthesis, few proposals including both a specification of the component quantum devices and the coupling between them have appeared [2].

Recently, a specific proposal for a quantum cellular automata (QCA) implementation has been made by the authors and coworkers [3,4]. The scheme is based on a quantum cell composed of several quantum dots and containing two electrons. Coulomb repulsion between the electrons causes the charge in the cell to align along one of two directions. These two alignment states, "polarizations", are used to encode binary information. The Coulomb coupling of the charge distribution in one cell to the charge in neighboring cells provides a physics-based local coupling between cells. The coupling leads to a highly bistable saturation behavior in the polarization, avoiding some of the criticisms of usual quantum interference-based device characteristics [5]. Specific arrangements of cells which can function as AND and OR gates have been proposed.

In this paper we examine in detail the linear arrays of such quantum dot cells which form the "wires" in the QCA scheme proposed. In the next section we review the physics of the basic cell and the model proposed in reference [3]. Section III presents the theoretical machinery, a Hartree self-consistency scheme, which we use to examine arrays of cells. Section IV contains the examination of the behavior of a linear array of cells. We show that for a large range of physical parameters, the linear array behaves as a binary wire. Section V contains a discussion of the results.

II. Coupled Quantum Cells

The quantum dot cell is shown schematically in Fig. (1a). It consists of four quantum dots on the corners of a square and one central dot [6]. The cell is occupied by two electrons [7, 8]. Tunneling occurs between near-neighbors and next-near neighbors but the barriers between cells are assumed sufficient to completely suppress electron tunneling between cells. We treat the quantum dots in the site representation, ignoring any degrees of freedom within the dot.

A. Cell polarization

The Coulomb interaction causes the two electrons to tend to occupy antipodal sites. The two-electron ground state may then consist of the electrons aligned along one of two perpendicular axes as shown in Fig. (1b). We define a quantity called the cell polarization which measures the extent to which the charge is aligned along one of these two axes. We denote the single-particle density at site i as ρ_i. The polarization is then defined as

$$P \equiv \frac{(\rho_1 + \rho_3) - (\rho_2 + \rho_4)}{\rho_0 + \rho_1 + \rho_2 + \rho_3 + \rho_4} \quad . \tag{1}$$

If the two electrons are entirely localized in site *1* and *3*, then the polarization $P=1$. If the electrons are on sites 2 and 4, $P=-1$. An isolated cell would have a ground state which is a linear combination of these two polarizations, hence a net polarization of zero[9].

B. The cell Hamiltonian

We construct a simple model of the cell using a tight-binding Hubbard-type Hamiltonian. For an isolated cell, the Hamiltonian can be written,

$$H_0^{cell} = \sum_{i,\sigma} E_0 n_{i,\sigma} + \sum_{i>j,\sigma} t_{i,j}(a_{i,\sigma}^{\dagger} a_{j,\sigma} + a_{j,\sigma}^{\dagger} a_{i,\sigma}) +$$

$$\sum_i E_Q n_{i,\uparrow} n_{i,\downarrow} + \sum_{i>j,\sigma,\sigma'} V_Q \frac{n_{i,\sigma} n_{i,\sigma'}}{|\vec{R}_i - \vec{R}_j|} \quad . \tag{2}$$

Here $a_{i,\sigma}$ is the annihilation operator which destroys a particle at site i (i=0,1,2,3,4) with spin σ. The number operator for site i and spin σ is represented by $n_{i,\sigma}$. The on-site energy for each dot is E_0, the coupling between the i^{th} and j^{th} dot is $t_{i,j}$, and the on-site charging energy (the Coulomb cost for two electrons of opposite spin occupying the same dot) is E_Q. The last term in the Hamiltonian represents the Coulombic potential energy for two electrons located at sites i and j at positions $\vec{R}_i$ and $\vec{R}_j$.

For our "standard cell", on which the most of the numerical results reported here are based, we obtain the values of the parameters in the Hamiltonian from a simple, experimentally reasonable model. We take E_0 to be the ground state energy of a circular quantum dot with diameter $D=10$ *nm* holding an electron with effective mass $m^* = 0.067\ m_0$. The near-neighbor distance between dot centers, a, is taken to be 20 *nm*. The Coulomb coupling strength, V_Q, is calculated for a material with a dielectric constant of *10*. We take $E_Q=V_Q/(D/3)$. The coupling energy between the outer dots and the central dot is $t \equiv t_{0,i} = 0.3$ *meV* (i=1,4), and the next-near neighbor coupling connecting the outer dots, t', is taken to be $t/10$ (consistent with one-dimensional calculations for reasonable barriers). The range of possible values of these parameters is explored systematically below.

The interaction of the cell with the surrounding environment, including other neighboring cells, is contained in a second term in the Hamiltonian which we write as H_{inter}^{cell}. We solve the time independent Schrödinger equation for the state of the cell, $|\Psi_n\rangle$, under the influence of the neighboring cells:

$$(H_0^{cell} + H_{inter}^{cell})\ |\Psi_n\rangle = E_n\ |\Psi_n\rangle. \qquad (3)$$

The spins of the two electrons in a cell can be either aligned or anti-aligned, with corresponding changes in the spatial part of the wavefunction due to the Pauli principle. We will restrict our attention to the case of anti-aligned spins here because that is the ground-state configuration; the spin-aligned case exhibits nearly identical behavior. The

Hamiltonian is diagonalized directly in the basis of few-electron states. We calculate single particle densities, ρ_i, from the two-particle ground-state wavefunction $|\Psi_0\rangle$,

$$\rho_i = \sum_{\sigma} \langle \Psi_0 | n_{i,\sigma} | \Psi_0 \rangle, \tag{4}$$

and from the densities, calculate the resultant polarization P from Eq. (1).

To maintain charge neutrality, a fixed positive charge, $\bar{\rho}$, with magnitude $(2/5)e$ is assumed at each site. If cells had a net total charge then electrons in cells at the periphery of a line of cells would tend to respond mostly to the net charge of the other cells. In a semiconductor realization, the neutralizing positive charge would be provided by ionized donor impurities and charge on the surface of metal gates.

C. Calculating the cell-cell response function.

To be useful in cellular automata-type architectures [11], the polarization of one cell must be strongly coupled to the polarization of neighboring cells. Consider the case of two nearby cells shown in the inset of Fig. (2). Suppose the charge distribution in the right cell, labeled cell 2, is fixed. We assume cell 2 has polarization P_2, and that the charge density on site 0 is negligible (this means the charge density is completely determined by the polarization). For a given polarization of cell 2, we can compute the electrostatic potential at each site in cell 1. This additional potential energy is then included in the total cell Hamiltonian. Thus the perturbing Hamiltonian is

$$H_{inter}^{cell} = H_1^{cell} = \sum_{i \in cell\ 1, \sigma} V_i^1 n_{i,\sigma}, \tag{5}$$

where

$$V_i^m = \sum_{k \neq m, j} V_Q \frac{(\rho_j^k - \bar{\rho})}{|\vec{R}_{k,j} - \vec{R}_{m,i}|} \tag{6}$$

is the potential at site i in cell m due to the charges in all other cells. We denote the position of site j in cell k as $\vec{R}_{kj}$, and the single particle density at site j in cell k as ρ_j^k. The total Hamiltonian for cell 1 is then

$$H^{cell} = H_0^{cell} + H_1^{cell} \tag{7}$$

The two-electron Schrödinger equation is solved using this Hamiltonian for various values of P_2. The ground state polarization of cell 1, P_1, is then computed as described in the previous section.

Figure (2) shows the splitting between the ground state and first excited state of cell 1 as a function of P_2. (Actually, each state is an exchange-split pair of spatially symmetric and antisymmetric states, but the splitting is hardly resolved at the energy scale shown here.) The perturbation rapidly separates states of opposite polarization. The excitation energy for a completely polarized cell to an excited state of opposite polarization is about *0.8* meV for our standard cell. Figure (3) shows P_1 as a function of P_2 — the cell-cell response function. A very small polarization in cell 2 causes cell 1 to be very strongly polarized. As the figure shows, the polarization *saturates* very quickly to either P=+1 or P=-1. This bistable saturation is the basis of the effects described in this paper.

As discussed at greater length in reference [4], the abruptness of the cell-cell response function depends on the ratio of the kinetic energy coupling parameter, t in Eq. (2), to the Coulomb terms in the Hamiltonian. The magnitude of t depends exponentially on both the distance between the dots and the height of the potential barrier between them.

III. Hartree Self-consistent solution for many cells

In the analysis of the previous section, the two-electron eigenstates were calculated for a single cell. It is important to note that for the Hamiltonian employed, these are exact two-particle eigenstates. This was possible because we could explicitly enumerate all possible two-electron states and diagonalize the Hamiltonian in this basis set. We now want to analyze linear arrays of many cells. Exact diagonalization methods then become intractable because the number of possible many-electron states increases rapidly as the number of electrons increases. We must therefore turn to an approximate technique.

The potential at each site of a given cell depends on the charge density at each site of all other cells. We will treat the charge in all other cells as the generator of a Hartree-type potential and solve iteratively for the self-consistent solution in all cells. This approximation, which we call the Intercellular Hartree Approximation (ICHA), can be stated formally as follows. Let Ψ_0^k be the two-electron ground-state wavefunction for cell k. We begin with an initial guess for the densities. Then, for each cell we calculate the potential due to charges in all other cells using Eq. (6). Although the neighboring cells will normally dominate, we do not restrict the analysis to near-neighbors only, but include the effect of all other cells. For a cell k, this results in a perturbation of the isolated cell Hamiltonian of Eq. (2):

$$H_k^{cell} = \sum_{i \in cell\ k,\sigma} V_i^k n_{i,\sigma}. \tag{8}$$

The Schrödinger equation for each cell is now solved for the two-electron ground-state eigenfunction:

$$(H_0^{cell} + H_k^{cell}) |\Psi_0^k\rangle = E_0^k |\Psi_0^k\rangle. \tag{9}$$

From the ground state eigenfunctions we calculate the improved single particle densities.

$$\rho_j^k = \sum_\sigma \langle \Psi_0^k | n_{j,\sigma} | \Psi_0^k \rangle \tag{10}$$

The improved densities are then used in Eq. (6) and the system is iterated until convergence is achieved. Once the system converges, the many-electron energy, E_{total}, is computed from the sum of the cell eigen-energies using the usual Hartree correction term to account for over-counting of the Coulomb interaction energy between cells:

$$E_{total} = \sum_{k} E_k^0 - \sum_{i>q,\, i,j} V_Q \frac{\rho_i^k \rho_j^q}{|\vec{R}_{k,i} - \vec{R}_{q,j}|}. \tag{11}$$

It should be stressed that the ICHA still treats Coulombic, exchange, and correlation effects between electrons in the *same* cell exactly. The Hartree mean field approach is used to treat self-consistently the interaction between electrons in *different* cells [10].

It is relevant to point out that we do not require coherence of the many-electron wavefunction across the whole array of cells. All that is required to support this analysis is that the wavefunction is *coherent across a single cell*. No information about the phase of the wavefunction in other cells is relevant to the wavefunction in a given cell — only the charge densities in other cells need be known.

IV. Lines of Quantum Cells

Figure (4) shows schematically a line of two-electron quantum cells. The distance between centers of adjacent cells is three times the near-neighbor distance between dots in a single cell. If the polarization of the end cell is fixed, say to $P=+1$, a polarization will be induced in the neighboring cells. The question we address in this section is whether the saturation is sufficiently nonlinear that the entire line of cells will be "locked in" to a positive polarization. If this occurs for physically reasonable values of the Hamiltonian parameters, then lines of cells can perhaps be viewed as "wires" which transmit information, coded in the cell polarization, from one place to another.

A. Line saturation

Figure (5) shows the polarization as a function of cell number for a line of 10 cells. The polarization of cell 1 is set to values of P = 0.9, 0.8, 0.6, 0.2 and 0.02, and the ground state of the electrons in the remaining nine cells is calculated self-consistently using the ICHA method described in the preceding section. The Hamiltonian parameters for these cells are those of the standard cell. These parameters yield a very bistable cell response. The result is that even a slight polarization in the driver cell results in essentially complete polarization of all other cells in the line, as is clear in the figure. Figure (5b) is a plot of the calculated particle densities on each site in the line of cells for the case when the driver cell is polarized with only P = 0.02. This figure is not a schematic representation, but a plot of the calculated single-particle densities. The radius of each dot shown is proportional to the particle density at the corresponding site. The squares around the cells are aids to the eye only; the driver cell is indicated with a darker square around it.

As the tunneling energies t and t' are increased, the two-particle ground state wavefunction in each cell becomes less localized in the antipodal sites — the kinetic energy term begins to balance and eventually dominate the Coulomb term in the Hamiltonian. Figure (6) shows the polarization of the line when t=1.0 meV and t'=0. Figure (7) illustrates the case

when t=1.0 meV and $t'=t/10$. (The polarization of the last cell is always slightly lower because it has only one near-neighbor.) Notice that in both cases, the polarization saturates at a constant value, we denote P_{sat}, several cells away from the driver cell. If the driver is polarized at a value larger than P_{sat}, the polarization decreases in successive cells until it reaches P_{sat}. If the driver is polarized at a value smaller than P_{sat}, the polarization increases in successive cells until it reaches P_{sat}. The value of P_{sat} depends on the physical parameters in the cell Hamiltonian, and on the distance between cells.

If the driver cell has a fixed *negative* polarization, then the line will polarize to a saturation value of $-P_{sat}$. The undriven line has two degenerate ground states of opposite polarizations. The perturbation of the driver essentially selects one of these states as the new ground state, although it also modifies it in the region near the driver. Since we can change the signs of all polarizations, *including* the fixed drivers, and obtain another ground state configuration (a mirror image of the original), we need here only consider situations with a positive polarization driver cell.

Figure (8) shows the cell polarization for a line of cells when the line "fails". The kinetic energy parameters for this case are t=1.5 meV and $t'=t/10$. Since the value of t is significantly larger than the Coulomb induced splitting between the energy of oppositely polarized states (about 1 meV), the bistable response of the cells is very small. Thus, even a completely polarized driver cell fails to polarize the line. The polarization drops precipitously to zero.

Two important conclusions follow from these results. First, for a line of cells for which the tunneling energies t and t', are small enough to yield strong bistable behavior, a line of cells acts like a binary wire. That is, it robustly transmits a P=+1 or -1 polarization from one end to another. In fact it has the very attractive feature that it restores degraded signals back to the signal rails ($P \equiv \pm 1$). Secondly, the behavior of the line as a whole is characterized by the saturation polarization, P_{sat}. If P_{sat} is close to unity, the line functions

very well as a binary wire. If the individual cells are not strongly enough bistable, the value of P_{sat} will be significantly less than unity and the line of cells will be less effective as a binary wire. If the bistability is sufficiently weak, P_{sat} is zero. In the next section we examine P_{sat} as a function of the physical parameters which specify the cell Hamiltonian.

B. Dependence of P_{sat} on physical parameters

For a long line of cells, all the cells sufficiently removed from the ends will have polarization P_{sat}. The infinite line contains two degenerate ground states with $P = \pm P_{sat}$. We calculate P_{sat} by considering a segment of an infinite line. Figure (9) shows a "target" cell with three neighbors on each side (more distant neighbors have a negligible influence on the target cell). In the ground state all these cells have the same polarization, P_{sat}. We solve for P_{sat} iteratively using the following scheme: All 6 neighbors are set to a polarization corresponding to an initial guess. The resulting potentials on the sites of the target cell are calculated and the two-electron Schrödinger equation is solved for the ground state of the target cell. The (induced) polarization of the target cell is then calculated from the two-electron wavefunction. The polarization of the 6 neighbors is now set to this calculated value and the process is iterated. The iteration converges to a fixed-point when all cells have the same polarization, P_{sat}. The saturation polarization calculated this way is identical to that obtained by considering a long (*e.g.*, twenty cell) line and finding the polarization of the innermost cells.

We focus on the dependence of the saturation polarization as a function of the physical parameters which enter the cell Hamiltonian (Eq. (2)). Figure (10) shows the variation of P_{sat} with the kinetic energy parameter t with $t'=t/10$. All other parameters are kept fixed. This is equivalent to changing the barrier height between the quantum dots. Higher values of t correspond to lower barrier heights. As the figure illustrates, for values of t above about 1 meV (a barrier height of roughly 100 *meV* for the standard cell), the saturation polarization falls quickly to zero. The transition occurs near $t=1$ meV because that is roughly the energy splitting between the ground state and the excited state of opposite

polarization (see Fig (2)). When the kinetic energy gain in hopping to neighboring sites balances this cost, the tendency of the cell to polarize is lost.

Figure (11) illustrates the variation of P_{sat} with a, the near-neighbor distance between quantum dots. As a is varied, t and t' are kept constant. The intercellular distance is always 3a. The variation of a then has principally the effect of changing the strength of the Coulomb interaction between the cells and between dots in the same cell. The larger a, the weaker the Coulomb interaction and hence the weaker the bistable cell behavior.

Although we have focused on results for a particular "standard" cell with the specific physical parameters stated above, the saturation behavior is clearly determined by the ratio of the physical parameters, not their absolute values. Consider again the cell Hamiltonian in Eq. (2). The value of E_0 will not affect the polarization behavior because it simply adds a constant shift to the total energy. We set $t'=t/10$ for the near-neighbor and next-near-neighbor kinetic energy terms. The value of P_{sat} is then determined by three values: the kinetic energy parameter t, the site-site Coulomb energy parameter V_Q/a, and the on-site Coulomb term E_Q.

Let us examine what these three energy parameters correspond to physically. The energy t is half the value of the splitting between the spatially symmetric and antisymmetric states of a system of two quantum dots. It can also be considered as a hopping energy between neighboring dots which lowers the total energy by allowing the wavefunction to spread out spatially. The energy V_Q/a is the Coulomb energy of two electrons separated by the distance a (the near-neighbor inter-dot separation). The energy E_Q is the Coulomb energy of two electrons of opposite spin occupying the same quantum dot. It is roughly inversely proportional to the dot diameter [12].

Consider the three-dimensional parameter space spanned by these three physical energies t, V_Q/a, and E_Q. Systems with the same ratio of $t : V_Q/a : E_Q$ have identical saturation

behavior. The locus of equivalent systems is a ray passing through the origin in parameter space. Therefore, to explore saturation polarization for the entire parameter space spanned by these three physical parameters, it is sufficient to calculate P_{sat} on the surface of a sphere in the parameter space. This is shown in Fig. (12). The t axis has been scaled by a factor of 10. The values of P_{sat} are plotted through the gray-scale map shown. The map is non-uniform and is chosen to accentuate the very abrupt transition between values of P_{sat} near unity and values very close to zero.

Figure (12) shows that the saturation behavior is not limited to an "island" in the Hamiltonian's parameter space but is "continental". Further, for "most" of the parameter space, P_{sat} is very close to *1* or *0*. The transition is quite abrupt. A detailed examination of the interplay between on-site charging effects and near-neighbor effects awaits further study.

V. Discussion

The results presented suggest that the lines of quantum cells discussed here are indeed capable of forming binary wires in the following sense. Information is encoded in the polarization of individual cells. Say a bit value of 1 is represented by a polarization P=+1 and a bit value of 0 is represented by a polarization of P=-1. Suppose the polarization of an end cell is fixed to 1 (perhaps electrostatically) and the line of cells is allowed to relax to its ground state. The ground state will be one for which all the cells have polarization 1 (bit value 1). If the end cell is switched, and the line again allowed to relax to its ground state, all the cells will switch to P=-1 (bit value 0). *This mechanism transports information, but not charge, from one end of the wire to another.* It has the additional feature that inputs with polarization less than one, but still positive, will be "reset" to be 1. Similarly, degraded negative input polarizations will be reset to -1. The strong nonlinear bistable response of the coupled-cell system performs a role similar to gain in conventional digital devices, constantly restoring signal levels.

Note that in this scheme we rely on the ground-state configuration of the system — not the transient response. We assume inelastic processes are sufficient to relax the system to its new ground state after the input is changed. The wire "transmits" information in the sense that after this relaxation has occurred, the new ground state is one in which the output end of the wire matches the input end.

In reference [4] we discuss the implementation of logical gates using the interacting quantum cells analyzed here. AND gates, OR gates and inverters have all been designed using these ideas.

To function well, the cells need to be small enough that the Coulomb interaction between electrons in different dots is significant. Additionally, the effective barriers to tunneling between dots must be large enough that the kinetic energy advantage of spreading out the wavefunction does not overcome the Coulombic advantage of keeping the electrons in

antipodal sites. As the results of Section IVB made clear, however, the relevant range of physical parameters is not a small, carefully balance set.

Fabrication of such coupled dot structures surely represents a significant challenge, but the dimensions involved make it possible to conceive of semiconductor realizations using nanolithographic techniques presently being developed. Setting and reading the individual cell states at input and output ends involves the challenging task of sensing the presence of a single electron.

The theoretical analysis in this paper is a zero-temperature treatment. At a non-zero temperature, entropy will become important. The excited states of a line have a much greater degeneracy (hence entropy) than the ground state. For long enough lines, this means that the thermodynamic expectation value of the polarization will decay as the distance from the end driver cell increases. These effects will ultimately limit the size of usable binary wires and the operating temperatures feasible. Nevertheless, if the size scale can be sufficiently reduced (our standard cell is relatively large), more practical operating temperatures could be obtained.

In conclusion we have examined the behavior of lines of interacting quantum-dot cells. The bistable saturation in the cell-cell interaction results in "binary wire" behavior in which information, encoded in the cell polarization, can be robustly transmitted from one end of a line to another.

Acknowledgments

We gratefully acknowledge stimulating conversations with Wolfgang Porod and Gary Bernstein of the Notre Dame Nanoelectronics Group. We would also like to thank Karl Hess and Anthony Leggett for helpful comments. This work was supported in part by the Air Force Office of Scientific Research. This material is based in part upon work supported under a National Science Foundation Graduate Fellowship.

References

[1] R. T. Bate, *Bull. Am. Phys. Soc.* 22, 407 (1977); J.N. Randall, M.A. Reed, and G.A. Frazier, *J. Vac. Sci. Technol.* B7, 1398, (1989); D. K. Ferry, L. A. Akers, and E. W. Greeneich, *Ultra Large Scale Interconnected Microelectronics* (Prentice Hall, Englewood Cliffs N.J., 1988).

[2] A general proposal involving coupled resonant-tunneling diodes has been discussed by the Texas Instruments group. See J.N. Randall, A.C. Seabaugh, Y. -C. Kao, J.H. Luscombe, and B.L. Newell, *J. Vac. Sci. Technol.* **B9**, 2893 (1991).

[3] C. S. Lent, P. Douglas Tougaw, and Wolfgang Porod, *Appl. Phys. Lett.*, February 15, 1993).

[4] C. S. Lent, P. Douglas Tougaw, Wolfgang Porod and Gary H. Bernstein, to appear in *Nanotechnology.*

[5] R. Landauer, *Phys. Today* 42, 119 (1989).

[6] The similar behavior of various other model cells is discussed elsewhere. See C. S. Lent, P. Douglas Tougaw, and Wolfgang Porod, (unpublished).

[7] The ability to control quantum dot occupancies over as many as 10^8 dots using a back-gating technique has recently been reported by B. Meurer, D. Heitmann, and K. Ploog, *Phys. Rev. Lett.* 68, 1371 (1992).

[8] A proposal for cells with single electron occupancy has been made, but lacks the requisite bistable character. See P. Bakshi, D.A. Broido, and K. Kempa, *J. Appl. Phys.* 70, 5150 (1991).

[9] The polarization so defined is not to be confused with a dipole moment. For the situations we consider here, the ground state of the cell has no dipole moment,

though it may have a quadrupole moment. The cell polarization simply measures the degree to which the electronic charge is aligned and the direction of that alignment.

[10] Since electrons in different cells are physically distinguishable (there being no wavefunction overlap), the exchange coupling between them is zero. The Hartree and Hartree-Fock approximations are therefore equivalent in this case.

[11] T. Toffoli and N. Margolus, *Cellular Automata Machines*: A New Environment for Modeling (MIT Press, Cambridge, Massachusetts, 1987)

[12] For the standard cell, we take E_Q to be the Coulomb energy of two electrons separated by one-third the dot diameter D, a physically reasonable first approximation. Varying E_Q is roughly equivalent to changing D.

FIGURE CAPTIONS

FIGURE 1. Schematic of quantum cell. The geometry of the cell is shown in a). The solid lines indicate tunneling between the quantum dots. The tunneling energy between the inner dot and the outer dots is t, and the tunneling energy between adjacent outer dots is t'. The Coulomb repulsion between the two electrons which occupy the cell results in ground-state configurations with the electrons aligned in the two orientations shown in b). The polarization defined by Eq. (1) takes the values 1 and -1 for these two configurations.

FIGURE 2. The eigenstate energies for cell 1 as a function of the polarization of adjacent cell 2. The polarization of the eigenstates is indicated by the inset diagrams. The low energy state is always the one with the same polarization as the "driver" cell 2. Slight exchange splitting (between the spatially symmetric and antisymmetric states) is evident for very small values of P_2.

FIGURE 3. The cell-cell response function (after reference [3]). The induced cell polarization P_1 is plotted as a function of the neighboring cell polarization P_2. The solid line shows the polarization of the spin-antisymmetric state and the dotted line shows the polarization of the (nearly degenerate) spin-symmetric state.

FIGURE 4. A linear array of interacting cells. Each cell holds two electrons. Hopping between cells is assumed to be completely suppressed.

FIGURE 5. The response of a line of cells. The polarization of cell 1 (shown in (b) with a darker outline) is fixed and the ground-state polarization induced in the line of cells is calculated. The plot shows the induced polarization for a driver polarization of P = 0.9, 0.8, 0.6, 0.4, 0.2 and 0.02. For the case of the weakest driver polarization, P = 0.02, the charge densities on each site are shown in b). The diameter of each dot is proportional to the charge density on that site. The Hamiltonian parameters used here

are those of the "standard cell" discussed in the text. The result shows that even a slight polarization in a driver cell induces nearly complete polarization in the line of cells.

FIGURE 6. The response of a line of cells for a different value of tunneling energy parameter. As in Figure 5, the polarization of cell 1 is fixed and the ground-state polarization induced in the line of cells is calculated. The plot shows the induced polarization for a driver polarization of P = 0.9, 0.8, 0.6, 0.4, 0.2 and 0.02. The model cells here differ from the standard cells used for Figure 5 in that the tunneling energy t is 1.0 meV and t' is neglected.The result shows that even a slight polarization in a driver cell induces a polarization in the line of cells but that the polarization saturates at a value P_{sat} (here about 0.85).

FIGURE 7. The response of a line of cells for a different value of tunneling energy parameter. The polarization of cell 1 is fixed and the ground-state polarization induced in the line of cells is calculated. The plot shows the induced polarization for a driver polarization of P = 0.9, 0.8, 0.6, 0.4, 0.2 and 0.02. For the case of the weakest driver polarization, P = 0.02, the charge densities on each site are shown in b). The diameter of each dot is proportional to the charge density on that site. The model cells here differ from the standard cells used for Figure 5 in that the tunneling energy t is 1.0 meV and $t'=t/10$. The result shows that even a slight polarization in a driver cell induces a polarization in the line of cells but that the polarization saturates at a value P_{sat} (here about 0.7) which characterizes the response of the line.

FIGURE 8. Failure of a driver cell to polarize the line. In this case the tunneling parameters are chosen so that t = 1.5 meV and t' = t/10. The result is that the kinetic energy term in the Hamiltonian (Eq. (2)) overwhelms the Coulombic terms. A driver cell which is completely polarized then induces only a small polarization in its neighbors, and the polarization decays quickly down the line. The charge densities are shown in b)

FIGURE 9. Schematic view of cell geometry used in self-consistent calculation of the polarization of an infinitely long chain of cells.

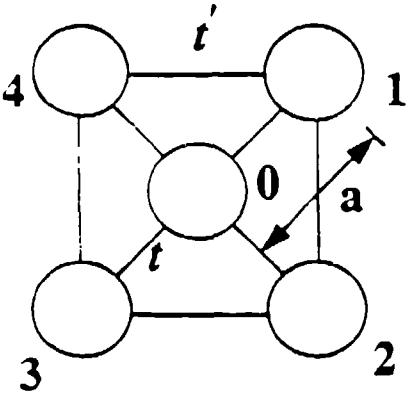
FIGURE 10. The saturation polarization for an infinite linear chain of cells as a function of the tunneling parameter t. Other cell Hamiltonian parameters are fixed at the "standard cell" values.

FIGURE 11. The saturation polarization for an infinite linear chain of cells as a function of the near-neighbor site distance a (see Figure 1). Other cell Hamiltonian parameters are fixed at the "standard cell" values.

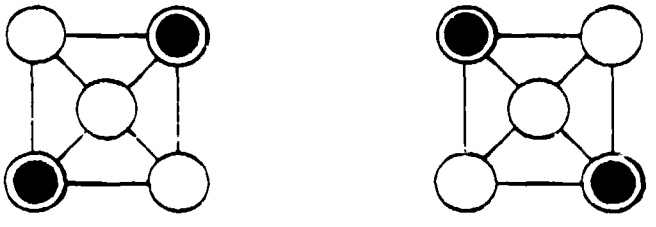
FIGURE 12. The values of P_{sat} for the parameter space spanned by the parameters in the Hamiltonian.

FIGURES

a)



b)



P = +1 P = -1

FIGURE 1. Schematic of quantum cell. The geometry of the cell is shown in a). The solid lines indicate tunneling between the quantum dots. The tunneling energy between the inner dot and the outer dots is t, and the tunneling energy between adjacent outer dots is t'. The Coulomb repulsion between the two electrons which occupy the cell results in ground-state configurations with the electrons aligned in the two orientations shown in b). The polarization defined by Eq. (1) takes the values 1 and -1 for these two configurations.

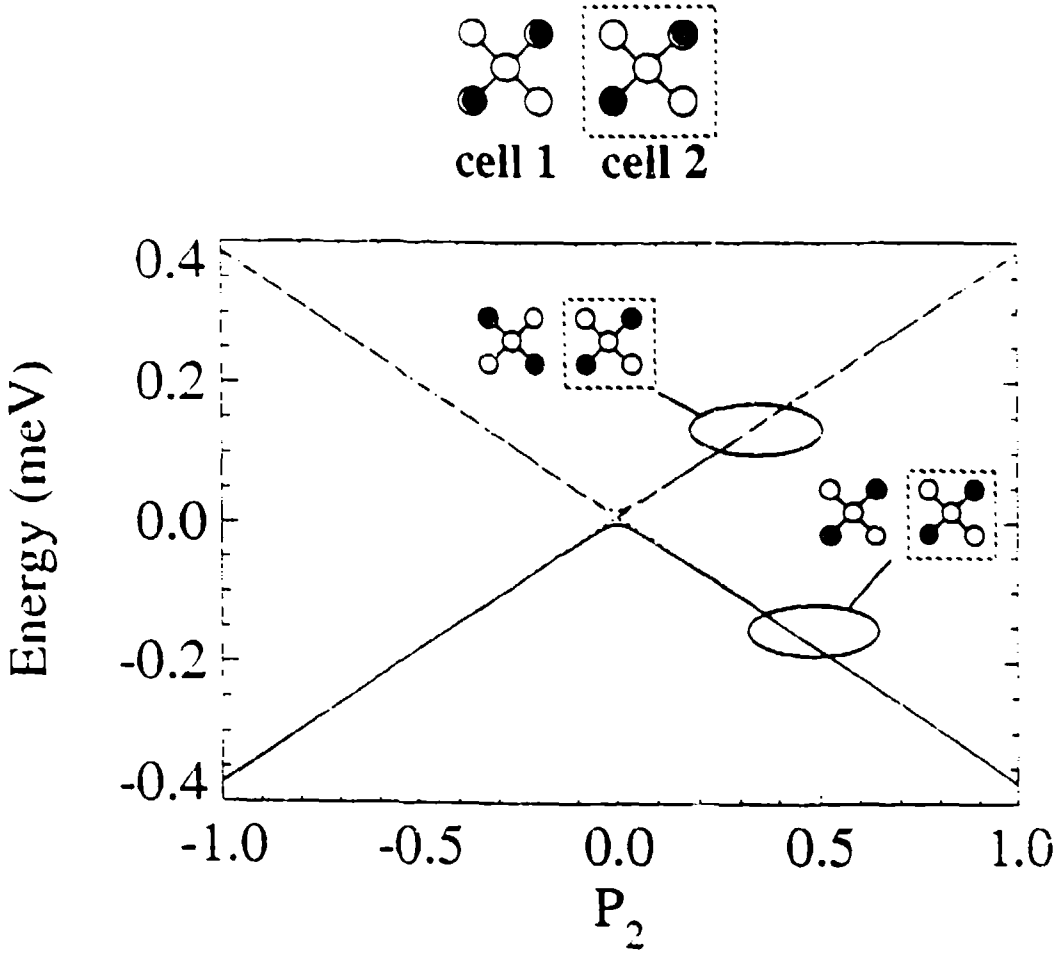


FIGURE 2. The eigenstate energies for cell 1 as a function of the polarization of adjacent cell 2. The polarization of the eigenstates is indicated by the inset diagrams. The low energy state is always the one with the same polarization as the "driver" cell 2. Slight exchange splitting (between the spatially symmetric and antisymmetric states) is evident for very small values of P_2.

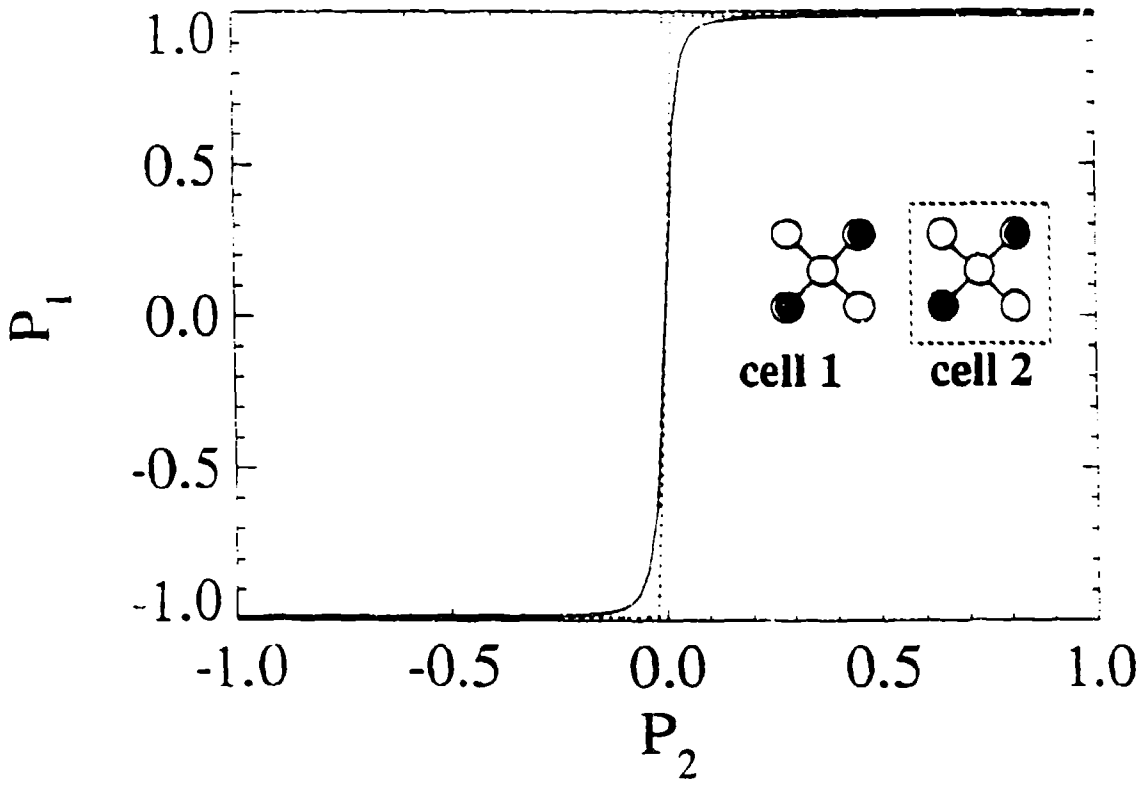


FIGURE 3. The cell-cell response function (after reference [3]). The induced cell polarization P_1 is plotted as a function of the neighboring cell polarization P_2. The solid line shows the polarization of the spin-antisymmetric state and the dotted line shows the polarization of the (nearly degenerate) spin-symmetric state.

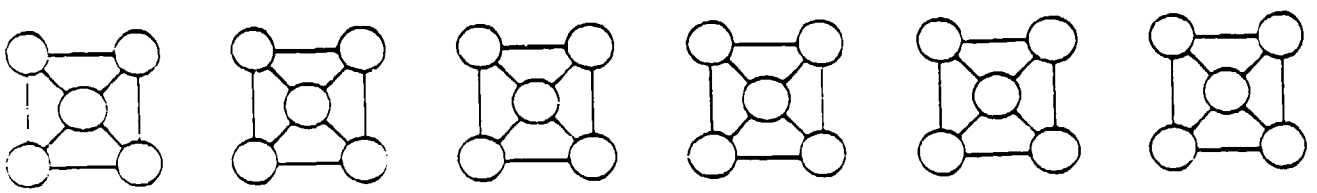


FIGURE 4. A linear array of interacting cells. Each cell holds two electrons. Hopping between cells is assumed to be completely suppressed.

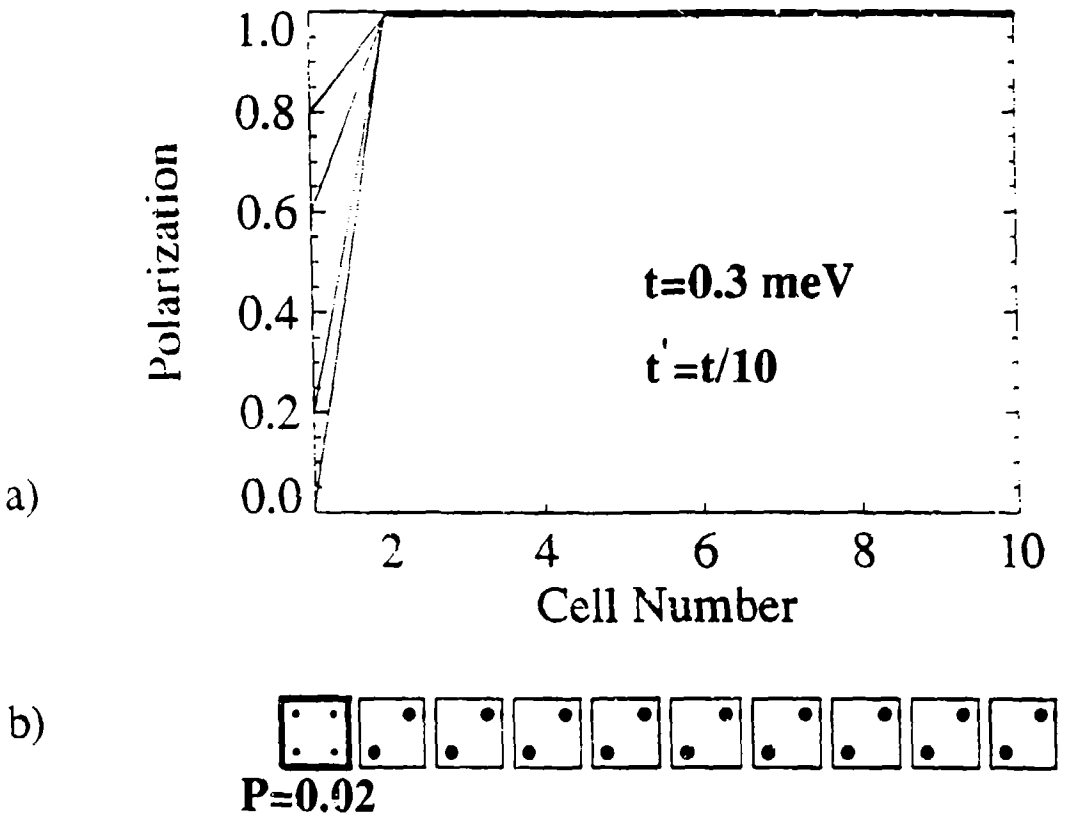


FIGURE 5. The response of a line of cells. The polarization of cell 1 (shown in (b) with a darker outline) is fixed and the ground-state polarization induced in the line of cells is calculated. The plot shows the induced polarization for a driver polarization of P = 0.9, 0.8, 0.6, 0.4, 0.2 and 0.02. For the case of the weakest driver polarization, P = 0.02, the charge densities on each site are shown in b). The diameter of each dot is proportional to the charge density on that site. The Hamiltonian parameters used here are those of the "standard cell" discussed in the text. The result shows that even a slight polarization in a driver cell induces nearly complete polarization in the line of cells.

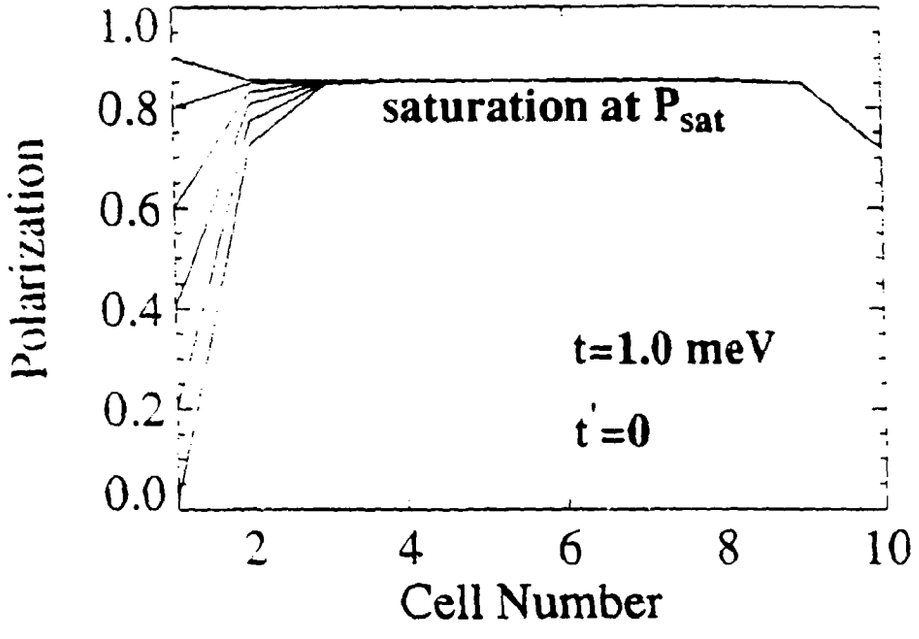


FIGURE 6. The response of a line of cells for a different value of tunneling energy parameter. As in Figure 5, the polarization of cell 1 is fixed and the ground-state polarization induced in the line of cells is calculated. The plot shows the induced polarization for a driver polarization of P = 0.9, 0.8, 0.6, 0.4, 0.2 and 0.02. The model cells here differ from the standard cells used for Figure 5 in that the tunneling energy t is 1.0 meV and t' is neglected. The result shows that even a slight polarization in a driver cell induces a polarization in the line of cells but that the polarization saturates at a value P_{sat} (here about 0.85).

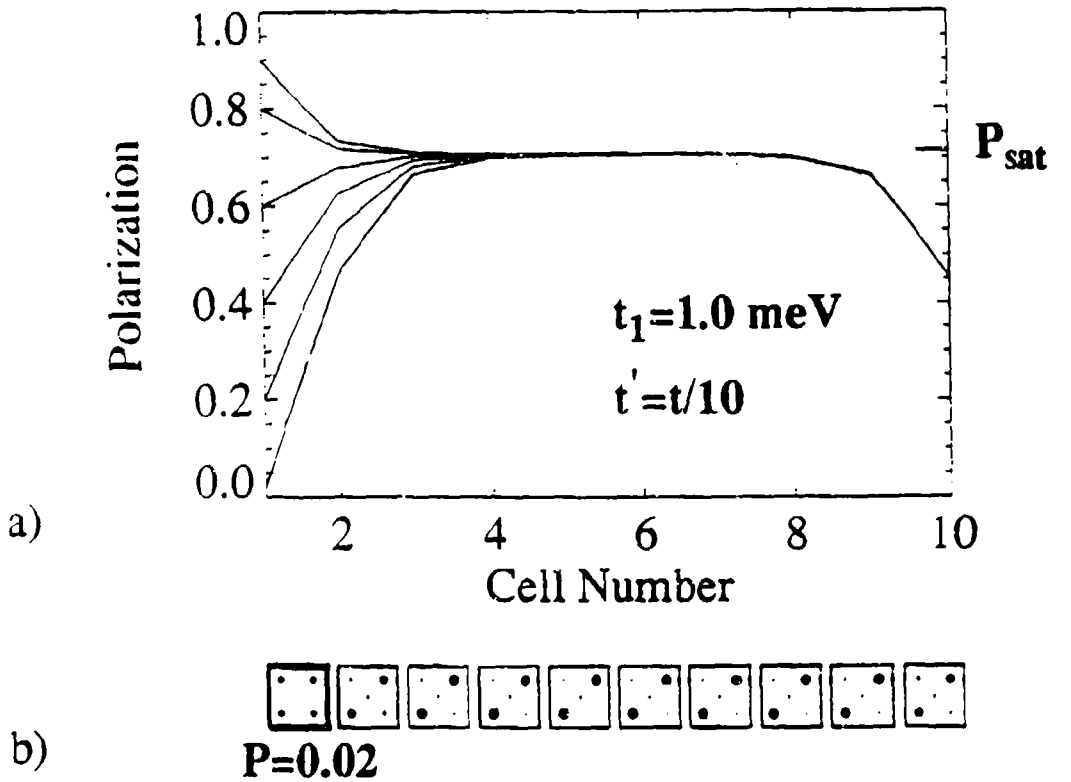


FIGURE 7. The response of a line of cells for a different value of tunneling energy parameter. The polarization of cell 1 is fixed and the ground-state polarization induced in the line of cells is calculated. The plot shows the induced polarization for a driver polarization of P = 0.9, 0.8, 0.6, 0.4, 0.2 and 0.02. For the case of the weakest driver polarization, P = 0.02, the charge densities on each site are shown in b). The diameter of each dot is proportional to the charge density on that site. The model cells here differ from the standard cells used for Figure 5 in that the tunneling energy *t* is 1.0 meV and *t'=t/10*. The result shows that even a slight polarization in a driver cell induces a polarization in the line of cells but that the polarization saturates at a value P_{sat} (here about 0.7) which characterizes the response of the line.

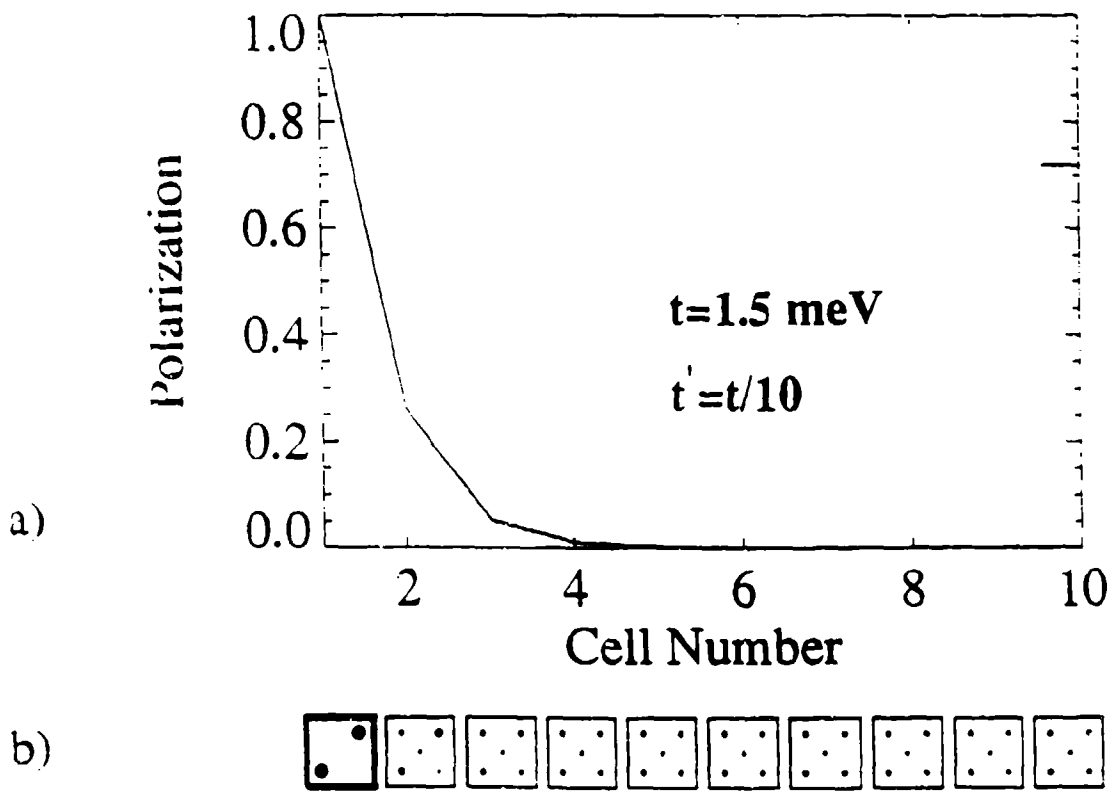
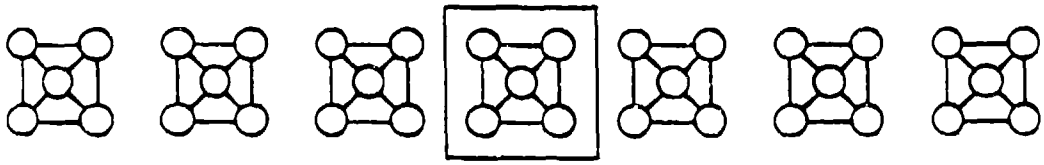


FIGURE 8. Failure of a driver cell to polarize the line. In this case the tunneling parameters are chosen so that $t = 1.5$ meV and $t' = t/10$. The result is that the kinetic energy term in the Hamiltonian (Eq. (2)) overwhelms the Coulombic terms. A driver cell which is completely polarized then induces only a small polarization in its neighbors, and the polarization decays quickly down the line. The charge densities are shown in b).



target cell

FIGURE 9. Schematic view of cell geometry used in self-consistent calculation of the polarization of an infinitely long chain of cells.

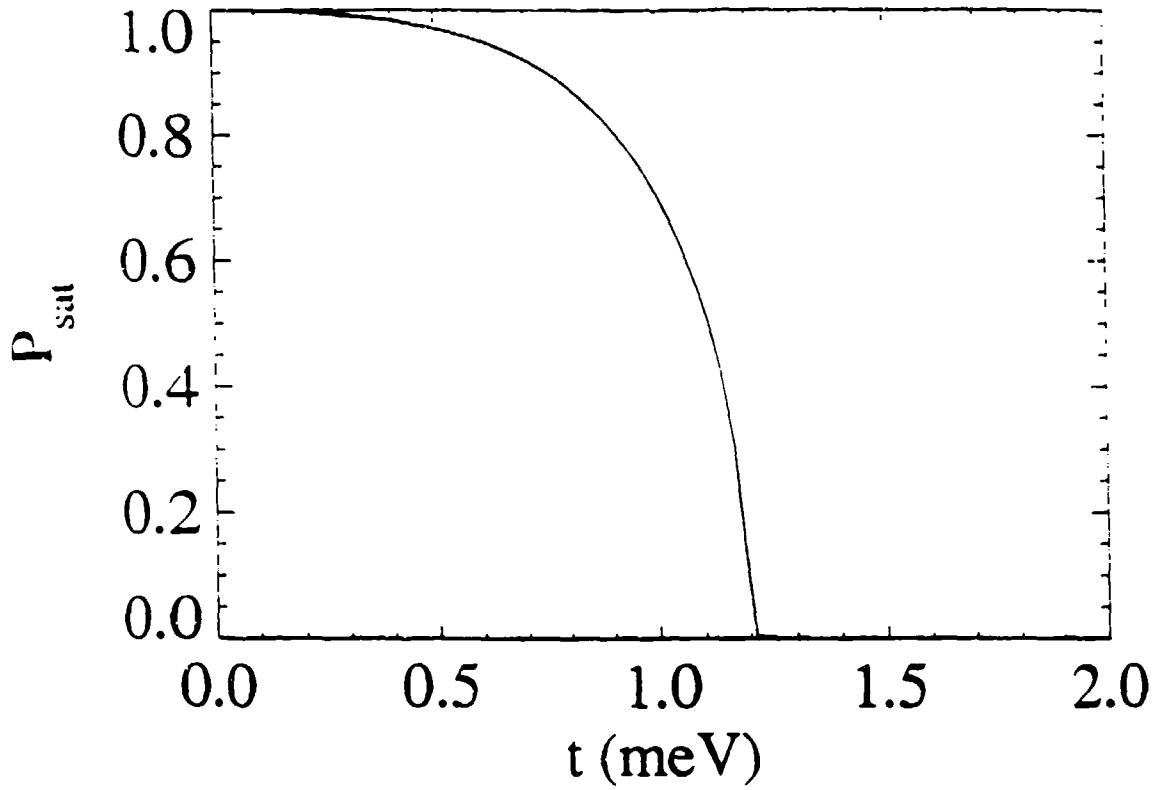


FIGURE 10. The saturation polarization for an infinite linear chain of cells as a function of the tunneling parameter *t*. Other cell Hamiltonian parameters are fixed at the "standard cell" values.

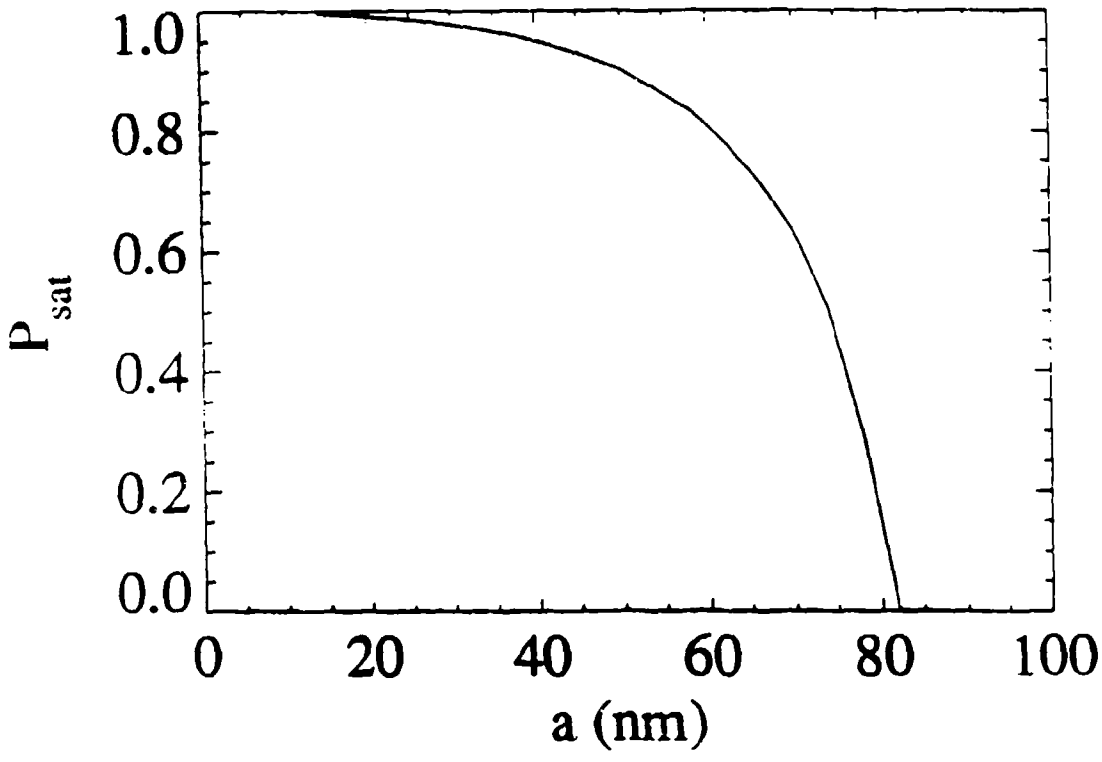


FIGURE 11. The saturation polarization for an infinite linear chain of cells as a function of the near-neighbor site distance a (see Figure 1). Other cell Hamiltonian parameters are fixed at the "standard cell" values.

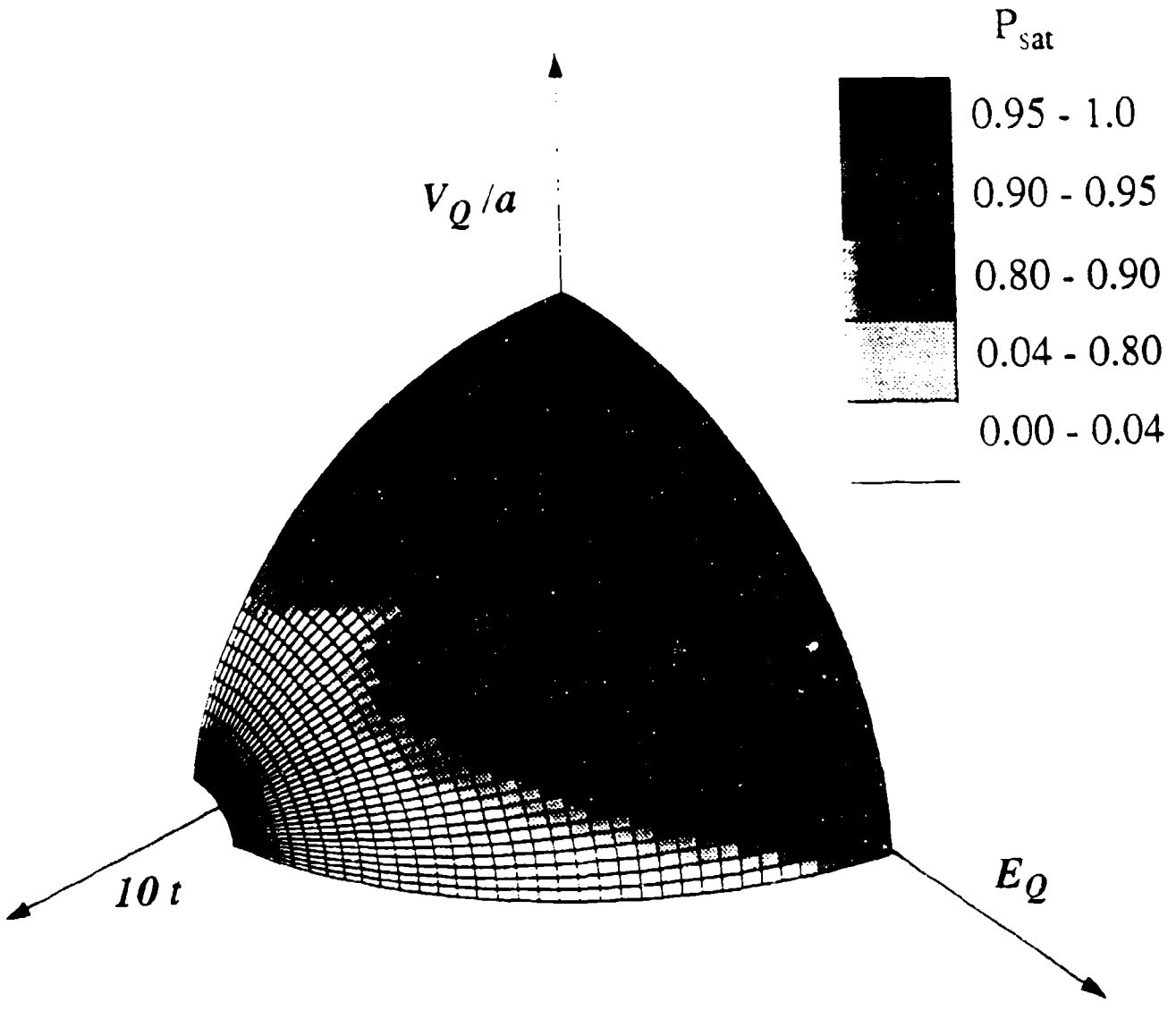


FIGURE 12. The values of P_{sat} for the parameter space spanned by the parameters in the Hamiltonian.

Bistable Saturation in Coupled Quantum-dot Cells

P. Douglas Tougaw, Craig S. Lent, and Wolfgang Porod
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

submitted to Journal of Applied Physics April 29, 1993

ABSTRACT

Model quantum dot cells are investigated as potential building blocks for quantum cellular automata architectures. Each cell holds a few electrons and interacts Coulombically with nearby cells. In acceptable cell designs, the charge density tends to align along one of two cell axes. Thus, we can define a cell "polarization" which can be used to encode binary information. The polarization of a cell is affected in a very nonlinear manner by the polarization of its neighbors. We quantify this interaction by calculating a cell-cell response function. Effects of non-zero temperature on the response of a model cell are investigated. We also examine the effects of multiple neighbors on a cell and discuss programmable logic gate structures based on these ideas.

PACS: 73.20.Dx, 85.90.+h, 71.90.+q

I. Introduction

For many years, the size of microelectronic devices has been shrinking, and this has led to faster, denser circuits. Despite these improvements, the basic computing paradigm has remained virtually unchanged because device operation has been largely unchanged apart from re-scaling. There is now much interest in extremely dense device arrays forming locally interconnected architectures like cellular automata (CA) [1] and cellular neural networks [2]. Such architectures could lead to changes in device structure of a less evolutionary and more revolutionary nature.

At the same time, many researchers have been investigating ways to use quantum structures as electronic devices. In the course of such research, a great deal has been learned about the behavior of electrons in very small structures. Because of the size of the structures involved, an outstanding difficulty is providing a scheme in which one of these quantum devices, which typically carry nanoamperes of current, could be used to drive several other similar devices. In addition, the capacitance of the wires needed to interconnect such structures would tend to dominate their behavior. Therefore, locally connected architectures like CA's may be an attractive paradigm for implementing quantum device architectures [3].

CA architectures composed of nanometer-scaled quantum devices which are coupled through the Coulomb interaction (no current flows between devices) have been proposed by the authors elsewhere [4-6]. We call such architectures quantum cellular automata (QCA). The QCA contains an array of quantum-dot cells which are connected locally by the interactions of the electrons contained within them. The quantum state of each multi-dot cell encodes the "logical" state of that cell. For this reason, each cell should ideally have exactly two stable states, since this will allow direct encoding of binary information [7]. Such two-state cells need also to exhibit bistable saturation to ensure that noise or small geometric variations do not overwhelm the signal.

To function as a CA, the state of each cell should be dependent on the states of its neighbors. In this paper, we compare the cell-cell coupling and bistable saturation of several different quantum cell designs which might form the basis of quantum cellular automata. All of these designs have certain characteristics in common: a few (typically four or five) quantum dots connected by coupling coefficients and populated by a total of one to three electrons. In these cells the required interaction between neighbors is caused by the mutual Coulombic repulsion of the electrons contained in the cells. We use a very simple model of each cell, neglecting details relating to exactly how the quantum dot structures are realized, but focussing on the charge distribution among the dots and the Coulomb coupling between cells. We define a cell-cell response function which characterizes the interaction between neighboring cells.

In the next section we will introduce the theoretical model of the "standard cell", on which much of the work of references 4, 5, and 6 is based. It is the most thoroughly investigated cell design because it displays strong bistable saturation. We discuss the model Hamiltonian used for the cell, the method used to calculate the cell-cell response function, and the effects of non-zero temperature. Section III compares various other cell designs. Among these are different geometric arrangements of the quantum dots, one and three electron cells, and continuous quantum dashes. In Section IV, we extend our results to include the effects of multiple neighbors on a cell. We show that such effects in a system with three nearest neighbors can be thought of as majority voting logic. We show how this behavior can be used to implement programmable logic gates, and then show other possible implementations for dedicated AND and OR gates. A discussion and conclusion follow in Section V.

II. A Model Quantum Cell

The model "standard cell" design, shown schematically in Fig. (1a), consists of five quantum dots located at the corners and the center of a square. Tunneling occurs between the central site and all four of the outer sites (near-neighbor tunneling) and to a lesser degree between neighboring outer sites (next-near-neighbor tunneling). It is assumed that the potential barriers between cells are high enough to completely suppress intercellular tunneling. The cell is occupied by a total of two electrons hopping among the five sites; these electrons tend to occupy antipodal outer sites within the cell due to their mutual electrostatic repulsion (see Figure 1(b)).

We will show that these two stable states are degenerate in an isolated cell, but an electrostatic perturbation in the cell's environment (such as that caused by neighboring cells) splits the degeneracy and causes one of these configurations to become the cell ground state. Altering the perturbation causes the cell to switch between the states in an abrupt and nonlinear manner. This very desirable bistable saturation behavior is due to a combination of quantum confinement, Coulombic repulsion, and the discreteness of electronic charge.

A. Cell polarization

Since Coulomb repulsion causes the electrons to occupy antipodal sites, the ground state charge density may have the electrons aligned along one of the two diagonal axes shown in Fig (1b). We therefore define the cell polarization, a quantity which measures the extent to which the charge density is aligned along one of these axes. The polarization is defined as

$$P \equiv \frac{(\rho_1 + \rho_3) - (\rho_2 + \rho_4)}{\rho_0 + \rho_1 + \rho_2 + \rho_3 + \rho_4} \quad . \tag{1}$$

where ρ_i denotes the electron probability density at site i. As in Fig. (1b), electrons completely localized on sites 1 and 3 will result in P=1, while electrons on sites 2 and 4

yield P=-1. An isolated cell would have a ground state which is a linear combination of these two states and would therefore have a net polarization of zero [8].

B. Model cell Hamiltonian

We employ a simple model of the quantum cell which uses a tight-binding Hubbard-type Hamiltonian. We represent the quantum dots as sites, ignoring any degrees of freedom internal to the dot. The Hamiltonian for a single isolated cell can be written as:

$$H_0^{cell} = \sum_{i,\sigma} E_0 n_{i,\sigma} + \sum_{i>j,\sigma} t_{i,j} (a_{i,\sigma}^{\dagger} a_{j,\sigma} + a_{j,\sigma}^{\dagger} a_{i,\sigma}) +$$

$$\sum_{i} E_Q n_{i,\uparrow} n_{i,\downarrow} + \sum_{i>j,\sigma,\sigma'} V_Q \frac{n_{i,\sigma} n_{j,\sigma'}}{|\vec{R}_i - \vec{R}_j|} \quad . \qquad (2)$$

Here $a_{i,\sigma}$ is the annihilation operator which destroys a particle at site i (i=0,1,2,3,4) with spin σ. The number operator for site i and spin σ is represented by $n_{i,\sigma}$. E_0 is the on-site energy for each dot, $t_{i,j}$ is the energy associated with tunneling between dots i and j, and E_Q is the on-site charging energy (the purely Coulombic cost for two electrons of opposite spin to occupy the same dot). The last term in (2) represents the Coulombic potential energy due to electrons on the i^{th} and j^{th} sites at positions $\vec{R}_i$ and $\vec{R}_j$. V_Q is an electrostatic parameter fixed by fundamental constants and the dielectric constant of the material used to form the dots.

For the cell described above we use values of the parameters in the Hamiltonian based on a simple, experimentally reasonable model. We take E_0 to be the ground state energy of a circular quantum dot of diameter 10 nm holding an electron with effective mass $m^* = 0.067\ m_0$. The near-neighbor distance between dot centers, a, is taken to be 20 nm. The Coulomb coupling strength, V_Q, is calculated for a material with a dielectric constant of 10, and E_Q is taken to be $V_Q/(D/3)$ [9]. The coupling energy between an outer dot and the central dot is $t \equiv t_{0,i} = 0.3$ meV (i=1,2,3,4), and the next-near neighbor coupling

connecting the outer dots, t', is taken to be $t/10$ (consistent with one-dimensional calculations for a barrier height of 150 meV) [10]

The spins of the two electrons in the cell can be either parallel or antiparallel. We consider here the case of electrons with antiparallel spins, since that is the ground state of the cell. Calculations with electrons having parallel spins yield qualitatively very similar results.

To maintain charge neutrality in the cell, a fixed positive charge $\bar{\rho}$, corresponding to a charge of $(2/5)e$, is also assumed on each site. In a single isolated cell this just renormalizes E_0, but the fixed charge is important when simulating systems with more than one cell. If each cell had a net negative charge, then electrons near the edges of a group of cells would respond mostly to the net negative charge of the other cells. In a semiconductor realization, the fixed positive charge would likely be provided by ionized donor impurities and charge on the surface of metal gates.

The interaction of a cell with its electrostatic environment (including neighboring cells) is contained in a second Hamiltonian term which we write as H_{inter}^{cell}. We solve the time-independent Schrödinger equation for the n^{th} eigenstate of the cell, $|\Psi_n\rangle$, under the influence of the neighboring cells:

$$(H_0^{cell} + H_{inter}^{cell})\,|\Psi_n\rangle = E_n\,|\Psi_n\rangle. \tag{3}$$

The total Hamiltonian is diagonalized directly in the basis of 2-particle site kets. We calculate the single-particle density, ρ_i, from the two-particle ground-state wavefunction $|\Psi_0\rangle$ by finding the expectation value of the number operator for site i:

$$\rho_i = \sum_{\sigma} \langle\Psi_0|\, n_{i,\sigma}\,|\Psi_0\rangle \tag{4}$$

We can then use these densities to calculate the cell polarization P as in Eq. (1).

C. Calculating the cell-cell response function.

To be useful in cellular automata-type architectures, the state of a cell must be strongly influenced by the states of neighboring cells. To demonstrate how one of these cells is influenced by the state of its neighbor, consider the two cells shown in the inset to Fig. (2). The cell centers are separated by a distance of $3a$ = 60 nm. We assume cell 2 has a given polarization P_2 and that the electron density on the central site is negligible. This means that the charge density is completely determined by the cell polarization. For the corresponding electron density on each site of cell 2, we calculate the electrostatic potential at each site of cell 1. This additional potential energy is then included in the Hamiltonian for cell 1. Thus the perturbing Hamiltonian component is:

$$H_{inter}^{cell} = H_1^{cell} = \sum_{i \in cell\ 1,\sigma} V_i^1 n_{i,\sigma}, \tag{5}$$

where

$$V_i^m = \sum_{k \neq m,\, j} V_Q \frac{(\rho_j^k - \tilde{\rho})}{|\vec{R}_{k,j} - \vec{R}_{m,i}|} \tag{6}$$

is the potential at site i in cell m due to the charges in all other cells. We denote the position of site j in cell k as $\vec{R}_{k,j}$, and the electron density at site j in cell k as ρ_j^k. The total Hamiltonian for cell 1 is then

$$H^{cell} = H_0^{cell} + H_1^{cell}. \tag{7}$$

The two-electron time-independent Schrödinger equation is solved using this Hamiltonian for a series of values of P_2 in the range [-1, +1]. The ground state polarization of cell 1, P_1, is then computed for each value of P_2 as described in the previous section. Thus, we can plot the *induced* polarization of cell 1 as a function of the polarization of cell 2. This function $P_1(P_2)$, which we call the cell-cell response function, is one measure of how well a cell will operate in a quantum cellular automaton architecture.

Fig. (2) shows the cell-cell response function for the standard cell. The highly nonlinear nature of the response indicates that a small polarization in cell 2 causes a very strong polarization in its neighbor, cell 1. The figure also shows that the polarization of cell 1 *saturates* very quickly to a value of +1 or -1. This bistable saturation is the basis of the quantum cellular automata since it means that we can encode bit information using the cell polarization. We assign the bit value of 1 to the P=+1 state and the bit value -1 to the P=-1 state. Since the cell is almost always in a highly polarized state ($|P| \cong 1$), the state of the cell will be indeterminate only if the electrostatic environment due to other cells is perfectly symmetric.

Fig. (3) shows the lowest four eigen-energies of cell 1 as a function of the polarization of cell 2. This shows that the perturbation due to the polarization of cell 2 quickly separates the states of opposite polarization. For a completely polarized standard cell, the excitation energy from the ground state to the first excited state with opposite polarization is about 0.8 meV. This corresponds to a temperature of about 9K.

The abruptness of the cell-cell response function depends on the ratio of the tunneling energy, t in Eq. (2), to the Coulomb energy for electrons on neighboring sites. The magnitude of the tunneling energy depends exponentially on both the distance between dots and the height of the potential barrier between them. Fig. (4) shows the cell-cell response function for different values of the tunneling energy t. The switching becomes more abrupt as t decreases. Of course, if t goes to zero, the tunneling would be completely suppressed and no switching would occur. Extremely small values of t would similarly slow the switching time. For t=0.3 meV, the standard cell value, we estimate the tunneling time as $\frac{\hbar}{t} = 2$ psec.

D. Non-Zero Temperature Cell-Cell Response

We extend the calculation of the cell-cell response function to nonzero temperatures by calculating the thermal expectation value of the electron density at each site of the cell.

$$\rho_i = \; < \; < \hat{n}_i > \; > \; = \frac{\sum_{\sigma}\sum_{n} \langle \Psi_n | \hat{n}_{i,\sigma} | \Psi_n \rangle \cdot e^{-E_n/(k_B T)}}{\sum_{n} e^{-E_n/(k_B T)}} \qquad (8)$$

Evaluating the thermal average requires knowledge of the excited states of the cell as well as the ground states. Using the results of Eq. (8), the polarization of the cell can be calculated as before using Eq. (1). The results of such a calculation for the standard cell (with the next-near-neighbor coupling t'=0) are shown in Fig. (5). The curve for T=0 is the same as in Fig. (2). The nonlinearity of the response degrades as the temperature increases. For temperatures up to 4.2 K, the response is good, but for higher temperatures it would probably be unacceptable for use in a QCA.

Note that this maximum operating temperature will increase as the size of the cell decreases. Thus, although a 20 nm design rule requires cryogenic temperatures for satisfactory operation, the design scales to smaller sizes easily and a much smaller, possibly macro-molecular implementation [11] would work at room temperature.

III. Alternative Quantum Cells

While the cell described above has demonstrated an excellent cell-cell response, there is no reason to discount other possible cell designs. Slight modifications to this cell give rise to a family of similar cells whose behavior can provide insight into the nature of the system.

In looking for other model cells, there are several approaches we can take. The most obvious of these is to alter the number of sites and their geometric arrangement. Alternatively, the cell occupation can be altered. Finally, tunneling between the intracellular sites can be increased, decreased, or effectively eliminated by varying the potential barriers between the sites.

A. Four quantum cells

In this section we will investigate the cell-cell response of four different quantum cells. While these four are representative of the sort of cells one might consider, they in no way exhaust the study of new cell designs.

The first cell, included mainly as a standard by which to judge the others, is the original cell described above with t=0.3 meV. This will be called cell A. Next will be the same cell with no tunneling between the outer neighbors ($t' = 0$). We will refer to this as cell B. Cell C omits the presence of the central site and allow tunneling only between the four outer sites. Finally, cell D inhibits tunneling even further, allowing it only between sites 1 and 2 and between sites 3 and 4. Schematic diagrams of these four cell designs are shown in Fig. (6a).

In a semiconductor realization of these cells, the minimum spacing between nearest neighbor sites will be limited by the fabrication technology. For this reason, these cells are designed with a constant 20 nm design rule (the nearest neighbors in each cell are 20 nm apart). While this makes cells C and D smaller, it is the most physically reasonable way to

compare their operation. The spacing between interacting cells is set at three times the near-neighbor dot spacing.

It is possible to consider cell B as an approximation to cell A which neglects tunneling between outer neighbors. In reality there will always be a certain amount of tunneling between outer sites, but this tunneling can be made arbitrarily small by selectively increasing the potential barriers between the outer sites. The same increase in potential barriers would be needed to suppress horizontal tunneling in cell D.

Fig. (6b) shows the cell-cell response functions for these four cells. This figure shows that cells A and B have very similar responses, and both are superior to cells C and D. Thus, elimination of the central site as in cells C and D degrades the response. This leads us back to the five-site cell we originally considered. Since the complete suppression of next-near neighbor coupling as in cell B might introduce additional fabrication difficulty with little improvement in the cell response, cell A may be the most practical of these four cell designs.

B. One and three electron cells

As an alternative to changing the geometry of the cell, we can also alter the electron occupancy. Fig. (7) shows the cell-cell response function for cell A occupied by a single electron, and Fig. (8) shows the response for the same cell with three electrons (two parallel spins, one antiparallel). These nearly linear response functions never become strongly polarized, even for fully polarized neighbors. This indicates that such cells would perform very poorly as the basis of a quantum cellular automaton [12].

C. Quantum dashes and double-wells

Proposals have been made for one-electron "quantum dash" cells which appear qualitatively similar to the cells we have discussed here [13]. In this section we investigate the cell-cell response function of single-electron quantum dashes and compare this to a

very similar double quantum well to show how important the discreteness of electronic charge is to the non-linearity of the response functions seen in the previous sections.

Since these cells are of a more spatially continuous nature than cells previously considered, the site representation is no longer useful. Each cell will be modeled as a one-dimensional hard-walled square well of width 30 nm. The two cells are separated by a distance of 20 nm. These dimensions are similar to those of the cells described above. We use the finite element method to solve the single-electron time-independent Schrödinger equation for each one-electron cell. The geometry used to calculate the cell-cell response function is shown schematically in Fig. (9a).

Since these cells have only a single axis along which to distribute the electronic charge, a new definition of polarization must be introduced. The new definition, which takes into account the continuous nature of the probability density, is:

$$P = \frac{\int_{-L/2}^{0} \rho(x)\,dx - \int_{0}^{L/2} \rho(x)\,dx}{\int_{0}^{L} \rho(x)\,dx} \tag{9}$$

Because of its continuous nature, the charge density in cell 2, the "driver" cell is no longer uniquely determined by specifying the cell polarization. We therefore fix the charge density to be constant in each half of cell 2.

The cell-cell response function calculated for such a system is shown in Fig. (9b). As this figure shows, the response is quite linear and cell 1 is virtually unpolarized even for a fully polarized neighbor. The electron probability density as a function of position for cell 1 with a fully polarized neighbor (P_2=1) is shown in Fig. (9c). The probability density is nearly symmetric about the center of the cell as we would expect for such a small polarization.

A related cell, the double well, is shown schematically in Fig. (10a). It is a quantum well of the same dimensions as in Fig. (9a), but the potential in the middle third of the well has been raised by 150 meV. This cell is also very similar to half of cell D from the last section, so we would expect its response to be much better than that of the simple quantum dash.

The calculated response, shown in Fig. (10b), is indeed much better than that of Fig. (9b). Its nonlinearity and saturation properties are very similar to those of cells C and D in Fig. (6b). This response shows that one-electron cells can be used to provide the required nonlinear response, but it is also possible to view each pair of these cells as a single 2-electron cell, which becomes geometrically very similar to cell D of Fig. (6a).

The fact that such a seemingly small change in the nature of the cell should cause such a profound change in the cell-cell response function is linked to the fact that electron charge is discrete in regions surrounded by high potential barriers. That is to say, the expectation value of the number operator approaches an integer value as the region becomes more and more isolated by the potential barriers surrounding it [14]. Therefore, almost the entire wavefunction will become localized in one half of the cell if a small asymmetry in the electrostatic environment is introduced. This fact is demonstrated in Fig. (10c), which shows nearly all the charge density in the right half of the cell. Since there is no barrier in the middle of the quantum dash to isolate the top and bottom of the cell, no such localization behavior is seen there and the charge density is always nearly symmetric about the center of the well.

IV. Multiple Neighbor Interactions

Thus far, we have only considered the interaction between a cell and a single neighboring cell. The natural extension of this is to investigate the effects of multiple neighbors on the state of a cell. Since this implies considering a system which contains several cells and therefore several electrons, we cannot use the direct solution method described earlier for treating a single cell. For the analysis of such systems, we treat the physics within each cell as before, including exchange and correlation effects exactly. The intercellular interaction is treated self-consistently using a Hartree approximation. This method, called the Intercellular Hartree Approximation (ICHA) is detailed in references 5 and 6.

Fig. 11 shows an arrangement of standard cells such that one cell has multiple neighbors. The charge densities of the cells on the top, left, and bottom are fixed, while those of the middle and right cells are free to react to the fixed charge. In an actual QCA, the states of the neighbors would not be fixed; they would be driven by the results of previous calculations or come from inputs at the edge of the QCA.

In the specific state shown in Fig. 11, two of the fixed neighbors are in the "one" state, and the other is in the "zero" state. When the ICHA is used to determine the ground state of this system, we find that the states of the center and right cells match the state of the majority of the fixed neighbors. We refer to this feature of the cell behavior, which is true for all combinations of the three inputs, as *majority voting logic*. Note that Figures 11, 12, and 13 are not schematic, but plots of the self-consistent electron density on each site. The radius of each dot is proportional to the single-electron density at that site.

While majority voting logic behavior is valuable by itself, its potential functionality is shown by a particular interpretation of the three inputs. In Fig. 12, we have singled out one of the three and called it the program signal. Note that any one of the three neighbors could serve as the program signal, but the one case we are showing (with the program line coming in from the left) is sufficient for illustration purposes. The four systems shown

include all possible combinations of signals on the two non-program lines. Since all four systems in Fig. 12 show the program line in the "one" state, the central cell can only be zero if the other two inputs are both zero. The system thus realizes the truth table of the OR operation. Likewise, if the program signal is zero, the result is zero unless both of the other inputs are one. This is a realization of the AND operation.

By interpreting any one of the inputs as a program line, we have implemented a programmable AND/OR gate. The nature of this gate (AND vs. OR) is defined by the state of the program line, and the other two inputs are applied to the gate thus defined.

The fact that the right-most cell always matches the central cell means that the result of this calculation can be propagated away from the gate, down a QCA "wire" [15], and eventually serve as the input to subsequent gates. It is necessary to distinguish between driving neighbors and driven neighbors in this system. Since the right-most cell is free to react to the states of its neighbors, it is a driven neighbor. Its state will always match that of the central cell, so only the three driving neighbors are involved in the majority voting. Of course, once the signal is propagated away from this gate, the outgoing cells are being driven and can be used as driving neighbors for subsequent gates.

A dedicated, non-programmable, implementation of the AND gate is shown in Fig. 13. This system has only two driving inputs; there is no program signal. The role previously played by the program signal, biasing the central cell so it can only be in the one state if both of its neighbors are, is performed by slightly enlarging the two quantum dots on sites 1 and 3 in the central cell [16]. This means that the ground state of the isolated cell is no longer a non-polarized state; the cell is biased toward the zero state and can only be persuaded to enter the one state if both of its driving neighbors are one. Again, the signal propagates away to the right and can be used to drive subsequent gates. A dedicated OR gate can similarly be implemented by enlarging sites 2 and 4, biasing the cell toward the one state. It will only be in the zero state if both of its driving neighbors are zeros also.

V. Discussion

With the above results, we have demonstrated several quantum-dot cells suitable for implementing a quantum cellular automata architecture. Examination of the cell-cell response function shows that for appropriate cell designs, the state of a cell is influenced very strongly by the state of its neighbors. The highly nonlinear response of the cell suggests that a signal that has become degraded by noise will be restored to full polarization by subsequent cells in the array [15]. In this way, the bistable saturation of the quantum cell is analogous to the gain in a conventional digital device.

We have assumed throughout that the many electron system is in its ground state. In general, a system will start in the ground state and then be driven into an excited state by externally changing the states of input cells near the edge of a QCA array. Inelastic processes, which are usually very detrimental to the operation of quantum devices, then drive the system back to a new ground state corresponding to the new inputs. The details of the temporal evolution of the many-electron system as it relaxes to its ground state are very complicated. In the QCA scheme, we rely on the properties of the system ground state and not the details of the relaxation process for doing the computation. This idea of "computing with the ground state" and the related concept of "edge-driven" systems are discussed more thoroughly in reference 5.

The behavior of lines of these cells, the most basic (and important) components of a quantum cellular automaton, is discussed in reference 6. The results show an excellent example of the restoration of full signal strength after degradation by noise. In addition, it shows that the particular set of parameters we chose in section IIB is not critical; there is a wide range of parameter values for which the cells transmit information from one cell to another.

Clearly, fabrication of these devices presents a major challenge in the realization of QCA devices, but semiconductor realizations of such systems using new nanolithographic

techniques should be possible. It is also possible that future realizations of these cells will be on a macromolecular basis. Another challenge, sensing the presence or absence of a single electron without disturbing the system, necessary for reading the output state of a QCA device, has been successfully addressed [17].

In conclusion, we have explored the interaction of neighboring quantum-dot cells. We have defined the cell-cell response function which characterizes the non-linearity of the coupling between cells and thus determines suitability of a particular cell design for quantum cellular automata implementations. Several cell designs that exhibit the required nonlinear response and bistable saturation have been examined. Temperature effects degrade cell performance but analysis in this simple model suggests that operation at 4.2K should be within the reach of semiconductor implementations. When a cell has several neighbors, its state is determined by the state of the majority of the neighboring cells. This majority voting logic make possible the construction of programmable AND/OR logic gates as well as dedicated AND and OR gates.

Acknowledgments

We gratefully acknowledge stimulating conversations with Gary H. Bernstein of the Notre Dame Nanoelectronic Group. This work was supported in part by the Air Force Office of Scientific Research and the Office of Naval Research. This material is based in part upon work supported under a National Science Foundation Graduate Fellowship.

References

[1] T. Toffoli and N. Margolus, *Cellular Automata Machines*: A New Environment for Modeling (MIT Press, Cambridge, Massachusetts, 1987)

[2] L.D. Chua and L. Yang, IEEE Trans. Circuits Syst. **35**, 1257 (1988), *ibid*, **35**, 1273 (1988).

[3] R. T. Bate, *Bull. Am. Phys. Soc.* **22**, 407 (1977); J.N. Randall, M.A. Reed, and G.A. Frazier, *J. Vac. Sci. Technol.* B7, 1398, (1989); D. K. Ferry, L. A. Akers, and E. W. Greeneich, *Ultra Large Scale Interconnected Microelectronics* (Prentice Hall, Englewood Cliffs N.J., 1988); J.N. Randall, A.C. Seabaugh, Y. -C. Kao, J.H. Luscombe, and B.L. Newell, *J. Vac. Sci. Technol.* **B9**, 2893 (1991).

[4] C. S. Lent, P. Douglas Tougaw, and Wolfgang Porod, *Appl. Phys. Lett.*, **62**, 714 (1993).

[5] C. S. Lent, P. Douglas Tougaw, Wolfgang Porod and Gary H. Bernstein, to appear in *Nanotechnology* **4**, (1993).

[6] C. S. Lent, P. Douglas Tougaw, submitted to Journal of Applied Physics, (unpub.).

[7] Others have attempted to apply our QCA architecture ideas (reference 4-6) to a system where the information is encoded in the spin of the electron. See S. Bandyopadhyay, B. Das, and A. Miller (unpublished).

[8] The polarization so defined is not to be confused with a dipole moment. For the situations we consider here, the ground state of the cell has no dipole moment, though it may have a quadrupole moment. The cell polarization simply measures the degree to which the electronic charge is aligned and the direction of that alignment.

[9] For the standard cell, we take E_Q to be the Coulomb energy of two electrons separated by one-third the dot diameter D, a physically reasonable first approximation.

[10] Reference [6] examines the behavior of lines of cells as all these Hamiltonian parameters are varied. The bistable behavior we examine here is not critically dependent on a particular choice of these parameters, but holds for a wide range of parameter choices.

[11] For a review of the current state, see Geoffrey J. Ashwell, ed., *Molecular Electronics* (Wiley, New York, 1992).

[12] Control of cell occupancy is clearly important in implementation. The ability to control quantum dot occupancies over as many as 10^8 dots using a back-gating technique has recently been reported by B. Meurer, D. Heitmann, and K. Ploog, *Phys. Rev. Lett.* **68**, 1371 (1992).

[13] P. Bakshi, D.A. Broido, and K. Kempa, *J. Appl. Phys.* **70**, 5150 (1991).

[14] See the discussion and two-electron calculation by C. S. Lent, in *Nanostructures and Mesoscopic Systems*, edited by Wiley P. Kirk and Mark A. Reed, 183 (Academic Press, Boston, 1992).

[15] See the discussion of the behavior of lines of cells as a binary wire in reference [6].

[16] Enlarging the quantum dots lowers the on-site energy E_0 and makes it energetically more favorable to occupy the larger dots.

[17] Noninvasive probing of single-electron charging in a semiconductor quantum dot has recently been reported by M. Field, C.G. Smith, M. Pepper, J.E.F. Frost, G.A.C. Jones, and D.G. Hasko, Phys. Rev. Lett. 70, 1311 (1993).

FIGURE CAPTIONS

FIGURE 1. Schematic of the basic five-site cell. (a) The geometry of the cell. The tunneling energy between the middle site and an outer site is designated by t, while t' is the tunneling energy between two outer sites. (b) Coulombic repulsion causes the electrons to occupy antipodal sites within the cell. These two bistable states result in cell polarizations of P=+1 and P=-1 (see Eq. (1)).

FIGURE 2. Cell-cell response function for the basic five-site cells shown in the inset. This shows the polarization P_1 induced in cell 1 by the fixed polarization of its neighbor, P_2. The solid line corresponds to antiparallel spins, and the dotted line to parallel spins. The two are nearly degenerate, especially for significantly large values of P_2.

FIGURE 3. The lowest four eigenstate energies of cell 1 induced by the polarization of cell 2. The insets show that the lowest two energy states always correspond to the same polarization direction as in the driver. Slight exchange splitting effects between spatially symmetric and spatially antisymmetric states breaks the fourfold degeneracy for very small values of P_2.

FIGURE 4. The cell-cell response function of the basic five-site cell for various values of the parameter t in Eq. (2) (after reference [4]). The induced polarization in cell 1 is plotted as a function of the polarization of its neighbor, cell 2. The curves correspond to t=0.2 meV (solid line), t=0.3 meV (dotted line), t=0.5 meV (dashed line), and t=0.7 meV (dot-dashed line). Note the horizontal axis only shows P_2 in the range [-0.1,+0.1].

FIGURE 5. The cell-cell response function for the standard cell with t'=0 at various temperatures. The response degrades as temperature increases. Above 4.2K, the response would be unacceptable for use in a QCA. This maximum operating

temperature is highly dependent on the physical size of the cell; molecule-sized cells would behave in a satisfactory manner up to room temperature.

FIGURE 6. Four geometric variations on the simple model quantum cell. (a) Schematic diagrams of the four cells. Cells C and D occupy less area, but all four cells are drawn with the same minimum spacing between neighbors. Cells B and D will require potential variation between the sites to selectively inhibit tunneling. (b) The cell-cell response function for each of these cell designs. Cell B has the best response, but the improvement over A is small.

FIGURE 7. The cell-cell response function for the basic five-site cell occupied by a single electron. The weak response indicates that such a cell is unsuitable as the basis of a QCA.

FIGURE 8. Cell-cell response function for the basic five-site cell occupied by three electrons. Such a cell is also unacceptable as the basis of a QCA.

FIGURE 9. The "quantum dash" as a QCA cell. (a) A schematic diagram of the cellular arrangement. The length and spacing is similar to that of the basic five-site cell in Fig. (1a). Each cell is modelled as a one-dimensional infinite square well. The cell-cell response is shown in (b). Note that the vertical axis only shows P_2 over the range [-0.1,+0.1]. (c) The one-dimensional charge density in cell 2 for a fully polarized neighbor (P_1=1). The nearly symmetric charge density yields a very low polarization.

FIGURE 10. The double well as a QCA cell. (a) A schematic diagram of the cellular arrangement. The total dimensions are identical to those of Fig. (8). The middle third of each cell contains a 150 meV barrier to isolate the top and bottom of the cell. (b) The cell-cell response function for such an arrangement. Note that the vertical axis now shows P_2 over the range [-1.0, +1.0]. (c) The one-dimensional charge density in cell 2

for a fully polarized neighbor ($P_1=1$). The highly asymmetric charge density results in a cell that is almost completely polarized.

FIGURE 11. Majority voting logic. The states of the center and right cells are always the same as the majority of the three fixed neighbors. The cells with heavy borders have fixed charge densities. These are not schematic diagrams; they are the actual results of the ICHA solution of the ground state charge densities in this system. The diameter of each dot is proportional to the charge density on that site.

FIGURE 12. The programmable AND/OR gate. The program line is set to one in each system, so the gate is displaying OR logic. All four combinations of the non-program line inputs are shown. The cells with heavy borders have fixed charge densities. Any one of the three inputs could be the program line; the left cell is not special. These are not schematic diagrams; they are the actual results of the ICHA solution of the ground state charge densities in each system. The diameter of each dot is proportional to the charge density on that site.

FIGURE 13. The non-programmable AND gate. All four combinations of the inputs are shown. The cells with heavy borders have fixed charge densities, while those with dotted borders are geometrically biased toward zero as shown in the inset. The bias is sufficient to decrease the on-site energy of the affected sites by 1%. Note that the output only equals one if both of the inputs are also one. These are not schematic diagrams; they are the actual ICHA results of the ground state charge densities in each system. The diameter of each dot is proportional to the charge density on that site.

FIGURES

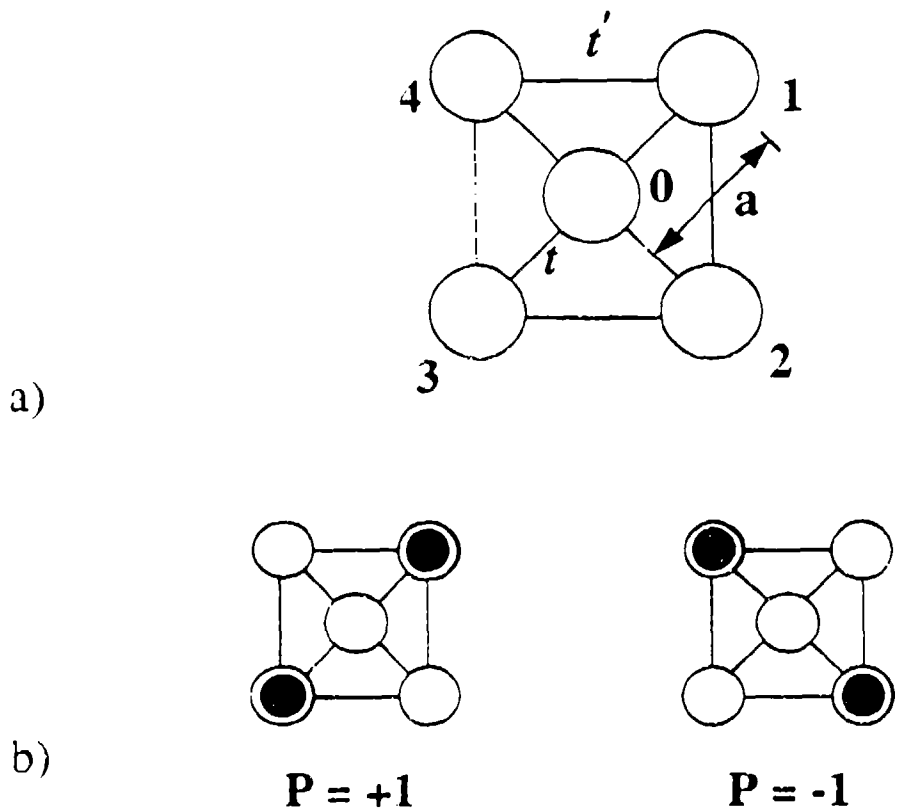


FIGURE 1. Schematic of the basic five-site cell. (a) The geometry of the cell. The tunneling energy between the middle site and an outer site is designated by t, while t' is the tunneling energy between two outer sites. (b) Coulombic repulsion causes the electrons to occupy antipodal sites within the cell. These two bistable states result in cell polarizations of P=+1 and P=-1 (see Eq. (1)).

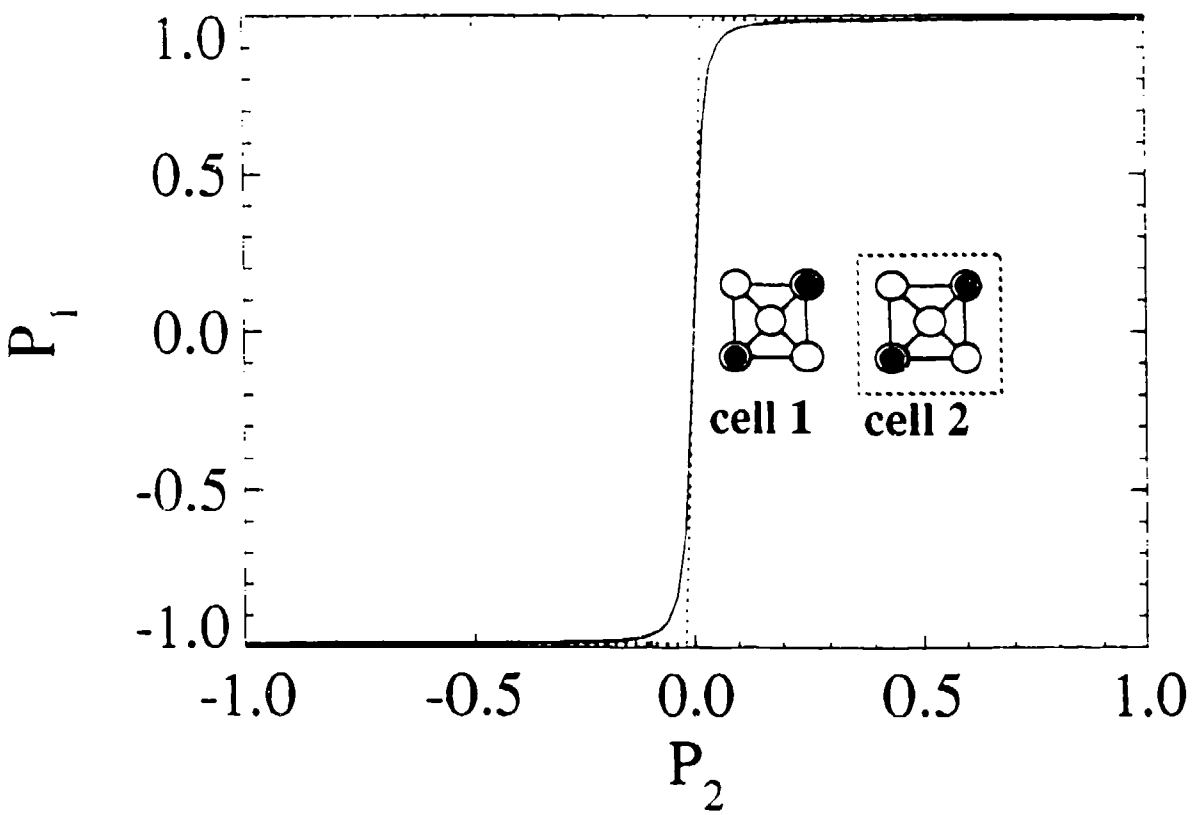


FIGURE 2. Cell-cell response function for the basic five-site cells shown in the inset. This shows the polarization P_1 induced in cell 1 by the fixed polarization of its neighbor, P_2. The solid line corresponds to antiparallel spins, and the dotted line to parallel spins. The two are nearly degenerate, especially for significantly large values of P_2.

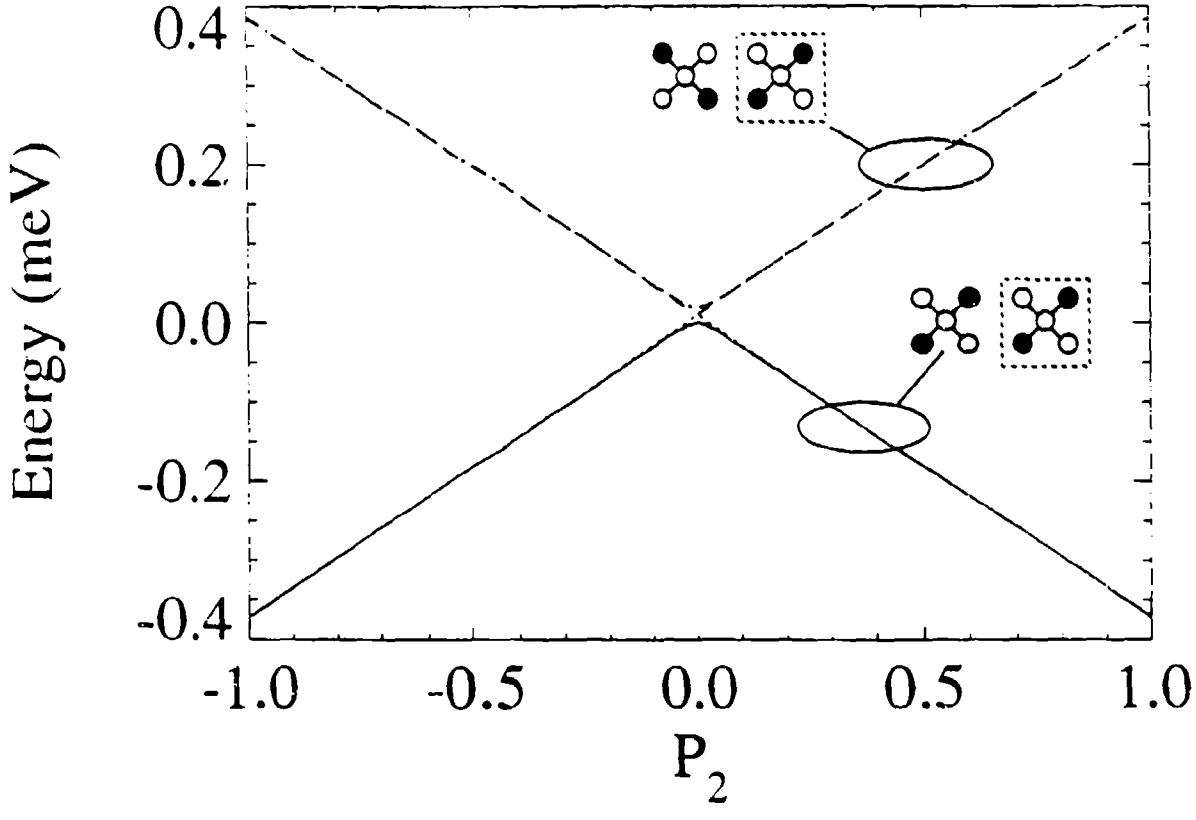


FIGURE 3. The lowest four eigenstate energies of cell 1 induced by the polarization of cell 2. The insets show that the lowest two energy states always correspond to the same polarization direction as in the driver. Slight exchange splitting effects between spatially symmetric and spatially antisymmetric states breaks the fourfold degeneracy for very small values of P_2.

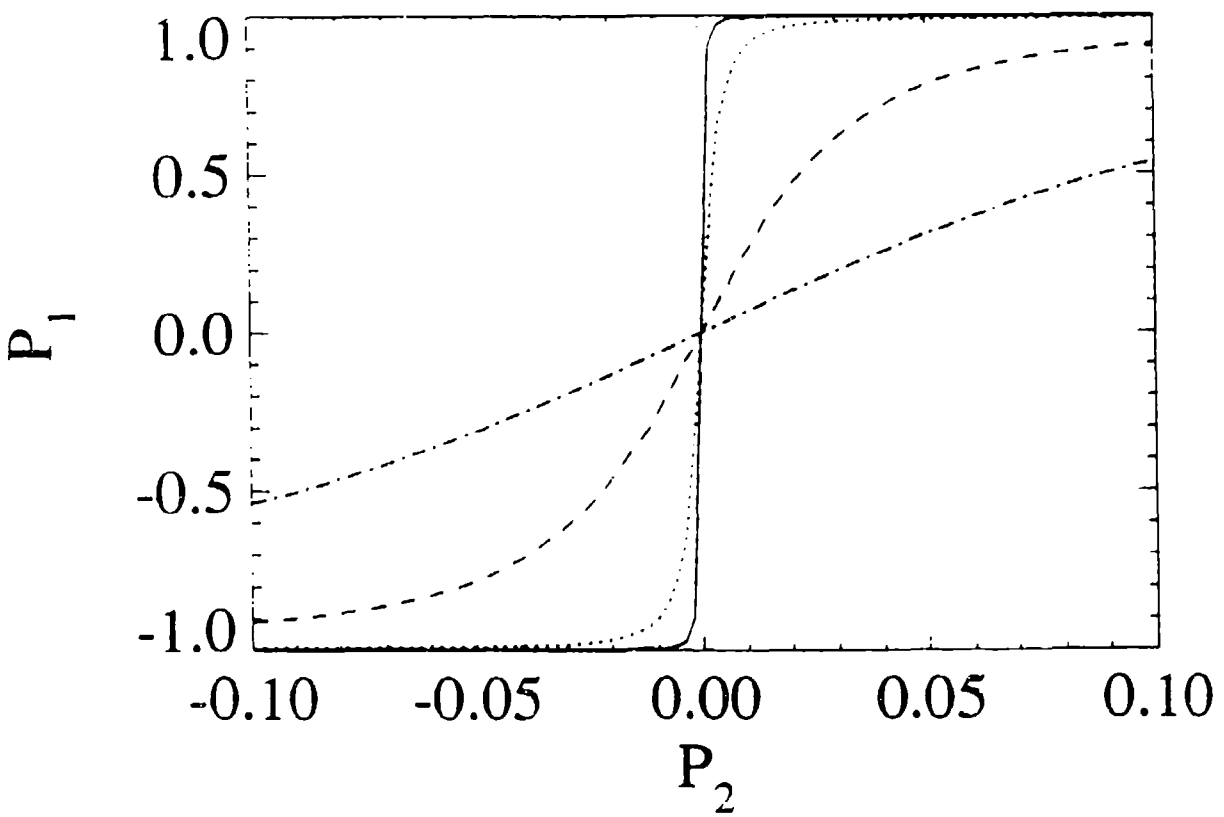


FIGURE 4. The cell-cell response function of the basic five-site cell for various values of the parameter t in Eq. (2) (after reference [4]). The induced polarization in cell 1 is plotted as a function of the polarization of its neighbor, cell 2. The curves correspond to t=0.2 meV (solid line), t=0.3 meV (dotted line), t=0.5 meV (dashed line), and t=0.7 meV (dot-dashed line). Note the horizontal axis only shows P_2 in the range [-0.1,+0.1].

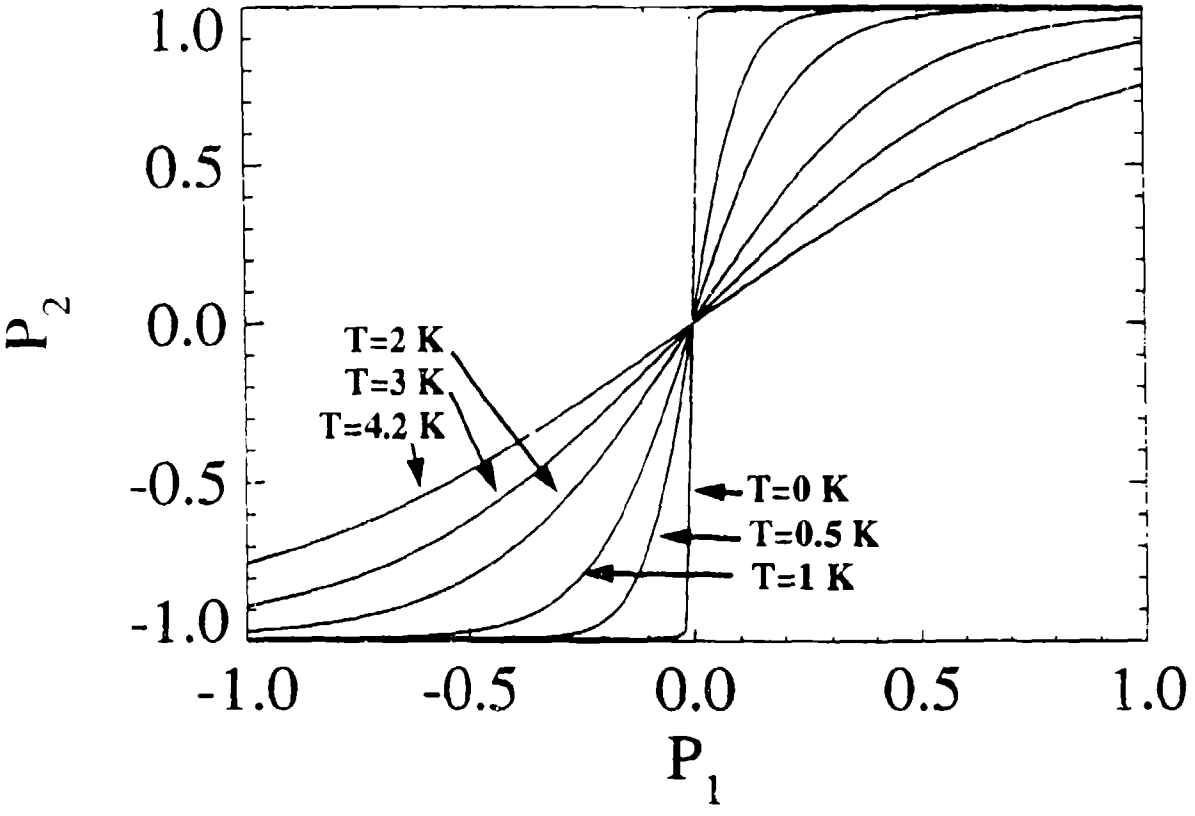
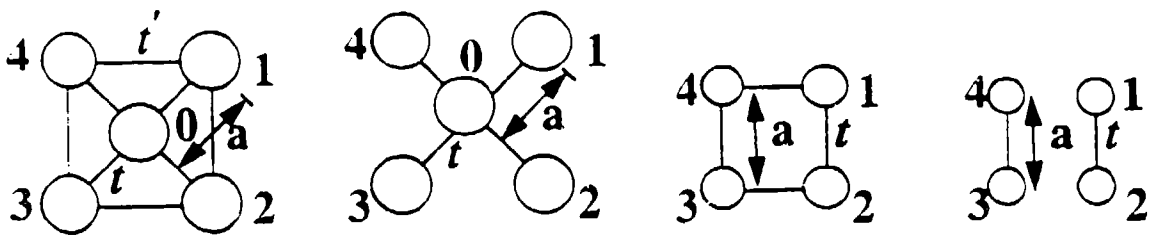


FIGURE 5. The cell-cell response function for the standard cell with t'=0 at various temperatures. The response degrades as temperature increases. Above 4.2K, the response would be unacceptable for use in a QCA. This maximum operating temperature is highly dependent on the physical size of the cell; molecule-sized cells would behave in a satisfactory manner up to room temperature.

a)



b)

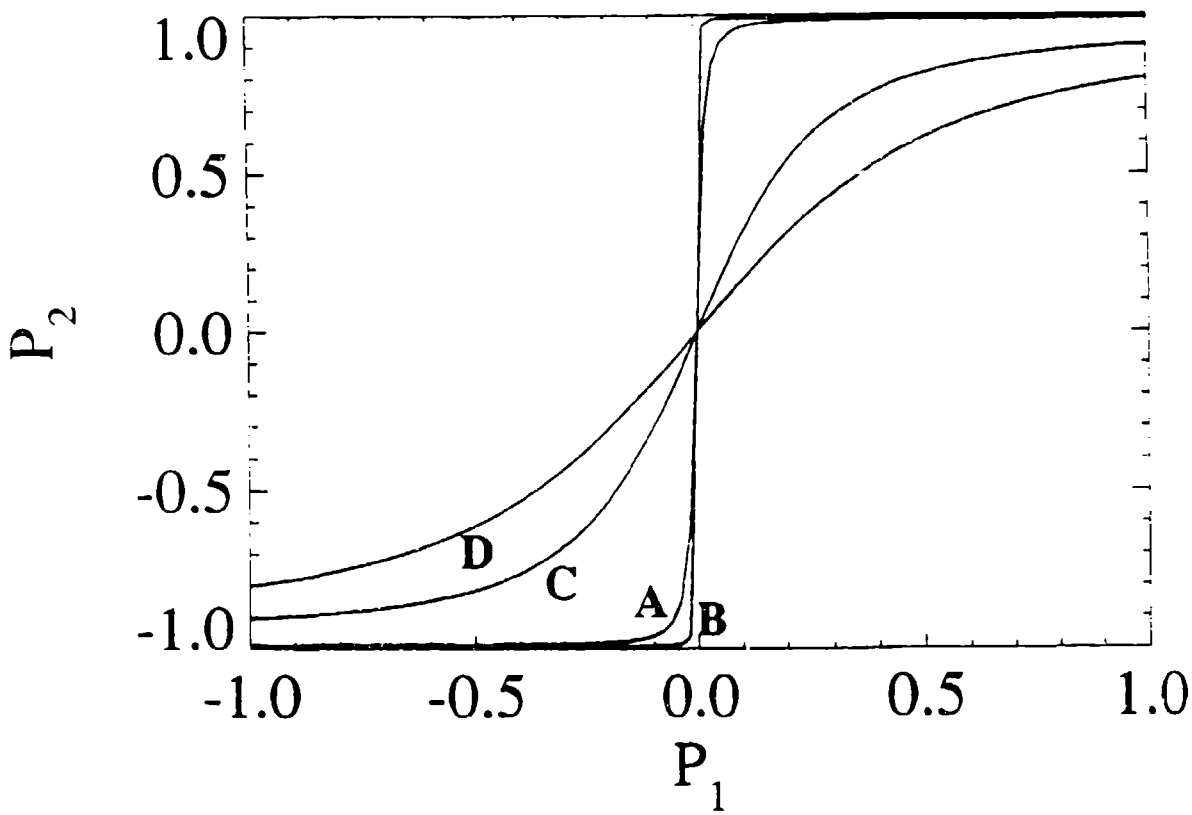


FIGURE 6. Four geometric variations on the simple model quantum cell. (a) Schematic diagrams of the four cells. Cells C and D occupy less area, but all four cells are drawn with the same minimum spacing between neighbors. Cells B and D will require potential variation between the sites to selectively inhibit tunneling. (b) The cell-cell response function for each of these cell designs. Cell B has the best response, but the improvement over A is small.

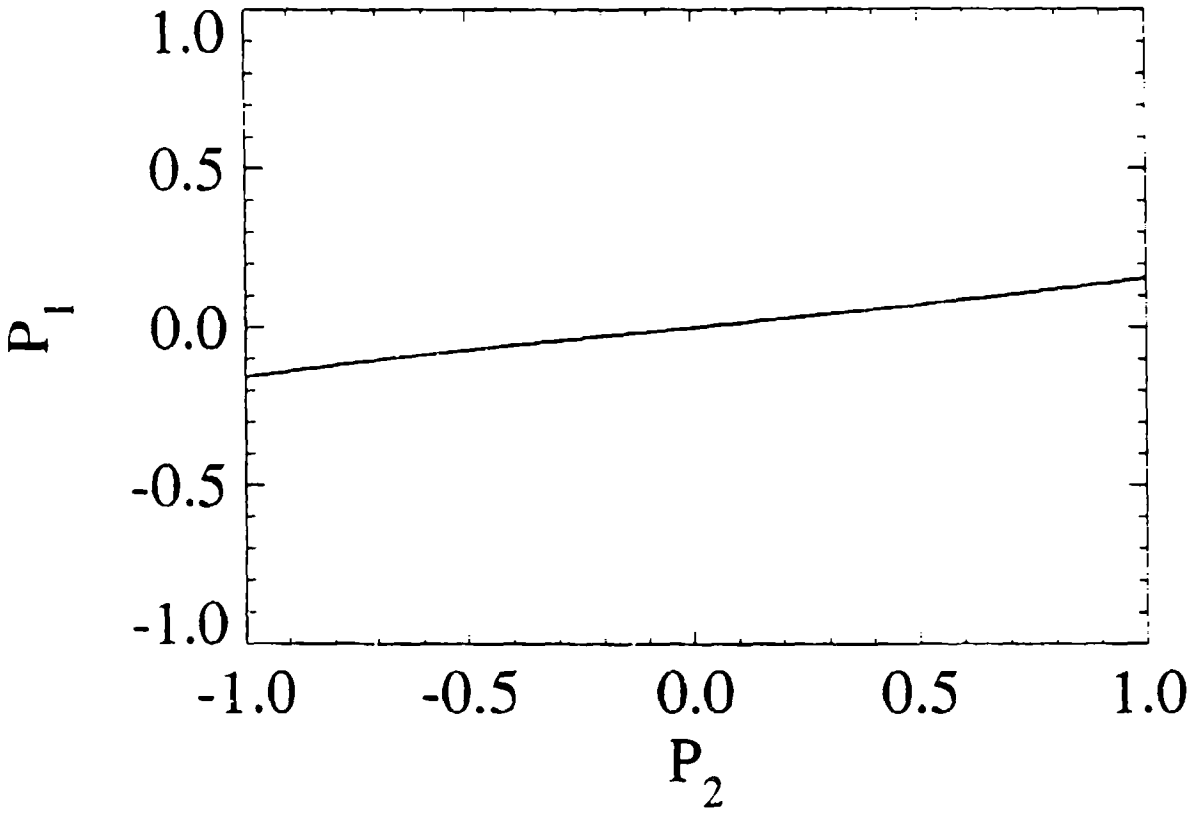


FIGURE 7. The cell-cell response function for the basic five-site cell occupied by a single electron. The weak response indicates that such a cell is unsuitable as the basis of a QCA.

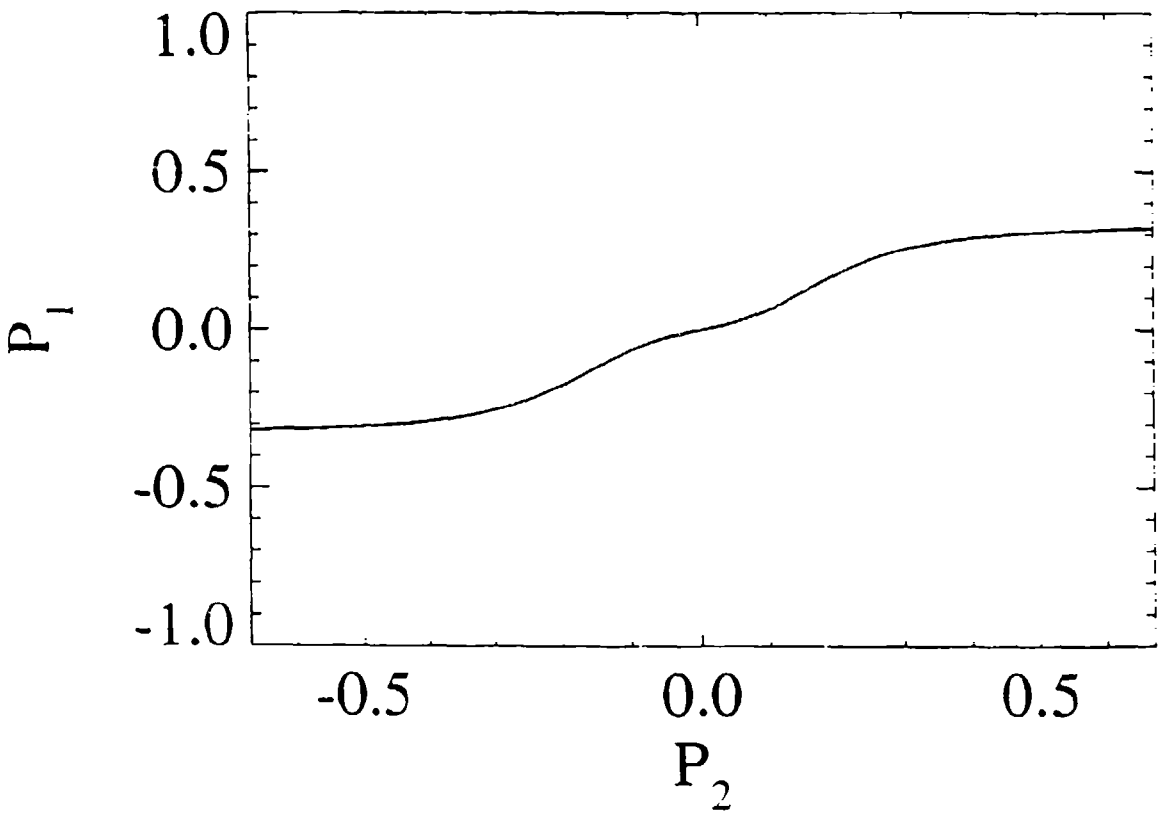


FIGURE 8. Cell-cell response function for the basic five-site cell occupied by three electrons. Such a cell is also unacceptable as the basis of a QCA.

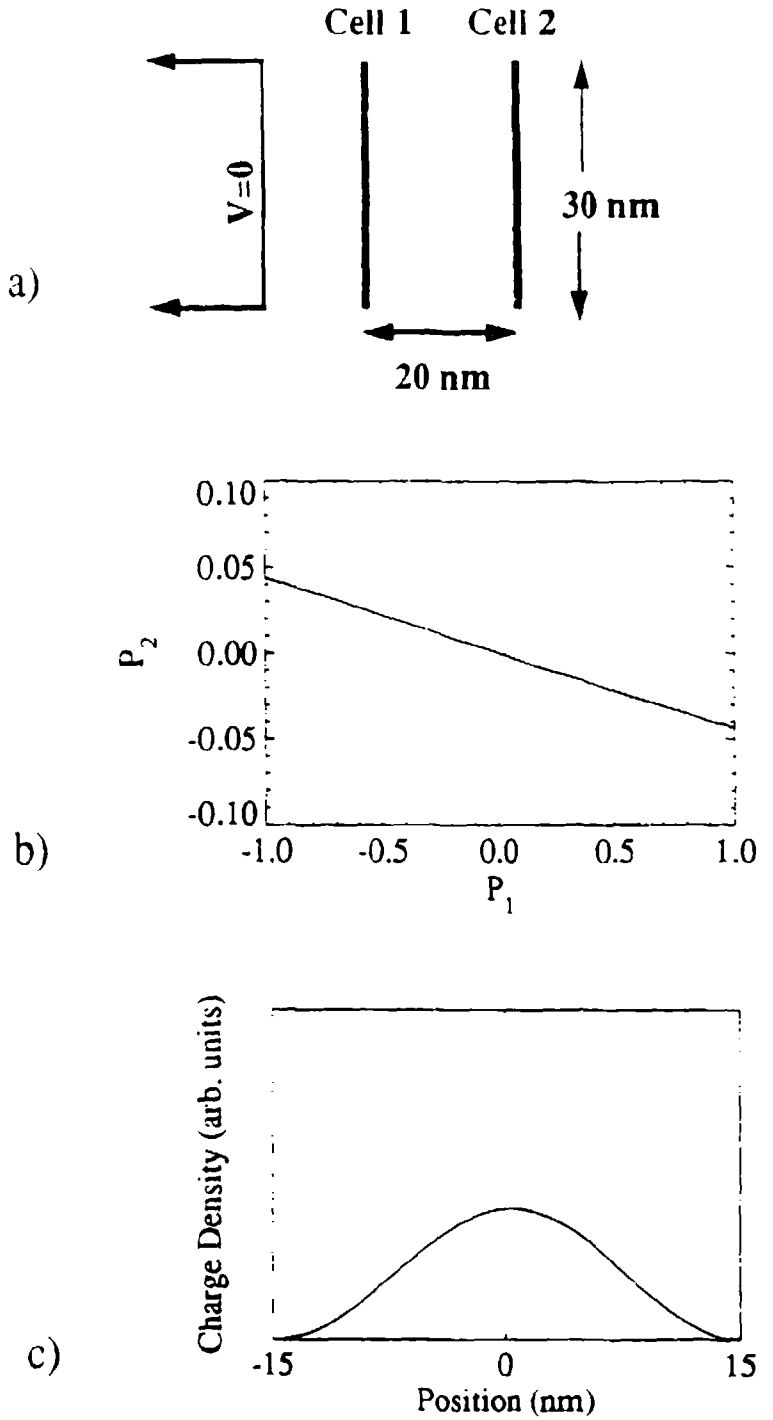


FIGURE 9. The "quantum dash" as a QCA cell. (a) A schematic diagram of the cellular arrangement. The length and spacing is similar to that of the basic five-site cell in Fig. (1a). Each cell is modelled as a one-dimensional infinite square well. The cell-cell response is shown in (b). Note that the vertical axis only shows P_2 over the range [-0.1,+0.1]. (c) The one-dimensional charge density in cell 2 for a fully polarized neighbor (P_1=1). The nearly symmetric charge density yields a very low polarization.

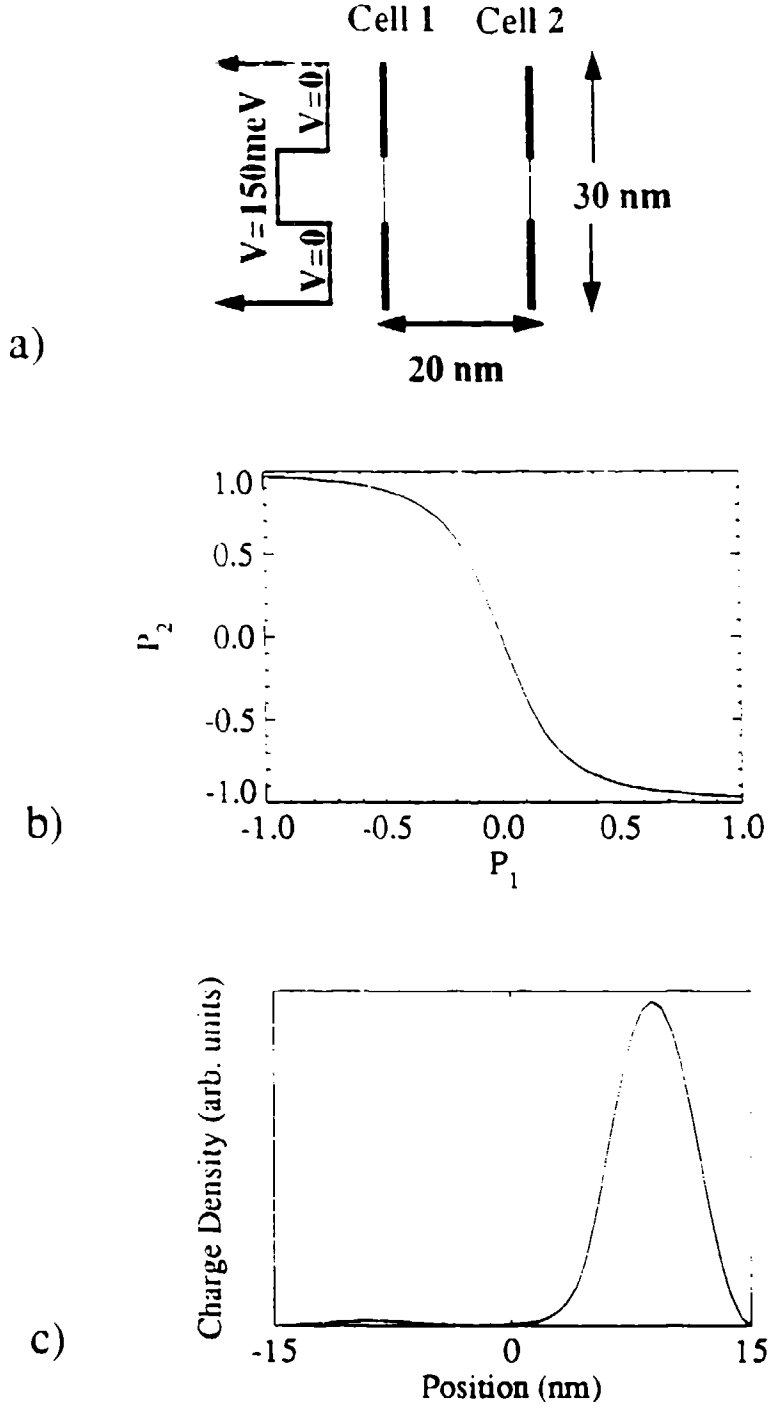


FIGURE 10. The double well as a QCA cell. (a) A schematic diagram of the cellular arrangement. The total dimensions are identical to those of Fig. (8). The middle third of each cell contains a 150 meV barrier to isolate the top and bottom of the cell. (b) The cell-cell response function for such an arrangement. Note that the vertical axis now shows P_2 over the range [-1.0, +1.0]. (c) The one-dimensional charge density in cell 2 for a fully polarized neighbor (P_1=1). The highly asymmetric charge density results in a cell that is almost completely polarized.

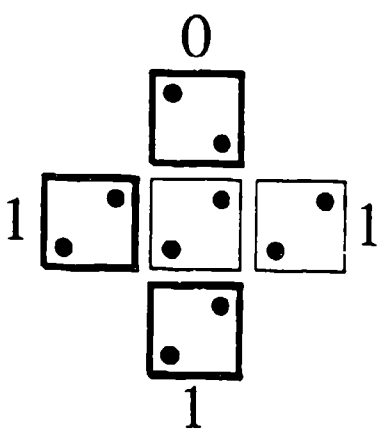


FIGURE 11. Majority voting logic. The states of the center and right cells are always the same as the majority of the three fixed neighbors. The cells with heavy borders have fixed charge densities. These are not schematic diagrams; they are the actual results of the ICHA solution of the ground state charge densities in this system. The diameter of each dot is proportional to the charge density on that site.

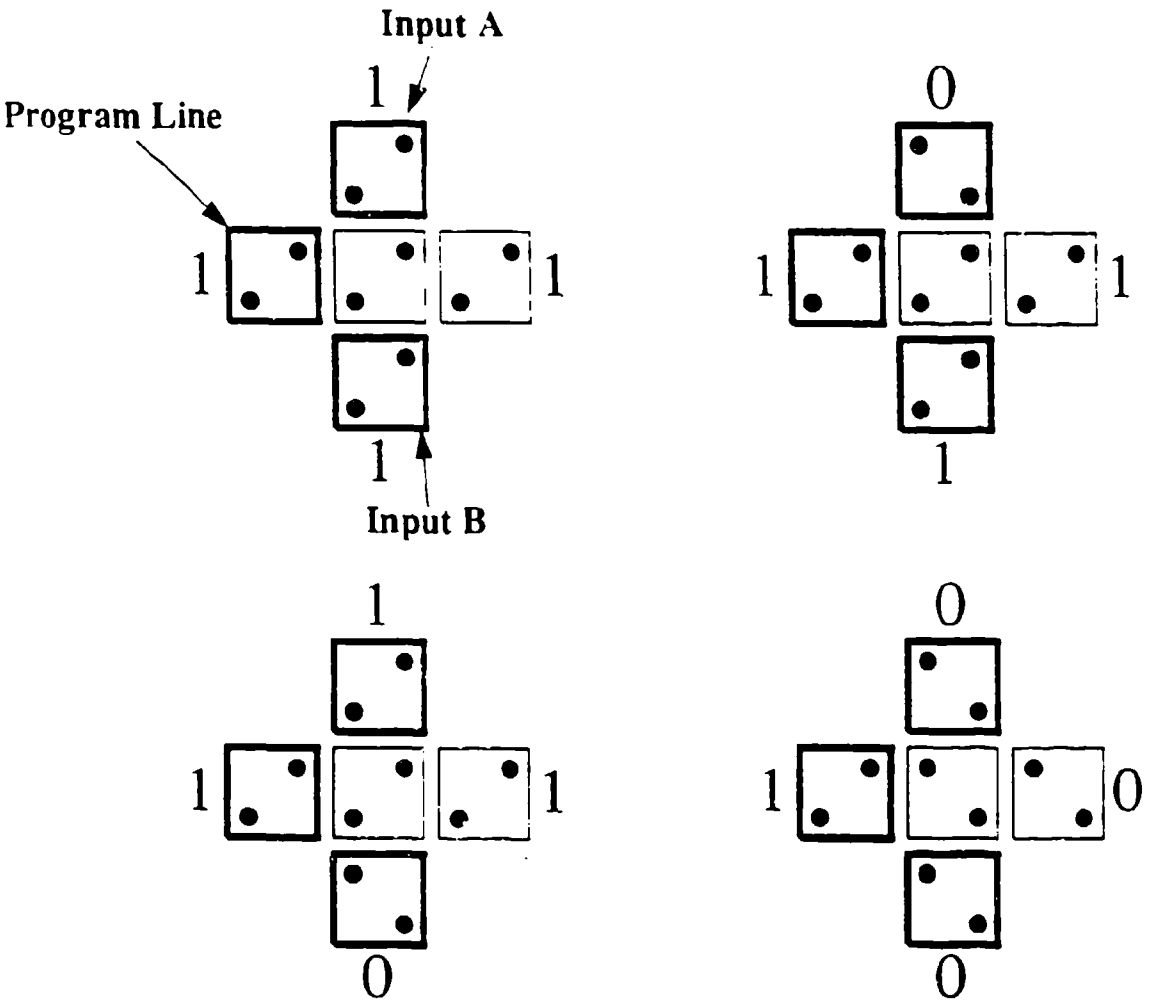


FIGURE 12. The programmable AND/OR gate. The program line is set to one in each system, so the gate is displaying OR logic. All four combinations of the non-program line inputs are shown. The cells with heavy borders have fixed charge densities. Any one of the three inputs could be the program line; the left cell is not special. These are not schematic diagrams; they are the actual results of the ICHA solution of the ground state charge densities in each system. The diameter of each dot is proportional to the charge density on that site.

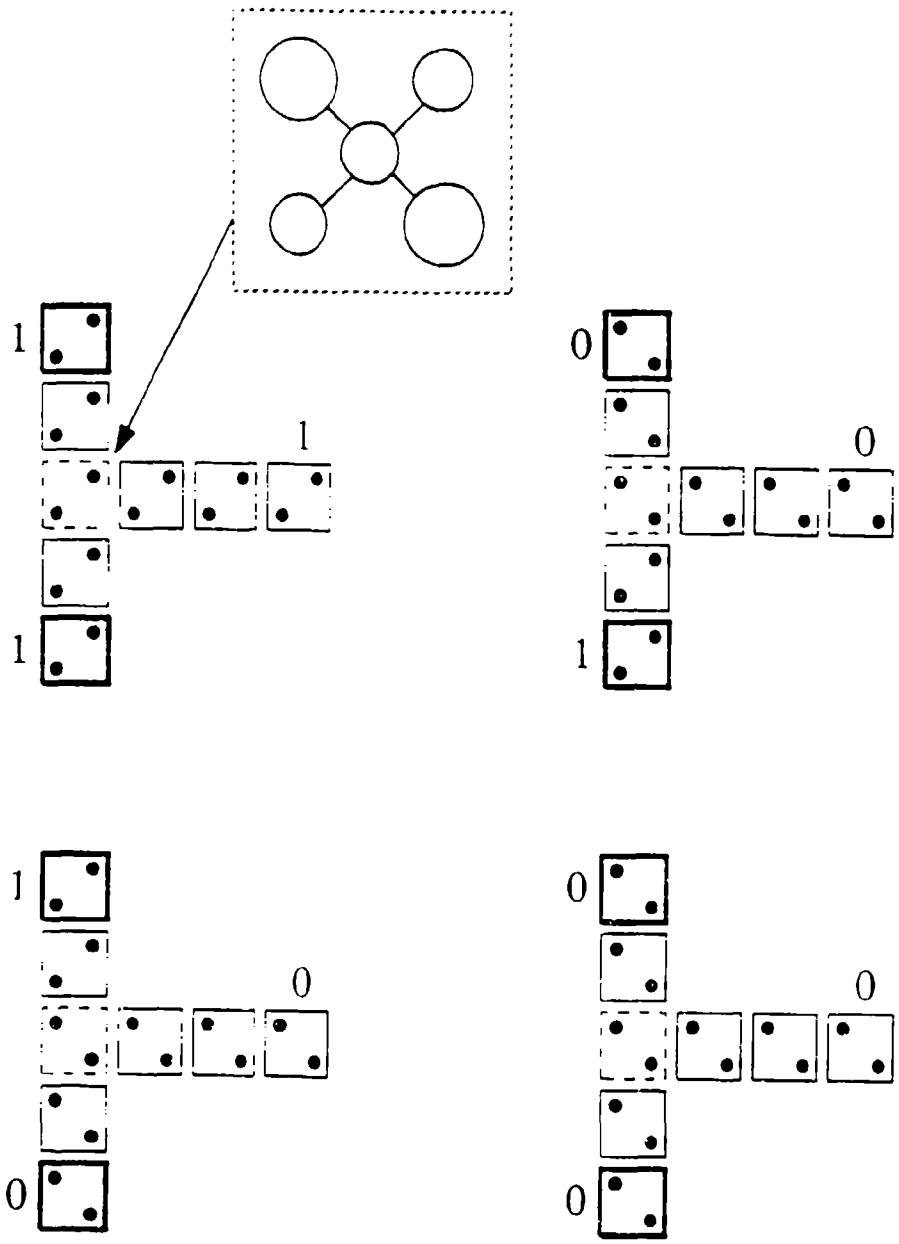


FIGURE 13. The non-programmable AND gate. All four combinations of the inputs are shown. The cells with heavy borders have fixed charge densities, while those with dotted borders are geometrically biased toward zero as shown in the inset. The bias is sufficient to decrease the on-site energy of the affected sites by 1%. Note that the output only equals one if both of the inputs are also one. These are not schematic diagrams; they are the actual ICHA results of the ground state charge densities in each system. The diameter of each dot is proportional to the charge density on that site.

Transmission resonances and zeros in quantum waveguides with resonantly coupled cavities

Wolfgang Porod, Zhi-an Shao, and Craig S. Lent
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 19 May 1992; accepted for publication 24 June 1992)

We report on the existence of transmission zeros in quantum waveguide structures with resonantly coupled cavities. Such zeros do *not* occur in the usual double-barrier resonant tunneling systems. For quantum waveguides, the transmission probability exhibits pairs of poles and zeros in the complex-energy plane. The observed sharp structure of the transmission resonances and zeros can be understood in terms of these zero-pole pairs.

Electronic transport in ultrasmall semiconductor structures resembles wave propagation in waveguides,[1,2] and device applications based on this analogy to microwave devices have been proposed. The transmission amplitude in these systems exhibits a rich structure which is related to resonance phenomena.[3-5] In this letter, we demonstrate how quasibound states in resonantly coupled cavities give rise to zero-pole pairs in the complex-energy plane. Transmission zeros are unique to quantum waveguide structures, and are absent for double-barrier resonant tunneling.

Resonances have long been studied in transmission through double-barrier resonant-tunneling structures.[6,7] It is well known that these resonances are related to the existence of quasibound states in the quantum-well region. Within the Breit–Wigner formalism, a quasibound state at energy E_0 and decay time $\tau=\hbar/\Gamma$ will give rise to a transmission resonance with a Lorentzian line shape, $T(E)=\frac{1}{4}\Gamma^2/[(E-E_0)^2+\frac{1}{4}\Gamma^2]$. In the complex-energy plane, this corresponds to a simple pole of the transmission amplitude at the complex energy $z=E_0-i\Gamma/2$.[8]

It is instructive to demonstrate why a quasibound state gives rise to a pole in the complex-energy plane. In analogy to an optical Fabry–Perot resonator, the total transmission amplitude across both barriers, $\mathcal{T}_{RL}$ (from left to right), may be related to partial transmission and reflection coefficients at each barrier:[9]

$$\mathcal{T}_{RL}=t_{RW}(e^{ikL}+e^{ikL}r_Re^{ikL}r_Le^{ikL}+e^{ikL}r_Re^{ikL}r_Le^{ikL}r_Re^{ikL}r_Le^{ikL}+\ldots)t_{WL} \quad (1)$$

$$=t_{RW}\frac{1}{e^{-ikL}-r_Re^{ikL}r_L}t_{WL}. \quad (2)$$

Here, t_{WL} denotes the transmission amplitude from the left to the well region, and r_R is the reflection coefficient at the right barrier. The phase accumulated by an electron with energy E moving from one side of the quantum-well resonator of length L to the other is represented by the phase factor $\exp(ikL)$, where $k=\sqrt{(2m^*E)/\hbar^2}$. Poles occur at those complex energies for which the denominator vanishes. For real-value reflection amplitudes, it is an easy matter to see that this occurs for wave numbers k whose real part is an integer multiple of π/L, i.e., for quasibound states. Note that this geometric series does not possess zeros.

As an illustration of the above arguments, we show in Fig. 1 the transmission amplitude in the complex-energy plane for a double-barrier resonant-tunneling structure which is schematically depicted in the inset. This structure supports true bound states for $E<0$, double-barrier resonances for $0<E<V_0$, and continuum resonances for $E>V_0$. Figure 1(a) shows the transmission coefficient on the real-energy axis ($E>0$), while Fig. 1(b) depicts a contour plot of the absolute value of the transmission amplitude in the complex-energy plane. Note that transmission resonances and poles occur at essentially the same energies.

We now proceed to the study of transmission in quantum waveguides in the presence of coupled resonators. The prototypical resonator structure we adopt is the resonantly coupled stub, as schematically shown in Fig. 2. The resonant cavity is created by closing off the side arm in a three-way branch [Fig. 2(b)]. The new feature here, with respect to quasi-one-dimensional double-barrier resonant-tunneling is the existence of wire branches which inherently makes this problem spatially two-dimensional.

We start out by considering transmission through the three-way splitter [Fig. 2(a)]. This problem has been studied by several workers,[10-12] and we follow them in choosing a scattering matrix approach which connects the out-going and in-coming waves in each wire branch,

$$\begin{pmatrix}O_L\\O_R\\O_S\end{pmatrix}=\begin{pmatrix}r_L & t_{LR} & t_{LS}\\ t_{RL} & r_R & t_{RS}\\ t_{SL} & t_{SR} & r_S\end{pmatrix}\begin{pmatrix}I_L\\I_R\\I_S\end{pmatrix}. \quad (3)$$

The elements of the scattering matrix are constrained by unitarity. The reflection and transmission amplitudes are the elements of a Shapiro–Büttiker matrix,[10,13] and their specific values depend upon the detailed geometry of the waveguide and junction. Typically, the elements of the scattering matrix are slowly varying functions of energy.

Making the sidearm into a resonator forces a standing wave in the stub [Fig. 2(b)]. This imposes an additional relationship between the out-going and in-coming amplitudes in the side branch, O_S and I_S, respectively,

$$O_S=\lambda(E)I_S, \quad (4)$$

where $\lambda=e^{i\Phi(E)}$ and the phase angle $\Phi(E)$ is a property of the resonator.[5]

The transmission and reflection amplitudes for the quantum waveguide with a resonantly coupled stub,

 0003-6951/92/361350-03$03.00

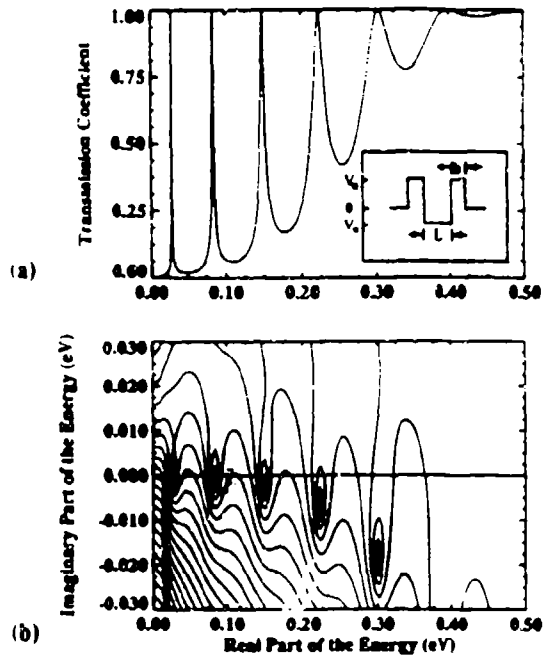


FIG. 1. Transmission amplitude for a double-barrier resonant-tunneling structure, which is schematically depicted in the inset ($V_0=0.2$ eV, $V_w=0.1$ eV, $L=30$ nm, and $b=5$ nm); (a) shows the transmission coefficient on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

and $\mathscr{R}$, may now be determined by combining Eqs. (3) and (4). For the case of transmission from left to right, they are given by

$$\mathscr{T}_{RL}=t_{RL}+\frac{t_{RS}t_{SL}}{\lambda-r_S}, \quad \text{and} \quad \mathscr{R}_L=r_L+\frac{t_{LS}t_{SL}}{\lambda-r_S}. \tag{5}$$

Note that the second terms contain a resonance denominator, which is reminiscent of double-barrier resonances, Eq. (2). An expansion of (5) yields

$$\mathscr{T}_{RL}=t_{RL}+t_{RS}\left(\frac{1}{\lambda}+\frac{1}{\lambda}r_S\frac{1}{\lambda}+\frac{1}{\lambda}r_S\frac{1}{\lambda}r_S\frac{1}{\lambda}+\cdots\right)t_{SL}, \tag{6}$$

which allows the following interpretation of the transmission and reflection amplitudes for the resonantly coupled stub. Each multiple reflection in the stub contributes a factors r_S for reflection back into the sidearm and a phase factor $1/\lambda$ for the roundtrip (up and down). The sum of all multiple reflections results in a geometric series, very much like a Fabry–Perot resonator. Structure in the transmission amplitude is due to two effects: (i) The resonance denominator which gives rise to poles, and (ii) the possibility of destructive interference between the first and the second terms which may result in transmission zeros. Note that in contrast to Eq. (1), zeros are now possible.

Next, we examine in more detail conditions for the existence of transmission zeros. As seen from Eq. (5), zeros in the transmission amplitude, $\mathscr{T}_{RL}=0$, occur if

$$\lambda=r_S-\frac{t_{RS}t_{SL}}{t_{RL}} \quad (\text{when } \mathscr{T}_{RL}=0). \tag{7}$$

The above condition relates a property of the resonator, λ, to the characteristics of the junction between the stub and the channel, t's and r's. In particular, Eq. (7) can only be true if the right-hand side is a phase factor on the unit circle, i.e., $|r_S-t_{RS}t_{SL}/t_{RL}|=1$. *It is, perhaps, a surprising consequence of unitarity, but easy to show, that always*

$$\left|r_S-\frac{t_{RS}t_{SL}}{t_{RL}}\right|=1. \tag{8}$$

Unitarity, therefore, ensures that both the left-hand side and the right-hand side of Eq. (7) are constrained to the unit circle, which implies that a transmission zero occurs when both phase angles are the same. This proves the existence of transmission zeros for transmission in waveguides.

In the following, we elucidate the above general arguments using specific model systems. We assume that the transmission channels are very thin wires, allowing us to consider only the coordinate along the wire. Transmission and reflection coefficients of such quasi-one-dimensional models have been studied in literature.[10,11,14] We match the values of the wave functions and the appropriate derivative boundary conditions at each branch point in the network.[15] For the symmetrical open branch, this results in reflection and transmission amplitudes which are independent of energy and orientation, $r_L=r_R=r_S=-1/3$ and $t_{LR}=t_{LS}=t_{SR}=2/3$.

The sidearm can now be made into a resonator of length L by erecting an infinite potential barrier. This forces the wave function to be zero there, and implies $\lambda(E)=-\exp(-2ikL)$. Thus, λ moves around the unit circle with an angular frequency proportional to the wave number. Using Eq (5), the transmission amplitude for this case of a strongly coupled stub can be given analytically, $\mathscr{T}=2/[2+i\cot(kL)]$. Figure 3 shows $\mathscr{T}(z)$ in the complex-energy plane: for a complex argument z. Note the appear-

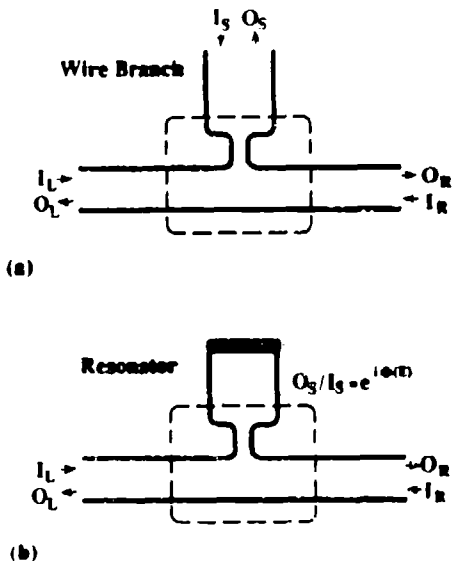


FIG. 2. Schematic drawing of a waveguide with a resonantly coupled cavity. (a) shows a wire branch with incoming and outgoing waves outside the junction region, which is indicated by the dashed box; (b) shows a resonant stub which is obtained by closing off the sidearm.

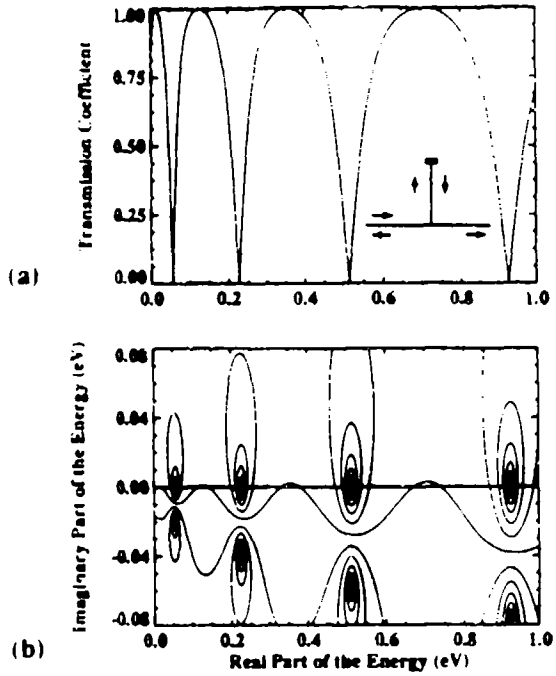


FIG. 3. Transmission amplitude for the strongly coupled stub, which is schematically depicted in the inset; (a) shows the transmission coefficient on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

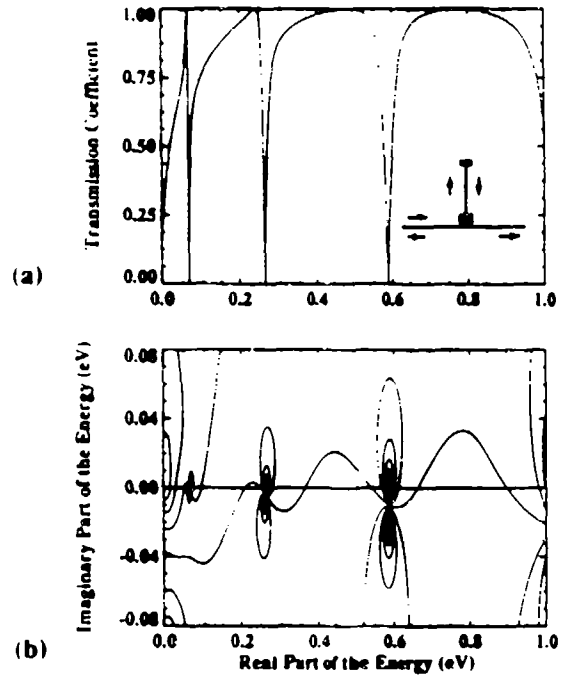


FIG. 4. Transmission amplitude for the weakly coupled stub, which is schematically depicted in the inset; (a) shows the transmission coefficient on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

ance of transmission zeros on the real-energy axis, and the existence of poles in the fourth quadrant of the complex-energy plane. The zeros occur at energies for which a standing wave forms in the stub, i.e., when $k = n \cdot \pi / L$ with $n = 1, 2, \ldots$. Also note that the maxima of the transmission coefficient do not occur at the location of the poles, as for double-barrier resonant tunneling. This is a particularly simple example since the elements of the scattering matrix are independent of energy.

An energy dependence in the elements of the scattering matrix may be introduced by weakly connecting the stub to the channel via a tunneling barrier, as schematically depicted in the inset of Fig. 4. We present numerical results for a tunneling barrier of 0.5-eV height and 1-nm thickness. Figure 4 shows the transmission amplitude for this weakly coupled stub in the complex energy plane. Note again the existence of transmission zeros as predicted by Eq. (8). The poles move closer to the real-energy axis which corresponds to the longer lifetime of the resonant states due to the confining barrier. The most striking feature is the occurrence of a zero-pole pair for each quasibound state. With increasing barrier height, the pole approaches the zero which leads to a sharper and sharper transition between a transmission zero and one on the real-energy axis (compare Figs. 3 and 4). In the limit of an infinitely high barrier, the poles and zeros merge which corresponds to $\mathcal{T} = 1$ for a completely decoupled stub.

In summary, we have demonstrated that quasibound states in cavities, which are resonantly coupled to quantum waveguides, lead to zero-pole pairs of the transmission amplitude in the complex-energy plane. The proximity of these zeros and poles leads to sharp variations of the transmission coefficient with energy. Zeros are a new "animal" not observed in the usual one-dimensional transmission resonances.

We are grateful for stimulating discussions with Dr. S. Bandyopadhyay, Dr. M. Büttiker, Dr. S. M. Goodnick, Dr. W. Pötz, Dr. P. Price, and Dr. M. Sain. This work was supported, in part, by AFOSR and ONR.

[1] *Nanostructure Physics and Fabrication*, edited by M. Reed and W. P. Kirk (Academic, Boston, 1989).
[2] *Analogies in Optics and Micro Electronics*, edited by W. Van Haeringen and D. Lenstra (Kluwer Academic, Dordrecht, 1990).
[3] F. Sols, M. Macucci, U. Ravaioli, and K. Hess, J. Appl. Phys. **66**, 3892 (1989).
[4] C. S. Lent, in *Computational Electronics*, edited by K. Hess, J. P. Leburton, and U. Ravaioli (Kluwer, Boston, 1990), p. 259.
[5] P. J. Price, IEEE Trans. Electron. Devices **39**, 520 (1992).
[6] P. J. Price, Phys. Rev. B **38**, 1994 (1988).
[7] M. Büttiker, IBM J. Res. Dev. **32**, 63 (1988).
[8] T. B. Bahder, C. A. Morrison, and J. D. Bruno, Appl. Phys. Lett. **51**, 1089 (1987).
[9] S. Datta, *Quantum Phenomena* (Addison-Wesley, Reading, MA, 1989).
[10] M. Büttiker, Y. Imry, and M. Ya. Azbel, Phys. Rev. A **30**, 1982 (1984).
[11] M. Cahay, S. Bandyopadhyay, and H. L. Grubin, Phys. Rev. B **39**, 12 989 (1989).
[12] D. Kowal, U. Sivan, O. Entin-Wohlman, and Y. Imry, Phys. Rev. B **42**, 9009 (1990).
[13] B. Shapiro, Phys. Rev. Lett. **50**, 747 (1983).
[14] C. H. Wu and G. Mahler, Phys. Rev. B **43**, 5012 (1991).
[15] J. E. Avron, A. Raveh, and B. Zur, Rev. Mod. Phys. **60**, 873 (1988).

