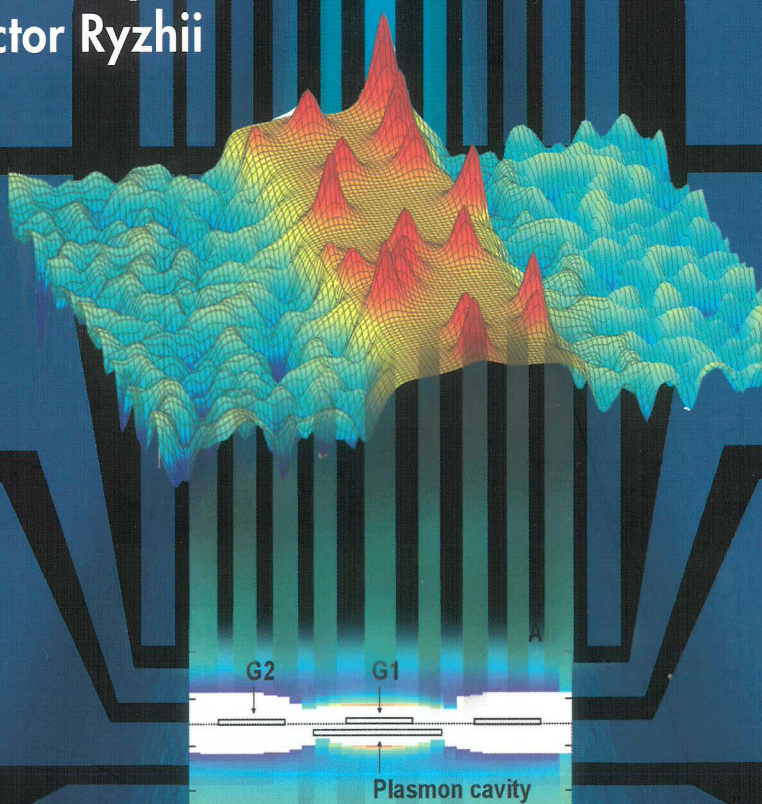


PHYSICS AND MODELING OF TERA- AND NANO-DEVICES

Editors

Maxim Ryzhii

Victor Ryzhii



World Scientific

PHYSICS AND MODELING OF TERA- AND NANO-DEVICES

SELECTED TOPICS IN ELECTRONICS AND SYSTEMS

Editor-in-Chief: **M. S. Shur**

Published

- Vol. 31: Advanced Device Modeling and Simulation
ed. *T. Grassler*
- Vol. 32: Terahertz Sensing Technology — Vol. 2
Emerging Scientific Applications and Novel Device Concepts
eds. *D. L. Woolard, W. R. Loerop and M. S. Shur*
- Vol. 33: GaN-Based Materials and Devices
eds. *M. S. Shur and R. F. Davis*
- Vol. 34: Radiation Effects and Soft Errors in Integrated Circuits and Electronic Devices
eds. *R. D. Schrimpf and D. M. Fleetwood*
- Vol. 35: Proceedings of the 2004 IEEE Lester Eastman Conference on High Performance Devices
ed. *Robert E. Leoni III*
- Vol. 36: Breakdown Phenomena in Semiconductors and Semiconductor Devices
M. Levinshtein, J. Kostamovaara and S. Vainshtein
- Vol. 37: Radiation Defect Engineering
Kozlovski V. and Abrosimova V.
- Vol. 38: Design of High-Speed Communication Circuits
ed. *R. Harjani*
- Vol. 39: High-Speed Optical Transceivers
eds. *Y. Liu and H. Yang*
- Vol. 40: SiC Materials and Devices — Vol. 1
eds. *M. S. Shur, S. Rumyantsev and M. Levinshtein*
- Vol. 41: Frontiers in Electronics
Proceedings of the WOFE-04
eds. *H Iwai, Y. Nishi, M. S. Shur and H. Wong*
- Vol. 42: Transformational Science and Technology for the Current and Future Force
eds. *J. A. Parmentola, A. M. Rajendran, W. Bryzik, B. J. Walker, J. W. McCauley, J. Reifman, and N. M. Nasrabadi*
- Vol. 43: SiC Materials and Devices — Vol. 2
eds. *M. S. Shur, S. Rumyantsev and M. Levinshtein*
- Vol. 44: Nanotubes and Nanowires
ed. *Peter J. Burke*
- Vol. 45: Proceedings of the 2006 IEEE Lester Eastman Conference on Advanced Semiconductor Devices
eds. *Michael S. Shur, P. Maki and J. Kolodzey*
- Vol. 46: Terahertz Science and Technology for Military and Security Applications
eds. *Dwight L. Woolard, James O. Jensen, R. Jennifer Hwu and Michael S. Shur*

PHYSICS AND MODELING OF TERA- AND NANO-DEVICES

Editors

Maxim Ryzhii

Victor Ryzhii

University of Aizu, Japan

 **World Scientific**

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

Selected Topics in Electronics and Systems — Vol. 47
PHYSICS AND MODELING OF TERA- AND NANO-DEVICES

Copyright © 2008 by World Scientific Publishing Co. Pte. Ltd

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN-13 978-981-277-904-5

ISBN-10 981-277-904-3

Editor: Tjan Kwang Wei

Printed in Singapore by Mainland Press Pte Ltd

PREFACE

The content of this issue is based on the invited and contributed papers presented by the researchers working in the field of physics and modeling of novel electronic and optoelectronic devices at the International Workshop “Tera- and Nano-Devices: Physics and Modeling” held on October 16–19, 2006 in Aizu-Wakamatsu, Japan. The workshop was organized by V. Ryzhii, G.P. Berman, V. Mitin, T. Otsuji, M. Ryzhii, A. Satou, and M.S. Shur.

The papers in this issue include devices based on carbon nanotubes, generation and detection of terahertz radiation in semiconductor structures including terahertz plasma oscillations and instabilities, terahertz photomixing in semiconductor heterostructures, spin and microwave-induced phenomena in low-dimensional systems, and various computational aspects of device modeling.

We thank the financial support from University of Aizu, Los Alamos National Laboratory, Air Force Office of Scientific Research/Asian Office of Aerospace Research and Development*, University at Buffalo, Sendai Section of the Institute of Electrical and Electronics Engineering, and Technical Group on Terahertz Application Systems of Institute of Electronics, Information, and Communication Engineers.

* Disclaimer: AFOSR/AOARD support is not intended to express or imply endorsement by the US Federal Government.

This page intentionally left blank

CONTENTS

Preface	v
Semiconductor Device Scaling: Physics, Transport, and the Role of Nanowires	1
<i>D. K. Ferry, R. Akis, M. J. Gilbert, A. Cummings and S. M. Ramey</i>	
Polaronic Effects at the Field Effect Junctions for Unconventional Semiconductors	13
<i>N. Kirova</i>	
Cellular Monte Carlo Simulation of High Field Transport in Semiconductor Devices	21
<i>S. M. Goodnick and M. Saraniti</i>	
Nanoelectronic Device Simulation Based on the Wigner Function Formalism	31
<i>H. Kosina</i>	
Quantum Simulations of Dual Gate MOSFET Devices: Building and Deploying Community Nanotechnology Software Tools on nanoHUB.org	41
<i>S. Ahmed, G. Klimeck, D. Kearney, M. McLennan and M. P. Anantram</i>	
Positive Magneto-Resistance in a Point Contact: Possible Manifestation of Interactions	51
<i>V. T. Renard, T. Ota, N. Kumada and H. Hirayama</i>	
Impact of Intrinsic Parameter Fluctuations in Nano-CMOS Devices on Circuits and Systems	57
<i>S. Roy, B. Cheng and A. Asenov</i>	
HEMT-Based Nanometer Devices Toward Tetrahertz Era	65
<i>E. Sano and T. Otsuji</i>	
Plasma Waves in Two-Dimensional Electron Systems and Their Applications	77
<i>V. Ryzhii, I. Khmyrova, M. Ryzhii, A. Satou, T. Otsuji, V. Mitin and M. S. Shur</i>	
Resonant Terahertz Detection Antenna Utilizing Plasma Oscillations in Lateral Schottky Diode	95
<i>A. Satou, V. Ryzhii, T. Otsuji and M. S. Shur</i>	

Terahertz Polarization Controller Based on Electronic Dispersion Control of 2D Plasmons	103
<i>T. Nishimura and T. Otsuji</i>	
Higher-Order Plasmon Resonances in GaN-Based Field-Effect Transistor Arrays	113
<i>V. V. Popov, M. S. Shur, G. M. Tsymbalov and D. V. Fateev</i>	
Ultra-Highly Sensitive Terahertz Detection Using Carbon-Nanotube Quantum Dots	123
<i>Y. Kawano, T. Fuse and K. Ishibashi</i>	
Generation of Ultrashort Electron Bunches in Nanostructures by Femtosecond Laser Pulses	127
<i>A. Gladun, V. Leiman, A. Arsenin, O. Mannoun and V. Tarakanov</i>	
Characterization of Voltage-Controlled Oscillator Using RTD Transmission Line	133
<i>K. Narahara, T. Yamaki, T. Takahashi and T. Nakamichi</i>	
Infrared Quantum-Dot Detectors with Diffusion-Limited Capture	141
<i>N. Vagidov, A. Sergeev and V. Mitin</i>	
Magnetoresistance in Fe/MgO/Fe Magnetic Tunnel Junctions	149
<i>N. N. Beletskii, S. A. Borysenko, V. M. Yakovenko, G. P. Berman and S. A. Wolf</i>	
Modeling and Implementation of Spin-Based Quantum Computation	155
<i>M. E. Hawley, G. W. Brown and G. P. Berman</i>	
Quantum Engineering for Threat Reduction and Homeland Security	163
<i>G. P. Berman, A. R. Bishop and B. M. Chernobrod</i>	
Strong Phase Shift Mask Manufacturing Error Impact on the 65 nm Poly Line Printability	175
<i>N. Belova</i>	

SEMICONDUCTOR DEVICE SCALING: PHYSICS, TRANSPORT, AND THE ROLE OF NANOWIRES

D. K. FERRY, R. AKIS, M. J. GILBERT,* and A. CUMMINGS

*Department of Electrical Engineering and Center for Solid State Electronics Research
Arizona State University, Tempe, AZ 85287-5706, USA
ferry@asu.edu*

S. M. RAMEY

Intel Corporation, Hillsboro, OR 97124, USA

Nanoelectronics (including nanomagnetics and nanophotonics) generally refers to nanometer scale devices, and to circuits and architectures which are composed of these devices. Continued scaling of the devices into the nanometer range leads to enhanced information processing systems. Generally, this scaling has arisen from three major sources, one of which is reduction of the physical gate length of individual transistors. Until recently, this has also allowed an increase in the clock speed of the chip, but power considerations have halted this to levels around 4 GHz in Si. Indeed, there are indications that scaling itself may be finished by the end of this decade. Instead, there are now pushes to seek alternative materials for nano-devices that may supplement the Si CMOS in a manner that allows both higher speeds and lower power. In this paper, we will cover some of the impending limitations, and discuss some alternative approaches that may signal continued evolution of integrated circuits beyond the end of the decade.

Keywords: Nanoelectronics; nanowires; discrete impurities; ballistic transport.

1. Introduction

As the density of integrated circuits continues to increase, a resulting shrinkage of the dimensions of the individual devices of which they are comprised has followed. Smaller circuit dimensions reduce the overall circuit area, thus allowing for more transistors on a single die without negatively impacting the cost of manufacturing. However, this reduction in device size is only one of three factors identified by Moore in the increased density of modern integrated circuits. Equally important are the two other factors of an increase of die size and an increase in circuit cleverness – the reduction in number of devices and chip area to implement a given circuit or function. All of these lead to the driving force for continued integration complexity is *the reduction in cost per function for the chip*. We will return to this later, but the purpose of this paper is primarily to

* Present address: Microelectronics Research Center, University of Texas at Austin, Austin, TX 78758, USA

examine some of the problems and options for continued reductions in device size and increase of functionality per chip.

As semiconductor feature sizes shrink into the nanometer scale regime, device behavior becomes increasingly complicated as new physical phenomena at short dimensions occur, and limitations in material properties are reached. In addition to the problems related to the actual operation of ultra-small devices, the reduced feature sizes require more complicated and time-consuming manufacturing processes. This fact signifies that a pure trial-and-error approach to device optimization will become impossible since it is both too time consuming and too expensive. Nevertheless, it is important to consider these new physical effects which will occur in small devices, as these effects may well eventually dominate device performance. In Sec. 2, we will examine the importance of discrete impurities in these devices.

The MOSFET (Metal Oxide Field Effect Transistor) has been a staple of the semiconductor industry for many years, and it is its inclusion in complementary MOS circuitry that has driven much of the rapid density increases of the past decade. Currently the gate length is about 35 nm, and will continue to be reduced in future generations. In fact, it is quite likely that the gate length will approach 10 nm before the end of this decade. With such a small channel length, it is then assumed that ballistic transport should be the dominant method of transport. However, recent experiments have suggested that the ballistic length in silicon may well be less than the assumed 10 nm.¹ In Sec. 3, we will examine ballistic transport, and show that modern MOSFETs are probably not dominated by ballistic transport, and that this is likely a good thing!

One of the most promising solutions for devices beyond the normal MOSFET is that of the tri-gate nanowire transistor.² This device offers improved control over the channel, nearly ideal sub-threshold slope, and excellent behavior when compared to the traditional bulk MOSFET. In fact, the tri-gate transistor is almost a nanowire with a wrapped gate. But, the use of nanowire transistors is more extensive than just those based in Si, and these devices have a great deal of promise for applications beyond those of the normal MOSFET. Indeed, some applications have been suggested that would allow them to implement reconfigurable architectures, which get at the third component of increased complexity on modern chips — circuit cleverness. We turn to a consideration of these nanowire devices in Sec. 4, where we mention a number of nanowire devices that have appeared as well as discuss the circuit implementation of vertical transistors. We also cover some novel processing approaches which have been suggested for nanowire devices and their circuits, such as spin processing using Rashba fields in heterostructure devices.

Finally, we summarize the paper and discuss possible future directions for nano-device research in Sec. 5.

2. Discrete Impurity Scattering Effects

As semiconductor devices scale to smaller sizes, averaging of the carriers and dopants into a density for a region becomes less appropriate. Rather, the granularity of the

dopants, and the interactions between electrons, becomes a significant contributor to overall device performance. The treatment of doping as discrete atoms leads to more accurate mobility calculations,³ and to threshold voltage fluctuations relating to their actual number and location.⁴ For example, the potential landscape, for a slice of a silicon-on-insulator (SOI) MOSFET, is shown in Fig. 1. The source and drain contain donors (attractive), while the channel contains acceptors (repulsive).

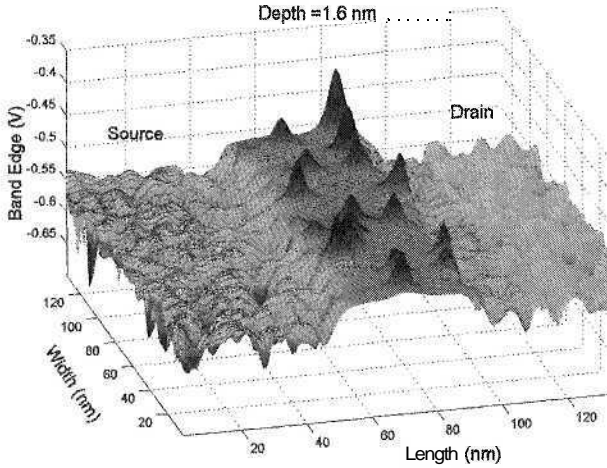


Fig. 1. The potential landscape for a slice of an SOI MOSFET.²¹ The potential minima in the source and drain arise from donors, while the peaks in the channel arise from acceptors.

Moreover, it is equally apparent that attempts to reduce the effect of random dopants by leaving the channel undoped will not completely solve this problem. It may be observed that the presence of random dopants in the source and drain regions mean that the boundary between e.g. source and channel is rather vague. This boundary is a randomly varying position depending upon just where the donors are sited near the boundary. Hence, this randomness will introduce an effective "line edge roughness" equivalent to that of lithography edge roughness in the gate definition.⁵

To accurately include these random dopant effects into Monte Carlo simulations, an additional routine must be included such as the molecular dynamics (MD) approach. Former implementations of MD within Monte Carlo simulations treated the interaction with a classical force description. However, quantum mechanical effects may also play a critical role in electron transport in these small devices. Quantum mechanical effects, along with a MD method for treating electron-ion interactions, have been incorporated in an ensemble Monte Carlo simulation of ultra-small SOI MOSFETs.⁶

Within a device simulator, the time-dependent self-consistent electric fields must be obtained, usually by solving the Poisson equation on a mesh. However, this solution typically will not have the necessary spatial resolution to describe the short-range nature of the electron-electron and electron-ion interactions. Therefore, it is desirable to explicitly include these interactions with a molecular dynamics routine. However, MD

methods also provide an electric field, derived from the Coulomb force. When including an MD routine within a device simulator, some provision must be made to avoid “double counting” the force from a discrete ion. One successful method involves calculating a corrected Coulomb force to be used in the MD routine. In principle, one can do a separation of the net inter-particle forces into long-range and short-range parts, as

$$V(r) \sim \frac{1}{r} \equiv \frac{F(r)}{r} + \frac{1-F(r)}{r}, \quad (1)$$

where $F(r)$ is a function which begins at 0 for $r = 0$, and increases to unity for large r . Hence, the first term in Eq. (1) is a long-range term, which can be incorporated into the Poisson equation, while the second term is a short-range term which is used within the MD computation.

This correction may be optimized by pre-computing the mesh force from a single ion and subtracting it from the Coulomb force. The accuracy of the long-range fit is then used to adjust the nature of the transition function $F(r)$. This corrected Coulomb force, the last term in Eq. (1), is then added to the actual mesh force obtained during the real simulation.

To avoid unnecessary computation, the two-particle mesh force is pre-computed for a given mesh spacing and then used in subsequent simulations. It was found that a spacing of 0.2 nm between these points provided adequate resolution for the two-particle mesh force. Since mesh fields also depend upon the simulation type (classical versus effective quantum potential), these mesh fields are also calculated for each type of simulation. More details on the method can be found in Ref. 6.

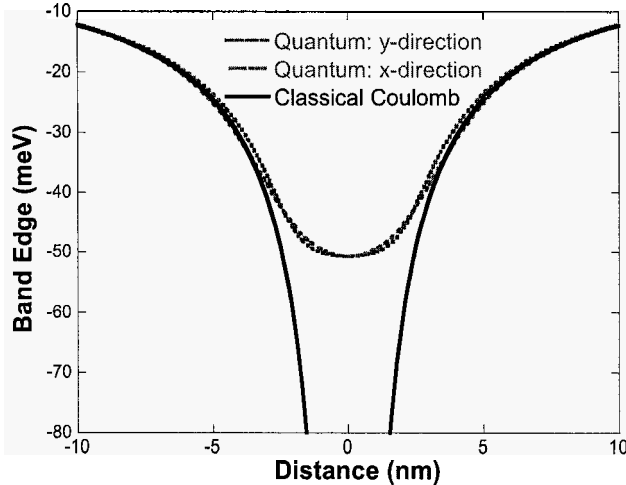


Fig. 2. Impurity potential with grid force corrected by the quantum potential. Combining the cut-off Coulomb potential with the effective potential gives the correct ionization energy for the donor.

As mentioned, one also needs to incorporate corrections for the onset of quantum effects in the nano-devices. Generally, the non-zero size of the electron wave packets

tends to smooth out any sharp potential. Hence, one can use such an effective potential, which is convolved with the actual solution of Poisson's equation to account for the onset of the initial quantum effects within the device. This leads to charge set-back from the oxide interface and the initial quantization of the energy within the channel. The two major approaches to incorporating these quantum corrections are the so-called density gradient method^{7,8} and the effective potential method.^{9,10} Careful comparison between these two approaches tends to show that they give qualitatively similar results in device simulations.¹¹ In Fig. 2, we illustrate how an impurity potential, here that of an attractive donor atom, can be smoothed with the effective potential, so that the mesh solution gives a proper minimum when smoothed in this way. The depth of this potential should be the ionization energy of the donor atom, here about 51 meV.

3. Ballistic Transport in Nano-Devices

Ballistic transport in semiconductors is a relatively old idea. It was first discussed in regard to mesoscopic structures, where the mean free path was comparable to the device size, in connection with the Landauer formula.¹² In fact, the ideas of ballistic transport are even older, and derive from the earliest treatments of transport in vacuum diodes. The Langmuir-Child law describes the ballistic transport of electrons in a thermionic diode, with space charge built up near the cathode (corresponding to our source in a MOSFET), after the two who developed it **independently**.^{13,14} Both of these men derived the expression for the current, finding that

$$I \sim \frac{V^{3/2}}{L^2} \quad , \quad (2)$$

and it is this relationship that has become known as the Langmuir-Child law. More recently, Shur and Eastman proposed that device performance could be improved by utilizing ballistic transport in ultra-small channel length semiconductor devices,¹⁵ but also showed that the current in an n^+-n-n^+ device would have the same space charge and current relationship as that of Eq. (1). It is important that the MOSFET has a space charge region, and potential maximum, between the source and the channel, and it is this that creates the connection to the Langmuir-Child law, as demonstrated by Shur and Eastman. In essence, the latter were suggesting use of high velocities due to transient velocity overshoot that can occur in many materials.¹⁶ More recently, there have been many suggestions that ballistic transport can occur in short-channel devices, and might improve the performance.¹⁷ This, in fact, is not the case, and a proper treatment of ballistic transport will show that it is detrimental to good device operation. In this section, we will outline the basic tenets that establish this point.

First, true ballistic transport occurs in the complete absence of scattering. This is, of course, found in vacuum tubes. There, electrons leave the cathode and form a space charge layer adjacent to this region. The solution of Poisson's equation for the region between the cathode and the plate yields the Langmuir-Child Law [Eq. (2)]. The importance of the Shur and Eastman result¹⁵ is that *exactly the same behavior* is found in n^+-n-n^+ semiconductor structures, which is the structure that is found in the n -channel

MOSFET. Electrons move out of the source into the channel, creating a space-charge region at the source-channel interface. It is modulation of this space-charge region by the gate potential that produces the normal device characteristics.¹⁸ Variation of the space-charge region by the gate (or by the grid in the vacuum tube) leads to a family of triode-like curves obeying Eq. (1) with different (gate voltage dependent) coefficients. These triode curves are not good for either logic or high frequency applications.

How then are the good characteristics, with current saturation, obtained? In the case of the vacuum tube, a “screen” grid (a metal grid) is inserted and held at a constant high potential so that the space-charge region is isolated from the anode potential. In the case of the MOSFET, similar screening occurs, but this time it is provided by the *scattering that occurs in the region between the space-charge layer and the drain*. One normally does not connect scattering with screening, but this is a common occurrence in, for example, quantum transport. Moreover, it has been seen in detailed simulation that scattering has a large effect on the actual potential distribution and therefore on the device characteristics.¹⁹ In fact, we can see this behavior in the simple device characteristics

$$\begin{aligned}
 I_{DS} &= \frac{\mu WC_{ox}}{L} \left(V_G - V_T - \frac{V_D}{2} \right) V_D \\
 &= \frac{\mu WC_{ox}}{2L} \left[(V_G - V_T)^2 - (V_G - V_T - V_D)^2 \right] .
 \end{aligned}
 \tag{3}$$

Saturation occurs when the second term in the square brackets vanishes due to pinchoff at the drain end. In this situation, the drain potential does not affect the source operation.

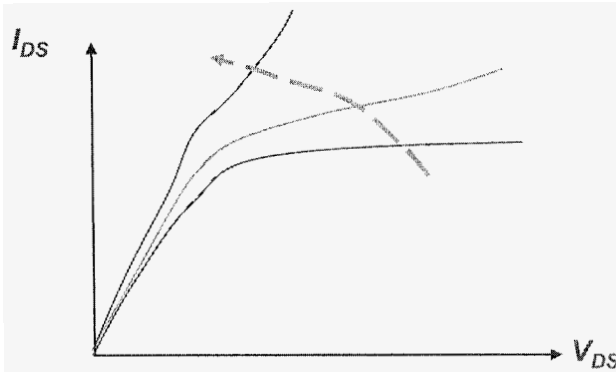


Fig. 3. Change in the output characteristics as ballistic transport becomes important. As scattering is reduced, the curves will transition to triode-like behavior.

However, when we begin to lose the scattering in the channel, e.g., when we begin to see quasi-ballistic transport, then we should begin to see a transition to triode-like characteristic curves, with a reduction in output drain resistance. This behavior is shown in Fig. 3. The saturation will disappear as this triode-like behavior becomes more and more prominent. The astute reader will notice that this behavior is exactly like drain-induced barrier lowering (DIBL). In fact, DIBL is the first precursor to ballistic transport.

DIBL occurs when there is insufficient scattering to screen the space-charge region from the drain potential variations. Since it is generally accepted that DIBL is detrimental to good device operation, we may safely conclude that we really don't want to have ballistic transport occurring in our devices.

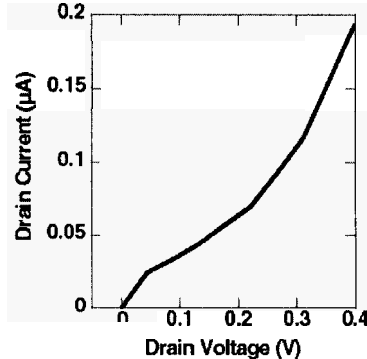


Fig. 4. Output characteristics for the InAs MOSFET device with a gate bias of 0.4 V.

Given that ballistic behavior is detrimental to the devices, how can we ascertain that it is not really occurring. This is difficult to achieve experimentally, but not so difficult to investigate in meaningful device simulations. As we point out in the next section, it is quite likely that future devices may well be built around the concept of nanowires. To that end, it is logical to investigate whether there is any ballistic behavior in such devices. Kotylar *et al.*²⁰ examined classical scattering in a Si quantum wire and concluded that the mobility would not be improved in this structure, contrary to many expectations. We pursued a different approach and utilize a fully quantum mechanical, and three-dimensional, simulation of small semiconductor quantum-wire MOSFETs.²¹ In this approach, the full Poisson equation solution is used to determine the local potential, and a recursive scattering matrix approach is used to determine the transport through the device. In this process, for each iteration from one transverse slice of the device to the next, a local Dyson's equation is solved with the slice Hamiltonian, a procedure equivalent to the scattering matrix solution of the Lippmann-Schwinger equation. In Fig. 4, we show one output characteristic for a 30 nm gate length InAs quantum wire MOSFET, in which there is no scattering in the channel.²² We have considered an InAs tri-gate MOSFET, whose structure is the same as in Fig. 5(a) (below), except the channel is 8.5 nm wide and 30.3 nm long. The source and drain are 26.3 nm wide. The InAs layer is taken to be 9.3 nm thick. The source and drain are doped to $6 \times 10^{18} \text{ cm}^{-3}$, and the channel is undoped but assumed to be weakly *p*-type. The oxide is taken to be HfO_2 . It is clear in this device that the ballistic behavior discussed above is operating.

As mentioned above, a local Dyson's equation is solved with the slice Hamiltonian. This means that we can modify this Hamiltonian by the direct inclusion of a slice self-energy as well as a self-energy coupling between the slices where that is appropriate.

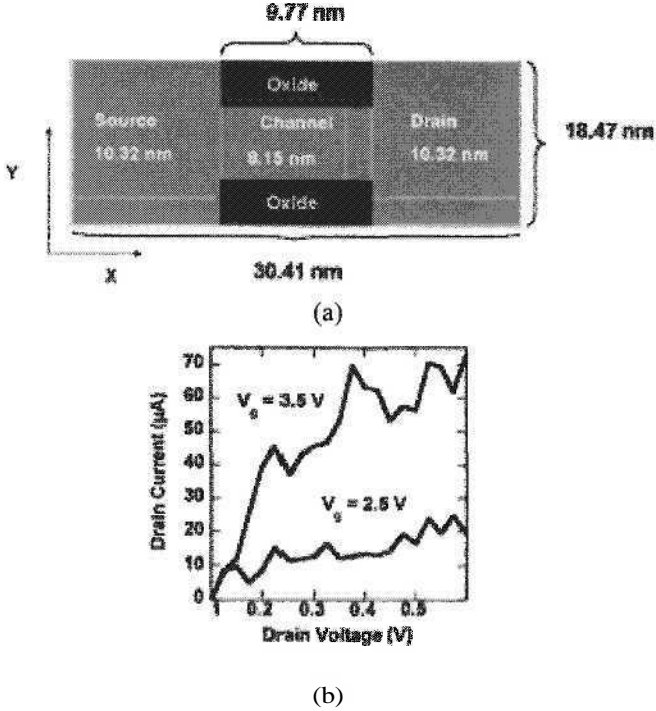


Fig. 5. (a) Structure of the Si quantum wire transistor. The SOI layer gives a Si thickness of 6.5 nm. The source and drain are doped to 10^{20} cm^{-3} and the channel is undoped. The oxide also covers the top, and the gate is on three sides, over the gate oxide. (b) Characteristics for a device whose gate length is 10.3 nm.

This self-energy term describes the dissipation within the device.²³ We have computed his self-energy for all the normal phonon scattering processes expected to occur in a Si nanowire (impurity scattering is included directly through the random impurity potential). This self-energy is now incorporated in the Hamiltonian to solve for the overall transport through the device. In Fig. 5, we show a typical set of device characteristics, which illustrates that these devices, even with such short channels, exhibit fairly good saturation in the output characteristics (there is very little parasitic source resistance due to the structure shown in the figure). The variations in the current arise from the quantum interferences that are still present in the devices, even in the presence of the strong phonon scattering.

Now, we can use this same structure, with varying gate length (and channel length) to study whether or not there is any ballistic behavior in the device. From Fig. 5(b), we do not see the characteristic power law behavior that should be present if ballistic transport were important here. However, there is another way to check this, and that is to vary the channel length at low drain bias. If the transport is ballistic, then Landauer's formula tells us that the conductance should be constant, and therefore the resistance should be independent of the length. On the other hand, if the transport is diffusive, then the resistance should depend linearly on the length of the channel.

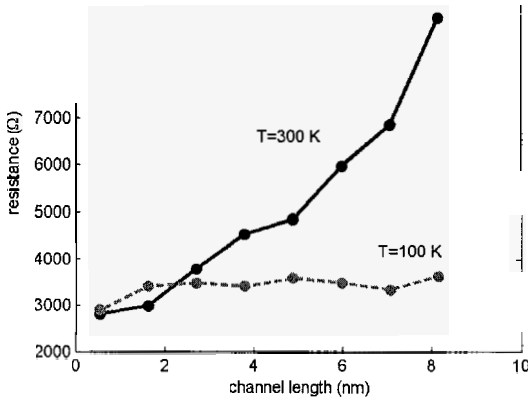


Fig. 6. Variation of the channel resistance, at a drain bias of 10 mV, as a function of the channel length. The constant behavior at 100 K is indicative of ballistic transport, while the linear rise at 300 K is indicative of diffusive transport.

In Fig. 6, we show the results for a Si nanowire SOI MOSFET, in which we plot the resistance as a function of the channel length.²⁴ At low temperatures (100 K), the resistance is independent of the length of the channel. This result is expected for ballistic transport, which arises because the phonons are frozen out at this low temperature. At high temperatures (300 K), however, the resistance exhibits the expected Ohm's Law linear dependence on device length. Below 2 nm, direct source-drain tunneling prohibits observation of the nanowire effects, and this is independent of temperature. Thus, it appears that there will be no real onset of important ballistic transport in future devices down to gate lengths below 5 nm, although there will continue to be problems with DIBL (while not shown, the results of Fig. 5 are sensitive to the drain potential that is imposed).

But, this is for silicon devices, which have relatively low mobilities and velocities. The scattering in Si is quite strong, and the high energy phonons give good momentum randomization, all of which serves to reduce the chances of ballistic behavior. Still, it is seen at low temperatures, as shown in Fig. 6. If we now move to a semiconductor with higher mobilities and velocities, and with scattering that is anisotropic, will the result change? In the III-V materials, the scattering is dominated by the polar LO phonon, which produces strongly anisotropic scattering, due to its Coulomb nature. As was seen above, the characteristics of the InAs device clearly show the trend toward the power law behavior, which can be indicative of the onset of ballistic behavior. The mobility in InAs is almost two orders of magnitude larger than that of Si, so that a similar increase in mean-free path is expected. Thus, a 2-4 nm limit in Si becomes several tens of nm in InAs, and the behavior seen in the figure is surely expected, even at room temperature.

4. Nanowire Devices

Of the long list of prospective devices for future technology that will allow a continuation of the trend of decreasing size in CMOS, one clear front runner is the tri-gate silicon nanowire transistor (Si NWT).² This device produces excellent output currents, good I_{on}/I_{off} ratio, and superior sub-threshold slope. SNWTs also appear to be superior to the bulk transistor in that the extra gates give a great deal of additional electrostatic control over the channel, creating *volume* inversion rather than *surface* inversion. However, Si NWTs still suffer from one of the main detracting elements of the silicon based transistors, low channel mobility. For this reason, Intel has proposed transistors based on III-V compounds,²⁵ which are known to have higher mobilities than silicon. In fact, both approaches may be pursued with the use of III-V-based NWTs.

Nanowires have been pursued for their intrinsic ability to make smaller devices for several years.²⁶ Vertically grown nanowires have been grown in several materials, and heterostructures have been embedded within these wires²⁷ to create quantum dots and resonant-tunneling diodes. In addition, carbon nanotubes (CNT) have been investigated for their ability to be used for a “semiconductor” device.²⁸ In many cases, the nanowires or CNTs are placed horizontal on an oxidized Si surface, and then source and drain contacts evaporated to create the device. Here, the back Si is used as the gate electrode. While making a primitive device, it has not stopped various authors from claiming fantastic performance from such devices. The problem with this approach is that a fair comparison requires recognition of the principle outlined at the beginning of this paper: *the cost per function is the driving force in VLSI*. Hence, it is not meaningful to compare the nanowire with an equivalent sized Si device; rather, the Si device with which the comparison should be made is one that occupies an equivalent amount of chip area. The nanowire devices usually have enormous source and drain contacts, and, if this size were used in the Si device, the nanowire devices would have no advantage at all.

Other problems arise from this as well. First, the intrinsic transconductance of any transistor is reduced by the parasitic source resistances, according to

$$g_{m,eff} = \frac{g_{m,int}}{1 + g_{m,int} R_S} \quad (4)$$

Here, $g_{m,int}$ is the intrinsic property of the device. Because the nanowire transistor, by its nature, carries only a few quantum modes, the resistance R_S is quite large. Consequently, the effective transconductance is usually only $1/R_S$. The cutoff frequency (which is essentially the reciprocal of the delay time in logic applications) is given by

$$f_T = \frac{g_{m,eff}}{2\pi(C_G + C_{par})} \quad (5)$$

Here, C_G is the actual effective gate capacitance while C_{par} is the parasitic capacitance. In the structures discussed above, the parasitic capacitance is much larger than the real gate capacitance, so that we find

$$f_T \sim \frac{1}{2\pi R_S C_{par}} \quad (6)$$

That is, the performance of the nanowire, or CNT, device is dominated by parasitic properties, not by any intrinsic properties of the nanowire device.

Does this mean that nanowire, and CNT, transistors have no future? Not at all! Rather, it means that the proper technology to incorporate these structures into high performance devices has not been utilized in most cases (but, the reader should look at the cases where this has been done²⁵). The proper method of comparison relies upon a well-designed set of experiments, and these have been described by Chau *et al.*²⁹ Moreover, we also need to remember that there is a third important factor in increasing chip device density, and that is clever circuit design.

A modern integrated circuit chip is a dense array of many different materials. While most of the devices sit at the Si surface, in the bottom-most level, there are several levels of metals and insulators lying above this. In fact, one of the most important uses of nanowire transistors may well be as vertical switches between levels of metal in these upper layers. This has been proposed for vertical CNT transistors,³⁰ and vertical transistors have been grown in semiconductors as well.³¹ The development of a vertical nanowire transistor, which can be integrated within the metallization layers of the integrated circuit will allow for active system reorganization, which can open the way to many novel new applications. Other applications of nanowires, whether vertically or horizontally oriented, may be in the area of novel computational approaches, such as quantum computing. The idea of using quantum wires for a qubit was apparently first suggested by Bertoni *et al.*³² Such processing uses the ideas of moving the qubits to various sites where manipulations are performed.^{33,34} This idea of “flying” qubits has been discussed in connection with the use of spin for the quantum state.³⁵ Here, the spin state accompanies an electron (or hole) moving through the quantum wire.

5. Conclusions

In this paper, we have presented some thoughts on the future of semiconductor devices intended for use in VLSI chips. We have discussed the roles of discrete impurities, ballistic transport, and quantum wires. While the end of devices as we know them may be in sight, it is not clear that no new ideas will evolve. In particular, even if device size scale reduction ends, there will continue to be advances in die size and in circuit cleverness. These will continue the exponential growth in chip complexity.

References

1. G. Timp, J. Bude, K. K. Bourdelle, J. Garno, A. Ghetti, H. Grossman, M. Green, G. Forsyth, Y. Kim, R. Kleiman, F. Klemens, A. Kornblit, C. Lochstampfor, W. Mansfield, S. Moccio, T. Sorsch, D. M. Tennant, W. Timp, R. Tung, The Ballistic Nano-Transistor, *IEDM Tech. Dig.* **1999**, 55-58 (1999).
2. H. S. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, *IEEE Electron Dev. Lett.* **24**, 263 (2003).

3. W. J. Gross, D. Vasileska, and D. K. Ferry, *IEEE Electron Dev. Lett.* **20**, 463 (1999).
4. D. Vasileska, W. J. Gross, and D. K. Ferry, *Superlatt. Microstruc.* **27**, 147 (2000).
5. T. D. Linton, Jr., S. Yu, and R. Shaheed, *VLSI Design* **13**, 103 (2001).
6. S. M. Ramey and D. K. Ferry, *IEEE Trans. Nanotech.* **2**, 193 (2003).
7. E. Madelung, *Z. Physik* **40**, 322 (1927).
8. J.-R. Zhou and D. K. Ferry, *IEEE Trans. Electron Dev.* **39**, 473 (1992).
9. D. K. Ferry, *Superlatt. Microstruc.* **27**, 61 (2000).
10. D. K. Ferry, R. Akis, and D. Vasileska, Quantum effects in MOSFETs: Use of an effective potential in 3D Monte Carlo simulation of ultra-short channel devices, *IEDM Tech. Dig. 2000*, 287-290 (2000).
11. A. Asenov, J. R. Watling, A. R. Brown, and D. K. Ferry, *J. Comp. Electron.* **1**, 503 (2002).
12. R. Landauer, *IBM J. Res. Develop.* **1**, 223 (1957).
13. C. D. Child, *Phys. Rev. (ser. 1)* **32**, 492 (1911).
14. I. Langmuir, *Phys. Rev.* **2**, 450 (1913).
15. M. S. Shur and L. F. Eastman, *IEEE Trans. Electron Dev.* **26**, 1677 (1979).
16. J. R. Barker, D. K. Ferry, and H. L. Grubin, *IEEE Electron Dev. Lett.* **1**, 209 (1980).
17. See, e.g., J. Saint Martin, A. Bournel, and P. Dollfus, *IEEE Trans. Electron Dev.* **51**, 1148 (2004).
18. M. Lundstrom, Z. Ren, and S. Datta, Essential physics of transport in nanoscale MOSFETs, *2000 SISPAD Tech. Dig.*, 1-5 (2000).
19. A. Svizhenko and M. P. Anantram, *IEEE Trans. Electron Dev.* **50**, 1459 (2003).
20. R. Kotylar, B. Obradovic, P. Matagne, M. Stettler, and M. D. Giles, *Appl. Phys. Lett.* **84**, 5270 (2004).
21. M. J. Gilbert and D. K. Ferry, *J. Appl. Phys.* **95**, 7954 (2004).
22. M. J. Gilbert and D. K. Ferry, *J. Appl. Phys.* **99**, 054503 (2006).
23. M. J. Gilbert, R. Akis, and D. K. Ferry, *J. Appl. Phys.* **98**, 094303 (2005).
24. R. Akis, M. J. Gilbert, and D. K. Ferry, *J. Phys. Conf. Series* **38**, 87 (2006).
25. R. Chau, J. Brask, S. Datta, G. Dewey, M. Doczy, B. Doyle, J. Kavalieros, B. Jin, M. Metz, A. Majumdar, and M. Radosavljevic, Emerging silicon and non-silicon nanoelectronic devices: Opportunities and challenges for future high-performance and low-power computational applications, *2005 IEEE VLSI-TSA Int. Symp. Tech. Dig.*, 13-16 (2005).
26. See, e.g., L. Samuelson, *Materials Today* **6**(10), 22 (2003).
27. M. T. Björk, C. Thelander, A. E. Hansen, L. E. Jensen, M. W. Larsson, L. R. Wallenberg, and L. Samuelson, *Nano Lett.* **4**, 1621 (2004).
28. M. Radosavljevic, H. J. Tersoff, and Ph. Avouris, *Appl. Phys. Lett.* **83**, 2435 (2003).
29. R. Chau, S. Datta, M. Doczy, B. Doyle, B. Jin, J. Kavalieros, A. Majumdar, M. Metz, and M. Radosavljevic, *IEEE Trans. Nanotechnol.* **4**, 153 (2005).
30. A. P. Graham, G. S. Duesberg, W. Hoenlein, F. Kreupl, M. Liebau, R. Martin, B. Rajesekharan, W. Palmer, R. Seidel, W. Steinhoegl, and E. Unger, *Appl. Phys. A* **80**, 1141 (2005).
31. H. T. Ng, J. Han, T. Yamada, P. Nguyen, Y. P. Chen, and M. Meyyappan, *Nano Lett.* **4**, 1247 (2004).
32. A. Bertoni, P. Bordone, R. Brunetti, C. Jacoboni, and S. Reggiani, *Phys. Rev. Lett.* **84**, 5912 (2000).
33. A. Bertoni, R. Ionicioiu, P. Zanardi, F. Rossi, and C. Jacoboni, *Physica B: Cond. Matter* **314**, 10 (2002).
34. A. Bertoni and S. Reggiani, *Semicond. Sci. Technol.* **19**, S113 (2004).
35. A. E. Popescu and R. Ionicioiu, *Phys. Rev. B* **69**, 245422 (2004).

POLARONIC EFFECTS AT THE FIELD EFFECT JUNCTIONS FOR UNCONVENTIONAL SEMICONDUCTORS

NATASHA KIROVA

*Laboratoire de Physique des Solides C.NRS, UMR 8502, Université Paris-Sud, Orsay, F-91405, France
kirova@lps.u-psud.fr*

We consider properties of junctions for the FET geometry where molecular crystals or conducting polymers are used as semiconducting layers. In the molecular crystal Coulomb interaction of free electrons with surface polar phonons of the dielectric layer can lead to selftrapping of carriers and to the formation of a strongly coupled long-range surface polaron. The effect is further enhanced at presence of the bias electric field. The pronounced polaronic effects in conducting polymers change drastically the contact properties of these materials with respect to traditional semiconductors. Instead of the usual band bending near the contact interface, new allowed electronic bands appear inside the band gap. As a result the bias electric field and the injected charge penetrate into the polymer via creation of the soliton lattice which period changes with the distance from the contact surface. The performed studies open the possibility to describe the stationary characteristics and the hysteresis of the FET junctions and the Schottky diodes as well as to explain the photoluminescence suppression or enhancement under the bias electric field.

Keywords: FET; polaron; molecular crystal; polymer.

1. Introduction

Novel synthetic conductors possess a unique possibility to vary drastically their electronic properties depending on the external parameters such as pressure, magnetic and electric fields, temperature, etc. One of the distinctive features of these materials is a strong anisotropy of their electronic properties that is caused by the anisotropy of the transfer integrals. As a result, the electron system confined in a bulk lattice demonstrates properties inherent in one-, two-, or three-dimensional systems. Besides, the organic compounds offer a unique possibility of continuously tuning the dimensionality of the electronic system, which results in a variety of novel phases. Increasing experimental activity is devoted to unconventional semiconductors: transition metal oxides and chalcogenides, molecular crystals, conjugated polymers. A new experimental dimension comes from the possibility to change carrier concentration under the applied gate at high electric field near the junction interface of the field effect transistors (FET).

We suggest theoretical considerations for conditions to maintain a high volume concentration of the injected charge near the junction interface, the active role of the gate dielectrics, the effects of electron-phonon coupling at the junction. Interaction of an injected electron in the semiconducting layer with specific surface phonon modes results in formation of a polaron, located near the interface between the

semiconductor and the gate dielectrics,^{1,2} such an interface polaron can exist already for non-polar semiconductors like molecular crystals, and it will be ultimately present in traditional oxides like SrTiO₃. The polaron formation is endorsed by the bias electric field. The existence of polarons shows up in enhanced effective mass, mid-gap states and in pseudo-gap regime in tunneling experiments. The mobility of the FET is not only the property of the active semiconductor, but it intrinsically depends on the gate dielectric interface.

The pronounced polaronic effects in conducting polymers^{3,4} change drastically the contact properties of these materials with respect to traditional semiconductors. Instead of the usual band banding near the contact interface, new allowed electronic bands appear inside the band gap. As a result the bias electric field and the injected charge penetrate into the polymer via creation of a soliton lattice, which period changes with the distance from the contact surface. This results in the branching of the band structure. New narrow allowed bands grow inside the original gap, also expanding in their turn. Single electron carriers (polarons) are pulled to the contact area forming induced surface states. In time resolved experiments, e.g. in optically assistant junction formation, the charge injection goes via two steps: (i) fast dynamic process: charge injection and polaron formation; (ii) slow kinetic process: majority carriers - polarons collide and are absorbed into the ground state - providing one more period in the soliton lattice. The minority carriers (polarons of opposite sign) recombine with preexisting solitons reducing their number. The depletion layer is formed via reduction of soliton concentration.

2. Junction with Isotropic Semiconductor

Our goal is to have the dielectrics - metal transition induced by the gate electric field. For this reason we have to create a surface layer of electrons with high enough density to overcome the Coulomb interactions. We need to know the distribution of the field E , potential Φ and the concentration of carriers; the depth l of the distribution, their values at the surface.

Let the junction surface is $\mathbf{r} = (x,y)$ plane and the semiconductor occupies the semi-space $z > 0$. We can write the free energy functional as follows:

$$L = \int_{-\infty}^{\infty} dx dy \int_0^z dz \left[w(n) - \mu_0 n + en\Phi - \frac{\varepsilon}{8\pi} \left(\frac{d\Phi}{dz} \right)^2 \right]. \quad (1)$$

Here n and $w(n)$ are the density and the energy of the injected charge carriers correspondingly, Φ is the electric field potential, ε is the dielectric susceptibility of the media, $\mu_0 = \text{const}$ is the chemical potential of the carriers outside of our semiconductor, at $z < 0$. The corresponding variational equations have the form:

$$\begin{aligned} \frac{\partial w}{\partial n} - \mu_0 + e\Phi - \frac{1}{8\pi} \left(\frac{d\Phi}{dz} \right)^2 \frac{\partial \varepsilon}{\partial n} &= 0, \\ \frac{1}{4\pi} \frac{\partial}{\partial z} \left(\varepsilon \frac{d\Phi}{dz} \right) + en &= 0. \end{aligned} \quad (2)$$

If ε doesn't depend on the charge density, we obtain:

$$e\Phi = \mu_0 - \mu, \quad \mu = \frac{\partial w}{\partial n},$$

$$\Omega(\mu) + \varepsilon \left(\frac{d\Phi}{dz} \right)^2 = \Omega(\mu_\infty) = \text{const.} \quad (3)$$

Here $\Omega(\mu) = w(n) - \mu n$ is the Gibbs thermodynamic potential, $\mu(n)$ is the local chemical potential. The injected charge is characterized by the constant electrochemical potential $\mu + e\Phi$. We are interested in the case when the gate voltage is completely screened by the injected charge, and no electric field penetrates to the bulk. Then the boundary conditions can be written as:

$$z = 0: \quad \Phi = 0, \quad \mu = \mu_0 \quad (4)$$

$$z = \infty: \quad \Phi = \Phi_\infty, \quad \mu = \mu_\infty. \quad (5)$$

For the isotropic 3D degenerate electron gas⁵

$$n = \frac{(2m\mu)^{3/2}}{3\pi^2\hbar^3}; \quad w = \frac{(2m)^{3/2}\mu^{5/2}}{5\pi^2\hbar^3}, \quad (6)$$

and we obtain the solution of the Eq. (3) with the boundary conditions (4) and (5) for the distribution of the electric potential:

$$e\Phi = \mu_0 \left(1 - \frac{1}{(1+z/l)^4} \right). \quad (7)$$

Where the characteristic length l is determined by

$$l^2 = \frac{15\pi\varepsilon\hbar^3}{2me^2\sqrt{2m\mu_0}} = a_B^2 \frac{15\pi}{2} \sqrt{\frac{Ry}{\mu_0}}. \quad (8)$$

The corresponding plot is presented at Fig.1.

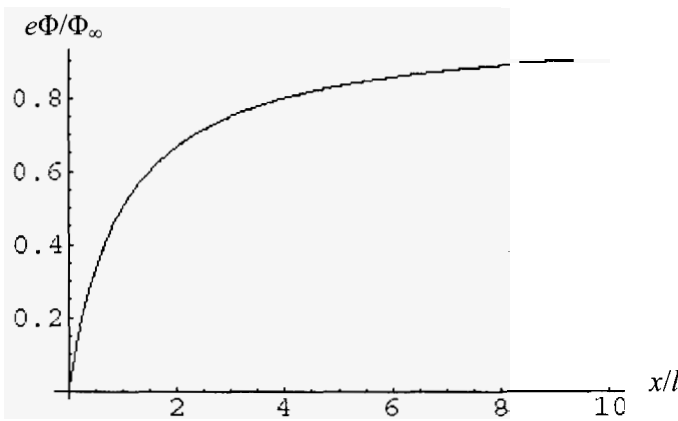


Fig.1. Electric potential distribution for a junction with an isotropic semiconductor.

The highest possible electric displacement and hence the injected charge on the junction is determined by the break down electric field E_g and the dielectric susceptibility ϵ_g of the gate dielectrics. The boundary condition at the interface gives $E(0) = E_g \epsilon_g / \epsilon$, and we can rewrite Eq. (8) as follows:

$$l = a_B \epsilon^{3/5} (15\pi)^{2/5} \left(\frac{m_e}{m} \right)^{3/5} \left(\frac{Ry}{2ea_B \epsilon_g E_g} \right)^{1/5} \quad (9)$$

The concentration of the electric charge per unit cell volume a_0^3 is:

$$\nu = \frac{D(0)a_0^3}{4\pi l} = \frac{a_0^3}{a_B^3} \frac{1}{\epsilon^{3/5}} \left(\frac{m}{m_e} \right)^{3/5} \left(\frac{ea_B \epsilon_g E_g}{Ry} \right)^{6/5}. \quad (10)$$

Below in Table 1 we present the breakdown fields, dielectric susceptibilities and estimated surface charge densities for some traditional gate materials. The data are taken from Ref. 6.

Table 1. Dielectric properties of various gate dielectrics

Dielectrics	E_g $10^6 V / cm$	ϵ_g	$\epsilon_g E_g$ $10^7 V / cm$	n $10^{13} cm^{-2}$
SiO_2	6	4-10	2.4-6	1.3 - 3.3
Al_2O_3	3	10-11	4 -4.5	2.2 - 2.5
Y_2O_3	5	9-13	4.5 - 6.5	2.5-3.6
HfO_2	50	15	75	41.4

Notice, that to avoid a Wigner crystallization, the picture of free electron gas requires rather high electron concentration.⁷ And this leads us to polar dielectrics. But in this case we face another problem: the formation of the surface long range polaron also in the case of the junction with non-polar semiconductor.

3. Surface Long Range Polarons in Molecular Crystals

The influence of surface phonons on the electron polaronic state inside the ionic crystal has been studied long time ago.⁸ Here we address the opposite situation when the electron resides in the non-polar media (e.g. a molecular crystal)^{1,2} and the polaron is formed by the interaction with the surface phonons of the polar dielectrics.