Revision 6 5/11/93

Transmission Resonances and Zeros in Quantum Waveguide Systems with Attached Resonators

Zhi-an Shao, Wolfgang Porod, and Craig S. Lent

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

(Received May , 1993)

Transmission phenomena in quantum waveguide structures are studied by examining the transmission amplitude in the complex-energy plane. We find that, similar to double-barrier resonant tunneling, there are transmission poles in the complex-energy plane for quantum waveguide structures which contain quasi-bound states in attached resonators, such as t-stubs and loops. In contrast to double-barrier resonant tunneling, however, we also find that the quantum wire networks also possess transmission zeros (antiresonances), which always occur on the real-energy axis. The existence of transmission zeros is a unique feature of quantum waveguide system with attached resonators, but is absent for double-barrier resonant tunneling, which contains the resonant cavity as part of the transmission channel. We demonstrate that each quasi-bound state of the resonantly-coupled quantum waveguide system leads to a zero-pole pair of the transmission amplitude in the complex-energy plane. The previously noted resonance/antiresonance behavior of the transmission coefficient which leads to its sharp variation as a function of energy, can be understood in terms of these zero-pole pairs. We discuss our findings in the context of Fano resonances (U. Fano, Phys. Rev. **124**, 1866 (1961)) which are known to occur when

two scattering channels are available, one correpsonding to a continuum of states and the other one to a discrete quasi-bound state.

PACS numbers: 72.10.Bg, 73.20.Dx, 73.40.Gk

I. INTRODUCTION

Electronic transport in ultra-small semiconductor structures reveals the quantum mechanical wave nature of the charge carriers. As shown in pioneering papers by Landauer and Büttiker [1,2], electronic conduction in this so-called mesoscopic regime [3] can be viewed as a transmission problem, and the conductance is related to the transmission coefficient by the quantum unit of conduction e^2/h. Much work has been inspired by analogous wave phenomena in optics and the possibility of utilizing resonant transmission behavior for electronic device applications [4]. Fabry-Perot-like transmission resonances in semiconductor superlattice structures have been studied since the seminal work of Esaki and Tsu [5]. The phenomenon of double-barrier resonant tunneling (DBRT) is well understood and practical devices based on this concept, even operating at room temperature, have been developed [6]. Recent work has focused on transmission in electronic waveguides and related quantum interference devices, and it has been noted that resonance phenomena in these structures give rise to rich features in the transmission coefficients [7].

In this paper we investigate resonance phenomena for transmission in quantum waveguide structures by studying the transmission amplitude in the complex-energy plane. In analogy to double-barrier resonant tunneling, we find that transmission resonances are related to the existence of quasi-bound states. There is one major

difference, however, as pointed out by us in a recent letter [8]. DBRT gives rise to the well-known Lorentzian-shaped Breit-Wigner transmission resonances, which correspond to poles in the complex-energy plane. The tranmission coefficient for resonantly coupled waveguides, on the other hand, exhibits resonance/antiresonace features, which correspond to zero-pole pairs in the complex transmission amplitude. We will demonstrate that these antiresonances, in fact, are zeros and that their existence is a consequence of unitarity for transmission in quantum waveguide structures.

Resonance-antiresonance features including sharp transmission minima have been seen in recent numerical work on conduction in quantum waveguide structures and for transmission through an oscillating barrier [9]. A much-studied system consists of a main transmission channel with an attached t-stub resonator [10–16]. Sharp drops to a minimum are observed in the transmission coefficient (we will show that those, in fact, are zeros), and forbidden bands are formed for multiple stub systems [17–20]. Using a scattering matrix approach, Price [22] shed light on the relationship between the transmission coefficient and the quasi-bound states in the resonant t-stubs. Another example of a resonantly-coupled waveguide consists of a channel which is connected to a circular cavity [21]. In this so-called "quantum whistle" the quasi-bound states in the cavity are excited at resonant energies by the "electron wind" in the channel, giving rise to sharp structure in the transmission coefficient. Impurities in a transmission waveguide also give rise to conduction in the presence of quasi-bound states, and resonance effects have been observed in several studies [23–28]. Geometric effects in conduction channels may also result in bound states [29], and the transmission coefficient has been studied for bends [30, 31], corners [32, 33], crosses [34–37], etc. The sharp drops of the transmission coefficient are also found in the loop structures [38–41]. Similar resonant behavior has also been reported for transmission in systems where several electronic subbands are available. Such

studies include electronic Γ — X conduction band minima [42-44] and heavy-hole — light-hole [45] interference effects for resonant tunneling in GaAs/AlAs double-barrier heterostructures. Common to all of the above examples is that resonance features are observed when more than one scattering channel is available [46]. When continuum states interact with localized states, two scattering channels are available, one belongs to a continuum and the other to a bound state. These so-called Fano resonances [47] have first been studied for the interaction between light and electrons from atoms and molecules. The two scattering channels may also belong to the two arms of a loop, or to two different electronic conduction band minima (for example, Γ and X).

The prototypical system studied in this paper is shown in Fig. 1(a). An incident wave impinges upon the resonator structure under study, as schematically represented by the dashed box, and transmitted and reflected waves emerge. The classes of systems investigated here comprise double-barrier resonant tunneling and quantum wire networks, including t-stubs and loops. The quasi-one-dimensional problem of DBRT can be thought of as transmission in a quantum wire which contains two barriers, as shown in Fig. 1(b) where the potential barriers on the channel are represented by the shaded boxes. Also shown is a "zoo" of so-called strongly- and weakly-coupled t-stubs and loops which will be investigated. As described in more detail in the appendix, we utilize several different numerical as well as analytical schemes to obtain the transmission and reflection amplitudes in the complex-energy plane. Our analysis concentrates on the analytical behavior of these functions, in particular on the zeros and/or poles which represent the quasi-bound states in the resonators. We assume perfectly ballistic transport and quasi-one-dimensional dynamics, which may be realized in the lowest subband of a very narrow quantum wire.

The body of this paper is organized as follows: Double-barrier resonant tunneling is reviewed and discussed in section II. A general treatment of quantum waveguides

with attached resonators is presented in section III, which includes the proof for the existence of transmission zeros as a consequence of unitarity. We study thin-wire networks in section IV and show results for t-stub and loop structures. Section V contains concluding remarks. Details of the mathematical formalism are given in the appendix.

II. DOUBLE-BARRIER RESONANT TUNNELING

It is well known that transmission resonances occur in double-barrier resonant-tunneling structures [48, 49], and that these resonances are related to the existence of quasi-bound states in the quantum well region. This relationship can be made explicit when viewing the transmission amplitude in the complex-energy plane [50, 51]. It has been shown that the poles of the propagator are the same as the poles of the transmission amplitude [51]. Consequently, a quasi-bound state at energy E_0 and decay time $\tau = \hbar/\Gamma$ gives rise to a simple pole in the transmission amplitude, $t(z)$, at the complex energy, $z = E_0 - i\Gamma/2$. If this pole is sufficiently close to the real-energy axis, the transmission probability, $T(E) = |t(E)|^2$, for a physical energy on the real-energy axis, E, is given by,

$$T(E) = \frac{\frac{1}{4}\Gamma^2}{(E - E_0)^2 + \frac{1}{4}\Gamma^2} \quad , \tag{1}$$

which is a Lorentzian line centered at energy E_0 with a full width at half maximum of Γ. This Breit Wigner formula [52, 53] describes the transmission resonance which is caused by the quasi-bound state at energy E_0 and decay time $\tau = \hbar/\Gamma$.

It is instructive to demonstrate how a quasi-bound state, which is the constructive superposition of multiply reflected waves, gives rise to a pole in the complex-energy plane. In analogy to an optical Fabry-Perot resonator [54], the total transmission amplitude, T, can be expressed in terms of the partial transmission and reflection amplitudes, t's and r's, at each barrier. For transmission from left to right,

$$\mathcal{T}_{RL} = t_{RW}(e^{ikL} + e^{ikL}r_R e^{ikL}r_L e^{ikL} + e^{ikL}r_R e^{ikL}r_L e^{ikL}r_R e^{ikL}r_L e^{ikL} + \cdots)t_{WL} \quad (2)$$

$$= t_{RW}\frac{1}{e^{-ikL} - r_R e^{ikL}r_L}t_{WL}. \quad (3)$$

Here, t_{WL} and t_{WR} denote the transmission amplitudes from the left to the well region and from the well region to the right, respectively. The reflection amplitudes at the right and the left barriers are denoted by r_R and r_L, respectively. An electron with energy E and wavenumber $k = \sqrt{(2m^*E)/\hbar^2}$ accumulates a phase factor of $exp(ikL)$ in traversing the width of the well, which is denoted by L. Poles in the transmission amplitude are thus possible at those energies and wavenumbers for which the resonance denominator in (3) vanishes. It is an easy matter to see that, for real-valued reflection amplitudes, this occurs for wavenumbers k with a real part which is an integer multiple of π/L. But this also is precisely the condition which determines the energies of the quasi-bound states. Therefore, transmission resonances and resonant states coincide in energy. We also emphasize that the geometric series in (2) does not posses zeros (except for the trivial case of zero transmission, i.e. $t_{LW} = 0$ or $t_{WR} = 0$.)

We illustrate the above arguments with a numerical example. Figure 2 shows the transmission coefficient in the complex-energy plane for a double-barrier resonant tunneling system which is schematically depicted in the inset (barrier height $V_0 = 0.2$ eV, barrier depth $V_W = -0.1$ eV, well width $L = 30$ nm, and barrier thickness $b = 5$ nm). This structure supports true bound states for $V_W < E < 0$, quasi-bound states for $0 < E < V_0$, and continuum resonances for $E > V_0$. Part (a) of Fig. 2 shows the transmission coefficient on the real-energy axis ($E > 0$), while part (b) depicts a contour plot of the absolute value of the transmission amplitude in the complex-energy plane. Note that transmission resonances and poles occur at essentially the same real energies and that no transmission zeros exist. The lowest three resonances

possess energies below the barrier height; they correspond to long-lived states with poles close to the real-energy axis. The continuum resonances at higher energies have shorter lifetimes and the corresponding poles move farther into the complex-energy plane. Alternately, increasing the strength of the barrier (height or thickness) moves the poles closer up to the real-energy axis, which implies a longer lifetime of the resonant state. In the limit of an infinite barrier, only bound states exist and the poles move onto the real axis at these bound-state energies.

It is well known that quasi-bound states lead to rapid variations in the phase of the transmission amplitude. Figure 3 represents the phase, modulo 2π, of $\mathcal{T}_{RL}$ in the complex-energy plane; in part (a) on the real-energy axis, and in part (b) as a contour plot in the plane. Going full circle around each pole, there is a phase change of 2π indicative of a simple pole of the form $1/(E-(E_0-i\Gamma/2))$. Note that lines of constant phase "stream downward" toward more negative imaginary parts.

Figure 4 shows the charge accumulated in the quantum well, which is obtained as the integral of the charge density between the barriers. Consistent with our interpretation, maximum accumulation is found at those energies which correspond to the quasi-bound states. The resonant charge buildup is due to the constructive superposition of reflected and transmitted waves.

III. QUANTUM WAVEGUIDES WITH ATTACHED RESONATORS

It is well known that the transmission coefficient in a quantum waveguide possesses structure due to changes in the waveguide geometry, such as bends, constrictions, crosses, etc. The related problem of conduction in waveguides in the presence of quasi-bound states has received less attention, although a few studies have been reported [21, 22]. Here, we specifically focus on transmission in quantum waveguide structures which are connected to resonant cavities. The isolated resonator possesses

bound states; attaching it to the conduction channel allows the wavefunction to leak out. We will show that this coupling introduces both finite lifetimes of the quasi-bound states and resonance effects in the transmission coefficient. These features are similar to the double barrier resonant tunneling problem considered in the previous section, however, the structure of the transmission amplitude in the complex-energy plane is different in each case. In particular, we will demonstrate that transmission zeros exist for the resonant waveguides, and that the complex transmission amplitude possesses zero-pole pairs which are related to the quasi-bound states.

Coupling between the quantum waveguide and the resonator is accomplished in two steps, as schematically shown in Fig. 5. First, the junction region is viewed as the branch point in a three-way splitter, as indicated by the dashed line in Fig. 5(a). The properties of this wire branch are described by a scattering matrix, which is further explained in section III.A. Second, the resonator is obtained by closing off the side arm, as schematically shown in Fig. 5(b). The standing wave in the resonant cavity is characterized by a phase factor, which is detailed in section III.B. Based on the unitarity of the scattering matrix, we are able to prove that the transmission coefficient for these structures exhibits zeros.

A. Wire Branch

For the three-way splitter, as shown in Fig. 5(a), the amplitudes of out-going and in-coming waves are related by the 3×3 scattering matrix:

$$\begin{pmatrix} O_L \\ O_R \\ O_S \end{pmatrix} = \begin{pmatrix} r_L & t_{LR} & t_{LS} \\ t_{RL} & r_R & t_{RS} \\ t_{SL} & t_{SR} & r_S \end{pmatrix} \begin{pmatrix} I_L \\ I_R \\ I_S \end{pmatrix} \tag{4}$$

The elements of the scattering matrix represent the reflection amplitudes in each branch, the r's, and the transmission amplitudes between the various arms, the t's;

the branches are denoted by L, R, and S, for left, right, and side (or stub), respectively. We have implicitly assumed single-moded behavior, i.e. there is only one relevant transverse mode in each branch to scatter to. Time reversal invariance, which applies here in the absence of magnetic fields, constrains the scattering matrix to be symmetric [55, 56]. Current conservation requires this scattering matrix also to be unitary, which implies the following relationships:

$$r_L r_L^* + t_{LR} t_{LR}^* + t_{LS} t_{LS}^* = 1 \tag{5}$$

$$t_{RL} t_{RL}^* + r_R r_R^* + t_{RS} t_{RS}^* = 1 \tag{6}$$

$$t_{SL} t_{SL}^* + t_{SR} t_{SR}^* + r_S r_S^* = 1 \tag{7}$$

And:

$$r_L t_{RL}^* + t_{LR} r_R^* + t_{LS} t_{RS}^* = 0 \tag{8}$$

$$r_L t_{SL}^* + t_{LR} t_{SR}^* + t_{LS} r_S^* = 0 \tag{9}$$

$$t_{RL} t_{SL}^* + r_R t_{SR}^* + t_{RS} r_S^* = 0 \tag{10}$$

The values of the elements in the scattering matrix depend upon both the carrier energy and the junction geometry. This, of course, offers us the possibility of controllable device operation. For the special case of a completely symmetrical three-way splitter with three identical arms, the scattering matrix is constant, which will be shown below. Generally, the transmission and reflection amplitudes for a branch are slowly varying functions of energy.

B. Resonator

Terminating the side arm completely, results in the formation of a standing wave in the stub. The amplitudes of the out-going and in-coming waves then no longer are linearly independent, but are constrained by an additional relationship.

$$O_S = \lambda(E)\, I_S. \tag{11}$$

Here, $\lambda = exp(i\Phi(E))$ is a phase factor which describes the standing wave. The energy-dependent phase $\Phi(E)$ depends upon the details of the resonator geometry.

Transmission in the presence of the resonator is described by a 2×2 scattering matrix. The condition (11) reduces the dimensionality of the original S-matrix by 1. Combining (4) and (11), it is an easy matter to show that,

$$\begin{pmatrix} O_L \\ O_R \end{pmatrix} = \begin{pmatrix} \mathcal{R}_L & \mathcal{T}_{LR} \\ \mathcal{T}_{RL} & \mathcal{R}_R \end{pmatrix} \begin{pmatrix} I_L \\ I_R \end{pmatrix}, \tag{12}$$

where $\mathcal{T}_{RL}$ denotes the transmission amplitude from left to right, and $\mathcal{R}_L$ the reflection amplitude for the left-hand side, which are given by,

$$\mathcal{T}_{RL} = t_{RL} + \frac{t_{RS}t_{SL}}{\lambda - r_S}, \tag{13}$$

$$\mathcal{R}_L = r_L + \frac{t_{SL}t_{SL}}{\lambda - r_S}. \tag{14}$$

Note that the total amplitudes depend upon the characteristics of the t-stub resonator, i.e. $\lambda(E)$, and the details of the three-way splitter, i.e. the partial amplitudes $t(E)$ and $r(E)$. In a fashion similar to double-barrier resonant tunneling, the second term of eqn. (13) can be expanded as,

$$\mathcal{T}_{RL} = t_{RL} + t_{RS}\left(\frac{1}{\lambda} + \frac{1}{\lambda}r_S\frac{1}{\lambda} + \frac{1}{\lambda}r_S\frac{1}{\lambda}r_S\frac{1}{\lambda} + \cdots\right)t_{SL}. \tag{15}$$

The resulting geometric series, contained in the brackets, describes multiple reflections in the side arm. Each roundtrip in the stub (up and down) contributes a phase factor $1/\lambda$, and each reflection back into the stub a factor r_S. The amplitudes for being scattered into (from the left) and out (to the right) of the resonator are denoted by t_{SL} and t_{RS}, respectively, and t_{RL} denotes the straight-through transmission path without a detour into the resonant stub.

The general form of $\mathcal{T}_{RL}$ is that it consists of two independent terms, as seen in eqns. (13) - (15). One term describes the straight-through transmission in the absence of the stub, and the second term is a geometric series which is due to multiple reflections in the resonator. This has as a consequence the following structure of the transmission amplitude in the complex-energy plane: (i) poles are possible because of the resonance denominator, in analogy to double-barrier resonant tunneling; and (ii) zeros are possible because the two terms may cancel each other. The existence of both terms is the major difference with respect to double-barrier resonant tunneling. In the latter case, as derived in the previous section, the transmsission amplitude is given by just the geometric series.

The form $\lambda - r_S$ for the denominator of $\mathcal{T}_{RL}$ also predicts that the sharpness of the resonance features will depend upon the strength of the coupling between the side arm and the channel. For a weakly coupled t-stub, there is a large probability for reflection back into the side arm, i.e. $|r_S| \to 1$. This implies that in this case r_S may approach λ on the unit circle (compare Fig. 6), thus giving rise to sharp resonance features. On the other hand, a strongly-coupled t-stub has $|r_S| = 1/3$ and therefore the amplitude r_S never gets close to λ on the unit circle, resulting in weak features in the transmission coefficient.

C. Proof of the Existence of Transmission Zeros

In the previous section we demonstrated that transmission zeros are possible, in principle. In this section we prove that, in fact, transmission zeros must occur and that their existence is a consequence of unitarity.

A zero of the transmission amplitude at a certain energy means that $\mathcal{T}_{RL}=0$ for that energy. As can be seen from eqn. (13), this implies that

$$\lambda = r_S - \frac{t_{RS}t_{SL}}{t_{RL}} \quad (when\ \mathcal{T}_{RL} = 0). \tag{16}$$

The above condition relates a property of the resonator, the phase factor λ, to a property of the junction bewteen the waveguide and the resonant cavity, the t's and r's which are the elements of the scattering matrix for the three-way splitter. Because λ is a phasor on the unit circle, (16) can only be true if the r.h.s. also lies on the unit circle for those energies at which transmission zeros occur. It is a perhaps surprising consequence of unitarity that *for all energies* the r.h.s. of (16) is constrained to be on the unit circle, *i.e. always*

$$\left| r_S - \frac{t_{RS} t_{SL}}{t_{RL}} \right| = 1. \tag{17}$$

The detailed proof of the above equation is given in Appendix B1. Unitarity, therefore, ensures that both the l.h.s. and the r.h.s. of (16) are constrained to the unit circle, as schematically shown in Fig. 6, which implies that a transmission zero occurs when both phase angles are the same. Usually, the r.h.s. is a slowly varying function of energy, and the l.h.s. is a phasor which moves around the unit circle with an angular dependence proportional to the wavenumber, as shown in the examples below [57].

Using similar arguments, one may also investigate the conditions for unity transmission, $\mathcal{T}_{RL}=1$. The unitarity of the scattering matrix by itself is not sufficent to ensure the existence of ones in the transmission amplitude. As shown in Appendix B2, $\mathcal{T}_{RL}=1$ holds true for certain energies if the structure is symmetric with respect to 'left' and 'right,' i.e. if the scattering matrix possesses the symmetry property that $t_{RS} = t_{LS}$, etc. This behavior is similar to double barrier resonant tunneling where it is known that perfect transmission only occurs for symmetric barrier structures [48, 51].

IV. THIN-WIRE NETWORKS

We illustrate the above general arguments by demonstrating several specific examples. As a model system, we choose networks of thin wires which are sufficiently

narrow that only motion in the direction of the wires is of interest. The motion perpendicular to the wire is frozen in the lowest transverse subband resulting in quasi one-dimensional dynamics. These thin-wire networks exhibit the essential behavior of the transmission amplitude in the complex-energy plane under study here, without additional geometric complications inherent in a true two-dimensional system. For certain structures our model is sufficiently simple to yield analytical answers. Other cases require numerical investigation. We show results for stronlgy- and weakly-coupled stubs and symmetric and asymmetric loops.

A. Three-way Splitter

The basic building block of the stub and loop structures is the simple wire branch, or three-way splitter. The wavefunction in the left, right, and stub branches are denoted by Ψ_L, Ψ_R, and Ψ_S, respectively. For a perfect wire, they are superpositons of in-coming and out-going plane waves,

$$\begin{aligned} \Psi_L &= I_L e^{ikx_L} + O_L e^{-ikx_L} \\ \Psi_R &= O_R e^{ikx_R} + I_R e^{-ikx_R} \\ \Psi_S &= O_S e^{ikx_S} + I_S e^{-ikx_S}. \end{aligned} \tag{18}$$

Here, x_L, x_R and x_S denote the spatial coordinates in each branch, i.e., left, right, and stub, respectively. The matching conditions at the branch point, $x_L = x_R = x_S = 0$, for the wavefunctions and the derivative are:

$$\begin{aligned} I_L + O_L &= O_R + I_R \\ I_L + O_L &= O_S + I_S \\ I_L - O_L &= O_R - I_R + O_S - I_S \end{aligned} \tag{19}$$

The last of these relations is the matching condition for the derivatives. We require the sum of the derivatives in all directions at a branch point to be equal to zero, which is a generalization of the current conservation condition. However, the above

derivative condition is stronger than current conservation; it implies the conservation of current, but not vice versa. Similar conditions have been used in the literature of quantum networks [58], where the sum of the derivatives is related to a so-called vertex potential which is a delta-function potential at a given branch point. In our case, the vertex potential is assumed to be zero and the general Sturm-Liouville boundary condition reduces to the above requirement of the derivatives adding to zero.

The above system of equations represents 3 conditions which constrain the 6 amplitudes for the incoming and outgoing waves in each of the three branches. These 6 amplitudes are also related by a scattering matrix when viewing the 3 amplitudes of the incoming waves as independent parameters, and the 3 amplitudes of the outgoing waves as the dependent variables. The 3 equations (19) thus are sufficient to completely specify the scattering matrix. It is an easy matter to show that,

$$\begin{pmatrix} O_L \\ O_R \\ O_S \end{pmatrix} = \begin{pmatrix} -1/3 & 2/3 & 2/3 \\ 2/3 & -1/3 & 2/3 \\ 2/3 & 2/3 & -1/3 \end{pmatrix} \begin{pmatrix} I_L \\ I_R \\ I_S \end{pmatrix} \tag{20}$$

Comparing with (4), we see that $r_L = r_R = r_S = -1/3$ and $t_{RL} = t_{RS} = t_{SL} = 2/3$ for this simple three-way splitter.

In Appendix C we show that the above result for the scattering matrix is a special case of previous work [38, 41, 12] in which the coupling of a ring to a lead was studied.

B. T-Stubs

We now create a t-stub resonator in the form of a dangling-wire of length L by truncating the side arm of the three-way splitter with an infinite potential barrier at $x_S = L$. The wavefunction has to be zero at the end of the side arm, $O_S e^{ikL} + I_S e^{-ikL} = 0$, which implies that the amplitudes of the in-coming and out-going waves are related

by a phase factor, $\lambda = O_S/I_S$, as previously disc .s geometry, the standing waves in the resonator are characterized by the phasor, $\lambda(E) = -exp(-2ikL)$, with a phase that changes linearly with wavenumber k.

The transmission amplitude for this so-called strongly-coupled stub, which is schematically shown in the inset of Fig. 7, can now be found by using in eqn. (13) the above form for λ and the elements of the scattering matrix for the wire branch (20). It is an easy matter to show that the transmission amplitude is given analytically by:

$$\mathcal{T} = \left[1 + \frac{i}{2}cot(kL)\right]^{-1}. \tag{21}$$

Figure 7 shows the transmission amplitude in the complex-energy plane for a strongly-coupled t-stub with stub length $L = 10$ nm; part (a) shows the transmission coefficient on the real-energy axis, and part (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane. Note the appearance of transmission zeros on the real-energy axis, and the existence of transmission poles in the complex-energy plane. Figure 8 represents the phase of the transmission amplitude in the complex-energy plane for the strongly-coupled stub. The phase, modulo 2π, is shown in part (a) on the real-energy axis, and in part (b) as a contour plot in the plane. Note that at each quasi-bound state, which coincides with the transmission zeros, the phase rapidly changes by π. The most striking feature of Fig. 8 is the demonstration of zero-pole pairs in the complex-energy plane. The zeros occur at energies for which standing waves form in the stub, i.e. when $k = n\pi/L$ with $n = 1, 2, \cdots$. The wavefunction at a transmission zero has to be zero at the branch point which forces the wavefunction in the resonantor to be zero at both ends of the stub. *Note also that the maxima of the transmission coefficient do not align with the location of the poles, as in the case of double-barrier resonant tunneling.*

Because of the symmetry of the structure, transmission maxima in this case are transmission ones. For $\mathcal{T} = 1$, the wavefunction at the branch point must be 1, which implies a standing wave in the stub with a maximum at the branch point and a zero at the end point: i.e. when $k = (2n+1)\pi/(2L)$ with $n = 0, 1, 2, \cdots$. In general, maxima for waveguide structures occur between two zeros at some intermediate energy which is determined by the proximity to the real-energy axis of neighboring poles. For this simple case, the elements of the scattering matrix are constants, i.e. independent of energy. The energy dependence of the transmission amplitude is provided by the resonance phasor $\lambda(E)$.

An energy dependence in the elements of the scattering matrix may be introduced by weakly coupling the stub to the channel via a tunneling barrier of length l and height V_0, as schematically depicted in the inset of Fig. 9. The transmission amplitude may be given in closed form,

$$\mathcal{T} = \left[1 + \frac{i}{2}\frac{K}{k}\frac{[k\cos(k\Delta)\cos(Kl) - K\sin(k\Delta)\sin(Kl)]}{[k\cos(k\Delta)\sin(Kl) + K\sin(k\Delta)\cos(Kl)]}\right]^{-1} \qquad (22)$$

where $k = \sqrt{2m^*E}/\hbar$, $K = \sqrt{2m^*(E - V_0)}/\hbar$, and $\Delta = L - l$. For the numerical example shown in Fig. 9, we choose a tunneling barrier at the branch point with a height of 0.5 eV and a thickness of 1 nm (The same dimensions will also be used for the barriers in subsequent examples). Figure 9 depicts the transmission amplitude in the complex-energy plane for this so-called weakly-coupled stub. Note again the existence of transmission zeros on the real-energy axis, as predicted by our general arguments in section III. Note also that, with respect to the previous example of a strongly-coupled stub, the poles in the complex plane now are closer to the real axis, which corresponds to the longer lifetime of the resonant states due to the confining barrier. As a consequence of the zeros and poles approaching each other in the complex plane, the maxima on the real axis move closer to the zeros, which results in the stronger

energy dependence observed in Fig. 9(a) when compared to Fig. 7(a). Each resonant state produces a zero-pole pair in the complex energy plane which gives rise to the observed energy dependence of the transmission coefficient on the real-energy axis. With increasing barrier height, the pole approaches the zero which leads to a sharper and sharper transition between a transmission zero and one on the real-energy axis. In the limit of an infinitely high barrier, the poles and zeros merge which corresponds to $\mathcal{T} = 1$, and independent of energy, for a channel with a completely decoupled stub.

The proximity of transmission zeros and ones on the real-energy axis, which results in the more or less sharp variations of the transmission coefficient with energy, can be understood from a wavefunction argument. The electronic states in the resonator are standing waves, which have to match to the wavefunctions in the channel at the branch point. This implies that for a transmission zero or one, the wavefunction at the branch point has to be zero or one, respectively. This circumstance is illustrated in Figures 10 and 11, where the branch point is labeled by 0 on the spatial coordinate. Shown are the absolute values of the wavefunctions in the stub at both transmission zeros and ones for the cases of strongly- and weakly-coupled t-stubs, respectively. As discussed above, and illustrated in Figs. 10(a) and 11(a), for the strongly-coupled stub the standing waves in the resonator differ by a quarter wavelength for $\mathcal{T} = 0$ and $\mathcal{T} = 1$. This implies that in this case the zeros and ones are well separated in energy, as seen in Fig. 7(a). For the weakly-coupled stub, the standing wave is connected to the wire branch via a tunneling barrier, as illustrated in Figs. 10(b) and 11(b). Now only a small change in the wavelength of the standing wave is needed as the transmission coefficient varies from a maximum to a minimum. Also note that the amplitudes of the standing waves at a zero are larger than those at a one. Figure 12 illustrates, as a function of energy, the charge build-up in the t-stub, which is measured by the integral of $|\Psi|^2$ over the length of the stub. The arrows at the top

axis denote the locations of the poles of the transmission amplitude in the complex-energy plane for the cases of strongly- and weakly-coupled stubs, which are identified in the insets; compare also Figs 7(b) and 9(b). The resonant states for the weakly-coupled stub are sharper in energy and they contain more charge than those of the strongly-coupled stub, which is due to the confinement provided by the tunneling barrier.

It is interesting to study double-barrier resonant tunneling in addition to t-stub resonances. The results of adding two tunneling barriers on the transmission channel for strongly- and weakly-coupled stubs are shown in Figs. 13 and 14, respectively. The separation between these two barriers with height $V_0 = 0.5$ eV is assumed to be $d = 4$ nm, and the length of the stub is again $L = 10$ nm. We notice in both cases that there is, in addition to the now familiar zero-pole pairs, another set of poles in the complex-energy plane, and a corresponding set of transmission ones on the real-energy axis. The additional ones are caused by resonant tunneling through the two tunneling barriers on the transmission channel, in complete analogy to the previously discussed case of double-barrier resonant tunneling. Comparing Figs. 7 and 13 for the strongly-coupled stub, we note that the transmission zeros occur at exactly the same energies. Because of the confinement provided by the barriers, the zero-pole pairs are closer together in Fig. 13 than in Fig. 7, which is also reflected in the sharper energy dependence on the real-energy axis. The additional pole at $E = 0.14$ eV, which is associated with the transmission one, is due to the first resonant state energy of two tunneling barriers on the transmission channel. Because of our choice for the separation between the barriers and the length of the stub, this energy is close to the energy of the second lowest t-sub resonant state. This leads to the apparent coupling of the two poles, and the resulting "stretching" of the second lowest zero-pole pair. The case of the weakly-coupled stub with the double barriers

on the transmission channel is illustrated in Fig. 14; compare to Fig. 9 without the barriers in the channel. Again, the additional double barrier resonant tunneling does not alter the location of the zeros which are a property of the quasi-bound states in the stub, cf. Figs. 14 and 9. Also, the additional barriers lead to the very close zero-pole pairs and sharp transmission peaks shown in Fig. 14. Because of the relative lengths chosen, the lowest double-barrier resonant state is close in energy to the second lowest quasi-bound state in the t-stub. As also explained above, this leads to an interaction between the pole which is due to double barrier resonant tunneling, and the pole which belongs to the t-stub's zero-pole pair.

The interaction of the poles due to the double-barrier resonances and a weakly-coupled t-stub is illustrated in Fig. 15. The separation between the double barriers, which are placed in the channel symmetrically around the stub, is $d = 4$ nm, $d = 6$ nm, and $d = 8$ nm for parts (a), (b), and (c), respectively. In all cases, the total length of the t-stub is $L = 10$ nm. For Fig. 15(a), lowest double-barrier resonant state is close in energy to the second quasi-bound state in the stub, which leads to the observed interference between the lowest isolated pole and the second zero-pole pair. No strong interaction is observed in Fig. 15(b), which shows sharp zero-pole pairs and isolated poles at intermediate energies. In Fig. 15(c), every double barrier resonance is close in energy to a t-stub state, and there is again strong interaction between the poles in the complex-energy plane. Figure 16 shows the phase in the complex-energy plane for the same structure as in Fig. 15(b). Note the rapid phase change for each zero-pole pair, compare Fig. 8, and the behavior of the isolated poles which is similar to that in Fig. 3.

All wire structures discussed so far are symmetrical about the t-stub, which implies that the scattering matrix possesses a left-right invariance. As shown in the previous figures and proven in the appendix, this symmetry property ensures that

transmission maxima are transmission ones. Non-symmetrical wire structures are shown in the insets of Figs. 17 and 18 for strongly- and weakly-coupled t-stubs, respectively. The left-right symmetry is broken by placing a single tunneling barrier on one side of the transmission channel at a distance of $d = 2$ nm from the branch point. While transmission maxima no longer correspond to perfect transmission, $\mathcal{T} < 1$, transmission zeros still persist in accordance with our theory. Note furthermore, that the asymmetrical barrier in the channel does not alter the location of the zeros, which is a property of the t-stub; compare Fig. 17 to Fig. 7 (strong coupling), and Fig. 18 to Fig. 9 (weak coupling). For the strongly-coupled t-stub, there are poles in the complex-energy plane which correspond to standing waves between the tunneling barrier and the end of the stub. An example is the additional pole between the lowest and second lowest zero-pole pair shown in Fig. 17. For the weakly coupled t-stub, these additional quasi-bound states are not possible because of the barrier at the branch point. Consequently, Fig. 18 does not show bouble barrier resonant tunneling poles in addition to the familiar zero-pole pairs.

C. Loops

We also investigated transmission in loop structures which has been the topic of numerous previous studies reported in the literature. Our key finding reported here is that loops also exhibit zero-pole pairs of the transmission amplitude in the complex-energy plane. Loops, in analogy to t-stubs, may be viewed as resonators with associated quasi-bound states, which give rise to transmission resonances and zeros.

Figures 19 and 20 show the transmission amplitude in the complex-energy plane for symmetrical and asymmetrical loops, respectively, as sketched in the insets. The lengths of the lower and upper arms are denoted by L_1 and L_2, respectively. For these

so-called strongly-coupled loops, the transmission amplitude may be found analytically by cascading the scattering matrices of the two branch points, which represent the coupling to the leads, and with values given by eqn. (20). A phase factor e^{ikL_1} represents the phase change in traversing the arm with length L_1, with a similar term for the other arm.

$$\mathcal{T} = \frac{4e^{ikL_1}(1 - e^{2ikL_2}) + 4e^{ikL_2}(1 - e^{2ikL_1})}{(3 - e^{ik(L_1+L_2)})^2 - (e^{ikL_1} + e^{ikL_2})^2}. \tag{23}$$

From the above analytical expression for $\mathcal{T}$, we see that transmission zeros occur when the numerator in eqn. (23) vanishes, i.e.

$$sin(kL_1) + sin(kL_2) = 0. \tag{24}$$

This requires the arguments of both sine-functions to satisfy,

$$k(L_1 + L_2) = 2n\pi, \tag{25}$$

or

$$k(L_1 - L_2) = (2n + 1)\pi. \tag{26}$$

The first relation, eqn. (25), is the condition for the formation of a standing wave around the loop. The second relation, eqn. (26), is the condition for destructive interference due to the Aharonov - Bohm effect. Note that for the symmetrical loop, a zero in the numerator is accompanied by a zero in the denominator of (23).

The strongly-coupled symmetric loop, as shown in Fig. 19 for $L_1 = L_2 = 10.5$ nm, exhibits transmission resonances with $\mathcal{T} = 1$, but no zeros are visible. Note the existence of poles in the plane which are due to multiple reflections between the two branch points, in a fashion analogous to double barrier resonant tunneling. The symmetric loop is special since it possesses true bound states which produce poles on

the real-energy axis. These poles exactly cancel the zeros of T, compare Fig. 19(a), which would otherwise be present. Transmission zeros appear for the strongly coupled asymmetric loop, as shown in Fig. 20 for $L_1 = 12$ nm and $L_2 = 10.5$ nm. Any slight asymmetry leads to decaying quasi-bound states with their poles removed from the real-energy axis. This produces the zero-pole pairs visible in Fig. 20.

The difference between symmetrical and asymmetrical loops can also be understood using wave functions arguments, schematically shown in Fig. 21. In general, the wavefunction in the loop is given by,

$$\psi_1 = A_1 e^{ikx_1} + B_1 e^{-ikx_1} \tag{27}$$

$$\psi_2 = A_2 e^{ikx_1} - B_2 e^{-ikx_1} \tag{28}$$

where ψ_1 and ψ_2 are the wave functions and x_1 and x_2 are the coordinates along the respective arm of the loop; the origin is chosen to be at the left branch point B_L (see inset of Fig. 21). The A's and the B's are the coefficients which will be determined by the particular boundary conditions. Utilizing the same matching conditions as in section IV.A at the branch points B_L and B_R, we obtain,

$$1 + r = A_1 + B_1 \tag{29}$$

$$1 + r = A_2 - B_2 \tag{30}$$

$$1 - r = (A_1 - B_1) + (A_2 - B_2) \tag{31}$$

and

$$t = A_1 e^{ikL_1} + B_1 e^{-ikL_1} \tag{32}$$

$$t = A_2 e^{ikL_2} + B_2 e^{-ikL_2} \tag{33}$$

$$t = (A_1 e^{ikL_1} - B_1 e^{-ikL_1}) + (A_2 e^{ikL_2} - B_2 e^{-ikL_2}) \tag{34}$$

A standing wave is formed around the loop when $k(L_1+L_2) = 2\pi$. For the symmetrical loop, we find from (23) that the transmission amplitude $t = -1$ and hence $r = 0$.

Substituting these values in (29) (31), it is an easy matter to show that $A_1 = A_2 = 3/4$ and $B_1 = B_2 = 1/4$. Thus, the wavefunction for the standing wave in the symmetrical loop is given by,

$$\psi(x) = cos(kx) + \frac{i}{2}sin(kx), \tag{35}$$

where x can be either x_1 or x_2, and its absolute value is plotted in Fig. 21(a). It can be seen that at this incident energy, the incoming amplitude at the branch point B_L is one, and the outgoing amplitude at the other branch point B_R is also one. Hence a standing wave around the symmetrical loop corresponds to perfect transmission, as opposed to the zero predicted by the numerator of (23). For the case of a standing wave in an asymmetrical loop, we find from (23) that $t = 0$ and $r = 1$, which yields,

$$A_1 = \frac{2}{1 - e^{2ikL_1}} \qquad B_1 = \frac{-2e^{2ikL_1}}{1 - e^{2ikL_1}} \tag{36}$$

$$A_2 = \frac{2}{1 - e^{2ikL_2}} \qquad B_2 = \frac{-2e^{2ikL_2}}{1 - e^{2ikL_2}}. \tag{37}$$

The corresponding wave functions in each arm of the loop are given by,

$$\psi_1 = \frac{2sin[k(L_1 - x_1)]}{sin(kL_1)} \tag{38}$$

$$\psi_2 = \frac{2sin[k(L_2 - x_2)]}{sin(kL_2)} \tag{39}$$

Plotting this wave function in Fig. 21(b), we see that in this case the wavefunction is zero at the out-going branch point B_R, but nonzero at the in-coming branch point B_L. Therefore, a standing wave around an symmetrical loop corresponds to a transmission zero, as also predicted by the zero numerator in (23).

The charge accumulated in the strongly-coupled loop is shown in Fig. 22. Part (a) and (b) represent the symmetrical and the asymmetrical cases, respectively, and the same units for the charge are used in both parts. The arrows at the top axis indicate the location of the poles. Note that for the strongly-coupled symmetrical loop in part

(a) the arrows also indicate the positions of transmission maxima, which are due to the existence of true-bound states leading to the cancellation of the zeros and poles on the real-energy axis. Maximum charge accumulation in the asymmetric loop also occurs at the formation of quasi-bound states, as shown in part (b) Note that this resonant charge build up is rather sharp in energy, and that it is more pronounced than for the symmetrical loop.

We have also investigated weakly-coupled loops which are formed by the addition of two small tunneling barriers at the two branch points to the leads. Figures 23 and 24 present the transmission amplitude in the complex-energy plane for the weakly-coupled symmetric and asymmetric loops, respectively, as schematically shown in the insets. Because of the confinement provided by the barriers, the poles move closer to the real-energy axis which corresponds to the increased lifetime of the quasi-bound states. This is true both for the isolated poles, as well as for the zero-pole pairs. Similar to the weakly-coupled t-stubs, the close proximity of the zero-pole pairs leads to a sharp variation of the transmission coefficient on the real-energy axis.

V. CONCLUSION

We studied transmission phenomena in quantum waveguide systems in the presence of resonant cavities. In particular, we investigated the analytical behavior of the transmission amplitude in the complex-energy plane. For the single-moded quantum wires under study here, the dynamics is quasi-one-dimensional. This also allows us to compare directly the much-studied problem of double-barrier resonant tunneling to the lesser understood case of transmission in quantum wire systems; DBRT can be viewed as transmission in a channel which also contains both barriers. One of our main conclusions is that the analytical behavior of the transmission amplitude is different depending upon whether or not the resonant cavity is part of the transmission

channel, as for DBRT, or is attached to the channel, like t-stubs or loops. If the main transmission path is directly through the resonantor, the transmission amplitude only exhibits poles in the complex-energy plane. These poles give rise to the well-known Lorentzian-shaped Breit-Wigner transmission resonances. If in addition to the direct transmission path there is an additional path due to an attached resonator, the transmission amplitude exhibits zero-pole pairs in the complex-energy plane. The vicinity of these zeros and poles produces resonance/antiresonance behavior of the transmission coefficient.

It was noted by several workers in previous studies that the transmission coefficient in quantum waveguide systems exhibited a qualitatively different behavior as compared to the familiar case of DBRT. Our research shows that these observed sharp variations of the transmission coefficient as a function of energy can be understood in terms of the zero-pole pairs in the complex-energy plane. The proximity of the zero and the pole which is produced by each quasi-bound state leads to the sharp energy dependence. Furtheremore, we showed that the existence of transmission zeros for resonantly-coupled waveguides is a consequence of unitarity, regardless of the symmetry of the system. In other words, reflection peaks with amplitude equal to 1 occur for symmetrical as well as for non-symmetrical structures. In contrast, transmission peaks with an amplitude equal to 1 only occur for symmetrical structures. This result is familiar from DBRT, where it is known that perfect transmission only is possible for symmetrical barriers. In related studies, Price [60] distinguishes between peaks in transmission, which he terms resonances of the first kind, and dips in the transmission coefficient, which he terms resonances of the second kind. He also shows that the peak value of the reflection probability, corresponding to the resonances of the second kind, is always 1,regardless of the symmetry of the system, in contrast to the transmission resonance case (resonances of the first kind). This behavior is shown to

persisit in the case of multichannel ballistic transport [61].

Another striking difference between DBRT and the waveguides is the location of the transmission peaks relative to the poles in the complex-energy plane. From DBRT, one is used to associate the location of a transmission maximum with the energy of a quasi-bound state, which are represented by the poles. For transmission in waveguides with attached resonators, the quasi-bound states still are given by the poles in the complex-energy plane, however, their location does *not* match the peaks in the transmission probability. Each quasi-bound state now is represented by a zero-pole pair, where the energy of the pole is close to the energy of the zero. Transmission maxima occur somewhere between the zeros where the exact location of the peaks depends upon the proximity of the poles to the zero and to the real-energy axis. Therefore, it is no longer valid for the quantum waveguide structures, to associate the energy of a transmission peak with the energy of a quasi-bound state.

In summary, our main conclusions are: (i) Transmission zeros exist in quantum waveguide structures with attached resonators, and their existence is a consequence of unitarity; this result is in contrast to DBRT where no transmission zeros are possible. (ii) For the quantum waveguide structures, each quasi-bound state of the attached resonator leads to a zero-pole pair of the transmission amplitude in the complex-energy plane. (iii) In a fashion similar to DBRT, symmetrical waveguide systems possess peaks with perfect transmission. (iv) In contrast to DBRT, the location of the transmission maxima in resonantly-coupled quantum waveguides does not give the energy of the quasi-bound resonantor states.

Acknowledgement: We are grateful for stimulating discussions with Dr. S. Bandyopadhyay, Dr. M. Büttiker, M. Chen, Dr. S. M. Goodnick, H. K. Harbury, M. Leng, Dr. W. Pötz, Dr. P. Price, and Dr. M. Sain. This work was supported, in part, by the Air Force Office of Scientific Research and the Office of Naval Research.

APPENDIX A: NUMERICAL CALCULATIONS

I. MODEL

Utilizing both wavefunction matching and finite element methods, the time-independent Schrödinger Equation is solved to obtain the transmission amplitude in the complex-energy plane for both double-barrier resonant tunneling and the quantum waveguide structures. We assume that an incoming wave with energy E incident upon the system from the left, results in a reflected and transmitted wave, as schematically shown in Fig. 1(a). The wavefunctions in the asymptotic regions on the left and right are given by:

$$\psi_L(x,k) = exp(ikx) + r(k)\ exp(-ikx) \tag{A1}$$

$$\psi_R(x,k) = t(k)\ exp(ikx) \tag{A2}$$

Here $k = \sqrt{2m^*E}/\hbar$ is the complex wavevector variable, and $r(k)$ and $t(k)$ denote the reflection and the transmission amplitudes, respectively. The effective mass m^*=0.067m_0 is used in the calculation, where m_0 is the free-electron mass. In this treatment, we neglect the effects of charge accumulation and inelastic scattering.

The transmission amplitude $t(k)$, or $t(E)$, for an in-coming wave with wavenumber k, or energy E, is then obtained from the out-going wavefunction ψ_R by

$$t(k) = \psi_R(x_0,k)\ e^{-ikx_0}$$

, with x_0 fixed. The transmission amplitude may also be obtained from the Green function. If we denote the out-going Green propagator by $G^+(x,x';k)$, then

$$t(k) = 2ik\ G^+(0,x_0;k)\ e^{-ikx_0}$$

. In a separate section below, we discuss how the Green function readily is obtained in the finite element method.

II. WAVEFUNCTION MATCHING METHOD

This technique is very useful if the problem domain consists of sections in which the potential is constant. One may then write the wavefunction in each section as a superposition of left- and right-going waves with unknown coefficients. The matching conditions for the wavefunction and its derivative at the endpoints of each section lead to a linear system of equations for the unknown expansion coefficients. Below, we show as examples the results of a weakly-coupled t-stub resonator and a strongly-coupled loop.

1. Weakly-Coupled T-Stub

The example t-stub structure is schematically shown in the inset of Fig. 9, where a t-stub is weakly-coupled to the main transmission channel by a tunneling barrier. The stub length is denoted by L, and a tunneling barrier of length l and height V_0 is located at the branch point. Choosing the branch point as the coordinate origin, the wavefunctions in each region are given by,

$$\psi_L = e^{ikx_L} + re^{-ikx_L} (x_L < 0), \tag{A3}$$

$$\psi_R = te^{ikx_R} (x_R > 0), \tag{A4}$$

$$\psi_B = A_B e^{iKx_S} + B_B e^{-iKx_S} (0 < x_S < l), \tag{A5}$$

$$\psi_S = A_S e^{ikx_S} + B_S e^{-ikx_S} (l < x_S < L). \tag{A6}$$

We denote the wavenumber of the propagating waves by $k = \sqrt{2m^* E}/\hbar$, and of the decaying waves by $K = \sqrt{2m^*(E - V_0)}/\hbar$, which is for the wavefunction in the barrier, ψ_B. The spatial coordinates x_L, x_R, and x_S represent the left, right, and stub branches, respectively. Using the matching conditions for the branch point, which are discussed in section IV. A, we obtain,

$$
\begin{aligned}
1 + r &= t \\
1 + r &= A_B + B_B \\
1 - r &= (K/k)(A_B - B_B) + t \\
A_S e^{ikl} + B_S e^{-ikl} &= A_B e^{iKl} + B_B e^{-iKl} \\
A_S e^{ikl} - B_S e^{-ikl} &= (K/k)(A_B e^{iKl} - B_B e^{-iKl}) \\
A_S e^{ikL} + B_S e^{-ikL} &= 0
\end{aligned}
\tag{A7}
$$

The above system of equations is sufficiently simple that it can be solved analytically, and the result for the transmission amplitude t is given by eq. (22) in IV. B. In general, the resulting linear system of quations has to be solved numerically.

2. Strongly-Coupled Loop

We also present as an example the so-called strongly-coupled loop structure which is schematically shown in the inset of Fig. 21(b). We choose the left branch point B_L as the coordinate origin for the left branch and the two arms of the loop, and the right branch point B_R as the origin for the right branch. Then the wavefunctions in each region can be written as,

$$\psi_L = e^{ikx_L} + re^{-ikx_L} (x_L < 0) \tag{A8}$$

$$\psi_R = te^{ikx_R} (x_R > 0) \tag{A9}$$

$$\psi_1 = A_1 e^{ikx_1} + B_1 e^{-ikx_1} (0 < x_1 < L_1) \tag{A10}$$

$$\psi_2 = A_2 e^{ikx_2} + B_2 e^{-ikx_2} (0 < x_2 < L_2) \tag{A11}$$

Here, x_L and x_R are the coordinates of the left and the right branch, and x_1 and x_2 are the coordinates of the respective arms of the loop with length L_1 and L_2, respectively. Using the matching conditions discussed in section IV. A for each branch point, we obtain,

$$
\begin{aligned}
1 + r &= A_1 - B_1 \\
1 + r &= A_2 - B_2 \\
1 - r &= (A_1 - B_1) + (A_2 - B_2) \\
t &= A_1 e^{ikL_1} + B_1 e^{-ikL_1} \\
t &= A_2 e^{ikL_2} + B_2 e^{-ikL_2} \\
t &= (A_1 e^{ikL_1} - B_1 e^{-ikL_1}) + (A_2 e^{ikL_2} - B_2 e^{-ikL_2})
\end{aligned}
\tag{A12}
$$

These particluar equations can be solved analytically, and the solution for the transmission amplitude is given by eq. (23) in section IV. C. Again, for more complicated loop structures the resulting linear system of equations has to be solved numerically.

III. FINITE ELEMENT METHOD

For systems with complicated potential variations, it is convenient to discretize the problem domain and to evaluate the wavefunction at a set of nodal points. Between nodal points, in each so-called finite element, the solution is approximated by linear basis functions. We typically divide the structure into 60 elements, and the energy mesh is 300×50 for quantum waveguide systems and 200×30 for double-barrier resonant tunneling. In the vicinity of some of the very sharp zero-pole pairs, the real-energy axis is further refined, typically by some 6000 nodal points.

1. Formalism

For the problem domain $[0, a]$, which contains the potential V, the one-dimensional Schrödinger equation,

$$
-\frac{\hbar^2}{2m^*}\psi'' + V\psi = E\psi, \tag{A13}
$$

can be written as,

$$
\psi'' + K^2\psi = 0, \quad K^2 = 2m^*(E - V)/\hbar^2 \quad (0 < x < a) \tag{A14}
$$

$$\psi'' - k^2\psi = 0, \qquad k^2 = 2m^*E/\hbar^2 \qquad (x < 0, x > a) \tag{A15}$$

The regions outside the problem domain are assumed to be potential free. For an incoming wave from the left, which gives rise to a reflected and a transmitted component (compare Fig. 1(a)), the boundary conditions at the edges of the problem domain are,

$$\frac{\psi(0)}{\psi'(0)} = \frac{1}{ik}\frac{1+r}{1-r} \tag{A16}$$

$$\frac{\psi(a)}{\psi'(a)} = \frac{1}{ik} \tag{A17}$$

Using a test function ϕ, the weak variational form of (A14) is,

$$\int_0^a (\psi''\phi - k^2\psi\phi)dx = 0. \tag{A18}$$

Integration by parts yields,

$$\psi'(a)\phi(a) - \psi'(0)\phi(0) - \int_0^a \psi'\phi'dx + \int_0^a k^2\psi\phi dx = 0. \tag{A19}$$

We now discretize the solution domain, and define n nodal point coordinates x_i, where x_1 corresponds to $x = 0$, and x_n to $x = a$. Furthermore, we expand the wavefunction ψ and the test function ϕ in terms of their values at the nodal points ψ_i and ϕ_i, respectively, and in terms of shape functions U_i,

$$\psi = \sum_{i=1}^{n} \psi_i U_i, \tag{A20}$$

$$\phi = \sum_{i=1}^{n} \phi_i U_i. \tag{A21}$$

Using the property of the shape functions that $U_j(x_i) = \delta_{ij}$, we obtain at the boundary of the solution domain,

$$\psi(a) = \psi_n, \qquad \phi(a) = \phi_n, \tag{A22}$$

$$\psi(0) = \psi_1, \qquad \phi(0) = \phi_1. \tag{A23}$$

With (A16)-(A17) and (A22)-(A23), (A19) can now be written as,

$$ik\psi_n\phi_n - ik[2 - \psi_1]\phi_1 - \int_0^a \psi'\phi' dx + \int_0^a K^2\psi\phi dx = 0. \tag{A24}$$

After substituting (A20)-(A21) and some rearranging, (A24) becomes,

$$2ik\phi_1 = ik\psi_n\phi_n + ik\psi_1\phi_1 - \sum_{j=1}^{n}\phi_j\left[\sum_{i=1}^{n}\int_0^a U_i'U_j' dx w_i - \sum_{i=1}^{n}\int_0^a K^2U_iU_j dx w_i\right] \tag{A25}$$

which can be written as,

$$\begin{aligned} &-2ik\sum_{j=1}^{n}\phi_j\delta_{1j} \\ &= \sum_{j=1}^{n}\phi_j\left[\sum_{i=1}^{n}\int_0^a U_i'U_j' dx w_i\right] - \sum_{j=1}^{n}\phi_j\left[\sum_{i=1}^{n}\int_0^a K^2U_iU_j dx w_i + ik\psi_n\delta_{nj} + ik\psi_1\delta_{1j}\right]. \end{aligned} \tag{A26}$$

Since ϕ is an arbitrary function, the above equation must hold individually for each term in the sum. Remembering that $K^2 = k^2 - 2m^*V/\hbar^2$, (A26) can be written in matrix form as,

$$[C - k^2Q - ikB]\psi = P. \tag{A27}$$

This is an equation for the wavefunction contained in the vector ψ. The vector P represents the forcing term due to the in-coming wave, and C, Q, B are $n \times n$ matrices, given by

$$C_{ij} = \int_0^a U_i'U_j' dx + (2m^*/\hbar^2)\int_0^a VU_iU_j dx \tag{A28}$$

$$Q_{ij} = \int_0^a U_iU_j dx \tag{A29}$$

$$B_{ij} = \delta_{nj}\delta_{ni} + \delta_{1j}\delta_{1i} \tag{A30}$$

$$P_j = -2ik\delta_{1j} \tag{A31}$$

Finally, we define the matrix A as,

$$A = C - k^2Q - ikB. \tag{A32}$$

and (A27) becomes,

$$A\psi = P. \tag{A33}$$

This is the finite-element discretized form of the Schrödinger equation. We can now find the wavefunction of the system by solving the above linear system of equations, and we can then determine from the solution vector ψ the transmission amplitude.

2. Matching Condition at the Branch Point, $\sum_{i=1}^{n} \psi' = 0$

The matching condition $\sum_{i=1}^{n} \psi' = 0$ at the branch point may be included in a natural way in the finite element method. The Schrödinger equation in each branch reads,

$$-\hbar^2/2m^* \psi_i'' + V\psi_i = E\psi_i, \tag{A34}$$

where i=1,2,3 $\cdots$, n labels the various branches. Applying the weak variational form of the finite element method yields,

$$\psi_i'(a_i)\phi(a_i) - \psi_i'(0)\phi(0) - \int_0^{a_i} \psi_i'\phi' dx + \int_0^{a_i} K^2\psi_i\phi dx = 0. \tag{A35}$$

Here, the coordinate at the branch point is chosen to be 0 and the other edge of the domain in the i^{th} branch is denoted by a_i. When adding the above equations for all branches, we obtain a term,

$$\sum_{i}^{n} \psi_i'(0)\phi(0). \tag{A36}$$

Since both the wavefunction and the test function are continuous, our matching condition at the branch point make this term vanish,

$$\phi(0)\sum_{i}^{n} \psi_i'(0) = 0. \tag{A37}$$

Therefore, the internal boundaries at the branch point do not introduce additional terms in the finite-element matrices, and only the external boundaries of the solution domain enter in the final matrices A and P.

3. Green Function $G = -A^{-1}$

Here we prove that the Green function G is given by $G = -A^{-1}$, where A is the coefficient matrix defined in (A32). The Green function is defined by,

$$G''(x, x', E) + K^2 G(x, x', E) = \delta(x - x'). \tag{A38}$$

As for the finite element method, we now integrate over the solution domain $[0, a]$,

$$\int_0^a \left[G''(x, x', E) + K^2 G(x, x', E) - \delta(x - x')\right] V(x, E) dx = 0 \tag{A39}$$

where $V(x, E)$ is a test function. Integration bt parts yields,

$$\int_0^a \left[G'(x, x', E)V'(x, E) - K^2 G(x, x', E)V(x, E)\right] dx \tag{A40}$$
$$+V(x', E) - G'(a, x', E)V(a, E) + G'(0, x', E)V(0, E) = 0.$$

Discretization of the solution domain introduces the values of the Green function G_{ij} at the nodal points x_i and x_j,

$$G(x, x', E) = \sum_{i=1}^{n} \sum_{j=1}^{n} G_{ij}(E) U_i(x) U_j(x'), \tag{A41}$$

$$V(x, E) = \sum_{l=1}^{n} V_l(E) U_l(x), \tag{A42}$$

where U_i is the same shape function as in (A20) and (A21). This yields,

$$\int_0^a G'(x, x', E)V'(x, E)dx = \sum_{ijl}^{n} V_l(E) U_j(x') Z_{li} G_{ij}, \tag{A43}$$

$$\int_0^a G(x, x', E)V(x, E)dx = \sum_{ijl}^{n} V_l(E) U_j(x') Q_{li} G_{ij}, \tag{A44}$$

with,

$$Z_{li} = \int_0^a U_l' U_i' dx, \tag{A45}$$

$$Q_{li} = \int_0^a U_l U_i dx. \tag{A46}$$

Note that Q is the same matrix as defined in (A29). For the boundaries at the edge of the solution domain, we use our knowledge of the Green function for free propagation,

$$G(x, x', E) = \frac{e^{ik|x-x'|}}{2ik} \qquad (x < 0, x > a). \tag{A47}$$

At the boundaries $x = 0$ and $x = a$, we then have the following conditions,

$$G'(a, x', E) = ikG(a, x', E), \tag{A48}$$

$$G'(0, x', E) = -ikG(0, x', E). \tag{A49}$$

Using (A37)-(A38) and the equality $U_i(a) = U_i(0) = 1$, which is a property of the shape functions, we obtain

$$G'(a, x', E)V(a, E) = \sum_{ij}^{n} V_i(E)U_j(x')[ik\delta_{in}\delta_{ln}G_{ij}], \tag{A50}$$

$$G'(0, x', E)V(0, E) = -\sum_{ij}^{n} V_i(E)U_j(x')[ik\delta_{i1}\delta_{l1}G_{ij}]. \tag{A51}$$

We now define as in (A30),

$$B_{li} = \delta_{in}\delta_{ln} + \delta_{i1}\delta_{l1}, \tag{A52}$$

and after substituting (A46)-(A47) and (A38)-(A40) into (A36), we finally obtain

$$\sum_{ji}^{n} V_i(E)U_j(x')\left[\sum_{i}^{n}(C_{li} - k^2Q_{li} - ikB_{li})G_{ij} + \delta_{lj}\right] = 0, \tag{A53}$$

where $C_{li} = Z_{li} + (2m^*/\hbar^2)\int_0^a VU_iU_j\,dx$ which is the same matrix as in (A28). We may now define $A_{li} = (C_{li} - k^2Q_{li} - ikB_{li})$ which is an element of the same matrix as in (A32). Since V is an arbitrary test function and the shape functions U are linearly independent, the Green function G and A are related by

$$AG = -I. \tag{A54}$$

Q.E.D.

APPENDIX B: UNITARITY AND ZEROS

1. Proof of the Existence of Transmission Zeros

In this section, we prove that, for all energies,

$$\left| r_S - \frac{t_{RS} t_{SL}}{t_{RL}} \right| = 1, \tag{B1}$$

which ensures the existence of transmission zeros.

Let $\alpha = r_S - t_{RS} t_{SL} / t_{RL}$. We will prove (B1) by demonstrating that $\alpha\alpha^* = 1$.

$$\alpha\alpha^* = r_S r_S^* + \frac{t_{RS} t_{RS}^* t_{SL} t_{SL}^*}{t_{RL} t_{RL}^*} - \frac{r_S^* t_{RS} t_{SL} t_{RL}^*}{t_{RL} t_{RL}^*} - \frac{r_S t_{RS}^* t_{SL}^* t_{RL}}{t_{RL} t_{RL}^*} \tag{B2}$$

The unitarity of the scattering matrix requires:

$$r_L r_L^* + t_{RL} t_{RL}^* + t_{SL} t_{SL}^* = 1, \tag{B3}$$

$$t_{RL} t_{RL}^* + r_R r_R^* + t_{RS} t_{RS}^* = 1, \tag{B4}$$

$$t_{SL} t_{SL}^* + t_{RS} t_{RS}^* + r_S r_S^* = 1, \tag{B5}$$

, and,

$$r_L t_{RL}^* + t_{RL} r_R^* + t_{SL} t_{RS}^* = 0, \tag{B6}$$

$$r_L t_{SL}^* + t_{RL} t_{RS}^* + t_{SL} r_S^* = 0, \tag{B7}$$

$$t_{RL} t_{SL}^* + r_R t_{RS}^* + t_{RS} r_S^* = 0, \tag{B8}$$

We now substitute (B7), $r_S^* t_{SL} = -(r_L t_{SL}^* + t_{RL} t_{RS}^*)$, into the third term of (B2) and the complex conjugate of (B8), $r_S t_{RS}^* = -(t_{RL}^* t_{SL} + r_R^* t_{RS})$, into the fourth term We define β as the sum of these two terms, which is given by,

$$\begin{aligned}
\beta &= -\frac{r_S^* t_{SL} t_{RL}^* t_{RS}}{t_{RL} t_{RL}^*} - \frac{r_S t_{RS}^* t_{RL} t_{SL}^*}{t_{RL} t_{RL}^*} \\
&= \frac{r_L t_{SL}^* t_{RL}^* t_{RS} + t_{RL} t_{RS}^* t_{RL}^* t_{RS}}{t_{RL} t_{RL}^*} + \frac{t_{RL}^* t_{SL} t_{RL} t_{SL}^* + r_R^* t_{RS} t_{RL} t_{SL}^*}{t_{RL} t_{RL}^*} \\
&= t_{RS}^* t_{RS} + t_{SL}^* t_{SL} + \frac{t_{SL}^* t_{RS} (r_L t_{RL}^* + r_R^* t_{RL})}{t_{RL} t_{RL}^*}.
\end{aligned} \tag{B9}$$

Substituting it back into (B2) and using (B5)-(B6) gives:

$$\begin{aligned}\alpha\alpha^* &= r_S r_S^* + t_{RS}^* t_{RS} + t_{SL}^* t_{SL} + \frac{t_{RS} t_{RS}^* t_{SL} t_{SL}^*}{t_{RL} t_{RL}^*} + \frac{t_{SL}^* t_{RS}(r_L t_{RL}^* - r_R^* t_{RL})}{t_{RL} t_{RL}^*} \\ &= 1 + \frac{t_{RS} t_{RS}^* t_{SL} t_{SL}^*}{t_{RL} t_{RL}^*} + \frac{t_{SL}^* t_{RS}(-t_{SL} t_{RS}^*)}{t_{RL} t_{RL}^*} \\ &= 1 \end{aligned} \tag{B10}$$

Q.E.D.

2. Proof of the Existence of Transmission Ones in Symmetrical Structures

In this section, we prove that, for a symmetrical structure,

$$\left| r_S - \frac{t_{SL} t_{SL}}{r_L} \right| = 1 \tag{B11}$$

which ensures the existence of transmission ones.

Instead of proving $\mathcal{T}_{RL} = 1$ directly, we prove $\mathcal{R}_L = 0$, which implies

$$\lambda = r_S - \frac{t_{SL} t_{SL}}{r_L}. \tag{B12}$$

If the modulus of the r.h.s. is one, which is condition (B11), zeros in reflection, and thus ones in transmission, must exist.

Let $\alpha = r_S - t_{SL} t_{SL}/r_L$. Condition (B11) then requires $|\alpha| = 1$, which we will prove by demonstrating that $\alpha\alpha^* = 1$.

$$\alpha\alpha^* = r_S r_S^* + \frac{|t_{SL}|^4 - r_S^* t_{SL}^2 r_L^* - r_S (t_{SL}^2)^* r_L}{|r_L|^2} \tag{B13}$$

We now substitute (B7), $r_S^* t_{SL} = -(r_L t_{SL}^* + t_{RL} t_{RS}^*)$, into the third term of (B13) and (B7), $r_L t_{SL}^* = -(t_{RL} t_{RS}^* + r_S^* t_{SL})$, into the fourth term. We define β as the sum of these two terms, which is given by,

$$\begin{aligned}\beta &= \frac{-r_S^* t_{SL}^2 r_L^* - r_S (t_{SL}^2)^* r_L}{|r_L|^2} \\ &= \frac{|t_{SL}|^2 |r_L|^2 + |t_{SL}|^2 |r_S|^2 + t_{RL} t_{RS}^* (t_{SL} r_L^* + r_S t_{SL}^*)}{|r_L|^2} \end{aligned} \tag{B14}$$

Substituting it back into (B13) and using (B7) gives,

$$\alpha\alpha^* = |r_S|^2 + |t_{SL}|^2 + |t_{RS}|^2 + \frac{|t_{SL}|^4 + |r_S|^2|t_{SL}|^2 - |r_L|^2|t_{RS}|^2 - |t_{RL}|^2|t_{RS}|^2}{|r_L|^2} \quad \text{(B15)}$$

$$= 1 + \frac{|t_{SL}|^4 + |r_S|^2|t_{SL}|^2 - |r_L|^2|t_{RS}|^2 - |t_{RL}|^2|t_{RS}|^2}{|r_L|^2}$$

Using (B5) in the first two terms of the above numerator, and (B3) in the last two terms, yields

$$\alpha\alpha^* = 1 + \frac{|t_{SL}|^2(1-|t_{RS}|^2) - |t_{RS}|^2(1-|t_{SL}|^2)}{|r_L|^2} \quad \text{(B16)}$$

$$= 1 + \frac{(|t_{SL}|^2 - |t_{RS}|^2)}{|r_L|^2}$$

For symmetrical systems, $|t_{SL}|^2 = |t_{RS}|^2$, which proves (B11). The symmetrical structures shown in Figs. 4-7 possess transmission ones.

For the non-symmetrical structures shown in Figs. 9-10, $|t_{SL}|^2 \neq |t_{RS}|^2$, and no transmission ones exist. Condition (B11) is not satisfied in this case.

APPENDIX C: SCATTERING MATRIX

Here, we show that our result for the scattering matrix in thin wire networks, which we derived in section IV.A, is a special case of previous work in which the coupling of a ring to a lead was studied [38, 41, 12]. It was shown in Ref. [38] that the scattering matrix is determined by three parameters in the form,

$$S = \begin{pmatrix} -(a+b) & \epsilon^{1/2} & \epsilon^{1/2} \\ \epsilon^{1/2} & a & b \\ \epsilon^{1/2} & b & a \end{pmatrix} \quad \text{(C1)}$$

Unitarity imposes the following constraints.

$$|a+b|^2 + 2\epsilon = 1, \quad \text{(C2)}$$

$$a^2 + b^2 + \epsilon = 1. \quad \text{(C3)}$$

The parameter ϵ, which ranges between $0 < \epsilon < 1/2$, measures the strength of the coupling between the lead and the ring. A completely detached ring corresponds to $\epsilon = 0$, and maximum coupling is assumed to occur for $\epsilon = 1/2$.

However, we noted in previous work [12] that the strongest coupling occurs for $\epsilon = 4/9$ (and not for $\epsilon = 1/2$). According to (C2) and (C3), $\epsilon = 4/9$ corresponds to the following values for the other two parameters, $a = -1/3$ and $b = 2/3$. The resulting scattering matrix for these values of a and b is exactly the one derived by us in section IV.A, i.e. eqn. (20). We conclude that our choice of the matching conditions at the branch point leads to a scattering matrix which is a special case of the general form (C1) for the parameter value $\epsilon = 4/9$.

REFERENCES

[1] R. Landauer, IBM J. Res. Develop. 1, 223 (1957); R. Landauer, Philos. Mag. **21**, 863 (1970).

[2] M. Büttiker, Phys. Rev. Lett. **57**, 1761 (1986).

[3] R. A. Webb, Oliver E. Buckley Price Lecture, March Meeting of the American Physical Society, Indianapolis, March 1992.

[4] *Analogies in Optics and Micro Electronics*, ed. by W. VanHaeringen and D. Lenstra (Kluwer Academic Publishers, Dordrecht, 1990).

[5] L. Esaki and R. Tsu, IBM J. Res. Develop. **14**, 61 (1970); R. Tsu and L. Esaki, Appl. Phys. Lett. **22**, 562 (1973).

[6] T. C. L. G. Sollner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker, and D. D. Peck, Appl. Phys. Lett. **43**, 588 (1983); T. C. L. G. Sollner, P. E. Tannenwald, D. D. Peck, and W. D. Goodhue, Appl. Phys. Lett. **45**, 1319 (1984).

[7] *Nanostructure Physics and Fabrication*, ed. by M. Reed and W.P. Kirk (Academic Press, Boston, 1989).

[8] W. Porod, Z. Shao, and C. S. Lent, Appl. Phys. Lett. **61**, 1350 (1992).

[9] P. F. Bagwell and R. K. Lake, Phys. Rev. B. **46**, 15 329 (1992); P. F. Bagwell, A Kumar, and R. Lake, in *Quantum Effect Physics, Electronics, and Applications*, ed. by K. Ismail (Adam Hilger, Bristol, 1992);

[10] F. Sols, M. Macucci, U. Ravaioli, and K. Hess, Appl. Phys. Lett. **54**, 350 (1989); and F. Sols, M. Macucci, U. Ravaioli, and K. Hess, J. Appl. Phys. **66**, 3892

(1989).

[11] S. Datta, Superlatt. Microstruct. 6, 83 (1989).

[12] S. Subramaniam, S. Bandyopadhyay, and W. Porod, J. Appl. Phys. **68**, 4861 (1990).

[13] K.-F. Berggren and Z.-L. Ji, Superlattices and Microstructures **8**, 59 (1990).

[14] Y. Avishai and Y. B. Band, Phys. Rev. B **41**, 3253 (1990).

[15] Y. Takagaki and D. K. Ferry, J. Appl. Phys. **72**, 5001 (1992).

[16] J.-B. Xia, Phys. Rev. B **45**, 3593 (1992).

[17] L. P. Kouwenhoven, F. W. J. Hekking, B. J. van Wees, C. J. P. M. Harmans, C. E. Timmering, and C. T. Foxon, Phys. Rev. Lett. **65**, 361 (1990).

[18] H. Wu, D. W. L. Sprung, J. Martorell, and S. Klarsfeld, Phys. Rev. B **44**, 6351 (1991).

[19] J. A. Brum, Phys. Rev. B **43**, 12082 (1991).

[20] Z.-L. Ji and K.-F. Berggren, Phys. Rev. B **45**, 6652 (1992).

[21] C. S. Lent, Computational Electronics: Semiconductor Transport and Device Simulation, ed. by K. Hess, J. P. Leburton, and U. Ravaioli, p. 259 (1991).

[22] P. J. Price, IEEE Trans. Electron. Dev. **39**, 520 (1992).

[23] C. S. Chu and R. S. Sorbello, Phys. Rev. B **40**, 5941 (1989); Z. Chen and R. S. Sorbello, Phys. Rev. B **44**, 12 857 (1991).

[24] P. F. Bagwell, Phys. Rev. B **41**, 10 354 (1990); J. Phys.: Condens. Matter **2**, 6179 (1990).

[25] E. Tekman and S. Ciraci, Phys. Rev. B **42**, 9098 (1990).

[26] Y. B. Levison, M. I. Lubin, E. V. Sukhorukov, Phys. Rev. B **45**, 11 936 (1992).

[27] Y. Takagaki and D. K. Ferry, Phys. Rev. B **45**, 6715 (1992); Phys. Rev. B **46**, 15 218 (1992).

[28] J. M. Mao, Y. Huang, and J. M. Zhou, J. Appl. Phys. **73**, 1853 (1993).

[29] P. Exner and P. Seba, J. Math. Phys. **30**, 2574 (1989); Y. Avishai, D. Bessis, B. G. Giraud, and G. Mantica, Phys. Rev. B **44**, 8028 (1991).

[30] C. S. Lent, Appl. Phys. Lett. **56**, 2554 (1990).

[31] Sois BENDS .....

[32] A. Weisshaar, J. Lary, S. M. Goodnick, and V. K. Tripathi, Appl. Phys. Lett. **55**, 2114 (1989).

[33] A. Weisshaar, J. Lary, S. M. Goodnick, and V. K. Tripathi, J. Appl. Phys. **70**, 355 (1991).

[34] R. L. Schult, D. G. Ravenhall, and H. W. Wyld, Phys. Rev. B **39**, 5476 (1988).

[35] C. S. Lent, Appl. Phys. Lett. **57**, 1678 (1990).

[36] K.-F. Berggren and Z.-L. Ji, Phys. Rev. B **43**, 4760 (1991).

[37] Y. Takagaki and D. K. Ferry, Phys. Rev. B. **44**, 8399 (1991).

[38] M. Büttiker, Y. Imry, and M. Ya. Azbel, Phys. Rev. A **30**, 1982 (1984); M. Büttiker, Phys. Rev. B. **32**, 1846 (1985).

[39] M. Cahay, S. Bandyopadhyay, and H. L. Grubin, Phys. Rev. B **39**, 12989 (1989).

[40] D. Kowal, U. Sivan, O. Entin-Wohlman, and Y. Imry, Phys. Rev. B **42**, 9009

(1990).

[41] C. H. Wu and G. Mahler, Phys. Rev. B **43**, 5012 (1991).

[42] D. Y. K. Ko and J. C. Inkson, Phys. Rev. B **38**, 9945 (1988); Semicond. Sci. Technol. **3**, 791 (1988).

[43] T. B. Boykin, B. Pezeshki, and J. S. Harris, Jr., Phys. Rev. B **46**, 12 769 (1992).

[44] D. Z.-Y. Ting and T. C. McGill, Phys. Rev. B **47**, 7281 (1993).

[45] R. Wessel and M. Altarelli, Phys. Rev. B **39**, 12 802 (1989).

[46] E. Tekman and P. F. Bagwell, preprint.

[47] U. Fano, Phys. Rev. **124**, 1866 (1961).

[48] P. J. Price, Phys. Rev. B **38**, 1994 (1988).

[49] M. Büttiker, IBM J. Res. Develop. **32**, 63 (1988).

[50] T. B. Bahder, C. A. Morrison and J. D. Bruno, Appl. Phys. Lett. **51**, 1089 (1987); J. D. Bruno, T. B. Bahder, and C. A. Morrison, Phys. Rev. B **37**, 7098 (1988).

[51] G. García-Calderón, A. Rubio, and R. Romo, J. Appl. Phys., **69**, 3612 (1991); G. García-Calderón and A. Rubio, Phys. Rev. B **36**, 4462 (1987); G. García-Calderón, Solid State Commun. **62**, 441 (1987).

[52] G. Breit and E. Wigner, Phys. Rev. **49**, 519 (1936).

[53] G. Gamow, Z. Phys. **51**, 204 (1928).

[54] S. Datta, *Quantum Phenomena*, Addison-Wesley, (1989).

[55] P. W. Anderson, D. J. Thouless, E. Abrahams, and D. S. Fisher, Phys. Rev. B

22, 3519 (1980).

[56] D. S. Fisher and P. A. Lee, Phys. Rev. B **23**, 6851 (1981).

[57] It appears that only in pathological cases, when both phases "chase" each other, there is no transmission zero. In such cases, a zero and pole coincide and cancel each other. Minute changes in the geometry will cause the zero-pole pair to separate.

[58] J. E. Avron, A. Raveh, and B. Zur, Rev. Mod. Phys. **60**, 873 (1988).

[59] B. Shapiro, Phys. Rev. Lett. **50**, 747 (1983).

[60] P. J. Price, Appl. Phys. Lett. **62**, 289 (1993).

[61] P. J. Price, "Transmission and reflection peaks in multichannel ballistic transport," preprint.

FIGURE CAPTIONS

Figure 1. Schematic diagram of resonator structures coupled to two leads; (a) shows an incident wave from the left with its transmitted and reflected components; (b) presents typical resonator structures, such as double barriers, t-stubs, and loops; the shaded boxes on the waveguides represent potential barriers.

Figure 2. Transmission coefficient for a double-barrier resonant tunneling structure which is schematically depicted in the inset ($V_0 = 0.2eV, V_w = 0.1eV, L = 30nm, and b = 5nm$); (a) shows the transmission probablility on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

Figure 3. Phase of the transmission amplitude for double-barrier resonant tunneling for the same structure as in Fig. 2: (a) shows the phase of the transmission amplitude on the real-energy axis, and (b) shows a contour plot of the phase in the complex-energy plane.

Figure 4. Charge accumulated in the well region for double-barrier resonant tunneling for the same structure as in Fig. 2.

Figure 5. Schematic drawing of a waveguide with a resonantly-coupled cavity: (a) shows a wire branch with incoming and outgoing waves outside the junction region, which is indicated by the dashed box; (b) shows a resonant t-stub which is obtained by closing-off the side arm.

Figure 6. Schematic representation of the condition for the existence of transmission zeros, which occur when both phasors coincide on the unit circle.

Figure 7. Transmission coefficient for the strongly-coupled t-stub structure which is schematically depicted in the inset: (a) shows the transmission probability on the

real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

Figure 8. Phase of the transmission amplitude for the strongly-coupled t-stub (same parameters as in Fig. 7); (a) shows the phase of the transmission amplitude on the real-energy axis, and (b) shows a contour plot of the phase in the complex-energy plane. Note the existence of zero-pole pairs.

Figure 9. Transmission coefficient for the weakly-coupled t-stub structure which is schematically depicted in the inset; (a) shows the transmission probability on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

Figure 10. Wave functions in the stubs corresponding to the lowest transmission zero (solid line) and transmission one (dotted line) for the t-stub structures shown in the insets; (a) strongly-coupled t-stub (same parameters as in Fig. 7), and (b) weakly-coupled t-stub (same parameters as in Fig. 9).

Figure 11. Wave functions in the stubs corresponding to the second lowest transmission zero (solid line) and transmission one (dotted line) for the t-stub structures shown in the insets; (a) strongly-coupled t-stub (same parameters as in Fig. 7), and (b) weakly-coupled t-stub (same parameters as in Fig. 9).

Figure 12. Charge accumulated in the stubs for the structures shown in the insets; (a) strongly-coupled t-stub (same parameters as in Fig. 7), and (b) weakly-coupled t-stub (same parameters as in Fig. 9); the arrows indicate the position of the poles which correspond to the quasi-bound states.

Figure 13. Transmission coefficient for the weakly-coupled t-stub structure which is schematically depicted in the inset; (a) shows the transmission probability on the

real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

Figure 14. Transmission coefficient of the weakly-coupled t-stub structure which is schematically depicted in the inset: (a) shows the transmission probability on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

Figure 15. Transmission coefficient for the weakly-coupled t-stubs shown in Fig. 14 for different separations between the two tunneling barriers on the transmission channel, which are $4, 6, 8nm$ for (a), (b), and (c), respectively; the left column shows the transmission probability on the real-energy axis, and the right column shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane; part (a) is for the same parameters as in Fig. 14.

Figure 16. Phase of the transmission amplitude of the weakly-coupled t-stub for the same parameters as in Fig. 15(b): (a) shows the phase change of transmission amplitude on the real-energy axis, and part (b) shows a contour plot of the phase in the complex-energy plane.

Figure 17. Transmission coefficient for the asymmetrical t-stub structure which is schematically depicted in the inset: (a) shows the transmission probability on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

Figure 18. Transmission coefficient for the asymmetrical t-stub structure which is schematically depicted in the inset: (a) shows the transmission probability on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

Figure 19. Transmission coefficient for the strongly-coupled symmetrical loop

which is schematically depicted in the inset: (a) shows the transmission probability on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

Figure 20. Transmission coefficient for the strongly-coupled asymmetrical loop which is schematically depicted in the inset: (a) shows the transmission probability on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

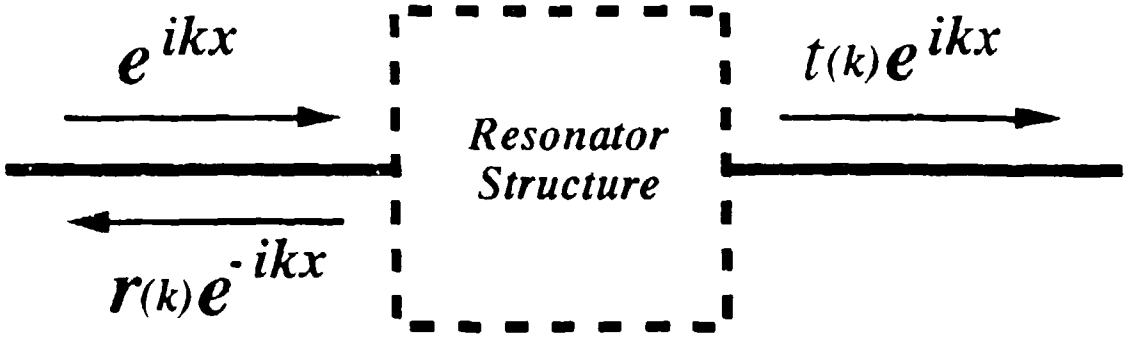
Figure 21. Wave functions for the strongly-coupled loop structures shown in Figs. 19 and 20; B_L and B_R denote the branch points at the left and right, respectively; (a) wave function for the lowest bound state in the symmetrical loop which results in a transmission one; (b) wave function for the lowest quasi-bound state in the asymmetrical loop which results in a transmission zero.

Figure 22. Charge accumulated in the strongly-coupled loops shown in Figs. 19 and 20: (a) symmetrical loop, and (b) asymmetrical loop. The arrows indicate the position of the corresponding poles.

Figure 23. Transmission coefficient for the weakly-coupled symmetrical loop which is schematically depicted in the inset; (a) shows the transmission probability on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

Figure 24. Transmission coefficient for the weakly-coupled asymmetrical loop which is schematically depicted in the inset: (a) shows the transmission probability on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

(a)

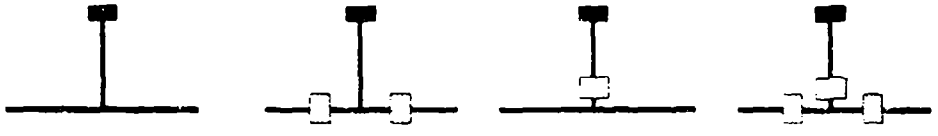


(b)

Resonator Structures

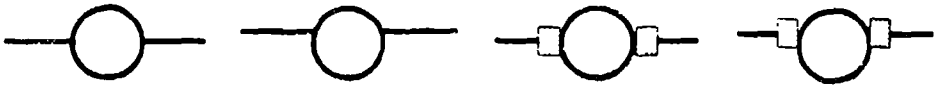
Double Barrier Resonant Tunneling

T-Stubs



Strongly-Coupled Stub Weakly-Coupled Stubs

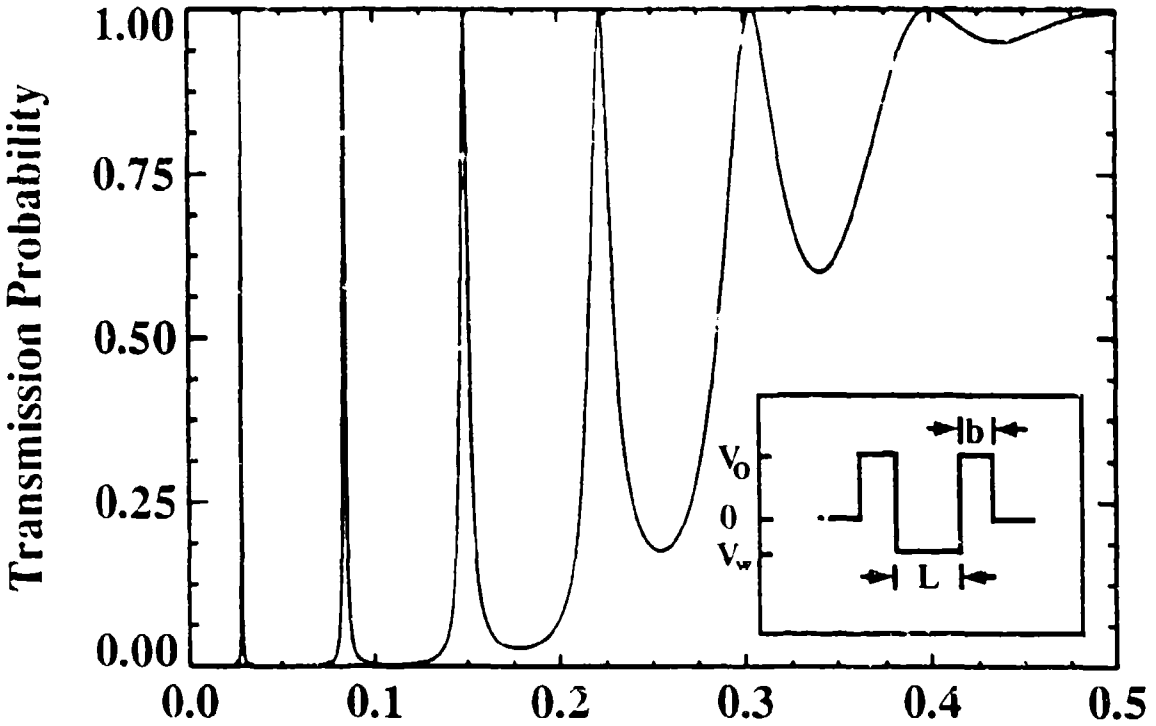
Loops



Strongly-Coupled Loops Weakly-Coupled Loops

Fig 1

(a)


Transmission Probability
1.00
0.75
0.50
0.25
0.00
0.0
0.1
0.2
0.3
0.4
0.5
b
V_0
0
V_w
L

(b)

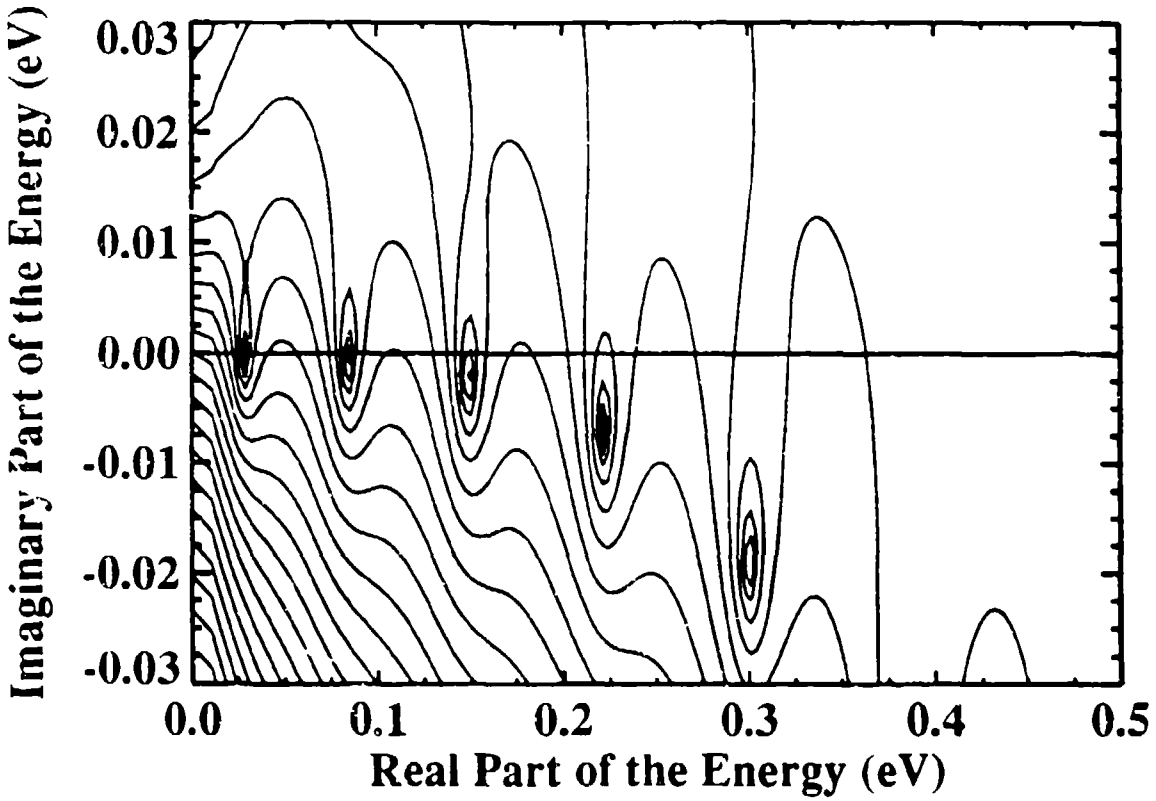
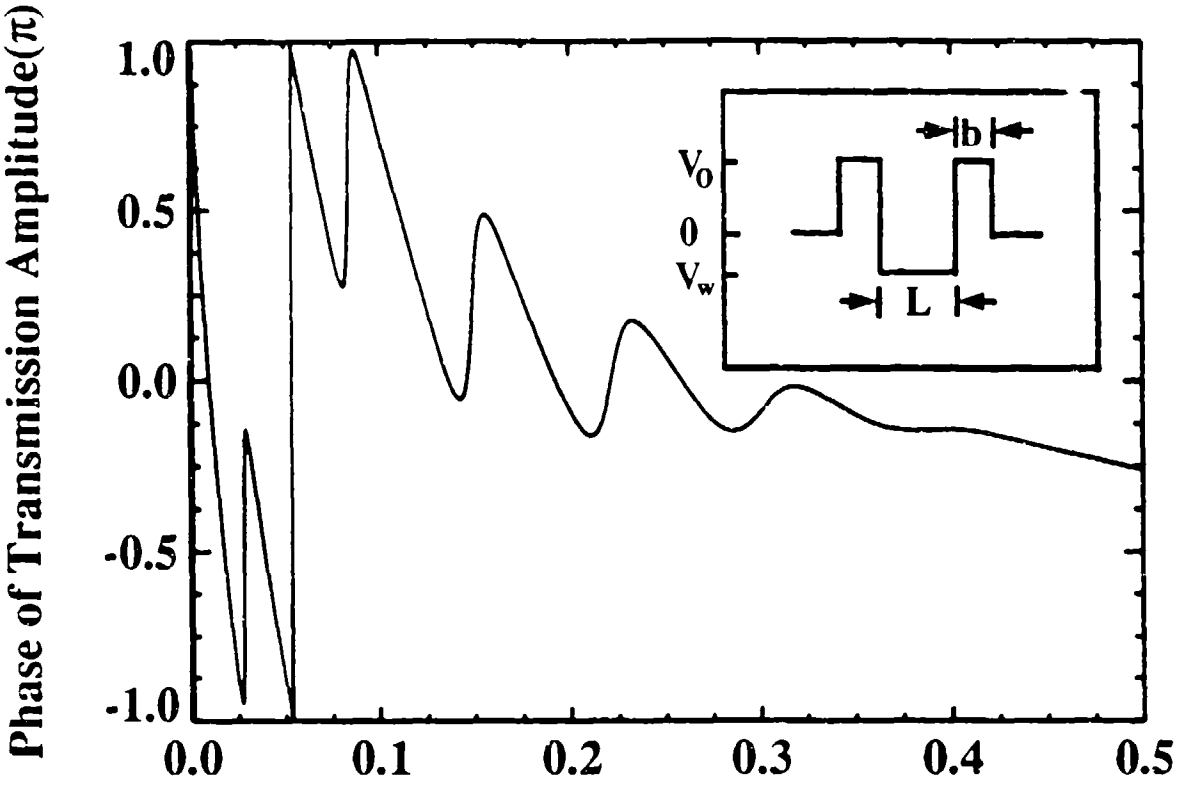

Imaginary Part of the Energy (eV)
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
0.0
0.1
0.2
0.3
0.4
0.5
Real Part of the Energy (eV)

Fig 2

(a)


Phase of Transmission Amplitude(π)
1.0
0.5
0.0
-0.5
-1.0
0.0
0.1
0.2
0.3
0.4
0.5
b
V0
0
Vw
L

(b)

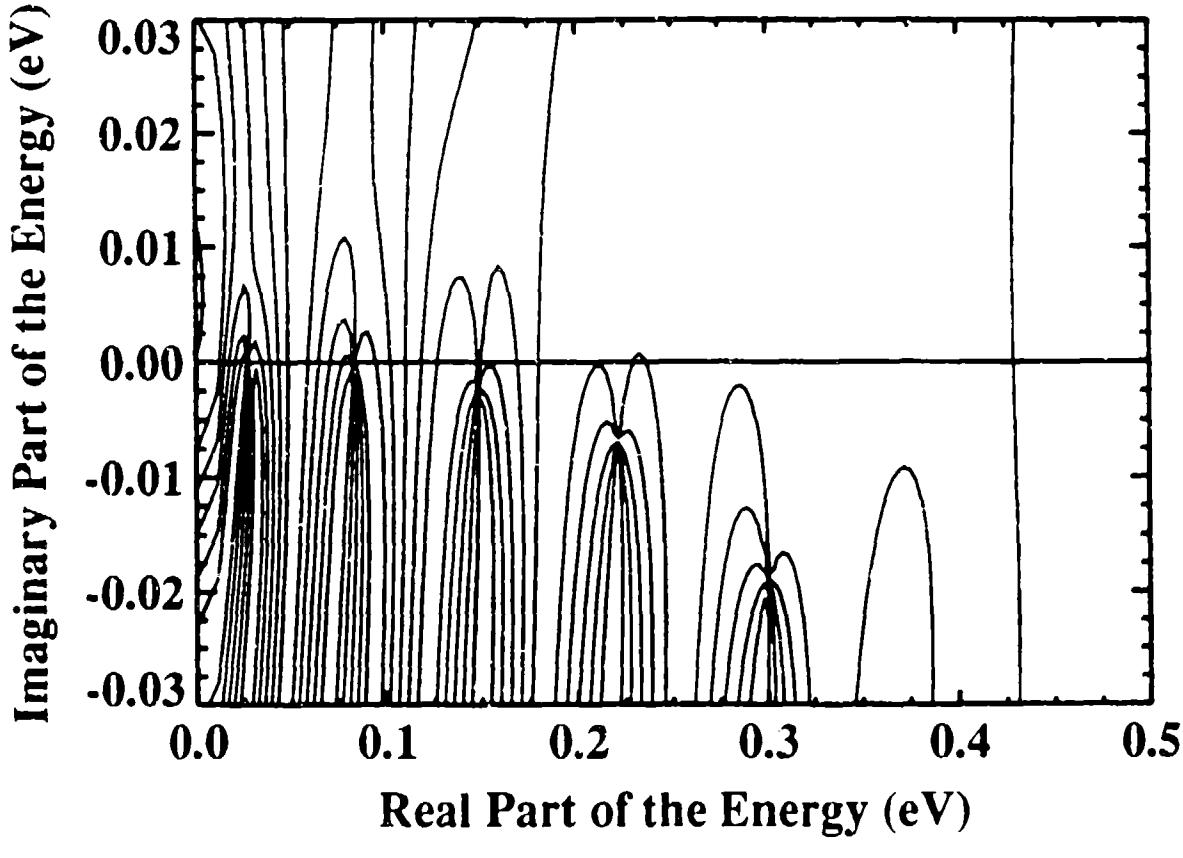

Imaginary Part of the Energy (eV)
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
0.0
0.1
0.2
0.3
0.4
0.5
Real Part of the Energy (eV)

Fig 3

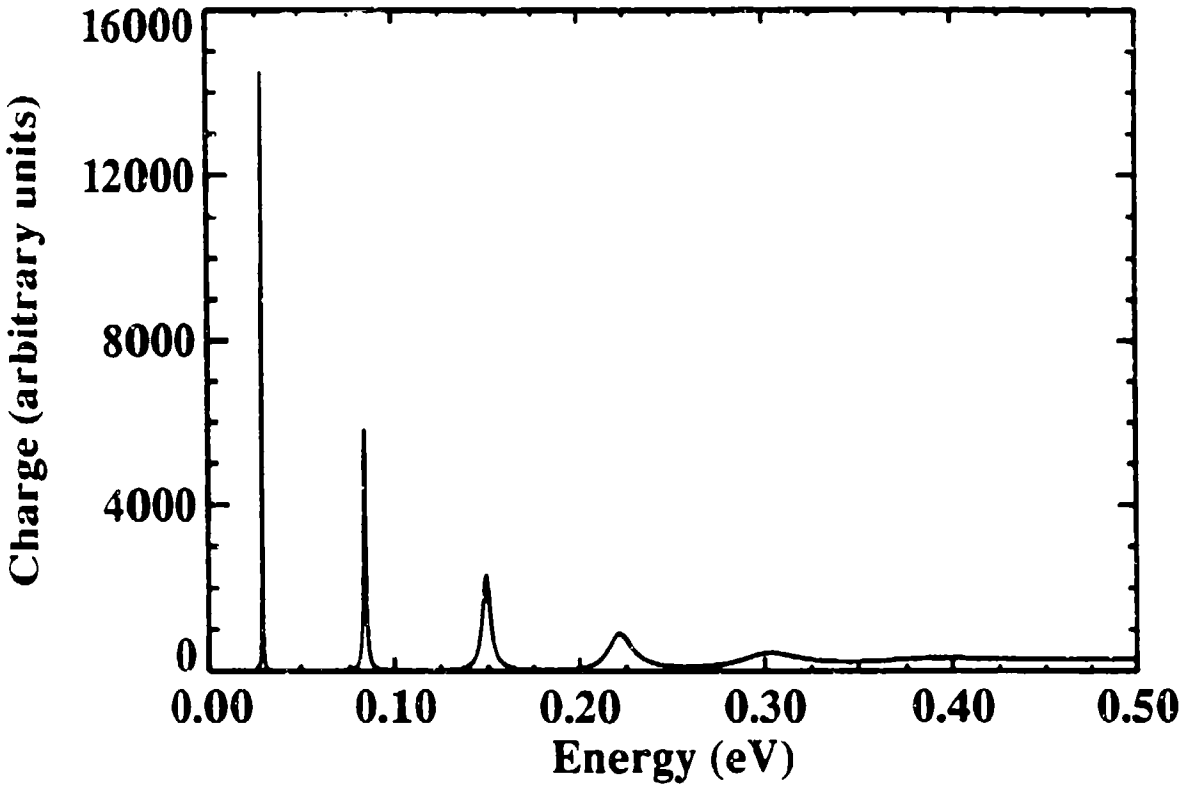
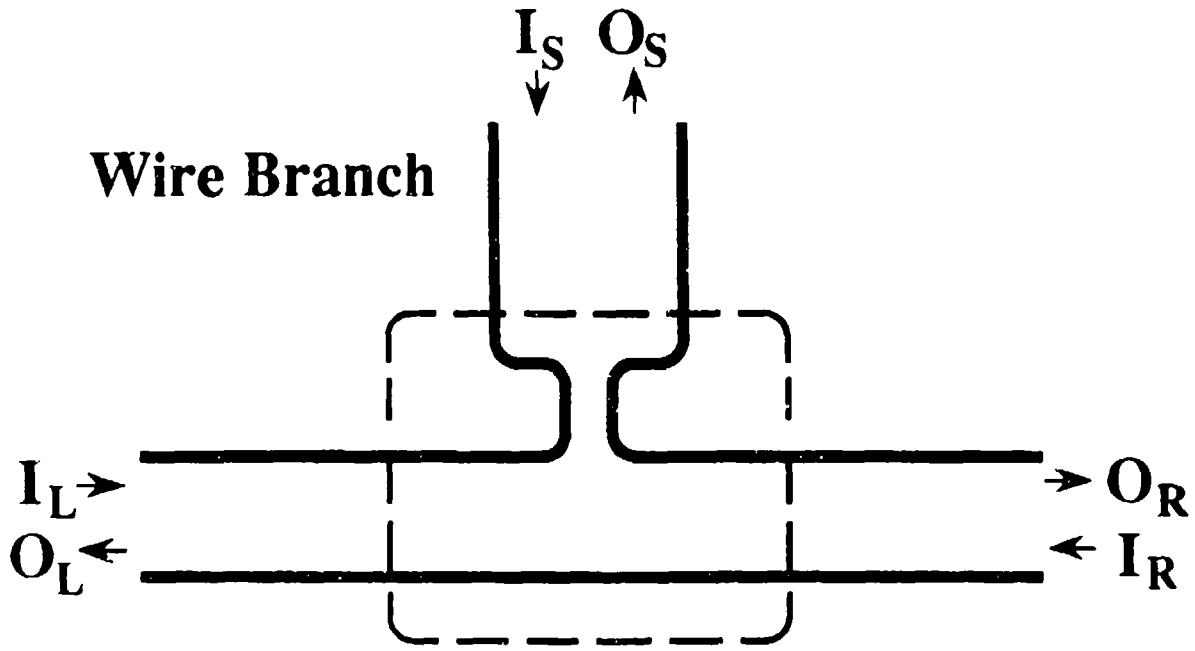
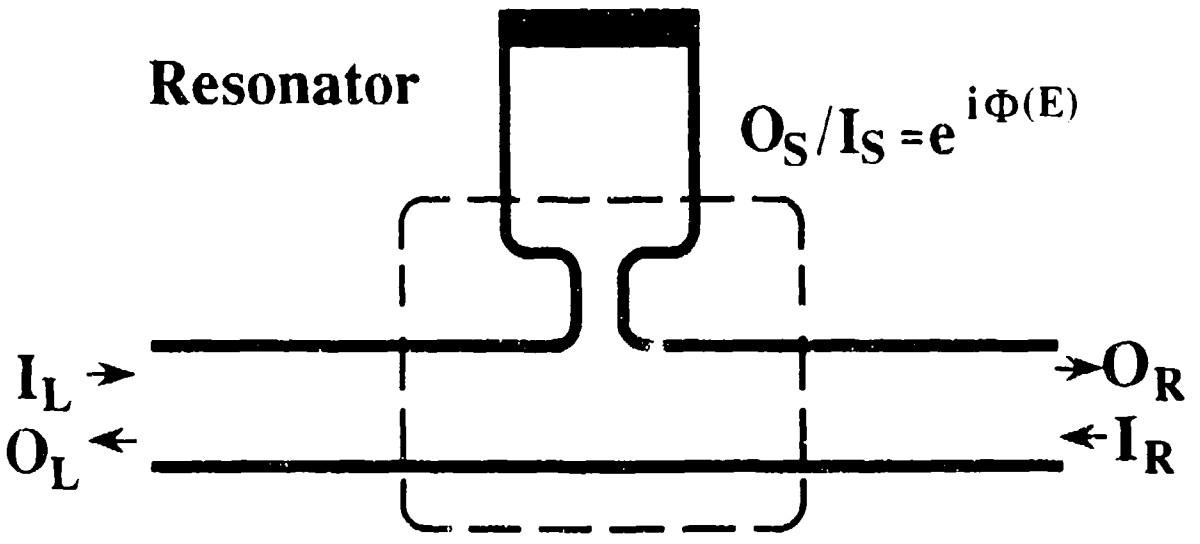

16000
12000
8000
4000
0
Charge (arbitrary units)
0.00
0.10
0.20
0.30
0.40
0.50
Energy (eV)

Fig 4

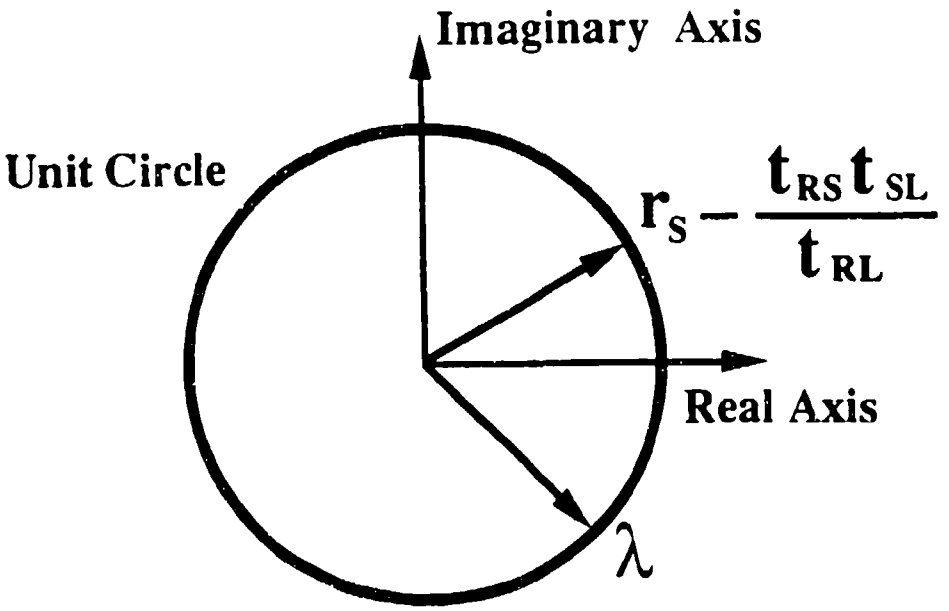


(a)

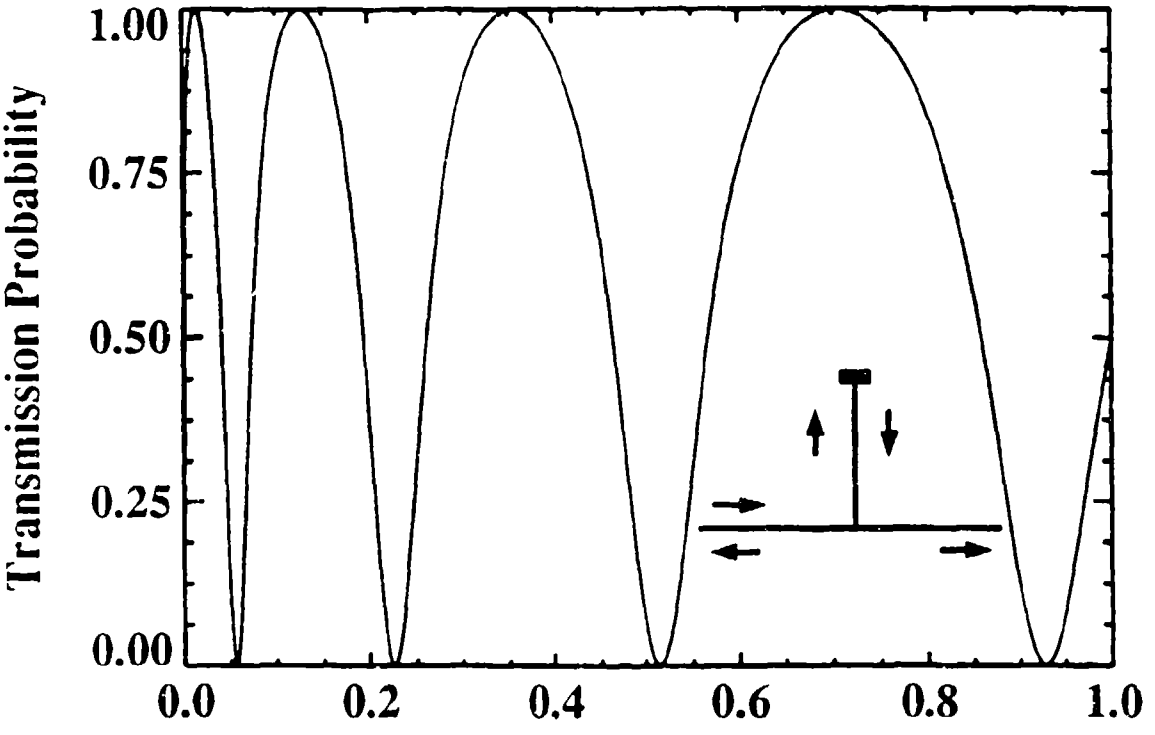


(b)

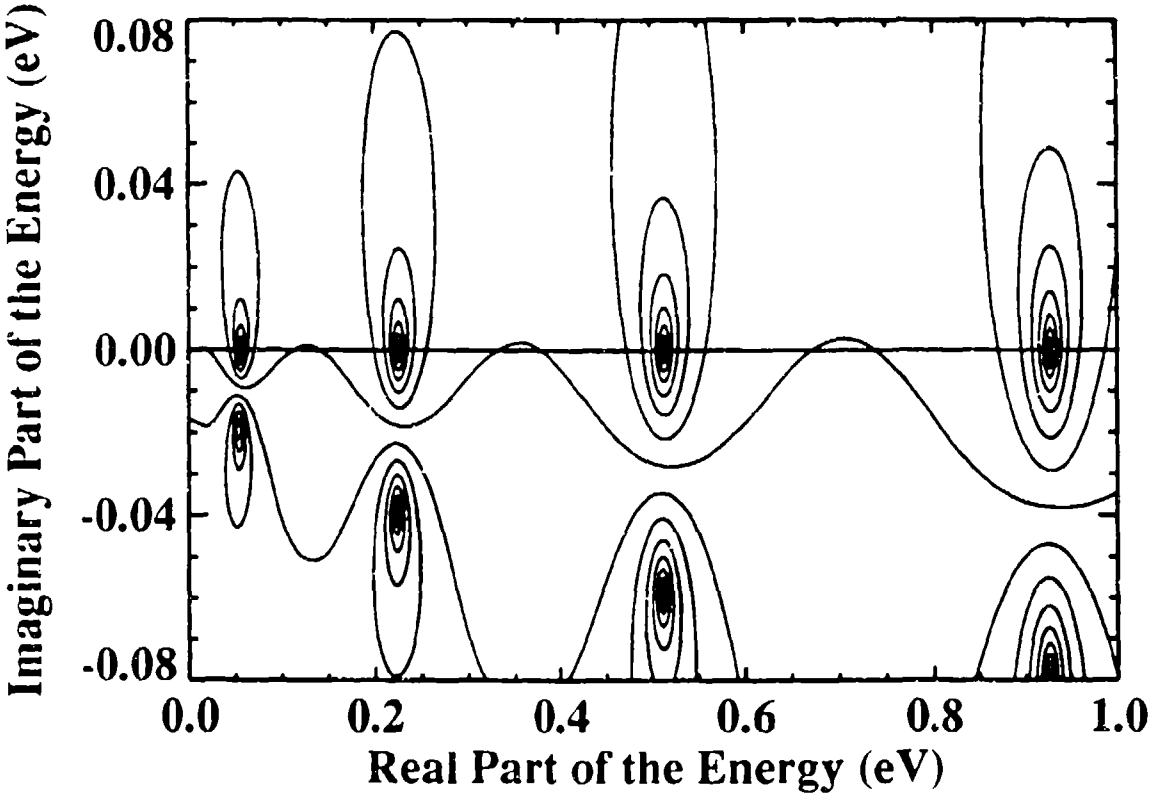
Fig. 5


Imaginary Axis
Unit Circle
$r_S - \frac{t_{RS} t_{SL}}{t_{RL}}$
Real Axis
λ

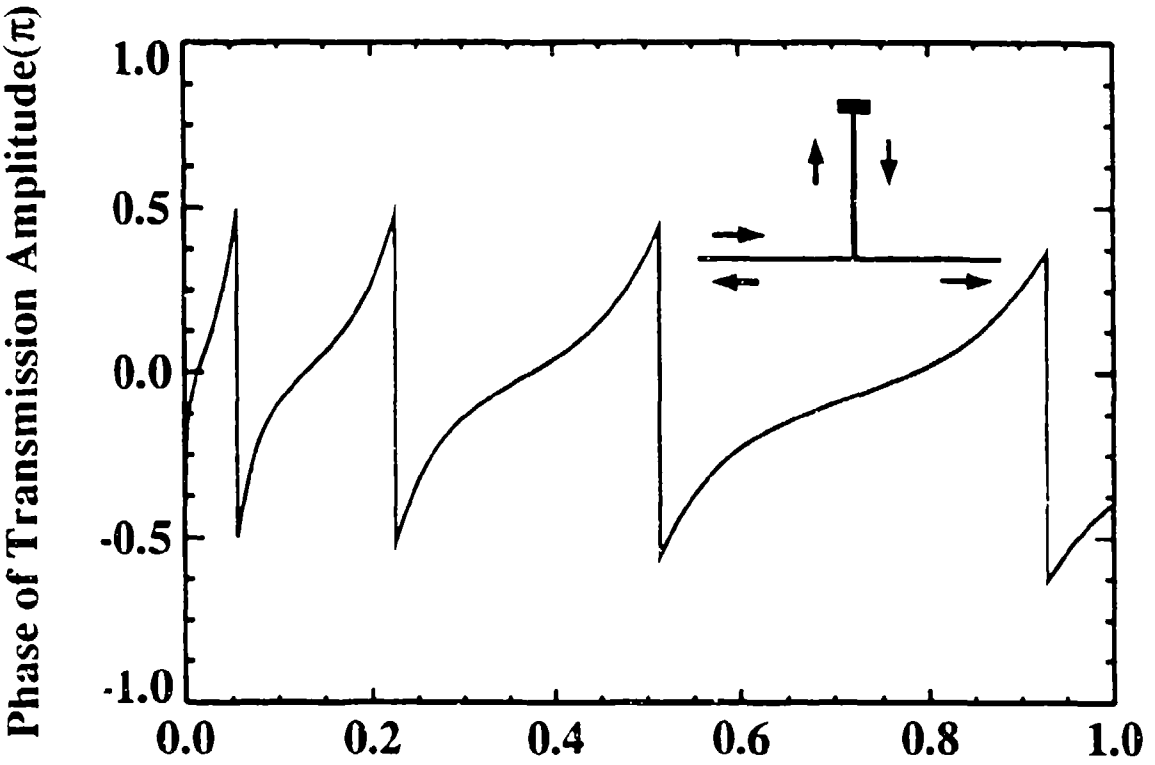
(a)


Transmission Probability
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

(b)


Imaginary Part of the Energy (eV)
0.08
0.04
0.00
-0.04
-0.08
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

(a)


Phase of Transmission Amplitude(π)
1.0
0.5
0.0
-0.5
-1.0
0.0
0.2
0.4
0.6
0.8
1.0

(b)

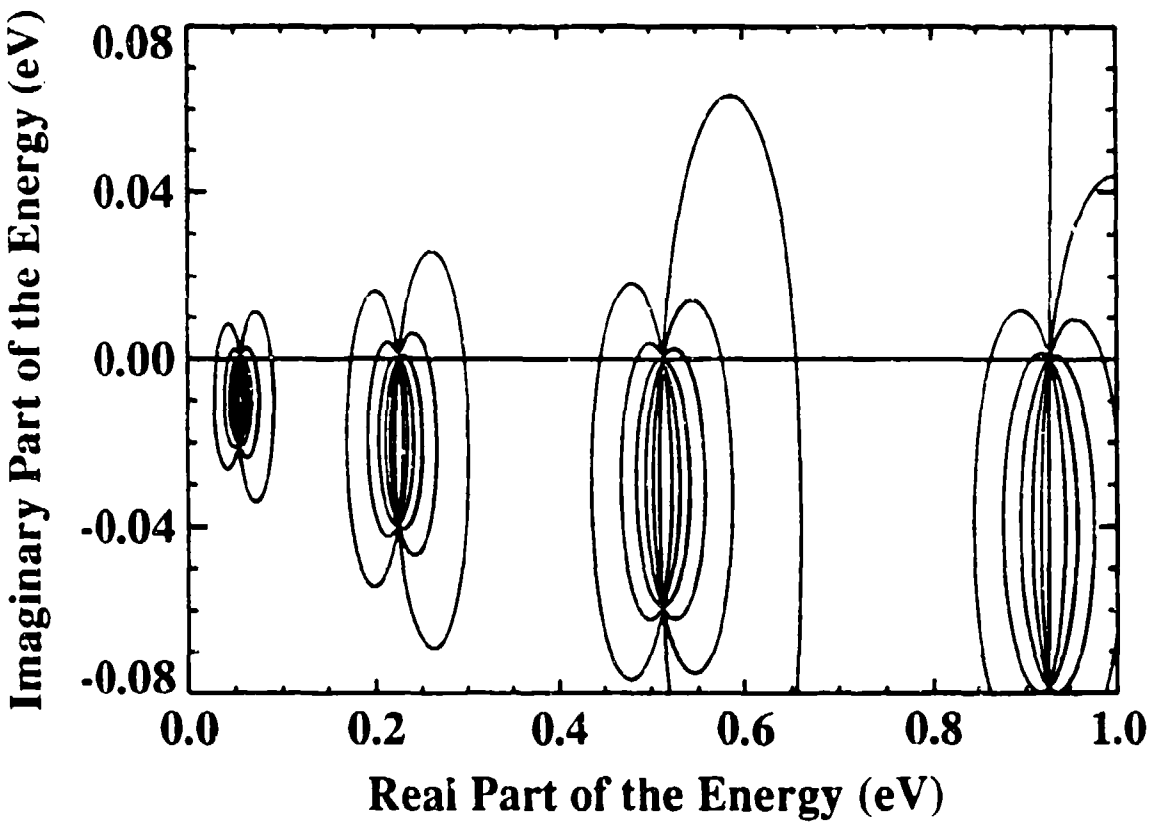
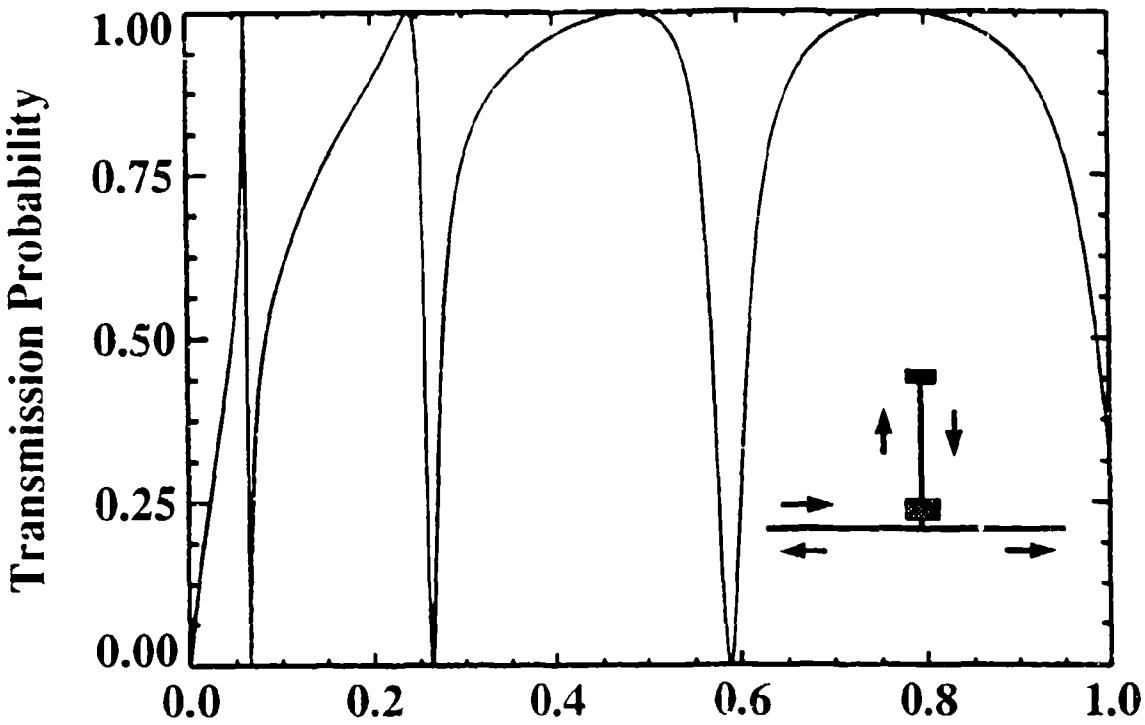
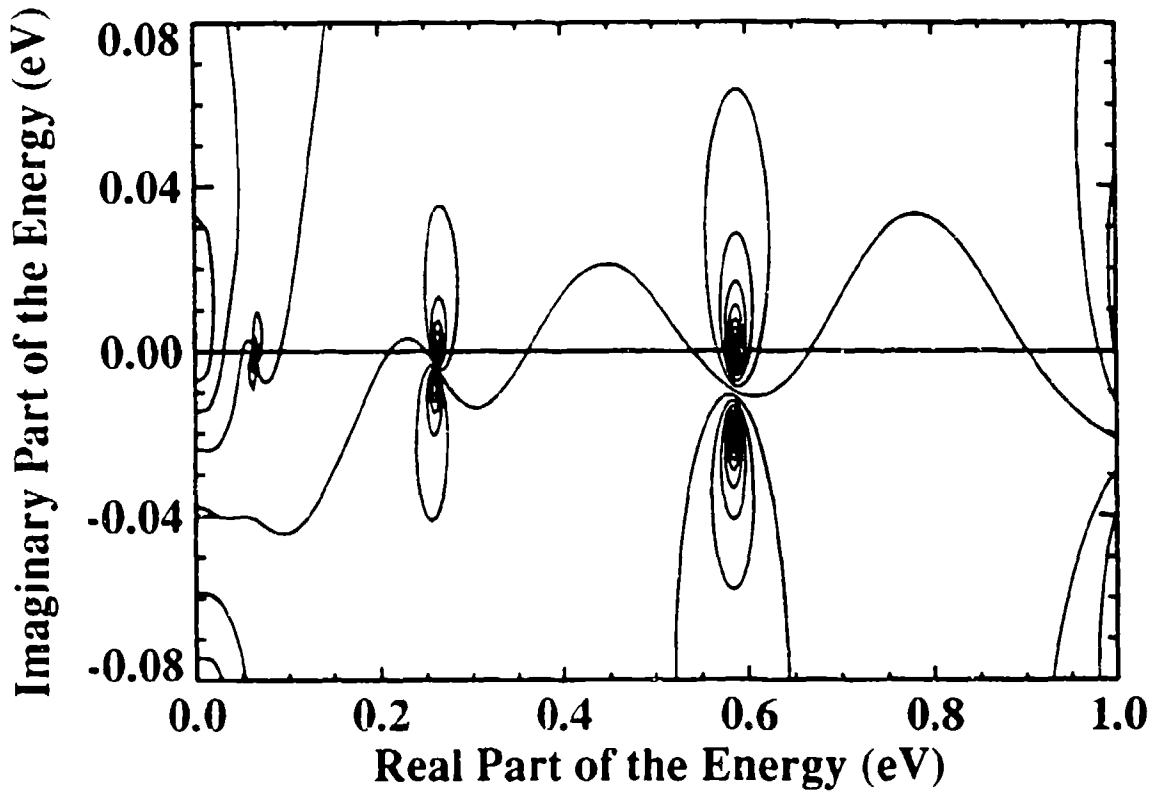

Imaginary Part of the Energy (eV)
0.08
0.04
0.00
-0.04
-0.08
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

Fig. 3

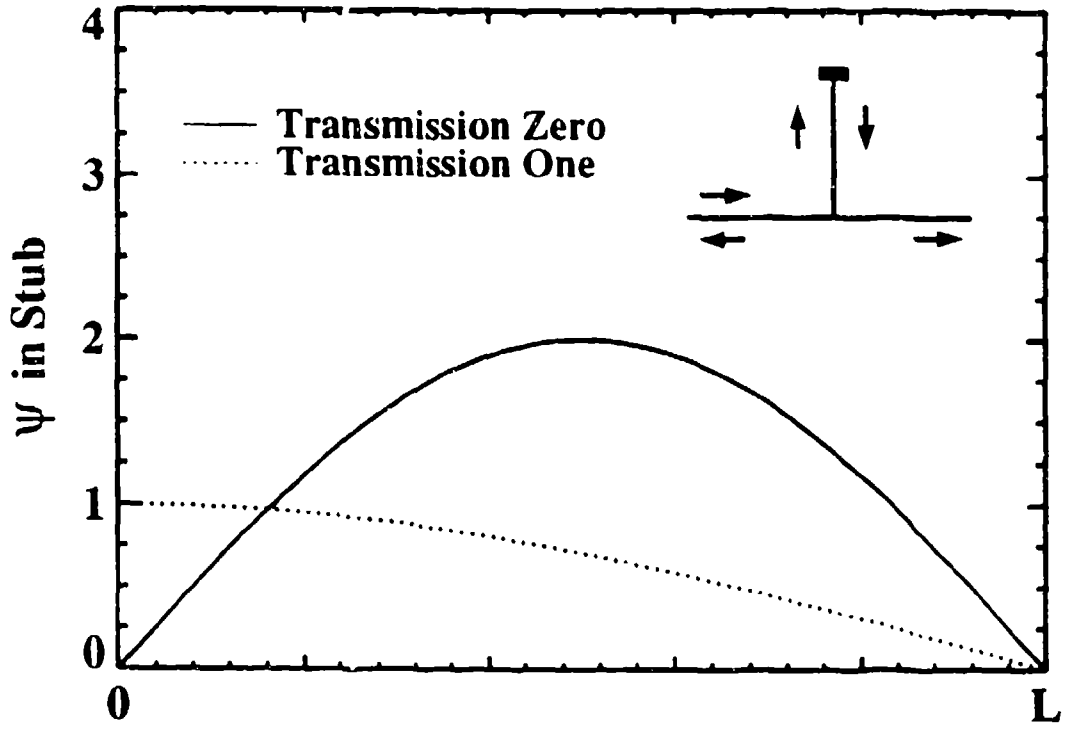

Transmission Probability
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

(a)

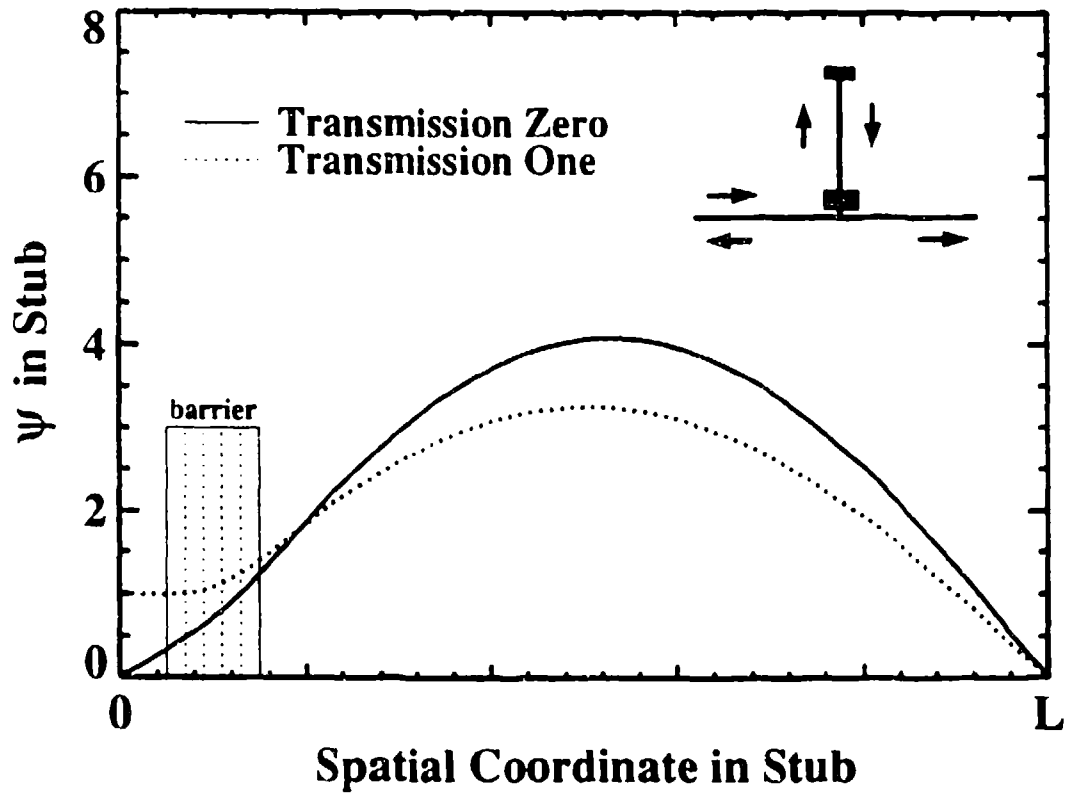

Imaginary Part of the Energy (eV)
0.08
0.04
0.00
-0.04
-0.08
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

(b)

Fig 9

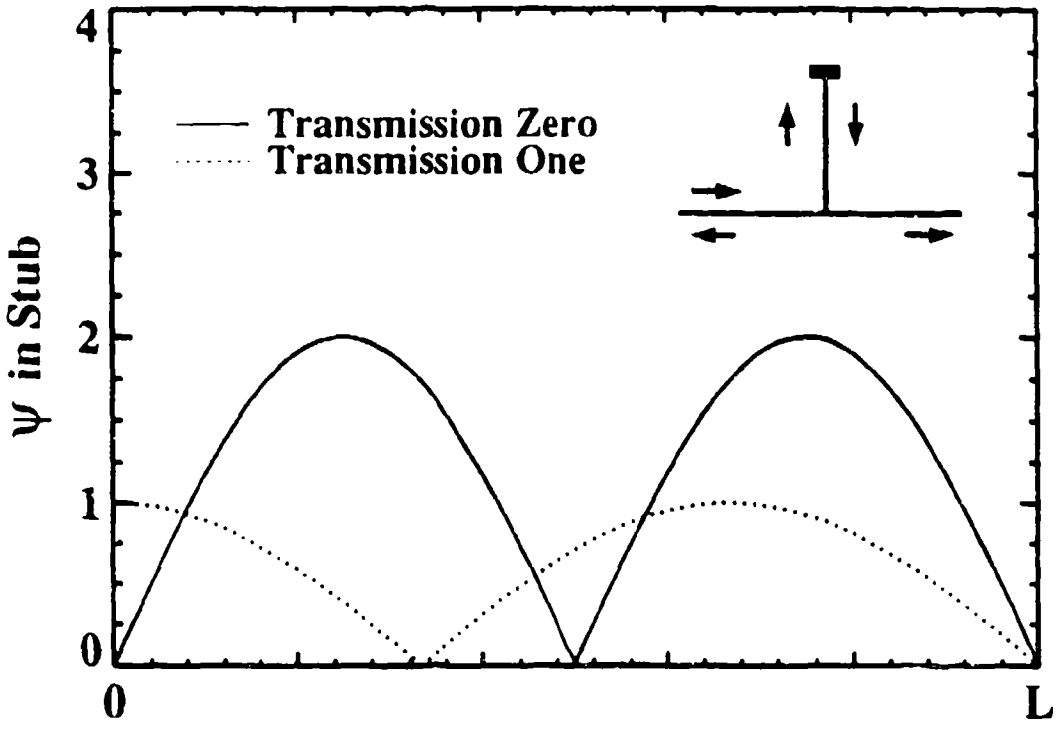

Transmission Zero
Transmission One
ψ in Stub
0
1
2
3
4
0
L

(a)

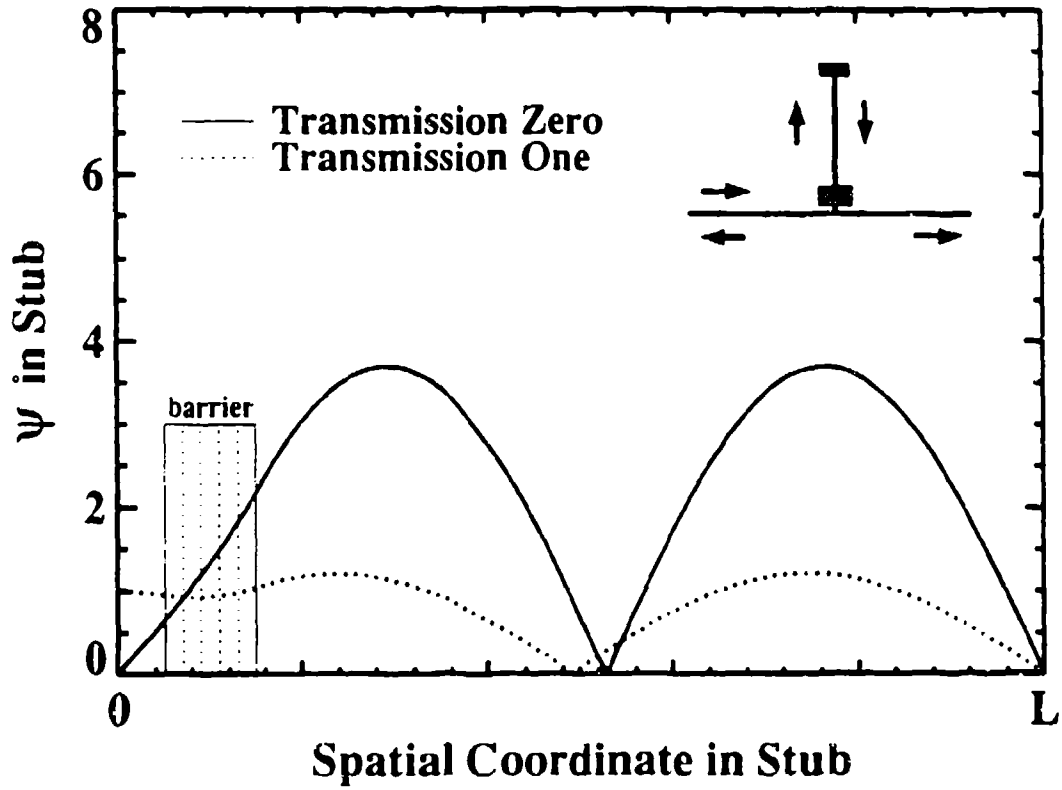

Transmission Zero
Transmission One
ψ in Stub
barrier
0
2
4
6
8
0
L
Spatial Coordinate in Stub

(b)

Fig. 10



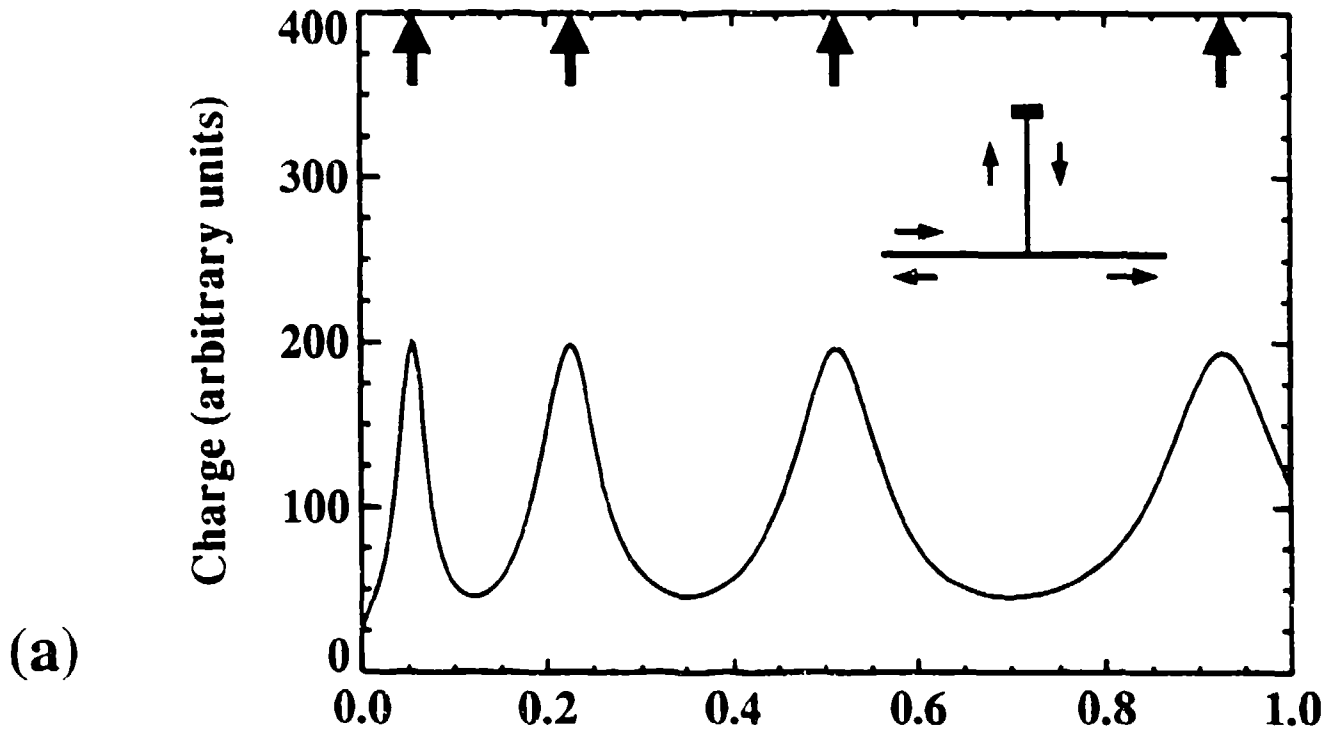
(a)



(b)

Spatial Coordinate in Stub

Fig. 11


(a)
Charge (arbitrary units)
400
300
200
100
0
0.0
0.2
0.4
0.6
0.8
1.0

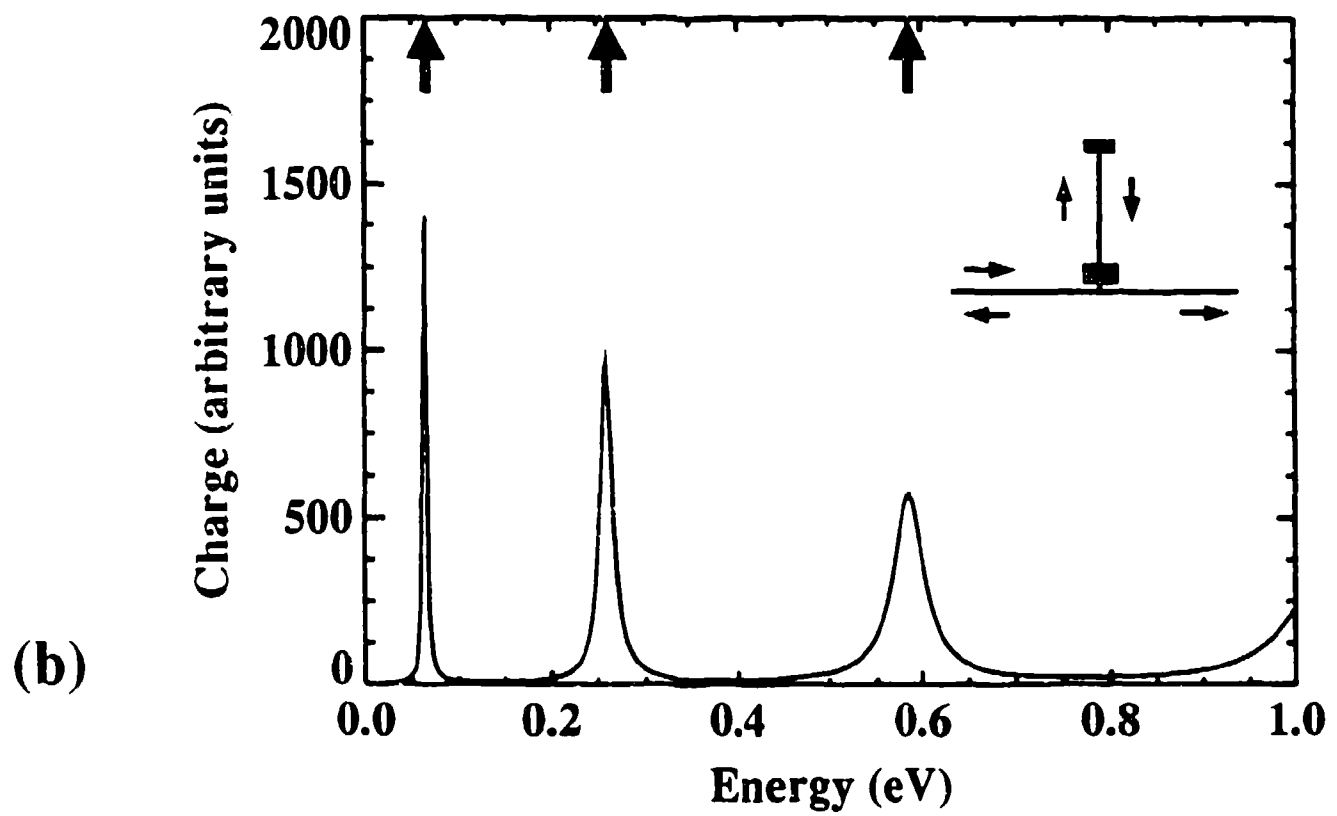
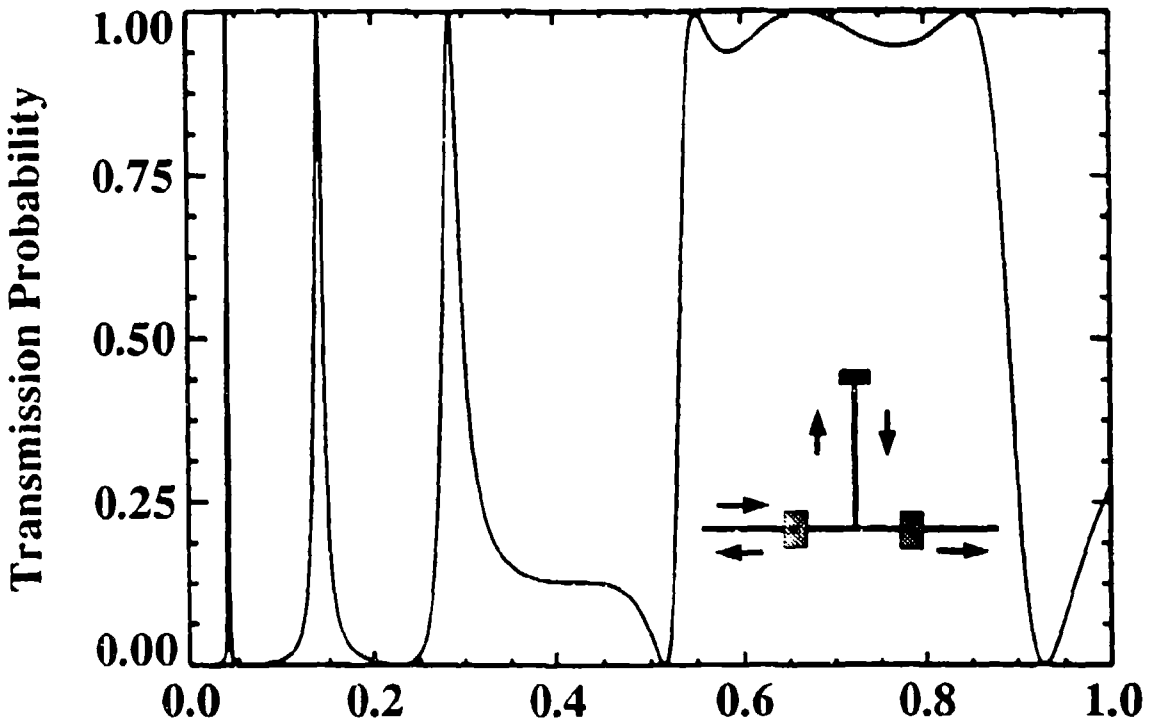

(b)
Charge (arbitrary units)
2000
1500
1000
500
0
0.0
0.2
0.4
0.6
0.8
1.0
Energy (eV)

Fig. 12

(a)


Transmission Probability
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

(b)

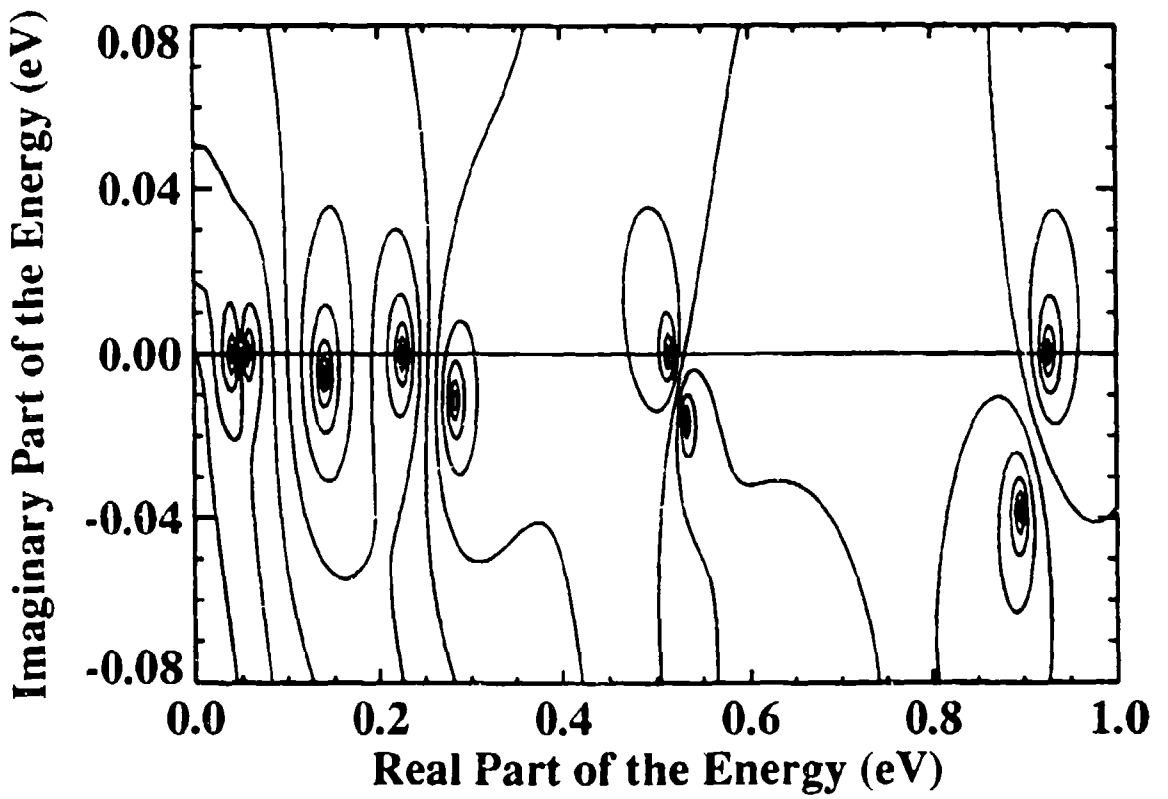
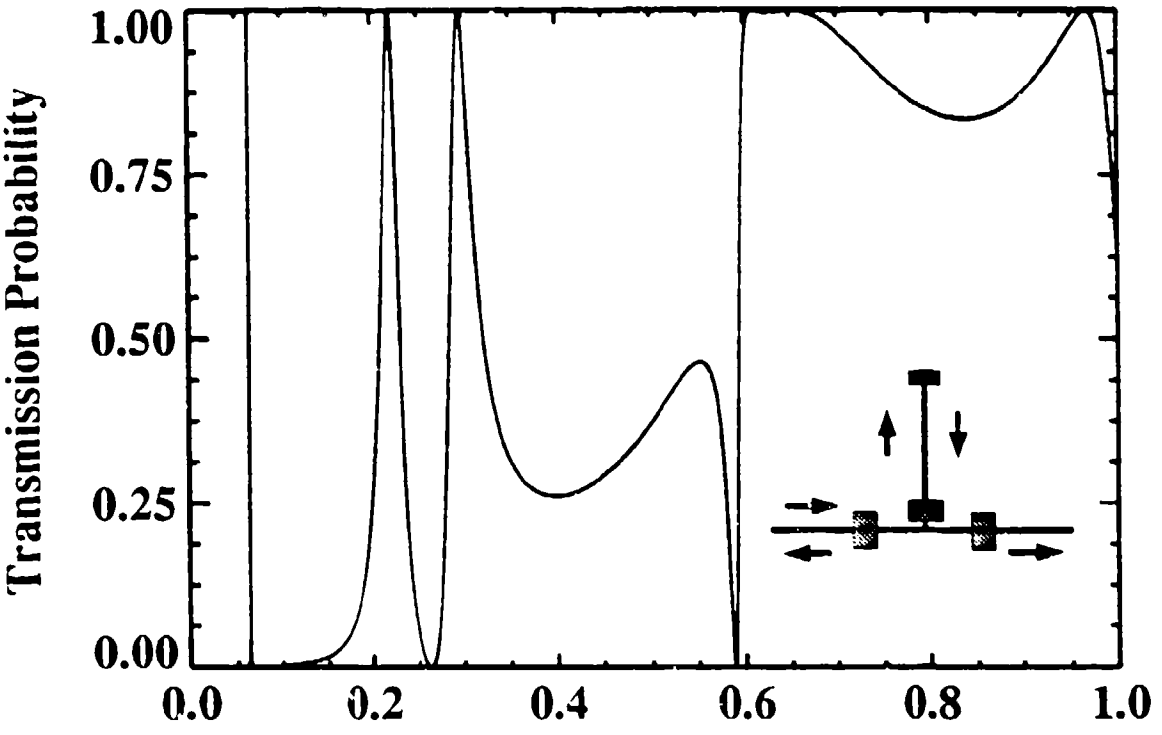
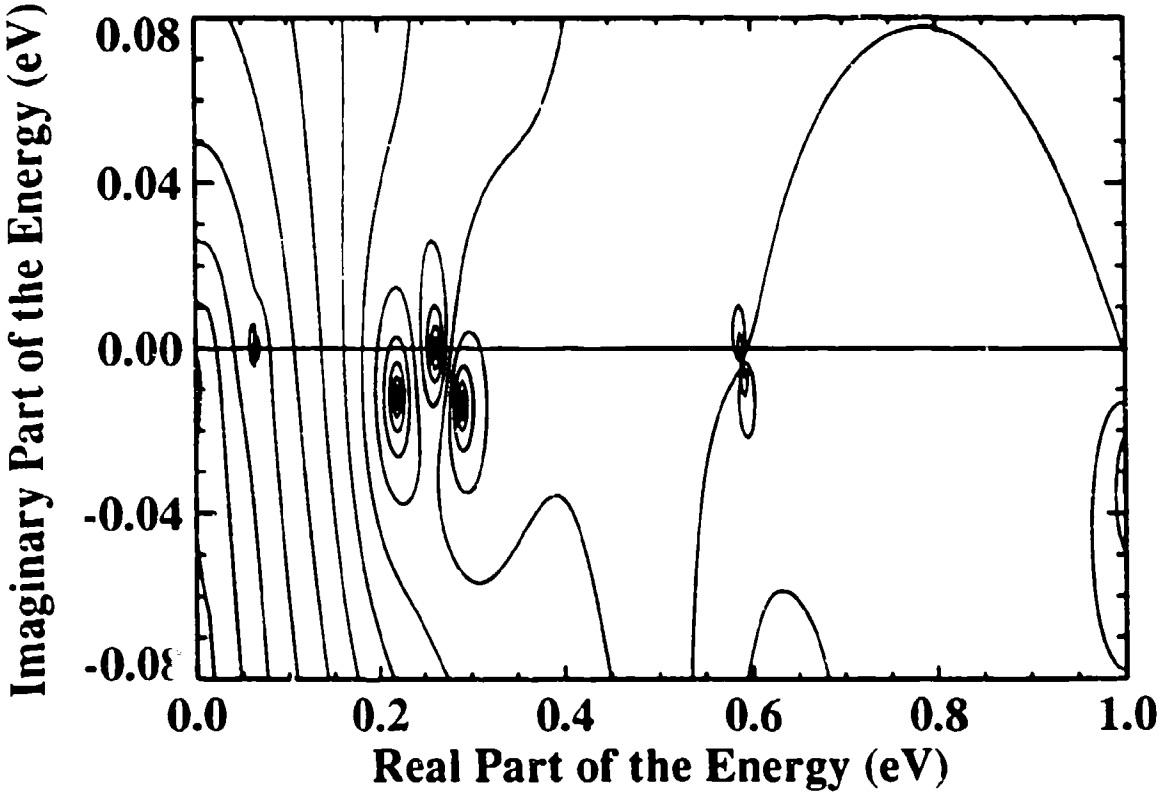

Imaginary Part of the Energy (eV)
0.08
0.04
0.00
-0.04
-0.08
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

Fig 13


Transmission Probability
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

(a)


Imaginary Part of the Energy (eV)
0.08
0.04
0.00
-0.04
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

(b)

Fig 14

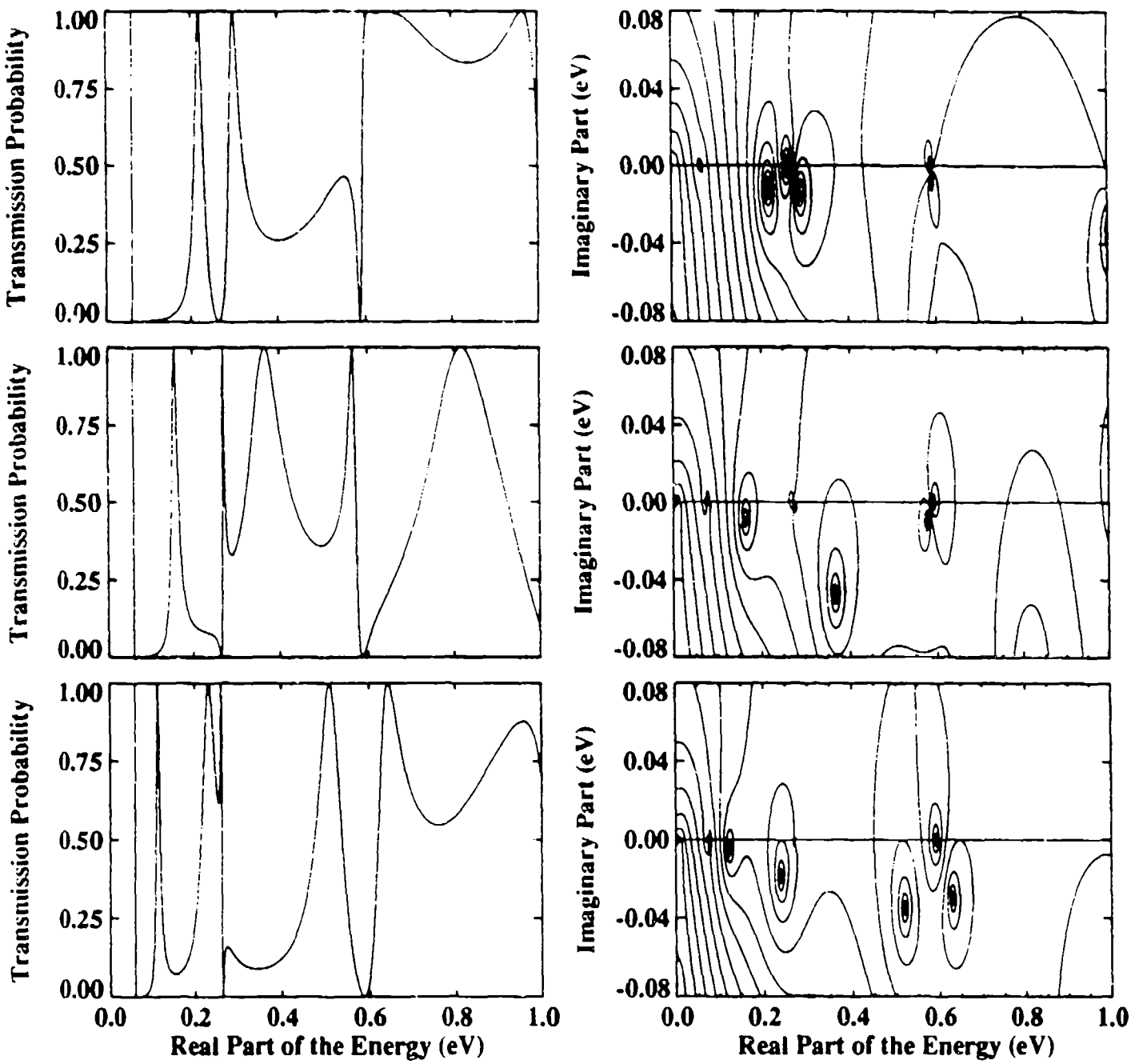
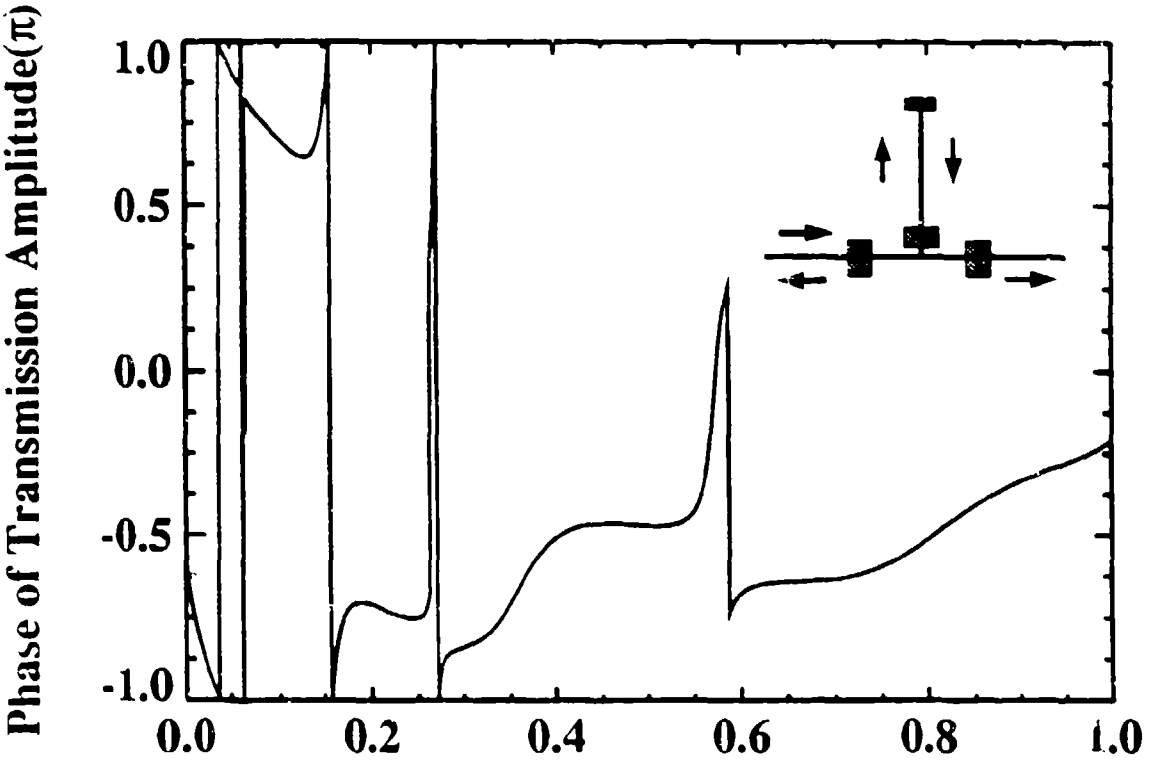

(a)
(b)
(c)
Transmission Probability
Imaginary Part (eV)
1.00
0.75
0.50
0.25
0.00
0.08
0.04
0.00
-0.04
-0.08
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

Fig. 15

(a)


Phase of Transmission Amplitude(π)
1.0
0.5
0.0
-0.5
-1.0
0.0
0.2
0.4
0.6
0.8
1.0

(b)

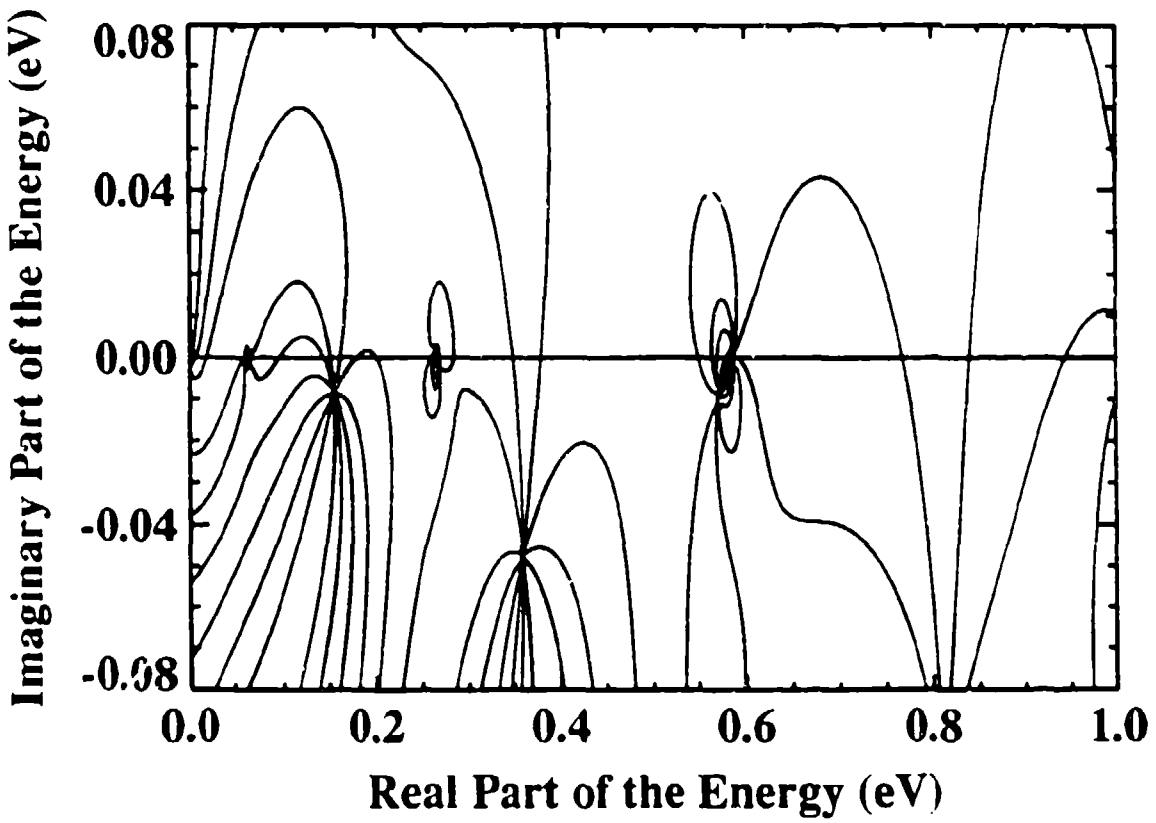
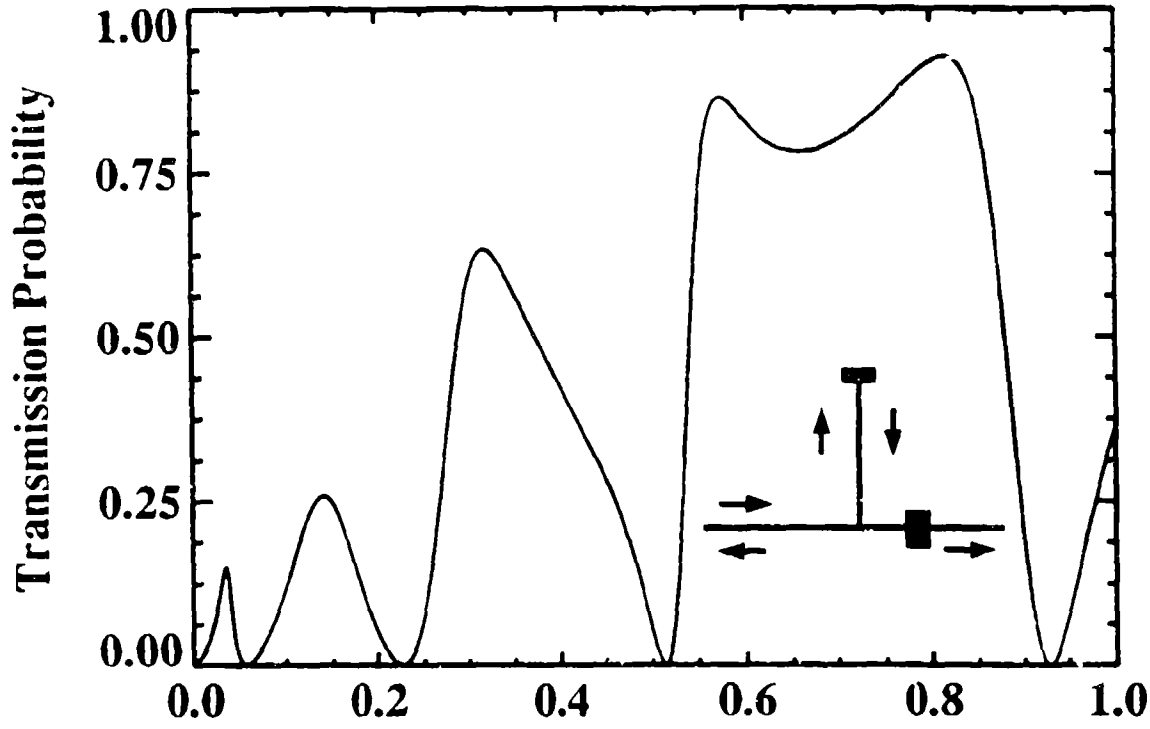
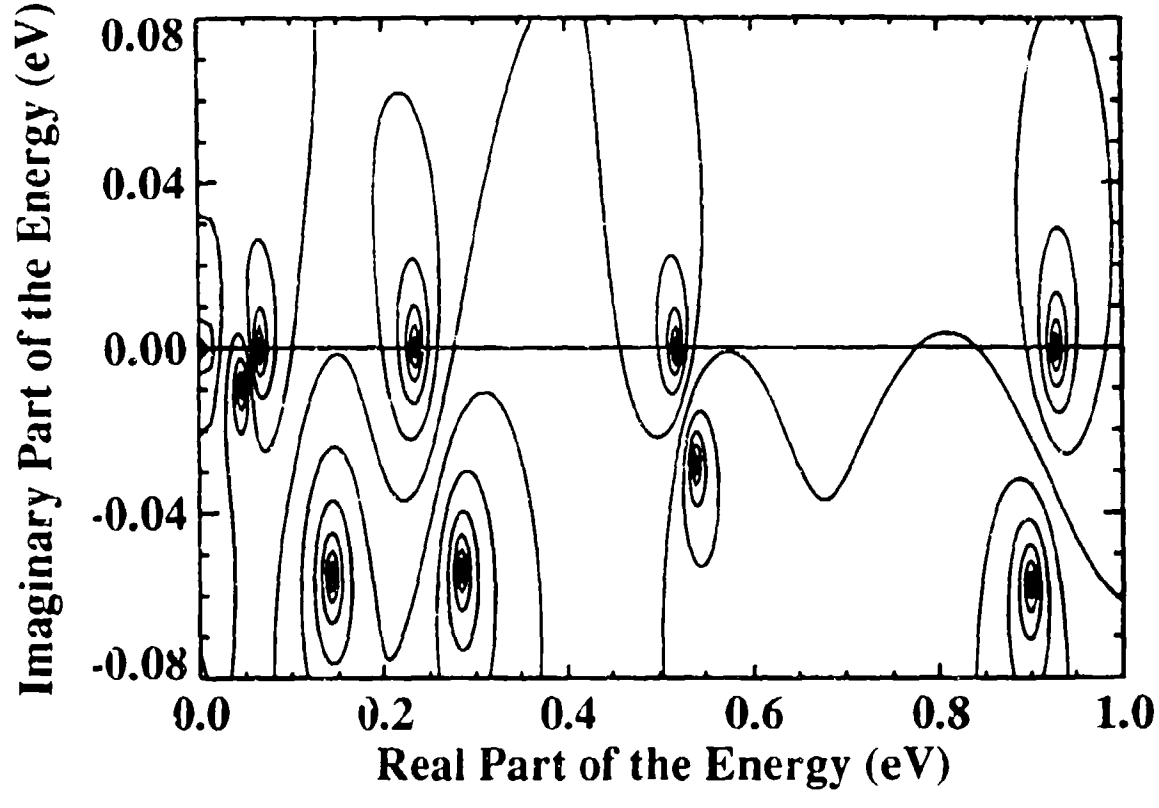

Imaginary Part of the Energy (eV)
0.08
0.04
0.00
-0.04
-0.08
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

Fig. 16


Transmission Probability
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

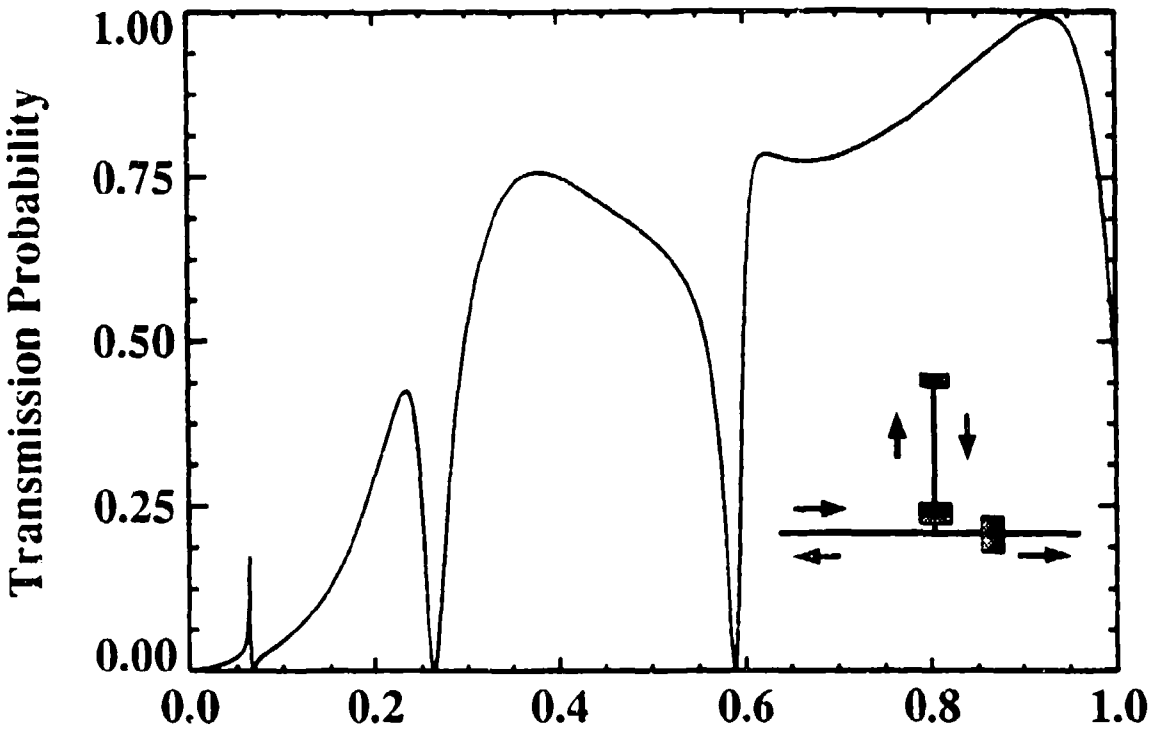
(a)


Imaginary Part of the Energy (eV)
0.08
0.04
0.00
-0.04
-0.08
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

(b)

Fig. 17

(a)


Transmission Probability
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

(b)

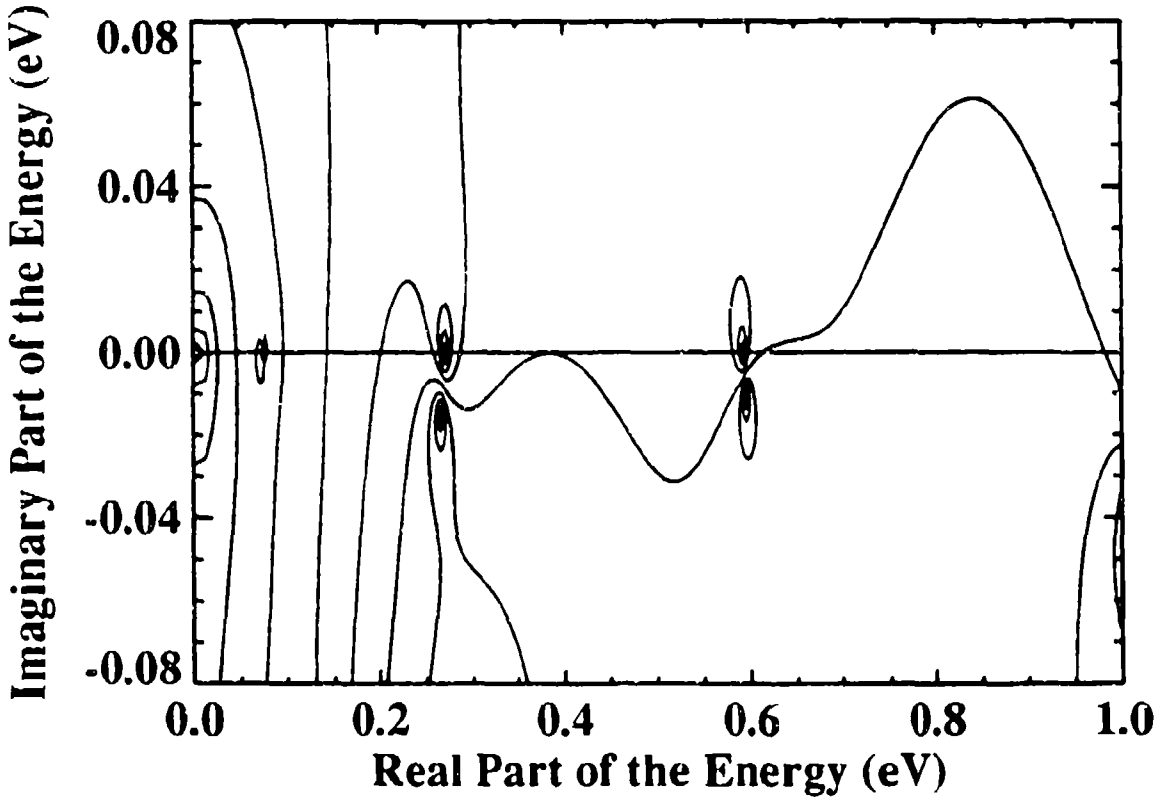
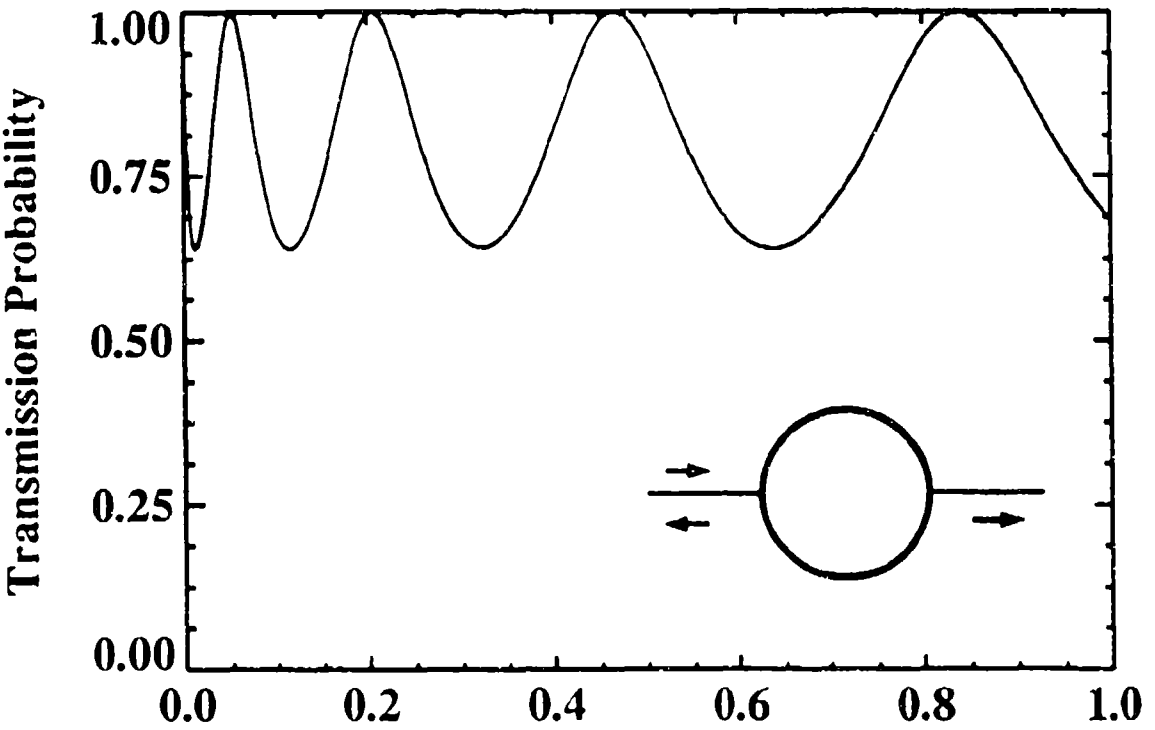
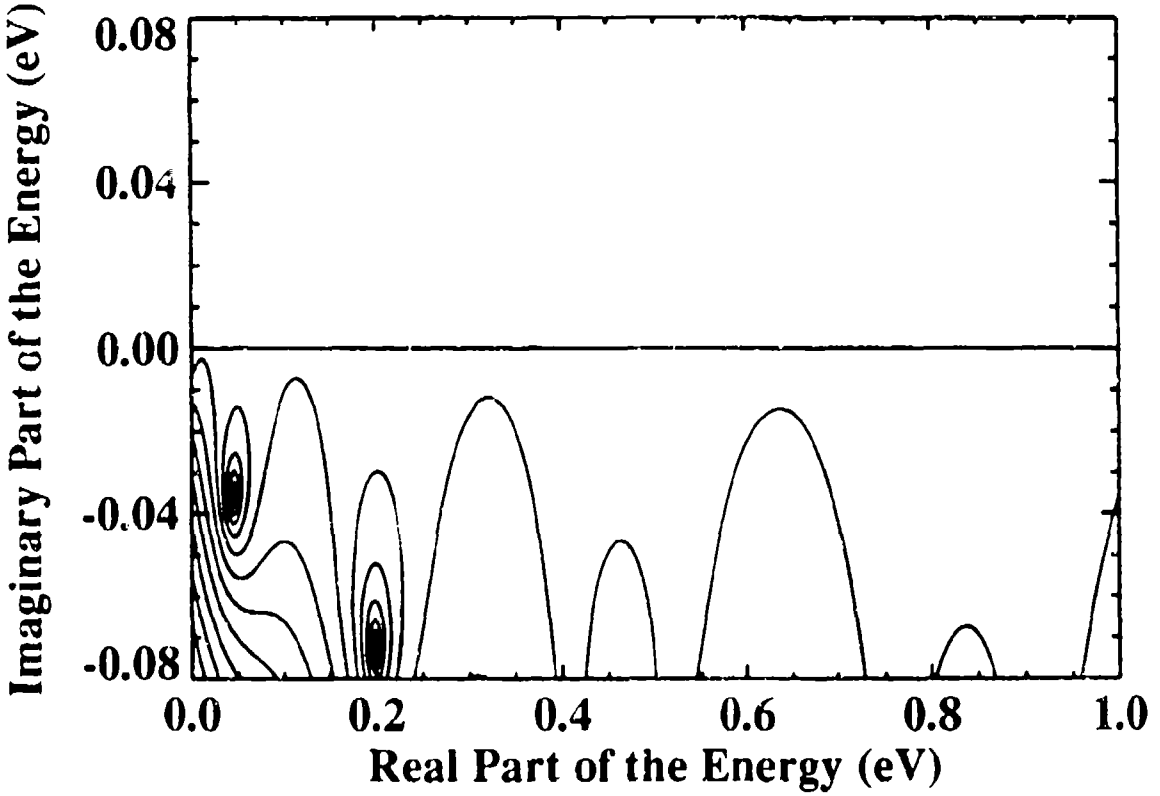

Imaginary Part of the Energy (eV)
0.08
0.04
0.00
-0.04
-0.08
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

Fig 18


Transmission Probability
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

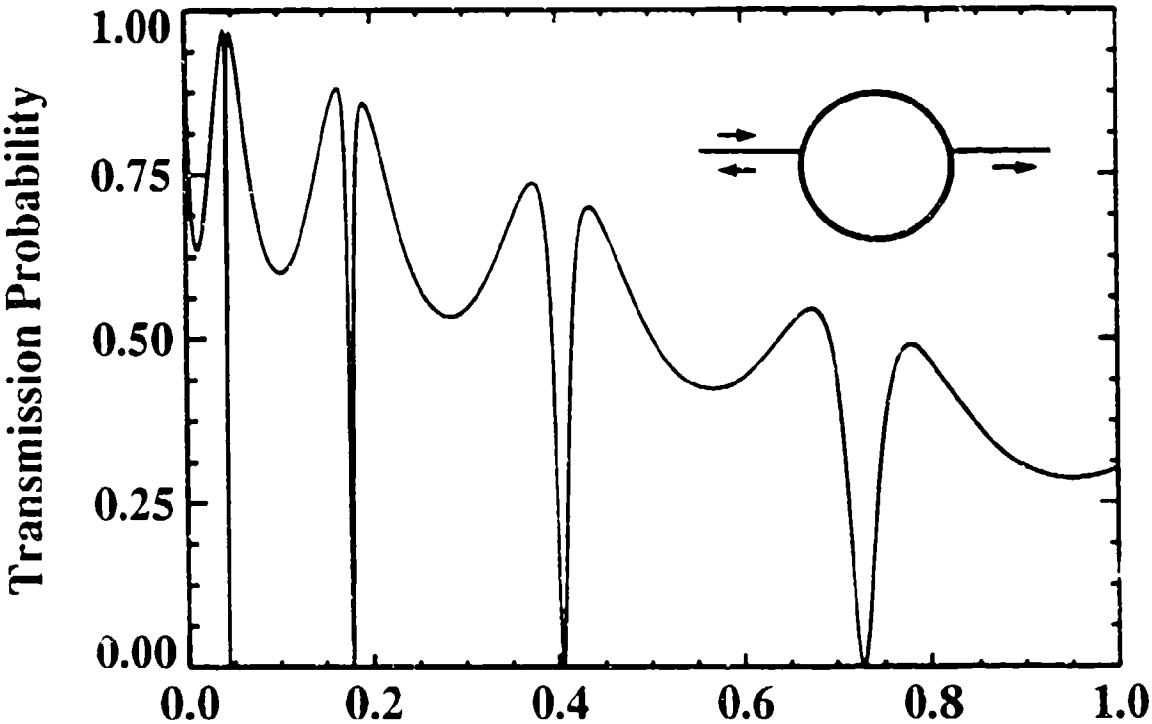
(a)


Imaginary Part of the Energy (eV)
0.08
0.04
0.00
-0.04
-0.08
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

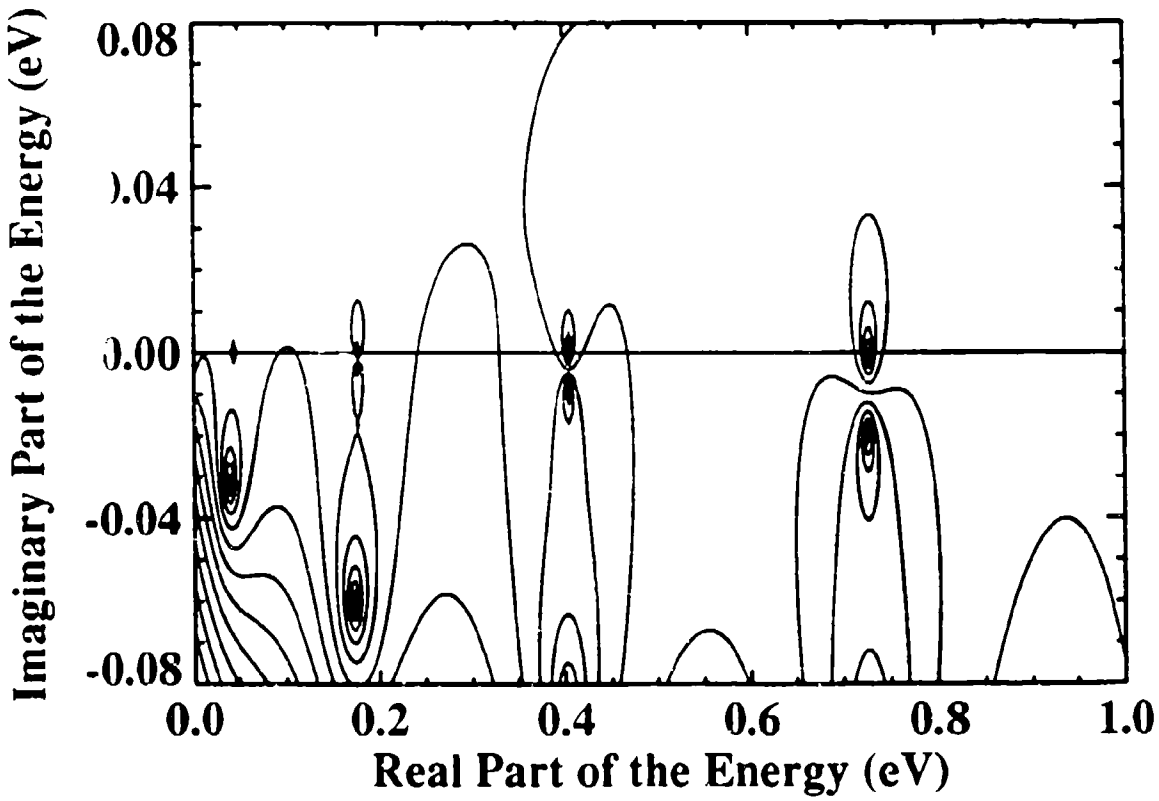
(b)

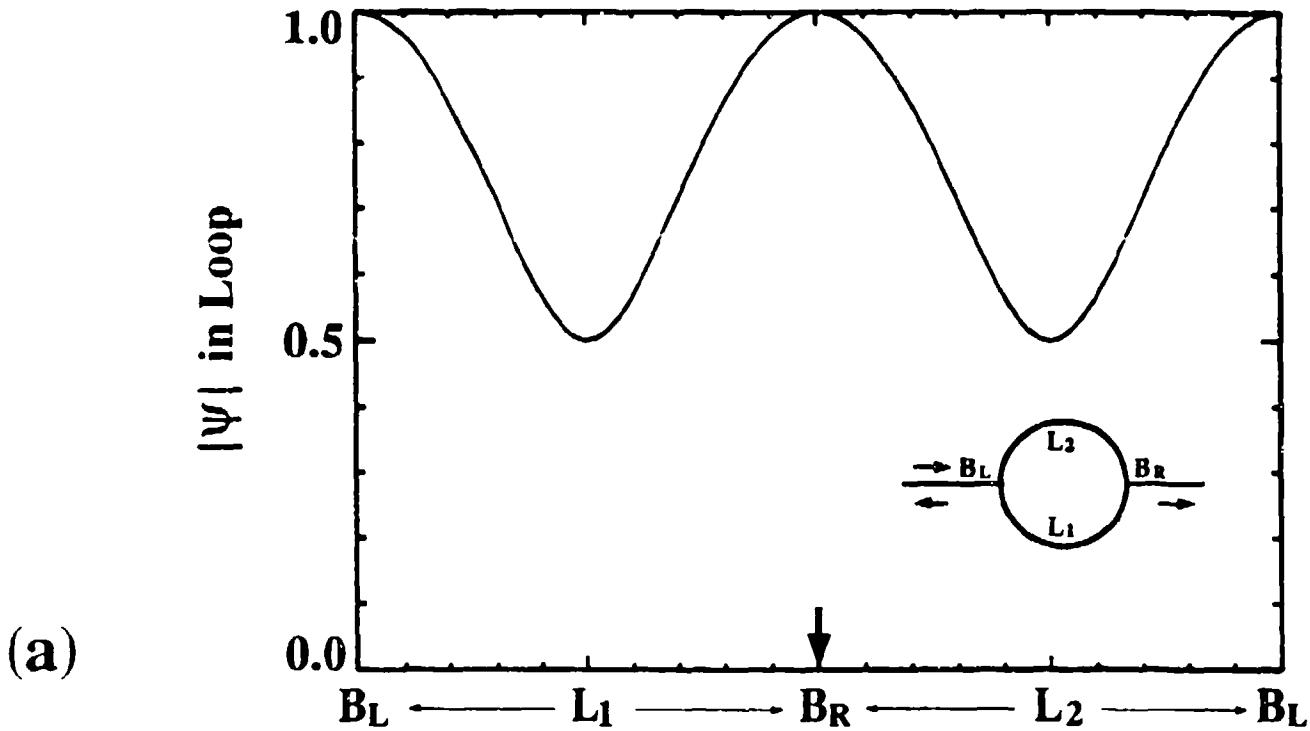
Fig. 19

(a)

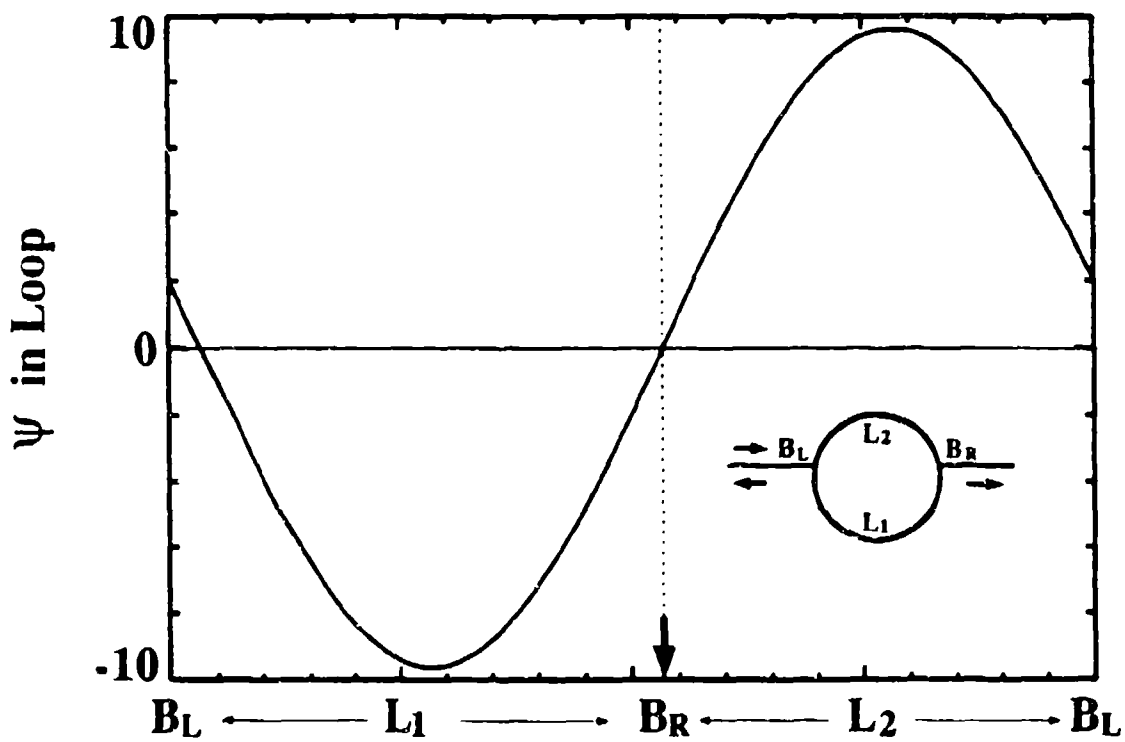

Transmission Probability
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

(b)


Imaginary Part of the Energy (eV)
0.08
0.04
0.00
-0.04
-0.08
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

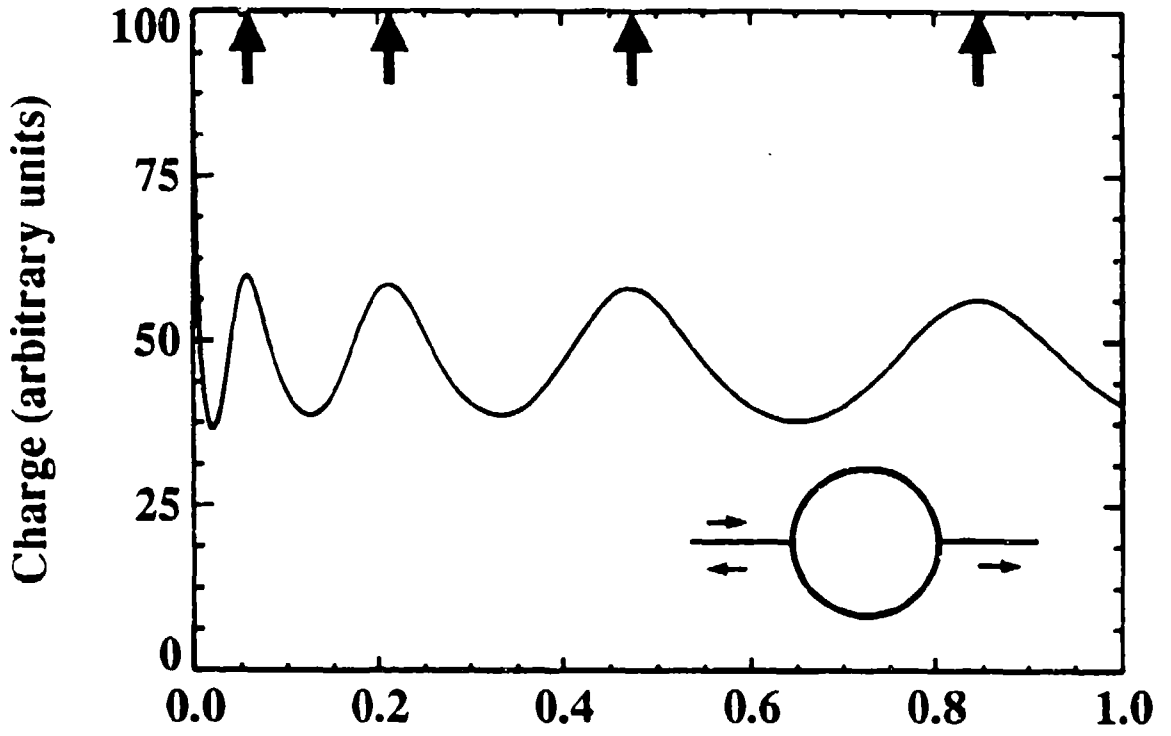


(a)

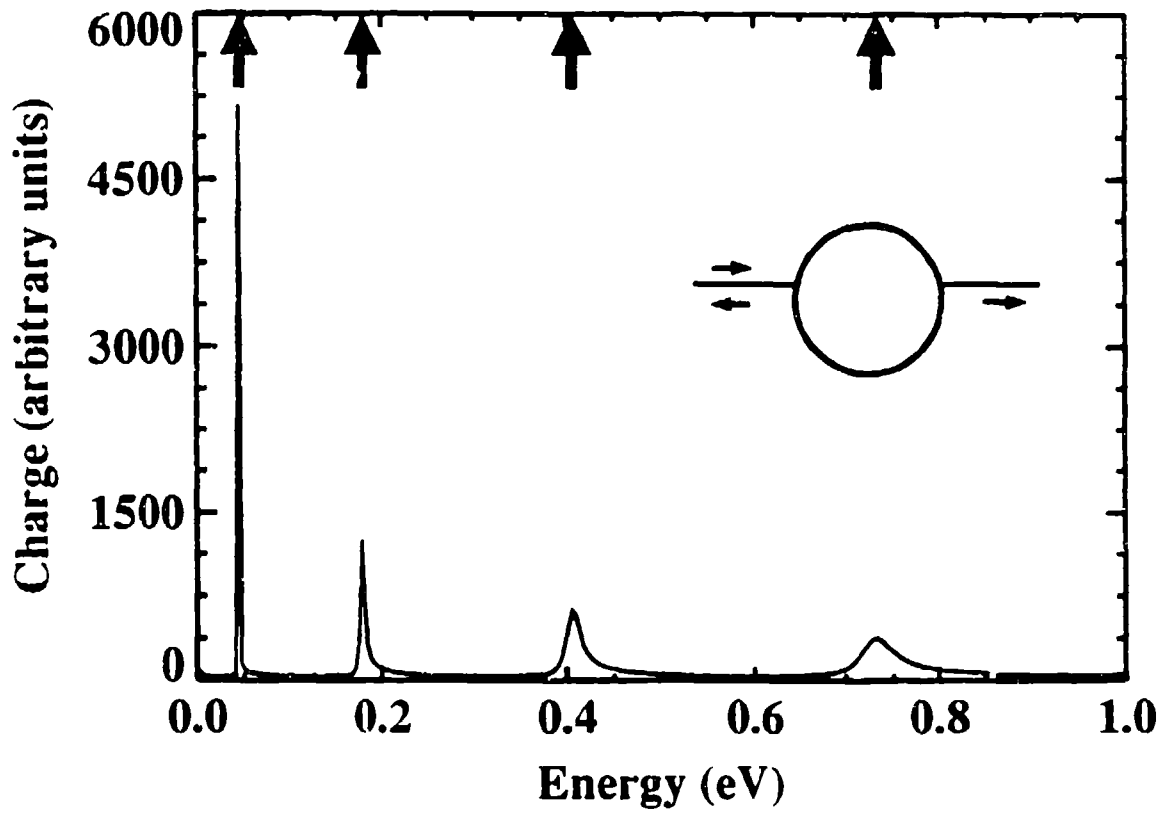


(b)

Spatial Coordinate in Loop


100
75
50
25
0
Charge (arbitrary units)
0.0
0.2
0.4
0.6
0.8
1.0

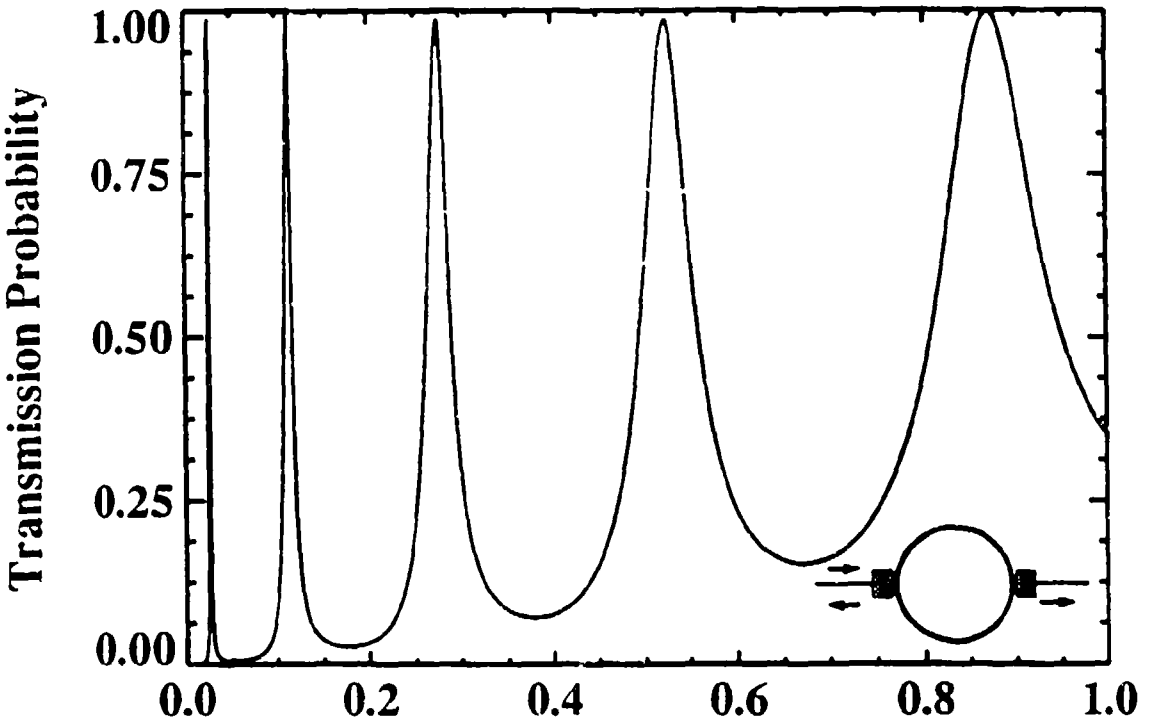
(a)


6000
4500
3000
1500
0
Charge (arbitrary units)
0.0
0.2
0.4
0.6
0.8
1.0
Energy (eV)

(b)

Fig. 22

(a)


Transmission Probability
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

(b)

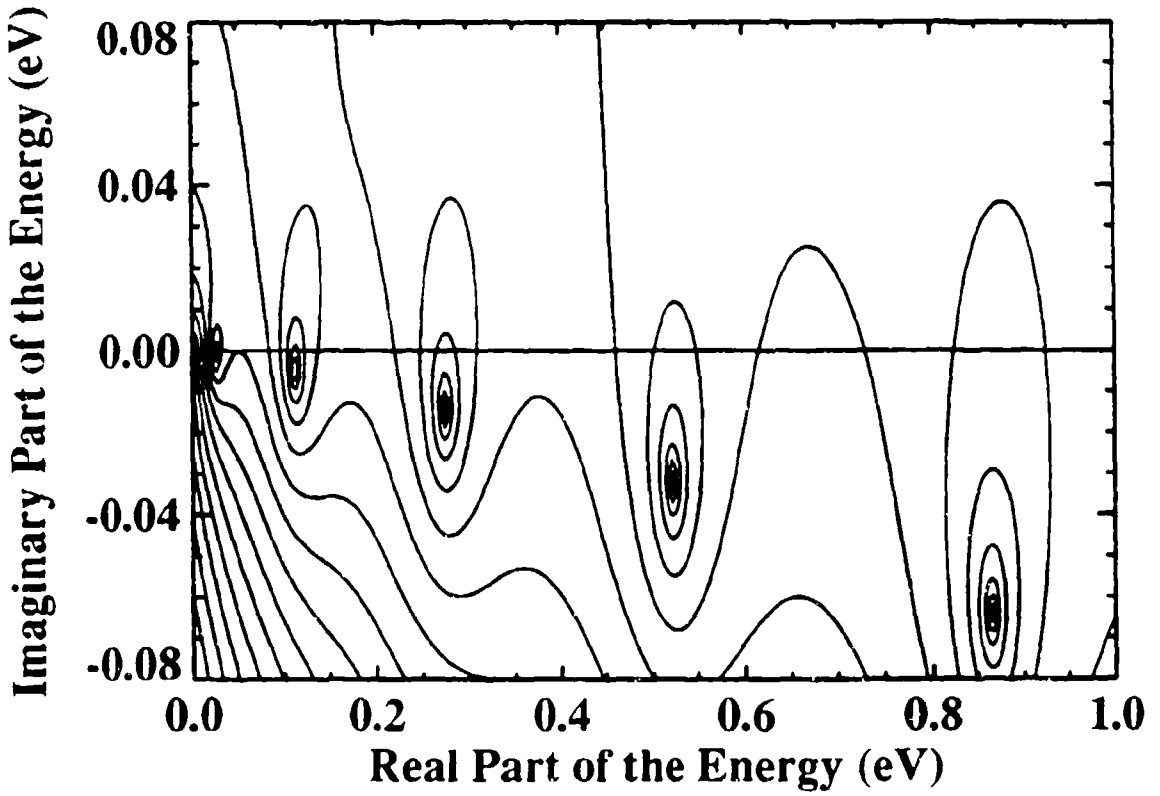
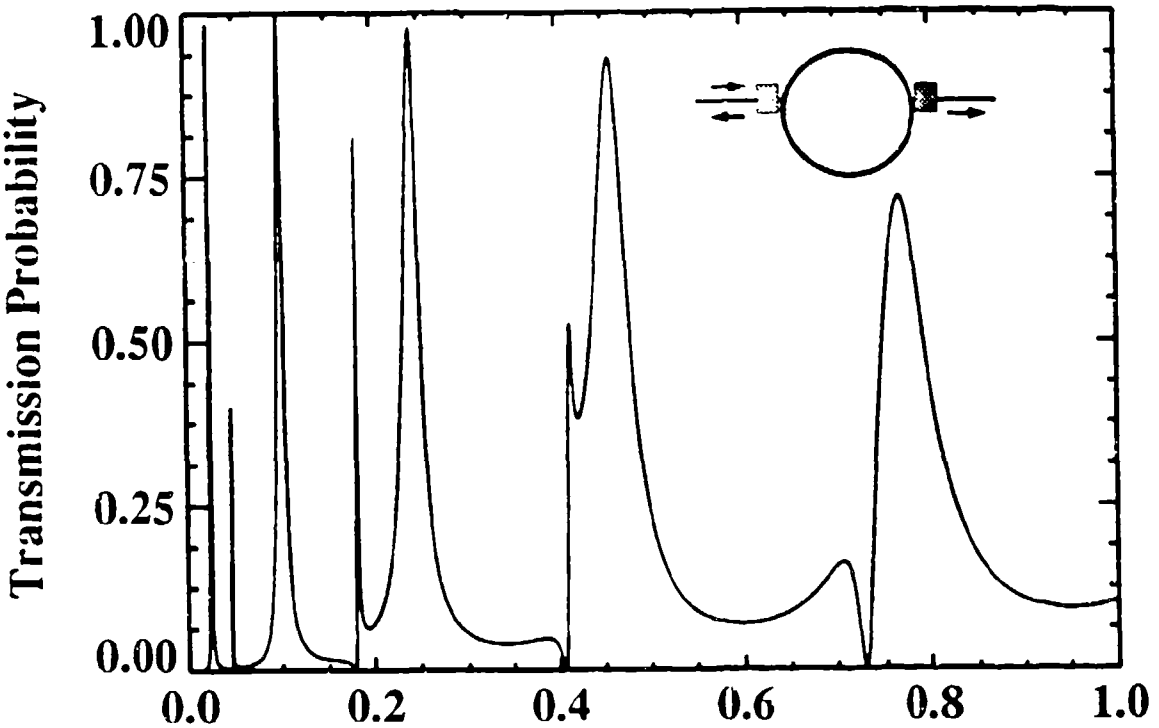

Imaginary Part of the Energy (eV)
0.08
0.04
0.00
-0.04
-0.08
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

Fig 22

(a)


Transmission Probability
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

(b)

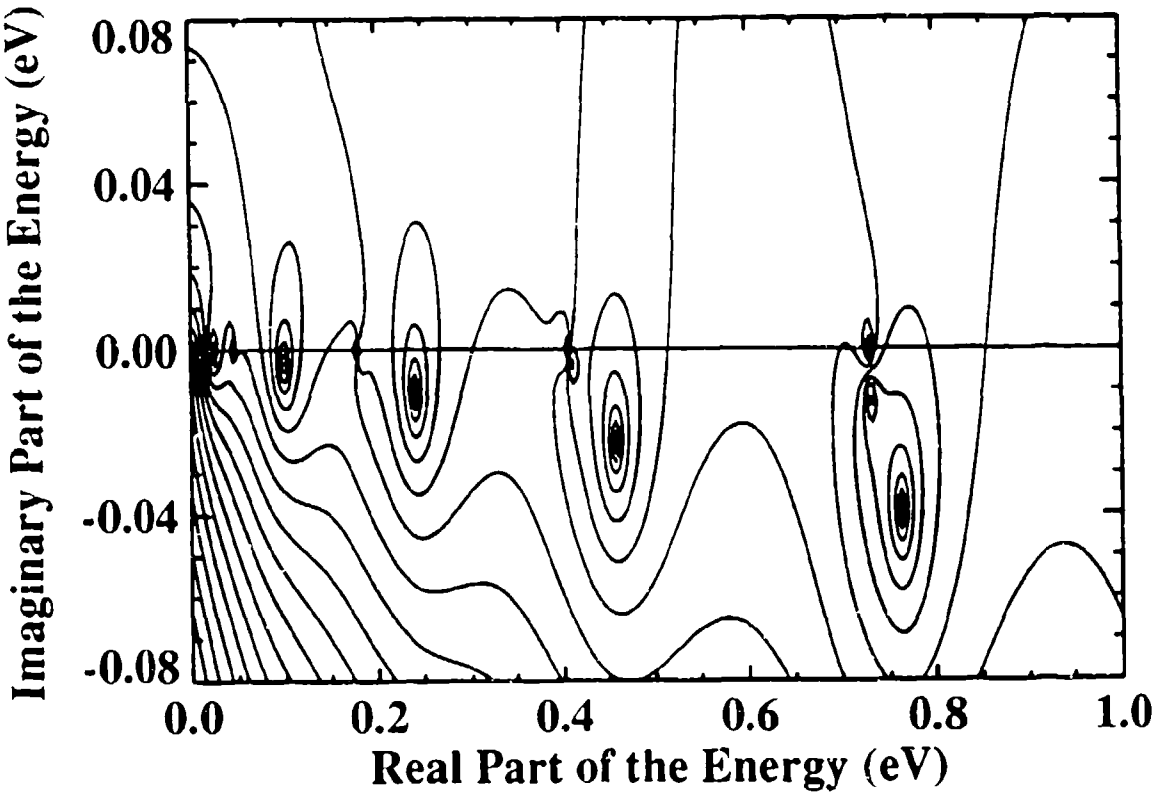

Imaginary Part of the Energy (eV)
0.08
0.04
0.00
-0.04
-0.08
0.0
0.2
0.4
0.6
0.8
1.0
Real Part of the Energy (eV)

Fig 24

Lateral *p–n* junctions between quasi-two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces

Henry K. Harbury and Wolfgang Porod
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Stephen M. Goodnick
Department of Electrical and Computer Engineering, Oregon State University, Corvallis, Oregon 97331

(Received 30 January 1992; accepted 5 March 1992)

We report on a numerical study of the potential profile and energy states of lateral *p–n* junctions at the GaAs/AlGaAs interface. The junctions arise from the amphoteric nature of Si doping during molecular-beam epitaxial growth on {100} versus {111} surfaces and have been previously realized experimentally through selective chemical etching. We find that the occurrence of a lateral *p–n* junction is sensitive to the doping of the overlayer and for Si doping concentrations less than 5×10^{17} cm^{-3} in $Al_{0.3}Ga_{0.7}As$, the *p–n* junction vanishes. We have studied the formation of a quantum wire in a V-groove structure and show that a one-dimensional system is in fact formed which may be controlled by the reverse bias applied between the *n* and *p* regions.

I. INTRODUCTION

During molecular-beam epitaxial (MBE) growth of GaAs and AlGaAs, Si doping is observed to be amphoteric depending on the substrate orientation. While growth on {100} surfaces invariably leads to *n*-type doping, Ballingall and Wood observed both *n*- and *p*-type electrical behavior on {110} surfaces depending on the growth temperature.[1] *p*-type behavior has been reported for {*N*11}A (*N* = 1,2,...) surfaces (Ga terminated) while *n* type is found for growth on GaAs {*N*11}B surfaces (As terminated).[2–4] This behavior in the case of {111}A surfaces may be understood from the preferential incorporation of Si onto As sublattice sites where Si behaves as an acceptor, rather than the more energetically favorable incorporation of Si into Ga sublattice sites (where it behaves as a donor) found in most cases.

Through the use of selective chemical etching, it is possible to produce {111}A facets on {100} surfaces. With epitaxial regrowth, one may then obtain regions of both *n*- and *p*-type doping on the same surface. This idea was used by Miller to fabricate lateral GaAs *p–n* junctions which demonstrated good diode *I–V* characteristics. Using Si-doped $Al_xGa_{1-x}As$ rather than GaAs during epitaxial regrowth, Ebner *et al.*[5] reported electroluminescence corresponding to the GaAs quantum well band gap, suggesting the existence of a quasi-two-dimensional (2D) *p–n* junction at the corrugated GaAs/AlGaAs interface. The ability to fabricate lateral *p–n* junctions between high mobility 2D hole and electron gases allows for a variety of novel complementary device structures. V-groove structures have been reported in the literature which result in quantum wire structures that may be employed, for example, in one-dimensional (1D) semiconductors.[6–8] By utilizing amphoteric Si in such a V-groove structure, a lateral *p–n–p* quantum wire may be realized.[9] This novel way of fabricating a quantum wire structure with lateral *n–p* junctions may offer some advantages compared to present unipolar structures which use metal gate electrostatic confinement or sidewall etching (see, e.g., Ref. 10).

We have theoretically investigated the properties of such quasi-2D *p–n* junctions for various layer structures and doping conditions. Using a 2D finite-element method, we determine the potential landscape, and the electron and hole charge densities at the interfaces by solving Poisson's equation within a semiclassical Thomas–Fermi screening model. Given a certain potential profile, we then solve Schrödinger's equation for the quantized electron and hole states at the heterointerfaces in one spatial dimension. We discuss the design of a novel quantum wire structure at the corner of an etched and overgrown V groove as mentioned above. Among the interesting properties of such a quantum wire structure, are the coexistence of quasi-2D electron *and* hole states next to a quasi-1D electron *or* hole system.

II. MODEL SYSTEM

The model geometry for the study of lateral *p–n* junctions is schematically depicted in Fig. 1. A 500 Å thick layer of Si-doped $Al_{0.3}Ga_{0.7}As$ is overgrown on a semi-insulating GaAs substrate with an etched V groove. The substrate terminates in a (100) plane, whereas the sidewalls of the V groove are members of the {111} family of planes. The angle between the (100) and (111)A plane is 54.74°. As shown in Fig. 1, *n*-type doping is present for the (100) plane, and *p*-type doping for the {111} planes. The composition and layer thicknesses are chosen to match those used by Ebner *et al.*[5] in fabricating a lateral light emitting structure. Also shown is the heterointerface between GaAs and AlGaAs. For suitable doping conditions, a 2D electron and hole gas (2DEG and 2DHG, respectively) forms along this interface.

We study the potential profile in equilibrium for this spatially 2D model geometry. Within a Thomas–Fermi

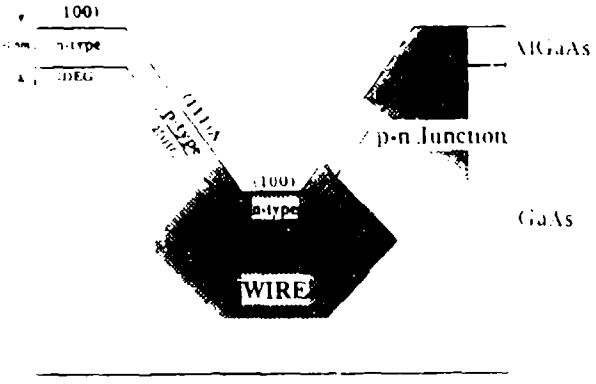


FIG. 1. Schematic diagram of the model geometry.

screening model, the electrostatic potential $\phi(\mathbf{r})$ determines the distribution of charge and, as a consequence, the band bending resulting from Poisson's equation,

$$\nabla^2\phi(\mathbf{r}) = \frac{e}{\epsilon\epsilon_0}[n(\mathbf{r}) - p(\mathbf{r}) + N_A^- - N_D^+] \quad (1)$$

The characteristics of the different materials enter through the doping densities and the band discontinuities at the heterointerfaces. The amphoteric Si-doping densities contribute N_A^- in the p-type layers and N_D^+ in the n-type layers. Complete ionization of the impurities is assumed. The electron and hole densities are denoted by $n(\mathbf{r})$ and $p(\mathbf{r})$, respectively. For degenerate statistics, the electronic charge density is given by the Fermi–Dirac integral of order $\frac{1}{2}$, $n(\mathbf{r}) = N_c(\pi/2)F_{1/2}(\eta)$, where η is the energy separation between the local conduction band edge $E_c(\mathbf{r})$ and the Fermi level E_f measured in units of k_BT, $\eta = -(E_c - E_f)/k_BT$. A similar relationship applies for the hole density. All our calculations are performed for room temperature.

We also study the quantum-confined electronic states at the heterointerfaces by solving the Schrodinger equation for a given potential landscape. We are particularly interested in the formation of 2DEGs and 2DHGs at the interfaces between the substrate and the overgrown layer. To this end, we solve Schrodinger's equation in the direction perpendicular to the heterointerfaces for the potential which we obtain from Poisson's equation,

$$-\frac{\hbar^2}{2m^*}\nabla^2\psi(\mathbf{r}) + [V(\mathbf{r}) - e\phi(\mathbf{r})]\psi(\mathbf{r}) = E\psi(\mathbf{r}). \quad (2)$$

Here, $V(\mathbf{r})$ includes external potentials and band offsets, and $\phi(\mathbf{r})$ is the electrostatic potential obtained from Poisson's equation.

III. NUMERICAL METHOD

The determination of the 2D conduction and valence band potential distribution is a challenging numerical problem, because of the complicated geometry and the different spatial scales in the problem. The size of the fabricated structure is on the order of several microns, while near the heterointerface the potential profile changes rapidly over a few nanometers. This requires a highly nonuniform grid which is coarse in the bulk and very fine close to the interfaces and in the overgrown layer. This nonuniform mesh is also designed to preserve the angle between crystallographic directions throughout the structure.

We use the finite element method and Newton–Raphson iteration to solve the nonlinear Poisson equation within a Thomas–Fermi screening model. Von-Neuman and Dirichlet boundary conditions are used at the edge of the problem domain. Far inside the substrate, the potential is fixed at a value set by the background doping and by the requirement of space charge neutrality in the bulk. The potential at the $Al_{0.3}Ga_{0.7}As$ (100) and (111) surfaces is assumed to be pinned due to surface states at the usual value of 0.8 eV (separation between the conduction band edge and the Fermi level). Dirichlet boundary conditions at the left and right sides of the solution domain force the potential to be flat, which simulates invariance along the heterointerfaces far from the p–n junctions. Our domain is a nonuniform triangular mesh with first order finite element Chapeau basis functions. We typically use a mesh of approximately 8500 nodes in which the largest element far from the heterointerface has an area of 1.2×10^{-3} Å^2, and the smallest element close to the heterointerface has an area of 70 Å^2.

The finite element method results in a large system of linear equations for the unknown potential at each nodal point. We employ band width optimization and sparse matrix methods for an efficient numerical solution. A standard LU decomposition method is used to solve the linear system after it has been reduced to skyline band-symmetric form. Most numerical computations are performed on a Convex C-2. The solution of Poisson's equation typically takes 925 s of CPU time.

We also solve Schrodinger's equation for the resulting potential profiles. Solutions are found for the 1D problem perpendicular to the heterointerfaces. This gives us information regarding quantum confinement in the 2DEG and 2DHGs.

IV. LATERAL 2D p–n JUNCTION

We now concentrate on the portion of our model structure which is shaded and labeled "p–n Junction" in the schematic drawing of Fig. 1. Depending upon the doping densities in the overgrown AlGaAs layer, the (100) and (111) heterointerfaces may display 2D electron and/or hole systems, respectively. A lateral p–n junction will form if the Si-doping density is high enough to produce, at the same time, 2D electron and hole systems along the heterointerfaces.

Figure 2 shows the band profiles at the n side of the p–n junction in the direction perpendicular to the heterointerface for several values of the Si-doping density. Shown are the conduction and valence band edges, and the Fermi level is displayed as the dotted line. As can be seen, a 2D electronic system forms for doping densities higher than $Al_{0.3}Ga_{0.7}As$. Figure 3 shows the corresponding informa-

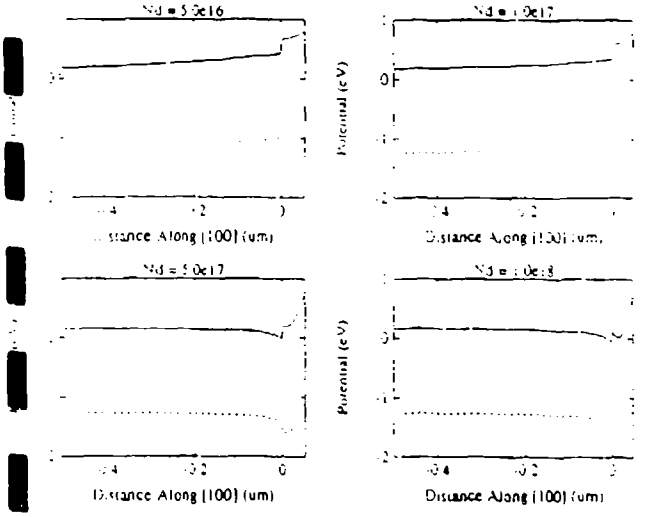


FIG. 2. Band diagrams on the n side of the lateral p–n junction for several doping concentrations. The solid line represents the conduction band, the dashed line the valence band, and the dotted line the Fermi level.

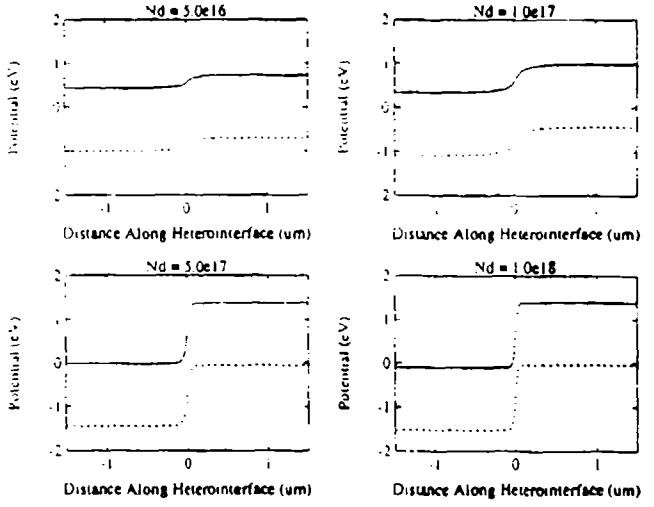
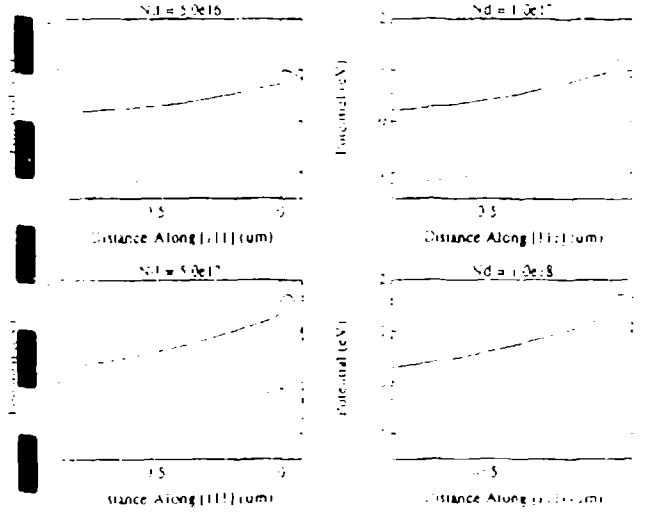


FIG. 4. Band diagrams along the p–n junction for several doping concentrations. The solid line represents the conduction band, the dashed line the valence band, and the dotted line the Fermi level.

...n for holes. The potential profiles are now plotted in a direction perpendicular to a {111} plane. Again, a 2D hole system emerges for doping densities in excess of $\sim 10^{17}$ cm^{-3}. Figures 2 and 3 illustrate the existence of quasi-2D carriers far from the lateral p–n junction. The band diagrams in Figs. 2 and 3 are shown at a distance of about 2 μm from the junction which is sufficiently far to ensure invariance of these profiles with distance.

Figure 4 demonstrates the lateral p–n junction. Figure 4 shows the conduction and valence band edges in a direction along the (100) and (111) heterointerfaces, respectively. The transition from p- to n-type doping occurs at $x = 0$. Negative values of the distance x correspond to the (111) interface, and positive values to the (100) interface. The emergence of a p–n junction is clearly visible as the doping density increases to levels greater than 5×10^{17} cm^{-3}. The transition from p- to n-type behavior occurs over a distance of about 0.25 μm.

We also determined the energies of the quantum-confined states at the n side by solving Schrödinger's equation for the band profiles shown in Fig. 2. For a doping concentration of 1×10^{18} cm^{-3}, we find the following values of the energies for the lowest three eigenstates which are localized in the channel: 0.034, 0.060, and 0.084 eV.



...Band diagrams on the p side of the lateral p–n junction for several ...ng concentrations. The solid line represents the conduction band, the ...shed line the valence band, and the dotted line the Fermi level.

V. LATERAL V-GROOVE QUANTUM WIRE STRUCTURE

As discussed earlier, a V-groove etch may be used to define a quantum wire structure based on the orientation-dependent properties of Si during MBE growth. We consider the structure shown in Fig. 1 where a notch has been formed using an anisotropic etch leaving {111}A surfaces on the sidewalls, and a (100) at the bottom of the trench. The $Al_{0.3}Ga_{0.7}As$ overlayer is assumed to be doped alternately n and p type (compare Fig. 1).

The solution domain for the investigation of the quantum wire is shown in Fig. 1 by the hatched area which is labeled "WIRE." Figure 5 shows the mesh used for the

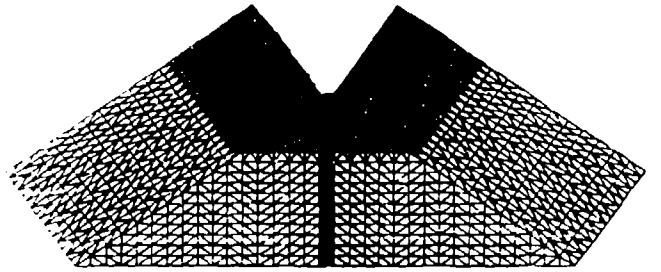


FIG. 5. The mesh used for the calculation of the potential defining the quantum wire structure.

FIG. 6. The potential surface for a doping concentration of 1×10^{18} cm^{-3}. Note the strong bending which supports a quantum wire at the p–n–p junction.

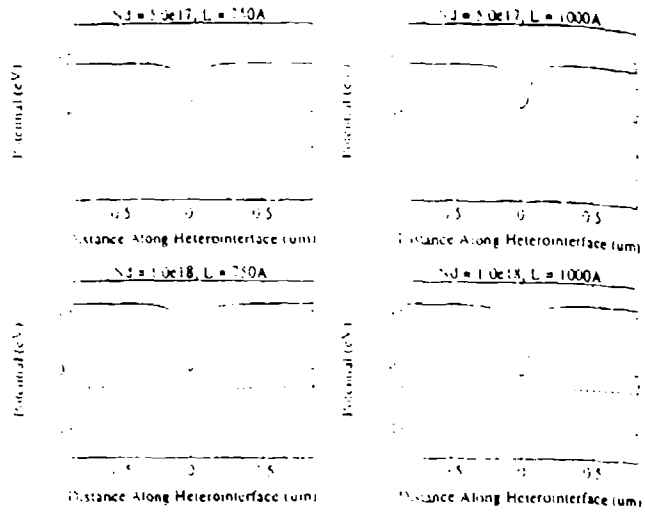


FIG. 8. Band diagrams along the p–n–p junction for several doping concentrations and lengths of the n-type section. The solid line represents the conduction band, the dashed line the valence band, and the dotted line the Fermi level.

calculation of the potential landscape. Note the extreme detail close to the heterointerfaces. This mesh contains 8500 nodal points.

The resulting potential surface for a doping concentration of 1×10^{18} cm^{-3} is displayed in Fig. 6. Note the strong band bending at the center which forms the wire and the weak potential variation in the bulk GaAs region. A contour plot of the same potential landscape is shown in Fig. 7 which shows a close-up view of the center n-type region.

Figure 8 shows the band diagram for the p–n–p junction for various lengths of the (100) n-type region (denoted L) and doping concentrations. The origin of the horizontal axis, which denotes distance, is chosen to be in the center of the n-type region. Note that a wire forms for a length of 1000 Å, since the conduction band approaches the Fermi level. If the central n-type region is shortened, the wires disappears due to side depletion from the adjacent p-type regions. This is shown to occur for a length of $L = 750$ Å.

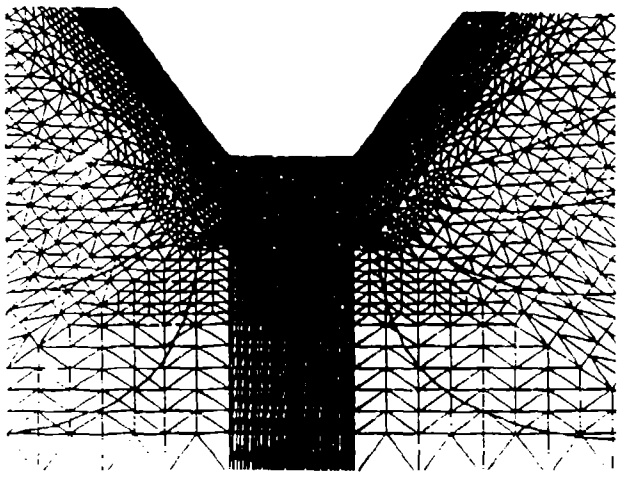


FIG. 7. A contour plot of the potential shown in Fig. 6.

VI. CONCLUSIONS

We have demonstrated the existence of lateral p–n junctions on corrugated GaAs/AlGaAs interfaces. The n- and p-type behavior is made possible by the amphoteric nature of Si dopants as a function of crystallographic plane orientations. At V-grooves etched into a semi-insulating GaAs substrate, a p–n junction forms at the intersection of (100) and {111} planes if the doping in the overlayer exceeds 5×10^{17} cm^{-3}. We have also demonstrated the possibility of basing this system for the design of quantum wires which may form at the bottom of the V groove. We find that a quantum wire is to be expected in this p–n–p structure for a length of the n-type region on the order of 1000 Å and a doping density of 1×10^{18} cm^{-3}.

ACKNOWLEDGMENTS

The authors would like to acknowledge fruitful discussions with J. Ebner, T. K. Plant, and C. S. Lent regarding this work. The authors would also like to acknowledge the support of A. Churchill for the preparation of the manuscript. This work has been partially supported by the Office of Naval Research and the Air Force Office of Scientific Research.

[1] J. M. Ballingall and C. E. C. Wood, Appl. Phys. Lett. 41, 947 (1982).
[2] W. I. Wang, E. E. Mendez, T. S. Duan, and L. Esaki, Appl. Phys. Lett. 47, 326 (1985).

[1] D. L. Miller, Appl. Phys. Lett. 47, 1309 (1985).
[2] S. Subbanna, H. Kroemer, and J. L. Merz, J. Appl. Phys. 59, 488 (1986).
[3] J. Ebner, J. Larry, G. Eliason, and T. K. Plant, in Proc. ESSDERC 90, (1990 Unpublished), p. 401.
[4] D. L. Miller, A. Shahar, U. Koren, and P. J. Corvini, Appl. Phys. Lett. 54, [illegible] (1989).
[5] E. Kapon, D. M. Hwang, and R. Bhat, Phys. Rev. Lett. 63, 430 (1989); and E. Kapon, S. Simhony, R. Bhat, and D. M. Hwang, Appl. Phys. Lett. 55, 2715 (1989).
[6] E. Colas, S. Simhony, E. Kapon, R. Bhat, D. M. Hwang, and P. S. D. Lin, Appl. Phys. Lett. 57, 914 (1990).
[7] Private communication.
[8] *Nanostructure Physics and Fabrication*, edited by M. A. Reed and W. P. Kirk (Academic, San Diego, 1989).

Lateral *p-n* junctions and quantum wires formed by quasi two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces

Wolfgang Porod and Henry K. Harbury
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Stephen M. Goodnick
Department of Electrical and Computer Engineering, Oregon State University, Corvallis, Oregon 97331

(Received 13 April 1992; accepted for publication 4 August 1992)

We report the results of modeling lateral *p-n* junctions and *p-n-p* quantum wire structures at corrugated GaAs/AlGaAs interfaces, using the surface orientation dependent amphoteric nature of Si doping. We determine the potential landscape and the electron and hole charge densities within a semiclassical Thomas–Fermi screening model, and then solve the two-dimensional Schrödinger equation using finite elements for the quantized electron and hole states at the heterointerfaces. We demonstrate the formation of a one-dimensional electron system confined between two lateral *p-n* junctions, and discuss the advantages of this structure compared to conventional electrostatic confinement schemes for fabricating quantum wires.

Recent studies[1-4] have shown that Si acts as an amphoteric dopant depending on substrate orientation during molecular beam epitaxial growth of GaAs and AlGaAs. In particular, growth on {100} surfaces invariably leads to *n*-type doping, while *p*-type behavior has been reported for Ga-terminated {111}A surfaces. These findings have led to the investigation of lateral *p-n* junctions at selectively etched and epitaxially regrown V grooves consisting of {111}A facets on {100} surfaces. Miller[3] demonstrated the feasibility of lateral GaAs *p-n* junctions which exhibited good diode current-voltage characteristics. By using Si-doped $Al_{0.3}Ga_{0.7}As$ rather than GaAs during epitaxial regrowth, one may also realize a lateral *p-n* junction between quasi-two-dimensional electron and hole systems at the corrugated GaAs/AlGaAs interface. Ebner *et al.*[5] demonstrated electroluminescence corresponding to the GaAs quantum well band gap in such a system, and Harbury *et al.*[6] reported calculations which confirmed the existence of a lateral *p-n* junction between two-dimensional electron and hole gas systems for Si doping densities in excess of 5×10^{17} cm^{-3}.

The ability to fabricate lateral *p-n* junctions between high mobility two-dimensional hole and electron gases allows for a variety of novel complementary device structures. In previous studies, V-grooves have been employed for the design of quantum wires.[7-10] By utilizing amphoteric Si in such a V-groove structure, a novel quantum wire system may be realized at lateral *p-n-p* junctions.[11] This new way of fabricating a quantum wire with lateral *p-n* junctions[12] may offer some advantages compared to present unipolar structures[13] which use metal gate electrostatic confinement or sidewall etching.

In this letter, we report our results of modeling both lateral *p-n* junctions and lateral *p-n-p* quantum wire structures at corrugated GaAs/AlGaAs interfaces. We determine the potential landscape and the electron and hole charge densities by solving Poisson's equation within a semiclassical Thomas–Fermi screening model. Given a certain potential profile, we then solve Schrodinger's equation for the quantized states at the heterointerfaces. Among the interesting properties of such structures are the coexistence of quasi-two-dimensional electron *and* hole states next to a quasi-one-dimensional electron *or* hole system.

Figure 1 shows a schematic drawing of the model corrugated GaAs/AlGaAs interface. The composition and layer thicknesses are similar to those used by Ebner *et al.*[5] in fabricating a lateral light emitting structure. A 50 nm thick layer of Si-doped $Al_{0.3}Ga_{0.7}As$ is overgrown on a semi-insulating GaAs substrate with an etched V-groove. The substrate terminates in a (100) plane, whereas the sidewalls of the V-groove are members of the {111}A family of planes. As indicated in Fig. 1, *n*-type doping is present for the (100) layers, and *p*-type doping for the (111) layers.[14] Also shown is the heterointerface between GaAs and AlGaAs. For suitable doping conditions, a two-dimensional electron and hole gas (2DEG and 2DHG, respectively) forms along this interface.

We use the finite element method and Newton–Raphson iteration to solve the nonlinear Poisson equation within a Thomas–Fermi screening model for room temperature. We assume surface pinning at the $Al_{0.3}Ga_{0.7}As$ (100) and (111) facets (0.8 eV separation between the conduction band edge and the Fermi level), and *n*-type background doping in the bulk of 1×10^{15} cm^{-3}. The determination of the two-dimensional conduction and valence band edges is a challenging numerical problem which requires a highly nonuniform mesh because of the complicated geometry and the different doping densities. The resulting large system of linear equations is solved utilizing band width optimized sparse matrix methods. We also study the quantum-confined electronic states at the heterointerfaces by obtaining solutions of the two-dimensional Schrodinger equation for the calculated potential landscape.

We now concentrate on the portion of our model structure which is shaded and labeled "*p-n* Junction" in the schematic drawing of Fig. 1. Depending upon the doping densities in the overgrown AlGaAs layer, the (100) and (111) heterointerfaces may induce two-dimensional electron and/or hole systems, respectively. A lateral *p-n* junc-

 0003-6951/92/401823-03$03.00

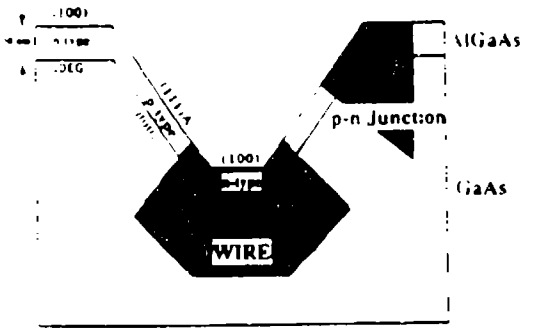


FIG. 1. Schematic diagram of the model geometry.

tion will form if the Si-doping density is high enough to simultaneously produce two-dimensional electron and hole systems along the heterointerfaces. Figure 2 shows, for several values of the Si-doping density in the overlayer, the calculated conduction band profiles at the *n*-side of the *p-n* junction in the direction *perpendicular* to the (100) heterointerface; the semiclassical Fermi level defines the zero of energy and is shown by the dashed line. The inset displays the conduction band minimum at the GaAs side of the heterointerface, and a quasi-two-dimensional electron gas forms for doping densities higher than 5×10^{17} cm^{-3}. Similar behavior is found for the valence bands in the direction *perpendicular* to a (111) interface. Figure 3 demonstrates the existence of lateral *p-n* junctions between 2DEGs and 2DHGs for two values of the overlayer doping density. Plotted are the conduction and valence band edges in a direction *parallel* to the (111) and (100) heterointerfaces, respectively, as schematically shown in the inset. Negative values of the distance correspond to the (111) interface, and positive values to the (100) interface. The emergence of a *p-n* junction is clearly visible, and the transition from *p*- to *n*-type behavior occurs over a distance of about 100 nm.

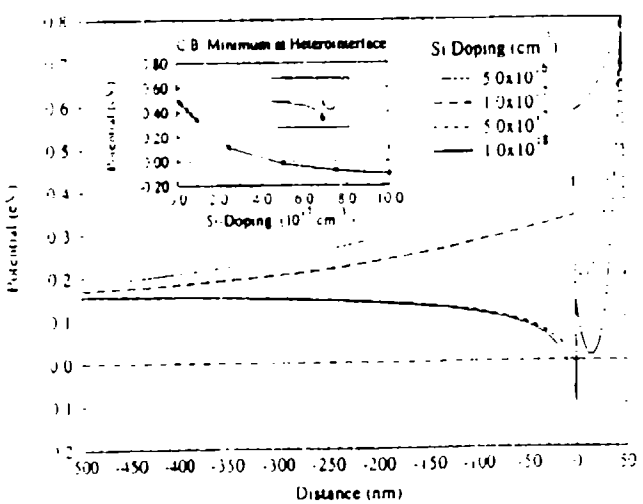


FIG. 2. Conduction band profile *perpendicular* to the (100) interface for different values of the Si-doping density in the overlayer. The inset shows the variation with doping of the conduction band minimum at the heterointerface.

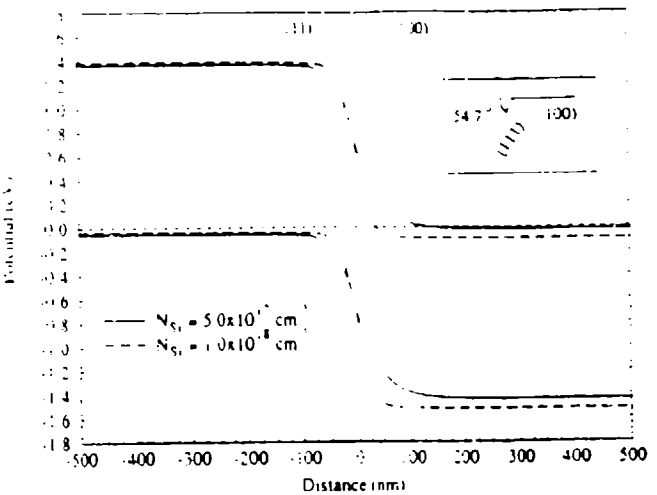


FIG. 3. Band diagram for the lateral *p-n* junction. Shown are the conduction and valence bands (for two values of the doping density) along the (111) and (100) interfaces, as indicated in the inset.

At the base of the V groove, a narrow (100) *n*-type region exists between two (111) *p*-type regions which results in lateral potential confinement in addition to that of the heterojunction. We have investigated the formation of a 1D quantum wire in the *n*-region by solving Poisson's and Schrödinger's equations in the solution domain shown in Fig. 1 by the hatched area which is labeled "WIRE." Figure 4 gives the band diagram for the *p-n-p* junction, where the center *n*-type section is modeled with a length of $L = 100$ nm and the overlayer doping is chosen to be 1×10^{18} cm^{-3}. The conduction and valence band edges are shown at the GaAs side of the junction and in a direction *parallel* to the heterointerfaces. Note that an accumulation of electrons occurs at the center of the *n*-type (100) section where the conduction band moves below the semiclassical Fermi level, which is again chosen to be the zero of energy

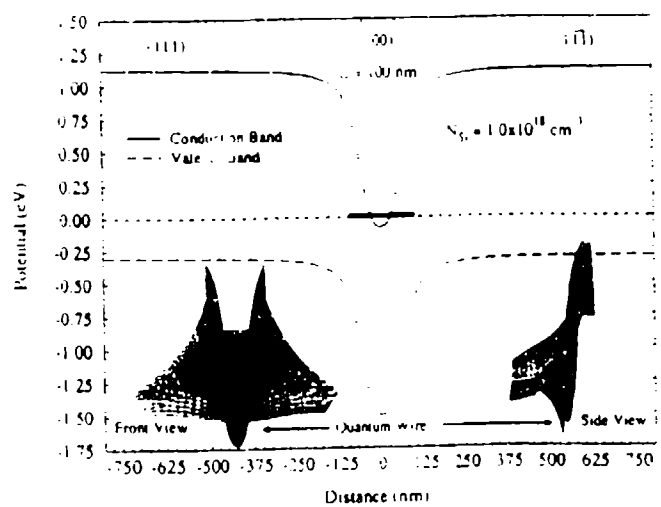


FIG. 4. Band diagram for the lateral *p-n-p* structure. Shown are the conduction and valence bands along the heterointerfaces for a 100 nm long center *n*-type section and a Si-doping density of 1×10^{18} cm^{-3}. The insets show front and side views of the two-dimensional conduction band profile. The potential "pocket" which holds the quantum wire is also indicated.

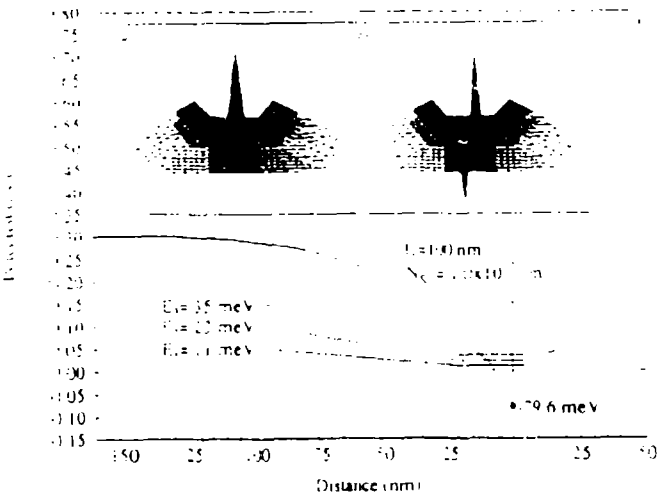


FIG. 5. Potential variation *perpendicular* to the heterointerface (at the center of the *n*-type region) for the two-dimensional conduction band profile shown in Fig. 4. The insets show the wave functions for the lowest two quantum wire states, which are confined at the GaAs side of the heterointerface (the darker top portion of the mesh corresponds to the higher-density AlGaAs overlayer).

and shown as the dashed line. The lateral *p-n* junctions on both sides confine the accumulated electrons in the direction parallel to the heterointerface. A better perspective of the two-dimensional potential variation can be gained by the insets which present front and side views of the two-dimensional conduction band profile. Note the dip in the center which defines the "pocket" holding the electrons.

Figure 5 shows the variation of the conduction band in the direction *perpendicular* to the heterointerface at the center of the *n*-type (100) region. The sharp dip below the semiclassical Fermi level is now clearly visible. The corresponding solutions to the two-dimensional Schrodinger equation for the ground and first excited one-dimensional subbands are shown by the insets in Fig. 5. The peak of the envelope function clearly lies in the GaAs side of the heterojunction localized between the two *p*-type regions. (The darker top portion corresponds to the denser mesh in the AlGaAs overlayer.) The confinement energies relative to the semiclassical Fermi energy are shown schematically on the band diagram for the first three levels. The energies shown all correspond to states originating from the lowest two-dimensional subband energy of the heterojunction itself. The spacing of the levels, here on the order of 10 meV, depends on the doping and the width of the *n*-region at the base of the V-groove. As the width is reduced, the subband separation should increase, although limitations are imposed by the lateral extent of the depletion region which eventually results in complete depletion of the wire if made too narrow.

The results above demonstrate the feasibility of realizing a quantum wire structure on a corrugated GaAs/AlGaAs surface. There are several possible advantages of such a structure compared to the current state of the art which relies on sidewall etching or electrostatic confinement from Schottky contacts on the surface of the AlGaAs. First, the actual definition of the width of the *n*-region is defined by the etch time of the anisotropic etch through an optically defined photoresist mask. Thus, electron or ion-beam lithography is not required to fabricate the wire structure. The bipolar nature of the structure may be utilized to inject minority carriers into the wire structure from the *p*-regions, which may lead to some interesting device applications. Also, if one additionally provided separate contacts to the *p*- and *n*-type regions using nanolithography, control of the wire transport properties could be achieved both by varying the width of the confinement (through reverse biasing the *p-n* junctions), and through varying the electron Fermi energy via a gate contact on the AlGaAs above the electron channel.

In summary, we have demonstrated the existence of lateral *p-n* junctions between 2DEGs and 2DHGs on corrugated GaAs/AlGaAs interfaces, which is made possible by the amphoteric nature of Si-dopants as a function of crystallographic plane orientations. At V-grooves etched into a semi-insulating GaAs substrate, a *p-n* junction forms at the intersection of (100) and (111) planes if the doping in the overlayer exceeds 5×10^{17} cm^{-3}. We have also demonstrated the possibility of basing this system for the design of quantum wires which may form at the bottom of the V groove. We find that a quantum wire is to be expected in this *p-n-p* structure for a length of the *n*-type region on the order of 100 nm and a Si-doping density of 1×10^{18} cm^{-3}.

The authors would like to acknowledge fruitful discussions with J. Ebner, D. J. Kirner, C. S. Lent, J. L. Merz, M. Mueller, and T. K. Plant. This work has been partially supported by the Office of Naval Research and the Air Force Office of Scientific Research.

[1] J. M. Ballingall and C. E. C. Wood, Appl. Phys. Lett. **41**, 947 (1982).
[2] W. I. Wang, E. E. Mendez, T. S. Kuan, and L. Esaki, Appl. Phys. Lett. **47**, 826 (1985).
[3] D. L. Miller, Appl. Phys. Lett. **47**, 1309 (1985).
[4] S. Subbanna, H. Kroemer, and J. L. Merz, J. Appl. Phys. **59**, 488 (1986).
[5] J. Ebner, J. E. Lary, G. W. Eliason, and T. K. Plant, *Proceedings of the 20th European Solid State Device Research Conference, ESSDERC 90*, edited by W. Eccleston and P. J. Rossner (Adam Hilger, Bristol, 1990), pp. 401–404.
[6] H. K. Harbury, W. Porod, and S. M. Goodnick, presented at the 19th International Conference on the Physics and Chemistry of Semiconductor Interfaces, Death Valley, CA, January 1992.
[7] B. I. Miller, A. Shahar, U. Koren, and P. J. Corvini, Appl. Phys. Lett. **54**, 188 (1989).
[8] E. Kapon, D. M. Hwang, and R. Bhat, Phys. Rev. Lett. **63**, 430 (1989); and E. Kapon, S. Simhony, R. Bhat, and D. M. Hwang, Appl. Phys. Lett. **55**, 2715 (1989).
[9] E. Colas, S. Simhony, E. Kapon, R. Bhat, D. M. Hwang, and P. S. D. Lin, Appl. Phys. Lett. **57**, 914 (1990).
[10] H. K. Harbury and W. Porod, J. Vac. Sci. Technol. B **8**, 923 (1990).
[11] J. Ebner (private communication).
[12] The use of *p-n* junctions for lateral confinement of quantum wires was pioneered by C. C. Dean and M. Pepper, J. Phys. C **15**, L1287 (1982), and by A. B. Fowler, A. Hartstein, and R. A. Webb, Phys. Rev. Lett. **48**, 196 (1982).
[13] *Nanostructure Physics and Fabrication*, M. A. Reed and W. P. Kirk, Eds. (Academic, San Diego, 1989).
[14] The details of the amphoteric nature of the silicon doping near the intersection of the (100) and (111) surfaces are not known, and likely to be more complicated than our model. This and possible compensation effects might tend to reduce the abruptness of the lateral *p-n* junction.

A novel quantum wire formed by lateral *p–n–p* junctions between quasi-two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces

Henry K. Harbury and Wolfgang Porod
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Stephen M. Goodnick
Department of Electrical and Computer Engineering, Oregon State University, Corvallis, Oregon 97331

(Received 6 August 1992; accepted for publication 19 October 1992)

Numerical modeling of a novel quantum wire structure formed by the confinement of electrons between lateral quasi-two-dimensional (Q2-D) *p–n* junctions in a corrugated GaAs/AlGaAs heterostructure is reported on. Such a quantum wire may be realized at the tip of a Si-doped AlGaAs overgrown *V* groove in a SI–GaAs substrate due to the surface orientation dependence of Si doping. The two-dimensional conduction and valence band potential profiles for the electron and hole charge densities are solved within a semiclassical Thomas–Fermi screening model. The quantized electronic wire states at the heterointerface are then obtained by solving the two-dimensional effective mass Schrödinger equation using the calculated potential profile. The parameter space of the one-dimensional electron system is explored to establish which features of the structure are dominant factors in controlling the electronic states. It is demonstrated that the energy level spacing of the quantum wire depends primarily on the lateral confinement width in the *n*-type region at the tip of the *V* groove. The ground state energy of the wire is shown to depend on both the lateral confinement width and the vertical heterointerface confinement width. The results of our initial calculations are also reported on to incorporate lateral gates on the surface to obtain direct control of the quantum wire transport properties. The advantages of fabricating quantum wires with this structure compared to conventional methods of electrostatic confinement are discussed.

I. INTRODUCTION

Recent experiments have demonstrated that Si behaves as a substrate orientation dependent amphoteric dopant in molecular beam epitaxial (MBE) grown GaAs and AlGaAs. This effect was first studied on As-terminated {111} and {110} planes; then further studies were reported on Ga-terminated {111}A, As-terminated {111}B, {100}, {211}, and higher index planes.[1–3] These studies show that, under suitable growth conditions, Si-doped overlayers grown on {100} GaAs substrate surfaces have donor behavior, whereas Si-doped overlayers grown on Ga-terminated {111}A GaAs substrate surfaces have acceptor behavior.

The controllable amphoteric nature of Si doping incited the investigation of fabricating lateral *p–n* junctions in GaAs. Bulk GaAs lateral *p–n* junctions with good diode current-voltage characteristics were reported by Miller in 1985.[4] HEMT-compatible heterostructure lateral *p–n* junctions, between quasi-two-dimensional electron and hole systems, were reported by Ebner *et al.* in 1990.[5] In this latter study, Si-doped $Al_{0.3}Ga_{0.7}As$ was epitaxially regrown on a semi-insulating corrugated GaAs substrate. The selectively etched and epitaxially regrown structure was fabricated with exposed Ga-terminated {111}A facets on the normally exposed {100} substrate. Ebner *et al.*[5] reported electroluminescence results that correspond to the GaAs quantum well band gap of such a system, and Harbury *et al.*[6] reported calculations that confirm the existence of a lateral *p–n* junction between two-dimensional electron and hole gas systems (2DEG and 2DHG, respectively) with Si-doping densities in excess of 5×10^{17} cm^{-3}.

A variety of novel complementary device structures may be based on lateral *p–n* junctions formed between high mobility 2DEG and 2DHG systems. In particular, exploiting the amphoteric dopant behavior of Si in a selectively etched and epitaxially regrown *V* groove structure presents a novel quantum wire system realized between lateral *p–n* junctions.[7] The depletion regions in such a HEMT-compatible *p–n–p*-type geometry provide lateral confinement of the carriers at the heterointerface of the *n*-type center region, thereby realizing a quantum wire. The use of *p–n* junctions for lateral confinement of quantum wires was pioneered by Dean and Pepper[8] and by Fowler and coworkers.[9] *V*-groove geometries have been previously explored for the design of quantum wires.[10–13] This new design, however, may offer some advantages compared to present unipolar structures,[14] which use metal gate electrostatic confinement or sidewall etching.

In this paper we report on numerical modeling of the lateral *p–n–p* quantum wire system realized in a corrugated GaAs/AlGaAs heterostructure. In Sec. II we present our model of the system, and in Sec. III we discuss our solution methods for obtaining the two-dimensional electrostatic potential for electrons and holes, and the electronic states of the quantum wire. We also present the details of the numerical method in the Appendix. Our results are presented in Sec. IV, and we conclude in Sec. V.

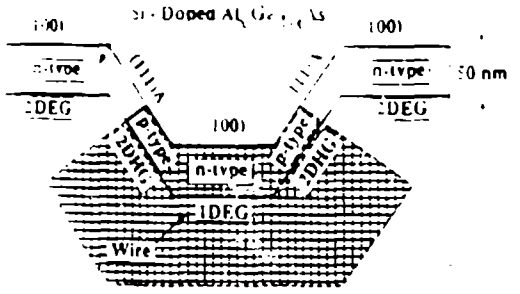


FIG. 1. Schematic diagram of the model V-groove corrugated heterostructure geometry. The n-type or p-type nature of the amphoteric silicon doped overlayer is labeled along with the corresponding crystallographic surface orientation. Possible formation of 2DEG's, 2DHG's, and a 1DEG is also noted. The hatched region represents the computation domain.

II. MODEL SYSTEM

Figure 1 shows a schematic diagram of the model corrugated GaAs/AlGaAs heterojunction system. The chosen overlayer thickness and composition are similar to those of a lateral light emitting structure fabricated by Ebner *et al.*[5] A 50 nm thick layer of Si-doped $Al_{0.3}Ga_{0.7}As$ is overgrown on an etched V groove of a semi-insulating GaAs substrate. The sidewalls of the V-groove expose the {111}A family of planes of the normally {100} terminated substrate. As indicated in Fig. 1, the amphoteric Si-doped overlayer exhibits n-type behavior on the {100} surfaces and p-type behavior on the {111}A surfaces. The two-dimensional electron and hole gas shown in Fig. 1 (labeled 2DEG and 2DHG, respectively) forms along the heterointerface for suitable doping conditions.

The hatched region in Fig. 1 represents the finite element calculation domain used to model the p–n–p quantum wire structures. The range of the ternary composition, the doping density, and the structural dimensions are the parameters that define the system. A two-dimensional equilibrium potential profile of the model geometry is sought for different material parameters. The band bending due to the electrostatic potential, $\phi(r)$, is obtained from the solution of Poisson's equation within a semiclassical Thomas–Fermi screening model,

$$\nabla^2\phi(r)=\frac{e}{\epsilon\epsilon_0}\left(n(r)-p(r)+N_A^-(r)-N_D^+(r)\right). \tag{1}$$

Complete ionization of the impurities is assumed at room temperature, such that the background doping of the semi-insulating GaAs, assumed to be lightly n type, contributes to N_D^+ in the substrate, and the amphoteric Si doping contributes to both N_A^- and N_D^+ in the overlayer region. The electron density, $n(r)$, is given by the Fermi–Dirac integral of order $\frac{1}{2}$,[15]

$$n(r)=4\pi\left(\frac{2k_BTm_e^*}{h^2}\right)^{3/2}\int_{E_c(r)}^{\infty}\frac{(E-E_c(r))^{1/2}\,dE/k_BT}{1+\exp((E-E_F)/k_BT)}$$

$$=N_cF_{1/2}(\eta), \tag{2}$$

where $N_c=2(2\pi k_BTm_e^*/h^2)^{3/2}$ is the effective conduction band density of states, $\eta=(E_F-E_c(r))/k_BT$ measures the separation between the Fermi level, E_F, and the position-dependent conduction band edge, $E_c(r)$, and $F_{1/2}(\eta)$ is the Fermi–Dirac integral of order 1/2. A similar term is obtained for the hole density, $p(r)=N_vF_{1/2}(\eta)$, where η is now the separation between the valence band edge and the Fermi level, and N_v is the effective valence band density of states.

We enforce bulk charge neutrality deep inside the semi-insulating GaAs substrate, assumed to be slightly n type with a background doping density of 1.0×10^{15} cm^{-3} which is completely ionized. A density of surface states along the {100} and {111} exposed facets is also assumed, such that the electrostatic potential is pinned to the near midgap value of 0.8 eV separation between the conduction band edge and the Fermi level. The problem domain extends far enough from the p–n junction regions, such that the electrostatic potential becomes invariant parallel to the heterointerface.

The confined states of the p–n–p quantum wire structure are also of interest. The bound state wave functions are sought by solving the two-dimensional Schrödinger equation for a computed potential profile,

$$-\frac{\hbar^2}{2m_e^*}\nabla^2\psi(\mathbf{r})+[V(\mathbf{r})-e\phi(\mathbf{r})]\psi(\mathbf{r})=E\psi(\mathbf{r}). \tag{3}$$

Here, $V(\mathbf{r})$ includes external potentials and band offsets and $\phi(\mathbf{r})$ is the electrostatic potential obtained from the solution of Poisson's equation for the V-groove geometry.

The decay of the wave function far from the potential "pocket" that forms the quantum wire provides the necessary boundary conditions to formulate Eq. (3) as an eigenvalue/vector problem. The wave functions of interest, however, are the lowest energy quantum wire states. The problem, therefore, can be reduced to a subspace of the full eigensystem.

III. NUMERICAL METHODS

The solution of the two-dimensional Poisson and Schrödinger equations is a challenging numerical problem that requires domains of highly nonuniform mesh design, bandwidth optimization, and sparse matrix methods. The general considerations of the solution method are discussed in Sec. III A, whereas the details of the numerical method are deferred to the Appendix. The interesting and often ignored subject of nonuniform mesh generation is discussed in Sec. III B.

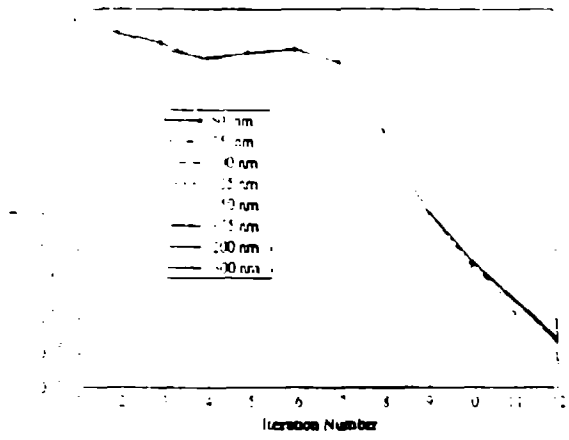


FIG. 2. Convergence of the Poisson solution Newton–Raphson iteration scheme for several different p–n–p geometries.

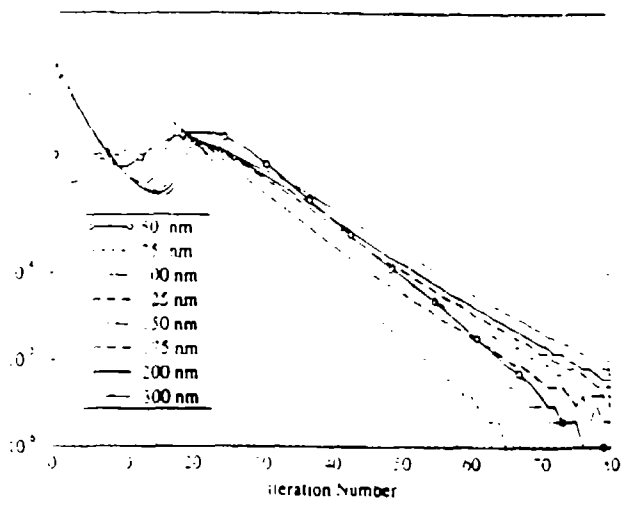


FIG. 3. Convergence of the Schrödinger solution subspace iteration scheme for several different p–n–p geometries.

A. The solution of Poisson and Schrödinger equations

The two-dimensional Poisson and Schrodinger equations are both solved with the finite element method. The Poisson equation is solved for the discretization domain schematically shown by the hatched region in Fig. 1, whereas the Schrodinger equation is only solved for the small subregion near the V-groove tip, where the electrostatic potential "pocket" defines the quantum wire.

For the Poisson solution, the bulk charge neutrality boundary conditions deep inside the GaAs substrate are implemented by fixing the electrostatic potential at a few points on the boundary to the value consistent with the background doping, and by applying zero-valued Neumann conditions on the remainder of the bulk boundary. This condition forces the boundary normal electric field to zero, and forces the potential to the proper equilibrium bulk value. Surface pinning on the exposed facets is implemented by Dirichlet conditions, which force the conduction band potential to the approximate midgap value of 0.8 eV above the Fermi level. Zero-valued Neumann boundary conditions are applied along the side boundaries to model the asymptotic invariance parallel to the heterointerface. The discontinuity in the electrostatic potential at the heterointerface is treated as a linear constraint on the system of equations and is implemented with the penalty element method. The nonlinear Poisson equation is linearized by the Newton–Raphson iterative method. The details of the finite element formulation and linearization are deferred to Appendix A.

The resulting system of linear equations is solved by standard L/U decomposition, backward substitution, and forward elimination. Bandwidth optimization of the discretized equation numbers in conjunction with the skyline storage technique provide an efficient solution of the linear system. The iterative error of the Newton–Raphson solution is plotted in Fig. 2 for typical runs of the lateral p–n–p quantum wire domain for various widths of the n-type region at the V-groove tip. The finite element domain consists of 16 168 elements and 8427 bandwidth-optimized nodes. The error drops by approximately 9 decades within 12 iterations with a typical cumulative solution time of 430 sec on an IBM RISC System/6000.

For the Schrodinger equation, the asymptotic decay of the wave function can be modeled by either zero-valued Neumann boundary conditions or zero-valued Dirichlet boundary conditions. Continuity of the wave function across the heterointerface is implemented with the penalty element method and provides compatibility with the Poisson solver. The details of the finite element formulation of the Schrödinger equation are deferred to Appendix B.

Because only the lowest few eigenstates are sought, the discretized Schrödinger equation is solved by a subspace iteration method. Efficient skyline storage and sparse matrix methods are used to obtain the lowest ten bound states. The iterative error of the subspace algorithm is plotted in Fig. 3 for the ground state of the lateral p–n–p quantum wire for the same selection of n-region V-groove tip widths, as were chosen for Fig. 2. After initial oscillation, the error exponentially decreases at approximately one decade every 15 iterations. For the finite element Schrodinger domain consisting of 3482 elements and 1876 bandwidth-optimized nodes, a solution of a 20 eigenstate subspace over 120 iterations typically consumes 240 cpu seconds on an IBM RISC System/6000.

B. Nonuniform mesh generation

Two-dimensional numerical modeling of realistic structures, such as those considered in this work, requires the development of sophisticated mesh domains. Bulk charge neutrality boundary conditions imply that the domain, Ω, must extend microns into the bulk region. At the heterointerface and in the p–n junction depletion regions, however, the potential profile may undergo large variations in nanometer distances. The disparity between characteristic length scales and the need to keep the problem size computationally tractable constrains the discretization scheme to highly nonuniform mesh designs. A further constraint on mesh design is the need to preserve the charac-

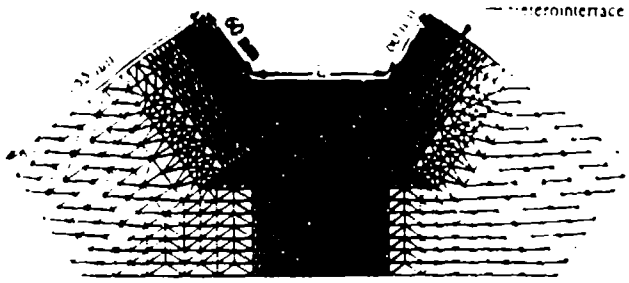


FIG. 4. Nonuniform, triangular element mesh domain used to solve the Schrodinger equation for the quantum wire states. The Schrodinger mesh is a small subdomain of the Poisson mesh cut from the region of the potential "pocket" formed at the bottom of the V groove. This mesh has 3482 triangular elements and 1876 nodes.

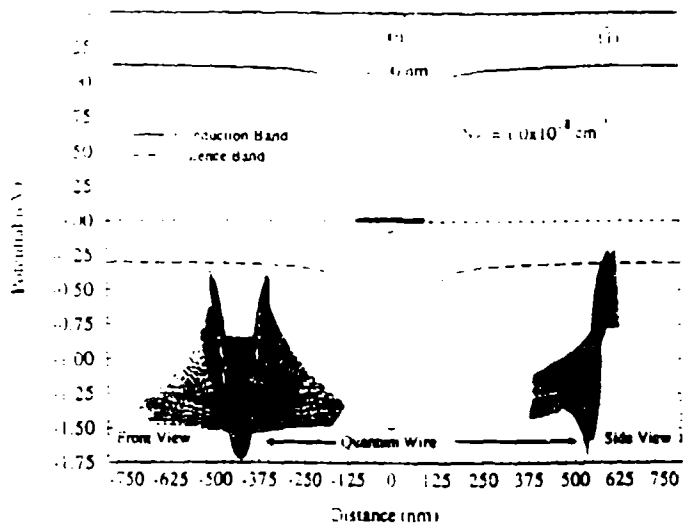


FIG. 5. Band diagram for the lateral p–n–p quantum wire structure. Shown are the conduction and valence bands along the heterointerfaces for a 100 nm long center n-type section and a Si-doping density of 1.0×10^{18} cm^{-3}. The insets show front and side views of the two-dimensional conduction band profile. The potential "pocket" that holds the quantum wire is also indicated.

teristic geometry of the structure. The 54.7° angle between the {100} and {111} crystallographic planes, for example, must be maintained throughout the domain.

With a rectangular master element design, it would be difficult to maintain arbitrary angles throughout the domain, and the resulting system of equations would likely be intractably large. With the triangular master element, however, it is easy to maintain arbitrary geometric angles in nonuniform mesh designs. Furthermore, standard bandwidth optimization algorithms exist for triangular element domains to efficiently number the nodes for sparse matrix methods.

At present there are few mesh layout tools available to the scientific community suitable for device modeling applications. Xmgredit,[16] however, is a powerful mesh generation tool that was originally designed for oceanographic land-margin studies, which is general enough for device modeling work. Xmgredit is a window based interactive mesh editing tool that allows arbitrary triangulation of a domain and provides all the necessary functions for grid refinement.

Figure 4 shows the mesh generated to solve the Schrodinger equation for the lateral p–n–p quantum wire structure. This mesh is a small subdomain of the larger mesh used to solve the Poisson equation, since the bound state wave functions are nonzero only near the potential "pocket" at the V-groove tip. We present only the smaller Schrodinger mesh in Fig. 4 because its ratio of largest to smallest element is small enough that all the elements can be clearly seen.

As mentioned above, the Poisson mesh consists of 16 168 elements and 8427 nodes, and the Schrodinger subdomain consists of 3482 elements and 1876 nodes. For the larger Poisson mesh, the maximum element area is 3.7×10^5 Å^2 and the minimum element area is 70 Å^2, while maintaining the proper crystallographic orientations throughout the domain. The sidewalls of the V groove extend approximately 800 nm for the Poisson mesh and approximately 60 nm in the Schrodinger mesh, as shown in Fig. 4.

In designing such a large nonuniform mesh, it is easiest to generate a separate mesh for each region of the heterostructure with different material properties and then connect the subdomains together. The special penalty elements can then be added along the heterointerfaces to force the potential discontinuity. We have found that a computationally efficient triangular element mesh can be generated for any two-dimensional device geometry with a minimum of effort.

IV. LATERAL V-GROOVE QUANTUM WIRE STRUCTURE

The results presented in this section concentrate on the p–n–p V-groove structure indicated by the hatched model domain labeled "wire" in the schematic diagram shown in Fig. 1. Interesting behavior is expected when back-to-back lateral quasi-2-D p–n junctions are formed with a common n-type region. In particular, if the n-type region is made narrow, one would expect the 2DEG formed at the heterointerface to be laterally "squeezed" into a 1DEG by the depletion regions of the p–n junctions on both sides. A conduction band "pocket" is expected to form at the tip of the V groove, which may support two-dimensionally confined quantum wire states.

Figure 5 is a plot of the band diagram for a lateral p–n–p junction with an overlayer Si-dopant concentration of 1.0×10^{18} cm^{-3} and with a narrow 100 nm n-type region at the V-groove tip. The electrostatic potential is plotted *parallel* to the heterointerface on the GaAs substrate side, along the {111}–{100}–{111} surfaces. The formation of an electron gas is expected in the region where the conduction band approaches and traverses the semiclassical Fermi energy, indicated by the dashed line, which is chosen as the zero of energy. The quantum wire potential "pocket" is more clearly visible in the insets to Fig. 5, where portions of the front and side views of the two-dimensional conduction band profile are shown. Only a small submesh of the full two-dimensional computation domain is shown in the

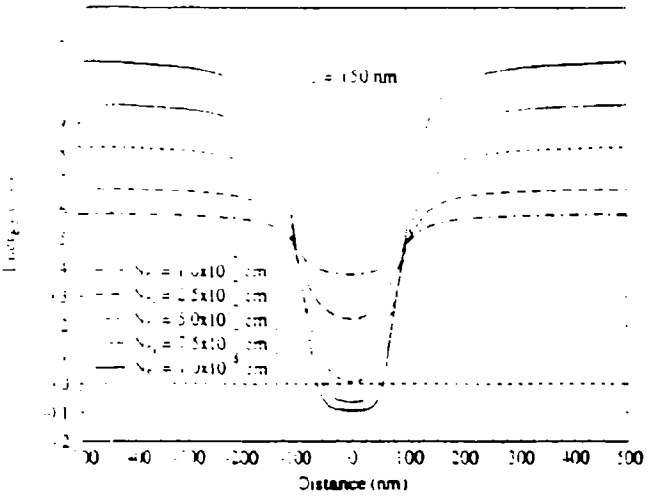


FIG. 6. Conduction band profiles for the p–n–p quantum wire structure. Shown are the conduction bands along the (111) and (100) heterointerfaces with a fixed 150 nm width n-type section for five different overlayer dopant concentrations.

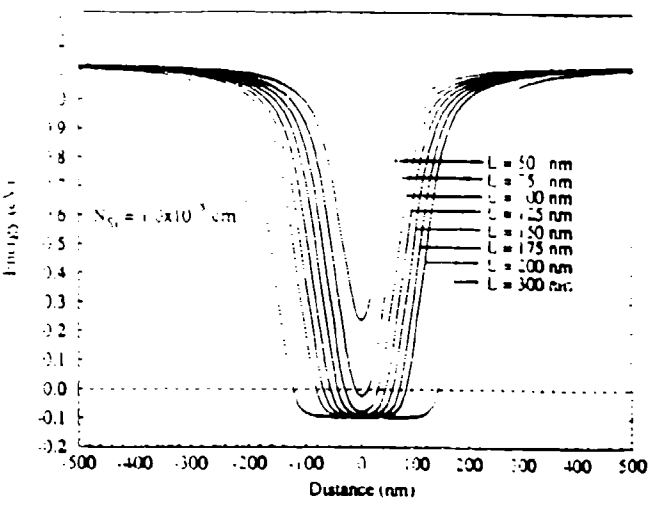


FIG. 7. Conduction band profiles for the p–n–p quantum wire structure. Shown are the conduction bands along the (111) and (100) heterointerfaces with a fixed 1.0×10^{18} cm^{-3} doped $Al_xGa_{1-x}As$ overlayer for eight different widths of the exposed (100) surface.

inset figures for better delineation of the potential well, although the calculations are performed on the full domain to ensure that bulk charge neutrality conditions are satisfied deep in the GaAs substrate. The lateral p–n–p junctions that form at the heterointerface are clearly visible in the left inset of Fig. 5 (front view), whereas the heterointerface and "pocket" are more clearly seen in the right inset (side view).

Shown in Fig. 6 is a comparison of the lateral conduction band profile for several different overlayer Si-dopant concentrations with constant 150 nm width n-type regions. Again, the semiclassical Fermi energy is chosen as the zero of energy and is indicated by the dashed line. As is expected, the p–n–p depletion widths are reduced and the quantum well depth in the n region increases with higher overlayer dopant concentration. For the case of a 150 nm wide V-groove tip and overlayer dopant concentration of 1.0×10^{18} cm^{-3}, flatband conditions are approached at the center of the n region, as shown by the solid curve in Fig. 6. This flatband condition occurs when the two lateral depletion regions do not overlap and therefore there is no *lateral* field component in the center (n-type) region of the V-groove tip.

The effect of the width of the n-type V-groove tip on the lateral conduction band profile is shown in Fig. 7 for a constant Si overlayer dopant concentration of 1.0×10^{18} cm^{-3}. The semiclassical Fermi energy is again indicated by the dashed line at 0.0 eV. Larger V-groove tip widths correspond to an increase in depth and width of the lateral quantum well. Flatband conditions are demonstrated in Fig. 7 for n-region widths greater than 150 nm. As in Fig. 6, *lateral* flatband conditions occur at the bottom of the quantum well "pocket" when the two lateral p–n depletion regions do not overlap. Once flatband conditions are established in the bottom of the well, the maximum depth of the "pocket" becomes independent of the V-groove tip width. However, the electronic states in the well are expected to be a strong function of the lateral well width. In the limit of a very wide n region, the behavior of the system will approach that of isolated p–n junctions and the quasi-1DEG becomes a quasi-2DEG in the n region. The V-groove tip width, therefore, controls the transition of the n region from a 1DEG quantum wire system to a 2DEG lateral p–n–p system.

Figures 6 and 7 demonstrate how control of the *lateral* quantum well electrostatic potential can be obtained through adjusting both the overlayer Si-dopant concentration and the lateral V-groove dimensions. It is also of interest to study the dependence of the *vertical* heterointerface confinement for the same parameters. Plotted in Fig. 8 are the conduction band profiles in a direction *perpendicular* to the heterointerface, at the center of a 150 nm wide n region, for the same dopant conditions used in Fig. 6. As before, the zero of energy is chosen to be the semiclassical Fermi energy and is indicated by the dashed line. Increasing negative values of distance correspond to increasing depth into the GaAs substrate, whereas increasing positive

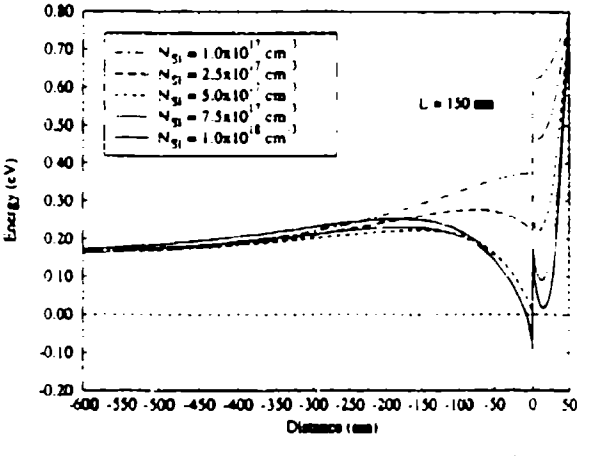


FIG. 8. Potential variation *perpendicular* to the heterointerface (at the center of a fixed 150 nm width n-type region) for five different overlayer dopant concentrations.

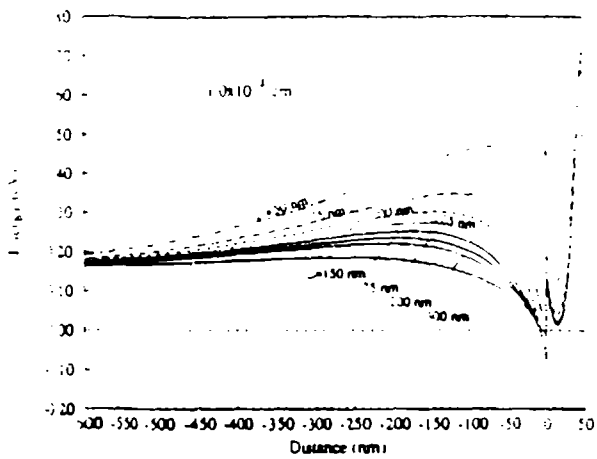


FIG. 9. Potential variation *perpendicular* to the heterointerface (at the center of the *n*-type region) with a fixed 1.0×10^{18} cm^{-3} Si-doped $Al_{0.3}Ga_{0.7}As$ overlayer for eight different widths of the *n*-type region.

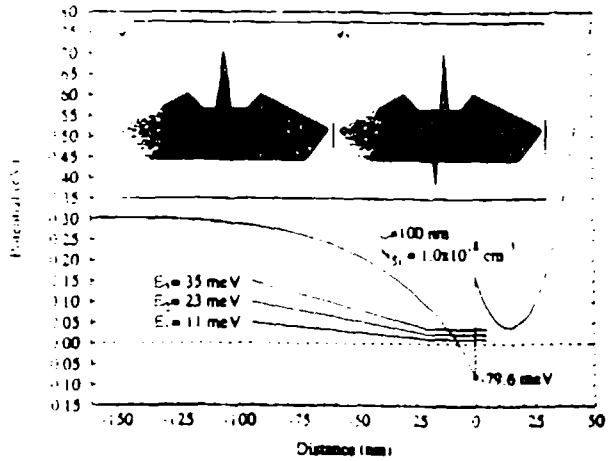


FIG. 10. Potential variation *perpendicular* to the heterointerface (at the center of a 100 nm wide *n*-type region) for the two-dimensional conduction band profile shown in Fig. 5. The insets show the wave functions for the lowest two quantum wire states, that are confined at the GaAs side of the heterointerface (the darker top portion of the mesh corresponds to the AlGaAs overlayer).

values of distance correspond to locations in the $Al_{0.3}Ga_{0.7}As$ overlayer approaching the surface. For clarity, the results are only plotted up to a depth of 600 nm into the GaAs substrate, although the calculation domain extends for several microns to ensure that bulk charge neutrality conditions are satisfied. Figure 8 shows that heterointerface confinement is only achieved for overlayer dopant concentrations above a critical value of approximately 5.0×10^{17} cm^{-3}. The conduction band potential on the GaAs side at the heterointerface has a strong dependence on the overlayer doping for concentrations below the critical value, marking the onset of the electron gas formation, and a much weaker dependence for concentrations above this critical value. The conduction band potential in Fig. 8 first *increases* from its bulk value for points approaching the heterointerface, and then it sharply decreases to form the heterointerface confinement "notch." This "hump" in the GaAs conduction band is more pronounced for higher overlayer dopant concentrations.

In a similar fashion to Fig. 8, Fig. 9 shows the corresponding conduction band profiles for the same selection of center *n*-region widths used in Fig. 7 with a constant overlayer dopant concentration of 1.0×10^{18} cm^{-3}. The most noticeable feature is the pronounced "hump" in the GaAs conduction band at narrower *n*-region widths. For the narrowest, $L=50$ nm case denoted by the dash–dotted line in Fig. 9, a sharp heterointerface confinement notch is formed with a wide "hump" barrier that extends several hundred nanometers into the GaAs. In this case, the barrier in the GaAs should prevent the leaking of any one-dimensional quantum wire states into the bulk substrate. As the *n*-region width is increased, the GaAs barrier decreases until, in the limit, the conduction band profile approaches that of the far *n* side of an isolated *p*–*n* junction.

It is of interest to verify that the quantum wire structure supports two-dimensionally confined electronic states. To this end, the calculated conduction band profile in a region enclosing the potential "pocket" is used as the electrostatic potential in the Schrödinger Hamiltonian, as discussed in Sec. III A. Presented in Fig. 10 are the lowest energy quantum states sought for the system with a 100 nm *n*-region width and 1.0×10^{18} cm^{-3} overlayer dopant concentration. The conduction band potential is plotted for a slice through the center of the *n*-type region in a direction *perpendicular* to the {100} surface. The lowest eigenenergy, labeled by E_1 in Fig. 10, is at 11 meV referenced to the semiclassical Fermi energy, which is denoted by the dashed line at 0.0 eV. The subsequent excited states each have a separation of about 12 meV associated with the contribution of the lateral confinement. The left and right insets in Fig. 10 show the full two-dimensional results for the ground state and first excited state, respectively. The $Al_{0.3}Ga_{0.7}As$ overlayer appears as the region of higher mesh density near the top of the inset figures and delineates the heterointerface. The wave functions are concentrated in the deep "pocket" on the higher mobility semi-insulating GaAs substrate side of the heterointerface.

It is clear from the insets of Fig. 10 that the first excited state corresponds to the first excited mode of the *lateral* confinement potential induced by the *p*–*n* junctions. The quantized level spacing for the lower eigenstates is therefore controlled by the nature of the *lateral* confinement. The nearly equal level spacing of the lowest states is due to the parabolic-like shape of the lateral potential profile of the 100 nm width *n*-region structure, as can be seen in Fig. 7. The higher-energy states, however, also include excited modes of the confinement perpendicular to the heterointerface. This is particularly true for narrow *n*-region *V*-groove structures in which the magnitude of both confinement dimensions are of similar order. The large separation of the lowest state above the conduction band minima at the heterointerface, labeled by the dot in Fig. 10, is attributed to both the lateral confinement energy and the confinement energy *perpendicular* to the heterointerface.

The details of the amphoteric nature of the silicon doping near the intersection of the (100) and (111) surfaces are not known, and are likely to be more complicated than our model. This and possible compensation effects might

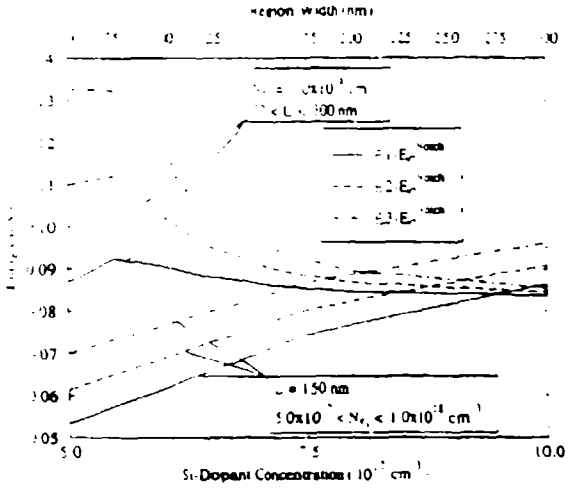


FIG. 11. Variation of the lowest three quantum wire states with overlayer dopant concentration and n-region width. The energies are referenced to the conduction band minimum at the heterointerface (i.e., the bottom of the "pocket"). The negative sloping curves correspond to independently changing the width of the n-type region for a fixed 1.0×10^{18} cm^{-3} overlayer doping. The positive sloping curves correspond to independently changing the overlayer doping for a fixed 150 nm wide n-type region.

tend to reduce the abruptness of the lateral p–n junction, which also would tend to reduce the energy level spacings.

The dependence of the eigenstates on both the overlayer doping and the n-region width is presented in Figs. 11 and 12. Simultaneously shown in Fig. 11 are the dependence of the lowest three bound state energies on the lateral n-region width and the overlayer Si-doping concentration. The bottom axis corresponds to independently changing the overlayer Si-dopant concentration for a fixed 150 nm width V-groove tip. The top axis corresponds to independently changing the lateral width of the V-groove tip for a fixed overlayer dopant concentration of 1.0×10^{18} cm^{-3}. The energies on the ordinate are plotted relative to the minimum of the conduction band profile at the heterointerface, i.e., relative to the bottom of the potential well, denoted by E_c^{Notch}. The ground state energy is lowered by independently widening the n region and saturates at the limiting value given by the heterointerface confinement. In approaching the wide n-region limit, the one-dimensional electron gas evolves into a two-dimensional electron gas, which is manifest by the pronounced decrease in the level spacing in Fig. 11. The change in behavior for narrow widths, i.e., below 75 nm, seen in Fig. 11, occurs when the magnitude of the lateral junction confinement is comparable to that of the heterointerface confinement. Independently increasing the overlayer dopant concentration has several effects on the potential well; the confining well in the direction *perpendicular* to the heterointerface becomes "deeper" and "thinner," whereas the *lateral* junction confinement becomes wider and more abrupt, with smaller side depletion regions. Figure 11 shows that the net effect of increased doping is to increase the ground state energy, relative to the bottom of the quantum well, with only a slight decrease in the energy level spacings. Simultaneously plotted in Fig. 12 are the energy level separations for the same parameters given in Fig. 11. The bottom and top axis represent the effect of independently changing the overlayer doping and independently changing the n-type region width, respectively. The nearly equal separation of the second mode from the first and the third mode from the second is due to the parabolic shape of the lateral junction confinement. When the lateral n-region width is increased, the pronounced drop in the level spacing is observed, whereas only a slight drop is observed for increased overlayer doping. The n-region width is therefore the dominant controlling parameter of the quantum wire energy level spacing.

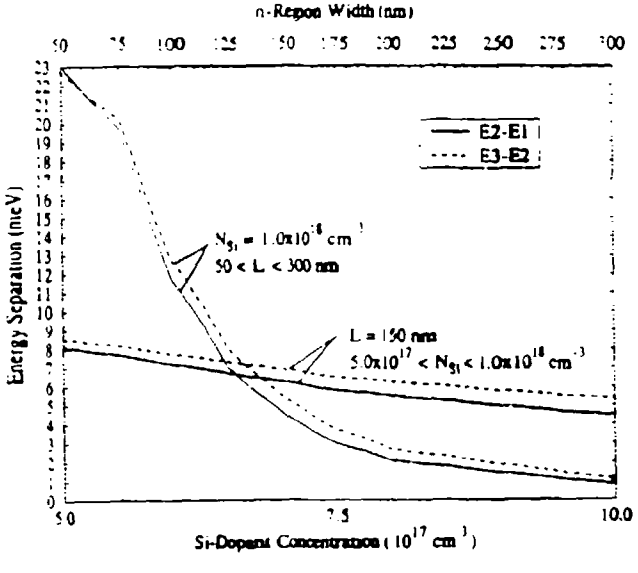


FIG. 12. Variation of the quantum wire state separations with overlayer dopant concentration and n-region width. The sharply sloped curves correspond to independently changing the width of the n-type region for a fixed 1.0×10^{18} cm^{-3} overlayer doping. The slightly sloped curves correspond to independently changing the overlayer doping for a fixed 150 nm wide n-type region. The solid and dashed lines are the separation between the second and first states and between the third and second states, respectively.

Control of the optical and transport properties of the active wire region is of primary interest for eventual device applications of this corrugated GaAs/AlGaAs heterojunction system. The previous results indicate that independent control of the quantum confinement dimensions may be obtained by the fabrication of gates on the corrugated surface. In this section we also present the method and results for our initial attempt to model the electrostatic potential profile with the addition of Schottky gates.

A simple gate geometry is chosen such that each wall of the V groove has a Schottky gate. The gates on the p-type {111} sidewalls stop 100 nm from the n-type {100} V-groove tip. The center gate, along the {100}, extends over the full n-region width. In these calculations, the same bias will be applied to the sidewall gates and the center gate will be referenced to ground.

The Thomas–Fermi screening model for the electron and hole concentrations, described in Sec. II, tacitly assumes an equilibrium carrier distribution. This model is valid for small perturbations from equilibrium, such that

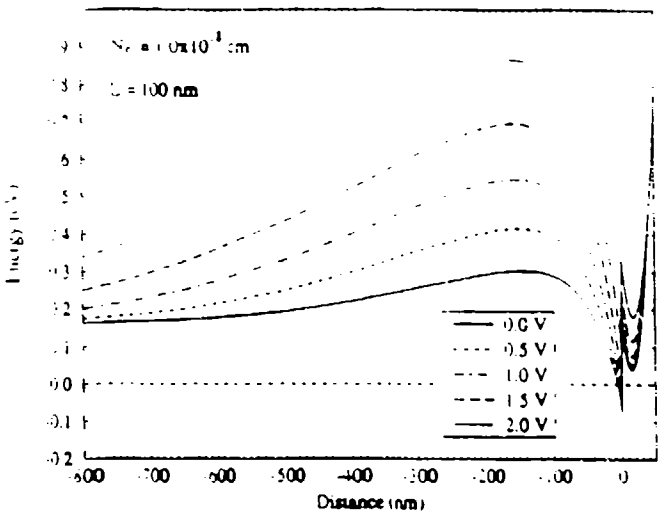


FIG. 13. Potential variation *perpendicular* to the heterointerface (at the center of a 100 nm wide *n*-type region, and 1.0×10^{18} cm^{-3} Si overlayer doping) for several different bias potentials applied to Schottky gates on the V-groove sidewalls.

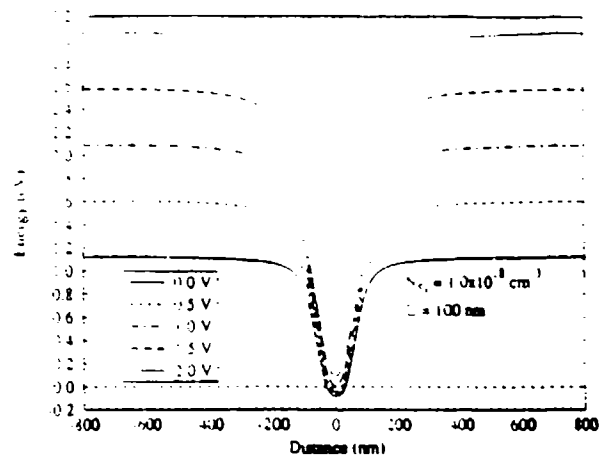


FIG. 14. Lateral conduction band profiles for the same structure and bias conditions, as shown in Fig. 13. Shown are the conduction bands at the heterointerface parallel to the surface along the (100) and (111) planes.

no appreciable hole concentration exists at the surface. However, under the bias conditions necessary to obtain control of the quantum wire, this charge model must be modified to extend its range of validity. To this end we assume two constant quasi-Fermi levels: one for electrons referenced to the Fermi level in the bulk GaAs substrate, and one for holes referenced to the Fermi level of the sidewall metal gates.[17] The electron and hole concentrations are therefore calculated most accurately in the regions, where they are the majority carrier, and least accurately in the minority carrier regions, where their contribution to the total charge density is least significant.

The boundary conditions on the corrugated AlGaAs surface are modified to include the gate bias. The boundary beneath the gates is modeled by Dirichlet conditions that specify the potential as the sum of the equilibrium surface pinned value plus the bias potential. Since the center gate is referenced to ground, the boundary along the *n*-type region remains pinned at the approximate midgap potential of 0.8 eV above the bulk quasi-Fermi level. The 100 nm exposed surfaces between the gates is modeled by a linear drop in the potential along the surface from the sidewall gates to the center electrode. This somewhat *ad hoc* assumption is justified because the overlayer is depleted of mobile carriers, as seen in the previous results.

The vertical conduction band profile through the center of the *n* region is plotted in Fig. 13 for five different side-gate biases. The overlayer Si dopant concentration is 1.0×10^{18} cm^{-3} and the *n*-region width is 100 nm. As before, increasing negative values of distance correspond to positions deeper in the bulk GaAs substrate referenced to the heterointerface. The quasi-Fermi level for electrons is specified by the bulk GaAs Fermi level and is denoted by the dashed line at 0.0 eV. Figure 13 demonstrates that, with this gate geometry, increasing bias reduces the confinement width of the notch and only slightly depletes the electron gas. The increase of the "hump" in the GaAs is a feature that should further confine the electronic states to the heterointerface.

The lateral *p–n–p* conduction band profile is shown in Fig. 14 for the same structure and sidegate bias conditions. Positions parallel to the heterointerface are referenced to the center of the *n* region, and the quasi-Fermi level for electrons is denoted by the dashed line at 0.0 eV: Increasing sidegate bias narrows the lateral confinement dimension, which was previously shown to control the quantum wire level spacing.

These initial results show that control of the electronic properties of the structure may be obtained by the addition of lithographically defined gates on the surface. It is intuitively clear that better control of the lateral confinement width can be achieved with only the sidegates and no center electrode. It is difficult, however, to formulate well-posed boundary conditions along the exposed surface without a center reference electrode. The nature of the Poisson equation requires Dirichlet, Neumann, or mixed boundary conditions along the exposed surface, none of which are simple without completely *ad hoc* assumptions. It is also of interest to form Ohmic contacts to the two-dimensional carriers in the structure, which should provide control of the transport properties through reverse-biased *p–n* junctions. Ohmic contacts also provide the interesting possibility of minority carrier injection into the quantum wire, which may give rise to new device applications, since this is an inherently bipolar structure. The model of an Ohmic contact, however, must include direct control of the two-dimensional carriers, which implies additional constraints along the heterointerface. We are working on a better model for an exposed surface, containing occupied surface trap states under zero current bias conditions, and the development of well-posed electrostatic boundary conditions.

V. SUMMARY

We have reported on our method for modeling a novel *p–n–p* quantum wire structure formed in a corrugated GaAs/AlGaAs heterostructure. We solve the two-dimensional Poisson equation, within a semiclassical

Thomas–Fermi screening model, for the conduction and valence band profiles. The quantum confined electronic states are subsequently obtained by solving the two-dimensional Schrodinger equation using the previously computed electrostatic potential.

We have demonstrated that it is feasible to realize a bipolar quantum wire structure on a HEMT-compatible corrugated GaAs/AlGaAs interface by exploiting the surface orientation dependent amphoteric nature of Si doping. These results indicate that the quantum wire level separations are controlled by the lateral confinement width, and the parabolic-like shape of the lateral confinement potential gives rise to nearly equal level spacings. It was also shown that the ground-state energy of the wire states can be controlled by the vertical confinement width. Initial results demonstrate that the electronic properties of the structure may be controlled by the addition of gates on the corrugated surface. This structure has several possible advantages compared to sidewall etched or Schottky contact electrostatically confined systems.[4] The definition of the n-region width is controlled by the anisotropic etch time through an optically defined photoresist mask; ion-beam or electron-beam lithography is not required to fabricate the structure. This inherently bipolar heterostructure might be used to inject minority carriers into the quantum wire from the p regions, possibly leading to new device applications. If the p-type and n-type regions are separately contacted, by further nanolithography, the lateral confinement width might be controlled by reverse biasing the p–n junctions and the electron Fermi energy might be controlled via a gate contact on the AlGaAs.

ACKNOWLEDGMENTS

The authors would like to acknowledge fruitful discussions with J. Ebner, D. J. Kirkner, C. S. Lent, J. L. Merz, and T. K. Plant regarding this work. The authors would also like to acknowledge the help of J. Westerink for obtaining mesh generation and bandwidth optimization tools and M. Mueller for creating Dore' graphics programs to display the results. This work has been partially supported by the Office of Naval Research and the Air Force Office of Scientific Research.

APPENDIX A: FORMULATION OF THE POISSON PROBLEM

1. Problem statement

The model structure is partitioned into regions, Ω_1, Ω_2,..., Ω_N, as shown in Fig. 15, defined by different material parameters, such as Al mole fraction and Si-doping concentration. We want to solve the two-dimensional Poisson equation on the whole domain $\Omega \equiv \Omega_1 \cup \Omega_2 \cdots \cup \Omega_N$. The total boundary of domain Ω will be denoted by Γ and is partitioned into three terms: Γ_S for that part of Γ that is on the surface of the structure, i.e., the boundary on which the Fermi level is surface pinned, $\hat{\Gamma}$ for that part of Γ on which bulk charge neutrality conditions apply, and last Γ_B for the remainder of Γ on which the normal electric field is forced to zero. Bulk charge neutrality can also be specified

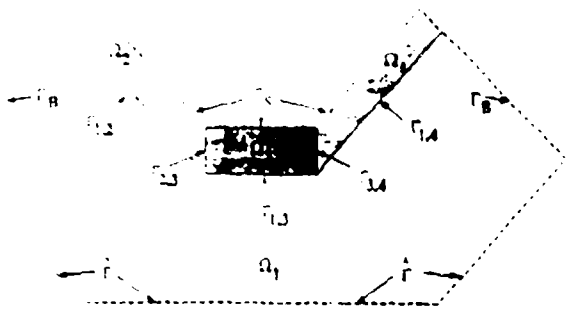


FIG. 15. Schematic diagram of the calculation domain. Domains of different material composition are indicated by different shades of grey and labeled Ω_1–Ω_4. The bulk substrate boundary is denoted by $\hat{\Gamma}$, the surface pinned boundary is denoted by Γ_S, and the side boundaries are denoted by Γ_B. The boundaries between adjacent materials Ω_i and Ω_j is denoted by $\Gamma_{i,j}$. In the quantum wire structure, the AlGaAs/GaAs heterointerface occurs along $\Gamma_{1,2}$, $\Gamma_{1,3}$, and $\Gamma_{1,4}$.

as zero-valued Neumann boundary conditions. We have found that specifying essential boundary conditions at a few points and zero-valued Neumann conditions on the remainder of the bulk boundary works best, since it provides a quick visual check of the solution for proper convergence. The boundaries of each subregion are denoted by Γ_m. The portion of Γ_m in common with another subdomain, Γ_n is denoted by $\Gamma_{m,n}$. All the heterointerfaces in Ω will therefore occur along the $\Gamma_{m,n}$ boundaries. The potential on the Γ_m side of a heterointerface formed between regions Ω_m and Ω_n is labeled $u_{\Gamma_{m,n}}$ and likewise labeled $u_{\Gamma_{n,m}}$ on the opposite side. We will also define $u=(E_c(x,y)-E_F)/k_BT$ as the conduction band edge energy relative to the Fermi energy in units of eV/k_BT. We now specify the zero of potential energy to correspond to the Fermi energy, $E_F=0$. The problem to be solved can now be written as follows:

Find:

$$u=u_1\cup u_2\cdots\cup u_N\in H^2(\Omega), \tag{A1}$$

such that

$$-\nabla^2 u_m(x,y)=\frac{\rho(x,y)}{\epsilon_m\epsilon_0}, \quad (x,y)\in\Omega_m \tag{A2}$$

and

$$u_m=\begin{cases}\hat{u}_m & \text{on } \Gamma_m\in\hat{\Gamma};\\ \hat{u}_s & \text{on } \Gamma_m\in\Gamma_S;\\ u_{\Gamma_{n,m}}-\Delta u_{m,n} & \text{on } \Gamma_{m,n},\ m<n.\end{cases} \tag{A3}$$

and

$$\nabla u_m\cdot\hat{n}_i=0, \quad \Gamma_m\in\Gamma_B, \tag{A4}$$

where $\rho(x,y)=-e[n(r)-p(r)+N_A^- -N_D^+)]$ is the total charge density used in Eq. (1) of Sec. II of the main text, $\hat{u}_m$ is the value of the potential set by the background doping and the requirement of space charge neutrality, $\hat{u}_s=0.8\ eV/k_BT$ is the near midgap surface pinned potential, and $\Delta u_{m,n}=u_{\Gamma_{n,m}}-u_{\Gamma_{m,n}}$ is the conduction band discontinuity at the heterointerface formed between regions Ω_m and Ω_n, and $\hat{n}_i$ is the unit surface normal. The charge

tensity term on the right-hand side of the Poisson equation will be denoted by $f_m = f(u_m)$, and is given by the Fermi–Dirac integral of order 1/2. The nonlinear nature of the Poisson equation will be discussed later.

We will use the method of variations to formulate the numerical problem. We first identify a total energy functional as

$$\Pi = \sum_m \left\{ \int_{\Omega_m} \left[\frac{1}{2} \nabla u_m \cdot \nabla u_m - f_m u_m \right] d\Omega_m \right\} - \sum_m \left\{ \int_{\Gamma_m} f_m u_m \, d\Omega_m \right\} + \sum_m \left\{ \frac{\alpha}{2} (u_{\Gamma_m} - \hat{u}_m)^2 \right\} + \sum_{\substack{m,n \\ m<n}} \left\{ \frac{\alpha}{2} (u_{\Gamma_{m,n}} - u_{\Gamma_{n,m}} - \Delta u_{m,n})^2 \right\}. \tag{A5}$$

We have used the penalty method to incorporate the linear constraint boundary conditions, where α is a large penalty factor that will be discussed later. We prefer the penalty method to the Lagrange multiplier method since it does not increase the dimensionality of the resulting linear system. We now invoke the principle of stationarity of Π with respect to the state variable u by evaluating $\delta\Pi = 0$. In the calculus of variations, this is the minimization of the energy functional. We denote arbitrary variations in the state variable, which satisfy the essential boundary conditions by δ. From the minimization we obtain

$$\delta\Pi = 0 = \sum_m \left\{ \int_{\Omega_m} \nabla u_m \nabla \delta u_m \, d\Omega_m \right\} - \sum_m \left\{ \int_{\Omega_m} f_m \delta u_m \, d\Omega_m \right\} + \sum_m \{ \alpha (u_{\Gamma_m} - \hat{u}_m) \delta u_{\Gamma_m} \} + \sum_{\substack{m,n \\ m<n}} \{ \alpha (u_{\Gamma_{m,n}} - u_{\Gamma_{n,m}} - \Delta u_{m,n}) \cdot (\delta u_{\Gamma_{m,n}} - \delta u_{\Gamma_{n,m}}) \}. \tag{A6}$$

The state variable, u_m, is subject to the boundary conditions specified by Eqs. (A3) and (A4) above. The third term on the right-hand side is the penalty method implementation of the bulk charge neutrality conditions. These essential boundary conditions could be implemented by Dirichlet conditions with a corresponding reduction of the number of unknowns. The solution to Eq. (A5) above is unique and equivalent to the solution of the original problem.

2. Finite element formulation

The domain Ω is nonuniformly discretized with M nodal points at locations $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$, for each of which is defined an orthonormal global shape function, $\phi_i(\mathbf{r})$, such that $\phi_i(\mathbf{r}) = \delta_{i,j}$. In our calculations we use linear Chapeau shape functions. The potential u is approximated by an expansion in this basis, $u(\mathbf{r}) \simeq \sum_{i=1}^{M} u_i \phi_i(\mathbf{r}) = \mathbf{N}(\mathbf{r}) \cdot \mathbf{u}$, where $\mathbf{N}(\mathbf{r})$ is the vector of M shape functions and $\mathbf{u}$ is the vector of M potentials at each nodal point. The gradient of the potential can be likewise approximated by the expansion $\nabla u(\mathbf{r}) \simeq \sum_{i=1}^{M} u_i \nabla \phi_i(\mathbf{r}) = \mathbf{B}(\mathbf{r}) \cdot \mathbf{u}$, where $\mathbf{B}(\mathbf{r})$ is the $3 \times M$ matrix of the gradients of the M global shape functions. The same approximations are made for the potential variations, $\delta u(\mathbf{r}) \simeq \delta \mathbf{u}^T \cdot \mathbf{N}^T$ and $\nabla \delta u \simeq \delta \mathbf{u}^T \cdot \mathbf{B}^T$. Using these expansions, the variational statement of the Poisson equation can be written as

$$\delta\Pi = 0 = \sum_m \left\{ \delta \mathbf{u}_m^T \left(\int_{\Omega_m} \mathbf{B}^T(\mathbf{r}) \mathbf{B}(\mathbf{r}) d\Omega_m \right) \mathbf{u}_m - \delta \mathbf{u}_m^T \left(\int_{\Omega_m} \mathbf{N}^T(\mathbf{r}) \cdot f(\mathbf{u}_m) d\Omega_m \right) \right\} + \sum_m \{ \alpha \delta \mathbf{u}_{\Gamma_m}^T \cdot \mathbf{u}_{\Gamma_m} \} - \sum_m \{ \alpha \delta \mathbf{u}_{\Gamma_m}^T \hat{u}_m \} + \sum_{\substack{m,n \\ m<n}} \left\{ \alpha \begin{pmatrix} \delta \mathbf{u}_{\Gamma_{m,n}} \\ \delta \mathbf{u}_{\Gamma_{n,m}} \end{pmatrix}^T \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{pmatrix} \mathbf{u}_{\Gamma_{m,n}} \\ \mathbf{u}_{\Gamma_{n,m}} \end{pmatrix} \right\} - \sum_{\substack{m,n \\ m<n}} \left\{ \alpha \begin{pmatrix} \delta \mathbf{u}_{\Gamma_{m,n}} \\ \delta \mathbf{u}_{\Gamma_{n,m}} \end{pmatrix}^T \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Delta u_{m,n} \right\}. \tag{A7}$$

We now define $\mathbf{e}_{\Gamma_m}$ as a $M \times 1$ vector containing a 1 in each position corresponding to $\hat{\Gamma}_m$, and zero elsewhere. We also define the $M \times M$ global matrix $\mathbf{K}'_{\Gamma_{m,n}}$ and the $M \times 1$ global vector $\mathbf{P}'_{\Gamma_{m,n}}$ as follows:

$$\mathbf{K}'_{\Gamma_{m,n}} = \begin{matrix} \\ \mathbf{u}_{\Gamma_{m,n}} \\ \mathbf{u}_{\Gamma_{n,m}} \end{matrix} \begin{matrix} \begin{matrix} \delta \mathbf{u}_{\Gamma_{m,n}} & \delta \mathbf{u}_{\Gamma_{n,m}} \end{matrix} \\ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \end{matrix}, \tag{A8a}$$

$$\mathbf{P}'_{\Gamma_{m,n}} = \begin{matrix} \\ \mathbf{u}_{\Gamma_{m,n}} \\ \mathbf{u}_{\Gamma_{n,m}} \end{matrix} \begin{matrix} \begin{matrix} \Delta u_{m,n} \end{matrix} \\ \begin{pmatrix} 1 \\ -1 \end{pmatrix} \end{matrix}. \tag{A8b}$$

We now define the following $M \times M$ matrix, $\mathbf{K}_m$, $M \times 1$ vector $\mathbf{f}_m(\mathbf{u}_m)$, and $M \times 1$ vector $\mathbf{P}_m$ as

$$\mathbf{K}_m = \left\{ \int_{\Omega_m} \mathbf{B}^T(\mathbf{r}) \mathbf{B}(\mathbf{r}) d\Omega_m \right\} + \alpha \mathbf{e}_{\Gamma_m}^T \mathbf{e}_{\Gamma_m} + \alpha \sum_{n>m} \mathbf{K}'_{\Gamma_{m,n}}, \tag{A9}$$

$$\mathbf{f}_m(\mathbf{u}_m) = \left\{ \int_{\Omega_m} \mathbf{N}^T(\mathbf{r}) f(\mathbf{u}_m) d\Omega_m \right\}, \tag{A10}$$

$$\mathbf{P}_m = \alpha \mathbf{e}_{\Gamma_m} \hat{u}_m + \alpha \sum_{n>m} \mathbf{P}'_{\Gamma_{m,n}}. \tag{A11}$$

The linear constraint boundary conditions are incorporated into the stiffness matrix, $\mathbf{K}_m$, by the penalty method. The penalty, α, must be several orders of magnitude larger than the largest diagonal element of the first term in Eq. (A9). Although the final solution will not depend upon the actual value of α, it must be sufficiently large to force the essential boundary conditions of the system.

We can simplify the problem by defining the global matrix $\mathbf{K}$, global vector $\mathbf{f}(\mathbf{u}_m)$, and global vector $\mathbf{P}$, such that $\sum_m \delta \mathbf{u}_m^T \mathbf{K}_m \mathbf{u}_m^T = \delta \mathbf{u}^T \mathbf{K} \mathbf{u}$, $\sum_m \delta \mathbf{u}_m^T \mathbf{f}_m(\mathbf{u}_m) = \delta \mathbf{u}^T \mathbf{F}(\mathbf{u})$, and

$\sum_m \delta u_m P_m = \delta u^T P$. The finite element formulation for the Poisson equation can now be written as the following system of equations:

$$\mathbf{K}\mathbf{u} = \mathbf{f}(\mathbf{u}) - \mathbf{P}. \tag{A12}$$

3. Newton–Raphson formulation

The charge density term in Eq. (A10), $\mathbf{f}(\mathbf{u})$, is a function of the potential, $\mathbf{u}$, through the semiclassical Thomas–Fermi screening model for the carrier density, as discussed in Sec. II of the main text, $n(\mathbf{u}) = N_c F_{1/2}(-\mathbf{u})$, and a similar term is obtained for the hole concentration. $N_c = 2(2\pi k_B T m_c^*/h^2)^{3/2}$ is the effective conduction band density of states and $F_{1/2}$ is the Fermi–Dirac integral of order $\frac{1}{2}$. We use the Newton–Raphson iteration method to solve the nonlinear system of equations. If we let n denote the iteration count, then we can represent the solution of the $n+1$ iteration by

$$\mathbf{u}^{n+1} = \mathbf{u}^n + \Delta\mathbf{u}^n, \tag{A13}$$

where $\Delta\mathbf{u}^n$ represents the change in the solution between iterations. The nonlinear function $\mathbf{f}$ can be approximated by a first-order Taylor series expansion:

$$\mathbf{f}(\mathbf{u}^{n+1}) = \mathbf{f}(\mathbf{u}^n) + \partial\mathbf{f}(\mathbf{u}^n)\Delta\mathbf{u}^n, \tag{A14}$$

where $\partial\mathbf{f}(\mathbf{u}^n)$ is the partial derivative of $\mathbf{f}$ with respect to $\mathbf{u}$ at the nth iteration and is given by the Fermi–Dirac integral of order $-\frac{1}{2}$.[15] We substitute this expansion for $\mathbf{f}$ into the system of equations as follows:

$$\mathbf{K}\mathbf{u}^{n+1} = \mathbf{f}(u^{n+1}) + \mathbf{P}, \tag{A15}$$

$$\mathbf{K}\mathbf{u}^n + \mathbf{K}\,\Delta\mathbf{u}^n = \mathbf{f}(\mathbf{u}^n) + \partial\mathbf{f}(\mathbf{u}^n)\Delta\mathbf{u}^n + \mathbf{P}, \tag{A16}$$

$$(\mathbf{K} - \partial\mathbf{f}(\mathbf{u}^n))\Delta\mathbf{u}^n = \mathbf{f}(\mathbf{u}^n) + \mathbf{P} - \mathbf{K}\mathbf{u}^n. \tag{A17}$$

The problem has now been reduced to a system of linear equations in $\Delta\mathbf{u}$ at each iteration. We use the bulk value of the potential given by the Si-dopant concentration and the charge neutrality condition in each domain Ω_m as an initial guess, and then solve the linear system for $\Delta\mathbf{u}^n$ and update the solution $\mathbf{u}$ until the mean-squared difference between iterations is below a specified tolerance, typically 1×10^{-9}. A standard L/U decomposition method is used to solve the sparse linear system using an efficient skyline storage technique.