The electron is confined within the molecular crystal $z > 0$, its wave function is distributed near the interface. The only dipole active excitations of the polar dielectrics, to which the electron is coupled, are the surface phonon modes. Following the general scheme of Ref. 9, the electron-phonon interaction can be divided into two parts. The first one comes from the interaction with high frequency phonons which is reduced to the classical limit of the image charge potential. Another part comes from the interaction

with low frequency phonons and it results in polaron formation. The total energy of such a surface polaron is $W_p = W_0 w$, where $W_0 = Ry(m/m_e)$:

$$w = \int dx dy dz \left[-a_B^2 \psi^* \Delta \psi - \frac{a_B}{2\epsilon_{\infty}^*} \psi^* \frac{1}{z} \psi - \frac{eE}{W_0} \psi^* z \psi \right] - \frac{a_B}{\epsilon_0^*} \int dq \left[\int_0^{\infty} \rho_q(z) e^{-qz} dz \right]. \quad (11)$$

Here

$$\epsilon_{\infty,0}^* = \epsilon \frac{\epsilon_{\infty,0} + \epsilon_b}{\epsilon_{\infty,0} - \epsilon_b}, \quad (12)$$

m_e is the bare electron mass, m is the effective band mass, ϵ_b is the dielectric susceptibility of the molecular crystal, ϵ_{∞} and ϵ_0 are the high and the low frequency dielectric susceptibilities of the gate dielectric and $\rho_q(z)$ is the in-plane Fourier component of the charge density, $\mathbf{q} = (q_x, q_y)$, $\mathbf{r} = (x, y)$

$$\rho_q(z) = \int e^{-i\mathbf{q}\mathbf{r}} \psi(\mathbf{r}, z) \psi^*(\mathbf{r}, z) d\mathbf{r}. \quad (13)$$

The polaron effective mass can be estimated as:

$$\frac{M_p}{m} = \frac{2Ry}{\hbar\omega_s} \frac{a_B^3}{\epsilon_0^*} \left(\frac{m_e}{m} \right)^2 \int_0^{\infty} q^2 dq \left[\int_0^{\infty} \rho_q(z) e^{-qz} dz \right]. \quad (14)$$

For numerical estimates we have considered the interface between Al_2O_3 and pentacene. At zero bias electric field the polaron wave function has comparable localization length in all three directions. The polaron energy W_p is rather small and the polaron mass $M_p \sim 1.1 m$ is only weakly enhanced. At presence of the bias field W_p and M_p increase rather slowly up to the $E = 10^5$ V/cm but then, at higher fields, there is the strong enhancement of both^{1,2} (see Figs. 2 and 3).

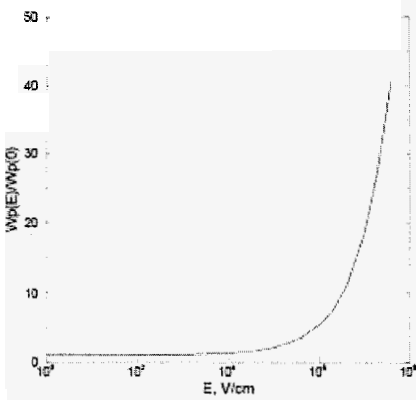


Fig.2. Energy for the surface polaron normalized on its value at zero bias electric field.

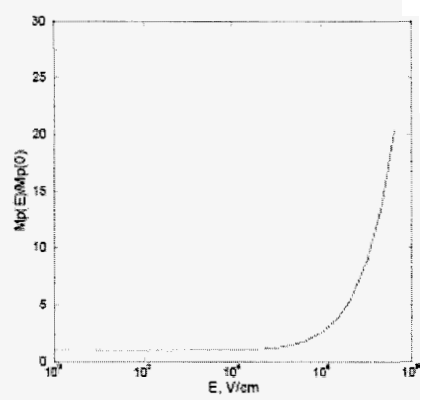


Fig.3. Effective mass for the surface polaron normalized on its value at zero bias electric field.

The wave function stays almost unchanged in the junction plane. At the same time, it is progressively squeezed in the perpendicular direction. So the gate electric field makes the polaron wave function distribution more two-dimensional. Notice the existence of the crossover bias electric field when the polaron energy and effective mass start drastically

to increase. Experimentally it should manifest itself in a decrease of the charge carrier mobility and in changes of the mobility regime from the band to the hopping one. The effect should be more pronounced in polar semiconducting oxides like SrTiO₃.

The polaron effective mass [Eq. (14)] and hence the measured mobility of the charge carriers in FET geometry strongly depends on the dielectric properties of the gate dielectrics. At Fig. 4 we present the comparison of the model predictions with the experimental data from Ref. 10. It should be mentioned that in spite we didn't take into account the variation of the surface phonon energies for various dielectrics, the agreement between theoretical curve and experimental points looks quite reasonable.

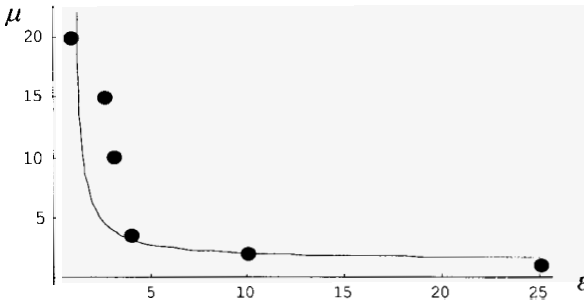


Fig. 4. Single parameter fit of presented theory (solid line) for mobility as a function of the dielectric susceptibility of the gate dielectrics. The experimental points are taken from Ref. 10.

4. Junction with Conducting Polymer

In common semiconductors the band bending near the metal-semiconductor interface brings the chemical potential of the semiconductor to the metal Fermi level. But in conducting polymers the deep self-trapping of free charge carriers results in the formation of soliton (bipolaron) lattice. This originates the band branching in addition to the band bending.^{10,11} The new allowed band grows inside the original gap, which also expands in its turn. The equilibrium between the metal and the polymer is defined now with respect to pairs of particles forming a lattice of solitons (bipolarons), rather than with respect to the single electrons. The single-electron levels are not matched at the junction interface. The tunneling injection between electronic levels occurs instantaneously at a given lattice configuration, hence the activation energy will be required. To estimate the possible charge carriers concentration and the electric field near the interface we have to use the self-consistent theory of screening by soliton (bipolaron) lattice.¹⁰ As before, we deal with the Eqs. (2), but now the injected charge carriers cannot be described as a Fermi electron gas. Instead for the carriers energy density $w(n)$ we have to use the expression derived for the periodic lattice of solitons¹²

$$w(n) = \frac{\Delta}{\pi \xi_0} \left(\frac{1}{2} - \frac{1}{r^2} + \frac{2}{r^2} \frac{E(r)}{K(r)} \right). \quad (15)$$

Here Δ is the gap in the undoped polymer, ξ_0 is the soliton size. $K(r)$, $E(r)$ are complete elliptic integrals of the first and the second kind,¹³ $0 < r < 1$. The electronic band boundaries are given as:

$$E_+ = \frac{\Delta}{r}, \quad E_- = \frac{\Delta}{r} \sqrt{1-r^2}. \quad (16)$$

The parameter r is related to the soliton density n as:

$$n(r) = \frac{1}{\xi_0} \frac{1}{rK(r)}. \quad (17)$$

The energy of the single electron excitation (polaron level) is defined as:

$$\frac{E_r}{\mu_0} = \frac{4}{\pi} K(r') \left[Z\left(\frac{\pi}{4}, r\right) - \frac{r^2}{\sqrt{2(2-r^2)}} \right] + \frac{8}{\pi} K(r') \Lambda_0\left(\frac{\pi}{4}, r'\right). \quad (18)$$

Z, Λ_0 are the Zeta function of Jacobi and Lambda function of Heuman correspondingly.¹³ And we obtain in parametric form:

$$\mu(r) = 2 \frac{\Delta}{r} \frac{E(r)}{r}; \quad z(r) = \sqrt{\frac{\varepsilon a^2 \hbar v_F}{2\pi^2 e^2}} \int_{r_0}^r \frac{K(r) dr}{r \sqrt{1-r^2}}; \quad e\Phi(r) = \mu_0 - \frac{2\Delta}{\pi} \frac{E(r)}{r}. \quad (19)$$

The corresponding band structure is presented at Fig. 5.

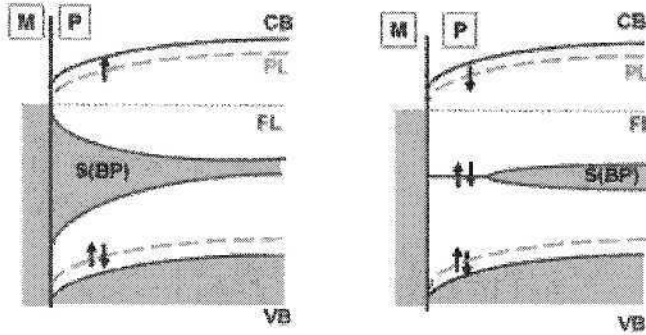


Fig. 5. Band structure near the metal-polymer junction for major (left) and minor (right) carriers' injection. Here M and P indicate the area of metal and polymer, VB and CB are the valence and conducting bands, S(BP) denotes the new soliton (bipolaron) band, FL- Fermi energy of the metal, PL - polaronic level.

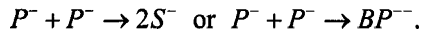
We can determine the injected charge surface density a , the junction area capacitance C and the penetration depth l :

$$a = e \frac{\sqrt{\varepsilon s}}{2\sqrt{2}\pi a_B} \frac{\Delta}{Ry} \frac{\sqrt{1-r_0^2}}{r_0}, \quad C = \frac{\sqrt{\pi \varepsilon s}}{2\sqrt{2}} \frac{\sqrt{1-r_0^2}}{E(r_0) - r_0}, \quad l = \sqrt{\frac{\varepsilon s \xi_0 \Delta}{4\pi^2 a_B Ry}} \int_{r_0}^1 \frac{K(r) dr}{r \sqrt{1-r^2}}.$$

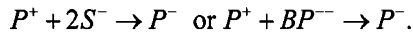
Here s is the area per chain in the perpendicular plane, e is the dielectric susceptibility (notice, that $e = \varepsilon_{\perp}$), a_B is the Bohr radius and $Ry = 13.6$ eV. Depending on the metal and polymer, we estimate $l \sim 20-60$ Å, $C \sim 10$ pF, and the built-in contact electric field as $E \sim 10^5 - 10^6$ V/cm.

The detailed analysis gives us that single-electron excitations created by charge injection or by interband absorption are attracted to the junction surface (see the polaron level at Fig. 3). The polymer may not have the surface states like those coming from dangling bonds in traditional covalent semiconductors, but the charge injection produces these states selfconsistently. Being injected, the majority carriers require an additional

energy to escape into the volume. Otherwise, after pair collisions they disappear in the ground state providing one more period in the soliton (bipolaron) lattice:



The minority carriers reverse their charge via recombination with preexisting solitons (bipolarons) reducing their number:



The depletion layer will be formed via reduction of the numbers of bipolarons (solitons) as indicated at Fig. 3, left.

5. Conclusions

There are special reservations on the physical processes at the junction. Thus the tempting choice of a gate material with a high dielectric constant brings to life the interface polaronic effects which can take place even semiconductors are non polar. Coulomb interaction of free electrons in molecular crystals with polar phonons of the gate dielectric results in selftrapping of carriers and in formation of the surface polaron. A bias electric field drastically enhanced the effect. Even for its typical values $\sim 10^6$ eV/cm, it profoundly stabilizes the surface long range polarons. The existence of polarons will show up in enhanced effective mass, mid-gap states and in pseudo-gap regime in case of tunneling experiments.

The mobility measured in organic FET is not only the property of specific organic molecule used, but it intrinsically depends on the organic/gate dielectric interface. Following special effects take place for the metal-polymer junction:

- band branching instead of usual band bending;
- non homogeneous soliton (bipolaron) lattice formation;
- transformation of the injected minority carriers into the majority ones;
- photoluminescence suppression for the LED.

References

1. N. Kirova and M.-N. Bussac, *Phys. Rev. B* **68**, 235312 (2003).
2. N. Kirova, *Curr. Appl. Phys.* **6**, 97 (2006).
3. Yu Lu, *Solitons and Polarons in Conducting Polymers* (World Scientific Publishing, 1988).
4. S. Brazovskii and N. Kirova, *Sov. Sci. Rev., Sec. A, Phys. Rev.*, I. M. Khalatnikov Ed., **5**, 99 (1984) (Harwood Acad. Publ.).
5. L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon Press 1963).
6. K. F. Young and H. P. R. Frederikse, *J. Phys. Chem. Ref. Data* **2**, 313 (1973).
7. B. Spivak, *Journ. de Physique IV* **114**, 337 (2004).
8. J. Sak, *Phys. Rev. B* **6**, 3981 (1972).
9. S. I. Pekar, *Sov. Phys. JETP* **16**, 333 (1946).
10. S. Brazovskii and N. Kirova, *Synth. Met.* **55-57**, 1254 (1993).
11. S. Brazovskii and N. Kirova, *Synth. Met.* **77**, 229 (1996).
12. S. Brazovskii, I. Dzyaloshinskii, and N. Kirova, *Sov. Phys. JETP* **54**, 120 (1982).
13. M. Abramovitz and I.A. Stegun, *Handbook of Mathematical Functions* (National bureau of standarts, New York 1965).

CELLULAR MONTE CARLO SIMULATION OF HIGH FIELD TRANSPORT IN SEMICONDUCTOR DEVICES

STEPHEN M. GOODNICK

*Department of Electrical Engineering, Arizona State University, P.O. Box 875706
Tempe, AZ 85287-5706, USA
Stephen.goodnick@asu.edu*

MARCO SARANITI

*Department of Electrical and Computer Engineering, 3301 S. Dearborn Street,
Chicago, IL 60616-3793, USA
Saraniti@iit.edu*

Here we discuss the use of the Cellular Monte Carlo (CMC) method for full band simulation of semiconductor transport and device modeling. The electronic band structure and phonon spectra are used as direct inputs to the program for both cubic, hexagonal, and strained crystal structures using both empirical and ab initio methods. As a particular example, this method is applied to study high field transport in GaN and GaN/AlGaIn heterostructures, where good agreement is obtained between the simulated results, and experimental pulse I-V measurements of transport. For device simulation, the CMC algorithm is coupled to an efficient 2D/3D multi-grid Poisson solver. We discuss the application of this algorithm to several technological problems of interest, including ultra-short channel Si/Ge MOSFETs, III-V compound HEMTs, and AlGaIn/GaN HEMTs.

Keywords: Monte Carlo; semiconductors; transport.

1. Introduction

Particle simulation based on the Ensemble Monte Carlo (EMC) method has been used now for over 30 years as a numerical method to simulate nonequilibrium transport in semiconductor materials and device.¹ In application to semiconductor transport problems, a random walk is generated for each particle of the ensemble, consisting of free flights under the influence of local forces acting on the particle, terminated by instantaneous, random scattering events due to scattering processes such as phonons, impurities, etc. This stochastic simulation algorithm may be shown to be equivalent to an exact solution of the semi-classical Boltzmann Transport Equation (BTE), and therefore fully accounts for non-stationary effects that are important for short channel devices when coupled with the appropriate set of field equations (e.g. Poisson's equation). The popularity of the method is based on both physical and numerical considerations. Sophisticated physical models may be implemented without affecting the algorithmic stability of the method, and the microscopic picture of charge motion is attractive in terms of the physical insight

given into device behavior. In addition, many body interactions beyond the simple one-electron BTE may be incorporated through for example molecular dynamics treatment of the particle-particle interaction. Quantum effects are being increasingly incorporated into the technique, including phenomena such as tunneling, quantization of motion due to reduce dimensionality, effective or quantum potentials, collision broadening, finite collision duration effects and even full quantum Monte Carlo algorithms.¹

One of the main drawbacks of EMC methods is the computational overhead associated with particle based methods compared to more approximate solutions to the BTE based on the drift-diffusion or hydrodynamic models. The computational burden continues to limit its use in commercial device modeling, even with the evolution of faster computers, particularly when sophisticated (and computationally demanding) physical models are implemented. A typical example of the model evolution is the increased use of fullband representation of the electronic **structure**,^{2,3} where the simple analytic representation of the band structure is replaced by a much more realistic (but computationally expensive) model based on, for example, the empirical pseudopotential method (EPM).⁴ New algorithmic approaches are clearly needed both to increase the modeling capabilities of these methods, and to reduce the simulation time so making them suitable as design tools.

The so-called fullband cellular Monte Carlo (CMC) method was developed to alleviate the computational burden of full band particle-based simulation without imposing severe physical approximations.⁵ The CMC model is based on a nonuniform discretization of the first Brillouin Zone (BZ) of the crystal lattice. The CMC approach allows the tabulation of the transition probability between all initial and final states on the mesh, greatly simplifying the final state selection of the conventional fullband EMC algorithm, which usually involves a search for final states in the BZ, which can be numerically intensive.

In the present paper, we discuss the application of the CMC method to several device technologies of current interest. After first discussing some details of the CMC device simulator implementation in Sec. 2, we discuss the application of this algorithm to several technological problems of interest, including ultra-short channel Si/Ge MOSFETs, and AlGaIn/GaN HEMTs. For heterostructure systems, we account for quantization of motion using the effective potential method. Good agreement is obtained in comparing to various experimental results for the DC current-voltage characteristics. We also investigate the high frequency performance of such devices using pulse and sinusoidal excitation.

2. Cellular Monte Carlo Method

As mentioned in the introduction, the fullband CMC method is basically a discrete k -space variant of a traditional EMC simulation that evolved from work on Cellular Automaton approaches to charge transport.⁶ The main idea in the fullband CMC is that the entire first Brillouin Zone (BZ) of the semiconductor(s) (or its irreducible wedge) is discretized on a non-uniform mesh in k -space, and that the total transition rate for every

pair of initial and final k values is **pre-tabulated** in a large look-up table. As illustrated in Fig. 1, the final state for scattering is chosen by the selection of one random number for the final state, rather than the conventional EMC algorithm of first selecting the type of scattering, then choosing the final state after scatter with a search throughout the entire BZ. Efficient algorithms to perform the latter have been developed in the literature,⁷ however clearly a lookup table is computationally much faster. The main limitation is the requirement for large memory (typically several gigabytes), in order to store all possible initial and final state transition rates, yet accurately represent the energy and momentum, particularly close to the principle band extremum.

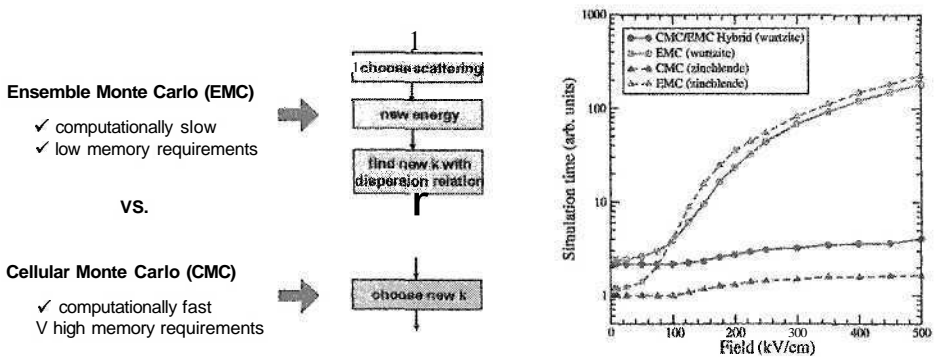


Fig. 1. Algorithmic differences of fullband EMC versus CMC particle simulation, and comparison of the simulation time for a k -space simulation of electron transport in GaN (both zincblende and wurtzite) for the CMC, EMC, and hybrid CMC/EMC methods.

The first step in the CMC algorithm is the calculation of the electronic states using an appropriate band structure model. While in principle fully *ab initio* methods may be used for the **bandstructure**, in practice semi-empirical methods are more typically employed such as the empirical pseudo-potential method (EPM) mentioned in the previous section.⁴ The electronic states are typically calculated in the irreducible wedge of the first BZ, and symmetry transformations used for equivalent states in other parts of the BZ. In order to calculate the electron-phonon scattering rates for acoustic and optical phonons, semi-empirical lattice dynamics methods are also used to tabulate the full phonon dispersion, using for example the Valence Shell Model.⁸ The full dispersion of the phonons is used, and two levels of model used for the electron-phonon rate, either a constant deformation potential model, or fully anisotropic rates calculated using the Rigid Ion Model (RIM).⁹ Impurity scattering and other defect type scattering are calculated from Fermi's Golden rule using the full electronic dispersion and overlap factors for the Bloch states.

In order to minimize the memory requirements in the CMC algorithm, several innovations have been introduced. One is the development of a hybrid algorithm, which uses the CMC over most of the critical region of k -space relevant for transport, but uses EMC close to the conduction band (CB) or valence band (VB) minimum or maximum, to obtain better accuracy of the final state energy, as well as at high energies where very few carriers exist.⁵ A comparison of the simulation time of a conventional fullband EMC

simulation (using a first order algorithm for the final state selection) and the CMC and hybrid CMC for high field transport in both wurzite and zincblende GaN is shown in Fig. 1. As can be seen, as the field increases, and carriers undergo increasing scattering, and the ratio of the EMC to CMC execution time increases, by as much as 100 times. More recent innovations in terms of memory usage have been introduced recently as well,¹⁰ where integer representation of the transition rates, and corresponding compression of the memory are achieved, at relatively little loss of fidelity.

3. High Field Transport

Using the tabulated scattering rates in across the first BZ, electronic transport is simulated in the usual way, in which a time-step is introduced, and the carrier motion is synchronized over every time step, and each particle checked as to whether scattering occurs. For simple k-space simulation of transport under a driving electric field, an ensemble is simulated, and the average velocity extracted after reaching steady state to calculate the velocity field characteristics.

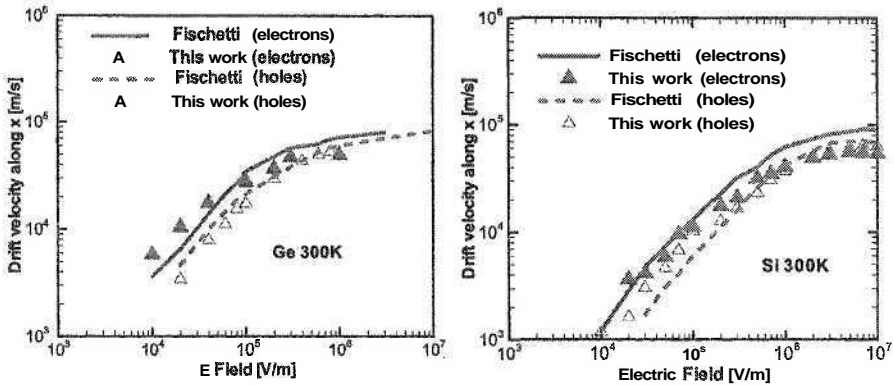


Fig. 2. Comparison of the simulation time for a k-space simulation of electron transport in GaN (both zincblende and wurzite) for the CMC, EMC, and hybrid CMC/EMC methods

For comparison with previous full-band EMC results using DAMOCLES³ (which was already benchmarked against available experimental velocity-field data), we first compare to the velocity field characteristics of both n-type and p-type Si and Ge, as shown in Fig. 2. Here the same EPM parameters and constant deformation potentials as Fischetti and Laux are used. A VSM for the phonon spectra is used. As seen in Fig. 2, the agreement between the CMC and the DAMOCLES results is generally good, and the basis for the SOI and GOI device simulation discussed in the next section.

Another material system that has been studied is wurzite phase GaN. The group III nitrides are of current interest for applications such as short wavelength optoelectronic and high power, high frequency electronic applications. These materials are generally wide-band gap materials, and can crystallize in both wurzite and zinc-blende polytypes, although the former is the dominant equilibrium structure.¹¹ The high field transport

properties of these materials are still relatively less well understood compared to cubic semiconductors such as Si and GaAs. Recent experimental studies of high field transport have indicated peak velocities in excess of $2\text{-}3 \times 10^7$ cm/s in bulk GaN and AlGaN/GaN heterostructures.^{12,13}

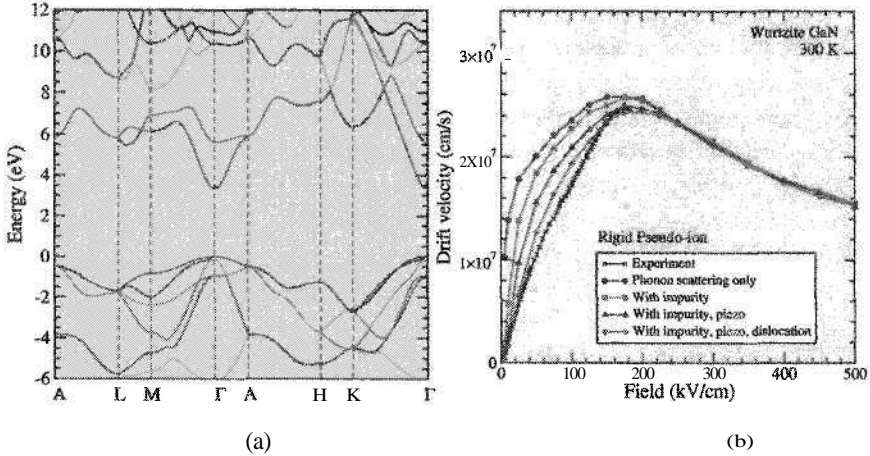


Fig. 3. (a) EPM bandstructure for wurtzite GaN, and (b) Calculated velocity field characteristics compared to experiment for bulk GaN including various scattering mechanisms.

The EPM bandstructure shown in Fig. 3(a) is calculated using local parameters from Brennan *et al.*¹⁴ The full phonon dispersion is calculated using both a Keating potential and valence-shell model, and full anisotropic electron-phonon deformation potential scattering is treated using the rigid ion model. Anisotropic polar optical phonon scattering (both LO and TO-like modes) for wurtzite GaN is also included in the simulation.¹⁵ Piezoelectric scattering is also known to be strong in nitride materials due to the lack of the inversion symmetry, which is included as well. Crystal dislocations are another important effect for GaN semiconductor devices. Elastic scattering rates due to threading dislocations, as well as ionized impurity scattering, are included in the transport model. Figure 3(b) shows the calculated velocity-field characteristics for bulk GaN at 300 K, with various mechanisms included, compared with the experimental pulsed I-V data of Barker *et al.*¹² While the peak velocity is mainly determined by phonon scattering (deformation, piezo- and polar mode scattering), the low field region is primarily dominated by elastic mechanisms including ionized impurity and dislocations.

4. CMC Device Modeling

In order to simulate semiconductor device behavior, the CMC transport kernel has been coupled to a 2D/3D multi-grid Poisson solver,¹⁶ which allows a faster simulation of different families of semiconductor devices with complex geometries and boundary conditions. Here we discuss two applications of this simulator of technological interest, one a comparison of the performance of Ge and Insulator (GOI) transistor technology

compared to Si on Insulator (SOI) technology, the second a study of the performance of AlGaN/GaN heterostructure field effect transistors (HFETs).

4.1. SOI and GOIMOSFETs

While Si SOI technology has become increasingly popular, scaling of Si technology below 90 nm requires new high- κ dielectric materials to replace SiO_2 in order to increase the effective capacitance and reduce tunneling. The replacement of SiO_2 with deposited dielectrics has spurred renewed interest in alternate materials to Si for nanoscale CMOS such as Ge, which has superior hole and electron mobilities to Si, as well as a much closer match of n-type and p-type properties.

Figure 4(a) shows a schematic of a generic SOI structure that has been simulated here to compare GOI and SOI performance. A high-K dielectric is assumed for the gate insulator with an equivalent SiO_2 oxide thickness of 0.8 nm and a dielectric constant of 21. The current voltage characteristics and high frequency scattering parameters for p-channel devices have been calculated and compared¹⁷ for both technologies using the CMC simulator for holes in Si and Ge, results of which were shown earlier in Fig. 2. Figure 4(b) shows the comparison of the calculated I_D - V_{DS} characteristics for 50 nm gate-length SOI and GOI p-channel devices. As can be seen, the GOI device leads to higher drive currents, which can be shown to be directly related to a higher channel velocity on average for Ge holes versus Si holes, due to the lower effective mass of holes in Ge.

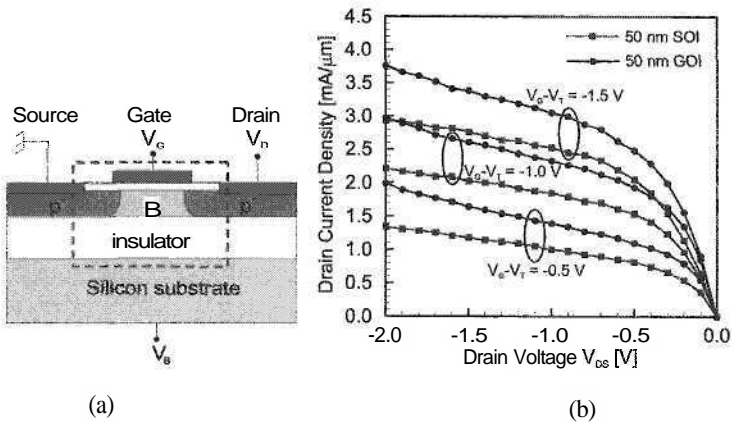


Fig. 4. (a) Simulated structure for both SOI and GOI devices, (b) simulated $I_{DS} - V_{DC}$ characteristics for different gate voltages for GOI (black) and SOI (dark gray) devices.

4.2. AlGaN/GaN HFETs

GaN/AlGaN heterostructure field-effect transistors (HFETs) have been demonstrated,¹⁸ which have shown promising performance as microwave power devices. In this heterostructure system, there is a difference in spontaneous polarization between AlGaN and GaN, in addition to the strain due to the difference of lattice constants, giving

rise to piezoelectric polarization. As a result, the discontinuity of the polarization at the interface results in a high density carrier without any intentional doping of the barrier layer.

Figure 5(a) shows a cross-section of a AlGaN/GaN HFET device, corresponding to the experimental microwave power device reported by Lee *et al.*¹⁹ Due to the strong polarization fields at the heterointerface in this system, strong carrier confinement and quasi-two-dimensional effects are important. To account for quantum effects semi-classically, the so-called ‘effective potential’ approach is used,²⁰ in which the effective potential, V_{eff} , is calculated as a convolution of a Hartree potential, V , obtained from Poisson’s equation, and a Gaussian function. In principle, the half-width of the Gaussian, a_0 , may be calculated from first principles. Here it is treated as a parameter to obtain the best least square fit between the charge density coming from the solution of the self-consistent 1D Schrödinger-Poisson equation, and the density using the effective potential. Figure 5(b) shows the comparison for several different gate voltages.

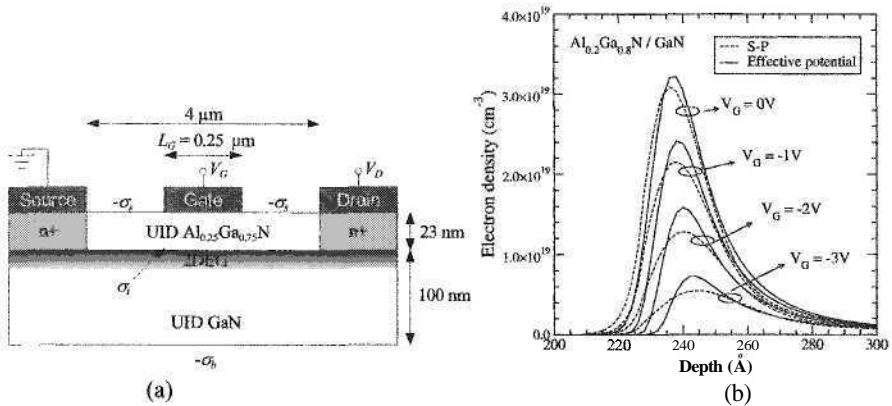


Fig. 5. (a) Schematic diagram of the simulated HFET structure, (b) Comparison of the interface charge density for various gate biases calculated self-consistently, and that calculated using the effective potential using a_0 as a fit parameter.

The solution to the 2D Poisson equation with the effective potential is solved at each time step, and the current-voltage characteristics simulated and compared with the experimental DC I-V characteristics and transconductance as shown in Fig 6. The experimental curves show some evidence of self-heating at high source-drain bias due to power dissipation; however, thermal simulation shows that the temperature rise in the channel is relatively weak for this particular structure. As can be seen, good agreement is obtained when the effective potential model is included, both in the I-V and transconductance versus gate voltage.

In order to calculate the unity gain frequency and unilateral power gain for this device, an AC excitation is superimposed on the DC bias, and the corresponding response at the device output (from a two-port model standpoint) is extracted, as illustrated in Fig. 7. The AC excitation can either be a step function, from which the frequency

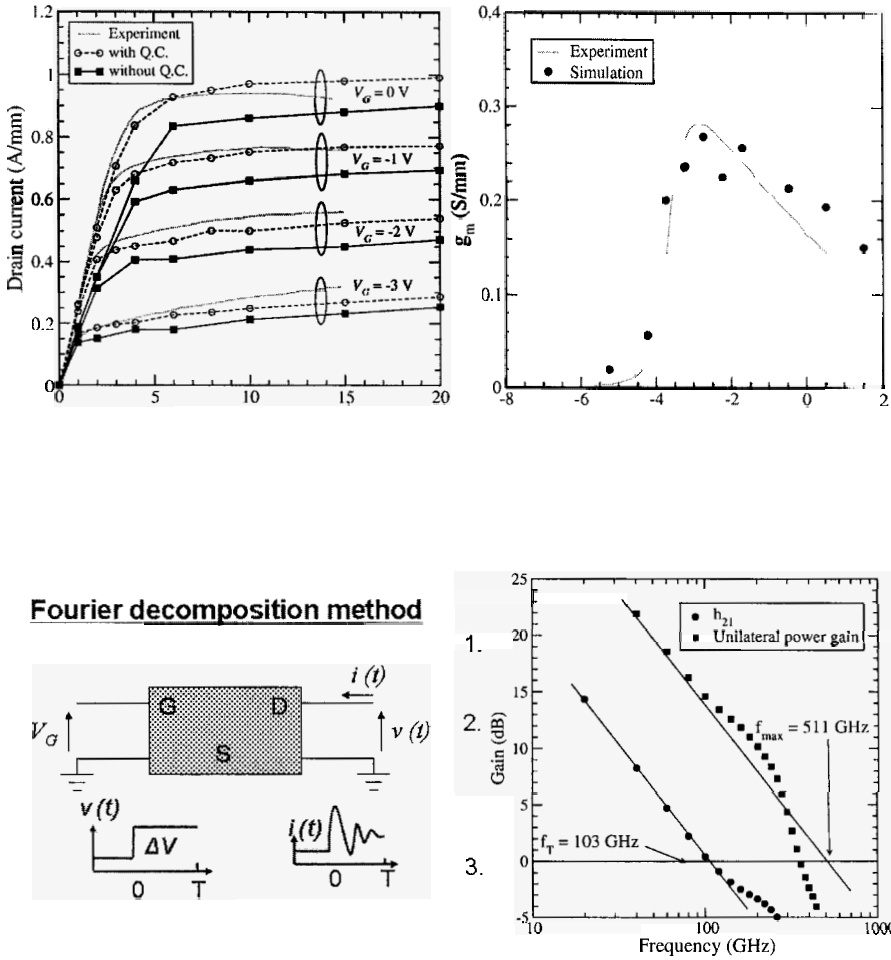


Fig. 7. Analysis of high frequency performance in device simulation and the simulated short-circuit and unilateral power gain versus frequency.

The calculated short circuit current gain, and unilateral power gain are shown in Fig. 7, where $f_T = 103$ GHz is predicted, where as the experimental results are lower (65 GHz for f_T for example). Some of this difference is due to simulation of only the active part of the device, whereas the extrinsic parasitics will limit the performance even more, and must be extracted from the experimental structure. However, due to the large density and high fields present at the interface, nonequilibrium phonon effects may limit transport as well,²¹ which reduces the peak carrier velocity, and subsequently reduces the frequency response. Such effects are the topic of future investigations using the CMC code.

Acknowledgments

The authors would like to thank the various persons who helped contribute to this work, including S. Aboud, J. Ayubi-Moak, S. Beysserie, J. Branlard, D. K. Ferry, and S. Yamakawa. The authors would also like to thank the support of the National Science Foundation and the DoD High Performance Computing Modernization Program.

References

1. C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer-Verlag, Vienna, 1989).
2. K. Hess, *Monte Carlo Device Simulation: Full Band and Beyond* (Kluwer Academic Publishing, Boston, 1991).
3. M. V. Fischetti and S. E. Laux, *Phys. Rev. B* **38**, 9721 (1988).
4. J. R. Chelikowsky and M. L. Cohen, *Phys. Rev. B* **14**, 556 (1976).
5. M. Saraniti and S. M. Goodnick, *IEEE Trans. Elec. Dev.* **47**, 1909 (2000).
6. K. Kometer, G. Zandler, and P. Vogl, *Phys. Rev. B* **46**, 1382 (1992).
7. J. Bude and R. K. Smith, *Semicond. Sci. Technol.* **9**, 840 (1994).
8. K. Kunc and O. H. Nielsen, *Computer Physics Communications* **17**, 413 (1979).
9. S. Zollner, S. Gopalan, and M. Cardona, *J. Appl. Phys.* **68**, 1682 (1990).
10. J. Branlard, *Ph.D. Dissertation* (2004).
11. S. J. Pearton, F. Ren, A. P. Zhang, and K. P. Lee, *Materials Science and Engineering* **R30**, 55 (2000).
12. J. M. Barker, D. K. Ferry, S. M. Goodnick, D. D. Koleske, A. E. Wickenden, and R. L. Henry, *Microelectronic Engr.* **63**, 193 (2002).
13. J. M. Barker, R. Akis, D. K. Ferry, S. M. Goodnick, T. J. Thornton, D. D. Koleske, A. E. Wickenden, and R. L. Henry, *Physica B* **314**, 39 (2002).
14. I. H. Oguzman, J. Kolník, K. F. Brennan, R. Wang, T. Fang, and P. P. Ruden, *J. Appl. Phys.* **80**, 4429 (1996).
15. C. Bulutay, B. K. Ridley, and N. A. Zakhleniuk, *Phys. Rev. B* **62**, 15754 (2000).
16. M. Saraniti, A. Rein, G. Zandler, P. Vogl, and P. Lugli, *IEEE Trans. CAD* **15**, 141 (1996).
17. S. Beysserie, J. Branlard, S. Aboud, Member, S. M. Goodnick, and M. Saraniti, *accepted for publication in IEEE Trans. Elec. Dev.* (2007).
18. J. P. Ibbetson, P. T. Fini, K. D. Ness, S. P. DenBaars, J. S. Speck, and U. K. Mishra, *Appl. Phys. Lett.* **77**, 250 (2000).
19. C. Lee, P. Saunier, J. Yang, and M. A. Khan, *IEEE Elect. Dev. Lett.* **24**, 616 (2003).
20. D. K. Ferry, *Superlatt. Microstruct.* **27**, 61 (2000).
21. L. Ardaravičius, A. Matulionis, J. Liberis, O. Kiprijanovic, M. Ramonas, L. F. Eastman, J. R. Shealy, and A. Vertiatchikh, *Appl. Phys. Lett.* **83**, 4038 (2003).

This page intentionally left blank

NANOELECTRONIC DEVICE SIMULATION BASED ON THE WIGNER FUNCTION FORMALISM

HANS KOSINA

*Vienna University of Technology, Institute for Microelectronics
Gusshausstrasse 27–29/E360, Vienna, A-1040, Austria
kosina@iue.tuwien.ac.at*

Coherent transport in mesoscopic devices is well described by the Schrödinger equation supplemented by open boundary conditions. When electronic devices are operated at room temperature, however, a realistic transport model needs to include carrier scattering. In this work the kinetic equation for the Wigner function is employed as a model for dissipative quantum transport. Carrier scattering is treated in an approximate manner through a Boltzmann collision operator. A Monte Carlo technique for the solution of this kinetic equation has been developed, based on an interpretation of the Wigner potential operator as a generation term for numerical particles. Including a multi-valley semiconductor model and a self-consistent iteration scheme, the described Monte Carlo simulator can be used for routine device simulations. Applications to single barrier and double barrier structures are presented. The limitations of the numerical Wigner function approach are discussed.

Keywords: Nanoelectronic devices, device simulation, Wigner function, kinetic equation, Monte Carlo method.

1. Introduction

For FETs with gate lengths below 10 nm quantum effects such as direct source-to-drain tunneling become important and start affecting the device characteristics.¹ Recent studies show that scattering will still affect the current² and that the transition to ballistic transport appears at much shorter gate lengths than previously anticipated.³ An accurate theory of MOSFETs near the scaling limit must therefore account for the interplay between coherent quantum effects and dissipative scattering effects. This mixed transport regime can suitably be treated by the Wigner equation. Early numerical solutions of the Wigner equation were obtained using finite difference methods, assuming simplified scattering models based on the relaxation time approximation.⁴ However, for realistic device simulation more comprehensive scattering models are required. With the advent of Monte Carlo (MC) methods for the Wigner equation^{5,6} it became feasible to include the full Boltzmann collision operator. The development of MC methods for the Wigner equation, however, is

hampered by the fact that, as opposed to the semi-classical case, the integral kernel is no longer positive. This so-called negative sign problem will lead to exponentially growing variances of the Markov Chain MC method. The Wigner potential operator can also be viewed as a generation term of positive and negative numerical particles. In this picture the sign problem shows up in the avalanche of numerical particles generated. A stable MC method can only be achieved by means of a suitable particle annihilation algorithm.

2. The Physical Model

Quantum transport is modeled by a time-independent, one-electron Wigner equation for a multi-valley semiconductor. The set of Wigner equations is coupled through the inter-valley phonon scattering terms:

$$\begin{aligned} \frac{1}{\hbar} \left(\nabla_{\mathbf{k}} \epsilon_v(\mathbf{k}) \cdot \nabla_{\mathbf{r}} + \mathbf{F}(\mathbf{r}) \cdot \nabla_{\mathbf{k}} \right) f_v(\mathbf{k}, \mathbf{r}) = & \\ & \sum_{v'} \int [1 - f_v^0(\mathbf{k}, \mathbf{r})] S_{vv'}(\mathbf{k}, \mathbf{k}') f_{v'}(\mathbf{k}', \mathbf{r}) d^3 k' \\ & - \left(\sum_{v'} \int [1 - f_{v'}^0(\mathbf{k}, \mathbf{r})] S_{v'v}(\mathbf{k}', \mathbf{k}) d^3 k' \right) f_v(\mathbf{k}, \mathbf{r}) \\ & + \int V_w(\mathbf{k} - \mathbf{k}', \mathbf{r}) f_v(\mathbf{k}', \mathbf{r}) d^3 k' \end{aligned} \quad (1)$$

Silicon: $v, v' = [100], [010], [001]$.

This equation determines the Wigner function f_v for valley v . A valley's energy dispersion relation $\epsilon_v(\mathbf{k})$ is assumed to be anisotropic and parabolic, resulting in a local diffusion term, $\nabla_{\mathbf{k}} \epsilon_v \cdot \nabla_{\mathbf{r}} f$. Note that a non-parabolic $\epsilon(\mathbf{k})$ relation in the single-electron Hamiltonian would give a non-local diffusion term of the form $\int \hat{\epsilon}(\mathbf{k}, \mathbf{r} - \mathbf{r}') f_v(\mathbf{k}, \mathbf{r}') d^3 r'$.

A spectral decomposition of the potential profile $V(\mathbf{r})$ is applied.⁷ The slowly varying component gives the classical force \mathbf{F} , whereas the rapidly varying component is taken into account through the Wigner potential V_w :

$$V(\mathbf{r}) = V_{\text{cl}}(\mathbf{r}) + V_{\text{qm}}(\mathbf{r}) \quad (2)$$

$$\mathbf{F}(\mathbf{r}) = -\nabla V_{\text{cl}}(\mathbf{r}) \quad (3)$$

$$V_w(\mathbf{q}, \mathbf{r}) = \frac{1}{i\hbar} \int \left[V_{\text{qm}}\left(\mathbf{r} + \frac{\mathbf{s}}{2}\right) - V_{\text{qm}}\left(\mathbf{r} - \frac{\mathbf{s}}{2}\right) \right] e^{-i\mathbf{q}\cdot\mathbf{r}} \frac{d^3 s}{(2\pi)^3}. \quad (4)$$

In Eq. (1) scattering is treated semi-classically through a Boltzmann collision operator, where the transition rate $S_{vv'}(\mathbf{k}, \mathbf{k}')$ from initial state (v', \mathbf{k}') to final state (v, \mathbf{k}) is given by Fermi's golden rule. It should be noted that usage of the Boltzmann collision operator in the Wigner equation represents some ad hoc assumption. A rigorous treatment of electron-phonon scattering would require a frequency-dependent Wigner function, $f(\mathbf{k}, \mathbf{r}, \omega)$. It is related to the non-equilibrium Green's

function $G^<$ by $G^<(\mathbf{r}, \mathbf{k}, \omega) = if(\mathbf{k}, \mathbf{r}, \omega)$ and can reasonably be approximated as⁸ $f(\mathbf{k}, \mathbf{r}, \omega) = f_w(\mathbf{k}, \mathbf{r})A(\mathbf{r}, \mathbf{k}, \omega)$. To arrive at Fermi's golden rule the spectral function A is reduced to the Dirac δ -function.

Furthermore, in Eq. (1) the Pauli blocking factor the equilibrium Fermi function f_v^0 is used. The assumption of a Boltzmann collision operator in Eq. (1) ensures that in the semiclassical regions, such as the highly doped contact regions, the conductivity is finite and that the mean energy increase due to degeneracy is taken into account.

3. Numerical Methods

To solve Eq. (1) numerically, a stationary MC method has been proposed, based on the interpretation of the potential operator $\Theta[f_w] = \int V_w(\mathbf{k} - \mathbf{k}')f_w(\mathbf{k}', \mathbf{r})d^3k'$ as a generation term of numerical particles.⁵ The mass conservation property of the potential operator can be exactly satisfied by the numerical particle model if one generates the numerical particles only pair-wise, for instance, with statistical weights +1 and -1. A suitable annihilation algorithm for numerical particles needs to be introduced in order to achieve a stable MC method. Since one can devise various algorithms for particle generation and, in particular, for particle annihilation, in the following the latest developments are described.

3.1. Particle Generation

A direct numerical representation of the Wigner potential $V_w(\mathbf{q}, \mathbf{r})$ would require the discretization of both momentum and space coordinates. The problem can be simplified by expressing the Wigner potential in terms of $\hat{V}(\mathbf{q})$, the Fourier transform of the potential $V_{qm}(\mathbf{r})$. The potential operator can be rewritten as follows:

$$\Theta_w[f_w](\mathbf{k}, \mathbf{r}) = \frac{1}{\hbar} \int |\hat{V}(\mathbf{q})| \sin[\varphi(\mathbf{q}) + \mathbf{q} \cdot \mathbf{r}] \times \left(f_w \left(\mathbf{k} - \frac{\mathbf{q}}{2}, \mathbf{r}, t \right) - f_w \left(\mathbf{k} + \frac{\mathbf{q}}{2}, \mathbf{r}, t \right) \right) \frac{d^3q}{(2\pi)^3}. \quad (5)$$

An advantage of this formulation is that no discretization of the spatial variable \mathbf{r} is needed. The expression can be evaluated at the actual position \mathbf{r} of a particle. Only the momentum variable \mathbf{q} needs to be discretized in order to numerically represent $|\hat{V}|$, the modulus, and φ , the phase of \hat{V} .

The structure of Eq. (5) suggests the usage of a rejection technique. As a normalization quantity one obtains an upper limit for the pair generation rate

$$\gamma_{\max} = \frac{1}{\hbar} \int |\hat{V}(\mathbf{q})| \frac{d^3q}{(2\pi)^3}. \quad (6)$$

At a rate of γ_{\max} the free flight of a particle is interrupted to check for particle pair-generation. From the distribution $|\hat{V}(\mathbf{q})|$ one generates randomly the momentum transfer \mathbf{q} . Then the sine function is evaluated at the actual particle position \mathbf{r} as

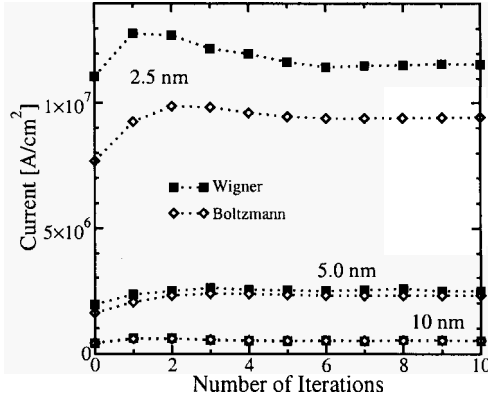


Fig. 1. Current through Si n - i - n diodes as a function of the number of self-consistent iterations with the width of the intrinsic region as a parameter.

$s = \sin[\varphi(\mathbf{q}) + \mathbf{q} \cdot \mathbf{r}]$. With probability $|s|$ the pair-generation event is accepted, otherwise a self-scattering event is performed. In the former case, two particle states are generated with momenta $\mathbf{k}_1 = \mathbf{k} - \mathbf{q}/2$ and $\mathbf{k}_2 = \mathbf{k} + \mathbf{q}/2$ and statistical weights $w_1 = w_0 \text{sign}(s)$ and $w_2 = -w_1$, respectively, where w_0 is the statistical weight of the initial particle. Since Eq. (5) is local in real space, the particle pair is generated at the position \mathbf{r} of the initial particle.

3.2. Particle Annihilation

Different variants of the single-particle Monte Carlo method can be devised.⁵ The variant discussed in Ref. 9 guarantees exact current conservation. The only input parameter required is the ratio of negative and positive trajectories, which makes the algorithm easy to control. The idea is that from the trajectory tree generated by a particle injected at the contact only one branch is actually traced. For steady state problems considered here a phase space mesh can be utilized, on which numerical particles are temporarily stored. An annihilation mesh is introduced for each valley-type. The meshes are defined in the three-dimensional phase-space, spanned by one spatial and two momentum coordinates.

3.3. Coupling to the Poisson Equation

A self-consistent iteration scheme between Wigner MC and the Poisson equation is implemented. The adopted scheme, which is similar to the Gummel iteration scheme for the basic semiconductor equations,¹⁰ is commonly used in classical one-particle MC simulations.¹¹ Figure 1 shows the iteration history of the current through a Si n - i - n diode for different widths of the intrinsic region. Currents computed using Wigner and classical MC show similar convergence behavior.

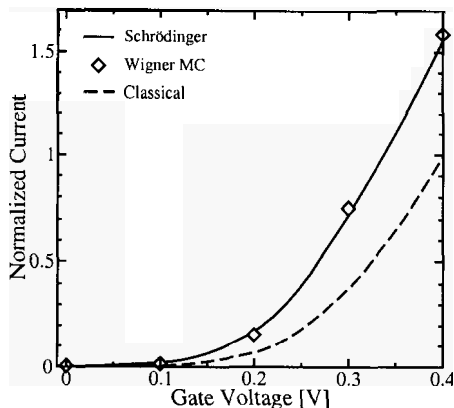


Fig. 2. Normalized ballistic currents calculated classically and quantum mechanically. Results from Wigner MC and the Schrödinger solver are in good agreement. The potential profile is obtained from a device simulation of a 10 nm gate length DG MOSFET.

4. Results

The described MC method can be used for routine device simulations. For the purpose of verification, the first example assumes a frozen potential profile from a 10 nm gate length double-gate MOSFET. Figure 2 compares the quantum ballistic currents as obtained from a collision-less Wigner MC simulation and from a numerical Schrödinger solver. Good agreement is observed. The quantum ballistic current is higher than the classical ballistic current due to an additional contribution from carriers tunneling through the potential barrier.

To study the effects of scattering and tunneling on the device characteristics we consider Si n - i - n diodes with the length W of the intrinsic region ranging from 20 nm down to 2.5 nm. The doping profile is assumed to increase gradually from the intrinsic region to the highly doped contact region over the same distance W . Three transport models are compared: Wigner equation and Boltzmann equation with electron-phonon and ionized-impurity scattering included, yielding currents I_{WIG} and I_{BTE} , respectively. The Wigner equation without scattering inside the intrinsic and transition regions gives the current I_{COH} (coherent). Figure 3 shows that the effect of scattering reflected in the difference $I_{\text{COH}} - I_{\text{WIG}}$ decreases with decreasing device length. However, even for $W = 2.5$ nm the relative difference in the currents is still of the order of 25%, indicating that scattering cannot be neglected. Also shown is the current difference due to tunneling, $I_{\text{WIG}} - I_{\text{BTE}}$. Clearly, this current component rises with reduced barrier width.

5. Discussion

When using the Wigner equation for numerical simulation one should be aware of several peculiarities inherent to that approach. This approach is useful for a certain class of problems only, as briefly outlined below.

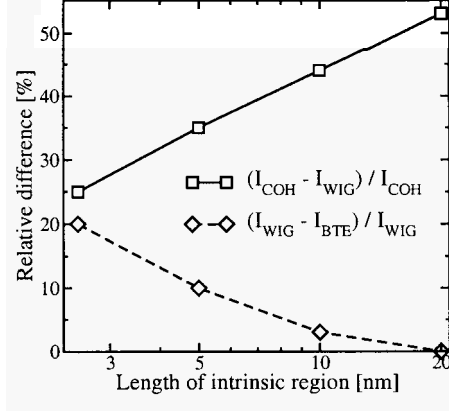


Fig. 3. Relative difference between currents of an n - i - n diode calculated using different transport models: Wigner MC with and without scattering in the intrinsic region (squares); Wigner MC and classical MC (diamonds).

5.1. Interpretation of the Results

In many cases the quantities of interest are certain moments of the phase space distribution and not the full distribution itself. As in the classical case, the carrier density is given by the zeroth order moment and the current density by the first order moment. The second order moment is an energy like quantity. Its physical interpretation, however, is not obvious in the Wigner representation, but can be appreciated in the wave function representation.

In terms of the density operator $\hat{\rho}(t)$ the carrier density is given by the diagonal element, $n(\mathbf{r}, t) = \langle \mathbf{r} | \hat{\rho}(t) | \mathbf{r} \rangle$, where $|\mathbf{r}\rangle$ is the single-particle position eigenstate. With the momentum operator $\hbar\hat{\mathbf{k}}$ the kinetic energy is related to the diagonal element of the operator $\hat{\mathbf{k}}^2\hat{\rho}$. This operator needs to be symmetrized to guarantee real valued diagonal elements. For the symmetrized operator $\hbar^2(\hat{\mathbf{k}}^2\hat{\rho} + \hat{\rho}\hat{\mathbf{k}}^2)/(2m^*)$ the diagonal element evaluate to¹²

$$w(\mathbf{r}) = \sum_i p_i (E_i - V(\mathbf{r})) |\Psi_i(\mathbf{r})|^2, \quad (7)$$

where ψ_i denote the wave functions, E_i the eigen-energies, and p_i the probabilities determining the mixed state of the system.

The energy density in Eq. (7) can become negative in tunneling regions where the energy of one or more states is below the band edge, $E_i < V(\mathbf{r})$. Transformation of Eq. (7) into the Wigner representation gives¹²

$$w(\mathbf{r}) = \int \frac{\hbar^2}{2m^*} \left(|\mathbf{k}|^2 - \frac{1}{4} \nabla_{\mathbf{r}}^2 \right) f_w(\mathbf{k}, \mathbf{r}, t) \frac{d^3k}{(2\pi)^3}. \quad (8)$$

Thus, the kinetic energy density is not simply given by the second moment of the

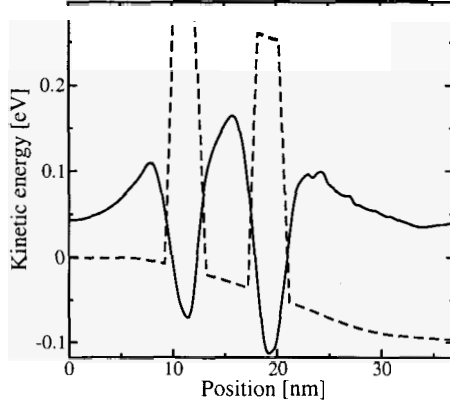


Fig. 4. Mean kinetic energy of electrons (solid line) in a resonant tunneling diode calculated from Eq. (8). In the tunneling barriers (dashed line) the mean kinetic energy is negative.

Wigner function. A correction proportional to the second derivative of the electron concentration is required. Figure 4 shows the kinetic energy density of electrons and the conduction band profile for a resonant tunneling diode.

5.2. The Bound-states Problem

For energy eigenstates, $\psi_n(\mathbf{r}, t) = \psi_n(\mathbf{r}, 0) \exp(-\epsilon_n t/\hbar)$ the density matrix is time-independent, $\rho_{nn}(\mathbf{r}_1, \mathbf{r}_2, t) = \psi_n^*(\mathbf{r}_1, 0) \psi_n(\mathbf{r}_2, 0)$. In this case the system Hamiltonian and the density operator commute, and the quantum Liouville equation reduces to

$$i\hbar \frac{\partial \hat{\rho}}{\partial t} = [\hat{H}, \hat{\rho}] = 0 \quad (9)$$

This trivial equation, which does not contain the system Hamiltonian any longer, cannot determine bound-state density matrix. However, any given bound-state density matrix is time-independent and will satisfy this equation.

Since the quantum Liouville equation and the Wigner equation are linked by the Wigner-Weyl transform, similar arguments hold for the Wigner equation. It has been shown in Ref. 13 that bound states cannot be obtained from the following Wigner equation for ballistic motion:

$$\left(\frac{\partial}{\partial t} + \frac{\hbar \mathbf{k}}{m^*} \cdot \nabla_{\mathbf{r}} \right) f_w(\mathbf{k}, \mathbf{r}, t) - \int V_w(\mathbf{k} - \mathbf{k}', \mathbf{r}) f_w(\mathbf{k}', \mathbf{r}, t) d^3 k' = 0 \quad (10)$$

$$V_w(\mathbf{q}, \mathbf{r}) = \frac{1}{i\hbar} \int \left\{ V\left(\mathbf{r} + \frac{\mathbf{s}}{2}\right) - V\left(\mathbf{r} - \frac{\mathbf{s}}{2}\right) \right\} e^{-i\mathbf{q} \cdot \mathbf{s}} \frac{d^3 s}{(2\pi)^3}.$$

The harmonic oscillator is an example clearly demonstrating this problem. If the potential is a quadratic function of position, $V(\mathbf{r}) = m^* \omega^2 |\mathbf{r}|^2 / 2$, the Wigner equation (10) reduces to the corresponding classical Liouville equation, with

$\mathbf{F}(\mathbf{r}) = -m^*\omega^2\mathbf{r}$ being the classical force:

$$\left(\frac{\partial}{\partial t} + \frac{\hbar\mathbf{k}}{m^*} \cdot \nabla_{\mathbf{r}} + \frac{\mathbf{F}(\mathbf{r})}{\hbar} \cdot \nabla_{\mathbf{k}} \right) f_w(\mathbf{k}, \mathbf{r}, t) = 0. \tag{11}$$

This equation propagates an initial distribution classically. Therefore, the single equation (10) is not completely equivalent to the Schrödinger equation. In the literature two independent solutions of this problem have been proposed.

Carruthers and Zachariassen¹³ start from the Schrödinger equation and derive an adjoint Wigner equation. If this adjoint equation is considered in addition, the usual Schrödinger eigenvalue problem can be reconstructed from the two Wigner equations. The adjoint equation is obtained from a density matrix equation involving the anti-commutator,¹³ $[\hat{H}, \hat{\rho}]_+ = 2\epsilon\hat{\rho}$, and takes the following form

$$\frac{\hbar^2}{2m^*} \left(|\mathbf{k}|^2 - \frac{1}{4}\nabla_{\mathbf{r}}^2 \right) f_{mn}(\mathbf{k}, \mathbf{r}) - \int \tilde{V}_w(\mathbf{k} - \mathbf{k}', \mathbf{r}) f_{mn}(\mathbf{k}', \mathbf{r}) d^3k' = \frac{\epsilon_m + \epsilon_n}{2} f_{mn}(\mathbf{k}, \mathbf{r}) \tag{12}$$

$$\tilde{V}_w(\mathbf{q}, \mathbf{r}) = \frac{1}{2i\hbar} \int \left\{ V\left(\mathbf{r} + \frac{\mathbf{s}}{2}\right) + V\left(\mathbf{r} - \frac{\mathbf{s}}{2}\right) \right\} e^{-i\mathbf{q}\cdot\mathbf{s}} \frac{d^3s}{(2\pi)^3}.$$

The ϵ_n are the eigenvalues of the Hamiltonian. For $m = n$ one obtains the bound-state Wigner functions, which are real valued. The case $m \neq n$ gives complex functions. The entire set of $f_{mn}(\mathbf{k}, \mathbf{r})$ form a complete orthonormal set for all Wigner functions.

Tatarskii¹⁴ takes as foundation the Wigner representation of quantum mechanics. The solutions of the Wigner equation have to be subjected to a necessary and sufficient condition which selects an allowed class of Wigner distributions describing quantum-mechanical pure states. The condition can be formulated in terms of the density matrix, obtained by the inverse Wigner-Weyl transform of the Wigner function, as follows:¹⁴

$$\nabla_{\mathbf{r}_1} \nabla_{\mathbf{r}_2} \ln \rho(\mathbf{r}_1, \mathbf{r}_2) = 0 \tag{13}$$

$$\rho(\mathbf{r}_1, \mathbf{r}_2) = \int f_w\left(\mathbf{k}, \frac{\mathbf{r}_1 + \mathbf{r}_2}{2}\right) e^{i\mathbf{k}\cdot(\mathbf{r}_1 - \mathbf{r}_2)} \frac{d^3k}{(2\pi)^3} \tag{14}$$

As opposed to the adjoint equation (12), this condition is a pure mathematical one and does not contain any information about the system Hamiltonian. When condition (13) is satisfied, one can reconstruct the wave function up to a phase factor from the Wigner function. Furthermore, it can be shown that from the Wigner equation (10), with an initial condition satisfying the restriction (13) on the allowable form of pure state Winger functions, the Schrödinger equation follows.¹⁴ In the case of the harmonic oscillator, the quantization condition for the energy does not follow from the Wigner equation, but from the supplementary condition.

The fact that the Wigner equation alone cannot provide the bound-states of a closed system has some implications on the numerical solution methods, especially

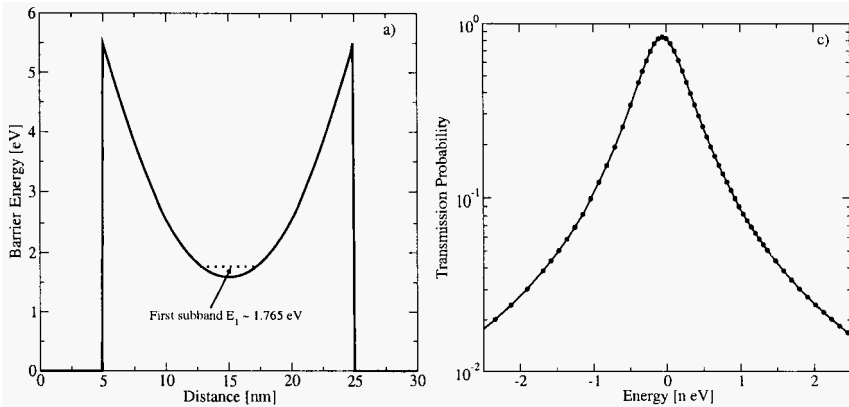


Fig. 5. Resonant tunneling structure with a quadratic potential well (a). The broadening of the first quasi-bound state is in the 10^{-9} eV range (b).

if the current through the system is determined by quasi-bound states of long life time. In this case the energy levels have very little broadening, which indicates that the system is almost closed. This situation is illustrated in Fig. 5 for the example of a resonant tunneling structure with a quadratic potential barrier. The spacing between resonance energies is constant for this potential. The first resonance peak of the transmission probability, calculated using a Schrödinger solver, is also shown in Fig. 5. To resolve this resonance a highly non-uniform energy grid with extremely small spacing around the resonance peaks is needed. The discrete Fourier transform utilized by a numerical Wigner equation solver, however, permits only equi-distant grids in momentum space. With such a grid the resonances for the above example cannot be resolved in practice, and the discrete Wigner equation would be ill-conditioned. From this discussion one can conclude that a numerical Wigner function approach is applicable only to sufficiently open systems, i.e., to systems with not too narrow resonances.

6. Conclusions

A Monte Carlo simulator performing a self-consistent numerical solution of the Wigner equation has been presented. Dissipation effects are included semi-classically through the Boltzmann collision operator. Algorithms for generation and annihilation of numerical particles have briefly been reviewed. The quantum MC method turns gradually into the classical MC method when the potential profile becomes smoother. Therefore, the simulation method can be used, for instance, to study the gradual emergence of quantum effects when a device structure is scaled down. While the numerical Wigner function approach can also handle the classical limit, it is unsuitable to describe closed quantum systems. The bound-state problem has been theoretically discussed in detail. It has been shown that the numerical

Wigner function approach should only be applied to structures exhibiting not too narrow resonances.

Acknowledgment

This work has been supported by the Austrian Science Fund, project I79-N16.

References

1. J. Wang and M. Lundstrom, Does source-to-drain tunneling limit the ultimate scaling of MOSFETs?, in *Int. Electron Devices Meeting*, 29.2.1–29.2.4 (IEEE 2002).
2. P. Palestri, D. Esseni, S. Eminent, C. Fiegna, E. Sangiorgi, and L. Selmi, Understanding quasi-ballistic transport in nano-MOSFETs: Part I - Scattering in the channel, and in the drain, *IEEE Trans. Electron Devices* **52**(12), 2727–2735 (2005).
3. M. Gilbert, R. Akis, and D. Ferry, Phonon-assisted ballistic to diffusive crossover in silicon nanowire transistors, *J. Appl. Phys.* **98**(9), 094303–1–8 (2005).
4. W. Frensley, Boundary conditions for open quantum systems driven far from equilibrium, *Rev. Mod. Phys.* **62**(3), 745–789 (1990).
5. H. Kosina, M. Nedjalkov, and S. Selberherr, A Monte Carlo method seamlessly linking classical and quantum transport calculations, *J. Comput. Electron.* **2**(2-4), 147–151 (2003).
6. L. Shifren, C. Ringhofer, and D. Ferry, Inclusion of nonlocal scattering in quantum transport, *Phys. Lett. A* **306**, 332–336 (2003).
7. A. Gehring and H. Kosina, Wigner-function based simulation of classic and ballistic transport in scaled DG-MOSFETs using the Monte Carlo method, *J. Comput. Electron.* **4**, 67–70 (2005).
8. W. Hänsch, *The Drift-Diffusion Equation and Its Applications in MOSFET Modeling*, Computational Microelectronics (Springer Verlag, Wien 1991).
9. H. Kosina, V. Sverdlov, and T. Grasser, Wigner Monte Carlo simulation: Particle annihilation and device applications, in *IEEE Proc.: Simulation of Semiconductor Processes and Devices*, Monterey, CA, USA, 357–360 (2006).
10. S. Selberherr, *Analysis and Simulation of Semiconductor Devices* (Springer Verlag, Wien, 1984).
11. F. Venturi, R. Smith, E. Sangiorgi, M. Pinto, and B. Ricco, A general purpose device simulator coupling Poisson and Monte Carlo transport with applications to deep submicron MOSFET's, *IEEE Trans. Computer-Aided Design* **8**(4), 360–369 (1989).
12. H. Kosina and M. Nedjalkov, Wigner function based device modeling, in *Handbook of Theoretical and Computational Nanotechnology* **10**, 731–763 (American Scientific Publishers, Los Angeles, 2006, in print).
13. P. Carruthers and F. Zachariasen, Quantum collision theory with phase-space distributions, *Rev. Mod. Phys.* **55**(1), 245–285 (1983).
14. V. Tatarskii, The Wigner representation of quantum mechanics, *Sov. Phys. Usp.* **26**(4), 311–327 (1983).

QUANTUM SIMULATIONS OF DUAL GATE MOSFET DEVICES: BUILDING AND DEPLOYING COMMUNITY NANOTECHNOLOGY SOFTWARE TOOLS ON NANO HUB.ORG

SHAIKH AHMED, GERHARD KLIMECK, DERRICK KEARNEY, MICHAEL MCLENNAN

*Network for Computational Nanotechnology,
Purdue University, West Lafayette, IN 4790, USA
ssahmed@purdue.edu*

M P ANANTRAM

*Department of Electrical and Computer Engineering,
University of Waterloo, Ontario, Canada
anant@ecemail.uwaterloo.ca*

Undesirable short-channel effects associated with the relentless downscaling of conventional CMOS devices have led to the emergence of new classes of MOSFETs. This has led to new and unprecedented challenges in computational nanoelectronics. The device sizes have already reached the level of tens of nanometers where quantum nature of charge-carriers dominates the device operation and performance. The goal of this paper is to describe an on-going initiative on nanoHUB.org to provide new models, algorithms, approaches, and a comprehensive suite of freely-available web-based simulation tools for nanoscale devices with capabilities not yet available commercially. Three software packages nanoFET, nanoMOS and QuaMC are benchmarked in the simulation of a widely-studied high-performance novel MOSFET device. The impact of quantum mechanical effects on the device properties is elucidated and key design issues are suggested.

Keywords: NanoHUB; MOSFETs; Quantum effects; Online simulation; NEGF

1. Introduction

Rapid progress in nanofabrication technology accelerating the continuous and aggressive downscaling of classical CMOS devices and the associated detrimental short-channel effects have inspired the emergence of new classes of MOSFETs. To fabricate devices beyond current scaling limits, IC companies are simultaneously pushing the planar, bulk silicon CMOS design while exploring alternative gate stack materials (high- k dielectric¹ and metal gates), band engineering methods (using strained Si or SiGe^{2,3,4,5}), and alternative transistor structures that include primarily partially-depleted (PD) and fully-depleted (FD) silicon-on-insulator (SOI) devices. However, there is a virtual consensus that the most *practically scalable* variety of all these devices, that are in the focus of many researchers' study today, are double-gate SOI MOSFETs with a sub-10 nm gate length, ultra-thin, intrinsic channels and highly doped (degenerate) bulk electrodes—see, e.g., recent reviews^{6,7} (and Figure 1). In such transistors, short channel effects typical for

their bulk counterparts are minimized, while the absence of dopants in the channel maximizes the mobility and hence drive current density.

These advanced MOSFETs may be practically implemented in several ways including planar, vertical, and FinFET geometries. However, several design challenges have been identified such as a process tolerance requirement of within 10% of the body thickness and an extremely sharp doping profile with a doping gradient of 1 nm/decade. The SIA forecasts that this new device architecture may extend MOSFETs to the 22 nm node (9-nm physical gate length) by 2016. Intrinsic device speed may exceed 1 THz and integration densities will be more than 1 billion transistors/cm².

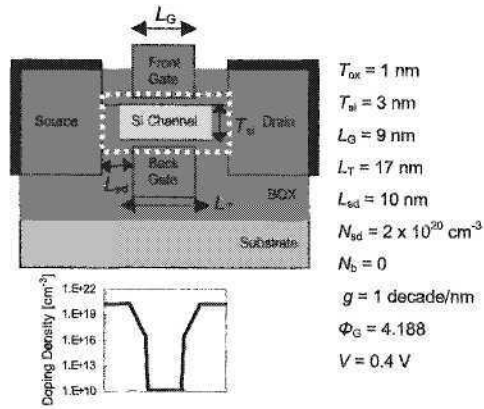


Fig. 1. The model of a widely-studied double-gate SOI MOSFET with ultra-thin intrinsic channel.

2. Nanoscale Device Simulation: Quantum Effects

Semiclassical Boltzmann transport has been the principal support in the field of modeling and simulations of semiconductor technology since its early development. To date, most *commercial* device simulations including the full-band Monte-Carlo (FBMC) method are based on the solution of the Boltzmann transport equation (BTE) and its simplified derivatives such as the hydrodynamic (HD) equations and the drift-diffusion (DD) model. However, for silicon nanoscale devices with active regions below 0.1 microns in diameter, macroscopic transport descriptions based on drift-diffusion (DD) models are clearly inadequate⁸. In the nanoscale regime, the transport is expected to be dominated by quantum effects throughout the active region. Quantum effects in the surface potential will have a profound impact on both, the amount of charge which can be induced by the gate electrode through the gate oxide, and the profile of the channel charge in the direction perpendicular to the surface (the transverse direction). Also, because of carrier confinement in the channel, mobility (or microscopically speaking, carrier scattering) will be different from the three-dimensional (3-D) case. A well known approach to study the impact of spatial confinement on mobility behavior is based on the self-consistent solution of the 2D Poisson–1D Schrödinger–2D Monte Carlo, and requires enormous computational resources as it requires storage of position dependent scattering tables that describe carrier transition between various subbands. However, in the smallest size devices, carriers experience very little or no scattering at all (*ballistic limit*). In such ballistic limit the question arises if a classical transport approach (BTE) or a full quantum mechanical approach is appropriate. Various quantum formalisms based on density matrices, Wigner functions, Feynman path integrals, and non-equilibrium Green's

functions (NEGF) have been developed and proposed. The Green's functions approach is the most exact, but at the same time appears, from the historical literature perspective, as the most difficult of all. The nanoHUB.org now features a webpage with introductory NEGF materials⁹. In contrast to, for example, the Wigner function approach (which is Markovian in time), the Green's functions method allows one to consider simultaneously correlations in space and time or space and energy, both of which are expected to be important in nanoscale devices. However, the difficulties in understanding the various terms in the resultant equations and the computational burden needed for its actual implementation are perceived as a great challenge. However, a successful utilization of the Green's function approach commercially is the NEMO (Nano-Electronics Modeling) simulator⁹, which is effectively 1-D and is primarily applicable to resonant tunneling diodes. This work presented here focuses mainly on the modeling and simulations of the size-quantization effect within a fully quantum mechanical NEGF approach and a quantum-corrected Monte Carlo transport framework for 2-D MOSFET structures.

3. Community Nanoscale MOSFET Softwares

The Network for Computational Nanotechnology (NCN, <http://www.ncn.purdue.edu>) is a multi-university, NSF-funded initiative with a mission to lead in nanotechnology research and education as well as outreach to students and professionals by offering a set of cyber services (accessible through the nanoHUB portal www.nanoHUB.org) including interactive online simulation, tutorials, seminars, and online courses packaged using e-learning standards. In the past 12 months³, the educational and outreach services were accessed by over 16,200 users. More than 3,500 users performed over 94,000 online simulations. Over 30 applications are available online ranging from toy models to sophisticated simulation engines not yet available commercially. All the NCN services are freely open to the public.

NanoHUB is dedicated to offer services to real users such as experimentalists and educators, not to computational scientists alone. Therefore, a user friendly GUI is required for these tools to be operated by non-experts. The tool should be available for anybody without any installation requirements. From a deployment perspective the tool development and GUI development must be streamlined, the codes must be benchmarked, and quality tested, and adequate computational resources must be available. Rappture is the new nanoHUB.org created toolkit that enables the rapid development of GUIs for applications. Two approaches can be followed: (1) The legacy application is not modified at all and wrapper script translates Rappture I/O to the legacy code. (2) Rappture is integrated into the source code to handle all I/O (see Figure 2). The first step is to declare the parameters associated with one's tool by describing Rappture objects in the Extensible Markup Language (XML). Rappture reads the XML description for a tool and generates the GUI automatically. The second step is that the user interacts with the GUI, entering values, and eventually presses the *Simulate* button. At that point,

³ Data as of September 2006. Updated information can be found at <http://nanoHUB.org/usage/>

Rapture substitutes the current value for each input parameter into the XML description, and launches the simulator with this XML description as the *driver* file. The third step shows that, using parser calls within the source code, the simulator gets access to these input values. Rapture has parser bindings for a variety of programming languages, including C/C++, Fortran, MATLAB, Octave, Python, Perl, and Tcl. And finally, the simulator reads the inputs, computes the outputs, and sends the results through *run* file back to the GUI for the user to explore.

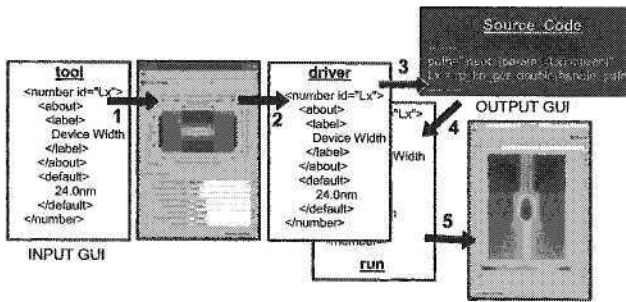


Fig. 2. Rapture: **Revolutionizing** tool development.

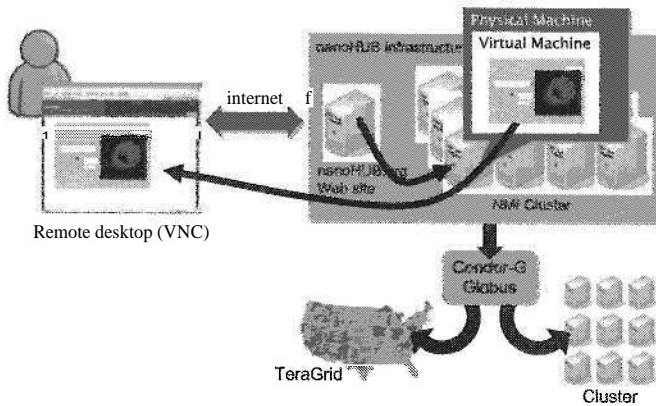


Fig. 3. Remote access to simulators and compute power.

The process of *web-based deployment* of these tools is depicted in Figure 3. A user visits the www.nanohub.org site and finds a link to a tool. Clicking on that link will cause our middleware to create a virtual machine running on some available CPU. This virtual machine gives the user his/her own private file system. The middleware starts an application and exports its image over the Web to the user's browser. The application looks like an Applet running in the browser. The user can click and interact with the application in real time taking advantage of high-performance distributed computing

power available on local clusters at Purdue University. Large scale calculations will soon be launched on the NSF TeraGrid or the open science grid. The following sections present three in-house software packages namely nanoFET, nanoMOS and QuaMC and benchmark these tools in the simulation of a widely-studied high-performance novel MOSFET device.

3.1 nanoFET

nanoFET is a newly developed tool, based on the research performed at NASA¹⁰, that simulates quantum mechanical size quantization in the inversion layer and phase coherent and ballistic transport properties in two-dimensional MOSFET devices. The overall simulation framework consists of the *real-space* effective mass NEGF equations solved self-consistently with Poisson's equation. Solution of this set of equations is computationally intensive. Hence, nonuniform spatial grids are essential to limit the total number of grid points while at the same time resolving physical features. A novel algorithm for efficient computation of electron density without complete solution of the system of equations even in the presence of nonzero self-energies throughout the device has been used in this simulator^{9,10}. The numerical problem consists in computing the diagonal elements of the matrix $G^r = [EI - H - \Sigma]^{\dagger}$ (retarded Green's function) and $G^< = G\Sigma^<G^{\dagger}$ (electron correlation Green's function), where E is the energy level, H is the device Hamiltonian matrix, and Σ and $\Sigma^<$ are self energies (\dagger denotes the transpose conjugate of a matrix). The algorithmic flow is based on Dyson's equation solved through recursive Green's function approach. nanoFET has been parallelized with Message Passing Interface (MPI) and ported to various computing platforms at Purdue University. The MPI is applied in the integration procedure to calculate the charge density over the energy spectrum while the Green's function at each energy point is calculated by a serial algorithm. The resulting speed-up factor (Figure 4) shows a satisfactory scaling behavior for up to 32 processors. The reason of a declining behavior beyond 32 processors is attributed to the increased amount of time used in various MPI communication calls (right panel of Figure 4) that are implemented immediately after the parallel computation in order to exchange and gather large amount and size of data among the processors.

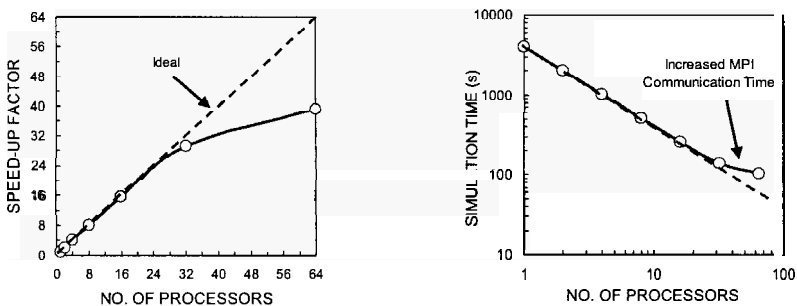


Fig. 4. (Left) nanoFET parallel performance. (Right) MPI timing diagram.

3.2 nanoMOS

nanoMOS is a 2-D simulator for thin body (less than 5 nm), fully depleted, double-gated n -MOSFETs. A choice of five transport models is available (drift-diffusion, classical ballistic, energy transport, quantum ballistic, and quantum diffusive). The transport models treat quantum effects in the confinement direction exactly and the names indicate the technique used to account for carrier transport along the channel. The quantum transport models are based on *mode-space* method within an effective mass approximation. The use of a mode space approach produces high computational efficiency that makes the quantum simulator practical for extensive device simulation and design. Scattering in the device can also be treated by a simple model that uses so-called Büttiker probes¹¹.

3.3 QuaMC

QuaMC is a quasi three-dimensional quantum-corrected semiclassical Monte Carlo transport simulator for conventional and non-conventional MOSFET devices. A parameter-free quantum field approach has been developed and utilized quite successfully in order to capture the size-quantization effects in nanoscale MOSFETs. The method is based on a perturbation theory around thermodynamic equilibrium and leads to a quantum field formalism in which the size of an electron depends upon its energy¹². This simulator uses different self-consistent event-biasing schemes for statistical enhancement in the Monte-Carlo device simulations has been presented¹³. Regarding the Monte Carlo transport kernel, the explicit inclusion of the longitudinal and transverse masses in the silicon conduction band is done in the program using the Herring-Vogt transformation. Intravalley scattering is limited to acoustic phonons. For the intervalley scattering, both g - and f -phonon processes have been included.

4. Simulation Results and Discussion

Figure 1 shows the simulated DG FET device structure, which is similar to the device reported in Ref. 7. For quantum transport simulation purposes only the dotted portion of the device which has been termed as the *intrinsic* device is taken into considerations. The device was originally designed in order to achieve the ITRS performance specifications for the year 2016. The effective intrinsic device consists of two gate stacks above and below a thin silicon film. For the intrinsic device, the thickness of the silicon film is 3 nm. Using a thicker body reduces the series resistance and the effect of process variation but it also increases the short channel effects (SCE) due to lesser control on the channel charge. From SCE point of view, a thinner body is preferable but it is harder to fabricate very thin films of uniform thickness, and the same amount of process variation ($\pm 10\%$) may give intolerable fluctuations in the device characteristics. For the simulated intrinsic device, source/drain width is 3 nm. The top and bottom gate insulator thickness is 1 nm, which is expected to be near the scaling limit for SiO₂. For the gate contact, a metal gate with tunable workfunction, Φ_G , is assumed, where Φ_G is adjusted to 4.188 to provide a

specified off-current value of $4 \mu\text{A}/\mu\text{m}$. The background doping of the silicon film is taken to be intrinsic, however, to take into account the diffusion of the dopant ions, the doping profile in the S/D extensions is graded with a coefficient of g which equals to 1 dec/nm (Figure 1). The doping gradient, g , affects both on-current and off-current. According to the ITRS roadmap, the high performance (HP) logic device would have a physical gate length of $L_G = 9 \text{ nm}$ at the year 2016. At this scale, 2-D electrostatics and quantum mechanical effects both play an important role. Values of all the structural parameters of the device are shown in Figure 1.

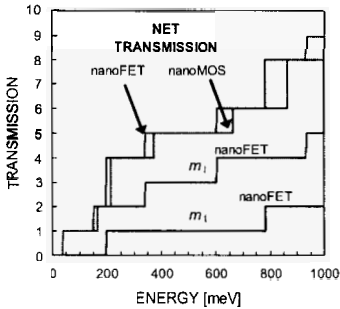


Fig. 5. Benchmarking nanoFET against nanoMOS: Transmission profile for a system with zero/flat potential applied in the device region.

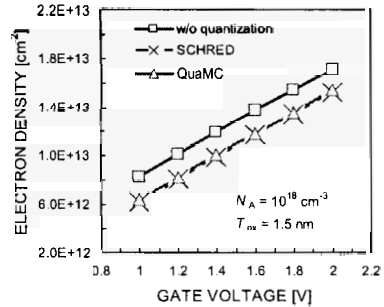


Fig. 6. Benchmarking of QuaMC against SCHRED: Variation of the sheet electron density versus gate voltage in a MOS capacitor structure.

The newly developed nanoFET simulator is first benchmarked against the nanoMOS simulator in the non-selfconsistent calculation of transmission under zero/flat potential distribution throughout the device region. The resulting transmission profile is shown in Figure 5. The curve marked as m_t on this plot corresponds to the (nanoFET derived) transmission in the primed valleys ($m_t = 0.19m_0$) and that marked as m_1 corresponds to the unprimed valleys ($m_1 = 0.98m_0$). The two topmost curves give the net transmission from nanoFET and nanoMOS. Noticeable is the reasonable agreement in the ground and excited state energy levels thus validating the approaches used to calculate the underlying Green's functions.

To verify the validity of the effective quantum field approach integrated in QuaMC simulator, self-consistent simulations of MOS capacitor structures have been performed and the reduction in the sheet electron density due to the band-gap widening effect is calculated. The simulation results are shown in Figure 6. Notice the excellent agreement between the SCHRED simulation data for the sheet electron density and the simulation results obtained from QuaMC. SCHRED is a 1-D Schrödinger-Poisson solver that is also available on the nanoHUB.org¹⁴.

Next, the *intrinsic* device is simulated, within a *ballistic* approximation, to study the impact of size-quantization effects on the DG MOSFET performance. Figure 7 shows the total transmission vs. energy profile for a bias of $V_G = 0 \text{ V}$ and $V_D = 0.4 \text{ V}$ computed in

nanoFET. Note that the steps in the plot correspond to the subband levels due to channel quantization. Two main features in this plot are: (a) the integer values at the total transmission plateaus, (b) existence of a slope (not sharp) near the subbands which has significant impact on determining the current. The corresponding density of states peaks at the lateral confinement resonance level in the channel region. Figure 8 shows the current vs. energy plot where the increase in current for $V_D = 0.4V$ ($\sim 0.49 \mu A/\mu m$) as compared to that for $V_D = 0.05V$ ($\sim 0.22 \mu A/\mu m$) is due to an increase in both the tunneling (compare the shaded areas) and the thermionic components.

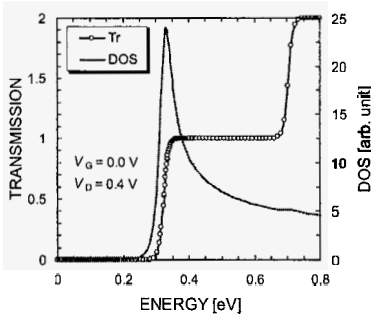


Fig. 7. nanoFET total transmission and density of states (DOS) vs. energy plot.

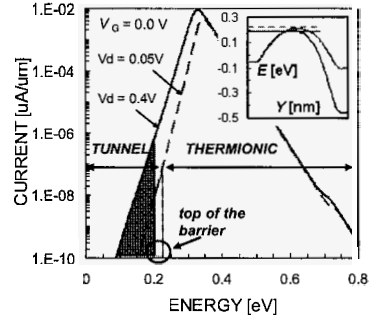


Fig. 8. nanoFET current vs. energy plot for two applied drain biases.

The device *ballistic* output characteristics are shown in Figure 9. All these simulations were performed on nanoHUB.org. For benchmarking purposes, apart from the device specifications, mentioned below is a list of key simulator parameters. (a) *nanoMOS*: A well-tempered DG MOSFET was chosen with low-field mobility = $300 \text{ cm}^2/V\text{s}$, Cauchy-Thomas parameter = 2, oxide-penetration was neglected, self-consistent convergence parameter = $1.0E-3$, Poisson convergence parameter = $1.0E-6$, all valleys were considered with 1 subband in each valley, the dual-gate and the Fermi-Dirac flags were activated. (b) *nanoFET*: nonuniform adaptive energy grid with a size (number) of 512 was used, resonance in the active region was included, full-simulation mode Poisson equation being solved self-consistently with an iteration number of 10, mass-anisotropy was included, both the $G^<$ and $G^>$ were calculated, backward recursion flag was turned on, 16 computing nodes with 1 processor/node were used. (c) *QuaMC*: Monte Carlo simulation time was 2ps with a time-step of 0.1fs, event-biasing flag was turned off, quantum correction flag was turned on for quantum-mechanical and off for classical calculations, for ballistic transport all the scattering switches were off and for dissipative transport they were turned on. As shown in Figure 9, for a device with 3nm film thickness, although at low drain biases there remains slight discrepancy in the results obtained from the nanoFET and nanoMOS simulations, they tend to overlap at higher drain voltages. The QuaMC simulations were also found to produce almost the same current as those obtained from fully quantum mechanical counterpart.

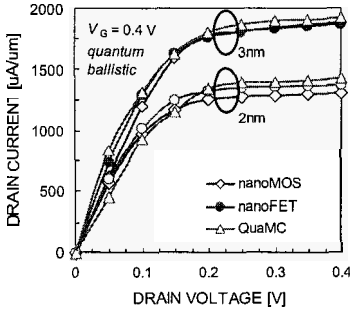


Fig. 9. Output characteristics for ballistic transport: Comparison of different simulators.

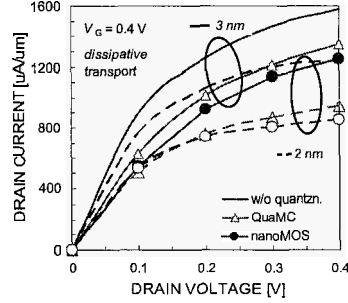


Fig. 10. Dissipative transport through nanoMOS and QuaMC: Impact of channel thickness.

Finally, simulation results for dissipative transport (a) without quantization from classical QuaMC and (b) with quantization obtained from nanoMOS and QuaMC are compared. The output characteristics are shown in Figure 10. Devices with both 3 nm and 2 nm channel thickness are used with a gate bias of 0.4 V. The salient features of this figure are: (1) Even with an undoped channel region, the devices achieve a significant improvement over the short channel effects (SCEs) as seen from the near-flat nature of the curves in the saturation region. This is due to the use of the two gate electrodes and an ultrathin SOI film which makes the gates gain more control on the channel charge. (2) Reducing the channel SOI film thickness to 2 nm further reduces the SCEs (saturation region has a reduced slope). (3) Comparing the classical and quantum mechanical simulations in both the devices, one can see that the impact of quantization effects (percentage change in drain current) reduces with an increase in drain voltage. Higher biases primarily open all the available channels and quantum mechanical and classical calculations agree more closely. (4) Reduction in the drain current is more in 2 nm case throughout the range of the applied drain bias because of the stronger physical confinement. (5) Comparing the 3 nm and the 2 nm cases (classical or quantum mechanical), the increase in the percentage reduction in the drive current at higher drain biases is due to the series resistance effect in the thinner device. (6) Finally, the comparison between the quantum potential formalism and the NEGF approach for the device with 3 nm film thickness shows reasonable agreement which further establishes the applicability of these methods in the simulations of different technologically viable nanoscale classical and nonclassical MOSFET device structures.

5. Conclusion

NCN aims at the development of new *community codes* that provide the nanoscience research community with new capabilities and lay a foundation for a new generation of CAD tools that will pave the way to ground breaking nanotechnology devices. In the past 12 months, the educational and outreach services were accessed by over 16,200 users.

More than 3,500 users performed over 94,000 online simulations. We have presented benchmarking results obtained from nanoFET, nanoMOS and QuaMC, validating the applicability of these nanoHUB.org tools in the simulations of future nanoscale MOSFETs, that show the importance of quantum effects in these novel devices.

Acknowledgments

Supported by the Indiana 21st Century Fund and National Science Foundation under Grant No. EEC-0228390. The computational resources of nanoHUB.org have been utilized. Discussions with Dragica Vasileska and Sayed Hasan are acknowledged.

References

1. W. Zhu, J. P. Han, T. P. Ma, "Mobility Measurement and Degradation Mechanisms of MOSFETs Made With Ultrathin High-k Dielectrics", *IEEE Trans. Elect. Dev.*, **51**, 98 (2004).
2. Welsler, J. L. Hoyt and J. F. Gibbons, "NMOS and PMOS transistors fabricated in strained silicon/relaxed silicon-germanium structures", *IEDM Tech. Dig.*, 1000 (1992).
3. D. Vasileska and S. S. Ahmed, "Narrow-Width SOI Devices: The Role of Quantum Mechanical Size Quantization Effect and the Unintentional Doping on the Device Operation", *IEEE Transactions on Electron Devices*, **52**, 227 (2005).
4. D. Vasileska, G. Formicone and D.K. Ferry, "Doping dependence of the mobility enhancement in surface-channel strained-Si layers," *Nanotechnology*, **10**, 147 (1999).
5. R. Oberhuber, G. Zandler, and P. Vogl, "Subband structure and mobility of two-dimensional holes in strained Si/SiGe MOSFET's," *Phys. Rev. B*, **58**, 9941 (1998).
6. Thomas J. Walls, Victor A. Sverdlov, and Konstantin K. Likharev, "Nanoscale SOI MOSFETs: a comparison of two options," *Solid-State Elect.*, **48**, 857 (2004).
7. Sayed Hasan, Jing Wang, and Mark Lundstrom, "Device design and manufacturing issues for 10 nm-scale MOSFETs: a computational study," *Solid-State Elect.*, **48**, 867 (2004).
8. N. Sano, A. Hiroki, K. Matsuzawa, "Device modeling and simulation toward sub-10nm semiconductor devices," *IEEE Trans. Nanotechnology*, **1**, 63 (2002).
9. https://www.nanohub.org/topics/negf_theory
R. Lake, G. Klimeck, R.C. Bowen, and D. Jovanovic, *J. Appl. Phys.*, **81**, 7845 (1997).
10. A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel and R. Venugopal, "Two Dimensional Quantum Mechanical Modeling of Nanotransistors", *Jour. of Appl. Phys.*, **91**, 2343 (2002).
http://www.nanohub.org/simulation_tools/nanoFET_tool_information
11. Z. Ren, R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom, "nanoMOS 2.5: a two-dimensional simulator for quantum transport in double-gate MOSFETs," *IEEE Trans. Electron. Dev.*, **50**, 9, 1914, (2003).
https://www.nanohub.org/simulation_tools/nanomos-3_tool_information
12. S. Ahmed, C. Ringhofer, D. Vasileska, "Parameter-Free Effective Potential Method for Use in Particle-Based Device Simulations", *IEEE Transactions on Nanotechnology*, **4**, 465 (2005).
http://www.nanohub.org/simulation_tools/quame2d_tool_information
13. M. Nedjalkov, S. Ahmed, and D. Vasileska, "A self-consistent event biasing scheme for statistical enhancement", *Journal of Computational Electronics*, **3**, 305 (2004).
14. SCHRED simulations performed on nanoHUB.org.
https://www.nanohub.org/simulation_tools/schred_tool_information

POSITIVE MAGNETO-RESISTANCE IN A POINT CONTACT : POSSIBLE MANIFESTATION OF INTERACTIONS.

V. T. RENARD

*NTT Basic Research Laboratories NTT corporation, 3-1 Morinosato Wakamiya,
Atsugi, Kanagawa 243-0198, Japan*

T. OTA

*NTT Basic Research Laboratories NTT corporation, 3-1 Morinosato Wakamiya,
Atsugi, Kanagawa 243-0198, Japan and
SORST-JST 4-1-8 Honmachi,
Kawaguchi, Saitama 331-0012, Japan*

N. KUMADA

*NTT Basic Research Laboratories NTT corporation, 3-1 Morinosato Wakamiya,
Atsugi, Kanagawa 243-0198, Japan*

H. HIRAYAMA

*NTT Basic Research Laboratories NTT corporation, 3-1 Morinosato Wakamiya,
Atsugi, Kanagawa 243-0198, Japan and
SORST-JST 4-1-8 Honmachi,
Kawaguchi, Saitama 331-0012, Japan and
Department of Physics, Tohoku University
Sendai 980-8578, Japan*

We report a non-monotonic and strongly temperature dependent magneto-resistance observed in clean quantum point contacts. At the same time the conductance of the point contact varies linearly with temperature. This unexpected behavior may be related to electron-electron interactions.

Keywords: Quantum point contact; electron-electron interaction.

1. Introduction

Since the first realization of semiconductor based systems^{1,2} experimentalists have been able to study point contacts in the quantum limit (both the mean free path and Fermi wave length are comparable to the contact dimensions). This achievement was soon followed by a huge amount of experimental and theoretical attention (see Ref. 3 and references therein). One of the most emblematic effect is the quan-

tization of the conductance of quantum point contact (QPC) in units of $2 \times e^2/h$ which can be explained by one-particle quantum physics³. Further support for this simple picture is the appearance of plateaus at odd multiples of e^2/h in a large magnetic field which evidence the Zeeman splitting of the 1D modes. At present the main activity concerning QPCs is concentrated on the first conductance plateau below which an additional step forms around $0.7 \times (2e^2/h)$. This plateau evolves continuously to the lowest spin-resolved plateau at e^2/h as in-plane magnetic field is increased. This suggests a possible spontaneous spin polarization, presumably due to electron-electron interactions.

The purpose of the present report is to show that in high quality QPCs the high conductance curves also reveals features that may be attributed to electron-electron interactions. We have measured a strongly temperature dependent and non-monotonic magneto-resistance (MR) which could not be described by the existing one particle description.

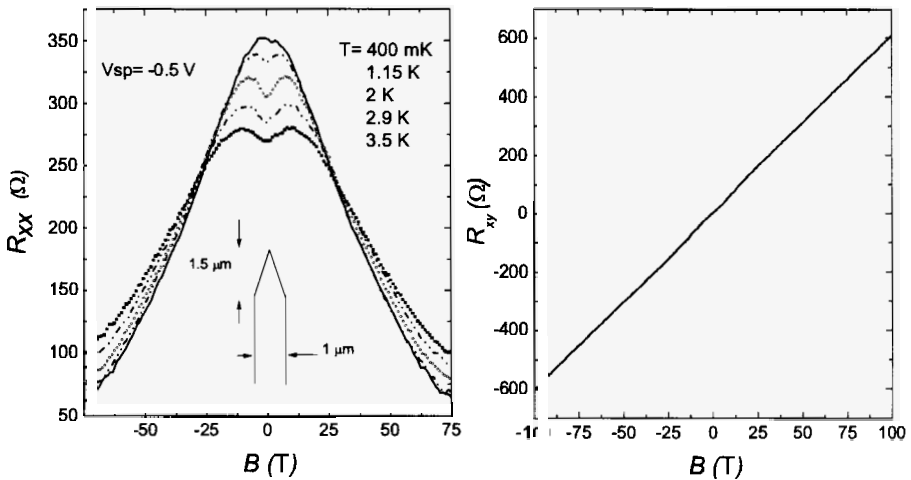


Fig. 1. a) Magneto-resistance of the sample for different temperatures. The arrow shows the temperature independent point. The geometry of the split gate is also displayed. b) Hall resistance of the sample for $T=400$ mK and $T=3.5$ K.

2. Experiment

The point contacts were defined using the split gate depletion **technique**^{1,2} to define a nano-scale point contact between two adjacent macroscopic two-dimensional electron gases (2DEG). The initial 2DEG had a low temperature mobility of 4×10^6 cm²/Vs for a density of $n_s = 10^{11}$ cm⁻². The top split gates were realized using up to date electron beam lithography. Several samples with different geometrical parameters were studied with a similar result. We will therefore present the results concerning one sample with 1.4 μ m spaced needle split gates (see Figure 1). The sample resistance was recorded in a four terminal configuration using low frequency techniques at low excitation voltage. The sample was studied down to 400 mK and up to $B = 10$ T in a He³ cryostat equipped with a super-conducting magnet.

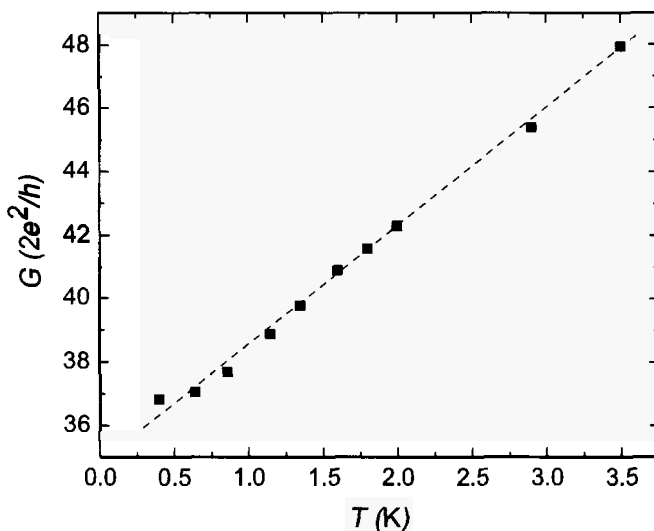


Fig. 2. Conductance of the sample as a function of temperature. The dashed line is a guide for the eyes.

3. Results and discussion

Figure 1 displays the magneto-resistance of the point contact for a split gate bias of -0.5 V which corresponds to a conductance approximately equal to $40 \times 2e^2/h$. It can be seen that the resistance increases when a small (less than 20 mT) magnetic field is applied at high temperatures. For larger fields the resistance decreases with magnetic field. This results in a maximum of resistance at about 20 mT. When the

temperature is decreased the resistance at zero magnetic field increases and the maximum shifts to lower fields. Eventually the magneto-resistance becomes entirely negative for the lowest temperature. Figure 2 shows that the conductance of the sample at $B = 0$ T varies linearly with temperature between 400 mK and 3.5 K. Signs of a possible saturation may be seen at the lowest temperature.

Let us now discuss these results starting from the magneto-resistance. It was shown early that applying a magnetic field perpendicular to the device leads to a decrease of the resistance of the quantum point contacts⁵. This can be explained by the fact that the constriction becomes gradually transparent to the edge states of the 2D adjacent wide regions. The associated negative magneto-resistance was written⁵ as follows :

$$R_{AT} = \frac{h}{2e^2} (N_{min}^{-1} - N_{wide}^{-1}), \quad (1)$$

where N_{min} and N_{wide} are the number of occupied states in the constriction and in the wide regions respectively. Their magnetic field dependence is expressed in Ref. 5. The negative magneto-resistance that we observe at larger fields is related to this phenomenon. Indeed Eq. 1 fits the high field ($B > 20$ mT) data very well. We used k_f (Fermi wave vector) and W (width of the constriction) as parameters and obtained $W = 1.2 \pm 0.02 \mu\text{m}$ and $k_f = (9.32 \pm 0.7) \times 10^7 \text{ m}^{-1}$, comparable to the lithographic sizes and density. Equation 1 however does not explain the small temperature dependence of the slope of the magneto-resistance at high fields which can be mainly attributed to an apparent variation of the Fermi wave vector. We have at the moment no explanation for this renormalization.

Let us now discuss the central result of this study : the non-monotonic magneto-resistance. Several origins have been already discussed in the literature^{6,7}. It was shown in Ref. 6 that a positive magneto-resistance could appear in long quasi one-dimensional systems if the probability of diffusive scattering by the boundaries is non zero. This effect can be discarded in our case for three reasons. First in this case the maximum of the resistance was precisely observed to occur at $W/l_c = 0.55$, where l_c is the cyclotron radius⁶. In our measurement the maximum occurs at $W/l_c < 0.2$. Also contrary to our measurement, the phenomenon was shown to be completely temperature independent for $T < 20$ K⁸. Finally the effect cannot be observed for short constrictions like ours since boundary scattering doesn't exist⁶. Another possible origin of the positive magneto-resistance is the hydrodynamic effects first discussed by Gurzhy *et al*⁷. They considered the interplay of electron-electron interactions and boundary scattering and showed that it could result in a positive magneto-resistance and a linear conductance as function of temperature. However, as already stressed boundary scattering is very unlikely to occur in our system. If a similar hydrodynamic effect appears in our devices electron-electron interactions have to be mediated by another yet unknown mechanism.

To our knowledge, the last possible origin of the magneto and temperature dependence of the resistance of our device is the Friedel oscillations. Oscillations of the electron density indeed arise close to the boundaries or impurities due to the

interference between incoming and reflected electron waves. These oscillations of density in turn create an oscillatory potential seen by conduction electrons. This acts as an extra scattering potential which is "switched on" when the temperature is decreased as the system becomes coherent. This interaction related scattering potential was shown to have a great importance in two dimensional systems^{9,10}. It could in principle explain the observed strong linear temperature dependence in our case. As for the field dependence more theoretical investigations has to be done.

4. Conclusion

We have studied clean quantum point contacts which showed unusual temperature and field dependencies. This result could not be explained by the single-particle picture and may require a many-body treatment.

Acknowledgments

We are very grateful to O. Tkachenko, V. Tkachenko, I. Gornyi and Y. Tokura for very interesting discussions.

References

1. D. A. Warham *et al*, "One-dimensional transport and the quantisation of the ballistic resistance", *J. Phys. C.* **21** (1988) L209-L214.
2. B. J. van Wees *et al*, "Quantized conductance of point contacts in a two-dimensional electron gas", *Phys. Rev. Lett.* **60** (1988) 848-850.
3. c. W. J Beenakker and H. van Houten, "Quantum transport in semiconductor nanostructures", *Solid state physics.* **44** (1991) 1-228.
4. K. J. Thomas *et al*, "Possible Spin Polarization in a One-Dimensional Electron Gas", *Phys. Rev. Lett.* **77** (1996) 135-138.
5. van Houten *et al*, "Four terminal magnetoresistance of a two-dimensional electron gas constriction in the ballistic regime", *Phys. Rev. B.* **37** (1988) 8534-8536.
6. T. J. Thornton, M. L. Roukes, A. Scherer and B. P. van de Gaag "Boundary scattering in quantum wires", *Phys. Rev. Lett.* **63** (1989) 2128-2131.
7. R. N. Gurzhy, A. N. Kalinenko and A. I. Kopeliovich, "Electron-electron collisions and a new hydrodynamic effect in two-dimensional electron gas", *Phys. Rev. Lett.* **74** (1995) 3872-3875.
8. F. Rahman, T. J. Thornton, B. L. Gallagher and R. A. Stradling, "Boundary scattering in a wet etched InAs/GaSb heterostructure wires : with and without magnetic field.", *Semicond. Sci. Technol.* **14** (1999) 478-483.
9. G. Zala, B. N. Narozhny, and I. L. Aleiner, "Interaction corrections at intermediate temperatures: Longitudinal conductivity and kinetic equation", *Phys. Rev. B.* **64** (2001) 214204-214235.
10. I. V. Gornyi and A. D. Mirlin, "Interaction-Induced Magnetoresistance: From the Diffusive to the Ballistic Regime", *Phys. Rev. Lett.* **90** (2003) 076801-076804.