APPENDIX B: FORMULATION OF THE SCHRODINGER PROBLEM

1. Problem statement

The solution domain for the Schrödinger equation is described in the same manner as the Poisson domain. The quantum mechanical problem of solving Eq. (3) in Sec. II of the main text, can be stated as follows:

Find:

$$\psi = \psi_1 \cup \psi_2 \cdots \cup \psi_N \in H^2(\Omega), \tag{B1}$$

such that

$$-\frac{\hbar^2}{2m_c^*}\nabla^2\psi_m(x,y) + [V(x,y) - e\phi_m(x,y)] = E\psi_m(x,y), \tag{B2}$$

and

$$\psi_m = \begin{cases} 0, & \text{on } \Gamma_m \in \hat{\Gamma}\cup\Gamma_S, \\ \psi_{\Gamma_{m,n}} & \text{on } \Gamma_{m,n},\ m<n, \end{cases} \tag{B3}$$

or

$$\nabla\psi_m \cdot \hat{n}_i = 0, \quad \text{on } \Gamma_m \in \hat{\Gamma}\cup\Gamma_S$$

and

$$\psi_m = \psi_{\Gamma_{n,m}} \quad \text{on } \Gamma_{m,n},\ m<n, \tag{B4}$$

where $V(x,y)$ includes the external potentials and $\phi_m(x,y) = u_m(x,y)\times k_B T$ is the electrostatic potential obtained from Poisson's equation and $\hat{n}_i$ is the unit boundary normal. As indicated above, either zero-valued essential or zero-valued Neumann boundary conditions can be used to specify the decay of the wave function.

We will use the method of variations, as we did with the Poisson equation, to formulate the numerical method. We identify the total functional as

$$\begin{aligned}\Pi = &\sum_m \left| \int_{\Omega_m} \frac{1}{2}\left(-\frac{\hbar^2}{2m_c^*}\nabla\psi_m\right)^2 d\Omega_m \right| \\ &+ \sum_m \left| \int_{\Omega_m} \frac{1}{2}(V_m - e\phi_m)\psi_m^2 d\Omega_m \right| \\ &- \sum_m \left| E\int_{\Omega_m} \frac{1}{2}\psi_m^2\, d\Omega_m \right| + \sum_m \left| \frac{\alpha}{2}\psi_{\Gamma_m}^2 \right| \\ &+ \sum_{\substack{m,n \\ m<n}} \left| \frac{\alpha}{2}(\psi_{\Gamma_{m,n}} - \psi_{\Gamma_{n,m}})^2 \right|.\end{aligned} \tag{B5}$$

The minimization of the total functional yields

$$\begin{aligned}\delta\Pi = 0 = &\sum_m \left| \int_{\Omega_m} \frac{\hbar^2}{2m_c^*}\nabla\psi_m\nabla\delta\psi_m\, d\Omega_m \right| \\ &+ \sum_m \left| \int_{\Omega_m} (V_m - e\phi_m)\psi_m\,\delta\psi_m\, d\Omega_m \right| \\ &- \sum_m \left| E\int_{\Omega_m} \psi_m\,\delta\psi_m\, d\Omega_m \right| + \sum_m \{\alpha\psi_{\Gamma_m}\delta\psi_{\Gamma_m}\} \\ &+ \sum_{\substack{m,n \\ m<n}} \{\alpha(\psi_{\Gamma_{m,n}} - \psi_{\Gamma_{n,m}})(\delta\psi_{\Gamma_{m,n}} - \delta\psi_{\Gamma_{n,m}})\}.\end{aligned} \tag{B6}$$

The state variable, ψ_m, is also subject to the boundary conditions specified by Eq. (B3) or (B4). The zero-valued

essential boundary conditions are implemented above by the penalty method, but could also be implemented by zero-valued Dirichlet conditions with a corresponding reduction of the number of unknowns. The last term in Eq. (B5) implements the continuity of the wave function across the heterointerface by the penalty method. The Schrodinger solution is therefore compatible with the Poisson solution of the electrostatic potential. The solution to Eq. (B6) above is unique and equivalent to the solution of the original problem stated in Eqs. (B1)–(B4).

2. Finite element formulation

The domain, Ω, for the Schrodinger problem uses the same generalized discretization scheme developed for the Poisson problem and is schematically shown in Fig. 15. Using linear Chapeau basis functions we expand $\psi(\mathbf{r})$ and $\delta\psi(\mathbf{r})$ in the finite element approximations: $\psi(\mathbf{r}) \simeq \mathbf{N}(\mathbf{r})\cdot\psi$, $\nabla\psi(\mathbf{r}) \simeq \mathbf{B}(\mathbf{r})\cdot\psi$, $\delta\psi(\mathbf{r}) \simeq \delta\psi^T\cdot\mathbf{N}(\mathbf{r})^T$, and $\nabla\delta\psi(\mathbf{r}) \simeq \delta\psi^T\cdot\mathbf{B}(\mathbf{r})^T$. With these expansions, the variational formulation of the problem becomes

$$\delta\Pi=0=\sum_m \left|\delta\psi_m{}^T\left(\frac{\hbar^2}{2m_c^*}\int_{\Omega_m}\mathbf{B}^T(\mathbf{r})\mathbf{B}(\mathbf{r})d\Omega_m\right)\psi_m\right| + \sum_m \left|\delta\psi_m{}^T\left(\int_{\Omega_m}(V_m-e\phi_m(\mathbf{r}))\mathbf{N}^T(\mathbf{r})\cdot\mathbf{N}(\mathbf{r})d\Omega_m\right)\psi_m\right|$$
$$-\sum_m \left|\delta\psi_m{}^T\left(E\int_{\Omega_m}\mathbf{N}^T(\mathbf{r})\cdot\mathbf{N}(\mathbf{r})d\Omega_m\right)\psi_m\right| + \sum_m\{\alpha\,\delta\psi_{\Gamma_m}^T\cdot\psi_{\Gamma_m}\} + \sum_{m,n\atop m<n}\left|\alpha\begin{pmatrix}\delta\psi_{\Gamma_{m,n}}\\ \delta\psi_{\Gamma_{n,m}}\end{pmatrix}^T\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\begin{pmatrix}\psi_{\Gamma_{m,n}}\\ \psi_{\Gamma_{n,m}}\end{pmatrix}\right|. \tag{B7}$$

Using similar definitions as those in the Poisson development in Appendix A, this set of linear equations can be reduced to a generalized eigenvalue problem,

$$\mathbf{A}\psi = E\mathbf{B}\psi. \tag{B8}$$

Because we are only interested in the lowest bound states of the system, we need not solve for all of the eigenenergies and wave functions. We use the subspace iteration method with efficient skyline storage sparse matrix methods to solve for the lowest ten bound states.

[1] M. Ballingall and C. E. C. Wood, Appl. Phys. Lett. **41**, 947 (1982).
[2] W. I. Wang, E. E. Mendez, T. S. Kuan, and L. Esaki, Appl. Phys. Lett. **47**, 826 (1985).
[3] S. Subbanna, H. Kroemer, and J. L. Merz, J. Appl. Phys. **59**, 488 (1986).
[4] D. L. Miller, Appl. Phys. Lett. **47**, 1309 (1985).
[5] J. Ebner, J. E. Lary, G. W. Eliason, and T. K. Plant, *Proceedings of the 20th European Solid State Device Research Conference*, ESSDERC 90, edited by W. Eccleston and P. J. Rossner (Hilger, Bristol, 1990), pp. 401–404.
[6] H. K. Harbury, W. Porod, and S. M. Goodnick, presented at The 19th International Conference on the Physics and Chemistry of Semiconductor Interfaces, Death Valley, CA, January 1992.
[7] W. Porod, H. K. Harbury, and S. M. Goodnick, Appl. Phys. Lett. **61**, 1823 (1992).
[8] C. C. Dean and M. Pepper, J. Phys. C **15**, L1287 (1982).
[9] A. B. Fowler, A. Hartstein, and R. A. Webb, Phys. Rev. Lett. **48**, 196 (1982).
[10] B. I. Miller, A. Shahar, U. Koren, and P. J. Corvini, Appl. Phys. Lett. **54**, 188 (1989).
[11] E. Kapon, D. M. Hwang, and R. Bhat, Phys. Rev. Lett. **63**, 430 (1989); E. Kapon, S. Simhony, R. Bhat, and D. M. Hwang, Appl. Phys. Lett. **55**, 2715 (1989).
[12] E. Colas, S. Simhony, E. Kapon, R. Bhat, D. M. Hwang, and P. S. D. Lin, Appl. Phys. Lett. **57**, 914 (1990).
[13] H. K. Harbury and W. Porod, J. Vac. Sci. Technol. B **8**, 923 (1990).
[14] *Nanostructure Physics and Fabrication*, edited by M. A. Reed and W. P. Kirk (Academic, San Diego, CA, 1989).
[15] C. M. Wolfe, N. Holonyak, Jr., and G. E. Stillman, *Physical Properties of Semiconductors* (Prentice-Hall, Englewood Cliffs, NJ, 1989).
[16] Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology.
[17] E. H. Rhoderick and R. H. Williams, *Metal-Semiconductor Contacts* (Oxford University Press, Oxford, 1988).

Reduced carrier cooling and thermalization in semiconductor quantum wires

L. Rota and F. Rossi
Dipartimento di Fisica, Universita di Modena, Via Campi 213/A, 41100 Modena, Italy

S. M. Goodnick
Department of Electrical and Computer Engineering, Oregon State University, Corvallis, Oregon 97331

P. Lugli
Dipartimento di Ingegneria Elettronica, II Universita di Roma, Via E. Carnevale, 00173 Roma, Italy

E. Molinari*
Istituto "O. Corbino," Consiglio Nazionale delle Ricerche, Via Cassia 1216, 00189 Roma, Italy

W. Porod
Department of Electrical and Computer Engineering, University of Notre Dame, Notre Dame, Indiana 46556
(Received 5 June 1992)

By using a Monte Carlo analysis of the carrier relaxation in GaAs quantum wires following laser photoexcitation, we show that carrier cooling due to phonon emission and internal thermalization due to electron-electron interaction are significantly decreased with respect to bulk systems. This decreased thermalization is mainly attributed to the reduced efficiency of intersubband processes and to the reduced effect of electron-electron intrasubband scattering.

Most studies of transport in quantum-wire (QWR) and other mesoscopic systems have been primarily concerned with the near-equilibrium linear-response regime where quantum coherence effects are evident (see, for example, Ref. 1 and references therein). However, Cingolani and co-workers[2] have recently reported ultrafast pump and probe absorption studies of a quantum-wire system in which carrier relaxation was observed to be considerably slower than found in bulk systems under similar excitation conditions. Under such conditions, the carrier distribution is far from equilibrium, and many subbands of the quasi-one-dimensional (quasi-1D) wire system are occupied. The average energy of the carriers is well above that of the lattice, and phonon emission is dominant as carriers relax their energy in the return to equilibrium.

In such quasi-1D structures, the electron-phonon-scattering rate is substantially modified by quantization of the carrier motion due to the two-dimensional confining potential, as well as by changes in the Fröhlich-interaction Hamiltonian caused by phonon confinement and localization.[3] Calculations of the electron-phonon-scattering rates in quasi-1D systems have been reported in the literature.[3-6] Both intrasubband and intersubband scattering rates for bulklike polar optical phonons interacting with quantized electrons were reported, including self-energy corrections to account for broadening of the final density of states.[3,6] More recently, scattering rates have been calculated for both quantized electrons and confined phonons in the 1D structure.[4] Monte Carlo (MC) simulation of the carrier dynamics in quasi-1D systems under an applied electric field have been reported, and show features in the nonequilibrium distribution function arising from the reduced dimensionality of the system[5,6] as well as real-space transfer out of the 1D system.[7]

The aim of the present Brief Report is to extend the analysis of nonequilibrium charge transport in quasi-1D systems by studying carrier relaxation following laser excitation. In particular, we consider the role of electron-electron scattering in the thermalization of the photoexcited carriers, an effect which has been neglected in previous treatments.[5-7] Here the carrier dynamics is modeled using a k-space ensemble Monte Carlo simulation,[8] which allows a semiclassical transient analysis of the short-time evolution of the one-particle carrier distribution function during and after the generation of electron-hole pairs by the time-dependent optical pulse.

As a model system, we consider QWR structures such as those fabricated from multiple-quantum-well (MQW) samples by chemical etching.[2] This technique gives a GaAs rectangular quantum wire confined in the vertical growth direction (z) by $Al_xGa_{1-x}As$ layers, and free standing along the transverse direction (y). In the present MC simulation of this structure, electron-polar optical phonon (EP) scattering as well as the electron-electron (EE) interaction are included for both intrasubband and intersubband transitions. The EP and the EE interactions have been shown to dominate the subpicosecond transient regime in bulk GaAs (Refs. 9 and 10) and in $GaAs/Al_xGa_{1-x}As$ MQW's (Refs. 11-13) for excitation energies below the threshold for intervalley transfer. We have deferred the study of scattering due to the photoexcited holes until later work to focus attention on the EE interaction alone.

Scattering due to the EE and EP interactions is treated using time-dependent perturbation theory within the ap-

proximation of Fermi's golden rule. To first order, the net scattering rate due to confined phonons in QW systems differs little from that calculated using bulk phonons, at least for well widths of about 100 Å or more. In the present work, we thus neglect phonon confinement effects due to the heterojunction interface and assume that the phonon modes are infinitely extended in the vertical z direction. Fuchs and Kliewer slab modes[24] are assumed in the lateral direction confined by air on either side. Broadening of the energy levels is also included to account for inhomogeneities in the wire width as discussed elsewhere.[24]

Carrier-carrier scattering in the QWR system is assumed to occur via a statically screened Coulomb interaction (the only affordable method in MC simulations), which is similar to that used in 3D (Ref. 5) and quasi-2D systems.[11] The unscreened matrix element of the bare Coulomb potential between particle 1 of momentum k_1 in subband i with a second particle 2 of momentum k_2 initially in subband j is given by

$$H^{ee}=\frac{e^2}{2\pi\epsilon L}\int dz\int dy\int dz'\int dy'\varphi_i^*(y,z)\varphi_m^*(y,z)\varphi_j^*(y',z')\varphi_n^*(y',z')$$
$$\times\delta(k_1+k_2-k_1'-k_2')K_0(q_x\sqrt{(y-y')^2+(z-z')^2}), \tag{1}$$

where $\varphi_i(y,z)$ is the free-electron wave function in subband i, K_0 is the zeroth-order modified Bessel function, q_x is the wave vector exchanged in the x direction, $k_1'-k_1$, and L is the simulated length of the wire. The final subbands after scattering are m for particle 1 and n for particle 2. The total scattering rate for an electron of wave vector k_1 in a given subband i is given by the sum over all final states for this electron and over all initial and final states of the partner electron. Neglecting for the moment degeneracy effects (degeneracy is accounted for later with a rejection technique), and using conservation of energy and momentum, the EE scattering rate becomes

$$\Gamma(k_1,E_1)=\frac{e^2}{2\pi\epsilon\hbar^2L}\sum_{jmn}\int\left[\frac{H^{ee}(q_1)^2}{\left|\frac{q_1}{\mu'}-\left|\frac{k_1}{m_1'}-\frac{k_2}{m_2'}\right|\right|}-\frac{H^{ee}(q_2)^2}{\left|\frac{q_2}{\mu'}-\left|\frac{k_1}{m_1'}-\frac{k_2}{m_2'}\right|\right|}\right]dk_2', \tag{2}$$

where μ is the relative mass of the two carriers after scattering, and q_1 and q_2 are the two possible values for the exchanged momentum:

$$q_x=-\mu'\left|\frac{k_1}{m_1'}-\frac{k_2}{m_2'}\right|\pm\mu'\left[\frac{k_1^2}{m_1'^2}-\frac{k_2^2}{m_2'^2}-\frac{2k_1k_2}{m_1'm_2'}-\frac{k_1^2}{\mu'}\left|\frac{1}{m_1'}-\frac{1}{m_1}\right|-\frac{k_2}{\mu'}\left|\frac{1}{m_2'}-\frac{1}{m_2}\right|-\frac{2\Delta}{\mu'\hbar^2}\right]^{1/2}, \tag{3}$$

where $\Delta=E_1'+E_2'-E_1-E_2$ is the difference in the subband energy levels. As is evident from Eq. (3), for every pair of interacting electrons there are only two final states available. If, at the end of the scattering, the two carriers remain in their respective subbands, the only final state available is that in which the electrons exchange their crystal momentum. Such scattering is irrelevant if both particles originate from the same subband, since the two particles are indistinguishable. If the two electrons are in different subbands, this interaction produces a significant energy exchange between the subbands. We shall still refer to intrasubband scattering as the interaction of two electrons that lie in different subbands, as long as they remain in the same subbands after scattering. Intersubband scattering then refers to the case in which one or both of the electrons change subbands. The relative strength of intrasubband versus intersubband scattering depends on the magnitude of the EE form factor given by Eq. (1). Figure 1(a) shows a plot of this form factor as a function of q_x for several different types of intrasubband and intersubband transitions illustrated schematically in Fig. 1(b). The continuous line represents the intrasubband scattering between electrons in the first and the second subbands. The form factor peaks at small q_x, and is much larger than the intersubband form factors (shown by the dashed and dotted lines, respectively) due to the orthogonality of the eigenfunctions. It is interesting to note that this type of scattering never creates new states in k space, and thus does not produce any thermalizing effect, only an energy redistribution between different subbands. The only scattering that enables a generation of new k states (and thus momentum relaxation) occurs for intersubband scattering. In this case, the difference of energy between the initial and final subbands enters into the conservation law and allows the creation of two completely new states. A typical transition is shown in Fig. 1(b) by the dashed line. As mentioned earlier, such intersubband scattering is greatly reduced due to the small form factor, and so thermalization due to carrier-carrier scattering is significantly reduced compared to that in higher-dimensional systems.

The screened scattering rate can be computed using the bare Coulomb potential dressed by the dielectric function of the carrier system:

$$H^{ee}(q)\rightarrow\frac{H^{ee}(q)}{\left|\epsilon_{j,n}^{i,m}(q,\omega)\right|},$$

where ϵ is the dielectric matrix of the multisubband system. The static, long-wave limit is taken which allows

the analytic determination of the determinant above.

The MC method adopted here is substantially the same as that used in the simulation of bulk GaAs and GaAS/$Al_xGa_{1-x}As$ QW's,[8] with free motion in only the x direction. A rejection technique has been applied to take into account the Pauli exclusion principle in the final state after scattering, similar to that used in the 3D and 2D cases.[8,11] As a model system, we have considered a 100-Å GaAs layer cladded by $Al_{0.3}Ga_{0.7}As$ layers on either side in the z direction. The lateral width is chosen to be 300 Å, which gives a total of ten 1D subbands in the well.

In order to simplify the analysis, we have considered only photoexcitation from the heavy-hole band to the conduction band. The laser excitation is assumed to have a time duration of 50 fs and a spectral width of 20 meV. All the simulations were performed at a lattice temperature of 10 K with an equivalent density of 10^{17} cm^{-3}. Electrons are generated in all the energetically accessible subbands (assuming that the $\Delta n = 0$ rule holds for transition between valence- and conduction-band subbands proportional to the final density of states).

In Fig. 2 we show the results of a simulation performed with a low excitation energy (43 meV) where only EE scattering is considered. The results are plotted versus the magnitude of k to illustrate the difference between the simulation results and those of a heated Maxwell-Boltzmann distribution (shown by the dotted line) of the same average energy as the injected distribution. In the QWR case [Fig. 2(a)], very little thermalization has occurred, even after 1 ps. In contrast to this result, the bulk case [Fig. 2(b)] is completely thermalized after 400 fs. The result illustrates the greatly reduced effect of intrasubband EE scattering in a 1D system in contrast to higher-dimensional systems due to the reduced phase space of the final state. Since thermalization in the presence of EE scattering alone may only proceed via intersubband events in a QWR system, the time scale is much longer than that required in 2D and 3D.

Figure 3 shows the combined effect of the EP and EE interactions, for both the QWR [Fig. 3(a)] and the bulk [Fig. 3(b)], respectively, when a higher excitation energy is used (170 meV). The peaks present in the QWR simulation correspond to the position of the subband minima indicated by the arrows shown in Fig. 3(a). The distribution function in the bulk is broadened with respect to the QWR case during the laser excitation, due to the strong effect of the EE interaction, and is completely thermalized after 300 fs. Again, the dotted lines in parts (a) and (b) correspond to a heated Maxwellian distribution at the temperature of the electron gas. In the QWR, the distribution function approaches that of heated Maxwellian only after 1 ps. The presence of both scattering mechanisms significantly increases the thermalization rate compared to that of EE scattering alone [Fig. 2(a)], since the emission of polar optical phonons allows the electrons to access new momentum states in conjunction with intrasubband EE scattering which redistributes energy among the various subbands. The combined thermalization rate, however, is still much slower than in an

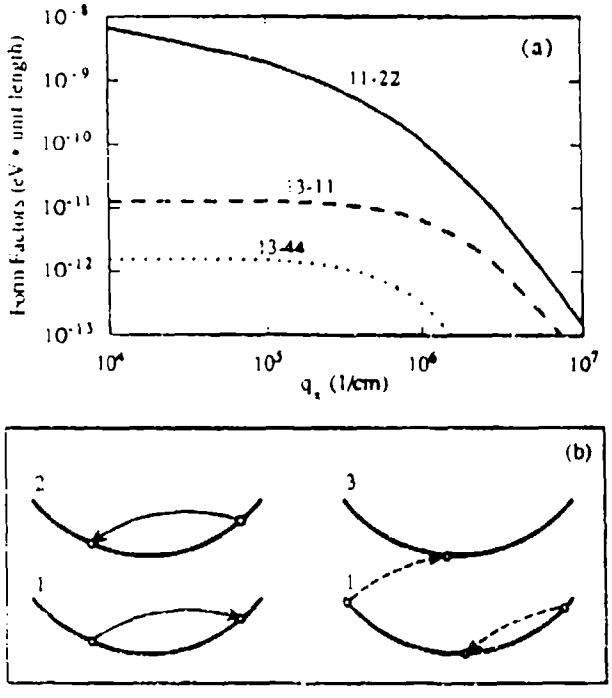


FIG. 1. (a) The form factor for EE transitions, Eq. (1), showing the intrasubband (solid line) contribution for two electrons lying in subbands 1 and 2, respectively, and the intersubband ones (dotted and dashed lines) for an electron jumping from subband 1 to 3 while the other remains in subband 1 (dashed line) or 4 (dotted line). (b) Schematic representations of the first two electronic transitions described above.

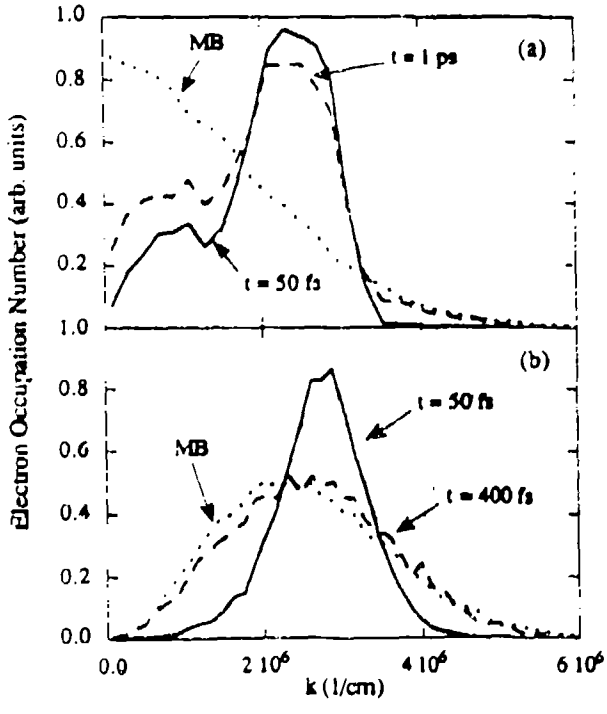


FIG. 2. (a) Electron occupation number in the QWR at two different times when only EE interaction is included as a function of k_x. The dotted line represents a heated Maxwellian at the temperature of the electron gas. (b) The same condition simulated in the bulk (in this case k is the modulus of the total momentum).

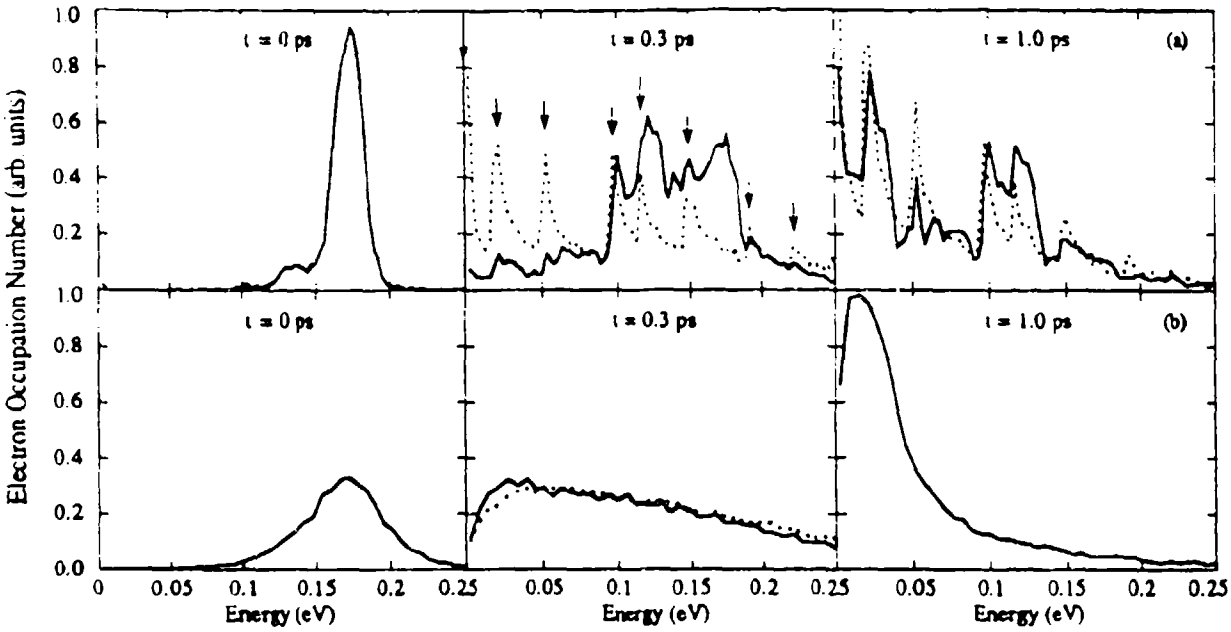


FIG. 3. (a) Electron occupation number in the QWR at three different times after the excitation including both EP and EE scattering. (b) The same conditions above repeated for the bulk.

equivalent bulk system, as illustrated by the comparison in Fig. 3. This slow thermalization in the QWR system is consistent with recent measurements[2] where a very slow thermalization was found.

In summary, we have presented a Monte Carlo study of photoexcited carrier relaxation in a quantum wire with realistic dimensions in which electron-electron and electron-polar optical phonon scattering is included. The energy relaxation rate due to electron-phonon scattering is reduced due to the relative strength of the intersubband versus intrasubband scattering. A strong reduction in the efficiency of the single-particle intercarrier interaction is shown, which results in a slow internal thermalization of the photoexcited electrons.

This work was partially supported by "Consorzio Nazionale di Fisica della Materia (INFM)," and by the EEC Commission under the Esprit Basic Science Project "NANOPT."

*Present and permanent address: Dipartimento di Fisica, Universita di Modena, via Campi 213/A, 41100 Modena, Italy.

[1] *Nanostructure Physics and Fabrication*, edited by M. A. Reed and P. Kirk (Academic, Boston, 1989).
[2] R. Cingolani, H. Lage, L. Tapfer, H. Kalt, D. Heitmann, and K. Ploog, Phys. Rev. Lett. 67, 891 (1991).
[3] U. Bockelmann and G. Bastard, Phys. Rev. B 42, 8947 (1990).
[4] M. A. Stroscio, Phys. Rev. B 40, 6428 (1989); K. W. Kim, M. A. Stroscio, A. Bhatt, R. Mickevicius, and V. V. Mitin, J. Appl. Phys. 70, 319 (1991).
[5] S. Briggs and J. P. Leburton, Phys. Rev. B 38, 8163 (1988).
[6] D. Jovanovic, S. Briggs, and J. P. Leburton, Phys. Rev. B 42, 11 108 (1990).
[7] S. Briggs and J. P. Leburton, Phys. Rev. B 43, 4785 (1991).
[8] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer-Verlag, Wien, 1989).
[9] T. Elsaesser, J. Shah, L. Rota, and P. Lugli, Phys. Rev. Lett. 66, 1757 (1991).
[10] W. Z. Lin, W. Schoenlein, J. G. Fujimoto, and E. P. Ippen, IEEE J. Quantum Electron. QE-24, 267 (1988).
[11] S. M. Goodnick and P. Lugli, Phys. Rev. B 37, 2578 (1988).
[12] S. M. Goodnick and P. Lugli, Phys. Rev. B 38, 10 135 (1988).
[13] S. M. Goodnick and P. Lugli, in *Hot Carriers in Semiconductor Nanostructures: Physics and Applications*, edited by J. Shah (Academic, New York, 1992), pp. 191–234.
[14] R. Fuchs and K. L. Kliewer, Phys. Rev. 140, A2076 (1965).

FIELD EFFECTS IN SELF-CONSISTENT TRANSPORT CALCULATIONS FOR NARROW SPLIT-GATE STRUCTURES*

Henry K. Harbury, Wolfgang Porod, and Craig S. Lent
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

(Received 19 May 1991)

We study local potential variations due to self-consistent space-charge effects in calculations of coherent transport in narrow split-gate structures. We present a numerical technique based on calculating the Hartree potential from the charge density obtained by solving the two-dimensional effective-mass Schrödinger equation for scattering states, and from the bound charge density obtained from a semi-classical Thomas-Fermi screening model. This method allows us to obtain the local self-consistent potential variations close to scattering centers exposed to an incident flux of electrons.

1. Introduction

Many recent studies have appeared in the literature which present calculations of conductances in narrow split-gate structures that can be realized using advanced fabrication techniques [1]. In these calculations the Schrödinger equation is solved for the channel geometry to obtain transmission coefficients, from which the conductance is obtained. It has been pointed out [2] that self-consistent effects may be significant in these structures, but it has been prohibitively difficult to include them in the numerical calculations.

We present a technique for capturing self-consistent space-charge effects in two-dimensional transport calculations in split-gate structures. We study the self-consistent variations in carrier density and local potentials close to scattering centers exposed to an incident flux of electrons.

2. Numerical Method

Self-consistent electronic states are obtained through an iterative solution of the 2-D Schrödinger equation using current carrying boundary conditions. Any bound electronic charge is approximated by the semi-classical Thomas-Fermi screening model. The Hartree potential due to the electron density is explicitly determined from the total charge density in each iteration, and is used to calculate the electronic states in the subsequent iteration. The self-consistent solution is obtained when the potential difference between iterations converges to within a desired tolerance.

A. Problem Domain

Our model of the device geometry consists of a discretized calculation domain, Ω_c, of mesh elements connected to input and output leads, Ω_1 and Ω_2, as shown in Fig. 1. The current carrying leads attach to the edges of the calculation domain, at $x = 0$ and $x = l$, and extend to infinity. Although the leads are not part of the calculation domain, the charge contained in the leads must be included in the numerical model for the Hartree potential. For simplicity, we assume hard-wall boundary conditions for the conducting channel. A split-gate structure is realized by adding a large fixed potential in the gate regions of the problem. The size of the calculation domain must be sufficiently large that the disturbance due to the constriction has decayed at $x = 0$ and $x = l$. Selecting device geometries that are symmetric across the channel enables us to solve the problem on only one half of the original domain. The resulting bandwidth reduction significantly lowers the computation time.

B. Effective Mass Schrödinger Equation

The 2-D effective-mass Schrödinger equation is repeatedly solved on the discretized spatial domain for the traveling wave "scattering states". V_0 denotes the confining

0749-6036/92/020189+05 $02.00/0

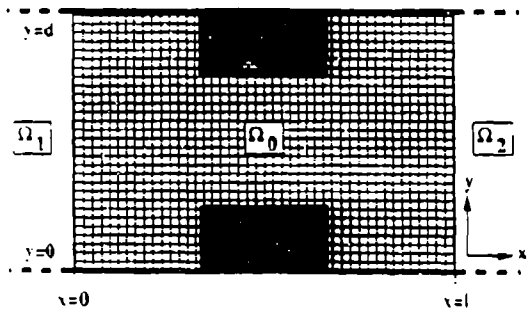


FIGURE 1. Schematic of the channel geometry showing the finite element domain, Ω_0, and the leads, Ω_1 and Ω_2. The shaded regions represent the confining gate potentials.

channel potential, V_H represents the self-consistent Hartree potential, and m^* is the electron effective-mass,

$$\left[-\frac{\hbar^2}{2m^*}\nabla^2 + V_{conf}(x,y) + V_H(x,y)\right]\psi_E(x,y) = E\psi_E(x,y). \quad (1)$$

Deep inside the leads, i.e. far from the scattering region, we assume free electron propagation. The electronic wave function can be factorized into a longitudinal plane-wave part, $e^{\pm ikx}/\sqrt{k}$, and a transverse part with quantized modes, $\chi_n(y)$, which are due to the confining potential of the channel. The scattering eigenstates for injection from the left-hand lead, Ω_1, in mode n are, therefore, of the form,

$$\psi_E(x,y) = \begin{cases} \frac{1}{\sqrt{k_n}}e^{ik_n x}\chi_n(y) + \sum_m r_{nm}\frac{1}{\sqrt{k_m}}e^{-ik_m x}\chi_m(y) & x \to -\infty \\ \sum_m t_{nm}\frac{1}{\sqrt{k_m}}e^{ik_m x}\chi_m(y) & x \to +\infty \end{cases} \quad (2)$$

The index n runs over all possible transverse modes with $E = E_n + \hbar^2k_n^2/2m^*$, which includes both traveling and evanescent states.

The finite element method is used to reduce the problem to a large system of coupled linear equations. The Quantum Transmitting Boundary Method [3] is used to implement the finite element boundary conditions appropriate for the current carrying leads.

C. Charge Density

The electron density in the channel due to current carrying states is obtained by explicitly summing over the contributions of all occupied scattering states. We assume the linear response regime and zero temperature. The scattering state electron density is then given by summing over all states with $E \le E_F$,

$$n_{SC}(x,y) = \sum_{E \le E_F} |\psi_E(x,y)|^2. \quad (3)$$

The sum includes the wave-functions due to injection from both lead 1 and lead 2. In the linear response regime, injection is symmetric for both leads and this can be used to reduce computation time. We typically find that twenty terms in the summation over electronic states are sufficient.

Bound electronic states can exist inside the scattering region if the total potential dips below its asymptotic value far into the leads. The quantum mechanical electron density obtained for traveling states does not contain any contribution that may arise from bound charge inside the problem domain, Ω_0. If this contribution is neglected the electron density would be underestimated in the region of a potential well, resulting in a larger attractive Hartree potential and subsequently an unstable runaway condition. As a first order approximation, we choose to model the bound-state electronic charge density with the semi-classical Thomas-Fermi screening model. If the potential in the solution domain is lower than its asymptotic value in the leads, $V_{tot}(x,y) < V_{tot}(0,y)$, then a semi-classical electron density is obtained at the point (x,y) by $n_{TF}(x,y) = \frac{m^*}{\pi\hbar^2}[V_{tot}(0,y) - V_{tot}(x,y)]$, where V_{tot} is the total potential used in solving the Schrodinger equation. This simple approximation over-estimates the total electron density in the region of a potential well because we fill a discrete quantum system with a continuous electron density. The over-estimation, however, will cause the potential well to become more "shallow" in the subsequent iteration resulting in a smaller semi-classical electron density.

To maintain charge neutrality, a compensating positive background charge, n_B, is needed. The background charge is obtained by enforcing charge neutrality in the leads. The local disturbance caused by any constriction is assumed to be completely contained within the problem domain, Ω_0, with negligible effect in the leads. This criterion allows us to model the charge density in the infinitely extending leads as a continuation of the charge density at the edges of the problem domain, $n_B = \int_0^1 n_{SC}(0,y)\,dy = \int_0^1 n_{SC}(l,y)\,dy$.

This model for the positive background neglects the fact that in most 2DEG systems the positive donors are in a parallel plane isolated from the carriers by a spacer layer. Although we do not model the spacer layer in these calculations, one could, in principle, use an "effective" positive background charge density without changing the algorithm [4].

D. Hartree Potential

The Hartree potential at any location (x,y) is obtained by explicitly summing over all charge contributions in the channel,

$$V_H(x,y) = \frac{e}{4\pi\varepsilon}\iint_{\Omega_0} \frac{n(x',y')\,dx'\,dy'}{\sqrt{(x'-x)^2 + (y'-y)^2}} \quad (4)$$

The integral can be broken into three distinct contributions, $V_H = V_H^{\Omega_0} + V_H^{\Omega_1} + V_H^{\Omega_2}$, corresponding to the domain, Ω_0, and the two leads, Ω_1 and Ω_2, respectively. The term due to the charge density inside the solution domain, $V_H^{\Omega_0}$, is explicitly calculated by Gaussian quadrature on the finite elements. The components of the Hartree potential due to the infinite leads, $V_H^{\Omega_1}$ and $V_H^{\Omega_2}$, must be, in part, calculated analytically. Using the conditions of charge neutrality across the leads, $\int_0^L \rho(l, \eta) = 0$, and of translational invariance down the leads, $\rho(x > l, \eta) = \rho(l, \eta)$, yields

$$V_H^{\Omega_2}(x, \eta) = \frac{\qquad\qquad}{-\frac{}{}\int_0^{\infty} dl' \rho(l, \eta) \ln(\quad)} \tag{5}$$

The above integral across the lead is evaluated numerically from the charge density at the edge of the finite element domain, $\rho(l, \eta)$. We obtain a similar Hartree term due to the charge density in the left lead, $V_H^{\Omega_1}$.

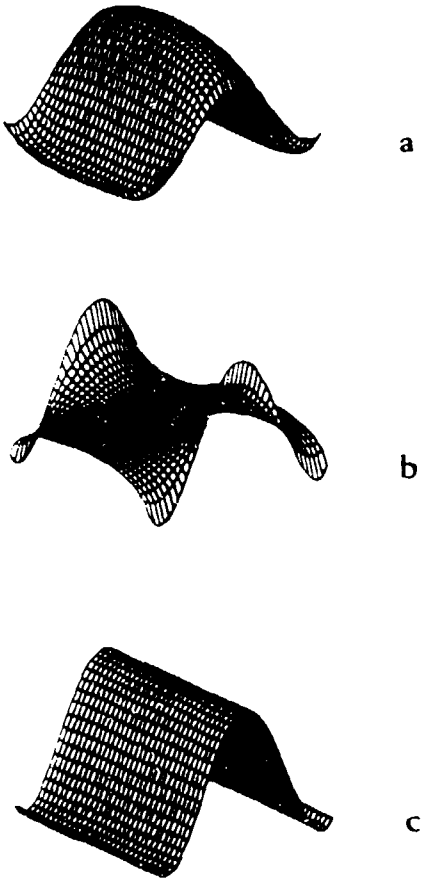


FIGURE 2. The components of the Hartree potential for the open channel after the first iteration: (a) $V_H^{\Omega_0}$, (b) $V_H^{\Omega_1} + V_H^{\Omega_2}$, (c) V_H.

E. Self-Consistency

The resulting Hartree potential is used to modify the total potential on which the Schrödinger equation is solved in the subsequent iteration. The self-consistent potential and charge density are obtained when the difference in the total potential between iterations converges to within a specified tolerance.

3. Results

A. Open Channel

We first investigate the electronic charge density and the Hartree potential for an open, unconstricted channel. The solutions must show translational invariance down the channel, namely $\rho(x, \eta) = \rho(\eta)$ and $V_H(x, \eta) = V_H(\eta)$. This provides a test for the Hartree potential obtained by summing $V_H^{\Omega_0}$, $V_H^{\Omega_1}$, and $V_H^{\Omega_2}$. We choose to study single mode injection by occupying states up to a Fermi energy that is below the second subband in the leads. Fig. 2 shows in part (a) the Hartree potential due to the solution domain $V_H^{\Omega_0}$ and in part (b) the sum of both lead contributions, $V_H^{\Omega_1} + V_H^{\Omega_2}$, after the first iteration. The total Hartree potential, which is the sum of (a) and (b), results in the potential shown in Fig. 2(c) and exhibits the desired trans-

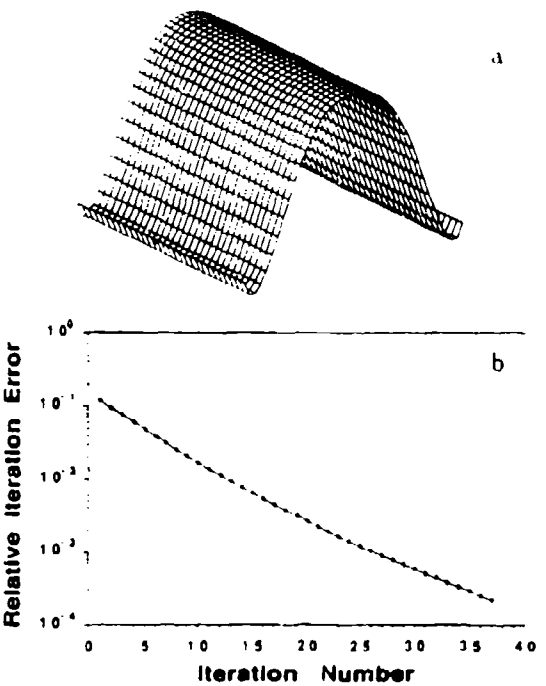


FIGURE 3. Shown is (a) the converged open channel potential, and (b) the iteration error.

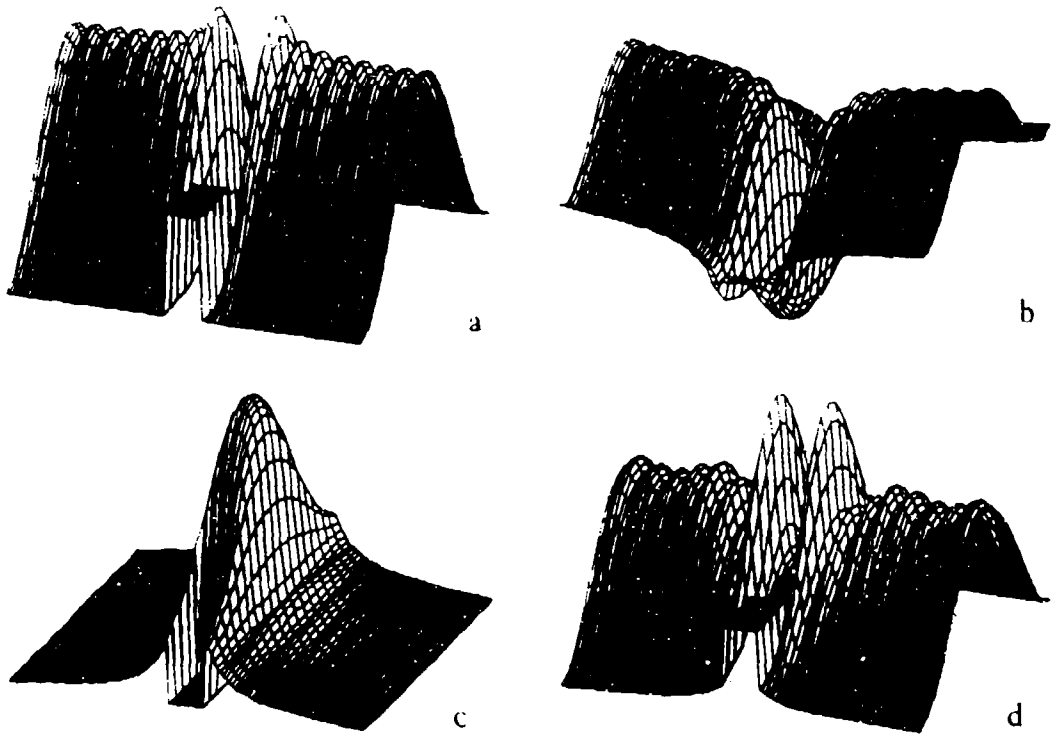


FIGURE 4 Shown is (a) the charge density for the constricted channel (WNW structure) after the first iteration, (b) the new total potential after the first iteration, (c) the semi-classical electron density in the second iteration, and (d) the total charge density after the second iteration.

tational invariance. This clearly shows that the Hartree potential due to charge in the leads exactly compensates for the potential due to charge within the domain, and is a very critical test of our numerical method. Because the total potential in the solution domain, U_0, does not drop below its asymptotic value in the leads, there are no bound electronic states in the open channel. The Hartree potential in Fig. 2(c), obtained from the completely quantum mechanical electron density, is used to modify the total potential with which the Schrodinger equation is solved on the subsequent iteration. Fig. 3(a) shows the converged solution after 37 iterations, and Fig. 3(b) plots the change in the total potential between iterations versus iteration number. The algorithm is stable and convergent for the open channel and, as Fig. 3(a) indicates, the "floor" of the quantum wire is "buckled" by several meV.

B. Constricted Channel

We finally investigate a partially constricted channel, our model for a narrow split-gate structure. We show in Fig. 4(a) the charge density after the first iteration for single mode injection, and in Fig. 4(b), (c) and (d) the new total potential used in the second iteration, the semi-classical electron density, and the total charge density after the second iteration respectively.

The first iteration charge density far from the constriction asymptotically approaches the case of an undisturbed system. The leakage of charge between the split gates in the first iteration is small enough that the charge density in Fig. 4(a) resembles the solution of the totally constricted channel [5]. The smaller electron density in the gate region in Fig. 4(a) results in the formation of a potential well through the Hartree term used to obtain the total potential shown in Fig. 4(b). This potential well in the gate region will, in subsequent iterations, cause an increase in the local electron density near the scattering center. This results both from an increase in occupied traveling states through the gate region and from the semi-classical model for the bound electron density, shown in Fig. 4(c), since the potential in the gate region dips below the po-

tential far into the leads. The total charge density obtained in the second iteration, due to both the electron density of the traveling states and the semiclassical bound electron density, is shown in Fig. 4(d). The increase in the electron density near the scattering center is clearly visible. Although not fully converged, it is clear from these calculations that the oscillations in the local field near the scattering center due to self-consistent space charge effects could be significant in transport calculations since they will effect the transmission amplitudes of the structure. The oscillations in the potential due to a scattering center in the 700Å channel can extend up to a half micron into the leads, and are on the order of a few meV. Although the potential well in the scattering region does not become deeper in later iterations, we have found it difficult to achieve convergence for split-gate structures. We are currently investigating various convergence schemes and more work is needed on this problem.

4. Conclusion

We have presented a method which incorporates self-consistent space charge effects in transport calculations for narrow split-gate structures. Our method uses an iterative solution of the effective-mass Schrödinger equation, with a semi-classical model for bound electronic charge, and explicit calculation of the Hartree potential. This method allows us to investigate the local self-consistent potential variations close to scattering centers exposed to an incident flux of electrons.

For the case of an open channel, or quantum wire, we have demonstrated the validity of our method and achieved a fully convergent solution. We presented some results for a partially constricted channel, or narrow split-gate structure, but for this case convergence has been difficult to achieve. More work is needed to ensure convergence of the self-consistent cycle for general split-gate structures.

5. References

[1] *Nanostructure Physics and Fabrication*, eds. M. A. Reed and W. P. Kirk, (Academic Press, 1989).

[2] L. Escapa and N. Garcia, *Appl. Phys. Lett.* **56**, 901 (1990).

[3] C. S. Lent and D. J. Kirkner, *J. Appl. Phys.* **67**, 6353 (1990).

[4] J. A. Nixon and J. H. Davies, *Phys. Rev. B* **41**, 7929 (1990).

[5] H. K. Harbury, W. Porod, C. S. Lent, *Proceedings of the Seventh International Conference on the Numerical Analysis of Semiconductor Devices and Integrated Circuits, NASECODE-7*, p. 136, ed. by J. J. H. Miller, (Front Range Press, 1991).

Spatial distribution of current and Fermi carriers around localized elastic scatterers in quantum transport

S. Chaudhuri and S. Bandyopadhyay
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

M. Cahay
Department of Electrical and Computer Engineering, University of Cincinnati, Cincinnati, Ohio 45221
(Received 26 September 1991)

The two-dimensional spatial distribution of the current and Fermi carriers around localized elastic scatterers in phase-coherent electron transport has been calculated using a generalized scattering-matrix approach. The distributions show dramatic differences depending on whether the scatterers are attractive (donorlike) or repulsive (acceptorlike). We find that attractive scatterers can produce strong *vortices* in the current, resulting in localized magnetic moments, while repulsive scatterers produce much weaker vortices and may not produce any at all in quasiballistic transport (few impurities). This is a significant difference between majority-carrier transport (when the scatterers are attractive) and minority-carrier transport (scatterers are repulsive). The vortices are caused by quantum-mechanical interference between scatterers and are accentuated by evanescent modes which have a stronger effect in the case of attractive scatterers owing to the formation of quasidonor states. We also examine the influence of the impurity configuration (positions of the scatterers) on the nature of the vortices.

I. INTRODUCTION

The spatial distribution of the current and electric field around localized scatterers in electron transport has been a subject of significant interest even in the late 1950s.[1] After a period spanning more than two decades, it is again receiving attention[2] following the advent of mesoscopic systems that exhibit phase-coherent electron transport. In this paper, we present results based on a rigorous solution of the Schrödinger equation that show the spatial distribution of current and Fermi carriers (i.e., carriers with the Fermi energy) in a two-dimensional semiconductor of finite width in the presence of δ scatterers. The analysis is fully quantum mechanical and incorporates the effects of phase coherence of the electron wave function—a feature that was neglected in the previous work of Refs. 1 and 2. Our results indicate that there are dramatic differences between the current distributions for attractive and repulsive scatterers. Similar differences are also observed in the distributions of Fermi carriers. In addition, we find that current vortices can form in the vicinity of impurities resulting in localized magnetic moments. These vortices are caused by quantum-mechanical interference between elastic scatterers and require the presence of evanescent modes. They also affect the Fermi carrier concentration. Finally, we discuss how the localized magnetic moments associated with the vortices may lead to interesting features in low field quantum magnetotransport.

This paper is organized as follows. In the next section, we briefly present the theory used in computing the two-dimensional current and Fermi carrier-concentration profiles. In Sec. III, we present results for a prototypical structure which consists of a disordered GaAs mesa (a two-dimensional structure confined along the width) whose width is comparable to the Fermi wavelength and in which transport is multichanneled. The results consist of the computed conductance versus Fermi energy, which show quantized steps associated with subband filling in quasiballistic transport.[3] The nature of these steps is very different for attractive and repulsive scatterers and depends significantly on whether or not evanescent modes (i.e., transport channels with imaginary wave vectors) associated with unfilled subbands above the Fermi level are included in the analysis. In addition, we have plotted the current and Fermi carrier-concentration profiles for injection from each filled subband individually as well as the total current and Fermi carrier-concentration profiles. These plots show that strong current vortices can form in the presence of attractive scatterers in both quasiballistic and diffusive transport. When the scatterers are repulsive, vortices usually do not form in the quasiballistic regime, but weak vortices can occur in the diffusive regime. We also show how the vortices are influenced by the impurity configuration and discuss the relationship they have with quasidonor states that form when the scatterers are attractive. Finally, in Sec. IV, we present our conclusions and discuss interesting features that may be observed in the presence of a magnetic field.

II. THEORY

We consider a mesa as shown in Fig. 1. The structure is single moded in the z direction, which means that the thickness along this direction is less than the Fermi wavelength, so that only one transverse subband in this direction is occupied. The width W along the y direction can be larger than the Fermi wavelength, so that multiple y-

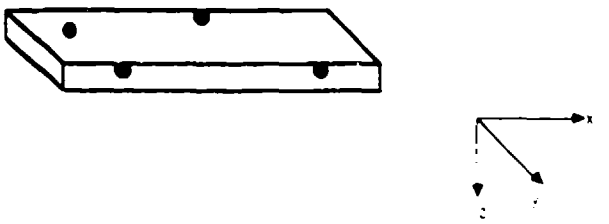


FIG. 1. A GaAs mesa of length 1160 Å and width 1000 Å with a random distribution of four impurities inside.

directed subbands may be filled. The structure contains a number of elastic δ-function scatterers whose positions (i.e., the impurity configuration) are generated by two independent uniform random number generators (for the x and y coordinates). In all our analysis we assume a temperature of $T=0$ K and linear-response transport.

The Schrödinger equation describing electron transport in this confined two-dimensional disordered structure is written as

$$-\frac{\hbar^2}{2m^*}\nabla^2\psi+V\psi=E\psi\ , \tag{1}$$

where m^* is an isotropic effective mass and the potential energy V consists of two terms

$$V\equiv V(x,y)=V_c(y)+V_{\rm imp}(x,y)\ . \tag{2}$$

The first term gives the effect of the confining potential in the y direction and the second term is due to impurities. For δ-function impurities, we have

$$V_{\rm imp}=\gamma\sum_{i=1}^{s}\delta(x-x_i)\delta(y-y_i)\ , \tag{3}$$

where s is the total number of impurities in the structure and the summation extends over all the impurities. The position of the ith impurity is denoted by (x_i,y_i).

The assumption of a δ-function shape for impurity potential is of course not realistic for modulation-doped samples where the scattering potential is smooth. More realistic scattering potentials have been treated in the context of quantum transport in narrow channels,[4] but here we are only interested in the basic physics of the spatial distributions. For this, a δ-function potential serves just as well.

The confining potential $V(y)$ gives rise to an orthogonal set of transverse modes in y (or subbands) whose wave functions satisfy the eigenequation

$$\left[-\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial y^2}+V_c(y)\right]\chi_n(y)=\varepsilon_n\chi_n(y)\ , \tag{4}$$

where ε_n is the energy at the bottom of the nth subband. We assume hardwall boundary conditions, which give

$$\chi_n(y)=\sqrt{2/W}\sin\left[\frac{n\pi y}{W}\right]\ , \tag{5}$$

$$\varepsilon_n=\frac{\hbar^2}{2m^*}\left[\frac{n\pi}{W}\right]^2\ . \tag{6}$$

Each subband (or mode) n has a parabolic energy dispersion given by

$$E_n(k_x)=\frac{\hbar^2k_x^2}{2m^*}+\varepsilon_n\ , \tag{7}$$

where k_x is the x-directed wave vector.

The x-directed wave vector k_F^n of the nth mode at the Fermi energy E_F is obtained from the above equation as

$$k_F^n=\frac{\sqrt{2m^*(E_F-\varepsilon_n)}}{\hbar}\ . \tag{8}$$

If $E_F>\varepsilon_n$, then k_F^n is real. We call such a mode a propagating mode, which corresponds to a subband that is occupied at $T=0$ K. If, on the other hand, $E_F<\varepsilon_n$, then the subband is empty and k_F^n is imaginary. We call such a mode an evanescent mode. In the linear-response regime, only the electrons at the Fermi energy carry current. Therefore, the evanescent modes will themselves not carry any current, since their wave vectors are imaginary. However, they can affect the transmission coefficients of the propagating modes that carry current and thus have an indirect but significant effect on the current in the structure.

The total wave function $\psi_m(x,y)$ for an electron injected from the subband m in the left contact with the Fermi energy can be written as

$$\psi_m(x,y)=\sum_{p=1}^{M}\frac{[A_{mp}(x)\exp(ik_F^px)+B_{mp}\exp(-ik_F^px)]\chi_p(y)}{\sqrt{k_F^p}}\ , \tag{9}$$

where the summation extends over both propagating and evanescent modes (the total number of modes is M). When the wave function is written in this form, all propagating modes carry the same current. We now only have to evaluate the coefficients $A_{mp}(x)$ and $B_{mp}(x)$ to determine the wave function. If we know these coefficients at any position (x_0,y) in the structure, then the coefficients at any other position (x_1,y) can be found from

$$\begin{bmatrix}\mathbf{r}_0\\ \mathbf{r}_1\end{bmatrix}=\begin{bmatrix}t_{11} & t_{12}\\ t_{21} & t_{22}\end{bmatrix}\begin{bmatrix}l_1\\ l_0\end{bmatrix}\ , \tag{10}$$

where l_1 is a column vector of length M whose elements are $A_{m1}(x_0)$, $A_{m2}(x_0)$, $A_{m3}(x_0)$, . . . , $A_{mM}(x_0)$, l_0 is a column vector of length M whose elements are $B_{m1}(x_0), B_{m2}(x_0), B_{m3}(x_0), \ldots, B_{mM}(x_0)$, $\mathbf{r}_0$ is a column vector of length M whose elements are $A_{m1}(x_1), A_{m2}(x_1), A_{m3}(x_1), \ldots, A_{mM}(x_1)$, and $\mathbf{r}_1$ is also a column vector of length M whose elements are $B_{m1}(x_1), B_{m2}(x_1), B_{m3}(x_1), \ldots, B_{mM}(x_1)$. The coefficients t_{11}, t_{12}, t_{21}, and t_{22} are each $M\times M$ matrices and the square matrix in Eq. (10) is the so-called transfer matrix for the section of the structure between $x=x_0$ and x_1. A pictorial depiction is shown in Fig. 2.

The above equation allows us to determine the coefficients $A_{mp}(x_1)$ and $B_{mp}(x_1)$ (and hence the wave function) at any arbitrary location x_1 if we know the wave function [or the coefficients $A_{mp}(x_0)$ and $B_{mp}(x_0)$]

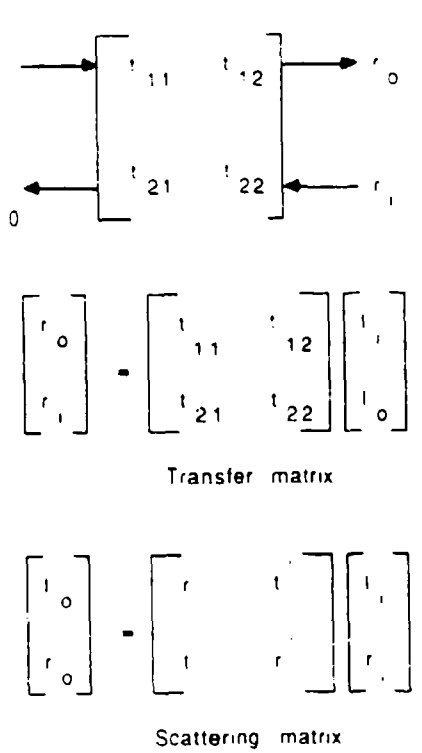


FIG. 2. Incident and reflected waves at any arbitrary section of the structure. The relationships between the waves are expressed through the transfer matrix [Eq. (10)] and the scattering matrix [Eq. (12)].

at any one point $x=x_0$ in the device. We do know the wave function at the left contact ($x=0-$). We assume that for injection into the mth mode, it is given by the scattering states[5] (this is our boundary condition)

$$\psi_m(0-,y)=\frac{\exp(ik_F^m x)\chi_m(y)}{\sqrt{k_F^m}}+\sum_{p=1}^{M}\frac{R_{mp}\exp(-ik_F^p x)\chi_p(y)}{\sqrt{k_F^p}}\ . \tag{11}$$

so that $A_{mp}(x=0-)=\delta_{mp}$ (Krönicker delta) and $B_{mp}(x=0-)=R_{mp}$. Therefore, we can find the wave function anywhere in the structure using Eqs. (10) and (11) if we know the transfer matrix between the left contact and the region of interest, as well as the coefficients R_{mp}. These are found from the appropriate scattering matrices. The scattering matrix for any section relates the amplitudes of the waves reflected from that section to the incident amplitudes (see Fig. 2) according to

$$\begin{vmatrix} l_0 \\ r_0 \end{vmatrix}=\begin{vmatrix} r & t' \\ t & r' \end{vmatrix}\begin{vmatrix} l_i \\ r_i \end{vmatrix}\ , \tag{12}$$

where t, t', r, and r' are the elements of the scattering matrix and are themselves $M\times M$ matrices. The procedure to derive the scattering matrix for any arbitrary section was described in Ref. 6 and will not be repeated here. The coefficients R_{mp} are simply the elements of the matrix r appearing in the total scattering matrix for the entire structure and can be found directly. The only other unknowns that we need are the elements of the transfer matrix appearing in Eq. (10) which are related to the elements of the scattering matrix for that section by the simple relations

$$t_{11}=t-r'(t')^{-1}r\ ,\quad t_{12}=r'(t')^{-1}\ ,$$
$$t_{21}=-(t')^{-1}r\ ,\quad t_{22}=(t')^{-1}\ . \tag{13}$$

Once we have evaluated the scattering matrices, we have therefore found all the quantities that we need to evaluate the wave function everywhere in the structure. Equations (9)–(13) are used to determine the wave function. This method of finding the wave function, although straightforward, is, however, computationally not robust. A problem arises when evanescent modes are included in the analysis, which causes the matrix t' to become near singular and makes its inversion [required in Eq. (13)] difficult. To avoid this problem, an alternate approach was used in obtaining our results. This is described below.

To find the wave function $\psi_m(x,y)$ at any position (x,y), we need to know the coefficients $A_{mp}(x)$ and $B_{mp}(x)$. For this, we break up the structure into two sections—one interposed between the left contact and x, and the other between x and the right contact. Referring to Fig. 3, we can relate the column vector α_m [containing the coefficients $A_{mp}(x)$] and the column vector β_m [containing the amplitudes $B_{mp}(x)$] to the column vector $\mathbf{a}_m$ [containing the amplitudes $A_{mp}(x=0-)$ at the left contact] and the column vector $\mathbf{b}_m$ [containing the amplitudes $B_{mp}(x=0-)$] through the scattering matrix

$$\begin{vmatrix} \mathbf{b}_m \\ \alpha_m \end{vmatrix}=\begin{vmatrix} r_1 & t_1' \\ t_1 & r_1' \end{vmatrix}\begin{vmatrix} \mathbf{a}_m \\ \beta_m \end{vmatrix}\ . \tag{14}$$

The above scattering matrix describes the first section. Similarly we can write a scattering matrix for the second section as (see Fig. 3)

$$\begin{vmatrix} \beta_m \\ \mathbf{c}_m \end{vmatrix}=\begin{vmatrix} r_2 & t_2' \\ t_2 & r_2' \end{vmatrix}\begin{vmatrix} \alpha_m \\ 0 \end{vmatrix}\ . \tag{15}$$

It is easy to see from the above two equations that

$$\alpha_m=t_1\mathbf{a}_m+r_1'\beta_m\ ,\quad \beta_m=r_2\alpha_m\ . \tag{16}$$

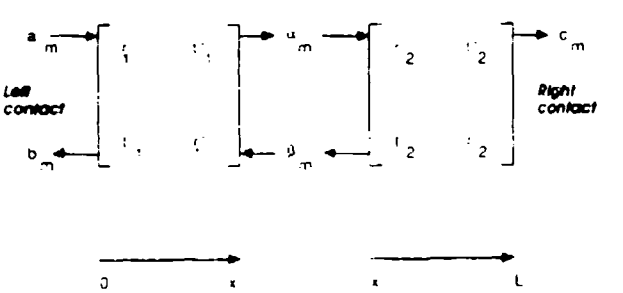


FIG. 3. A structure is decomposed into two sections with a common boundary at the coordinate x. The wave amplitudes at x are related to the wave amplitudes at the left and right contacts through the two scattering matrices describing the two sections.

Substituting the second equality in Eq. (16) into the first equality of that equation, we get

$$\alpha_m = I - r_1' r_2 (I - r_1' r_2)^{-1} t_1 \mathbf{a}_m , \tag{17}$$

$$\beta_m = r_2 (I - r_1' r_2)^{-1} t_1 \mathbf{a}_m , \tag{18}$$

where I is the identity matrix.

Since the scattering matrix elements for both sections can be determined from the recipe of Ref. 6 and, in addition, $\mathbf{a}_m$ is also known from the boundary conditions ($A_{mp}(x=0)=\delta_{mp}$), we can easily find α_m and β_m from the above relations and this gives us the wave function at any arbitrary position x. This method does not suffer from the numerical instability of the previous method when evanescent modes are included in the analysis, provided, of course, that r_1' and r_2 are not too large. Unless transport is strongly diffusive and approaches the strong localization regime, r_1' and r_2 remain small enough to prevent numerical instabilities from affecting this method.

The wave function $\psi_m(x,y)$ allows us to calculate the current density $\mathbf{J}_m(x,y)$ and the Fermi carrier concentration n_{sm} for injection from each mode m

$$\mathbf{J}_m(x,y) = \frac{iq\hbar}{2m^*}\{\psi_m(x,y)\nabla\psi_m^*(x,y) - \psi_m^*(x,y)\nabla\psi_m(x,y)\} , \tag{19}$$

$$n_{sm}(x,y) = |\psi_m(x,y)|^2 . \tag{20}$$

The total current density is found by vectorially adding the contributions of all propagating modes and the total Fermi carrier concentration is found by the scalar addition of the contributions of all modes, both propagating and evanescent.

$$\mathbf{J}_{\text{total}}(x,y) = \sum_{\text{propagating modes}} \mathbf{J}_m(x,y) ,$$

$$n_s^{\text{total}}(x,y) = \sum_{\text{propagating + evanescent}} n_{sm}(x,y) . \tag{21}$$

The above superposition is justified since we assume that there is no phase coherence between different modes in the contacts, so that the total current or Fermi carrier concentration is the incoherent superposition of the contributions of each mode. This picture is consistent with the Landauer-Büttiker formalism of linear response[5] in which the contacts are viewed as ideal reservoirs.

III. RESULTS

A. Conductance versus Fermi energy for attractive and repulsive impurities in the quasiballistic regime

In Fig. 4, we show the conductance versus Fermi energy of a GaAs mesa 1000 Å wide and 1160 Å long. The structure contains four impurities that are randomly distributed within the sample. In all cases the strength of each δ-function impurity was assumed to be such that the quantity γ [see Eq. (3)] ≈ 3 feV cm^2.

The conductance was calculated from the two-probe Landauer-Büttiker formula[5] in which the transmission was calculated directly from the scattering matrix following the prescription of Ref. 6. The curves show the usual quantized conductance steps associated with subband filling. Even though there is disorder in the channel, the quantization of the conductance is not destroyed as predicted by some authors[9–11] because there are not enough impurities for this to happen. The effect of the impurities

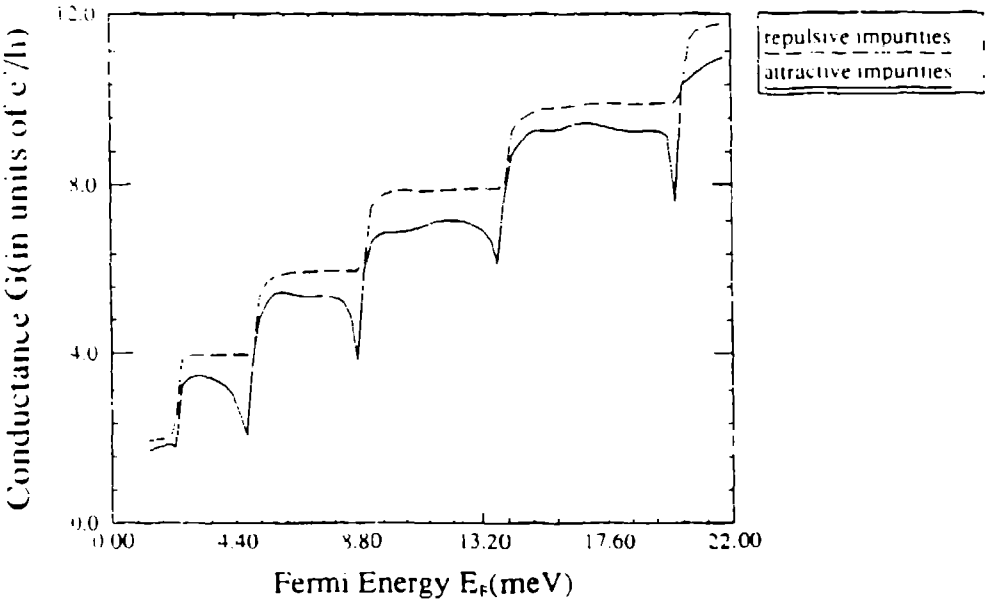


FIG. 4. The two-terminal linear-response conductance vs Fermi energy for the mesa of Fig. 1, showing quantized conductance steps. The steps for attractive impurities have lower heights and dips preceding a step, due to the formation of quasidonor states associated with the impurities. Forty evanescent modes were required to obtain convergence of the results even though the largest number of propagating states was six.

in this quasiballistic regime is twofold: it reduces the height of the conductance steps since is reduces the transmission through the channel, and it causes ringing in the characteristics because of wave interference effects caused by multiple reflection from the impurities.[10,11]

The conductance characteristics are plotted for both attractive and repulsive impurities and show pronounced differences for the two cases (majority- and minority-carrier transport). Attractive scatterers lead to a lower conductivity and also observable dips in the conductance both of which are caused by quasidonor states forming just below the subband minima. In fact, the dip energies correspond to the energies of these states. These have been reported by other authors as well.[12–14] Repulsive scatters do not produce dips since there are no quasibound states associated with them. In calculating these curves, we had to include 40 evanescent modes to obtain convergence of the results, even though the largest number of propagating modes (corresponding to the highest Fermi energy considered) was only six.

B. Spatial distribution of current in the quasiballistic regime

In Fig. 5, we plot the current profile corresponding to a Fermi energy of 4.8 meV, which is slightly below the third subband bottom. At this Fermi energy, there are two propagating modes in the sample. Figure 5(a) corresponds to the current pattern for the first mode ($\mathbf{J}_1$) and Fig. 5(b) corresponds to the pattern for the second mode ($\mathbf{J}_2$) when the scatterers are attractive. Figure 5(c) shows the resultant current pattern (total current density) obtained by vectorially adding $\mathbf{J}_1$ and $\mathbf{J}_2$.

It can be seen from the figures that a vortex has formed in the left bottom corner of the figure. There are not enough impurities to form strong and multiple vortices. To show the formation of strong multiple vortices, we plotted the total current profile in the case of six impurities, and the result is shown in Fig. 5(d). The vortices are now very clear and strong. These vortices, however, are not necessarily centered around impurities, nor are they pinned by the impurities.

The current pattern changes dramatically if we alter the impurity configuration. An example is shown in Fig. 5(e), where we have kept the number of impurities the same (six), but changed their positions. The vortices have disappeared. This dependence on the configuration can only be the result of quantum interference effects (interference between the scatterers). This indicates that the vortices are caused by quantum interference.

In Fig. 5(f) we show how the current pattern changes when the scatterers are made repulsive instead of attractive. For a few impurities (quasiballistic transport) we see no vortices at all; however for a large number of impurities corresponding to the diffusive regime, we see some evidence of vortices forming. We will discuss the current patterns in the diffusive regime later.

C. Spatial distribution of Fermi carriers in the quasiballistic regime

In Figs. 6(a)–6(f) we plot the concentration of Fermi electrons for the cases corresponding to Figs. 5(a)–5(f), respectively. It is obvious that the electron concentration is larger where the vortices form, since the electrons are partly localized in these regions. In Fig. 6(e), although the vortices have nearly disappeared, the concentration of Fermi electrons peak near the impurities because of electrostatic attraction (the impurities are attractive). The concentration profile is much smoother for repulsive impurities, since no vortices form in this case. Because the concentration variation is much more rapid in the case of majority-carrier transport, we expect space-charge effects to be more important in this case as well. However, we have not considered space-charge effects in any detail in this paper.

D. The influence of evanescent modes in the quasiballistic regime

In Fig. 7(a) we show the current pattern for four impurities in the absence of evanescent modes, while in Fig. 7(b) we show the corresponding electron concentration profiles. These should be compared with Figs. 5(c) and 6(c), where evanescent modes were included. The vortices do not appear when evanescent modes are neglected. This can be understood as follows. The wave function must be peaked where vortices form [this is apparent from the Fermi carrier concentration profile in Fig. 6(d)] since the electron is partially localized around the vortex center. The wave function everywhere is a linear superposition of particle-in-a-box wave functions, which are the normal modes of the system. In order to obtain a localized wave function by superposing particle in a box states, one would need several such states. Therefore merely the propagating states (which in our case are the lowest two states) will not suffice to create the vortices; one must include many higher states all of which are evanescent.

Evanescent modes also have other serious effects, especially when the tail of the mode around one impurity does not decay sufficiently before the next impurity is reached. It has been pointed out that in the case of attractive impurities, a local buildup of evanescent modes occurs around the impurities, due to the formation of quasidonor states (quasibound states).[13] This buildup will play an important role in the formation of residual resistivity dipoles discussed in Refs. 1 and 2. The study of such dipoles would, however, require a self-consistent treatment involving the simultaneous solution of the Schrödinger and Poisson equations. This is beyond the scope of the present work and is reserved for future investigation.

E. Conductance versus Fermi energy in the diffusive regime

In Fig. 8 we show the conductance versus Fermi energy for a 1000-Å-wide and 2.52-μm-long mesa with 60 impurities. The plots are for both attractive and repulsive impurities. The conductance steps are barely discernible, since 60 impurities are enough to seriously degrade the quantization.[15] As before, the attractive scatterers produce dips in the conductance just before the opening of an additional channel because of the quasidonor states. However, there is an additional set of dips right after the

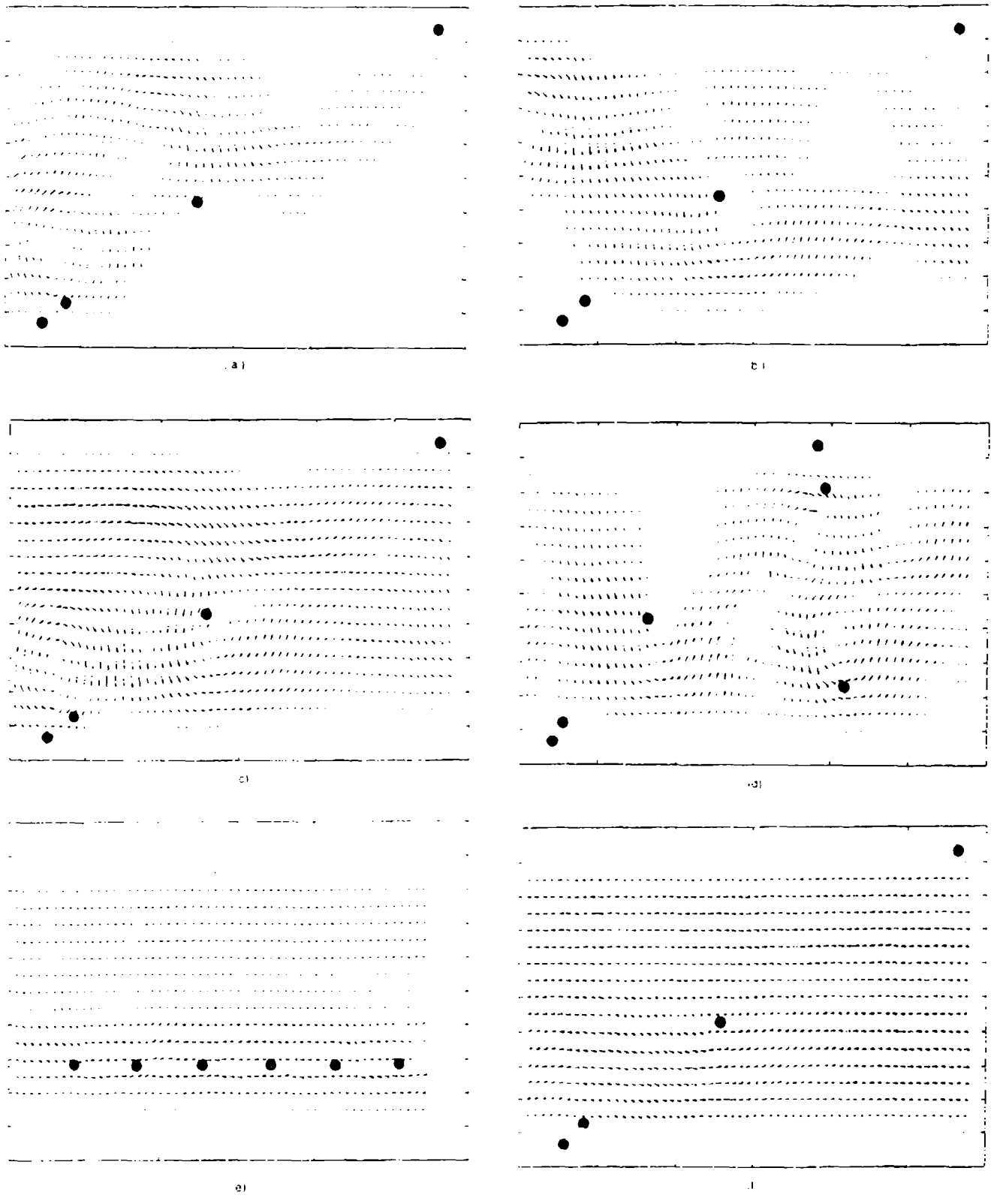


FIG. 5. The two-dimensional current density profile in the GaAs mesa in Fig. 1. There are four attractive impurities whose positions coincide with the centers of the solid circles. The horizontal direction is along the x axis. The Fermi energy $E_F = 4.8$ meV so that there are two propagating modes, i.e., the lowest two subbands are filled. (a) Injection is from the first subband only. (b) Injection is from the second subband. (c) The resultant current density due to injection from both subbands obtained by the vector addition of the current densities in (a) and (b). (d) The profile of the total current density in a mesa of length 1780 Å containing six impurities (all attractive). The vortices are now clear and strong. (e) The total current density profile in the same mesa [of (d)], but for a different impurity configuration. The vortices do not appear for this configuration. (f) Profile of the total current density in the structure of (e), with the sign of the impurity potential reversed (the impurities are repulsive).

opening of another channel, which has also been observed by other authors.[11,36] It was argued that this set of dips is due to the fact that the longitudinal wave vector of the newly opened subband is small, so that electrons injected with this wave vector suffer strong impurity scattering into other subbands), thereby decreasing the conductance.

F. Current profile in the diffusive regime

In Figs. 9(a) and 9(b), we show the total current profile for a section of the structure containing the first four impurities. The injection energy (Fermi energy) is 2.3 meV. The two plots are for attractive and repulsive impurities. Vortices are now observed in both cases and their radii are quite large. This shows that although the impurity potential is very short range (δ-function potential), its effect on the current distribution is quite long range.

When there are a large number of impurities, the sign of the impurity potential (i.e., whether attractive or repulsive) appears to play little part in the formation of vortices. Hence it appears that the vortices are *not caused* by quasidonor states which are associated only

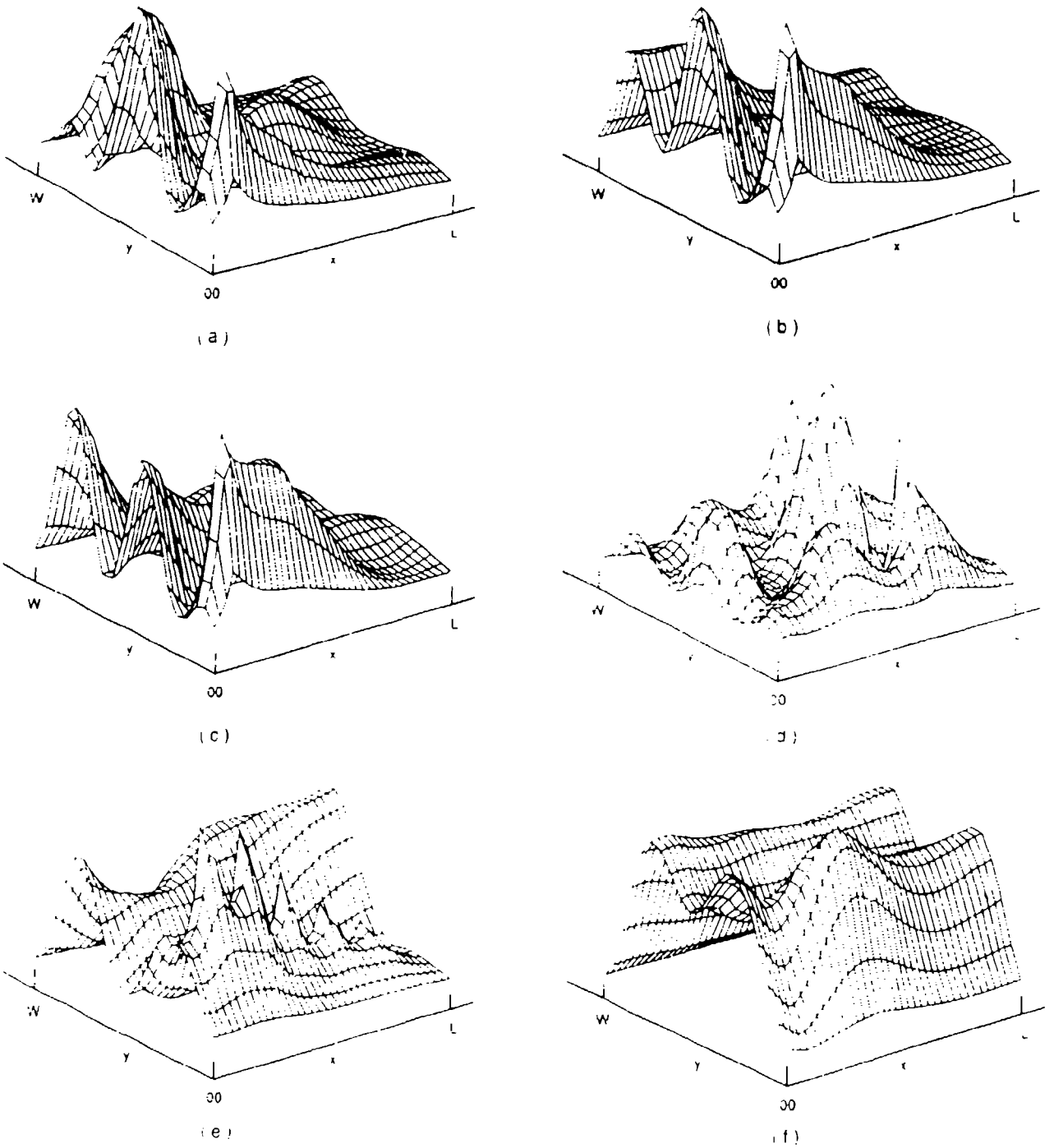


FIG. 6. Two-dimensional profile of the concentration of Fermi carriers. (a)–(f) correspond to the parameters and conditions of Figs. 5(a)–5(e).

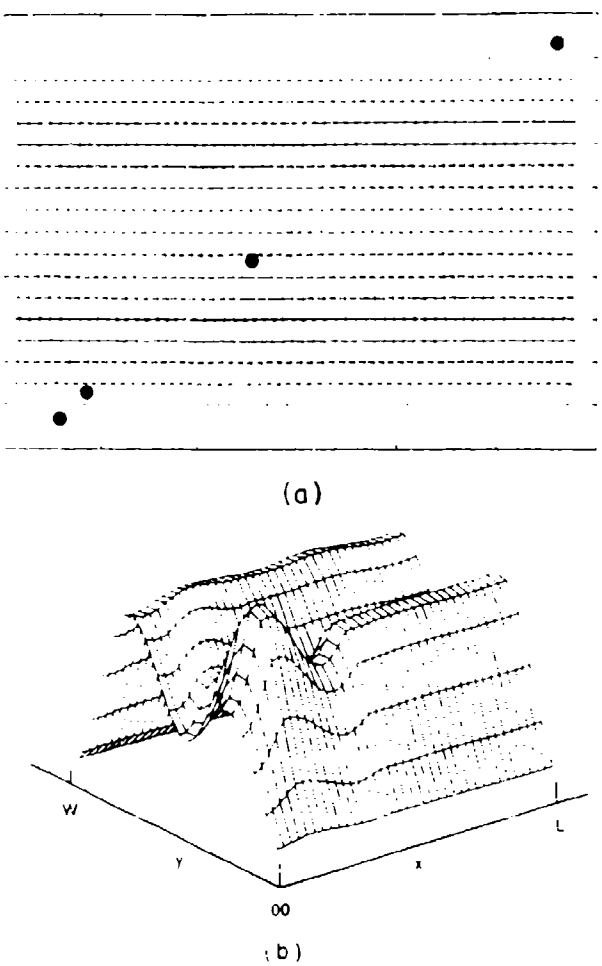


FIG. 7. Plots in the absence of evanescent modes for the structure of Figs. 5(c) and 6(c). (a) Profile of the total current density; (b) the concentration profile for Fermi electrons.

with attractive impurities; however, the donor states may have an effect on the vortices since the nature of the vortices certainly changes when the sign of the scattering potential is reversed. We discuss quasidonor states and vortex formation in the next subsection.

G. Quasidonor states and vortices

To examine the relationship between the quasidonor states and vortex formation, we plotted the total current density profiles for three different Fermi energies which are above (4.8 meV), below (4.3 meV), and at (4.7 meV) the quasidonor-state energy associated with the third subband in a structure identical with that which gave the result of Fig. 5(c). The plots for Fermi energies of 4.3 and 4.7 meV are shown in Figs. 10(a) and 10(b), respectively. Comparing Figs. 5(c), 10(a), and 10(b), we find that the vortices are not any more prominent when carriers are injected at the quasidonor-state energy (4.7 meV) than when they are injected at energies above or below the donor-state energy. Therefore, the vortices do not seem to be associated with donor states, even though the donor states may have an effect on them.

When the injection of electrons is at a quasidonor-state energy, evanescent states build up around the impurities[13] and therefore the Fermi carrier concentration shows large spikes at the impurity locations. These spikes are certainly the result of electron capture into bound states around attractive impurities, but they are not in any way associated with vortices. Hence it appears that the vortices are caused by quantum-mechanical interference effects and require the presence of evanescent modes, but they are not the result of quasidonor states. This is further confirmed by the fact that they show up even when the scatterers are repulsive [Fig. 5(f)] when there are no quasidonor states at all.

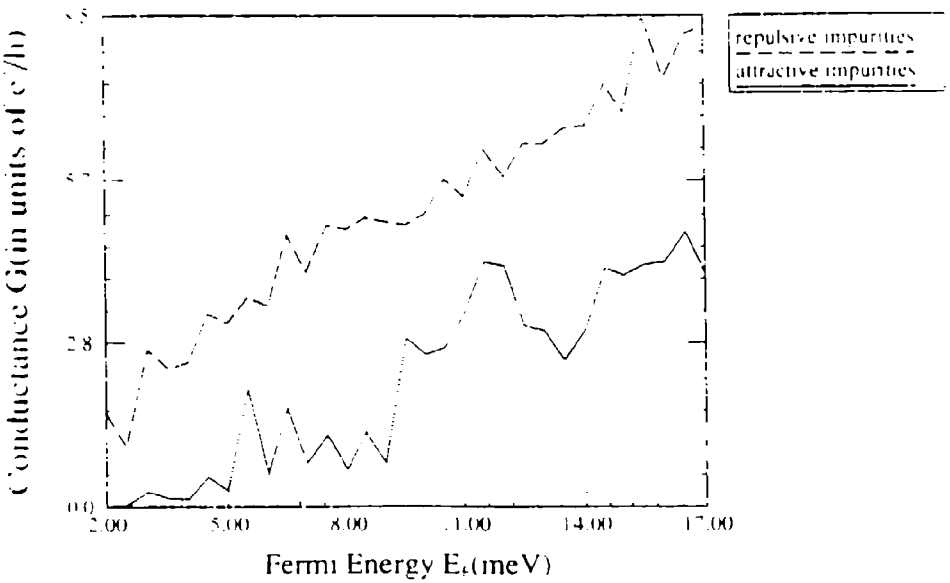


FIG. 8. The two-terminal linear-response conductance vs Fermi energy for a 2.52-μm-long mesa of width 1000 A having 60 impurities randomly distributed. Transport is now diffusive instead of quasiballistic.

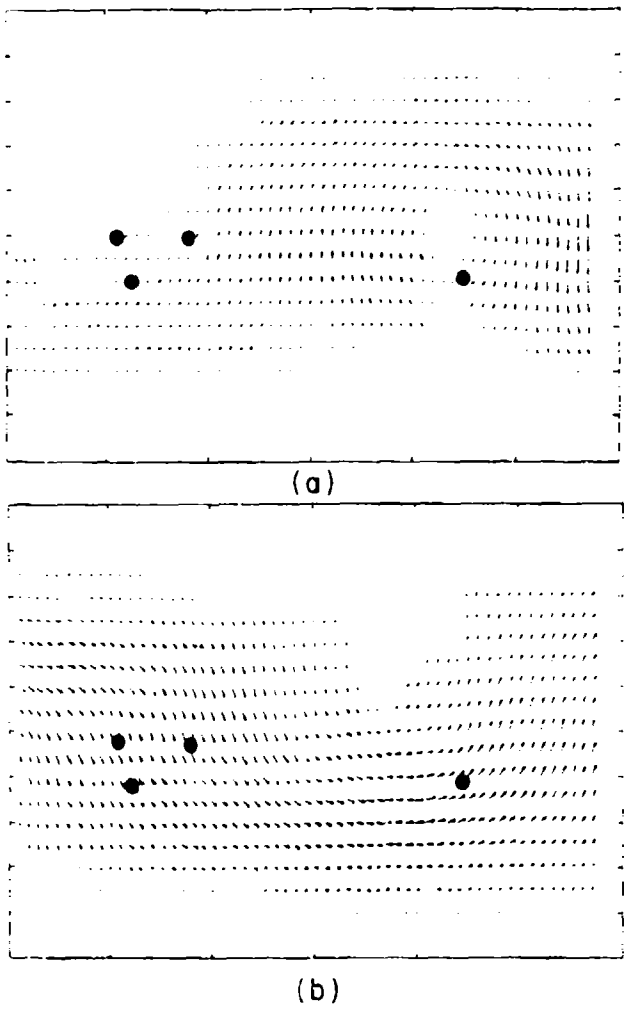


FIG. 9. Profile of the total current density in the mesa of Fig. 8 plotted for a section containing the first four impurities. The injection energy (Fermi energy) is 2.4 meV. (a) For attractive impurities; (b) for repulsive impurities.

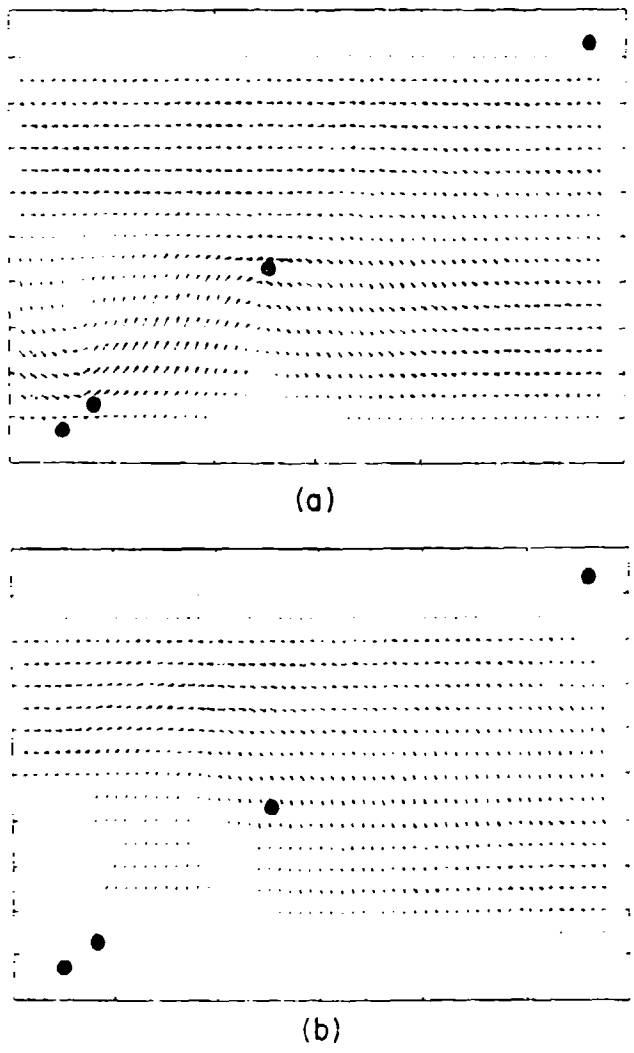


FIG. 10. Profile of the total current density in the structure of Fig. 5(c) with the Fermi energy slightly altered while keeping the number of propagating modes the same. (a) $E_F = 4.3$ meV (below the quasidonor level splitting off from the third subband); (b) $E_F = 4.7$ meV (the energy of the quasidonor level).

IV. CONCLUSION

In this paper, we have examined the spatial distributions of current density and Fermi carrier concentration around localized elastic scatterers (impurities) in phase-coherent transport. We found that vortices occur in the current for both attractive and repulsive impurities, and their strength is typically larger for attractive impurities (majority-carrier transport). These vortices are not caused by quasidonor states which form when the impurities are attractive; instead, they seem to be caused by quantum-mechanical interference between the waves reflected multiply from impurities. We also showed that evanescent modes are necessary for the formation of these vortices and that the nature (or even occurrence) of the vortices depends on the impurity configuration, which is consistent with the conclusion that they are caused by quantum-mechanical interference of waves reflected from the impurities.

In concluding this paper, we wish to make a remark about the significance of these vortices. Because the vortices involve a circulating current pattern, they will give rise to localized magnetic moments.[17] A magnetic field applied perpendicular to the mesa will therefore either quench or accentuate the vortices, depending on the direction of the field. Diffusive transport in the presence of a magnetic field has been treated in the past within the scattering matrix formalism,[18] but the treatment has examined only terminal characteristics and not internal distributions of current or carriers. It would be interesting to examine what effect a magnetic field has on the two-dimensional current and carrier-concentration profiles. This is presently being investigated and will be reported in a future publication.

ACKNOWLEDGMENTS

This work was supported by the U.S. Air Force Office of Scientific Research under Grant No. AFOSR91-0211 and by IBM.

[1]R. Landauer, IBM J. Res. Dev. 1, 223 (1957).
[2]R. Landauer, IBM J. Res. Dev. 32, 306 (1988).
[3]B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon, Phys. Rev. Lett. 60, 848 (1988); D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. C 21, L209 (1988).
[4]J. A. Nixon, J. H. Davies, and H. U. Baranger, Phys. Rev. B 43, 12 638 (1991).
[5]R. Tsu and L. Esaki, Appl. Phys. Lett. 22, 562 (1973); S. Datta and R. Frohne, J. Appl. Phys. 64, 4086 (1988).
[6]S. Datta, M. Cahay, and M. McLennan, Phys. Rev. B 36, 5655 (1987); M. Cahay, M. McLennan, and S. Datta, *ibid.* 37, 10 255 (1988); M. Cahay, S. Bandyopadhyay, M. A. Osman, and H. L. Grubin, Surf. Sci. 228, 301 (1990).
[7]S. Ramo, J. R. Whinnery, and T. van Duzer, *Fields and Waves in Communication Electronics* (Wiley, New York, 1965).
[8]R. Landauer, in *Localization, Interaction and Transport Phe-*[illegible] by G. Bergmann and Y. Bruynseraede (New York, 1985); M. Büttiker, IBM J. Res. [illegible]; see also Refs. 1 and 2.
[illegible]d M. Jonson, Phys. Rev. B **44**, 3810 (1991).
[illegible] S. Das Sarma, Phys. Rev. B **40**, 3379 (1989).
[11]I. Kander, Y. Imry, and U. Sivan, Phys. Rev. B **41**, 12 941 (1990).
[12]C. S. Chu and R. S. Sorbello, Phys. Rev. B **40**, 5941 (1989).
[13]P. F. Bagwell, Phys. Rev. B **41**, 10 354 (1990); A. Kumar and P. F. Bagwell, *ibid.* **43**, 9012 (1991).
[14]E. Tekman and E. Ciraci, Phys. Rev. B **42**, 9098 (1990).
[15]There are some vestiges of the quantization that remain even when the transport is diffusive, as long as strong localization does not set in.
[16]A. Kumar and P. F. Bagwell (unpublished).
[17]C. S. Lent, Appl. Phys. Lett. **57**, 1678 (1990).
[18]H. Tamura and T. Ando (unpublished); S. Chaudhuri, S. Bandyopadhyay, and M. Cahay (unpublished).

PHYSICAL REVIEW B VOLUME 47, NUMBER 0 15 MAY 1993

Current, potential, electric field, and Fermi carrier distributions around localized elastic scatterers in phase-coherent quantum magnetotransport

S. Chaudhuri and S. Bandyopadhyay
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

M. Cahay
Department of Electrical and Computer Engineering, University of Cincinnati, Cincinnati, Ohio 45221

(Received 3 April 1992; revised manuscript received 28 December 1992)

The spatial distributions of the current, Fermi carriers, electric field (due to space charges) and electrostatic potential in a disordered mesoscopic structure are calculated in the presence of arbitrary magnetic fields. These distributions are useful in elucidating many features of quantum magnetotransport, such as the formation of edge states at high magnetic fields and their near-perfect transmittivity, the evolution of the integer quantum Hall effect, the creation of magnetic bound states around an impurity, the magnetic response of current vortices that form as a result of quantum interference between scatterers and the walls of a quantum wire, the dependence of the quantized-conductance steps in a backgated quantum wire on an applied magnetic field, the behavior of residual-resistivity dipoles and the electrostatic space-charge potential in a magnetic field, the dependence of the sign of the magnetoresistance on the impurity configuration, etc. We examine the current, Fermi carrier concentrations, electric field, and both chemical- and electrostatic-potential profiles associated with each of these phenomena, and relate them to the observed terminal characteristics in each case.
PACS number(s): 72.10.Fk, 72.20.Fr

I. INTRODUCTION

Many features of quantum magnetotransport in a mesoscopic structure can be understood by examining the spatial distributions of the current, potential, electric field, and Fermi carriers around elastic scatterers in the presence of a magnetic field. In this paper, we have calculated such spatial patterns from a fully microscopic quantum-mechanical formalism based on the Schrödinger equation. These spatial distributions are, of course, not "precise" in the context of the Heisenberg uncertainty principle which prohibits simultaneous definition of both the position and the momentum of an electron; instead they should be viewed as spatial distribution of quantum-mechanical averages (or expected values) of the corresponding quantities. These patterns help in understanding several quantum transport phenomena and provide direct visualization of the associated physics. For example, we show the spatial distribution of the current carried by edge states in a quantum wire at high magnetic field. The current patterns clearly show the high transmittivity which is responsible for the quantization of the Hall resistance in the quantum Hall regime. We also show how current circulates around an impurity when a magnetic bound state forms. The recovery of the quantization of conductance steps in a dirty quantum wire by the application of a magnetic field is known to be associated with the suppression of backscattering, and the current patterns that we obtain show this very clearly. The onset of the integer quantum Hall effect is caused by the edges of a quantum wire becoming perfectly smooth equipotential surfaces and our chemical potential profiles demonstrate this strikingly. The magnetic response of current vortices, electrostatic potential, and electric field all show quenching phenomena. We also found why the sign of the magnetoresistance of a quantum wire could depend on the position of a single impurity inside the wire. These, and the nature of other phenomena, become very clear when one examines the spatial patterns of current, carrier concentration, and electrostatic, and/or chemical potentials. The spatial distributions are not only of great importance in understanding the origin and the detailed physics of many quantum magnetotransport phenomena, but they often determine the nature of certain physical phenomena such as electromigration.

This paper is organized as follows. In Sec. II, we briefly present the quantum-mechanical theory used in computing the two-dimensional spatial distributions of current, Fermi carrier concentration, potential, and internal electric field profiles around localized scatterers in the presence of a magnetic field. In Sec. III, we present the results for a disordered GaAs quantum wire containing a variable number of impurities. We show the spatial patterns associated with various magnetotransport phenomena in both single-channeled and multichanneled transport. Such spatial patterns were presented by us in the limit of coherent diffusive transport (in the absence of any magnetic field) earlier.[1] Others have presented such patterns in the limit of coherent ballistic transport with no scattering whatsoever.[2] To our knowledge, we are the first to present such patterns in the *coherent diffusive* regime *in the presence of a magnetic field.*

II. THEORY

The theory for our calculations of the current and carrier density patterns in a quantum wire in the absence of a magnetic field has been described in Ref. 1. Here, we

repeat only the basic features. We consider a mesa as shown in Fig. 1 which is single moded in the z direction (i.e., only a single subband is occupied in that direction), but multimoded in the y direction.

The Schrödinger equation describing electron transport in this confined quasi-one-dimensional disordered structure under a magnetic field is

$$\frac{(\mathbf{p}-e\mathbf{A})^2}{2m^*}\psi+V\psi=E\psi\,, \tag{1}$$

where $\mathbf{p}$ is the momentum operator, m^* is an isotropic effective mass, and the potential energy V consists of two terms.

$$V\equiv V(x,y)=V_c(y)+V_{\mathrm{imp}}(x,y)\,. \tag{2}$$

The first term gives the effect of the confining potential in the y direction and the second term is due to impurities. We assume hardwall boundary conditions in the y direction and δ potentials for the impurities so that we have

$$V_{\mathrm{imp}}(x,y)=\gamma\sum_{i=1}^{s}\delta(x-x_i)\delta(y-y_i)\,, \tag{3}$$

where s is the total number of impurities in the structure and the summation extends over all the impurities. The position of the ith impurity is denoted by (x_i,y_i).

The wave function ψ^0 in the absence of impurities can be written as

$$\psi^0(x,y)=e^{ik_x x}\phi(y)\,, \tag{4}$$

where k_x is the x-directed wave vector and $\phi(y)$ satisfies the eigenequation

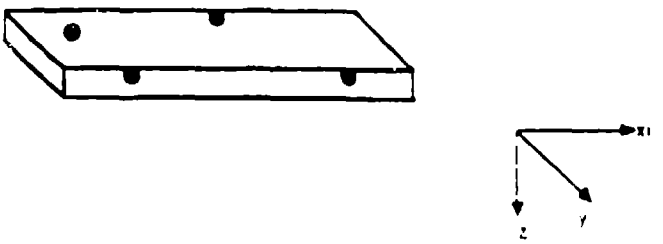


FIG. 1. A quasi-one-dimensional quantum wire containing a random array of elastic scatterers. The structure is single moded in the z direction but multimoded in the y direction. The magnetic field is applied in the z direction.

$$\frac{\partial^2\phi}{\partial y^2}+\frac{2m^*}{\hbar^2}E\phi(y)-\left(\frac{y}{l^2}\right)^2\phi(y)+2\frac{y}{l^2}k_x\phi(y)-k_x^2\phi(y)=0\,, \tag{5}$$

with l being the magnetic length given by $l=\sqrt{\hbar/eB}$.

The wave function $\phi(y)$ is found by solving the above equation numerically using a finite difference scheme as described in Ref. 3. This method also gives the energy dispersion relations E vs k_x for the various hybrid magnetoelectric subbands that arise from the confining effects of the magnetic field and the confining potential in the y direction. A set of computed energy dispersion relations are shown in Fig. 2. After calculating the energy dispersion relations, we choose a certain energy E (which we call the Fermi energy E_F) and find the corresponding x-directed wave vectors k_p^F for various magnetoelectric subbands (indexed by p) from the dispersion relations. These wave vectors k_p^F and the wave functions $\phi_p(y)$ are used to determine the total wave function $\psi_m(x,y)$ (in the presence of impurities) for an electron injected with the Fermi energy E_F from the subband m in the left contact. The wave function is given by

$$\psi_m(x,y)=\sum_{p=1}^{M}\frac{A_{mp}(x)\exp(ik_p^F x)\phi_p(y)+B_{mp}(x)\exp(ik_{-p}^F x)\phi_{-p}(y)}{\sqrt{|v_p^F|}}\,, \tag{6}$$

where the summation extends over both propagating and evanescent modes (the total number of modes considered is M). The subscript $-p$ indicates that the quantity corresponds to a wave function with an oppositely directed velocity as compared to the one with a subscript p.

When the wave function is written in this form (i.e., each mode is normalized by the square root of the corresponding group velocity), all propagating modes at the only have to evaluate the coefficients $A_{mp}(x)$ and $B_{mp}(x)$ to determine the wave function of Fermi electrons everywhere. If we know these coefficients at any position (x_0,y) in the structure, then the coefficients at any other position (x_1,y) can be found from

$$\begin{bmatrix} r_0 \\ \cdot \end{bmatrix}=\begin{bmatrix} t_{11} & t_{12} \\ \cdot & \cdot \end{bmatrix}\begin{bmatrix} I_1 \\ \cdot \end{bmatrix} \tag{7}$$

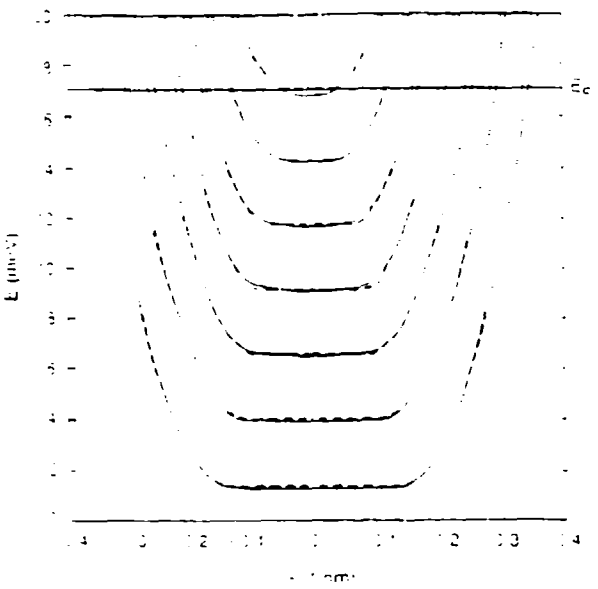


FIG. 2. The energy dispersion relations for the hybrid magnetoelectric subbands in a quantum wire of width 2000 Å at a magnetic flux density of 1.5 T. The solid lines are the results of exact numerical calculations, whereas the broken lines are the results of the semiclassical Bohr-Sommerfeld quantization rule.

column vector of length M whose elements are $B_{m1}(x_0)$, $B_{m2}(x_0)$, $B_{m3}(x_0)\cdots B_{mM}(x_0)$, r_0 is a column vector of length M whose elements are $A_{m1}(x_1)$, $A_{m2}(x_1)$, $A_{m3}(x_1)\cdots A_{mM}(x_1)$, and r_1 is also a column vector of length M whose elements are $B_{m1}(x_1)$, $B_{m2}(x_1)$, $B_{m3}(x_1)\cdots B_{mM}(x_1)$. The coefficients t_{11}, t_{12}, t_{21}, and t_{22} are each $M\times M$ matrices and the square matrix in Eq. (7) is the so-called transfer matrix for the section of the structure between $x=x_0$ and x_1.

The above equation allows us to determine the coefficients $A_{mp}(x_1)$ and $B_{mp}(x_1)$ [and hence the wave function] at any arbitrary location x_1 if we know the wave function [or the coefficients $A_{mp}(x_0)$ and $B_{mp}(x_0)$] at any one point $x=x_0$ in the device as well as the transfer-matrix elements. We do know the wave function at the left contact ($x=0$). We assume that for injection into the mth mode, the wave function is given by scattering states[5] (this is our boundary condition)

$$\psi_m(0-,y)=\frac{\exp(ik_m^F x)\phi_m(y)}{\sqrt{v_m^F}}-\sum_{p=1}^{M}\frac{R_{mp}\exp(ik_p^F x)\phi_p(y)}{\sqrt{v_p^F}} \tag{8}$$

so that $A_{mp}(x=0-)=\delta_{mp}$ (Kronecker delta) and $B_{mp}(x=0-)=R_{mp}$. Therefore, we can find the wave function anywhere in the structure using Eqs. (6) and (7) if we can find the elements of the appropriate scattering matrices in Eq. (7). In the absence of a magnetic field, these elements are fairly easy to find,[1] but for nonzero magnetic fields, the wave functions of the various subsequently, the elements of the scattering matrix cannot be found by the normal prescription of Ref. 1. The method for finding these elements is described in the Appendix.

Once we have evaluated the scattering matrices, we have found all the quantities that we need to evaluate the wave function of Fermi electrons everywhere in the structure. From this wave function, we calculate all quantities of interest.

In linear-response transport at low temperatures, only electrons at the Fermi level carry a net current. Therefore, the two-dimensional wave function $\psi_m(x,y)$ of electrons in the mth subband allows us to calculate the contribution of that subband to the two-dimensional current density. This is given by

$$J_m(x,y)=\frac{iq\hbar}{2m^*}[\psi_m(x,y)\nabla\psi_m^*(x,y)-\psi_m^*(x,y)\nabla\psi_m(x,y)]-\frac{1}{m^*}e\mathbf{A}|\psi_m|^2 \tag{9}$$

The Fermi carrier concentration due to the contribution from the mth subband is defined by us as

$$n_m(x,y)=|\psi_m(x,y)|^2 . \tag{10}$$

The total current density is found by vectorially adding the contributions from all propagating modes and the total Fermi carrier concentration is found by the scalar addition of the contributions from all propagating modes. Note that no weighting by the density of states is necessary in the summation since the wave functions have been normalized by the square root of velocity which takes care of the density of states.

$$J_{\text{total}}(x,y)=\sum_{\text{propagating modes}} J_m(x,y),\qquad n_i^{\text{total}}(x,y)=\sum_{\text{propagating}} n_m(x,y) \tag{11}$$

Finally, we find the electrostatic space-charge potential $V(x,y)$ from the relation[4,6,7]

$$eV(x,y)=\frac{\sum_{\text{propagating}}|\psi_{1,m}(x,y)|^2\mu_1+|\psi_{2,m}(x,y)|^2\mu_2}{\sum_{\text{propagating}}|\psi_{1,m}|^2+|\psi_{2,m}|^2}, \tag{12}$$

where μ_1 and μ_2 are the chemical potentials in the two contacts of the quantum wire, $\psi_{1,m}$ is the wave function corresponding to injection in mode m (with energy μ_1) from the left contact, and $\psi_{2,m}$ is the wave function for injection in mode m (with energy μ_2) from the right contact. The above equation is derived from the Poisson equation and is valid only for (a) linear-response transport, (b) situations where the potential variation of $V(x,y)$ is smooth on the scale of the screening length, and (c) when the Fermi wavelength is considerably smaller than the screening length. When these conditions are not satisfied, the actual Poisson equation must be solved simultaneously with the Schrödinger equation, rather than using the simplified equation (12). That is numeri-

tion we have to account for carriers at all energies, not just the Fermi energy. This means that we must solve the Schrödinger equation for all energies which is a more demanding task than the present approach of solving it only at the Fermi energy. Fortunately, in semiconductors and in linear-response transport, the present approach may work well since space-charge effects are not tremendously important.[5] Ehrenreich, et al.

It was pointed out by McLennan, Lee, and Datta that $eV(x,y)$ is also what is actually measured at a chemical potential probe. In reality, the electrostatic potential is the convolution of the chemical potential with a screening function.[3] In the limit of strong screening (δ impurities), this screening function approaches a δ function so that the electrostatic and chemical potential become equivalent.

Once the electrostatic potential is determined, the corresponding electric field is found simply from

$$\vec{E}(x,y) = -\nabla V(x,y) . \tag{13}$$

In the next section, we present the spatial distributions of the current, Fermi carrier concentration, electrostatic or chemical potential, and electric field patterns associated with various magnetotransport phenomena.

III. RESULTS

A. Spatial distributions around a single impurity in single-channeled magnetotransport: Local effects

In Figs. 3(a) and 3(b), we show the spatial distribution of the current around a single scatterer placed in the middle of a GaAs quantum wire in the absence of any magnetic field. The two figures are for an attractive and a repulsive scatterer, respectively. The wire is 800 Å long and 1000 Å wide. The Fermi energy is 2.054 meV and only a single subband is occupied in both y and z directions so that transport is single channeled. There is pronounced current crowding near the impurity when the impurity potential is attractive and diverges away from the impurity when the impurity potential is repulsive. There are also two weak vortices above and below the scatterer for an attractive impurity. If the impurity is positioned close to the edges of the wire instead of at the center, such a pronounced effect does not occur. There are two reasons for this. First, the wave function diminishes in amplitude near the edges of the wire so that the interaction between the electron and the impurity is weakened when the impurity is closer to the walls. Second, the confining potential of the walls plays a greater role nearer the walls so that the current sees not only the impurity, but also the walls. As a result, the current pattern is determined by the combined interaction of the impurity and the wall.

It is obvious that when the impurity is right at the center, the interaction between the impurity and the electrons is strongest since the wave function (of the lowest subband) peaks at the center. A magnetic field will skew the wave function away from the impurity and hence decrease the interaction or scattering. This will lead to a positive magnetoresistance in this case. On the other hand, if the impurity is close to one of the walls, then a magnetic field will either bring the electrons closer to the impurity or take it further away depending on the direction of the magnetic field. Therefore we can observe either positive or negative magnetoresistance in that case (depending on the field direction) and the magnetoresistance will be asymmetric in the magnetic field. This is a rather striking effect which might be observable in mesoscopic structures at very low temperatures when impurity scattering is the dominant mechanism for resistance.

In Figs. 4(a) and 4(b), we have plotted the Fermi carrier concentration around the attractive and the repulsive

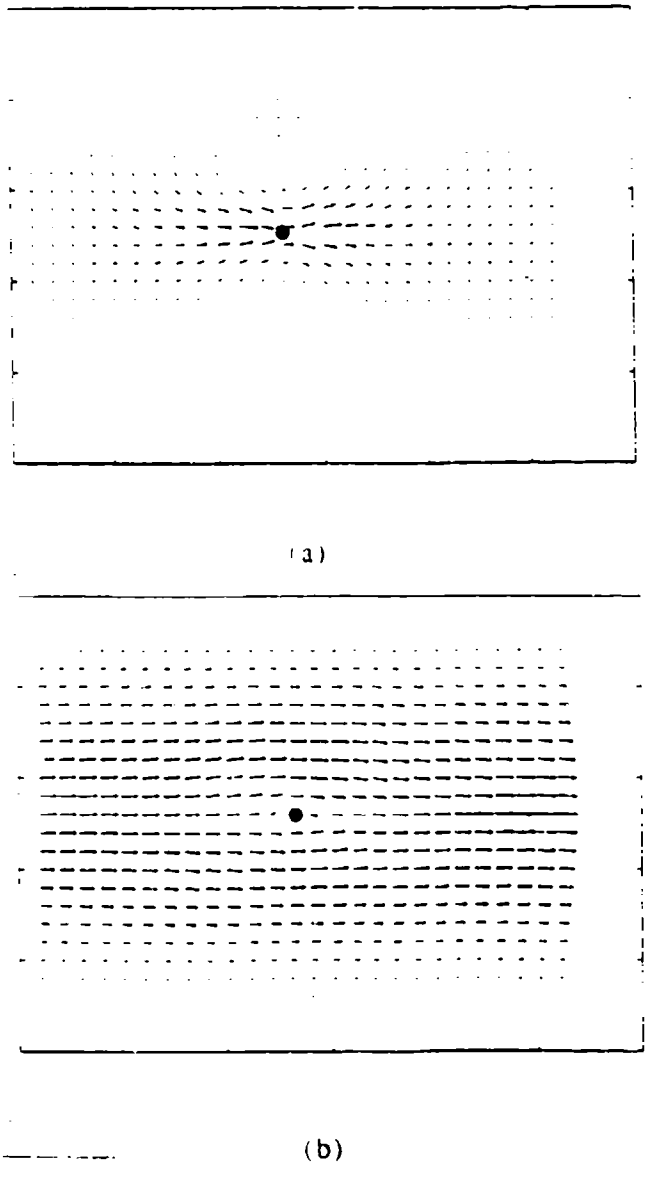


FIG. 3. The current distribution inside an 800-Å-long and 1000-Å-wide structure in the absence of any magnetic field. (a) The impurity is attractive and there is significant current crowding around the impurity which is shown by the solid circle. (b) The impurity is repulsive and the current detours around the impurity. This figure shows local effects of the impurity poten-

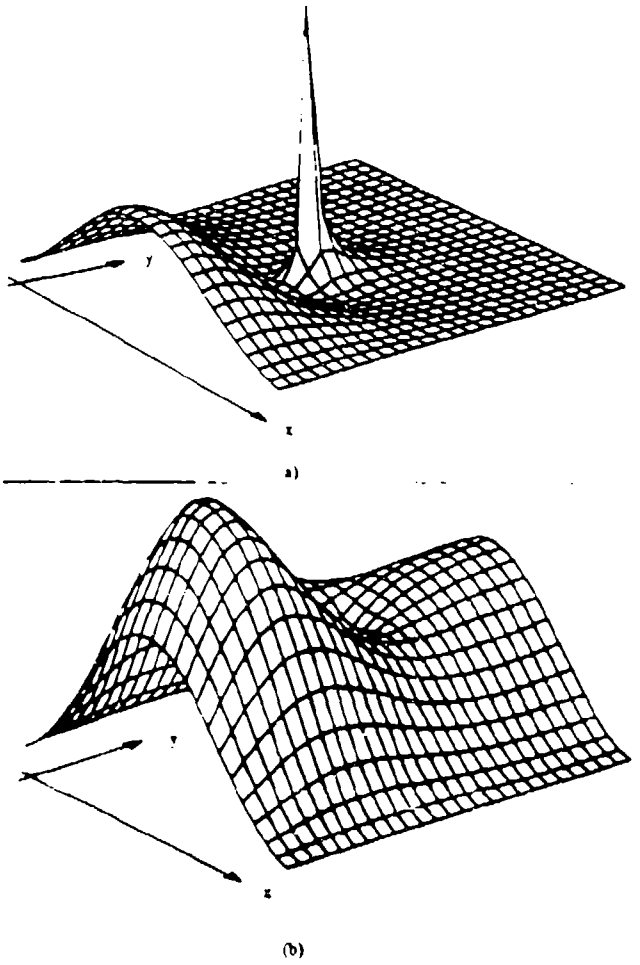


FIG. 4. The Fermi carrier concentration profile for the structure of Fig. 3 in the absence of any magnetic field. (a) The impurity is attractive; (b) the impurity is repulsive.

impurity of Figs. 3(a) and 3(b) in the absence of any magnetic field. As expected, in the case of the attractive impurity, we see a sharp spike (accumulation of electrons) around the impurity where the current crowding occurs. The electrons are attracted to the impurity by electrostatic attraction. In the case of repulsive impurity, the concentration is depleted where the impurity is positioned because of electrostatic repulsion. In Fig. 4(a), we see that there is a building up of electrons to the left of the scatterer and a deficit to the right. This happens because the impurity strongly reflects the incoming electrons which build up to the left. Because of the excess negative charge on one side and a deficit on the other, we expect a dipole to form. This is the residual-resistivity dipole as discussed by Landauer.[9] We can contrast these figures with the profiles that were presented in Ref. 1 which considered only multiple-impurity systems. When a large number of impurities are placed inside the wire, the local charge accumulation and depletion effect that causes a dipole to form is usually diminished because of quantum-mechanical interference between waves reflected from the various impurities. This quantum-mechanical interference formation in most cases, although it may sometimes also enhance the dipole formation depending on the impurity configuration, i.e., the exact phase relationships between the interfering waves. Therefore, even though the dipole formation is a local effect in itself, the size and orientation of the dipoles are significantly affected by nonlocal quantum interference effects.

In Figs. 5 and 6, we plot the electrostatic (or chemical) potential and electric field profiles. For these figures, we assume that $\mu_1 = 2.056$ meV and $\mu_2 = 2.052$ meV. We do not exactly see the characteristic electric field distribution of a dipole since the "poles" of the dipole are not "isolated" charges; they are "extended" charges. The plots are presented for both attractive and repulsive scatterers. They are very different when the impurity potential is attractive (which corresponds to majority carrier transport) as opposed to the case when it is repulsive (minority carrier transport). In the case of attractive impurities, there can be a building up of evanescent states around the impurities associated with the formation of quasi bound donor states.[10] These may help in dipole formation.

The application of a magnetic field causes dramatic changes in the current, carrier concentration, electrostatic potential, and electric field profiles. We show these

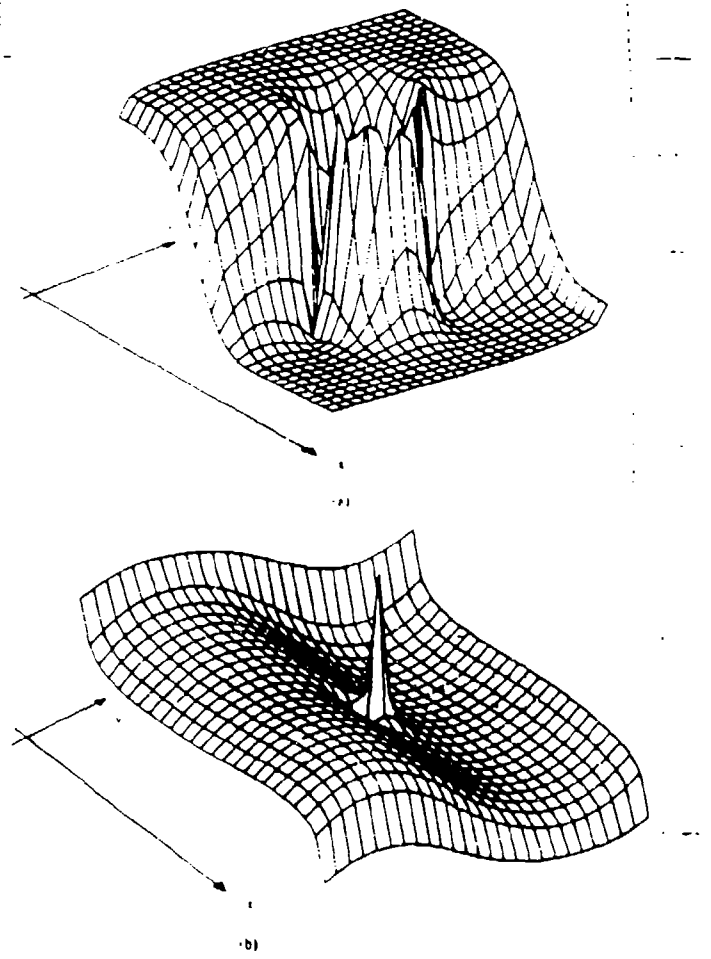


FIG. 5. The electrostatic (or chemical) potential profile inside the structure of Fig. 3 in the absence of any magnetic field.

effects for the single attractive impurity in Figs. 7(a), 7(b), 7(c), and 7(d) which are to be compared with Figs. 3(a), 4(a), 5, and 6, respectively.

In Fig. 7(a), we have plotted the current inside the wire in the presence of a magnetic flux density of $B = 2.18$ T. All other conditions correspond to the case of Fig. 3(a). Comparing with Fig. 3(a), we find that current crowding has disappeared and the two weak vortices have been quenched. The sample now no longer exhibits localized magnetic moments due to the vortices. In addition, there is a significant reduction of the backscattered current component. A similar reduction in backscattering is observed for a repulsive impurity also, except that it takes a lower magnetic field to reduce the backscattering by the same amount. The current pattern in Fig. 7(a) clearly shows that edge states have formed and carry the bulk of the forward current. There is also, however, a small reverse-traveling current component close to the walls. This component was explained in Ref. 11 (which dealt with edge states in ballistic transport as opposed to diffusive transport) as being due to the classical cyclotron trajectories (skipping orbits) having their velocities directed opposite to the net current flow adjacent to the edges. At higher magnetic fields, this reverse component diminishes so that backscattering is suppressed even more. In other words, the transmittivity of the edge states are very high at high magnetic fields. This agrees with Büttiker's picture of the integer quantum Hall effect[2] in which perfect transmission of edge states was invoked to explain quantization of the Hall resistance and vanishing of the longitudinal resistance.

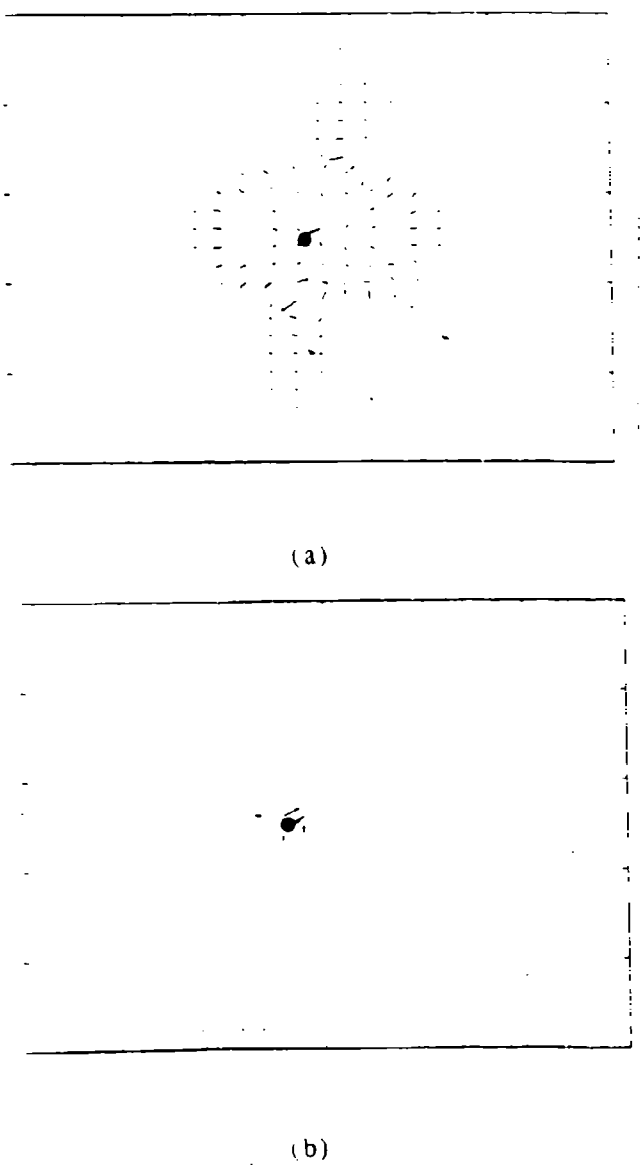


FIG. 6. The electric field profile inside the structure of Fig. 3 in the absence of any magnetic field. (a) The impurity is attrac-

In Fig. 7(b), we show the concentration profile of the Fermi carriers in the presence of a magnetic field of 2.18 T. The concentration piles up near an edge because of the Lorentz force skewing the wave function towards that edge. Note that there is no longer a buildup of electrons to the left of the scatterer and a deficit to the right since the buildup is due to reflection (backscattering) and this has diminished significantly. In fact, at high enough magnetic fields, backscattering is always reduced. Therefore, in general, residual-resistivity dipoles will be destroyed by an external magnetic field if it is of high enough magnitude.

In Fig. 8(a), we show the conductance of the structure of Fig. 3(a) versus the Fermi energy at a magnetic field of 3.5 T. There is only one attractive impurity in the structure placed at the center. A pronounced dip in the conductance is seen at an energy of 3.82 meV. At this energy, a magnetic bound state[3] forms around the impurity which causes the transmission (and hence the conductance) to decrease sharply. This state forms at about the same energy for a repulsive scatterer. The corresponding current profile is shown in Fig. 8(b). We find that there is only a small amount of net current traveling from one contact to another when a magnetic bound state forms and the bulk of the current circulates around the impurity. These magnetic bound states are quite different from the quasi donor states[10] which are also bound states giving rise to dips in the conductance characteristics. There are two major differences between magnetic bound states and quasi donor bound states. First, the former occurs regardless of whether the impurity is attractive or repulsive, whereas the latter forms only when the impurity is attractive. Second, the current patterns are very different. In the case of magnetic bound states, the current circulates around the impurity but in the case of quasi donor states, no such circulating current pattern needs to form.[1] Finally, the carrier concentration profiles are also quite different. In the case of magnetic bound states, the concentration builds up around the impurity but right at the impurity it drops sharply. This is shown in Fig. 8(c). Although the carriers may accumulate around the impurity even in the case of quasi donor state formation (owing to the buildup of evanescent states), the effect is much less pronounced. Finally, Fig. 8(d) shows the electrostatic-potential profile when a magnetic bound

B. Spatial distributions for multiple impurities in single-channeled transport: Nonlocal effects

In structures where multiple impurities are present, vortices form in the current pattern as a result of quantum interference between waves reflected from the walls and various impurities.[9] These vortices are a striking aftermath of nonlocality in phase-coherent quantum transport. They are not centered around impurities since they are a consequence of nonlocal effects, and their positions, as well as strengths, depend on the impurity configuration. In Fig. 9(a), we show the current distribution in a sample with two impurities without any magnetic field present. The structure is the same as that in Figs. 3–8 and the Fermi energy is 2.41 meV. There are two vortices in the current pattern. They have opposite circulations and produce antiparallel localized magnetic moments. As we turn on a magnetic field [see Fig. 9(b)], we see immediate quenching of one vortex accompanied by a slight reduction in the strength of the other. This different behavior for the two vortices can be explained by the fact that for one vortex, the Lorentz force tends to produce a circulation in a sense opposite to that of the vortex. Consequently, even a very low field is sufficient to quench this vortex. The resultant change in the quantum interference between the scatterers causes the other vortex to weaken as well, even though it has a circulation pattern in the same direction as that produced by the

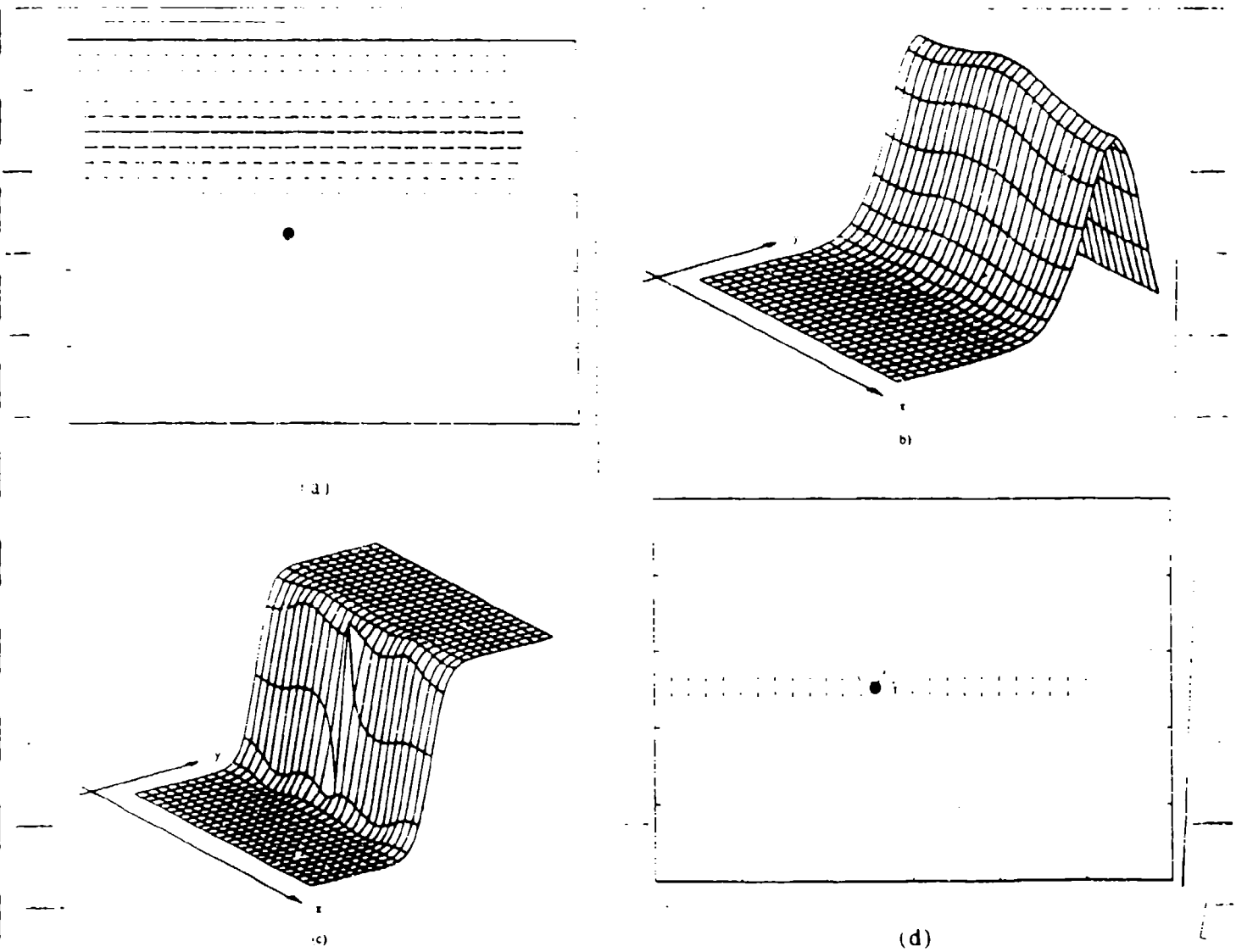


FIG. 9. The spatial distributions in the structure of Fig. 3 in a magnetic flux density of 2.18 T. The impurity is attractive. (a) The current distribution. The bulk of the forward current is carried by edge states. The small backscattered component near the walls was explained in Ref. 9. This figure should be compared with Fig. 3(a). Note that the local effects of the impurity potential are inhibited by the magnetic field. (b) The Fermi carrier concentration distribution which peaks near an edge because of the Lorentz force skewing the wave function towards that edge. (c) The electrostatic (or chemical) potential profile which [illegible] surfaces μ_1 along one wall and μ_2 along the other. The longitudinal four-terminal resistance that [illegible]

Lorentz force. As the magnetic field is increased slightly, the circulation of one vortex completely reverses while the other weakens further. This is seen in Fig. 9(c). At very high fields, the vortices completely disappear and the current is carried by the edge states from one terminal to the other.

C. Multichanneled transport

The previous results were for single-channeled transport in which an electron can scatter into only one propagating mode (subband). Some of the features observed in the single-channel case may really be consequences of single modality (i.e., only intrasubband scattering is allowed as opposed to intersubband scattering). In multichanneled conduction, there are more propagating modes available for scattering so that both intrasubband and intersubband scattering are allowed. This increases the total scattering probability and the current vortices that form require a larger magnetic field to be quenched. There are also more subtle effects. For injection from the second subband, the vortices that form are actually strengthened by a magnetic field. The net magnetic moment of the structure is increased by an external magnetic field so that the response is "paramagnetic." This is in contrast with the situation seen in single-channeled trans-

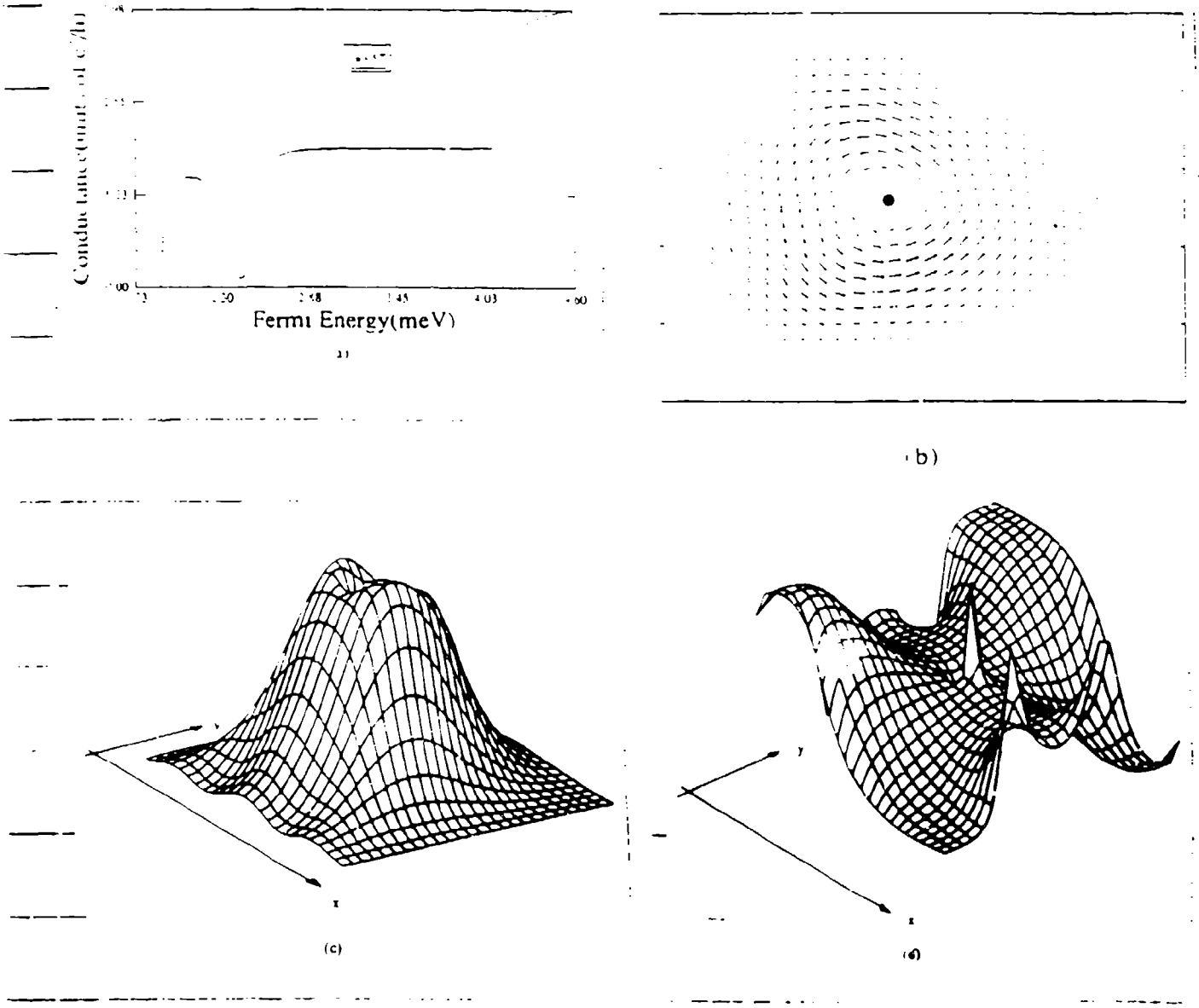


FIG. 8. Results for the formation of a magnetic bound state. (a) The conductance vs Fermi energy for the structure of Fig. 3 containing an attractive impurity. A magnetic flux density of 3.5 T has been applied. The dip in the conductance occurs at the energy at which a magnetic bound state forms. (b) The current distribution during the formation of a bound state. The current circulates

port Figs. 9) in which the response was "diamagnetic."

In Fig. 10, we have plotted the conductance versus Fermi energy of a GaAs quantum wire of length 900 Å and width 1000 Å containing four attractive impurities. The conductance shows quantized steps associated with subband filling.[4] The results are plotted for magnetic flux densities of $B = 0$ and 3.5 T. The quantization of the conductance steps is quite poor in the absence of a magnetic field because of significant backscattering from the impurity.[5,16] When a magnetic field of 3.5 T is turned on, the steps become wider and the lowest steps become much smoother. The steps become wider since the subband separation in energy increases with magnetic field and the widths of the steps are equal to these separations. In addition, the lowest steps become smoother since the transmission probabilities of the lowest subbands increase at high magnetic fields owing to the formation of edge states with high transmittivity.[5,18] The quantization for the higher steps is generally worse and does not improve very much with increasing magnetic field since the transmittivities of the higher subbands are always lower. The real-space edge state trajectories at the two edges corresponding to the higher subbands are spatially closer to one another than those corresponding to the lower subbands. This causes increased coupling between the higher subband edge states and hence increased backscattering.

In Figs. 11(a), 11(b), and 11(c), we plot the current distributions around a single attractive impurity when two subbands are occupied. The three figures are the current patterns due to injection from the first subband, the second subband, and the resultant current pattern, re-

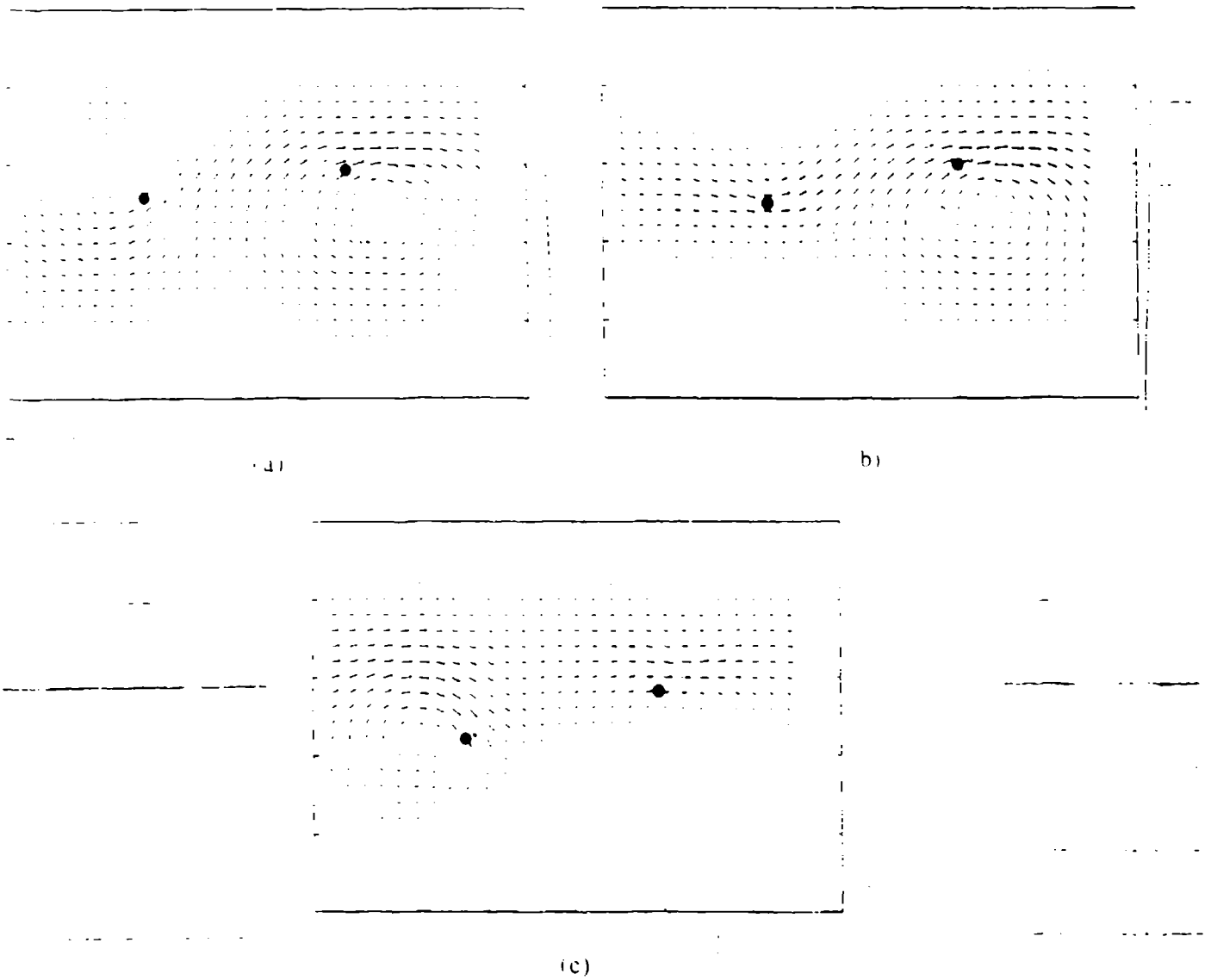


FIG. 9. The spatial distributions for a 900-Å-long and 1000-Å-wide structure containing multiple impurities. (a) The current profile in zero magnetic field showing the formation of two strong vortices due to quantum interference between the scatterers and [illegible] of the other by a magnetic flux density of 0.08 T. (c) Complete change

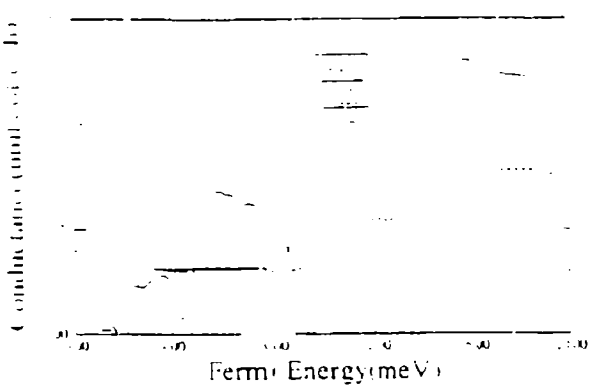


FIG. 10. Recovery of conductance quantization in a disordered structure by a magnetic field. The structure is 900 Å long and 1000 Å wide and contains two attractive impurities. In the absence of a magnetic field, the quantization is quite poor. The transmission of the lowest subband is very small so that the height of the lowest conductance step is also very small. A magnetic flux density of 3.5 T improves the quantization of the lowest steps significantly and also widens the steps. The quantization of the higher steps is always worse than that of the lower ones because of the relatively lower transmission of the higher subbands.

spectively. The current pattern of the second subband [Fig. 11(b)] does not show any significant backscattering component since the wave function of the second subband has a null at the center of the wire where the impurity is located. Consequently, electrons from the second subband do not interact with the impurity and therefore there is hardly any backscattering. A magnetic field of 1 T skews the wave function of the second subband and causes the electrons from this subband to interact with the impurity. This increases backscattering and hence we expect to see positive magnetoresistance in this case. The current profile for the second subband at a field of 1 T is shown in Fig. 11(d). There is a large backscattered component in the current clearly visible in the upper section of the device which was not present when the magnetic field was absent. Of course, at high enough magnetic fields, edge states form and backscattering is reduced.

D. The integer quantum Hall effect

In Figs. 12(a) and 12(b), we show the electrostatic-potential profiles in a 900-Å-long and 1000-Å-wide quantum wire with two impurities at magnetic fields of 0 and 3.5 T. Only one subband is occupied. We assume that $\mu_1 = 2.413$ meV and $\mu_2 = 2.408$ meV. Figures 12(c) and 12(d) show the corresponding electric field profiles. Note that in the presence of the magnetic field, both edges of the wire become extraordinarily smooth equipotential surfaces. Consequently, if we attached two voltage probes at any of the edges, they will measure a relative voltage difference of zero and hence the four-terminal effect. The voltage difference between the two edges is the Hall voltage drop. This figure should be contrasted with Fig. 8(d) which shows the potential profile when a magnetic bound state forms. In that case, the edges are no longer equipotential surfaces and the quantum Hall effect is not observed. It is well known that if the radius of the magnetic bound state is equal to or larger than the width of the wire, then the edges states at the two edges of the wire are coupled by the bound state and electrons can resonantly tunnel from one edge to the other resulting in backscattering. This destroys the quantum Hall effect as predicted by Jain and Kivelson[19] and later verified by Lee, McLennan, and Datta.[20] We also observe the same effect in Fig. 8(d).

IV. CONCLUSION

In this paper, we have plotted the current, Fermi carrier concentration, electrostatic-potential, and electric field profiles associated with various magnetotransport phenomena. These results are very helpful in elucidating many features of these effects.

ACKNOWLEDGEMENTS

The authors are indebted to Supriyo Datta, Craig S. Lent, S. M. Goodnick, and Henry Harbury for helpful comments. This work was supported by the U.S. Air Force Office of Scientific Research under Grant No. AFOSR 91-0211 and the Office of Naval Research under Contract No. N00014-91-J-1505.

APPENDIX: DERIVATION OF THE SCATTERING MATRIX FOR A SINGLE IMPURITY IN THE PRESENCE OF A MAGNETIC FIELD

In this Appendix, we derive the scattering matrix describing electron propagation across an elastic δ scatterer (impurity) in the presence of an arbitrary magnetic field. The *real-space matching method* discussed in Ref. 21 is employed. We take a section through the impurity across the width of the structure (y direction) and break this section into a number of mesh points along the width. Then, we enforce continuity of the wave function and its first derivative across the section.

If we assume the impurity to be located at position $(0,y)$, then the wave function for the incoming wave from the left side of the impurity can be written as

$$\psi_n^l(x,y) = \xi_n(x,y) + \sum_{p=1}^{M} r_{pn} \xi_{-p}(x,y), \quad x \le 0 ,$$

$$\psi_n^r(x,y) = \sum_{p=1}^{M} t_{pn} \xi_p(x,y), \quad x \ge 0 ,$$

where

$$\xi_{\pm p}(x,y) = \frac{\exp(ik_{\pm p}^x x)\phi_{\pm p}(y)}{\sqrt{|v_p^x|}}$$

coefficient from the jth mode to the ith mode.

For the incoming wave from the right side of the impurity, the wave function can be written as

$$\psi_n^+(x,y) = \xi_{-n}(x,y) + \sum_{p=1}^{M} r_{pn}\xi_p(x,y), \quad x \le 0,$$

$$\psi_n^+(x,y) = \sum_{p=1}^{M} t_{pn}\xi_{-p}(x,y), \quad x \le 0.$$

The scattering matrix across the impurity is defined as

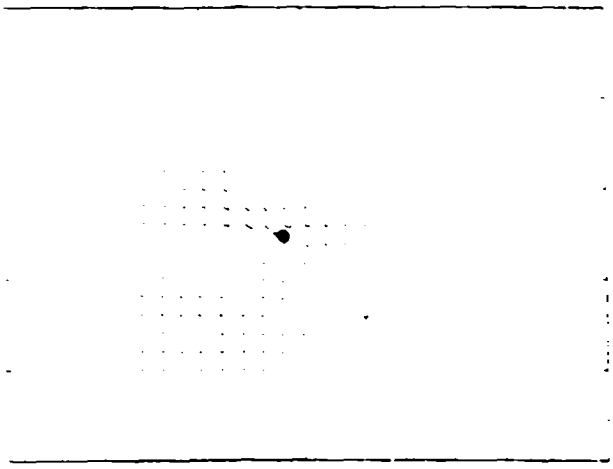
$$S = \begin{pmatrix} r & t' \\ t & r' \end{pmatrix}$$

where the individual elements r, r', etc., are themselves $M \times M$ matrices and M is the total number of modes (propagating plus evanescent).

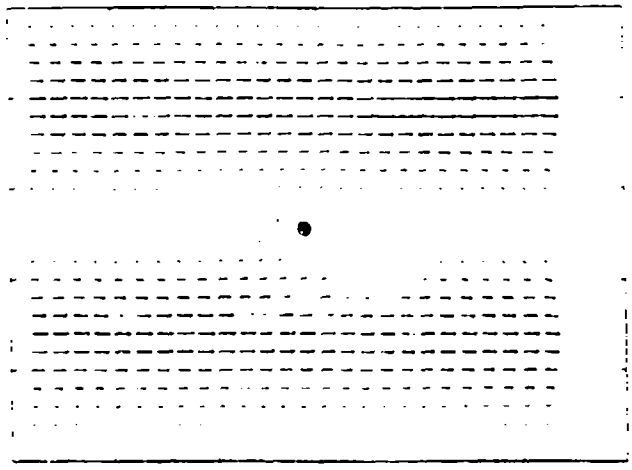
We now enforce the continuity of the wave function and its first derivative at all mesh points. This allows to write the set of real-space mode-matching equations for the incoming wave from the left

$$\psi_n^+(0,y_i) = \psi_n^+(0,y_{i'}), \quad \text{for all possible } i,$$

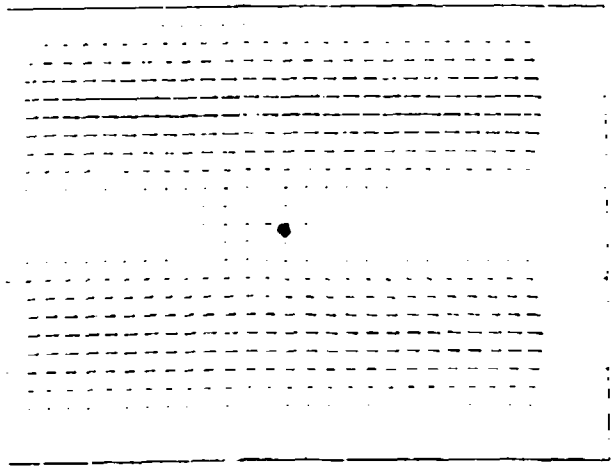
$$\frac{\partial}{\partial x}\psi_n^+(0,y_i) - \frac{\partial}{\partial x}\psi_n^+(0,y_{i'}) = \frac{2m^*}{\hbar^2}\delta(y-y_0)\psi(0,y_i),$$



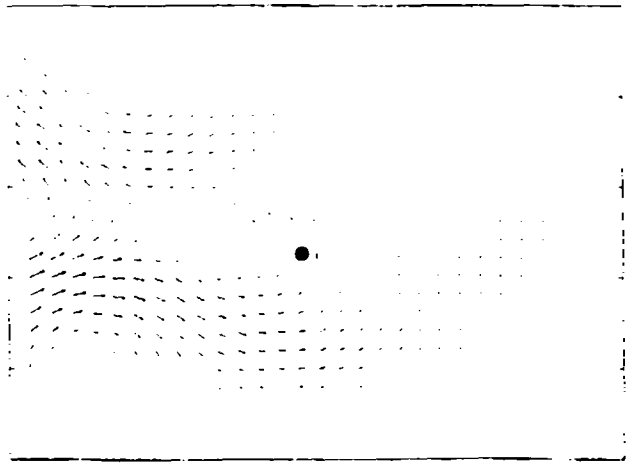
(a)



(b)



(c)



(d)

FIG. 11. Spatial distribution of multichanneled conduction in a structure of length 800 Å and width 1000 Å with one attractive impurity. (a) Current distribution for injection from the lowest subband only at zero magnetic field. (b) Current distribution for injection from the second subband at zero magnetic field. Not much backscattering is observed since the second subband wave function has a null at the impurity location. (c) The total current distribution at zero magnetic field. (d) The current distribution at a flux density of 1 T for injection from the second subband only. Significant backscattering is observed compared to (b) since the wave function of the second subband is skewed towards the impurity by the magnetic field which increases the momentum of the electrons

where y_i is the ith mesh point along the y direction.