This page intentionally left blank

IMPACT OF INTRINSIC PARAMETER FLUCTUATIONS IN NANO-CMOS DEVICES ON CIRCUITS AND SYSTEMS

SCOTT ROY, BINJIE CHENG, and ASEN ASENOV

*Department of Electronics & Electrical Engineering, University of Glasgow,
Glasgow, G12 8QQ, United Kingdom
S.Roy@elec.gla.ac.uk*

Device parameter fluctuations, which arise from both the stochastic nature of the manufacturing process and more fundamentally from the intrinsic discreteness of charge and matter, are a dominant source of device mismatch in nano-CMOS devices, and a bottleneck to the future yield and performance of circuits and systems. The impact of such parameter fluctuations is investigated for circuits — with a specific exemplar of 6-T SRAM — whose devices scale from 35 nm gate length. We posit a change in design approach to include the use of statistical compact models as a starting point for the development of cell libraries containing fluctuation information necessary for design under the constraints of parameter fluctuations, and novel Technology Aided System Design tools.

Keywords: Intrinsic parameter fluctuations; device simulation; circuit simulation

1. Introduction

As nano-CMOS device dimensions continue to scale, the relative magnitudes of the measured parameter fluctuations within device ensembles inexorably increase, and become a dominant sources of device mismatch in the decananometer regime.¹ Intrinsic parameter fluctuations (IPF), a result of the granular nature of charge and matter are, for a given device architecture, unaffected by improved process control. They are therefore recognized as a crucial bottleneck to the future yield and performance of CMOS devices (using present and anticipated device architectures) and corresponding circuits / systems. Figure 1 shows the fluctuations in gate characteristics due to one source of IPF, the random configuration of dopants in the channel of a 35 nm bulk MOSFET.

Historically, the two domains of device and circuit / system design have benefited from remaining somewhat orthogonal, but in the decananometer regime it is becoming increasingly difficult to perform optimised circuit design without a comprehensive understanding of device technology, and vice versa. Under these conditions, with the two design areas inextricably entangled, it becomes vitally important to understand the effects of IPF not only on devices, but also to assess their effects on circuits and systems. TCAD is already an essential tool in the development of new technology nodes, saving up to 25-30% on technology development costs.² If analysis of IPF can be performed by predictive simulation, and the TCAD and circuit analysis / ECAD toolflows updated to take account

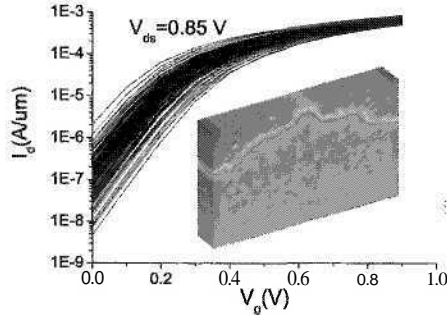


Fig. 1. Gate Characteristics of 200 macroscopically identical 35 nm MOSFETs obtained by ‘atomistic’ device simulation. (Inset: potential in single FET in which the detailed positions of dopants are considered).

of statistical effects, then extrinsic and intrinsic fluctuation sources can be investigated, individually or combined, and their effect on circuits and systems (including on the circuit and systems design process) predicted. Such assessments can issue an early warning to industry, dependant on device architecture, as to when the levels of IPF will become unacceptable under traditional circuit design constraints, and therefore when new electronics design methodologies and philosophies,³ such as probabilistic design, IPF aware / tolerated design flows, should be introduced in order to achieve the full benefits from continued device scaling.

In this paper, a hierarchical, statistical device to circuit simulation methodology, which can push ‘design for manufacture’ to a new level, is presented, and used to study the impact of IPF on fundamental circuits. In the next section each hierarchical level of the statistical process, device and circuit simulations are discussed. The impact of a key source of IPF, random dopant fluctuations, on exemplar 6-T SRAM cells is then described - using devices carefully calibrated to a 35 nm physical gate length bulk FET - and conclusions drawn.

2. Statistical Device Simulation

Physics based simulation is of crucial importance in understanding and predicting the impact of IPF on the characteristics of each new generation of devices, and corresponding impact on circuits and systems. We have developed a hierarchical device simulation methodology which involves two distinct numerical device simulation stages and a statistical compact modelling stage. Firstly, device models constructed in the Taurus simulation suite from Synopsys⁴ are calibrated against a real device structure and process flow - in this case a 35 nm physical gate length transistor from Toshiba.⁵ From this physical device, realistic device structures scaled to 25, 18 and 13 nm physical gate length (which will be referred to as the “25 nm generation”, “18 nm generation” and “13 nm generation” respectively below) are derived based on the guidelines for the 65, 45 and 32 nm technology nodes from the 2003 ITRS.² The devices have EOTs of 0.65, 0.5 and 0.3 nm, and junction depths of 13, 9 and 8 nm respectively. We assume V_{dd} of 1.0 V for the 25 and 18 nm generations and 0.9 V for the 13 nm generation. Doping profiles

and physical parameters are fed to the Glasgow 3D ‘atomistic’ simulator,⁶ which incorporates Density Gradient corrections for both electrons and holes to provide resolution of discrete dopants.⁷ Sources of fluctuations incorporated into the simulator include: line edge roughness,¹ interface roughness,⁸ the granular structure of the polysilicon gate⁹ and variation in high- κ gate stack morphology¹⁰ – allowing statistical analysis of the key sources of IPF in both traditional MOSFET architectures, partially depleted, and ultra-thin-body SOI devices.

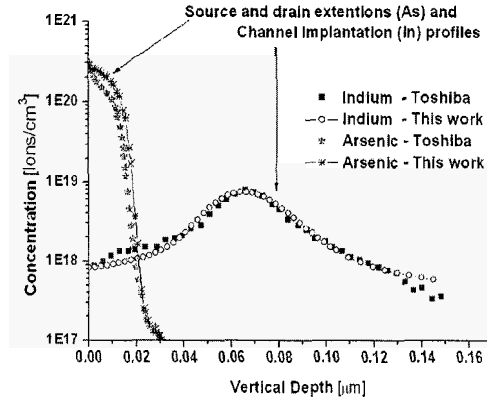


Fig. 2. The calibrated doping profiles of the 35 nm physical gate length nMOSFETs used in this study.

‘Atomistic’ simulation results are shown in Fig. 2, based on a simulation ensemble of 200 devices, populated with random configurations of dopants, and therefore subject to random dopant fluctuations (RDF). These cause significant drain current and subthreshold slope variation in weak inversion due to the lack of significant inversion layer screening. Simulation results show the off-current of 35 nm channel length devices spreading from $2 \mu\text{A}/\mu\text{m}$ to $8 \text{nA}/\mu\text{m}$, more than two orders of magnitude in value, and enough to cause serious difficulties for standby power dissipation control in any circuit constructed from these devices. Secondly, RDF causes the well documented threshold voltage fluctuations. Simulation results show that the threshold voltage standard deviation ΔV_{Th} for 35 nm channel length devices is more than 20% of the mean V_{Th} , indicating that significant on-current fluctuations will occur. Thirdly, there will be significant variations in the high drain bias characteristics due to electrical field fluctuations in saturation because of the lack of inversion layer screening near the drain / pinch-off. Finally, RDF causes DIBL variation through effective doping profile fluctuations. These last two factors can be considered as high field fluctuation effects and they have a joint impact on the saturation behaviour. If we use $\Delta V_{Th}/\Delta V_d$ as a figure of merit to describe the electrostatic characteristics at high drain bias, the standard deviation caused by RDF is around 10% of the mean value for 35 nm devices.

In order to investigate the impact of IPF at circuit level, statistical device simulation results should be transferred into compact models. The purpose of compact model is to

reproduce (not predict) the device characteristics in a computationally efficient way in order to make accurate circuit simulation practicable. The reproduction of device characteristics is achieved through a compact model parameter extraction procedure. Although physical understanding is crucially important in compact model development, it is unfeasible to build an efficient compact model based on a detailed physical device description (for example, surface potential based models are usually based on the classic physical description of a 1D ideal MOSFET structure, which should be a completely inappropriate basis for nano-scale devices). No matter whether the compact model is so called physically based or not, the ability to accurately reproduce real device characteristics, computation efficiency, and model flexibility are three essential requirements for statistical compact modelling. In this study, BSIM¹¹ models have been chosen to demonstrate the feasibility of the hierarchical simulation methodology.¹²

Statistical compact model extraction results from two extreme cases of devices subject to random dopant fluctuations are illustrated in Fig. 3. Although only device gate characteristics are inputs to this statistical parameter extraction stage, good agreement for drain characteristics has been achieved for all device generations, which confirms that gate characteristics contain most of the information on ‘atomistic’ fluctuations important to device operation.

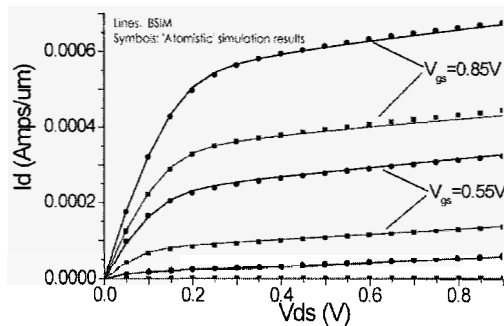


Fig. 3. Comparison between BSIM3v3 and ‘atomistic’ simulator result for two extreme 35 nm cases

3. Impact of Intrinsic Parameter Fluctuations on Circuits

A set of statistical compact model libraries are built based on the statistical parameter extraction results for 35, 25, 18, and 13 nm physical gate length devices. This forms the basis of Monte Carlo circuit simulation, carried out to investigate the impact of IPF at circuit level, with the speed of standard SPICE simulation, but with similar accuracy over the ensemble to that which can be obtained by 3D ‘atomistic’ simulation at a device level.

3.1. Impact of RDF on 6-T SRAM Cells

CMOS SRAM memory plays a critical role in modern microprocessors and System-on-Chip platforms. Due to their constrained cell area SRAM cells are notoriously sensitive to IPF. The result of static transfer characteristics for SRAM with cell ratio = 1 under the

influence of RDF is shown in Fig. 4, with the schematic of a 6-T SRAM cell illustrated inset. M2 and M4 are driver transistors, with M5 and M6 access transistors, and M1 and M3 load transistors. We assume that both NMOS and PMOS transistors have similar fluctuation properties, with average PMOS drive half that of NMOS. For a cell ratio of 1, due to RDF, nearly 10% of the 6-T SRAM cells will malfunction at the 35 nm generation due to RDF, even under conditions entirely free of supply or signal noise.

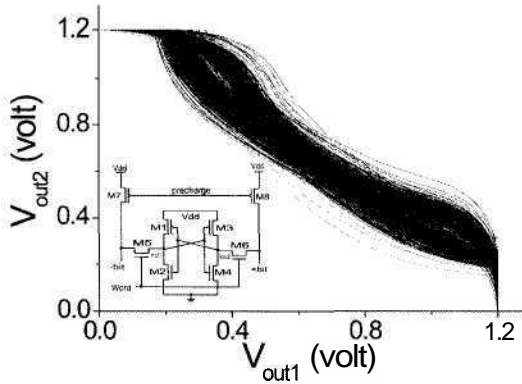


Fig. 4. The static transfer characteristics of 200 statistical SRAM circuit simulations (Inset: Circuit schematic of CMOS SRAM).

Increasing the SRAM cell ratio has two benefits for improved static noise margin (SNM) behaviour. Firstly, a larger cell ratio directly improves cell stability - reflected in an improved mean SNM value, μ . Secondly, a larger cell ratio implies a larger driver W/L ratio, which will reduce the magnitude of the characteristic fluctuations caused by random doping effects in the driver transistors, reducing somewhat the deviation σ of the SNM. Figure 5 clearly shows these benefits.

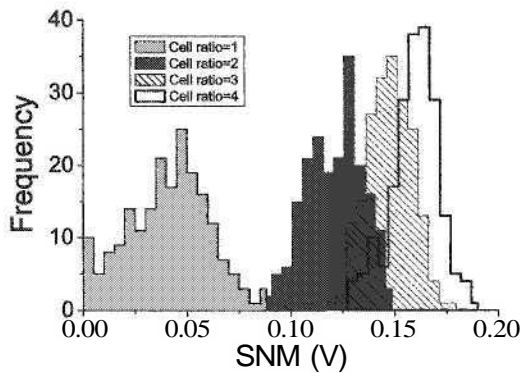


Fig. 5 Distribution of SNM over an ensemble of 200 SRAM cells for the 35 nm generation.

As a guideline, “ $\mu-6\sigma$ ” is required to exceed approximately 4% of the supply voltage to achieve 90% yield for 1Mbit SRAMs.¹³ If we consider only the fluctuations caused

solely by RDF (ignoring all other IPFs), the cell ratio should be at least 3. If other intrinsic fluctuation sources are taken into account, a larger cell area will be required in order to achieve reasonable yield. This implies that new design concepts, such as bias control technology,¹⁴ may be needed in order for SRAM to continuously gain the benefits of further bulk CMOS scaling.

The results for scaled SRAM are shown in Fig. 6. For the 13 nm generation, around 2% of SRAM cells with cell ratio of 2 are not readable even under ideal conditions. The normalized a of SNM increases from 0.13 for the 25 nm generation, to 0.19 for the 18 nm generation, and reaches 0.45 for the 13 nm generation. Although the 25 nm generation is three times better with respect to SNM fluctuation performance compared with the 13 nm generation at cell ratio of 2, it still cannot meet the “ $\mu-6\sigma$ ” test on yield control.

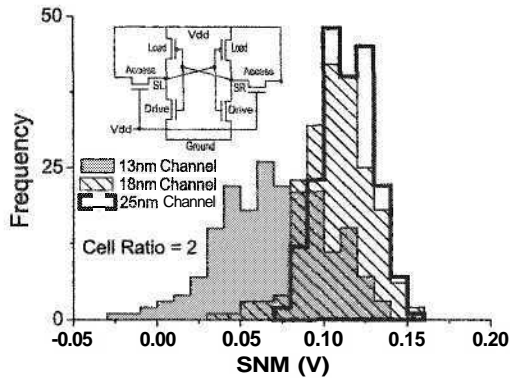


Fig. 6. Distribution of SNM over an ensemble of 200 SRAM cells for 25, 18 and 13 nm generations.

Previously we have demonstrated that if both driveability ratio and cell ratio are used as design parameters to achieve a trade-off between the write noise margin (the minimum DC noise voltage needed to fail to flip the cell state during a write period where the intention was to change that cell state) and SNM,¹⁴ the scalability of SRAM can be improved in the decananometer regime. In this study, instead of changing physical driveability of the access transistor, an access transistor gate bias control technique is applied to modify the driveability ratio during the reading period. As a result, this approach does not compromise the write noise margin performance since during the writing period the circuit can be restored to the optimized value for the writing operation. Figure 7 shows the improvement in the SNM performance introduced by the access transistor gate bias control. For 25 nm and 18 nm generations, every 10% decrease in the access transistor gate bias voltage will introduce approximately 20% improvement of the mean SNM value and the normalised a . For the 13 nm generation, approximately 30% SNM improvement can be achieved in exchange of a 10% decrease in the access transistor gate voltage.

Although issues associated with memory read time are not as critical as SNM for SRAM functionality, they will affect the overall system performance. SRAM read time

degradation is the major setback of the access transistor gate bias control approach, and needs to be taken into account in SRAM cell design. In the transient simulation of SRAM intrinsic reading performance, we assume 0.05 pF bit capacitance. In order to have sufficient noise margin, we also assume that the threshold for the sense amplifier is half of V_{dd} . The read time is measured as the time taken for the bit line voltage to drop to the sense threshold (intrinsic reading time). The simulation results are illustrated in Fig. 8.

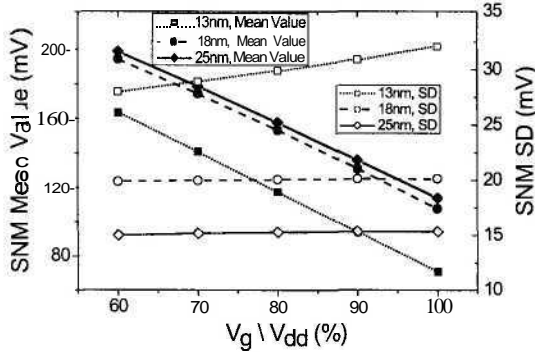


Fig. 7. Mean value and SD of SNM against access transistor gate bias, where cell ratio is 2.

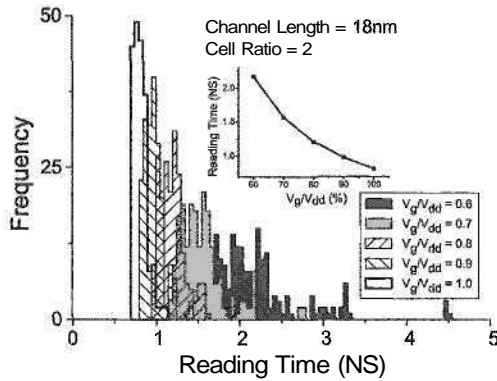


Fig. 8. Intrinsic SRAM cell reading time distribution against access transistor gate bias. Inset: mean value of reading time against access transistor gate bias.

It is clear that the access transistor gate bias has a dramatically impact on the SRAM cell intrinsic read time performance. For the 18 nm generation, a 40% drop in gate voltage will introduce approximately 150% degradation in the mean value and the σ of the read time. Unlike the impact on SNM performance, the impact of gate bias condition on the read time has the non-linear behavior, as illustrated in the inset. The 25 and 13 nm generations show a similar trend. This however provides some trade-off space. In this study, 80% of V_{dd} is chosen as the optimized access transistor gate bias, which roughly can transfer the similar level of loss on *intrinsic* reading speed into much needed SNM improvement. Usually the intrinsic reading speed is much higher than overall reading

speed, and the benefit in terms of SNM improvement outweighs the detrimental impact on overall reading speed loss. This also indicates that new system design philosophy is needed in order to tackle the IPF challenge in decananometer regime design: Instead of increasing system speed, the emphasis could be shifted into improving the system functionality by increasing the device density.

4. Conclusions

IPF is one of the main sources of device mismatch in decananometer regime. A hierarchical device to circuit simulation methodology has been presented which allows us to investigate the impact of IPF on circuit functionality and performance. The effects of RDF on SRAM has been discussed in detail and results indicate that new design concepts are needed in order to tackle the challenges associated with IPF. It is proposed that this hierarchical simulation methodology can provide a vital contribution in the development of such new design concepts.

Acknowledgments

This work was supported by the UK Engineering and Physical Sciences Research Council GR/R47325, and has been partly funded by the European Commission under the frame of the Network of Excellence "SINANO" (Silicon-based Nanodevices, IST-506844).

References

1. A. Asenov *et al.*, *IEEE Trans. Electron Devices* **50**, 1254-1260 (2003).
2. <http://public.itrs.net>.
3. D. Blaauw and K. Chopra, CAD tools for variation tolerance, *Proc. 42nd Design Automation Conf. (DAC)*, Anaheim CA, USA, 766 (2005).
4. *Taurus User's manual*, (Synopsys, 2003).
5. S. Inaba, K. Okano, S. Matsuda *et al.*, *IEEE Trans. Electron Devices* **49**, 2263-2270 (2002).
6. A. Asenov, M. Jaraiz, S. Roy, G. Roy *et al.*, Integrated atomistic process and device simulation of decananometre MOSFETs, *Proc. SISPAD 2002*, Kobe, Japan, 87 (2002).
7. G. Roy, A. R. Brown, A. Asenov, and S. Roy, *Superlattices and Microstructures* **33**, 327-334 (2003).
8. A. R. Brown, F. Adamu-Lema, and A. Asenov, *Superlattices and Microstructures* **34**, 283 (2004).
9. A. R. Brown, G. Roy, and A. Asenov, *Proc. ESSDERC 06*, Montreaux, Switzerland (2006).
10. G. Ferrari, J. R. Watling, S. Roy *et al.*, *J. Comput. Electronics*, in press (2006).
11. X. Xi, <http://www-device.eecs.berkeley.edu/~bsim3>.
12. B. Cheng, S. Roy, and A. Asenov, Compact model strategy for studying the impact of intrinsic parameter fluctuations on circuit performance, *Proc. 11th Int. Conf. MIXDES Design of Integrated Circuits and Systems*, Szczecin, Poland (2004).
13. P. A. Stolk, H. P. Tuinhout, R. Duffy *et al.*, CMOS device optimisation for mixed-signal technologies, *IEDM Tech. Digest*, 215 (2001).
14. B. Cheng, S. Roy, and A. Asenov, Low power, high density CMOS 6-T SRAM cell design subject to 'atomistic' fluctuations, *Proc. 7th European Workshop on Ultimate Integration of Silicon (ULIS2006)*, Grenoble, France (2006).

HEMT-BASED NANOMETER DEVICES TOWARD TERAHERTZ ERA

EIICHI SANO

*Research Center for Integrated Quantum Electronics, Hokkaido University, North 13, West 8, Kita-ku,
Sapporo 060-8628, Japan
esano@rciqe.hokudai.ac.jp*

TAIICHI OTSUJI

*Research Institute for Electronic Communications, Tohoku University, Katahira 2-1-1, Aoba-ku,
Sendai 980-8577, Japan
otsuji@riec.tohoku.ac.jp*

The terahertz region is one of the unexplored bands. This paper first reviews the present status of conventional high-speed devices, especially InP-based high electron mobility transistors (HEMTs), and addresses the technological problems facing the goal of terahertz operation. As an alternative approach to solve these problems, we developed a plasmon-resonant photomixer for realizing a coherent terahertz continuous-wave source. Preliminary results on electromagnetic response to impulsive photoexcitation at room temperature are reported briefly.

Keywords: HEMT; plasmon; photomixer.

1. Introduction

Millimeter waves are becoming more popular in our daily lives. For example, the 76-GHz radars used as the “eyes” in adaptive-cruise-control systems have been installed in several types of luxury cars. Millimeter waves have the advantages over microwaves of shorter wavelengths and higher frequencies. A shorter wavelength improves the spacial resolution of radars. In addition, the radars using millimeter waves can see objects through heavy rain or thick fog, conditions in which ordinary visible cameras cannot see. Millimeter-wave imaging systems can be used for applications such as assisting airplane landing in fog or detecting concealed weaponry, among others.¹

The higher frequency of millimeter waves over microwaves is indispensable for increasing transmission bitrates in wireless communication systems. Future ubiquitous networks will deploy a variety of information-processing and storage devices in the home, office, and public environments. These devices will be fixed or mobile and will get their energy from power lines, batteries, or even from the environment. Continued development of wireless communications, device-to-access point communications, and photonic IP networks will thus be crucial for realizing ubiquitous networks. In particular, a broad bandwidth will be needed for transmitting a huge amount of data over the

network. The bitrate per line is the metric for wired (i.e., metallic or optical fiber) communication systems. However, the bitrate per unit area (or volume) is an important factor in evaluating the performance of wireless communication systems. To expand the metric for wireless communications systems, it is necessary to increase the carrier frequency and spectral efficiency. The millimeter-wave band and future terahertz band are very attractive because of their high frequency and high directivity. Although high directivity is unsuitable for broadcasting, it can help to improve spectral efficiency when space multiplexing is employed.

Monolithically integrated circuits are indispensable parts of communications equipment, especially mobile wireless terminals. Device scaling has been the primary guideline for enhancing device performance (i.e., speed). However, this guideline faces difficulties of further reducing the device size when the few-tens-of-nanometers region has been reached. In addition to trying to solve these problems in the case of conventional devices, a great deal of effort has been expended in producing terahertz devices based on novel operation principles. Of these devices, plasmon-based ones are very attractive candidates from the viewpoint of room-temperature operation.²⁻⁶

This paper first reviews the present status of millimeter-wave monolithically-integrated circuits (MMICs) and high-speed device technologies for fabricating MMICs. It then addresses the technological problems in nanometer-scale high electron mobility transistors (HEMTs). Finally, our recent work on a newly developed plasmon-resonant photomixer is described.

2. Present Status of Millimeter-Wave MMICs

Figure 1 shows the evolution of high-frequency analog and high-speed digital integrated circuits fabricated with semiconductors. Although the operating frequency of analog circuits has stayed at around 200 GHz for a decade, the operating bitrate of digital circuits

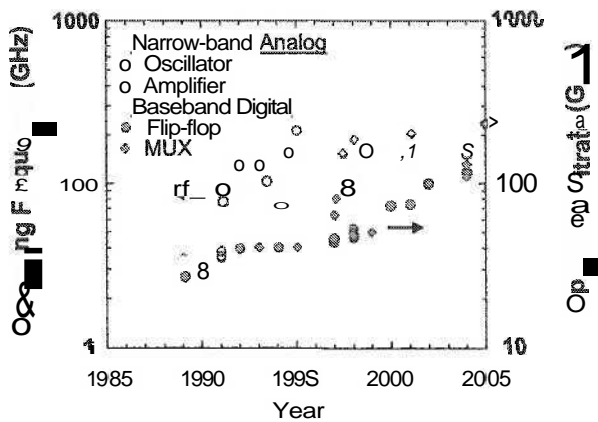


Fig. 1. Evolution of high-frequency analog and high-speed digital integrated circuits.

has gradually increased and reached 100 Gb/s.⁷⁻¹¹ Despite the rapid increase in the operating bitrate of digital circuits, de-terahertz digital processing will be impossible in the near future. This means that the terahertz band will be used as RF carriers for wireless communication systems or intermediate frequencies for optical fiber communication systems.⁵

A coherent continuous-wave (CW) source is one of the key devices for constructing all the electronic and photonic systems. Figure 2(a) shows the room-temperature output power versus oscillation frequency from the microwave to optical regions, and Fig. 2(b) shows the output power versus frequency characteristics of transistor oscillators. As is well known, a terahertz gap exists between electronic and photonic oscillators. Laser diodes using energy-level transition find it difficult to produce terahertz waves which correspond to room-temperature thermal energy. The output power of transistor oscillators is degraded over 100 GHz. This is partly due to the increase in interconnection loss between the active devices as well as insufficient RF device performance. Oscillators operating at over 100 GHz, with sufficient output power, are still fabricated with old-fashioned two-terminal devices such as Gunn diodes. With these devices, however, it is very difficult to achieve circuit functions (like amplification and mixing) other than oscillation. A high level of integration will be necessary for fabricating low-cost and compact wireless equipment. InP-based high electron mobility transistors (HEMTs) and heterojunction bipolar transistors (HBTs) are superior to Si- and GaAs-based devices in terms of speed performance.

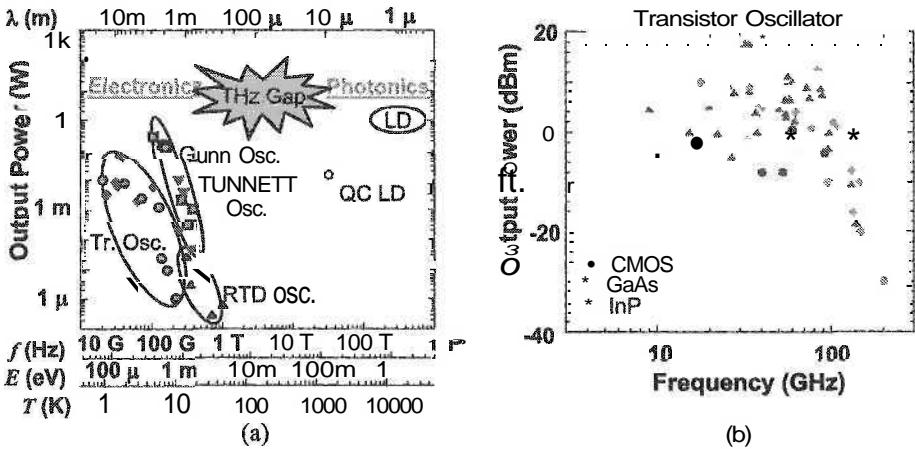


Fig. 2. (a) Room-temperature output power versus oscillation-frequency characteristics from microwave to optical regions and (b) detail of transistor oscillators.

3. Problems in Nanometer-Scale HEMTs

The current-gain cutoff frequency (f_T) and maximum oscillation frequency (f_{max}) are commonly used as measures for the performance of high-speed devices. Figure 3 shows the evolution of f_T and f_{max} of InP-based HEMTs along with their counterparts. f_T 's have increased gradually and reached 560 GHz¹² and 600 GHz¹³ for HEMTs and HBTs, respectively. On the other hand, f_{max} of InP-based devices has saturated in recent years.

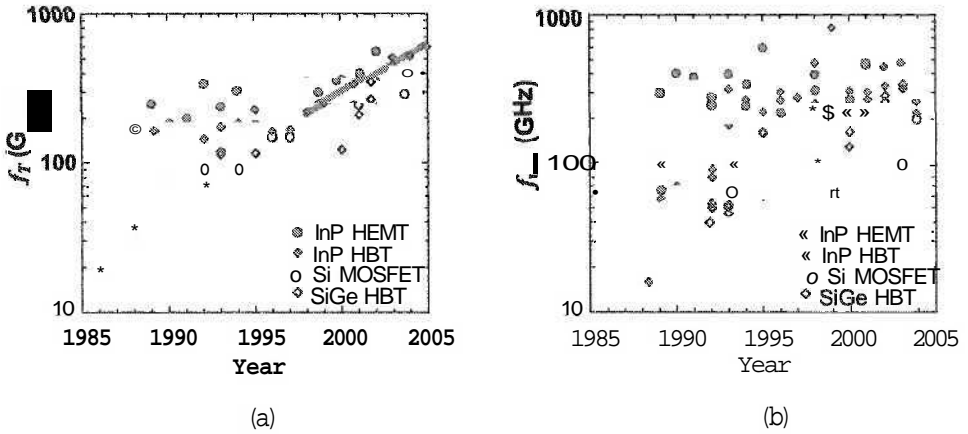


Fig. 3. Evolution of (a) f_T and (b) f_{max} .

When developing high-speed-device technologies, we must understand which device parameter is the key to enhancing circuit performance. The maximum transition frequency, at which the real part of the two-terminal impedance vanishes, is given by

$$f_{tr} \sim \frac{f_{max}}{2} \sqrt{\frac{3C_{GD}}{C_{GS}} + 1} \tag{1}$$

for Colpitts oscillators using FETs.¹⁴ Here C_{GS} and C_{GD} are the gate-to-source and gate-to-drain capacitances, respectively. Similar equations may be derived for other circuit configurations of oscillators. It is very natural that the transition frequency depends on f_{max} of the device. On the other hand, the maximum operating speed for flip-flops depends on f_T as well as the transconductance g_m and C_{GD} .¹⁵ These dependences of oscillators and flip-flops on device parameters in conjunction with the evolution of f_T and f_{max} shown in Fig. 3 may be one of the reasons for the trend in circuit performance shown in Fig. 1.

Device scaling has been the primary guideline for enhancing device speed. Figure 4 shows the well-known dependences of f_T and f_{max} on the gate length of FETs. It is clear

that f_T and, especially, f_{max} tend to saturate below 100 nm. This saturation is the short-channel effects as will be discussed later. According to the International Technology Roadmap for Semiconductors (ITRS 2003 Edition), f_T target value for Si nMOS will be 1.4 THz by 2018.¹⁶ However, solutions for attaining this high value are still unclear.

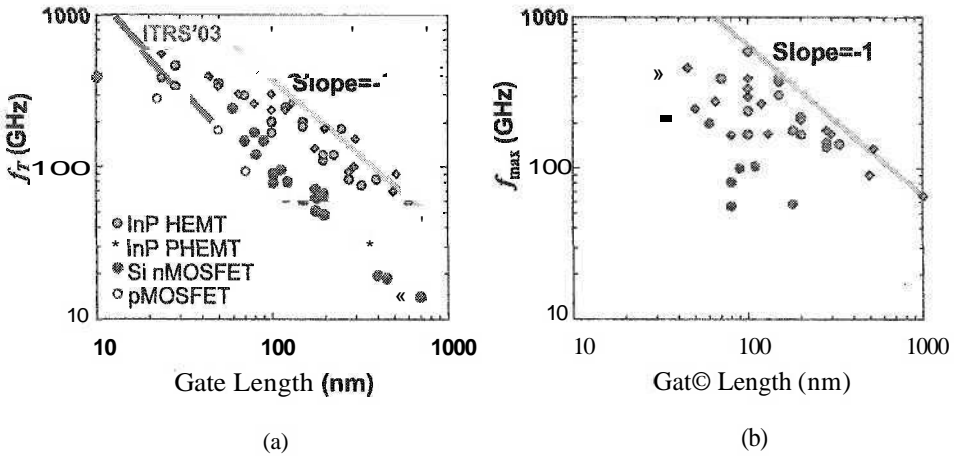


Fig. 4. Gate-length dependences of (a) f_T and (b) f_{max} .

We analyzed InP-based HEMTs using a drift-diffusion (DD)-based simulator for clarifying problems for nanometer-scale HEMTs. Figure 5 shows a simplified geometric model for HEMTs. An effective potential given by Ferry¹⁷ was used for taking into account of energy quantization in the channel layer and the electron transport through the cap layer.¹⁸ Figure 6 shows the threshold-voltage shift versus gate length characteristics. Open symbols denote simulated results while solid symbols denote experimental results reported by Suemitsu *et al.*¹⁹ and Endoh *et al.*²⁰ It is clear that 20-nm gate-length is the critical length restricted by the threshold-voltage shift for this material system.

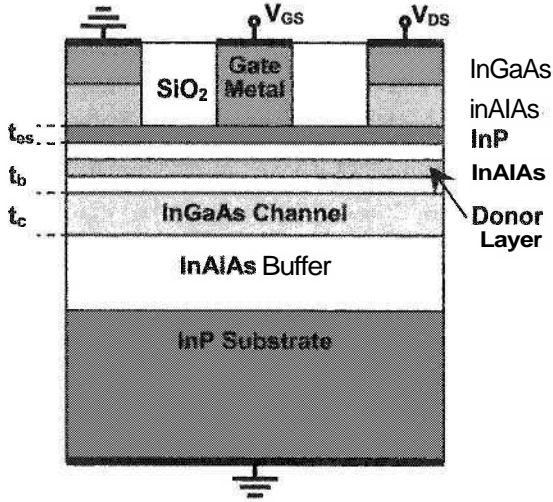


Fig. 5. Simplified geometric model for InP-based HEMTs.

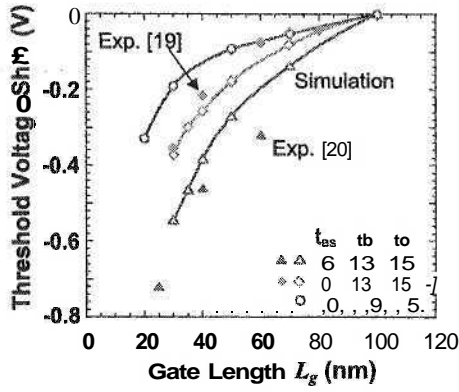


Fig. 6. Threshold-voltage shift versus gate length characteristics.

The time constant determining f_T is given by

$$\begin{aligned} \tau_T &= \frac{(C_{GS} + C_{GD})[1 + g_d(R_S + R_D)]}{g_m} + C_{GD}(R_S + R_D) \\ &= \tau_i + (\tau_i g_d + C_{GD})(R_S + R_D) \equiv \tau_i + \tau_{ext} \end{aligned} \quad (2)$$

where g_d is drain conductance, and R_S and R_D are source and drain parasitic resistances, respectively. The first part in the last equation is intrinsic and the second is extrinsic. Figure 7 shows the ratio of the intrinsic time constant to the extrinsic one reported in the literature. The extrinsic part is dominant for a gate length below 40 nm. A decreased ratio results in the f_T saturation as shown in Fig. 4(a).²¹ This means that a reduction in the extrinsic part is indispensable for increasing f_T in the few-tens-of-nanometers region.

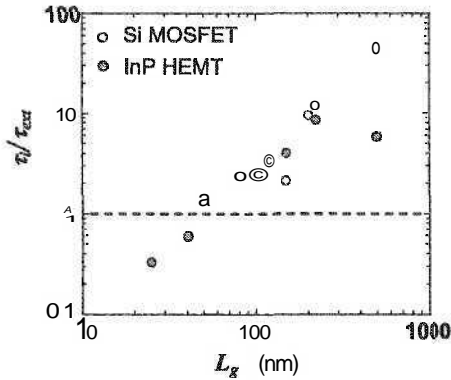


Fig. 7. Ratio of intrinsic time constant τ_i to extrinsic one τ_{ext} .

The gate electrode geometry was examined in terms of maximizing f_{max} . An expression given by Rohdin *et al.*²² was used for the gate resistance R_G . The gate-to-drain capacitance COD was calculated by solving Laplace's equation for the geometry shown in Fig. 8(a). Other device parameters were assumed to be the same as those for 0.1- μm gate HEMT.²¹ An expression for the maximum available gain given by Das²³ was used for determining f_{max} . Figure 8(b) shows dependence of calculated f_{max} on gate geometry, where the distance between the center of the gate and the left edge of the drain is kept constant. It is clear that the geometry dependence is very small. We believe that state-of-the-art HEMTs have dimensions similar to those shown in Fig. 8(b) and are almost completely optimized for maximizing f_{max} .

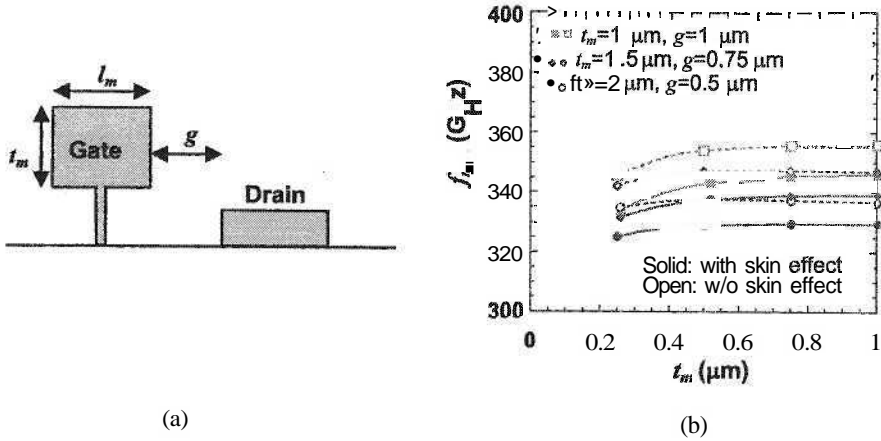


Fig. 8, (a) Gate-electrode model and (b) f_{\max} dependence on l_m as a function of l_m and g .

4. Plasmon Resonant Photomixer

In addition to trying to solve the above-mentioned problems in regards to conventional devices, a great deal of effort has been expended in achieving terahertz devices based on novel operation principles. Of these, plasmon-based devices are very attractive candidates from the viewpoint of room-temperature operation.²⁻⁶ Stimulated by theoretical work done by Mikhailov²⁴, we devised a novel grating-bicoupled plasmonic structure for radiating terahertz electromagnetic waves.²⁵ Figure 9 illustrates our plasmon-resonant photomixer.²⁵ When two co-linearly-polarized CW laser beams having slightly different frequencies f_0 and $f_0 + \Delta f_0$ are absorbed in the 2D channel, the generated electrons induce plasma instability. Doubly interdigitated grating gates convert the non-radiative plasma waves to radiative waves. The vertical cavity formed in between an optically transparent mirror and the 2D channel may serve like that in conventional lasers.

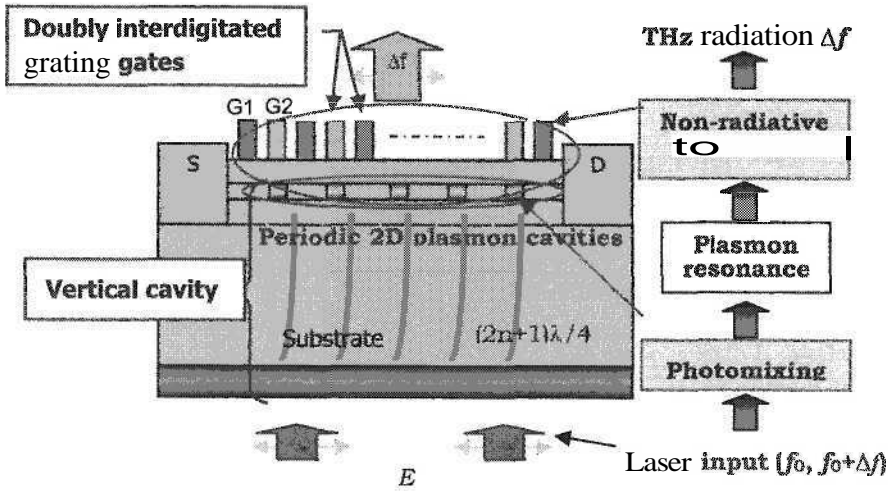


Fig. 9. Schematic cross-sectional view of plasmon-resonant photomixer.

We analyzed distributions of electrostatic potential, electron density, and electron velocity in a dark condition by using a modified version of our DD-based simulator. Only a 1.5 unit cell could be simulated due to a restriction on computer resources. An example of the simulation results is shown in Fig. 10. The results show that electrostatic potential, electron concentration, and velocity distributions can be modulated by changing gates G1 and G2. From these results, we propose here the following possible mechanism for photo-induced plasmon excitation in our device. Photoelectrons are dominantly generated in the regions under G1 and are then injected into the plasmon cavities under G2. This process can extensively accelerate the plasmon instability in both region G1 (due to electron outflux) and region G2 (due to electron influx).

Applying the theoretical work mentioned above and electromagnetic fullwave simulations,²⁵ we designed and fabricated our first sample of plasmon-resonant photomixer.²⁶ The epitaxial layer structure was based on the GaAs material system. Gate lengths for G1 and G2 were 70 nm and 350 nm, respectively. We measured the impulse response of the photomixer by electrooptic sampling. A 1.5- μm , 1-mW, 70-fs laser pulse was used as pump and probe beams. Figure 11 is an example of the measurement results. Figure 11(a) shows transient responses, and Fig. 11(b) shows Fourier-transformed spectra. Marked emission can be seen when the device is biased. Comparing a sample without an ITO mirror (not shown here), we found that the field strength was amplified by the vertical cavity. Radiation peaks shown in Fig. 11(b) correspond to the plasmon modes. Further experiments are required to clarify the photomixer operation in detail. We are now conducting such experiments, which will be reported in the near future. It should be noted that our plasmon-resonant photomixer is based on HEMT-fabrication technology and can be monolithically integrated with IF and baseband circuits using

conventional HEMTs. We believe that the plasmon-resonant photomixer is one of the most promising candidates for terahertz CW sources.

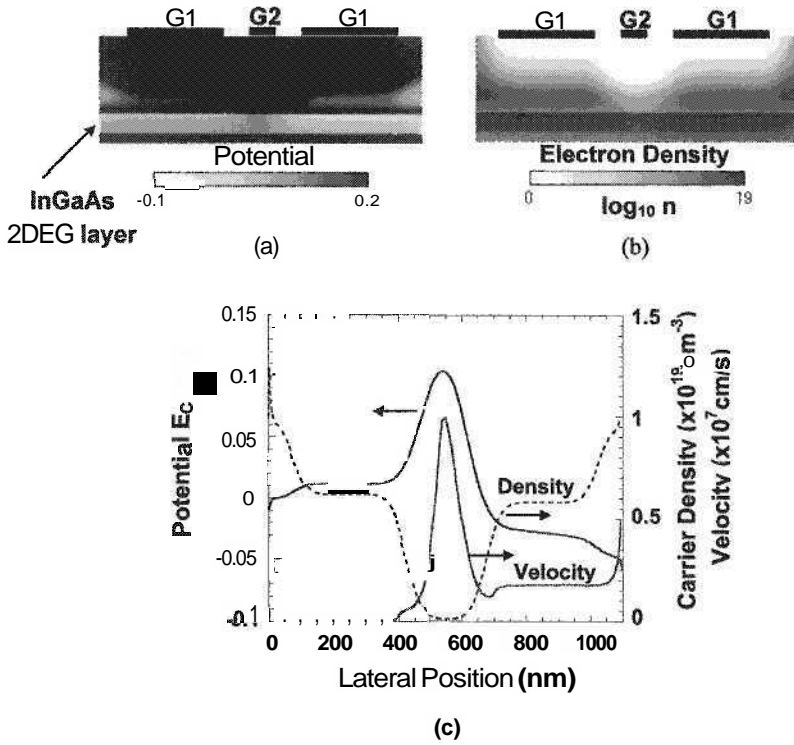
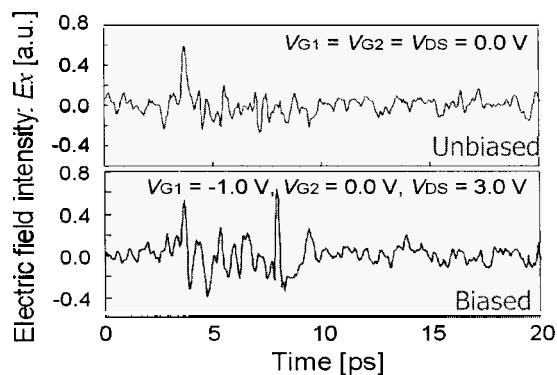
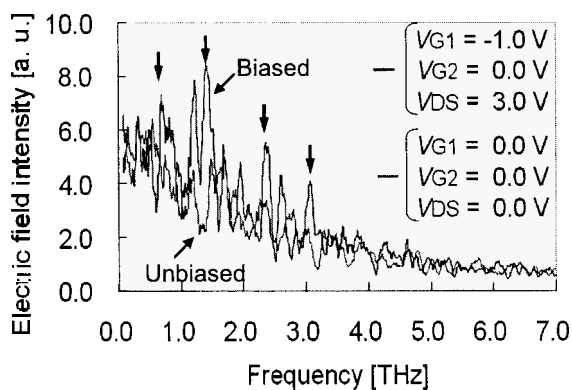


Fig. 10. (a) Conduction-band edge potential distribution, (b) electron density distribution, and (c) lateral distributions. Lengths of G1 and G2 are 300 nm and 100 nm, respectively. $V_{G1}=0$ V, $V_{G2}=-1.5$ V, and $V_{DS}=0.05$ V.



(a)



(b)

Fig. 11. (a) Transient waveform of fabricated plasmon-resonant photomixer and (b) Fourier-transformed spectra.

5. Conclusions

We have reviewed the present status of conventional high-speed devices, especially InP-based high electron mobility transistors (HEMTs), and addressed the technological problems regarding terahertz operation. As an alternative approach to attain terahertz operation, we devised a novel grating-bicoupled plasmonic structure for radiating terahertz electromagnetic waves. Electromagnetic response to impulsive photoexcitation was measured at room temperature by electrooptic sampling. We successfully observed emission of terahertz electromagnetic radiation stimulated by photo-induced plasmon instability at room temperature. These results demonstrate that the plasmon-resonant photomixer is one of the most promising candidates for terahertz CW sources.

Acknowledgments

The authors would like to thank Profs. V. Ryzhii, M. Dyakonov, S. Mikhailov, and T. Asano for their valuable comments, Dr. T. Nagatsuma for the electrooptic sampling, and Dr. Y. M. Meziani, M. Hanabe, and T. Nishimura for their support.

This work was financially supported in part by the Strategic Information and Communications R&D Promotion Programme (SCOPE) from MIC, Japan, and by Grand-in-Aid for Science Research (S) and (B) made available by MEXT, Japan.

References

1. L. Yujiri, M. Shoucri, and P. Moffa, *IEEE Microwave Magazine* **39** (2005).
2. M. Dyakonov and M. Shur, *Phys. Rev. Lett.* **71**, 2465 (1993).
3. K. Hirakawa, K. Yamanaka, M. Grayson, and D. C. Tsui, *Appl. Phys. Lett.* **67**, 2326 (1995).
4. M. Dyakonov and M. Shur, *IEEE Trans. Electron Devices* **43**, 380 (1996).
5. T. Otsuji, S. Nakae, and H. Kitamura, *IEICE Trans. Electron.* **E84-C**, 1470 (2001).
6. W. Knap, J. Lusakowski, T. Parenty, S. Bollaert, A. Cappy, V. V. Popov, and M. S. Shur, *Appl. Phys. Lett.* **84**, 2331 (2004).
7. D. Dawson, L. Samoska, A. K. Fung, K. Lee, R. Lai, R. Grundbacher, P. -H. Liu, and R. Raja, *IEEE Microwave Wireless Components Lett.* **15**, 874 (2005).
8. S. E. Rosenbaum, B. K. Kormanyos, L. M. Jelloian, M. Matloubian, A. S. Brown, L. E. Larson, L. D. Nguyen, M. A. Thompson, L. P. B. Katehi, and G. M. Rebeiz, *IEEE Trans. Microwave Theory Tech.* **43**, 927 (1995).
9. K. Murata, K. Sano, H. Kitabayashi, S. Sugitani, H. Sugahara, and T. Enoki, *IEDM Tech. Dig.*, 937 (2002).
10. Y. Suzuki, Z. Yamazaki, Y. Amamiya, S. Wada, H. Uchida, C. Kurioka, S. Tanaka, and H. Hida, *IEEE J. Solid-State Circuits* **39**, 2397 (2004).
11. Z. Griffith, M. Dahlsrom, M. J. W. Rodwell, X. -M. Fang, D. Lubyshev, Y. Wu, J. M. Fastenau, and W. K. Liu, *IEEE Electron Device Lett.* **26**, 11 (2005).
12. Y. Yamashita, A. Endoh, K. Shinohara, K. Hikosaka, T. Matsui, S. Hiyamizu, and T. Mimura, *IEEE Electron Device Lett.* **23**, 573 (2002).
13. W. Hafez and M. Feng, *Appl. Phys. Lett.* **86**, 152101 (2005).
14. E. Sano, *IEICE Trans. Electron.* **J87**, 413 (2004) (in Japanese).
15. E. Sano, *IEICE Trans. Electron.* **E86-C**, 1879 (2003).
16. <http://public.itrs.net/Files/2003ITRS/Home2003.htm>.
17. D. K. Ferry, *Superlat. Microstruct.* **28**, 419 (2000).
18. E. Sano, *Jpn J. Appl. Phys.* **42**, 4261 (2003).
19. T. Suemitsu, T. Ishii, H. Yokoyama, Y. Umeda, T. Enoki, Y. Ishii, and T. Tamamura, *IEDM Tech. Dig.*, 223 (1998).
20. A. Endoh, Y. Yamashita, K. Shinohara, M. Higashiwaki, K. Hikosaka, T. Mimura, S. Hiyamizu, and T. Matsui, *IPRM Conf. Proc.*, 448 (2001).
21. T. Enoki, H. Yokoyama, Y. Umeda, and T. Otsuji, *Jpn J. Appl. Phys.* **37**, 1359 (1998).
22. H. Rohdin, N. Moll, C. Su and G. S. Lee, *IEEE Trans. Electron Devices* **45**, 2407 (1998).
23. M. B. Das, *IEEE Trans. Electron Devices* **35**, 604 (1988).
24. S. A. Mikhailov, *Phys. Rev.* **B58**, 1517 (1998).
25. T. Otsuji, M. Hanabe, T. Nishimura, and E. Sano, *Opt. Exp.* **14**, 4815 (2006).
26. T. Otsuji, M. Hanabe, Y. M. Meziani, and E. Sano, *Device Research Conf.*, 193 (2006).

PLASMA WAVES IN TWO-DIMENSIONAL ELECTRON SYSTEMS AND THEIR APPLICATIONS

VICTOR RYZHII, IRINA KHYROVA, MAXIM RYZHII, and AKIRA SATOU

*Computer Solid State Physics Laboratory, University of Aizu
Aizu-Wakamatsu 965-8580, Japan*

TAICHI OTSUJI

*Research Institute of Electrical Communication, Tohoku University
Sendai 980-8577, Japan*

VLADIMIR MITIN

*Department of Electrical Engineering
University at Buffalo, SUNY, Buffalo, NY 14260, USA*

MICHAEL S. SHUR

*Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute, Troy, NY 12180, USA*

We overview the plasma properties of a two-dimensional electron gas in semiconductor heterostructures as well as new concepts and proposals of novel terahertz devices (detectors, sources utilizing plasma instabilities, and photomixers) based on these heterostructures.

Keywords: Two-dimensional electron gas; plasma oscillations; terahertz detector, oscillator, and photomixer.