At the location of the impurity, the derivative of the wave function is discontinuous by a δ function; elsewhere, it is continuous. Exact treatment of this condition is not possible. We average the quantities over the interval containing the impurity by multiplying every equation by some averaging function and then integrating over the interval. For our case, we simply choose the averaging function as

$$\langle n| = \frac{1}{y_{n-1} - y_n} \int_{y_n}^{y_{n-1}} dy \, .$$

Substituting for ψ_n and ξ_n, the above equations may be rewritten in matrix notation as

$$\begin{vmatrix} r1 & r1 & \cdots \\ r2 & r2 & \cdots \\ \cdots & \cdots & \cdots \end{vmatrix} \begin{vmatrix} \cdots \\ \cdots \\ \cdots \end{vmatrix} = \begin{vmatrix} c \\ d \\ \cdots \end{vmatrix} ,$$

where

$$r1_{ij} = \frac{1}{\sqrt{v_i}} I_{i,j} \, ,$$

$$r1_{ij} = -\frac{1}{\sqrt{v_{-j}}} I_{i,-j} \, ,$$

$$r2_{ij} = \left| \frac{ik_i}{\sqrt{v_i}} I_{i,j} - \frac{m^* \gamma}{\hbar^2 \sqrt{v_i}} \phi_{-j}(y_0) \delta_{i,0} \right| ,$$

$$r2_{ij} = \left| \frac{-ik_{-j}}{\sqrt{v_{-j}}} I_{i,-j} - \frac{m^* \gamma}{\hbar^2 \sqrt{v_{-j}}} \phi_{-j}(y_0) \delta_{i,0} \right| ,$$

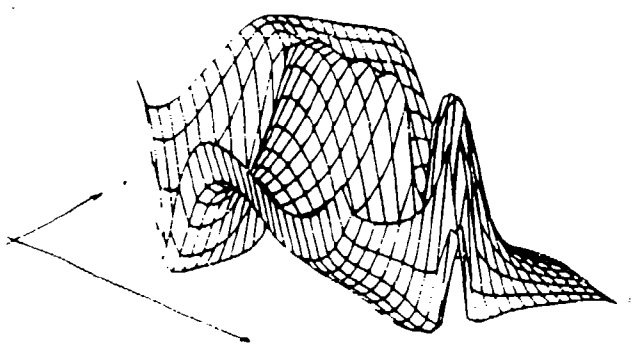
$$c_{ij} = \frac{I_{i,j}}{\sqrt{v_i}}$$

$$d_{ij} = \frac{ik_i}{\sqrt{v_i}} I_{i,j} + \frac{m^* \gamma}{\hbar^2 \sqrt{v_i}} \phi_j(y_0) \delta_{i,0} \, ,$$

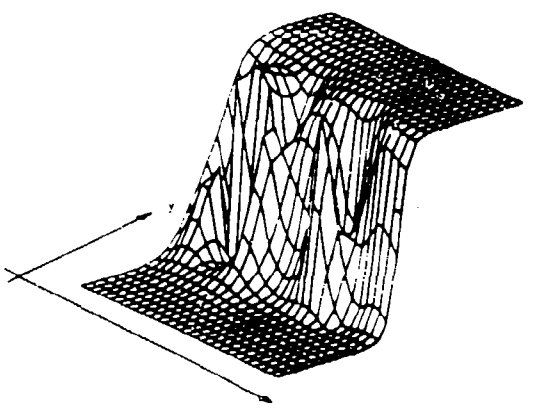
$$I_{i,j} = \langle i | j \rangle \, ,$$

$$\delta_{i,0} = 1 \quad (y_i = y_0)$$

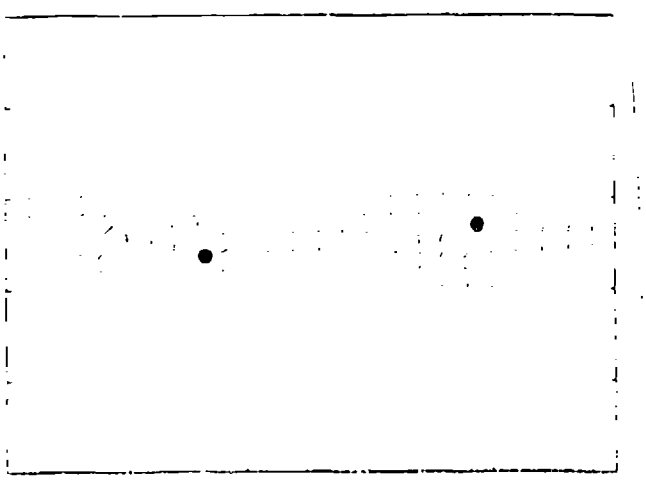
$$= 0 \quad (y_i \neq y_0) \, .$$



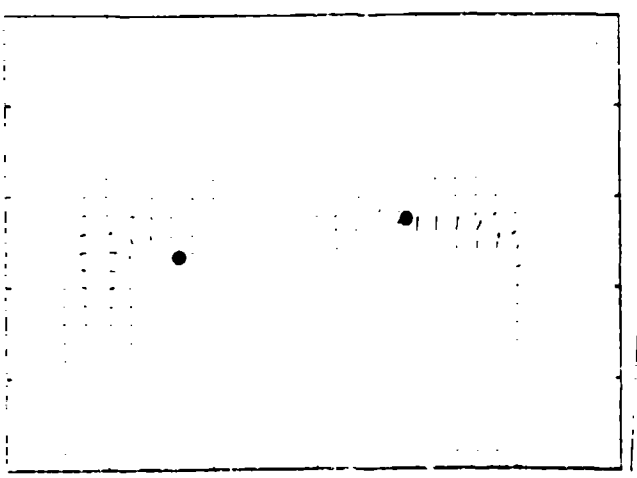
(a)



(b)



(c)



(d)

FIG. 12. Electrostatic (or chemical) potential and electric field profiles inside a structure with two attractive impurities. (a) Potential profile in the absence of any magnetic field. Significant potential variations are observed. (b) The potential profile at a magnet-

In a similar manner, the equations for the transmission and reflection coefficients for the incoming waves from the right can be derived,

$$\begin{pmatrix} t1' & r1' \\ t2' & r2' \end{pmatrix} \begin{pmatrix} c' \\ d' \end{pmatrix} = \begin{pmatrix} t' \\ r' \end{pmatrix},$$

where

$$t1'_{ij} = \frac{1}{\sqrt{v_i}} I_{i,-j},$$

$$r1'_{ij} = -\frac{1}{\sqrt{v_j}} I_{i,j},$$

$$t2'_{ij} = \left[\frac{ik_i}{\sqrt{v_i}} I_{i,-j} - \frac{m^* \gamma}{\hbar^2 \sqrt{v_j}} \phi_i(y_0) \delta_{i,0} \right],$$

$$r2'_{ij} = \left[\frac{-ik_i}{\sqrt{v_i}} I_{i,j} - \frac{m^* \gamma}{\hbar^2 \sqrt{v_j}} \phi_i(y_0) \delta_{i,0} \right],$$

$$c'_{ij} = \frac{I_{i,-j}}{\sqrt{v_j}},$$

$$d'_{ij} = \frac{ik_i}{\sqrt{v_j}} I_{i,-j} - \frac{m^* \gamma}{\hbar^2 \sqrt{v_j}} \phi_i(y_0) \delta_{i,0}.$$

For a free propagating region of length x which contains no impurity, we can write down the expressions for transmission and reflection coefficient matrices directly as

$$t_{ij} = \exp(ik_i x) \delta_{ij},$$

$$r_{ij} = r'_{ij} = 0,$$

$$t'_{ij} = \exp(ik_{-i} x) \delta_{ij}.$$

Current conservation requires that the scattering matrix be unitary and we must check for it at all impurity locations. The composite scattering matrix for the whole structure can be found by cascading the individual scattering matrices according to the law of composition of scattering matrices:[22,23]

$$S = S_1^{\text{free}} \otimes S_1^{\text{imp}} S_2^{\text{free}} \cdots \otimes S_N^{\text{imp}} \otimes S_{N+1}^{\text{free}},$$

where N is the total number of impurities.

A few statements about the numerics may be worthwhile. At high magnetic fields, the coefficient matrix required for finding the values of t and r becomes nearly singular and hence special numerical techniques are needed for matrix inversion. However, we found that the real-space matching technique is always numerically much better behaved than the momentum space mode-matching methods followed in Ref. 24. In fact, the real-space matching method ensures that current continuity is preserved to a much better approximation than momentum space mode-matching methods.

From the set of equations above, it is easy to show that the current across the δ scatterer is conserved and that the scattering matrix is mathematically unitary. To do this, we take the continuity equation and the derivative equations at any point, multiply the right side and the left side together, and then sum them over all points. This immediately shows the current continuity. We mention that the following relationships are useful in deriving the current continuity:

$$\int_{y=0}^{y=W} \phi_i(y) \phi_k(y) \left[k_j^F + k_k^F - \frac{2\gamma}{\hbar^2} \right] dy = 0, \quad j \neq k,$$

$$\int_{y=0}^{y=W} \phi_i(y)^2 dy = 1, \quad 1 \leq j \leq N.$$

[1] S. Chaudhuri, S. Bandyopadhyay, and M. Cahay, Phys. Rev. B 45, 11 126 (1992).

[2] C. S. Lent, Appl. Phys. Lett. 67, 1678 (1990); Zhen-Li Ji and K. F. Berggren, Phys. Rev. B 45, 6652 (1992); Zhen-Li Ji, Semicond. Sci. Technol. 7, 198 (1992).

[3] S. Chaudhuri and S. Bandyopadhyay, J. Appl. Phys. 71, 3027 (1992).

[4] For a review, see M. Büttiker, IBM J. Res. Dev. 32, 317 (1988).

[5] A. M. Kriman, N. C. Kluksdahl, and D. K. Ferry, Phys. Rev. B 36, 5953 (1987).

[6] C. S. Chu and R. S. Sorbello, Phys. Rev. B 42, 4928 (1990).

[7] M. J. McLennan, Y. Lee, and S. Datta, Phys. Rev. B 43, 13 846 (1991). This paper presents numerous spatial patterns that are calculated quantum mechanically in the limit of incoherent transport (i.e., in the presence of phase-randomizing scattering).

[8] O. Entin-Wohlman, C. Hartzstein, and Y. Imry, Phys. Rev. B 34, 921 (1986).

[9] R. Landauer, IBM J. Res. Dev. 1, 223 (1957); Z. Phys. B 21, 247 (1975).

[11] C. S. Lent, Phys. Rev. B 43, 4179 (1991).

[12] M. Büttiker, Phys. Rev. B 38, 9375 (1988).

[13] T. Ando and Y. Uemura, J. Phys. Soc. Jpn. 36, 959 (1974); H. Tamura and T. Ando, Phys. Rev. B 44, 1792 (1991).

[14] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon, Phys. Rev. Lett. 60, 848 (1988); D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. C 21, L209 (1988).

[15] S. He and S. Das Sarma, Phys. Rev. B 40, 3379 (1989); E. G. Haanapel and D. van der Marel, *ibid.* 39, 5435 (1989); C. S. Chu and R. S. Sorbello, *ibid.* 40, 8559 (1989); E. Tekman and S. Ciraci, *ibid.* 42, 9098 (1990).

[16] I. Kander, Y. Imry, and U. Sivan, Phys. Rev. B 41, 12 941 (1990).

[17] B. J. van Wees, L. P. Kouwenhoven, H. van Houten, C. W. J. Beenakker, J. E. Mooij, C. T. Foxon, and J. J. Harris, Phys. Rev. B 38, 3625 (1988).

[18] T. Ando, Phys. Rev. B 44, 8017 (1991); K. Shepard, *ibid.* 44

[3] Y. Lee, M. J. McLennan, and S. Datta, Phys. Rev. B **43**, 4333 (1991).
[4] R. Fronne and S. Datta, J. Appl. Phys. **64**, 4086 (1988).
[5] M. Cahay, M. McLennan, and S. Datta, Phys. Rev. B **37**, 10125 (1988); M. Cahay, Ph.D. thesis, Purdue University 1987 (unpublished).
[6] R. Redheffer, J. Math. Phys. **41**, 1 (1962).
[7] A. M. Kriman and P. Ruden, Phys. Rev. B **32**, 8013 (1985); D. G. Ravenhall, H. W. Wyld, and R. L. Shult, Phys. Rev. Lett. **62**, 1780 (1989); Y. Takagaki and D. K. Ferry, Phys. Rev. B **44**, 8399 (1991).

Numerical calculation of hybrid magneto-electric states in an electron waveguide

S. Chaudhuri and S. Bandyopadhyay
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 30 October 1991; accepted for publication 6 December 1991)

We have performed a numerical calculation of the energy dispersion relation of hybrid magneto-electric states (both propagating and evanescent) in an electron waveguide subjected to a magnetic field. Our results are considerably different from those obtained through the Bohr–Sommerfield quantization condition. We have also calculated the density of the magneto-electric states as a function of energy and the velocity versus energy relationships. Finally, we show how the wavefunctions of these states evolve with increasing magnetic field from particle in a box states to edge states. These results are useful in the analysis of numerous recent magnetotransport experiments performed in electron waveguides.

Electron waveguides (quasi-one-dimensional wires in which carrier transport is partially ballistic) has been the focus of many recent theoretical and experimental investigations.[1] Of particular interest in these structures is the nature of magneto-electric states that are formed under an applied transverse magnetic field.[2] These states are important in understanding numerous magnetotransport experiments such as the conductance oscillations of quasi-one-dimensional electron gases in a magnetic field,[3] the integer quantum Hall effect,[4] etc. In the past, magnetoelectric subbands and their wave functions were calculated in an approximate manner using analytical functions such as Weber or hypergeometric functions.[5] In this paper, we have calculated the wave function, the energy dispersion relation, the velocity versus energy relationships, and the density of these states starting from the Schrödinger equation. Our calculation is described below.

We consider an electron waveguide (a quasi-one-dimensional structure) as shown in Fig. 1. The magnetic field is directed along the z direction along which the confinement is complete in that only one transverse subband is occupied. The Schrödinger equation describing electrons in this structure is

$$\frac{(\mathbf{p}-e\mathbf{A})^2}{2m^*}\psi(x,y)+V(y)\psi(x,y)=E\psi(x,y), \tag{1}$$

where $\mathbf{A}$ is the magnetic vector potential, m^* is the electron's effective mass, and $V(y)$ is the electric confining potential in the y direction. The simultaneous presence of a magnetic potential and an electric potential hybridizes the electronic states causing them to be magneto-electric states.

The potential $V(y)$ is chosen to be consistent with hardwall boundary conditions

$$V(y)=0 \quad |y|<d$$
$$=\infty \quad |y|>d, \tag{2}$$

where $2d$ is the width of the structure in the y direction. We choose the Landau gauge

$$\mathbf{A}=(-By,0,0), \tag{3}$$

where B is the z directed magnetic flux density.

The wave function ψ can be written as

$$\psi(x,y)=e^{ik_x x}\phi(y), \tag{4}$$

where k_x is the x directed wave vector and $\phi(y)$ satisfies the eigenequation

$$\frac{\partial^2\phi(y)}{\partial y^2}+\frac{2m^*}{\hbar^2}E\phi(y)-\left(\frac{y}{l^2}\right)^2\phi(y)+2\frac{y}{l^2}k_x\phi(y)$$
$$-k_x^2\phi(y)=0 \tag{5}$$

with l being the magnetic length given by $l=\sqrt{\hbar/eB}$. To find the wave function $\phi(y)$ of the magneto-electric states, we have to solve the above equation subject to the boundary conditions

$$\phi(y=d)=\phi(y=-d)=0 \tag{6}$$

which follow from Eq. (2).

In order to find the energy dispersion relation E vs k_x for the magneto-electric states, we have to find the values of the wave vector k_x for a given energy E satisfying Eqs. (5) and (6). We then repeat this for various values of E to obtain the energy dispersion relation. Unfortunately, this is not straightforward since Eq. (5) is not an eigenequation in k_x for a given E due to the fact that it is nonlinear in k_x.

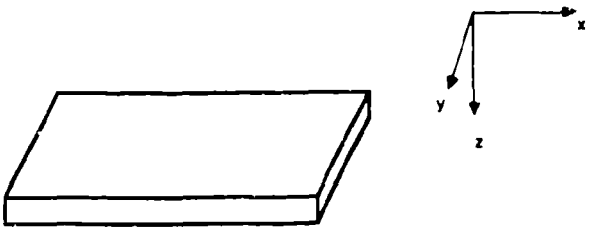


FIG. 1. An electron waveguide subjected to a magnetic field along the z axis.

 0021-8979/92/063027-03$04.00

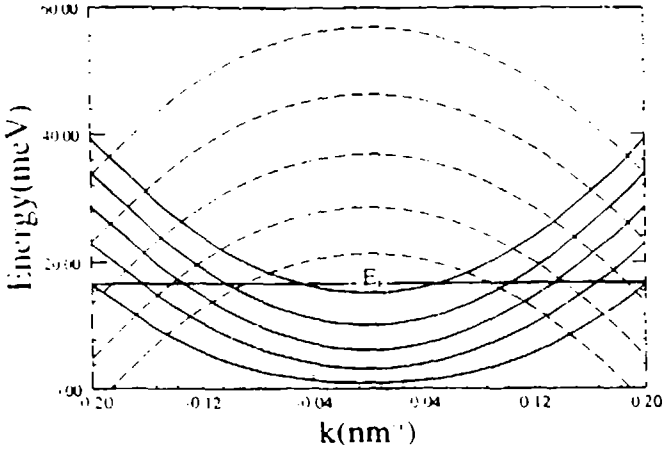


FIG. 2. Energy dispersion E vs k_z of magneto-electric subbands in a waveguide subjected to a magnetic field. The waveguide transverse width is 100 Å and the magnetic field is 1 T. The Fermi energy is taken to be 17 meV. If the subband bottoms are below the Fermi energy, the states are propagating states and have real wave vectors; otherwise they are evanescent states and have imaginary or complex wavevectors. For the parameters (well width and magnetic field) chosen, the lowest evanescent states have purely imaginary wave vectors. The evanescent states are shown with negative curvatures. The energy separation between the subband bottoms depends on the magnetic field as well as the width of the waveguide.

We therefore have to convert Eq. (5) into an eigenequation in k_x using the following transformation.

Let

$$\zeta(y) = k_x \phi(y) \tag{7}$$

Equation (5) can now be recast as

$$\begin{pmatrix} 0 & 1 \\ \frac{\partial^2}{\partial y^2} + \frac{2m^*}{\hbar^2}E - \left(\frac{y}{l^2}\right)^2 & \frac{2y}{l^2} \end{pmatrix} \begin{pmatrix} \phi \\ \zeta \end{pmatrix} = k_x \begin{pmatrix} \phi \\ \zeta \end{pmatrix}. \tag{8}$$

We discretize Eq. (8) within a finite difference scheme breaking the y domain into N grid points where N is some suitably large integer. We can now readily solve for N different k_x eigenvalues ($k_x^1, k_x^2 \ldots k_x^N$) for any given energy E and find the corresponding eigenvectors which give the wave function $\phi_n(y)$. Each value of n corresponds to a magneto-electric subband. This gives us the energy E versus the k_x^n ($n = 1,2,3,\ldots N$) relation, or the dispersion relation for N subbands. We vary E up to an arbitrary maximum which we call the "Fermi energy" E_F. If the bottom of a subband is above E_F, then the corresponding state is an evanescent state; otherwise, it is a propagating state. Evanescent states have imaginary or complex wave vectors (k_x values) whereas propagating states have real k_x values. It is interesting to note that while in the absence of a magnetic field evanescent states can only have imaginary wave vectors, in the presence of a magnetic field they can have both imaginary and complex wave vectors. The states with complex wave vectors can interfere since their wave functions are complex and they have nonzero current densities associated with them *even though they are evanescent*. It is interesting to note that because of this feature one could observe, say, the magnetostatic Aharonov–Bohm effect with the evanescent states but not the electrostatic Aharonov–Bohm effect.

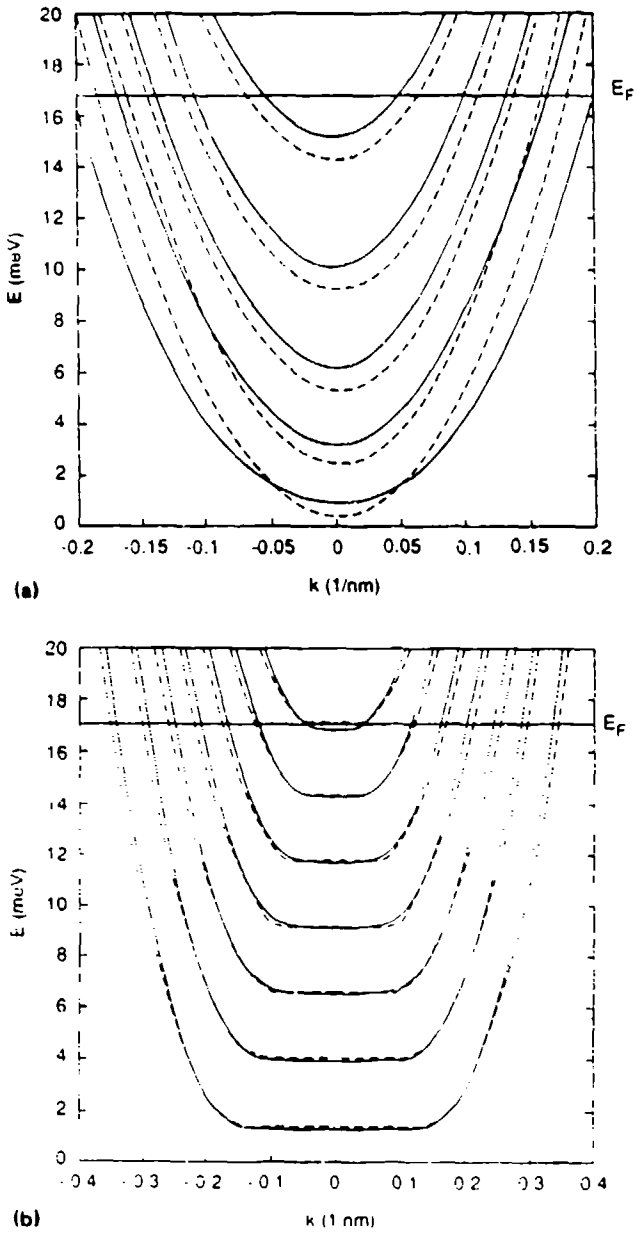


FIG. 3. Energy dispersion relations (for the propagating states only) obtained from our results (solid curves) and the Bohr–Sommerfield quantization rule (broken curves). (a) The waveguide width is 1000 Å and the magnetic field is 1 T; (b) the waveguide width is 2000 Å and the magnetic field is 1.5 T.

In Fig. 2 we show the energy dispersion relations including both propagating and evanescent states. In Figs. 3(a) and 3(b) we compare the dispersion relations of the propagating states with those obtained from the Bohr–Sommerfield quantization rule.[6,7] We find that there is a significant difference between the two results. In Fig. 4, we show the velocity versus energy relations for different subbands. The velocity in the nth subband is defined as $v_n = (1/\hbar)(\partial E_n/\partial k_x)$.[6] In Fig. 5, we show the density of states versus energy where the density of states is defined as $\Sigma_n (4/hv_n)\theta(E_n - E_n^0)$ with θ being the Heaviside step

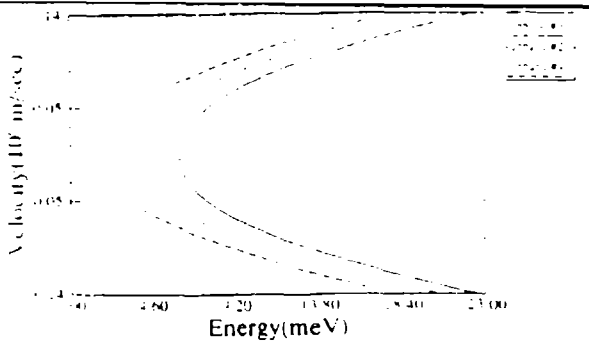


FIG. 4. Velocity vs energy relationships for the magneto-electric states in a waveguide width of 1000 Å and a magnetic field of 1 T.

function and E_n^0 the energy at the bottom of the nth subband. The summation is carried out over the propagating states only. The density of magnetoelectric states is useful in calculating the optical absorption spectra of a quantum wire subjected to a magnetic field, or the scattering rates for hot electrons in the presence of a magnetic field, and various other quantities.

Finally, in Figs. 6(a), 6(b), and 6(c), we show how the probability density $|\phi(y)|^2$ of the lowest subband at the energy $E = E_F$ evolves with increasing magnetic field. At zero magnetic field, the state is a particle in a box state, whereas at high magnetic fields, it peaks near the edges of the waveguide and becomes an edge state. For a given sign of the wave vector k_x, the probability density peaks near one of the edges of the waveguide whereas for the opposite sign of k_x, it will peak near the opposite edge. This behavior can be understood as skewing of the wavefunction towards one of the edges due to the Lorentz force applied on the electrons.

In conclusion, we have shown in this paper the nature of magneto-electric states in an electron waveguide. These states have been invoked to explain several experiments and phenomena such as magnetoconductance oscillations in electron waveguides and the quantum Hall effect. We

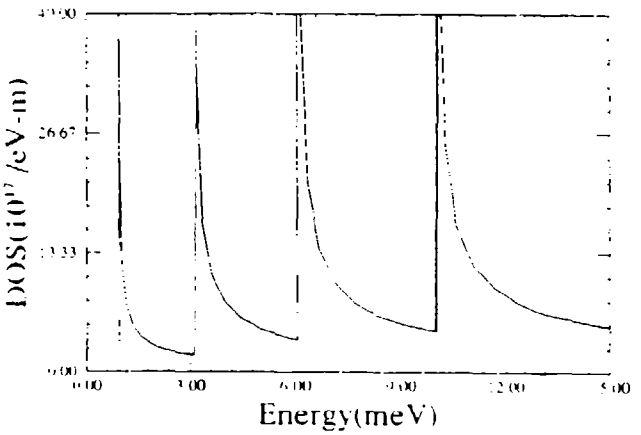


FIG. 5. Density of magneto-electric states in a 1000 Å wide waveguide at a field of 1 T.

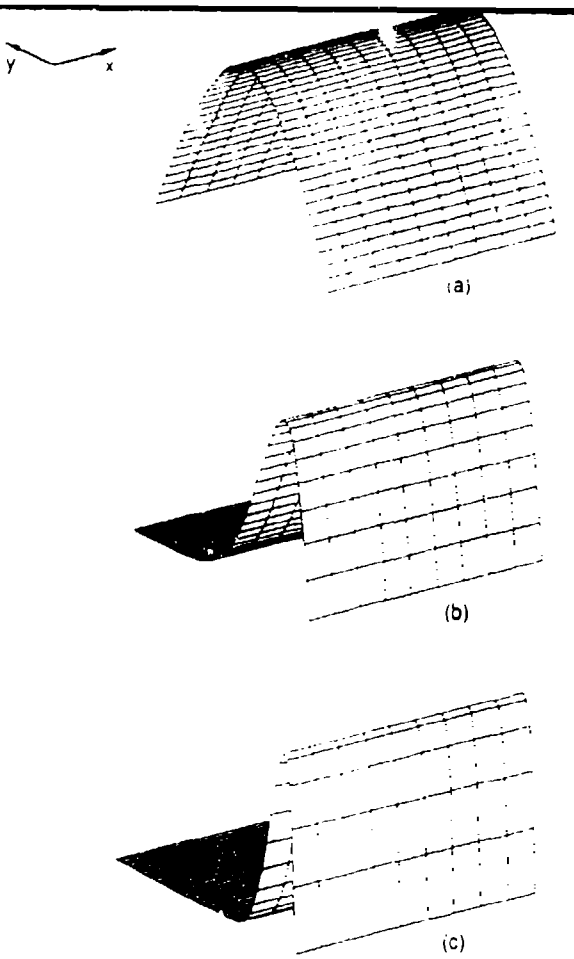


FIG. 6. Probability density in the lowest magneto-electric subband at the Fermi energy of 17 meV. The waveguide width is 1000 Å. (a) The magnetic field is zero; (b) the magnetic field is 1 T; (c) the magnetic field is 10 T.

believe that our results will be useful in the analysis of these and other experiments dealing with magneto-electric states.

This work was supported by the U.S. Air Force Office of Scientific Research under grant No. AFOSR 91-0211 and by the U.S. Office of Naval Research under grant number N00014-91-J-1505.

[1] For recent reviews, see, for example, *Nanostructure Physics and Fabrication*, edited by M. A. Reed and W. P. Kirk (Academic, Boston, 1989).
[2] See, for example, W. Zawadski, S. Klahn, and U. Merkt, Phys. Rev. B **33**, 6916 (1986); J. Yoshino, H. Sakaki, and T. Hotta, Surf. Sci. **142**, 326 (1984).
[3] S. B. Kaplan and A. C. Warren, Phys. Rev. B **34**, 1346 (1986).
[4] M. Büttiker, Phys. Rev. B **38**, 9375 (1988).
[5] A. H. MacDonald and P. Středa, Phys. Rev. B **29**, 1616 (1984).
[6] C. W. J. Beenakker, H. van Houten, and B. J. van Wees, Superlatt. and Microstructures **5**, 127 (1989).
[7] C. W. J. Beenakker and H. van Houten, Solid State Phys. **44**, 228 (1991).

A Monte Carlo Study of Correlations Between Impurity Scattering Events in a Two dimensional Electron Gas Giving Rise to Inhomogeneous Magnetoresistance

N. Telang and S. Bandyopadhyay
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

(Received 30 August 1991)

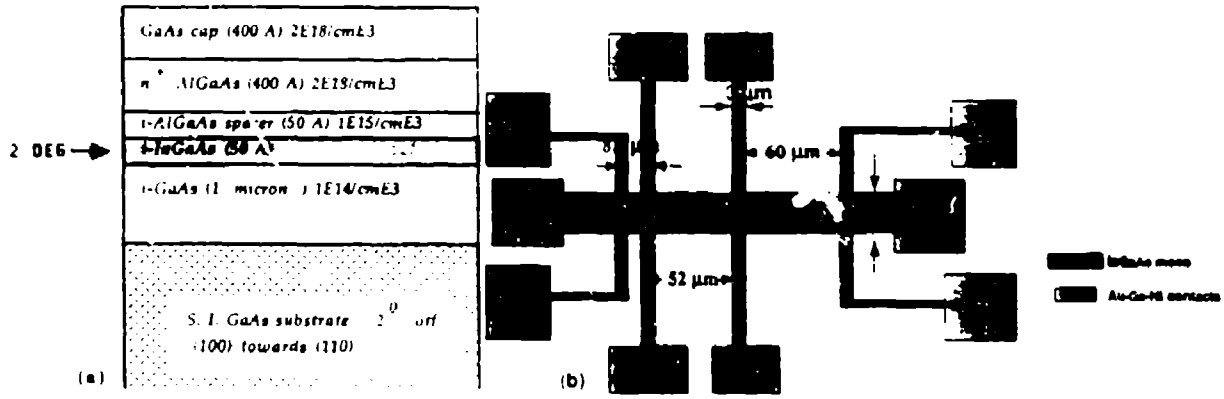
In a recent experiment, the low temperature linear response magnetoresistance of an InGaAs quantum well patterned into a Hall bar exhibited strikingly different behavior when measured between different voltage probes of the Hall bar. This anomaly was explained as being caused by subtle correlations between impurity scattering events (in the presence of a magnetic field) which are beyond Fermi's Golden Rule. In this paper, we present a Monte Carlo simulation of electron transport in which these correlations are captured by retaining information about the impurity configuration, i.e. the precise locations of the scattering centers within the sample. The results of the calculation agree qualitatively with the experimental data.

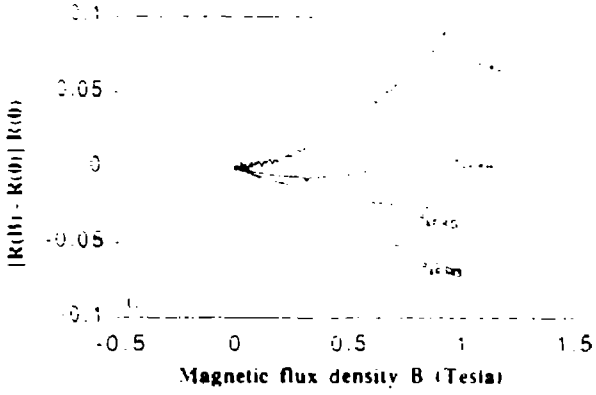
Introduction

In a recent experiment[1], the magnetoresistance of an InGaAs quantum well patterned into a Hall bar was measured at a temperature of 4.2 K under low bias ($\sim 4\ \mu$V). The magnetoresistance exhibited different behavior when measured between different voltage probes of the Hall bar and even showed different signs near zero magnetic field. The test structure and the data are shown in Fig. 1.

The observation of radically different magnetoresistance behavior in the *same* sample depending on the locations of the measuring probes (i.e where the sample is probed) is a rather unusual occurence since all well-known magnetoresistance mechanisms are *homogeneous* effects that cannot cause different behavior in different regions of a sample. However, it is possible to explain these results if the magnetoresistance is influenced by correlations between impurity scattering events that depend on the precise locations of the impurities (the impurity configuration). Since the impurity configuration is different in different regions of a sample, one would expect to see different magnetoresistance in different sample regions. This can explain the experimental observations. In Fig. 2 we elucidate the nature of these correlations by showing that a magnetic field can either enhance backscattering (positive magnetoresistance) or reduce backscattering (negative magnetoresistance) depending on how the scatterings between the different impurities are correlated. Such an effect is not masked by ensemble averaging over electron trajectories or by anisotropic inelastic scattering events (time varying perturbations) such as electron-electron scattering which do not perturb the trajectories significantly. A similar mechanism for magnetoresistance was proposed by Beenakker and van Houten in narrow samples[2].

0749-6036/92/010099 - 04 $02.00/0



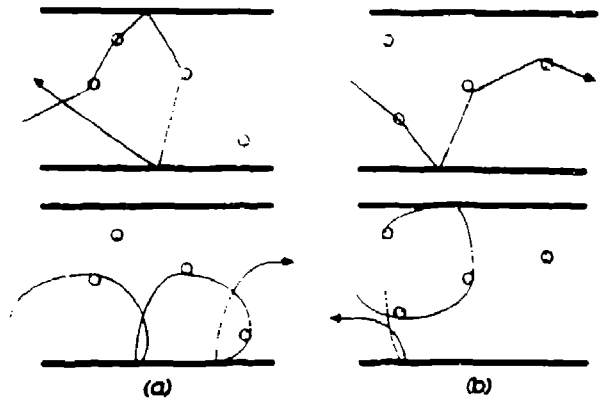


• Fig. 1: The experimental structure used in Ref. 1. (a) The quantum well of InGaAs, (b) the Hall bar showing the current and voltage contacts, and (c) the experimental magnetoresistance curves (the resistance $R_{mn,ij}$ stands for the four-terminal reistance measured by passing current between contacts m and n while measuring voltage between contacts i and j).

In this paper, we present a simple Monte Carlo model of electron transport in a confined two-dimensional electron gas to study the correlations between impurity scattering events in the presence of a magnetic field. Impurity scattering is the only scattering mechanism that is treated since it is overwhelmingly dominant over all other scattering mechanisms under the experimental conditions. The correlations between impurity scattering events are accounted for by keeping track of the exact locations of the impurities, i.e. the precise coordinates at which the scattering events take place. In the next section, we describe our model.

The Monte Carlo Model

In the Monte Carlo simulation, electrons are injected into the simulation domain with the Fermi velocity v_F at arbitrary angles. The injections at Fermi velocity are necessitated by the fact that at low bias and low temperature, it is only the electrons at the Fermi energy that carry current. The electrons are acted on by the Lorentz force $-ev_F \times B$ where B is the magnetic flux density. The Lorentz force, impurity scattering and scattering from the boundaries of the sample are the only three mechanisms that affect the direction of the electron velocity within the simulation domain and determine the classical trajectories. The only other scattering mechanism that could be considered is electron-electron scattering (although it is known to be ten times weaker than impurity scattering from experimental data). However, electron-electron scattering is very anisotropic and favors small angle scattering. For this reason, it is not at all effective in deflecting the trajectories significantly.

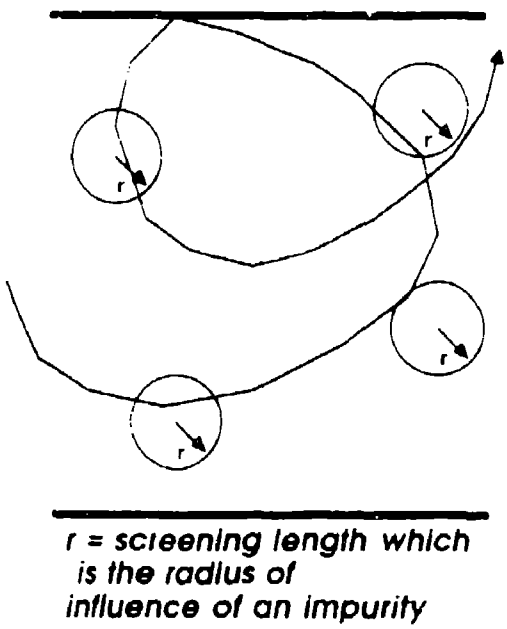


• Fig. 2: Possible electron trajectories inside a disordered sample in the absence (upper figure) and presence (lower figure) of a magnetic field. The impurities are denoted by shaded circles: (a) for this impurity configuration the magnetic field suppresses backscattering and enhances transmission so that the resistance decreases and the magnetoresistance is negative, (b) for this configuration the opposite happens to make the magnetoresistance positive.

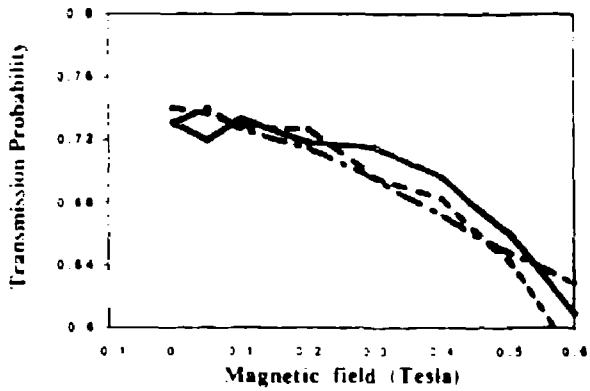
During the simulation, the electron position is sampled regularly with a period that is much smaller than the mean time between collisions estimated from the measured mobility. During each sampling, the direction of the Lorentz force is updated. If during a sampling, the electron is found within a screening distance from an impurity, it is made to scatter. Since the scattering is elastic, it does not change the magnitude of the electron's velocity, but it changes the direction. The scattering angle is determined within the framework of Fermi's Golden Rule, i.e. the anisotropy of the scattering is explicitly taken into account.

The impurity locations (impurity configuration) are generated by two independent random number generators that give the x- and y-coordinates. The average distance between the impurities is maintained at the mean-free-path calculated from the measured mobility and carrier concentration (the latter determines the Fermi velocity).

Boundary scattering is treated as follows. If an electron collides with a boundary of the sample, it is made to undergo either specular reflection (angle of incidence is equal to angle of reflection) or diffuse reflection (angle of reflection is arbitrary). Again, these elastic collisions do not change the magnitude of the electron's velocity, but change the direction depending on whether specular or diffuse reflection is chosen. The boundaries are placed at locations corresponding to the physical dimensions of the sample. A typical Monte Carlo trajectory is shown in Fig. 3.



• Fig. 3: A typical Monte Carlo trajectory traced out by an electron in the simulation. The centers of the circles are occupied by the impurities and the radii of the circles are all equal to the screening length.



• Fig. 4: Transmission probability versus magnetic flux density. The three traces are for three different impurity configurations. Note that the slopes of the curves near zero magnetic field are different for different configurations which means that the low-field magnetoresistance can be positive or negative depending on the configuration.

From the Monte Carlo simulation, we estimate the classical transmission probabilities of electrons through the simulation domain at various values of the magnetic field. The transmission probability is simply the fraction of electrons injected at the left edge of the simulation domain (left contact of the sample) that emerge through the right edge (right contact). In the Landauer-Büttiker picture of linear response transport[1] this transmission probability is related to the resistance of the sample so that if the transmission probability is significantly different for different impurity configurations, we can establish that the resistance can be significantly different for different impurity configurations as well. This, then, can explain the experimental observations.

Simulation Results

The simulation parameters that we used for our study are the following: the number of particles simulated is greater than 4000 which ensures statistical reliability. The Fermi velocity $v_F = 1.1 \times 10^8$ cm/sec (estimated from the measured carrier concentration of 1×10^{12} cm^{-2} in samples used in the experiment), the average distance between impurities = 7000 Å as estimated from the mean free path (calculated from the measured mobility of 18,000 $cm^2/V-sec$). The screening length was estimated to be 60 Å based on a Fermi-Thomas model. The sample domain is chosen large enough to contain 40 impurities.

In Fig. 4, we plot the transmission probability versus magnetic field for three different impurity configurations. Note that for some configuration, the transmission probability at first increases with the magnetic field which corresponds to negative magnetoresistance near zero field.

and for some other configuration the transmission probability decreases which is indicative of positive magnetoresistance. At high enough fields, the transmission always decreases with increasing field as the cyclotron radius becomes comparable to and then less than the separation between impurities. In this high field regime, the magnetoconductance is always negative as given by classical theory. All these features are observed in the experimentally measured characteristics. Although a quantitative comparison with the experimental data is impossible because of the uncertainties in several parameters, it is obvious that there is qualitative agreement between the simulation results and the experimentally observed characteristics.

Conclusion

In this paper, we have presented a Monte Carlo calculation of electron transport in a confined two-dimensional electron gas in the presence of impurity scattering and under the influence of a magnetic field. We have accounted for correlations between the impurity scattering events by keeping track of the impurity positions within the sample, i.e. the exact locations where the scattering events occur. The simulation shows that the dependence of the transmission probability through a disordered sample on the magnetic field is sensitive to the impurity configuration because of the correlations between the impurity scattering events. Since the transmission probability determines the resistance of the sample, this means that the magnetoresistance is a function of the impurity configuration. This dependence of the magnetoresistance on the impurity configuration explains the observations made in the experiment of Ref. 1.

Acknowledgement - This work was supported by the Air Force Office of Scientific Research under Grant No. AFOSR 91-0211, by the Jesse H. Jones Foundation and by IBM.

REFERENCES

1. S. Subramaniam, S. Bandyopadhyay, B. Das, G. H. Bernstein and P. A. Sekula Moisé, submitted to *Physical Review B*

2. C. W. J. Beenakker and H. van Houten, *Solid State Physics* Vol. 44, eds. H. Enrenreich and D. Tuimbull (Academic Press, New York, 1991).

3. M. Büttiker, *IBM Journal of Research and Development*, Vol. 32, 317 (1988).

Coupling and Crosstalk Between High Speed Interconnects in Ultralarge Scale Integrated Circuits

Supriyo Bandyopadhyay, *Senior Member, IEEE*

***Abstract*—The advent of sophisticated lithographic techniques has made it possible to fabricate densely packed ultralarge-scale-integrated (ULSI) circuits. In these chips, interconnect lines are so narrow and spaced in such close proximity that signal from one line could easily get coupled to another causing interference and crosstalk. This paper presents a general theory to model coupling between optical interconnects (waveguides) and quantum-mechanical coupling between narrow and very closely spaced silicide interconnects embedded in dielectrics (SiO_2).**

I. Introduction

RECENT advances in nanolithography [1]-[6] have made it possible to delineate electronic devices with feature sizes of a few hundred angstroms in both silicon and GaAs wafers. The ability to make such small features has led to the development of ultralarge-scale-integrated (ULSI) circuits with packing densities far greater than those of very-large-scale-integrated (VLSI) circuits. In 1980, Hewlett-Packard produced a single chip microprocessor with 0.5 million devices on a 1 sq cm chip using 1.25 μm feature sizes [7]. Today, we are envisioning chips with 10^9 devices. Simple scaling law shows that the feature sizes required for such ULSI chips will be a few hundred angstroms causing extremely dense packing.

In ULSI chips not only are the devices densely packed, but so are the interconnects. The close proximity of interconnect lines resulting from this dense packing could lead to severe problems with mutual interference and crosstalk. In this paper, we study crosstalk due to coupling between optical interconnects (GaAs waveguides embedded in AlGaAs) and coupling between narrow and closely spaced silicide interconnects surrounded by a dielectric. In the case of silicide interconnects, coupling is caused by quantum-mechanical tunneling of electrons from one line to another. This kind of coupling will be especially severe if the dielectric in which the interconnects are embedded is leaky (such as porous Si_3N_4 or SiO_2 grown by wet oxidation). In addition to causing crosstalk, tunneling can give rise to a unique problem. In multilayered interconnects, there can be crossings of two lines with a very thin dielectric layer sandwiched between them. If the thickness of the lines is a few hundred angstroms, then at the crossing we have a crossover capacitor whose effective plate area is a few hundred angstroms square and the plate separation is also of the same order. The corresponding capacitance can be estimated from standard formula [8]. If the linewidths are 300 Å and the plate separation is 100 Å, then the crossover capacitance is $\sim 10^{-17}$ F. Since the dielectric layer between the plates is thin enough, an electron can tunnel through this layer from one interconnect to another. Such tunneling can charge up the capacitor to 10 mV per electron! Effects such as these have already been observed in single discrete capacitors [9]. Obviously, stray voltages of this nature are undesirable in an integrated circuit and can cause reliability problems, logic errors, etc., especially if the supply voltages have been scaled down with the device sizes.

The close physical proximity of neighboring lines is not the only cause of increased coupling in ULSI. The increasing length of interconnects with increasing chip size also contributes to increased coupling since a larger region is available for interaction when the interconnects are long. Suffice it to say, then, that crosstalk and coupling can be a serious problem in large ULSI chips because of the dense packing and long interconnect length. To understand the nature of this coupling and finally to find ways of countering it, one needs to be able to describe and model such coupling effects within a basic microscopic formalism. In this paper, we present such a model derived from general coupled mode theory [10]-[18]. The formalism is perfectly general and applies for coupling between both optical and silicide interconnects, i.e., for both optical coupling and quantum mechanical coupling (tunneling). The applicability of the model for both types of coupling is a rather fortuitous coincidence which happens because the basic equations describing the two types of coupling at the microscopic level are the electromagnetic wave equation (derived from Maxwell's equation) and the Schrödinger equation which are mathematically similar. In the next section, we present the theory and then the Section III we estimate coupling coeffi-

Manuscript received August 23, 1991; revised December 4, 1991. This work was supported by the Office of Naval Research by Grant N00014-91-J-1505 and by IBM through a Faculty Development Award.

The author is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556.

IEEE Log Number 9107782.

cients for both types of coupling. In Section IV we present results for GaAs-AlGaAs optical interconnects and silicide interconnects. Finally, in Section V, we present the conclusions.

II. Theory

A. Coupling Between Optical Interconnects

To model electromagnetic coupling between a set of closely spaced optical interconnect lines, we view and the interconnects as optical waveguides and start from the wave equation that governs the propagation of an electromagnetic signal in a waveguide. We assume that the interconnects are nonlossy and nondispersive. *This is a very good assumption for optical interconnects comprising GaAs waveguides surrounded by AlGaAs cladding*. The scalar wave equation for a TE mode propagating in the $\vec{x}$ direction in any one waveguide reads [10]

$$\nabla^2 E_y(\vec{r}, t) = \mu\epsilon \frac{\partial^2 E_y}{\partial t^2} + \mu \frac{\partial^2}{\partial t^2} [P_{coup}(\vec{r}, t)]_y \quad (1)$$

where E_y is the y component of the electric field in the interconnect (waveguide) and $[P_{coup}]_y$ is the y component of a distributed polarization source caused by the coupling of signal from other interconnects. The quantities μ and ϵ are the permeability and permittivity of the interconnects.

To solve for the field E_y in the above equation, we invoke standard coupled mode theory. The solution $E_y(\vec{r}, t)$ can be written as a linear superposition of the normal modes (unperturbed fields) in the individual interconnects

$$E_y(\vec{r}, t) = \sum_n C_n(x) \mathcal{E}_y^{(n)}(y, z) e^{i\omega t} \quad (2)$$

where $\mathcal{E}_y^n(y, z)$ is the y component of the electric field in the isolated nth interconnect (in the absence of coupling) and ω is the signal frequency. The field $\mathcal{E}_y^{(n)}(y, z)$ satisfies the unperturbed wave equation so that

$$\left[\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right] \mathcal{E}_y^{(n)}(y, z) e^{i\omega t} = \mu\epsilon \frac{\partial^2}{\partial t^2} \mathcal{E}_y^{(n)}(y, z) e^{i\omega t} = -\omega_n^2 \mu\epsilon \mathcal{E}_y^{(n)}(y, z) e^{i\omega t} \quad (3)$$

where ω_n is the signal frequency in the nth interconnect.

Substituting (2) in (1), we get

$$\sum_n \left\{\frac{\partial^2 C_n}{\partial x^2} + C_n \left[\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right]\right\} \mathcal{E}_y^{(n)}(y, z) e^{i\omega t} = -\mu\epsilon\omega^2 \left[\sum_n C_n \mathcal{E}_y^{(n)}(y, z)\right] e^{i\omega t} + \mu \frac{\partial^2}{\partial t^2} [\dot{P}_{coup}(\vec{r}, t)]_y \left(\sum_n C_n \mathcal{E}_y^{(n)}(y, z)\right) e^{i\omega t} \quad (4)$$

where $[\dot{P}_{coup}(\vec{r}, t)]_y$ is an operator such that $[\dot{P}_{coup}(\vec{r}, t)]_y \mathcal{E}_y = [P_{coup}(\vec{r}, t)]_y$.

Using (3), we can replace the terms within the square brackets in the left-hand side of (4) to get

$$\sum_n \left[\frac{\partial^2 C_n}{\partial x^2} - \mu\epsilon\omega_n^2 C_n\right] \mathcal{E}_y^{(n)} = -\mu\epsilon\omega^2 \sum_n [C_n \mathcal{E}_y^{(n)}] + \mu \frac{\partial^2}{\partial t^2} [\dot{P}_{coup}]_y \left(\sum_n C_n \mathcal{E}_y^{(n)}\right). \quad (5)$$

Finally, multiplying the above equation by $\mathcal{E}_y^{(m)*}$ (the asterisk denotes complex conjugate) and integrating over all space, we obtain the coupled set of equations for the coefficients C_n

$$\sum_n \left\{\frac{\partial^2 C_n}{\partial x^2} O_{mn} + \mu\epsilon(\omega^2 - \omega_n^2) C_n O_{mn} - \kappa_{m,n}^2 C_n\right\} = 0 \quad (m = 1, 2, 3, \cdots, n) \quad (6)$$

where $\kappa_{m,n}^2 = \int d^3r \mathcal{E}_y^{m*} \mu (\partial^2/\partial t^2) [\dot{P}_{coup}]_y \mathcal{E}_y^n$ and O_{mn} is the overlap between the fields in the mth and nth interconnect ($O_{mn} = \int d^3r \mathcal{E}_y^{m*} \mathcal{E}_y^n$).

Equation (6) is a set of n coupled second order differential equations. The difference between (6) and the equations of conventional coupled mode theory is that we have not assumed $O_{mn} = 0$ for $m \neq n$. In fact, this assumption would be incorrect in the limit of strong coupling where the overlap between the fields in neighboring interconnects can be quite significant and neglecting this overlap may result in violation of energy conservation [13]–[18].

Let us now make the following substitution

$$\hat{D}_n(x) = C_n(x) \exp\left(-i \int k_n \, dx\right) \quad (7)$$

where

$$(k_n)^2 = \mu\epsilon(\omega^2 - \omega_n^2). \quad (8)$$

Substituting (7) in (6) and using (8), we get

$$\sum_n \left[\frac{\partial^2 \hat{D}_n}{\partial x^2} O_{mn} + 2ik_n \frac{\partial \hat{D}_n}{\partial x} O_{mn} - \kappa_{mn}^2 \hat{D}_n\right] \cdot \exp\left(i \int k_n \, dx\right) = 0 \quad (m = 1, 2, 3, \cdots, n). \quad (9)$$

Equation (9) can be recast in a matrix form

$$[\mathbf{A}] \frac{\partial^2}{\partial x^2} [\hat{\mathbf{D}}] + 2i[\mathbf{B}] \frac{\partial}{\partial x} [\hat{\mathbf{D}}] = [\mathbf{K}] [\hat{\mathbf{D}}] \quad (10)$$

where $[\hat{\mathbf{D}}]$ is an $n \times 1$ matrix whose elements are the coefficients $\hat{D}_1, \hat{D}_2, \cdots, \hat{D}_n$, $[\mathbf{A}]$ is an $n \times n$ matrix whose elements are $A_{mn} = O_{mn} \exp(i \int k_n \, dx)$, $[\mathbf{B}]$ is an $n \times n$ matrix whose elements are $B_{mn} = k_n O_{mn} \exp(i \int k_n \, dx)$ and $[\mathbf{K}]$ is an $n \times n$ matrix whose elements are given by $K_{mn} = \kappa_{mn}^2 \exp(i \int k_n \, dx)$.

Equation 10 is the general coupled mode equations for a system of n optical interconnects. To illustrate the usefulness of these equations, we now proceed to solve them

for the case of just *two* interconnects. For this, we first note that ratio of the second to the first term in the LHS of (10) is of the order of the ratio of the distance scale over which significant coupling occurs to the wavelength of the signal in the interconnects. For electromagnetic coupling, the signal wavelength is the wavelength of the optical or electromagnetic signal which is between 1 and 100 μm typically. (We are only concerned with high speed interconnects, i.e., optical interconnects or those that carry ultrashort pulses or millimeter waves.) For quantum mechanical coupling the wavelength is the DeBroglie wavelength of electrons which is between 10 and 100 Å. Since even for the most densely packed interconnects, we do not expect significant coupling to occur over such small scales, we can always neglect the first term in the LHS of (10) in comparison with the second term. This allows us to obtain closed analytical solutions of (10) in the case of two interconnects.

If we neglect the first term in the LHS (10), we get

$$\frac{\partial \dot{D}_1}{\partial x} = -i\Delta_1 \dot{D}_1 - i\Omega_{12}\dot{D}_2 \exp\left\{-i\int [k_1 - k_2]\,dx\right\}$$

$$\frac{\partial \dot{D}_2}{\partial x} = -i\Delta_2 \dot{D}_2 - i\Omega_{21}\dot{D}_1 \exp\left\{-i\int [k_2 - k_1]\,dx\right\} \tag{11}$$

where

$$\Delta_1 = \frac{1}{2}\left\{\frac{\kappa_{11}^2 O_{22}}{k_1(O_{11}O_{22} - |O_{12}|^2)} - \frac{\kappa_{21}^2 O_{12}}{k_2(O_{11}O_{22} - |O_{12}|^2)}\right\}$$

$$\Delta_2 = \frac{1}{2}\left\{\frac{\kappa_{22}^2 O_{11}}{k_2(O_{11}O_{22} - |O_{12}|^2)} - \frac{\kappa_{12}^2 O_{21}}{k_1(O_{11}O_{22} - |O_{12}|^2)}\right\}$$

$$\Omega_{21} = \frac{1}{2}\left\{\frac{\kappa_{21}^2 O_{11}}{k_2(O_{11}O_{22} - |O_{12}|^2)} - \frac{\kappa_{11}^2 O_{21}}{k_1(O_{11}O_{22} - |O_{12}|^2)}\right\}$$

$$\Omega_{12} = \frac{1}{2}\left\{\frac{\kappa_{12}^2 O_{22}}{k_1(O_{11}O_{22} - |O_{12}|^2)} - \frac{\kappa_{22}^2 O_{12}}{k_2(O_{11}O_{22} - |O_{12}|^2)}\right\} \tag{12}$$

We now make another transformation of variables

$$D_1 = D_1 \exp\left(-i\int \Delta_1\,dx\right)$$

$$\dot{D}_2 = D_2 \exp\left(-i\int \Delta_2\,dx\right)$$

$$k_1 = k_1' + \Delta_1$$

$$k_2 = k_2' + \Delta_2. \tag{13}$$

This reduces (11) to

$$\frac{\partial D_1}{\partial x} = -i\Omega_{12}D_2 \exp\left\{-i\int [k_1' - k_2']\,dx\right\}$$

$$\frac{\partial D_2}{\partial x} = -i\Omega_{21}D_1 \exp\left\{-i\int [k_2' - k_1']\,dx\right\}. \tag{14}$$

The above equations can be decoupled to yield

$$\frac{\partial^2 D_1}{\partial x^2} + i[k_1' - k_2']\frac{\partial D_1}{\partial x} + \Omega_{12}\Omega_{21}D_1 = 0 \tag{15}$$

whose solution is

$$D_1(x) = e^{i\delta x}[Pe^{i\nu x} + Qe^{-i\nu x}] \tag{16}$$

where P and Q are constants, $2\delta = k_2' - k_1'$ and $\nu = \sqrt{\kappa^2 + \delta^2}$ where $\kappa = \Omega_{12} = \Omega_{21}$. The last equality follows from the fact that the two interconnects are assumed to be identical.

Assuming that the wave vectors are independent of position x, we get from (7), (13), and (15)

$$C_1(x) = e^{ik_0 x}[Pe^{i\nu x} + Qe^{-i\nu x}] \tag{17}$$

where $k_0 = k_2' + k_1'$.

Similarly, we get

$$C_2(x) = -e^{ik_0 x}[P'e^{i\nu x} + Q'e^{-i\nu x}] \tag{18}$$

where $P' = Pk^+/\kappa$ and $Q' = Qk^-/\kappa$ with $k^+ = 2\delta + 2\nu$ and $k^- = 2\delta - 2\nu$.

To evaluate the constants P and Q, we need to apply the boundary conditions. Let

$$C_1(x = 0) = P + Q = A_1^+$$

$$C_2(x = 0) = -\frac{k^+P + k^-Q}{\kappa} = A_2^+ \tag{19}$$

This gives

$$P = -\frac{k^-}{2\nu}A_1^+ - \frac{\kappa}{2\nu}A_2^+$$

$$Q = \frac{k^+}{2\nu}A_1^+ + \frac{\kappa}{2\nu}A_2^+ \tag{20}$$

Now let

$$C_1(x = L) = B_1^+$$

$$C_2(x = L) = B_2^+ \tag{21}$$

Using (16)–(20), we get

$$B_1^+ = e^{ik_0 L}\left\{\left(\cos(\nu L) + \frac{i\delta}{\nu}\sin(\nu L)\right)A_1^+ + \left(-\frac{i\kappa}{\nu}\sin(\nu L)\right)A_2^+\right\}$$

$$B_2^+ = e^{ik_0 L}\left\{\left(\cos(\nu L) - \frac{i\delta}{\nu}\sin(\nu L)\right)A_2^+ + \left(-\frac{i\kappa}{\nu}\sin(\nu L)\right)A_1^+\right\} \tag{22}$$

This equation can be written in a matrix form as

$$\begin{pmatrix} B_1^+ \\ B_2^+ \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \cdot \begin{pmatrix} A_1^+ \\ A_2^+ \end{pmatrix} \tag{23}$$

where

$$a = e^{i\hat{\alpha}L}\left(\cos(\nu L) + \frac{i\hat{\delta}}{\nu}\sin(\nu L)\right)$$

$$b = c = e^{i\hat{\alpha}L}\left(-\frac{i\kappa}{\nu}\sin(\nu L)\right)$$

$$d = e^{i\hat{\alpha}L}\left(\cos(\nu L) - \frac{i\hat{\delta}}{\nu}\sin(\nu L)\right). \quad (24)$$

Note that the 2×2 matrix in (23) is in the form of a transmission matrix. The matrix element b ($= c$) is an indication of the coupling from one interconnect to another. It is the fraction of the signal in interconnect 1 that gets coupled to interconnect 2 after a distance L. Substituting back the values of κ and ν, we find that the quantity b is given by

$$|b| = |c| = \frac{\kappa}{\sqrt{\kappa^2 + \hat{\delta}^2}}\sin\left[\sqrt{\kappa^2 + \hat{\delta}^2}L\right]. \quad (25)$$

It is easy to show that this coupling is maximum if $\hat{\delta} = 0$ as long as $\tan(\kappa L) > \kappa L$ or as long as $\kappa L \leq \pi/2$. To have $\hat{\delta} = 0$ would require that the two interconnects be identical and carry the same signal frequency. Note that when $\hat{\delta}$ is zero it is possible for 100% of the signal in one interconnect to get coupled to the other and this happens at a distance

$$L_{100\%} = \frac{\pi}{2\kappa}. \quad (26)$$

In the case of $\hat{\delta} = 0$, the coupling over a distance L is simply given by

$$|b| = |c| = |\sin(\kappa L)| \quad (27)$$

Therefore, the fraction of the signal power from one interconnection that is coupled into another is given by

$$|b|^2 = |c|^2 = |\sin(\kappa L)|^2. \quad (28)$$

The most important step now is to derive an expression for the coupling constant κ. But before we proceed to do so, we show that the Schrödinger equation governing quantum mechanical coupling between interconnect lines is mathematically similar to the wave equation governing electromagnetic coupling between optical interconnects so that the Schrödinger equation yields similar solutions for the coupling parameter b ($= c$). In fact, the expression for b (for quantum mechanical tunneling) is identical to the expression given in (27).

B Quantum Mechanical Coupling

The equations governing quantum mechanical transport of electrons through an interconnect (which can be viewed as an electron waveguide) is the Schrödinger equation which reads

$$i\hbar\frac{\partial\psi}{\partial t} = (H_0 + H')\psi = \left(-\frac{\hbar^2}{2m}\nabla^2 + H'\right)\psi \quad (29)$$

where $\hbar$ is the reduced Planck's constant, m is the electron's effective mass, ψ ($= \psi(\bar{r}, t)$) is the electronic wave function, H_0 is the unperturbed Hamiltonian, and H' is the perturbation in the Hamiltonian arising from coupling.

We will assume that the perturbation is static (this neglects phonon-assisted or incoherent tunneling) and use time-independent perturbation theory. The wave function is written as

$$\psi(\bar{r}, t) = \psi(x, y, z, t) = [C_1(x)\Phi_1(y, z) + C_2(x)\Phi_2(y, z)]e^{-iEt/\hbar} \quad (30)$$

where $\Phi_{1,2}(y, z)e^{-iE_{1,2}t/\hbar}$ are the unperturbed wave functions in interconnects 1 and 2 in the absence of coupling. These wave functions are eigen functions of H_0 (which is Hermitean) and therefore orthonormal.

Since $\Phi_{1,2}(y, z)e^{-iE_{1,2}t/\hbar}$ are eigen functions of H_0, i.e., they satisfy the unperturbed Schrödinger equation, it is easy to see that

$$H_0\Phi_{1,2}(y, z) = E_{1,2}\Phi_{1,2}(y, z). \quad (31)$$

Substituting (30) in (29), we get

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2 C_1}{\partial x^2} + C_1\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\right)\Phi_1$$
$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2 C_2}{\partial x^2} + C_2\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\right)\Phi_2$$
$$= E[C_1\Phi_1 + C_2\Phi_2] - H'(C_1\Phi_1 + C_2\Phi_2) \quad (32)$$

where $\Phi_{1,2} \equiv \Phi_{1,2}(y, z)e^{-iEt/\hbar}$.

Comparing the above equation with (5), we find that they are mathematically similar. This is not surprising since both the electromagnetic wave equation and the Schrödinger equation describe the propagation of a wave through a medium. Because of this mathematical similarity we can write down the coupling coefficients $|b|$ and $|c|$ from analogy without any further derivation.

$$|b| = |c| = \sin(\zeta L) \quad (33)$$

where

$$\zeta = \frac{(m|H'_{12}|/\hbar^2)O_{11} - (m|H'_{11}|/\hbar^2)O_{21}}{\beta(O_{11}O_{22} - |O_{12}|^2)}$$
$$= \frac{(m|H'_{21}|/\hbar^2)O_{22} - (m|H'_{11}|/\hbar^2)O_{12}}{\beta(O_{11}O_{22} - |O_{12}|^2)} \quad (34)$$

while $H'_{m,n} = \int d^3r\Phi_m^* H'\Phi_n$ and

$$\beta = \sqrt{2m(E - E_1)}/\hbar = \sqrt{2m(E - E_2)}/\hbar \quad (35)$$

where E_1 and E_2 are the unperturbed energy levels in interconnect 1 and 2.

III. Calculation of Coupling Coefficients

A. Optical Coupling Coefficient

To calculate the optical coupling coefficient κ, we consider two identical parallel or crossing interconnects which we view as optical waveguides. The configuration is shown in Fig. 1(a) and (b). We concentrate on TE_{10} wave propagation and one particular transverse mode. Fig. 1(c) shows the refractive index profile $\eta(z)$ along the z direction where $\eta(z) = \sqrt{\mu_r(z)\epsilon_r(z)}$ and the subscript r denotes

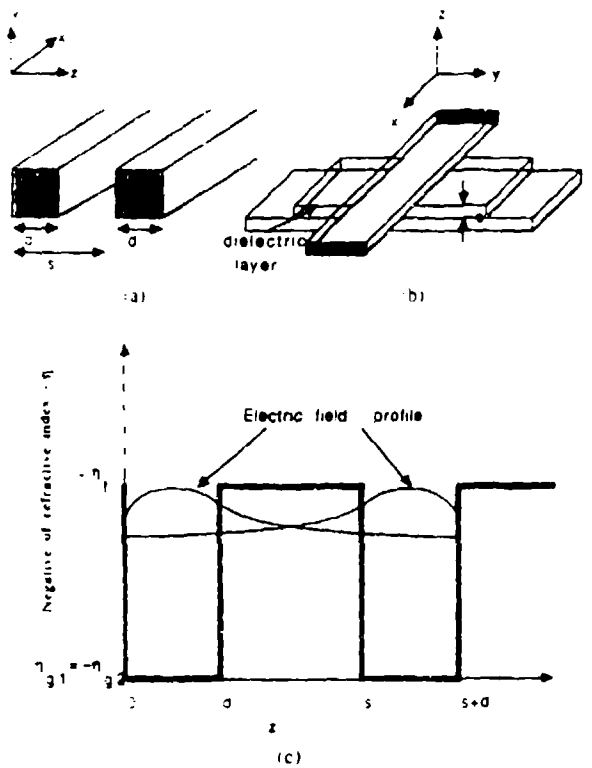


Fig. 1. (a) Two parallel interconnects of width d and separation s. (b) two crossing interconnects, and (c) the refractive index profile along with the electric field distribution in two neighboring optical interconnects that act as waveguides.

relative permeability or permittivity. The refractive indexes in the two interconnects are η_{g1} and η_{g2}, respectively, while the refractive index of the intervening medium is η_f.

We now calculate the quantities κ_{11} ($= \kappa_{22}$) and κ_{12} ($= \kappa_{21}$) which appear in the expressions for $\Omega_{12} = \Omega_{21} = \kappa$.

$$\kappa_{11}^2 = \omega^2 \mu_0 \epsilon_0 \int dz \mathcal{E}_v^{(1)}(z)(\eta_f - \eta_{g1}) \mathcal{E}_v^{(1)}(z)$$

$$\kappa_{12}^2 = \omega^2 \mu_0 \epsilon_0 \int dz \mathcal{E}_v^{(1)}(z)(\eta_f - \eta_{g1}) \mathcal{E}_v^{(2)}(z) \quad (36)$$

where μ_0 and ϵ_0 are the free space permeabilities and permittivities.

We will assume that the two interconnects are identical so that $\eta_{g1} = \eta_{g2} = \eta_g$. The electric field amplitudes are given by [see Fig. 1(c)] [12]

$$\mathcal{E}_v^{(1)}(z) = \begin{cases} E_f e^{\alpha_f z} & z \le 0 \\ E_g \cos(\gamma_g z + \phi) & 0 \le z \le d \\ E_f \exp[-\alpha_f (z - d)] & z \ge d \end{cases} \quad (37)$$

$$\mathcal{E}_v^{(2)}(z) = \begin{cases} E_f \exp[\alpha_f (z - s)] & z \le s \\ E_g \cos(\gamma_g (z - s) + \phi) & s \le z \le s + d \\ E_f \exp[-\alpha_f (z - s - d)] & z \ge s + d \end{cases} \quad (38)$$

where

$$\alpha_f^2 = k^2 - \eta_f^2 \omega^2 \mu_0 \epsilon_0$$

$$\gamma_g^2 = \eta_g^2 \omega^2 \mu_0 \epsilon_0 - k^2$$

$$E_g^2 (\eta_g^2 - N^2) = E_f^2 (\eta_g^2 - \eta_f^2)$$

$$\phi = -\tan^{-1}\left(\frac{\alpha_f}{\gamma_g}\right)$$

$$N = k/\omega \sqrt{\mu_0 \epsilon_0} \quad (39)$$

and k is defined in (8).

The quantities O_{11}, O_{22}, O_{12}, κ_{11}, and κ_{12} can now be computed easily.

$$O_{11} = O_{22} = \frac{E_g^2}{4\gamma_g} \left[4 \frac{\eta_g^2 - N^2}{\eta_g^2 - \eta_f^2} \cot\phi + 2\gamma_g d + \sin(2\gamma_g d + 2\phi) - \sin 2\phi \right] \quad (40)$$

$$O_{12} = O_{21} = e^{-\alpha_f s} E_g^2 \left[\frac{\eta_g^2 - N^2}{\eta_g^2 - \eta_f^2} \left\{ \frac{1}{\alpha_f} + e^{\alpha_f d}(s - d) \right\} \right] + e^{-\alpha_f s} E_g^2 \left[\sqrt{\frac{\eta_g^2 - N^2}{\eta_g^2 - \eta_f^2}} \left\{ \frac{\alpha_f (e^{\alpha_f d} \cos(\gamma_g d + \phi) - \cos\phi)}{\alpha_f^2 + \gamma_g^2} \right\} \right] + e^{-\alpha_f s} E_g^2 \left[\sqrt{\frac{\eta_g^2 - N^2}{\eta_g^2 - \eta_f^2}} \left\{ \frac{\gamma_g (e^{\alpha_f d} \sin(\gamma_g d + \phi) - \sin\phi)}{\alpha_f^2 + \gamma_g^2} \right\} \right] + e^{-\alpha_f s} E_g^2 \left[\sqrt{\frac{\eta_g^2 - N^2}{\eta_g^2 - \eta_f^2}} \left\{ \frac{\alpha_f (e^{\alpha_f d} \cos\phi - \cos(\gamma_g d + \phi))}{\alpha_f^2 + \gamma_g^2} \right\} \right] + e^{-\alpha_f s} E_g^2 \left[\sqrt{\frac{\eta_g^2 - N^2}{\eta_g^2 - \eta_f^2}} \left\{ \frac{\gamma_g (\sin(\gamma_g d + \phi) - e^{\alpha_f d} \sin\phi)}{\alpha_f^2 + \gamma_g^2} \right\} \right] \quad (41)$$

$$\kappa_{12}^2 = e^{-2\alpha s}\omega^2\mu_0\epsilon_0\left(\frac{\eta_e^2 - N^2}{\sqrt{\eta_e^2 - \eta_i^2}}\right)^2 E_e^2 \left[\frac{\sin(\gamma_e d)}{\sqrt{\alpha_e^2 + \gamma_e^2}}\right]^2 \quad (42)$$

$$\kappa_{11}^2 = e^{-2\alpha s}\omega^2\mu_0\epsilon_0\left(\frac{\eta_e^2 - N^2}{\eta_e + \eta_e}\right)^2 \frac{E_e^2}{2\alpha_e}\left[e^{2\alpha_e d} - 1\right]. \quad (43)$$

One can now use the results of (40)–(43) in (12) to obtain κ and then use it in (27) to find the coupling parameter b for electromagnetic coupling.

B. Quantum Mechanical Coupling Coefficient

To estimate the coupling coefficient ζ associated with quantum mechanical tunneling between two interconnects, we first refer to Fig. 2 which shows two parallel interconnects [Fig. 2(a)] or two crossing interconnects [Fig. 2(b)] as a system of two coupled quantum wires. Fig. 2(c) plots the potential energy profile $V(z)$ that an electron sees along the z direction. The electron's kinetic energy of motion in the z direction is quantized into subband states that are labeled by n (E_n is the energy of the nth subband state as shown in Fig. 2). The height of the potential barrier is $V - V_1$ for the first interconnect and $V - V_2$ for the second interconnect. The distance between the interconnects is s and the width of each interconnect (dimension along z direction) is d.

To calculate ζ, we first find the quantities H'_{12}, O_{12}, O_{11}, and O_{22} which appear in the expression for ζ [see (34)]. Following [15] we calculate these quantities as follows

$$H'_{12} = \left|\int_{-\infty}^{+\infty} \psi_1^*[V - V_2]\psi_2\, dz\right|. \quad (44)$$

The wave functions $\psi_1(z)$ and $\psi_2(z)$ are given by [18]

$$\psi_1(z) = \begin{cases} e^{\alpha z}\cos\theta & z \le 0 \\ \cos(\gamma z + \theta) & 0 \le z \le d \\ \exp[-\alpha(z-d)]\cos(\gamma d + \theta) & z \ge d \end{cases} \quad (45)$$

$$\psi_2(z) = \begin{cases} e^{\alpha(z-s)}\cos\theta & z \le s \\ \cos(\gamma(z-s) + \theta) & s \le z \le s + d \\ \exp[-\alpha(z-s-d)]\cos(\gamma d + \theta) & z \ge s + d \end{cases} \quad (46)$$

where

$$\alpha = \sqrt{2m/\hbar^2(V - E_1')}$$

$$\gamma = \sqrt{2m/\hbar^2(E_1' - V_1)}$$

$$\theta = -\tan^{-1}(\alpha/\gamma)$$

$$C = \frac{2}{\sqrt{d + 2/\alpha}}. \quad (47)$$

The energy E_n is the energy of an electron in the nth bound state in a quantum well and is found from the eigen equation

$$\gamma d = 2\tan^{-1}(\alpha/\gamma) + n\pi. \quad (48)$$

The quantities O_{11} ($= O_{22}$), O_{12} ($= O_{21}$), H'_{11} and H'_{12} can now be calculated easily

$$O_{11} = O_{22} = 1$$

$$O_{12} = O_{21} = \frac{2e^{-\alpha s}}{d + 2/\alpha}\left[\frac{\cos^2\theta + \cos^2(\gamma d + \theta)}{2\alpha} + e^{\alpha d}\cos(\gamma d + \theta)\cos\theta(s - d)\right] + \frac{2e^{-\alpha s}}{d + 2/\alpha}\cdot\cos\theta\left[\frac{\alpha(e^{\alpha d}\cos(\gamma d + \theta) - \cos\theta)}{\alpha^2 + \gamma^2}\right] + \frac{2e^{-\alpha s}}{d}\frac{2/\alpha}{}\cos\theta\left[\frac{\gamma(e^{\alpha d}\sin(\gamma d + \theta) - \sin\theta}{\alpha^2 + \gamma^2}\right] + \frac{2e^{-\alpha s}}{d + 2/\alpha}\cos(\gamma d + \theta)\cdot\left[\frac{\alpha(e^{\alpha d}\cos\theta - \cos(\gamma d + \theta))}{\alpha^2 - \gamma^2}\right] + \frac{2e^{-\alpha s}}{d + 2/\alpha}\cos(\gamma d + \theta)\cdot\left[\frac{\gamma(\sin(\gamma d + \theta) - e^{\alpha d}\sin\theta)}{\alpha^2 + \gamma^2}\right] \quad (50)$$

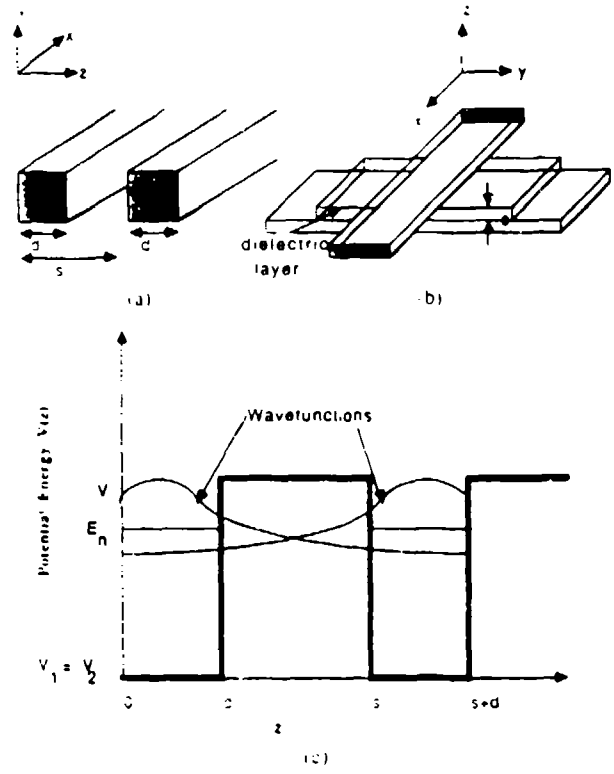


Fig. 2. (a) Two parallel interconnects of width d and separation s, (b) two crossing interconnects, and (c) the potential energy profile along with the wavefunctions in two neighboring narrow interconnects that act as quantum wires.

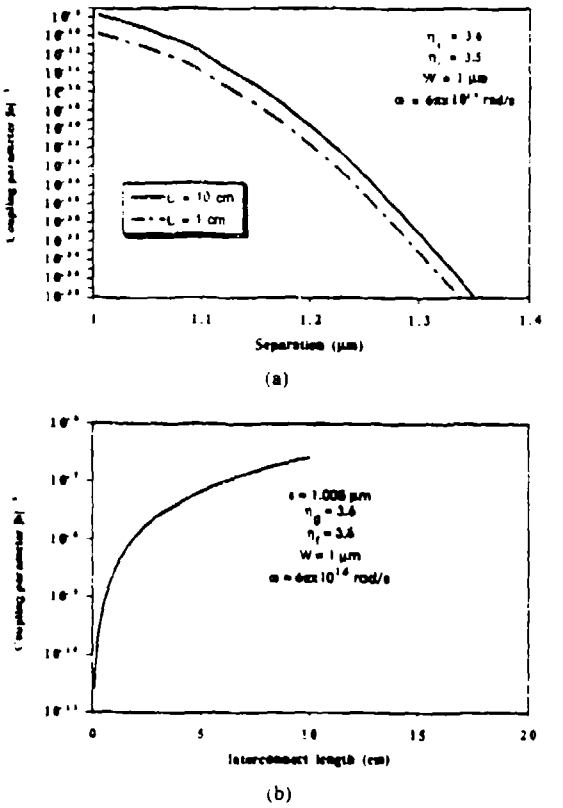


Fig. 3. (a) The coupling parameter $|b|^2$ (which is the fraction of the power in one interconnect coupled to the other) for electromagnetic coupling versus the separation between the interconnects. The interconnects are 1 μm wide and the refractive indexes of the interconnect material and the surroundings are 3.6 and 3.5, respectively. The results are plotted for two different lengths of the interconnects, namely 1 and 10 cm. The angular frequency is $6\pi \times 10^{14}$ rad/s. (b) The coupling parameter $|b|^2$ versus the length of the interconnect for a spacing of 1.005 μm. All other parameters are the same as in Fig. 3(a).

$$|H'_{1,2}| = e^{-\alpha s} \frac{2(V - V_2)}{d + 2/\alpha} \frac{\cos(\gamma d + \theta) \sin(\gamma d)}{\sqrt{\gamma^2 + \alpha^2}} \tag{51}$$

$$|H'_{1,1}| = e^{-2\alpha s} \frac{2(V - V_2)}{d + 2/\alpha} \frac{\cos^2(\gamma d + \theta)}{2\alpha} [e^{2\alpha d} - 1]. \tag{52}$$

We can use the results of (49)–(52) in (33)–(35) to obtain the coupling parameter b in the case of quantum mechanical coupling.

IV. Results

In Fig. 3(a) we show the coupling parameter $|b|^2$ due to optical coupling as a function of the spacing s between two identical optical interconnects for two different lengths of the interconnects. The coupling parameter $|b|^2$, as mentioned before, is the fraction of the power in one interconnect that is coupled to the other. The interconnects are assumed to be 1 μm wide (this is of the order of the wavelength of light emitted by semiconductor lasers and light emitting diodes), the material is GaAs with a refractive index of 3.6 and the isolating medium is AlGaAs with a refractive index of 3.5. Fig. 3(b) shows the coupling parameter due to electromagnetic coupling as a function of interconnect length when the spacing between the optical interconnects is 1.005 μm. The signal frequency is assumed to be $2x10^{14}$ Hz which roughly corresponds to the signal frequency of a GaAs diode laser. Other parameters regarding the interconnects are shown in figure legends. In Fig. 4(a) we show the coupling parameter due to quantum mechanical coupling as a function of spacing between two identical silicide interconnects embedded in silicon dioxide. We assume that the potential barrier between silicide and silicon dioxide is 3 eV which is close to the potential barrier between silicon and silicon dioxide. The interconnects are assumed to be 50 Å wide (the limit of present-day lithographic capability). The data is presented for two different lengths L of the interconnects (L = 1 cm and L = 10 cm). In Fig. 4(b), we show the coupling parameter as a function of the length of the interconnect for two different spacings of 100 and 150 Å. All other relevant parameters about the interconnects are displaced in the legend.

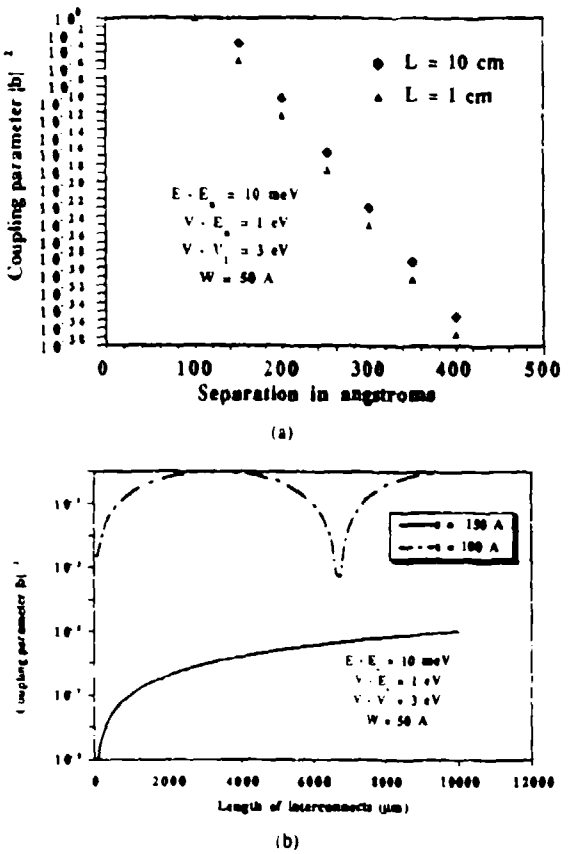


Fig. 4. (a) The coupling parameter $|b|^2$ for quantum mechanical coupling versus the separation between interconnects. The interconnects are assumed to be made of polysilicon embedded in solicon dioxide. The energy barrier between polysilicon and oxide is assumed to be 3 eV, the electron subband energy is 1 eV below the barrier and the longitudinal kinetic energy is 10 meV. The width W of the interconnects is 50 Å. The results are plotted for two different lengths of the interconnects, namely 1 and 10 cm. (b) The coupling parameter $|b|^2$ versus the length of the interconnect for two different spacings of 100 and 150 Å. All other parameters are the same as in Fig. 4(a).

V. Conclusion

It is evident from the results that coupling in optical interconnects is not a serious problem in ULSI. This is because GaAs-AlGaAs waveguides provide excellent confinement of the optical signal. However, quantum mechanical tunneling can be quite serious in silicide dielectrics embedded in silicon dioxide. As shown in Fig. 4(a), the coupling parameter $|b|^2$ for quantum mechanical coupling approaches 1 if the spacing is 100 Å. The coupling is always larger for a longer length of the interconnects since a larger region for coupling is provided. As shown in Fig. 4(b), the coupling is also larger when the spacing between interconnects is smaller. In Fig. 4(b), we see that the coupling oscillates with increasing length of the interconnect L when the spacing is small enough. This is easily understood from (33) which shows that $|b|^2 = \sin^2(\zeta L)$ which is an oscillatory function in L. When the spacing is 300 Å, the coupling and ζ are large enough that a period of the oscillation occurs when $L = 6300$ Å. For a spacing of 150 Å, the coupling and ζ are much smaller so that a much larger length L would be required for a full period of the oscillation. Therefore, we doe not see the oscillation for the spacing of 150 Å in Fig. 4(b).

Fig. 4 shows the quantum mechanical tunneling between silicide interconnects can be a serious problem in ULSI. Tunneling, of course, can be a coherent process or an incoherent process (phonon-assisted tunneling, for instance, is incoherent) and our analysis has dealt with coherent tunneling only. In our analysis, we assumed that the electron wavefunction is coherent over the entire width of the interconnect which is a good assumption for very narrow polysilicon interconnects (< 50 Å wide) even at room temperature and certainly below. Consequently, tunneling may cause significant crosstalk in ULSI circuits at room temperature and at cryogenic temperatures. Since many ultrafast electronic devices such as high electron mobility transistors (HEMT's) or quantum coupled devices utilizing resonant tunneling transistors may be constrained to operate at liquid nitrogen temperatures (77 K), quantum mechanical tunneling between interconnects could pose a serious problem at these temperatures. To circumevent this problem, one may devise ways of destroying the coherence of the electron wave function. One possibility is to impregnate the intervening dielectric isolating the interconnects with a soft magnetic material with strong spin-orbit scattering. Since spin orbit scattering is very efficient in destroying coherence, this may inhibit significant tunneling between neighboring interconnects and reduce crosstalk. The other obvious way of countering this problem is to use as isolating dielectrics those insulators that present a large energy barrier to tunneling. In this respect, silicon dioxide is better than silicon nitride since the energy barrier to tunneling is usually larger with silicon dioxide which has an energy gap of ~9 eV compared to ~5 eV for silicon nitride. The small energy gap of silicon nitride may preclude its use as isolating dielectric in ULSI in spite of its other attractive properties such as high resistance to Na^+ diffusion.

In conclusion, we have presented a model to calculate crosstalk between optical interconnects using waveguides and between narrow and closely spaced silicide interconnects embedded in dielectrics. We find that the crosstalk between silicide interconnects arising from quantum mechanical tunneling could be serious in ULSI circuits. We have also suggested some remedies for that problem.

References

[1] H. G. Craighead, "Nanolithography for ultrasmall structure fabrication," in *Physics of Quantum Electron Devices*, F. Capasso, Ed. New York: Springer-Verlag, 1989.

[2] H. I. Smith, K. Ismail, W. Chu, A. Yen, Y. C. Ku, M. L. Schattenburg, and D. A. Antoniadis, "Fabrication of quantum-effect electronic devices using x-ray nanolithography," in *Nanostructure Physics and Fabrication*, ed. M. A. Reed and W. P. Kirk, Eds. Boston, MA: Academic, 1989.

[3] E. E. Ehrichs, and A. L. deLozanne, "Fabrication of nanometer features with a scanning tunneling microscope," in *Nanostructure Physics and Fabrication*, M. A. Reed and W. P. Kirk, Eds. Boston, MA: Academic, 1989.

[4] S. P. Beaumont, "Fabrication and overgrowth of quantum wires and dots for optoelectronic applications," in *Nanostructure Physics and Fabrication*, M. A. Reed and W. P. Kirk, Eds. Boston, MA: Academic, 1989.

[5] ——, "Nanofabrication and optical assessment of quantum wires and dots," in *Granular Nanoelectronics*, D. K. Ferry, J. R. Barker, and C. Jacobini, Ed., NATO ASI Series. New York: Plenum, 1991.

[6] D. K. Ferry, "Lateral surface superlattices and the future of ULSI microelectronics," in *Granular Nanoelectronics*, D. K. Ferry, J. R. Barker, and C. Jacobini, Eds. NATO ASI Series. New York: Plenum, 1991.

[7] J. M. Mikkelson, L. A. Hall, A. K. Malhotra, S. D. Seccombe, and M. S. Wilson, *IEEE J. Solid State Circ.*, vol. 16, p. 542, 1981.

[8] S. I. Long and S. E. Butner, *GaAs Digital Integrated Circuit Design*. New York: McGraw Hill, 1990.

[9] T. A. Fulton and G. J. Dolan, "Observation of single electron charging effects in small tunnel junctions," *Phys. Rev. Lett.*, vol. 59, p. 109, 1987.

[10] A. Yariv, "Coupled mode theory for guided wave optics," *IEEE J. Quantum Electron.*, vol. QE-9, pp. 919-933, 1973.

[11] A. Ishimaru, *Electromagnetic Wave Propagation, Radiation and Scattering*. Englewood Cliffs, NJ: Prentice Hall, 1991.

[12] T. Tamir, *Ed., Integrated Optics*. New York: Springer Verlag, 1975.

[13] A. Hardy and W. Streifer, "Coupled mode theory of parallel waveguides," *J. Lightwave Technol.*, vol. LT-3, p. 1135, 1985.

[14] ——, "Coupled mode solutions of multiwaveguide systems," *IEEE J. Quantum Electron.*, vol. QE-22, p. 528, 1986.

[15] L. Tsang and S. L. Chuang, "Improved coupled mode theory for reciprocal anisotropic waveguides," *J. Lightwave Technol.*, vol. LT-6, p. 304, 1988.

[16] S. L. Chuang, "A coupled mode formulation by reciprocity and a variational principle," *J. Lightwave Technol.*, vol. LT-5, p. 5, 1987.

[17] ——, "A coupled mode theory for multiwaveguide systems satisfying reciprocity," *J. Lightwave Technol.*, vol. LT-5, p. 174, 1987.

[18] S. L. Chuang and B. Do, "Electron states in two coupled quantum wells—A strong coupling of modes approach," *J. Appl. Phys.*, vol. 62, p. 1290, 1987.

Supriyo Bandyopadhyay was born in Calcutta, India. He received the B.Tech. in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, the M.S. degree in electrical engineering from Southern Illinois University, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN.

From 1986 to 1987 he was a Visiting Assistant Professor at Purdue University. He joined the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, in 1987 where he is now an Associate Professor. His research interests are in quantum and hot electron transport and cryogenic measurements.

FEATURES OF QUANTUM MAGNETOTRANSPORT AND ELECTROMIGRATION IN MESOSCOPIC SYSTEMS

S. Bandyopadhyay, S. Chaudhuri and B. Das
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

M. Cahay
Department of Electrical and Computer Engineering
University of Cincinnati
Cincinnati, Ohio 45221

(Received 4 August 1992)

We have calculated the two-dimensional spatial distributions of several transport variables such as the current density, carrier concentration, chemical potential, space-charge electric field, residual resistivity dipole potential and electromigration forces in disordered mesoscopic structures subjected to arbitrary magnetic fields. These fully quantum mechanical calculations shed new light on such magnetotransport phenomena as the integer quantum Hall effect, the formation of magnetic bound states, magnetic response of current vortices caused by quantum interference between scatterers, the nature of residual resistivity dipoles in phase coherent transport, the difference between majority- and minority-carrier mobilities in the quantum mechanical regime, the electron wind force and direct force of electromigration in a mesoscopic system, and a variety of other phenomena.

Introduction

The spatial distribution of transport variables (such as the current density, carrier concentration, chemical potential, residual resistivity dipole field, etc.) are important in understanding many quantum transport phenomena[1] in mesoscopic structures and also electromigration effects[2,3]. In this paper, we present a microscopic calculation of such spatial distributions in a quasi one dimensional structure containing localized impurities and subjected to arbitrary magnetic fields. The calculations are fully quantum-mechanical and are based on the Schrödinger equation describing the system. Our results elucidate the nature of many different types of quantum magnetotransport phenomena (e.g. the integer quantum Hall effect, current vortex formation, etc.), and they also shed light on electromigration forces, particularly the wind force and the direct force on an ion that cause electromigration in solids. Electromigration is an especially important phenomenon in narrow one-dimensional mesoscopic samples, not only in its extreme effect of causing catastrophic device failure as it does in integrated circuits, but also in that it causes impurity motion which radically alters sample characteristics and leads to 1/f noise.

This paper is organized as follows. In the next section, we describe the theory for calculating spatial distributions of transport variables in mesoscopic structures. We then present results dealing with the two-dimensional spatial distribution of the current density in a disordered structure at various magnetic fields. The current distributions show a number of dramatic effects such as *current crowding* near an attractive impurity, *current detour* around a repulsive impurity, the formation of edge states with perfect transmittivity at the onset of the *integer quantum Hall effect*, circulating current patterns around an impurity characteristic of the formation of *magnetic bound states*, formation of current *vortices* (in the absence of any magnetic field) as a result of quantum interference between waves reflected from various impurities in a sample, the quenching of such vortices by a magnetic field, and so on. These results and others are accompanied by the corresponding modulations in the carrier concentration profile caused by current flow. From these profiles, one can clearly see the building up of an excess electron concentration around an attractive scatterer and a deficit around a repulsive one owing to electrostatic interaction of the electrons with the scatterer. Also, electrons accumulate around an impurity (irrespective of whether it is attractive or repulsive) when

0749-6036/92/050123 + 10 $08.00/0

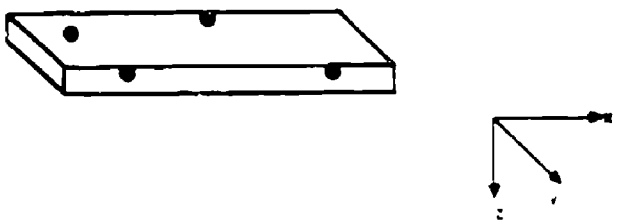
a magnetic bound state forms. These accumulations and depletions lead to inhomogeneities in the charge distribution which, in turn, lead to inhomogeneities in the electric field within the sample. Such inhomogeneities have a profound effect on the driving forces of electromigration acting on an impurity located inside the sample.

In addition to the spatial distributions of the current density and carrier concentration modulation, we also show the spatial distribution of the chemical potential inside a sample. In the absence of any magnetic field, the chemical potential profiles show very different features depending on whether the scatterers are attractive or repulsive. When a magnetic field is applied, the chemical potential profiles for both attractive and repulsive scatterers change dramatically. At high enough magnetic fields, the edges of the structure (along the direction of current flow) become strikingly smooth equipotential surfaces. This happens even when the potential at the center of the structure is exhibiting chaotic behavior. As a result, the longitudinal resistance measured by voltage probes attached to the edges of the structure will always read precisely zero (in a four-probe measurement). This is a clear demonstration of the quantum Hall effect. Also, we find that the chemical potential difference between the two edges of the structure (i.e. the Hall voltage drop) is exactly equal to the chemical potential drop between the two contacts of the structure when a single spin degenerate subband is occupied. Since the quantum mechanical transmission probability through the structure is exactly unity when this happens, it means that the Hall resistance is precisely quantized to $h/2e^2$ which is an unequivocal demonstration of the integer quantum Hall effect. It should be noted that these results pertaining to the integer quantum Hall effect are found from a rigorous microscopic calculation and *do not invoke any phenomenological construct*.

From the electrostatic potential profile, we have calculated the electric field distribution in the structure. The field shows significant inhomogeneities which have a profound influence on electromigration. Electromigration is caused by forces acting on charged impurities (ions) within the structure which tend to cause ion motion. Such motion can drastically alter the behavior of the structure in the phase coherent regime, lead to 1/f noise, and, in the extreme case, physically disrupt the structure leading to total device failure. The two main forces of electromigration are the "wind force" caused by collisions with electrons flowing past the ion and the "direct force" due to the electric field acting on the ion. We have calculated the spatial distribution of the so-called residual resistivity dipole potential which has a profound effect on the "wind force". In addition, the directions of both the "wind force" and the "direct force" at various ion locations and at varying magnetic field have also been calculated. To our knowledge, these are the first calculations of the effects of a magnetic field on electromigration forces in mesoscopic samples in the quantum coherence regime.

Theory

We consider a quasi one-dimensional structure as shown in Figure 1. The thickness is small enough that only a single



• Fig. 1: A quasi one-dimensional structure containing a random distribution of elastic scatterers. The structure is thin enough in the z-direction to allow the occupation of only one transverse subband in that direction. The magnetic field is in the z-direction.

transverse subband can be occupied in the z-direction (i.e. the confinement is complete in the z-direction). However, multiple transverse subbands may be occupied in the y-direction.

The Schrodinger equation describing this system under a z-directed magnetic field is

$$\frac{(\vec{p} + e\vec{A})^2}{2m^*}\psi + V\psi = E\psi , \qquad (1)$$

where $\vec{p}$ is the momentum operator, m^* is an isotropic effective mass, $\vec{A}$ is the magnetic vector potential, $\psi(x,y)$ is the two-dimensional wavefunction, and the potential energy V consists of two terms

$$V \equiv V(x,y) = V_1(y) + V_{impl}(x,y) . \qquad (2)$$

The first term gives the effect of the confining potential in the y-direction and the second term is due to impurities. We assume hardwall boundary conditions in the y-direction and delta potentials for the impurities so that we have

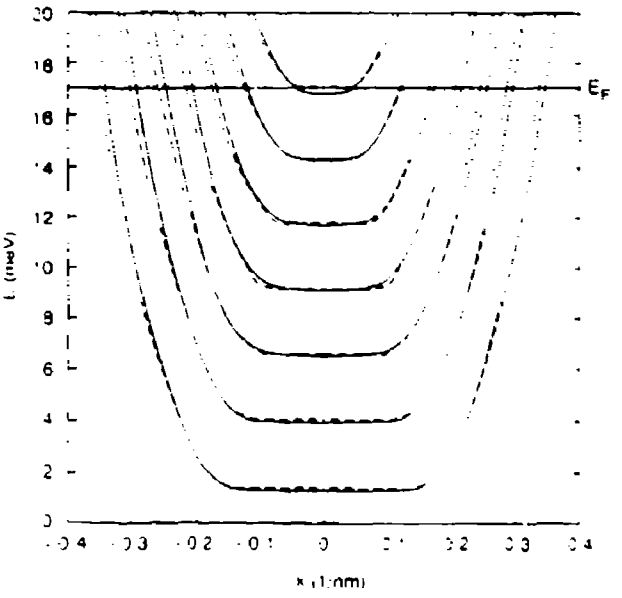
$$V_{impl}(x,y) = \sum_{i=1}^{s} \alpha_i \delta(x - x_i)\delta(y - y_i) , \qquad (3)$$

where s is the total number of impurities in the structure and the summation extends over all the impurities. The position of the ith impurity is denoted by (x_i, y_i). In all our calculations, we neglect effects associated with the spin of the electron.

The confining potential in the y-direction and the z-directed magnetic field give rise to a set of hybrid magnetoelectric subbands in the y-direction. Each such subband corresponds to a mode or a transport channel. Calling $\phi_m(y)$ the y-component of the wavefunction in the mth magnetoelectric subband, we can write the total wavefunction $\psi_{E_F,m}(x,y)$ for an electron injected from the left contact with the Fermi energy E_F into the mth magnetoelectric subband as

$$\psi_{E_F,m}(x,y) = \sum_{p=1}^{M} \frac{\left[A_{mp}(x)\exp(ik_p^F x)\phi_p(y)\right]}{\sqrt{|v_p^F|L}} + \sum_{p=1}^{M} \frac{\left[B_{mp}(x)\exp(ik_p^F x)\phi_{-p}(y)\right]}{\sqrt{|v_p^F|L}} , \qquad (4)$$

where L is the length of the structure, and k_n^F and v_n^F are the x-directed wavevector and x-directed group velocity correspond-



• Fig. 2: The energy-dispersion relations for the hybrid magnetoelectric subbands in a 2000 Å wide structure at a magnetic flux density of 1.5 Tesla. The solid curves are the result of an exact numerical calculation and the broken curves are those obtained from the application of Bohr-Sommerfeld quantization rule.

ing to the Fermi energy in the nth subband (see Fig. 2). These quantities are found from the dispersion relation of the hybrid magnetoelectric subbands. The method for finding the dispersion relations, as well as the wavefunctions $\varphi_n(y)$ in the nth magnetoelectric subband, has been described in reference 4 and will not be repeated here. Finally, the subscript $-p$ in Equation (4) indicates that the quantity corresponds to a wavefunction with an oppositely directed velocity as compared to the one with a subscript p.

In Equation (4), the summation over p (subband index) extends over both occupied and unoccupied subbands (i.e. both propagating and evanescent modes). It is important to include enough evanescent modes in the summation so that the computed wavefunction does not depend on the number of terms (M) in the summation. Even though the evanescent modes do not carry current, they have very significant effects on all transport variables, and especially their spatial distributions[1].

When the wavefunction is written in the form of Equation (4), (i.e. each mode indexed by $\pm p$ is normalized by the square root of the corresponding group velocity), all propagating modes at the Fermi energy carry the same current. This is consistent with the well-known cancellation of the velocity and density of states in quasi one-dimensional structures which causes every mode to carry the same current. To find the total wavefunction given by Equation (4), we now only have to evaluate the coefficients $A_{mp}(x)$ and $B_{mp}(x)$. These are found from a scattering matrix method that has been described in the first two citations in reference 1. Once the wavefunction has been evaluated, we can calculate all quantities of interest such as the spatial distribution of the current, carrier concentration profiles, residual resistivity dipole potential, chemical potential, electric field, and the forces of electromigration. The recipes for calculating these quantities from the wavefunction are described in the following paragraphs.

Spatial distribution of the current: In linear response transport and at low temperatures, only electrons with the Fermi energy carry a net current. Therefore, the two-dimensional wavefunction $\psi_m(x, y)$ of Fermi electrons in the mth magnetoelectric subband allows us to calculate the contribution of that subband to the two-dimensional current density $\vec{J}_m(x, y)$.

$$\begin{aligned}\vec{J}_m(x,y) &= \frac{ie\hbar}{2m^*}\left\{\psi_{E_F,m}(x,y)\nabla\psi^*_{E_F,m}(x,y)\right\} \\ &- \frac{ie\hbar}{2m^*}\left\{\psi^*_{E_F,m}(x,y)\nabla\psi_{E_F,m}(x,y)\right\} \\ &+ \frac{1}{m^*}\left\{e\vec{A}|\psi_{E_F,m}|^2\right\} \qquad (5)\end{aligned}$$

The total current density is then found by vectorially adding the contributions from all occupied subbands (propagating modes).

$$\vec{J}_{total}(x,y) = \sum_m^{M'} \vec{J}_m(x,y) . \qquad (6)$$

The above expression provides the current density at every coordinate point (x, y). Note that in calculating the total current density by summing over subbands, no weighting by the density of states is necessary in the summation since the wavefunctions have been normalized by the square root of velocity which takes care of the density of states factor. The above equation gives the two-dimensional spatial distribution of the current density (J_{total} as a function of coordinates (x, y)) for any arbitrary concentration and configuration of scatterers and for any magnetic field.

Carrier concentration change due to transport and its spatial distribution: To calculate the change in the carrier concentration caused by a current, we follow the method of Chu and Sorbello[4]. At low enough temperatures, the deviation of the semiclassical electron distribution function from the equilibrium value in the mth subband (as caused by a small applied electric field $\mathcal{E}$ driving the current) is given by

$$\Delta f_m(E) = -e\tau v(E,m)\mathcal{E}\delta(E - E_F) , \qquad (7)$$

where τ is a constant relaxation time describing a uniform background scattering due to all elastic collisions and $v(E, m)$ is the group velocity corresponding to electron energy E in the mth subband.

The change in the carrier concentration due to the current is then given by

$$\delta n_e(x,y) = \sum_m^{M'} \int dE N_{1d,m}(E)\Delta f_m(E)\left|\psi_{E,m}(x,y)\right|^2 , \qquad (8)$$

where $N_{1d,m}(E)$ is the one-dimensional density of states in the mth magnetoelectric subband (the prescription for calculating this was provided in reference 4), M' is the total number of propagating modes or occupied subbands, and $\psi_{E,m}$ is the wavefunction of Fermi electrons in the mth magnetoelectric subband. This wavefunction is not exactly the same as that given

by Equation (4); the only difference is that there should be no normalization of the various modes by the square-root of velocity in calculating $\Psi_{E,m}$. Therefore, if we use Equation (4) for the wavefunction, then $\delta n_s(x, y)$ should be be written as

$$\delta n_s(x, y) = -\frac{4e\tau\mathcal{E}}{\hbar}\sum_{m}^{M'} v_{E,m}|\psi_{E,m}(x, y)|^2 . \quad (9)$$

where v is now given by equation (4). The above equation gives the spatial distribution of the change in the carrier concentration as a consequence of transport.

Spatial distribution of the residual resistivity dipole potential: When an impurity is introduced in a solid, the resistance increases. This increase is associated with the formation of a "residual resistivity dipole" around the impurity which arises purely from the interaction of electrons with the impurity (scattering)[2]. Landauer[3] has pointed out that the residual resistivity dipole potential is an important source of electric field inhomogeneities in a sample and affects the so called "wind force" of electromigration. The wind force is the force that acts on an impurity owing to collisions with electrons during current transport.

To calculate the spatial distribution of the residual resistivity dipole potential, we follow Chu and Sorbello[2]. According to these authors, the electrostatic potential arising from δn_s and the induced screening charge which attempts to locally neutralize δn_s is the residual resistivity dipole potential. Within a Fermi-Thomas model, self-consistent screening gives this potential as

$$\delta U(x, y) = \frac{1}{e}\sum_{m}^{M'} N_{1D,m}^{-1}\delta n_{s,m}(x, y)$$
$$= -\pi\mathcal{E}\sum_{m}^{M'} v_m^2 E_{F,1D}|\psi_{E,m}(x, y)|^2 , \quad (10)$$

where $\delta n_{s,m}$ is the contribution to δn_s from the mth magnetoelectric subband.

From the above equation, we can readily obtain the spatial distribution of the residual resistivity dipole potential. The residual resistivity dipole *field* is the spatial gradient of this potential. In contrast to the claim of reference 3, this field does not have to be always directed along the direction of current flow, especially when quantum confinement (quantum size effect) and quantum interference (coherent effect) are important.

The "wind force" of electromigration. The wind force is the force that an ion or impurity in a solid experiences owing to an "electron wind" flowing past it during current transport. This electron wind imparts a force on the ion because the electrons collide with the ion and transfers momentum to the ion. The wind force causes electromigration (motion of ions or impurities) which has several effects in mesoscopic structures such as alteration of sample conductance, 1/f noise, and electrical failure caused by physical disruption of the sample at critical regions.

To calculate the wind force on an ion, we follow the work of Fiks and Huntington and Grone[4] who invoke Newton's law. The wind force is given by

$$\bar{F}_{wind} = \left(\frac{d\bar{p}}{dt}\right)_{coll} , \quad (11)$$

where the right hand side is the rate of loss of electron momentum due to collisions with the ion. It can be shown[5] from a rigorous semiclassical treatment based on the Boltzmann transport equation that the right hand side is $-\bar{J}_{total}/n\eta$ where n is the volume concentration and η is a semiclassical mobility associated with scattering by the ion. In the linear response regime, the above quantity is also equal to the negative of the spatial gradient of the chemical potential (quasi Fermi level). Therefore, the wind force on an ion located at coordinate (x_0, y_0) can be found directly from the chemical potential profile around the ion.

$$\bar{F}_{wind}(x_0, y_0) = -\bar{\nabla}\mu(x_0, y_0) , \quad (12)$$

where μ is the chemical potential

The above calculation of the wind force relies on the basic theory of reference 5 which does not distinguish between crystal momentum and real momentum in a solid. More sophisticated theories[6,7] require knowledge of the exact shape of the ion potential and also the exact spatial distribution of the charge pile up (or charge deficiency in the case of repulsive scatterers) around ions. These sophisticated theories could actually benefit from our quantum mechanical calculations of the change in the charge density around the ions caused by current flow (described previously in this section). A fully quantum mechanical, self-consistent and rigorous calculation of the wind force, starting from the Schrodinger-Poisson equations, is reserved for a future publication.

The "direct force" of electromigration: The direct force on an ion in a solid subjected to an electric field is the electrostatic force given by

$$\bar{F}_{direct} = Ze\bar{\mathcal{E}} , \quad (13)$$

where Z is the effective valency of the ion and $\bar{\mathcal{E}}$ is the local electric field at the ion site.

Bosvieux and Friedel[8] claimed that the difference in the charge of an interstitial ion from the background will be completely screened by the electrons so that the local electric field and hence the direct force on such an ion will be exactly zero. Gupta and co-workers[10] concluded in a similar vein that the direct force on a migrating ion in the saddle point position is exactly zero.

It was correctly pointed out by Das and Peierls[11] and also Landauer[12] that Bosvieux and Friedel could not be correct since their conclusion leads to inconsistencies. For instance, if an interstitial proton were completely screened and the local field is zero, then it should be also true of electrons and consequently a metal should have no electrical conductivity.

The correct equation to use for the direct force is Equation (13), but the electric field in that equation is not the space-averaged electric field as assumed by Huntington[13] and by Lou et. al.[14]. It was stressed by Landauer that the field is affected by the "charge density modulation" around an impurity that occurs during current flow. Therefore, the field must be calculated self-consistently by solving the Poisson and Schrödinger equation. In the linear response regime, this field can be obtained without much difficulty as described below.

spatial distribution of the chemical potential and electric field. Any net charge distribution $\rho(x,y)$ gives rise to an electrostatic potential $V(x,y)$ and an electric field $\mathcal{E}(x,y)$ which are related to the charge distribution through the Poisson equation

$$-\nabla^2 V(x,y) = \nabla \cdot \mathcal{E}(x,y) = \frac{\rho(x,y)}{\epsilon} \qquad (2)$$

It was shown by Büttiker[13] and Entin-Wohlman et al.[9] that in the linear response regime, and in situations where the spatial variation in the potential $V(x,y)$ is smooth on the scale of the screening length (i.e. the screening is very strong), the potential is given by

$$V(x,y) = \frac{\mu_1 - \mu_2}{2e} \frac{\sum_{n,m} T_{nm}(E_F) |\psi_{1n}(x,y)|^2 - \sum_{n,m} T_{nm}(E_F) |\psi_{2m}(x,y)|^2}{\sum_{n,m} T_{nm}(E_F)} \qquad (3)$$

where μ_1 and μ_2 are the chemical potentials (Fermi energy) in the two contacts of the structure under investigation. ψ_{1n} is the wavefunction of electrons injected from the left contact in subband n with energy ϵ_{1n} and ψ_{2m} is the wavefunction of electrons injected in mode m with energy ϵ_{2m} from the right contact. It was pointed out by Entin-Wohlman et al.[9] that $eV(x,y)$ is also what is actualy measured at a chemical potential probe so that it is also the chemical potential. In other words, in this regime, $eV \equiv \mu$.

Once the electrostatic (or chemical) potential is determined, the corresponding electric field is found simply from

$$\mathcal{E}(x,y) = -\nabla V(x,y) = -\nabla \mu(x,y)/e \qquad (4)$$

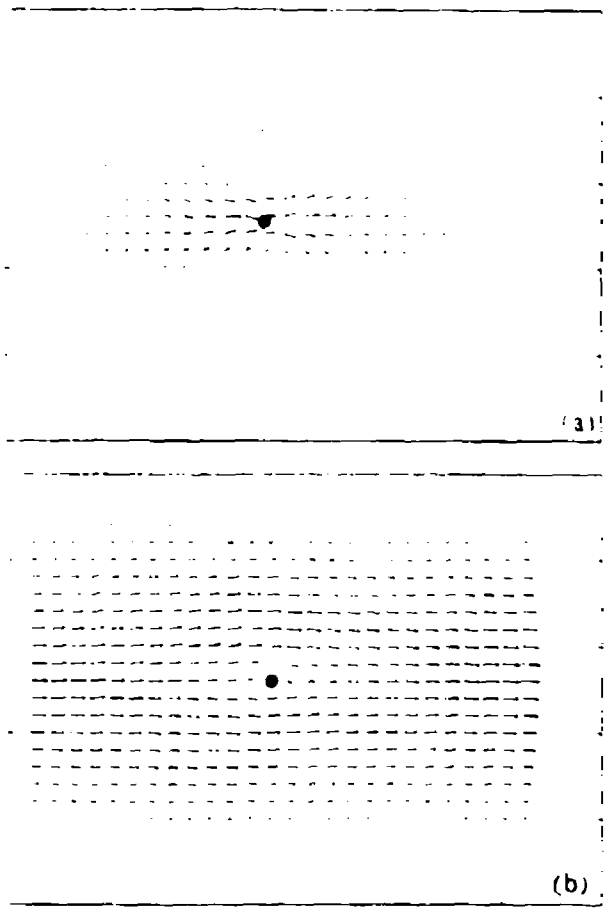
This is the electric field that should enter Equation (3) for the direct force in the linear response regime. Note that this means that the wind force and the direct force are equal in this regime if the valency Z of the ion is unity.

In the next section, we present results dealing with the spatial distribution of the current, carrier concentration change caused by the current, the chemical potential profile, the residual resistivity dipole potential, and the wind force and direct force on impurities in a prototypical GaAs quantum wire containing elastic delta-scatterers. We consider both attractive and repulsive scatterers and show some typical examples of how the spatial patterns respond to an external magnetic field.

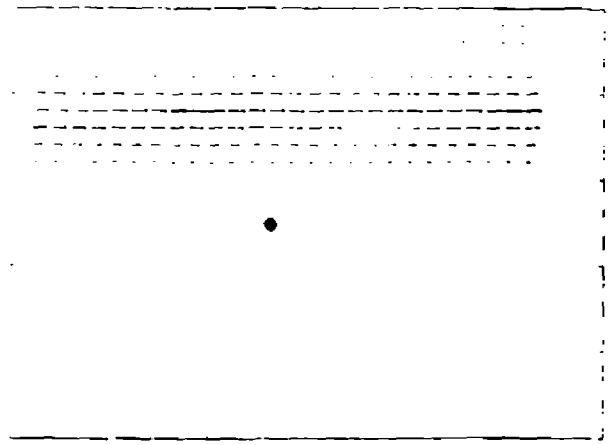
Results

Spatial distribution of the current around a single scatterer - local effects. We consider a GaAs wire which is 800 Å long and 1000 Å wide. The Fermi energy is 2.054 meV corresponding to an electron concentration of 3.5×10^{17}/cm^3. For this low concentration, only one subband is occupied in both transverse directions so that transport is strictly single-channeled.

In Figures 3(a) and 3(b), we show the current profiles around a single attractive and repulsive scatterer respectively in the absence of any magnetic field. In the case of an attractive scatterer, the current is drawn closer to the scatterer, whereas for a repulsive scatterer, the current detours around the scatterer. This



• Fig. 3. The spatial distribution of the current density $j_{total}(x,y)$ in a 800 Å long and 1000 Å wide structure in the absence of any magnetic field. The Fermi energy is 2.054 meV. (a). Current crowding around an attractive impurity. (b). current detour around a repulsive impurity.



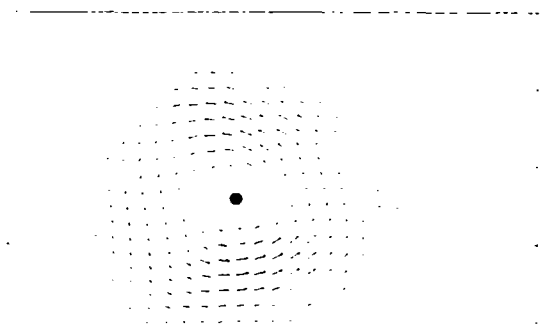
• Fig. 4. The spatial distribution of the current density $j_{total}(x,y)$ in the structure of Fig. 3(a) when a magnetic flux density of 2.18 Tesla is applied. The current flows along the top edge indicative of the formation of "edge states". This situation corresponds to the onset of the *integer quantum Hall effect*.

behavior is expected from electrostatic attraction and repulsion between electrons and the scatterer.

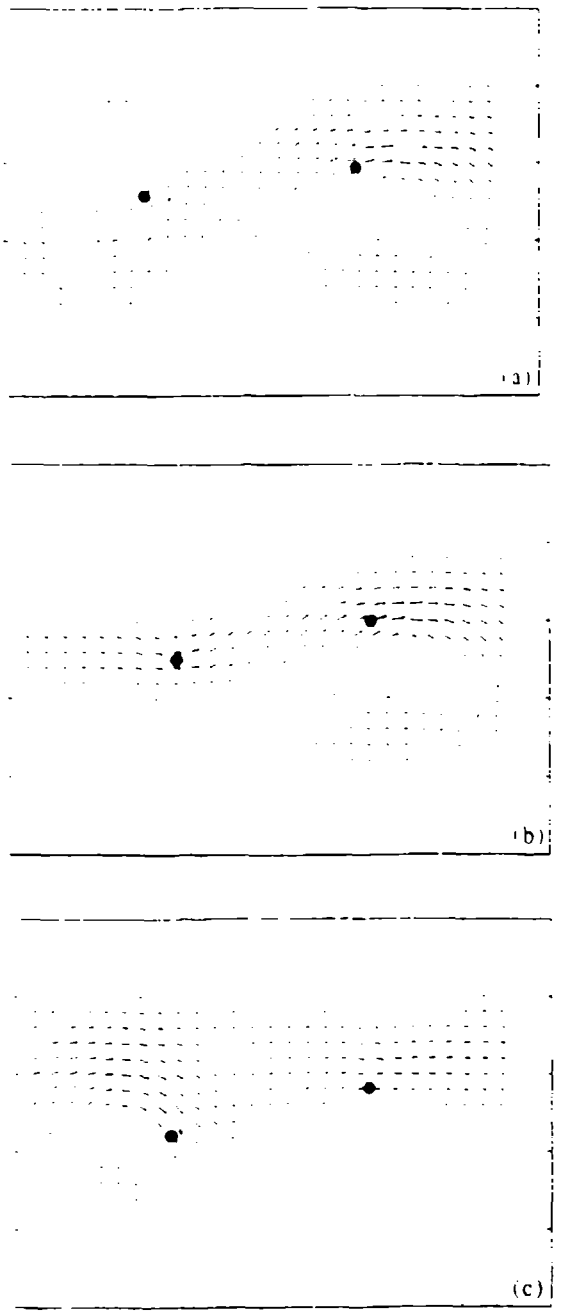
In Fig. 4, we show how the current pattern changes when a magnetic flux density of 2.18 Tesla is applied to the structure. The impurity is attractive. The current now flows entirely along one of the edges indicative of the formation of "edge states". There is practically no backscattering and the current flows straight through (the small reverse traveling component near the very top of the figure is not due to backscattering, it is due to skipping orbits and its origin was explained in reference 17). This current pattern is characteristic of the onset of the integer quantum Hall effect. The role of perfectly transmitting edge states in the quantization of Hall resistance was elucidated in reference 18.

In Fig. 5, we show the circulating current pattern around an attractive impurity when a magnetic bound state forms. The magnetic flux density is 3.5 Tesla. Note that the circulating current patterns hardly carry any net current in the x-direction. Therefore, the conductance of the structure will be very low when a magnetic bound state forms. It was shown in reference 1 that the conductance does indeed drop abruptly when a magnetic bound state forms.

Spatial distribution of the current around two scatterers - non-local quantum interference effects: In Fig. 6(a), we show the current pattern when two attractive impurities are present and no magnetic field is applied. Strong vortices form due to quantum interference between waves reflected from the two impurities. These vortices are purely a result of non-local quantum effects. The physics of these vortices were dealt with at length in the first two citations in reference 1. These vortices are interesting since they give rise to localized magnetic moments. In Fig. 6(b), we show how one of the vortices is destroyed at a flux density of 0.08 Tesla while the other remains. At a flux density of 1 Tesla, (Fig. 6(c)), the first vortex reap-



• Fig. 5. The spatial distribution of the current in the structure of Fig. 3(a) when the magnetic flux density is increased to 3.5 Tesla. At this field a magnetic bound state forms around the impurity which causes the circulating current pattern. The net transmission of current through the structure is very low so that the conductance of the structure drops abruptly when a magnetic bound state forms.



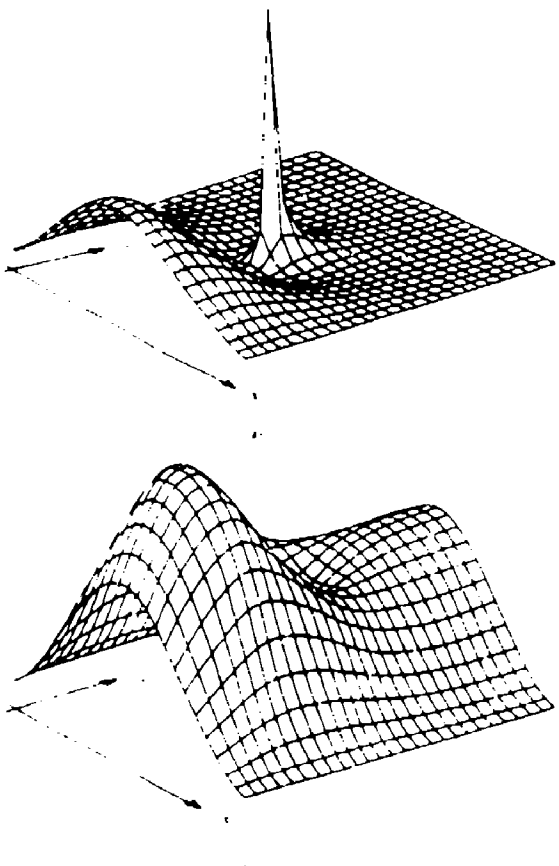
• Fig. 6. The spatial distribution of the current in a 900 Å long and 1000 Å wide structure containing two attractive impurities. (a). No magnetic field is present and two vortices form as a result of quantum interference between waves reflected from the two scatterers and the walls of the structure, (b). a magnetic flux density of 0.08 Tesla is applied and one of the vortices is quenched, (c). the magnetic flux density is increased to 1 Tesla and the first vortex reappears at a different location while the second vortex disappears.

pears at a different location and the second vortex is quenched with the simultaneous beginning of the formation of edge states.

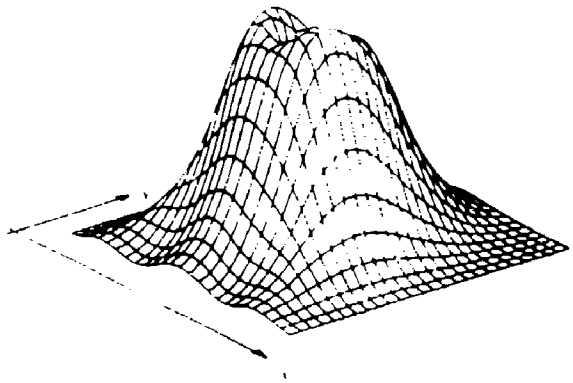
Carrier concentration change due to current flow: In Fig. 7(a) and 7(b), we show the change in the carrier concentration (from the equilibrium value) around the attractive and repulsive scatterers of Figs. 3(a) and 3(b) respectively. Carriers pile up around the attractive impurity and are depleted around the repulsive impurity because of electrostatic interaction.

In Fig. 8, we show the carrier pile up around the impurity when a magnetic bound state forms. The formation of such a state can cause significant inhomogeneities in the carrier profile and the associated electric field.

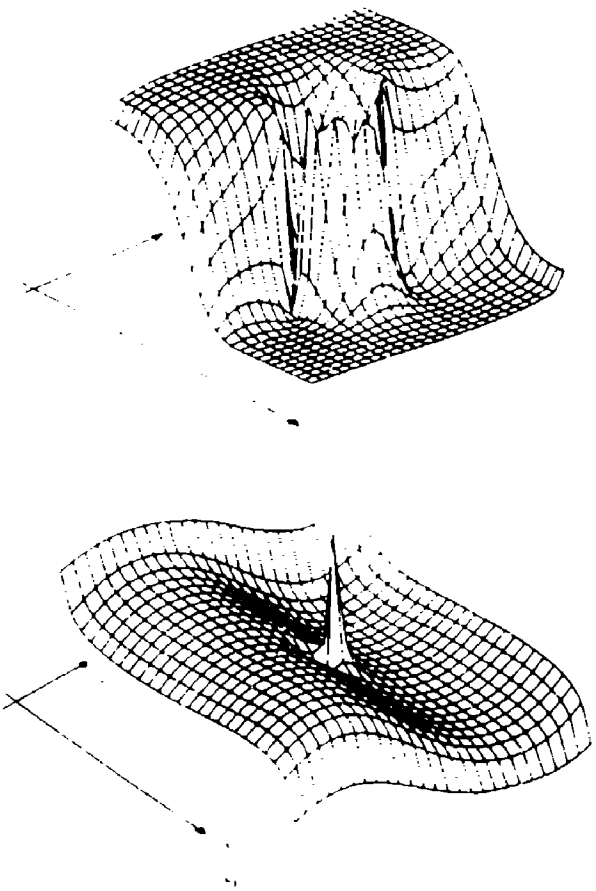
Chemical potential profile: In Figure 9(a), we show the chemical potential profile around the attractive impurity of Fig. 3(a). We assume that the chemical potentials at the two contacts of the structure (at x=0 and x=L) are μ_1 = 2.056 meV and μ_2 = 2.052 meV. Almost all of the chemical potential difference ($\mu_1 - \mu_2$) is dropped at the center where the impurity is located. This is expected since the impurity that is the major cause of the resistance for this sample. There are some oscillations in the chemical potential around the impurity but these are just signatures of quantum interference effects. The case of a repulsive impurity is very different. In this case (Fig. 9(b)), we see that



• Fig. 7. The spatial distribution of the change in the carrier concentration $\Delta n(x, y)$ from the equilibrium value due to the current. No magnetic field is applied. (a) the scatterer is attractive, (b) the scatterer is repulsive.



• Fig. 8. The spatial distribution of $|\psi|^2$, $\psi(x,y)$ at a magnetic flux density of 3.5 Tesla when a magnetic bound state forms.



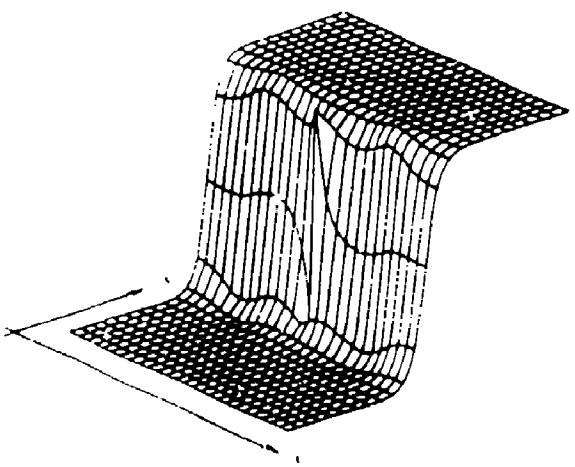
• Fig. 9. The spatial distribution of the chemical potential in the absence of any magnetic field. (a) the scatterer is attractive (corresponding to the case of Fig. 3(a)), (b) the scatterer is repulsive (corresponding to the case of Fig. 3(b)). Note that in the first case, most of the potential drop occurs around the impurity which is the dominant cause of resistance. In the second case, the potential is dropped more at the contacts showing that the contact resistances are dominant over the residual resistance of the impurity. This is a major difference between attractive and repulsive scatterers.

the potential drop occurs more near the contacts than around the impurity. There is only a narrow peak around the impurity but not much of a net drop from one end of the impurity to the other. This means that the major contributor for the resistance in this case is not the impurity, but the contact resistance. A repulsive impurity appears to be much less effective than an attractive impurity *of the same scattering cross-section* in reflecting electrons and causing resistance. We believe that this is related to the fact that quasi-donor states (quasi bound states) form around an attracive impurity[19] but not around a repulsive one. When quasi donor states form, evanescent waves build up around the impurity which cause a tremendous amount of reflection and hence resistance. *This difference between the resistances caused by an attractive impurity and a repulsive impurity of the same scattering cross-section is a major difference between majority carrier transport (when most scatterers are attractive) and minority carrier transport (when most scatterers are repulsive).* Note that this difference is purely a consequence of quantum mechanics since semiclassical scattering theories, such as the Born approximation or Fermi's Golden rule, do not discriminate between attractive and repulsive scatterers. An experimental verification of this difference would be to demonstrate a large difference between majority carrier mobilities (measured by Hall effect) and minority carrier mobilities (measured by Shockley-Haynes method).

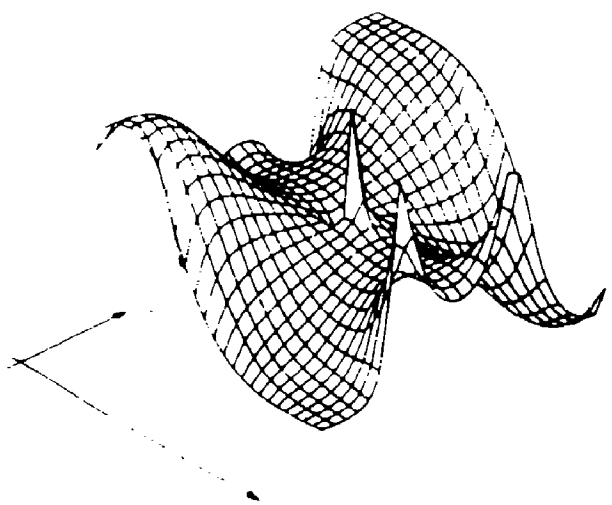
In Figure 10, we show how the chemical potential profile of Figure 9(a) changes when a magnetic flux density of 2.18 Tesla is turned on. The edges of the wire (along the length) become strikingly smooth equipotential surfaces. Consequently, the longitudinal resistance measured by attaching two voltage probes at any of the two edges will read exactly zero since the potential difference measured by these probes will be exactly zero. This is a manifestation of the *quantum Hall effect*. Also the Hall voltage drop (the difference between the chemical potentials at the two edges) is found to be exactly $\mu_1 - \mu_2$. Therefore, the Hall resistance is $(\mu_1 - \mu_2)/I$ where I is the current flowing in the structure. The current I is given by $I = G(\mu_1 - \mu_2)$, where G is the two-terminal conductance. In reference 1, we calculated G for this structure and found it to be precisely $2e^2/h$. Therefore, the Hall resistance is exactly $h/2e^2$ which is again a manifestation of the integer quantum Hall effect.

In Fig. 11, we show how the chemical potential changes when the flux density is increased to 3.5 Tesla. At this flux density, a magnetic bound state forms. The bound state couples electrons between the two sets of edge states carrying current along the two edges of the wire and causes backscattering. This destroys the quantum Hall effect[1,4] - a fact that is clearly seen from the chemical potential profile. The edges of the sample are no longer equipotential surfaces so that the longitudinal resistance will no longer measure zero. Also, the Hall voltage drop is different from $(\mu_1 - \mu_2)$ so that the Hall resistance is not quantized to sub-multiples of h/e^2. *This is an unequivocal demonstration that a magnetic bound state in a narrow wire destroys the integer quantum Hall effect.*

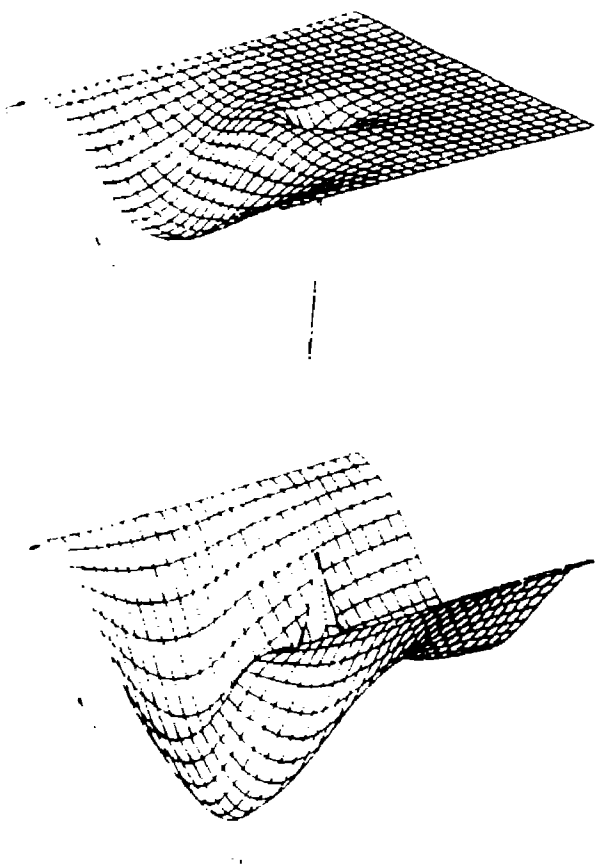
Spatial distribution of the residual resistivity dipole potential: In Figs. 12(a) and 12(b), we show the residual resistivity dipole potential in the case of the single attractive scatterer and the single repulsive scatterer of Figs. 3(a) and 3(b). The potential has an approximately $1/r$ decay or rise (r where r is the radial distance from the scatterer), as predicted by Chu and Sorbello for the case of a two-dimensional electron gas[2].



• Fig. 10. The spatial distribution of the chemical potential at a magnetic flux density of 2.18 Tesla. Note that the edges of the structure (along the x-direction) are exceptionally smooth equipotential surfaces so that the longitudinal resistance measured by attaching two probes at either edge will be exactly zero since the chemical potential difference between these two probes will be zero. The potential drop along the y-direction (Hall voltage) is independent of the x-coordinate and is exactly equal to the potential drop between the two contacts of the structure. The Hall resistance is exactly $h/2e^2$. This figure is a direct visualization of the integer quantum Hall effect.



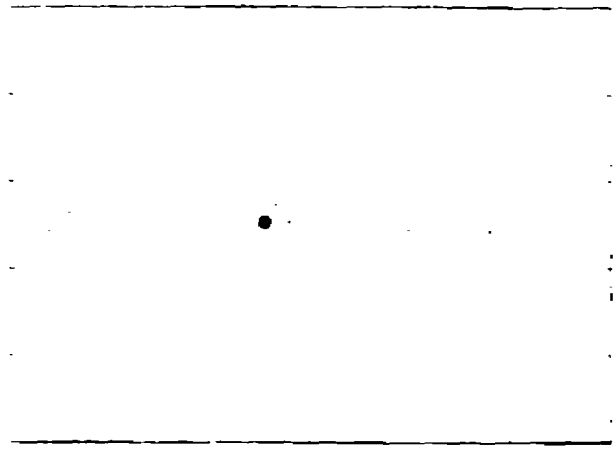
• Fig. 11. The spatial distribution of the chemical potential when the magnetic flux density is increased to 3.5 Tesla and a magnetic bound state forms. The edges are no longer equipotential surfaces and the Hall voltage drop is very different. The longitudinal resistance is no longer zero and the transverse resistance is no longer quantized to sub-integral multiples of h/e^2. This figure is a direct visualization of the destruction of the integer quantum Hall effect in a narrow wire by a magnetic bound state.



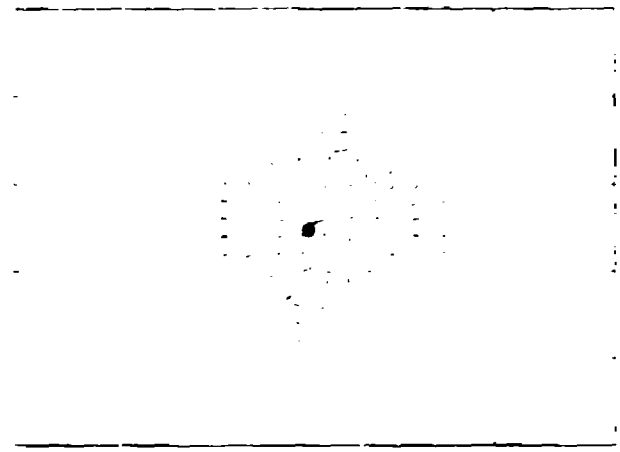
• Fig. 12. The spatial distribution of the residual resistivity dipole potential in the absence of any magnetic field. (a) The scatterer is attractive, (b) the scatterer is repulsive.

Spatial distribution of the electric field: In Fig. 13, we show the electric field distribution in the quantum Hall regime corresponding to the situation in Fig. 10. All the field is concentrated at the center of the wire and is basically the Hall field. When the magnetic field is turned off, the electric field profile changes to that shown in Fig. 14. There is no net field in the y-direction since there is no Hall voltage drop in this case.

The forces of electromigration: In Figs. 15(a) and 15(b), we show the directions of the wind forces and also the direct forces (recall that the two forces are equal) on two impurities in a 900 Å long and 1000 Å wide structure at zero magnetic field and at a magnetic flux density of 3.5 Tesla. Unlike in the case of a single impurity, no magnetic bound state forms at the flux density of 3.5 Tesla when two impurities are present. Note that at zero magnetic field, the wind forces on neither impurity is in the direction of the electric field (x-direction) in contrast to the predictions of semiclassical theories. This difference in the orientation of the forces is a quantum mechanical effect. In addition, magnetic field can drastically change the magnitude and the direction of the electromigration forces on both impurities. In the case of the first impurity on the left, the magnetic field almost completely removes the forces. In the case of the second impurity, the forces are enhanced by the magnetic field. Therefore, in the quantum coherent regime, whether a magnetic field



• Fig. 13. The spatial distribution of the electric field in the quantum Hall regime corresponding to the situation depicted in Fig. 10. The magnetic flux density is 2.18 Tesla. All the field is at the center and is directed along the y-direction (Hall field).



• Fig. 14. The spatial distribution of the electric field when the magnetic field is turned off.

increases or decreases the direct force on an ion (or whether it has any effect at all), is determined by the *locations* of the ions within the sample. This is purely a quantum-mechanical phenomenon.

Conclusions

In this paper, we have presented the spatial distributions of a large number of transport variables associated with various quantum magnetotransport phenomena in mesoscopic samples. These distributions elucidate the integer quantum Hall effect, the formation of magnetic bound states and its effect on the integer quantum Hall effect, the difference between resistances caused by attractive and repulsive scatterers of the same scattering cross-section and the corresponding difference between minority and majority carrier mobilities, the magnetic response of current vortices formed as a result of quantum interference between scatterers, the electric field distribution in the quan-

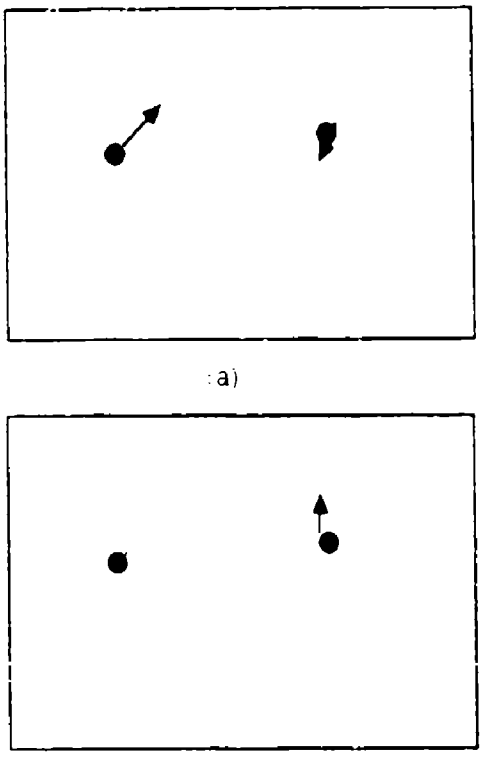


Fig. 15. The forces of electromigration acting on two impurities in a 900 Å long and 1000 Å wide structure. (a). No magnetic field is present. (b) The magnetic flux density is 3.5 Tesla.

tum Hall regime, the residual resistivity dipole potential and its spatial variations, and finally, the direct and wind forces of electromigration and their dependences on a magnetic field as well as the location of the ion. We believe that these results will be useful in interpreting many different types of transport phenomena and electromigration in mesoscopic structures.

Acknowledgement - This work was supported by the US Air Force Office of Scientific Research under Grant No. AFOSR 91-0211, by the US Office of Naval Research under contract no. N00014-91-J-1505, and by the National Science Foundation under grant no. ECS 9108932.

REFERENCES

1. S. Chaudhuri, S. Bandyopadhyay and M. Cahay, *Physical Review B*, Vol. 45, 11126 (1992); S. Chaudhuri, S. Bandyopadhyay and M. Cahay, to appear in *Physical Review B*; C. S. Lent, *Applied Physics Letters*, Vol. 67, 1678 (1990); Zhen-Li Ji and K. F. Berggren, *Physical Review B*, Vol. 45, 6652 (1992); K. F. Berggren, C. Besev and Zhen-Li Ji, to appear in *Physica Scripta*; Zhen-Li Ji, *Semiconductor Science and Technology*, Vol. 7, 198 (1992).

2. R. S. Sorbello, *Physical Review B*, Vol. 23, 5119 (1981); R. S. Sorbello and C. S. Chu, *IBM Journal of Research and Development*, Vol 32, 58 (1988).

3. R. Landauer, *IBM Journal of Research and Development*, Vol. 1, 223 (9157); R. Landauer and J. W. F. Woo, *Physical Review B*, Vol. 10, 1266 (1974); R. Landauer, *Z. Physik B*, Vol. 21, 247 (1975).

4. S. Chaudhuri and S. Bandyopadhyay, *Journal of Applied Physics*, Vol. 71, 3027 (1992).

5. V. B. Fiks, *Soviet Physics - Solid State*, Vol. 1, 14 (1959); H. B. Huntington and A. R. Grone, *Journal of Physics and Chemistry of Solids*, Vol. 20, 76 (1960).

6. There are subtle issues regarding whether the electron momentum should be taken as the crystal momentum or the real momentum or a combination of the two. This question has not been resolved satisfactorily and we will assume for the sake of simplicity that the relevant momentum is the crystal momentum. Since, in this paper, we will only calculate the direction and not the exact magnitude of the wind force, we believe that the distinction between crystal momentum and real momentum is not very important in this context.

7. S. Bandyopadhyay, M. E. Klausmeier-Brown, C. M. Maziar, S. Datta and M. S. Lundstrom, *IEEE Transactions on Electron Devices*, Vol. ED-34, 392 (1987).

8. C. Bosvieux and J. Friedel, *Journal of Physics and Chemistry of Solids*, Vol. 23, 123 (1962).

9. R. S. Sorbello, *Journal of Physics and Chemistry of Solids*, Vol. 34, 937 (1973); A. H. Verbruggen, *IBM Journal of Research and Development*, Vol. 32, 93 (1988).

10. R. P. Gupta, Y. Serruys, G. Brebec and Y. Adda, *Physical Review B*, Vol. 27, 672 (1983).

11. A. K. Das and R. Peierls, *Journal of Physics C: Solid State Physics*, Vol. 6, 2811 (1973).

12. R. Landauer, *Journal of Physics C: Solid State Physics*, Vol. 8, 761 (1975); R. Landauer, *Physical Review B*, Vol. 14, 1474 (1976).

13. H. B. Huntington, *Transactions of Metallurgical Society*, Vol. AIME 245, 2571 (1969).

14. L. Lou, W. L. Schaich and J. C. Swihart, *Physical Review B*, Vol. 33, 2170 (1986).

15. M. Büttiker, *IBM Journal of Research and Development*, Vol. 32, 317 (1988).

16. O. Entin-Wohlman, C. Hartzstein and Y. Imry, *Physical Review B*, Vol. 34, 921 (1986).

17. C. S. Lent, *Physical Review B*, Vol. 43, 4179 (1991).

18. M. Büttiker, *Physical Review B*, Vol. 38, 9375 (1988).

19. P. F. Bagwell, *Physical Review B*, Vol. 41, 10354 (1990).

20. J. K. Jain and S. Kivelson, *Physical Review Letters*, Vol. 60, 1542 (1988); Y. Lee, M. J. McLennan and S. Datta, *Physical Review B*, Vol. 43, 14333 (1991).

PHYSICAL REVIEW B VOLUME 47, NUMBER 0 15 APRIL 1993

Effects of collision retardation on hot-electron transport in a two-dimensional electron gas

N. Telang and S. Bandyopadhyay
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556
Received 1 May 1992; revised manuscript received 20 November 1992)

The effect of collision retardation on hot-electron transport in a two-dimensional electron gas is examined using an ensemble Monte Carlo simulation. We find that collision retardation (i.e., a nonzero collision duration) tends to make the electrons hotter by suppressing energy-relaxing collision events. Collision retardation also increases the steady-state drift velocity and high-field mobility by suppressing momentum-relaxation events. Finally, it also increases velocity overshoot somewhat.
PACS number(s): 71.45.Gm, 03.65.Sq, 71.10.+x, 78.47.+p

Hot-electron transport within the semiclassical formalism has traditionally been modeled by the Boltzmann transport equation (BTE). The assumption made in applying the BTE is that the *duration* of individual collision events suffered by electrons is vanishingly small compared to quasiparticle lifetimes or mean times between successive collisions.[1,2] If this assumption is to be avoided, then one must either resort to the full quantum kinetic equation[3] (such as the Kadanoff-Baym-Keldysh equation, which is much more difficult to solve than the BTE) or incorporate the effects of a finite collision duration (collision retardation) in the BTE in some heuristic but appropriate manner.

Recently, the effects of collision retardation were incorporated phenomenologically in the BTE.[4] The BTE is solved by Monte Carlo simulation, and in the simulation a scattering event is treated as a true scattering event only if a uniform random number in the interval [0,1] is larger than the quantity $\exp[-t/\tau_d]$, where t is the time that elapsed since the previous collision and τ_d is the collision duration time for the event. If the random number is smaller than this quantity, then the event is considered a self-scattering event. The collision duration time τ_d is assumed to be equal to $\hbar/(E-E_{th})$, where E is the initial energy of the colliding electron and E_{th} is the threshold energy for the scattering process. This expression for τ_d is derived from Landau's model for metals[5] (Fermi-liquid theory). In addition, Lipavsky *et al.*[6] have shown that τ_d calculated from this expression is identical with the quasiparticle formation time associated with the single-particle propagator. Using this expression to calculate τ_d as was done in Ref. 4), we have studied the effects of a finite collision duration on hot-electron transport in a two-dimensional electron gas, using the algorithm proposed in Ref. 4.

The test system that we chose for our simulation is a rectangular quantum well of length 1 μm, width 10 μm, and well thickness 100 Å. The confining potentials in both transverse directions are infinite (hardwall boundary conditions). The well material is GaAs and the lattice temperature is assumed to be 40 K. Electrons are injected from the left contact from a Maxwellian distribution and the simulation proceeds just as described in Refs. 7 and 10. There are, however, two differences between our approach and that of Ref. 7. We do not include space-charge effects by solving the Poisson equation at every time step in the Monte Carlo simulation, and instead of using a full-band Monte Carlo, we chose an approximate analytical model for the band structure of GaAs, which gives the energy dispersion relation as

$$\frac{\hbar^2 k^2}{2m^*} = E(1+\alpha E)\,, \tag{1}$$

where E and k are the energy and wave vector, respectively, m^* is the effective mass at the band bottom, and α is the nonparabolicity factor. The parameters m^* and α are different for the three different conduction-band valleys in GaAs and their values are chosen from Ref. 8. Since the electric field in our simulation is quite low (only 500 V/cm), we believe that the above approximate analytical relation for the band structure is adequate for our purpose. Note that it is necessary to keep the electric field low in order to ensure that the collision retardation time τ_d is typically much smaller than the mean time between collisions. This situation is necessary for the algorithm of Ref. 4 to be valid.[9]

In the simulation, we considered intrasubband and intersubband nonpolar acoustic-phonon scattering, intrasubband and intersubband polar optical-phonon scattering, electron-electron scattering, and intervalley scattering. Piezoelectric (polar acoustic-phonon) scattering, nonpolar optical-phonon scattering, and remote ionized impurity scattering were neglected since they are not very important in modulation-doped GaAs quantum wells at the lattice temperature of 40 K. Also, plasmon scattering was not included. Electron-electron scattering is modeled after Goodnick and Lugli,[11] who have calculated the rates for two-dimensional electron gases. Phonon scattering was treated by Ridley's model[12] for quantum wells, which assumes the phonon modes to be bulk modes rather than confined slab modes and neglects surface modes altogether. This is not a bad approximation. Since the amplitudes of the slab modes decay at the interface while those of the surface modes increase at the interface, the sum of all modes will appear approximately bulklike.[10] In fact, the scattering rates calculated by using bulk modes do not differ greatly[13] from those calculated by using more sophisticated models (including mi-

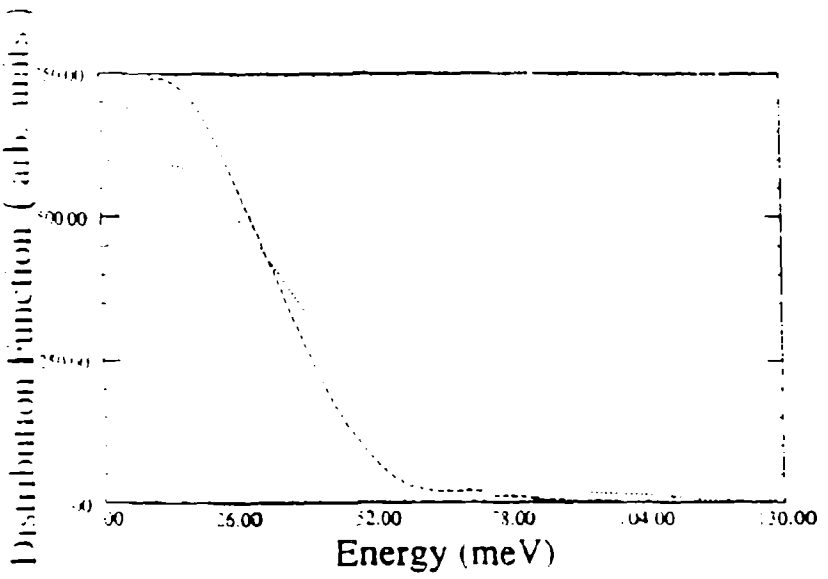


FIG. 1. The steady-state electron-distribution functions in energy for an electric field of 500 V/cm. The short-dashed line corresponds to the case when collision retardation is included and the long-dashed line corresponds to the case when collision retardation is neglected.

croscopic models).[14] Finally, we assume that the phonon modes are decoupled from plasmon modes, which is a good approximation[15] at the low carrier concentration of $10^{11}/cm^3$. We also neglect hot-phonon effects, the role of the Pauli exclusion principle,[16] self-consistent (space-charge) effects,[7] and many-body effects (exchange/correlation)[17] in the simulation.

In Fig. 1, we show the steady-state electron distribution functions in energy for an applied electric field of 500 V/cm with and without collision retardation. Collision retardation shifts electrons from low-energy states to the high-energy tail, thereby causing a relative depopulation of low-energy states. Both distribution functions are approximately drifted Maxwellians but with very different

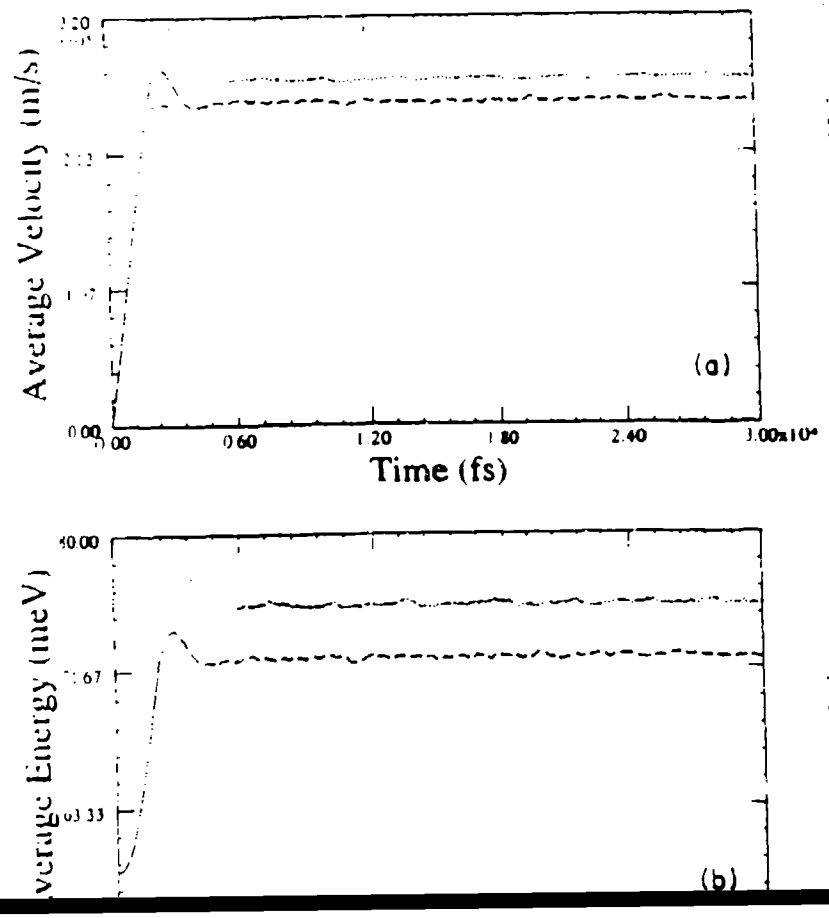


FIG. 2. (a) Velocity and (b) energy vs time for an electric field of 500 V/cm. The short-dashed line and the long-dashed line correspond, respectively, to the situations when collision retardation is included and neglected.

temperatures. When collision retardation is neglected (long-dashed line), the electron temperature is calculated to be ~98 K (the lattice temperature is 40 K). However, when collision retardation is included (short-dashed line), the electron temperature rises to ~147 K. Therefore, collision retardation makes the electron distribution much hotter. This is obviously due to the fact that collision retardation suppresses scattering by rejecting many scattering events near the thresholds. The scatterings that are suppressed are those for which the collision duration exceeds the time elapsed since the previous collision. For instance, retardation completely suppresses optical-phonon emission at just above the emission threshold, since the collision duration for such an event is infinitely long. This robs the electron ensemble of many energy-relaxing scattering events and makes the distribution hotter.

In Figs. 2(a) and 2(b), we show the transient response of the ensemble average velocity and the average energy to an applied electric field of 500 V/cm. Collision retardation (short-dashed line) *increases* the velocity and energy overshoot somewhat, and also *increases* the steady-state velocity and energy. The increase in the steady-state velocity is obviously caused by the suppression of momentum-relaxing collisions due to retardation, and the increase in energy is caused by the suppression of energy-relaxing events. It is interesting to note that collision retardation has the beneficial effects of increasing both the steady-state velocity and the velocity overshoot, which have serious implications for high-speed device applications. However, the increase is only slight; it is merely ~10%.

In conclusion, we have studied the effects of collision retardation on hot-electron transport in quantum-well samples. The results show that retardation increases the steady-state drift velocity, average energy, and the high-field mobility. These have important implications for high-speed devices.

This work was supported by the U.S. Air Force Office of Scientific Research under Grant No. AFOSR 91-0211.

[1] L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Addison-Wesley, New York, 1989).

[2] D.K. Ferry, *Semiconductors* (Macmillian, New York, 1991); J. R. Barker and D. K. Ferry, Phys. Rev. Lett. 42, 1779 (1979); J. R. Barker and D. K. Ferry, Solid-State Electron. 23, 531 (1980); J. R. Barker, J. Phys. C 6, 880 (1973).

[3] M. J. McLennan, Y. Lee, and S. Datta, Phys. Rev. B 43 13 486 1991).

[4] D. K. Ferry, A. M. Kriman, H. Hida, and S. Yamaguchi, Phys. Rev. Lett. 67, 633 (1991).

[5] R. Peierls, *Quantum Theory of Solids* (Oxford Universty Press, London, 1955).

[6] P. Lipavsky, F. S. Khan, A. Kalvova, and J. W. Wilkins, Phys. Rev. B 43, 6650 (1991).