1. Introduction

There is an increasing demand in compact devices operating in the terahertz (THz) range of frequencies. These devices include detectors and modulators of THz radiation, frequency multipliers, and coherent and tunable sources. The creation of compact semiconductor sources of THz radiation appears to be the most challenging task. The conventional classical semiconductor devices such as transistors of different types, Gunn diodes, tunneling diodes, and diodes exploiting electron-transit-time effect can only reach the low bound of the THz frequency range and are fairly ineffective in this range. The devices utilizing the electron transitions between quantized energy levels (quantum cascade lasers) can generate the THz radiation with the frequencies of few THz and above. Thus, a significant portion of the THz

range is not covered yet by compact and effective devices, so that the so-called THz gap still exists between 300 GHz and 3 THz. The attempts to use electron plasma wave effects in different semiconductor structures for the generation of the THz radiation have started in late 50's - early 60's. The main idea was to take advantage of the plasma wave self-excitation (plasma wave instability) under the conditions of sufficiently strong deviation of an electron system from equilibrium. This idea was stimulated by the fact that hot gaseous plasmas in natural and artificial systems are unstable toward the excitation of plasma waves of different types. However, in three-dimensional semiconductor structures the plasma instabilities are usually suppressed by electron collisions with impurities and phonons. Hopes to overcome this obstacle are associated with two-dimensional electron gas (2DEG) heterostructure with high electron mobility. Due to spatial separation of the donors and electrons in such systems, the electron collisions can be substantially weakened, particularly at low temperatures. Characteristic frequencies of the electron plasma waves in 2DEG systems with realistic parameters fall within the THz range. This leads to the feasibility of realizing of the conditions at which the collisional damping of plasma waves is weak. In such a case, 2DEG systems can exhibit pronounced resonant response associated with high quality factor of the plasma waves, so that these systems can be used as resonant cavities (resonators) in different THz devices. Apart from this, different mechanisms of the plasma wave self-excitation become possible that opens up the prospects of novel THz sources.^{1,2} The generation of THz radiation in 2DEG transistor-like heterostructures attributed to a plasma instability and plasmon-assisted photomixing observed in the experiments by different research groups.

This paper deals with the following:

First, we discuss the plasma properties of 2DEG semiconductor heterostructures (based on III-V compounds) with selective doping. The plasma waves in heterostructures with graphene are considered as well. Graphene, i.e., a monolayer of carbon atoms packed into a dense honeycomb crystal structure, attracted substantial attention recently as a unique representative of the thinnest crystalline layer to confine 2DEG. The specific of the plasma effects in graphene-based heterostructures is associated with the massless (neutrino-like) energy spectrum of electrons and holes.³

Second, we review our concepts and proposals of novel THz devices based on 2DEG systems and utilizing plasma effects. We consider mechanisms of plasma wave instabilities in 2DEG systems (high-electron mobility transistors, transistor structures with resonant-tunneling, lateral Schottky junction diodes and transistors, and **others**).⁴⁻⁷ After that, we discuss novel devices for detection and frequency multiplication of THz radiation.⁸⁻¹⁰ Then, we discuss the operation of devices utilizing resonant photomixing due to the excitation of the plasma oscillations by modulated light. These devices are based on the integration of a plasma resonant ungated or gated cavity with a p-i-n photodiode, uni-travelling-carrier photodiode,¹¹ or quantum-well infrared photodetector.¹² Finally, we analyze prospects of the appli-

cations of graphene gated heterostructures for THz applications.

2. Plasma Waves and Oscillations in 2DEG Channels

The 2DEG system is governed by a quantum or classical kinetic equation coupled with the Poisson equation for the self-consistent electric potential ψ :

$$\frac{\partial^2 \psi}{\partial z^2} + \frac{\partial^2 \psi}{\partial x^2} = \frac{4\pi e}{k} (\Sigma_d - \Sigma) \cdot \delta(z), \quad (1)$$

where e is the electron charge, k is the dielectric constant, Σ_d is the donor sheet concentration, $\delta(z)$ is the Dirac delta function, axis x is directed in the 2DEG channel plane, and axis z is directed perpendicular to this plane. Since the electron concentration in 2DEG systems in most interesting heterostructures and devices based on these heterostructures is rather high, the electron-electron interactions play an important role. To increase the electron mobility in 2DEG, the donors are usually placed at some distance from the 2DEG channel so the channel and the doped layer are separated by a spacer layer. As a result, the frequency of electron-electron collisions can markedly exceed the frequency of electron collisions with the donors, residual impurities, and phonons. In this case, 2DEG system can also be described to a good approximation by a hydrodynamic electron transport model. The equations of the latter model comprise the Euler equation and the continuity equation. In the case of 2DEG channel with a highly conducting plane (gate) placed parallel to the channel at sufficiently short distance W (in HEMT-like structures, W is the thickness of the gate structure between the channel and the gate contact, see Fig. 1), the above 2D Poisson equation can be replaced by its simplified consequence

$$\frac{\varphi}{W} = \frac{4\pi e}{k} (\Sigma - \Sigma_d), \quad (2)$$

Here, $\varphi = \varphi(x, t)$ is the electric potential in the 2DEG channel, i.e., $\varphi = \psi|_{z=0}$.

To study the plasma wave propagation and the response of the 2DEG system to external perturbations (incoming signals) the small-signal analysis is used. In such an approach, the ac components of the electron hydrodynamic velocity $u_\omega(x) \exp(-i\omega t)$, the electron sheet concentration and the self-consistent potential are presented as $u_\omega = u_\omega(x) \exp(-i\omega t)$, $\Sigma_\omega = \Sigma_\omega(x) \exp(-i\omega t)$, and $\varphi_\omega = \varphi_\omega(x) \exp(-i\omega t)$, respectively, where ω is the signal frequency. In particular, the linearized equations of the model for a uniform 2DEG channel with the dc electron sheet concentration Σ_0 result in the following dispersion equations for the plasma waves with the frequency ω and the wavenumber q propagating in the x -direction (so that $\varphi_\omega \propto \exp[i(qx - \omega t)]$):

$$\omega \simeq \sqrt{\frac{2\pi e^2 \Sigma_0}{km}} q - i \frac{\nu}{2} \quad (3)$$

in the ungated 2DEG channels, and

$$\omega \simeq \sqrt{\frac{4\pi e^2 \Sigma_0 W}{km}} q - i \frac{\nu}{2} = s q - i \frac{\nu}{2} \quad (4)$$

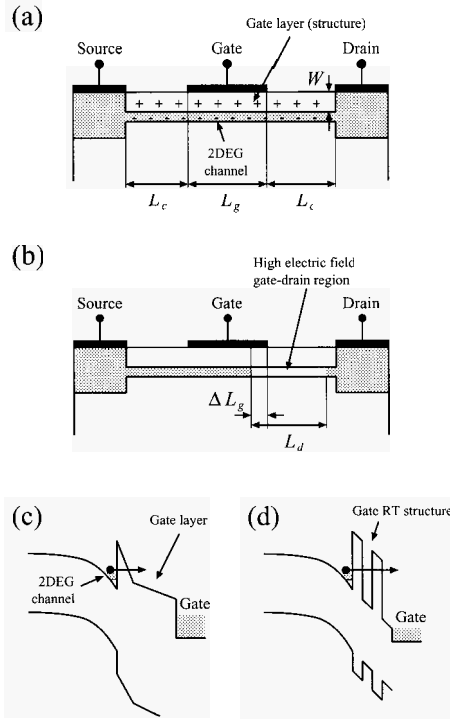


Fig. 1. Schematic view of HEMT structure (a) at moderate drain voltage and (b) in saturation regime (high drain voltage); (c) band diagram of HEMT with doped layer and (d) band diagram of HEMT with resonant-tunneling structure between the channel and the gate. Arrows show electron tunneling from the channel into the gate.

in the gated 2DEG channels with $W \ll L_g$ (L_g is the length of the gated section, i.e., the gate length). Here, $s = \sqrt{4\pi e^2 \Sigma_0 W / km}$ is the plasma wave velocity (phase and group velocity of relatively long plasma waves in the gated 2DEG channels). The second terms in Eqs. (3) and (4) stand for the plasma wave damping due to electron collisions.

As follows from Eqs. (3) and (4), in 2DEG channels with relatively small collision frequency ($\nu \ll \omega$), $\text{Re}\omega \propto \sqrt{\Sigma_0} \sqrt{q}$ ¹³ and $\text{Re}\omega \propto \sqrt{\Sigma_0 W} q$,^{14,15} respectively. The spectra of plasma waves in both ungated and gated 2DEG channels are strongly different from the spectrum of plasma waves in 3DEG, where $\text{Re}\omega$ is virtually independent of q . The damping of the plasma waves in the channels of both types is determined mainly by the electron collision frequency ν : $\text{Im}\omega \simeq \nu/2$. The electron leakage from (injection into) the 2DEG channel due to tunneling or resonant tunneling (RT) through the gate layer or more complex structure [see Figs. 1(c) and 1(d)] can lead to an additional damping or, vice versa, to its suppression. In some cases, the latter processes can result in a negative damping, i.e., in the plasma wave instability (see Sec. 3).

The 2DEG channels usually have limited length and are supplied with highly conducting contacts (for example, the source and drain contacts in the case of field-effect transistor). The edges of the 2DEG channel result in the reflection of the propagating plasma waves and the formation of the standing plasma waves (plasma oscillations). Due to this, some limitations on the plasma wave spectrum (quantization of the spectrum) appear. If the potential at the highly conducting contacts separated by the spacing L is fixed (so that, for example, $\varphi_\omega|_{x=0} = \varphi_\omega|_{x=L} = 0$), a simplified quantization rule gives $q = q_n = \pi n/L$, where $n = 1, 2, 3, \dots$ is the index of the oscillation modes. In particular, Eq. (4) for a HEMT structure with the gate length $L_g \simeq L$ for the fundamental mode ($n = 1$) yields $\omega \simeq \Omega$, where

$$\Omega = \sqrt{\frac{4\pi^3 e^2 \Sigma_0 W}{kmL_g^2}} \quad (5)$$

is the fundamental plasma frequency. In the case of the boundary conditions corresponding to a highly conducting contact at $x = 0$ ($\varphi_\omega|_{x=0} = 0$) and to a free 2DEG channel edge at $x = L$ ($d\varphi_\omega/dx|_{x=L} = 0$), the wavenumbers are $q_n = \pi n/2L$. More detailed and careful consideration shows that actually the frequencies of plasma oscillations depend on the shape and conductivity of the contacts^{16–19} as well as the properties of the substrate.²⁰ Although the effect of the boundary conditions on the plasma frequencies is not so strong and, therefore, can be neglected in many practical applications, the boundary conditions can be essential for the damping and growth (instability) of the plasma oscillations.^{4,21,22}

The quantization rule for the partially gated 2DEG channel (this is common for many field-effect transistor structures in which the channel sections adjacent to the source and drain are ungated) is also complicated by the channel nonuniformity. The ungated sections of the 2DEG channel do not affect significantly the spectrum of plasma oscillations until their length L_c is small in comparison with L_g .⁹ However, when L_c becomes comparable with L_g , the fundamental plasma frequency markedly decreases with increasing L_c . The spectra of plasma oscillations in the 2DEG channels with a periodic system of highly conducting gates (metal grating) depend on the period $L_c + L_g$ and the ratio L_c/L_g .

The plasma oscillation frequencies in the 2DEG channels of the real heterostructure devices fall into the THz range. Indeed, assuming for an ungated channel $\Sigma_0 = 1 \times 10^{12} \text{ cm}^{-2}$ and $L = 1 \times 10^{-4} \text{ cm}$, for a GaAs 2DEG channel, for the fundamental plasma frequency ($n = 1$) one obtains $\omega/2\pi \simeq 1.26 \text{ THz}$. For a similar gated channel with the gate length $L_g \simeq L = 1 \times 10^{-4} \text{ cm}$ and $W = 1 \times 10^{-5} \text{ cm}$, one can arrive at $\omega/2\pi \simeq 1 \text{ THz}$. Since Σ_0 in the gated channels depends on the gate voltage V_g , the plasma frequencies in such channels can be tuned by this voltage. Figure 2 shows the variation of the fundamental plasma frequency $\Omega/2\pi$ and the quality factor $Q = 4\Omega/\pi\nu$ with changing gate voltage V_g calculated for HEMT structures with relatively short ungated gate-source and drain-gate regions ($L_c \ll L_g$).

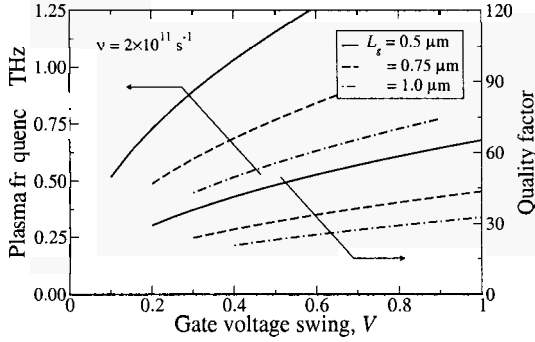


Fig. 2. Fundamental plasma frequency and quality factor vs applied gate voltage calculated for HEMT structures with different gate length ($L_g \gg L_c$).

Owing to nonlinearity of the hydrodynamic equations governing the plasma waves and oscillations, different nonlinear plasma effects in 2DEG channels can be essential and used for practical applications.^{1,4,23–25}

3. Plasma Instabilities

The self-excitation of plasma oscillations due to instability of a steady-state flow of electrons in nonequilibrium 3DEG and 2DEG systems in different heterostructures has been considered as a very attractive method of generating THz radiation. Many proposals to achieve the plasma wave self-excitation are based on the assumption of realizing sufficiently high electron drift velocity of a marked portion of electrons under the applied dc electric field. However, the realization of relatively high electron drift velocity inevitably requires a strong electric field that leads to a significant heating of electron system. The latter promotes a strong suppression of the plasma instability. Dyakonov and Shur have proposed a mechanism of plasma instability in HEMTs in the saturation regime,^{1,4} in which the abovementioned obstacle appears to be eliminated. In this regime, the variations of the drain current J_d with varying drain voltage V_d are rather small, so that the drain-source differential conductivity $\sigma_d = \partial J_d / \partial V_d |_{V_g = const}$ tends to zero when V_d increases. The feature of the conditions of the electron flow at the point between the quasi-neutral section of the 2DEG channel and the section adjacent to the drain contact depleted due to a high electric field [this point is located somewhere near the drain side of the gate edge, so that length, L_d , of the high-electric field region is close to the drain-gate spacing; see Fig. 1(b)], results in an effective reflection of the plasma waves propagating in the HEMT channel from the source to the drain. Such a reflection provides a positive feedback, which can lead to the plasma wave amplification, i.e., to the plasma instability. The condition of the Dyakonov-Shur (DS) instability is given by $\gamma_{DS} = u_g / L_g > \nu / 2$, where u_g is the dc electron drift velocity in the 2DEG channel under the gate. However, at high frequencies, the drain ac current

is determined not only by σ_d but also by the electron transit-time effects due to the delay of electrons propagating in the high-electric field region between the gate and the drain. The electron delay time can be estimated as $\tau_d = L_d/u_d$, where u_d is the average dc electron drift velocity in the HEMT high-electric field region ($u_d \gg u_g$). Taking into account the transit-time effects in question,²² the criterion of the plasma instability in HEMTs can be presented in more general form:

$$\gamma_{DS} + \gamma_{TT} > \nu/2. \quad (6)$$

Here, $\gamma_{TT} = u_d/KL_g$, where K is a numerical coefficient ($K > 1$). The THz emission, which was observed in some experiments,^{26–28} can be attributed to the DS mechanism complicated by the electron-transit-time effects. The self-excitation of plasma oscillations due to the electron-transit-time mechanism is possible because the dynamic conductivity of the high-electric field region becomes negative at certain values of the transit angle $\omega\tau_d$ and because of resonant properties of the 2DEG channel. Along with this, there are speculations that the negative dynamic conductivity of the high-electric field region associated with the quasi-periodic motion of electrons due to their scattering on polar optical phonons and due to inter-valley transitions (similar to those considered theoretically^{29–32} can play a marked role in the THz emission observed experimentally.

As shown theoretically,^{5,33,34} the self-excitation of plasma oscillations in HEMT-like structures is possible if the channel-gate structure exhibits the negative dynamic conductivity. This can occur in HEMTs with electron tunneling through the gate layer [through the top of the gate barrier as shown in Fig. 1(c)] provided that the transit time across this layer corresponds to a proper value of the transit angle.^{33,34} Another option is associated with the realization of a HEMT-like device with resonant-tunneling (RT) structure between the channel and the gate [see Fig. 1(d)].⁵ In these devices, the gate structure plays the role of a "distributed" diode with negative dynamic conductivity, whereas the 2DEG channel serves as a resonant cavity.

Figure 3 shows the structure of a device based on a 2DEG channel with a lateral Schottky junction adjacent to a metal (Schottky) contact.⁷ The operation of this device as a THz oscillator is possible due to combining of dynamic negative conductivity (associated with the delay of the electrons, which tunnel from the metal contact into the 2DEG channel, in the Schottky junction depletion region of thickness l) and a resonant response of the quasi-neutral portion of the channel. The condition of plasma instability associated with the mechanism in question can be expressed via the differential tunneling conductivity, g_t , of the lateral Schottky junction:

$$g_t > \nu/S, \quad (7)$$

where $S > 1$ is a number determined by the structural parameters.

As seen from Eqs. (6) and (7), the realization of the plasma oscillations self-excitation and, hence, the feasibility of THz emission from the devices under con-

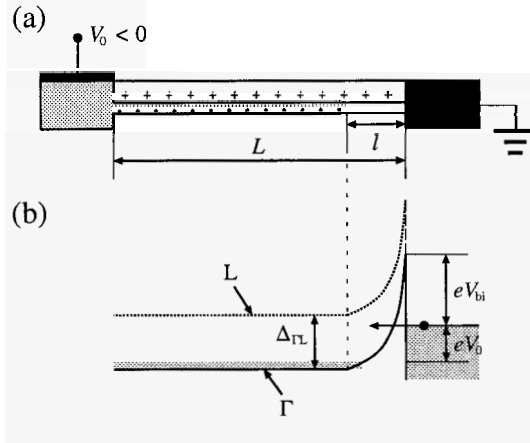


Fig. 3. Schematic view of lateral Schottky junction tunneling-injection transit-time oscillator under reverse bias ($V_0 > 0$). Arrow shows electron tunneling from the metal contact into the Γ -valley of the channel (intervalley separation $\Delta_{\Gamma,L} < eV_0$).

sideration in this section require a low collision frequency, ν , of electrons with impurities and phonons. This implies that the electron mobility in the 2DEG channel ($\mu \propto \nu^{-1}$) must be sufficiently high.

4. Detection of THz Radiation and Frequency Multiplication Using Resonant Excitation of Plasma Oscillations

The plasma effects in HEMT structures like that shown in Figs. 1(a) and 1(b) can be used for resonant detection of THz radiation. As proposed previously,¹ the excitation of plasma oscillations by incoming THz radiation in the HEMT 2DEG channel results in variation of the dc current in the channel or of the dc voltage between the source and drain. This is due to a rectification effect associated with nonlinearity of the hydrodynamic equations governing the plasma oscillations. The observation of the resonant detection of THz radiation in HEMTs was reported in several publications (see, for instance, Refs. 35 and 36). The THz detectors based on HEMT-like structures utilizing the excitation of plasma oscillations associated with different mechanisms of nonlinearity were proposed in Refs. 9, 37, and 38. In these devices, the rectified component of the dc current through the gate structure with a nonlinear dependence of the leakage channel-gate current is assumed to be used for the detection. The pertinent nonlinearity can be associated with strong (exponential) nonlinear dependences of the leakage current on the local potential in the 2DEG channel due to thermionic, tunneling, or resonant-tunneling origin of this current. Since the nonlinearities associated with thermionic, tunneling, and resonant tunneling mechanisms are fairly strong, the responsivity of the detectors using former mechanisms can significantly exceed the responsivity of the resonant

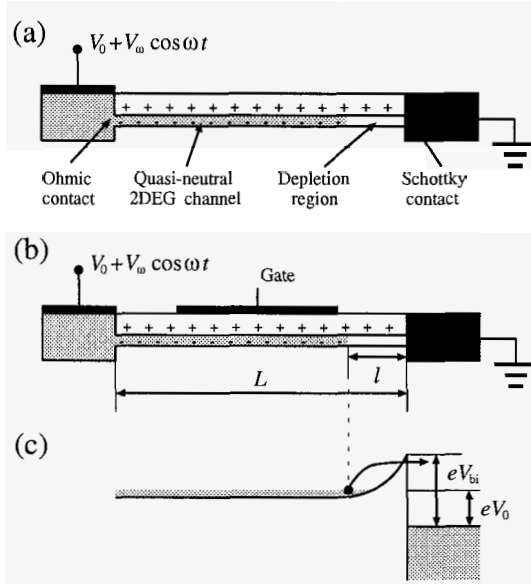


Fig. 4. Schematic view THz detectors on the base of lateral Schottky junction with (a) an ungated and (b) gated channel and (c) device band diagram under forward bias ($V_0 < 0$).

detectors using hydrodynamic nonlinearities. The resonant detectors of THz radiation utilizing the nonlinearity of the electron transverse transport can be based also on hot-electron transistors with both thermionic and tunneling injection.³⁹

Recently, THz detectors based on an ungated or gated 2DEG channel with a lateral Schottky junction (which provides a nonlinearity necessary for the rectification) have been **proposed**^{10,40} (see Fig. 4). Due to a large amplitude of the plasma oscillations forced by incoming THz signal, the nonlinearities in question lead not only to the occurrence of the rectified component of the terminal current (or the pertinent voltage) but to the occurrence of higher harmonics. Hence, the devices under discussion in this section can be used for both plasma-assisted resonant detection and frequency multiplication.

Calculations of the frequency-dependent responsivity \mathcal{R}_ω^{RSD} of the resonant Schottky detector in question result in^{10,40}

$$\mathcal{R}_\omega^{RSD} = \frac{R_\omega^{SD}}{\sinh^2(\pi\nu\omega/2\Omega) + \cos^2(\pi\omega/\Omega)}, \quad (8)$$

for the detector with a **gated channel** and

$$\mathcal{R}_\omega^{RSD} = \frac{R_\omega^{SD}}{\sinh^2(\pi\nu/2\Omega_{ung}^2) + \cos^2(\pi\omega^2/2\Omega_{ung}^2)} \quad (9)$$

for the detector with ungated channel. In Eq. (9),

$$\Omega_{ung} \simeq \sqrt{\frac{\pi^2 e^2 \Sigma_0}{kmL}} \quad (10)$$

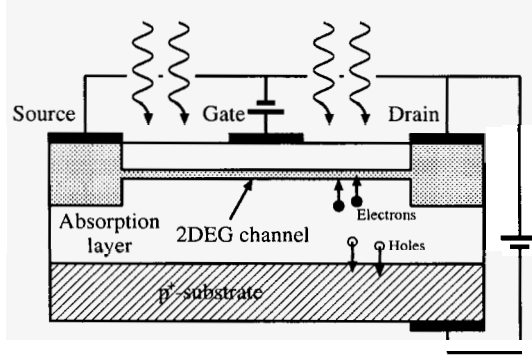


Fig. 5. Schematic view of a THz photomixing device based on a HEMT with absorption layer between the 2DEG channel and p^+ -doped substrate. Circles with arrows show photoelectrons and photoholes generated by optical signal in the absorption layer and directions of their propagation.

is the fundamental plasma frequency for the ungated 2DEG system [compare Eqs. (5) and (10)]. The quantity R_ω is the responsivity of a Schottky detector without resonant plasma cavity.

5. Resonant Photomixing Using Plasma Oscillations

Figure 5 demonstrates the structure of a photomixing device based on a HEMT. The operation of this photomixer is associated with the generation of photoelectrons and photoholes by transient optical radiation, in particular, by radiation of two laser beams with close photon frequencies Ω_1 and Ω_2 with $\omega = |\Omega_2 - \Omega_1| \ll \Omega_1, \Omega_2$ or by ultrashort optical pulses). The photoelectrons and photoholes are accelerated by the built-in electric field in the absorption region between the 2DEG channel and p^+ -doped substrate and, hence, induce the transient current in the 2DEG channel. When the signal frequency ω of the incoming optical radiation is close to one of the plasma resonant frequencies, the amplitude of the plasma oscillations forced by the transient current of photoelectrons and photoholes can be rather large. This occurs if the quality factor of the 2DEG channel as a resonant cavity is sufficiently large. In this case, the oscillating electron charge in the 2DEG channel induces large ac currents in the gate-substrate (and other) circuit, which feeds an antenna included into this circuit (antenna is not shown). The p^+ substrate is needed to collect the photoholes. It also serves as a contact to apply a large voltage to provide a sufficiently high electric field in the absorption region.^{8,41} The photomixers of such a type can be characterized by the frequency-dependent responsivity \mathcal{R}_ω , which is defined as the ratio of the Fourier component of the ac terminal current, J_ω , to the the Fourier component of incident optical power, P_ω . Figure 6 shows the responsivity of the HEMT photomixer as a function of gate voltage at fixed signal frequencies $\omega/2\pi$. In principle, a similar HEMT-based device but without the p^+ -substrate can also operate as a photomixer. In such a case, the photoholes can be

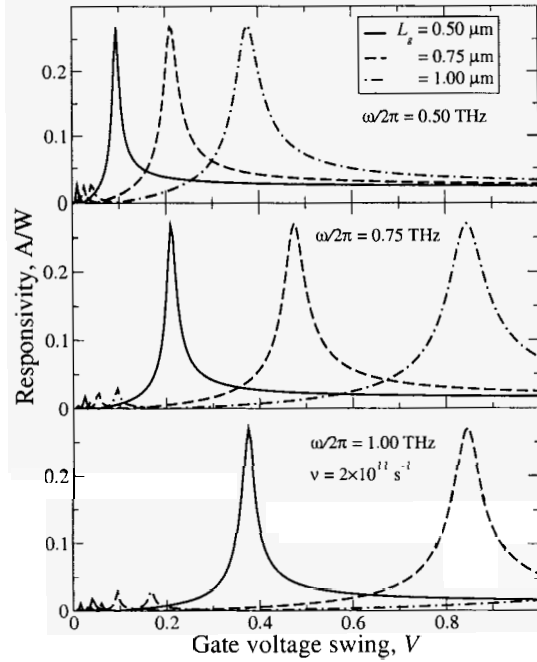


Fig. 6. Responsivity of HEMT photomixer as a function of gate voltage at fixed signal frequencies $\omega/2\pi$.

collected by the HEMT source. The resonant excitation of THz in HEMTs (without the p^+ substrate) irradiated by two lasers observed by Otsuji and co-workers.⁴² The THz plasmon-assisted emission from ungated 2DEG systems stimulated by optical signals was observed in some earlier papers.^{43,44} Similar principle of resonant photomixing can be used in devices with the 2DEG channel and different photosensitive sections, for example, uni-traveling-carrier (UTC) photodiode or quantum-well infrared photodetector (QWIP).^{11,12}

Figure 7 shows structures and band diagram of THz photomixers based on the integration of a 2DEG system (resonator) and a UTC-photodiode (photosensitive element).¹¹ The THz photomixing using infrared signals can be realized in a HEMT integrated with a QWIP. One can calculate analytically the response of the gated 2DEG channel, the frequency-dependent responsivity \mathcal{R}_ω^{PMD} of the photomixing device based on a HEMT integrated with a photodiode can be presented as¹²

$$\mathcal{R}_\omega = \frac{R_\omega^{PD}}{\sinh^2(\pi\nu\omega/2\Omega) + \cos^2(\pi\omega/\Omega)}. \quad (11)$$

Here, R_ω^{PD} is the responsivity of the photodiode portion of the device. The latter quantity steeply drops with increasing signal frequency. At fairly high signal frequencies, some effects associated with a significant deviation of the photoelectron distribution from equilibrium can be essential. In particular, the so-called near-

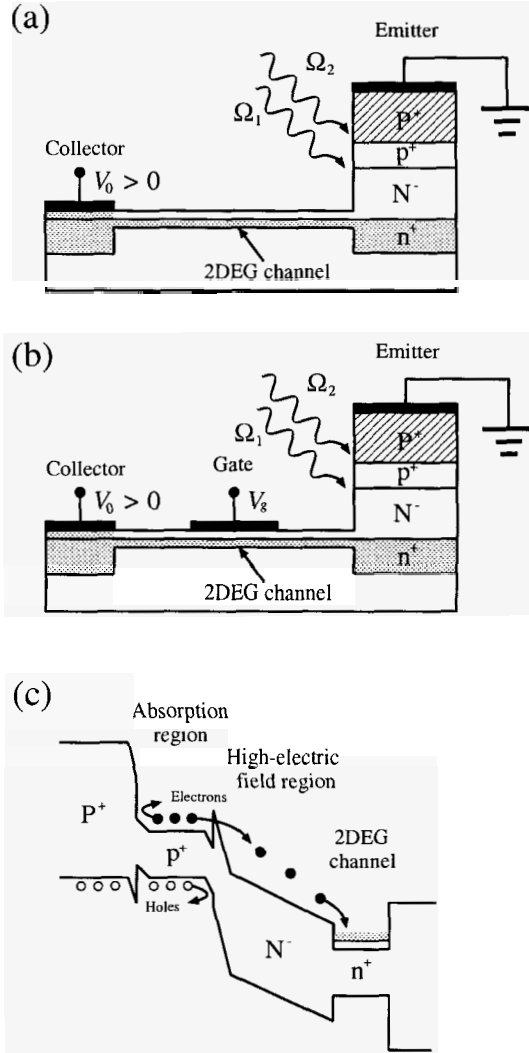


Fig. 7. Structures of THz photomixers based on integration of UTC-photodiode with (a) ungated 2DEG channel and (b) gated 2DEG channel and (c) device band diagram.

ballistic transport and velocity overshoot effect can slow down the diode frequency-dependent responsivity roll-off.⁴⁵ As seen from Eq. (11), \mathcal{R}_ω can exhibit sharp resonances at the fundamental plasma frequency and its odd harmonics, i.e., at $\omega = \Omega(2n - 1)$ provided the quality factor Q is large. An example of the calculated dependence of the responsivity of the HEMT-QWIP photomixer on the signal frequency is shown in Fig. 8.

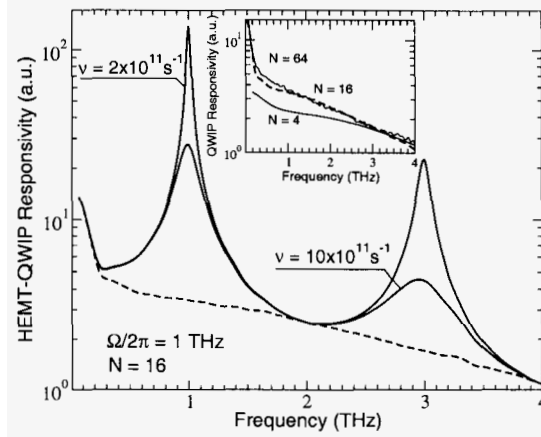


Fig. 8. Frequency-dependent responsivity of HEMT-QWIP photomixers. Dashed line corresponds to responsivity of a QWIP with 16 QWs. Inset shows the QWIP frequency-dependent responsivity calculated using Monte Carlo particle modeling with different numbers of QWs N .

6. Plasma Waves in Graphene-Based Heterostructures

The massless "relativistic" (neutrino-like, conical) energy spectrum of the particles (electrons and holes) in graphene can lead not only to specific features of the transport properties, photon-assisted transport, and quantum Hall effect, but to the unusual high-frequency properties and collective behavior of the two-dimensional electron or hole gas (2DEG or 2DHG) systems in the graphene-based heterostructures. The energy spectrum of electrons and holes is given by

$$\varepsilon_{\mathbf{p}} = \pm v_F |\mathbf{p}|, \quad (12)$$

where v_F is the characteristic velocity ($v_F \simeq 10^8$ cm/s) and \mathbf{p} is the electron (hole) momentum. The signs "+" and "-" correspond to electrons and holes, respectively.

Since the plasma properties of 2DEG become more pronounced with decreasing effective mass of electrons and increasing electron mobility, the THz devices based on graphene heterostructures with massless 2DEG can significantly surpass those made of relatively standard semiconductor heterostructures. The plasma waves in graphene heterostructures were studied in Refs. 3 and 46. The gated graphene heterostructures (with n^+ -Si substrate serving as the gate and the gate layer made of SiO_2), at different gate voltages were considered. The heterostructure in question is shown schematically in Fig. 9. As shown,⁴⁶ the dependence of the plasma wave frequency on the wavenumber in graphene is similar to that in the standard heterostructures. However, the parameters characterizing the plasma wave spectra are markedly different. In particular, although for a gated graphene heterostructure one obtains the dependence like Eq. (4), the real effective mass m should be replaced

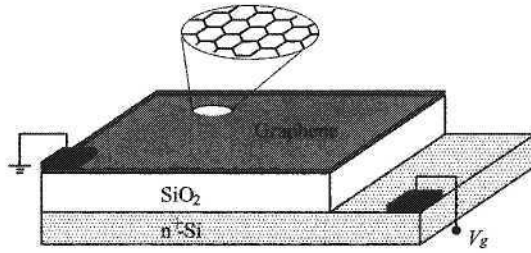


Fig. 9. Example of graphene-based heterostructure with substrate serving as a gate.

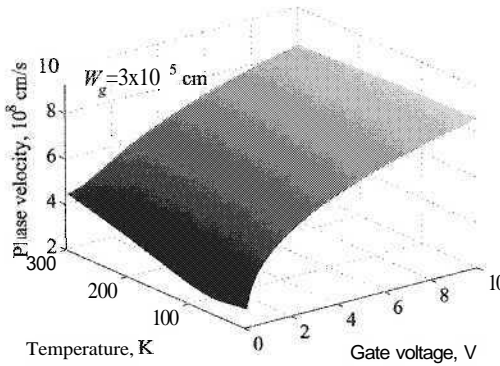


Fig. 10. Plasma wave phase velocity in gated graphene as a function of gate voltage and temperature.

by some fictitious mass

$$m = \frac{\varepsilon_F}{v_F^2} = \frac{\hbar}{v_F} \sqrt{\frac{\varkappa V_g}{2eW}} \propto W^{-1/2} v_g^{1/2}, \quad (13)$$

which depends on the electron (hole) concentration and, therefore, on the gate voltage. Here, ε_F is the electron (or hole) Fermi energy. At low gate voltages when both electrons and holes contribute to the plasma properties, the plasma frequency depends also on the temperature. Since the fictitious effective mass can be made rather small and due to relatively low dielectric constant of the substrate (SiO₂) for graphene, the plasma wave velocity s in graphene-based heterostructures can be 3-5 times larger than that in AlGaAl/GaAs heterostructures. The voltage and temperature dependencies of the plasma wave velocity s are shown in Fig. 10.⁴⁶

7. Conclusions

We discussed concepts which can result in the development of novel THz heterostructure devices including sources of THz radiation utilizing the self-excitation

of plasma oscillations (plasma instabilities), detectors and frequency multipliers of THz radiation, and THz sources using plasma-assisted photomixing of optical signals. However, the device proposals considered do not exhaust all interesting new ideas (see, for example, Refs. 24 and 47–50). The specific features of plasma waves in 2DEG channel can be used not only in the devices like those discussed above. The linearity of plasma wave dispersion ($\text{Re}\omega \propto q$), relatively large plasma wave phase and group velocity (up to $v \sim 10^8$ cm/s), a possibility to control the plasma wave propagation and the interaction plasma waves by locally applied voltage open up wide prospects to create new systems for processing of THz signals.

Acknowledgments

This work at University of Aizu and Tohoku University was financially supported by the Grant-in-Aid for Scientific Research (S) from the JSPS, Japan.

References

1. M. I. Dyakonov and M. S. Shur, Plasma wave electronics: Novel terahertz devices with two-dimensional electron fluid, *IEEE Trans. Electron Devices* **43**, 1640–1645 (1996).
2. M. S. Shur and V. Ryzhii, Plasma wave electronics, *Int. J. High Speed Electronics and Systems* **13**, 575–600 (2003).
3. V. Ryzhii, *Jpn. J. Appl. Phys.* Terahertz plasma waves in gated graphene heterostructures, **45**, L923–L925 (2006).
4. M. Dyakonov and M. Shur, Shallow water analogy for a ballistic field effect transistor: New mechanism of plasma wave generation by dc current, *Phys. Rev. Lett.* **71**, 2465–2468 (1993).
5. V. Ryzhii and M. Shur, Plasma instability and nonlinear terahertz oscillations in resonant-tunneling structures, *Jpn. J. Appl. Phys.* **40**, 546–550 (2001).
6. V. Ryzhii, A. Satou and M. S. Shur, Plasma instability and terahertz generation in HEMTs due to electron transit-time effect, *IEICE Trans.* **E89-C**, 1012–1019 (2006).
7. V. Ryzhii, A. Satou, I. Khmyrova, M. Ryzhii, T. Otsuji, V. Mitin, and M. S. Shur, Plasma effects in lateral Schottky junction tunneling transit-time terahertz oscillator, *J. Phys.: Conf. Ser.* **38**, 228–233 (2006).
8. V. Ryzhii, I. Khmyrova, A. Satou, P. O. Vaccaro, T. Aida and M. Shur, Plasma mechanism of terahertz photomixing in high-electron mobility transistor under interband photoexcitation, *J. Appl. Phys.* **92**, 5756–5760 (2002).
9. A. Satou, I. Khmyrova, V. Ryzhii, and M. S. Shur, Plasma and transit-time mechanisms of the terahertz radiation detection in high-electron-mobility transistors, *Semi-cond. Sci. Technol.* **18**, 460 (2003).
10. A. Satou, V. Ryzhii, T. Otsuji, and M. S. Shur, Resonant terahertz detectors with lateral Schottky junctions, *IEICE Trans.* (unpublished).
11. V. Ryzhii, A. Satou, I. Khmyrova, M. Ryzhii, and T. Otsuji, Terahertz photomixing in UTC-photodiode with electron resonant cavity, in *Tech. Dig. 9th Int. Symp. on Contemporary Photonics Technology (CPT 2006)*, Tokyo, Japan, 100–101 (2006).
12. V. Ryzhii, M. Ryzhii, I. Khmyrova, T. Otsuji, and M. S. Shur, Resonant terahertz photomixing in integrated HEMT-QWIP device, *Jpn. J. Appl. Phys.* **45**, 3648 (2006).
13. F. Stern, Polarizability of two-dimensional electron gas, *Phys. Rev. Lett.* **18**, 546–548 (1967).

14. A. V. Chaplik, Possible crystallization of charge carriers in low-density inversion layers, *Sov. Phys. JETP* **35**, 395–398 (1972).
15. M. Nakayama, Theory of surface waves coupled to surface carriers, *J. Phys. Soc. Japan* **36**, 393–398 (1974).
16. V. Ryzhii, A. Satou, and M. S. Shur. Admittance of a slot diode with a two-dimensional electron channel, *J. Appl. Phys.* **94**, 10041 (2003).
17. V. Ryzhii, A. Satou, I. Khmyrova, A. Chaplik, and M. S. Shur, Plasma oscillations in a slot diode structure with a two-dimensional electron channel, *J. Appl. Phys.* **96**, 7625 (2004).
18. A. Satou, V. Ryzhii, and A. Chaplik, Plasma oscillations in two-dimensional electron channel with non-ideally conducting side contacts, *J. Appl. Phys.* **98**, 034502 (2005).
19. A. Satou, I. Khmyrova, A. Chaplik, V. Ryzhii, and M. S. Shur, Spectrum of plasma oscillations in slot diode with two-dimensional electron channel, *Jpn. J. Appl. Phys.* **44**, 2592–2596 (2005).
20. A. Satou, V. Vyurkov, and I. Khmyrova. Plasma oscillations in terahertz photomixing high-electron-mobility transistor structure on p⁺-substrate, *Jpn. J. Appl. Phys.* **43(4B)**, L566–L568 (2004).
21. J. Crowne, Contact boundary conditions and the Dyakonov-Shur instability in high electron mobility transistors, *J. Appl. Phys.* **82**, 1242–1254 (1997).
22. V. Ryzhii, A. Satou, and M. S. Shur, Transit-time mechanism of plasma instability in high electron mobility transistors, *Phys. Stat. Sol. (a)* **202**, R113–R115 (2005).
23. M. Dyakonov and M. Shur, Detection, mixing, and frequency multiplication of terahertz radiation by two dimensional electronic fluid, *IEEE Trans. Electron Devices* **43**, 380–387 (1996).
24. A. O. Govorov, V. M. Kovalev, and A. V. Chaplik, Solitons in semiconductor microstructures with a two-dimensional electron gas, *JEPT Lett.* **70**, 488–490 (1999).
25. S. Rudin and G. Samsonidze, Formation and propagation of shock waves in the conduction channel of a field-effect-transistor structure, *Phys. Rev. B* **58**, 16369–16373 (1998).
26. W. Knap, J. Lusakowski, T. Parently, S. Bollaert, A. Cappy, V. V. Popov, and M. S. Shur, Terahertz emission by plasma waves in 60 nm gate high electron mobility transistors, *Appl. Phys. Lett.* **84**, 2331–2333 (2004).
27. Y. Deng, R. Kersting, J. Xu, R. Ascazubi, X.-C. Zhang, M. S. Shur, R. Gaska, G. S. Simin, M. Asif Khan, and V. Ryzhii, Millimeter wave emission from GaN high electron mobility transistor, *Appl. Phys. Lett.* **84**, 70–72 (2004).
28. J. Lusakowski, W. Knap, N. Dyakonova, L. Varani, J. Mateos, T. Gonzales, Y. Roelens, S. Bpillaert, A. Cappy, and K. Karpietz, Voltage tunable terahertz emission from a ballistic nanometer InGaAs/InAlAs transistor, *J. Appl. Phys.* **97**, 064307-1–7 (2005).
29. V. L. Kustov, V. I. Ryzhii, and Yu. S. Sigov, Nonlinear plasma instabilities in semiconductors subjected to strong electric fields in the case of inelastic scattering of electrons by optical phonons, *Sov. Phys. JETP* **52**, 1207–1212 (1980).
30. N. A. Bannov and V. I. Ryzhii, High-frequency pulsation of the current in diodes under quasiballistic conditions of electron motion, *Sov. Phys. Semicond.* **17**, 439–441 (1983).

31. V. I. Ryzhii, N. A. Bannov, and V. A. Fedirko, Ballistic and quasiballistic transport in semiconductor structures (review), *Sov. Phys. Semicond.* **18**, 481–493 (1984).
32. V. I. Ryzhii and V. A. Fedirko, Negative differential conductivity in thin semiconductor films in the case of intervalley scattering of electrons, *Sov. Phys. Semicond.* **17**, 850–851 (1983).
33. V. Ryzhii, and M. S. Shur, Analysis of tunneling-injection transit-time effects and self-excitation of terahertz plasma oscillations in high-electron mobility transistors, *Jpn. J. Appl. Phys.* **41**, L922–L924 (2002).
34. V. Ryzhii, and M. S. Shur, Tunneling- and barrier-injection transit-time mechanisms of terahertz plasma instability in high-electron mobility transistors, *Semicond. Sci. Technol.* **17**, 1168–1171 (2002).
35. J.-Q. Lü and M. S. Shur, Terahertz detection by high-electron-mobility transistor: Enhancement by drain bias, *Appl. Phys. Lett.*, **78**, 2587–2589 (2001).
36. F. Tepe, D. Veksler, V. Yu. Kacharovskii, A. P. Dmitriev, S. Romyantsev, W. Knap, and M. S. Shur, Plasma wave resonant detection of femtosecond pulsed terahertz radiation by a nanometer field-effect transistor, *Appl. Phys. Lett.* **87**, 022102–022104 (2005).
37. I. Khmyrova and V. Ryzhii, Resonant detection and frequency multiplication in barrier-injection heterostructure transistors, *Jpn. J. Appl. Phys.* **39-1**, 4727–4732 (2000).
38. V. Ryzhii, I. Khmyrova, and M. Shur, Resonant detection and frequency multiplication of terahertz radiation utilizing plasma waves in resonant-tunneling transistors, *J. Appl. Phys.* **88**(5), 2868–2871 (2000).
39. V. Ryzhii, Resonant detection and mixing of terahertz radiation by induced base hot electron transistors, *Jpn. J. Appl. Phys.* **37-1**(11), 5937–5944 (1998).
40. V. Ryzhii and M. S. Shur, Resonant terahertz detector utilizing plasma oscillations in two-dimensional electron system with lateral Schottky junction, *Jpn. J. Appl. Phys.* **45**(42), L1118–L1120 (2006).
41. A. Satou, V. Ryzhii, I. Khmyrova, M. Ryzhii, and M. S. Shur, Characteristics of a terahertz photomixer based on a high-electron mobility transistor structure with optical input through the ungated regions, *J. Appl. Phys.* **95**, 2084–2089 (2004).
42. T. Otsuji, Y. Kanamaru, H. Kitamura, M. Matsuoka, and O. Ogawara, Effect of heterostructure 2D electron confinement on the tunability of resonant frequencies of terahertz plasma-wave transistors, *IEICE Trans. Electron.* **E86-C**, 1985–1993 (2003).
43. N. Sekine, K. Yamanaka, K. Hirakawa, M. Vossenbuerger, P. Harring-Bolivar, and H. Kurz, Observation of terahertz radiation from higher-order two-dimensional plasmon modes in GaAs/AlGaAs single quantum wells, *Appl. Phys. Lett.* **74**, 1006–1008 (1999).
44. N. Sekine, K. Hirakawa, M. Vossenbuerger, P. Harring-Bolivar, and H. Kurz, Crossover from coherent to incoherent excitation of two-dimensional plasmons in GaAs/Al_xGa_{1-x}As single quantum wells by femtosecond laser pulses, *Phys. Rev. B* **64**, 201323-1-4 (2001).
45. M. Ryzhii and V. Ryzhii, Monte Carlo analysis of ultrafast electron transport in quantum well infrared photodetectors, *Appl. Phys. Lett.* **72**, 842–844 (1998).
46. V. Ryzhii, A. Satou, and T. Otsuji, Plasma waves in two-dimensional electron-hole system in gated graphene heterostructures, *J. Appl. Phys.* **101**, 024509-1-5 (2007).
47. A. O. Govorov, S. A. Studenikin, and W. R. Frank, Low-frequency plasmons in coupled electronic microstructures, *Phys. Sol. State* **40**, 499–502 (1998).
48. S. A. Mikhailov, Plasma instability and amplification of electromagnetic waves in low-dimensional electron systems, *Phys. Rev. B* **58**, 1517–1532 (1998).

49. V. Ryzhii, I. Khmyrova, and M. Shur, Terahertz photomixing in quantum well structures using resonant excitation of plasma oscillations, *J. Appl. Phys.* **91**(4), 1875–1881 (2002).
50. M. Hanabe, T. Otsuji, T. Ishibashi, T. Uno, and V. Ryzhii, Modulation effects of photocarriers on the terahertz plasma-wave resonance in high-electron mobility transistors under interband photoexcitation, *Jpn. J. Appl. Phys.* **44**, 3842–3847 (2005).

RESONANT TERAHERTZ DETECTION ANTENNA UTILIZING PLASMA OSCILLATIONS IN LATERAL SCHOTTKY DIODE

AKIRA SATOU and VICTOR RYZHII

*Computer Solid State Physics Laboratory, University of Aizu
Aizu-Wakamatsu 965-8580, Japan*

TAIICHI OTSUJI

*Research Institute of Electrical Communication, Tohoku University
Sendai 980-8577, Japan*

MICHAEL S. SHUR

*Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute, Troy, NY 12180, USA*

Keywords: THz; plasma oscillation; detector.

1. Introduction

Plasma oscillations in heterostructure devices with two-dimensional electron gases (2DEGs) in their channels can be utilized for generation and detection of terahertz (THz) radiation.^{1,2} Recently, the emission,^{3,4} detection,^{citeKnap-2002,Teppe-2005} and photomixing^{7,8} of THz radiation using such devices have been reported, demonstrating their potential for THz sources and detectors.

In recent papers,^{9,10} we proposed and studied THz resonant detectors based on 2DEG channels with lateral Schottky junctions (LSJs). These devices have the structures similar to high-electron-mobility transistors (either gated or ungated), except that they have LSJs at either source or drain contacts. Plasma oscillations in the channels are excited by THz signals supplied by external antennas or coupled via the contact pads. Due to the nonlinear current-voltage characteristics at LSJs, the THz signals induce rectified currents. Detected signals can be resonantly large when the frequency of signals is close to one of the resonant plasma frequencies, which can be in the THz range for the submicrometer channel with sufficiently large electron concentration.

In this paper, we extend our theory^{9,10} of the THz detector with LSJ by accounting for the effects of the contact pads serving as an antenna. We use another

approach, with a full set of the Maxwell equations instead of the Poisson equation (the same approach was used for example in Refs. 11 and 12). which is in principle more general than that in Refs. 9 and 10. For the first attempt, we consider the ungated structure, i.e., the lateral Schottky diode (Fig. 1), with LSJ at its drain side. We assume that the structure diode under consideration (Fig. 1) has the 2DEG

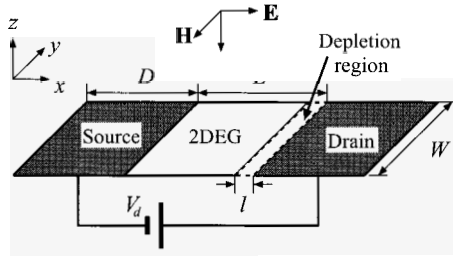


Fig. 1. Schematic view of a lateral Schottky diode

channel together with the source and drain contacts of two-dimensional strips. The lateral size of the device, W , and the contact length, D , are much larger than the channel length, L , and large enough to receive the incident wave. Other geometries such as a bow-tie¹³ are also possible (and even can be more effective than the simple strip contacts). The depletion region is formed at the drain side of the channel. A forward DC bias voltage between source and drain, V_d , leads to an electron current due to thermionic/field or field emission.

The mechanism of detection is as follows. The incident THz wave upon the device excites plasma oscillations in the 2DEG channel, and they produce an AC voltage drop across the LSJ depletion region. The AC voltage is rectified by the nonlinearity of the current-voltage characteristics of the LSJ. As a result, a DC photocurrent can be obtained. The photocurrent becomes resonantly large at the frequencies corresponding to the plasma frequencies and its harmonics.

2. Theory

From the assumptions about the device geometry made in the previous section, one can consider the device to comprise an infinite 2DEG channel strip, a depletion region, and semi-infinite source and drain two-dimensional contacts, all in the same plane. We set the coordinates as in Fig. 1. The origin is at the source edge of the channel.

We assume that the incident THz wave is linearly polarized in the x -direction: $E_x^{ext}(x, z) = E_0 \exp(-i\omega\sqrt{\epsilon}z/c - i\omega t)$, where ω is the frequency of the incident wave, ϵ is the dielectric constant of the surrounding medium and assumed to be uniform in all space, and c is the speed of light.

The response of the device to the incident wave can be described by the Maxwell

equations coupled with the constituent equation which governs the motion of electrons in the 2DEG channel. The former can be written as:

$$\nabla \times \mathbf{H}^{ind} = \frac{4\pi}{c} \mathbf{j}\delta(z) + \frac{\epsilon}{c} \frac{\partial \mathbf{E}^{ind}}{\partial t}, \quad (1)$$

$$\nabla \times \mathbf{E}^{ind} = -\frac{1}{c} \frac{\partial \mathbf{H}^{ind}}{\partial t}, \quad (2)$$

where \mathbf{E}^{ind} and \mathbf{H}^{ind} are the electric and magnetic fields induced by the incident wave, and $\mathbf{j} = j_x \hat{\mathbf{x}}$ is the AC current in the z -plane. Due to the symmetry in the y -direction, the only nonzero components of electric and magnetic fields are E_x^{ind} , E_z^{ind} , and H_y^{ind} . Combining Eqs. (1) and (2) and performing the Fourier transform for each vector component, we obtain $E_x^{ind}(q, z)$, which is the Fourier component of $E_x^{ind}(x, z)$,

$$E_x^{ind}(q, z) = -\frac{2\pi i \kappa_q}{\omega \epsilon} j_x(q) e^{-\kappa_q |z|}. \quad (3)$$

Here $\kappa_q^2 = q^2 - \omega^2 \epsilon / c^2$ and $j_x(q)$ is the Fourier component of $j_x(x)$. The total electric field at the z -plane is given by

$$E_x^{tot}(q, 0) = -\frac{2\pi i \kappa_q}{\omega \epsilon} j_x(q) + E_0 \delta(q). \quad (4)$$

From Eq. (4), we obtain the relation between the current and E_x in the z -plane

$$j_x(x) = \frac{\omega \epsilon i}{4\pi^2} \int_{-\infty}^{\infty} dq \frac{e^{iqx}}{\kappa_q} \int_0^L dx' e^{-iqx'} E_x^{tot}(x', 0) + \frac{c\sqrt{\epsilon}}{2\pi} E_0. \quad (5)$$

Here, the integral with respect to x' in Eq. (5) is taken only in the range of the 2DEG channel and depletion region because the contacts are highly conducting and hence $E_x^{tot}|_{x<0} = E_x^{tot}|_{x>L} = 0$. The branch of the square root in κ_q must be chosen so that the induced field $E_x^{ind}(x, z)$, i.e., the inverse Fourier transform of Eq. (3), correctly contains both the evanescent field and the scattered outgoing wave. This condition is met when we choose $\kappa_q = -i\sqrt{\omega^2 \epsilon / c^2 - q^2}$ for $|q| < \omega\sqrt{\epsilon}/c$ and $\kappa_q = \sqrt{\omega^2 \epsilon / c^2 - q^2}$ for $|q| > \omega\sqrt{\epsilon}/c$, where the square roots here take positive real values.

When the channel has a relatively large electron concentration, the 2DEG can be described by the hydrodynamic electron transport model. Neglecting the nonlinear terms, hydrodynamic equations can be reduced to the Drude formula

$$j_x(x) = \sigma E_x^{tot}(x, 0), \quad (6)$$

where $\sigma = \Sigma_0 e^2 i / m(\omega + i\nu)$ is the dynamic conductivity of the 2DEG, $e = |e|$ is the electron charge, Σ_0 , m , and ν are the DC electron concentration, effective mass, and collision frequency with phonons and impurities.

On the other hand, the AC current in the depletion region is represented by the AC component of

$$J = J_s \left\{ \exp \left[\frac{e(V_d + V_{LSJ})}{n_i k_B T} \right] - 1 \right\}, \quad (7)$$

where J_s is the saturation current, k_B is the Boltzmann constant, T is the temperature, n_i is the ideality factor for the LSJ, which varies from 10 to 100 (depending on temperature for typical LSJs),¹⁴ and V_{LSJ} is the voltage across the depletion region

$$V_{LSJ} = -\text{Re} \left[e^{-i\omega t} \int E_x^{tot}(x, 0) dx \right] \quad (8)$$

To simplify the analysis, we neglect the AC current in the depletion region in calculating the field distribution, and hence V_{LSJ} . Note that we *do not* mean to neglect this AC current at all, for it determines the detector responsivity! We only neglect the influence of this AC current on the field distribution. Taking into account this and Eq. (6), Eq. (5) becomes the integral equation in the interval $x = [0, L]$

$$\sigma\theta(L-l-x)E_x^{tot}(x, 0) = \frac{\omega\epsilon i}{4\pi^2} \int_{-\infty}^{\infty} dq \frac{e^{iqx}}{\kappa_q} \int_0^L dx' e^{-iqx'} E_x^{tot}(x', 0) + \frac{c\sqrt{\epsilon}}{2\pi} E_0, \quad (9)$$

where θ is the step function. To solve Eq. (9), we expand E_x^{tot} in the form

$$E_x^{tot}(x, 0) = \theta(L-l-x) \sum_{n=0}^{\infty} A_n^{(2DEG)} \cos\left(\pi n \frac{x}{L-l}\right) + \theta(x-L+l) \sum_{n=0}^{\infty} A_n^{(dep)} \cos\left(\pi n \frac{x-L+l}{l}\right) \quad (10)$$

and reduce Eq. (9) to the matrix equation for unknown coefficients $A_n^{(2DEG)}$ and $A_n^{(dep)}$. Then, the matrix equation can be numerically solved by truncating the matrix up to size N .

After calculating the coefficients, we can obtain the voltage across the depletion region, V_{LSJ} , by substituting them into Eq. (8). Then, the DC component in Eq. (7) gives the DC drain current:

$$J_d \simeq J_0 \left[1 + \frac{e^2 \overline{(V_{LSJ})^2}}{2(n_i k_B T)^2} \right], \quad (11)$$

where $J_0 = J_s \exp(eV_d/k_B T)$ and $\overline{(V_{LSJ})^2}$ is the time-average of the square of V_{LSJ} . The first term in Eq. (11) is the dark current, whereas the second term is the rectified photocurrent. The responsivity of the detector is defined as the ratio of the rectified photocurrent to the radiation power P_ω

$$R_\omega = \frac{e^2 \overline{(V_{LSJ})^2}}{2(n_i k_B T)^2 P_\omega}. \quad (12)$$

3. Results and Discussion

The responsivity Eq. (12) has resonant peaks near at resonant plasma frequencies⁹

$$\omega = \sqrt{\frac{2n-1}{2}} \omega_0 \quad (n = 1, 2, 3, \dots), \quad (13)$$

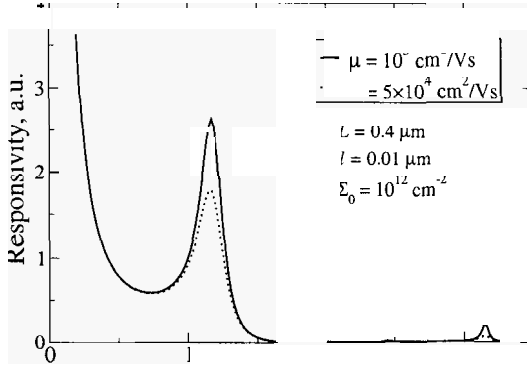


Fig. 2. The normalized responsivity vs. frequency of the incident wave for $L = 0.4 \mu\text{m}$, $l = 0.01 \mu\text{m}$, and $\Sigma_0 = 10^{12} \text{cm}^{-2}$, with different mobilities.

where

$$\omega_0 = \sqrt{\frac{2\pi^2\Sigma_0 e^2}{\epsilon m(L-l)}} \simeq \sqrt{\frac{2\pi^2\Sigma_0 e^2}{\epsilon m L}} \quad (14)$$

is the fundamental plasma frequency. For the AlGaAs/GaAs heterostructure with $\Sigma_0 = 10^{12} \text{cm}^{-2}$ and $L = 0.4 - 1.6 \mu\text{m}$, the fundamental frequency is in the range $\omega_0/2\pi \simeq 1 - 2 \text{THz}$.

Equation (13) has the factor $\sqrt{(2n-1)/2}$ because in the system under consideration plasma waves have odd number of quarter-waves in the 2DEG channel due to asymmetric boundary conditions: short-circuited at the source end and open-circuited at the drain. These boundary conditions are valid if the depletion region has a small capacitance, i.e., a large impedance; otherwise, a displacement current through the depletion region becomes large and the second condition cannot be achieved.^{9,15} In fact, the capacitance of the contact between a 2DEG and a 2D metal (even bulk metal) has the logarithmic dependence on the depletion region length,¹⁶ not like the inversely proportional dependence of a bulk-bulk contact. Therefore, the impedance of the depletion region is much larger than that of the 2DEG channel, and we can obtain the asymmetric boundary conditions.

Figure 2 shows the frequency dependence of responsivity for $L = 0.4 \mu\text{m}$, $l = 0.01 \mu\text{m}$, and $\Sigma_0 = 10^{12} \text{cm}^{-2}$ with different mobilities. The mobility μ is related to the collision frequency by the relation $\nu = em/\mu$.

One can clearly see in Fig. 2 that the responsivity has resonant peaks. The first peak occurs at $\omega/2\pi = 1.17 \text{THz}$, smaller than that predicted by Eq. (13). The second peak is hardly seen but at 2.44THz , and the third one is at 3.14THz , both values being almost equal to those calculated from Eq. (13). The linewidth of peaks are broadened compared with peaks obtained in Refs. 9 and 10 with the same mobilities due to the radiative decay.¹⁷

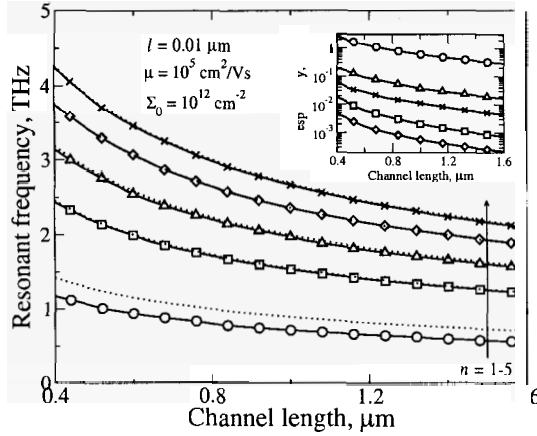


Fig. 3. Resonant frequencies vs. channel length for $n = 1 - 5$ (solid lines with markers). Dotted lines are plasma frequencies calculated from Eq. (13). Responsivities at resonant peaks are also shown in the inset.

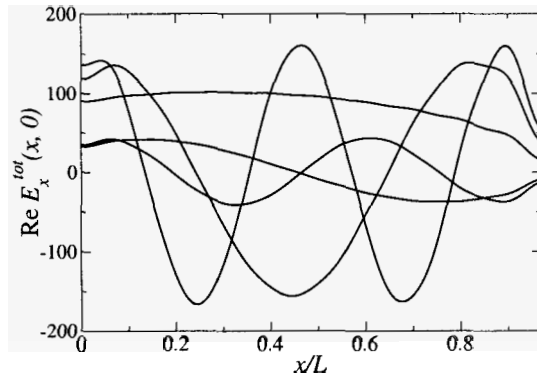


Fig. 4. The distributions of E_x^{tot} in the channel at the resonant frequencies. The same parameters as used in Fig. 2 and the mobility $\mu = 10^5 \text{ cm}^2/\text{Vs}$ are chosen.

Figure 3 shows the dependence of the resonant frequencies on the channel length (solid lines) together with the plasma frequencies calculated from Eq. (13) (dotted lines). The first mode ($n = 1$) has smaller frequency as predicted by Eq. (13), although resonant frequencies of higher modes almost coincide with Eq. (13). Such a strong softening of the first mode is also discussed in Ref. 11. The difference between the first mode and higher modes can be also seen in electric field distributions in the channel (Fig. 4). The distributions for higher modes are almost symmetric or anti-symmetric (except near edges of the channel), but the first mode has the distribution in between the half wave and quarter wave. This decrease in the frequency of the first mode is associated with the antenna capacitance under consideration.

The charge distribution in the channel for the first mode, which can be seen from Fig. 4 by differentiating the electric field distribution with respect to x , has the same sign everywhere except the source edge. Such an oscillation can be easily influenced by the external resources. On the other hand, charge distributions for higher modes are divided by areas which have different signs. In such oscillations restoring forces exerted on electrons in a certain area mainly come from electrons in neighboring areas. Thus, oscillations are more self-contained than the first mode and their frequencies are less affected by the antenna capacitance.

It is also shown from the inset of Fig. 3 that the peaks of modes with even n are suppressed compared with those of modes with odd n contrary to the result in Refs. 9 and 10, where peaks decrease in order as the mode index increases. The modes with odd n have symmetric distributions, whereas modes with even n have anti-symmetric distributions (Fig. 4).

Due to the geometrical restriction, plasma oscillations with both (nearly) symmetric and anti-symmetric distributions of the electric field with respect to the center of the channel can be allowed. However, the incident wave is symmetric, so that symmetric modes are much more strongly excited than anti-symmetric modes.

We have neglected the nonlinearity in hydrodynamic equations and used Eq. (6). One can account for it by including higher frequency terms ($e^{-i2\omega t}$, $e^{-i3\omega t}$, etc.) in the solutions of hydrodynamic equations. Then one can obtain the nonlinear relation between the current and electric field. From Eq. (8), it is clear that these higher order terms in the electric field add the positive contribution to Eq. (12). We expect that the higher order terms also have resonant peaks and increase the values of total peaks.

4. Conclusions

In conclusions, we have extended our theory^{9,10} of the lateral Schottky diode accounting for the effects of the contact pads serving as an antenna. We have demonstrated that the responsivity of the lateral Schottky diode antenna can have resonant peaks. Compared with the responsivity of the lateral Schottky diode connected with an external antenna,^{9,10} the lateral Schottky diode antenna has the first peak with smaller resonant frequency. This effect is associated with an explicit accounting for the antenna capacitance under consideration. Also, the even modes are less pronounced in the latter due to the symmetry of the structure.

Acknowledgments

This work was financially supported by the Grant-in-Aid for Scientific Research (S) from the Japan Society for the Promotion of Science (JSPS) and by the Grant-in-Aid for JSPS Fellows. One of Authors (A. S.) thanks Prof. Vyacheslav Popov for his valuable comments and discussions on the work.

References

1. M. Dyakonov and M. Shur, Shallow water analogy for a ballistic field effect transistor: New mechanism of plasma wave generation by dc current, *Phys. Rev. Lett.* **71**, 2465–2468 (1993).
2. M. Dyakonov and M. Shur, Detection, mixing, and frequency multiplication of terahertz radiation by two dimensional electronic fluid, *IEEE Trans. Electron Devices* **43**, 380–387 (1996).
3. J. Lusakowski, W. Knap, N. Dyakonova, L. Varani, J. Mateos, T. Gonzales, Y. Roelens, S. Bollaert, A. Cappy, and K. Karpierz, Voltage tunable terahertz emission from a ballistic nanometer InGaAs/InAlAs transistor, *J. Appl. Phys.* **97**, 064307 (2005).
4. N. Dyakonova, A. El Fatimy, J. Lusakowski, W. Knap, M. I. Dyakonov, M.-A. Poisson, E. Morvan, S. Bollaert, A. Shchepetov, Y. Roelens, Ch. Gaquiere, D. Theron, and A. Cappy, Room temperature terahertz emission from nanometer field-effect transistors, *Appl. Phys. Lett.* **88**, 141906 (2006).
5. W. Knap, V. Kachorovskii, Y. Deng, S. Rumyantsev, J.-Q. Lü, R. Gaska, G. Simin, X. Hu, M. Asif Khan, C. A. Saylor, and L. C. Brunel, Nonresonant detection of terahertz radiation in field effect transistors, *J. Appl. Phys.* **91**, 9346–9353 (2002).
6. F. Teppe, W. Knap, D. Veksler, M. S. Shur, A. P. Dmitriev, V. Yu. Kachorovskii, and S. Rumyantsev, Room temperature plasma waves resonant detection of sub-terahertz radiation by nanometer field-effect transistor, *Appl. Phys. Lett.* **87**, 052107 (2005).
7. T. Otsuji, M. Hanabe, and O. Ogawara, Terahertz plasma wave resonance of two-dimensional electrons in InGaP/InGaAs/GaAs high-electron-mobility transistors, *Appl. Phys. Lett.* **85**, 2119–2121 (2004).
8. Y. M. Meziani, M. Hanabe, T. Otsuji, and E. Sano, Threshold behavior of photoreponse of plasma waves by new photomixer devices, *Tech. Dig. Int. Conf. on Solid State Devices and Materials*, Yokohama, Japan, 864–865 (2006).
9. V. Ryzhii and M. S. Shur, Resonant terahertz detector utilizing plasma oscillations in two-dimensional electron system with lateral Schottky junction, *Jpn. J. Appl. Phys.* **45**, L1118–L1120 (2006).
10. A. Satou, V. Ryzhii, T. Otsuji, and M. S. Shur, Resonant terahertz detectors with lateral Schottky junctions, *Semicond. Sci. Technol.* (submitted).
11. V. V. Popov, G. M. Tsymbalov, M. S. Shur, and W. Knap, The resonant terahertz response of a slot diode with a two-dimensional electron channel, *Semiconductors* **39**, 142–146 (2005).
12. S. A. Mikhailov and N. A. Savostianova, Influence of contacts on the microwave response of a two-dimensional electron stripe, *Phys. Rev. B* **74**, 045325 (2006).
13. H. Tomioka, M. Suhara, and T. Okumura, Effects of geometrical structures on broadband characteristics in finite-sized on-chip self-complementary antennas, *Tech. Dig. Asia-Pacific Workshop on Fundamentals and Applications of Advanced Semiconductor Devices*, Sendai, Japan, 281–284 (2006).
14. M. S. Shur (private communication).
15. V. V. Popov (private communication).
16. V. Ryzhii, A. Satou, I. Khmyrova, M. Ryzhii, T. Otsuji, V. Mitin, and M. S. Shur, Plasma effects in lateral Schottky junction tunneling transit-time terahertz oscillator, *J. Phys.: Conf. Series* **38**, 228–233 (2006).
17. S. A. Mikhailov and N. A. Savostianova, Microwave response of a two-dimensional electron stripe, *Phys. Rev. B* **71**, 035320 (2005).

TERAHERTZ POLARIZATION CONTROLLER BASED ON ELECTRONIC DISPERSION CONTROL OF 2D PLASMONS

Takuya Nishimura and Taiichi Otsuji

*Research Institute of Electrical Communication, Tohoku University, 2-1-1, Aoba,
Sendai 980-8577, Japan
nishimura@riec.tohoku.ac.jp, otsuji@riec.tohoku.ac.jp*

We numerically investigated the possibility of terahertz polarization controller based on electronic dispersion control of two dimensional (2D) plasmon gratings in semiconductor heterostructure material systems. Taking account of the Mikhailov's dispersive plasmonic conductivity model, the electromagnetic field emission properties of the *gated* 2D plasmon gratings were numerically analyzed with respect to the density (n) of electrons by using in-house Maxwell's FDTD (finite difference time domain method) simulator. When n is low under a constant drift-velocity condition, the fundamental plasmon mode is excited, being coupled with the *radiative* zeroth mode of transverse electric (TE) waves. When n exceeds a threshold level, the second harmonic mode of plasmon is predominantly excited, being coupled with the *non-radiative* first mode of TE waves. We numerically demonstrated that if a grating mesh of 2D plasmons is formed where two independent 2D plasmon gratings are combined orthogonally, the structure can act as a polarization controller by electronically controlling the two axial plasmonic dispersions.

Keywords: Plasmon resonance; 2D plasmon gratings; dispersion control.

1. Introduction

The terahertz frequency band has been attracting attention due to the potentiality of opening up a new frequency resource which can bridge a gap between light and microwaves to bloom many new technological applications.¹⁻⁶ As a new principle of operation enabling terahertz oscillation, two-dimensional (2D) plasmon resonance in a conventional high electron mobility transistor (HEMT) was proposed by Dyakonov and Shur.⁷ When the density of electrons becomes very high, the 2D electron system behaves as plasma fluid. The plasmon resonant frequency produced by plasma fluid can be controlled by the gate bias voltage because the gate bias strongly depends on the electron density of the 2D channel. If one assumes a sub micron gate length, the resonant frequency falls in the terahertz range. Once the plasmon is excited, harmonic resonances as well as the fundamental one are produced due to the nonlinear nature of plasma hydrodynamics.

So far, various analytical¹⁶⁻¹⁵ and experimental studies¹⁶⁻³² on emission and detection of plasma wave resonance in HEMT's or similar field effect transistors (FET's) have been reported. Some of them proposed and/or demonstrated photomixing operation as terahertz coherent sources¹²⁻¹⁵ including observation of third harmonic resonances.^{28,29} The resonant oscillation of the 2D plasmon is a *non-radiative* mode so that a mode conversion mechanism like a grating antenna structure is needed in order to realize highly efficient emission of terahertz electromagnetic radiation. Very recently a grating-bicoupled plasmon-resonant terahertz emitter co-functioning photomixing operation has been proposed and has demonstrated emission of coherent terahertz waves at room temperature.³³⁻³⁶

In this paper, we focus on the field emission properties of a 2D plasmon grating which is a core element of the plasmon resonant emitter³⁵ we have recently developed. We investigate how the field emission profiles are affected by the plasmonic dispersion relations by using a FDTD (finite difference time domain) method, and discuss the possibility of polarization control based on the electronic dispersion control of 2D plasmon gratings in semiconductor heterostructure material systems.

2. Plasmon Resonant Terahertz Emitter

Figure 1 illustrates the cross sectional view of terahertz emitter³⁵ we have been recently developed. The terahertz emitter is based on a conventional high electron mobility transistor (HEMT). The terahertz emitter has a doubly interdigitated grating gates (G1 and G2) periodically localizes the 2D plasmons in sub 100-nm regions with a micron interval. A doubly interdigitated grating gates works as a terahertz antenna. The cladding guide with a low permittivity is an option for better confinement of the vertical cavity.

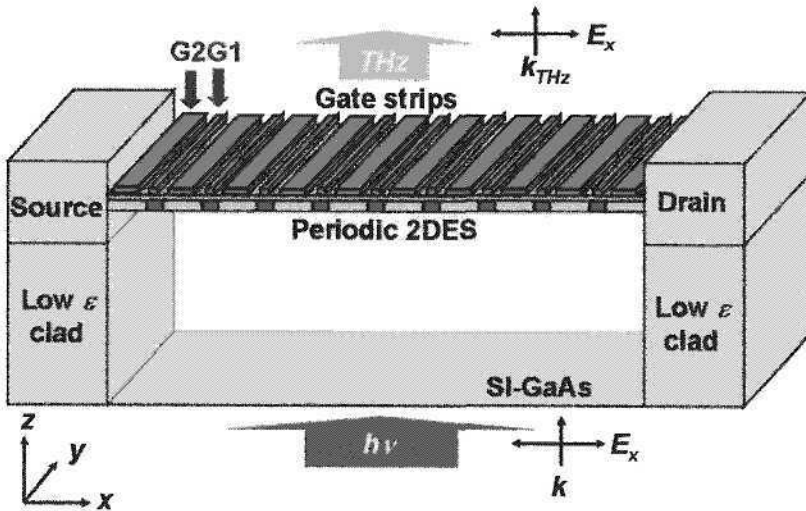


Fig. 1. Cross sectional view of plasmon resonant terahertz emitter.

When electromagnetic waves having the frequency ν in the terahertz range are absorbed in the 2D electron channel, photoexcited carriers¹² and/or phonon-polariton modes⁶ can excite the periodically localized 2D plasmons at the frequency ν . If the excitation frequency matches to the standing wave condition, the 2D plasmon makes resonant oscillation. According to the Dyakonov-Shur model, the plasma wave behaviour is formulated by the hydrodynamic equations.^{7,8} The resonant frequency is determined by the plasma wave velocity divided by the resonant plasmon cavity length. Assuming a sub-micron gate length and standard III-V compound material systems, the resonant frequency falls in the terahertz range. Since the plasma velocity is proportional to the square-root of the carrier density which is controlled by the gate bias, the resonant frequency would be electrically tunable.

The plasma wave itself is a *non-radiative mode*. The periodically localized 2D plasmon, however, can be coupled with plasmons in neighbour regions and make in-phase resonant oscillation. The 2D plasmon gratings can convert the *non-radiative* plasma waves to *radiative* electromagnetic waves. In the following analysis, we focus especially on the 2D plasmon gratings can improve mode conversion gain and radiation power.

3. Simulation

Physical phenomena of 2D plasmons are described by the hydrodynamic equations.⁷⁻¹¹ According to the Mikhailov's theory,⁹ the conductivity of 2D plasmons is led to the following form.

$$\sigma(k, \omega) = \frac{ne^2}{m} \frac{i\omega}{(\omega - kV_d) \left(\omega - kV_d + i\frac{1}{\tau} \right)}. \quad (1)$$

where n the density of electrons, e the electronic charge, m the effective mass, V_d the drift velocity, τ the total momentum relaxation time, ω the angular frequency, k the wave vector of the grating geometry. This Mikhailov's dispersive plasmonic conductivity is similar to the Drude-optical conductivity³⁷ including n and ω , but is different and featured by two distinctive parameters having k and V_d . Based on Eq. (1), we can electronically control the dispersion by changing n and V_d due to the relation of the gate and drain bias voltages. We will implement the conductivity equation (1) into our in-house Maxwell's FDTD simulator.

According to the standard Yee algorithm,³⁸ Maxwell's equations are differentiated in the time-space dimension. The series of the FDTD simulation reveals the temporal response of the 2D plasmon gratings, i.e. terahertz electromagnetic radiation from the 2D plasmon excitation. In the following subsections, we first focus on the analysis of field emission properties by using a Maxwell's FDTD simulator, and then, discuss the spatial field emission properties from the view point of polarization control.

4. Device Model

Figure 2 illustrates the device model of 2D plasmon gratings. We assume InGaP/InGaAs heterostructure material systems on a SI-GaAs substrate and room-temperature operation. The electron channel acting as a plasmonic cavity is formed with a 50-nm thick InGaAs layer sandwiched with an upper 100-nm thick InGaP carrier supplying layer and a lower 500-nm thick GaAs buffer layer. Nine periods of grating structures are modelled for the 2D plasmon gratings. The density of electrons in the plasmon cavity regions is a parameter ranging from 10^9 to 10^{12} cm^{-2} so as to make the fundamental plasmon resonance fall in the terahertz range. The conditions of the electron density for the connecting cavity regions are fixed at 2.0×10^7 cm^{-2} so as to obtain good confinement of electrons in the plasmon cavities. The length of the plasmon cavity and connecting cavity regions are set at 200-nm and 1.0- μm , respectively. The doubly interdigitated gate strips are assumed to be quantum wires placed on top of the electron channel in a 100-nm distance whose the density of electrons was set at $n = 1.75 \times 10^{11}$ cm^{-2} . The thickness of the 2DES and the quantum wired gate gratings are assumed to be the numerical lattice constant of 10 nm. In this simulation, the back ground drift velocity is fixed at 2.0×10^7 cm/s . The momentum relaxation time τ is set at 1.0×10^{-13} s. All the plasmon cavities are excited simultaneously with a Gaussian impulsive current signal.

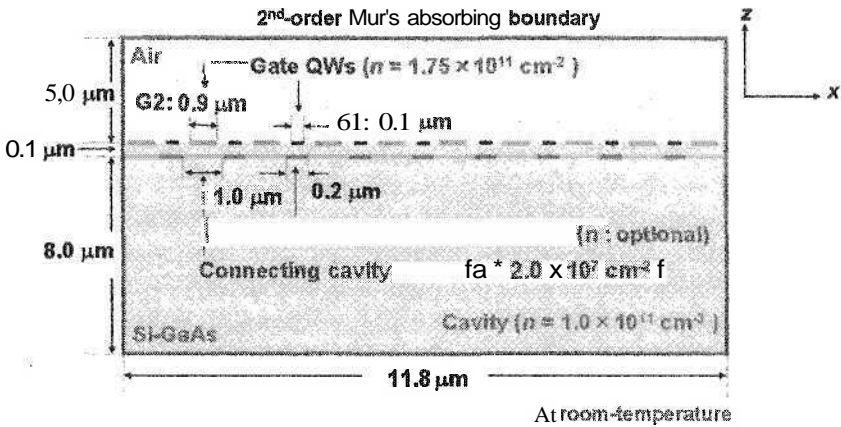


Fig. 2. The device model of 2D plasmon gratings for numerical simulation. The analytic space is closed by the 2nd order Mur's absorbing boundary. QW: quantum wire.

5. Results and Discussion

We investigated the field emission properties in terms of n for the 2D plasmon gratings shown in Fig. 2. Figure 3 shows typical spatial field distributions of the electric field intensity (E_x) on the x - z cross section at a specific time step. When n is set at a low value of 2.2×10^9 cm^{-2} , the electric field intensity distributes monotonically so that a radiative mode of transverse-electric (TE) waves is excited. The white color indicates an out of range: extremely high intensity. On the other hand, when n is set at a relatively high value

of $1.6 \times 10^{12} \text{ cm}^{-2}$, anti-parallel electric field is excited with respect to the center of the channel. The electric field intensity is cancelled out and *non-radiative* mode of TE waves is developed. These results imply that the *radiative* or *non-radiative* mode is directly reflected by the density of electrons in the plasmon cavities.

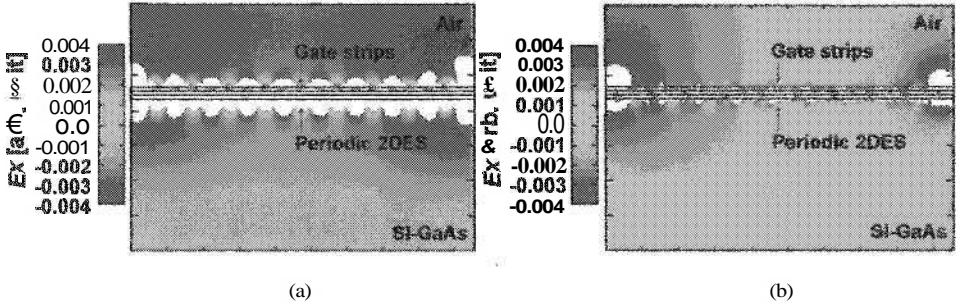


Fig. 3. The spatial field distribution at (a) $n = 2.2 \times 10^9 \text{ cm}^{-2}$, and (b) $n = 1.6 \times 10^{12} \text{ cm}^{-2}$.

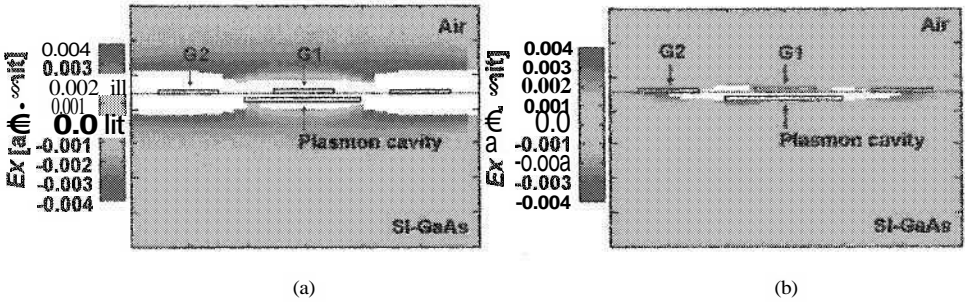


Fig. 4. The spatial field distribution of a single-core element $n = 2.2 \times 10^9 \text{ cm}^{-2}$, $n = 1.6 \times 10^{12} \text{ cm}^{-2}$.

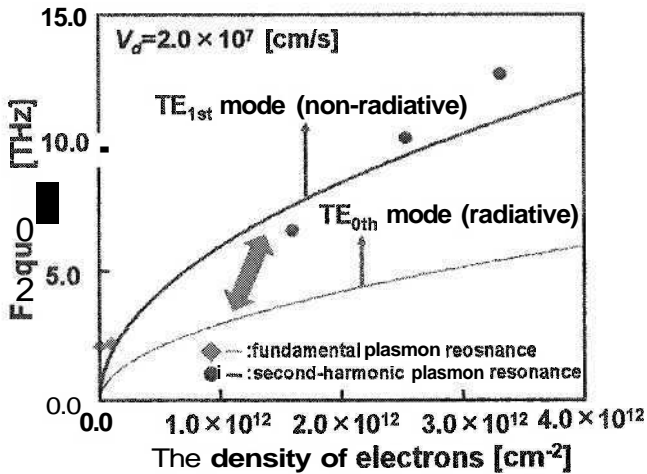


Fig. 5. Frequencies of 2D Plasmon modes as function of electron densities.

Figures 4 (a) and (b) close up a single-core area of Figs. 3 (a) and (b), respectively. When the density of electrons is very low, the fundamental mode of plasmons is excited to be coupled with the *zeroth* mode of TE waves. As a result, *radiative* electric waves are developed from the 2D plasmon gratings. On the contrary, when the density of electrons becomes high, the second harmonic mode of plasmons is predominantly excited to be coupled with the *first* mode of the TE waves. These results indicate that the electromagnetic radiation has a threshold property in terms of the density of electrons in the 2D plasmon cavities.

As a result of FDTD simulation, several peak emission frequencies reflecting the plasmon modes were observed. Figure 5 plots the peak emission frequencies versus the density of electrons in the 2D plasmon cavities. The solid lines show the fundamental and second harmonic plasmon resonant frequencies, respectively. According to the simulated results, when the density of electrons is very low, the results should trace the curve for the fundamental plasmon resonance that is *radiative* mode. On the other hand, when the density of electrons becomes very high, the peak frequencies should trace the curve for the second harmonic plasmon resonance that is a *non-radiative* mode. These results indicate that the mode of electromagnetic radiation can be electronically controlled by the density of electrons, thus, by the gate bias voltages.

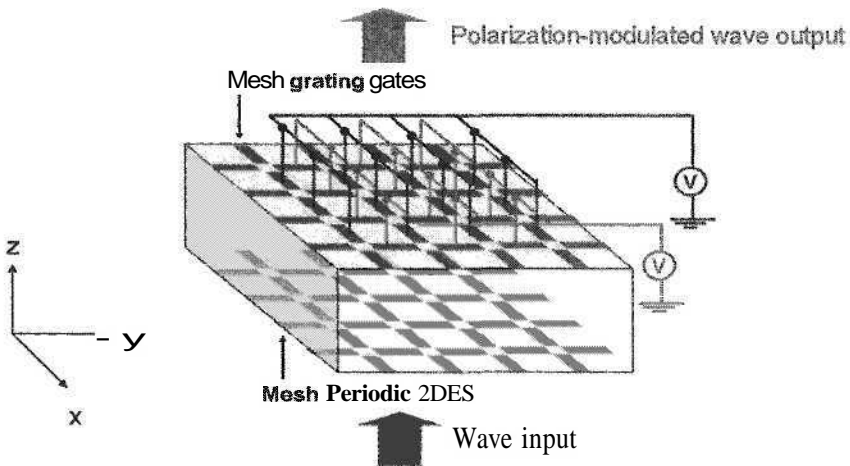


Fig. 6. Terahertz polarization controller based on electronic dispersion control of 2D plasmons by mesh grating gates.

Finally, Fig. 6 illustrates a conceptual image for a novel terahertz polarization controller by mesh grating gates. If the grating geometry is designed and fixed along with a specific direction, we can control the *radiative* or *non-radiative* propagation mode of electromagnetic waves by tuning the density of electrons. Such a unique advantage is adopted into Fig. 6. This device has two grating dimensions. When one can independently give the different bias voltages to x direction and y direction, one direction can work as a *radiative* mode while another direction can work as a *non-radiative* mode.

By making a specific mode distribution, this device is utilized as a kind of polarization controller.

6. Conclusions

We numerically investigated the possibility of terahertz polarization controller based on electronic dispersion control of 2D plasmon gratings in semiconductor heterostructure material systems. Taking account of the Mikhailov's dispersive plasmonic conductivity model, the electromagnetic field emission properties of the gated 2D plasmon gratings were numerically analyzed with respect to the density of electrons n by using in-house Maxwell's FDTD simulator. When n is low under a constant drift velocity condition, the fundamental plasmon mode is excited, being coupled with the radiative zeroth mode of TE waves. When n exceeds a threshold level, the second harmonic mode of plasmon is predominantly excited, being coupled with the non-radiative first mode of TE waves. We numerically demonstrated that if a grating mesh of 2D plasmons is formed where two independent 2D plasmon gratings are combined orthogonally, the structure can act as a polarization controller by electronically controlling the two axial plasmonic dispersions.

Acknowledgments

The authors thank Profs. K. Narahara, E. Sano, and V. Ryzhii for their valuable discussion. This work was financially supported in part by the SCOPE program from the MIC, Japan and the Grant in Aid for Scientific Research (S) from the JSPS, Japan.

References

1. D. M. Mittleman, R. H. Jacobsen, and M. C. Nuss, T-ray imaging, *IEEE J. Sel. Top. Quantum Electron.* **2**(3), 679-691 (1996).
2. S. P. Mickman and X. -C. Zhan, T-ray sensing and imaging, *Int. J. High Speed Electron. Syst.* **13**(2), 601-676 (2003).
3. C. M. Mann, Towards terahertz communications systems, *Terahertz Sources and Systems*, ed. R. E. Miles, P. Harrison, and D. Lippens, 261-267 (Kluwer Academic Publishers, Netherlands, 2001).
4. S. Verghese, K. A. McIntosh, and E. R. Brown, Highly tunable fiber-coupled photomixers with coherent terahertz output power, *IEEE Trans. Microw. Theory Tech.* **45**(8), 1301-1309 (1997).
5. H. Ito, F. Nakajima, T. Furuta, K. Yoshino, Y. Hirota, and T. Ishibashi, Photonic terahertz-wave generation using antenna-integrated uni-travelling-carrier photodiode, *Electron. Lett.* **39**(25), 1828-1829 (2003).
6. T. Otsuji, S. Nakae, and H. Kitamura, Numerical analysis for resonance properties of plasma-wave field-effect transistors and their terahertz applications to smart photonic network systems, *IEICE Trans. Electron.* **E84-C**(10), 1470-1476 (2001).
7. M. Dyakonov and M. Shur, Shallow water analogy for a ballistic field effect transistor: New mechanism of plasma wave generation by DC current, *Phys. Rev. Lett.* **71**(15), 2465-2468 (1993).
8. M. Dyakonov and M. Shur, Detection, mixing, and frequency multiplication of terahertz radiation by two-dimensional electron fluid, *IEEE Trans. Electron Dev.* **43**(3), 380-387 (1996).

9. S. A. Mikhailov, Plasma instability and amplification of electromagnetic waves in low-dimensional electron systems, *Phys. Rev. B* **58**(3), 1517-1532 (1998).
10. F. J. Crowne, Dyakonov-Shur plasma excitations in the channel of a real high-electron mobility transistor, *J. Appl. Phys.* **87**(11), 8056-8063 (2000).
11. F. J. Crowne, Contact boundary conditions and the Dyakonov-Shur instability in high electron mobility transistors, *J. Appl. Phys.* **82**, 1242-1254 (1997).
12. V. Ryzhii, I. Khmyrova, A. Satou, P. O. Vaccaro, T. Aida, and M. Shur, Plasma mechanism of terahertz photomixing in high-electron mobility transistor under interband photoexcitation, *J. Appl. Phys.* **92**(10), 5756-5760 (2002).
13. V. Ryzhii, I. Khmyrova, and M. Shur, Terahertz photomixing in quantum well structures using resonant excitation of plasma oscillations, *J. Appl. Phys.* **91**(4), 1875-1881 (2002).
14. V. Ryzhii and M. Shur, Analysis of tunnelling-injection transit-time effects and self-excitation of terahertz plasma oscillations in high-electron-mobility transistors, *Jpn. J. Appl. Phys.* **41**, L922-L924 (2002).
15. A. Satou, V. Ryzhii, I. Khmyrova, M. Ryzhii, and M. S. Shur, Characteristics of a terahertz photomixer based on a high-electron mobility transistor structure with optical input through the ungated regions, *J. Appl. Phys.* **95**, 2084-2089 (2004).
16. S. J. Allen, Jr., D. C. Tsui, and R. A. Logan, Observation of the two-dimensional plasmon in silicon inversion layers, *Phys. Rev. Lett.* **38**, 980-983 (1977).
17. D. C. Tsui, E. Gornik, and R. A. Logan, Far infrared emission from plasma oscillations of Si inversion layers, *Solid State Comm.* **35**, 875-877 (1980).
18. R. J. Wilkinson, C. D. Ager, T. Duffield, H. P. Hughes, D. G. Hasko, H. Armed, J. E. Frost, D. C. Peacock, D. A. Ritchie, A. C. Jones, C. R. Whitehouse, and N. Apsley, Plasmon excitation and self-coupling in a bi-periodically modulated two-dimensional electron gas, *J. Appl. Phys.* **71**(12), 6049-6061 (1992).
19. K. Hirakawa, K. Yamanaka, M. Grayson, and D. C. Tsui, Far-infrared emission spectroscopy of hot two-dimensional plasmons in $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$ heterojunctions, *Appl. Phys. Lett.* **67**, 2326-2328 (1995).
20. J. -Q. Lu, M. Shur, J. L. Hesler, L. Sun, and R. Weikle, Terahertz detector utilizing two-dimensional electronic fluid, *IEEE Electron Device Lett.* **19**(10), 373-375 (1998).
21. N. Sekine, K. Yamanaka, K. Hirakawa, M. Voseburger, P. Haring-Bolivar, and H. Kurz, Observation of terahertz radiation from higher-order two-dimensional plasmon modes in $\text{GaAs}/\text{AlGaAs}$ single quantum wells, *Appl. Phys. Lett.* **74**, 1006-1008 (1999).
22. M. Shur and J. -Q. Lu, Terahertz sources and detectors using two-dimensional electronic fluid in high-electron mobility transistors, *IEEE Trans. Microw. Theory Tech.* **48**(4), 750-756 (2000).
23. X. G. Peralta, S. J. Allen, M. C. Wanke, N. E. Harff, J. A. Simmons, M. P. Lilly, J. L. Reno, P. J. Burke, and J. P. Einstein, Terahertz photoconductivity and plasmon modes in double-quantum well field-effect transistors, *Appl. Phys. Lett.* **81**, 1627-1629 (2002).
24. W. Knap, Y. Deng, S. Romyantsev, and M. S. Shur, Resonant detection of subterahertz and terahertz radiation by plasma waves in sub-micron field-effect transistors, *Appl. Phys. Lett.* **81**, 4637-4639 (2002).
25. T. Otsuji, Y. Kanamaru, H. Kitamura, M. Matsuoka, and O. Ogawara, Effects of heterostructure 2D-electron confinement on the tunability of resonant frequencies of terahertz plasma-wave transistors, *IEICE Trans. Electron.* **E86-C**(10), 1985-1993 (2003).
26. W. Knap, J. Lusakowski, T. Parenty, S. Bollaert, A. Cappy, V. V. Popov, and M. S. Shur, Terahertz emission by plasma waves in 60nm gate high electron mobility transistors, *Appl. Phys. Lett.* **84**, 2331-2333 (2004).
27. D. Seliuta, E. Sirmulis, V. Tamosiunas, S. Balakauskas, S. Asmontas, A. Suziedelis, J. Gradauskas, G. Valusis, A. Lisauskas, H. G. Roskos, and K. Kohler, Detection of

- terahertz/sub-terahertz radiation by asymmetrically-shaped 2DEG layers, *Electron. Lett.* **40**, 631-632 (2004).
28. T. Otsuji, M. Hanabe, and O. Ogawara, Terahertz plasma wave resonance of two-dimensional electrons in InGaP/InGaAs/GaAs high-electron-mobility transistors, *Appl. Phys. Lett.* **85**(11), 2119-2121 (2004).
 29. M. Hanabe, T. Otsuji, T. Ishibashi, T. Uno, and V. Ryzhii, Modulation effects of photocarriers on the terahertz plasma-wave resonance in high-electron-mobility transistors under interband photoexcitation, *Jpn. J. Appl. Phys.* **44**(6A), 3842-3847 (2005).
 30. Y. M. Meziani, J. Lusakowski, N. Dyakonova, W. Knap, D. Seliuta, E. Sirmulis, J. Devenson, G. Valusis, F. Boeuf, and T. Skotnicki, Non resonant response of terahertz radiation by submicron CMOS transistors, *Dig. 6th Topical Workshop on Heterostructure Microelectronics*, Awaji, Japan, 54-55 (2005).
 31. F. Teppe, D. Veksler, V. Y. Kachorovski, A. P. Dmitriev, X. Xie, S.-C. Xhang, S. Romyantsev, W. Knap, and M. Shur, Plasma wave resonant detection of femtosecond pulsed terahertz radiation by a nanometer field-effect transistor, *Appl. Phys. Lett.* **87**, 022102 (2005).
 32. M. Lee, M. C. Wanke, and J. L. Reno, Millimeter wave mixing using plasmon and bolometric response in a double-quantum well field-effect-transistor, *Appl. Phys. Lett.* **86**, 033501 (2005).
 33. T. Otsuji, M. Hanabe, J. Shigenobu, S. Takahashi and E. Sano, A novel terahertz plasma-wave photomixer with resonant-cavity enhanced structure, *Conf. Dig. the Joint 29th Int. Conf. on Infrared and Millimeter Waves and 12th Int. Conf. on Terahertz Electronics*, Karlsruhe, Germany, 331-332 (2004).
 34. T. Otsuji, M. Hanabe, T. Nishimura, and E. Sano, A grating-bicoupled plasma-wave photomixer with resonant-cavity enhanced structure, *Optics Express* **14**(11), 4815-4825 (2006).
 35. T. Otsuji, Y. M. Meziani, M. Hanabe, T. Ishibashi, T. Uno, and E. Sano, A grating-bicoupled plasmon-resonant terahertz emitter fabricated with GaAs-based heterostructure metamaterial systems, *Tech. Dig. OSA Int. Topical Meeting on Photonic Metamaterials*, Grand Bahama Islands, The Bahamas, No. WD-29 (2006).
 36. T. Otsuji, M. Hanabe, Y. M. Meziani, and E. Sano, Terahertz emission of radiation from InGaP/InGaAs/GaAs grating bi-coupled plasmon-resonant photomixer, *Conf. Dig. the 64th Device Research Conf.*, Pennsylvania, USA, No. VA-2, 193-194 (2006).
 37. P. G. Huggard, J. A. Cluff, G. P. Moore, C. J. Shaw, S. R. Andrews, S. R. Keiding, E. H. Linfield, and D. A. Ritchie, Drude conductivity of highly doped GaAs at terahertz frequencies, *J. Appl. Phys.* **87**(5), 2382-2385 (2000).
 38. K. S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, *IEEE Trans. Antennas Propag.* **AP-14**(3), 302-307 (1966).

This page intentionally left blank

HIGHER-ORDER PLASMON RESONANCES IN GAN-BASED FIELD-EFFECT TRANSISTOR ARRAYS

V. V. POPOV^a and M. S. SHUR

*Department of Electrical, Computer, and System Engineering, Rensselaer Polytechnic Institute,
Troy, New York 12180, USA,
and RPI/IBM Center for Broadband Data Transfer, CII 9015, Rensselaer Polytechnic Institute,
Troy, New York 12180, USA
shurm@rpi.edu*

G. M. TSYMBALOV and D. V. FATEEV

*Institute of Radio Engineering and Electronics (Saratov Branch), ul. Zelyonaya 38,
Saratov 410019, Russia
popov@soire.renet.ru*

Terahertz (THz) response spectra of GaN-based field-effect transistor (FET) arrays are calculated in a self-consistent electromagnetic approach. Two types of FET arrays are considered: (i) FET array with a common channel and a large-area grating gate, and (ii) array of FET units with separate channels and combined intrinsic source and drain contacts. It is shown that the coupling between plasmons and THz radiation in the FET array can be strongly enhanced as compared to a single-unit FET. The computer simulations show that the higher-order plasmon modes can be excited much more effectively in the array of FET units with separate channels and combined source and drain contacts than in FET array with a common channel and a large-area grating gate.

Keywords: Field-effect transistor; plasmons; terahertz detection.

1. Introduction

High-frequency response of field-effect transistors (FETs) with two-dimensional (2D) electron channels is strongly affected by plasma oscillations excited in the channel under the gate electrode.^{1,2} This phenomenon in its various manifestations can be used for the detection,³⁻⁸ mixing⁹⁻¹¹ and generation¹²⁻¹⁷ of terahertz (THz) radiation.

There are two different types of plasma oscillations exist in a FET-like device. They are plasma oscillations excited in either ungated or gated regions of the electron channel. The dispersion relation for ungated plasma oscillation in an infinite 2D electron sheet (with neglect of the electron scattering in 2D electron layer) has a form^{18,19}

^a On leave from: Institute of Radio Engineering and Electronics (Saratov Branch), ul. Zelyonaya 38, Saratov, 410019, Russia. Electronic mail: popov@soire.renet.ru

$$\omega_p^2 = \frac{e^2 N}{m^* \epsilon_0 (\epsilon_1 + \epsilon_2)} k, \tag{1}$$

where ω_p and k are the frequency and wavevector of plasma wave, respectively, N is the areal electron density, ϵ_0 is the dielectric permittivity of vacuum, ϵ_1 and ϵ_2 are the dielectric constants of surrounding materials (substrate and barrier materials, respectively, see Fig. 1), e and m^* are the charge and effective mass of electron, respectively. The values of wavevector k of the ungated plasmons are quantized according to the length of ungated portion of the electron channel. The dispersion described by Eq. (1) is similar to that of gravitational waves on the surface of liquid in a deep basin. Thus this type of plasma waves may be referred to as the "deep water plasmons".²⁰

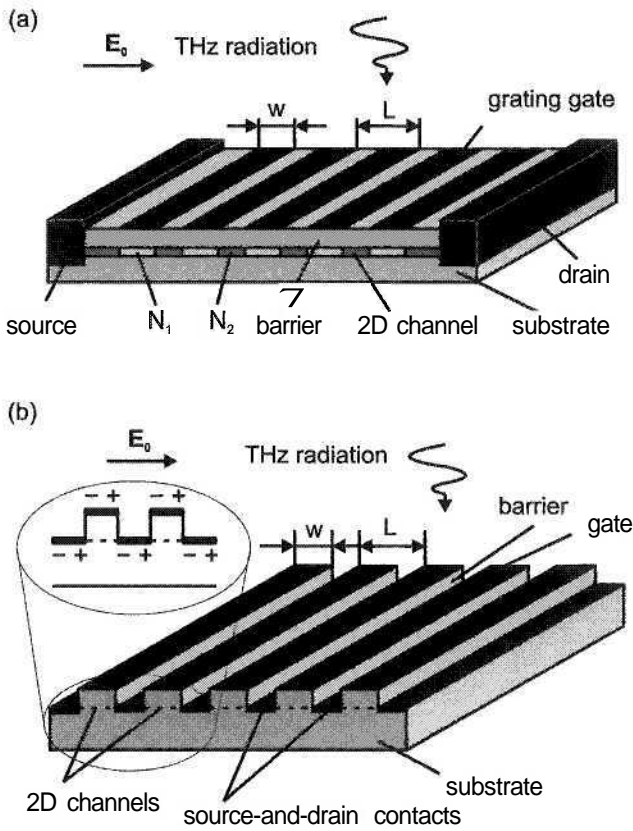


Fig. 1. Schematics of the FET arrays with (a) a common 2D electron channel and a grating gate and (b) with separate electron channels. Polarization of the electric field E_0 of the incident terahertz radiation is shown.

Plasma oscillations of the other type are excited in gated regions of the channel. If the infinite perfectly conductive gate plane is located at distance d from the infinite 2D electron channel then the dispersion relation for this kind of plasma waves is²¹

$$\bar{\omega}_n^{-2} = \frac{e^2 N}{m^* \varepsilon_0 [\varepsilon_1 + \varepsilon_2 \coth(kd)]} k. \quad (2)$$

If the gate plane is close enough to the electron channel, i.e., $kd \ll 1$, one obtains

$$\bar{\omega}_p = \sqrt{\frac{e^2 N d}{m^* \varepsilon_0 \varepsilon_2}} k, \quad (3)$$

which means that the plasma waves in the gated region have frequency-independent phase velocity $s = \bar{\omega}_p / k$. The values of the wavevector of the gated plasmons are quantized according to the length of the gate W , $k = n\pi / W$ with n being an integer. However, plasmon modes with even numbers n can not be excited by THz radiation under the gate contact with symmetric boundary conditions at the gate edges (because of zero net dipole moment of such modes). The plasma wave dispersion described by Eq. (3) is similar to that of gravitational waves on the surface of liquid in a shallow basin. Thus this particular type of plasma waves is referred to us the “shallow water plasmons”.^{1,2} The gated plasmons may be also referred to as the “acoustic plasmons”, since their linear dispersion, Eq. (3), similar to that of the acoustic waves.

If $d \ll W$, the areal electron density in the gated region of the channel is related to the gate voltage by the parallel plate capacitor formula

$$N = \frac{\varepsilon_0 \varepsilon_2}{ed} U_0, \quad (4)$$

where $U_0 = U_g - U_{th}$ is the difference between the gate voltage U_g and the channel threshold voltage U_{th} . Then Eq. (3) for the odd plasmon modes acquires a simple form^{1,2}

$$\bar{\omega}_p = k \sqrt{\frac{eU_0}{m^*}} \quad (5)$$

with the gated-plasmon phase velocity $s = \sqrt{eU_0 / m^*}$ and $k = (2n - 1)\pi / W$ ($n = 1, 2, 3, \dots$).

Plasmon modes excited under the gate contact (gated plasmons) are more attractive for practical applications because their frequencies can be effectively tuned by varying the gate voltage. However, simple estimations show that the frequency of the fundamental plasmon mode ($n = 1$) in a GaN-based HEMT may exceed 10 THz only if the gate length is shorter than 100 nm, which sets a limitation for designing the single-gate FET plasmon devices in the high-frequency THz range. (It should be also noted that HEMT's with such a short gate typically has quite long ungated access regions of the 2D electron layer connecting the gated portion of the channel to the source and drain contacts.²²) Another problem to be solved is how to effectively couple the gated plasmons to THz radiation. Unfortunately, the gated plasmons in a single-gate FET are weakly coupled to THz radiation²³ because: (i) the gated plasmons are strongly screened by the metal gate contact, (ii) they are have a vanishingly small net lateral dipole moment due to their acoustic nature (in this mode, electrons oscillate out-of-phase in the gate contact and

in the channel under the gate, which results in vanishingly small net lateral dipole moment), and (iii) gated plasmons strongly leak into ungated access regions of the channel.²³ Higher-order gated-plasmon modes ($n > 1$) have greater frequencies but even a smaller net dipole moment as compared to the fundamental plasmon mode, and they are more prone to leaking from the gated portion of the channel into the access regions. Hence, the higher-order gated plasmon modes in a single-gate FET can hardly be used for increasing the operation frequency of a single-gate THz plasmonic transistor.

It was suggested in Ref. 23 that the coupling efficiency of gated plasmons to THz radiation can be dramatically enhanced due to their interaction with the ungated portions of the electron channel. A sequence of gated and ungated portions of the channel can be realized, for example, in a FET with a grating gate. In this paper, we calculate the THz plasmon absorptions spectra of the FET with a common channel and a large-area grating gate [see Fig. 1(a)] in frame of the first principle self-consistent electrodynamic approach and show that the gated-plasmon resonances in such a structure increase by several orders of magnitude due to excitation of plasma oscillations in the ungated portions of the channel. We also show that pronounced higher-order plasmon resonances can be excited in a slit-grating gate FET.