[7] U. Ravaioli, in *Monte Carlo Device Simulation: Full Band and Beyond*, edited by K. Hess (Kluwer, Boston, 1991), Chap. 9.

[8] T. H. Glisson, C. K. Williams, J. R. Hauser, and M. A. Littlejohn, *VLSI Microstructure Science* (Academic, New York, 1982), Vol. 5, Chap. 3, p. 95.

[9] D. K. Ferry (private communication).

[10] M. Artaki and K. Hess, Phys. Rev. B 37, 2933 (1988).

[11] S. M. Goodnick and P. Lugli, Phys. Rev. B 37, 2578 (1988).

[12] B. K. Ridley, J. Phys. C 15, 5899 (1982).

[13] P. Yu, (unpublished).

[14] K. Huang and B. Zhu, Phys. Rev. B 38, 13 377 (1988).

[15] Xiaoguang Wu, F. M. Peeters, and J. T. Devreese, Phys. Rev. B 32, 6982 (1985).

[16] P. Lugli and D. K. Ferry, IEEE Trans. Electron Devices ED-32, 2431 (1985).

[17] A. M. Kriman, M. J. Kann, D. K. Ferry, and R. Joshi, Phys, Rev. Lett. 65, 1619 (1990).

To appear in PRL

Quenching of electron-acoustic phonon scattering in quantum wires by a magnetic field

N. Telang and S. Bandyopadhyay
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

ABSTRACT

We demonstrate the possibility of a dramatic quenching of acoustic phonon scattering in semiconductor quantum wires due to an external magnetic field. This quenching is the consequence of a reduction in the overlap between initial and final state electron wavefunctions (which reduces the scattering matrix element) in the presence of an external magnetic field. Such behavior is in sharp contrast to the case for optical phonons where the opposite trend is observed. The quenching phenomenon is important in understanding the suppression of *inelastic* backscattering at high magnetic fields which leads to the quantum Hall effect (in the Büttiker picture). Additionally, such quenching can give rise to strong negative magnetoresistance in narrow quantum wires which may have important applications in electronic devices and magnetic field sensors.

Electron-phonon scattering in quantum wires has been studied by a number of researchers in the past. Most of the theoretical work[1-5] (with some exception) has addressed electron interactions with confined optical phonons and surface phonons only, while neglecting interactions with either non-polar or polar acoustic phonons. Virtually no work has appeared in the literature that deals with the effect of a magnetic field on phonon scattering rates.

Recently, the effect of a magnetic field on confined polar optical phonon and surface phonon scattering rates was studied by us[5]. That study revealed that an external magnetic field *increases* the electron scattering rate for both polar optical phonon and surface phonon interactions, thereby giving rise to a positive magnetoresistance in quantum wires.

We have now extended our study to the case of electron interactions with both polar and non-polar *acoustic* phonons. These phonons are more numerous than optical phonons in most technologically important semiconductors at or below room temperature. We find the following intriguing features: (1) In contrast to the case of optical phonon scattering, the acoustic phonon scattering rate in a quantum wire is *decreased* by an external magnetic field. (2) the decrease can be dramatic (scattering rates can be diminished by up to *six* orders of magnitude at a magnetic flux density of 10 tesla), and (3) this decrease is relatively independent of the lattice temperature. In the following paragraphs, we describe our theoretical model and then present the results.

Electron-phonon scattering rate is calculated from Fermi's Golden Rule. Although the application of this rule at electron energies corresponding to subband minima in quantum wires has been criticized[7], we still use the Golden Rule because the quenching occurs mostly at energies far away from subband minima. In fact, the quenching increases with increasing energy

separation from the subband minimum.

The scattering rate for scattering from an energy state E_ν in subband ν in a quantum wire to an energy state $E'_{\nu'}$ in subband ν' is given by

$$S(E_\nu, E'_{\nu'}, \pm\vec{q}) = (2\pi/\hbar)\,|M(E_\nu, E'_{\nu'}, \pm\vec{q})|^2\,\delta(E'_{\nu'} - E_\nu \pm \hbar\omega_{\vec{q}}) \qquad (1)$$

where $M(E_\nu, E'_{\nu'}, \pm\vec{q})$ is the matrix element for transition due to an acoustic phonon of energy $\hbar\omega_{\vec{q}}$ and wavevector $\vec{q}$. The $\pm$ sign stands for emission (+) and absorption (-) respectively. For acoustic phonons, we assume a linear dispersion relation $\hbar\omega_{\vec{q}} = \hbar v_s q_x$ where v_s is the longitudinal velocity of sound (electrons only interact with longitudinal phonons).

The matrix element is given by[8]

$$M(E_\nu, E'_{\nu'}, \pm\vec{q}) = V^{\pm}\delta_{k-k',q_z}\int_{-L_y/2}^{L_y/2}\psi^*_{E'_{\nu'}}(y)\psi_{E_\nu}(y)e^{\pm iq_y y}dy\int_{-L_z/2}^{L_z/2}\phi^*_{E'_{\nu'}}(z)\phi_{E_\nu}(z)e^{\pm iq_z z}dz \qquad (2)$$

where $\psi_{E'_{\nu'}}$ ($\phi_{E'_{\nu'}}$) and ψ_{E_ν} (ϕ_{E_ν}) are the transverse y- (z-) components of the electron wavefunctions of the final and initial states in the presence of a magnetic field. k' and k are the longitudinal electron wavevectors of these states. V is the interaction potential and the δ is a Kronecker delta which represents momentum conservation. We have assumed the acoustic phonon modes to be bulk modes (plane waves) that are unaffected by a magnetic field.

In the case of polar acoustic phonon interaction (piezoelectric scattering), the interaction potential is given by[9]

$$V^{\pm} = i\left(\frac{\hbar e^2 e_{pz}^2\left(N_q + \frac{1}{2} \pm \frac{1}{2}\right)}{2\Omega\rho v_s \epsilon^2 q}\right)^{1/2} \qquad (3)$$

where e is the electronic charge, e_{pz} is the piezoelectric constant, ϵ is the dielectric constant, Ω is the normalizing volume in the phonon Brillouin

zone and N_q is the phonon occupation probability which we assume to be the Bose-Einstein factor.

In the case of non-polar (deformation potential) scattering, the interaction potential is given by[9]

$$V^{\pm} = i\left(\frac{\hbar D^2}{2\Omega\rho v_s} q\left(N_q + \frac{1}{2} \pm \frac{1}{2}\right)\right)^{1/2} \tag{4}$$

where D is the deformation potential.

The total scattering rate for phonon emission or absorption $S^{\pm}(E_\nu)$ for electrons with energy E in subband ν is obtained by integrating over all possible final states and phonon wavevectors.

$$S^{\pm}(E_\nu) = \int_0^{\infty} \int_0^{q_{max}} dE'_{\nu'} d^3\vec{q} D(E'_{\nu'}) S(E_\nu, E'_{\nu'}, \pm\vec{q}) \tag{5}$$

where q_{max} is the maximum phonon wavevector and $D(E'_{\nu'})$ is the density of hybrid magnetoelectric states that form in a quantum wire under an applied magnetic field. Both $D(E'_{\nu'})$ and the wavefunctions of the magnetoelectric states ψ (which appear in Equation (2)) are found by solving the Schrödinger equation for the quantum wire in a magnetic field. The solution employs a numerical finite difference method. This scheme has been described in Ref. 10.

The structure that we chose for illustration is a GaAs quantum wire of rectangular cross-section having width 500 Å in the y-direction and thickness 40 Å in the z-direction. The magnetic field is oriented along the z-direction. Only one subband is occupied in the z-direction even for the highest electron energy considered, but many are occupied in the y-direction.

In Fig. 1(a) and 1(b), we plot the piezoelectric emission and absorption rates $S^{+}(E_1)$ and $S^{-}(E_1)$ as a function of electron energy E_1 in the lowest hybrid magnetoelectric subband in the wire. There is a dramatic decrease

in the intra-subband scattering rates with increasing magnetic flux densities. This means that in narrow enough quantum wires (where only intra-subband scattering is allowed even for the highest energy that an electron can reach), dramatic quenching of acoustic phonon scattering can occur in a magnetic field leading possibly to strong *negative* magnetoresistance. In Fig. 2(a) and 2(b), we plot the scattering rates for deformation potential acoustic phonon scattering. Again, the same feature is observed.

The origin of this quenching is the following. Since both energy and momentum must be conserved in the scattering process, there are certain restrictions on the angle of scattering. For intra-subband scattering, primarily backward (large angle) scattering is allowed (in contrast to the case for polar optical phonon and impurity scattering). This means that the initial and final states of the electron (before and after scattering) will tend to have oppositely directed velocities. In the absence of a magnetic field, two such states will have a large wavefunction overlap (essentially 100 %) so that the matrix element for scattering (Equation 2) is large. However, when a magnetic field is applied, the wavefunctions $\psi_{E_y'}(y)$ and $\psi_{E_y}(y)$ of two such states will be skewed in opposite directions since the Lorentz force acts in opposite directions on oppositely traveling electrons. This decreases the overlap, and therefore the matrix element, which causes the quenching. To illustrate this effect, we have plotted in Fig. 3, the wavefunction of two oppositely traveling states (in the lowest magnetoelectric subband) at a magnetic flux density of 0.1 and 10 tesla. Note that the wavevefunctions are skewed towards opposite edges of the wire by the magnetic field which dramatically reduces the overlap between them and therefore the scattering matrix element given by Equation (2). This explains the quenching. Supression of acoustic phonon

emission in superlattices due to a magnetic field have been observed experimentally before[11] but was attributed to a different cause. To our knowledge, the present mechanism has never been reported before.

The quenching of acoustic phonon scattering by a strong magnetic field has serious implications for the integer quantum Hall effect. At very high magnetic fields and low temperatures (quantum Hall regime), current in a wire is carried exclusively by "edges states"[12] that are localized along the edges of the wire. Edge states localized along the same edge of the wire have the same direction of the electron velocity and carry current in the same direction, while edge states flowing along the opposite edge have opposite velocities and carry current in the opposite direction. In the light of the previous discussion, we can see that the probability of acoustic phonon scattering between oppositely traveling edge states is extremely low. This contributes to our understanding as to why backscattering is suppresed at high magnetic fields.

The suppression of backscattering is the central ingredient in the Büttiker picture of the quantum Hall effect[12]. In the past, we and others had demonstrated this suppresion in the quantum Hall regime considering only *elastic* (phase preserving) scattering events[13]. We now demonstrate this suppression for *inelastic* (phase breaking acoustic phonon) events as well. This is a very important extension of the Büttiker picture beyond the coherent linear response transport regime (where only elastic scattering is permitted). The quantum Hall effect is known to survive at finite voltages and currents when both elastic and inelastic scattering occur. This has been demonstrated theoretically[14] as well as experimentally[15]. Previous models attributed the breakdown of the quantum Hall effect at high current densities to the on-

set of acoustic phonon emission[7]. In that light, the suppression of acoustic phonon emission has important implications for the breakdown. It has also been claimed that the critical current for breakdown increases with increasing magnetic field. We believe that this may be due to increased suppression of phonon scattering with increasing field which is consistent with our result.

We have performed the calculation of the scattering rates at various temperatures (4.2 K, 77 K and 300 K). We found that the relative decrease in the scattering rate is fairly insensitive to temperature. The cause of the quenching is the decrease in the initial and final state wavefunctions in a magnetic field and this phenomenon of course does not depend on temperature. The experimental manifestation of the quenching however will be most prominent at those temperatures where acoustic phonon scattering is the dominant scattering mechanism in the quantum wire.

In conclusion, we have showed that a magnetic field can cause a dramatic decrease in acoustic phonon scattering (specifically intra-subband scattering) in a quantum wire. In narrow enough quantum wires (where only one subband is accessible in energy even at the highest applied electric field), this can cause a significant negative magnetoresistance if the wire material is such that acoustic phonon scattering is the dominant scattering mechanism at the temperature of interest. This effect can then be used to realize magnetic field sensors and other types of electronic devices where the resistance can be modulated by an external magnetic field. Finally, we have shown that acoustic phonon scattering between opposite sets of edge states is suppressed at high magnetic fields which reinforces the Büttiker picture of the integer quantum Hall effect.

Acknowledgement: This work was supported by the US Air Force

Office of Scientific Research under Grant No. AFOSR 91-0211.

References

1. S. Briggs and J. P. Leburton, *Phys. Rev. B*, **38**, 8163 (1988).

2. D. Jovanovic, S. Briggs and J. P. Leburton, *Phys. Rev. B*, **42**, 11108 (1990).

3. K. W. Kim, M. A. Stroscio, A. Bhatt, R. Mickevicius and V. V. Mitin, *J. Appl. Phys.*, **70**, 319 (1991).

4. Bangtien Zhu (unpublished).

5. N. Mori, H. Momose and C. Hamaguchi, *Phys. Rev. B*, **45** 4536 (1992).

6. N. Telang and S. Bandyopadhyay, *Proc. of the International Workshop on Computational Electronics*, (Beckman Institute, University of Illinois, 1992) p. 237.

7. P. F. Bagwell, *Phys. Rev. B*, **43**, 9012 (1990).

8. D. Jovanovic and J. P. Leburton, *Monte Carlo Device Simulation: Full Band and Beyond*, ed. K. Hess, (Kluwer Academic Press, Boston, 1991).

9. B. K. Ridley, *Quantum Processes in Semiconductors*, (Oxford Science Publications, Oxford, UK, 1988).

10. S. Chaudhuri and S. Bandyopadhyay, *J. Appl. Phys.*, **71** 3027 (1992).

11. A. Kastalsky and A. L. Efros, *J. Appl. Phys.*, **69**, 841 (1991).

12. M. Büttiker, *Phys. Rev. B*, **38**, 9375 (1988).

13. See, for example, S. Bandyopadhyay, S. Chaudhuri, B. Das and M. Cahay, *Superlattices and Microstructures*, **12**, 123 (1992); Also, see, T. Martin and

S. Feng, *Phys. Rev. Lett.*, **64**, 1971 (1990); J. J. Palacios and C. Tejedor, *Phys. Rev. B*, **44**, S157 (1991). The last two references show the suppression of acoustic phonon emission as well but do so phenomenologically. They also assume a parabolic confining potential when the very definition of edge states becomes nubulose.

14. Y. Lee, M. J. McLennan, G. Klimeck, R. K. Lake and S. Datta, *Superlattices and Microstructures*, **11**, 137 (1992). This paper theoretically demonstrates the quantum Hall effect in the linear but incoherent response regime.

15. See, for example, G. Ebert, K. von Klitzing, K. Ploog and G Weimann, *J. Phys. C*, **16**, 5441 (1983); M. E. Cage, R. F. Dzuiba, B. F. Field, E. R. Williams, S. M. Girvin, A. C. Gossard, D. C. Tsui and R. J. Wagner, *Phys. Rev. Lett.*, **51**, 1374 (1983); F. Kuchar, G. Bauer, G. Weimann and H. Burkhard, *Surface Science* **142**, 196 (1984).

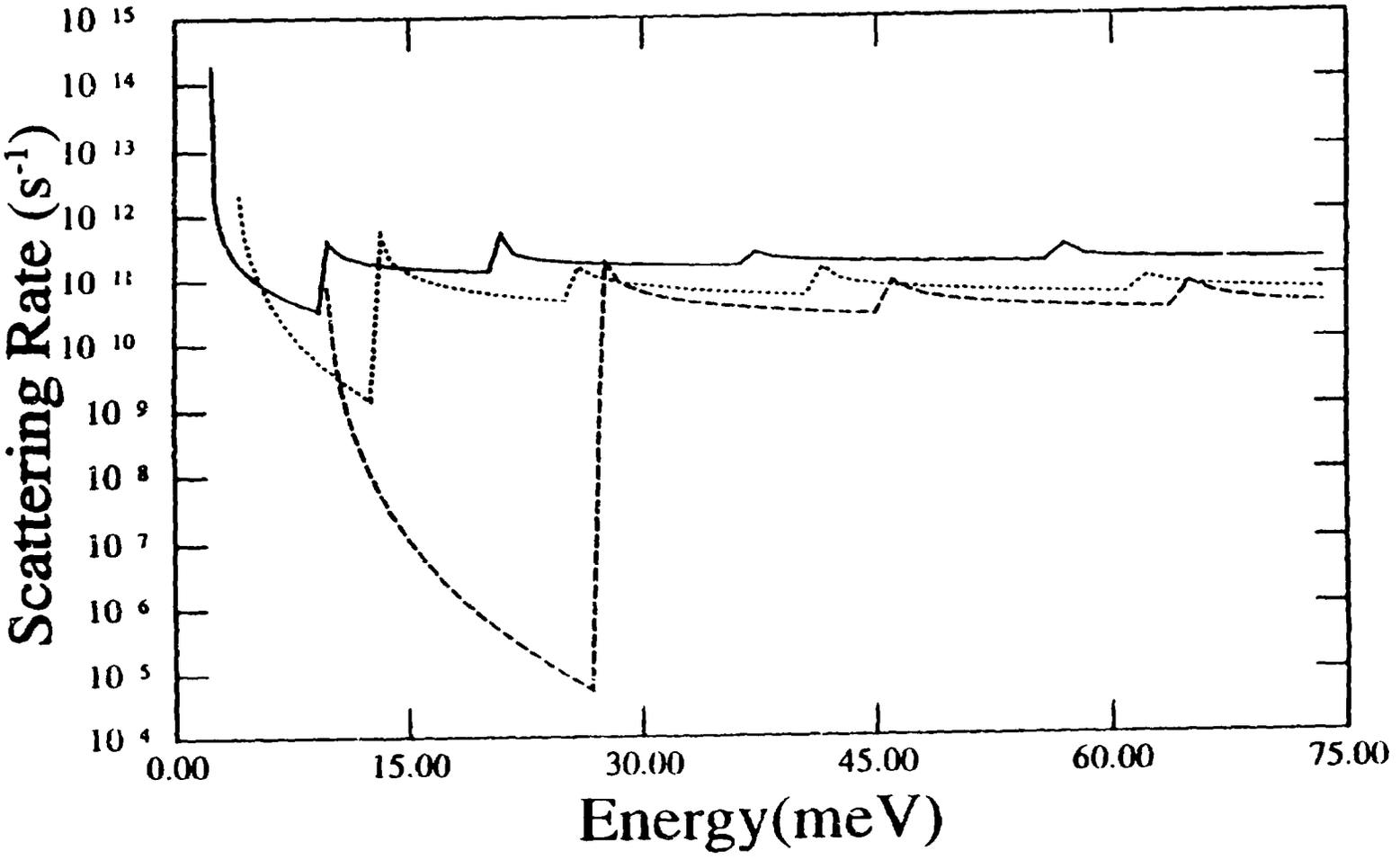
16. See, for example, O. Heinonen, P. L. Taylor and S. M. Girvin, *Phys. Rev. B*, **30**, 3016 (1984); P. Středa and K. von Klitzing, *J. Phys. C*, **17**, L483 (1984); H. L. Störmer, A. M. Chang, D. C. Tsui and J. C. M. Hwang, *Proc. of the 17th Int. Conf. on the Phys. of Semiconductors*, eds. D. J. Chadi and W. A. Harrison, (Springer-Verlag, Berlin, 1985), p. 267.

FIGURE CAPTIONS

Figure 1: Electron-polar acoustic phonon (piezoelectric) scattering rate as a function of energy for electrons in the lowest magnetoelectric subband in a GaAs quantum wire. Energy is measured from the bulk conduction band edge. The width of the wire is 500 Å (along the y-direction) and the thickness is 40 Å (along the z-direction). The lattice temperature is assumed to be 300 K. The solid line represents no magnetic field, the shortdashed line a magnetic flux density of 4 tesla and the long dashed line a magnetic flux desnity of 10 tesla. (a) emission rate, and (b) absorption rate.

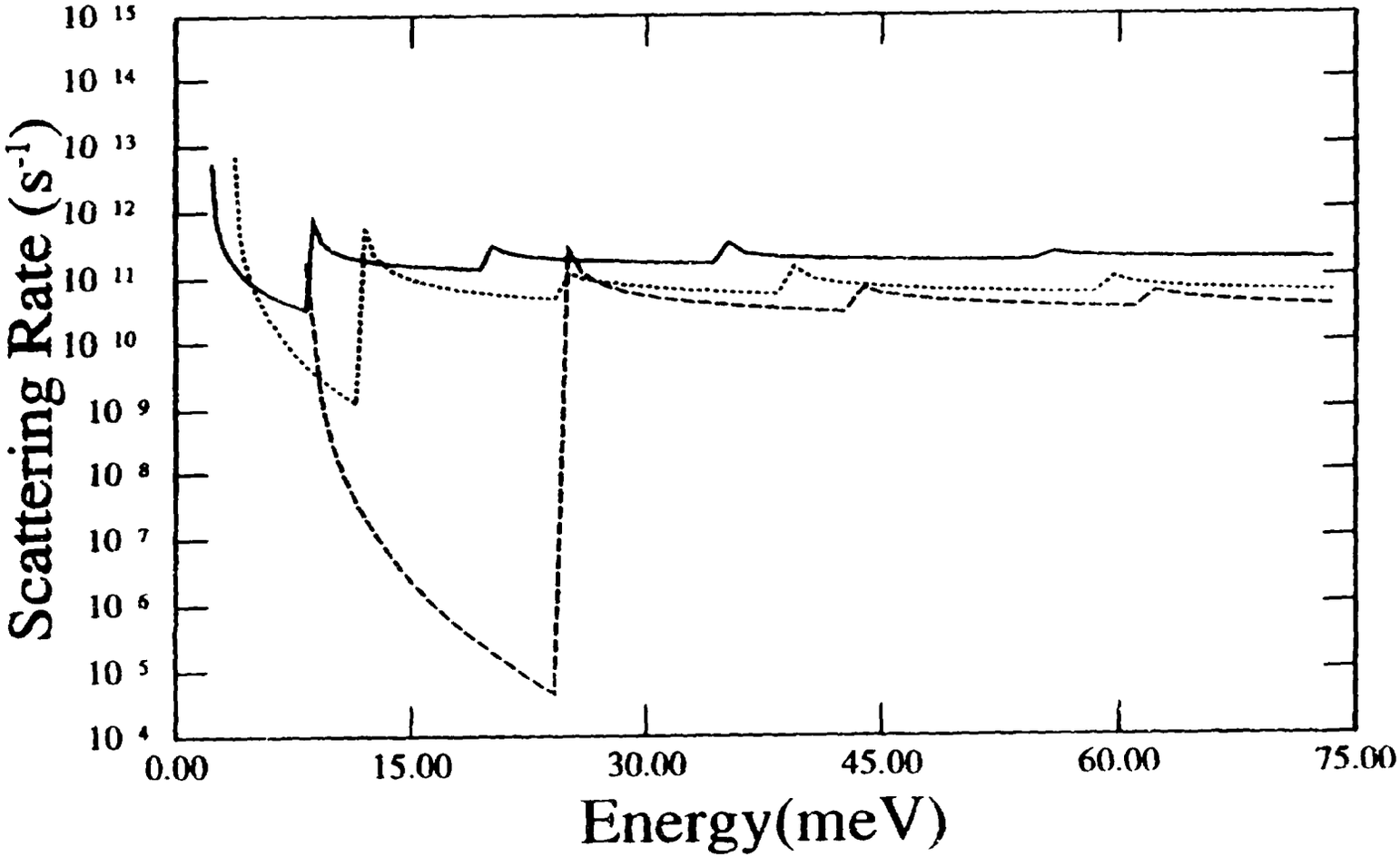
Figure 2: Electron-non polar acoustic phonon (deformation potential) scattering rate as a function of energy for electrons in the lowest magnetoelectric subband in a GaAs quantum wire. The width of the wire is 500 Å (along the y-direction) and the thickness is 40 Å (along the z-direction). The lattice temperature is assumed to be 300 K. The solid line represents no magnetic field, the shortdashed line a magnetic flux density of 4 tesla and the long dashed line a magnetic flux desnity of 10 tesla. (a) emission rate, and (b) absorption rate.

Figure 3: The y-component of the wavefunctions of two oppositely traveling states in the quantum wire at a magnetic flux density of (a) 0.05 tesla and (b) 10 tesla. The states are in the lowest magnetoelectric subband at an energy of 20 meV above bulk conduction band edge. The overlap between these two states is reduced significantly by the magnetic field which skews the wavefunctions towards opposite edges of the wire.


Scattering Rate (s^{-1})
10^{15}
10^{14}
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
0.00
15.00
30.00
45.00
60.00
75.00
Energy(meV)

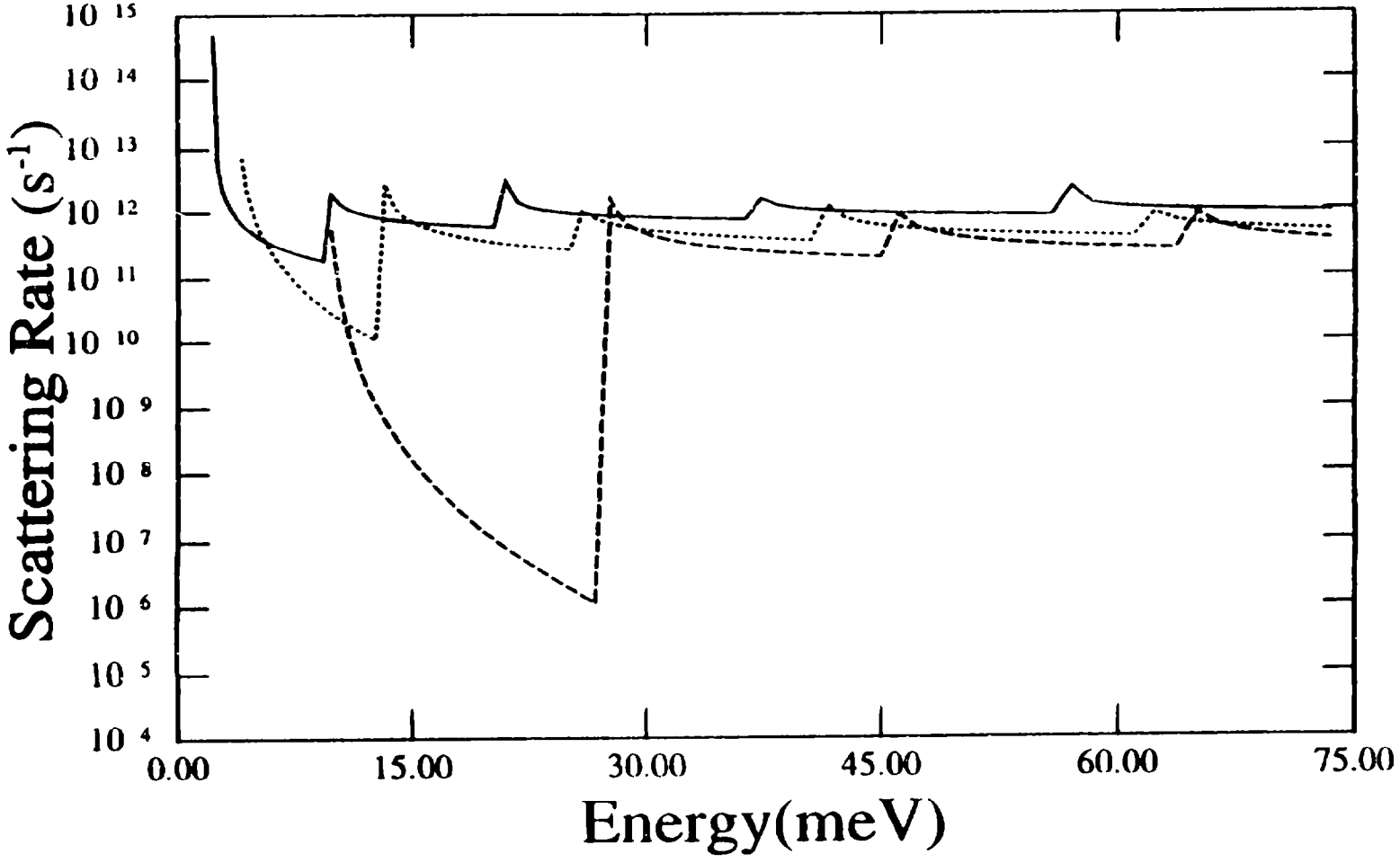
(a)

Fig 1a

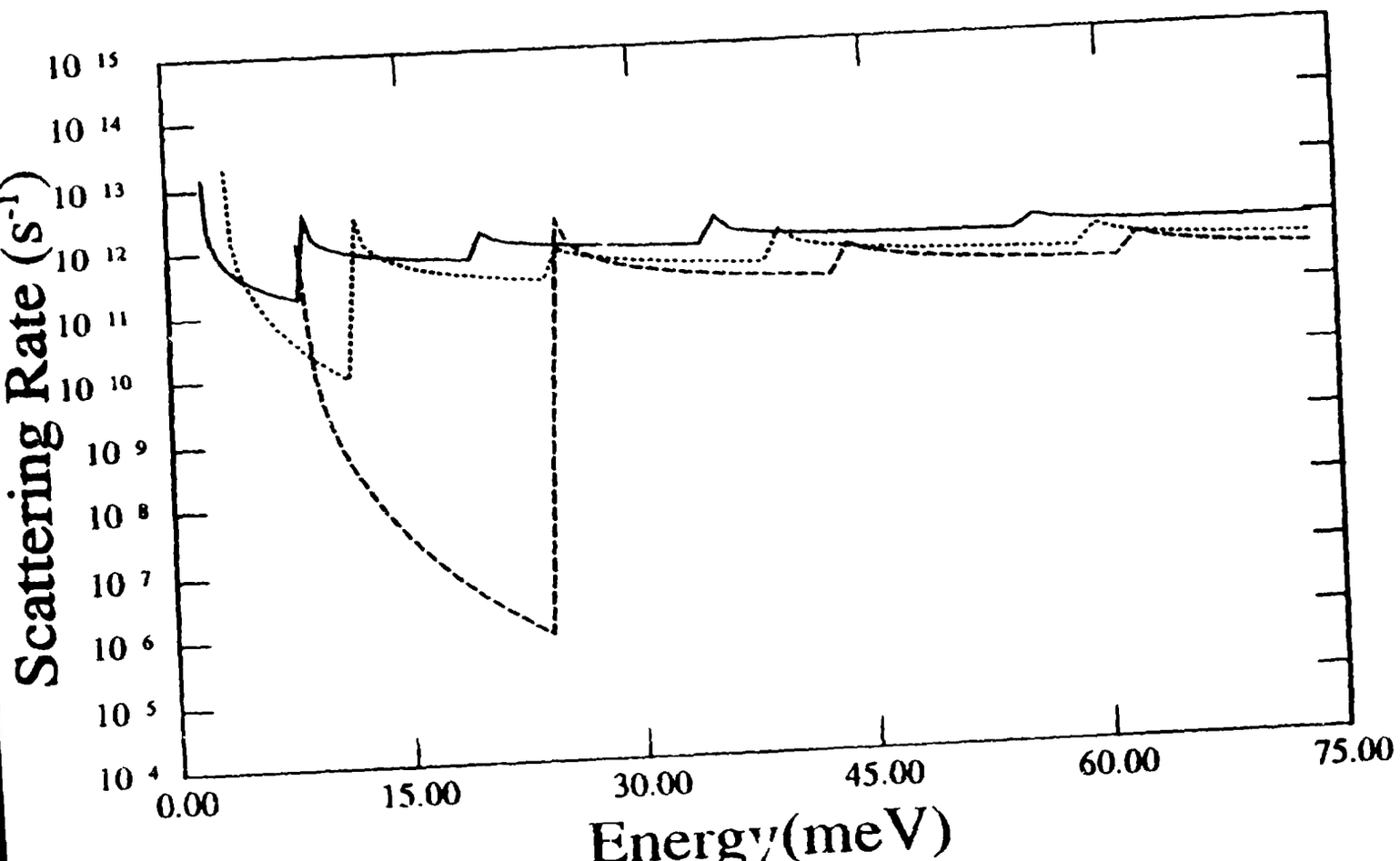

Scattering Rate (s^{-1})
10^{15}
10^{14}
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
0.00
15.00
30.00
45.00
60.00
75.00
Energy(meV)

(b)

fig 1b

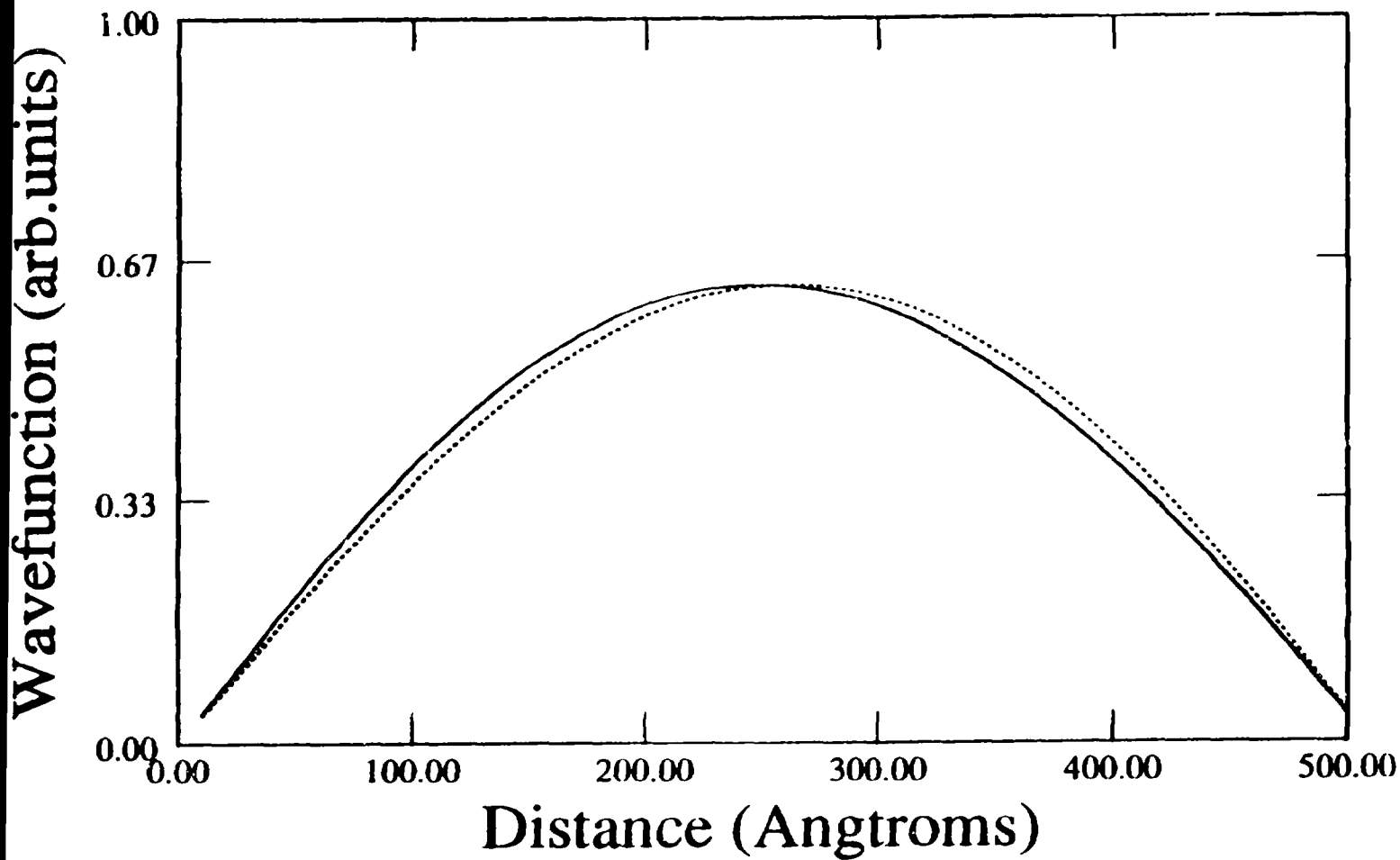

Scattering Rate (s^{-1})
10^{15}
10^{14}
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
0.00
15.00
30.00
45.00
60.00
75.00
Energy(meV)

(a)

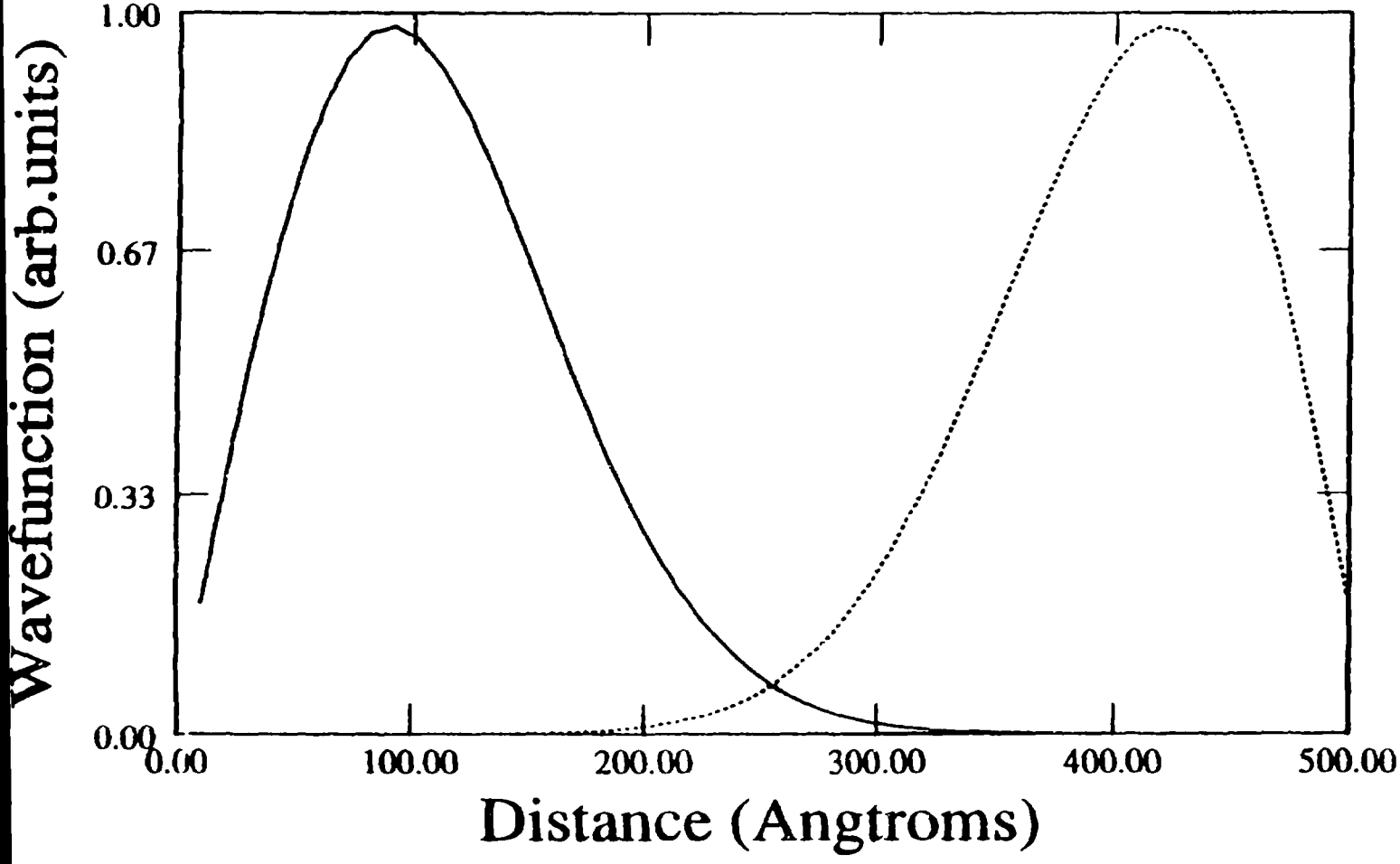

Scattering Rate (s^{-1})
10^{15}
10^{14}
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
0.00
15.00
30.00
45.00
60.00
75.00
Energy(meV)

(b)

1.9-6


1.00
0.67
0.33
0.00
Wavefunction (arb.units)
0.00
100.00
200.00
300.00
400.00
500.00
Distance (Angtroms)

(a)


Wavefunction (arb.units)
1.00
0.67
0.33
0.00
0.00
100.00
200.00
300.00
400.00
500.00
Distance (Angtroms)

(b)

Submitted to PRB

The effects of a magnetic field on electron-phonon scattering in quantum wires

N. Telang and S. Bandyopadhyay
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

We have calculated electron-phonon scattering rates in semiconductor quantum wires subjected to an external magnetic field. A magnetic field has several interesting effects on the scattering rates. It drastically reduces acoustic phonon scattering, but increases longitudinal optical and surface optical phonon scattering rates. The decrease in the acoustic phonon scattering has important implications for the integer quantum Hall effect. In addition, we have found that a magnetic field significantly enhances the difference between the scattering rates at energies just below and above a subband minimum and this may cause negative differential mobility to appear in quantum wires at electric fields far below the threshold for intervalley transfer. We have also observed other interesting features associated with one-dimensional confinement of electrons and phonons.

[1] PACS Indices: 72.10.Di, 72.20.Dp, 72.20.Fr, 72.20.Ht

I. Introduction

Recently, there has been a growing interest in the study of electron scattering in quasi one-dimensional structures[1,2,3] motivated by the belief that these structures can exhibit exceptionally high mobilties at reduced temperatures. Field effect transistors with quantum wire channels have been fabricated and show very large transconductances as a result of this enhanced mobility[4]. The increase in mobility accrues primarily from a suppression of impurity scattering which is the dominant scattering mechanism at cryogenic temperatures and low electric fields. However, at room temperature and above, or at high electric fields, scattering in semiconductor quantum wires is mostly due to phonons. Phonon scattering not only determines the low field mobility at room temperature, but it also determines the high field saturation velocity of electrons, the homogenoeous linewidth broadening of optical transitions, the relaxation rate for photoexcited carriers and a host of other phenomena in quantum wires.

In this paper, we have rigorously calculated electron phonon scattering rates (involving all types of phonon modes) in quantum wires using Fermi's Golden Rule. The application of this rule in quantum wires has been criticized in the past[5], but the criticism is valid only for electron energies at a subband bottom. At these energies, the density of states in quasi one-dimensional systems diverges which makes the Fermi's Golden Rule prescription (or more correctly, the Born approximation) invalid. Attempts at circumventing this problem by introducing arbitrary broadening of the density of states (presumably associated with surface roughness) have been reported in the literature[5]. We do not adopt this approach in our work since it is quite ad hoc; instead, we compute the scattering rates from the usual Fermi's Golden Rule, but with the caveat that it is not valid for electron energies corresponding to subband minima.

Fermi's Golden Rule based calculation of electron-phonon scattering rates in quantum wires have been reported by a number of researchers in the past. Leburton et. al.[7,8,9] have calculated electron-longitudinal polar optical phonon scattering rates in one-dimensional structures but without considering phonon confinement effects (the electrons are confined, but the phonon modes are asumed to be bulk modes). Recent experimental results, however, have revealed that phonon confinement may be important in quantum wires. Signatures of *surface* modes in cylindrical wires[10] and *confined* optical modes[11] in rectangular wires have been observed. Phonon confinement effects were explicitly taken into account in calculating optical phonon scattering rates by some other researchers recently[12–15]. It turns out (not quite unexpectedly) that the scattering rates calculated by assuming bulk modes actually do not differ greatly from the sumtotal of scattering rates due to surface and confined optical modes. Presumably, this is because the confined optical modes have nulls at the wire surfaces, whereas the surface modes peak at the surfaces, so that the superposition of both modes looks approximately bulklike[16]. As a result, phonon confinement effects are generally not of paramount importance in electron-optical phonon interaction in quantum wires.

While the above is true generally, there is one serious exception. That case corresponds to the situation when a magnetic field is present. *In a magnetic field, phonon confinement effects assume an added importance.* In fact, it is *because of* phonon confinement that a magnetic field has any effect at all on *optical* phonon scattering rates. If the optical phonon modes were unconfined bulk modes, *all such effects would be absent.* Therefore, any experimental manifestation of a dependence of optical phonon scattering rates in quantum wires on an external magnetic field is effectively a demonstration of phonon confinement as well.

While optical phonons are confined, acoustic phonons, by their very na-

ture, are almost always unlike. Nonetheless, a magnetic field has a significant effect on acoustic phonon scattering as well and this effect is not linked to phonon confinement. Interestingly, the effect on acoustic phonon scattering is actually much more dramatic than the effect on optical phonon scattering. The optical phonon scattering rate increases by less than an order of magnitude at magnetic field strengths achievable in a superconducting magnet ($\sim$ 10 tesla), but the scattering rate due to acoustic phonons can decrease by *several orders of magnitude* at the same field strength. This has important implications for the integer quantum Hall effect, particularly with respect to the breakdown of this effect at high current densities. The quantum Hall effect is manifested at temperatures where acoustic phonon scattering is the dominant phonon scattering event[1] and the breakdown of this effect is associated with acoustic phonon emission. A magnetic field also has other interesting effects. For instance, it can cause negative differential mobility to occur in a quantum wire at electric fields far below the threshold for intervalley transfer. This is associated with the Riddoch-Ridley mechanism.

This paper is organized as follows. In section II we will discuss the calculation of phonon scattering rates in a quantum wire subjected to a magnetic field. The results will be presented along with the appropriate interpretations in section III. Finally, in section IV, we will present the conclusions.

II. Theory

Acoustic phonon scattering in a magnetic field

We consider a quantum wire as shown in Fig. 1. Only one transverse subband is occupied along the z-direction (for even the highest energy an electron can reach), but many are occupied in the y-direction. A magnetic field is applied along the z-direction so that the electronic states are hybrid magnetoelectric states. In all calculations, we will assume that the phonons are unaffected by the magnetic field and are in thermodynamic equilibrium so that the phonon occupation probability is given by the Bose Einstein factor.

The acoustic phonon scattering rate for an electron with energy E_ν in the hybrid magnetoelectric subband ν to an energy state $E'_{\nu'}$ in subband ν' is given by Fermi's Golden Rule

$$S(E_\nu, E'_{\nu'}, \pm\vec{q}) = \frac{2\pi}{\hbar}|M(E_\nu, E'_{\nu'}, \pm\vec{q})|^2\delta(E'_{\nu'} - E_\nu \pm \hbar\omega_{\vec{q}}) , \qquad (1)$$

where $M(E_\nu, E'_{\nu'}, \pm\vec{q})$ is the matrix element for the transition due to an acoustic phonon of energy $\hbar\omega_{\vec{q}}$ and wavevector $\vec{q}$. It should be noted that the subband indices refer to y-directed subbands only since the confinement in the z-direction is complete. The $\pm$ sign in the delta function represents emission(+) or absorption(-) of an acoustic phonon respectively. For acoustic phonons, we assume a linear dispersion relation $\hbar\omega_{\vec{q}} = \hbar v_s q_x$ where v_s is the longitudinal velocity of sound (electrons only interact with longitudinal phonons).

The matrix element can be written as

$$M(E_\nu, E'_{\nu'}, \pm\vec{q}) = V^{\pm}\delta_{k-k'\pm q_x}\int_{-L_y/2}^{L_y/2}\psi^*_{E'_{\nu'}}(y)\psi_{E_\nu}(y)e^{\pm iq_y y}dy\int_{-L_z/2}^{L_z/2}\phi^*_{E'_{\nu'}}(z)\phi_{E_\nu}(z)e^{\pm iq_z z}dz , \qquad (2)$$

where $\psi_{E'_{\nu'}}(\phi_{E_{\nu'}})$ and $\psi_{E_\nu}(\phi_{E_\nu})$ are the transverse y- (z-) components of the electron wavefunction of the final and initial magnetoelectric states. k' and k

are the longitudinal electron wavevectors of these states. $V^{\pm}$ is the interaction potential and the δ is the Kronecker delta which represents momentum conservation. We have assumed the acoustic phonon modes to be bulk modes (plane waves) unaffected by the magnetic field.

In the case of piezoelectric (or polar acoustic phonon) scattering, the interaction potential $V^{\pm}$ is given by[18]

$$V^{\pm} = i\left(\frac{\hbar e^2 e_{pz}^2 (N_q + \frac{1}{2} \pm \frac{1}{2})}{2\Omega\rho v_s \epsilon q}\right)^{1/2} \quad . \tag{3}$$

where e is the electronic charge, e_{pz} is the piezoelectric constant, ϵ is the dielectric constant, Ω is the normalizing volume in the phonon Brillouin zone, ρ is the mass density and N_q is the phonon occupation probability. The plus sign refers to emission and the minus sign to absorption.

In the case of non-polar acoustic (deformation potential) scattering, the interaction potential $V^{\pm}$ is given by[18]

$$V^{\pm} = i\left(\frac{\hbar D_a^2}{2\Omega\rho v_s} q (N_q + \frac{1}{2} \pm \frac{1}{2})\right)^{1/2} \quad . \tag{4}$$

where D_a is the acoustic deformation potential.

The total scattering rate $1/\tau_{ac}$, for acoustic emission or absorption, is obtained by integrating over all possible final energy states.

$$\frac{1}{\tau_{ac}^{\pm}(E_\nu)} = \int_0^\infty \int_0^{q_{max}} dE'_{\nu'} d^3\vec{q} D(E'_{\nu'}) S(E_\nu, E'_{\nu'}, \pm\vec{q}) \tag{5}$$

where q_{max} is the maximum phonon wavevector and $D(E'_{\nu'})$ is the density of hybrid magnetoelectric states at the final energy $E'_{\nu'}$ given by the well-known relation

$$D(E'_{\nu'}) = (2/\pi) / (\partial E'_{\nu'}/\partial k) \quad . \tag{6}$$

To calculate the above expression, we need to know the eigenenergies E_ν, wavefunctions ψ_{E_ν} and density of hybrid magnetoelectric states $D(E_\nu)$. These are calculated exactly by solving the Schrödinger equation in a quantum wire subjected to a magnetic field. This is accomplished using a finite difference scheme that we have described in a previous publication[19].

Optical phonon scattering in a magnetic field

To calculate polar and non-polar longitudinal optical (LO) as well as surface optical (SO) phonon scattering rates, we have followed Stroscio[12] and Kim et. al.[14] whose models assume confined (slab) phonon modes. A more accurate model would require calculation of the phonon modes from a microscopic model[15,20,21]. This is reserved for future work.

The longitudinal polar optical phonon scattering rate of an electron with energy E_ν in the νth magnetoelectric subband is given by[20,21]

$$\frac{1}{\tau_{POP}(E_\nu)} = \sum_{\nu'=1}^{\nu_{max}} \frac{e^2}{4\pi\epsilon_o}\omega_{POP}(N_{\omega_{POP}} + 1/2 \pm 1/2) I_{POP} D(E'_{\nu'}) \,. \qquad (7)$$

where ω_{POP} is the longitudinal polar optical phonon frequency, $N_{\omega_{POP}}$ is the phonon occupation probability (Bose-Einstein factor) and ν_{max} is the index of the highest phonon mode considered. Again, the plus sign in the above expression refers to emission and the minus sign to absorption. The quantity $D(E'_{\nu'})$ is the density of states at the final energy $E'_{\nu'}$ in the magnetoelectric subband ν' and the quantity I_{POP} is given by

$$I_{POP} = \left(\frac{1}{\epsilon_r(\infty)} - \frac{1}{\epsilon_r(0)}\right)\frac{(2\pi)^2}{L_y L_z}\sum_{m'=1,2,3..}\sum_{n'=1,2,3..}\left\{\frac{4P_{m'n'}}{\left[k_x^2 + (\frac{m'\pi}{L_y})^2 + (\frac{n'\pi}{L_z})^2\right]^{1/2}}\right\}^2 , \qquad (8)$$

where $\epsilon_r(0)$ and $\epsilon_r(\infty)$ are the low- and high-frequency relative permittivities, k_x is the phonon wavevector along the length of the quantum wire, m' and n' are the transverse phonon mode indices along the y- and z-directions, L_y and L_z are the width and thickness of the quantum wire and $P_{m'n'}$ is the overlap integral which can be written as

$$P_{m'n'} = \int_0^{L_y}\int_0^{L_z} \frac{dy}{L_y/2}\frac{dz}{L_z/2} v^*_{fin}(y) w_{in}(y) \phi^*_{fin}(z) o_{in}(z) \sin\left(\frac{m'\pi y}{L_y}\right)\sin\left(\frac{n'\pi z}{L_z}\right) , \qquad (9)$$

In the above, $v_{in}(y)$ ($o_{in}(z)$) and $w_{fin}(y)$ ($\phi_{fin}(y)$) are the initial and final y- (z-) components of the wavefunctions of the electron in the presence of

a magnetic field as before and the sine functions are the confined phonon (slab) modes.

The non-polar longitudinal optical phonon scattering rate is given by[18]

$$\frac{1}{\tau^{\pm}_{NPO}(E_\nu)} = \sum_{\nu'=1} \frac{\pi D_o^2}{\rho \omega_{NPO} L_y L_z}(N_{\omega_{NPO}} + 1/2 \pm 1/2) I_{NPO} D(E'_{\nu'}) \,. \qquad (10)$$

where D_o is the deformation potential for the optical phonon, ω_{NPO} is the frequency of the non-polar longitudinal optical phonon and I_{NPO} is the overlap integral given by

$$I_{NPO} = \sum_{m'=1,2,3..} \sum_{n'=1,2,3.} \left(\int_0^{L_y} \int_0^{L_z} \frac{dy}{L_y/2} \frac{dz}{L_z/2} \psi^*_{f\nu}(y) \psi_{in}(y) \phi^*_{f\nu}(z) \phi_{in}(z) \right) \sin\left(\frac{m'\pi y}{L_y}\right) \sin\left(\frac{n'\pi z}{L_z}\right) \Bigg)^2 \,. \qquad (11)$$

The scattering rate due to surface optical (SO) phonons is given by[14],

$$\frac{1}{\tau^{\pm}_{SO}(E_\nu)} = \sum_{\nu'=1}^{\nu' max} \frac{e^2}{4\pi\epsilon_o} \omega_{SO}(N_{\omega_{SO}} + 1/2 \pm 1/2) I_{SO} D(E'_{\nu'}) \,. \qquad (12)$$

where ω_{SO} is the surface optical phonon frequency and

$$I_{SO} = \left(\frac{2\pi C' P_s}{\omega_{SO}}\right)^2 \,. \qquad (13)$$

with C' being the normalization constant[20] for phonon modes and P_s an overlap integral given by

$$P_s^s = \frac{1}{\cosh(\alpha L_y/2)\cosh(\beta L_z/2)} \int_{-L_y/2}^{L_y/2} \int_{-L_z/2}^{L_z/2} \frac{dy}{L_y/2} \frac{dz}{L_z/2} \psi^*_{f\nu}(y) \psi_{in}(y) \phi^*_{f\nu}(z) \phi_{in}(z) \cosh(\alpha y)\cosh(\beta z), \qquad (14)$$

for symmetric surface phonon modes, and

$$P_s^a = \frac{1}{\sinh(\alpha L_y/2)\sinh(\beta L_z/2)} \int_{-L_y/2}^{L_y/2} \int_{-L_z/2}^{L_z/2} \frac{dy}{L_y/2} \frac{dz}{L_z/2} \psi^*_{f\nu}(y) \psi_{in}(y) \phi^*_{f\nu}(z) \phi_{in}(z) \sinh(\alpha y)\sinh(\beta z). \qquad (15)$$

for antisymmetric surface phonon modes.

The quantities α, β have been defined and the dispersion relation for SO phonons given in Ref. 14.

In our example, only one subband is occupied in the z-direction. In that case, the SO phonon scattering is due to only the symmetric modes since the overlap integral in the z-direction becomes zero for antisymmetric modes. This happens since the z-component of the wavefunction (in the lowest subband) has even parity. Had multiple modes been occupied in the z-direction (some of which would have odd parity), the SO phonon scattering would have had contributions from both the symmetric and antisymmetric phonon modes. This feature is not affected by a magnetic field as long as the field is directed along the z-direction.

In this section, we have provided analytical expressions for the scattering rates. We now present some results for a prototypical quasi one-dimensional structure.

III. Results and Discussion

We consider a GaAs quantum wire (surrounded by AlGaAs) of width 500 Å (along the y-direction) and thickness 40 Å (along the z-direction). All phonon scattering rates are calculated for this system.

Acoustic phonon scattering rates

In Figures 2(a) and 2(b), we show the polar acoustic phonon (piezoelectric) emission and absorption rates as a function of electron energy in the lowest magnetoelectric subband. The results are shown for various magnetic field strengths. In Figures 3(a) and 3(b), we show the non-polar acoustic phonon (deformation potential) emission and absorption rates, again at different magnetic field strengths. The peaks in the scattering rates are associated with the divergence in the density of (final) states in a quasi one-dimensional structure.

The salient feature to note is that there is a dramatic quenching of the *intra-subband* scattering rate caused by a magnetic field. The rate decreases by *six orders of magnitude* at a magnetic flux density of 10 tesla. The inter-subband scattering rate also decreases, but the decrease is much less dramatic. In narrow enough wires, where only one transverse subband is occupied in both the y- and the z-direction (for even the highest energy an electron can reach), the only allowed scattering is intra-subband scattering since higher subbands are not accessible in energy. In such wires, a magnetic field can cause dramatic quenching of acoustic phonon scattering. This effect can then cause large negative magnetoresistance in a narrow wire, especially at temperatures where acoustic phonon scattering is the dominant scattering mechanism. Such a phenomenon has applications in magnetic field sensors, magnetic recording heads, etc.

Another interesting feature seen in Figs. 2 and 3 is the very large difference between the scattering rates at energies just below and above the first subband minimum. This difference is not large in the absence of a magnetic field, but becomes *very large in the presence of a magnetic field.* Even in the absence of a magnetic field, such a difference was once considered capable of triggering negative differential mobility in quantum confined systems (Riddoch-Ridley mechanism)[22]. Later simulations failed to reveal this effect in GaAs wires[3], but we believe that a magnetic field may cause it to appear. Only a complete Monte Carlo simulation of electron transport or some other solution of the Boltzmann Transport Equation can answer this question satisfactorily. This is important to know since (a) the threshold electric field for such negative differential mobility is presumably much lower than that associated with intervalley transfer (Ridley-Gunn-Hilsum effect) which makes it possible to realize low power microwave oscillators, and (b) the threshold electric field can be engineered at will by altering the dimensions (width and thickness) of the wire.

We now explain the origin of the magnetic field induced quenching of the intra-subband acoustic phonon scattering rate. Because of the simultaneous conservation of both energy and momentum in the scattering process, there are certain restrictions on the angle of scattering. For *intra-subband* scattering, primarily backward (large angle) scattering is allowed (in contrast to the case for polar optical phonon and impurity scattering, or even, as we shall see later, for *inter-suband* acoustic phonon scattering). This means that the initial and final states of the electron (i.e the states before and after scattering) will tend to have oppositely directed velocities. In the absence of a magnetic field, two such states will have a large wavefunction overlap (essentially 100 %) so that the matrix element for scattering (Equation 2) is large. However, when a magnetic field is applied in the z-direction, the y-components of the wavefunctions $\psi_{E_y}(y)$ and $\psi_{E_y}(y)$ of two such states will be skewed in opposite directions since the Lorentz force acts in opposite directions on oppositely traveling electrons. This decreases the overlap between the wavefunctions, and therefore the matrix element, which causes the quenching. To illustrate this effect, we have plotted in Fig. 4, the wavefunction of two oppositely traveling states (in the lowest magnetoelectric subband) at a magnetic flux density of 0.05 and 10 tesla. Note that the wavevefunctions are skewed towards opposite edges of the wire by the magnetic field which drastically reduces the overlap between them and therefore the scattering matrix element given by Equation (2). This explains the quenching[23].

The less pronounced decrease in the inter-subband rate has a compietely different origin. It can be understood as follows. The energy separation between hybrid magnetoelectric subbands increases with increasing magnetic field. Therefore, at higher magnetic fields, one needs higher energy (and therefore larger wavevector) phonons to effect inter-subband transitions. However, the scattering rate for higher energy phonons is less than that for lower energy phonons because of two reasons: (a) the phonon oc-

upation probability (Bose Einstein factor) decreases exponentially with increasing phonon energy, and (b) the interaction potential for scattering is inversely proportional to the phonon wavevector (and therefore phonon energy) in the case of polar acoustic phonon scattering (see Equation 3). In the case of non-polar acoustic phonon scattering, the interaction potential is directly proportional to the phonon wavevector (see Equation 4), but this linear increase is more than offset by the exponential decrease of the phonon occupation probability with increasing phonon energy. All this reduces the inter-subband scattering rate in the presence of a magnetic field. Recently, this effect was observed experimentally in a superlattice[24].

The quenching of acoustic phonon scattering by a strong magnetic field has serious implications for the integer quantum Hall effect. This effect is manifested at low temperatures under high magnetic field when acoustic phonon scattering is the dominant phonon scattering mechanism. In this regime, current in a wire is carried exclusively by "edges states"[25] that are localized along the edges of the wire. Edge states along opposite edges have opposite velocities and carry current in the opposite direction. In the light of the previous discussion, we can see that the probability of acoustic phonon scattering between oppositely traveling edge states (backscattering) is extremely low. This shows why backscattering is suppresed at high magnetic fields.

The suppression of backscattering is the central ingredient in the Büttiker picture for the origin of the integer quantum Hall effect[25]. In the past, we and others had demonstrated this suppresion in the quantum Hall regime considering only *elastic* (phase preserving) scattering events[26]. We have now demonstrated this suppression for *inelastic* (phase breaking acoustic phonon) events as well. This is a very important extension of the Büttiker picture beyond the coherent linear response transport regime (where only elastic scattering is permitted). The quantum Hall effect is known to survive at

finite voltages and currents when both elastic and inelastic scattering occur. This has been demonstrated theoretically[37] as well as experimentally[38]. Previous models attributed the breakdown of the quantum Hall effect at high current densities to the onset of acoustic phonon emission[39]. In that light, the suppression of acoustic phonon emission has important implications for the breakdown. It has also been claimed that the critical current for breakdown increases with increasing magnetic field. We believe that this may be due to increased suppression of acoustic phonon scattering with increasing field which is consistent with our result.

We have calculated the magnetic field dependences of the acoustic phonon scattering rates at various temperatures (4.2 K, 77 K and 300 K). We found that the relative decrease in the scattering rate is fairly insensitive to temperature. This is not surprising since the physical phenomena underlying the quenching are not temperature dependent.

Optical Phonon Scattering Rates

In Figures 5(a) and 5(b), we have plotted the longitudinal polar optical phonon emission and absorption rates as a function of electron energy in the lowest magnetoelectric subband for various magnetic field strengths. The non-polar optical phonon scattering rates are shown in Figs. 6(a) and 6(b). The latter are for the L-valley of GaAs since this scattering mechanism is forbidden in the Γ valley. Finally, the surface optical phonon (SO) scattering rates are shown in Figs. 7(a) and 7(b).

Unlike in the case of acoustic phonon scattering, all optical phonon scattering rates increase with magnetic flux density. This increase becomes more evident in the plots of the scattering rates versus magnetic field which are shown in Figures 5(c), 6(c) and 7(c).

The increase for both polar and non-polar longitudinal optical phonon scattering can be explained as follows. In the absence of a magnetic field, the

electron wavefunction in the ith mode is orthogonal to the jth phonon mode if $i \neq j$. This makes the integrals I_{POP} and I_{VPO} (and the corresponding scattering rates $1/\tau^{\pm}_{POP}$ and $1/\tau^{\pm}_{VPO}$) vanish for $i \neq j$. This means that the jth phonon mode will not contribute to the scattering of an electron in the ith subband. However, when a magnetic field is present, it skews the electron wavefunctions towards one edge of the wire (owing to the Lorentz force) and this breaks the orthogonality between the electron wavefunction in the ith mode and the jth phonon mode ($i \neq j$). As a result, all phonon modes now contribute to the scattering of an electron. This opens up many new scattering channels which were previously forbidden. The result is an increase in the total scattering rate when a magnetic field is present.

In the case of SO phonon, the increase in the scattering rate can be explained as follows. The magnetic field skews the electron wavefunctions towards one of the edges of the wire. This increases the overlap integral P_i since the SO phonon modes are localized at the edges and decay away from the edges. This effect is the dominant effect in a magnetic field. Consequently, the scattering rate increases.

Another interesting feature in Fig. 7(b) is that, in the absence of any magnetic field, one cannot observe the peaks in the absorption rate associated with the divergence of the one-dimensional density of (final) states. This was also noted in Ref. 6 and 14. It was claimed in Ref. 6 that this happens because the peaks are due to inter-subband scattering as opposed to intra-subband scattering. For SO phonon absorption, intra-subband scattering rate dominates over inter-subband scattering rate and therefore the peaks in the inter-subband rate are not discernible against the strong background of intra-suband scattering. However, when a magnetic field is turned on, the peaks appear (the two sets of peaks are associated with two phonon branches[14]). This indicates that a magnetic field *promotes inter-subband SO phonon absorption over intra-subband SO phonon absorption*. We believe

that this happens because the magnetic field breaks the orthogonality between the electron wavefunctions in two different subbands and increases the overlap integral P_z for inter-subband transitions. This is true of SO phonon emission as well. Therefore, a magnetic field brings out the peaks in the scattering rate.

IV. Conclusion

In this paper, we have investigated rigorously, for the first time, the effect of a magnetic field on phonon scattering in a quantum wire. We have demonstrated that a magnetic field can cause a dramatic reduction in intra-subband scattering due to acoustic phonons and a much less pronounced reduction in the inter-subband scattering. This has important implications for the quantum Hall effect. It may cause also negative differential mobility in a quantum wire at threshold electric fields far below that required for intervalley transfer. Finally, it may give rise to giant negative magnetoresistance in narrow enough wires at certain temperatures and electric field strengths which has applications in magnetic field sensors.

For optical phonon scattering, we found that the scattering rate increases with increasing magnetic field although the increase is nowhere near as dramatic as the decrease in the case of intra-subband acoustic phonon scattering. Optical phonon scattering determines the saturation velocity in most materials at electric fields below the threshold for intervalley transfer. It also determines the relaxation rate for photo-excited carriers and the homogeneous linewidth broadening of the photoluminsence spectra. Therefore, it is important to understand the effect of a magnetic field on both acoustic and optical phonon scattering rates. This work is an important step in that direction.

Acknowledgement: This work was supported by the Air Force Office of Scientific Research under grant number AFOSR-91-0211.

References

1. H. Sakaki, *Jpn. J. Appl. Phys.*, **19**, L735 (1980).

2. V. K. Arora, *Phys. Rev. B*, **23**, 5611 (1981).

3. A. Ghosal, D. Chattopadhyay and A. Bhattacharyya, *J. Appl.Phys.*, **5611**, 1981.

4. H. Sakaki (unpublished).

5. P. F. Bagwell, *Phys. Rev. B*, **43**, 9012 (1990); P. F. Bagwell, *ibid.* **41**, 10345 (1990).

6. R. Mickevicius, V. V. Mitin, K. W. Kim, M. A. Stroscio and G. J. Iafrate, (unpublished).

7. S. Briggs and J. P. Leburton, *Phys. Rev. B*, **38**, 8163 (1988).

8. D. Jovanovich, S. Briggs and J. P. Leburton, *Phys. Rev. B*, **42** 11108 (1990); S. Briggs, D. Jovanovich and J. P. Leburton, *Appl. Phys. Lett.*, **54**, 2012 (1989).

9. D. Jovanovich and J. P. Leburton, *Monte Carlo Simulation: Full Band and Beyond*, ed. K. Hess, (Kluwer Academic Publishers, Boston, 1991).

10. M. Watt, C. M. Sotomayer-Torres, H. E. G. Arnot and S. P. beaumont, *Semiconduc. Sci. Technol.*, **5**, 285 (1990); K. T. Tsen, K. R. Wald, T. Ruf, P. Y. Yu and H. Morkoc, *Phys. Rev. Lett.*, **67**, 2557 (1991).

11. G. Fasol, M. Tanaka, H. Sakaki and Y. Horikosh, *Phys. Rev. B*, **38**, 6065 (1988).

12. M. A. Stroscio, *Phys. Rev. B*, **40**, 6428 (1989).

13. M. A. Stroscio, K. W. Kim, M. A. Littlejohn and H. Chuang, *Phys. Rev. B*, **42**, 1488 (1990).

14. K. W. Kim, M. A. Stroscio, A. Bhatt, R. Mickevicius and V. V. Mitin, *J. Appl. Phys.*, **70**, 319 (1991). This reference only considers intra-subband scattering and not inter-subband scattering.

15. Bangfen Zhu (unpublished).

16. P. Y. Yu (unpublished).

17. *The Quantum Hall Effect*, eds. R. E. Prange and S. M. Girvin (Springer Verlag, Heidleberg, 1987).

18. B. K. Ridley, *Quantum Processes in Semiconductors*, (Oxford Science Publications, Oxford, 1988).

19. S. Chaudhuri and S. Bandyopadhyay, *J. Appl. Phys.*, **71**, 3027 (1992).

20. K. Huang and B. Zhu, *Phys. Rev. B*, **38**, 13377 (1988).

21. S. F. Ren and Y. C. Chang (unpublished).

22. F. A. Riddoch and B. K. Ridley, *Surf. Sci.*, **142**, 260 (1984).

23. N. Telang and S. Bandyopadhyay (unpublished).

24. A. Kastalsky and A. L. Efros, *J. Appl. Phys.*, **69**, 841 (1991).

25. M. Büttiker, *Phys. Rev. B*, **38**, 9375 (1988).

26. T. Martin and S. Feng, *Phys. Rev. Lett.*, **64**, 1971 (1990); J. J. Palacios and C. Tejedor, *Phys. Rev. B*, **44**, 8157 (1991); S. Bandyopadhyay, S. Chaudhuri, B. Das and M. Cahay, *Superlattices and Microstructures*, **12**, 123 (1992).

27. Y. Lee, M. J. McLennan, G. Klimeck, R. K. Lake and S. Datta, *Superlattices and Microstructures*, **11**, 137 (1992).

28. G. Ebert, K. von Klitzing, K. Ploog and G. Weimann, *J. Phys. C*, **16**, 5441 (1983); M. E. Cage, R. F. Dzuiba, B. F. Field, E. R. Williams, S. M. Girvin, A. C. Gossard, D. C. Tsui and R. J. Wagner, *Phys. Rev. Lett.*, **51**,

1374 (1983); F. Kuchar, G. Bauer, G. Weimann and H. Burkhard, *Surf. Sci.*, **142**, 196 (1984).

29. O. Heinonen, P. L. Taylor and S. M. Girvin, *Phys. Rev. B*, **30**, 3016 (1984); P. Středa and K. von Klitzing, *J. Phys. C*, **17**, L483 (1984); H. L. Störmer, A. M. Chang, D. C. Tsui and J. C. M. Hwang, *Proc. 17th. Intl. Conf. Semiconduc.*, eds D. J. Chadi and W. A. Harrison, (Springer Verlag, Heidleberg, 1985), p. 267.

FIGURE CAPTIONS

Figure 1: A quasi one-dimensional structure subjected to a magnetic field in the z-direction. The thickness is small enough that only one subband is occupied in the z-direction at all electron energies. Several subbands are occupied in the y-direction.

Figure 2 (a) Electron-polar acoustic phonon (piezoelectric) emission rate as a function of energy for electrons in the lowest magnetoelectric subband in a $GaAs$ quantum wire. (b) the absorption rate. Energy is measured from the bulk conduction band edge. The width of the wire is 500Å (along the y- direction) and the thickness is 40Å (along the z-direction). The lattice temperature is assumed to be 300 K. The solid line represents the rates in the absence of a magnetic field, the short-dashed line the rates in a magnetic flux density of 4 tesla and the dotted line the rates in a magnetic flux density of 10 tesla. (c) The scattering rate as a function of magnetic flux density. The upper frame is for emission and the lower for absorption. In the upper frame, the electron energy is assumed to be 37 meV and in the lower frame, it is 25 meV.

Figure 3 (a) Electron-non polar acoustic phonon emission rate as a function of energy for electrons in the lowest magnetoelectric subband in a $GaAs$ quantum wire. (b) the absorption rate. The structure is the same as that in Fig. 2. The solid, short-dashed and long-dashed lines represent tha same magnetic flux densites as in Fig. 2. (c) The scattering rate as a function of magnetic flux density. The upper frame is for emission and the lower for absorption. In the upper frame, the electron energy is assumed to be 37 meV and in the lower frame, it is 25 meV.

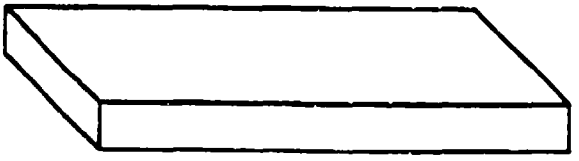
Figure 4: The y-component of the wave-functions of the two oppositely traveling states in the quantum wire at a magnetic flux density of (a) 0.05 tesla and (b) 10 tesla. The states are in the lowest magnetoelectric subband

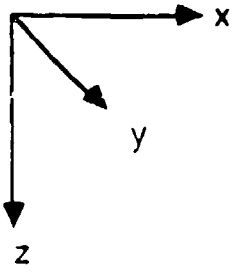
at an energy of $20meV$ above the bulk conduction band edge. The overlap between these two states is reduced significantly by the magnetic field which skews the magnetic field towards opposite edges of the quantum wire.

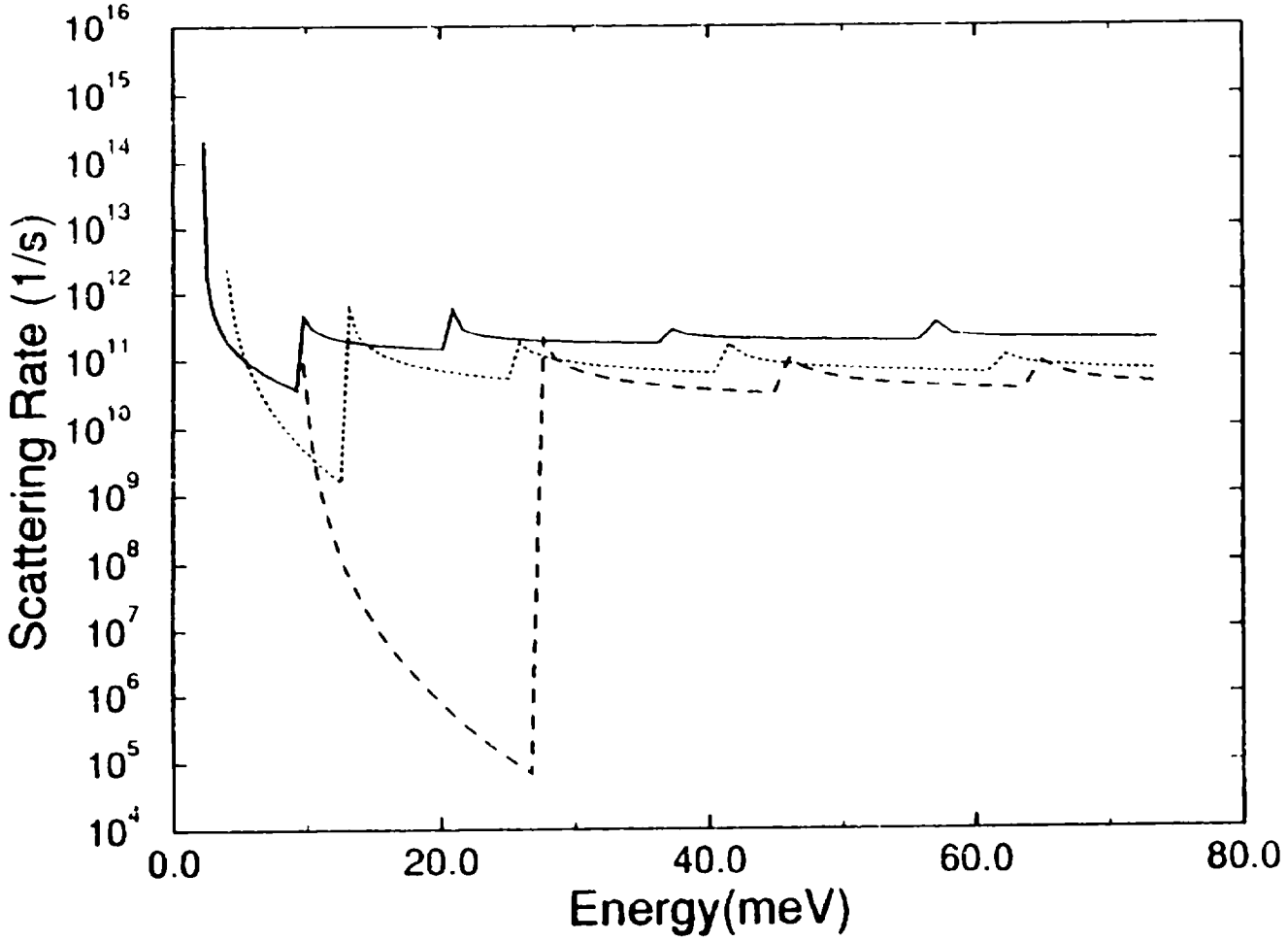
Figure 5: Electron-longitudinal polar optical phonon scattering rates as a function of energy for electrons in the lowest magnetoelectric subband. (a) emission rate and (b) absorption rate. Again, energy is measured from the bottom of the bulk conduction band edge. The solid, short-dashed and long-dashed lines have the same interpretation as in Fig. 2. (c) The upper frame is the emission and the lower frame is the absorption rate as a function of magnetic flux density. In the upper frame, the electron energy is 60 meV and in the lower frame it is 40 meV.

Figure 6: Electron-longitudinal non polar optical phonon scattering rates as a function of energy for electrons in the lowest magnetoelectric subband. (a) emission rate and (b) absorption rate. Again, energy is measured from the bottom of the bulk conduction band edge. The solid, short-dashed and long-dashed lines have the same interpretation as in Fig. 2. (c) The upper frame is the emission and the lower frame is the absorption rate as a function of magnetic flux density. In the upper frame, the electron energy is 45 meV and in the lower frame it is 10 meV.

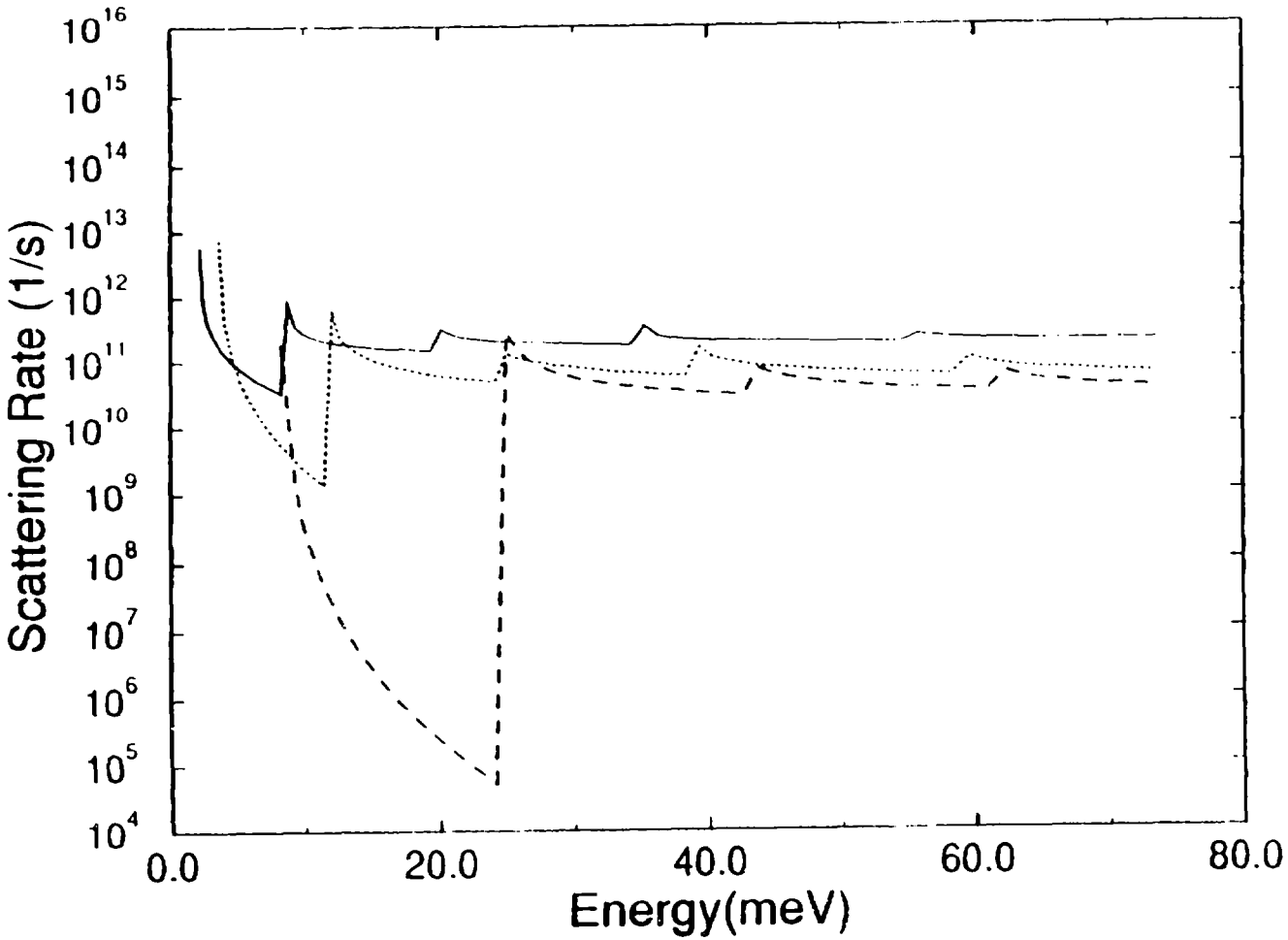
Figure 7: Electron-surface optical phonon scattering rates as a function of energy for electrons in the lowest magnetoelectric subband. (a) emission rate and (b) absorption rate. Again, energy is measured from the bottom of the bulk conduction band edge. The solid, short-dashed and long-dashed lines have the same interpretation as in Fig. 2. (c) The upper frame is the emission and the lower frame is the absorption rate as a function of magnetic flux density. In the upper frame, the electron energy is 60 meV and in the lower frame it is 40 meV.




x
y
z

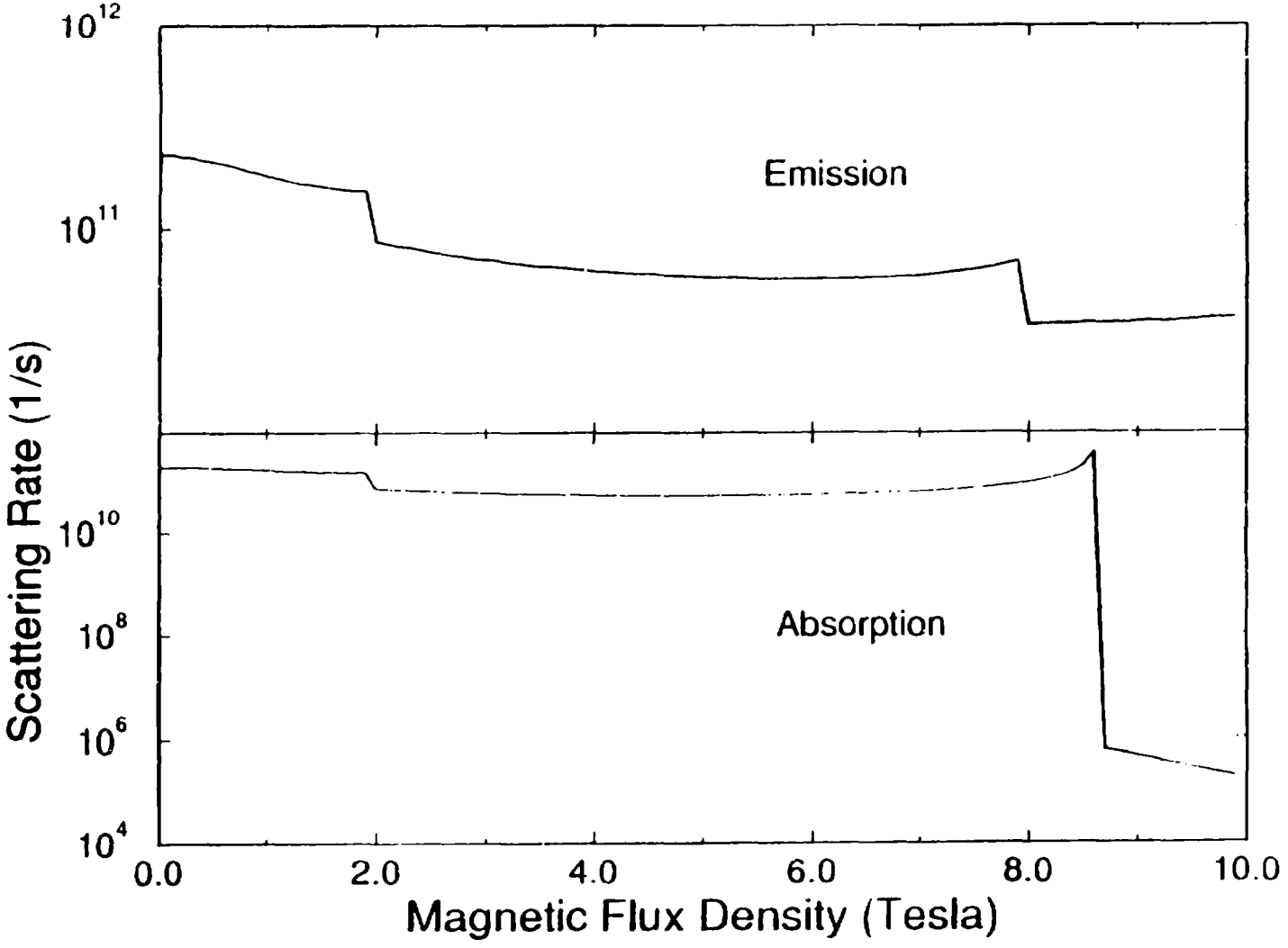

Scattering Rate (1/s)
10^{16}
10^{15}
10^{14}
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
0.0
20.0
40.0
60.0
80.0
Energy(meV)

(a)


Scattering Rate (1/s)
10^{16}
10^{15}
10^{14}
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
0.0
20.0
40.0
60.0
80.0
Energy(meV)

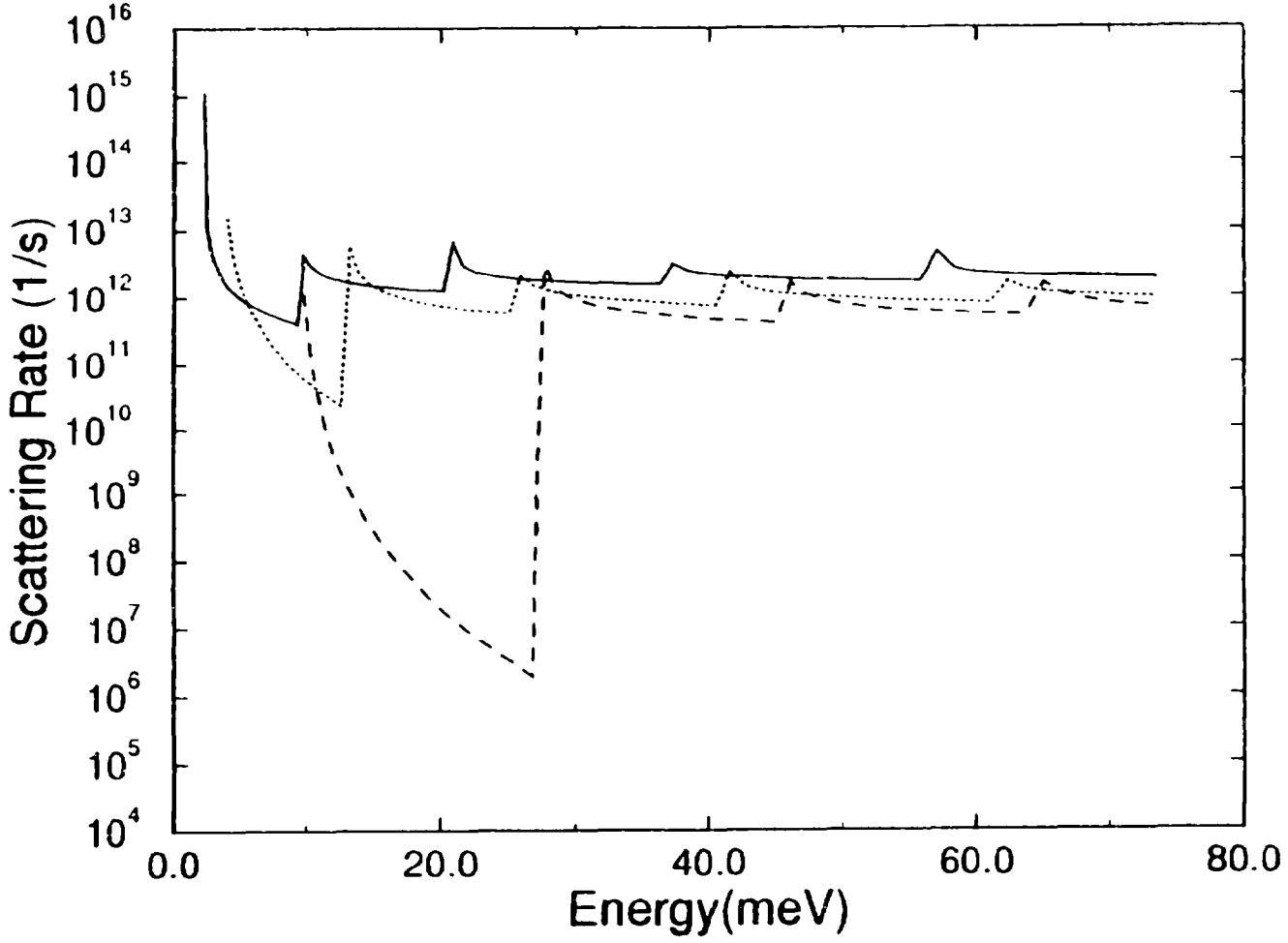
(b)

Fig 2(b)


10^{12}
10^{11}
10^{10}
10^{8}
10^{6}
10^{4}
Scattering Rate (1/s)
Emission
Absorption
0.0
2.0
4.0
6.0
8.0
10.0
Magnetic Flux Density (Tesla)

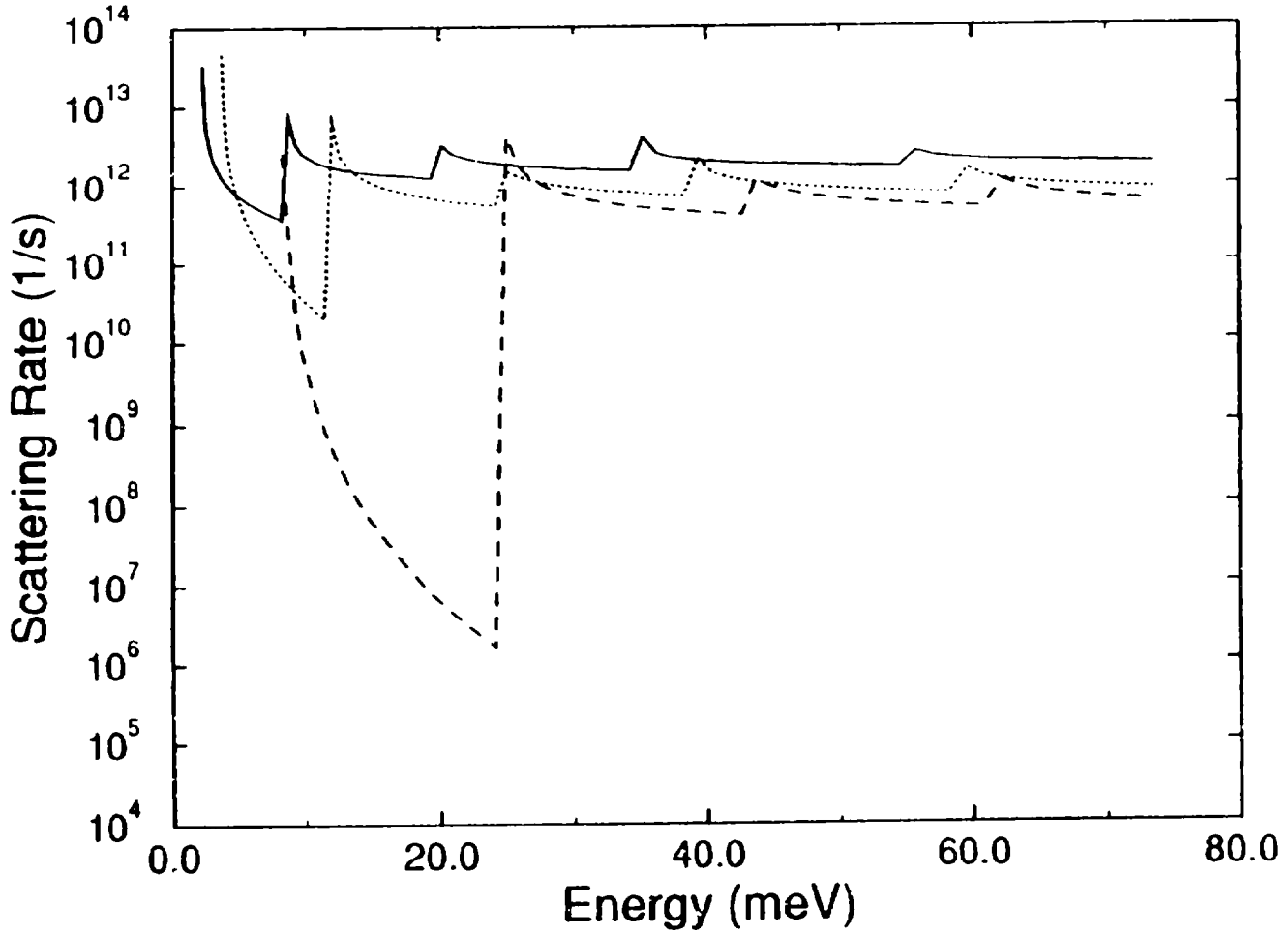
(c)

Fig 2(c)


Scattering Rate (1/s)
10^{16}
10^{15}
10^{14}
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
0.0
20.0
40.0
60.0
80.0
Energy(meV)

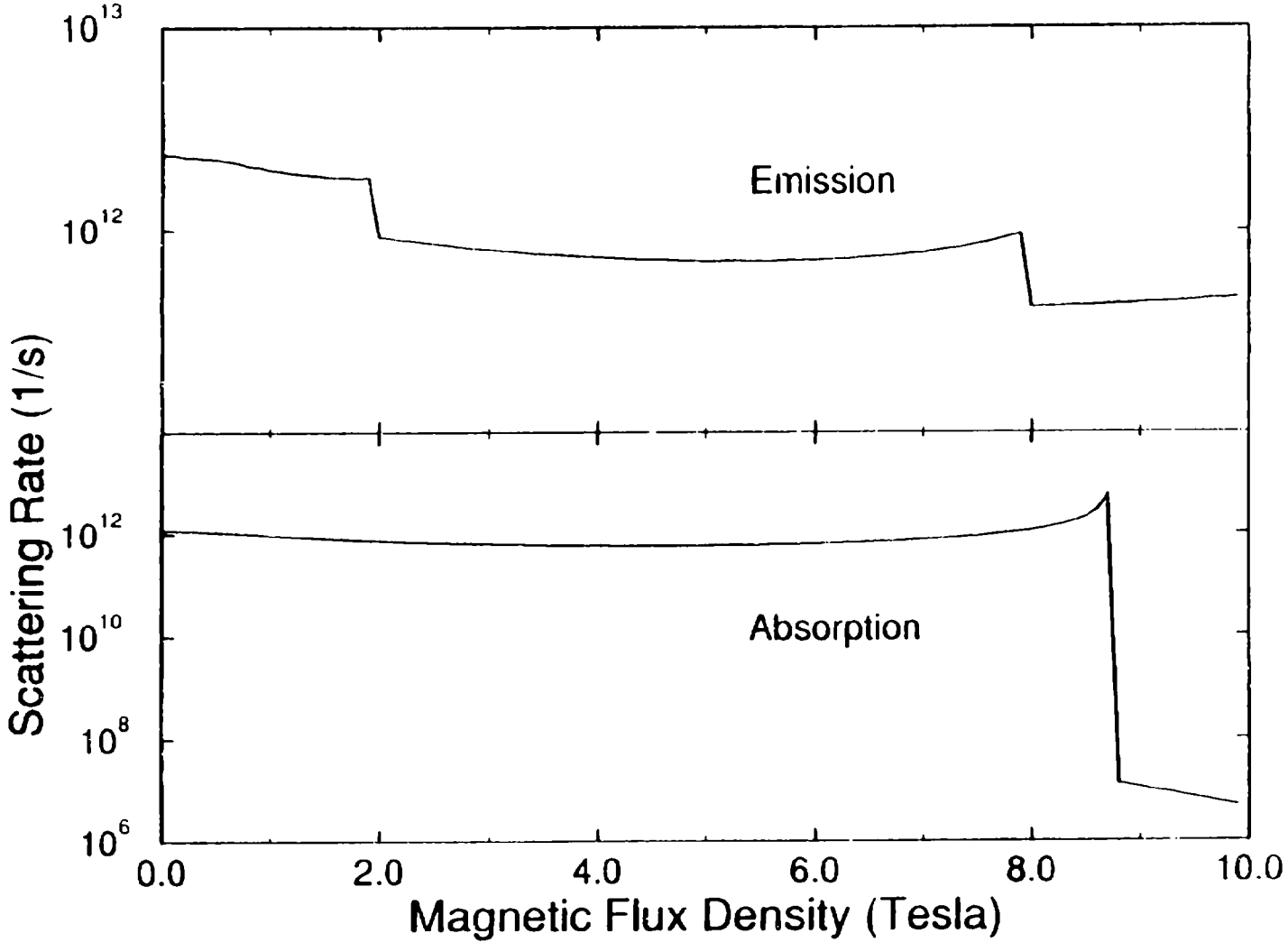
(a)

Fig 3(a)


Scattering Rate (1/s)
10^{14}
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
0.0
20.0
40.0
60.0
80.0
Energy (meV)

(b)

Fig 3(b)


Scattering Rate (1/s)
10^{13}
10^{12}
10^{12}
10^{10}
10^{8}
10^{6}
Emission
Absorption
0.0
2.0
4.0
6.0
8.0
10.0
Magnetic Flux Density (Tesla)

(c)

Fig 3(c)

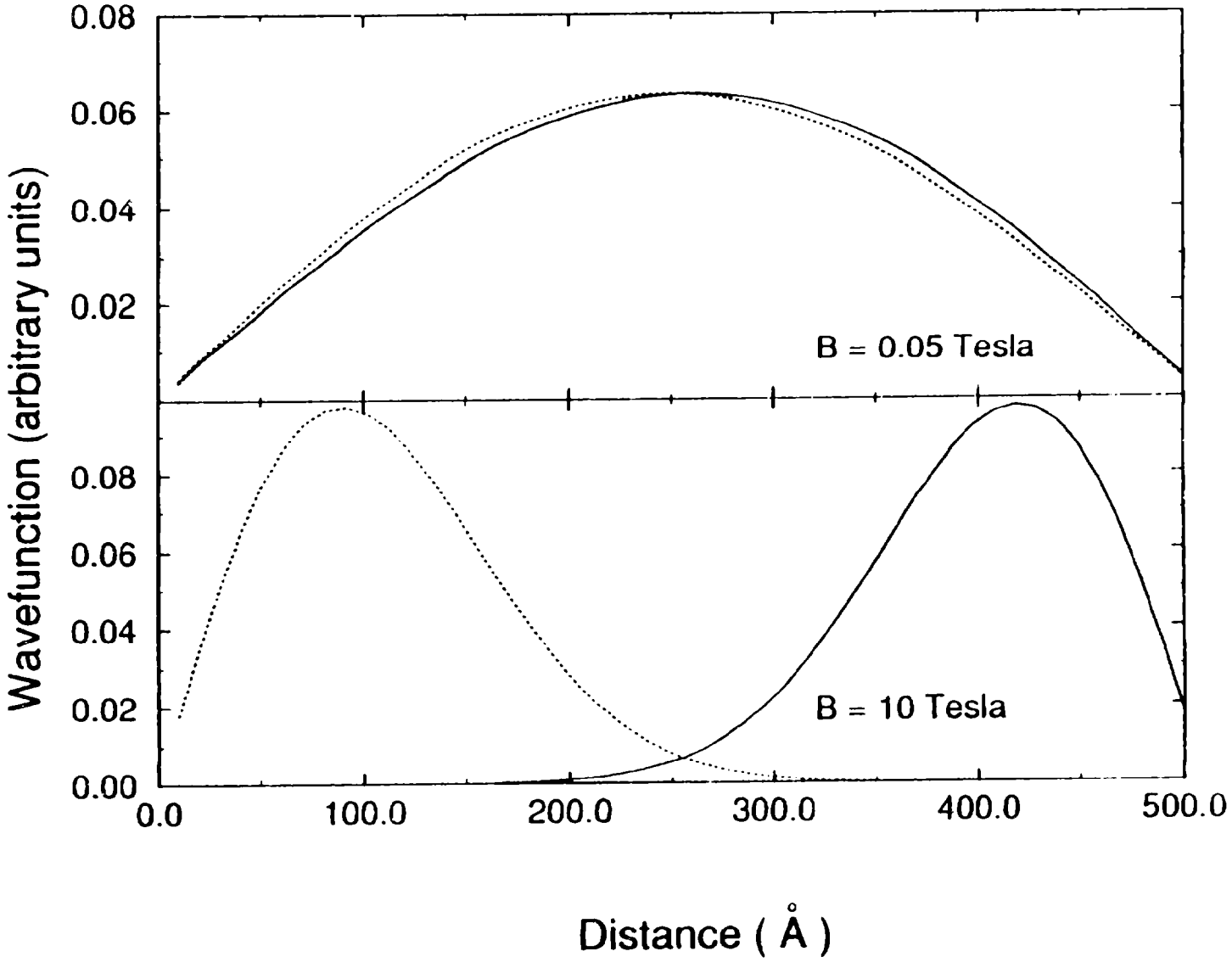
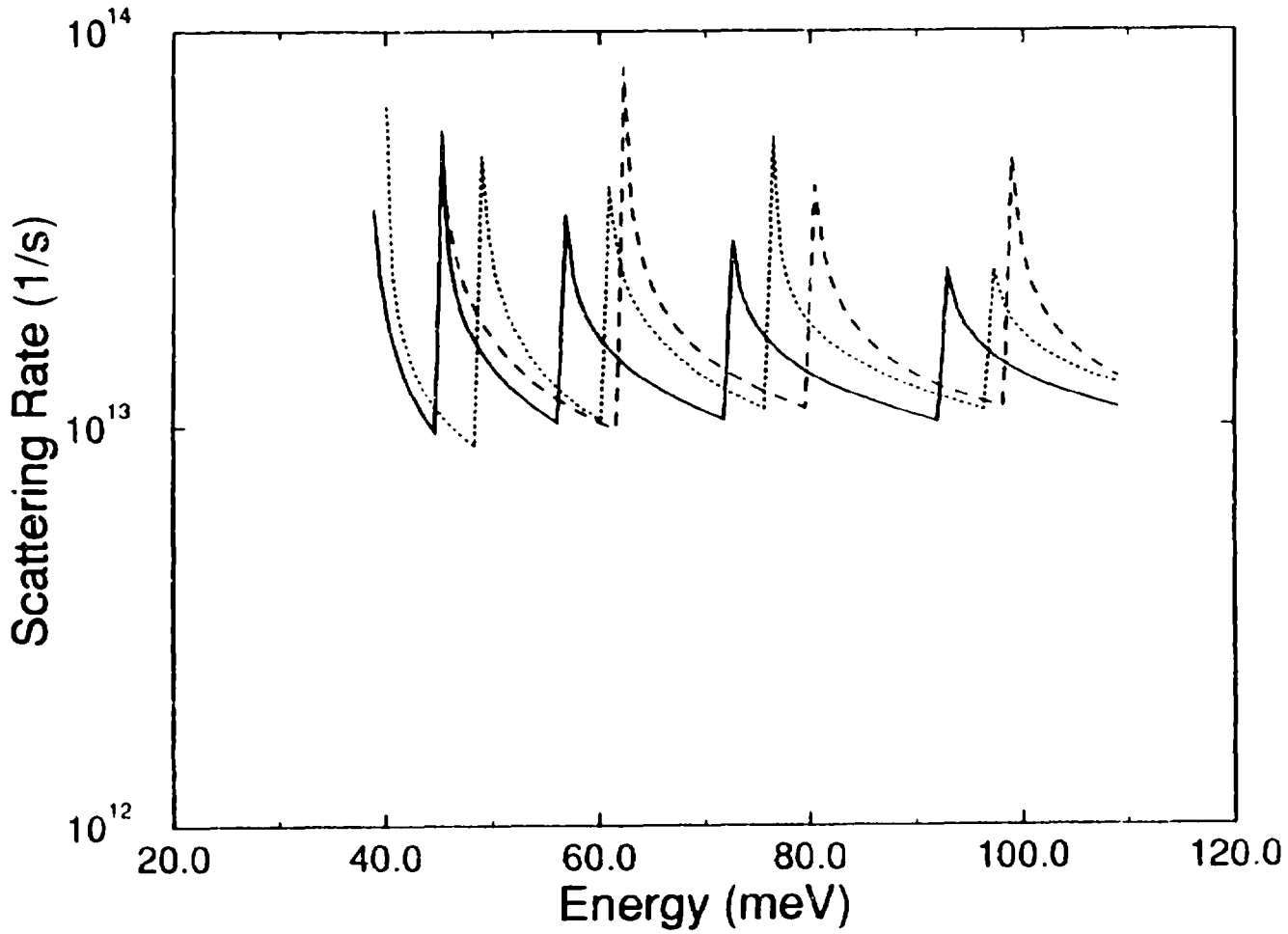
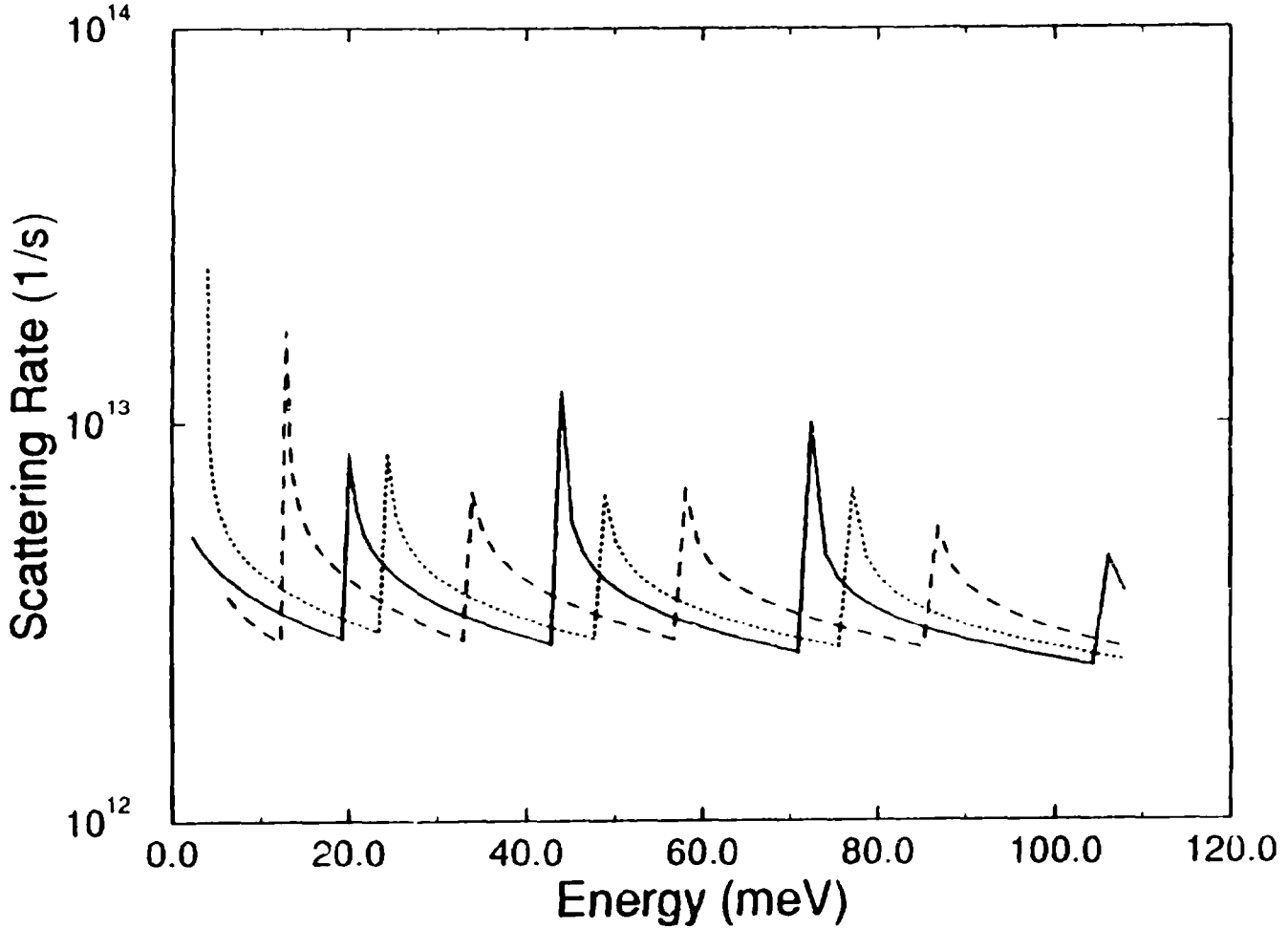

0.08
0.06
0.04
0.02
B = 0.05 Tesla
0.08
0.06
0.04
0.02
0.00
B = 10 Tesla
Wavefunction (arbitrary units)
0.0
100.0
200.0
300.0
400.0
500.0
Distance (Å)

Fig 4


10^{14}
10^{13}
10^{12}
Scattering Rate (1/s)
20.0
40.0
60.0
80.0
100.0
120.0
Energy (meV)

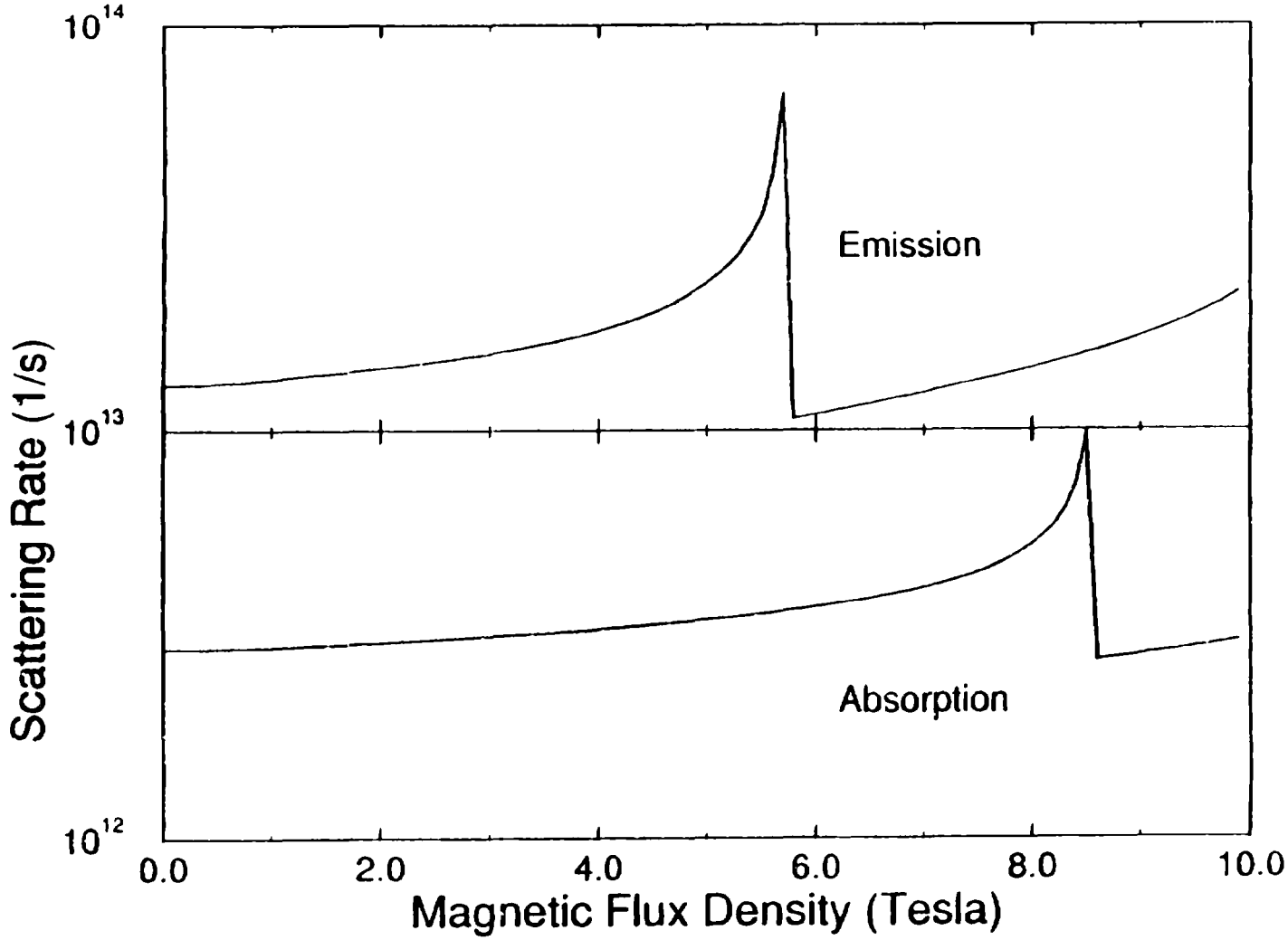
(a)

Fig 1 (a)


Scattering Rate (1/s)
10^14
10^13
10^12
0.0
20.0
40.0
60.0
80.0
100.0
120.0
Energy (meV)

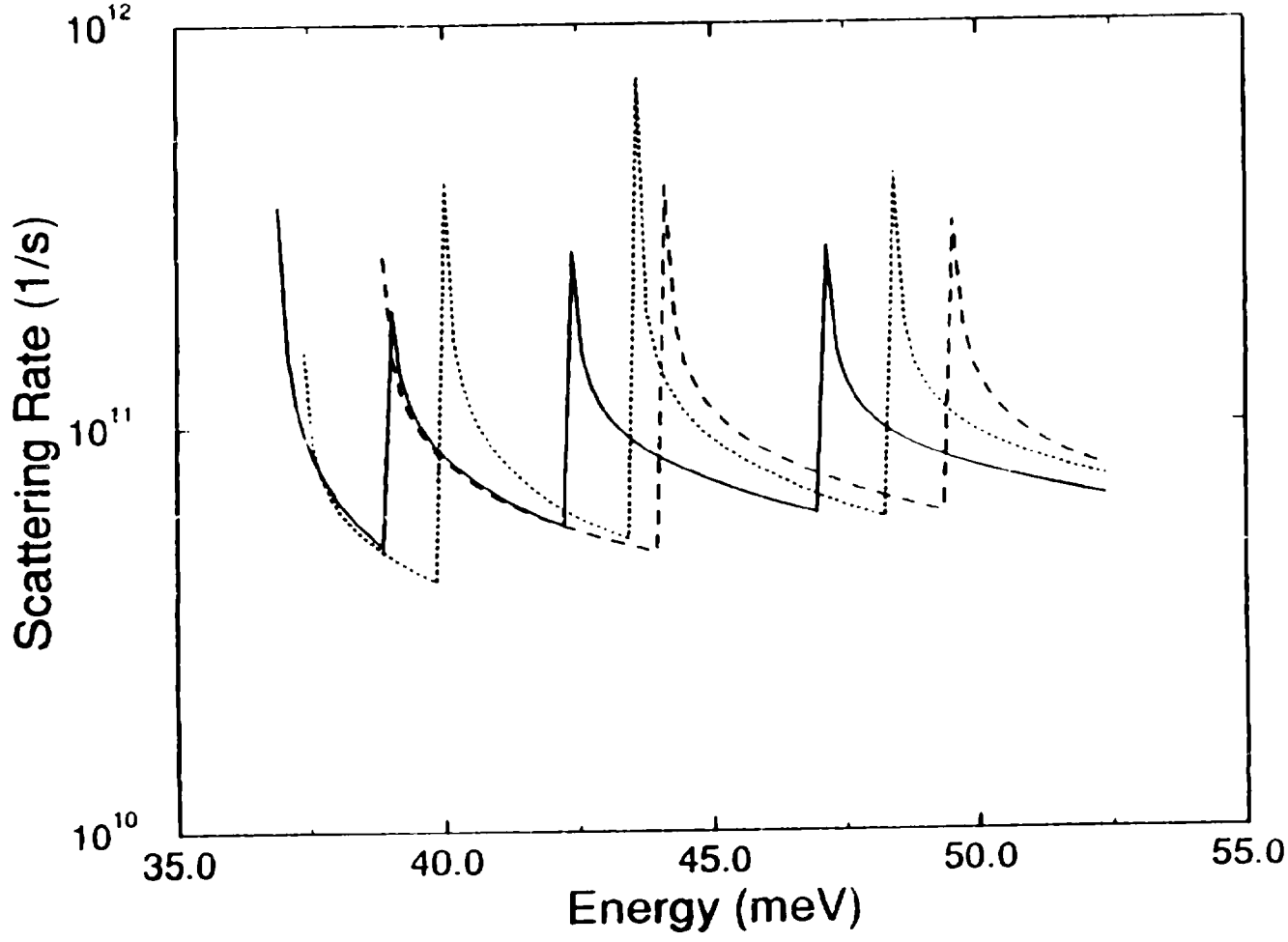
(b)

Fig 5(b)


10^{14}
10^{13}
10^{12}
Scattering Rate (1/s)
Emission
Absorption
0.0
2.0
4.0
6.0
8.0
10.0
Magnetic Flux Density (Tesla)

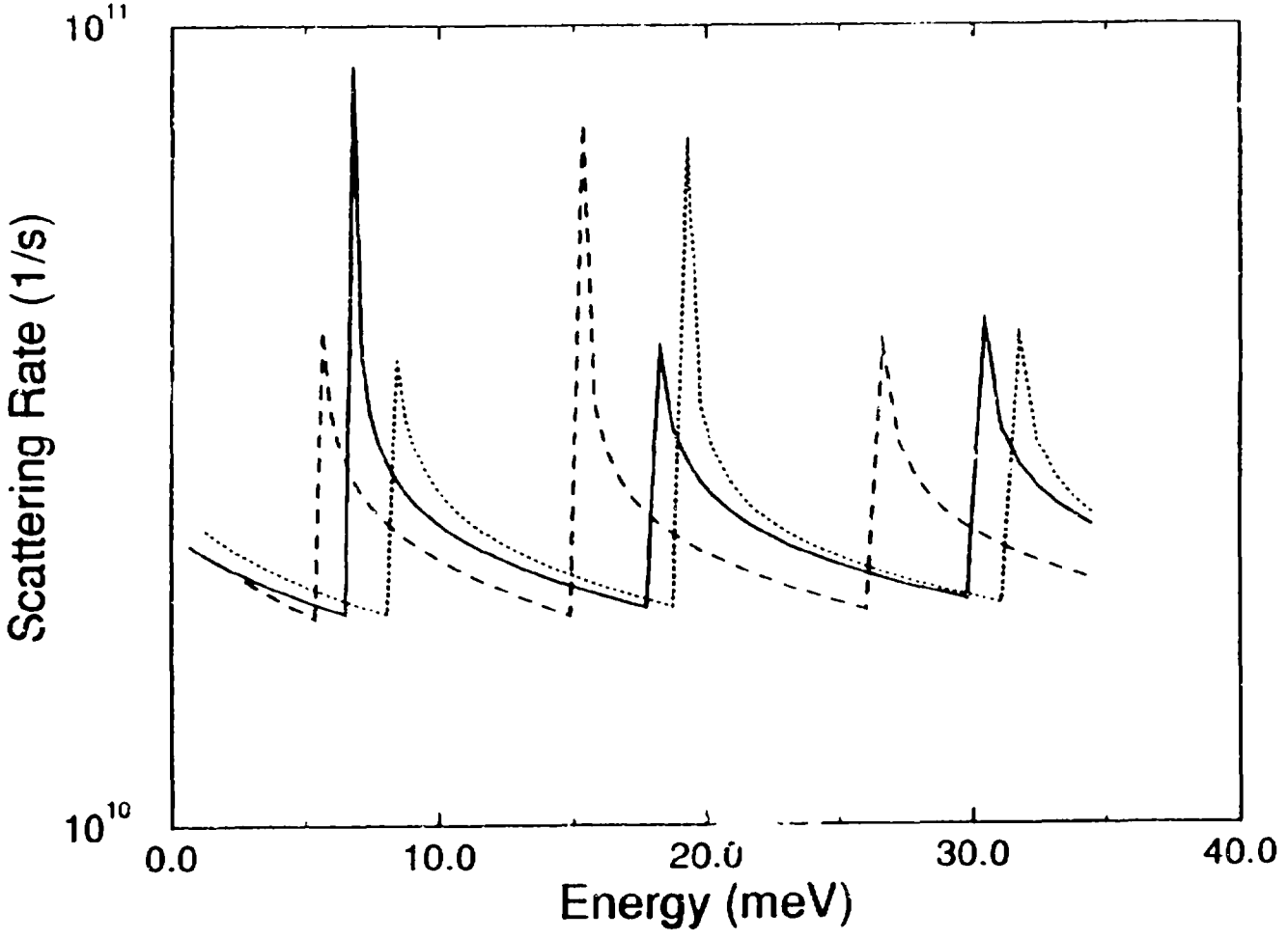
(c)

Fig 5(c)


10^{12}
10^{11}
10^{10}
Scattering Rate (1/s)
35.0
40.0
45.0
50.0
55.0
Energy (meV)

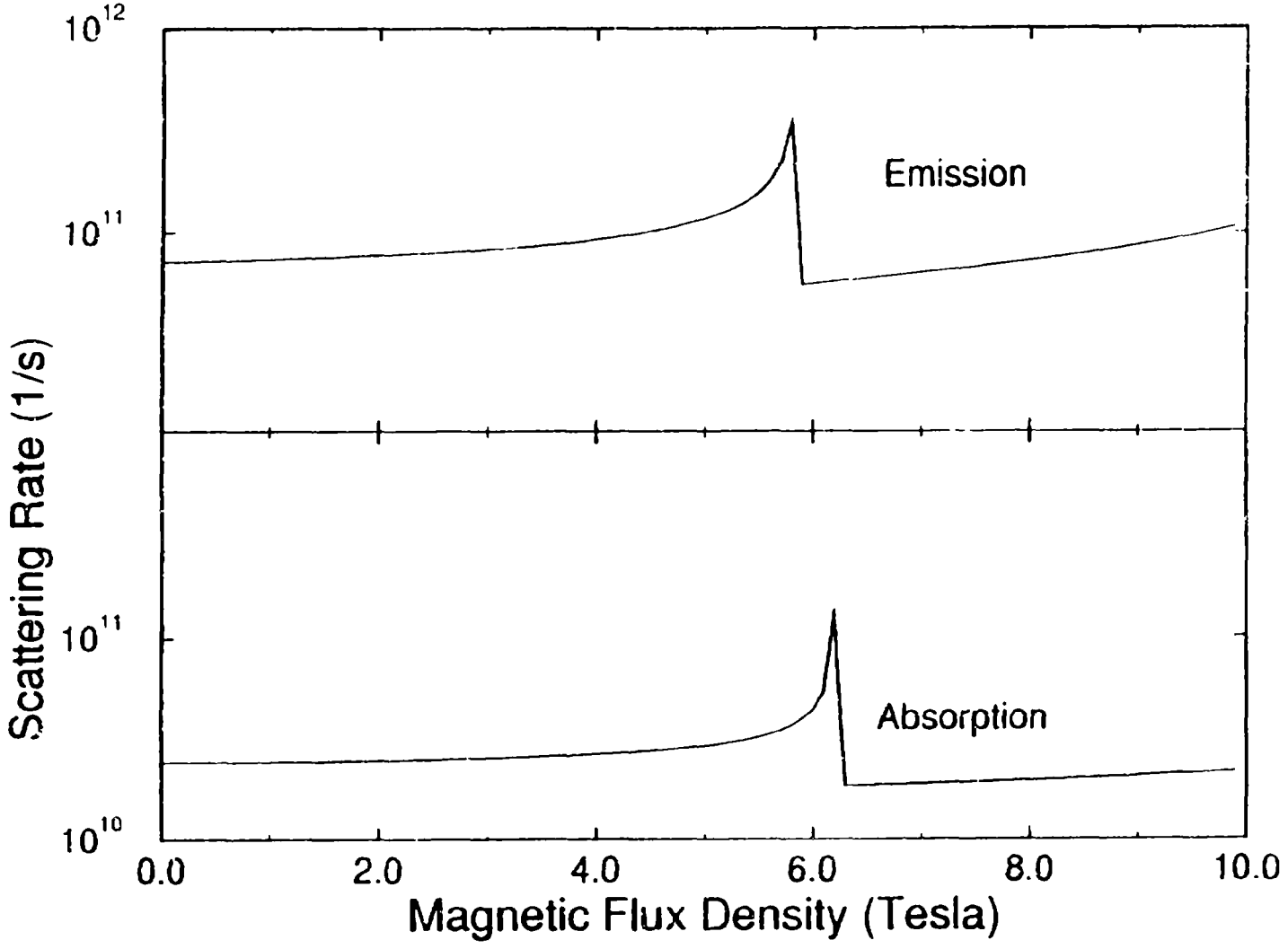
(a)

Fig 6(a)



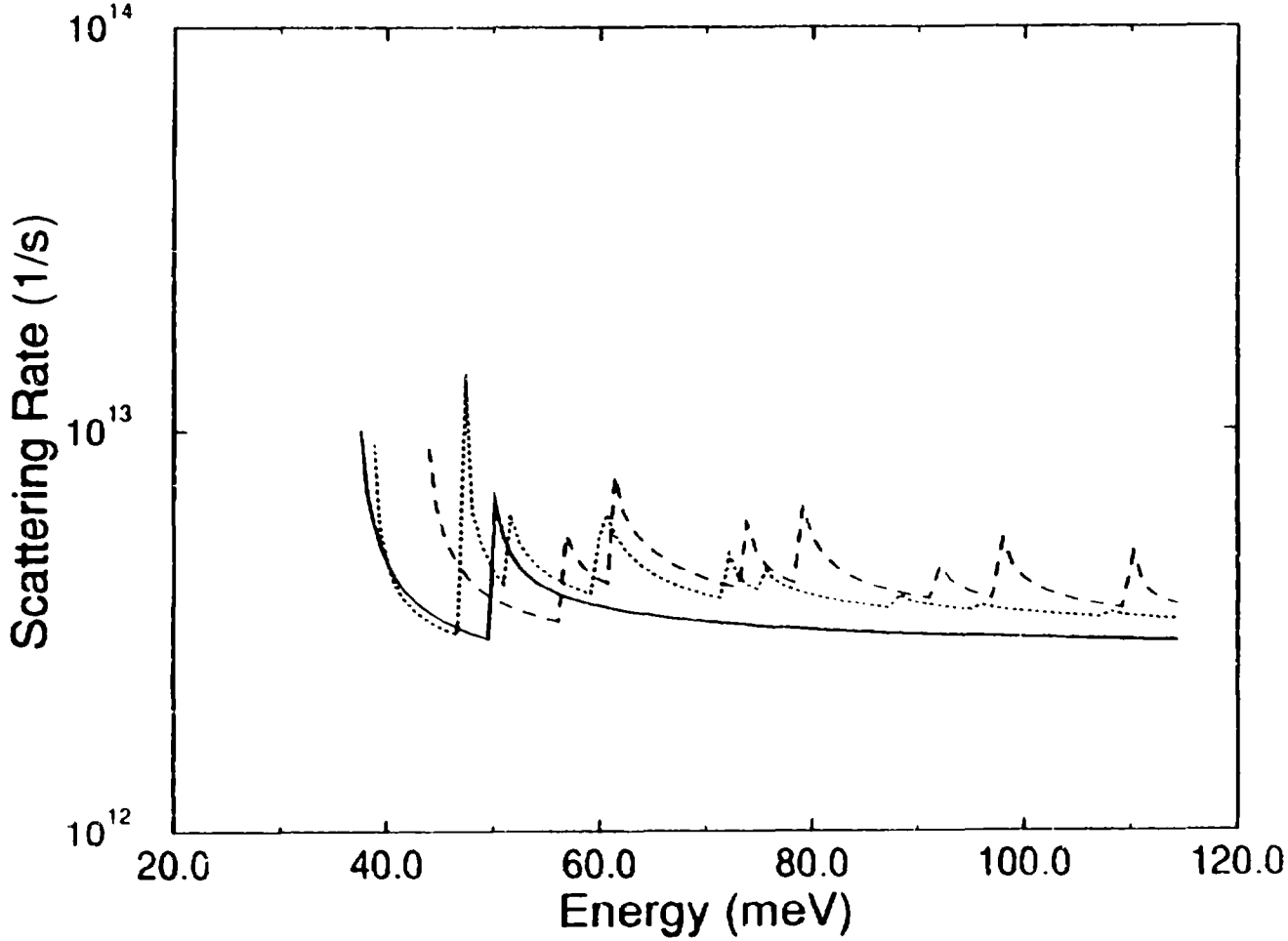
(b)

Fig. 6(b)


10^{12}
10^{11}
10^{11}
10^{10}
Scattering Rate (1/s)
Emission
Absorption
0.0
2.0
4.0
6.0
8.0
10.0
Magnetic Flux Density (Tesla)

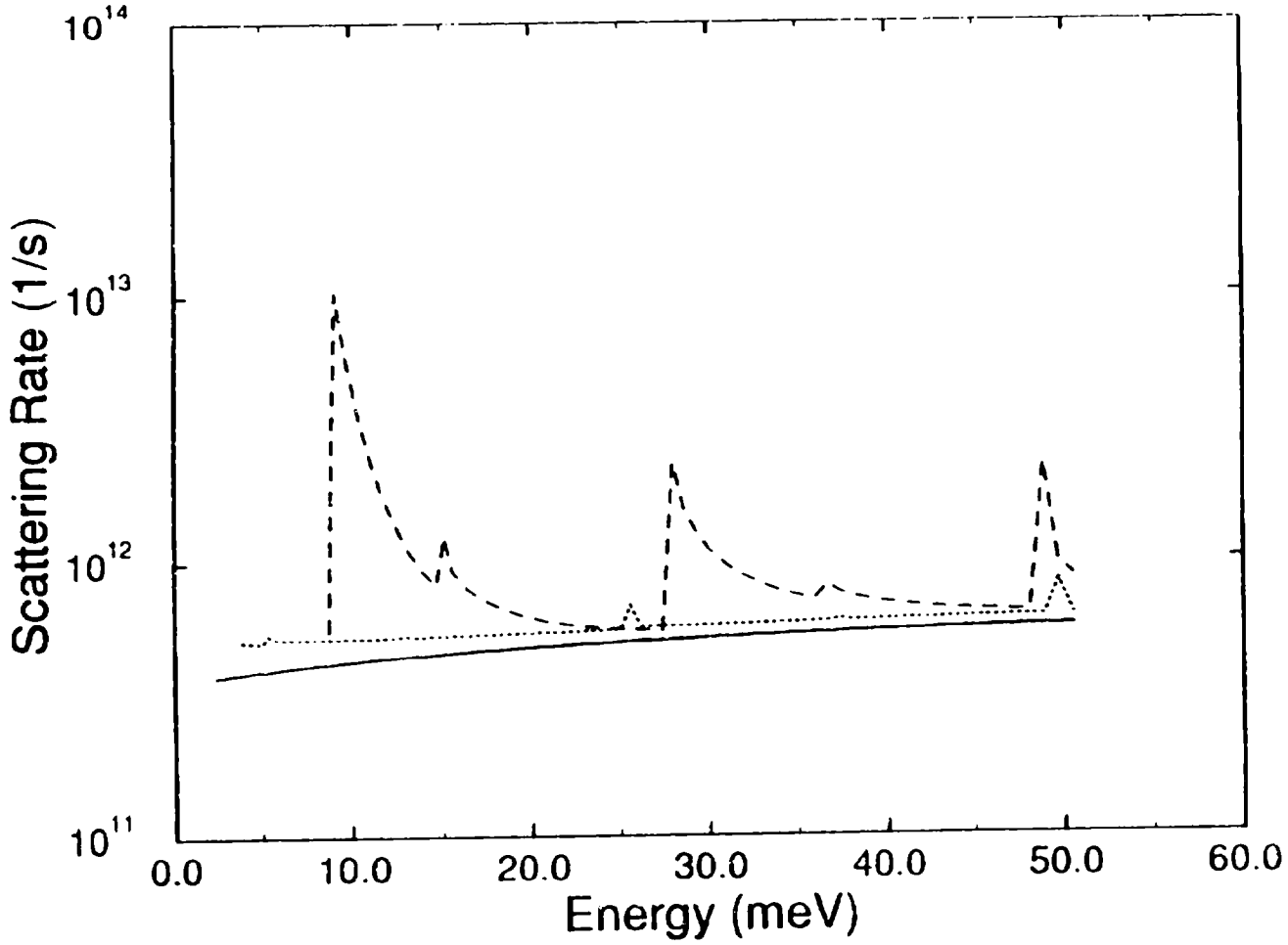
(c)

Fig. 6.(c)


10^{14}
10^{13}
10^{12}
Scattering Rate (1/s)
20.0
40.0
60.0
80.0
100.0
120.0
Energy (meV)

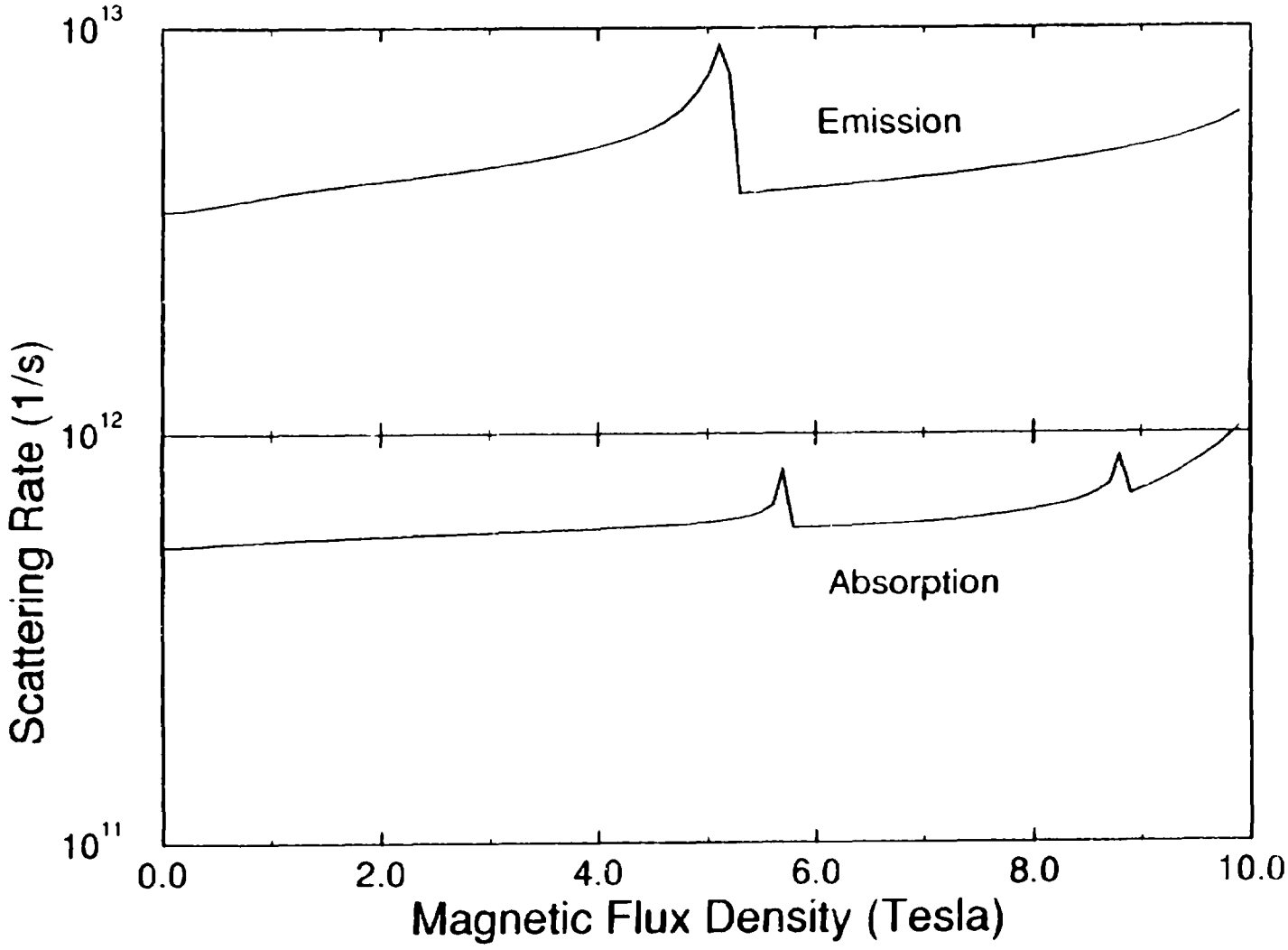
(a)

Fig 7(a)


Scattering Rate (1/s)
10^14
10^13
10^12
10^11
0.0
10.0
20.0
30.0
40.0
50.0
60.0
Energy (meV)

(b)

Fig 7(b)


10^{13}
10^{12}
10^{11}
Scattering Rate (1/s)
Emission
Absorption
0.0
2.0
4.0
6.0
8.0
10.0
Magnetic Flux Density (Tesla)

(c)

Fig 7(c)

Submitted to JAP

Supercomputing with spin polarized single electrons in a quantum coupled architecture

S. Bandyopadhyay, B. Das and A. E. Miller
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

We describe a novel quantum technology for ultrafast, ultradense and ultra-low power supercomputing. The technology utilizes *single electrons* as binary logic devices in which the *spin* of the electron encodes the bit information. The architecture mimics two dimensional cellular automata. It is realized by laying out on a wafer regimented arrays of nanophase particles each hosting an electron. Various types of logic gates, combinational circuits for arithmetic logic units, and sequential circuits for memory can be realized.

The technology has many advantages such as (1) the absence of physical interconnects between devices (inter-device interaction is provided by quantum mechanical coupling between adjacent electrons), (b) ultrafast switching times of $\sim$ 1 picosecond for individual devices, (c) extremely high bit density approaching 10 Terabits/cm^2, (4) non-volatile memory, (5) robustness and possible *room temperature* operation with very high noise margin and reliability, (6) a very low power delay product for switching a single bit ($\sim 10^{-20}$ Joules), and (7) a very small power dissipation of a few tens of nanowatts per bit.

We describe how the devices, circuits and architecture work, and then propose a new fabrication technology for realizing these chips.

I. INTRODUCTION

Recent advances in material growth techniques and nanolithography have spawned numerous proposals for novel electronic devices with superior computing performance. The most notable of these are "quantum devices" (and associated "quantum coupled architectures") whose operations rely on quantum mechanical effects as opposed to classical effects[1-9]. These new device and architecture ideas were motivated by the belief that conventional silicon technology will reach its limiting performance sometime around the end of this century[9], whereupon entirely new concepts will be required for the next generation of electronics.

While the belief regarding silicon technology may be true, none of the quantum devices proposed so far has succeeded in replacing the conventional silicon transistor. This, in turn, has prompted some skepticism[11,12,13] and raised serious questions about the viability of practical quantum device technology. It is currently believed that quantum devices are not suitable for integrated circuit implementation[11,12,13] since that would tax existing material and fabrication technology beyond the realm of the realistic. Moreover, almost all quantum devices, with perhaps the sole exception of resonant tunneling devices, are restricted to operate well below room temperature (typically ~ 1 K) which makes them impractical.

To overcome the drawbacks of generic quantum devices, we propose a new idea. In this scheme, the logic device (computing element) is a *single electron* whose "spin" encodes the bit information. The two possible polarizations of the spin represent the two binary bits. Although this is also a "quantum device", it is very different from the more familiar quantum *interference* devices where the bit information is encoded by the interference state of an electron and switching is achieved by modulating the electron's phase. Unlike the phase, an electron's spin is robust and cannot be flipped

easily by any external perturbation (noise) except a strong magnetic field. Therefore "spin devices" are *inherently reliable* and are practically *immune* to electrical noise! This attribute sets them apart from other quantum devices. In addition, switching a bit requires just toggling an electron's spin with *no physical movement of charges at all.* This is a great advantage since it eliminates transit time limitations on the switching speed and also resistance capacitance (RC) time constant limits. Most importantly, it eliminates problems arising from *charge trapping* by material defects. The latter is a severe problem in devices that rely on the movement of charge packets for switching. Since trapped charges are immobile, trapping will reduce the switching speed drastically. This is a devastating problem in charge coupled devices[14] and we believe that it will be even more serious in recently proposed *granular electronic* (quantum) devices[15] whose switching requires the movement of precisely *one* or a few electrons, usually by tunneling. Such devices will be extremely vulnerable to trapping (unfortunately, this problem is not often recognized by proponents of granular electronic devices). Trapping is a problem in any device that relies on the transfer of charge packets[16] and it will be more severe in single- or few-electron devices simply because if the lone electron (or one out of a few) gets trapped, the failure will be catastrophic. Therefore, single electron devices *must never* rely on the *physical movement of charges* for switching. Our single electron device meets that criterion. In fact, we believe that it has all the advantages of quantum devices such as high density, high speed, low power dissipation, etc. while having none of the usual disadvantages. This is a remarkable advancement.

This paper is organized as follows. In the next section, we describe quantum coupled architectures in which physical interconnects between devices are replaced by quantum mechanical coupling. The advantages of such an architecture are highlighted. We then describe how the proposed spin polarized single electron logic devices realize this architecture and how specific circuit

functions are implemented. This is exemplified by the designs of various logic gates, combinational digital systems for arithmetic logic units, and sequential digital systems for computer memory. We also provide a description of how to input and output binary data in such chips (reading and writing operations). In section IV, we describe a novel fabrication process for creating such a chip. Finally, in section V, we compare this new, revolutionary technology with other existing technologies, and then in section VI present the conclusions.

II. QUANTUM COUPLED ARCHITECTURES

The idea of quantum coupled interconnectless chips in which quantum mechanical coupling between quantum devices replaces physical interconnects (wires) dates back to at least 1987[6]. The preferred architecture for such chips is *cellular automata* since it is most synergistic with quantum coupling. Quantum coupling is short-range and could only work in architectures that have nearest neighbor interaction. Cellular automata not only meets this requirement, but it also affords massive parallelization and is known to be fault tolerant[7-10]. Moreover, it is ideal for single electron devices which necessarily have *small fan in and fan out* because they contain only a single electron. Finally, the cellular automata architecture can perform any desired memory or logic function, and in some special cases is more efficient than other architectures.

Researchers at Texas Instruments have proposed specific implementations of quantum coupled integrated circuits utilizing cellular automata architectures[5-11]. These circuits have no physical interconnects between devices and inter device communication is achieved through a variety of coupling mechanisms. In one scheme, the devices are quantum dot resonant tunneling devices and signal is transferred from one device to the next via tunneling of electrons[5]. Switches and logic gates (INVERTERs, NAND and NOR gates) have been designed using this scheme[6]. Also more complicated logic functions, such as Shannon cells, have been designed[8]. In addition to tunneling, other possible coupling mechanisms that have been proposed are electrostatic (or capacitive) coupling[7], optical coupling, acoustic coupling[9], etc.

The elimination of physical interconnects between devices is an important conceptual leap in the area of ultra large scale integrated circuits since physical interconnects pose the ultimate obstacle to further miniaturization[20].

However, it must also be understood that not all connecting wires on a chip can be eliminated. Even though the ones between devices can be eliminated, the chip itself must communicate with the external world (power supply, terminals, other chips etc.) through physical wires. Therefore, at least some devices on a chip must be connected to wires that will carry information back and forth from the external world. These devices will act as input/output ports. The Texas Instruments (TI) group's suggestion was that these devices be placed at the *periphery* of the chip for a simple technological reason. Bonding pads typically consume the area occupied by several hundred thousand devices. Therefore, it would not be possible to bond to individual ultra-small devices at the center of the chip (where packing is densest to facilitate short range quantum coupling) since even the most sophisticated bonding technique would not have that kind of spatial resolution. Therefore, it is advisable to wire only devices at the edges of a chip (where packing is relatively sparse) and provide (write) or retrieve (read) data only to and from these devices. This elegant scheme is adopted in almost all interconnectless architectures, including ours. In our scheme, data is input to electrons (devices) at the edges of the chip which communicate them to internal electrons via quantum coupling. The internal electrons then perform the computation and convey the result to other peripheral electrons (again, by quantum coupling) from which the result is read. This avoids the necessity to access internal electrons.

Another important idea that was implicit in the TI scheme was the concept of using devices that are in their *physical ground states* to perform the computation. The internal devices on a chip are in their ground states since they are not connected to any power source. External energy is provided to them only when they receive input from their nearest neighbors through quantum mechanical or other types of coupling. After receiving such inputs, the devices again decay to their ground states (which may be a different

ground state from the one that they were in prior to receiving the data; many distinct but energetically degenerate ground states are possible). It is the new ground state configuration of the devices that contain the result of the computation[21]. The advantages of ground state computing are the inherent stability of the device, improved noise immunity (noise perturbations may cause the device to stray from the ground state, but it will ultimately relax back to the ground state) and the elimination of the need to provide refresh cycles that are responsible for about 80 % of the power consumption of a conventional chip. These are major advantages.

A scheme that virtually takes the concept of ground state computing in an interconnectless architecture to its perfection was recently proposed by Bakshi and co-workers[22]. In this scheme, regimented arrays of elongated semiconductor quantum dots (termed "quantum dashes") are fabricated on a wafer with dimensions small enough that each dash contains a single electron. With the right arrangement and geometry, the ground state of this array will be such that the lone electron in each dash will be displaced towards one or the other edge of the dash giving rise to a net charge polarization. The polarizations in neighboring dashes are opposite when the system is in the ground state. This mimics antiferroelectricity. Each dash can now act as a binary logic device with the two possible directions of the charge polarization representing the two binary bits. Computation is performed by providing input to (orienting the polarizations in) peripheral dashes. These polarization modify those in the neighboring dashes and the effect propagates in a domino-type fashion through the entire array. The final configuration of the individual polarizations in the dashes (the new ground state) will represent the result of the computation. Our own computing paradigm is identical to this scheme, except that we deal with electron spin rather than charge polarization. The advantages of this are that we obtain much better bistability of the individual logic devices, much faster switching, improved

reliability, higher noise margin, etc.

Another scheme[23], inspired by the work of Bakshi et. al., proposed using five semiconductor quantum dots arranged in the shape of an X to form a single logic device. Each limb of the X ends in a dot and a central dot is placed where the limbs cross. There are only two electrons in the entire device which occupy either the two pivotal dots in one limb or those in the other. This gives rise to two possible (mutually perpendicular) charge polarizations. These two polarizations represent the two binary bits. Switching a bit would require the two electrons in a limb to tunnel through the central dot and occupy the pivotal dots in the other limb. This would reverse the polarization. Presumably such a scheme affords better bistability than the scheme of Bakshi et. al.[23]. Unfortunately, such a scheme is also replete with difficulties and has more disadvantages than advantages. Firstly, it requires physical movement of *precisely one* electron from a dot to another and, as stated before, trapping of that electron in any one dot would result in catastrophic failure. To avoid trapping, the devices must be switched (electrons transferred) extremely slowly[16] - slower than trapping/detrapping times which in most semiconductors is about 1 μsec). This will make such devices orders of magnitude slower than even conventional devices. Secondly, devices that rely on tunneling inherently suffer from *irreproducibility*[11]. Tunneling probability and tunneling speed are exponentially sensitive to barrier heights and widths which cannot be controlled with absolute precision[11]. Consequently, tunneling devices will have widely varying characteristics, even when they are fabricated by the same process on the same wafer. Devices with such wide variability are unsuitable for high density integrated circuits where billions of devices with nominally identical characteristics must be fabricated with a high degree of reproducibility. This problem of course affects the TI scheme as well since it proposes to use resonant tunneling devices. However, there the trapping problem at least is not serious since the devices are not single

electron devices. On the other hand, the scheme of Bakshi et al. uses single electron devices that do require some physical movement of charge. Therefore, they are not entirely immune to trapping, but the problem may not be as serious as in some other single electron devices. In Bakshi's scheme, electrons have to move within the *same* quantum dash, not between two different quantum dashes and more importantly, the movement of charges is not accomplished by tunneling. It is the ***combination*** of single electron transfer and transfer by tunneling that may be lethal. Considering these, the scheme of Bakshi et. al.[22] appear to be the best among the three described so far.

Although Bakshi's scheme is quite elegant, it has some drawbacks. The major drawback is that the transfer characteristic for switching is not sufficiently non-linear (in other words, the bistability is not sufficiently "hard"). This reduces the noise margin and device gain. In our scheme, which is similar to that of Bakshi et. al., we resolve this particular problem by representing a logic bit with an inherently *binary* quantity, namely electron spin. Unlike charge polarization, which is an analog variable, spin polarization in a quantum system is a binary digital variable. It can have only two possible polarizations - "up" and "down", and nothing in between. The immediate advantage of this is that device gain is no longer a relevant consideration and the reliability/noise margin improves dramatically.

III. COMPUTING WITH SPIN POLARIZED SINGLE ELECTRONS

The concept of using electron spin (an inherently bistable quantity) represent binary bits originated with Feynman. However, in order to realize actual logic functions and computing circuits, we have to meet four other requirements that are beyond the simple bistability. These four requirements are: (a) a mechanism for switching between the logic states, (b) a layout scheme to realize different circuit topologies for different logic functions (these will perform the desired computations), (c) a communication method (interconnection) between different logic devices to transfer information back and forth for computation (for this, the state of one logic device must determine that of its nearest neighbor if we adopt a cellular automaton approach), and (d) the ability to read and write bit information in selected (input/output) devices which provide the link between the chip and the external world.

The first requirement, namely switching a device from one logic state to another, is accomplished by flipping an electron's spin, either with a locally applied magnetic field (externally induced), or by magnon coupling between two nearest neighbor electrons (internally induced). The former is the mechanism for writing bits (input operation) and the latter occurs whenever the computation process requires one device to switch its neighbor. A local magnetic field can be applied by a spin polarized scanning tunneling microscope (SPSTM)[24,25,26] tip. We will discuss this in more detail later on.

The second requirement is met by fabricating single electron cells and then laying out these cells in various two dimensional patterns (by some patterning scheme) to realize different circuit topologies. The spatial arrangement of these cells determines how the single electrons (logic devices) are connected to each other and this realizes various circuits. The entire chip can be fabricated in this way.

The third requirement of inter-device communication is accomplished by quantum mechanical (spin-spin or magnon) coupling. The spin of one electron affects that of its neighbor in the following way. It is energetically favorable for two nearest neighbors to have opposite spins (we will show this later). Therefore, when the spin of one electron is switched, the neighbor feels it since the system goes to an excited state. The system can relax to the ground state only if the neighbor flips its own spin by emitting a phonon. This, in turn, flips the spin of the next electron, and so on. The effect again propagates in a domino-like fashion until the entire system of electrons has achieved a new ground state spin configuration. Information is therefore transmitted across the entire chip by perturbations that act like "spin waves"[27]. These waves typically have speeds that are about two to three orders of magnitude smaller than the speed of light.

Finally, the last requirement of reading and writing bit information in selected elements is fulfilled through the use of spin polarized scanning tunneling microscope (SPSTM) tips[24,25,26]. These tips can orient (write) the spin of an electron in a chosen cell by creating a localized magnetic field with atomic resolution and also measure (read) the spin polarization in an isolated cell from the magnitude of the spin-dependent tunneling current. In the next subsections, we show in detail this is actually accomplished.

Single electron logic devices, circuits and architecture

In this subsection, we explain how single electron logic devices, circuits and architecture work. To do this, we first show that any two nearest neighbor electrons (whose wavefunctions overlap) tend to have their spins antiparallel. This is caused by Coulomb, exchange and correlation interactions. This *antiferromagnetic order* (which realizes a natural inverter) is the only feature that is necessary for implementing any logic function. It forms the sole basis of the computing paradigm.

Antiferromagnetic ordering in a two electron system (in which the two electrons are confined to two well-separated core potentials) is actually well known[28,29]. It is the famous Heitler London result and its popular example is the Hydrogen molecule. In contrast, a system of many (more than two) electrons is not necessarily antiferromagnetic. However, if we consider a one-dimensional array (the so-called Heisenberg chain), then the net spin of the array must vanish[30] which is compatible with (although not a sufficient condition for) antiferromagnetism. The pure antiferromagnetic (Néel) state in a one-dimensional chain may be unstable against a variety of other states such as Anderson's resonanting valence bond state[31], the spin flop state[32], the magnetic discommensurate state[33], and the spin-Peierls state[34]. All of these are also possible ground states in one dimension[35]. Nonetheless, antiferromagnetism is the likely ground state in *finite* linear chains of *quantum confined* electrons (confined in isolated cells) because quantum confinement keeps the electrons apart and inhibits dimerization which is required for the other states.

Many logic circuits and logic gates require linear arrangements (chains) of electrons (these chains may be rectilinear or curvilinear). In these chains, the spins of neighboring electrons will be antiparallel. Such antiferromagnetic chains have been analyzed by Bethe[29]. Other circuits may require two dimensional arrangements of electrons. A two dimensional array of spins is basically a two dimensional Ising model. If we describe it by the Heisenberg spin Hamiltonian with only nearest neighbor interactions, then all that is required to obtain antiferromagnetism in such a system is that the exchange splitting (energy difference between the singlet and triplet states of two neighboring electrons) be *negative*[29]. We know this to be the case, so that antiferromagnetism is the likely ground state in a two dimensional array of single electron cells as well.

The fact that antiferromagnetism is preferred does not also guarantee

that it is stable. For a long time it was believed that antiferromagnetism in a one- or two-dimensional Ising model is not stable against lattice perturbations (phonons) and spin waves (magnons)[29,36]. However, it is now understood that the instability occurs only in *infinite* systems, whereas finite systems are theoretically stable[37]. Therefore, it is possible to sustain stable antiferromagnetism in a *finite* Ising system of restricted size.

The phase transition to antiferromagnetism in a two dimensional Ising system occurs at a temperature T_c which is given by the Onsager relation[29]

$$kT_c = \frac{2|J|}{\ln(1+\sqrt{2})} \quad (1)$$

where J is the exchange splitting. We would want this phase transition temperature to exceed room temperature so that the antiferromagnetism may be sustained at room temperature. This would allow room temperature operation of the circuits. Obviously, this can be ensured only by making the exchange splitting $|J|$ sufficiently large. This also causes a large energy difference between antiferromagnetic and ferromagnetic ordering which improves the noise margin since it decreases the probability of erratic spin flips. This probability is $\sim \exp(-|J|/kT)$ at a temperature T. For it to be small, we require the magnitude of J to be large.