The slit-grating-gate coupler is a conventional tool for exciting the higher-order plasmon modes in a 2D electron channel.^{24,25} However, a grating coupler with extremely narrow (sub-100-nm-wide) slits has to be fabricated in order to achieve a measurable strength of the higher-order plasmon resonances^{24,25}. Because of that, we suggest in this paper an alternative FET array design with FET units having separate channels and combined intrinsic source and drain contacts [see Fig. 1(b)]. Such a design allows one to effectively excite higher-order plasmon resonances in a structure having the characteristic dimensions of a micron scale. In earlier paper,²⁶ it was suggested that the coupling between plasmons in the FET channel and THz radiation might be more effective if FET units were arranged in an array. (Also see even earlier relevant paper,²⁷ where periodic ohmic contacts were alloyed into 2D electron channel producing an array of 2D electron diodes.) It was anticipated²⁶ that the plasmons in an array of FET's should absorb (or emit) THz radiation at least by a factor of the number of FET units in the array as stronger as a single-unit FET. In a recent paper,²⁸ we showed that the coupling between plasmons and THz radiation in the FET arrays can be strongly enhanced well beyond that obvious estimation due to a cooperative effect of synchronizing the plasma oscillations in all FET units in the array. In this paper, we demonstrate by computer simulation that intensive higher-order plasmon resonances can be effectively excited in high-frequency THz range (up to 20 THz) in the FET array with separate channels and combined source and drain contacts due to the strong coupling between plasmons and THz radiation. It is shown that in such a device the higher-order plasmon modes are excited much more effectively than in the FET structure with a large-area 2D electron channel coupled to THz radiation by a slit-grating gate.

2. Theoretical Approach

We have calculated the THz plasmon absorption spectra of two different FET arrays shown schematically in Fig. 1 in frame of the first principle self-consistent electrodynamic approach, using the Maxwell equations and the integral equation method developed earlier.²⁹⁻³¹ This approach involves the following steps:

- (1) the Maxwell equations are re-written in the Fourier series representation,
- (2) using the electrodynamic boundary conditions in the channel and gate planes, the amplitudes of the Fourier harmonics of the oscillating-electron-current density are related to those of the in-plane THz electric field in the gate as well as electron channel planes,
- (3) equalizing the sheet electron current density obtained in the previous steps to that in the Ohm's law yields the system of integral equations for the in-plane THz electric field in each (gated and ungated) portion of 2D electron channel as well as in metal contact strips within a period of the structure,
- (4) the system of integral equations is solved numerically by the Galerkin method³² through its projection on an orthogonal set of the Legendre polynomials within the respective interval.

Within this approach we describe the response of electron fluid in each portion of the 2D electron channel, having the equilibrium electron density either N_1 or N_2 [see Fig. 1(a)], by the sheet conductivity in the local Drude model as

$$\sigma(\omega) = i \frac{e^2 N_{1,2}}{m^* (1 - i\omega\tau)}, \quad (6)$$

where τ is the electron relaxation time. The sheet electron density under the gate strips N_1 is calculated in the parallel-plate capacitor model by formula Eq. (4) with d being the gate-to-channel distance (the barrier-layer thickness). Calculations were performed for the characteristic parameters of AlGaIn/GaN HEMT: $\epsilon_2 = 9$, $U_{th} = -(3\div 8)$ V, $d = (8\div 20)$ nm. Surface conductivity of the metal gate strips as well as the source and drain contacts in the FET array structure shown in Fig. 1(b) was assumed to be 2.5 S. Numerical calculations showed that the plasmon absorption spectra do not change noticeably at small (compared to the saturation current) DC currents through the channel (if one neglects the effect of the gated-channel-portion length modulation by the drain current, which does not change the physics of the effects under consideration). Because of that all numerical results below are presented for zero DC drain current. Notice that Eq. (6) describes an inductive admittance arising from a kinetic inductance of the 2D electron system responding to the total in-plane THz electric field (and having a resistive contribution caused by the electron scattering in the 2D channel). A capacitive contribution (which is necessary to produce a resonance in the whole system responding to the external THz electric field) describing the oscillating-charge accumulation in the system is provided by the gate-to-channel capacitance and is accounted for in our approach self-consistently by applying the proper electrodynamic boundary conditions in the channel and gate planes.

3. Grating-Gate FET with a Common Channel

Figure 2 shows the THz absorption spectra of AlGaIn/GaN HEMT with a common channel and a large-area grating gate [see Fig. 1(a)] having $1\text{-}\mu\text{m}$ -wide grating-gate strips for three different grating-gate-slit widths and zero gate voltage. These results demonstrate quite pronounced and well-resolved higher-order plasmon resonances even at room temperature in the structure with sub-micron slits. For narrower slits the higher resonances are excited with larger amplitudes, since narrow grating-gate slits generate strong higher Fourier harmonics of the incident THz wave. As a result, quite pronounced higher-order plasmon resonances may be excited at high THz frequencies up to 7th resonance at about 10 THz (not shown in Fig. 2).

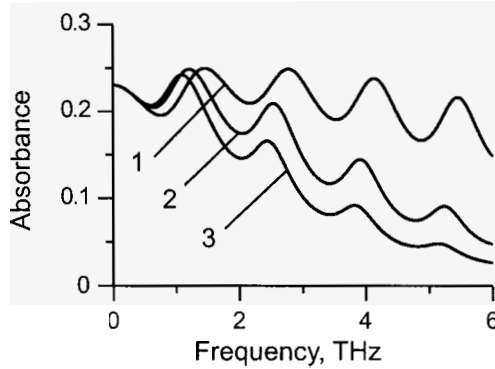


Fig. 2. Absorption spectra of AlGaIn/GaN HEMT with $1\text{-}\mu\text{m}$ -wide grating-gate strips for three different slit widths: $0.1\text{ }\mu\text{m}$ (curve 1), $0.3\text{ }\mu\text{m}$ (curve 2) and $0.5\text{ }\mu\text{m}$ (curve 3) at room temperature ($\tau = 2.27 \times 10^{-13}$ s) for zero gate voltage. $U_{\text{th}} = -3$ V, $d = 8$ nm.

The equidistant spectrum of plasmon resonances shown in Fig. 2 is characteristic of the acoustic-like dispersion, see Eq. (3), which evidences the excitation of the gated-plasmon modes in the structure (the ungated-plasmon modes with their dispersion given by Eq. (1) have much higher frequencies). The radiative damping and dissipation of plasmon oscillations contribute comparably to the total linewidth of the plasmon resonance in the FET with a slit-grating gate, which makes all plasmon resonances stronger. The maximum absorbance of 0.5 at the plasmon resonance is reached when the radiative and dissipative contributions to the resonance linewidth become equal (this condition for the maximum absorption can be also readily understood as matching the free-space impedance to the FET impedance at the plasmon resonance).³⁰ It is worth noting that for narrow slit widths the frequencies of plasmon resonances are the multiples of $2\pi/L$ and not the multiples of π/W as for a single-gate FET structure [see Eq. (3)]. This leads to a blue shift of all plasmon resonances seen in Fig. 2 for shorter periods of the structure. The calculations show that all plasma resonances *become weaker by two orders of magnitude* when the electron density under the grating-gate openings tends to zero. This fact demonstrates a crucial role of the ungated portions of the channel in excitation of the gated plasmons.

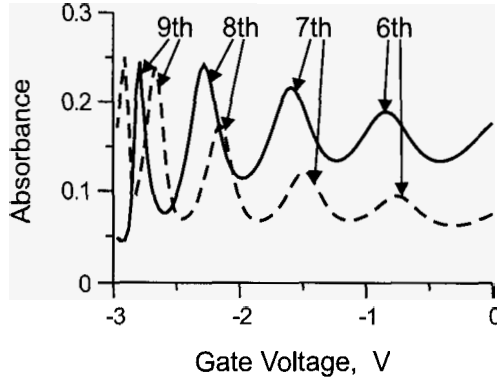


Fig. 3. The same as in Fig. 2 as a function of the gate voltage for two different slit widths: $0.1 \mu\text{m}$ (solid curve), $0.2 \mu\text{m}$ (dashed curve) at frequency 6.86 THz . Plasmon mode numbers are indicated.

Figure 3 shows the terahertz absorption of AlGaN/GaN HEMT with $1\text{-}\mu\text{m}$ -wide grating-gate strips and sub-micron grating-gate slits as a function of the gate voltage. The most intriguing result here is that the intensity of higher-order plasmon resonances (up to 9th order) increases at more negative gate voltage. It happens because the radiative damping of higher-order plasmon modes increases with increasing the electron density modulation along the electron channel (hence, a value of the radiative damping becomes closer to that of the dissipative damping of plasmons, which ensures more favorable condition for excitations of the plasmon resonances).³⁰ Notice also that the plasma resonances excited at more negative gate voltages exhibit a narrower resonance linewidth, ΔU_g . The more negative the resonant value of U_g , the narrower the resonance is. One can easily understand this fact by performing the differentiation of Eq. (5):

$$\Delta U_g \propto \Delta \bar{\omega}_p \sqrt{U_g - U_{th}}, \quad (7)$$

where $\Delta \bar{\omega}_p$ is the linewidth of the plasmon resonance in the frequency domain. Obviously, for a given $\Delta \bar{\omega}_p$, ΔU_g tends to zero when U_g approaches the threshold voltage U_{th} .

4. FET Array with Separate 2D Electron Channels

In Fig. 4 the calculated THz absorption spectra of the FET array with separate electron channels [see Fig. 1(b)] are shown for two different array periods and two different gate-strip width for zero gate voltage. (One can easily imagine a biasing scheme allowing to apply the identical gate and drain bias voltage to every FET in the array.) These results demonstrate that in such a structure, the intensive fundamental and all higher-order plasmon resonances up to the 8th resonance at 20 THz (not shown in Fig. 4) can be excited with comparable amplitudes even at room temperature. All resonances are shifted to higher frequencies for narrower gate-strip width as predicted by Eq. (5).

Plasma oscillations in all unit cells of the periodic FET array are excited with the same phase (and amplitude) dictated by the phase (and amplitude) of the incoming THz

wave. Even without the incoming THz wave, the plasmons, once excited (either by a thermal or stimulated mechanism), oscillate in phase in all unit cells of the FET array shown in Fig. 1(b) because the side metal contacts with high conductivity act as effective synchronizing elements between adjacent unit cells. Applying a mechanical analogy, one may think of rigid crossbars connecting oscillating springs arranged in a chain. Therefore, the plasma oscillations in the FET array behave as a single plasmon mode distributed over the entire area of the array. The formation of this cooperative mode ensures strong coupling between plasmons in the FET array and external THz radiation.²⁸

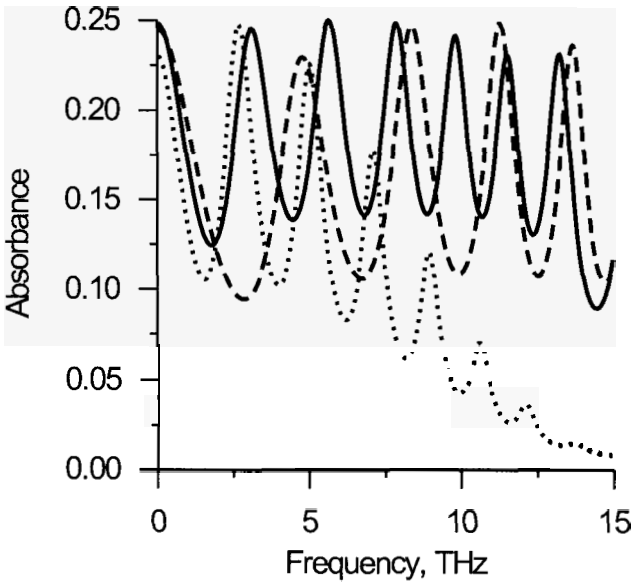


Fig. 4. Terahertz absorption spectra of the array of AlGaIn/GaN HEMT's with separate channels for two different gate lengths of each HEMT unit, $0.8 \mu\text{m}$ (solid curve) and $0.5 \mu\text{m}$ (dashed curve), at room temperature ($\tau = 2.27 \times 10^{-13}$ s) and zero gate voltage. Period of the array is $1.2 \mu\text{m}$. Dotted curve shows the THz absorption spectrum of AlGaIn/GaN HEMT with a homogeneous 2D electron channel and a slit-grating gate with the period $0.9 \mu\text{m}$ and the slit width $0.1 \mu\text{m}$. $U_{\text{in}} = -7.5$ V, $d = 20$ nm (from Ref. 28).

One can see from Fig. 4 that, although there is no by far a sub-100-micron characteristic lateral dimension in the FET array with separate electron channels, the higher-order plasmon modes are excited much more effectively in such a structure than in a FET structure with a common electron channel and a grating gate having the 100-nm-wide slits (see Sec. 3 and the dotted curve in Fig. 4). The physical mechanism of the higher-order plasmon mode excitation is entirely different in these two different structures. In the grating-gate coupler, narrow slits generate strong higher Fourier harmonics of the incident THz field with wavevectors $k = 2n\pi/L$, where L is the period of the grating gate with n being an integer. Then these strong Fourier harmonics excite higher-order plasmon modes with the same wavevectors in the channel when the frequency of the incoming THz wave coincides with the plasmon mode frequency. In contrast, in the FET array with separate channels, the incoming THz radiation induces the

oscillating charges of opposite sign across the vertical (typically sub-100-nm-wide) gap between the gate and side contacts at each edge of the gate contact in every FET unit [see inset in Fig. 1(b)]. These oscillating charges induce the electric fields with the same symmetry as in the plasmon modes excited under the gate contact with symmetric boundary conditions at the gate-contact edges. Hence, plasmon modes are effectively excited when their eigen-frequency coincide with the frequency of the incoming THz radiation.

5. Conclusions

It is shown that the coupling between plasmons and THz radiation in the FET arrays with either joint or separate 2D electron channels can be strongly enhanced as compared to a single-gate FET. Computer simulations show that the higher-order plasmon modes can be excited much more effectively in the FET array with separate channels as compared to the FET array with a common 2D electron channel and a slit-grating gate. Such enhancement is due to a cooperative effect of synchronizing the plasma oscillations in all separate channels in the FET array. These results open a gateway to designing the FET plasmonic devices in the high-frequency THz range up to 15 THz and even higher.

Acknowledgments

This work has been supported in part by the Russian Foundation for Basic Research (Grant No. 06-02-16155), Russian Academy of Sciences program "Quantum Nanostructures", by ONR (Program Managers Scott Steward and Dave Masters) and CRDF (Grant No. 2681).

References

1. M. Dyakonov and M. Shur, *IEEE Trans. Electron Devices* **43**, 380 (1996).
2. M.S. Shur and J.-Q.Lü, *IEEE Trans. Microwave Theory and Techniques* **48**, 750 (2000).
3. W. Knap, Y. Deng., S. Rumyantsev, J.-Q. Lu, M. S. Shur, C. A. Saylor, and L. C. Brunel, *Appl. Phys. Lett.* **80**, 3433 (2002).
4. X. G. Peralta, S. J. Allen, M. C. Wanke, N. B. Harff, J. A. Simmons, M. P. Lilly, J. A. Reno, P. J. Burke, and J. P. Eisenstein, *Appl. Phys. Lett.* **81**, 1627 (2002).
5. A. Satou, I. Khmyrova, V. Ryzhii, and M. S. Shur, *Semicond. Sci. Technol.* **18**, 460 (2004).
6. F. Teppe, W. Knap, D. Veksler, A. P. Dmitriev, V. Yu. Kachorovskii, S. Rumyantsev, and M. S. Shur, *Appl. Phys. Lett.* **87**, 052107 (2005).
7. E. A. Shaner, M. Lee, M. C. Wanke, A. D. Grine, J. L. Reno, and S. J. Allen, *Appl. Phys. Lett.* **87**, 193507 (2005).
8. D. Veksler, F. Teppe, A. P. Dmitriev, V. Yu. Kachorovskii, W. Knap, and M. S. Shur, *Phys. Rev. B* **73**, 125328 (2006).
9. V. Ryzhii, I. Khmyrova, A. Satou, P. O. Vaccaro, T. Aida, and M. S. Shur, *J. Appl. Phys.* **92**, 5756 (2002).
10. A. Satou, V. Ryzhii, I. Khmyrova, and M.S. Shur, *J. Appl. Phys.* **95**, 2084 (2005).
11. M. Lee, M. C. Wanke, and J. L. Reno, *Appl. Phys. Lett.* **86**, 033501 (2005).
12. W. Knap, J. Lusakowski, T. Parenty, S. Bollaert, A. Cappy, V. V. Popov, and M. S. Shur, *Appl. Phys. Lett.* **84**, 2331 (2004).

13. T. Otsuji, M. Hanabe, and O. Ogawara, *Appl. Phys. Lett.* **85**, 2119 (2004).
14. Y. Deng, R. Kersting, J. Xu, R. Ascazubi, X.-C. Zhang, M. S. Shur, R. Gaska, G. S. Simin, M. A. Khan, and V. Ryzii, *Appl. Phys. Lett.* **84**, 70 (2004).
15. N. Dyakonova, F. Teppe, J. Lusakowski, W. Knap, M. Levinshtein, A. P. Dmitriev, M. S. Shur, S. Bollaert, and A. Cappy, *J. Appl. Phys.* **97**, 114313 (2005).
16. V. Ryzhii, A. Satou, and M. S. Shur, *Phys. Stat. Sol. (a)* **202**, 113 (2005).
17. N. Dyakonova, A. El Fatimy, J. Lusakowski, W. Knap, M. I. Dyakonov, M.-A. Poisson, E. Morvan, S. Bollaert, A. Shechetov, Y. Roelens, Ch. Gaquiere, D. Theron, and A. Cappy, *Appl. Phys. Lett.* **88**, 141906 (2006).
18. R. H. Ritchie, *Phys. Rev.* **106**, 874 (1957).
19. F. Stern, *Phys. Rev. Lett.* **18**, 546 (1967).
20. M. Dyakonov and M. S. Shur, *Appl. Phys. Lett.* **87**, 111501 (2005).
21. A. V. Chaplik, *Zh. Eksp. Teor. Fiz.* **62**, 746 (1972) [*Sov. Phys. JETP* **35**, 395 (1972)].
22. T. Parenty, S. Bollaert, J. Mateos, X. Wallart, and A. Cappy, *Proc. 13th Int. Conf. Indium Phosphide and Related Materials (IPRM 2001)*, Nara, Japan, 626 (IEEE, New York, 2001).
23. V. V. Popov, O. V. Polischuk, and M. S. Shur, *J. Appl. Phys.* **98**, 033510 (2005).
24. T. N. Theis, *Surface Science* **98**, 515 (1980).
25. E. Batke, D. Heitmann, and C. W. Tu, *Phys. Rev. B* **34**, 6951 (1986).
26. M. I. Dyakonov and M. S. Shur, *Appl. Phys. Lett.* **67**, 1137 (1995).
27. S. J. Allen, F. Derosa, R. Bhat, G. Dolan, and C. W. Tu, *Physica B+C* **134**, 332 (1985).
28. V. V. Popov, G. M. Tsymbalov, D. V. Fateev, and M. S. Shur, *Appl. Phys. Lett.* **89**, 123504 (2006).
29. O. R. Matov, O. F. Meshkov, and V. V. Popov, *Zh. Eksp. Teor. Fiz.* **113**, 988 9 (1998) [*JETP* **86**, 538 (1998)].
30. V. V. Popov, O. V. Polischuk, T. V. Teperik, X. G. Peralta, S. J. Allen, N. J. M. Horing, and M. C. Wanke, *J. Appl. Phys.* **94**, 3556 (2003).
31. V. V. Popov, G. M. Tsymbalov, and N. J. M. Horing, *J. Appl. Phys.* **99**, 124303 (2006).
32. G. Korn, T. Korn, *Mathematical Handbook*, 2nd ed. (McGraw-Hill, New York, 1968).

ULTRA-HIGHLY SENSITIVE TERAHERTZ DETECTION USING CARBON-NANOTUBE QUANTUM DOTS

YUKIO KAWANO

*Advanced Device Laboratory, RIKEN, 2-1, Hirosawa, Wako, Saitama 351-0198, Japan
PRESTO, Japan Science and Technology Agency (JST), Kawaguchi, Saitama, Japan
ykawano@riken.jp*

TOMOKO FUSE

*Advanced Device Laboratory, RIKEN, 2-1, Hirosawa, Wako, Saitama 351-0198, Japan
Department of Information Processing, Interdisciplinary Graduate School of Science & Engineering, Tokyo
Institute of Technology, 4259, Nagatsuta-cho, Midori-ku, Yokohama 226-8503, Japan
tfuse@riken.jp*

KOJI ISHIBASHI

*Advanced Device Laboratory, RIKEN, 2-1, Hirosawa, Wako, Saitama 351-0198, Japan
CREST, Japan Science and Technology Agency (JST), Kawaguchi, Saitama, Japan
kishiba@riken.jp*

We have observed that Terahertz (THz) irradiation to a carbon nanotube quantum dot (CNT-QD) leads to the generation of an excess current in the Coulomb blockade regime. It was found that this THz detected signal survives even when the incident THz wave is extremely weak (~ 1 fW). This means that the CNT-QD could work as a highly sensitive THz detector.

Keywords: Carbon nanotube; Quantum dot; Terahertz detector

1. Introduction

A highly sensitive detector in the Terahertz (THz) range is now in strong demand in various research fields, such as radio astronomy, biochemical spectroscopy, and medicine as well as solid-state physics. Nevertheless, the photon energy of the THz wave, typically ~ 10 meV, is two to three magnitudes lower than that of the visible light, thus making the development of a high-performance THz detector a difficult task. Recent progress in the fabrication of nanoscale devices, however, opens up a new possibility of significantly enhancing the sensitivity of THz detection. The uses of single electron transistors fabricated from GaAs/AlGaAs heterostructures^{1,2} or an Aluminium³ have made it possible to achieve single-THz-photon counting. The operation of these devices, however, requires a very low temperature environment below 0.3 K using a $^3\text{He}/^4\text{He}$ dilution refrigerator or a ^3He refrigerator. This restricts a range of practical use.

In this regard, a quantum dot (QD) based on a carbon nanotube (CNT)^{4,5} is a promising candidate of a practical and powerful THz detector, because its much smaller size allows higher-temperature (in principle, room-temperature) operation of a single electron transistor, as compared to the previously developed GaAs- or Al-based single electron transistors. Recently, Fuse and coworkers have observed a THz photon-assisted tunneling (PAT) in CNT-QDs at 1.5 K using a standard ⁴He refrigerator.⁶ Their observation provides the interesting possibility that this device can be utilized as a THz photon detector with high-temperature operation. Here, we present another type of THz detection by CNT-QDs, which is different from the PAT. The present sensing mechanism has allowed us to achieve ultra-highly sensitive THz detection with a noise-equivalent power of $\sim 10^{-18}$ W/Hz^{1/2} at 1.5 K.

2. Experimental Setup

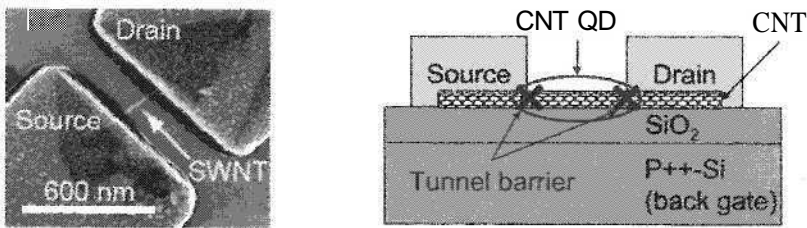


Fig. 1. (a) A scanning electron micrograph image of the CNT-QD. (b) Schematic of the CNT-QD structure.

Figure 1(a) displays a scanning electron microscope image of our CNT-QD. The CNT with a diameter of ~ 1 nm is processed into the QD device by depositing source and drain electrodes at interval of ~ 300 nm. As depicted in Fig. 1(b), this device is placed on a highly-doped silicon substrate with a thermally oxidized surface, and a metal sheet attached in the back surface of the Si substrate works as a gate electrode. In this experiment, CNT-QD devices were immersed into a ⁴He cryostat and single-electron transport measurements were performed at 1.5 K without and with THz irradiation. As a THz light source, we used a quasi-monochromatic and frequency tunable cyclotron emission radiated from a high-mobility two-dimensional electron gas in GaAs.⁷

3. Experimental Results

Figure 2 displays a source-drain current I_{SD} versus a gate voltage V_G in the CNT-QD, and compares the features between without and with THz irradiation. The blue curve (without THz irradiation) clearly shows a periodic oscillatory behavior as a function of V_G , so-called Coulomb blockade oscillation, indicating a single-electron charging effect. On the other hand, when the THz wave is irradiated onto the CNT-QD, it is found that a finite current emerges originally in the Coulomb blockade regime. We observed that this

current change with THz irradiation survives even when the THz power incident on the CNT-QD becomes extremely low (~ 1 fW). This means that the CNT-QD has a very high sensitivity to the incident THz wave.

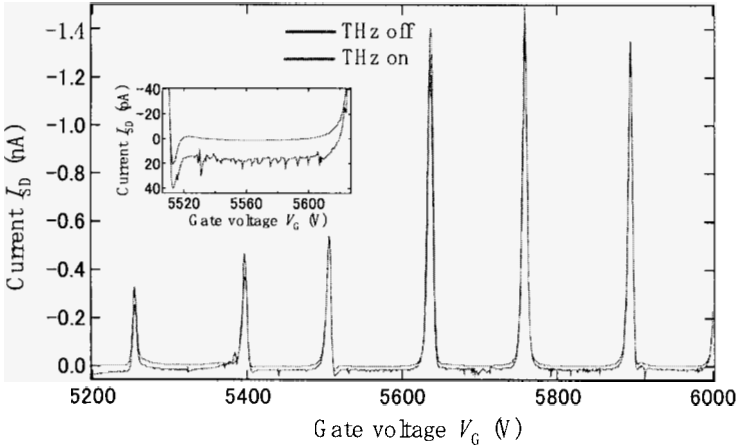


Fig. 2. The source-drain current I_{SD} as a function of the gate voltage V_G in the CNT-QD without and with THz irradiation.

4. Discussion and Remark

The mechanism of the THz detection observed in this work cannot be explained in terms of the PAT, because the PAT manifests itself as an appearance of new-side peaks in the Coulomb blockade regime.⁶ At present, we imagine that, by absorbing the incident THz photon energy, the electrons in the source electrode jump up to the drain electrode due to low tunnel barrier (several meV), formed at the metal-nanotube interfaces.

We would like to make a remark on future application of the CNT-QD to a THz detector. From the minimum of the observable THz power, we roughly estimate that the noise-equivalent power of the CNT-QD-based THz detector reaches the order of 10^{-18} W/Hz^{1/2}. This value is more than three orders of magnitude improved, compared to that of a THz detector based on GaAs/AlGaAs quantum Hall devices that we have previously developed.⁸ We expect that further improvement will be possible by shaping source and drain electrodes into antenna structures to effectively couple the CNT-QD with the incident THz wave.

5. Summary

In summary, we have studied THz irradiation effect on CNT-QDs, and observed, in the Coulomb blockade regime, the generation of the excess current by the THz irradiation. From the minimum of the observable THz intensity, we estimate that the noise-equivalent

power of the present THz detector reaches the order of 10^{-18} W/Hz^{1/2}. We expect that this value will be improved by using source and drain electrodes with the antenna structures.

Acknowledgments

This work was supported by Precursory Research for Embryonic Science and Technology (PRESTO) and Core Research for Evolutional Science and Technology (CREST) of Japan Science and Technology Agency (JST) and by a Grant-in-Aid for Young Scientists (B) (18740176) from the Ministry of Education, Culture, Sports, Science and Technology.

References

1. S. Komiyama *et al.*, *Nature* **403**, 405 (2000).
2. O. Astafiev *et al.*, *Appl. Phys. Lett.* **80**, 4250 (2002).
3. H. Hashiba *et al.*, *Phys. Rev. B* **73**, 081310(R) (2006).
4. S. J. Tans *et al.*, *Nature* **394**, 761 (1998).
5. S. Moriyama *et al.*, *Phys. Rev. Lett.* **94**, 186806 (2005).
6. T. Fuse, Y. Kawano, T. Yamaguchi, Y. Aoyagi, and K. Ishibashi, (submitted).
7. Y. Kawano *et al.*, *Phys. Rev. B* **59**, 12537 (1999).
8. Y. Kawano *et al.*, *J. Appl. Phys.* **89**, 4037 (2001).

GENERATION OF ULTRASHORT ELECTRON BUNCHES IN NANOSTRUCTURES BY FEMTOSECOND LASER PULSES

A. GLADUN, V. LEIMAN, A. ARSENIN, and O. MANNOUN

Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia

V. TARAKANOV

Institute for High Energy Density Associated Institute for High Temperatures RAS, Moscow, Russia

We present numerical investigation of anomalous internal photoelectric effect which is realized in thin film (< 100 nm) structures by surface plasmon (SP) excitation and its interaction with primary laser radiation. SP electric field gain and electron temperature in the SP field have been calculated.

1. Introduction

During last two decades ultrashort laser pulses with high intensities up to $\sim 10^{16}$ W/cm² have been used for electron acceleration. High electric field intensity in such pulses ($\sim 10^9$ V/cm) allows to accelerate electrons up to the energy of 100 MeV.¹ Recently it was demonstrated that laser pulses with relatively moderate intensity ($\sim 10^{10}$ W/cm²) may accelerate electrons up to energy 1 keV with pulse width less than that of the laser pulse. This anomalous effect of electron acceleration has been observed in the experiment of the excitation of a surface plasmon in a thin film structure by ultrashort laser pulse. This effect opens up new prospects in the development of high-performance ultrashort-pulse electron generators for biomedical applications, micro- or nanoelectronics.

There are two schemes for realization of anomalous photoelectric effect in a thin film structure. The first one is a classical Kretschmann scheme,² which has been used for a long time to excite surface plasmons at the interface metal/dielectric (see Fig. 1). Femtosecond electron pulses with energy up to 2 keV^{3,4} have been generated with this scheme. An acceleration gradient about 5 GW/cm was realized using a Ti:sapphir laser ($\lambda = 800$ nm) with an intensity of laser pulse 30 GW/cm² and pulse width 30 fs.

Another scheme (see Fig. 2) was proposed in Ref. 5. In this scheme a metal/metal interface (for example Al/Au) was used instead of classical metal/dielectric one. With this scheme femtosecond electron pulses with the energy of 20 eV were generated by laser pulses with the intensity of 1 GW/cm² and pulse width 400 fs at wavelength $\lambda = 800$ nm.

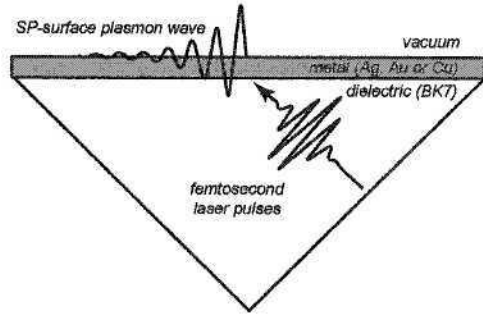


Fig. 1. Kretschmann scheme for anomalous electron acceleration by resonant surface plasmons. The surface of glass prism is coated with a thin film of noble metal like Au, Ag or Cu. A femtosecond laser pulse incident through glass prim at an angle exceeding total internal reflection angle excites surface plasmons.

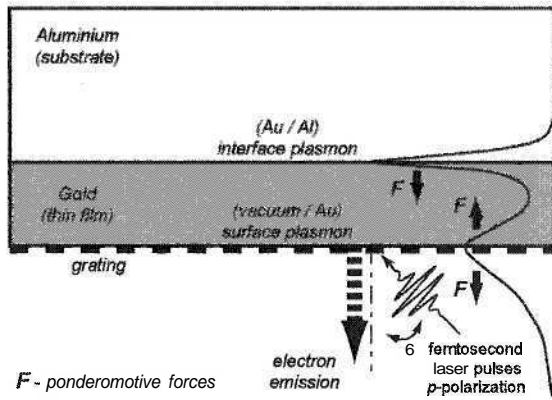


Fig. 2. Scheme of anomalous electron acceleration using a bimetal structure. Maximum of field gradient of excited surface plasmons is at the interface gold - aluminium.

Generation of such electrons from metal surface and their consecutive acceleration may be interpreted as an anomalous photoelectric effect. This effect is a result of ponderomotive force, which is related to the high electric field gradient of resonantly excited surface plasmon.²⁻⁵

A theoretical model for electron generation and acceleration has been proposed and used for numerical simulation.⁵ In that model the Drude formula for the frequency-dependent complex dielectric function was used

$$\epsilon_m(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega - i\nu_d)}, \tag{1}$$

where ω_p is plasma frequency, and ν_d is the damping rate. However for the adequate description of the system under consideration not only temporal but also spatial dispersion and nonlinear processes should be taken into account.

In this paper we develop a model of the interaction of a relatively low intensity ($\sim 10 \text{ GW/cm}^2$) femtosecond laser pulses with a thin metallic film deposited on a glass

prism (Kretschmann scheme), in which plasma is described by plasma current \mathbf{j}_p , which can be obtained from additional nonlinear equations. As analytical solution of the equations of the developed model was not possible two-dimensional numerical simulation has been performed.

2. Equations of the Model and Results of Numerical Simulation

For comprehensive description of processes in conductive and dielectric layers we introduce additional current \mathbf{j}_a into Maxwell's equations:

$$\begin{cases} \text{rot}\mathbf{B} = \frac{4\pi}{c}(\mathbf{j} + \mathbf{j}_a) + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t}, \\ \text{rot}\mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} \end{cases}, \tag{2}$$

where \mathbf{E} and \mathbf{B} – the intensity of electric field and magnetic induction, \mathbf{j} and $\mathbf{j}_a = \mathbf{j}_e + \mathbf{j}_p$ are the densities of normal and additional currents, and

$$\mathbf{j}_e = \frac{\varepsilon - 1}{4\pi} \frac{\partial \mathbf{E}}{\partial t}. \tag{3}$$

To account for metal films properties we use well-known Drude-Lorentz model, in which the metal is interpreted as dense high-collision cold plasma. Based on this model, plasma is described by plasma current, \mathbf{j}_p which can be obtained from the following equations

$$\frac{d\mathbf{j}_p}{dt} + \nu_e \mathbf{j}_p = \frac{\omega_{pe}^2}{4\pi} \mathbf{E} + \frac{e}{mc} [\mathbf{j}_p \times \mathbf{B}], \tag{4}$$

$$\frac{\partial \rho_a}{\partial t} = -\text{div}\mathbf{j}_a, \tag{5}$$

where $\omega_{pe} = \sqrt{\frac{4\pi n_e e^2}{m_e}}$ is the plasma frequency, n_e is concentration of free electrons in the metal, $\rho_e = en_e$ and ν_e is the collision frequency.

The resonance angle of surface plasmons coupling is determined by the following relation:

$$\theta_{sp} = \arcsin \sqrt{\frac{\varepsilon_m / \varepsilon_p}{\varepsilon_m + 1}}, \tag{6}$$

where ε_m – dielectric constant of metal, ε_p – dielectric constant of prism glass (for example, for glass BK7: $\varepsilon_p = 2.296$ at $\lambda = 625$ nm).

Numerical simulation has been performed in x-z geometry with electromagnetic code KARAT.⁶ This code self-consistently solves Maxwell's equations and also equations of relativistic motion by means of "particle-in-cell" (PIC/MCC) method.

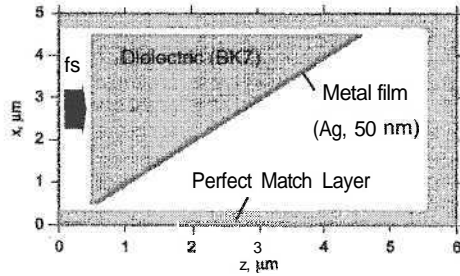


Fig. 3. Computational domain for Kretschmann scheme. To compensate non-physical effects related to the electromagnetic wave reflection at the numerical area boundaries, a special Berenger conditions are used (perfectly matched layer PML)⁷.

The computational domain (Fig. 3) has the following dimensions: $x = 5 \mu\text{m}$, $z = 6 \mu\text{m}$. Number of calculated cells along x - and z - axes are $N_x = 1001$ and $N_z = 1201$, respectively. The parameters of laser: wavelength $\lambda = 625 \text{ nm}$ (second harmonics of Cr^{+3} foresterite laser), a Gaussian pulse duration 30 fs, electric-field value $E = 10^7 \text{ V/cm}$, p-polarization. Refraction index of glass prism $n = 1.456$ ($\epsilon_p = 2.121$) at $\lambda = 625 \text{ nm}$. Thickness of silver film $d = 50 \text{ nm}$. Dielectric constant (at $\lambda = 625 \text{ nm}$) $\epsilon_m = -17.5 + i \cdot 0.5$,⁸ so $\omega_{pe} = 1.26 \times 10^{16} \text{ rad/sec}$ ($n_e = 5.01 \times 10^{22} \text{ cm}^{-3}$), and $\nu_e = 1.37 \times 10^{13} \text{ Hz}$. The refractive index of prism glass is assumed to provide a 45° resonance angle of the coupling of excited surface plasmon.

Figures 4 and 5 show the preliminary results of numerical simulation of femtosecond laser pulses interaction with metallic thin film based on Kretschmann scheme. Two-dimensional distribution of the electric field intensity (Fig. 4) shows that the force lines are concentrated along the interfaces. The field lines originate and terminate on film surface; the electric field has components, which are normal to surface in the vacuum region. Two-dimensional potential distribution depicted in Fig. 5 reveals the propagation of surface plasmons along the surface of metallic film (maxima appear at the interface metal/vacuum). Evaluation of electric field enhancement due to SP excitation gives for the laser field gain a value of about 5 (Fig. 4).

The same structure of the electric field and the same rate of laser field gain were obtained in papers^{3,4} by means of UPAS code, which is based on linear plasma model. However, realization of anomalous photoemission and electron acceleration (in vacuum) requires not only such a structure of electric field but also electric field gain of the order of 100. Although experimental investigations^{3,5} provided clear evidence of the existence of such processes the gain mechanism is not clear yet. Authors of above-mentioned works supposed that in addition to the considered propagating SP (PSP) we should take into account localized SP (LSP) on surface roughness with nanometer size (about 10 nm). Resonant excitation of LSP can produce laser field gain in the range of 100 - 1000.

On the other hand in early works Bloembergen and co-workers⁹ showed that at similar conditions electron temperature may reach 0.5 eV and hence thermoelectronic mission in vacuum becomes possible. This phenomenon can explain anomalous

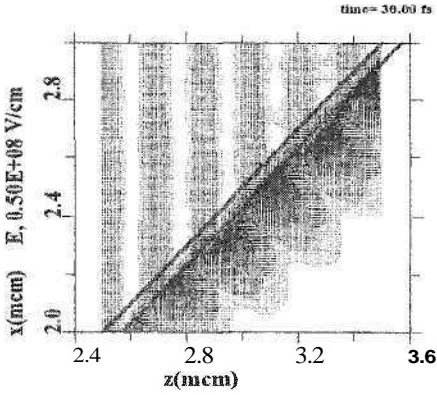


Fig. 4. Two-dimensional (x - z) distribution of the electric field intensity.

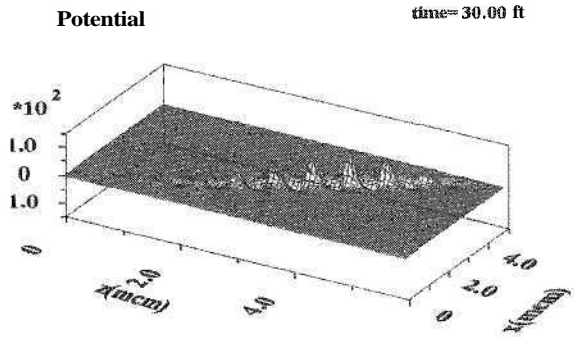


Fig. 5. Two-dimensional (x - z) potential distribution.

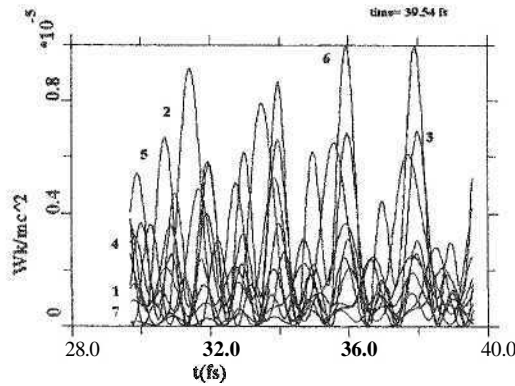


Fig. 6. Kinetic energy of probe electrons at the metal film surface. One division is about 1 eV,

photoelectric effect and electron acceleration. Figure 6 shows time dependences of kinetic energy of probe electrons at the metal film surface obtained from numerical simulation. These dependences demonstrate that average kinetic energy of probe electrons reach a value about 0.5 eV. We believe that electron heating and following thermoelectric emission reduce SP electric field gain.

3. Conclusions

A model of the interaction of a relatively low intensity ($\sim 10 \text{ GW/cm}^2$) femtosecond laser pulses with a thin metallic film deposited on a glass prism in the Kretschmann configuration was developed.

Two-dimensional numerical simulation has been performed using the electromagnetic code KARAT. Such characteristic as the structure of plasmon resonance, magnitude and distribution of electrical field and potential have been determined. Electric field gain by SPs and electron temperature in the SP field have been calculated. We believe that

thermoelectric emission may decrease the electric field gain. This assumption is based on our numerical simulation.

In conclusion it should be noted that external anomalous photoelectric effect indicates that internal anomalous photoelectric effect can be realized in multi-layered semiconductor nanostructures which allows to produce high-energy femtosecond electron pulse in different nano-devices.

Acknowledgments

The authors would like to thank Profs. V. Ryzhii for useful discussions and I. Khmyrova for assistance in the manuscript preparation. This work was supported by the Russian Foundation for Basic Research (Grant No. 05-01-00790) and Russian Federal Agency of Education.

References

1. R. Bingham, *Phil. Trans. R. Soc. A* **364**, 559–575 (2006).
2. T. Tsang, T. Srinivasan-Rao, and J. Fischer, *Phys. Rev. B* **43**(11), 8870-8878 (1991).
3. S. E. Irvine, A. Dechant, and A. Y. Elezzabi, *Phys. Rev. Lett.* **93**(18), 184801 (2005).
4. S. E. Irvine and A. Y. Elezzabi, *Phys. Rev. A* **73**(1), 013815 (2006).
5. J. Kupersztych and M. Raynaud, *Phys. Rev. Lett.* **95**(14), 147401 (2005).
6. V. P. Tarakanov, *User's Manual for Code KARAT* (Springfield, VA: Berkley Research Associate Inc., 1992).
7. J.-P. Berenger, *J. Comput. Phys.* **114**(2), 185-200 (1994).
8. P. G. Johnson and R. W. Christy, *Phys. Rev. B* **6**(12), 4370-4379 (1972).
9. J. G. Fujimoto, J. M. Liu, E. P. Ippen, and N. Bloembergen, *Phys. Rev. Lett.* **53**(19), 1837-1840 (1984).

CHARACTERIZATION OF VOLTAGE-CONTROLLED OSCILLATOR USING RTD TRANSMISSION LINE

KOICHI NARAHARA[†], TAKUO YAMAKI,
TATSUNORI TAKAHASHI, and TORU NAKAMICHI

*Department of Electrical Engineering, Yamagata University, 4-3-16
Yonezawa, Yamagata 992-8510, Japan*

[†]narahara@yz.yamagata-u.ac.jp

Widely-tunable voltage-controlled oscillator using resonant tunneling diodes in a distributed manner is discussed. The circuit configuration and the principle of operation of the VCO are described together with several results of numerical calculations, which include necessary condition of permanent oscillation and how much the oscillation frequency is tuned by the voltage.

Keywords: Oscillator; RTDs; nonlinear transmission line.

1. Introduction

The terahertz electromagnetic waves have attracted a great deal of attention because they are expected to contribute to the applications such as measurement systems with picosecond temporal resolution, over-100-Gbit/s communication systems and spectroscopic imaging systems. Recently, many promising devices are investigated to develop terahertz oscillators, which include quantum cascade lasers,^{1,2} resonant tunneling diodes (RTDs),³ tunnel injection transit time (TUNNETT) diodes,⁴ plasma-wave transistors⁵ and photo-mixing devices.⁶⁻⁸ Among these, the development of electronic THz ICs operating at room temperature that doesn't need photonics is very important, because it enables cost-effective, convenient and small-size THz application systems. In this respect, RTDs are best expected to produce breakthroughs. As for the continuous-wave generation, controllability of the oscillation frequency widens THz applications, including data coding, spectroscopy and chirped pulse radar. We thus investigate the method of voltage-controlled oscillation using RTDs.

A peculiar characteristics of an electromagnetic wave on a RTD line, which means a transmission line periodically loaded with small-size RTDs, is expected in that the step-pulse wave on the line travels forward at the beginning, and then returns to the input. The returned wave is reflected at the input so that this back-and-forth behavior continues permanently. By extracting part of waves from a RTD

line, we can generate an oscillatory electromagnetic wave. The oscillation frequency is determined by the round-trip time of the traveling wave, thus it turns out to lie in terahertz by decreasing the traveling distance or increasing the wave velocity as long as the cutoff of the operation frequency of the loaded RTDs doesn't limit. Moreover, the oscillation frequency can be tuned by the amplitude of applied step pulse.

The propagation of a step pulse on a RTD line is first discussed by Richer in 1966.⁹ He derived an exact solution for the dynamic-steady-state (dss) waveform on an idealized RTD line or a *switch line* to show that it propagates a kink-like waveform without attenuation and with constant time delay per section. His main purpose is to clarify the mechanism of the nondistorted signal propagation on the nerve axon by the simulation using a switch line, so that he paid special attention to the dss solutions. We reconsider a switch line for use in high-speed electronics, especially for the generation of ultra-high-frequency electrical continuous waves.

In this paper, we describe the circuit configuration of the voltage-controlled oscillator (VCO), and its principle of operation using idealized RTD model. Then we discuss several results of numerical calculations, which include necessary condition of permanent oscillation and how much the oscillation frequency is tuned by the voltage.

2. Characteristics of Idealized RTD Lines

Figure 1(a) shows three sections of a RTD line, where L and R show series inductor and series resistor of the unit section, respectively. Length of unit section is dx and the small-letter symbols show the per-unit-length quantities. A shunt RTD is modeled by a capacitor symbolized by $C = cdx$ and a bias-dependent conductance aligned in parallel. The typical current-voltage characteristics is shown by the solid curve in Fig. 1(b). For a while, we idealize it as to be the dotted curve in Fig. 1(b) for simplicity. The voltage levels where RTDs exhibit negative differential resistance (NDR) are degenerated to a threshold, V_{th} , and the finite conductance above V_{th} is neglected. We can then write the current through RTDs as $G\theta(V_{th} - V)V$ [$G = gdx$], where $\theta(V)$ shows the Heaviside function. This nonlinear transmission line is exactly called a switch line by Richer. Hereafter, we call the voltage range greater (less) than V_{th} region I (II). Then, the voltage wave is influenced by finite shunt conductance in region II and is not affected by conductance in region I.

The VCO requires the input pulse crosses V_{th} as shown in Fig. 1(c). It is a rising step pulse having an amplitude of V_0 . In general, the transmission equation of the linear dispersive line allows exponential solutions together with sinusoidal ones.¹⁰ It is possible for the exponential waves to develop on a nonlinear transmission lines including a switch line, although the exponential solutions are usually discarded because they diverge at infinity so that they cannot satisfy any physically meaningful boundary conditions. As for Richer's dss solution of a switch line, this exponential waves develop at the voltage range above and below the threshold of the loaded

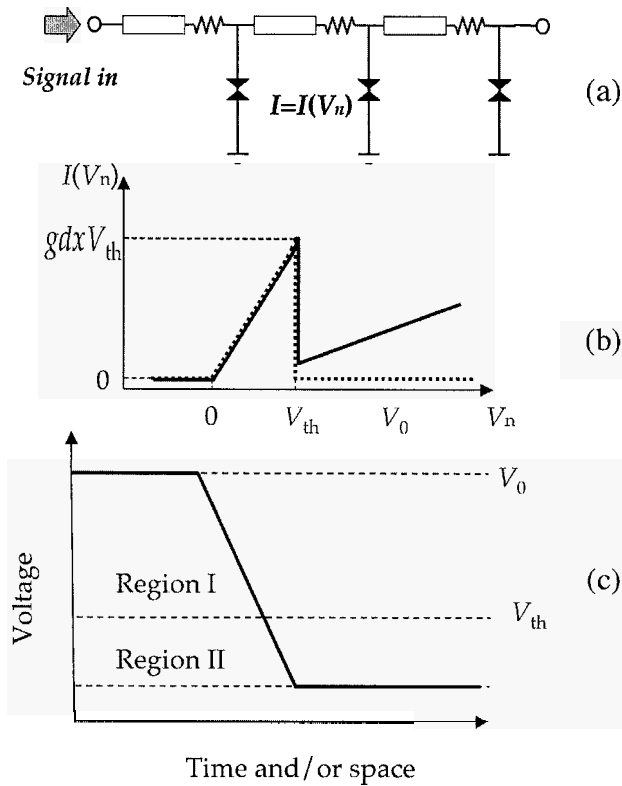


Fig. 1. Schematic of VCO. (a) Circuit diagram, (b) current-voltage characteristics of RTD, (c) signal application.

switch. These two exponential waves are continuously united at the threshold so that they form a kink-like pulse.

When a rising step pulse, whose bottom and top voltage levels respectively lie in the voltage range below and above the threshold of switches, is input to a switch line, we found that the ordinary sinusoidal wave develops at the voltage range above the threshold instead of the exponential counterpart. The sinusoidal part can unite continuously with the leading exponential wave. This exotic pulse is not stable so that it becomes attenuated with the transmission on a switch line and finally disappears. Then, a stable dss kink-like pulse develops at the point where the forward pulse disappears and starts to travel backward. In the following, we clarify the reason why the pulse on a switch line exhibits this back-and-forth motion.