Fortunately, unlike in a natural system, the exchange splitting J in an artificially structured two dimensional array can be *engineered*. The magnitude of the splitting depends on two factors: (a) the separation between adjacent single electron cells, and (b) the size of the cells which determines the degree of quantum confinement for each electron. By adjusting these parameters, we can make J sufficiently large to allow room temperature operation.

To determine the right range of parameters for making $|J|$ sufficiently large, we have calculated $|J|$ exactly for a model two electron system by solving the pertinent Schrödinger equation numerically.

The Schrödinger equation describing two electrons in two quantum boxes

single electron cells) is

$$i\hbar\frac{\partial\psi(\vec{r}_1,\vec{r}_2,t)}{\partial t}=\left[-\frac{\hbar^2}{2m^*}\nabla^2_{\vec{r}_1}-\frac{\hbar^2}{2m^*}\nabla^2_{\vec{r}_2}+V(\vec{r}_1)+V(\vec{r}_2)+\frac{q^2}{4\pi\epsilon|\vec{r}_1-\vec{r}_2|}\right]\psi(\vec{r}_1,\vec{r}_2,t) \quad (2)$$

where V is the potential profile seen by the electrons and we model it by the profile shown in Fig. 1. The above equation treats the effects of Coulomb, direct exchange and correlation interactions exactly. Since we are dealing with well-localized electrons and not treating magnetic ions, we can neglect superexchange, indirect exchange, itenerant exchange and dipole interactions totally[28].

The above equation and its simpler time-independent version (for steady-state dynamics) have been solved numerically by a number of researchers including one of us[28,39]. From these solutions, we have found that $|J|$ can be larger than 100 meV if the cells have a diameter of 20 Å and are separated from each other by 10 Å wide and 1 eV high barriers. Fortunately, these parameters can be realized in nanophase systems that we will describe later. We also find that with these parameters, the critical temperature for phase transition to antiferromagnetism (see Equation (1)) is $\sim$ 2600 K (theoretically) which is much above room temperature. This allows room temperature operation.

Logic gate realization

We now demonstrate the design of various logic circuits. For these designs, all that we need to assume is one basic property: *nearest neighbor electrons have opposite spins (antiferromagnetic ordering)*. We will adopt the convention that the "up" spin state is logic level 1 and the "down" spin state is logic level 0. Interconnectless logic gates in quantum coupled architecture were also designed in Ref. 6, 7, 22 and 23. We not only design logic gates, but ultimately show how complete combinational and sequential digital systems such as half-adders and flip-flops can be constructed. To

our knowledge, this is the first design of such complete systems in quantum coupled architectures.

NOT gates (inverters):

It is obvious that a system of just two coupled electrons (in closely spaced cells) constitutes a natural inverter. If the spin of one is "up", then the spin of the other must be "down" and vice versa since the ordering is antiferromagnetic. Therefore, if we consider the electron spin in one cell to be the input and the other to be the output, the output will always be the inverted version of the input. This realizes a NOT gate which is schematically depicted in Fig. 2.

AND and NAND gates:

To construct a NAND gate, consider three equally spaced cells in a linear chain (Fig. 3(a)). The two extreme cells are the two input ports and the one in the middle is the output port. If the spins in the two extreme cells are oriented "up" (i.e. both inputs are held at logic level 1), then the spin in the middle cell must be "down" for antiferromagnetic ordering. Similarly, it is easy to see that when the inputs are held at logic level 0, the output will be at 1.

Now, if one of the inputs is 1 and the other is 0, then the output can be either 1 or 0 since these two possibilities are energetically degenerate. We can resolve this degeneracy by applying a weak external magnetic field (*globally on the entire chip*) which induces a small Zeeman splitting (smaller than the exchange splitting J) between the "up" and "down" spin states . The direction of this field is such that the "up" spin state is favored. Therefore, if any one of the two inputs is at logic level 1, then the output will also be at logic level 1.

We have now realized the following *truth table*:

Input 1	Input 2	Output
1	1	0
1	0	1
0	1	1
0	0	1

It is easy to see that the above table is that of a NAND gate. Therefore, we have realized a NAND gate.

The NAND gate can be easily converted to an AND gate by directing the output of the NAND gate through an inverter. This requires four cells in a non-linear chain. The two extreme cells are the input ports and the one off the line is the output port. The spin orientations in the various cells for various inputs are also shown in Fig. 3(b). It can be easily verified from this diagram (which is essentially the "truth table") that this system is an AND gate.

OR and NOR gates

The OR gate can be realized from NAND gates and inverters through an application of De Morgan's law of Boolean algebra. This law states $\overline{\overline{A}\,\overline{B}} = A + B$, where A and B are two binary Boolean quantities. The right hand side of the above equality is the OR function of two quantities A and B. Therefore, an OR gate can be realized by realizing the left hand side of the equality using NAND gates and inverters. The nanophase realization is shown in Fig. 4.

Exclusive OR gates

Exclusive OR gates can be realized by using the exclusive OR relation $Y = (A + B)(\overline{AB}) = (\overline{\overline{AB}})(\overline{AB})$, where A and B are two inputs and Y is the output. The nanophase array in Fig. 5 is an exclusive OR gate.

Combinational digital systems for arithmetic logic units

A digital computer is required to perform only two basic types of functions: logic operations and memory storage[40]. Logic operations are acheived through combinational digital systems (consisting of logic gates) while random access memory (RAM) can be realized through sequential digital systems. In the following example, we design the most basic combinational digital system used in an arithmetic logic unit. It is the binary *half adder*.

In a half adder, if A and B are two binary addends, S the sum, D the digit indicating the last digit of the sum and C the carry, then D is the exclusive OR function of A and B, (i.e. $D = (\overline{\overline{A}\,\overline{B}})(\overline{AB})$[40]) while C is the AND function of A and B. The schematic realization of a half adder is shown in Fig. 6(a) and the actual realization with single electron cells is shown in Fig. 6(b). The chip area consumed by such a system is only about 3000 $\AA^2$ which promises extremely high functional density.

In a similar fashion, one can construct code converters, parity checkers, parity encoders, multiplexers, etc. These circuits are of course more complicated and are not presented in this paper.

Sequential digital systems for random access memory

As an example of a sequential digital system for memory, we show the design of an SR flip flop. The nanophase realization is shown in Fig. 7. The reader can verify that it indeed performs as required. Other types of flip-flops such as J-K and master-slave J-K can also be constructed in a similar fashion. From these flip-flops, all basic memory circuits such as shift registers and counters can be constructed[40].

Reading and writing operations: orienting and detecting electron spins in single electron cells

We now address a very critical component of any computing scheme, namely the READ/WRITE mechanism. The WRITE operation will align an electron's spin in a single electron cell (which will be a few atom particle) to the desired orientation, while the READ operation will detect the orientation. For both operations, we need to access individual cells with virtually *atomic resolution*. This can be achieved with a spin-polarized scanning tunneling microscope (SPSTM) which offers the atomic resolution as well as the ability to couple to an electron's spin.

Since their inception[41], scanning tunneling microscopes (STM) have been extensively used for surface analysis of atomic arrangements as well as nanofabrication[42,43]. A SPSTM is a special type of STM in which the probe is constructed from a magnetic material. The tunneling current in this case depends on the magnetization of the probe (the spin orientation at the very tip) as well as the magnetization of the surface (spin orientation of the surface atoms). In the last few years, there has been considerable interest in SPSTM for fundamental studies of surface magnetism as well as developing techniques for magnetic recording[44]. Using SPSTM, spin polarized electrons have been observed on the surfaces of SiO_2 and Cr[24,25]. Very recently, the imaging of magnetite (Fe_3O_4) was reported using an Fe tip[26]. Even though SPSTM is a relatively new technique, the results obtained in the past few years show that it has a great potential for magnetic recording and detection in an atomic scale. We believe it to be the ideal technique for spin READ/WRITE (spin alignment/spin detection) operations in single electron cells.

Reading mechanism

The READ operation will be performed by detecting the spin polarization in selected cells (at the periphery of the chip) with SPSTM tips. A number of tips will be used for this purpose as schematically shown in Fig. 8. Each output port will have an individual tip attached to it. The tips themselves will be fixed in space with respect to the chip since mechanical motion of probe tips during reading and writing would slow down the computational speed unacceptably.

The way an SPSTM detects spin is the following. In an SPSTM, the tunneling current depends on the relative spin polarizations of the probe tip and the atom on the surface. The probe tip has a fixed known polarization. The atom on the surface can have either up- or down- spin. Thus, from a measure of the tunneling current, the spin polarization in a surface atom (or a single electron cell) can be determined at any time. Theoretical estimates show[45] that the difference between the currents for the two spin polarizations can differ by a factor of 3 which is sufficiently large for unambiguous spin detection. This allows us to perform the READ operation.

Of course, certain dificulties can be encountered in the READ operation. These are due to magnetostriction and variations in cell size and shape. Magnetostriction can cause a change in the thickness and the shape of the cell (particle) itself. Since the gap between the probe tip and the particle is very small ($\sim$ 1 Å), even a small amount of magnetostriction will affect the tunneling current. This problem needs to be investigated in detail to estimate its importance. A possible way to alleviate the problem is to use a non-magnetic probe tip to measure the particle height under two different magnetizations and then use this data to calibrate the reading operation. The effect of particle shape on the tunneling current will have to be also properly taken into consideration.

Next, we should address the selection of probe materials for SPSTM. A number of criteria have to be satisfied in this regard. One important consideration is that the magnetization of the probe tip should not be affected by the magnetization of the device and vice versa. This will be difficult since the two tips will be in very close proximity. One solution to this problem is to use an antiferromagnetic tip like chromium (Cr) which does not influence the studied surface through the magnetostatic field[45]. Since the devices are expected to operate at room temperature, an additional restriction on the tip material is that it should have Néel or Curie temperature above room temperature. A list of electrically conducting antiferromagnets with Néel temperatures higher than room temperature are given in ref. 44. A few of these (MnPt, MnNi and Cr) have been already used in SPSTM studies. We feel that Cr probes are the optimum choice.

Finally, the question that needs to be answered is what the size of the SPSTM peripheral input/output devices would be. Regular STMs are very much larger than a chip so that the use of STMs may seem to defeat the very purpose of integration. Actually, this is not true. We do not use the SPSTMs as microscopes and we do *not* scan them. All we need are the atomically sharp probe tips. These can be embedded or vertically positioned on the chip by a variety of nanofabrication techniques. We do not expect the input/output devices to consume any more area than typical bonding pads and input/output pins in a regular chip.

Writing mechanism

The WRITE mechanism will polarize the spin of the electron in an input cell to the desired orientation. Again, SPSTM tips will be used for this purpose. To polarize an electron's spin to a desired orientation, a strong enough magnetic field needs to be generated locally (with atomic resolution and range). This will be achieved by the technique described below.

A soft magnetic probe will be placed in close contact with the cell (particle). During the writing operation, it will be magnetized electrically. The magnetostatic force experienced (the basis for Magnetic Force Microscope) is expected to alter the spin polarization of the particle. Probe material has to be selected such that the polarization of the probe is not altered by the magnetostatic force. Since Fe has been demonstrated to perform well as an SPSTM tip[26], soft Fe will be the first choice.

Unidirectional isolation between input and output

In all electronic logic devices, a necessary requirement is *unidirectional isolation* between input and output. The output of one device should drive the input of the next; but the logic state of this next device must not influence the logic state of the preceding one. In other words, signal must flow unidirectionally. This is accomplished by somehow providing unidirectional isolation between the input and output ports of a logic device. This means that the input signal determines the output signal but not vice versa.

In conventional devices, this is accomplished through the device *gain*. The input signal is amplified on its way to the output port, whereas the output signal is attenuated in propagating to the input port. Unfortunately, spin polarized single electron logic devices have no gain like all other granular electronic devices. While this does not pose a problem with noise margin or signal restoration (since electron spin is robust and practically immune to electrical noise), it still poses a problem with unidirectional isolation.

To understand why this is so, we refer to Fig. 9. There are two NOT gates in series and Fig. 9(a) shows the equilibrium configuration of spins. Now imagine that the input of the first NOT gate is flipped by an external source such as by an SPSTM (Fig. 9(b)). At this point, the spin state in the central cell becomes indeterminate since the spin the right cell favors the "upspin" state while the spin in the left cell favors the "downspin" state. If

the external magnetic field (refer to the discussion of AND gates) favors the "upspin" state, then the spin in the central cell will not flip in response to the external input. In other words, the first NOT gate fails. In fact, if the external input (SPSTM) is removed, the spin in the leftmost cell (input port) will flip back to "upspin".

The above is an example of the input being determined by the state of the output port rather than the reverse. This is due to the lack of unidirectional isolation between input and output. This problem was probably never recognized and to our knowledge was never addressed before in the context of granular electronic devices. *We believe that this is the major problem with such devices and may ultimately limit their applicability.*

A possible solution to this problem is shown in Fig. 9(c). We change the spacing betwen cells as shown in Fig. 9(c). Since the right cell is farther from the central cell than the left cell, the left cell has dominant sway. This provides effective unidirectional isolation. Unfortunately, this type of solution is problem specific. Also, increasing the separation cannot be carried on indefinately since increasing separation also decreases the strength of magnon coupling. Ultimately, this limits the number of logic devices that can be used on a chip. We are presently investigating this problem in greater detail.

IV. FABRICATION OF SINGLE ELECTRON LOGIC CHIPS

Realization of single electron logic chips with nanophase materials

In this section, we address the fabrication of single electron logic chips. The schematic view of such a chip is shown in Fig. 10. There is a dense arrangement of cells at the center and a sparse arrangement at the periphery. The peripheral cells are the input/output ports for reading and writing.

The obvious way to realize single electron cells would be to use conventional nanolithography such as electron beam or X-ray lithography. These techniques could also lay them out in specific patterns to implement various circuit topologies. However, we are convinced that this will *not* work because of the damages that such processes inflict on the structures. The reasons are elucidated later on. Therefore, we have chosen an entirely new fabrication technique that utilizes nanophase particles for single electron cells. The patterning scheme for laying out these cells is also novel. We believe that this has an excellent chance of success.

The fabrication of single electron chips requires two basic abilities: (a) deposition of nanophase particles (single electron cells) with good control over size, and (b) arranging them at selected positions on a wafer to realize the various circuit topologies. We call the latter component "patterning" in analogy with the term commonly used in integrated circuit delineation.

Nanophase material deposition

The ability to deposit nanophase particles on a wafer with good control over size is well developed technique. There are many methods by which this can be achieved, but the gas condensation technique has been the most

broadly adopted. The properties and deposition techniques have been described by a number of authors[60-62] and we refer the reader to this literature for further elucidation. Nanophase particles of γ-Al_2O_3 have been produced at Argonne National Laboratory and have shown that a significant number of particles have diameters below 30 Å. We show a high resolution TEM image of ZnO particles deposited on Carbon films in Fig. 11. The particle diameter is around 90 Å. For even smaller particle size, one can employ rf sputtering of the source material for deposition which can produce particles in the 20 - 50 Å range[63-64]

Patterning by selective area nano deposition

To achieve the desired arrangement of particles on a wafer for realizing specific circuit configurations, we propose to employ a new technique that we call *selective area nano deposition* (SAND). In this technique, the deposited particles are made to nucleate only at pre-selected sites on the wafer. These sites are chosen according to the desired arrangement and this realizes the patterning.

There are two ways to pre-select nucleation sites with atomic resolution. They are described below.

- **Creation of charged nucleation sites**

  In this method, minute amounts of charge are deposited in ultrasmall ($\sim$ 30 Å diameter) sites on a wafer using a *field emission* STM. An STM can be made to operate in the field emission mode by applying a relatively large voltage ($\sim$ 150 V) between the probe tip and the substrate. As a result, a very narrow electron beam is extracted from the tip. STMs operating in the field emission mode have been demonstrated to provide 3 nm resolution for imaging[62].

  For the creation of a nucleation site, the STM probe tip will be moved

to the desired location and brought close to the substrate. Following this, a large voltage will be applied to the probe tip which will result in a very fine electron beam emanating from the tip. The location where the electron beam strikes the substrate will become negatively charged. Since the beam diameter is very small, the dimension of this charged location is expected to be of nanometer scale. The STM probe tip will then be moved to the next location and another negatively charged site will be created. This is repeated (stepped across the entire wafer by computer control) to delineate the desired array pattern. If the substrate is insulating, the charges in isolated spots will not leak out. The situation is analogous to the creation of charged spots during SEM imaging of non-conducting materials.

After the patterning is complete, the nanophase particles are deposited. They will preferentially occupy (or migrate to) the charged sites owing to electrostatic attraction, especially if they themselves had been charged with the opposite polarity beforehand. This method is very similar to xerography and is illustrated in Fig. 12.

Once the particles have migrated to and settled down in their nucleation sites, they remain stuck there owing to electrostatic attraction. This may be viewed as electrostatic "bond" formation. The excess (stray) particles which did not find nucleation sites to form such bonds, will be removed by passing the wafer underneath a charged stripping plate whose charge is opposite in sign to that of the particles. Finally, we will be left only with particles in the desired arrangement. This accomplishes the patterning.

One disadvantage of this scheme is the following. The wafer or substrate on which a single electron logic chip is fabricated must be conducting since STM reading and writing requires conducting substrates.

However, during patterning, it has to be made temporarily insulating to prevent leaking away of the deposited charges. A possible mechanism for achieving this is to use *cooled* GaAs or silicon substrates. At low temperatures, they are insulating owing to dopant "freeze out", but they become conducting at room temperature. Therefore, cooling the substrate with liquid nitrogen during patterning and deposition will be necessary.

- **Creation of nanometer size (uncharged) holes for nucleation sites**

  In this scheme, nanometer sized holes are created by indentation using an STM. The technique has already been demonstrated on grapnite substrates[63]. A relatively large current and voltage is applied between the probe tip and the substrate to generate these holes. In ref. 63, it was also claimed that gold clusters deposited on the substrate preferentially occupy these holes. *This is exactly the SAND technique* (therefore, there is already some existing evidence that the SAND technique will be successful). The advantage of this particular technique (as opposed to the previous technique of creating *charged* sites) is that (a) it is easier and has already been demonstrated[63], and (b) one does not need insulating substrates which eliminates the need for substrate cooling unlike in the previous scheme.

The final issue that needs to be addressed is the stability of the particle arrangement. Even without electrostatic attraction, deposited particles stick to the surface owing to surface tension. Therefore, once a configuration is attained, it may be maintained indefinitely. The stability under adverse circumstances, such as under elevated temperatures that promote Brownian motion, needs to be investigated.

Mass production

The process of pre-patterning with a scanning tunneling microscope is similar to "direct writing". Although effective, it has the same drawback as all direct writing processes - it is very time consuming and not suitable for mass production. For the latter purpose, it may be more convenient to replace the direct write process with a technique that involves *exposure through a mask*. This may indeed be possible. X-ray, electron beam or focused ion beams can be focussed to $\sim$ 10 Å and therefore can pass through masks with similar feature sizes without significant diffraction. We can expose selected areas of a wafer to these beams through masks. Irradiated areas will become temporarily charged or indented and desired patterns can be realized in the same way as before. The only difference is that this process is suitable for mass production and will have a high throughput, although it requires a much larger capital investment.

Advantages of Selective Area Nano Deposition (SAND)

It may appear unusual that we have chosen an entirely new fabrication scheme for fabricating arrays of ultrasmall particles in apparent neglect of the well-known and time honored techniques of electron beam or x-ray lithography followed by post processing. This is because we believe that *lithography followed by post processing is totally unsuitable for creating structures that are a few tens of angstroms in size as opposed to a few thousands of angstroms in size*. To our knowledge, no attempt at creating *undepleted* semiconductor structures of a few tens of angstroms in size has been successful with lithography and post processing. The reason for this is that electron beam or x-ray lithography introduces an abundance of material defects[64] during exposure which are further increased during post procesing (such as reactive ion etching). These defects that are induced by radiation damage severly degrade

ne structures and deplete them of all mobile carriers by Fermi level pinning. it is curious that in spite of this, lithography and post processing continue to be the preferred techniques for fabricating single and few electron structures which must accomodate a precise number of electrons with no tolerance for even one extra electron! In Fig. 10, one can see the virtually undisturbed lattice planes in most particles. Only one particle in this melee has a stacking fault that is visible. The material quality is astoundingly good and shows few, if any, disiocations. We believe that the Selective Area Nano Deposition SAND) technique, which we have proposed, is the ideal technique for making ultrasmall structures with ~ 10 Å resolution in all three dimensions. It may indeed become the dominant technique for nanofabrication in the future.

V. COMPARISON WITH EXISTING AND OTHER PROPOSED TECHNOLOGIES

In this section, we will compare our proposed computing technology with existing and other proposed technologies. To do this, we first provide an estimate of the figures of merit for single electron logic chips.

Estimation of the switching speed, power dissipation and allowable bit density

In single electron logic, a bit is switched by flipping an electron's spin. This is achieved either by locally applying a magnetic field (during writing) or by magnon coupling (during computation). In the former case, the switching time will be of the order of $\hbar/g\mu B$ where B is the flux density of the locally applied field, g is the g-factor (which can be very large in some semimagnetic semiconductors) and μ_B is the Bohr magneton. For a flux density of 1 tesla and a g-factor of 10, the switching time is $\sim$ 1 picosecond.

In the second case, magnon coupling flips the spin of an electron by emitting a phonon. Spin-phonon coupling can be quite strong in pyroelectric materials (uniaxial crystals without inversion symmetry) where electric dipole spin resonance[65,66] can increase spin flip rates significantly. In some materials like HgTe, spin-phonon transition linewidths of 0.4 meV have been predicted[67] which gives a switching time of $\sim \hbar/0.4meV \approx 1$ picosecond.

The power dissipation for switching a single bit can be estimated as follows. If the energy splitting between the triplet and singlet state is 100 meV and the switching time is 1 picosecond, then the power dissipation for switching a single bit is $\sim$ 100 meV/1 picosecond = 16 nanowatts. This is a few orders of magnitude smaller than what can be acheived in conventional devices. The power delay product is then $\sim 10^{-20}$ Joules which is of the same order as that achievable with quantum interference devices. It is orders

of magnitude smaller than what can be achieved with conventional devices, including Josephson junctions.

Next, we calculate the bit density that can be realized. Each nanophase particle (or each bit) occupies an area of $\sim 50 \times 50$ A^2. Therefore, the bit density will be ~ 25 terabits/cm^2. Such a high bit density poses a problem with cooling. Since the power dissipation per bit is 16 nanowatts, the *maximum* power dissipation from a 1 cm^2 chip will be $\sim$ 400,000 Watts! Removal of 1000 W/cm^2 from a silicon chip was demonstrated more than ten years ago[8] and it may be possible to improve this. However, acquiring a capability of removing 400 kilowatts/cm^2 at room temperature will not be easy. This problem can of course be eliminated altogether by reducing the operating temperature from room temperature to 77 K. The device sizes can then be increased to reduce the energy splitting between the triplet and singlet states (energy difference between logic levels) to 10 meV. This is still larger than the thermal energy kT at 77 K and therefore allows 77 K operation. Also, at 77 K, the phonon assisted spin flip rate (switching speed) may decrease by a factor of 10 since phonon assisted scattering rates are proportional to the Bose-Einstein factor which has an exponential dependence on temperature. This reduces the power dissipation per bit by a factor of 100 thereby reducing the total dissipation to 4000 Watts/cm^2 which is more manageable.

It must be emphasized that low temperature (77 K) operation is not required because of device or circuit limitations. Rather, it is required because current heat removal technology cannot perform at the required level. Once heat sinking technology has improved enough, room temperature operation can be restored.

We now compare our proposed technology with both other quantum device technologies and conventional technologies.

Comparison with quantum devices

It has been pointed out several times[11,12,13] that quantum interference devices are impractical for integrated circuits. This is because their characteristics are extremely sensitive to a few angstroms variation in size, or a few millivolts variation in voltage, or a few nanoamperes variation in current. Because of the lack of fabrication tolerance, these devices are not *reproducible*. Consequently, they cannot be used in integrated circuits where hundreds of millions of devices must be fabricated reproducibly with reasonably high yield.

In addition to having no fabrication tolerance, quantum interference devices also have practically no noise tolerance. Such delicate devices cannot work in integrated circuits where voltage variations will inevitably occur owing to reflection, attentuation and distortion of signals communicated between various devices.

There are some other fundamental shortcomings of quantum interference devices. For instance, the lack of non-linear operating characteristics (the only exception is the resonant tunneling device) and the lack of intrinsic device gain make these devices totally unsuitable for digital and logic applications. Finally, the extremely low current carrying capability (quantum devices must operate at low currents to avoid dephasing interactions) causes these devices to be actually quite slow in their overall switching response ($\sim$ 100 psec), sometimes slower than even conventional silicon devices[39].

Granular quantum devices whose switching relies on the transfer (usually via tunneling) of one or a few electrons from one region of space to another are often worse than quantum interference devices. Examples of these devices are the single electron transistor based on Coulomb blockade and the scheme in Ref. 23. In addition to having most of the disadvantages of quantum interference devices, they also have the additional disadvantage of being

extremely slow in their response. The switching speed is reduced dramatically by lack of tolerance to trapping[6]. In CCDs, trapping is known to cause switching delays of 1 μs - typically[16], resulting in an extremely slow bit rate of 1 Mbit/sec. In single or few electron devices, the problem is bound to be worse since the charge packets are extremely small (one or a few electrons instead of about 10,000 in conventional CCDs). Therefore, no single electron device should ever rely on charge transfer for switching. Our single electron device does not rely on such charge movement.

Finally, almost all quantum devices (with the sole exception of resonant tunneling devices) have a serious drawback. They cannot operate even at 77 K, let alone room temperature. This feature makes them impractical. In contrast, our devices can operate at room temperature.

Comparison with conventional technologies

Conventional technologies including bipolar junction transistors (BJTs and HBJTs), complementary metal oxide semiconductor field effect transistors (CMOS), Josephson junctions, magnetic bubble memory, etc. have their own limitations. It is believed that the major problems of BJT and MOS technology are associated with scaling. Power supply voltages cannot be scaled down indefinitely. The minimum they can reach is the thermal voltage below which the noise margin becomes unacceptably poor. When device sizes are scaled down without scaling the voltages, the electric field increases proportionately with decreasing length. Ultimately, the electric field will reach the critical value for breakdown which sets a limit to device scaling.

The product of the power supply voltage and the unity gain frequency in conventional devices cannot exceed $F_{max}v_{max}$ where F_{max} is the breakdown field and v_{max} is the saturation velocity at that field. This is known as the Johnson limit[70] and it has never been surmounted. For silicon, this limits the maximum unity gain frequency to 2×10^{11} Hz and the switching speed

to 5 psec if the voltage is 1 volt. This is a fundamental material limit and no amount of clever innovativeness can surmount it[71]. Therefore, it becomes necessary to explore alternate means of realizing ultrafast and ultradense computing devices.

We now examine our proposed spin polarized single electron logic devices. The advantages of these devices are the following:

- The devices can operate at room temperature.
- The devices have fabrication tolerance. The size and shape of the particles are not critical as long as the particles are small enough to host single electrons. In fact, even single electron occupancy is not really necessary. All that is required is that the particles have a resultant spin moment determined by the interplay of Coulomb, exchange and correlation forces. The spin moment must also respond to magnons.
- The *extrinsic* switching speed of a device is $\sim$ 1 psec. Such ultrafast switching is made possible by the fact that no charge movement is necessary so that we are not limited by transit time or resistance-capacitance (RC) time constants. This switching speed is better than that of quantum interference devices, far better than that of granular electronic devices relying on charge transfer, and comparable to the switching speed of the highest performing conventional devices such as complementary metal oxide semiconductors (CMOS) and Josephson junctions[69].
- The power dissipation is only tens of nanowatts per bit which allows extremely dense integration. The power delay product is $\sim 10^{-20}$ Joules which is comparable to that of quantum interference devices.
- The logic variable is spin which is a robust physical variable and is practically immune to electrical noise!. Therefore these devices can

operate with extremely high noise margin and reliability.

- Memory elements are non-volatile, basically for the same reason that magnetic bubble memories are non-volatile. Once the spin of a particle has been oriented, it remains in that configuration unless perturbed by a magnetic field.

- Memory can be extremely dense (~ 10 Terabits/cm^2). However, this density may not be achievable immediately because of limitations imposed by the maximum obtainable rate of heat removal from a chip. The effective memory density is further enhanced by circuit compaction. In conventional circuits, at least four transistors are required to make a latch for storing a single bit. Here, a single electron can store a single bit.

- The architecture is *interconnectless* with interconnection between devices provided by quantum coupling. This removes the major hurdle to miniaturization.

- Since all logic devices are in the ground state, we do not need frequent *refresh cycles*. Conventional devices always operate in excited states and therefore require constant input and refresh cycles through individual address lines and interconnects. We do not need these cycles which eliminates 80 % of the energy requirement.

The above are some of the major advantages of the proposed technology. Needless to say, they are so attractive that the scheme merits a thorough theoretical and experimental investigation. The rewards of such an endeavor may be well worth the effort.

VI. CONCLUSION

In conclusion, we have proposed and described a novel quantum technology for ultrafast, ultradense and ultra low power supercomputing. It utilizes single electrons as logic devices in a quantum coupled cellular automata architecture. For fabrication, we have proposed a new technology (SAND) and a new kind of material system (nanophase materials) which we believe are ideal for atomic scale nanofabrication.

Acknowledgement: The authors are grateful to Dr. J. A. Eastman and Profs. P. Bakshi, M. D. Lemmon and P. F. Bagwell for many fruitful discussions. We are especially grateful to Prof. J. A. Cooper, Jr. for discussions regarding unidirectional isolation. This work was supported in part by the Air Force Office of Scientific Research under grant number AFOSR-91-0211.

References

[1]. *Quantum Electron Devices*. Springer Series in Electronics and Photonics 28. ed. F. Capasso. (Springer-Verlag, New York. 1990).

[2]. S. Datta, M. R. Melloch, S. Bandyopadhyay and M. S. Lundstrom. *Appl. Phys. Lett.*, **48**. 487 (1986).

[3]. S. Datta and B. Das. *Appl. Phys. Lett.*, **56**. 665 (1990).

[4]. C. C. Eugster, J. A. del Alamo, M. J. Rooks and M. R. Melloch. *Appl. Phys. Lett.*, **60**. 642 (1992).

[5]. S. Datta. *Superlattices and Microstructures*, **6**. 83 (1989); F. Sols, M. Macucci, U. Ravaioli and K. Hess *Appl. Phys. Lett.*, **54**. 350 (1989).

[6]. R. T. Bate, G. A. Frazier, W. R. Frensley, J. W. Lee and M. A. Reed, Proc. of SPIE, Vol. 792, 26 (1987)

[7]. R. T. Bate. *Solid State Technology*, , 101 (1989); R. T. Bate, G. A. Frazier, W. R. Frensley and M. A. Reed, *Texas Instruments Technical Journal*, 6. 13 (1989); J. Randall, J. Luscombe, M. Reed and A. Seabaugh. *ibid.* 6. 49 (1989).

[8]. J. N. Randall, M. A. Reed and G. A. Frazier, *J. Vac. Sci. Technol. B*, **7**. 1398 (1989).

[9]. J. N. Randall, M. A. Reed and Y. C. Kao, *J. Vac. Sci. technol. B*, **8**. 1348 (1990).

[10]. P. K. Chatterjee, P. Yang and H. Shichijo, *IEE Proc.*, **130** pt. 1. 105 (1983).

[11]. R. Landauer, *Physics Today*, **42**, 119 (1989); R. Landauer, in *Nanostructure Physics and Fabrication*, eds. M. A. Reed and W. P. Kirk, (Academic Press, Boston, 1989); p.

[12]. R. W. Keyes, *Physics Today*, **45**, 42 (1992).

[13]. S. Subramaniam, S. Bandyopadhyay and W. Porod, *J. Appl. Phys.*, **68**, 4861 (1990).

[14]. D. K. Schröder, *Advanced MOS Devices*, Modular Series on Solid State Devices, edited by R. F. Pierret and G. W. Neudeck, Vol. VII, (Addison-Wesley, Reading, Massacnusetts, 1989).

[15]. *Granular Nanoelectronics*, eds. D. K. Ferry, J. R. Barker and C. Jacobini (Plenum, New York, 1991).

[16]. See, for example, J. E. Carnes and W. F. Kosonocky, *Solid State Technol.*, **67**, ... (1974); see also R. F. Pierret, "MOS Deep Depletion Devices" *Purdue University Technical Report*, **TR-EE-78-50** (1978).

[17]. Peter Gacs, *MIT Workshop on Cellular Automata*, June 1986.

[18]. M. Harao and S. Noguchi, *J. Comp. and Sys. Sci.*, **11**, 171 (1975).

[19]. Special Issue on Fault Tolerant Computing, *IEEE Trans. on Computers*, **C-35**, Feb 1982.

[20]. D. K. Ferry, L. A. Akers and E. W. Greeneich, *Ultra Large Scale Integrated Microelectronics*, (Prentice Hall, Englewood Cliffs, New Jersey, 1988).

[21]. Computing with devices in their ground state is not entirely a new concept since it was well-known in the context of Hopfield type neural networks.

[22]. P. Bakshi, D. Broido and K. Kempa, *J. Appl. Phys.*, **70**, 5150 (1991); P. Baskhi (unpublished).

[23]. C. S. Lent, P. D. Tougaw and W. Porod, *Proc. International Workshop on Computational Electronics*, Urbana, Illinois, May, 1992, p. 163; C. S. Lent, P. D. Tougaw and W. Porod, *Appl. Phys. Lett.*, **62**, 714 (1993); C. S. Lent, P. D. Tougaw, W. Porod and G. H. Bernstein (unpublished).

[24]. Y. Manassen, R. J. Hamers, J. E. Demuth, and A. J. Castellano, Jr., *Phys. Rev. Lett.*, **62**, 2531 (1989).

[25]. R. Wiesendanger, H.-J. Guntherodt, G. Guntherodt, R. J. Gambino, and R. Ruf, *Phys. Rev. Lett.*, **65**, 247 (1990).

[26]. R. Wiesendanger, I. V. Shvets, D. Burgler, G. Tarrach, H.-J. Guntherodt, and J. M. D. Coey, *Europhys. Lett.*, **19**, 141 (1992).

[27]. F. Keffer, *Am. J. Phys.*, **21**, 250 (1953); G. G. Low, *J. Appl. Phys.*, **35**, 998 (1964).

[28]. Neal W. Ashcroft and N. David Mermin, *Solid State Physics*, (Saunders College, Philadelphia, 1976), Chap. 32.

[29]. J. M. Ziman, *Principles of the Theory of Solids*, Second edition, (Cambridge University Press, Cambridge, UK, 1972), Chap. 10.

[30]. E. Lieb and D. Mattis, *Phys. Rev.*, **125**, 164 (1962).

[31]. P. W. Anderson, *Mat. Res. Bull.*, **8**, 153 (1973).

[32]. See, for example, S. Inagaki and H. Fukuyama, *J. Phys. Soc. Jpn.*, **53**, 4836 (1984).

[33]. See, for example, T. W. Hijmans, H. B. Brom and L. J. De Jongh, *Phys. Rev. Lett.*, **54**, 1714 (1985).

[34]. See, for example, J. W. Bray, L. V. Interrante, I. S. Jacobs and J. C. Bonner, *Extended Linear Chain Compounds*, Vol. III, edited by J. S. Miller (Plenum, 1983).

[35]. See, for example, H. Fukuyama and S. Inagaki, in *Magnetic Properties of Low Dimensional Systems*, edited by L. M. Falicov and J. L. Morán Lopez, (Springer Verlag, Heidelberg, Germany, 1986), p. 156.

[36]. R. E. Peierls, *Ann. Inst. Henri Poincaré*, **Helv.** 5, 177 (1935); L. D. Landau, *Phys. Z. Sowjet.* **11**, 26 (1937).

[37]. See, for example, L. Gunther, *Phys. Lett.* **25A**, 649 (1967); G. Allan in Ref. 35, p. 2.

[38]. C. S. Lent, *Nanostructures and Mesoscopic Systems*, eds. M. A. Reed and W. P. Kirk, (Academic Press, Boston, 1991), p. 183; L. F. Register and K. Hess, *ibid.* p. 369.

[39]. M. Cahay, T. Singh and S. Bandyopadhyay, *Proc. Intl. Workshop on Computational Electronics*, Beckman Institute, University of Illinois, p. 147 (1992).

[40]. See, for example, Jacob Millman, *Microelecronics*, (McGraw Hill, New York, 1979).

[41]. G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, *Appl. Phys. Lett.*, **40**, 178 (1982); also *Phys. Rev. Lett.*, **49**, 57(1982).

[42]. U. Staufer, R. Weisendanger, L. Eng, L. Rosenthaler, H. R. Hidber, and H. J. Guntherodt, *J. Vac. Sci. Technol.*, **A 6**, 537(1988).

[43]. E. E. Ehrichs, R. M. Silver, and A. L. deLozanne, *J. Vac. Sci. Tech.*, **A 6**, 540 (1988).

[44]. I. V. Shvets, R. Wiesendanger, D. Burgler, G. Tarrach, H.-J. Guntherodt, and J. M. D. Coey, *J. Appl. Phys.*, **71**, 5489 (1992).

[45]. A. A. Minakov, and I. V. Shvets, *Surf. Sci. Lett.*, **236**, L377 (1990).

[46]. M. Uda, *Scripta Metall. et. Mater.* (in press).

[47]. H. Gleiter, *Deformation of Polycrystals: Mechanisms and Microstructures*, eds. N. Hansen, A. Horseweil and T. Lefters, Ris National Laboratory, Roskilde, p. 15.

[48]. R. C. Baetzold and J. F. Hamilton, *Prog. Solid State Chem.*, **15**, 1 (1983).

[49]. H. Gleiter, *Prog. in Mat. Sci.* **33**, 223 (1989).

[50]. R. Uyeda, *Prog. in Mat. Sci.* **35**, 1 (1991).

[51]. R. W. Siegel, *Atomic Level Properties of Interface Materials*, eds. D. Wolf and S. Yip, (Chapman and Hall, London, 1991).

[52]. R. W. Siegel, *Processing of Metals and Alloys*, ed. R. W. Cahn, Vol. 15, (VCH Verlags-gesellschaft, Weinheim, 1991).

[53]. R. W. Siegel and J. A. Eastman, *Mat. Res. Soc. Symp. Proc.*, **132**, 3 (1989).

[54]. R. P. Andres, et. al., *J. Mat. Res.*, **4**, 704 (1989).

[55]. H. Oya, T. Ichihashi and N. Wada, *Jpn. J. Appl. Phys.*, **21**, 554 (1982).

[56]. S. Yatsuya, K. Yamauchi, T. Kamakura, A. Yanagada, H. Wakaiyama and K. Mihima, *Surface Sci.*, **156**, 1011 (1985).

[57]. S. Yatsuya, T. Kamakura, K. Yamauchi and K. Mhama, *Jpn. J. Appl. Phys.*, Part 2 **25**, L42 (1986).

[58]. H. Hahn and R. S. Averback, *J. Appl. Phys.*, **67**, 1113 (1990).

[59]. G. M. Chow, R. L. Holtz, A. Pattnaik, A. S. Edelstein, T. E. Schlesinger and R. C. Cammerata, *Appl. Phys. Lett.*, **56**, 1853 (1990).

[60]. R. S. Bowles, J. J. Kolstad, J. M. Calo and R. P. Andres, *Surface Sci.*, **106**, 117 (1981).

[61]. C. Hayashi, *J. Vac. Sci. Technol.*, **A 5**, 1375 (1987).

[62]. R. Allenspach, and A. Bischof, *Appl. Phys. Lett.*, **54**, 587 (1989).

[63]. K. Terashima, M. Kondoh, and T. Yoshida, *J. Vac. Sci. Technol.*, **A 8**, 381 (1990).

[64]. See, for example, C. M. Knoedler, *J. Appl. Phys.*, **68**, 1129 (1990); T. Fink, D. D. Smith and W. D. Braddock, *IEEE Trans. Elec. Dev.*, **ED-37**, 1422 (1990); T. B. Stellwag, M. R. Melloch, J. A. Cooper, Jr., S. T. Sheppard and D. D. Nolte, *J. Appl. Phys.*, **71**, 4509 (1992); B. Das, S. Subramaniam and M. R. Melloch (unpublished).

[65]. R. Romestain, S. Geschwind and G. E. Devlin, *Phys. Rev. Lett.*, **39**, 1583 (1977).

[66]. M. Dobrowolska, H. D. Drew, J. K. Furdyna, T. Ichiguchi, A. Witowski and P. A. Wolff, *Phys. Rev. Lett.*, **49**, 845 (1982).

[67]. W. Zawadzki, G. Bauer, W. Racek and H. Kahlert, *Phys. Rev. Lett.*, **35**, 1098 (1975).

[68]. D. B. Tuckerman and R. F. W. Pease, *IEEE Elec. Dev. Lett.*, **2**, 126 (1981).

[69]. T. Kobayasni, M. Miyaka, Y. Okazaki, T. Matsuda, M. Sato, K. Digucni, S. Ohki and M. Oda. *Technical Digest of the IEEE Electron Device Meeting* (IEEE, New York, 1988), p. 881; S. Kotani, T. Imamura and S. Hasuo, *ibid*, p. 884; Y. Yamane, T. Enoki, S. Sugitani and M. Hirayama, *ibid*, p. 894.

[70] E. O. Johnson, *RCA Rev.*, **26**, 163 (1965).

[71]. M. Nagata, *IEEE J. Sold State Cir.*, **27**, 465 (1992).

[72]. S. Bandyopadhyay, S. Datta and M. R. Melloch, *Superlattices and Microstructures*, **2**, 539 (1986).

FIGURE CAPTIONS

Figure 1: The potential profile seen by two electrons in two isolated nanophase particles. The confining potential is assumed to be rectangular for convenience. The basic property that the spins of the two electrons are antiparallel (in the ground state) does not depend on the potential shape. However, the splitting between the singlet and triplet state (or the exchange energy J) depends on the potential shape, the width of the wells and the height of the potential barrier that separates them.

Figure 2: A nanophase realization of an INVERTER (NOT gate). The spin state (logic level) in one particle is the inverted version of the spin state or logic level in the other.

Figure 3: A nanophase realization of (a) a NAND gate, and (b) an AND gate. Also shown are the four possible spin configurations of the array which correspond to the Boolean truth table.

Figure 4: A nanophase realization of (a) an OR gate, and (b) a NOR gate.

Figure 5: A nanophase realization of an exclusive OR gate.

Figure 6: (a) A realization of a binary half adder using exclusive OR and AND gates, (b) a nanophase realization.

Figure 7: (a) A realization of an SR flip flop using NAND gates, (b) a nanophase realization.

Figure 8: Reading the spin of a nannophase particle using a SPSTM. The tunneling current measured by the SPSTM tip depends on the relative spin polarizations of the tip and the particle.

Figure 9: An example of failure due to the lack of unidirectional isolation between input and output. (a) The equilibrium configuration of electron

spins (logic states) in two NOT gates in series. (b) The logic state at input of the first NOT gate (left cell) is changed by an SPSTM (external source), but the state at the output does not change in response because of the previous state at the output of the second NOT gate. The failure occurs because of a lack of unidirectional isolation between the input and output. (c) A possible solution. The left cell is closer to the central cell and therefore holds dominant sway.

Figure 9: Schematic layout of a single electron logic chip. The packing is dense at the center and sparse at the edges where the reading and writing operations are done.

Figure 10: A TEM micrograph of nanophase particles of ZnO. The particles are unpatterned and form a dense random array. The average particle size is 90 Å. The virtually undisturbed lattice planes are visible and show practically no crystal defect in most cases. Note the uniformity in size and the astounding material quality.

Figure 11: Figure illustrating the three steps in the Selective Area Nano Deposition procedure.

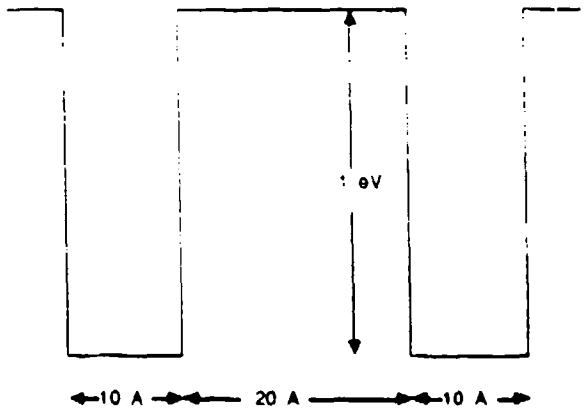


Fig 1
Bandyopadhyay et. al

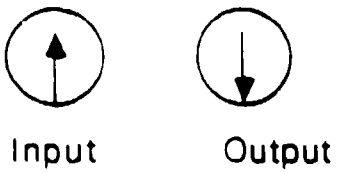
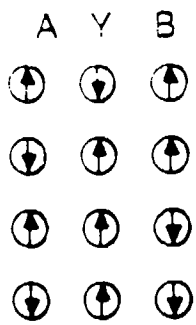


Fig 2

NAND gate

Nanophase realization



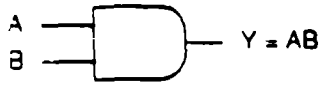
Possible spin configurations of the array

(Truth table)

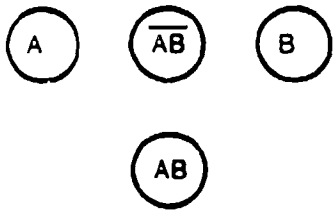
Boolean Truth Table

A	Y	B
1	0	1
0	1	1
1	1	0
0	1	0

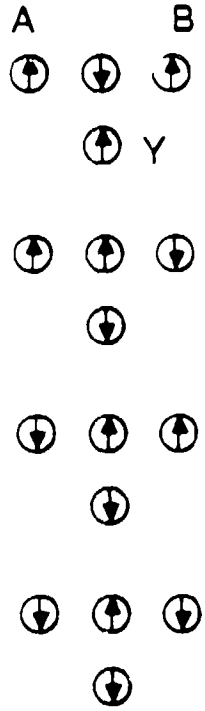
(a)



AND gate



Nanophase realization



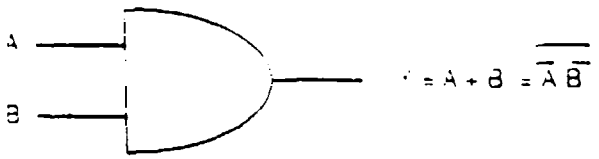
Possible spin configurations of the array

(Truth table)

Boolean Truth Table

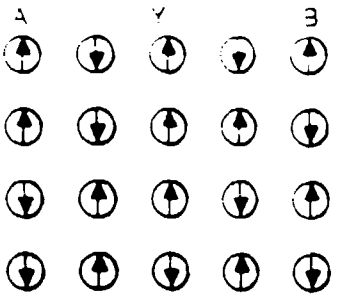
A	Y	B
1	1	1
0	0	1
1	0	0
0	0	0

(b)



An OR gate

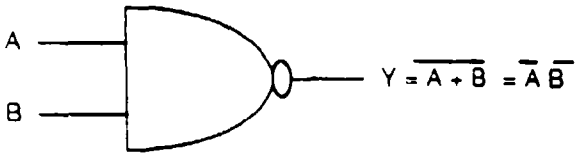
A nanophase realization



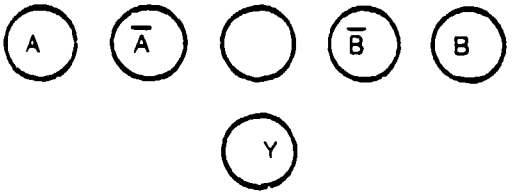
Possible spin configurations of the array

(Truth Table)

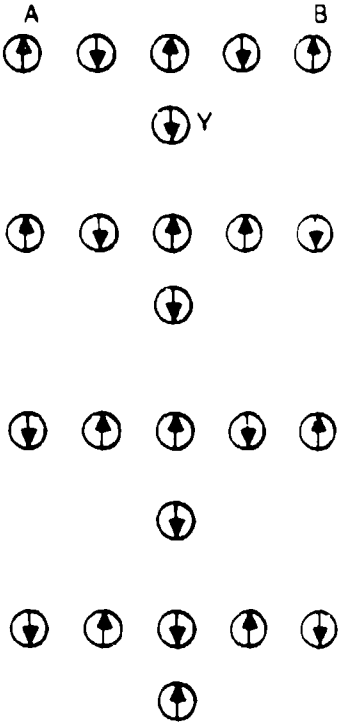
(a)



A NOR gate



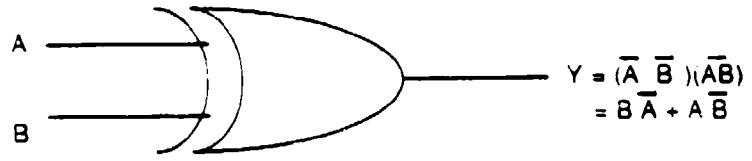
A nanophase realization



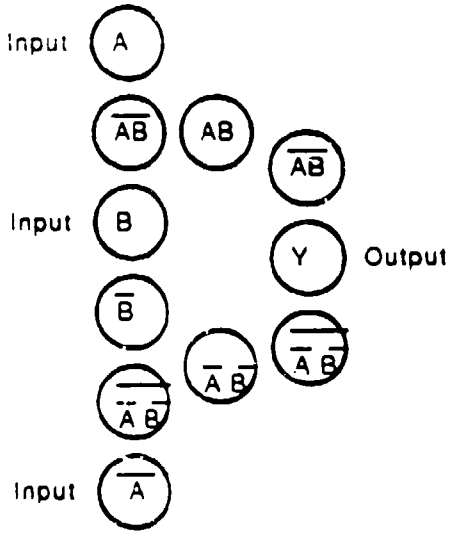
Possible spin configurations of the array

(Truth table)

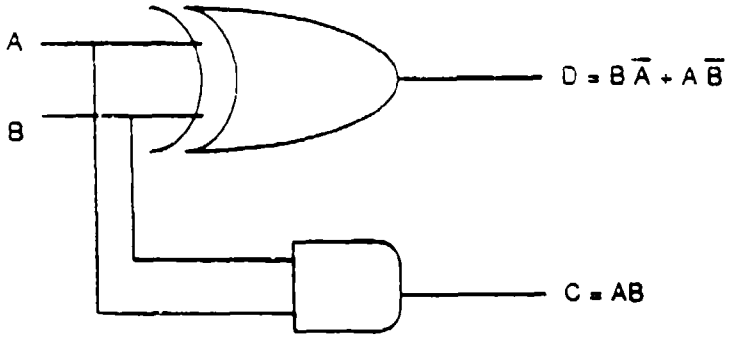
(b)



An exclusive OR gate

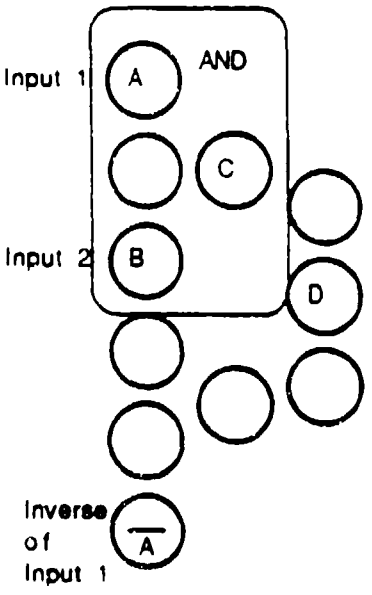


A nanophase realization



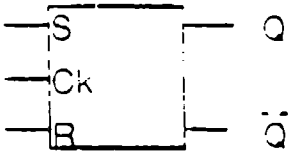
A half adder realized from an exclusive OR and an AND gate

(a)



A nanophase realization. The AND gate is shown within the box

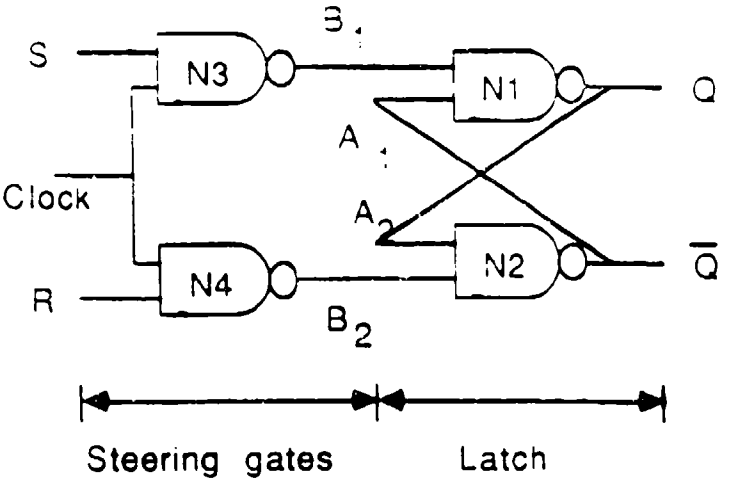
(b)



An SR flip flop

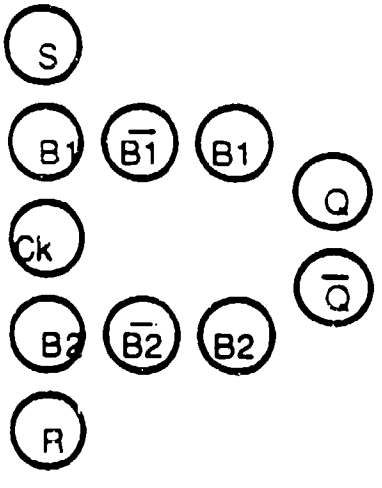
S_n	R_n	Q_{n+1}
0	0	Q_n
1	0	1
0	1	0
1	1	?

Truth table



A NAND gate realization of an SR flip flop

(a)



A nanophase realization

(b)

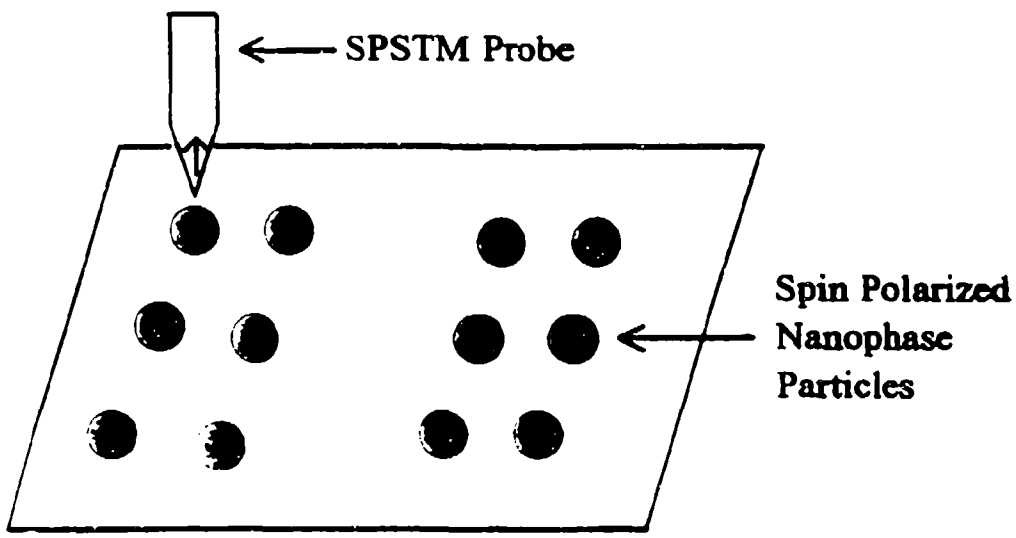
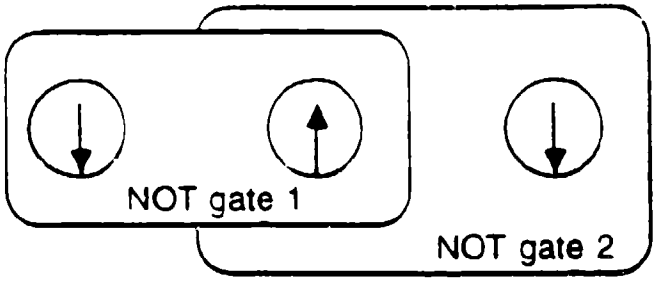
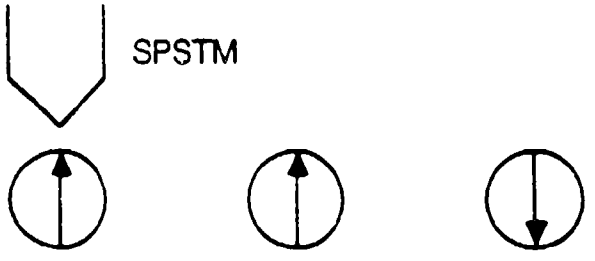


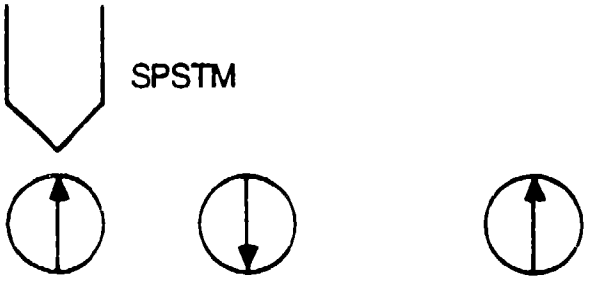
Fig 8



(a)



(b)



(c)

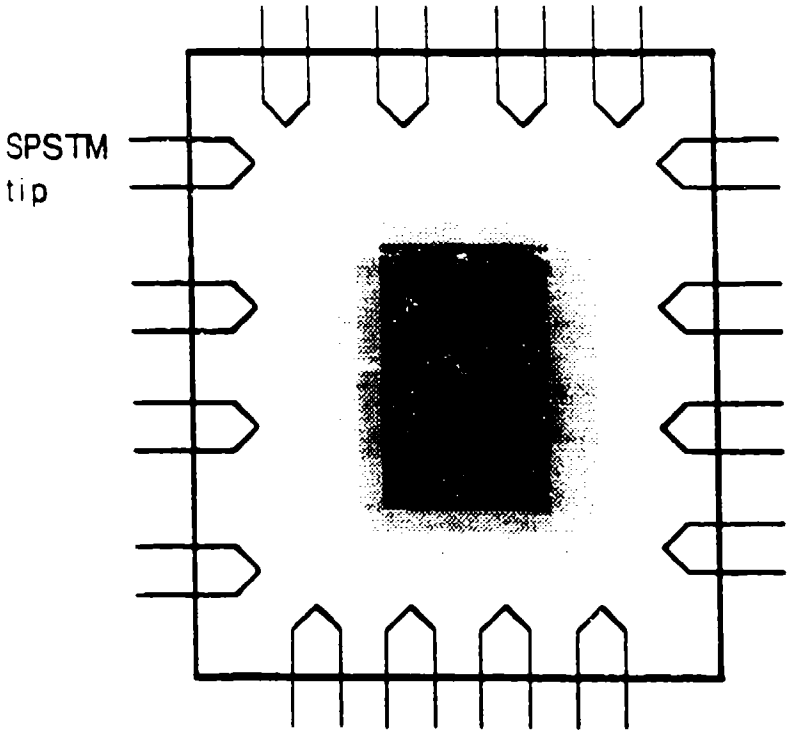
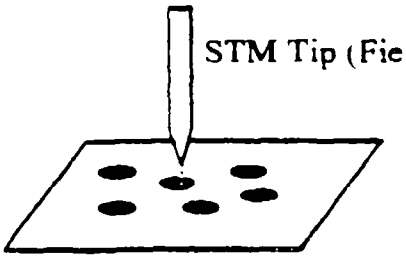
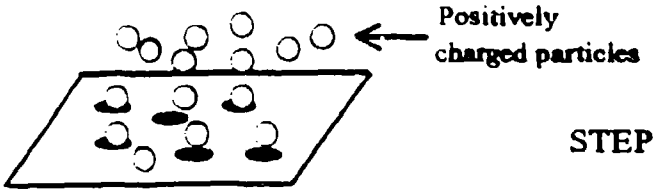


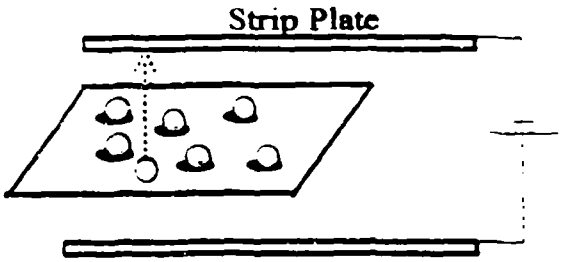
Fig 10



STEP 1. Creation of negatively charged nucleation sites



STEP 2. Deposition of Nanophase Particles



STEP 3. Stripping of Undesired Particles

Fig 12

Spatial Density of Lines Exposed in PMMA by Electron Beam Lithography

by

Xiaokang Huang[1], Gary H. Bernstein[1], Greg Bazán[1], Davide A Hill[2]
University of Notre Dame
Notre Dame, IN 46556

1 Department of Electrical Engineering
2 Department of Chemical Engineering

Abstract- In this paper we report the experimental investigation of the density of patterns exposed with electron beam lithography (EBL). A linear relationship was found between the minimum width of PMMA walls and the given resist thickness (height of the walls). Below the minimum width, PMMA walls become wavy or fail due to the internal stresses caused by PMMA swelling during development. Previous research pertaining to proximity effects at feature sizes and spacing below the 0.25 μm range is limited. We have experimentally investigated proximity effects in very high density gratings in the pitch range from 50 to 330 nm. The relationship between electron beam dose and grating pitch, where proximity effects during electron beam lithography play a major role, was achieved. By fitting the experimental results with a triple Gaussian model, the contributions of the different electron distributions in proximity effects were determined. It was found that fast secondary electrons dominate the proximity effects in the range we studied and they limit the density of patterns fabricated by EBL.

I. INTRODUCTION

The fabrication of nanostructures, and in particular quantum devices, depends on the resolution and achievable pattern density of lithographic techniques. In electron beam lithography (EBL), the lithographic resolution relies on that of the electron beam generator and the contrast of the resist/developer system. Poly (methyl methacrylate) (PMMA) is still the most popular choice of resist for nanostructure fabrication because of its extremely high resolution.

In the lift-off process commonly used in nanofabrication, PMMA exposed with an electron beam is followed by a develop step, after which the unexposed resist remains on the substrate and acts as a shadow for metal evaporation. For very high density patterns, the remaining PMMA is in the form of a thin wall between the developed lines, which serves as the spacer between deposited metal lines. Although PMMA is usually considered to be a non-swelling resist, accounting partially for its high-resolution properties, some absorption of developer is necessary to aid in the development process [1]. A very small amount of swelling can induce internal stresses in thin PMMA walls which weaken its mechanical properties and cause the deformation of thin resist walls during development in the fabrication of very close lines. In the limit of very high resolution lithography, lines are placed so closely together that the resist walls can either become wavy, fall over, or fail completely [2,3]. Therefore, the buckling of the PMMA walls limits the ultimate density of patterns in the EBL process. Although Chen [2] predicted that the limit of line pitch is ~ 75 nm, our results show that for 100-nm-thick PMMA walls, PMMA can be as narrow as 20 nm without resist failure [3], resulting in higher density than expected by them. Sub-50-nm pitch semiconductor-metal photodetectors have also been achieved by Chou et al. [4].

Another serious limitation on very high density patterns with electron beam lithography is the proximity effect. This is due to the distribution of the primary electron beam, forward scattered electrons, secondary electrons scattered in the resist film on the substrate, and backscattered electrons from the substrate. It is important because those scattered electrons expose regions of the resist that are not originally written by the electron beam. For the schematic grating shown in Fig. 1a, the electron beam exposes the pattern on each line. In the ideal

case, the resulting energy distribution in the resist is very spatially confined (Fig. 1b). However, the actual energy distribution for very high density gratings is less distinct [5], as shown in Fig. 1c, and proximity effects decrease the modulation transfer function of the exposure for closely spaced patterns. (The modulation transfer function (MTF) is defined in Fig. 1c.) Assume the peak dose received by each line is E_p and the lowest dose received at the areas between lines is E_v. The resolution of gratings depends on the capability of a given developer to distinguish between E_v and E_p. The distribution of electrons in e-beam lithography is complicated by the fact that it depends on almost all parameters in lithography, e. g. accelerating potential, beam size, photoresist thickness and type, and substrate properties. Investigations of the energy distributions have been reported elsewhere [5-10]. A double Gaussian model has mostly been used for the electron distributions [5,6,11-14]. Computer aided proximity effect correction and other methods reducing the proximity effects during e-beam lithography have been pursued [15-20]. The feature size and spacing in most of the published work are in the half-micron regime, which is quite useful for fabricating ULSI circuits. In this case, backscattered electrons dominate the proximity effects because of the typically low beam energies utilized (~ 25 keV). The width of the backscattered electron distribution is in the order of the feature size and spacing. However, previous research pertaining to proximity effects at feature sizes and spacing below the 0.25 μm range is limited [9-11].

In this paper, we report the experimental investigations of the density of the electron beam lithography limited by the physical strength of PMMA and proximity effects. Proximity effects in very high density gratings were experimentally studied in the pitch range from 50 to 330 nm. We will discuss the relationship between electron beam dose and grating pitch, where proximity effects during electron beam lithography play a major role.

II. EXPERIMENTAL PROCEDURE

The EBL system used in this study consisted of an Amray 1400 SEM with a maximum beam energy of 50 keV and a minimum beam width of 5 nm using a W cathode, controlled by an IBM PS/2 personal computer interfaced through a Pragmatic Instruments 2201A 16-bit

arbitrary waveform generator [21]. Special care was taken to minimize all noise sources. PMMA (950,000 a. m. u.) was spun on Si and SiO_2/Si wafers and baked at 170 °C for 4 hours. Mixtures of methyl isobutyl ketone (MIBK):2-propanol (IPA) (1:3) with the addition of 1% and 1.5% methyl ethyl ketone (MEK) by volume, were used to develop samples. This mixture has been shown to have very high development contrast properties [22]. For both the experiments on proximity effect and the strength of PMMA, different process parameters were used. A summary of the experimental parameters is given in Table 1. The resist thickness in the proximity effect measurement was kept to 50 ~ 70 nm in order to avoid PMMA failure during the processes. Gratings with different pitch in the proximity effect experiments are 3 μm x 3 μm in size and separated by 2 μm with each other in order to eliminate the proximity effects caused by the scattered electrons from adjacent gratings [23].

III. RESULTS AND DISCUSSION

A. Strength of PMMA

In order to achieve lift-off of evaporated metal films, it is necessary that PMMA walls remain between the developed areas. The width of PMMA walls was determined by measuring the spaces between resulting metal patterns after lift-off. The undercutting for thin resists was neglected because it is small for thin resists and high beam energies. The height of the walls after development is assumed to be the same as the initial resist thickness because the unexposed PMMA of very high molecular weight has very low solubility [22]. One might suppose that in the case of high density gratings, areas between lines might be partially developed because of proximity effects. For grating pitch larger than 70 nm, we find in the next section that $E_v/E_p < 0.6$. For the developer with contrast, γ^*, larger than 10, the amount of the resist that is removed through developing is small compared with the original thickness of the resist, and therefore can be ignored in the final resist thickness determination.

PMMA can be considered a strong, glassy material after baking. However, during development, the absorption of the developer causes swelling, and internal stresses result in the deformation of PMMA walls. The wall aspect ratio (height-to-width) is an important parameter

in delineating the stability threshold, since a tall, thin wall will be strongly susceptible to small dimensional perturbations around the erect position, while a short and thick wall will be able to sustain a higher degree of swelling before buckling.

Figure 2 shows experimental results of the minimum width of a PMMA wall achievable for a given starting resist thickness. Error bars represent variations in metal line separation (and therefore wall thickness) over the length of the lines. These data were obtained by smoothly varying the line pitch for a given resist thickness and observing the point at which lines either failed completely or became noticeably wavy. The data indicate a direct proportionality between critical wall height and width, implying a constant aspect ratio for buckling. The area above the critical line is the regime in which straight PMMA walls can be achieved. The area below it is the regime in which the walls become wavy or fail.

The stability of an infinitely long wall of width, w, and height, h, with a rigidly anchored bottom and a top free end can be modeled with linear plate theory [24]. The details of the theoretical calculation are given in Ref. [25]. According to this model, the critical wall aspect-ratio for stability, $(h/w)^*$, is given as a function of the swelling strain:

$$\left(\frac{h}{w}\right)^* = \sqrt{\frac{3}{5(1-\nu^2)\epsilon_{sw}}} \qquad (1)$$

where ν is Poisson's modulus and ϵ_{sw} is the hypothetical swelling strain that the wall would undergo if allowed to expand freely. For $(h/w) > (h/w)^*$ the wall will buckle as shown in Fig. 3a. The bright areas in Fig. 3 are evaporated gold, and the dark areas are the areas where PMMA walls stood before lift-off. Equation (1) predicts a linear relationship between the critical wall height and width for a constant ϵ_{sw}, which is consistent with our experimental results as shown in Fig. 2. From the slope of the line in Fig. 2, the critical aspect ratio for stability can be calculated, giving $(h/w)^* = 4.75$. Then $\epsilon_{sw} = 5.1\%$ for $\nu = 0.5$ from Eqn. (1). This indicates that 5% swelling can lead to failure of PMMA walls with aspect ratio greater than 5.

In making gratings, a set of parallel PMMA walls forms during development. The

buckling of one of the walls can lead to the catastrophic failure of the whole grating, because surface forces can induce cross-correlation between deformation of adjacent structures, producing a pattern of adjacent sinusoidal waves 180° out-of-phase with respect to each other as shown in Fig. 3b. Figure 3c shows a 50-nm-pitch grating with straight resist walls between lines.

B. Proximity Effects

Proximity effects must be strongly considered in the fabrication of very high density patterns. It can be expected that when lines get closer together, the dose needed for exposure should decrease because inter-line scattering of electrons causes additional energy deposition. Experimental results of normalized dose versus grating pitch are shown in Fig. 4 along with theoretical simulations. In the figure, doses for various pitch values are normalized to that needed for the same linewidth in the 150 nm pitch grating. Three samples with different doses and development time were measured. The data of all three samples are very consistent. Besides exposure dose, many factors affect line width in experiments, such as contrast of developers, developing time and temperature. Our results show that although line width may change with other factors, their effects on the relationship between normalized dose and grating pitch are small and can be neglected. It can also be seen that normalized dose decreases slowly with pitch until about 100 nm when the change becomes rapid. As we will show, this effect is due to the different distributions of scattered electrons in the resist during exposure. Models of the exposure energy distribution in the resist employed for fitting the experimental data include the double Gaussian [6]:

$$f(r) = \frac{1}{\pi(1+\eta)}\left[\frac{1}{\alpha^2}\exp\left(-\frac{r^2}{\alpha^2}\right) + \frac{\eta}{\beta^2}\exp\left(-\frac{r^2}{\beta^2}\right)\right], \qquad (2)$$

and the triple Gaussian model [9]:

$$f(r) = \frac{1}{\pi(1+\eta)}\left[\frac{1}{\alpha^2}\exp\left(-\frac{r^2}{\alpha^2}\right) + \frac{\eta}{\beta^2}\exp\left(-\frac{r^2}{\beta^2}\right) + \frac{\eta'}{\gamma^2}\exp\left(-\frac{r^2}{\gamma^2}\right)\right]. \qquad (3)$$

The simulation curves are shown in Fig. 4. In Eqns. (2) and (3), the first term is used to

describe the primary beam and forward scattered electron distribution. The second term accounts for the behavior of the backscattered electrons. The third term in the triple Gaussian model is added to describe all other exposures that are not included in the first two terms. The third term may result from large angle forward scattered electrons, secondary electrons, or even broad tails in the primary beam distribution [9,10]. The results of Monte Carlo calculations by Joy [5] and Murata et al. [7] clearly show broad tails to the energy distribution due to fast secondary electrons, which we will show play an important role in proximity effect calculations over the spatial range discussed here. In Eqns. (2) and (3), α, β, and γ are the widths of each Gaussian distribution, and η and η' are the ratios of exposures of the second and third term to the forward exposure, respectively. According to Ref. [8], $\beta = 10$ μm and $\eta = 0.8$ for 50 keV electrons in solid silicon substrates. α, γ, and η' were chosen for the apparent best fit. The dashed line in Fig. 4 shows the calculation in which only the first Gaussian term, that is, only the contribution of forward scattered electrons, was considered. Since the width of this term is only a few tens of nanometers, forward scattered electrons only cause a short range proximity effect and cannot change with dose over large pitch.

When the second Gaussian term was added in the calculation (the double Gaussian model), the weak change of dose over large changes in pitch (solid line in Fig. 4) was caused by the long range proximity effect of backscattered electrons. The small difference between the dashed curve and the solid curve shows that the proximity effect caused by backscattered electrons at 50 keV is weak. $\alpha = 0.037$ μm was used in these two curves. The dash-dot line in Fig. 4 was calculated with the triple Gaussian model. This curve fits the data very well with $\alpha = 0.029$ μm, $\gamma = 0.4$ μm, and $\eta' = 0.35$. We conclude, therefore, that the proximity effect from the γ-term electrons is very important in high density patterns in EBL. Since these electrons cause proximity effects in a broad range relative to the forward scattered electrons, and their sources are still not clear, we call them simply "broad range electrons."

The ratios of the contributions of scattered electrons from the inter-line exposures to the total exposure received by each line are shown in Fig. 5. The solid line, the dashed line, and dash-dot line correspond to the inter-line exposure contributions of forward scattered electrons,

backscattered electrons, and broad range electrons, respectively. From Fig. 5, we see that for the inter-line proximity effects: (i) The contribution of backscattered electrons weakly depends on the grating pitch because of their wide distribution (~ 10 μm [8,26]) at 50 keV. The backscattered coefficient is a constant for certain materials when the beam energy is larger than 20 keV [27]. Hence the intensity of the backscattered electron distribution for 50 keV is much lower than for 20 keV because it spreads much further. Therefore, for high density patterns, the proximity effect of backscattered electrons is like a weak background. It can be ignored in the total exposure because the contribution of backscattered electrons is less than 2% and much smaller than the contribution of the broad range electrons in the whole pitch range over which our calculations were performed. (ii) The effect of forward scattered electrons is not important until the pitch of gratings approaches the width of its distribution. Their effect increases dramatically when the pitch are smaller than 50 nm. Therefore, the proximity effect of forward electrons will limit the resolution of electron beam lithography. (iii) Broad range electrons dominate the proximity effect in the pitch range from 50 to 350 nm or even larger in our experiments. This range is on the order of feature sizes of quantum devices and ULSI circuits. Therefore it is very important to include this term in the proximity effect correction. Since the contribution of the broad range electrons to the total exposure is larger than 20% when grating pitch are less than 50 nm, their limitation on the resolution and density of electron beam lithographically defined patterns cannot be ignored.

We further consider here the role of fast secondary electrons in proximity effects. Fast secondary electrons have a complicated distribution; it is a narrow peak with a wide shoulder [5,7]. The width of the narrow peak is about 10 nm in free standing resist. Its role in proximity effects was included in the forward scattered electrons [5] which are described by the first term of the double or triple Gaussian models. However, the proximity effect caused by the wide shoulder was neglected by previous researchers. From [5] we note that the half width of the shoulder is about 100 nanometers and is on the order of the distribution of the broad range electrons. Therefore, it can be concluded that fast secondary electrons are part of the broad range electrons, but they are not the only part. Therefore, the fast secondary electrons introduce

both short and long range proximity effects, and thereby limit the resolution of EBL.

As mentioned above, the total dose at the region between the electron beam scanned areas increases when lines are placed closer together. The resolution of gratings created with EBL depends on how well a developer can distinguish between the deposited energy densities between lines and at the lines, E_v and E_p respectively. Fig. 6 shows E_v/E_p as a function of grating pitch with different Gaussian parameters. The dotted line results from the double Gaussian model with the same parameters used in the calculation of Fig. 4. For 50-nm-pitch gratings, the ratio E_v/E_p is about 0.98. This small difference between exposure energies cannot be distinguished with any practical developers, in contradiction to the experimental results. Therefore, although the double Gaussian model can be used to qualitatively simulate the normalized dose changing with grating pitch, it fails in predicting the density and resolution limitation in electron beam lithography at these very small dimensions. The solid line is calculated with the triple Gaussian model with the same parameters that fit the experimental data in Fig. 4. It can be seen that E_v/E_p approaches unity when the grating pitch decrease. $E_v/E_p = 1$ means that the grating patterns are completely washed out because of the proximity effects. For 50-nm-pitch gratings, $E_v/E_p \approx 0.87$. The contrast, γ^*, of our developer (1.5% MEK) is larger than 10 and corresponds to $D_i/D_f \approx 0.8$ [22], where D_i is defined as the critical dose below which the resist is not developed at all, and D_f is the dose above which the resist is totally dissolved. (Note that the exposure energy, E, is proportional to the received dose, D, so contrast curves yield useful information about the limits on absorbed energy.) The resist will be partially dissolved when the dose is less than D_f, but greater than D_i. This means that the height of the PMMA walls was decreased for 50-nm-pitch gratings during development because of partial developing. Since the initial PMMA thickness in our case was 60 ~ 70 nm, the height of remaining PMMA walls is assumed to be larger than 40 nm, which we have found is sufficient for good lift-off of 15-nm metal patterns. This also shows that the development is very critical in making very high density patterns; a slight over-development may cause the PMMA walls to be dissolved.

From Fig. 5, we saw that forward scattered electrons, including fast secondary electrons, cause the proximity effects to dramatically increase when grating pitch approach α, the width

of the distribution of forward scattered electrons. Thus the resolution of EBL can be improved when α decreases. This is clearly shown by the dashed line ($\alpha = 20$ nm) and the dash-dot line ($\alpha = 10$ nm) in Fig. 6. This requires using a very high energy system with a non-Gaussian-shaped beam or a resist that depends less on low energy electrons in forward electron scattering. Assuming $\alpha = 10$ nm for a very high resolution electron beam, a grating with 20 nm pitch can be expected using a developer with the contrast $\gamma^* = 10$, as shown in Fig. 6. This agrees with predictions of Joy [5] that 10 nm represents the ultimate resolution for PMMA as electron beam resist. Our results restate those of Joy that the narrow peak of the energy deposited by secondary electrons is practically flat out to a distance of 5 nm, and therefore so low in contrast (over the peak) as to be indistinguishable to developers with reasonable contrast values.

We saw that the pattern density is also limited by the failure of PMMA walls. As discussed above, the lower the achieved thickness of the PMMA walls, the closer is the achievable line spacing. However, the resist thickness cannot be lowered arbitrarily because, for good lift-off, the resist thickness should be about 3 times the metal thickness, which depends on the ultimate desired device characteristics. According to Fig. 2, 10-nm spacing between metal lines can be achieved with 40-nm-thick PMMA, and is sufficient for successful lift-off of 10- to 15-nm-thick metal lines. Therefore, we conclude that the limits to the density of metal lines due to the proximity effect and the strength of PMMA walls are consistent for the achievement of lines with pitch as small as 20 nm.

IV. SUMMARY

The proximity effects in electron beam lithography were experimentally investigated with different grating pitch. The results were successfully simulated with a triple Gaussian model. Proximity effects caused by backscattered electrons were insignificant in our case because of their wide distribution. Forward scattered electrons were very important when the grating pitch were about the width of this electron distribution. Broad range electrons, described by the third term in the triple Gaussian model, dominate the proximity effect in the pitch range from 50 nm

to 350 nm. Secondary electrons cause both short and long range proximity effects. Their effects can be included in forward scattered electrons as well as broad range electrons. Their limitation on the EBL resolution cannot be ignored.

Upon exposure to suitable solvents, such as common developing and rinsing solutions, narrow PMMA walls were found to become unstable as a result of two effects: (1) swelling, and (2) lateral surface forces between adjacent structures. We investigated the relationship between minimum line spacing and maximum PMMA thickness for EBL fabrication using a high contrast developer. We also showed experimental evidence that supported our predictions, thus demonstrating the existence and severity of such instabilities in nanolithography.

From our calculation of E_v/E_p as a function of grating pitch with the triple Gaussian model, a 20-nm-pitch grating can be expected using a very high resolution electron beam generator and high contrast resist/developer system. The closest spacing between metal lines can be as narrow as 10 nm for 10- to 15-nm thick metals, implying that 20 nm gratings with 10-nm line width and 10-nm line spacing are achievable.

ACKNOWLEDGMENTS

We thank D. T. Leighton for helpful discussions, and S. Subramaniam for help in sample preparation and useful discussion. We gratefully acknowledge IBM, AFOSR, ONR and the University of Notre Dame for their support of this research. D. A. H. thanks the Donors of the Petroleum Research Fund, administered by the American Chemical Society, for partial support of this research.

References

1. A. Weill, in The Physics and Fabrication of Microstructures and Microdevices, ed. by M. J. Kelly and C. Weisbuch, Springer-Verlag, New York, p. 58 (1986).
2. Zhong Wei Chen, in Advances in Electronics and Electron Physics, 83, ed. by Peter W. Hawkes, Academic Press, New York, p. 187, (1992).
3. Xiaokang Huang, Greg Bazán, Davide A. Hill, and Gary H. Bernstein, to appear in the *J. of the Amer. Chem. Soc.*
4. S. Y. Chou, Y. Liu, and P. B. Fischer, *IEDM-91*, 745 (1991).
5. David C. Joy, *Microelectronic Engineering*, **1**, 103, (1983).
6. T. H. P. Chang, *J. Vac. Sci. Technol.*, **12**(6), 1271 (1975).
7. Kenji Murata, David F. Kyser, and Chiu H. Ting, *J. Appl. Phys.*, **52**(7), 4396 (1981).
8. R. M. Mankiewich. L.D. Jackel, and R. E. Howard, *J. Vac. Sci. Technol.*, *B* **3**(1), 174 (1985).
9. S. J. Wind, M. G. Rosenfield, G. Pepper, W. W. Molzen, and P. D. Gerber, *J. Vac. Sci. Technol.*, *B* 7(6), 1507 (1989).
10. R. J. Bojko, B. J. Hughes, *J. Vac. Sci. Technol.*, *B* **8**(6), 1909 (1990).
11. S. A. Rishton, D. P. Kern, *J. Vac. Sci. Technol.*, *B* **5**(1), 135 (1987).
12. A. S. Chen and A. R. Neureuther, *J. Vac. Sci. Technol.*, *B* **3**(1), 148 (1985).
13. Akio Misaka, Kenji Harafuji and Noboru Nomura, *J. Appl. Phys.*, **68**(12), 6472 (1990).
14. Geraint Owen, *J. Vac. Sci. Technol.*, *B* **8**(6), 1889 (1990).
15. Geraint Owen and Paul Rissman, *J. Appl. Phys.*, **54**(6), 3573 (1983).
16. S. W. J. Kuan, C. W. Frank, Y. H. Yen Lee, T. Eimori, D. R. Allee, R. F. W. Pease, and R. Browning, *J. Vac. Sci. Technol.*, *B* 7(6), 1745 (1989).
17. D. G. L. Chow, J. F. McDonald, D. C. King, W. Smith, K. Molnar, and A. J. Steckl, *J. Vac. Sci. Technol.*, *B* **1**(4), 1383 (1983).
18. R. L. Kostelak, E. H. Kung, M. G. R. Thomson, and S. Vaidya, *J. Vac. Sci. Technol.*, *B* **6**(6), 2042 (1988).

19. E. Nishimura, T. Takigawa, T. Abe, and Y. Katoh, *J. Vac. Sci. Technol.*, *B* **4**(1), 164 (1986).

20. Robert C. Frye, *J. Vac. Sci. Technol.*, *B* **9**(6), 3054 (1991).

21. Greg Bazán and Gary H. Bernstein, submitted for publication.

22. Gary H. Bernstein, Davide A. Hill, and Wen-Ping Liu., *J. Appl. Phys.*, **71**(8), 4066, (1992).

23. K. L. Lee and H. Ahmed, *J. Vac. Sci. Technol.*, **19**(4), 946 (1981).

24. S. P. Timoshenko, J. M. Gere, Theory of Elastic Stability, 2nd ed., p. 363, McGraw-Hill, New York (1961).

25. Davide A. Hill, Xiaokang Huang, Greg Bazán, and Gary H. Bernstein, to appear in *J. Appl. Phys.*

26. J. S. Greeneich, *J. Vac. Sci. Technol.* **16**, 1749 (1979).

27. Ludwig Reimer, in Scanning Electron Microscopy, Springer-Verlag, New York, p. 137 (1985).

Figure Captions

Fig. 1. Exposure energy distribution in very high density gratings. (a) Schematic of the beam scan for a grating. (b) Energy distribution in the near ideal case for the beam scan shown in (a). (c) Actual energy distribution in very high density gratings.

Fig. 2. Minimum width of PMMA walls achievable versus starting resist thickness (wall height).

Fig. 3. Remaining metal patterns after lift-off. (a) Failure of parallel lines caused by wavy PMMA walls between them. (b) A typical failure of gratings resulting from PMMA wall swelling and buckling. (c) A 50-nm grating with straight PMMA walls between lines. Light areas are gold, and dark areas are the shadows of PMMA walls which have been removed during lift-off.

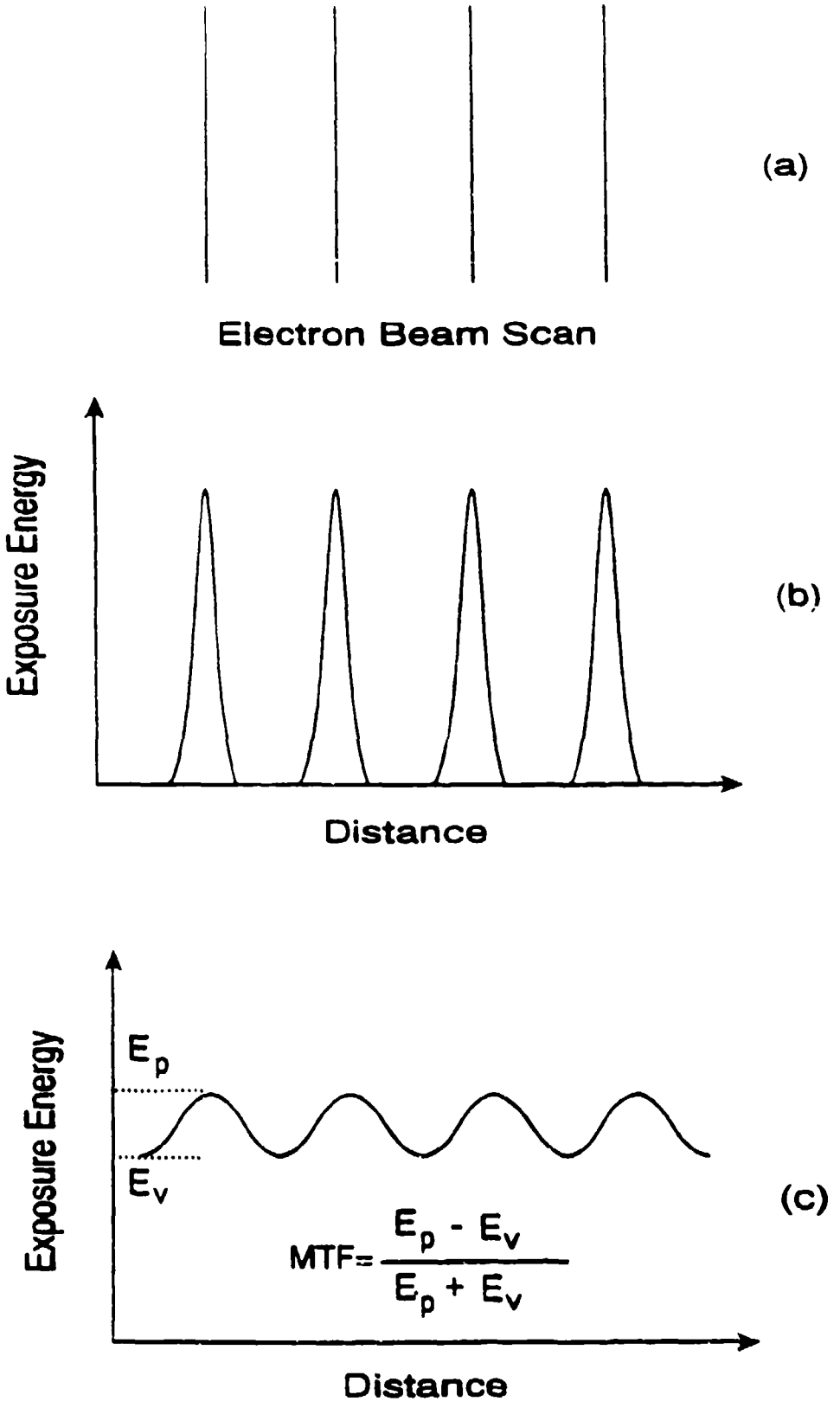
Fig. 4. The relationship between normalized dose and grating pitch. Doses are normalized to the dose required for the same line width at a pitch of 150 nm.

Fig. 5. Inter-line proximity effects caused by scattered electrons versus pitch of gratings.

Fig. 6. E_v/E_p versus grating pitch. E_v is the lowest exposure energy received in the areas between lines. E_p is the maximum exposure energy received by each line.

Table 1. Experimental Parameters.

Experiment	Proximity effect	Strength of PMMA
Patterns	Arrays of gratings	Two parallel lines or gratings
Line Pitch	50 - 330 nm	70 - 150 nm
Accelerating Voltage	50 kV	45 kV
Resist Thickness	60 - 70 nm	105 - 180 nm
Developer	MIBK:IPA:MEK =1:3:0.06 (1.5% MEK)	MIBK:IPA:MEK =1:3:0.04 (1% MEK)
Developer Temp. (°C)	26	25
Developing Time	6 - 10 seconds	15 seconds
Metal Thickness	15 nm	30 nm


(a)
Electron Beam Scan
Exposure Energy
(b)
Distance
Exposure Energy
E_p
E_v
(c)
$MTF = \frac{E_p - E_v}{E_p + E_v}$
Distance

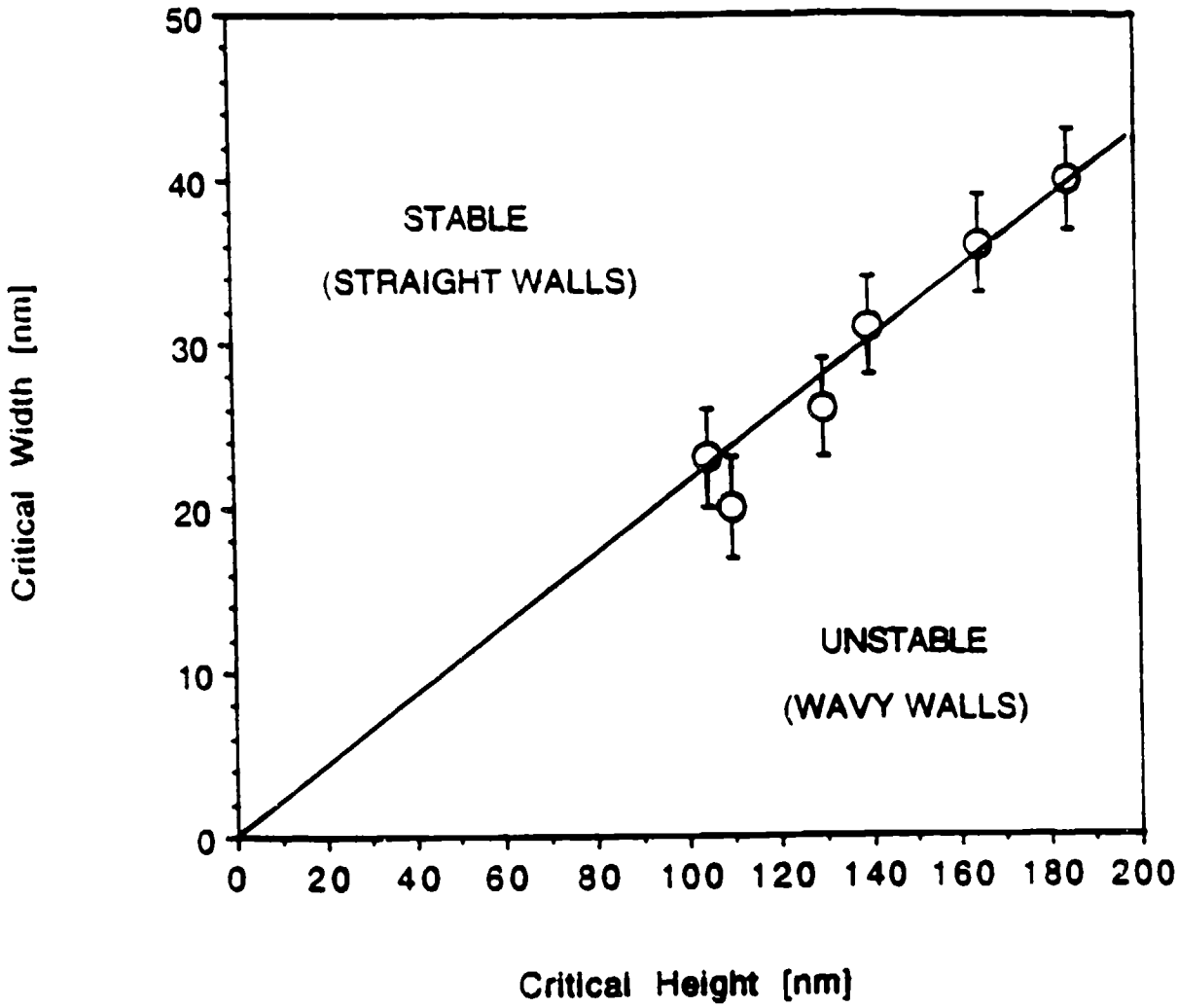
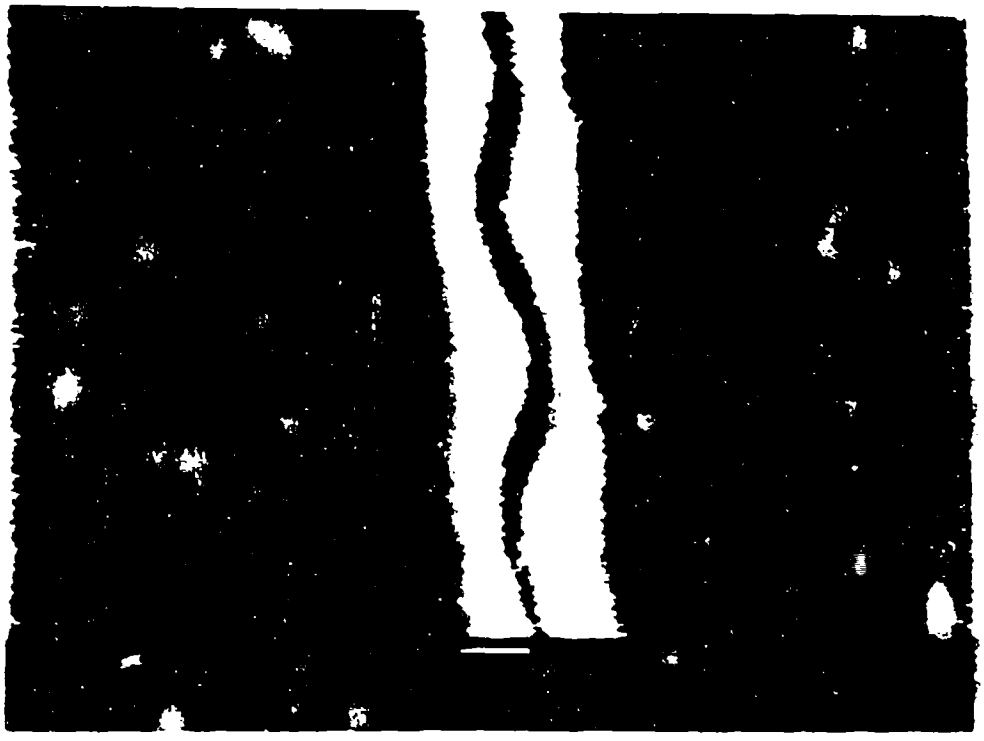
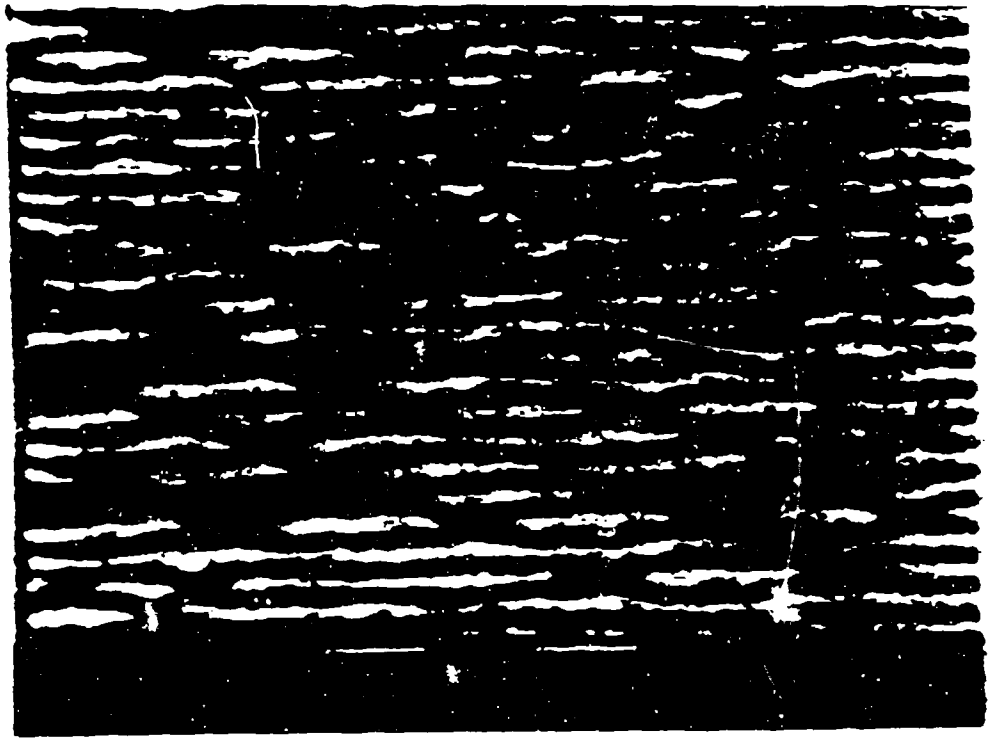


Fig. 2 Hwang et al.



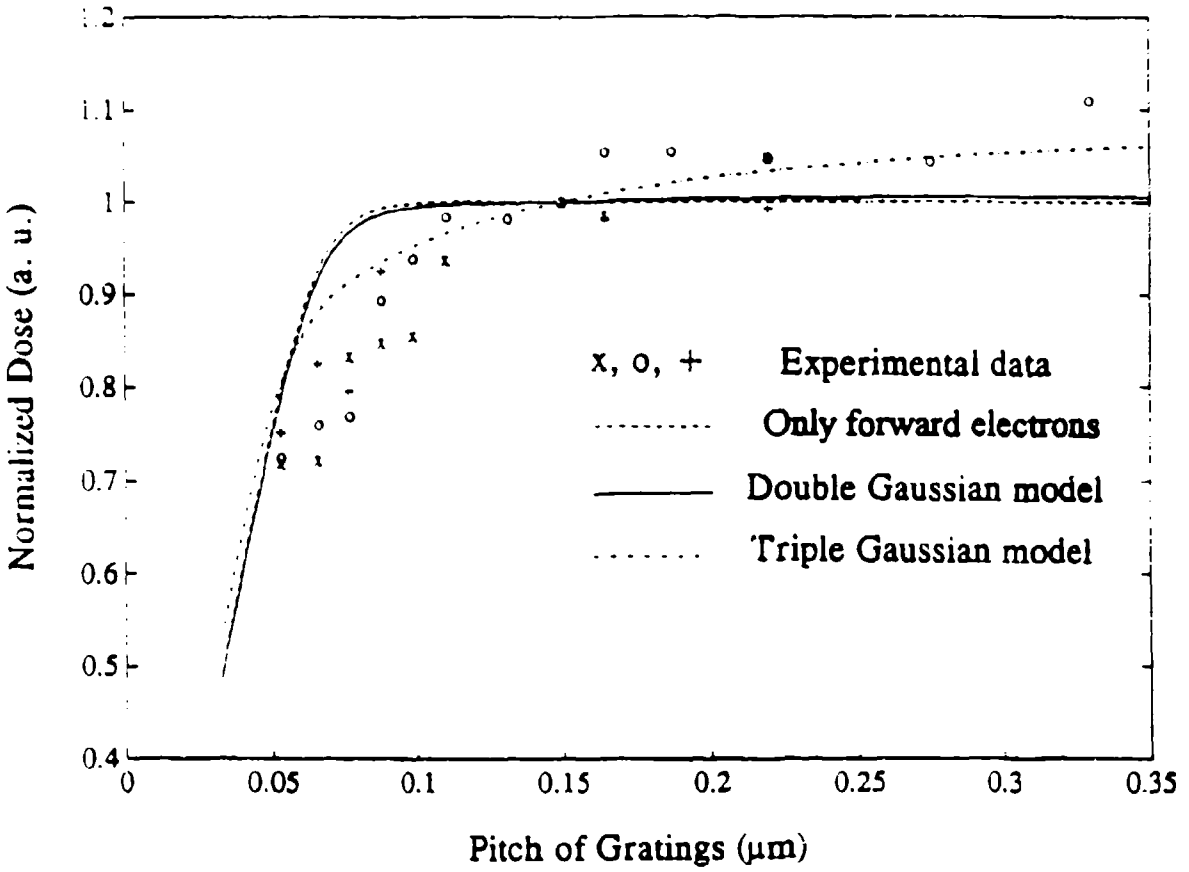
(a)



(b)

(c)

Fig. 3 c


Normalized Dose (a. u.)
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
Pitch of Gratings (μm)
x, o, +
Experimental data
Only forward electrons
Double Gaussian model
Triple Gaussian model

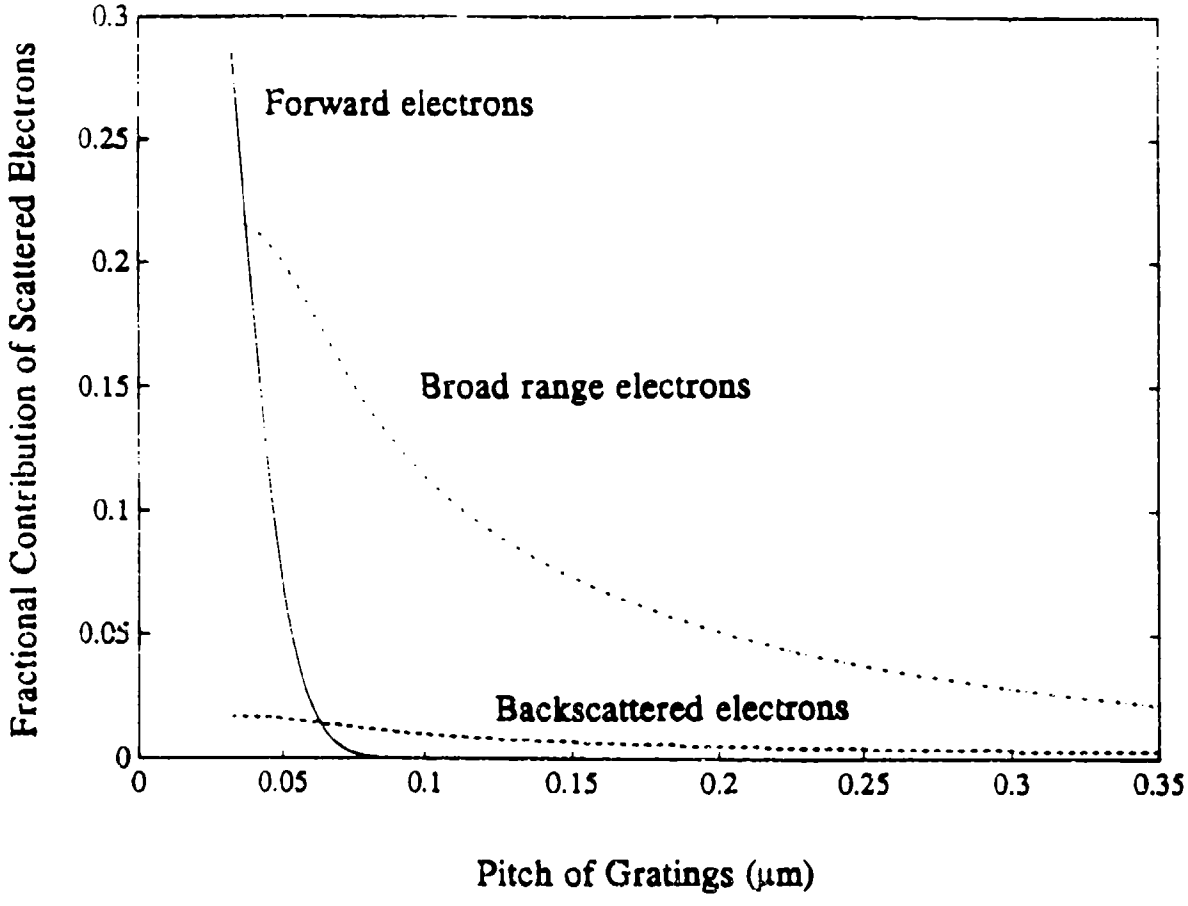
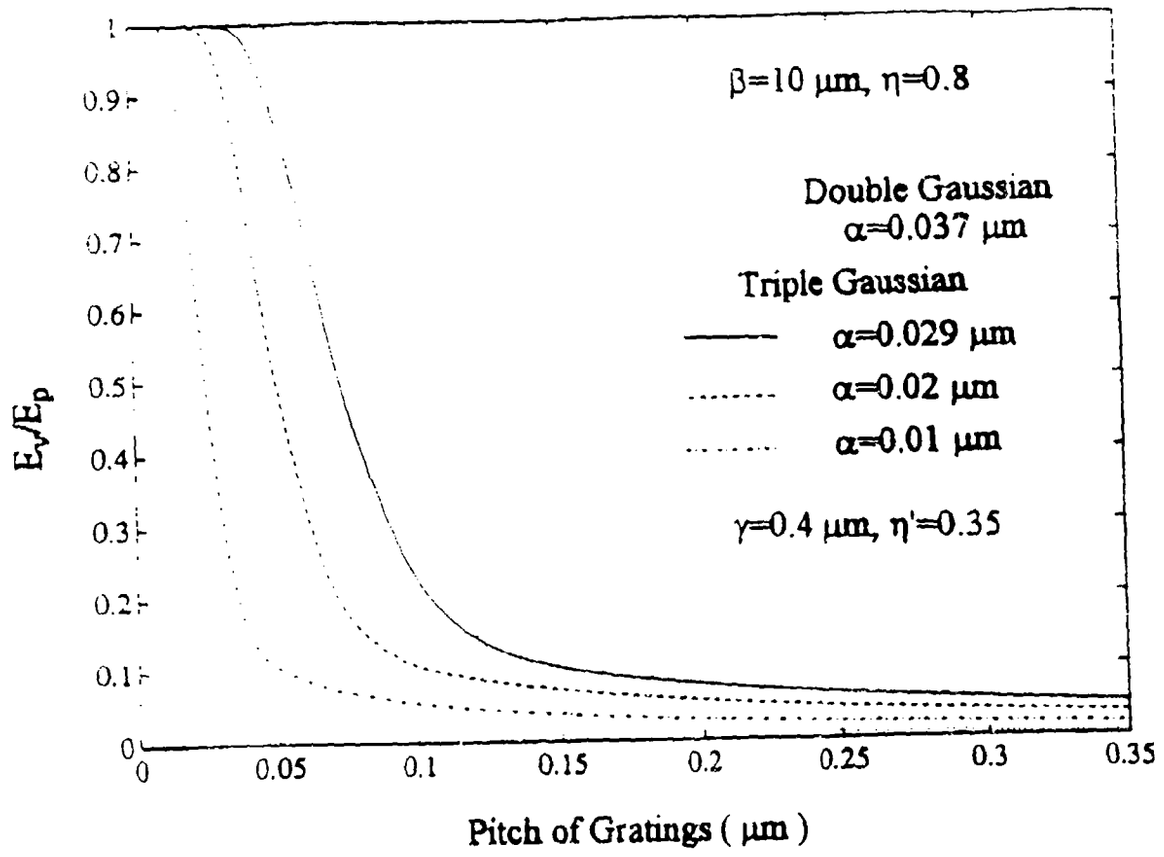


Fig. 5 Huang et al


β=10 μm, η=0.8
Double Gaussian
α=0.037 μm
Triple Gaussian
α=0.029 μm
α=0.02 μm
α=0.01 μm
γ=0.4 μm, η'=0.35
E_v/E_p
Pitch of Gratings (μm)
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35

Electron Beam Lithography Over Large Scan Fields

Greg Bazan and Gary H. Bernstein*
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

Abstract - We have developed an advanced, research-oriented electron beam lithography system using a 3 channel arbitrary waveform generator and a scanning electron microscope. The system is capable of exposing patterns with features less than 100 nm over scan fields greater than 1.0 mm, as well as features less than 20 nm in smaller scan fields. Other system features include capabilities for registration and for generating exposure matrices for evaluating new electron beam resists and optimizing exposure parameters. We report the details of the system and show results made possible by our system.

*corresponding author

Introduction

For more than a decade, intense effort has been put towards the construction of dimensionally confined systems for the study of quantum effects in mesoscopic structures [1]. To this end, simple devices have been constructed using processing methods similar to those used to fabricate GaAs MESFETs. To study quantum effects, minimum feature sizes within these structures must be less than 1 μm, and often less than 100 nm [2-4]. Several approaches for fabricating structures with features less than 50 nm have been implemented [5-7], but none combines the versatility and resolution capabilities of electron beam lithography (EBL) systems. X-ray lithography systems are capable of such resolution but are not readily available. Some commercial EBL systems have the resolution necessary to produce the required sizes yet are not readily available due to the their cost. Using scanning electron microscope (SEM)-based EBL systems can provide a low-cost, versatile, research exposure tool whose only disadvantage is throughput. Other cost effective methods include the modification of scanning tunnelling microscopes (STM) [8], but this technology currently is limited by the overhead exposure time and the small exposure area.

Several EBL system designs have been implemented and described by others [9-20]. However, our work extends the utility of this technique into the regime of very large scan fields with nanometer resolution. Our motivation behind the development of this system was the ability to fabricate nanostructures and nano-devices incorporating all exposure steps (mesa, contacts, gates, bonding areas) into one tool as well as gate level exposure for quantum based circuit architectures [21]. Also recent optical experiments require nanostructures in large scan fields due to the difficulty in focusing lasers onto small areas [22]. We designed our system with three specific goals. The first was to build a system capable of exposing any arbitrary pattern with a resolution less than 100 nm over scan fields as large as 1.4 mm. The second was to allow the exposure tool to perform accurate pattern registration of 100 nm over scan fields as large as 1.4 mm. Finally, the system was designed so that

cumulative exposure parameters could be stored for future analysis and images printed and saved easily.

The most common method used to convert an SEM to an EBL system incorporates a computer fitted with two digital to analog converters (DACs) whose outputs are connected through amplifiers to the SEM scan coils. More sophisticated conversions involve computer controlled beam blanking and stage movement. Nabity and Wybourne [9] give an especially thorough review of the design issues involved in an SEM-based EBL system. In all cases, the computer controls the beam position by the application of voltages via the DACs. Ideally, the beam position resolution of these DAC-based systems is strictly limited by the DAC resolution. Earlier designs used 12-bit DACs while most current systems use 16-bit DACs and, presumably, future systems will have even higher resolution. When 16-bit resolution is used, the beam can be positioned anywhere on a 65,536 by 65,536 point grid. The spatial resolution of this grid is determined by the SEM magnification. To move the beam, successive positions in the form of digital integers or pixels are input to the DAC at a predetermined rate calculated from the feature size, the exposure dose, and the beam current. This rate is calculated to include the exposure time at that location as well as any overhead time necessary to account for scan coil settling [9-11].

The minimum feature size of such systems is not only determined by the DAC spatial resolution, but also by the temporal resolution. For systems with poor temporal resolution, a low exposure dose can only be achieved by having fewer pixels in an exposure to result in the desired dose at a given magnification and beam current. In this case, if the pixel spacing is large compared to the beam diameter a discrete pixel exposure will result. Therefore, if faster pixel outputs are possible, pixels can be placed closer together on the grid resulting in smoother lines, more uniform exposure doses, and improved overall performance.

System Overview

Control of the lithography system is achieved through an in-house software program [23] written specifically for electron beam lithography. The program we developed is capable of generating two types of files: layout exposure and alignment (registration). Each file must be assigned a magnification which is used to determine the scan field by the relation that an SEM magnification times the scan field equals a system constant. Our system constant has been set at 140,000 where the scan field is in units of microns. Having such a large system constant allows for the possibility to expose patterns over 7 mm. Both files allow for an exposure to occur only if the beam current and dose are specified. A layout file can consist of dot, line, and area exposures defined as a set of primitive shapes. Currently the set of shapes includes dot grids, angled lines, gratings, grids, rings, rectangles, parallelograms, and triangles. The registration process involves continuously scanning or imaging any features designated as the alignment marks. These features must be positioned at the corners of a rectangle so that adjustments can be made to reduce rotational, horizontal, and vertical overlay errors as well as to select the correct magnification. These marks are scanned simultaneously or individually through a window whose size can be varied to improve the alignment precision.

A complete schematic diagram of our lithography system is shown in Fig. 1. The lithography system is controlled by an IBM PC Model 30/286 [24] equipped with a 32 megabyte hard drive, 4 megabytes of RAM, a Hewlett-Packard GPIB interface card, and a Computer Boards, Inc. CIO-DIO24 digital input/output card. The GPIB interface controls the operation of the waveform generator and the CIO-DIO24 card monitors the TTL channel synchronization outputs of the waveform generator and controls the synchronization of the beam blanker. Connected to the computer through the serial port is a switch box used to direct information to a plotter for layouts or to the SEM stage control system. The computer is also connected to a laserprinter to record SEM exposure parameters as well as the size and position of exposures within a layout. Our SEM is an AMRAY 1400 with a 50 kV acceleration

potential, tungsten source, electrostatic beam blanker, and a Robinson back-scattered electron detector. Such a large acceleration potential gives rise to a wide back-scattered electron distribution in the resist, thus reducing proximity effects during the exposure of dense layouts [25]. The Robinson back-scattered electron detector is used during the registration process to image marks through the resist. Also connected to the SEM is a Keithley 614 electrometer used to measure beam current collected in a Faraday cup mounted to the sample holder.

The beam position and blanking are controlled by a Pragmatic Instruments Model 2201A High-Definition Arbitrary Waveform Generator (AWG). Current advances in waveform generation electronics [26, 27] make the use of a high resolution AWG a superior choice for controlling SEM-based EBL systems. The 2201A is a three channel waveform generator with each channel capable of generating a waveform with a maximum length of 65,535 samples and sampled at a selectable rate from 0.100 Hz to 2.000 MHz with 4-digit resolution. Each waveform sample has 16-bit resolution within a selectable voltage range from 0.01 to 10.2 V with 4-digit resolution. The selectable channel amplitude is extremely useful for SEMs whose scan amplifiers compensate for the size of micrographs. By scaling the output of the 2201A accordingly, the SEM scan amplifier ratio can be balanced resulting in a square scan field. Each channel also has a synchronization output (TTL) which is used to monitor the end of waveform generation, to provide the pulse signals needed for registration, and to control the beam blanking. The three waveform outputs are connected to the SEM with channel 1 operating as a TTL signal to control beam blanking, and channel 2 and 3 functioning as analog outputs to control the horizontal and vertical beam positions. Channel signal outputs are referred to as BBK, XSCAN, and YSCAN respectively, and channel synchronization outputs are referred to as BBKSYNC, XSYNC, and YSYNC, respectively.

The Quantel Crystal Digital Image Processor is used to capture video information from the SEM. To capture images, the image processor requires video input as well as line and frame pulse signals. These pulse signals correspond to the

end of a horizontal scan and the end of a vertical scan respectively. To capture an image, the processor synchronizes its capture rate to that of the line and frame pulses. When the lithography system is used in an imaging mode, the image processor receives line and frame pulses from the SEM and the scanned image is digitized. When the system is used in a registration mode, the processor receives line and frame pulses from the 2201A (XSYNC and YSYNC) so that only the scanned areas over the registration marks are digitized.

Layout Exposure

We have taken great care in designing an EBL system capable of large scan field exposures with nanometer resolution. For us, the primary issue for large scan field exposures is the signal to noise (S/N) ratio in the signals used to control the SEM scan coils. Noise in the generated signal can result in a deflection of the beam which depends on the gain of the scan amplifiers. In order to complete patterns in large scan fields where the gain of the scan amplifiers is large, the noise must be less than the minimum feature size desired. For example, in a scan field of 0.7 mm, for a 30 nm feature to be exposed, the S/N ratio of the signal generator must be greater than 87 dB. We carefully analyzed the scan coil signals at various points in the lithography system using a Hewlett-Packard 35660A dynamic signal analyzer and were able to improve the S/N ratio measured at the SEM scan coils from 78 dB to 100 dB using proper shielding and grounding techniques [28, 29]. A 100 dB S/N ratio represents a peak-to-peak noise signal equivalent to only 14 nm in a 1.4 mm scan field.

Every stage of the system from the computer to the scan coils was modified. Substantial modifications were required to eliminate signals emanating from the filament current power supply. By isolating the supply from the scan coil electronics power supply, the S/N ratio improved by 10 dB. The primary source of noise was identified as a ground loop problem resulting from the General Purpose Interface Bus (GPIB) connection between the 2201A and the computer. Two commercial GPIB optical isolation units were tested with no improvement noted. Therefore, a ground

network was designed and implemented resulting in a 20 dB improvement in the S/N ratio. Fig. 2 shows a line pattern exposure before and after improvements were made. For large scan fields, the S/N ratio becomes much more of a factor as the noise in the scan coil signals directly correlates to an observable physical displacement of the beam as shown in Fig. 2(a). For small scan fields, the S/N ratio becomes much less of a limiting factor.

Layout and pattern exposure require the user to specify a beam current and magnification for the SEM, and exposure doses for the primitive shapes in the layout. The layout is exposed by programming the 2201A and CIO-DIO24 for each shape. The shape exposure order can be changed before pattern exposure. First, the 2201A is programmed with the necessary signals calculated to give the correct dose, for a specific magnification and beam current. Because of the flexibility of the 2201A, the shape algorithms are written so that the smallest time per pixel is calculated, allowing the highest resolution waveforms to be generated. Before a shape is exposed, a channel is selected to trigger the output of the other channels. The synchronization signal of this channel is then monitored by the CIO-DIO24 to detect the end of shape exposure. The AWG is then set to be triggered through the GPIB once the BBK circuit is programmed. The BBK circuit consists of fast logic gates controlled by the AWG synchronization outputs and the CIO-DIO24. The CIO-DIO24 selects which synchronization signals are to be used to control the SEM beam blanker. Also, the CIO-DIO24 is programmed to monitor the synchronization channel selected to signal the end of shape exposure. By continuously polling the CIO-DIO24 during shape exposure, the completion of an exposure is detected when a logic LO is read from the triggering channel synchronization output.

The commands used to program the 2201A to expose a horizontal rectangle are shown in Table 1. YSCAN is selected as the triggering channel for BBK and XSCAN signals. Fig. 3 shows an example of the waveforms programmed to expose a horizontal rectangle from the instructions in Table 1. The rectangle is exposed when YSCAN is triggered by the computer. As YSCAN is generated, when the second

waveform sample of YSCAN is reached, BBK and XSCAN are triggered. This sequence guarantees that BBK and XSCAN are synchronized to YSCAN. The end of the exposure is detected by polling YSYNC through the CIO-DIO24. When YSCAN waveform output begins, YSYNC changes from logic LO to logic HI. At the last waveform sample of YSCAN, YSYNC changes to a logic LO signaling the end of the shape exposure and the 2201A is programmed for the next shape exposure.

Use of the 2201A allows us to minimize the total overhead time for an exposure, as well as maximize control of exposure dose. Because SEM parameters such as focus and astigmatism can drift, large overhead times can result in poor exposure conditions for complex, lengthy patterns. We have optimized the software for all line types to take advantage of this property. When a waveform is calculated, the 2201A is instructed to load its memory with the waveform samples. Next, a sampling rate is selected for each memory location and the waveform is then triggered for output. For the next exposure, instead of loading a new waveform, the start and stop pointers in channel memory are moved to correspond to the next exposure and triggered. With the resolution and range of sampling rates, the exposure dose can be varied by 0.1% from one exposure to another. Such resolution is needed when attempting to characterize new resists and developer solutions as well as to optimize exposure parameters.

Registration

We have developed an accurate technique for achieving pattern registration, relying on the 2201A and the image processor. The 2201A is used to control the acquisition rate of the processor such that only the information scanned from the alignment marks is digitized. The synchronization signals of the 2201A are configured to emulate line and frame pulse signals with one frame consisting of 480 lines of video information. Shown in Fig. 4, our technique involves scanning one mark or scanning four marks simultaneously. The four-mark scan can be used to correct rotational errors while single mark scans are used to correct magnification errors as well as horizontal and vertical displacements. Once a mark has been

scanned, the video information is collected by the processor and displayed. Examples of single and multiple mark scans are provided in Fig. 5. The superimposed dark areas in the scanned images are achieved through beam blanking and serve as guides during alignment. To improve registration accuracy, the size of the scan windows can be scaled, resulting in a magnified view of the registration mark.

Results

We have tested our system in a variety of scan fields, exposing arbitrary patterns with many exposure doses. Studies investigating the properties of PMMA [30, 31] and charges in oxides [32] have been performed using this system. For this work, exposures were performed at 50 kV on thick silicon substrates with 100 nm thick PMMA of molecular weight 950,000 a.m.u. baked for 10 minutes at 170 °C using a hot plate. Development was performed at 21°C for 15 seconds in a 1:3 mixture of methyl isobutyl ketone : isopropanol with 1% methyl ethyl ketone [33]. Metal films of Au and Ti/Au were evaporated at normal incidence in an electron beam evaporator. Lift-off was performed by soaking in acetone for 2 to 3 minutes.

At conventional scan fields around 250 μm^2, dots approximately 13 nm in diameter were fabricated [34]. Some of our results demonstrating the flexibility of the system, as well as resolution, are shown in Fig. 6. Fig. 7 shows the same line pattern exposure at the center and top edge of a 1.0 mm scan field. Fig. 8 shows magnified views of the vertical grating and spoke pattern at the edge of the 1.0 mm scan field demonstrating high resolution features in large scan fields. In Fig. 7(b), we see that the innermost ring is displaced in the vertical direction, which is an example of the effects of scan coil settling. The inner most ring is drawn first in this pattern. Once the scan coils have settled, the remaining pattern is exposed without any noticeable errors. Such errors are corrected by choice of pattern order and addition of delays into the 2201A programming sequence. To test line width variation over a 1.0 mm scan field, the pattern in Fig. 9 was exposed. The line variation was estimated to be 5 nm for 65 nm wide lines. A more accurate investigation would

require micrographs with more detail which, at this time, are currently not available. To test the registration accuracy, the double exposure pattern in Fig. 10 was used with our best results shown in Fig. 11. We believe we have achieved an overlay accuracy of less than 100 nm.

Summary

We have shown that using an AWG to control an SEM for EBL can provide a versatile nanolithographic exposure tool capable of large scan field exposures. With the flexibility of the 2201A, we are able to precisely control the exposure dose, accurately position the electron beam and maintain beam position as well as control the image processor to facilitate device registration. The 16-bit resolution in voltage and 4-digit resolution in sampling rates allows for optimal control of exposure dose. The variable output of the waveform generator also simplifies the adjustment of scan field aspect ratios. We have shown that we are capable of exposing patterns in large scan fields with nanolithographic features. Having the flexibility to pattern the fine structures as well as other device features in one system reduces the overhead associated with device design and fabrication.

Acknowledgments

This work supported in part by IBM, AFOSR, ONR, NASA, and the University of Notre Dame. We thank Xiaokang Huang for the micrographs of exposures completed in small scan fields. We especially thank Meinrad Steiner at Pragmatic Instruments, Inc. for his GPIB programming tips.

References

1. T. K. Gaylord, E. N. Glytsis, G. N. Henderson, K. P. Martin, D. B. Walker, D. W. Wilson, K. F. Brennan, *Proc. of IEEE* **79**, 1159 (1991).
2. J. M. Ryan, J. Han, A. M. Kriman, and D. K. Ferry, *SPIE* 1284, 39 (1990).
3. G. Bernstein and D. K. Ferry, *IEEE Trans. Elect. Dev.* **35**, 887 (1988).
4. G. Bernstein and D. K. Ferry, *Superlattices and Microstructures* **2**, 373 (1986).
5. N. Giordano, *Phys. Rev. B* **22**, 5635 (1980).
6. E. H. Anderson, C. M. Horwitz, and H. I. Smith, *Appl. Phys. Lett.* **49**, 874 (1983)
7. D. C. Flanders, *Appl. Phys. Lett.* **36**, 93 (1980).
8. M. A. McCord and R. F. W. Pease, *J. Vac. Sci. Technol. B* **6**, 293 (1988).
9. J. C. Nabity and M. N. Wybourne, *Rev. Sci. Instrum.* **60**, 27 (1988).
10. S. Thoms, S. P. Beaumont, and C. D. W. Wilkinson, *J. Vac. Sci. Technol. B* **7**, 1823 (1989).
11. C. C. Andrews and W. P. Kirk, *J. Vac. Sci. Technol. B* **8**, 1858 (1990).
12. A. J. Speth, A. D. Wilson, A. Lern, and T. H. P. Chang, *J. Vac. Sci. Technol.* **12**, 1235 (1975).
13. J. Bellessa, F. Carcenac, A. Izrael, H. Launois, and D. Mailly, *Micro. Elec. Engr.* **6**, 175 (1987).
14. S. Mackie and S. P. Beaumont, *Solid State Technol.* **28**, 117 (1985).
15. B. L. McIntyre and C. L. Dennis, *Applied Optics* **27**, 196 (1988).
16. B. McIntyre, R. Perman, and B. Gold, *Res. and Develop.* **31**, 44 (1989).
17. C. F. Cook, T. F. AuCoin, and G. J. Iafrate, *Solid State Technol.* **28**, 125 (1985).
18. C. P. Umbach, A. N. Broers, R. H. Koch, C. G. Willson, R. B. Laibowitz, *IBM J. Res. Develop.* **32**, 454 (1988).
19. A. Muray, M. Issacson, I. Adesida, and B. Whitehead, *J. Vac. Sci. Technol. B* **1**, 1091 (1983).
20. A. N. Broers, J. M. E. Harper, and W. W. Molzen, *Appl. Phys. Lett.* **35**, 392 (1978).
21. C. S. Lent, P. D. Tougaw, W. Porod, and G. H. Bernstein, to appear in *Nanotechnology.*

22. B. McCombe, private communication.

23. Turbo Pascal version 5.5, Borland International.

24. A high speed computer is not necessary because the timing is performed by the high speed electronics of the AWG.

25. X. Huang, G. Bazán, G. H. Bernstein, and D. A. Hill, submitted for publication.

26. A. Pini, *Eval. Engr.* **30**, 26, July (1991).

27. J. Sauer, *Eval. Engr.* **31**. 26, March (1992).

28. R. Morrison and W. H. Lewis, *Grounding and Shielding in Facilities*, John Wiley & Sons, Inc., New York, (1990).

29. R. Morrison, *Noise and Other Interfering Signals*, John Wiley & Sons, Inc., New York, (1992).

30. D. A. Hill, X. Huang, G. Bazán, and G. H. Bernstein, *J. Appl. Phys.* **72**, 4088 (1992).

31. X. Huang, G. Bazán, D. A. Hill, and G. H. Bernstein, *J. Electrochem. Soc.* **139**, 2952 (1992).

32. G. H. Bernstein, S. W. Polchlopek, R. Kamath, and W. Porod, ~~submitted for publication.~~ Scanning, in press

33. G. H. Bernstein, D. A. Hill, and W. P. Liu, *J. Appl. Phys.* **71**, 4066 (1992).

34. As imaged with a field emission SEM courtesy of Topcon Technologies, Inc.

Tables

TABLE 1. Commands used to configure 2201A for horizontal rectangle exposure.

Figures

FIGURE 1. Schematic diagram of electron beam lithography system.

FIGURE 2. Improvements made in noise reduction for an exposure completed in a 1.4 mm scan field. (a) Pattern processed before noise reduction. (b) Pattern after improvements were made to the system.

FIGURE 3. Example of the 2201A waveform output when exposing a horizontal rectangle. Shown here is an example of three passes across the rectangle. YSCAN triggers BBK and XSCAN. Notice the beam is off during the retrace of XSCAN. With this, each sample in YSCAN corresponds to one XSCAN across the rectangle. The time scale for BBK and XSCAN is kept constant at 0.5 µsec and the time scale of YSCAN is equal to the total time for one cycle of XSCAN (xretrace + xsamples). To ensure channel synchronization, YSCAN waveform output begins on its SECOND waveform sample.

FIGURE 4. Registration mark layout with definitions. Marks are numbered in a clockwise pattern. A scan of all marks is completed by scanning marks 1, 2, 3, and 4 in order. Scanning of individual marks is selected from a menu within the program. The size of the marks is enlarged for clarity.

FIGURE 5. Processed back-scattered electron images of alignment marks for (a) four-mark and (b) single mark scans.

FIGURE 6. Miscellaneous patterns demonstrating flexibility and resolution of the lithography system. (a) Au dots on 40 X 50 nm grid and (b) Au grating with lines on 55 nm centers exposed in the center of a 240 µm scan field. (c) Au ring with inner diameter less than 60 nm with a width of 18 nm. (d) Au line structure with line widths less than 100 nm demostrating arbitrary pattern exposure capability.

FIGURE 7. Micrographs of line patterns exposed (a) in the center and (b) at the top edge of a 1 mm scan field.

FIGURE 8. Magnified view of lines in Fig. 7(b). (a) Vertical grating and (b) spoke

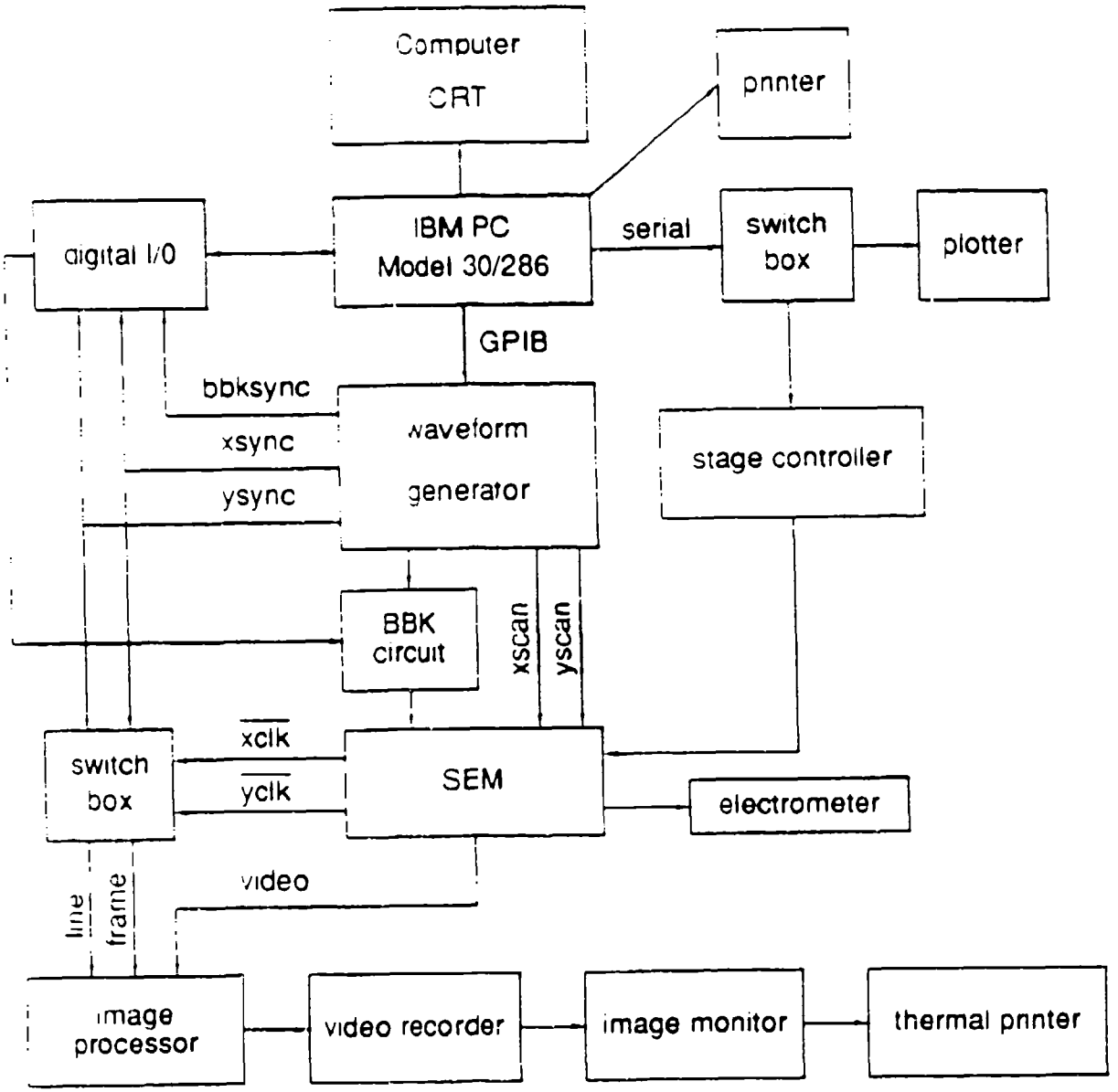
pattern demonstrating features under 100 nm in a 1.0 mm scan field.

FIGURE 9. Exposure pattern used to investigate the line width variation across a 1.0 mm scan field. The pattern was exposed both horizontally and vertically. Over 40 micrographs were taken of fabricated lines. Micrographs (a) and (b) correspond to location of boxes in figure drawn.

FIGURE 10. Exposure pattern used to determine registration accuracy. The dark lines of patterns A and B were fabricated first. Having aligned the sample, the mirror image of patterns A and B was then fabricated and the registration accuracy determined.

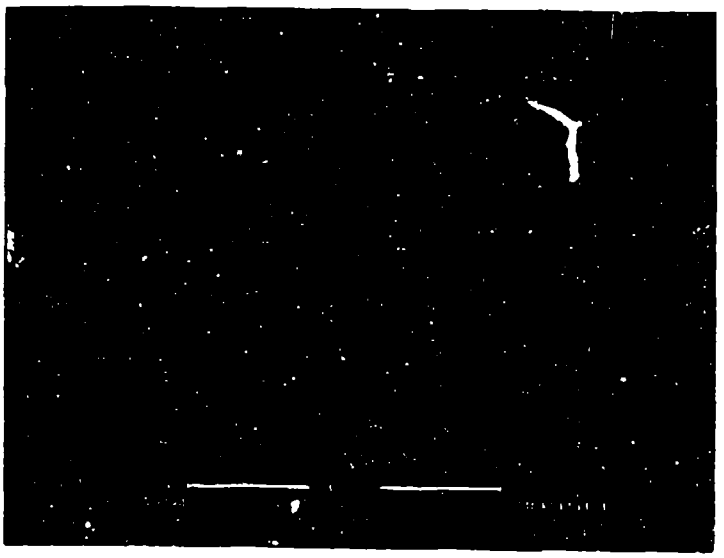
FIGURE 11. Micrographs of alignment results for Pattern A (a) and Pattern B (b) of Fig. 10 located in the center of the scan field.

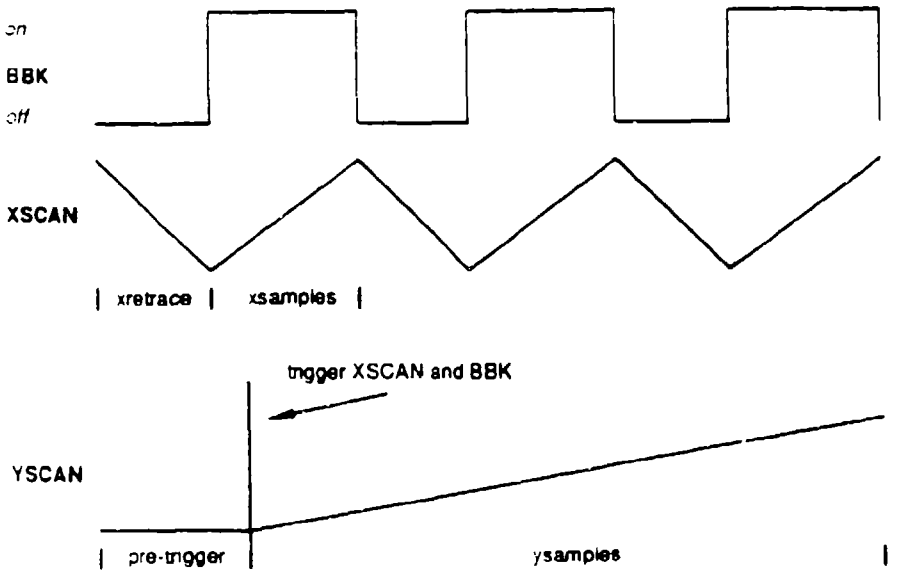
Commands	BBK	XSCAN	YSCAN
amplitude	2.5	10.0	10.0
offset	2.5	0	0
mode	burst	burst	trig
start address	0	0	0
stop address	xsamples + xretrace-1	xsamples + xretrace -1	ysamples+1
burst	ysamples	ysamples	-
clock sel	twomeg	twomeg	synth
filter	-	on	on
deglitch	-	off	off
sync start	-	-	1
sync length	-		ysamples
output switch	on	off/on	off/on
trigger sel	chan3	chan3	man
trig1 start	-	-	1
trig1 length	-	-	1
trig2 start	-	-	1
trig2 length	-	-	1

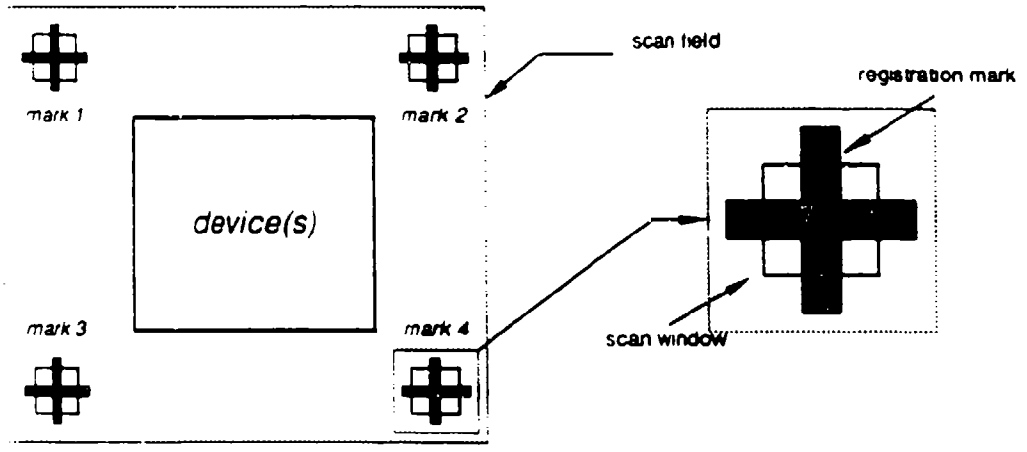

Computer
CRT
pnnter
digital I/0
IBM PC
Model 30/286
serial
switch
box
plotter
GPIB
bbksync
xsync
ysync
waveform
generator
stage controller
BBK
circuit
xscan
yscan
switch
box
xclk
yclk
SEM
electrometer
line
frame
video
image
processor
video recorder
image monitor
thermal pnnter

(a)

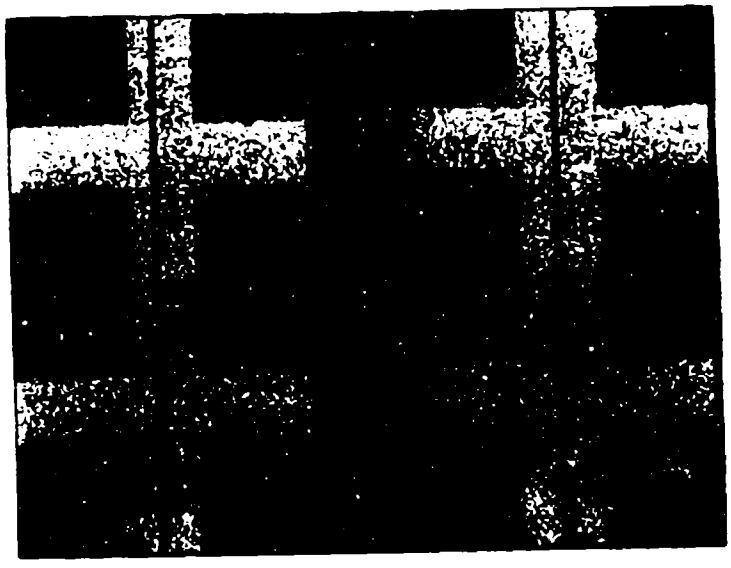
(b)



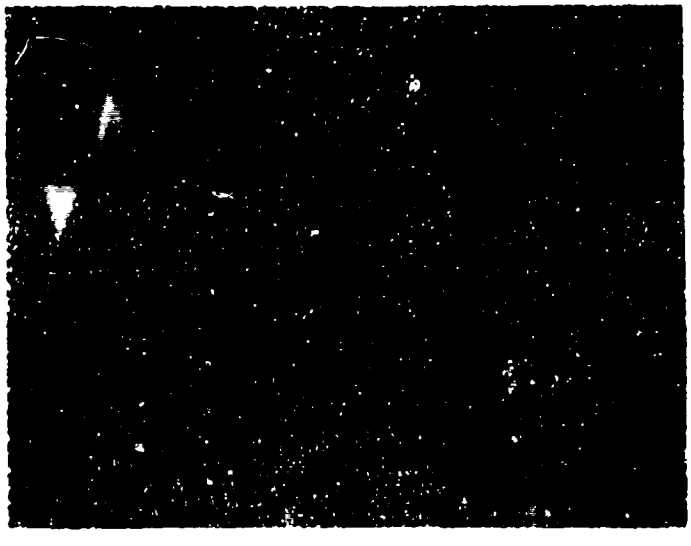

on
BBK
off
XSCAN
xretrace
xsamples
trigger XSCAN and BBK
YSCAN
pre-trigger
ysamples

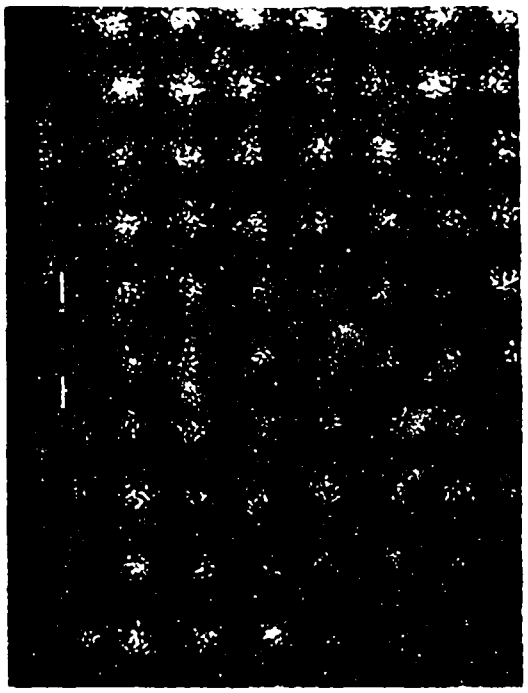

scan field
registration mark
mark 1
mark 2
device(s)
mark 3
mark 4
scan window

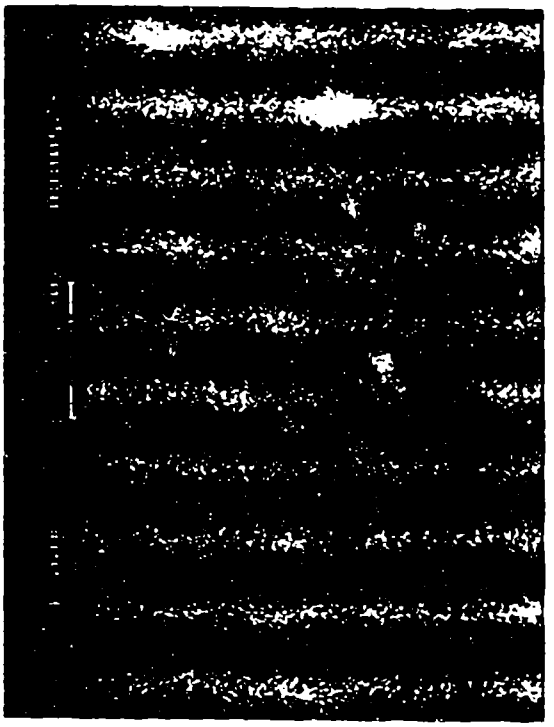
a)

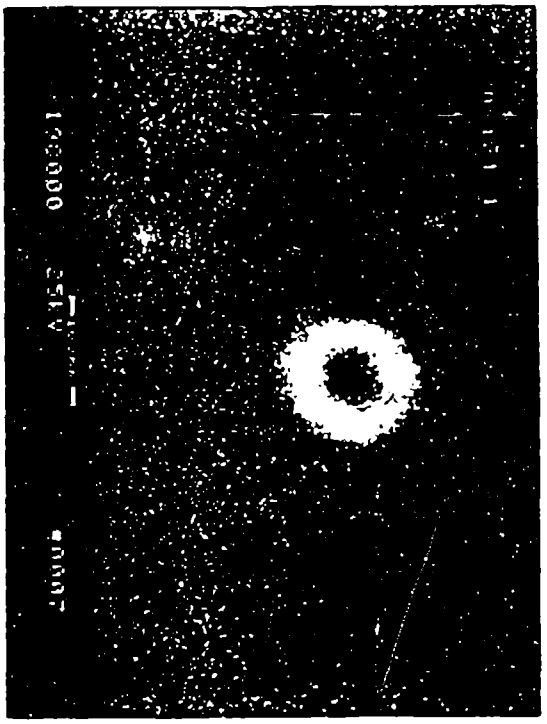


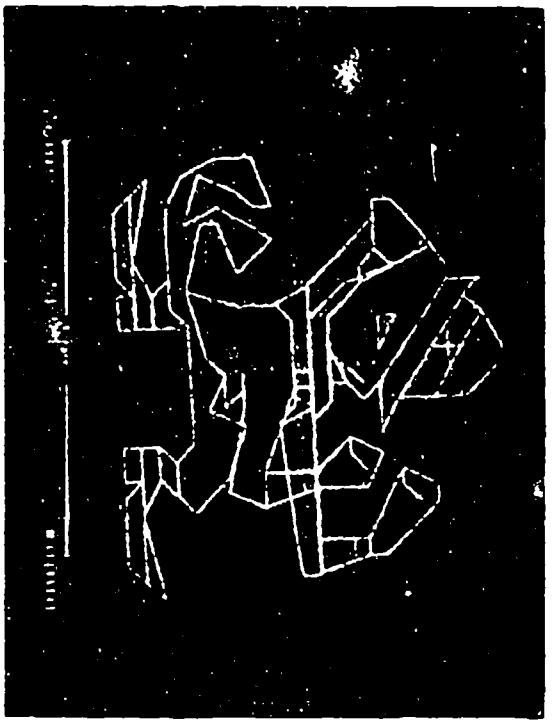
(b)



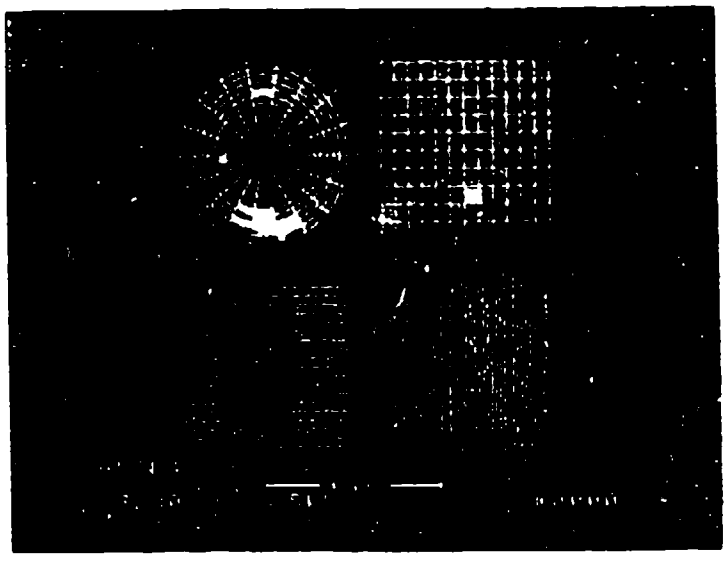




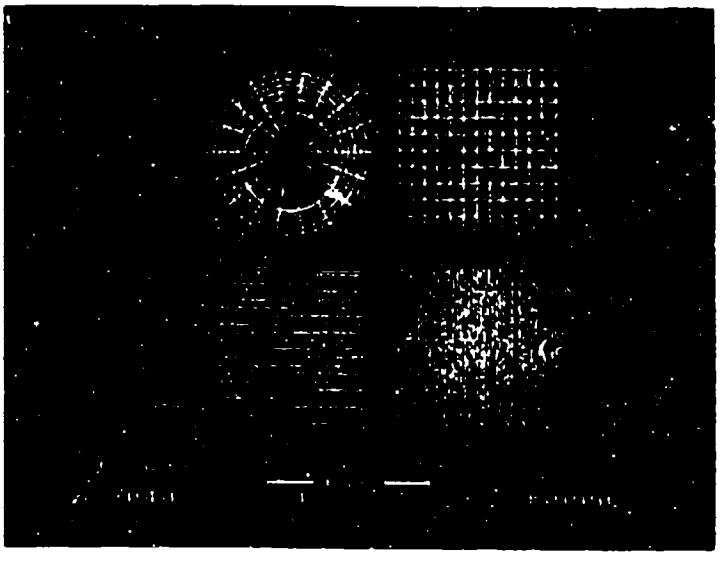




(a)

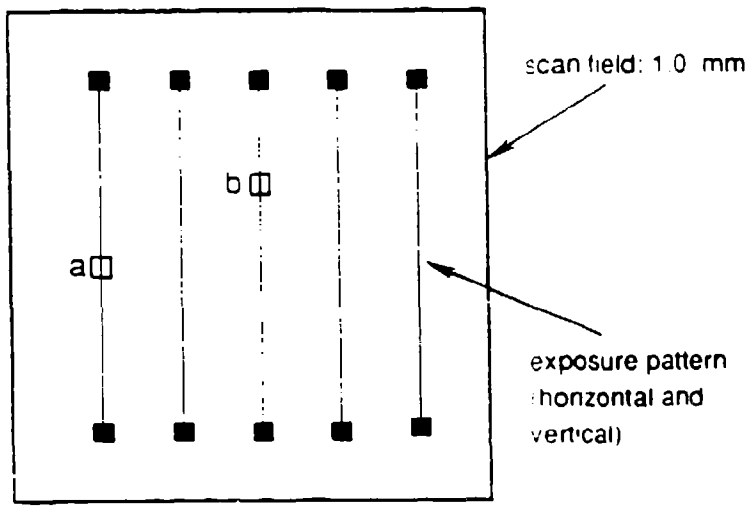


(b)

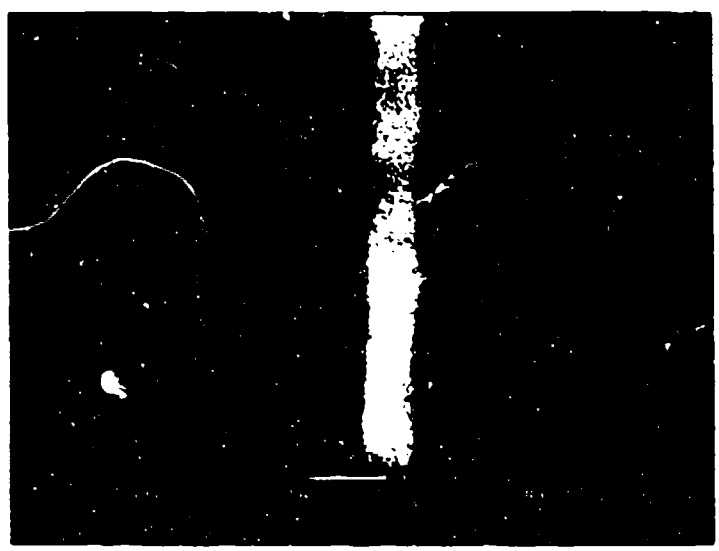


a)

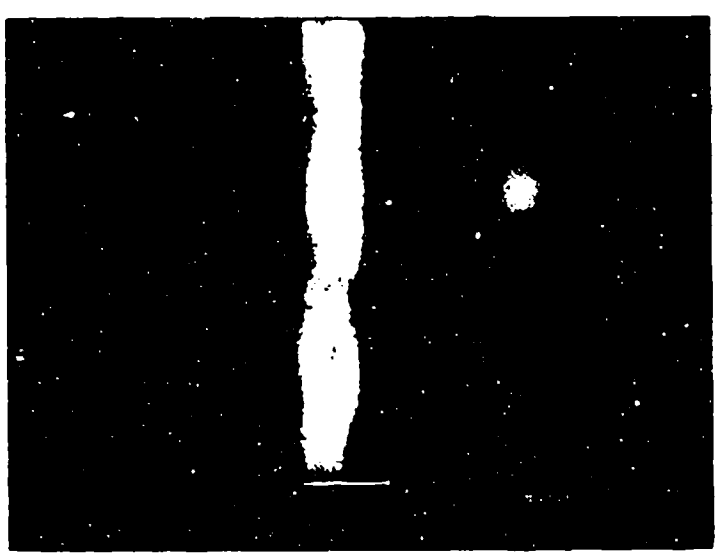
(b)

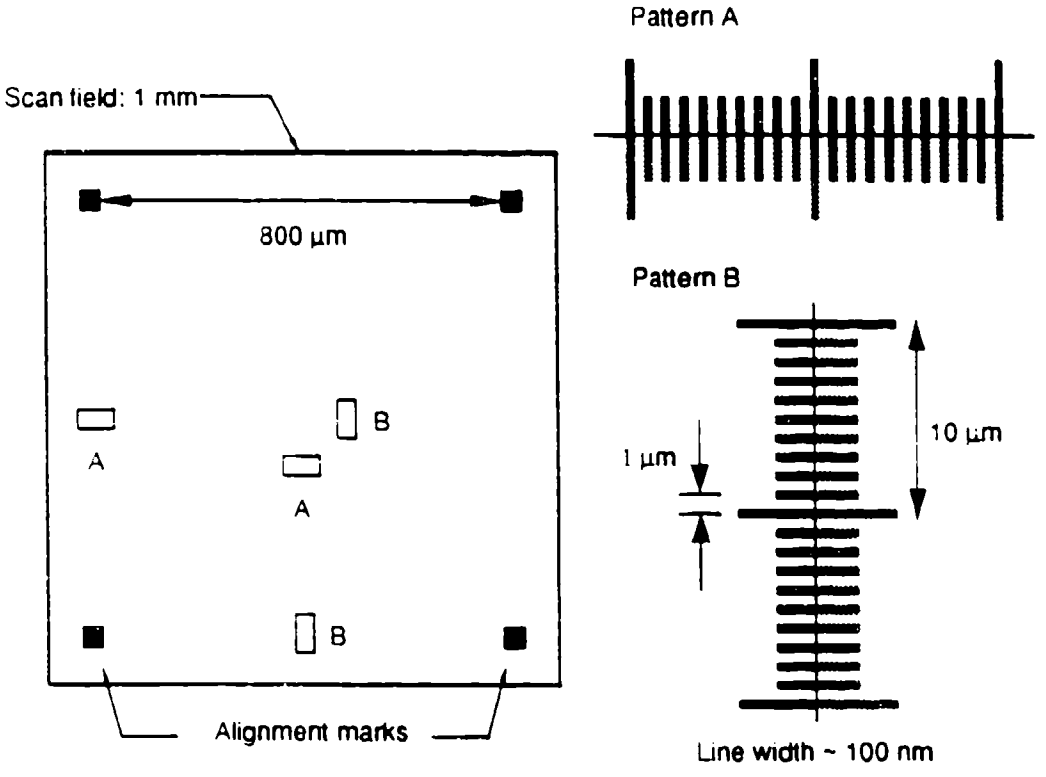

scan field: 1.0 mm
b
a
exposure pattern
(horizontal and
vertical)

(a)

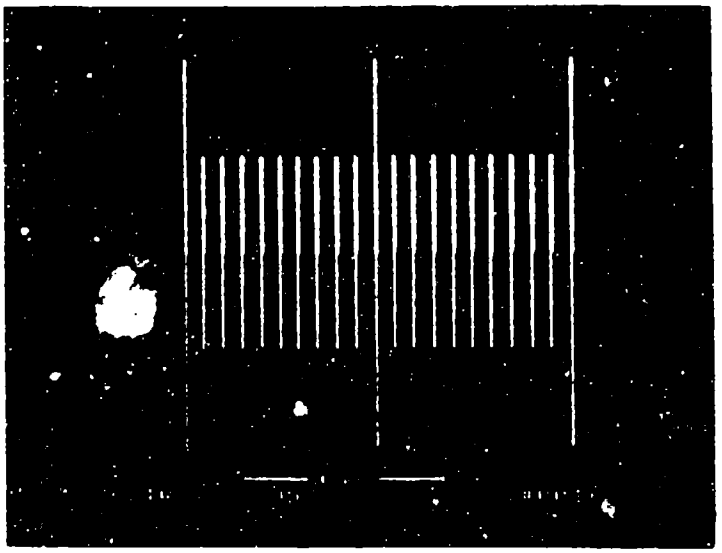


(b)

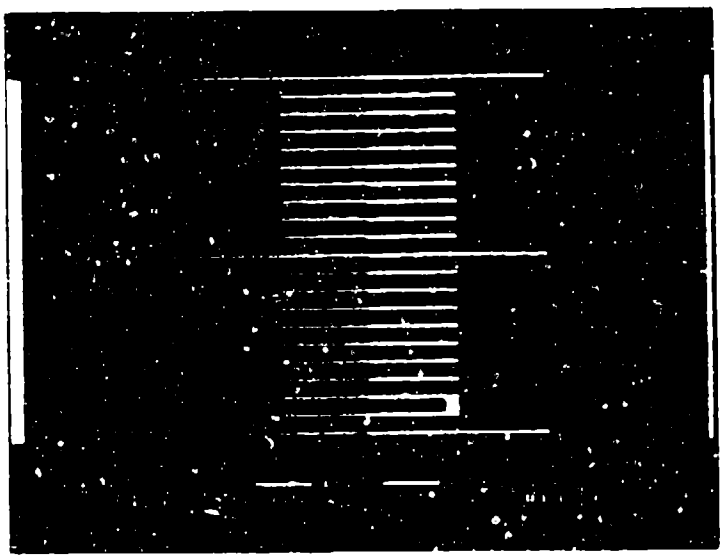



Scan field: 1 mm
800 µm
A
B
A
B
Alignment marks
Pattern A
Pattern B
1 µm
10 µm
Line width ~ 100 nm

(a)



(b)



Swelling and surface forces-induced instabilities in nanoscopic polymeric structures

Davide A. Hill[a)]
Department of Chemical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Xiaokang Huang, Greg Bazán, and Gary H. Bernstein
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 1 May 1992; accepted for publication 12 July 1992)

For closely spaced, nanolithographically defined lines, a thin wall of resist remains to act as the metal line spacer. When exposed to a developer, and then rinsing solution, closely spaced resist walls may become unstable as a result of two effects: (1) internal stresses due to swelling, and (2) lateral surface forces between adjacent walls. In this article we perform a linear stability analysis of a thin polymer wall under the simultaneous action of internal stresses and lateral surface forces. We calculate a stability boundary, and show that internal stresses are necessary for the formation of deformation patterns of finite wavelength. We find that, for slightly subcritical swelling stresses a small lateral force can induce buckling, while, for slightly subcritical surface tractions large internal stresses are necessary to induce instability. The theoretical predictions are in good agreement with experimental data on poly(methyl-methacrylate) walls produced by electron beam lithography.

INTRODUCTION

In addition to ultrahigh resolution, an important aspect of nanofabrication for nanoelectronics is the achievable spacing of metal features. Extremely closely spaced features have been utilized in such devices as lateral tunneling transistors,[1] split gates,[2] and optical gratings.[3] In applications such as these, metal features are required with spaces as small as a few hundred Å. The fabrication of very narrow gaps over appreciable distances (a few μm) requires very good control of lithographic processing parameters.

Most pattern transfer in nanolithography for quantum devices is effected by the lift-off process in which a thin metal layer is deposited by perpendicular, thermal evaporation onto a resist [usually poly (methyl methacrylate) (PMMA)] such that metal falls either on the surface of the undeveloped (positive) resist or onto the surface of the substrate where exposed and developed resist has been removed. Subsequent removal of unexposed resist also removes all metal except that on the surface of the substrate. When metal patterns are to be placed very closely together, the space is formed by the shadowing of the deposited metal by a thin "wall" of resist. We report here on the fundamental limitations of solvent/PMMA interactions which lead to failure of the thin features necessary to perform lift-off of very closely spaced features.

For thin resists at moderate energies, undercutting by beam scattering within the resist can approximately be described by:[4]

$$b = 625(Z/E_0)t^{3/2}(\rho/A)^{1/2}, \tag{1}$$

where b is the increase of the diameter of the beam at the bottom surface of the resist, Z, A, ρ, and t are the atomic number, atomic weight, density, and thickness of the resist, respectively, and E_0 is the beam energy in keV. For PMMA, $Z=3.6$, $A=6.7$ g/mol, and $\rho=1.2$ g/cm^3.[5] For 45 keV electrons and a PMMA thickness of 130 nm, undercutting is about 5 nm at each sidewall of our p[illegible], which aids in the lift-off process.

As discussed above, the unexposed resist spacer is in the form of a thin wall between the developed lines. When lines are placed too closely together and high resolution is achieved, the resist walls can either become wavy, fall over, or even lose adhesion to the surface, in which case they fail completely. In this regard, internal stresses due to swelling and lateral surface forces between adjacent structures can play dominant roles. Although PMMA suffers little swelling, which accounts partially for its ultrahigh resolution properties, some absorption of developer is necessary to aid the development process.[6] Swelling and solubilization of polymeric resists have been extensively investigated.[7,8] In marginally poor solvents, solvation forces favoring mixing are insufficient to promote dissolution of the polymer. Appreciable swelling can occur, however, leading to weaker mechanical properties, and to the buildup of internal stresses in constrained structures. Amorphous polymers of high molecular weight may typically experience between 0.1% and 30% swelling, without solubilization, depending on solvent quality and temperature.[8] With mixtures of solvents, the swelling behavior can be of considerable complexity, since preferential partitioning of one or more components of the mixture in the polymer can occur. This would be the case for most resist developer solutions, which usually comprise more than a single component. Notwithstanding their valuable insight, thermodynamic models have enjoyed only limited success in the quantitative prediction of solubility properties of multicomponent polymer-solvent systems, and such an attempt will not be made here. Nevertheless, as a discriminator of the qual[illegible]

a)Author to whom correspondence should be addressed.

 0021-8979/92/214088-07$04.00

of the model, our predictions will be required to compare favorably with typical swelling data.

Recent experiments have considerably clarified the mechanism of interaction between polymeric surfaces. Studies of this nature have recently become possible thanks to the perfecting of the surface force apparatus by Israelachvili and Tabor.[9] In this system, the separation between two mutually orthogonal, cylindrical mica surfaces onto which polymer is adsorbed can be measured with resolution better than 5 Å. The separation measurements rely on interferometry, while knowledge of the flexural rigidity of the instrument enables calculation of the force between the surfaces. Normalization of the force by 2π times the mean radius of curvature of the mica sheets directly yields the surface energy of interaction.[10] Experiments show that in good solvents, polymer surfaces always repel each other due to the tendency of the chains to mix most favorably with solvent molecules. Conversely, in poor solvents chain-chain interactions are more favorable, leading to strong long-range attraction. Experimentally, the attractive forces between polymer layers become significant within a distance of about twice the radius of gyration of the chains, which can extend to several hundred Å for high molecular weight polymers.[10] Common post-development rinsing solutions are, by design, of poor thermodynamic quality. Therefore, in addition to swelling, attractive forces are expected to play a vital role in the failure of polymeric nanostructures.

EXPERIMENT

Our electron beam lithography (EBL) system consists of an Amray 1400 scanning electron microscope (SEM) with a minimum spot size of 35 Å controlled by a personal computer interfaced through a Pragmatic Systems 2201A 16-bit arbitrary wave form generator. We used PMMA with molecular weight 950 000 amu baked at 170 °C for 4 h for all exposures. Resist and developer parameters were very tightly controlled. Resist thickness was 105–180 nm as measured with an Alpha-Step 200 surface profiler. The developer used in all cases was a solution of MIBK:IPA (1:3) with the addition of 1% (MEK) by volume.[11] It has been shown that contrast results are extremely sensitive to variations in developer temperature due to the high activation energy of the PMMA in the developer solvents.[11] Using a temperature controller, our developer temperature was maintained at 25 ± 0.5 °C. Develop time for all samples was 15 s followed by a 15 s rinse in methanol. After development, all samples were dried in air at 90 °C for 5 min.

All exposures were performed at a beam energy of 45 keV at a spot size of 5 nm and 120 μm scan field. Substrates used in all cases were 200 nm SiO_2 on thick silicon wafers. Doses were very critical for making high-density patterns because of proximity effects during e-beam lithography. They were not only related to the line pitch, but also to the size of the patterns. For grating pitches ranging from 75 to 150 nm, the doses varied from 1.5 to 2.5×10^{-9} C/cm. For the case of only two parallel lines, doses ranged from 2.5 to 3.5×10^{-9} C/cm for pitches from 80 to 120 nm.

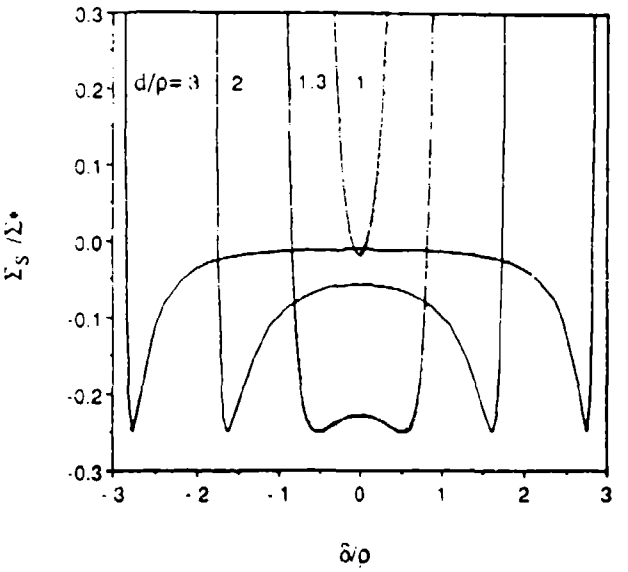


FIG. 1. Schematic representation of the potential energy of interaction of a polymer surface confined between and equidistant from two other surfaces. As the dimensionless separation d/ρ increases, the curves broaden and the equilibrium position around $\delta/\rho=0$ becomes indifferent to small lateral displacements. At very small separations the potential becomes purely repulsive. At intermediate separations the erect position may either be stable or unstable depending on the rigidity of the wall.

All metal evaporations were gold from an electron beam evaporator with thicknesses of 30 ± 1 nm. Lift-off was performed by soaking the samples in acetone for about 5 min, followed by the application of acetone using a hypodermic syringe, followed by a 15 s methanol rinse, and final blow drying in nitrogen.

THEORY

The interaction energy between two opposing surfaces, Σ, can be obtained by linear superposition of long-range, attractive, and short-range, repulsive interactions. Generally, the potentials are individually modeled as a power-law of the reciprocal of the separation distance, r, giving (within an arbitrary additive constant):[10,12]

$$\Sigma=\Sigma^*[(\rho/r)^a-(\rho/r)^b], \tag{2}$$

where Σ^* and ρ characterize the strength and width of the potential well, respectively, and a and b $(a>b)$ prescribe the rate of growth and decay of repulsive and attractive interactions, respectively. For a polymer wall confined between, and equidistant from two parallel surfaces with separation d, the following approximation is appropriate:

$$\Sigma_s=\Sigma^*\{[\,|\rho^2/(d^2-\delta^2)|\,]^{a/2}-[\,|\rho^2/(d^2-\delta^2)|\,]^{b/2}\}$$

$$(\text{for } -d<\delta<d), \tag{3}$$

where Σ_s is the surface energy (energy/area), and δ any virtual lateral displacement away from the planar configuration. Note that δ, and hence Σ_s, will in general vary from point to point along the surface of the wall. Figure 1 schematically illustrates the shape of Σ_s for different values of the dimensionless separation, d/ρ. It is seen that for

large values of d/ρ the erect position is indifferent, while it becomes progressively more unstable as the surface-to-surface separation decreases. (Note that we are ignoring for the moment the elasticity of the wall, which, by counteracting deformation, would stabilize the system.) Interestingly, as the separation further decreases repulsion becomes dominant and the equilibrium reverts to unconditionally stable. Experimentally,[10] the transition should occur around $d \approx R/2$, where R is the radius of gyration of the polymer chains. Apparently, this situation could be exploited to greatly improve resolution in nanolithography. However, it may be difficult in practice to take advantage of the phenomenon, since typically, $R/2$ is on the order of 5–10 nm. Additionally, interpenetration of the chains will, in fact, impede complete separation of the walls after solvent evaporation.

Differentiation of the interfacial energy [Eq. (3)] with respect to δ provides the interfacial stress acting on the lateral surface of the wall, σ:

$$\sigma = -\frac{d\Sigma_s}{d\delta}. \tag{4}$$

In the limit of small displacements, the detailed structure of the potential is irrelevant, and σ can be linearized around the origin ($\delta = 0$) to give

$$\sigma = \Theta\delta, \tag{5}$$

where

$$\Theta \equiv -\left(\frac{d^2\Sigma_s}{d\delta^2}\right)_{\delta=0}. \tag{6}$$

Henceforth, Θ will be referred to as the "surface stress coefficient." Equation (6) can be used to calculate Θ from surface force measurements, for any given polymer-solvent system. Note that both the magnitude and sign of Θ will generally depend on the dimensionless separation d/ρ. Therefore, in the following we shall attempt to estimate only the order of magnitude of Θ.

Dimensional analysis of Eq. (6) and experiments[10] suggest the following scaling:

$$\Theta \approx \Delta\Sigma_r/(R)^2, \tag{7}$$

where $\Delta\Sigma_r$ is the depth of the attractive potential well, and R the end-to-end distance of the polymer chains. As mentioned previously, surface-force-apparatus experiments directly yield Σ_r as the measured force normalized by 2π times the mean radius of curvature of the mica sheets.[10] Unfortunately, experimental data of this nature are not available for PMMA. Polystyrene "brushes" in various solvents have been extensively investigated, however. Because of the osmotic nature of the interactions, the detailed chemical structure of the polymer is believed to be only of secondary importance.[10] Therefore, data on polystyrene should prove useful for obtaining estimates of the general magnitude of $\Delta\Sigma_r$. Figure 1 of Ref. 10 gives plots of normalized force versus separation for polystyrene brushes in an aliphatic solvent below the θ temperature. The data show a pronounced attractive well with a depth of about

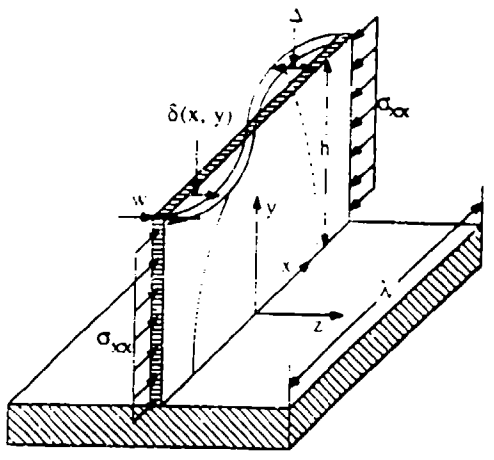


FIG. 2. Diagram of tall, thin wall demonstrating geometrical parameters used in the analysis. The wall has been hypothetically sectioned at the nodal points of the distortion.

$\Delta\Sigma_r \approx 160$ (μN/m) (after proper normalization by 2π). Roughly, we may assume $R \approx 10$–30 (nm) in Eq. (7), thus placing Θ in the range: $10^{11} < \Theta < 10^{12}$ (N/m^3). We show later that these values are sufficient to cause instabilities in arrays of closely spaced, thin polymer walls. In the following, we shall focus on the case $\Theta > 0$ (signifying an unstable equilibrium), and make use of Eq. (5) to describe the lateral surface traction on the wall.

We consider now the stability of an infinitely long wall of width w and height h under the simultaneous action of lateral forces and internal stresses due to swelling. The system of interest is shown in Fig. 2, which also gives the characteristic geometrical parameters. Mathematically, this problem is conveniently formulated within the framework of linear plate theory, leading to the following homogeneous boundary value problem for the lateral displacement, δ:[13]

$$\frac{\partial^4\delta}{\partial x^4} + 2\frac{\partial^4\delta}{\partial x^2\partial y^2} + \frac{\partial^4\delta}{\partial y^4} = -\Lambda\frac{\partial^2\delta}{\partial x^2} + \Omega\delta, \tag{8}$$

$$\delta = 0 \quad \text{at} \quad y = 0 \quad \text{(no displacement at the bottom)}, \tag{9}$$

$$\frac{\partial\delta}{\partial y} = 0 \quad \text{at} \quad y = 0 \quad \text{(built-in bottom end)}, \tag{10}$$

$$\frac{\partial^2\delta}{\partial y^2} + \nu\frac{\partial^2\delta}{\partial x^2} = 0 \quad \text{at} \quad y = 1 \quad \text{(torque-free top)}, \tag{11}$$

$$\frac{\partial^3\delta}{\partial y^3} + (2-\nu)\frac{\partial^3\delta}{\partial x^2\partial y} = 0 \quad \text{at} \quad y = 1$$

$$\text{(shear force-free top)} \tag{12}$$

where the dimensionless variables and parameters are defined as

$$x = X/h; \quad y = Y/h, \tag{13}$$

$$\Lambda = \frac{\sigma_{xx}12(1-\nu^2)h^2}{Ew^2}. \tag{14}$$

$$\Omega = \frac{\Theta 12(1-\nu^2)h^4}{Ew^3}. \tag{15}$$

In Eqs. (14) and (15), $\sigma_{xx} > 0$ is the internal compressive stress, Θ is defined in Eq. (5), ν is Poisson's modulus, E Young's modulus, and h and w the wall height and width, respectively.

For an infinitely long wall, the solution to Eq. (8) is of the form:[13]

$$\delta(x,y) = \cos(\xi x)\{A[\sinh(\alpha y) - \sin(\beta y)] + B[\cosh(\alpha y) - \cos(\beta y)]\}, \tag{16}$$

where A and B are integration constants, and

$$\alpha = \sqrt{\xi^2 + \sqrt{\Lambda\xi^2 + \Omega}}$$
$$\beta = \sqrt{-\xi^2 + \sqrt{\Lambda\xi^2 + \Omega}}. \tag{17}$$

Variable ξ is related to the wavelength of the prevailing distortion in the x direction, λ, as

$$\lambda = 2\pi h/\xi. \tag{18}$$

Imposition of the boundary conditions generates the following equation for the eigenvalues:

$$2ts + (s^2 + t^2)\cosh(\alpha)\cos(\beta) = \frac{1}{\alpha\beta}(\alpha^2 t^2 - \beta^2 s^2) \times \sinh(\alpha)\sin(\beta), \tag{19}$$

where

$$t = \beta^2 + \nu\xi^2$$
$$s = \alpha^2 - \nu\xi^2. \tag{20}$$

Note that in order for β in Eq. (17) to be real, we must have

$$0 \le \epsilon \le 1$$

where

$$\epsilon \equiv 2\xi^2/(\Lambda + \sqrt{\Lambda^2 - 4\Omega}). \tag{21}$$

The use of ϵ, instead of ξ, provides a more efficient strategy for a systematic search of the eigenvalues.

Equation (19) implicitly defines a relation among Λ, Ω, and ϵ. In the search for the roots, the following limiting cases are important: (1) $\Lambda = 0$ and (2) $\Omega = 0$. These cases will be treated separately, since they provide clues for the solution of the general problem in which lateral forces and internal stresses act simultaneously.

Case (1): $\Omega = 0$

As is customary in stability theory, Eq. (19) (with $\Omega = 0$) is satisfied only by a discrete set of ϵ (or ξ) values, provided Λ is greater than or equal to a minimum critical value, Λ^*. Physically, Λ^* is the smallest (dimensionless) stress necessary to reach the marginal stability limit (indifferent equilibrium), and can be calculated by solving Eq. (19) numerically. We find: $\Lambda^* = 10.40$, and $\xi^* = 1.917$. Combining these results with Eqs. (14) and (18), the smallest critical stress, σ^*_{cr}, and the associated wavelength, λ, are calculated to be

$$\sigma^*_{cx} = 26Ew^2/[30(1-\nu^2)h^2], \tag{22}$$

$$\lambda = h2\pi/1.917. \tag{23}$$

For swelling induced stresses, σ^*_{cx} can in turn be obtained from

$$\sigma^*_{cx} = E\varepsilon_{sw}, \tag{24}$$

where ε_{sw} is the hypothetical swelling strain that the wall would undergo if allowed to expand freely. Equations (22) and (24) enable calculation of the critical wall aspect ratio for stability, $(h/w)^*$, as a function of the swelling strain:

$$\left(\frac{h}{w}\right)^* = \sqrt{26/30(1-\nu^2)\varepsilon_{sw}}. \tag{25}$$

Equation (25) is convenient for comparison with experiments, since it does not contain the unknown modulus E (which presumably changes upon swelling). As expected, a short wall can sustain a higher degree of swelling before buckling. Additionally, for a constant ε_{sw}, Eq. (25) predicts a direct proportionality between critical wall height and width. As shown later on, our experimental findings confirm this prediction.

Case (2): $\Lambda = 0$

Treatment of this case is similar to the previous one except that, here the smallest critical lateral "force," Ω^*, corresponds to $\xi^* = 0$. This implies an infinite wavelength, λ [Eq. (18)], which means that the wall does not buckle, but simply bends over sideways. Solving Eq. (19) numerically, we find: $\Omega^* \approx 12.363$, and $\xi^* = 0$. Making use of Eq. (15), the critical surface-stress coefficient, Θ^*, is obtained as

$$\Theta^* \cong \frac{1.03Ew^3}{(1-\nu^2)h^4}. \tag{26}$$

If estimates of E and ν were available, Eq. (26) could be used to determine Θ^* experimentally.

The search for the eigenvalues in the combined case (i.e., $\Lambda \neq 0$ and $\Omega \neq 0$) is now restricted to the domain $(0 < \Lambda < \Lambda^*, 0 < \Omega < \Omega^*, 0 < \epsilon < 1)$. Delineation of the stability boundary implies the search of the minimum values $\bar{\Lambda}$ and $\bar{\Omega}$ that simultaneously satisfy Eq. (19). A convenient numerical strategy is to discretize the interval $0 < \bar{\Lambda} < \Lambda^*$, and then find the corresponding minimum values of $\bar{\Omega}$ (or vice versa). The results of these calculations are given in Table I, while Fig. 3 shows the stability boundary. The wavelength of the distortion, λ, as a function of Ω is shown in Fig. 4. Note that λ is relatively insensitive to lateral forces up to about 80% of Ω^*. It is interesting to note in Fig. 3 that, for $\Lambda < 5$, a slightly subcritical lateral stress, say $\Omega = 0.95\ \Omega^*$, does not cause instability. On the other hand, for a slightly subcritical value of Λ, a small lateral force can be catastrophic.

RESULTS AND DISCUSSION

Plotted in Fig. 5 are the experimental critical heights for buckling versus width, for PMMA walls fabricated by electron beam lithography. A line through the origin satisfactorily fits the data, indicating a direct proportionality

TABLE I. Minimum values of Λ and $\Omega=\Omega(\Lambda)$ for stability, and the associated normalized wavelength, λ/h (the calculations assume $\nu=0.50$).

Λ	Ω	λ/h
10.40	0.000	3.277
10.19	1.000	3.396
9.990	2.000	3.582
9.737	3.000	3.623
9.474	4.000	3.653
8.932	6.000	4.002
8.272	8.000	4.319
7.450	10.00	5.097
6.700	11.30	6.398
6.000	12.08	9.721
5.000	12.362	59.2
0.000	12.363	∞

between critical wall height and width. This finding is in agreement with Eq. (25), assuming a constant degree of swelling. The critical aspect ratio for buckling can be calculated from the slope of the line, and we find that $(h/w)^*=4.75$. Calculation of ε_{sw} from Eq. (25) requires knowledge of Poisson's modulus, ν. Because the polymer is swollen, and the internal stresses are small, the assumption $\nu=0.50$ is appropriate, giving $\varepsilon_{sw}=5.1\%$. This value is well within the expected range for glassy polymers.

Another easily accessible observable is the wavelength of the distortion after buckling, λ, [Eq. (23)]. Figure 6 shows an electron micrograph of the shadow of a wall formed from a 140-nm-thick PMMA layer. The light areas are gold films created by lift-off. The measured wavelength of the distortion is 460 nm, in excellent agreement with the value of 459 nm given by Eq. (23) with $h=140$ nm.

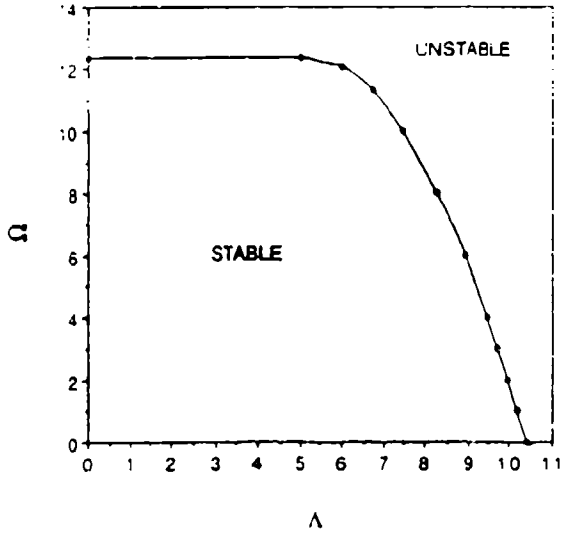


FIG. 3. Stability envelope for the case in which internal stresses (Λ) and lateral attractive forces (Ω) act simultaneously [Eqs. (14) and (15)]. For $\Lambda\leq 5$ the curve is practically horizontal, implying that for slightly subcritical values of Ω, large internal stresses are needed to cause instability. On the other hand, for slightly subcritical values of Λ a small lateral force is sufficient to induce buckling.

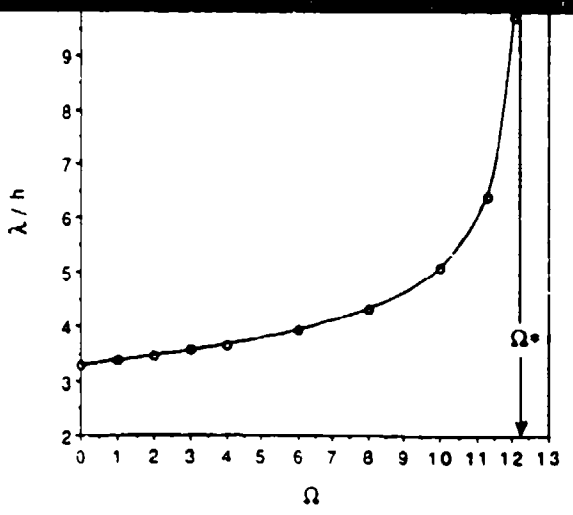


FIG. 4. Plot of the dimensionless wavelength of the distortion of the wall, λ/h, as a function of the dimensionless critical lateral stress, Ω.

The maximum amplitude of the lateral displacement of the wall Δ (see Fig. 2) can also be measured from the figure. The hypothesis that the deformation profile remains unaffected by solvent evaporation allows us to obtain an independent estimate of ε_{sw} from the observed value of Δ. The curvilinear contour length between nodal points at the top of the wall after buckling can be calculated from

$$L=\lambda(1+\varepsilon_{sw})=\frac{\lambda}{2\pi}\int_0^{2\pi}\sqrt{1+\left(\frac{2\pi\Delta}{\lambda}\right)^2\sin^2(z)}\;dz. \tag{27}$$

For $(2\pi\Delta/\lambda)^2\ll 1$ (subject to verification), the integration kernel in Eq. (27) can be expanded up to second order terms and integrated to give

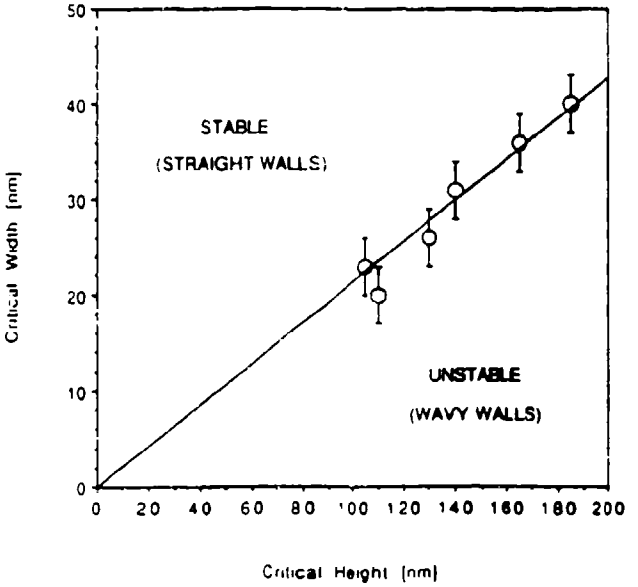


FIG. 5. Minimum width of PMMA walls achievable for a given starting resist thickness (wall height).

FIG. [illegible] Remaining metal after lift-off. The dashed lines mark light areas and dark wavy one is the shadow cast by PMMA wall. The scale marker is 500 nm.

$$\frac{\Delta}{\lambda} = \frac{\Delta}{\lambda}$$

This equation can be used to calculate [illegible] with the measured values of Δ and λ from Fig. [illegible]. We estimate $\lambda \approx$ [illegible] nm, giving with $\gamma = 0.4(a)$ nm $\gamma \approx$ [illegible]. The agreement with the previous estimate of [illegible] is satisfactory. These findings confirm that swelling is the dominant mechanism for instability in our experiments. It is remarkable that a [illegible] swelling can lead to the failure of a wall less than 10 times higher than its width.

In the above experiments the effects of lateral surface forces were presumably small due to the relatively large separation between features. As pointed out at the end of the previous section, for slightly supercritical values of h a small lateral force is sufficient to induce instability. In this regard, the magnitude of parameter Θ in Eq. [illegible] is expected to be of crucial importance. Theoretical estimates of Θ ranged from [illegible] N/m. Therefore, we may verify now whether the instabilities can be triggered by physically meaningful values of Θ.

The worst possible scenario envisions surface forces alone as sufficient to induce lateral collapse of the walls ($\gamma = 0$). Equation (26) can then be used to calculate a

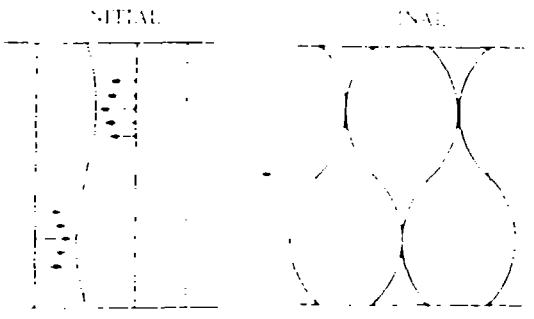


FIG. [illegible] Schematic illustration of metal islands in between closely spaced features immediately after metal deposition. One of the walls

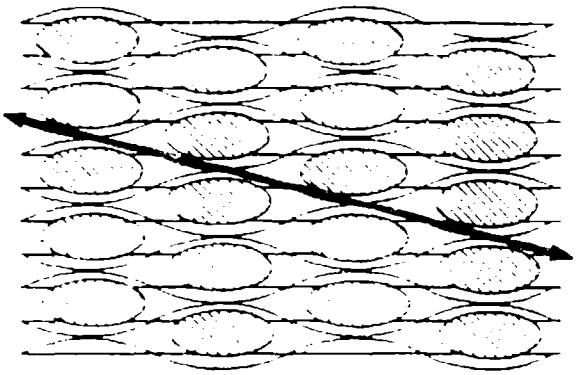


FIG. [illegible] Schematic of the final topology of the pattern in Fig. [illegible] The shaded sinusoidal areas represent metal patches left after metal evaporation and lift-off.

critical wall height, h^*, for assumed values of E, [illegible] and Θ. Choosing $E = 10^9$ MPa, [illegible], $\lambda = 500$ nm, and $\Theta = 10^{-2}$ N/m, we estimate $h^* \approx 250$ nm, which is in the range of interest for nanolithography.

The implications of these findings for fabrication of arrays of closely spaced walls can be substantial. Indeed, if surface forces can induce cross-correlation between deformation of adjacent features, with important consequences on the final topology of the pattern. Figure [illegible] schematically illustrates this effect. It is seen that the swelling induced buckling of one of the walls can lead to the catastrophic failure of the entire array, producing a pattern of adjacent sinusoidal waves 180° out-of-phase with respect to each other. Metallization followed by lift-off would then cause the appearance of a pattern of metal islands, as shown in Fig. 8. Experimental data supporting these ideas are shown in Fig. 9. Therefore, attractive surface interactions can be extremely important in the fabrication of thin, closely spaced features. This issue

FIG. [illegible] Scanning electron micrograph showing a failed pattern of an array of close [illegible]ed walls. The dark diagonal bands are the [illegible] the points where failed resist walls have come into contact prior to metallization. The scale marker corresponds to [illegible] μm. This figure should be compared with Fig. 8.

is especially crucial when dealing with marginally poor solvents, as is the case for most post-development, rinsing solutions.

As a final note on the modeling of wall stability, the following points must be considered. First, in the theoretical treatment we have ignored the effect of hydrodynamic forces which, at least in principle, could also be invoked as a source of instability (e.g., agitation during development). In fact, such effects are unimportant within the context of the stability problem, since upon cessation of flow a hypothetically deformed structure would always "spring back" to its most stable configuration (as dictated by the equilibrium criteria derived previously). Additionally, hydrodynamic effects cannot explain the correlation between deformation of adjacent structures observed in Fig. 9. Second, our simplified analysis assumes that the walls are: (1) strongly anchored at the bottom and (2) of constant width. The validity of the first assumption is easily inferred from the experiments, since floppy, weakly anchored walls would be incapable of retaining a deformed configuration of high strain energy such as the one shown in Fig. 6. In regards to the second assumption, we know that it cannot be entirely accurate, since undercutting inevitably leads to walls of decreasing thickness from top to bottom. A full analysis of this effect goes beyond the scope of the present work, but simple qualitative arguments in support of our approximation can be offered for the case of mildly tapered, strongly anchored walls in the swelling dominated regime (i.e., $\Omega \ll \Lambda$). Indeed, due to the anchoring boundary condition, the deformation at the bottom of the wall is small [see Eq. (16)]. Therefore, as a result of buckling, regions of the wall closer to the surface experience neither an appreciable release of strain energy nor a sizable decrease of surface free energy. In other words, the bottom portion of the wall contributes negligibly to the onset of instability, since its contribution to the overall energy balance is small. The relevant wall width, w, controlling the onset of the instability [i.e., the value to be used in Eqs. (14) and (22)] should then be closest in value to that at the top, free end. This argument is also consistent with the fact that, for $\Omega \ll \Lambda$, the wavelength of the distortion, λ [Eq. (23)], is independent of w.

SUMMARY AND CONCLUSIONS

Upon exposure to suitable solvents such as common developing and rinsing solutions, nanofabricated polymeric walls may become unstable as a result of two effects: (1) swelling and (2) lateral surface forces between adjacent structures. By performing a linear stability analysis we calculated the stability envelope and wavelength of the distortion for the general case in which internal swelling stresses and lateral surface forces act simultaneously. We also showed experimental evidence that supports our predictions, thus demonstrating the existence and severity of such instabilities in nanolithography. In the limiting case of small surface forces, predictions and data agreed quantitatively. Order-of-magnitude calculations of the strength of the lateral interactions between polymer walls (based on experimental surface force data) confirmed that lateral attractive forces can substantially enhance the action of swelling stresses, leading eventually to failure. Practical implications of these instabilities in the fabrication of arrays of closely spaced polymer walls were discussed. Formulation of developers that maximize contrast, but minimize swelling and attractive interfacial forces, is thus relevant to attempts to achieve higher resolution in nanostructure fabrication.

ACKNOWLEDGMENTS

We would like to thank D. T. Leighton for helpful discussions and S. Subramaniam for help in sample preparation. This research was supported in part by IBM, AFOSR, ONR, and the University of Notre Dame. D. A. H. gratefully acknowledges partial support by the Donors of The Petroleum Research Fund, administered by the American Chemical Society.

[1] D. R. Allee, S. Y. Chou, J. S. Harris, Jr., and R. F. W. Pease, J. Vac. Sci. Technol. B 7, 2015 (1989).
[2] J. C. Wu, M. N. Wybourne, W. Yindeepol, A. Weisshaar, and S. M. Goodnick, Appl. Phys. Lett. **59**, 102 (1991).
[3] S. Y. Chou, Y. Liu, and P. B. Fischer, IEDM **91**, 745 (1991).
[4] J. I. Goldstein, in *Introduction to Analytical Electron Microscopy*, edited by J. J. Hren, J. I. Goldstein, and D. C. Joy (Plenum, New York, 1979), p. 101.
[5] G. Owen, Repts. Prog. Phys. **48**, 795 (1985).
[6] A. Weill, in *The Physics and Fabrication of Microstructures and Microdevices*, edited by M. J. Kelly and C. Weisbuch (Springer, New York, 1986), p. 58.
[7] W. M. Moreau, *Semiconductor Lithography* (Plenum, New York, 1988), chap. 10.
[8] G. Park and K. Ueberreiter, in *Diffusion in Polymers*, edited by J. Crank and G. Park (Academic, New York, 1968), pp. 140–162, 219–257.
[9] J. N. Israelachvili and D. Tabor, Proc. R. Soc. London Ser. A **331**, 19 (1972).
[10] S. S. Patel and M. Tirrell, Ann. Rev. Phys. Chem. **40**, 597 (1989).
[11] G. H. Bernstein, D. A. Hill, and W. P. Liu, J. Appl. Phys. **71**, 4066 (1992).
[12] H. J. Taunton, C. Toprakcioglu, L. J. Fetters, and J. Klein, Macromolecules **23**, 571 (1990).
[13] S. P. Timoshenko and J. M. Gere, *Theory of Elastic Stability*, 2nd ed. (McGraw-Hill, New York, 1961), p. 363.

Reprinted from JOURNAL OF THE ELECTROCHEMICAL SOCIETY
Vol. 139, No. 10, October 1992
Printed in U.S.A.

Stability of Thin Resist Walls

Xiaokang Huang, Greg Bazán, Davide A. Hill,* and Gary H. Bernstein

Department of Electrical Engineering and Department of Chemical Engineering, University of Notre Dame, Notre Dame, Indiana 46556*

ABSTRACT

The physical strength of poly (methyl methacrylate) (PMMA) resist is an important factor for making very dense patterns with most types of lithography. For lines placed closely together using the lift-off technique, a thin wall of PMMA remains to act as the metal line spacer. When placed in contact with a suitable developer, thin polymer walls may swell and buckle, depending on the desired aspect ratio of the thin walls. We use electron beam lithography and a high-contrast developer solution to study the relationship of the maximum height-to-width ratio of the PMMA walls necessary to avoid buckling, and the resulting effects on metal thickness after lift-off. Maximum achievable aspect-ratios for our developer system were found to be constant (nearly 5:1) for widths from 20 to 45 nm and lengths of tens of microns. Theoretical predictions based on the hypothesis of a swelling-induced elastic instability are in excellent agreement with the experiments.

In the past few years, quantum devices based on ultrasmall geometries have received considerable attention. The usual case involves some feature, *e.g.*, gate length or metal line width, which is reduced in size in order to observe size quantization. In some cases, reduction of line widths plays only half the role, whereas a decrease in spacing between features plays an equally significant part. For example, split gates with 100 nm gaps in the form of bends have been made for the study of transmission properties at corners.[1] Also, lateral resonant tunneling transistors have been demonstrated with distances between gates of about 100 nm.[2] More recently, optical gratings have been fabricated with a metal line pitch of 50 nm over several microns.[3] All such devices require excellent control of line spacing.

Because of its uniquely high resolution, poly (methyl methacrylate) (PMMA) is among the most important resists for nanostructure fabrication. Utilizing the lift-off process, the resolution of resulting metal patterns depends not only on the width of the defined lines, which relates to the resolution of the beam formation system, but also to the quality of the edge profile of the resist, which depends to a great extent on the contrast of the resist/developer system. The

* Electrochemical Society Active Member.

most common developers for nanolithography are methyl isobutyl ketone:isopropyl alcohol (MIBK:IPA) [1:3][4] and cellosolve:methanol (3:7),[5] both demonstrating very high resolution. One weakness of these developers is the inability to dissolve high molecular weight components left behind from doses just at the threshold of complete development. This can lead to scum layers which can interfere with the lift-off process.[4] Higher doses necessary to decrease the molecular weight of the scum layer lead to a loss of resolution due to such factors as beam tails and energy scattering. It has been shown[5] that adding a minute percentage of methyl ethyl ketone (MEK) to the common developers can increase contrast and improve lift-off results, and therefore allow denser patterns to be fabricated.

In this paper we discuss the fabrication of extremely dense patterns on a solid substrate at energies below 50 keV without the use of such techniques as shadowing of a single pass exposure[3] or re-alignment of separate patterns.[4] Although patterns as small as 15 nm have been achieved with double tilt shadow evaporation,[3] a major difficulty is that of connecting the resulting lines to other structures within a device. Double alignment techniques[4] have shown equally impressive results with 16 nm lines on 24 nm pitch, but this method is inherently difficult and complicated, relies on the use of a scanning transmission electron microscope (STEM) for aligning the patterns, and requires multiple exposure steps. So far, only membranes, as opposed to solid substrates, have been utilized in this technique since alignments are facilitated using transmission imaging. Although adding significant complexity to the process, membranes offer an inherent advantage of reduced proximity effects, and thus a higher exposure modulation between the lines and spaces. Unfortunately, however, membranes are not amenable to device fabrication or suitable device performance. The developer used here helps to improve contrast so that very dense patterns can be fabricated on solid substrates even without very high beam energies offered by the use of TEMs.

Both the beam diameter of the exposure tool and the contrast of the resist/developer system have a significant effect on the achievable density of narrow metal lines. In the limit of very high resolution and contrast, the mechanical strength of the PMMA used for lift-off is the final limitation to making extremely closely spaced patterns. The lift-off process is described in Fig. 1. Exposing energy, in our case electrons, causes bond-breaking in the PMMA (Fig. 1a). After development (Fig. 1b), metal is evaporated perpendicularly to the surface, so that metal coats only the top of the resist and the exposed substrate surface (Fig. 1c). For thin resists at moderate energies, undercutting can approximately be described by[9]

$$b = 625\,(Z/E_0)t^{1.5}(\rho/A)^{1.5} \qquad [1]$$

where b is the increase of the diameter of the beam at the bottom surface of the resist, Z, A, ρ, and t are the atomic number, atomic weight, density, and thickness of the resist, respectively, and E_0 is the beam energy in keV. For PMMA, $Z = 3.6$, $A = 6.7$ g/mol, and $\rho = 1.2$ g/cm^3.[10] For 45 keV electrons and a PMMA thickness of 130 nm, undercutting is about 5 nm at each sidewall of our pattern, which aids in the lift-off process. During the lift-off step, dissolution of the resist in a suitable solvent causes removal of all metal except that on the substrate surface (Fig. 1d).

As seen above, the unexposed resist spacer is in the form of a thin wall between the developed lines. When lines are placed too closely together and high resolution is achieved, the resist walls can either become wavy, fall over, or even lose adhesion to the surface, in which case they fail completely. In this regard, internal stresses due to swelling can play a dominant role. Although PMMA suffers little swelling, which accounts partially for its ultrahigh resolution properties, some absorption of developer is necessary to aid the development process.[11] Swelling of polymeric resists has been extensively investigated.[12] In poor solvents, solvation forces favoring mixing are insufficient to promote dissolution of the polymer. Partial plasticization and

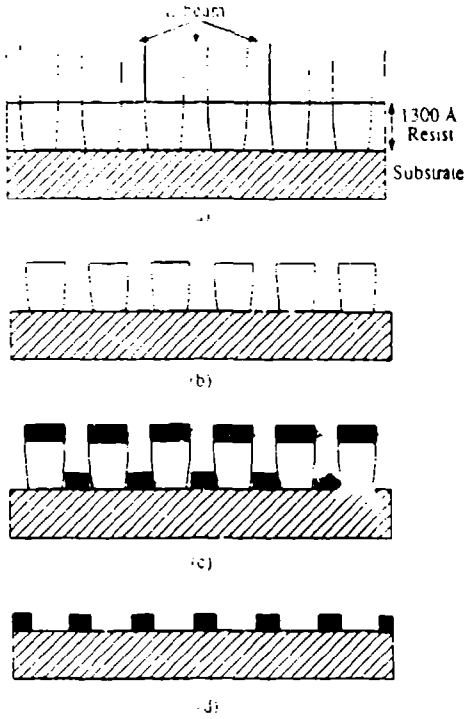


Fig. 1. Explanation of the lift-off process. (a) Exposure of resist by electron beam before development, resulting in slight undercutting. (b) Resist wall profile after development. (c) Resist walls after metallization from small, perpendicular source. (d) Metal pattern after dissolution of remaining resist in suitable solvent.

swelling can occur, however, leading to weaker mechanical properties, and to the build up of internal stresses in constrained systems. Typically, glassy polymers can experience between 1 and 30% swelling, without dissolution, depending on solvent quality and temperature.[13] Such a low degree of expansion may be irrelevant in most cases, but, as shown below, can lead to severe dimensional stability problems in the fabrication of thin polymer walls.

Experimental

Our EBL system consists of an Amray 1400 scanning electron microscope (SEM) with a minimum spot size of 35 Å controlled by an IBM PS/2 personal computer interfaced through a Pragmatic Systems 2201A 16-bit arbitrary waveform generator. The details of the system are described elsewhere.[14]

We used PMMA with molecular weight 950,000 amu baked at 170°C for 4 h for all exposures. Resist and developer parameters were very tightly controlled. Resist thickness was 105 to 180 nm as measured with an Alpha-Step 200 surface profiler. The developer used in all cases was a solution of MIBK:IPA (1:3) with the addition of 1% MEK by volume.[5] It has been shown that contrast results are extremely sensitive to variations in developer temperature due to the high activation energy of the PMMA in the developer solvents.[6] Using a temperature controller, our developer temperature was maintained at 25 ± 0.5°C. Develop time for all samples was 15 s followed by a 15 s rinse in methanol. After development, all samples were dried in air at 90°C for 5 min.

All exposures were performed at a beam energy of 45 keV at a spot size of 5 nm and 120 μm scan field. Substrates used in all cases were 200 nm SiO_2 on thick silicon wafers. Doses were very critical for making high-density patterns because of proximity effects during E-beam lithography. They were not only related to the line pitch, but also to the size of the patterns. For grating pitches ranging from 70 to 150 nm, the doses varied from 1.5 to 2.5 nC/cm. For the case of only two parallel lines, doses ranged from 2.5 to 3.5 nC/cm for pitches from 80 to 120 nm.

Fig. 2. Remaining metal after lift-off of a typical failed wall. Light areas are gold, and dark, wavy line is the shadow of failed PMMA wall.

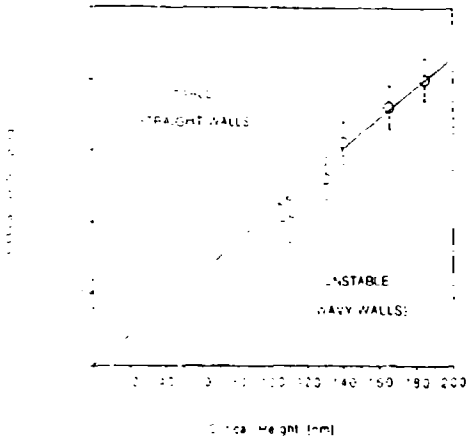


Fig. 3. Minimum width of PMMA walls achievable for a given starting resist thickness (wall height).

All metal evaporations were gold from an electron beam evaporator with crucible size of 2.5 cm and a distance of 40 cm from the source to wafer. Metal thicknesses were closely calibrated with a film thickness monitor and a surface profilometer. All metal thicknesses were 30 ± 2 nm. Lift-off was performed by soaking the samples in acetone for about 5 min, followed by the application of acetone using a hypodermic syringe, followed by a 15 s methanol rinse, and final blow drying in nitrogen.

Results

In order to achieve lift-off of evaporated metal films, it is necessary that PMMA walls remain between the developed areas. Assuming minor undercutting for thin resists and high beam energies, it is expected that the spaces of the resulting metal patterns closely reflect the width of the remaining PMMA walls, especially at the top. Because of the very low solubility of unexposed PMMA of very high molecular weight,[7] the height of the walls after development is still approximately the same as the initial resist thickness.

After application and bake, PMMA can be considered a strong, glassy material. Exposure to the developer, however, causes swelling of the polymer, leading to the buildup of internal stresses in strongly anchored structures. Compressive stresses in the wall can, in turn, lead to buckling.

The wall aspect-ratio (height-to-width) is an important parameter in delineating the stability threshold, since a tall, thin wall will be strongly susceptible to small dimensional perturbations around the erect position, while a short wall will be able to sustain a higher degree of swelling before buckling. Figure 2 shows the remaining metal after lift-off of a typical failed wall. The bright areas are evaporated gold, and the dark, wavy center line is the shadow of the remaining wall over which gold was deposited and removed. The wall was about 30 nm wide and 140 nm high. We show later that a direct relationship exists between the original wall height and the wavelength of the distortion.

Figure 3 shows the minimum width of a PMMA wall achievable for a given starting resist thickness. Error bars represent variations in metal line separation (and therefore wall thickness) over the length of the lines. These data were obtained by smoothly varying the line pitch for a given resist thickness and observing the point at which lines either failed completely or became noticeably wavy. The data indicate a direct proportionality between critical wall height and width, implying a constant aspect-ratio for buckling.

Although the lower the thickness of the PMMA, the closer the achievable line spacing, the resist thickness cannot be lowered arbitrarily because, for different processes, different thickness of metal are required depending on step coverage, line continuity, multiple metal layers, device scaling requirements, electrical resistance constraints, opacity to radiation (e.g., x-rays), etc. We have determined that to achieve reliable lift-off, the thickness of PMMA should be at least three times the desired thickness of the small metal pattern. This rule of thumb will depend to some extent on the metal evaporation system, since the physical size of the evaporant source, the source-to-sample spacing, and the care taken to place the sample perpendicular to and directly above the source will all affect the amount of metal deposited on the inside surfaces of the PMMA etched by the developer. Using this relationship, and guided by the "critical-aspect-ratio" criterion, we fabricated double gold lines with an approximately 20 nm space and a metal thickness of 30 nm, from 105 nm thick resist layers. The lines were continuous and straight over a distance of 25 μm. Figure 4 shows a portion of these lines at high magnification.

Discussion

We begin our discussion with some simple theoretical considerations about the stability of a long wall with a rigidly anchored bottom and a top free end subject to internal stresses. Comparison with experiments will then closely follow the theoretical predictions.

The system of interest is shown in Fig. 5, which also gives the characteristic geometrical parameters. Mathematically, the problem is conveniently formulated within the framework of linear plate theory,[8] leading to the following homogeneous boundary value problem for the lateral displacement, δ

$$\frac{\partial^4\delta}{\partial x^4} + 2\frac{\partial^4\delta}{\partial x^2\partial y^2} + \frac{\partial^4\delta}{\partial y^4} = -\frac{\sigma_x 12(1-\nu^2)}{Et^2}\frac{\partial^2\delta}{\partial x^2} \qquad [2]$$

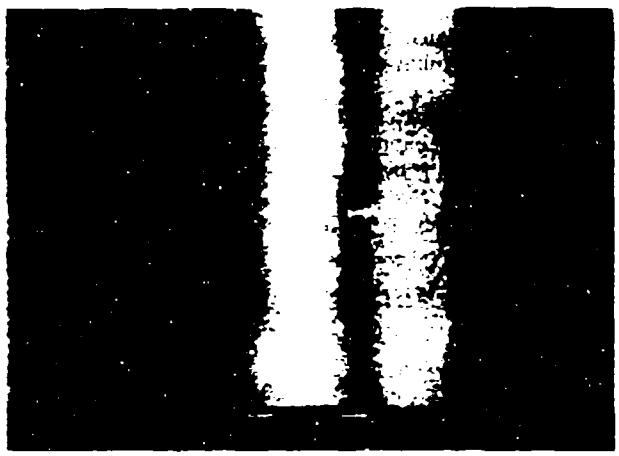


Fig. 4. Very closely spaced gold lines at high magnification. The distance between lines is 20 nm over a distance of 25 μm.

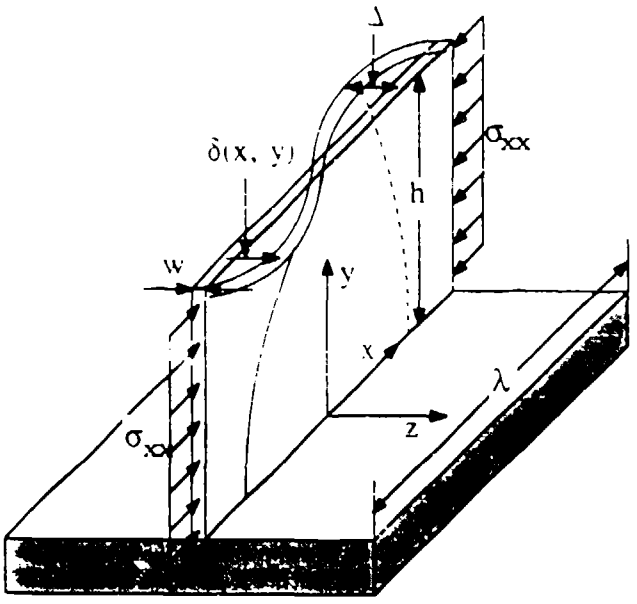


Fig. 5. Diagram of tall, thin wall demonstrating geometrical parameters used in the analysis

$\delta = 0$ at $y = 0$ (no displacement at the bottom) [3]

$\frac{\partial \delta}{\partial y} = 0$ at $y = 0$ (built-in bottom end) [4]

$\frac{\partial^2 \delta}{\partial y^2} + \nu \frac{\partial^2 \delta}{\partial x^2} = 0$ at $y = h$ (torque-free top) [5]

$\frac{\partial^3 \delta}{\partial y^3} + (2 - \nu) \frac{\partial^3 \delta}{\partial x^2 \partial y} = 0$ at $y = h$ (shear force-free top) [6]

where $\sigma_{zz} \geq 0$ is the internal compressive stress, ν Poisson's modulus, E Young's modulus, and w and h the wall thickness and height, respectively. For an infinitely long wall the solution to Eq. 2 is of the form

$$\delta(x, y) = \cos(\xi x/h)[A(\sinh(\alpha y/h) - \sin(\beta y/h)) + B(\cosh(\alpha y/h) - \cos(\beta y/h))] \quad [7]$$

where A and B are integration constants, and

$$\alpha = \sqrt{\xi^2 + \sqrt{\Lambda \xi^2}} \quad [8]$$

$$\beta = \sqrt{-\xi^2 + \sqrt{\Lambda \xi^2}}$$

$$\Lambda = \frac{\sigma_{zz} 12(1 - \nu^2) h^2}{E w^3}$$

Variable ξ is related to the wavelength of the prevailing distortion, λ, as follows

$$\lambda = 2\pi h/\xi \quad [9]$$

Imposition of the boundary conditions provides the following equation for the eigenvalues

$$2ts + (s^2 + t^2)\cosh(\alpha)\cos(\beta) = \frac{1}{\alpha\beta}(\alpha^2 t^2 - \beta^2 s^2)\sinh(\alpha)\sin(\beta) \quad [10]$$

where

$$t = \beta^2 + \nu\xi^2$$

$$s = \alpha^2 - \nu\xi^2 \quad [11]$$

As customary in stability theory, Eq. 10 is satisfied only by a discrete set of ξ values, provided Λ is greater than or equal to a minimum critical value, Λ^*. Physically, Λ^* is the smallest (dimensionless) load necessary to reach the marginal stability limit (indifferent equilibrium), and can be calculated by solving Eq. 10 numerically. We obtain $\Lambda^* = 19.40$, and $\xi^* = 1.917$. Combining these results with Eq. 8 and 9, the smallest critical stress, σ^*_{zz}, and the associated wavelength, λ, are calculated to be

$$\sigma^*_{zz} = 26\, E w^3/[30(1 - \nu^2)h^2] \quad [12]$$

$$\lambda = h\, 2\pi/1.917 \quad [13]$$

For swelling-induced stresses, σ^*_{zz} can in turn be obtained from

$$\sigma^*_{zz} = E\varepsilon_{sw} \quad [14]$$

where ε_{sw} is the hypothetical swelling strain that the wall would undergo if allowed to expand freely. Equations 12 and 14 allow us to calculate the critical wall aspect-ratio for stability, $(h/w)^*$, as a function of the swelling strain

$$\left(\frac{h}{w}\right)^* = \sqrt{\frac{26}{30(1 - \nu^2)\varepsilon_{sw}}} \quad [15]$$

For $(h/w) > (h/w)^*$ the wall will buckle. Equation 15 is more convenient for comparison with experiments than Eq. 12, since it does not contain the unknown modulus E (which presumably changes upon swelling). As expected, a short wall can sustain a higher degree of swelling before buckling. Additionally, Eq. 15 predicts that, for a constant ε_{sw}, the critical wall height and width should be directly proportional to each other. This prediction is in excellent agreement with our data, as shown in Fig. 3.

The slope of the line in the figure provides the critical aspect ratio for buckling: $(h/w)^* = 4.75$. Calculation of ε_{sw} from Eq. 15 requires knowledge of Poisson's modulus, ν. Because the polymer is swollen, and the stresses are relatively small, the assumption $\nu = 0.50$ is appropriate, giving: $\varepsilon_{sw} = 5.1\%$. This value is well within the expected range for glassy polymers.[9]

Another easily accessible observable is the wavelength of the distortion after buckling, λ, (Eq. 13). In Fig. 2, the measured wavelength of the distortion is 460 nm, which is in excellent agreement with the value of 459 nm given by Eq. 13 with $h = 140$ nm.

The maximum amplitude of the lateral displacement of the wall, Δ (see Fig. 5), can also be measured by SEM. If we assume that the deformation profile is unaffected by solvent evaporation, then the observed value of Δ can be used to obtain an independent estimate of ε_{sw} as follows.

The curvilinear contour length between nodal points at the top of the wall after buckling can be calculated from

$$L = \lambda(1 + \varepsilon_{sw}) = \frac{\lambda}{2\pi}\int_0^{2\pi} \sqrt{1 + \left(\frac{2\pi\Delta}{\lambda}\right)^2 \sin^2(z)}\, dz \quad [16]$$

For $(2\pi\Delta/\lambda)^2 \ll 1$ (subject to verification), the kernel in Eq. 16 can be expanded up to second-order terms and integrated to give

$$\varepsilon_{sw} = \left(\frac{\pi\Delta}{\lambda}\right)^2 \left[1 - \frac{3}{2}\left(\frac{\pi\Delta}{\lambda}\right)^2\right] \quad [17]$$

This equation can be used to calculate ε_{sw}, with the measured values of Δ and λ. From Fig. 2 we estimated $\Delta \approx 33$ nm, giving (with $\lambda = 460$ nm) $\varepsilon_{sw} = 4.7\%$. This value is in good agreement with the previous estimate of 5.1%. These findings confirm that swelling is the dominant mechanism for instability in our experiments. It is remarkable that a 5% swelling can lead to the failure of a wall less than five times higher than its width.

Summary and Conclusions

We have investigated the relationship between minimum line spacing and maximum metal thickness for EBL-fabricated patterns using high-contrast resists. Both theory and experiments confirm that a small degree of swelling may lead to the catastrophic failure of PMMA walls of relatively modest aspect-ratio. For subcritical aspect-ratios, line spacings as small as 20 nm, and metal thicknesses as high as 700 nm were reliably achieved without loss of adhesion or

distortion. We believe that the formulation of developers that maximize contrast while minimizing swelling is desirable for attaining higher resolution in nanostructure fabrication. Hopefully, the results of this paper will permit more efficient screening of improved developer solutions through the initial observation of swelling properties.

Acknowledgments

We would like to thank D. T. Leighton for helpful discussions and S. Subramaniam for help in sample preparation. This research was supported in part by IBM, AFOSR, ONR, and the University of Notre Dame. D.A.H. gratefully acknowledges the Donors of The Petroleum Research Fund, administered by the American Chemical Society, for partial support of this research.

Manuscript submitted April 30, 1992; revised manuscript received June 22, 1992.

The University of Notre Dame assisted in meeting the publication costs of this article.

REFERENCES

1. J. C. Wu, M. N. Wybourne, W. Yindeepol, A. Weisshaar, and S. M. Goodnick, *Appl. Phys. Lett.*, 59, 102 (1991).
2. D. R. Allee, S. Y. Chou, J. S. Harris, Jr., and R. F. W. Pease, *J. Vac. Sci. Technol. B*, 7, 2015 (1989).
3. S. Y. Chou, Y. Liu, and P. B. Fischer, *IEDM Tech. Dig.*, p. 745 (1991).
4. S. P. Beaumont, P. G. Bower, T. Tamamura, and C. D. W. Wilkinson, *Appl. Phys. Lett.*, **38**, 436 (1981).
5. H. G. Craighead, R. E. Howard, L. D. Jackel, and P. M. Mankiewich, *ibid.*, **42**, 38 (1983).
6. G. H. Bernstein, D. A. Hill, and W. P. Liu, *J. Appl. Phys.*, 71, 4066 (1992).
7. S. Y. Chou and P. B. Fischer, *J. Vac. Sci. Technol. B*, 8, 1919 (1990).
8. K. Y. Lee, J. Frost, C. Staniey, W. Patrick, W. S. Mackie, S. P. Beaumont, and C. D. W. Wilkinson, *ibid.*, 5, 322 (1987).
9. J. I. Goldstein, in *Introduction to Analytical Electron Microscopy*, J. J. Hren, J. I. Goldstein, and D. C. Joy, Editors, p. 101, Plenum Press, New York (1979).
10. G. Owen, *Reps. Prog. Phys.*, **48**, 795 (1985).
11. A. Weill, in *The Physics and Fabrication of Microstructures and Microdevices*, M. J. Kelly and C. Weisbuch, Editors, p. 58, Springer-Verlag, New York (1986).
12. W. M. Moreau, *Semiconductor Lithography*, Chap. 10, Plenum Press, New York (1988).
13. G. Park and K. Ueberreiter, in *Diffusion in Polymers*, J. Crank and G. Park, Editors, pp. 140-162, 219-257, Academic Press, Inc., New York (1968).
14. G. Bazan and G. H. Bernstein, In preparation.
15. S. P. Timoshenko and J. M. Gere, *Theory of Elastic Stability*, 2nd ed., p. 363, McGraw-Hill, Inc., New York (1961).

ON THE ATTAINMENT OF OPTIMUM DEVELOPER PARAMETERS FOR PMMA RESIST

Gary H. Bernstein[a] and Davide A. Hill[b]

[a]Department of Electrical Engineering
[b]Department of Chemical Engineering
University of Notre Dame
Notre Dame, IN 46556

(Received 19 May 1991)

Because of the uniquely high resolution offered by poly(methyl methacrylate) (PMMA) for practically all forms of lithography, its technological significance remains strong. We have found significantly higher contrast developers for PMMA which have shown improved resolution, less susceptibility to proximity effects and better lift-off properties with no loss of sensitivity to exposure. We report data which demonstrate an improvement in contrast by more than 35% with the addition of as little as 1.3% methyl ethyl ketone (mek) to other common developers. We discuss a model which explains the observed contrast enhancement and predicts that for small changes, an increase in either mek concentration, temperature or develop time increases contrast. We include experimental evidence in support of temperature and mek related improvements.

1 Introduction

Of all materials utilized in the fabrication of nanostructures, poly (methyl methacrylate) (PMMA) is among the most important. PMMA remains so popular as a positive resist that virtually all laterally defined quantum devices rely on PMMA for patterns below 100 nm regardless of the lithographic technique employed. In fact, PMMA is the only proven positive resist usable into the 10 nm regime. (Contamination resist is utilized for etch masks in this dimensional regime, but is slower and more difficult to work with.)

The typical manner in which PMMA is employed is that of exposure to electrons, X-rays or deep UV light followed by development, metallization and lift-off. The resolution of the resulting metal pattern depends on both the width of the defined line and the quality of the edge profile in the resist. In turn, edge profile depends on the contrast, γ, of the resist/developer system. Contrast is defined as the maximum slope of the curve of normalized resist thickness remaining after development as a function of the log of the exposure dose. As contrast increases, such properties as lift-off reliability, proximity effect immunity and, to some extent, resolution, are all improved.

We have found that several parameters can increase the contrast of PMMA. First, we found that adding a very small percentage of methyl ethyl ketone, (mek) to other common developers effects a drastic improvement in contrast, dimensional consistency and lift-off quality. Second, we have found that, over a narrow temperature range, an increase in temperature also improves contrast. Finally, as discussed in further sections, we propose that an increase in develop time is also predicted to improve contrast.

2 Experimental

Contrast results for a variety of developer formulations were determined using a standard technique[1,2]. Using electron beam lithography (EBL) at 40 kV, we exposed 25 50 μm x 50 μm squares in 950,000 amu PMMA, 0.4 to 1 μm thick with a series of doses, developed for various times and temperatures, and determined the normalized resist

0749-6036/92/020237 + 04 $02.00/0

thicknesses by surface profilometry using a Sloan Dektak II profilometer. The quality of the data was excellent, and an example of the raw data is given in Fig. 1. Variations in resist thickness were shown to have no effect on the results. Experimental parameters and contrasts are summarized in Table I.

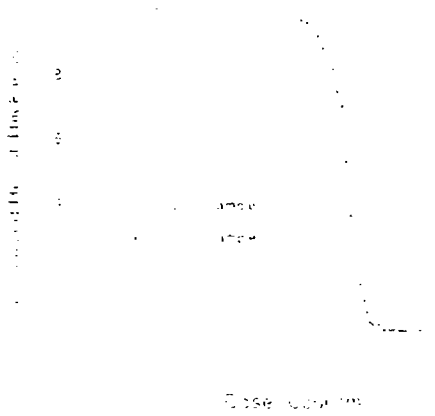


Figure 1. Example of data from which contrast is calculated.

The primary ingredients of the developer mixtures were chosen from solutions commonly employed as developers for PMMA. These are methyl ethyl ketone ("mek"), ethyl alcohol (EtOH) 26.5/73.5 mixture referred to here as "MEK", methyl isobutyl ketone (mibk), isopropyl alcohol (IPA) 1/3 mixture referred to here as "MIBK", and 2-ethoxy ethanol (Cellosolve or "cs"), methanol (MeOH) mixture referred to here as "CS". These mixtures were chosen in their complete forms due to their known properties as developers of PMMA. It should be noted that solubility properties of mixtures of several solvents are extremely difficult or impossible to predict, and such an exercise will not be attempted here. In subsequent sections we offer a qualitative explanation of the role of the harshest component, mek, in improving contrast and offer guidelines toward future improvements in contrast.

3. Results and Discussion

For each resist given in Table I, we calculated contrast and plotted the results as a function of development temperature, as shown in Fig. 2. [illegible] increasing contrast with temperature [illegible]

Table I. Summary of contrast experiment data.

Sample I.D.	Volume percent MIBK	Volume percent CS	Volume percent MEK	Develop Time (sec)	Temp (°C)	γ
1	0	100	0	60	19	1.6
2	0	100	0	60	21	1.7
3	100	0	0	60	19	1.1
4	100	0	0	60	21	1.1
5	66.4	32.0	1.6	60	19	2.1
6	66.4	32.0	1.6	60	21	2.1
7	49.6	48.0	2.4	60	19	3.2
8	49.6	48.0	2.4	60	21	3.3
9	50.1	45.4	4.5	60	19	7.1
10	50.1	45.4	4.5	60	19	7.6
11	50.1	45.4	4.5	60	19	9.4
12	50.1	45.4	4.5	60	19	9.4
13	50.1	45.4	4.5	60	19	9.1
14	50.1	45.4	4.5	60	19	8.2
15	0	95	5	60	21	12.6
16	0	0	100	60	19	4.7

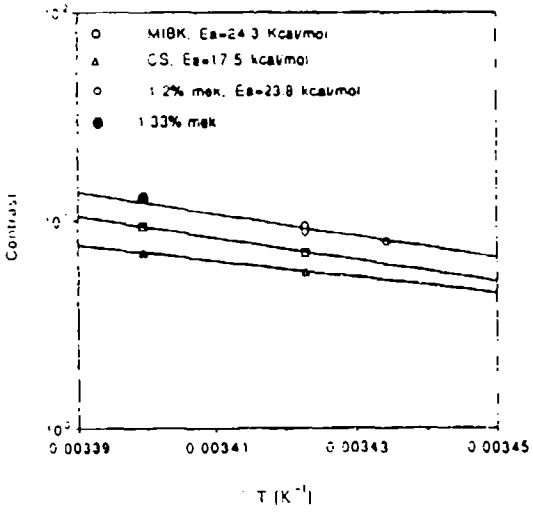


Figure 2. Arrhenius plot of contrast for various formulations.

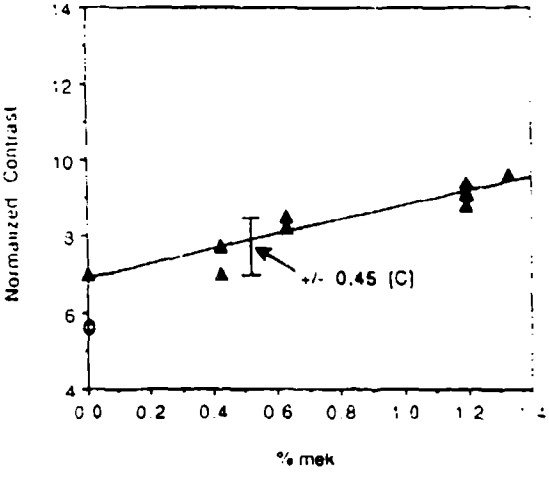


Figure 3. Plot of contrast as a function of % mek formulation normalized to 19 C.

between 18 C and 21 C is clearly evident. This range is especially important since room temperature variations coupled with evaporative cooling of the developer solvents are shown to be especially critical. It can be shown[3] that γ can be expressed as

$$\gamma = \left[r_d \frac{\partial \log r_d}{\partial \log M} \frac{\partial \log M}{\partial \log D} \frac{\Delta t_d}{h_0} \right]_{max} \quad (1)$$

where r_d is the average rate of dissolution, M is the molecular weight, D is the dose, Δt_d is the development time and h_0 is the initial resist thickness. The first term, r_d, depends on $\exp(-E_a/RT)$ where E_a is the activation energy, R is the gas constant and T is the temperature. This theory also predicts that γ depends linearly on develop time. This has not yet been investigated and remains for future work on this topic.

Figure 3 gives the dependence of contrast on percent mek (pure methyl ethyl ketone component) normalized to 19°C. The CS data are omitted since the activation energy of Cellosolve is significantly different from that of MIBK and MIBK/MEK mixtures (see Fig. 1) and therefore do not normalize to the plot. Note that data corresponding to 21°C for pure MIBK and that of 18°C at 1.19% mek have collapsed onto the straight line and their multiplicity is obscured.

An explanation of the time, temperature and mek concentration dependence follows. Figure 4 is the qualitative dependence of dissolution rate on the

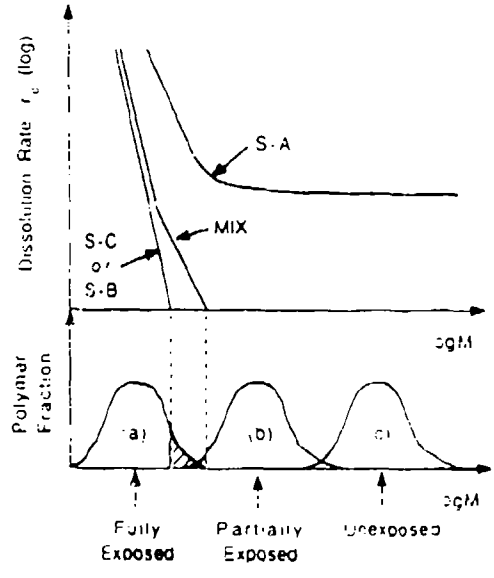


Figure 4. Dependence of dissolution rate on temperature, time and solvent strength superimposed on fragmented molecular weight distribution.

strength of a solvent, time or temperature as a function of molecular weight, M. This dependence has been demonstrated by Greeneich for MIBK/IPA solutions[4]. Below this curve is the dependence of the M distribution after three exposure doses[2]. First, as any of the above parameters is slightly increased, the solvent quality changes in such a manner as to dissolve more efficiently the higher M tails remaining

after normal exposures[3]. For the case of very slight amounts of MEK added to MIBK, the mek acts independently of the MIBK to remove efficiently high M components left behind by the MIBK, thus improving contrast, side wall profiles and proximity effect immunity without decreasing sensitivity. The effect on lower M components is not increased enough to affect deleteriously the contrast[3].

The above results suggest a way in which developers might be further formulated to increase contrast even higher. We suggest that solvents be chosen for their range of effectiveness in dissolving various molecular weights of PMMA. Assuming these solvents retain their basic properties in solution, the solvents could be mixed in the percentages which reflect the percentages of molecular weights of the PMMA remaining after exposure (as shown in Fig. 4). These tailored developers would be strong enough to remove only the proportions of the PMMA in the particular ranges of M affected by the individual components but would not act to remove higher M components in adjoining, less exposed areas or in unexposed areas. In a very simplified way, this is what has been accomplished by the addition of mek, a much stronger solvent for PMMA, to mibk. Clearly, much work in this area is still to be done.

4 Summary

We have demonstrated that new developer solutions containing small amounts of a strong developer mixed with a weaker one can enhance all aspects of PMMA development with no deleterious side effects. In addition, we demonstrated experimentally that good control of temperature is critical for consistent contrast control, where in fact, at least slightly higher temperatures are desirable. New directions for research in this area were suggested.

Acknowledgements

The authors wish to thank Wen-Ping Liu for sample preparation and D. K. Ferry for helpful discussions. This work was supported in part by IBM and ONR (G.H.B.) and the University of Notre Dame.

References

1. M. A. Mohsin and J. M. G. Cowie, *Polymer* **29**, 2130 (1988).
2. V. K. Sharma, R. A. Pethrick and S. Affrossman, *Polymer* **23**, 1732 (1982).
3. D. A. Hill and G. H. Bernstein, unpublished.
4. J. S. Greeneich, *J. Electrochem. Soc. Sol. St. Sci. and Techn.* **122**, 970 (1975).
5. S. Mackie and S. P. Beaumont, *Sol. St. Tech.* August, 117 (1985).

New high-contrast developers for poly(methyl methacrylate) resist

Gary H. Bernstein[a)]
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Davide A. Hill
Department of Chemical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Wen-Ping Liu[b)]
Center for Solid State Electronics Research, Arizona State University, Tempe, Arizona 85287-6206

(Received 3 October 1991; accepted for publication 8 January 1992)

New developers for poly(methyl methacrylate) consisting of mixtures of common developing components have been carefully investigated. It has been found that adding a small percentage of methyl ethyl ketone to methyl isobutyl ketone and Cellosolve results in a significant increase in contrast. Results of contrast experiments as well as improvements in electron-beam lithographic exposures are reported. An explanation of the mechanism of contrast and resolution enhancement is offered.

I. INTRODUCTION

Very large scale integration (VLSI) technology continues to push toward smaller geometries with advancement in the $\frac{1}{3}$ μm range and experimental circuits such as monolithic microwave integrated circuits (MMICs) in the 0.1 μm range. The fabrication of nanostructures and quantum devices below 0.1 μm especially relies on ultrahigh lithographic resolution, and as such continues to utilize primarily poly(methyl methacrylate) (PMMA) as the resist of choice. PMMA continues to find applications in electron-beam lithography (EBL), focused ion-beam lithography, x-ray lithography, deep uv optical lithography, and, more recently, scanning tunneling lithography.[1] It is almost universally the choice for the development of new lithographic techniques since PMMA offers the highest known resolution of any organic resist. A drawback to the use of PMMA is that it is generally not sensitive enough for manufacturing purposes, but it does offer extremely high resolution. For this reason, PMMA remains technologically important.

The generally accepted criterion of resolution prediction of a resist-developer system has fallen on contrast γ. High contrast can be thought of as the ability of a developer to discriminate between very small differences in exposure dose from one small area to an adjoining one. Exposure in electron-beam lithography is not precisely localized due to factors including the Gaussian beam cross section, forward scattering in the resist, backscattering from the resist and substrate (proximity effect), and the generation of secondary electrons in the resist.[2] These factors decrease the modulation transfer function of the exposure for closely spaced patterns.

In PMMA, a positive resist, bonds are broken by incident radiation (e.g., an electron beam), decreasing the average molecular weight M of the long-chain molecules. This increases the solubility of the resist in a suitable developer allowing the exposed pattern to be indented into the resist. In further processing steps, the pattern is transferred by various techniques to the substrate material. An important limitation on the density of the exposed pattern, and in some respects the individual line-widths, is the proximity effect which causes exposure of the resist up to several micrometers away from the area of the primary exposure. As lines are placed closer together, the total dose increases between the primary exposed regions, thus leading to a decreased differential in exposure.[2] The requirements of the developer to distinguish between differences in dose become more stringent if the pattern is to be resolved.

Ideally, a developer will remove all positive resist that has been exposed to a dose above a certain threshold and none of the resist that has been exposed below that threshold. In practice, a range of exposures results in partial development of the resist over that range. For this reason, higher contrasts allow a narrower area to be developed completely even in the presence of those factors that broaden the exposure. For other lithography schemes mentioned above, similar energy scattering limitations to resolution exist. It is expected that for all types of lithography, higher resolution will result from higher contrast, and the improvements discussed below will be equally relevant to any application of PMMA.

Contrast is a measure of the sharpness of the developing threshold of the developer-resist system and is defined as the slope of the plot of normalized thickness of the resist remaining after development versus the logarithm of the exposure dose. Figure 1 illustrates the definition of γ. The onset of development D_i is defined to be the dose at the intersection of the steepest part of the contrast curve extrapolated to the 100% level of the curve. The sensitivity D_f is defined to be the dose of the extrapolated curve at zero thickness.

As discussed above, γ is a measure of the ability of the developer to distinguish between gradations of dose. Since contrast is expressed as

$$\gamma = [\log(D_f/D_i)]^{-1}, \tag{1}$$

a)Author to whom correspondence should be addressed.
b)Present Address: VLSI Research Department, Lishan Microelectronics Co., P. O. Box 19, Lintong, Shaanxi 710600, Peoples Republic of China.

 0021-8979/92/084066-10$04.00

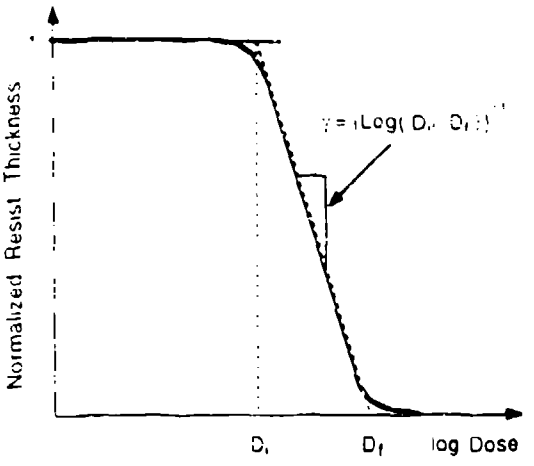


FIG. 1. Plot of normalized positive resist thickness remaining after development as a function of log of exposure dose (the contrast curve) illustrating the definitions of D_i and D_f.

a higher absolute value of γ represents a higher contrast developer. (In this report, the absolute value of all contrast values will be used, but are all negative.) The ideal situation discussed above in which there is an absolute dose threshold for dissolution represents an infinite contrast, whereas in practice the contrast is always finite. Contrast is commonly measured experimentally for a given development time by varying the exposure dose in small increments through the region of dose that causes the resist to dissolve until total dissolution is achieved. A common method for measuring γ is to use a surface profilometer to measure the thickness of resist remaining after exposure and development and plot the data as discussed above.[3-9]

As we discuss later, the contrast parameter is entirely dictated by the dissolution behavior of the polymer. Papanu *et al.* have produced valuable fundamental work on the mechanism of PMMA dissolution in one- and two-solvent developers, and for different polymer molecular weights.[10] Their experiments greatly clarify the role of thermodynamic quality of the solvent on the rate and mechanism of polymer dissolution, but only for one- and two-component developer mixtures. Here we approach the more complex case of a multicomponent developer and, on more practical grounds, we study the effect of solvent quality (in our case various mixtures) on the rate of dissolution of electron-beam-exposed PMMA resists. We explore combinations of common developers to ascertain whether or not improvements in the lithographic process could be promoted simply by using readily available mixtures of solvents.

These new developer solutions are mixtures of chemicals that have been commonly employed individually as developers for PMMA, but which when combined exhibit higher performance than when alone. Our combination of these chemicals has shown improvements in achievable contrast of PMMA as well as improved lithographic results. Some of our developer solutions give improvements of up to 35% in contrast and no loss in sensitivity.

Three chemical solutions are often employed as PMMA developers. These are (i) methyl ethyl ketone:ethanol (EtOH) in a ratio of 26.5:73.5; (ii) Cellosolve (2-ethoxyethanol):methanol (MeOH) in a ratio of 3:7; and (iii) methyl isobutyl ketone:2-propanol (IPA) in a ratio of 1:3. These solutions are referred to here simply as solution A, or S-A, solution B, or S-B, and solution C, or S-C, respectively. Pure methyl ethyl ketone, Cellosolve, and methyl isobutyl ketone will be represented by MEK, CS, and MIBK, respectively. S-C has been reported to provide very high resolution at exposures of 50 kV electrons with a development time of 15–45 s[11,12] while S-B has been reported to provide very high resolution in EBL exposures between 20 and 120 kV with a development time of 5 s.[13,14] It is mixtures of these three common developer solutions which we have found to give markedly improved results in contrast.

II. EXPERIMENT

For resist exposures, we used an ISI-100B scanning electron microscope (SEM) modified for EBL in a vector-scanned mode with an exposure spot (pixel) spacing and beam diameter of 100 nm each.[15] All test exposures were performed at 40 kV on thick silicon substrates. All PMMA films were of molecular weight 950 000 a.u., spin coated and baked at 160 °C for 4 h, with resist thicknesses ranging from 0.4 to 1.0 μm. Exposures were arranged in a 6×6 array of 50×50 μm^2 squares. Resist thicknesses remaining after development were obtained using a Dektak II surface profilometer.

S-A, S-B, and S-C were tested alone and in combination using development times that yielded approximately the same dose for complete development and also corresponded to typical values from the literature. These times were 45 s for S-C,[11,12] 5 s for S-B,[13,14] and 2 s for S-A. The development time for almost all mixtures of these components was 10 s. This time was based upon the relative strength and concentrations of the components used in the mixtures and was estimated initially to yield the same time to complete development for approximately similar doses. The range of developer temperatures represents common fluctuations caused by varying room conditions, evaporative cooling, etc.

Development was performed by holding the samples with tweezers and dipping in the developer with light agitation for the required time. Temperatures during development were determined by inserting a mercury-bulb thermometer, calibrated to 1 °C, into the beaker. Developer times were controlled as much as possible by careful attention to the process. Significant error occurred only in the possible case of S-A with a 2 s develop (estimated at less than 20%), but this data was used only for crude comparison as will be discussed below. All samples were rinsed for 15 s in MeOH immediately after development and blown dry in nitrogen.

The unexposed dissolution properties of the three developer components were tested to ensure that normalized

TABLE I. Results of contrast experiment identified by sample number, concentrations, and development temperature.

Sample no.	Volume Percent S-C	S-B	S-A	MEK	Develop time (s)	Temperature (C)	γ
1	0	100	0	0	5	19	5.7
2	0	100	0	0	5	21	7.0
3	100	0	0	0	45	19	7.0
4	100	0	0	0	45	21	9.3
5	66.4	32.0	1.6	0.424	10		7.7
6	66.4	32.0	1.6	0.424	10		7.0
7	49.6	48.0	2.4	0.636	10		8.2
8	49.6	48.0	2.4	0.636	10		8.5
9	50.1	45.4	4.5	1.19	10	18	7.9
10	50.1	45.4	4.5	1.19	10	18	7.9
11	50.1	45.4	4.5	1.19	10	19	9.4
12	50.1	45.4	4.5	1.19	10	19	9.4
13	50.1	45.4	4.5	1.19	10	19	9.1
14	50.1	45.4	4.5	1.19	10	19	8.8
15	0	95	5	1.19	5	21	12.8
16	0	0	100	100	2	19	4.7

thicknesses were not compromised by excessive dissolution rates. This was necessary since the profilometer technique can only measure resulting resist thicknesses with respect to the unexposed portions. By soaking for several hours, S-B and S-C were found to yield unexposed dissolution rates of 2.8 and 0.5 A/min, respectively. The harshest developer, S-A, displayed the highest dissolution rate of 51 A/s. However, even this rate was still small enough to be negligible given our initial thickness and short develop times for mixtures containing even small amounts of MEK.

Data were taken for the three developer components alone as well as for a variety of mixtures of the three components. Contrast was determined from the steepest slope on the graphs of percent resist remaining versus the log of exposure dose. The steepest slopes were determined using a linear regression routine applied to the three steepest data points. Special care was taken to apply exactly the same criterion for the determination of steepest slope in all cases.

The mixtures of the developers were investigated to show the effects of varying primarily the S-A concentration. It was deemed that S-A is such a harsh developer that its overall concentration should be very low, and also that the properties of S-B and S-C are similar enough that their concentrations should also be similar. However, higher concentrations of S-C were included in the mixtures since it yields generally higher contrast than S-B. Since S-A was found to give low contrast by itself at an MEK concentration of only 26.5%, we investigated primarily the range of very small MEK concentrations. For these reasons, the three mixtures investigated were S-C:S-B:S-A in ratios of 42:20:1, 21:20:1, and 11:10:1. We consider the most important parameter to be the S-A concentration, which in the above mixtures is 1.59%, 2.38%, and 4.55%, respectively. This corresponds to an absolute concentration of MEK in the total mixture of only 0.42%, 0.64%, and 1.2%, respectively. Please note that mixture ratios are expressed in units of S-A volumes, and percentages are provided in Table I.

III. RESULTS AND DISCUSSION

The results of our contrast experiment are shown in Figs. 2 and 3. Figure 2 shows contrast curves for the three original developers, S-A, S-C, and S-B. For the data shown, the contrast for S-A is 4.7, for S-C 7.0, and for S-B 5.7. It is important to note, as will be demonstrated below, that small differences in contrast can be very important to lithographic results, given the logarithmic nature of the plot.

Figure 3 shows contrast curves for two mixtures of the three common solutions. It can be seen by comparing Figs. 2 and 3 that the sensitivity of the mixtures is approximately the same as that of S-C, falling midway between that of S-A and S-B. This is an extremely desirable trait of a developer, namely, that an increase in contrast does not lead to a decrease in sensitivity of the resist-developer system. This is more significant considering that the develop time for the mixtures is only 10 s compared with 45 s for S-C.

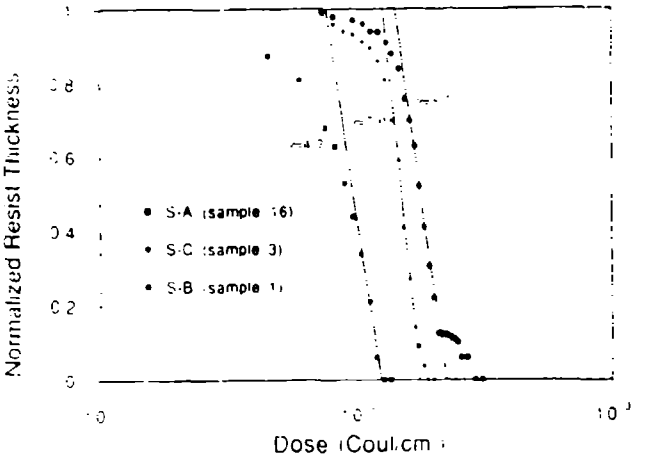


FIG. 2. Contrast curves for the three original developers.

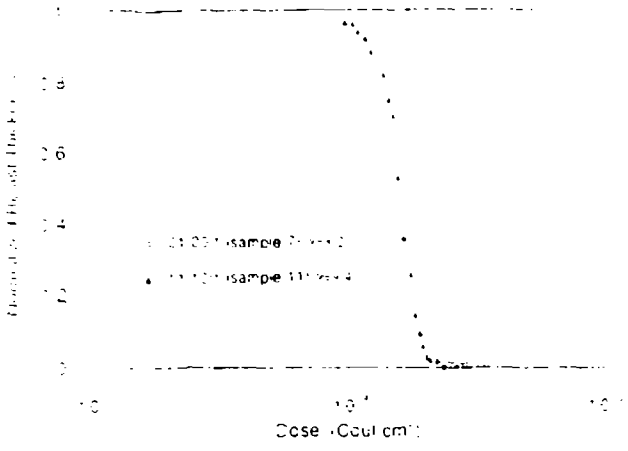


FIG. 3. Contrast curves for mixtures of the original developer components. Ratios are S-C:S-B:S-A.

This implies that the mixture must offer higher sensitivity than S-C or S-B alone if development times are increased. In addition, it is shown in Appendix A that contrast is predicted to increase for increased develop times but only for an increased solubility, as in our case of added MEK. This implies that further study would indicate a possibility of achieving even higher contrasts coupled with improved sensitivity for longer develop times. In practice, the optimum exposures for highest resolution correspond to the dose at D_f of Fig. 1, where D_f must decrease for increased develop time. This has strong implications for the viability of improving the throughput of processes utilizing PMMA for high-resolution commercial applications.

We have observed markedly improved lithographic performance using the new developer mixtures. Figure 4 shows photomicrographs of test patterns consisting of sets of parallel lines of decreasing pitch to 0.1 μm in the direction of the outside of the pattern, and overlapping to form a variable-spaced grid in the corner. Both test patterns were exposed on a single sample under the identical conditions of resist thickness, bake temperature, bake time, dose, line spacing, and beam energy. After exposure the patterns were separated and developed as discussed above using S-C for pattern 4(a) and S-C:S-B:S-A 11:10:1 for pattern 4(b).

The resulting improved contrast can be observed from the variations in light and dark areas. One expects that as the parallel lines are exposed closer together, the proximity effects cause some exposure of the spaces between the lines.[2] A lower-contrast developer will dissolve the PMMA at the lines of exposure and also remove much of the resist between the lines. This is the cause of the light areas in Fig. 4(a) as the line pitch decreases. However, a sufficiently higher-contrast developer will remove mostly the resist in areas of direct exposure and less of the resist between the lines. This is the case in Fig. 4(b) where there is clearly less light area, showing that the resist remains intact between the lines even in the presence of identical proximity effects to part 4(a). These results cannot be explained as a

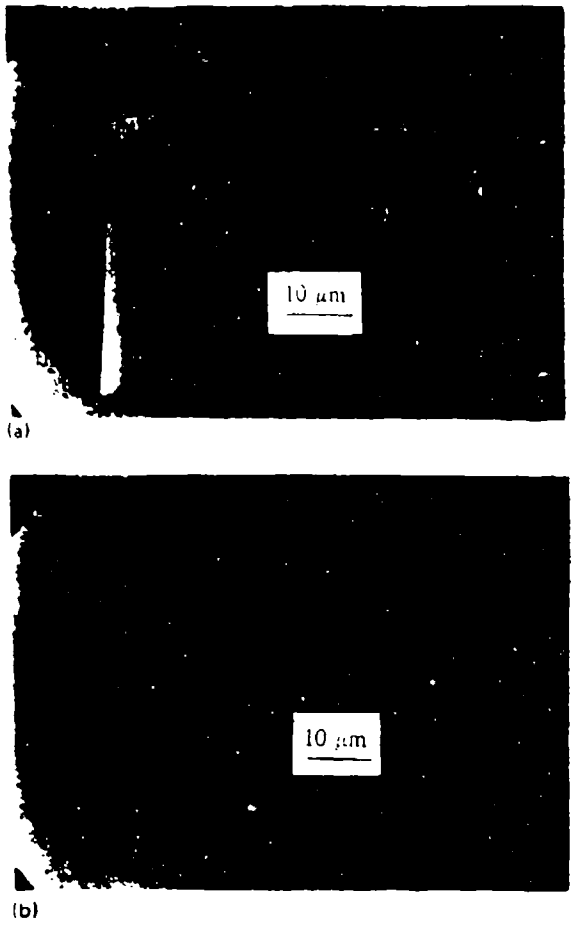
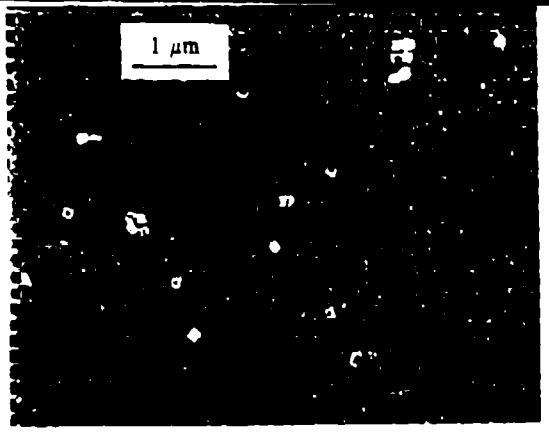


FIG. 4. Photomicrographs of test patterns illustrating improved contrast from new developers. The test patterns consist of parallel lines of decreasing pitch in the outward directions, crossing in a variable-pitch grid pattern at the corner. Light areas indicate severe loss of resist between the exposed lines. Both patterns were exposed under identical conditions, but developed by (a) S-C for 45 s and (b) S-C:S-B:S-A 11:10:1 for 10 s. The appearance of less light area in (b) implies that the mixture has provided much higher contrast than that of S-C in (a).

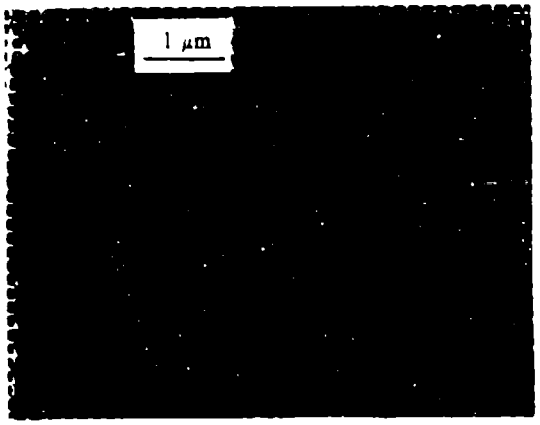
simple matter of sensitivity and overdeveloped exposure, since the sensitivity of the 11:10:1 mixture is almost exactly the same as that of S-C as can be seen by comparing Figs. 2 and 3.

Figure 5(a) shows a grid pattern developed by the S-B solution. The linewidths vary by about 25% toward the interior of the pattern. Figure 5(b) shows a pattern developed in S-C:S-B:S-A 11:10:1. There is almost no evidence of linewidth variations over the pattern due to the higher contrast of the 11:10:1 mixture versus that of the S-B. In addition, lift-off properties are affected by contrast. Figure 5(a) shows a typical example of metal flakes left from the unexposed areas after a lift-off step in acetone. Figure 5(b), however, shows the consistently improved lift-off achieved due to steeper sidewalls resulting from the higher contrast. The patterns of Figs. 4 and 5, then, help demonstrate that the new mixtures provide usably higher contrast compared with conventional developers.

Table I shows the contrast for various percentages of developer solutions and temperatures. A plot of contrast as



(a)



(b)

FIG. 5. Grid patterns formed by EBL exposure at 40 kV on GaAs substrates followed by development and lift-off of 40 nm gold. (a) developed in S-B and (b) developed in S-C:S-B:S-A 11:10:1. Note decreased line broadening in (b) vs (a) indicating higher contrast of the mixture over S-B. Also note typical improved lift-off as a result of steeper edge profiles due to higher contrast.

a function of percent S-A, uncorrected for temperature dependence, is shown in Fig. 6. A trend of increasing γ as S-A is increased is indicated. The same data, but normalized to 19 °C, are shown in Fig. 7. (Note that contrast data shows no correlation with the initial resist thickness, which is accounted for in the process of normalizing the remaining thickness after development.) The temperature normalization assumes an Arrhenius behavior with an activation energy of 24.3 kcal/mol, found by fitting contrast data at a fixed S-A percentage as a function of temperature. This value is intermediate between that found by Cooper, Krasicky, and Rodriguez [16] of 25 kcal/mol for MEK and that reported by Greeneich[17] of 24 kcal/mol for MIBK. In the case of Fig. 7, a trend of increasing contrast for increasing percent MEK is now clearly evident. The data, including S-C, collapse to a straight line of the least-squares fit very closely except for the case of S-B at 0% MEK. (Precise temperatures were not available for the case of 0.42% and 0.64% MEK. The bar in the figure is not an error bar, but represents calculated uncertainty of plus or minus 0.45 °C based on the known value of the activation energy centered about 19 °C.) It is striking that within this linear region,

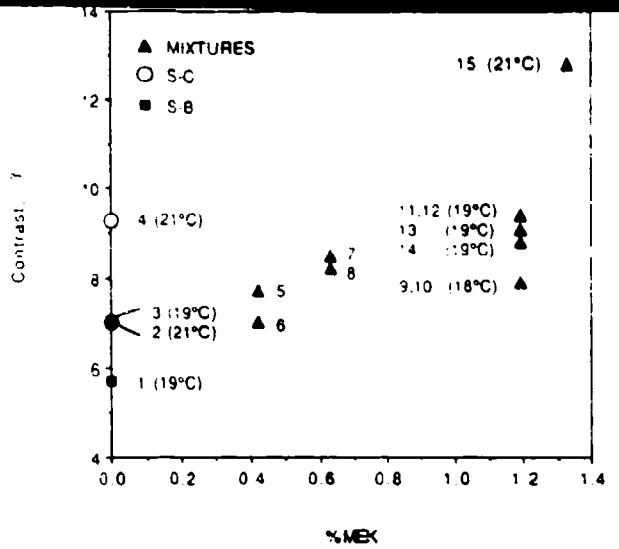


FIG. 6. Plot of γ as a function of percent MEK, uncorrected for temperature dependence. Sample number and temperature are noted for each datum.

about a 1.3% increase of MEK concentration in the mixture results in more than a 35% increase in contrast, and that further improvements can be promoted by simply increasing the temperature.

The fact that S-B does not fall on the temperature-compensated line can be qualitatively understood by noting

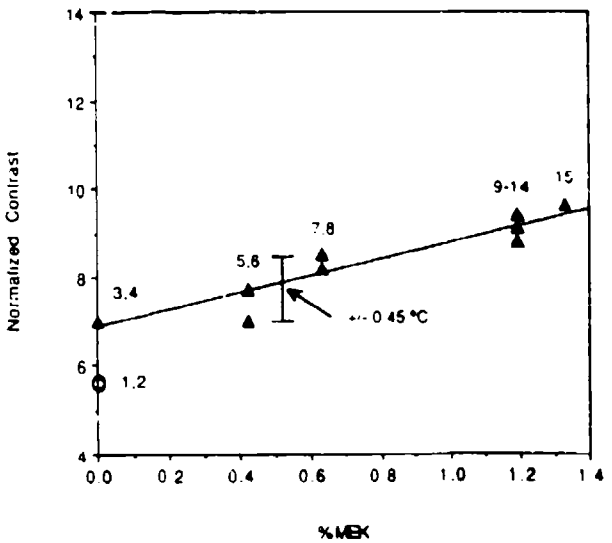


FIG. 7. Plot of γ as a function of percent MEK, corrected for temperature dependence by normalizing to 19 °C. The normalization assumes an Arrhenius dependence with an activation energy of 24.3 kcal/mol. Notice that the points in Fig. 6 corresponding to sample no. 4 (21 °C) for pure S-C and that of samples no. 9 and 10 (18 °C) at 1.19% MEK have collapsed onto the straight line and are now obscured on the graph by other data points. The bar in the figure is not an error bar, but represents calculated uncertainty of ±0.45 °C based on the known value of the activation energy centered about 19 °C.

that both MEK and MIBK are ketones, while CS (2-ethoxyethanol) contains both alcohol and ether groups. The chemical compositions of MIBK and MEK are very similar, differing only by two aliphatic carbons. In order to understand the dynamics completely, one would need to consider a full thermodynamic and transport analysis involving six components. A complete analysis is beyond the scope of this work, but a simple explanation of the role of the components is given below.

Previous analyses of the development of EBL-exposed resist show that measured decrements of normalized resist thickness upon exposure (i.e., the contrast curve) should be directly proportional to the rate of polymer dissolution (see also Appendix A). For a given developer system, this rate depends on the average molecular weight of the exposed polymer, which in turn is a function of the dose. Appendix A shows that contrast is expected to exhibit the same temperature behavior as that of dissolution rate and explains our observation of increased γ for developments performed at slightly higher temperatures where dissolution rates are in general higher. We found that at all S-A concentrations, γ increased with temperature between 18 and 21 °C. With the exception of only the S-B data, all other data properly collapsed onto the best-fit line as a function of temperature as suggested by the above theory.

The main factors that affect the rate of solubilization of a polymer in a solvent are the plasticization and thermodynamic compatibility, or solvent "quality." Plasticization refers to the ability of the solvent to penetrate the polymer and increase the free volume,[18,19] making it more vulnerable to dissolution. The rate at which a solvent can plasticize a polymer is related to its molecular size and therefore its molecular weight. Gipstein *et al.*[20] reported a noticeable solvent-size effect in the dissolution rate of atactic PMMA in a homologous series of *n*-alkyl acetates. They found a precipitous drop in dissolution rate for increased penetrant molecular weight. As an approximation we will use molecular weights as an estimator of the relative ability of the solvent components to plasticize the exposed PMMA.

The solvent quality is related to the match between three thermodynamic parameters of the solvent and the polymer. The closer these are, the higher is the thermodynamic compatibility resulting in a higher dissolution rate.[19] These parameters are the three components of the global solubility parameter δ_t, namely the dispersive, polar, and hydrogen-bonding components, δ_d, δ_p, and δ_h, respectively, where

$$\delta_t^2 = \delta_p^2 + \delta_h^2 + \delta_d^2. \tag{2}$$

We have found it useful to compare all of these components in an attempt to explain our results. In comparing the similarity of the three solubility parameters between the developers and PMMA, we have defined the parameter

$$\Delta = [(\delta_{pi} - \delta_{pj})^2 + (\delta_{hi} - \delta_{hj})^2 + (\delta_{di} - \delta_{dj})^2]^{1/2}, \tag{3}$$

where i indicates a developer component and j the resist. Since δ_t is the length of the solubility vector in solubility parameter space, Δ is the distance between the end points of these vectors. We therefore refer to Δ as simply "distance."

TABLE II. Molecular weight and solubility parameters for each component of the mixtures. The distance Δ is included for each component.

Component	M (g/mol)	δ_t [(MPa)$^{1/2}$]	δ_p [(MPa)$^{1/2}$]	δ_h [(Mpa)$^{1/2}$]	Δ [(MPa)$^{1/2}$]
MEK	72	14.0	9.3	9.5	1.4
EtOH	46	12.6	11.2	20.0	12
MIBK	100	14.4	8.1	5.9	3.4
IPA	60	14.0	9.8	16.0	7.5
CS	90	13.0	9.1	15.2	6.8
MeOH	32	11.6	13.0	24.0	16
MMA	90	13.5	10.1	8.5	...

It is expected that the greater the distance Δ is, then the less efficient the component is at dissolving PMMA, and therefore the lower is the solvent quality. Such measure of compatibility is consistent with the idea that polar, dispersive, and hydrogen-bonding forces can, in principle, act independently. (For example, two chemical species, one strongly polar but nonhydrogen bonding, and the other weakly polar but strongly hydrogen bonding, could have the same global solubility parameter and yet be immiscible. This is the same as saying that the length of δ_t in solubility space is the same for each species, but their directions vary greatly.) Thus only those solubility parameter components pertaining to congruent energetic interactions should be compared.

The calculated values of Δ for all pure solvents compared with MMA, along with the molecular weights of each component, are given in Table II. Again, the plasticization, or molecular weight, of the individual components must also be considered in explaining their role in the solutions. Table III indicates the order of each developer component in its ranking of plasticization and solvent quality. We can see that MEK is indeed the best solvent but rates fourth as a plasticizer. In S-A, the EtOH serves as a diluent and to help in plasticization, being ranked second in this regard. These two facts help to explain why S-A is such an effective solvent of PMMA at 3060 A/min for a molecular weight as high as 950 000 a.u. This also helps to explain why S-A exhibits the worst contrast, since it strongly dissolves PMMA of all molecular weights. Although CS ranks third in solvent quality, it appears that the presence of MeOH in S-B, ranked first for plasticization, helps to increase the effectiveness of CS as a solvent. Regarding S-C dissolution, the component MIBK is rated

TABLE III. Ranking of plasticization and solvent quality for each component of the developer solutions.

	MEK	EtOH	MIBK	IPA	CS	MeOH
Distance Δ [(MPa)$^{1/2}$]	1.4	12	3.4	7.5	6.8	16
M (g/mol)	72	46	100	60	90	32
Solvent quality (rank)	1	5	2	4	3	6
Plasticization (rank)	4	2	6	3	5	1

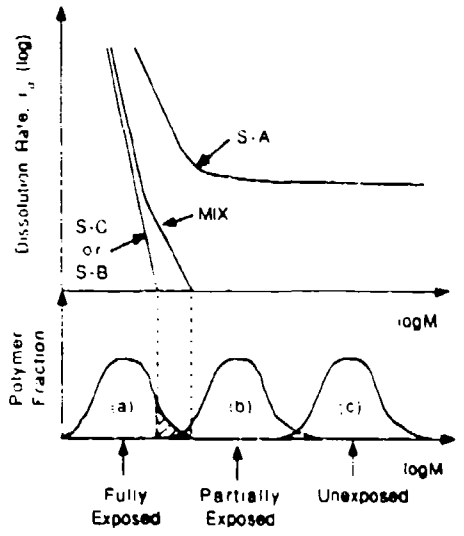


FIG. 8. Explanation of mechanism by which trace amounts of S-A in S-C increases contrast. The solubility curve for S-C alone is compared to the expected curve after the addition of a small amount of S-A (MIX) and S-A alone (see also Ref. 17). Contrast is improved through the increase of r_d for higher molecular weights with only slight change to r_d at low ones. Below the graph are molecular weight distributions for (a) fully exposed, (b) partially exposed, and (c) unexposed PMMA. The shaded region corresponds to those molecular weights that require large exposure doses to be further fragmented. It is preferable to use the lowest exposure dose possible to achieve the highest resolution. The addition of traces of S-A increases the solubility of these fragments.

second as a solvent but is the poorest plasticizer. Isopropanol being ranked only third as a plasticizer helps to explain why S-C dissolves the unexposed PMMA at only 0.5 Å/min.

As mentioned above, prediction of the solubility properties of systems containing many solvents can be very complex, and only a simple qualitative explanation of those properties can be offered here. Within the exposed resist, a distribution of molecular weights exists whose average depends on the exposure dose.[12] Under conditions of low-level exposure, the inability of a relatively high-contrast developer, such as S-C, to completely remove the PMMA in the primary exposed area is due to the presence of high-molecular-weight polymers that have not been exposed sufficiently to be broken into sufficiently small strands. The presence of a small percentage of MEK increases the overall thermodynamic quality of the solvent, causing those remaining long-chain PMMA strands to be more selectively removed without a major change in the solubility rate of the unexposed or less exposed areas. This argument is supported in Appendix B.

The above argument is plausible when discussed in light of Figure 5 of Ref. 17, which gives solubility rates of MIBK solutions as a function of fragment molecular weight M. We discuss here a two-component system that can be extended to more components. Figure 8 shows a qualitative curve of dissolution rate r_d versus fragment molecular weight M after exposure for a single developer, e.g., S-C, and also a proposed curve for the two-developer mixture S-C + S-A (MIX). Since the dissolution rate for pure MEK is much higher than for either S-C or dilute MEK in solution, and given the lower value of γ for S-A, r_d for S-A is expected to extend to higher molecular weight with a lower slope and approach a constant for very high molecular weights (similar to MIBK in Ref. 17). Portrayed below in the figure are qualitative molecular weight distributions of a fully, partially, and unexposed polymer. By adding small amounts of MEK to S-C, we have affected the solubility properties of the S-C as shown in the MIX curve to dissolve the higher-molecular-weight fragments far more quickly while changing the slope of the curve only slightly for the lower molecular weight. To first order, we expect the solubility properties to be affected little for lower molecular weights, but increased substantially for higher. (A thermodynamic model that qualitatively supports this assertion is presented in Appendix B where we analyze a two-component developer system by assuming that the rate enhancement derives exclusively from an increase in polymer solubility.) Highest benefit to contrast and resolution is achieved when the cutoff of solubility for the new mixture is above the highest molecular weight of the fragments in the primary exposed area but is below that of the marginally exposed resist for a slightly lower effective dose in a closely adjoining area. This can never be achieved totally in practice, as the distribution of fragment molecular weights in the adjoining areas is a continuum and overlaps that of the polymer in the primary exposed area.

In our case of mixing the three components S-C, S-B, and S-A, Fig. 7 indicates that the contrast improvement is dominated by the presence of MEK. We believe, however, that a more complete study would indicate that each component enhances the solubility in a range of molecular weights yielding an effective characteristic in which all fragmented molecular weights in the primary exposed areas are removed, but little is removed outside. It must be kept in mind that the average molecular weights of the PMMA fragments increase quickly with distance away from the beam. The goal in choosing a correct exposure for achieving high resolution is that the top of the range of fragment molecular weights in the primary exposed area lies just in the range dissolved at an appreciable rate by the trace amounts of MEK in the mixture, so that high-molecular-weight fragments are selectively removed from the area of primary exposure. It is possible to conceptualize an ideal developer whose properties are described in Fig. 9. This developer would have "break points" in the r_d vs M curve that tailor the dissolution rate to the molecular weight distribution curve. This ideal developer would then be assured of dissolving no more of the higher-molecular-weight fragments than was absolutely necessary for a given exposure dose, thereby minimizing the erosion of the lower-molecular-weight fragments (relative to the unexposed areas) in the less-exposed adjoining area. (These arguments are further supported by the theoretical predictions for a binary developer as discussed in Appendix B, and in Fig. 10.) Our three component system is a crude, but demonstrably effective, attempt at achieving this goal.

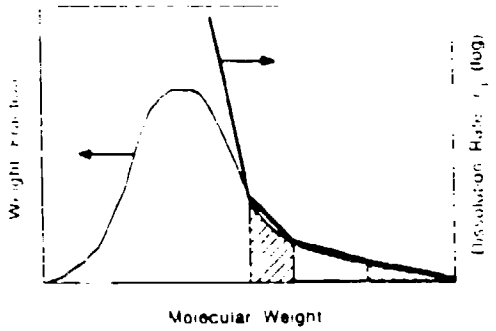


FIG. 9. Solubility curve for a conceptualized "ideal" developer superimposed on the distribution of fragment molecular weights for PMMA. The shape of the solubility curve ensures that dissolution rates for each molecular weight are no higher than necessary to dissolve all molecular weight fractions completely, thereby minimizing erosion of areas adjoining those of primary exposure.

IV. SUMMARY AND CONCLUSIONS

We have investigated new developer mixtures for PMMA positive resist. We demonstrated that adding a small percentage of methyl ethyl ketone to methyl isobutyl ketone/Cellosolve solutions increased contrast substantially. This was illustrated in practice by a comparison of EBL exposures each demonstrating decreased proximity effect.

The mechanism for increased contrast by adding trace amounts of MEK was explained in terms of selective broadening of the range of solubilities for S-A in S-C/S-B solutions. It was concluded that the presence of MEK increased the range of dissolution of fragmented molecular weights in the primary exposed regions to more efficiently clear the exposed PMMA without affecting unexposed areas.

We feel that in view of the substantial improvements of the contrast of PMMA and demonstrated improvements in minimizing the proximity effect, the use of PMMA can be further justified for ultrasmall geometries in future specialized integrated circuits and optical and x-ray mask fabrication applications. In addition, as future x-ray sources increase in brightness, the feasibility of using PMMA will become more favorable with the trade-off in support of resolution versus speed becoming more advantageous.

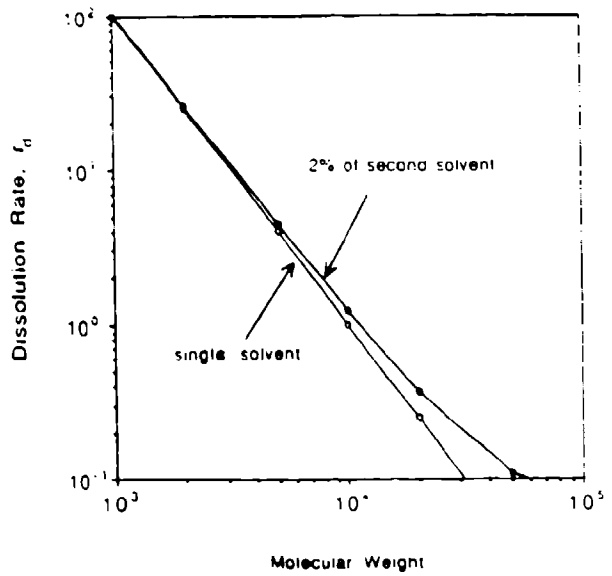


FIG. 10. Calculated rate of dissolution for a single solvent and also that solvent combined with 2% of a second solvent as a function of molecular weight. The mixed curve has been calculated from Eqs. (B7) and (B8) assuming $\beta(x_1/x_2) = M/100$.

ACKNOWLEDGMENTS

The authors would like to thank J. P. Kohn and D. K. Ferry for helpful discussions. Also, we are grateful to KTI Chemicals for providing materials. This research was supported in part by IBM, AFOSR, ONR, and the University of Notre Dame.

APPENDIX A

Here, we show that the contrast γ is proportional to the rate of polymer dissolution. We consider the dissolution of a polymer film of uniform initial thickness h_0. The polymer has an average molecular weight M. Exposure to a developer solution for a time Δt_d will cause partial erosion of the film, and the removal of a layer of thickness Δh. In general, we may write for $\Delta h \ll h_0$,

$$\Delta h/h_0 = r_d \Delta t_d / h_0, \tag{A1}$$

where r_d is the average rate of dissolution, which depends on M, temperature T, and solvent quality. Experiments[16,17] indicate that

$$r_d = f(M, T, \text{solvent quality}) \exp(-E_d/RT) M^{-\alpha}, \tag{A2}$$

where E_d is the activation energy for dissolution, R the ideal-gas constant, and f a function of solvent quality, polymer molecular weight, and T. The exponent α, a positive real number, may also be a function of M, T, and the solvent. Within certain ranges of M, however, such dependencies of f and α are weak, so that the main effects of T and M are accounted for predominantly by the Arrhenius and power law, respectively.

For a polymer exposed to radiation, the average molecular weight depends on the exposure dose D, i.e. $M = M(D)$. Thus the after-development normalized thickness of an exposed resist will depend on the average polymer molecular weight M, which, in turn, is a function of exposure dose D. The contrast γ, defined as

$$|\gamma| = \max\left(\left| \frac{\partial(\Delta h/h_0)}{\partial \log D} \right| \right), \tag{A3}$$

where the term on the right-hand side can be expressed as

$$\frac{\partial(\Delta h/h_0)}{\partial \log D} = \frac{\partial(r_d \Delta t_d/h_0)}{\partial \log D} = r_d \left(\frac{\partial \log r_d}{\partial \log M} \right) \left(\frac{d \log M}{d \log D} \right) \frac{\Delta t_d}{h_0} \tag{A4}$$

Equation (A4) shows explicitly a proportionality between γ and r_d. Furthermore, for exclusively Arrhenius temperature dependence and a power-law molecular weight dependence of r_d, as in Eq. (A2), we will have

$$\left(\frac{\partial \log r_d}{\partial \log M}\right) = -\alpha$$

and

$$\left(\frac{d \log M}{d \log D}\right) = \text{function of } D \text{ only,} \tag{A5}$$

which are both independent of temperature. In this case, the contrast will show an Arrhenius temperature dependence, with the same activation energy as the dissolution rate. This supports our observations of increased γ for higher developer temperatures.

Interestingly, from Eq. (A4) we see that for a fixed rate, temperature, and $M(D)$ functionality, a lower value of α will always give lower contrast. Greeneich[17] has shown experimentally for dissolution of PMMA in MIBK:IPA that the slopes of $\log(r_d)$ vs $\log(M)$ decrease appreciably with increasing solvent quality (i.e., for larger MIBK:IPA ratios); indeed, for pure MIBK, after a relatively sharp transition in slope at M around 10^4 a.u., the curves eventually approach a constant (i.e., $\alpha \approx 0$). In this case, according to Eqs. (A4) and (A5) the contrast should either approach zero, or, more realistically, become very small. This is consistent with our finding that films developed in pure S-A, an extremely strong (good) solvent, always showed poor contrast, in fact, poorer than those developed in either S-C or S-B.

APPENDIX B

We propose here a simple explanation of our findings, based on the hypothesis that our observed enhancement of contrast with traces of MEK is due to purely thermodynamic factors. This hypothesis is reasonable, since MEK is such a strong solvent. We assume the ratio of the dissolution rate (and thus contrast) of the system with MEK to that without it to be proportional to the corresponding ratio of polymer volume fractions. This functionality is consistent with the idea that any multiplicative factor quantifying rate enhancement should increase with polymer solubility, and approach unity at zero MEK concentration.

Our task, then, is to determine the increase in polymer volume fraction in solution caused by the presence of traces of a third component in an otherwise binary (polymer + solvent) and biphasic (solid polymer + solution) system. This can be done within the framework of solution thermodynamics.

We consider two equilibrium systems. The first one, a two-phase two-component system, comprises a pure, solid polymer (phase S) in equilibrium with a solution of polymer (component 1) in a solvent (component 2, phase L). The second system (denoted here by an asterisk) is identical to the first one, except for the presence of traces of a second solvent (component 3) in the liquid. We assume that the polymer is monodisperse, and that the solid precipitate is always made exclusively of polymer, with no appreciable amounts of either solvents.

Thermodynamic equilibrium requires that the chemical potential μ of the polymer in the solution and in the solid must be equal, i.e.,

$$\mu_1^S = \mu_1^L \quad \text{and} \quad \mu_1^{S^*} = \mu_1^{L^*}, \tag{B1}$$

where, again, component 1 is the polymer, and S and L stand for solid and liquid, respectively.

In both cases, the solid is a pure polymer, i.e.,

$$\mu_1^S = \mu_1^{S^*}, \tag{B2}$$

thus

$$\mu_1^L = \mu_1^{L^*}. \tag{B3}$$

Expressions for the chemical potentials of the various components in a ternary mixture of a polymer with two solvents have been given by Flory;[21] for the polymer we have

$$\begin{aligned}\mu_1 - \mu_1^0 = RT[&\ln \phi_1 + (1 - \phi_1) - \phi_2(x_1/x_2) - \phi_3(x_1/x_3) \\ &- (\chi_{12}\phi_2 + \chi_{13}\phi_3)(\phi_2 - \phi_3) \\ &- \chi_{23}(x_1/x_3)\phi_2\phi_3)],\end{aligned} \tag{B4}$$

where the superscript 0 refers to the pure, amorphous, polymer at the same temperature T and pressure of the solution, R is the ideal-gas constant, and x_1, x_2, and x_3 are quantities proportional to the molecular volumes of the polymer 1, and solvents 2 and 3, respectively. Additionally, ϕ_i is the volume fraction of species i ($i = 1,2,3$), and the interaction parameters χ_{ij} are given by

$$\chi_{ij} = z\Delta w_{ij}x_i/RT, \tag{B5}$$

where z is a lattice coordination number and Δw_{ij} the molar interaction energy between species i and j.

Note that Eqs. (B4) and (B5) have been proven to be of only limited accuracy, even for simple binary systems, and especially for dilute polymer solutions.[22] The predicted trends are generally correct, however, and consistent with most experimental evidence. Thus, this model should be able to capture, at least qualitatively, the main physics of our problem.

Equations (B4) and (B5) can be used to calculate the chemical potential of the polymer both in the presence and absence of component 3. Because component 3 is very dilute, the following approximations can be made:

$$|\phi_1^{L^*} - \phi_1^L| \ll 1, \quad \phi_2^L \cong \phi_2^{L^*}, \quad \phi_3^L = 0, \quad \phi_3^{L^*} \ll 1. \tag{B6}$$

From Eqs. (B3)–(B5) we obtain

$$\begin{aligned}\ln(\phi_1^{L^*}/\phi_1^L) = \phi_3^{L^*}(x_1/x_3)[1 + \phi_2^{L^*}(zx_3/RT) \\ \times(\Delta w_{23} - \Delta w_{12} - \Delta w_{13})].\end{aligned} \tag{B7}$$

Thus, an exponential enhancement in polymer solubility is predicted to occur upon addition of traces of component 3, provided that:

$$\beta = [1 + \phi_2^{L^*}(zx_2/RT)(\Delta w_{23} - \Delta w_{12} - \Delta w_{13})] > 0. \quad \text{(B8)}$$

For $\beta \sim 1$, the increase can be substantial, since (x_1/x_3) is proportional to the ratio of the molecular weight of the polymer to that of component 3. Notice that an exponential increase of the ratio $(\phi_1^{L^*}/\phi_1^L)$ with polymer molecular weight does not necessarily imply an unbounded increase of polymer volume fraction, since for higher, insoluble, molecular weights, the decay to zero of ϕ_1^L will be faster than any concomitant growth of the enhancement factor. Flory, in his analysis of polymer fractionation by selective precipitation from binary mixtures, derived an equation [see Eq. (23) of Ref. 21] that is very similar to Eq. (B7).

In order to obtain the sign of β in Eq. (B8) it is necessary to estimate the differences of the interaction energies between the various components. Following Gordon and Taylor[23] (see also Ref. 18) in a semiempirical approach, we assume the following proportionality [see Eq. (3)]:

$$\Delta w_{ij} \propto \Delta_{ij}^2 = [(\delta_{pi} - \delta_{pj})^2 + (\delta_{di} - \delta_{dj})^2 + (\delta_{hi} - \delta_{hj})^2], \quad \text{(B9)}$$

where i and j denote the particular substances, and δ_p, δ_d, and δ_h are the polar, dispersive, and hydrogen-bonding components of the solubility parameter, respectively. In our case

$$(\Delta w_{23} - \Delta w_{12} - \Delta w_{13}) \propto (\Delta_{23}^2 - \Delta_{12}^2 - \Delta_{13}^2). \quad \text{(B10)}$$

Hence, the sign of the interaction energy term on the left-hand side can be estimated from the properties of the pure components. (Further theoretical refinements would require that the factor of proportionality between Δw_{ij} and Δ_{ij}^2 be a function of the molar volumes of the various components. Such refinements are not deemed necessary here, where only rough estimates are of interest.) For a positive right-hand side in Eq. (B10), β in Eq. (B8) will also be positive, indicating an enhancement of polymer solubility. On the other hand, a negative term in Eq. (B10) may either signify an increase or a decrease of solubility, depending on the relative magnitude of the various terms in Eq. (B8).

It is interesting to estimate the sign and magnitude of the quantity in Eq. (B10) for PMMA in MIBK with traces of MEK. This system should be reasonably close to the one we achieved experimentally using the pure-component solubility parameters in Table II (MMA = 1, MIBK = 2, MEK = 3), we obtain, in units of cal/ml,

$$\Delta_{23}^2 = 14.49, \quad \Delta_{12}^2 = 11.57, \quad \Delta_{13}^2 = 2,$$

giving

$$\Delta_{23}^2 - \Delta_{12}^2 - \Delta_{13}^2 = 0.92 \text{ cal/ml}.$$

We see that the overall interaction energy is positive, so that we would indeed expect an enhancement of solubility in our case, consistent with observations.

Furthermore, it is possible to extract from our data with Eq. (B7) an "experimental" value of $\beta(x_1/x_3)$. The linear increase of contrast with percent MEK in Fig. 7 is, in fact, consistent with Eq. (B7), since for small values of the argument, the exponential can be approximated as a straight line. From the temperature-normalized contrast data we obtain $\beta(x_1/x_3) = 24$, which is reasonable, in light of our many approximations, and, possibly, some more fundamental limitations inherent to the thermodynamic model itself. Again, it is worth noticing here that, given the crudeness of our model, such "theoretical" results should be considered only for their qualitative value.

A rate-enhancement curve, calculated from Eq. (B7) by assuming $\phi_3 = 0.02$, $\beta(x_1/x_3) = (M)/100$, and an inverse-square dependence of the dissolution rate on M, is given in Fig. 10 (Note that we have arbitrarily truncated the enhanced-rate curve, since it would have shown an unrealistic upturn for higher molecular weights, due to our assumption of a fixed inverse-square dependence of the original rate for all molecular weights. In reality, due to the finite solubility of higher fractions, the dissolution rate eventually drops to zero much faster than M^{-2}, and no upturn in the net rate occurs). Figure 10 should be compared with curves 1 and 3 in Fig. 5 of Ref. 17.

[1] M. A. McCord and R. F. W. Pease, J. Vac. Sci. Technol. B. 6, 293 (1988).
[2] D. C. Joy, Microelectron. Eng. 1, 103 (1983).
[3] W. J. Cooper, P. D. Krasicky, and F. Rodriguez, Polymer 26, 1069 (1985).
[4] W. M. Moreau, Opt. Eng. 22, 181 (1983).
[5] S. Mackie and S. P. Beaumont, Solid State Technol. 28, 117 (1985).
[6] B. K. Daniels, P. Trefonas, and J. C. Woodbrey, Solid State Technol. 31, 105 (1988).
[7] M. A. Mohsin and J. M. C. Cowie, Polymer 29, 2130 (1988).
[8] K. G. Chiong, M. B. Rothwell, S. Wind, J. Bucchignano, and F. J. Hohn, J. Vac. Sci. Technol. B 7, 1771 (1989).
[9] V. K. Sharma, R. Pethrick, and S. Affrossman, Polymer 23, 1732 (1982).
[10] J. S. Papanu, D. W. Hess, D. S. Soane, and A. T. Bell, J. Electrochem. Soc. 136, 3077 (1989).
[11] A. N. Broers, J. M. E. Harper, and W. W. Molzen, Appl. Phys. Lett. 33, 392 (1978).
[12] S. P. Beaumont, P. G. Bower, T. Tamamura, and C. D. W. Wilkinson, Appl. Phys. Lett. 38, 436 (1981).
[13] H. G. Craighead, R. E. Howard, L. D. Jackel, and P. M. Mankiewich, Appl. Phys. Lett. 42, 38 (1983).
[14] A. Muray and M. Isaacson, J. Vac. Sci. Technol. B 1, 1091 (1983).
[15] All exposures were performed at the Center for Solid State Electronics Research at Arizona State University.
[16] W. J. Cooper, P. D. Krasicky, and F. Rodriguez, J. Appl. Polym. Sci. 31, 65 (1986).
[17] J. S. Greeneich, J. Electrochem. Soc. Solid State Sci. Technol. 122, 970 (1975).
[18] C. A. Kumins and T. K. Kwei, in *Diffusion in Polymers*, edited by J. Crank and G. S. Park (Academic, New York, 1968), p. 107.
[19] F. Rodriguez, *Principles of Polymer Systems*, 3rd ed. (Hemisphere, New York, 1989), p. 36.
[20] E. Gipstein, A. C. Ouano, D. E. Johnson, and O. U. Need III, Polymer Eng. Sci. 17, 396 (1977).
[21] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, NY, 1953), p. 549.
[22] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, NY, 1953), p. 559.
[23] J. Gordon and J. S. Taylor, J. Appl. Chem. 2, 493 (1952).

SCANNING Vol. 14, 345–349 (1992)
©FAMS, Inc.

Received September 16, 1992

Determination of Fixed Electron-Beam-Induced Positive Oxide Charge

G.H. BERNSTEIN, S.W. POLCHLOPEK, R. KAMATH, W. POROD

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana, USA

Summary: Contrary to previous beliefs, electron-beam-induced positive charges in insulators persist where created for at least several weeks without discernible movement. Formerly, coating with a thin metal overlay was thought to allow the charge to leak away. Coating with a conductor is shown to shield electric fields from affecting the imaging probe, but to remove no charge from the specimen. A new technique is introduced for the evaluation of the properties of electron-beam-induced positive charges in metal-oxide-semiconductor (MOS) capacitors. MOS structures were subjected to partial area exposure in a scanning electron microscope. These exposures resulted in the creation of areas of localized positive charge within the oxide, which was observed as steps in the capacitance-voltage data. A systematic study was performed. It related the exposed area to the step height and the amount of induced charge to the voltage shift of the step. A model describing the observed phenomenon is presented, followed by a comparison of theoretical and experimental results. The progress of the charge over time was studied by performing capacitance-voltage analysis 30 min after electron beam exposure and up to 4 weeks later.

Introduction

Charging of specimens is a common problem encountered in the imaging of all insulating materials in a scanning electron microscope (SEM). In general, the details of the charging mechanism can be a complex function of beam and material parameters. Krause and co-workers (Krause *et al.* 1989) have demonstrated that in thin films, trapped charge can be either positive or negative, depending on the beam-accelerating voltage, the thickness of the insulating layer, and the nature of the substrate. Typically, at higher voltages, SEM images of thin insulators on conducting substrates appear dark because of trapped positive charges left behind by secondary electrons ejected by the primary beam. It is commonly believed that the coating of insulating specimens by a conductive layer causes this positive charge to be drained away from the imaging area, thus eliminating distortions of the probe and restoring usable resolution (Goldstein *et al.* 1981). A nuisance in SEM imaging, excess positive charge in the gate oxide of metal-oxide-semiconductor (MOS) devices (McLean *et al.* 1989, Nicollian and Brews 1982) can be created by electron beam processing used in certain integrated circuit fabrication and analysis techniques. With electron beam lithography, metal over- or under-layers are used to reduce the charging effects during beam writing of nonconductive resists (Henderson 1980). The precise properties and behavior of the induced charges in insulating films are uncertain. Although in MOS structures much is known about the creation, minimization, and annealing of e-beam-induced bulk oxide damage (Balasinski *et al.* 1988, Keery *et al.* 1976, Sah *et al.* 1983, Shimaya *et al.* 1983), details of the behavior of localized positive charge within oxide layers are not fully understood. This is partly because an easy and accurate method to monitor localized oxide charge has not been established.

For the first time we can report that electron-beam-induced positive charges remain fixed where created. To demonstrate this, we have developed a technique that uses a certain structure in the capacitance-voltage (C-V) curves of MOS capacitors. This structure can be used to yield valuable information regarding the location and amount of charge deposited in SiO_2 in a MOS structure. Using electron beam lithography, we accurately exposed partial areas of MOS capacitors with known electron doses. This served to effectively create two capacitors in parallel, with one metal gate, having different threshold values due to the induced positive charge in the oxide. By performing C-V measurements before exposure, 30 min afterward, and later, it was possible to electrically monitor the induced positive charge. The results of these experiments, along with conclusions about the electrical behavior of the charge within the oxide, are reported. A model relating the exposed capacitor area to the theoretical C-V results is proposed and compared with experimental results.

This work was supported in part by ONR, AFOSR, IBM, and the University of Notre Dame.

Address for reprints:

G.H. Bernstein
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556, USA

Experiment

The starting material used for MOS capacitor fabrication was <100> orientation, p-type silicon with doping concentration $N_a = 1.5\times10^{15}$ cm^{-3}. After careful cleaning, the silicon wafers were oxidized at 1000°C for 20 min, resulting in oxide thicknesses ranging from 738–833 Å, as deduced from the accumulation capacitance of the individual C-V curves. In order to minimize the effects of mobile ion contamination, oxygen was bubbled through a solution of 2% HCl in deionized water. Circular aluminum dots of an area 4.9×10^{-3} cm^2 were evaporated in a high vacuum through a shadow mask onto the surface of the oxidized wafers. The thickness of the Al for all of the dots was 1,100 Å as measured with an Alpha-step 200 profilometer. The top surfaces of the metallized wafers were coated with a protective layer of positive photoresist, then immersed in a buffered hydrofluoric acid (HF) chemical etching solution which removed the oxide from the back surface. Following photoresist removal, aluminum was evaporated onto the back surface of the wafers, after which the wafers were annealed in flowing nitrogen gas for 20 min at 420°C.

Partial areas of the completed MOS capacitors were subjected to electron beam exposure as shown schematically in Figure 1a, changing the device properties where exposed and thus effectively dividing the device into two separate capacitors in parallel, as described in Figure 1b. Exposures were performed at 20 kV and 0.8 nanoamperes in an Amray model 1400 scanning electron microscope. The SEM was controlled using a Pragmatic Instruments model 2201A arbitrary waveform generator and custom software to ensure dose and area accuracy. Exposures were executed by quickly imaging a single capacitor at a very low current, blanking the beam, setting the current to the desired value, and scanning (using e-beam lithography) a square of known area entirely within the MOS capacitor. The dose was fixed at 4×10^{-5} C/cm^2, and exposed areas ranged from 10.4–62.4% of the total area of the capacitor. The upper limit on areal exposures was determined by the size of the largest possible inscribed square within a circle (with area ratio $2/\pi$).

Results and Analysis

Total electron beam exposure of high quality MOS devices is known to cause a negative shift of C-V characteristics (Johnson 1975, Keery *et al.* 1976, McLean *et al.* 1989, Nicollian and Brews 1982, Sah *et al.* 1983, Shimaya *et al.* 1983). Free electrons created in the oxide during exposure either recombine with holes or leave the insulator; thus they are electrically eliminated (Keery *et al.* 1976, Zaininger and Holmes-Siedle 1967). Holes, however, behave differently because of their much lower mobility in SiO_2, as some will recombine with electrons, while a significant fraction is captured in stationary traps, resulting

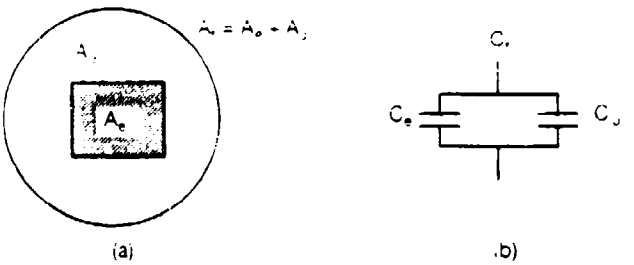


FIG. 1. Schematic diagram of the partial area exposure of a MOS capacitor. (a) Definition of the exposed area, A_e, and the unexposed area A_u. (b) The exposed and unexposed portions give a total capacitance, C_t, which is the sum of the two parallel capacitances, C_e and C_u.

in a fixed positive charge contained in the oxide (Keery *et al.* 1976, Zaininger 1966, Zaininger and Holmes-Siedle 1967). The result of this trapped positive charge is depletion and inversion of the device at more negative voltages.

C-V measurements were performed on a series of MOS capacitors both before and 30 min after partial area exposure in an SEM, as described above. Figure 2 shows typical results obtained from two of these capacitors that have fractional exposure areas, A_e/A_t, where A_e is the exposed area and A_t is the total area, of 10.4 % and 62.4%, respectively. The relevant features in this figure are the additional steps in the C-V curves observed after exposure. Comparison of Figures 2a and 2b shows that the step height increases for larger fractional area exposure but the location of the step is approximately the same for constant areal dose.

We demonstrate below that our observed steps are due to trapped positive charge in the exposed area, which causes that area to deplete and invert at more negative voltages than the unexposed portion of the capacitors. Although steps have been previously reported in C-V data after electron beam irradiation, they were of different origin. In a study of Zaininger (1967), steps were observed after total capacitor exposure and were attributed to the enhancement of existing interface states concentrated around a single energy level in the band gap. When the degree of perfection of the interface was high, no step was observed, but the entire C-V curve was shifted to more negative voltages. In order to test whether our steps were due to such a surface state phenomenon, we performed whole area exposures of our capacitors. The results of our test are shown in Figure 3. The shift of the C-V characteristics with no step present indicates that our steps in the partial exposure experiments are not related to the enhancement of preexisting interface states but are a result of the localized positive charge created in the exposed area.

We now develop a quantitative model, which relates the size of the observed steps to the areas of exposures, assuming that the created positive charge is fixed where created. A single capacitor subjected to partial area exposure can be modeled as two capacitors connected in parallel: one capacitor, C_e, with an area equal to that of the exposed area, and another, C_u, with an area equal to that of the un-

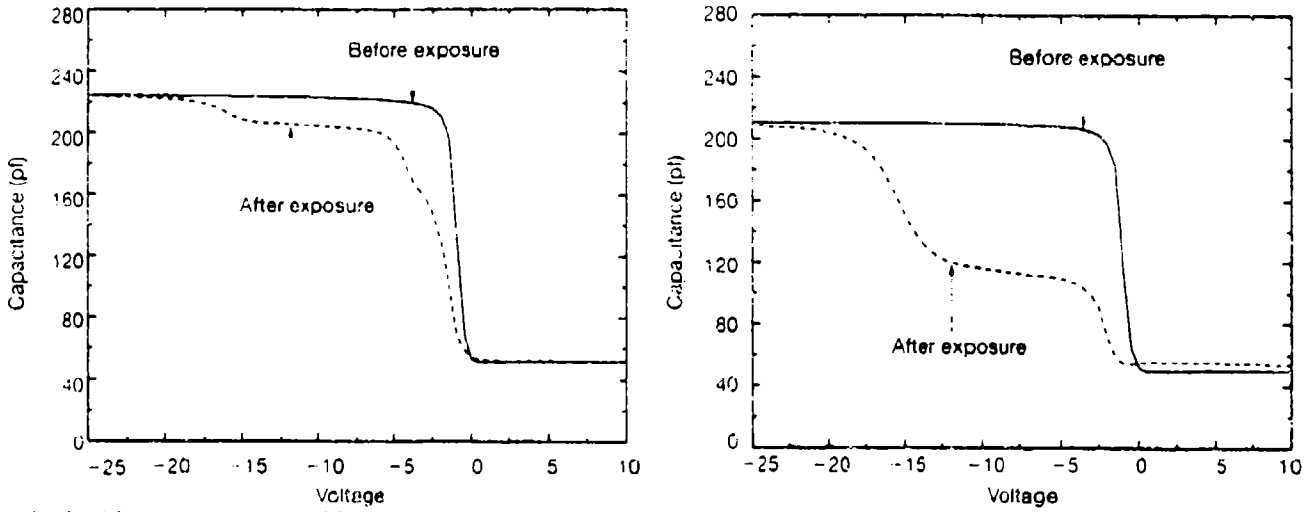


FIG. 2 C-V characteristics of MOS capacitors before and after e-beam partial area exposure. The exposure areas were (a) 10.4% and (b) 62.4% of the total area.

exposed area. Since the exposed area has a more negative threshold voltage, V_{te}, because of the presence of the positive charge in the oxide, it undergoes depletion and inversion, while the unexposed area is still in accumulation. The unexposed area inverts at the unchanged threshold voltage, V_{tu}. Figure 4 shows schematically a C-V curve, which is a superposition of the individual C-V curves for the capacitors C_e and C_u. The additional step observed in the experiments is expected to occur at the threshold voltage of the exposed area. The height of the step should be equal to the difference between the accumulation capacitance, C_{ae}, and the inversion capacitance, C_{ie}, of the exposed area.

To determine the theoretical relationship between the step height, $C_{ae} - C_{ie}$, and the exposed area, A_e, we consider the following relations as derived from basic MOS C-V theory (Pierret 1983). The accumulation capacitance of the exposed area is given by,

$$C_{ae}=\frac{\kappa_{ox}\varepsilon_0 A_e}{d_{ox}} \tag{1}$$

where κ_{ox} is the dielectric constant of SiO_2, ε_0 is the vacuum permittivity, and d_{ox} is the oxide thickness. Similarly, the semiconductor capacitance of the exposed area in inversion can be expressed as:

$$C_{ie}=\frac{\kappa_{Si}\varepsilon_0 A_e}{W} \tag{2}$$

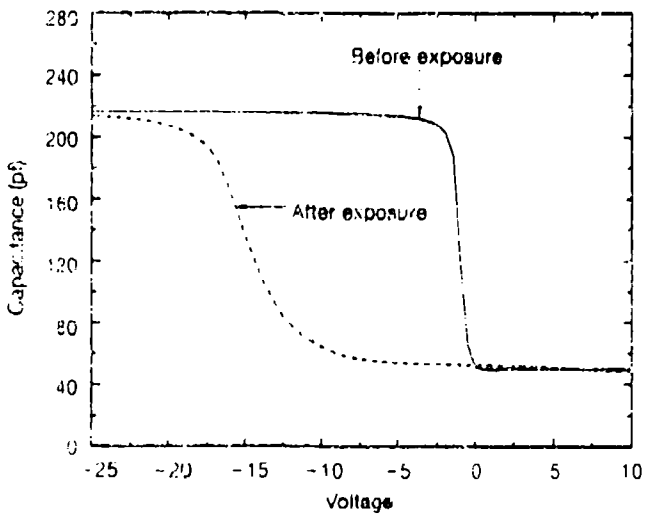


FIG. 3 C-V characteristics of a MOS capacitor before and 30 min after total e-beam area exposure.

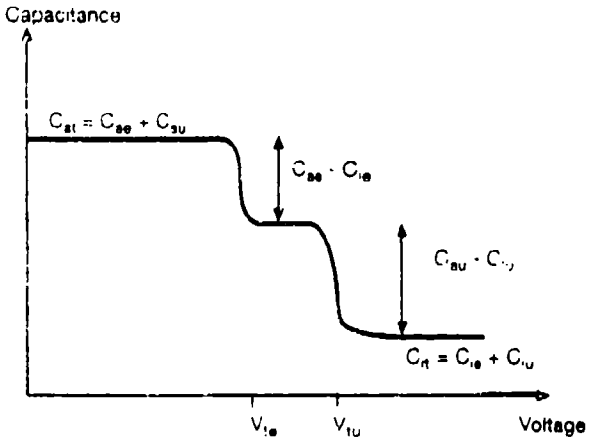


FIG. 4 Schematic diagram of the C-V curve resulting from partial area exposure. Key: C = capacitance, with subscripts a = accumulation, t = total, e = exposed, u = unexposed, and i = inversion. V_{te} and V_{tu} are threshold voltages for exposed and unexposed portions, respectively.

where K_s is the dielectric constant of silicon and W is the maximum depletion-layer width into the semiconductor. The capacitance of the exposed area in inversion can now be expressed as the series combination of the above two capacitances:

$$C_{ie} = \frac{C_{ox} C_{se}}{C_{ox} + C_{se}} \quad (3)$$

The total accumulation capacitance is:

$$C_{at} = \frac{\kappa_{ox} \varepsilon_0 A_t}{d_{ox}} \quad (4)$$

where A_t is the total capacitor area. Thus, combining all of the above equations, we obtain a linear relationship between the step height and the exposed area:

$$\frac{C_{at} - C_{ie}}{C_{at}} = K \frac{A_e}{A_t} \quad (5)$$

where

$$K = \left[\frac{1}{1 + \frac{\kappa_s d_{ox}}{\kappa_{ox} W}} \right]$$

Figure 5 compares the experimental data, obtained 30 min after exposure, and the theoretical predictions from Eq. (5). Shown is a plot of the normalized step height, $(C_{at} - C_{ie})/C_{at}$, versus the normalized exposure area, A_e/A_t. The error bars in the experimental data represent uncertainty in the sharpness of the steps. For each individual capacitor with a certain exposed area, we also show the theoretical prediction from Eq. (5) for the step height, taking into account the precise value of the oxide thickness for the calculation of the constant, K. The straight line shown in the figure represents the relationship of Eq. (5) for a value of K corresponding to an average oxide thickness of 766 A.

Figure 5 gives excellent agreement between our experimental results and the above theoretical model, which assumes that the positive charge remains fixed in the area where it is created. The step height in the C-V data may thus be used as a measure for the areal extent of positive charges in insulators. *In particular, the out-diffusion of positive oxide charges may be monitored as a change of the step height in time.*

To gain further insight into the time-dependence of the positive oxide charge, we performed an experiment to determine its position and magnitude a significant time after its creation. We analyzed the C-V data 30 min, 2 weeks, and 4 weeks after partial area exposure of an MOS capacitor, where the exposed area was 41.6% of the total area. The chip containing this capacitor was stored at room temperature in a dark shielded box during the 4-week time period. Figure 6 shows the before- and after-exposure curves. We see that the step observed 30 min after exposure has moved approximately 0.8 volts in the positive direction after 2 weeks, but only 0.2 volts more after an additional 2 weeks. Note that the step height did not

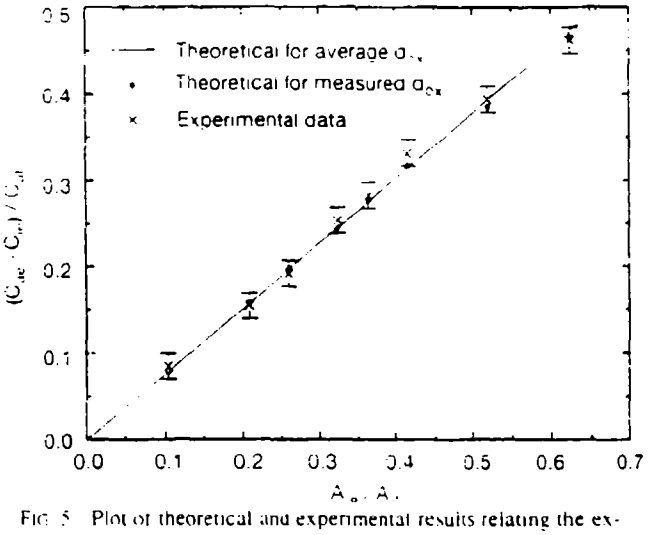


FIG. 5. Plot of theoretical and experimental results relating the exposure area to the resulting step in the C-V characteristics. The crosses with error bars indicate experimental data. The filled circles represent the predictions of our theoretical model using the actual oxide thickness for each case. The solid line shows the linear relationship of Eq. (5) for the average oxide thickness of d_{ox} = 766 A.

change over the 4-week period, implying that *no measurable diffusion of charge has taken place.*

Even though the charge did not spread out, its magnitude decreased over time. The amount of charge dissipation can be determined from simple MOS theory. The shift in the threshold voltage of the exposed area, ΔV_{fe}, is related to a change in the oxide charge, ΔQ_{ox}, by:

$$\Delta V_{fe} = \frac{-\Delta Q_{ox}}{C_{ox}} \quad (6)$$

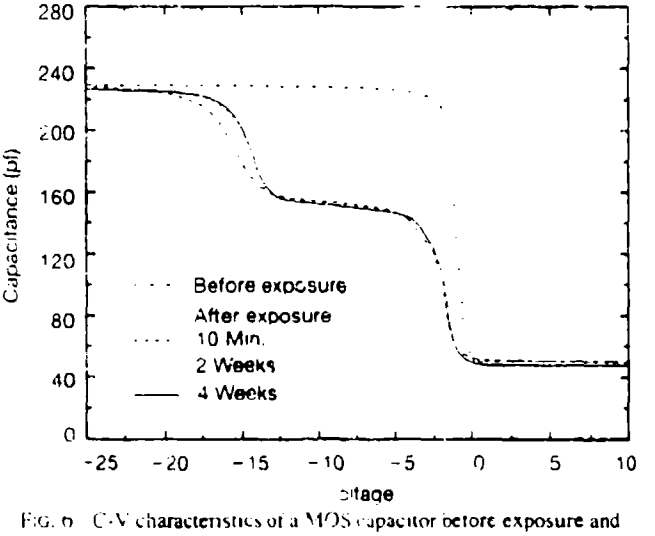


FIG. 6. C-V characteristics of a MOS capacitor before exposure and 30 min, 2 weeks, and 4 weeks after 41.6% of the total area was exposed with an electron beam.

The shift in the threshold voltage of −14.5 volts 30 min after the exposure corresponds to an induced charge density of 6.8×10^{-7} C/cm^2. After 2 weeks, the charge is reduced by only about 6% to 6.4×10^{-7} C/cm^2. After an additional 2 weeks, the charge is further reduced a mere 1.5%. Positive fixed charges in the oxide can be considered nearly "permanent" in the absence of any annealing treatments. In fact, Zaininger (1967) observed that positive charge can endure for several months in capacitors wholly exposed to electron radiation.

One might theorize a mechanism by which induced positive charge is not fixed in the oxide but, in fact, leaks out over a small distance onto the ungrounded metal (i.e., electrons into the oxide), causing a bending of the band structure for the two materials and resulting in a cessation of charge transfer after some time. Such a mechanism might not be detectable by our technique, as our lateral resolution is obviously limited and the oxide is very thin. However, if this were the case, the metal probe used in the C-V measurement would quickly drain the charge for the metal and allow a continued leakage of charge to the C-V measurement system. Presumably, such a leakage would be expected to act rapidly since in an SEM, sample charging is apparently controlled on time scales of the order of screen refresh rates, that is, a matter of a few seconds or less.

Conclusion

We have shown that partial area exposure of MOS capacitors using electron beam irradiation results in the creation of an area of *immobile* positive charge within the SiO_2 layer. This charge causes depletion and inversion of the exposed area of the capacitor at more negative voltages than the unexposed area, which resulted in a step in the C-V characteristics. A systematic study of exposed area as a function of step height indicated that step height increased with increasing exposure area. After a theoretical model had been proposed, experimental results were compared with those obtained from this model. The good agreement we obtained demonstrated a direct correlation between the exposed area and the step height in the C-V data. We exploited this phenomenon to show that the positive oxide charge does not diffuse over long time periods. This demonstrates that in electron microscopy, the common technique of coating nonconductive substrates with a metal overlayer in the case of positive charging is accomplished not by the draining of excess charge away from the imaging site but rather by the electrical shielding of the site by a ground plate.

Acknowledgments

The authors thank Mr. Robert J. Minniti, Jr., and Mr. Malay Kundu for assistance with measurements and data acquisition, and Mr. Greg Bazán for electron beam lithography exposures. We are grateful to Prof. Stephen J. Krause and Prof. David C. Joy for helpful discussions.

References

Balasinski A, Jakubowski A, Swit A: The influence of electron beam energy on defect density in MOS device quality oxides. *Vacuum* 38, 1041–1043 (1988)

Goldstein JI, Newbury DE, Echlin P, Joy DC, Fiori C, Lifshin E: *Scanning Electron Microscopy and X-Ray Microanalysis*. Plenum Press, New York (1981) 461–494

Henderson RC: Device fabrication by electron beam lithography. In *Electron Beam Technology in Microelectronic Fabrication*. (Ed. Brewer GR) Academic Press, New York (1980) 226–227

Johnson WC: Mechanisms of charge build-up in MOS insulators. *IEEE Trans Nucl Sci* NS-22, 2144–2150 (1975)

Keery WJ, Leedy KO, Galloway KF: Electron beam effects on microelectronic devices. *Scanning Electron Microscopy, Proc Workshop on Microelectronic Device Fabrication with the SEM*. IIT Research Institute, Chicago (1976) 507–513

Krause SJ, Mohr J, Bernstein G, Ferry DK, Joy DC: A new method for measuring the thickness of thin film insulators. *Proc Microbeam Analysis Soc* 14, 459–462 (1989)

McLean FB, Boesch HE Jr, Oldham TR: Electron-hole generation, transport, and trapping in SiO_2. In *Ionizing Radiation Effects in MOS Devices and Circuits* (Eds. Ma TP, Dressendorfer PV) John Wiley & Sons, New York (1990) 87–92

Nicollian EH, Brews JR: *MOS (Metal Oxide Semiconductor) Physics and Technology*. John Wiley & Sons, New York (1982) 549–553

Pierret RF: Field effect devices. *Modular Series on Solid State Devices* 4, Addison-Wesley, (1983) 43–80

Sah CT, Sun JYC, Tzou JJT: Effects of keV electron radiation on the avalanche-electron generation rates of three donors on oxidized silicon. *J Appl Phys* 54, 1563–1568 (1983)

Shimaya M, Shiono N, Nakajima O, Hashimoto C, Sakakibara Y: Electron beam induced damage in poly-Si gate MOS structures and its effect on long-term stability. *J Electrochem Soc* 130, 945–950 (1983)

Zaininger KH: Irradiation of MIS capacitors with high energy electrons. *IEEE Trans Nucl Sci* NS-13, 237–247 (1966)

Zaininger KH, Holmes-Siedle AG: A survey of radiation effects in metal-insulator-semiconductor devices. *RCA Review* June 208–240 (1967)