When denoting the voltage and current at the n -th node as V_n and I_n , respec-

tively, the transmission equation of a switch line¹¹ is given by

$$l \frac{dI_n}{dt} = -rI_n + \frac{V_n - V_{n+1}}{dx}, \quad (1)$$

$$c \frac{dV_n}{dt} = -g\theta(V_{th} - V_n)V_n + \frac{I_{n-1} - I_n}{dx} \quad (2)$$

The line is linear and dispersive when considering only regions I and II. Thus, it is meaningful to analyze the dispersion curves in each region. The phase velocities normalized by $1/\sqrt{lc}$ of the modes having wave number k are explicitly given by

$$v_{I,\sin} = \frac{1}{2kdx} \sqrt{-\alpha^2 + 16 \sin^2 \frac{kdx}{2}}, \quad (3)$$

$$v_{I,\exp} = \frac{1}{2kdx} \left(-\alpha + \sqrt{\alpha^2 + 16 \sinh^2 \frac{kdx}{2}} \right) \quad (4)$$

$$v_{II,\exp} = \frac{1}{2kdx} \left(-\alpha - \beta + \sqrt{(\alpha - \beta)^2 + 16 \sinh^2 \frac{kdx}{2}} \right) \quad (5)$$

where $v_{I,\sin}$, $v_{I,\exp}$ and $v_{II,\exp}$ show the normalized phase velocities of the sinusoidal mode in region I, the exponential mode in region I and the exponential mode in region II, respectively. Moreover, α and β are normalized series resistance and shunt conductance given by

$$\alpha = rdx \sqrt{\frac{c}{l}} \quad (6)$$

$$\beta = gdx \sqrt{\frac{l}{c}}. \quad (7)$$

We employ the sign convention: $\omega t - kdxn$ for positive propagation. Thus, in region I, k has to be negative, while in region II, it has to be positive for exponential modes. Figure 2 shows the dispersion curves for $\alpha = 0.18$ and $\beta = 2.0$. The horizontal axis shows the wave number with a unit of the inverse of dx , and the vertical axis shows the normalized phase velocities. For steady or quasi-steady pulse propagation, the velocity of the pulse edge in region I has to be the same as that in region II. When the pulse propagate forward, the exponential mode in region II can only couple with the sinusoidal mode in region I, because no exponential counterpart is present. At point P , both the velocity and the steepness are coincident, so that the forward pulse occupies the region around this point in Fig. 2. Note that the wavenumber at P is relatively large, so that for the pulse propagating forward, short-wavelength oscillatory waves are supposed to be observed in region I, which is supported by an exponential wave developed in region II. On the other hand, when the pulse travels backward, the exponential mode in region II can only couple with the exponential mode in region I. Points Q_1 and Q_2 , where both modes have a coincident velocity, correspond to this backward-traveling pulse. By the continuity of the voltage and

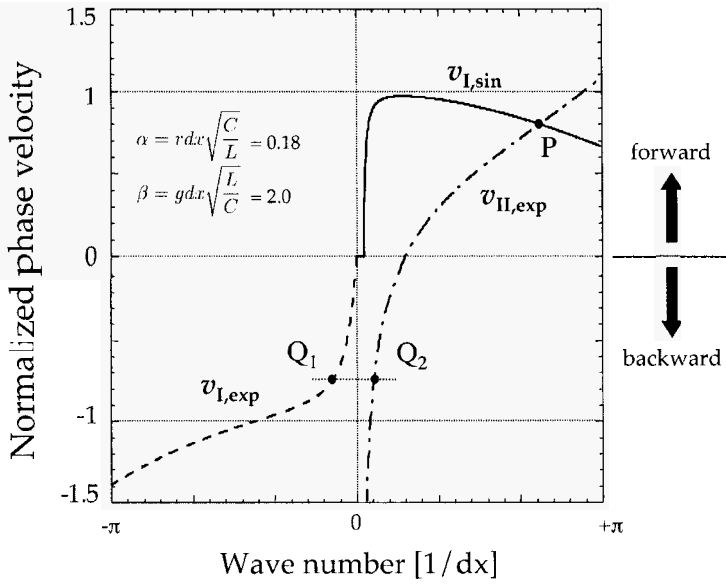


Fig. 2. Dispersion curves of idealized RTD line. Solid curve: sinusoidal mode in region I, dashed curve: exponential mode in region I, dotted curve: exponential mode in region II.

current at V_{th} , we obtain the condition,

$$\frac{V_{th}}{V_{th} - V_0} = \frac{f(\omega_2, k_2)}{f(\omega_1, k_1)}, \quad (8)$$

$$f(\omega, k) = \frac{r + l\omega}{1 - e^{-kdx}} \quad (9)$$

where $k_{1,2}$ and $\omega_{1,2}$ show the wave number and the frequency of $Q_{1,2}$, respectively. This condition determines the vertical position of $Q_{1,2}$. Note that the vertical position depend on both the pulse amplitude and the threshold. Moreover, the evanescent waves are developed in both regions, so that no oscillation will be observed, if

$$\beta < \frac{4}{\alpha} \sinh^2 \left[\sin \left(\frac{\alpha}{4} \right) \right] \quad (10)$$

This process is illustrated in Fig. 3. The spatial position on the line is shown horizontally, and the voltage is shown vertically. Figure 3(a) shows the behavior of the forward pulse. Because the forward pulse combining a sinusoidal mode in region I and an exponential mode in region II does not have any permanent profile, it is not stable and becomes attenuated and finally disappears, so that only exponential pulse edge is left in region II, as shown in Fig. 3(b). This edge develops an exponential modes in region I to form a dss pulse, and then it starts to travel backward as shown in Fig. 3(c). When the backward dss pulse reaches the input end, it is reflected as in Fig. 3(d), so that it again starts to travel forward

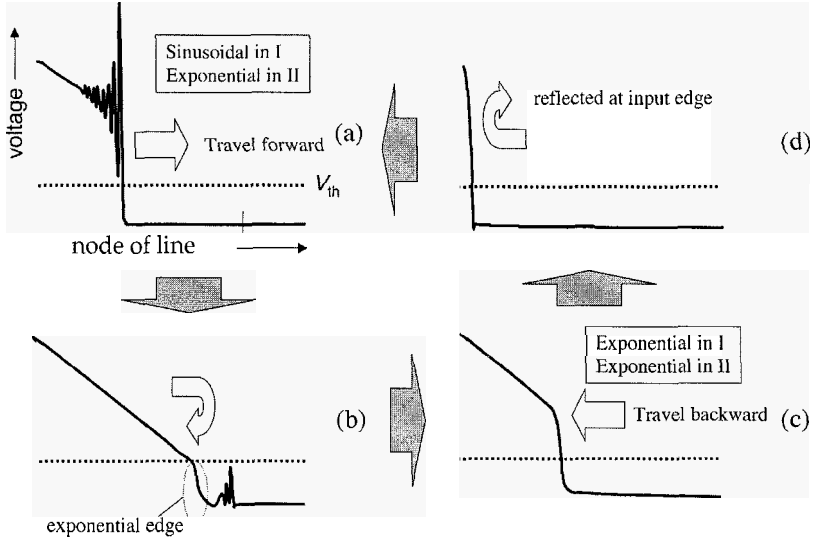


Fig. 3. Principle of operation of VCO

as an unstable sinusoidal-exponential wave. This process continues permanently with proper boundary condition at the input, thus succeeding in the generation of electromagnetic continuous waves.

3. Numerical Results

For the characterization of the RTD line as a VCO, we numerically solved Eqs. (1) and (2). We set c , l , V_{th} and dx to 1.0 pF/mm, 1.0 nH/mm, 0.2 V and 5 μm , respectively. The rise time of the step pulse was 1.0 ns with the form of raised cosine. We carried out calculations with several different values of α , β and V_0 . Figure 4(a) shows the dependence of the oscillation frequency on α and β at $V_0 = 1.5V$. As α and β increases, the oscillation frequency increases. This is because the round-trip distance decreases by the large attenuation. Moreover, the dark region fails to keep oscillation. As Eq. (10) shows, the oscillation is not universal but is sensitive to the line parameters including attenuating elements and the threshold. On the other hand, Fig. 4(b) shows the voltage controllability of the oscillation frequency. The horizontal axis shows V_0 . The left-vertical axis shows the oscillation frequency, and the right-vertical axis shows the distance to the turning point. The threshold was set to 0.2 V. As the amplitude decreases, the oscillation frequency increases. In particular, it rapidly increases at the amplitude of around 0.4 V. The oscillation frequency becomes maximal at 0.3 V, and then decreases. The oscillation is possible at the voltage amplitude in the very neighborhood of the threshold. Although this calculation is based on idealized RTD model, the oscillation frequency potentially

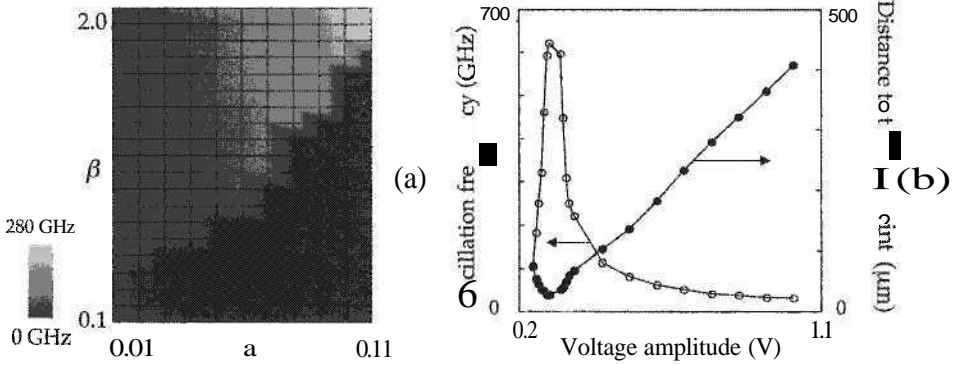


Fig. 4. Numerical characterization of VCO. (a) Dependence of oscillation frequency on line parameters, (b) voltage controllability of VCO.

covers very wide bandwidth of over 600 GHz.

For the permanent oscillation, the degree of reflected pulse has to be kept constant at each point when the pulse edge returns the input interface. In fact, some transmission line connecting a voltage source with the VCO is required practically, so that it seems difficult to ensure the invariance of the reflection condition. However, the voltage at the interface restores the DC level of the source at every temporal point when the pulse edge reaches the turning point. Thus, every reflected edge at the interface has the fixed amplitude, keeping constant reflection.¹² In the case of practical RTDs, the finite conductance in region I is also taken into consideration. The dispersion curves vary not so much through the introduction of finite conductance in region I that the fundamental property of the line is preserved. Voltage-controlled oscillation is thus expected to be possible for practical RTD line.

4. Conclusions

We investigated the voltage-controlled oscillation in a RTD line. The operation bandwidth of the state-of-the-art RTD covers terahertz band. By employing this high-speed RTDs in the line, we think the proposed method will contribute to the development of an electronic terahertz oscillator to no small extent.

Acknowledgment

This work was financially supported by the Grant-in-Aid for Scientific Research (S) from the Japan Society for the Promotion of Science, Japan.

References

1. R. Köhler, A. Tredicucci, F. Beltram, H. E. Beere, E. H. Linfeld, A. G. Davies, D. A. Ritchie, R. C. Iotti, and F. Rossi, *Nature* 417, 156 (2002).

2. B. S. Williams, S. Kumar, Q. Hu, and J. L. Reno, *Electron. Lett.* **40**, 431 (2004).
3. N. Ohashi, S. Hattori, S. Suzuki and M. Asada, *Jpn. J. Appl. Phys.* **44**, 7809 (2005).
4. J. Nishizawa, P. Plotka, H. Makabe, and T. Kurabayashi, *IEEE Microwave and Wireless Components Letters* **15**, 597 (2005).
5. M. I. Dyakonov and M. S. Shur, *IEEE Trans. Electron. Devices* **43**, 1640 (1996).
6. V. Ryzii, I. Khmyrova, and M. Shur, *J. Appl. Phys.* **91**, 1875 (2002).
7. H. Ito, T. Furuta, S. Kodama, and T. Ishibashi, *Electron. Lett.* **36**, 1809 (2000).
8. M. Hanabe, T. Otsuji, T. Ishibashi, T. Uno, and V. Ryzii, *Jpn. J. Appl. Phys.* **44**, 3842 (2005).
9. I. Richer, *IEEE Trans. Circuit Theory* **CT-13**, 388 (1966).
10. R. Hirota, *Soriton no Suuri (Mathematical Theory of Soliton via Direct Method)* (Iwanami Shoten, Tokyo, 1992) Chap. 1 (in Japanese).
11. A. Scott, *Active and Nonlinear Wave Propagation in Electronics* (Wiley-Interscience, New York, 1970).
12. K. Narahara, *J. Appl. Phys.* **100**, 64908 (2006).

INFRARED QUANTUM-DOT DETECTORS WITH DIFFUSION-LIMITED CAPTURE

N. VAGIDOV, A. SERGEEV, and V. MITIN

*ECE Department, University at Buffalo, The State University of New York, 312D Bonner Hall,
Buffalo, NY 14260, USA
nizami@eng.buffalo.edu
asergeev@eng.buffalo.edu
vmitin@buffalo.edu*

Employing Monte-Carlo simulations we investigate parameters and optimize geometry of IR quantum-dot detectors with diffusion-limited capture into the dots surrounded by potential barriers. Our results show that structures with modulation doping of interdot matrix provide an effective separation of the localized and conducting electron states. In these structures, the capture time is mainly determined by the quantum dot concentration and the height of potential barriers around dots. The capture is not sensitive to the dot positions. It also weakly depends on the electric field up to the characteristic value, at which significant electron heating allows hot electrons to overcome the barriers. Optimizing the carrier capture and transit times, we show that quantum-dot structures have a lot of potentials for increasing the photoconductive gain and for the development of IR room-temperature detectors.

Keywords: Quantum-dot photodetector; diffusion-limited capture; modulation doping; gain.

1. Introduction

Starting from the pioneering work,¹ various quantum-dot structures have been intensively studied as promising candidates for development of effective infrared photodetectors with numerous applications in middle and long-wavelength IR ranges.²⁻⁵ It is expected that the quantum-dot structures can provide long lifetime of excited electrons, which in turn increases the photoconductive gain and leads to higher responsivity and higher operation temperature.⁴⁻⁶ Capture processes in quantum-dot structures can be significantly suppressed by a proper choice of geometry of the structure and specific doping. The carrier lifetimes would strongly increase, if the dots would be separated from the “conducting channels” by the potential barriers created by means of the modulation doping. In our previous works⁶⁻⁸ we considered various realizations of potential barriers and their effects on the detector performance.

In the current paper, we investigate a model of the photodetector operating at the room temperature, where the electron mean free path is relatively short and the photoelectron capture is determined by the electron diffusion in the field of potential barriers surrounding the charged dots.⁸ In Ref. 8, we obtained a set of analytical results

related to the trapping cross-section at small electric fields applied to the structure. Here, we employ Monte-Carlo simulations to optimize the geometry of the IR quantum-dot detectors with diffusion-limited capture into the dots surrounded by potential barriers. We study the capture and transit times as functions of the dot positions, sensor geometry, and electric field applied. Finally, we calculate the photoconductive gain and discuss the optimal structures.

2. Theoretical Model

In theoretical model quantum dots are considered as ideal spheres of radius a that are located at distances $2b$ from each other in x -, y -, and z -directions. The interdot matrix is doped by donors, and electrons from ionized donors move into the quantum dots. The negative charge of electrons in the dots and positive charge of ionized donors outside of the dots ($r > a$) form the barrier potential, $V(r)$, shown in Fig. 1. The number, N , of trapped by quantum dots electrons is determined by the concentration of donors. For the case of $N \gg 1$, we can exploit the model of the spherically-symmetric distribution of the charge and quasi-classical

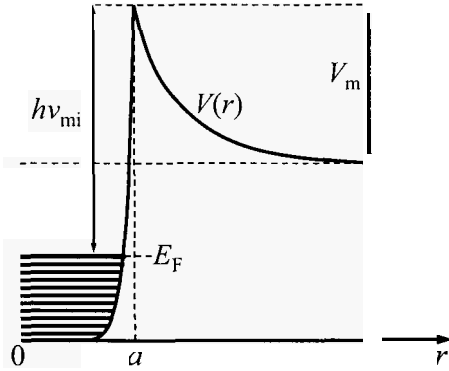


Fig. 1. Energy band diagram.

approximation for the electron Fermi energy. The electrostatic potential can be described in this approximation as

$$V(r) = V_0 \left(\frac{b}{r} + \frac{r^2}{2b^2} - \frac{3}{2} \right), \quad V_0 = \frac{eN}{\epsilon b [1 - (a/b)^3]}, \quad (1)$$

where ϵ is the dielectric permittivity of the interdot matrix. From Eq. (1) it is obvious that $V(r)$ is defined by several parameters, such as quantum dot radius, a , interdot distance, $2b$, number of electrons in quantum dot, N , that can be varied in wide ranges.

We consider room temperatures when the relaxation in quantum dots is fast enough to use the theoretical model with diffusion limited capture. The cross-section of electron capture is described by the following formulas:

$$\sigma = \pi \alpha a^2 \tilde{\nu} \exp\left(-\frac{eV_m}{kT}\right) \left(1 + \frac{3}{4} \frac{\alpha a}{\ell} F(V)\right)^{-1}, \quad (2)$$

$$F = a \exp\left(-\frac{eV_m}{kT}\right) \int_a^b \frac{dr}{r^2} \exp\left(\frac{eV(r)}{kT}\right), \quad (3)$$

where \tilde{v} is the electron thermal velocity, ℓ is the electron mean free path, α is the probability for an electron at $r \leq a$ to be captured by the quantum dot, and $V_m = V(a)$. Equations (2) and (3) were obtained for zero voltage applied to the structure. They are valid only if the positions of the turning points for the electron moving in the potential $V(r)$ are close enough to the quantum dot area. Satisfactorily this is reached at $a \geq 3$ nm. The formula for the electron capture time, τ_{capt} , is

$$\tau_{capt} = (N_d \sigma \tilde{v})^{-1}, \quad (4)$$

where N_d is the dot concentration. We use the same model for Monte-Carlo computations to investigate the dependence of electron capture times and gain of the photodetector on the concentration of dots and on the type of three-dimensional interdot space configurations.

3. Results of Monte-Carlo Simulations

We exploit Monte-Carlo program to simulate transport of three-dimensional electrons in InAs/GaAs photodetectors with the same concentration of InAs dots but with their

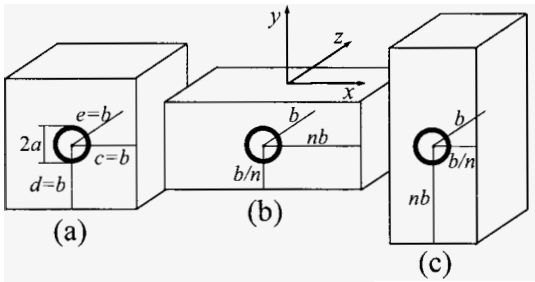


Fig. 2. Possible quantum-dot photodetector configurations: (a) the cube configuration: $c = d = e = b$; (b) the flattened parallelepiped configuration: $c = nb$, $d = b/n$, and $e = b$; and (c) the elongated parallelepiped configuration: $c = b/n$, $d = nb$, and $e = b$. Number n defines deviation from square configuration.

different space locations in GaAs interdot matrix: (a) the cube configuration when all three distances between quantum dots $2c$, $2d$, and $2e$ in x -, y -, and z -directions, respectively, are equal to $2b$; (b) the parallelepiped configuration, which is flattened out in the direction of the voltage applied; and (c) the parallelepiped configuration, which is elongated in the direction of the voltage applied, (see Fig. 2). The concentration of quantum dots, $N_d = 1/cde = 1/b^3$, stays the same for all of three

configurations. As we consider InGaAs dots embedded into the GaAs matrix, all parameters were taken for these materials. The Monte-Carlo simulation program is based on the model with electrons that may populate Γ -, L-, and X-valleys. It includes three major types of electron scatterings on: 1) acoustic, 2) polar optical, and 3) intervalley phonons. The electric field, applied in the y -direction provides strongly non-equilibrium electron distributions, but, at the same time, the increase of electron energy is not enough to initiate the avalanche processes or to modify significantly the potential barriers.

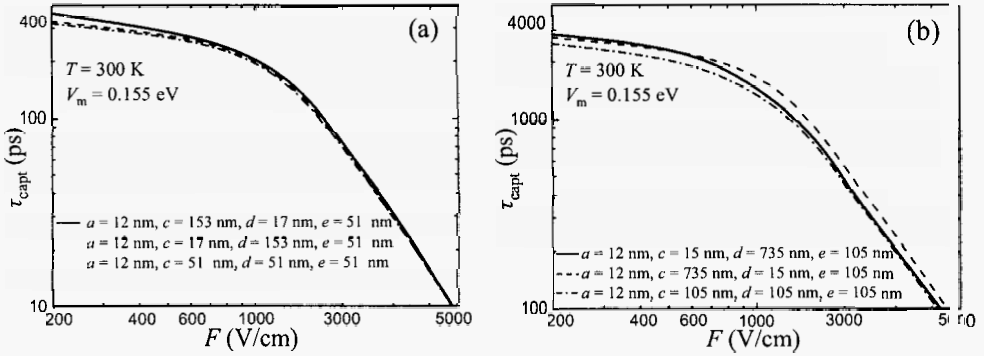


Fig. 3. Dependences of the photoelectron capture time on the electric field applied to the structures with the same value of dot radius and three different space configurations: (a) high and (b) low dot concentrations.

The results of simulations show that for the chosen variation of geometry the electron capture time, τ_{capt} , weakly depends on dot arrangement and that it is primarily determined by the dot concentration. In Figs. 3(a) and 3(b), we present dependences of the photoelectron capture times on the electric field applied to the structures with three space configurations (see Fig. 2) and dot concentrations of $1/(51 \text{ nm})^3 = 7.54 \times 10^{15} \text{ cm}^{-3}$ (a) and $1/(105 \text{ nm})^3 = 8.63 \times 10^{14} \text{ cm}^{-3}$ (b). As it is expected, the electron capture times for the cube configurations are slightly smaller (around 10 %) than that for parallelepiped configurations. The capture times for flattened and elongated parallelepiped configurations are almost the same. Thus, the capture time is not sensitive to dot arrangement. As it is seen in the Figs. 3(a) and 3(b), the capture time strongly depends on dot concentration and electric field, F . At electric fields greater than 1 kV/cm, the capture time, τ_{capt} , decreases by two orders of magnitude, while the electric field increases several times and such a behavior is practically insensitive to the structure geometry. Note, that the abrupt decrease of the carrier lifetime in the electric field explains the negative differential photoconductivity in quantum-dot detectors.⁹

Calculated dependences on the height of the potential barrier, V_m , for the considered three space configurations of quantum dots show similar behavior. At electric fields below 1 kV/cm, the dependence of the electron capture time on V_m is close to the exponential law that was obtained in the absence of the electric field (Eq. 2). In strong electric fields and low potential barriers, the capture time is practically independent on the barrier height [see Figs. 4 (a), 4(b), and 4(c)].

The main reason of such behavior is the heating of electrons in the applied electric field. At the characteristic field, average electron energy becomes comparable with the potential barrier. It is clarified in Fig. 5, where average energies, $\tilde{\epsilon}$, and population of Γ - and L-electrons are presented for electric fields up to 5 kV/cm. At room temperature the Γ -valley is mainly populated and at low electric fields, $F < 1 \text{ kV/cm}$, electrons there stay weakly heated. At higher fields, above 3 kV/cm, the average energy of Γ -electrons, which mostly contribute to the electron transport (population of Γ -valley is almost 80 % even at

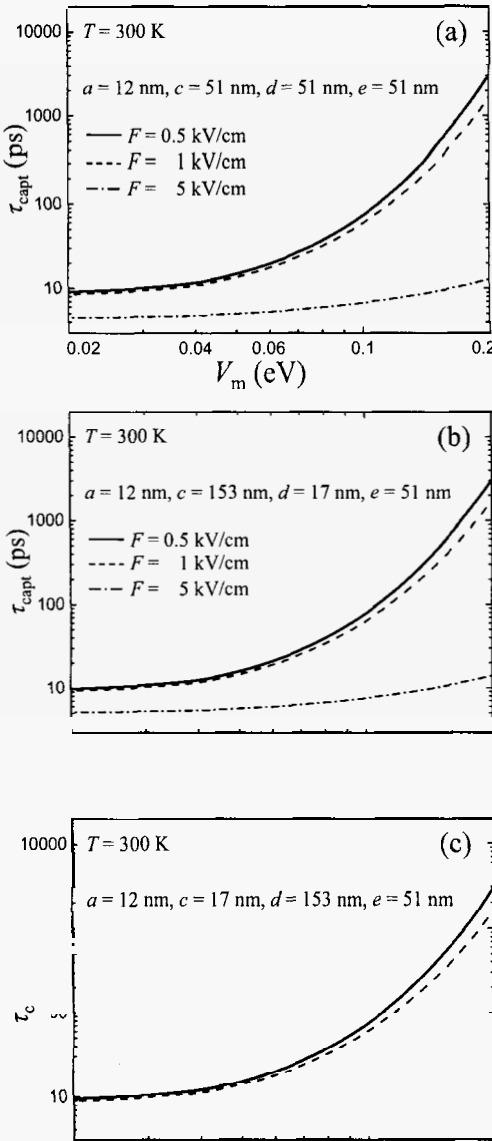
electric field $F = 5$ kV/cm; see Fig. 5), is almost linearly proportional to the electric field. As the average energy of Γ -electrons becomes comparable with the potential barrier, V_m ,

the electrons can be captured by quantum dots with much higher probability. As we can see from Figs. 3 and 4 electron capture times decrease with the increase of applied electric field, which causes the electron heating. The considered effects are fully described by Eqs. (2) and (3) if we will replace parameter kT by $\tilde{\varepsilon}$.

It is also important to note that at the room temperature, the electron mean free path ℓ is controlled by the inelastic electron scattering on the polar optical phonons. It is around 7 nm, which is smaller than the dot diameter. So, the electron capture is not sensitive to the relaxation processes inside of the dot. At the same time the mean free path is substantially smaller than the interdot distance. This is why the electron heating does not change the hierarchy of the relaxation lengths and the characteristic distances and, therefore, heating is practically independent on geometry of the structure.

The calculated electron capture time dependences on the electric field, F , for three space configurations allow us to analyze the gain dependences on electric field. The photoconductive gain is defined as a ratio of capture time, τ_{capt} , to the transit time across the active region of the device, τ_{tr} :

$$g = \tau_{capt} / \tau_{tr}. \quad (5)$$



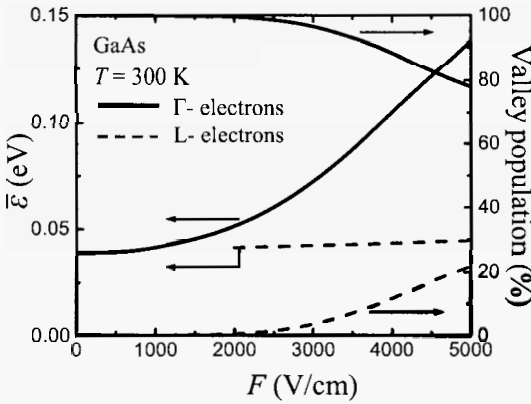


Fig. 5. The average electron energy and population of Γ - and L-valleys of GaAs by electrons as a function of the electric field, F .

Found in the process of Monte Carlo modeling average drift velocities and capture times for two quantum dot concentrations and three space configurations (Fig. 3) are used to calculate dependences of the photoconductive gain on the electric field, F at the fixed value of the potential barrier (see Fig. 6). Despite the differences in the dot concentrations and dot space configurations, the behavior of gain dependences in Figs. 6(a) and 6(b) are very similar. The gain reaches maximum value at

electric field of ~ 1 kV/cm, which is also the characteristic field for the dependences shown in Fig. 3. At low electric fields, where capture time, τ_{capt} , is almost constant, the gain increases due to a decrease of the transit time, τ_{tr} ; at high electric fields, the gain decreases due to exponential decrease of the capture time.

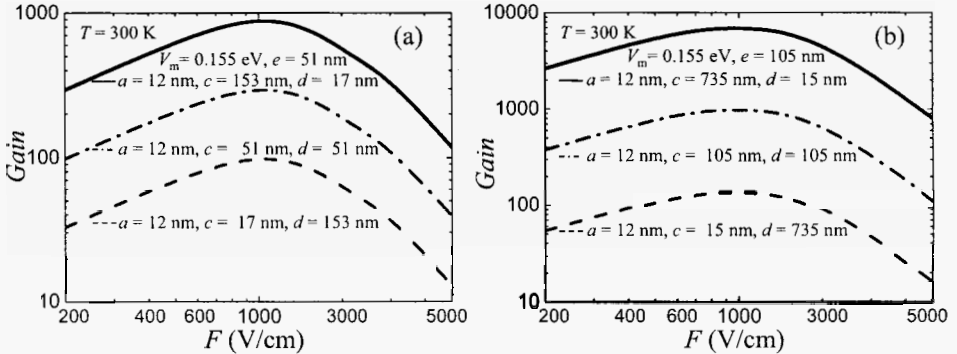


Fig. 6. Photoconductive gain per layer as a function of the electric field for two quantum dot concentrations and three space configurations that are shown in the Fig. 2.

In summary, the quantum dot structures with modulation doped barriers that separate localized and conducting electrons are promising candidates for the development of room-temperature IR photodetectors. Monte-Carlo simulations demonstrate weak dependence of capture times on the space geometry and electric field up to 1 kV/cm. The transit time substantially decreases in the configuration, which is flattened out in the direction of the voltage applied. Thus, the photoconductive gain is mainly limited by the electron heating. The probability of carrier capture substantially increases, when the

average electron temperature becomes comparable with the height of the potential barriers. The detector performance can be significantly improved by properly engineered potential barriers as well as by the optimal choice of quantum dot concentration, space arrangement, and operating regime. In particular, high barriers and structures flattened in the direction of the applied voltage [as in Fig. 2(a)] lead to a substantial increase of gain.

Acknowledgements

The research was supported by NYSTAR and PRF grants.

References

1. V. Ryzhii, *Semicond. Sci. Technol.* **11**, 759 (1996).
2. Lin Jiang, Sheng S. Li, Nien-Tze Yeh *et al.*, *Appl. Phys. Lett.* **82**, 1986 (2003).
3. W. Zhang, H. Lim, M. Taguchi *et al.*, *Appl. Phys. Lett.* **86**, 191103 (2005).
4. P. Bhattacharya, X. H. Su, S. Chakrabarti *et al.*, *Appl. Phys. Lett.* **86**, 191106 (2005).
5. J. Phillips, *J. Appl. Phys.* **91**, 4590 (2002).
6. V. V. Mitin, V. I. Pipa, A. V. Sergeev *et al.*, *Infrared Phys. and Technol.* **42**, 467 (2001).
7. V. Ryzhii, I. Khmyrova, V. Mitin *et al.*, *Appl. Phys. Lett.* **78**, 3523 (2001).
8. A. Sergeev, V. Mitin, and M. Stroscio, *Physica B* **316-317**, 369 (2002).
9. V. Ryzhii, *Appl. Phys. Lett.* **78**, 3346 (2001).

This page intentionally left blank

MAGNETORESISTANCE IN Fe/MgO/Fe MAGNETIC TUNNEL JUNCTIONS

N. N. BELETSKII, S. A. BORYSENKO, and V. M. YAKOVENKO

*Usikov Institute of Radiophysics and Electronics of the National Academy of Sciences of
Ukraine, 12 Acad. Proskura Str., 61085 Kharkov, Ukraine
beletski@ire.kharkov.ua*

G. P. BERMAN

*Theoretical Division, MS B213, Los Alamos National Laboratory, Los Alamos, New Mexico
87545, USA
gpb@lanl.gov*

S. A. WOLF

*Department of Physics and Materials Science, University of Virginia, Charlottesville, Virginia
22903, USA
saw6b@virginia.edu*

The magnetoresistance of Fe/MgO/Fe magnetic tunnel junctions (MTJs) was studied taking into consideration image forces. For MTJs with an MgO insulator, explanations are given of the giant tunneling magnetoresistance (TMR) effect and the effect of the increasing TMR with an increase in MgO insulator thickness. It is demonstrated that the electron current density through MTJs can be high enough to switch the magnetization of a ferromagnetic electrode.

Keywords: Tunneling magnetoresistance; magnetic tunnel junctions.

1. Introduction

At present, “ferromagnetic metal/MgO/ferromagnetic metal” magnetic tunnel junctions (MTJs) have demonstrated high values of the tunneling magnetoresistance (TMR) at room temperature.^{1–10} Such MTJs are characterized by low potential barriers of the order of 0.34–0.50 eV.^{3,8,6,10} The low-barrier MTJs are very favorable for developing magnetoresistive random-access-memory (MRAM).^{11,12}

In this paper the influence of image forces on the TMR and on the electron current density in low barrier MTJs is taken into account. We show that in such MTJs the electron current density is enough for switching the magnetization of the ferromagnetic electrodes.^{7,10}

2. Theoretical Model

Consider MTJs consisting of two identical ferromagnetic metals separated by a non-magnetic insulator. The magnetizations of the two ferromagnetic electrodes are collinear [parallel (P) and antiparallel (AP)] and directed along the boundaries. We use a two-band model of free electrons in ferromagnetic metals. We assume, for simplicity, that the electron masses in the ferromagnetic electrodes and in the insulator are equal. Then, the transmission coefficient of electrons through the MTJs depends only on the transverse electron energy E_z .

The transverse parts of the Hamiltonians for the ferromagnetic electrodes and the insulator, which correspond to the movement of electrons through the MTJs (z axis), have the following form:

$$H_L = -\frac{\hbar^2}{2m} \frac{d^2}{dz^2}, \quad z < 0 \quad (1)$$

$$H_B = -\frac{\hbar^2}{2m} \frac{d^2}{dz^2} + U_{b\sigma_L} - \frac{eV_a z}{d} + \Phi(z), \quad 0 < z < d \quad (2)$$

$$H_R = -\frac{\hbar^2}{2m} \frac{d^2}{dz^2} + (\sigma_L - \sigma_R)\Delta - eV_a, \quad z > d, \quad (3)$$

where $U_{b\sigma_L} = U_b + \sigma_L\Delta$, U_b and d are the height and width of the potential barrier, e and m are the charge and mass of the electron, V_a is the applied voltage, Δ is the half-width exchange splitting of two bands with different directions of electron spin in ferromagnetic electrodes, $\sigma_{L,R} = \pm 1$ are conventional spin indexes corresponding to orientation of the electron spin along (the sign $+$ or \uparrow) and opposite to (the sign $-$ or \downarrow) the direction of magnetization of ferromagnetic electrodes.

The image force potential $\Phi(z)$ has the following form:¹³

$$\Phi(z) = -\frac{e^2}{2\varepsilon_b} \left(\frac{1}{2z} + \sum_{n=1}^{\infty} \left[\frac{nd}{(nd)^2 - z^2} - \frac{1}{nd} \right] \right), \quad 0 < z < d, \quad (4)$$

where ε_b is the permittivity of the barrier material.

We assume that in the left ferromagnetic electrode, the electrons with $\sigma_L = 1$ have a larger Fermi energy ($E_{F\uparrow} = \mu + \Delta$, μ is the electrochemical potential) than the electrons with $\sigma_L = -1$ ($E_{F\downarrow} = \mu - \Delta$). As a result, we have $k_{F\uparrow} > k_{F\downarrow}$, where $k_{F\uparrow,\downarrow} = \sqrt{2m_F E_{F\uparrow,\downarrow}}/\hbar$ is the Fermi electron momentum. The electrons with $\sigma_L = 1$ are called majority electrons, and the electrons with $\sigma_L = -1$ are called minority electrons. Later on we assume that the magnetization changes its orientation only in the right ferromagnetic electrode. We have chosen iron (Fe) as the ferromagnetic metal, for which $k_{F\uparrow} = 1.09 \text{ \AA}^{-1}$, $k_{F\downarrow} = 0.42 \text{ \AA}^{-1}$.¹⁴ Using these numerical values we find ($m = m_0$, m_0 is the free electron's mass): $\mu \approx 2.6 \text{ eV}$, $\Delta \approx 1.93 \text{ eV}$, $E_{F\uparrow} \approx 4.53 \text{ eV}$, $E_{F\downarrow} \approx 0.67 \text{ eV}$.

Following the paper,¹⁴ we describe the height of the potential barrier U_b by the non-dimensional parameter $u_b = (U_b - \mu)/E_{F\uparrow}$.

Figure 1 shows the influence of ε_b on the shape of the potential barrier at $V_a = 0.5 \text{ V}$, $u_b = 0.1$, $d = 1.5 \text{ nm}$ for majority (a) and minority (b) electrons. It

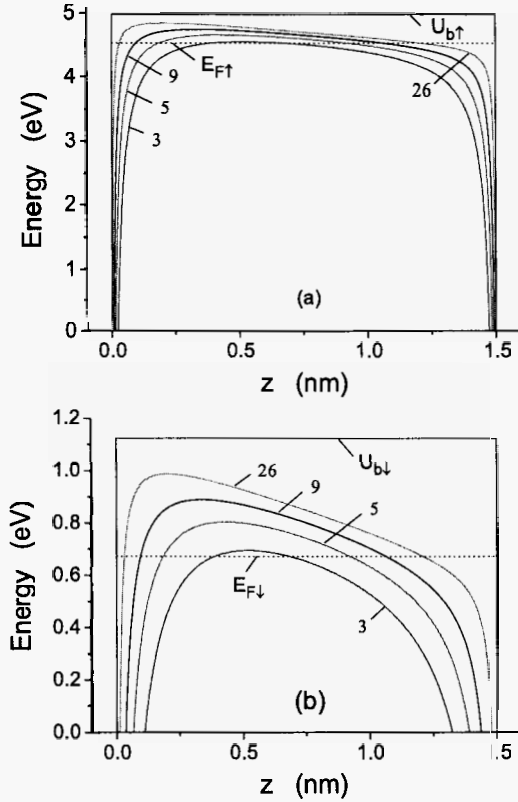


Fig. 1. The shape of the potential barrier at $V_a = 0.5$ V, $u_b = 0.1$, $d = 1.5$ nm for majority (a) and minority (b) electrons. The numbers next to the curves show the values we considered for ε_b .

is evident from Fig. 1 that the shape of the potential barrier depends significantly on the value of the insulator permittivity ε_b . With decreasing ε_b , the height and thickness of the potential barrier decrease.

Assume that an electron characterized by the σ_L -state in the left ferromagnetic electrode, tunnels through the potential barrier and occupies the σ_R -state in the right ferromagnetic electrode. Then, the densities of the partial electron currents $J_{\sigma_L\sigma_R}$ through the MTJs can be described by the following equation¹⁵

$$J_{\sigma_L\sigma_R} = \frac{emk_B T}{4\pi^2 \hbar^3} \int_0^\infty T_{\sigma_L\sigma_R} \ln \left\{ \frac{1 + \exp[(E_{F\sigma_L} - E_z)/k_B T_0]}{1 + \exp[(E_{F\sigma_L} - E_z - eV_a)/k_B T_0]} \right\} dE_z, \quad (5)$$

where k_B is the Boltzmann constant, T_0 is the temperature, $f_{\sigma_L}(E) = \{1 + \exp[(E - E_{F\sigma_L})/k_B T_0]\}^{-1}$ is the Fermi function, $E_{F\sigma_L}$ is the Fermi energy of the electrons in the σ_L -state. The partial transmission coefficients $T_{\sigma_L\sigma_R}(E_z, V_a)$ of electrons can be obtained using the propagation matrix formalism.¹⁶

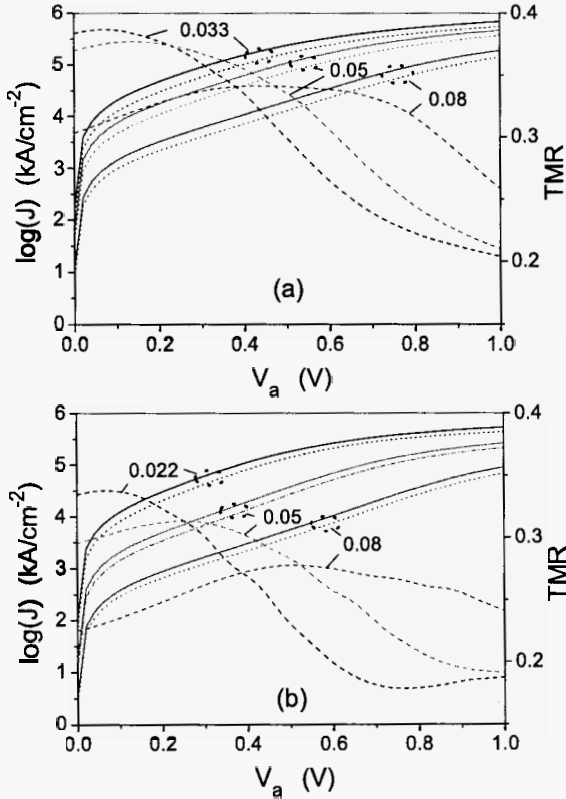


Fig. 2. $J_P(V_a)$ (solid lines), $J_{AP}(V_a)$ (dot lines) (the left axis of ordinates) and $TMR(V_a)$ (dashed lines, the right axis of ordinates) for $d = 1.5$ nm, $T_0 = 300$ K, (a) $\epsilon_b = 5$, (b) $\epsilon_b = 9$. The numbers next to the curves show the corresponding values of u_b .

The magnetoresistance of the tunneling junction can be presented in the form

$$TMR = (J_P - J_{AP})/J_P, \tag{6}$$

where $J_P = J_{\uparrow\uparrow} + J_{\downarrow\downarrow}$ and $J_{AP} = J_{\uparrow\downarrow} + J_{\downarrow\uparrow}$ are the densities of the electron currents through the MTJs for the parallel and antiparallel magnetization orientations of the ferromagnetic electrodes, correspondingly.

3. Numerical Results

We will consider the bias voltage dependencies of the electron current densities through MTJs and accompanied TMR for a number of dimensionless heights u_b of the potential barrier. In Fig. 2 the left ordinate axis corresponds to dependencies of $J_P(V_a)$ (solid lines) and $J_{AP}(V_a)$ (dot lines), and the right ordinate axis is related to $TMR(V_a)$ dependencies (dashed lines) at $d = 1.5$ nm, (a) $\epsilon_b = 5$, (b) $\epsilon_b = 9$.

It follows from Fig. 2 that electron current densities through MTJs depend on ϵ_b .

With decreasing ε_b they grow. At a chosen value of ε_b , electron current densities through MTJs increase with decreasing u_b .

It should be noted that electron current densities rise sharply with increasing V_a . They can be quite sufficient for current-driven magnetization switching. Thus, for example, the critical current densities required for current-driven magnetization switching in CoFeB/MgO/CoFeB MTJs annealed at 300° C are found to be as low as 8.8×10^5 A/cm² with accompanying TMR = 0.42.⁷ It follows from Fig. 2 that critical current densities can already be reached at $V_a \gtrsim 0.1$ V. For example, at $V_a = 0.5$ V, $u_b = 0.08$ we find that $J_P = 1.8 \times 10^7$ A/cm² ($J_{AP} = 1.2 \times 10^7$ A/cm²) for $\varepsilon_b = 5$ and $J_P = 5.5 \times 10^6$ A/cm² ($J_{AP} = 3.8 \times 10^6$ A/cm²) for $\varepsilon_b = 9$.

4. Conclusions

The general properties of low-barrier MTJs containing MgO insulator can be described by a two-band model of free electrons in ferromagnetic electrodes by taking into account image forces. The image forces change the shape of the potential barrier considerably, and in that way they increase the TMR and the electron current density in MTJs. To obtain simultaneously high values of both electron current densities and TMR, it is necessary to use low-barrier MTJs with small values of barrier material permittivity. The results obtained here can be used to develop MRAM with desirable characteristics.

Acknowledgment

This work was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396.

References

1. S. Parkin, Spin-polarized current in spin valves and magnetic tunnel junctions, *MRS Bulletin* **31**, 389–394 (2006).
2. S. S. P. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, and S.-H. Yang, Giant tunneling magnetoresistance at room temperature with MgO(100) tunnel barrier, *Nature Materials* **3**, 862–867 (2004).
3. S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions, *Nature Materials* **3**, 868–871 (2004).
4. S. Yuasa, A. Fukushima, T. Nagahama, K. Ando, and Y. Suzuki, High tunnel magnetoresistance at room temperature in fully epitaxial Fe/MgO/Fe tunnel junctions due to coherent spin-polarized tunneling, *Jpn. J. Appl. Phys.* **43**, L588–L590 (2004).
5. D. D. Djayaprawira, K. Tsunekawa, M. Nagai, H. Maehara, S. Yamagata, N. Watanabe, S. Yuasa, Y. Suzui, and K. Ando, 230% room-temperature magnetoresistance in CoFeB/MgO/CoFeB magnetic tunnel junctions, *Appl. Phys. Lett.* **86**, 092502 (2005).
6. K. Tsunekawa, D. D. Djayaprawira, M. Nagai, H. Maehara, S. Yamagata, N. Watanabe, S. Yuasa, Y. Suzui, and K. Ando, Giant tunneling magnetoresistance effect in

- low-resistance CoFeB/MgO(001)/CoFeB magnetic tunnel junctions for read-head applications, *Appl. Phys. Lett.* **87**, 072503 (2005).
7. J. Hayakawa, S. Ikeda, Y. M. Lee, R. Sasaki, T. Meguro, F. Matsukura, H. Takahashi, and H. Ohno, Current-driven magnetization switching in CoFeB/MgO/CoFeB magnetic tunnel junctions, *Jpn. J. Appl. Phys.* **44**, L1267–L1270 (2005).
 8. J. Hayakawa, S. Ikeda, F. Matsukura, H. Takahashi, and H. Ohno, Dependence of giant tunnel magnetoresistance of sputtered CoFeB/MgO/CoFeB magnetic tunnel junctions on MgO barrier thickness and annealing temperature, *Jpn. J. Appl. Phys.* **44**, L587–L589 (2005).
 9. S. Ikeda, J. Hayakawa, Y. M. Lee, T. Tanikawa, F. Matsukura, and H. Ohno, Tunnel magnetoresistance in MgO-barrier magnetic tunnel junctions with bcc-CoFe(B) and fcc-CoFe free layers, *J. Appl. Phys.* **99**, 08A907 (2006).
 10. G. D. Fuchs, J. A. Katine, S. I. Kiselev, D. Mauri, K. S. Wooley, D. C. Ralph, and R. A. Buhrman, Spin torque, tunnel-current spin polarization, and magnetoresistance in MgO magnetic tunnel junctions, *Phys. Rev. Lett.* **96**, 186603 (2006).
 11. S. Parkin, X. Jiang, C. Kaiser, A. K. Roche, and M. Samant, Magnetically engineered spintronic sensors and memory, *Proc. IEEE* **91**, 661–680 (2003).
 12. S. A. Wolf, D. Treger, and A. Chtchelkanova, Spintronics: the future of data storage?, *MRS Bulletin* **31**, 400–403 (2006).
 13. J. G. Simmons, Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film, *J. Appl. Phys.* **34**, 1793–1803 (1963).
 14. J. C. Slonczewski, Conductance and exchange coupling of two ferromagnets separated by a tunneling barrier, *Phys. Rev. B* **39**, 6995–7002 (1989).
 15. R. Tsu and L. Esaki, Tunneling in a finite superlattice, *Appl. Phys. Lett.* **22**, 562–564 (1973).
 16. J.-G. S. Demers and R. Maciejko, Propagation matrix formalism and efficient linear potential solution to Schrodinger's equation, *J. Appl. Phys.* **90**, 6120–6129 (2001).

MODELING AND IMPLEMENTATION OF SPIN-BASED QUANTUM COMPUTATION

M. E. HAWLEY and G. W. BROWN

*Materials Science and Technology Division, Los Alamos National Laboratory, MS G755,
Los Alamos, NM 87545, USA**
hawley@lanl.gov

G. P. BERMAN

*Theory Division, Los Alamos National Laboratory, MS B213,
Los Alamos, NM 87545, USA*
gpb@lanl.gov

Quantum computation is a revolutionary new paradigm that has experienced tremendous growth since the mid 1990s and has inspired a number of ingenious schemes whose long-term goal is to realize a large-scale, fast, parallel, and easily fabricated quantum computer (QC). Silicon-based solid-state proposals using spins of dopants, such as phosphorus, as qubits are attractive because of the long spin relaxation times and the potential for scaling the device to a large number of qubits and integrating the QC with existing silicon technology. We propose a modified Kane¹ architecture in which we address the difficult fabrication and addressing problems by simplifying the device structure and utilizing an existing, reliable optical detection method. The device consists of linear arrays of P atoms, which will be entangled through their exchange interactions. Scanning tunneling microscopy is used to fabricate the device and an external magnetic field and a large field gradient enable individual spin addressability. The lone P electron spin is used to control the nuclear spin qubit orientation. Excitons, generated in the substrate via a laser, are used to probe the spin states. A He-3 cryostat-based spin manipulation and readout system has been developed but many challenges exist in making a nano-scale quantum device in a real materials system.

Keywords: Quantum computing; solid-state; spin qubits.

1. Introduction

Quantum computation is a revolutionary new paradigm that has experienced tremendous growth since the mid 1990s. The motivating force for this growth has been the well-recognition potential that quantum computation could have for solving certain problems currently intractable on conventional computers. This dramatic computational potential is achieved through exploiting quantum superposition to attain massive parallelism.

Initially, theoretical efforts led the way in this area by developing new quantum-

* MST-8, Materials Science and Technology Division, MS G755, Bikini-Atoll Road, P.O. Box 1663, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States of America.

based computer algorithms. Although the experimental efforts to develop and implement viable quantum computer (QC) concepts lagged behind their theoretical counterparts, this quest has none-the-less inspired a number of ingenious schemes over the last 10 years based on cold ion traps,² neutral atoms in micro-optical elements,³ nitrogen-vacancy centers in diamond,⁴ and superconducting flux⁵ or charge qubits,⁶ for example. With current QC technologies, primitive algorithms (including Shor's quantum factoring algorithm to factor the number 15)⁷ have been demonstrated by room temperature liquid-state nuclear magnetic resonance techniques, a system that is known to be inherently unscalable for large computations. However, the long-term goal in this quest is to achieve a large-scale, fast, and easily fabricated QC. Silicon-based solid-state schemes using either nuclear or electron spins of dopants, such as phosphorus, as qubits are attractive for a number of reasons: phosphorus nuclei in silicon have long spin relaxation times (up to hours); such a device is potentially scaleable; and a silicon-based device can potentially be integrated with existing silicon technology. The long decoherence times relax the time constraints on how fast quantum operations and readout need to be carried out before the prepared quantum states are lost to environmental interactions.

2. Quantum Computation

Classical computer memory is stored in bits, which can have two values, 0 or 1, while a QC stores information in a 2-level quantum system. For each quantum bit or "qubit", zero and one quanta are denoted $|0\rangle$ and $|1\rangle$. One attractive possibility is a spin- $1/2$ nucleus in a static magnetic field where the equivalent to these states is spin up $|\uparrow\rangle$ or spin down $|\downarrow\rangle$. In addition, each *qubit* can exist in both states simultaneously, i.e. in a superpositional state:

$$|\psi_{qubit}\rangle = a|0\rangle + b|1\rangle. \tag{1}$$

Thus, for N qubits one can store 2^N numbers simultaneously. One of the simplest examples of a quantum gate is a combination of a Hadamard operation followed by a controlled-not (CNOT),

$$|\uparrow_1\uparrow_2\rangle \xrightarrow{\pi/2, \omega_1} \frac{1}{\sqrt{2}}|\uparrow_1 + \downarrow_1\rangle\uparrow_2 = \frac{1}{\sqrt{2}}|\uparrow_1\uparrow_2 + \downarrow_1\uparrow_2\rangle \xrightarrow{\pi, \omega_2 (spin\downarrow)} \frac{1}{\sqrt{2}}|\uparrow_1\uparrow_2 + \downarrow_1\downarrow_2\rangle. \tag{2}$$

The first operation is a Hadamard $\pi/2$ operation at the Larmor frequency of spin 1, ω_1 , that puts spin 1 in a superpositional state. The second CNOT π operation at the Larmor frequency of spin 2, ω_2 , is a conditional flip that rotates spin 2 if and only if spin 1 is down. These operations combine to create the entangled state of the two spins, i.e. either both spins are up or both spins are down.

D. J. DiVincenzo enumerated a set of general criteria for implementation of a suitable candidate for a functional QC in 1997.⁸ They include:

- (1) A scaleable physical system with well-characterized qubits,

- (2) Environmental isolation, i.e. a long decoherence time,
- (3) An external mechanism and universal set of operations to manipulate individual qubits for controlling qubit-qubit interactions to create complex multi-qubit quantum states,
- (4) An initialization procedure to produce a ground state such as $|000\dots\rangle$,
- (5) A reliable readout method to determine the final state of the system.

3. Background - Silicon-Based Solid-State Quantum Computer

In 1998 Bruce Kane put forward the first concept for a phosphorus-in-silicon solid-state quantum computer (SSQC)¹ that met the DiVincenzo criteria listed above. The architecture consists of a square array of phosphorus spin $\frac{1}{2}$ atoms 20 nm apart buried 20 nm below the surface. The phosphorus atoms, which are shallow donors, interact through electron-electron exchange interactions. The phosphorus array is fabricated in an ultra-high vacuum chamber using scanning tunneling microscope (STM) methods, described below. The overlayer used to bury the array consists of a thick layer of homoepitaxial silicon covered by a thin insulating barrier layer, both of which are grown *in situ* by molecular beam epitaxy. Electron beam lithography is used to fabricate “A-gates” and “J-gates” on the surface. An “A-gate” is positioned over each qubit while a “J-gate” is placed between each pair of “A-gates”. Since the qubits are only 20 nm apart, the electrodes must be much less than 10 nm wide to prevent interference from and cross-talk between neighboring electrodes. Each “A-gate” controls the hyperfine interaction between a phosphorus nucleus and its lone electron and, therefore, determines its Larmor frequency in an external magnetic field. This provides individual addressability of each qubit. Voltages applied to the “J-gates” regulate the exchange interaction between electrons on adjacent atoms, i.e. turn on and off the interaction between the qubits. Originally, magnetic resonance force microscopy (MRFM) was proposed as a viable individual spin readout method. Although there has been successful detection of an individual electron spin by MRFM,⁹ the measurement took over almost two weeks to complete and single nuclear spin detection has not been demonstrated. Even if the measurement time could be reduced significantly, the sensitivity would have to be improved by nearly 2000 times to detect a nuclear spin. This readout technique is impractical for detecting a single nuclear spin for QC applications.

The Kane approach, described above, requires a complicated set of fabrication steps. Both precise atomic-scale placement of qubits in a solid-state material and achieving nanometer-scale fabrication of surface gates have been demonstrated. A major challenge is to register the surface electrodes with the buried qubits. A second barrier to success is the development of a reliable readout method. Surmounting both of these challenges is a necessary condition for even testing the concept and achieving demonstrations of even the simplest quantum operations. Completion of this entire fabrication procedure and The difficulties with implementing this scheme, that is the alignment of surface electrodes with the buried phosphorus atoms, have not resulted in success in dealing with other issues such decoherence effects due to the gate electrodes, and development of other spin

readout methods. Although an alternative charge readout method has been proposed, little progress has been made toward demonstrations of quantum gate operations.

4. New SSQC Approach

To address these issues, we have simplified the device structure, eliminated the need for surface gate electrodes, and identified an existing, reliable readout method. The new QC device concept is shown in Fig. 1. It consists of a linear array of phosphorus nuclear spin qubits placed 35 nm apart and entangled through *weak* exchange interactions. The separation b between the linear QCs is much larger than a , the spacing within each QC, to prevent spin interactions between spins on different QCs. The use of duplicate QCs increases the number of photons emitted to ensure a signal above the spectrometer **signal-to-noise** ratio.

To simplify the spin interactions, initially each linear QC will consist of only 2 phosphorus atoms fabricated by STM methods, which consists of using the sharp STM tip as a precision lithographic tool to define the position of the atoms in the QC array in a hydrogen monolayer covering on an ultra-high vacuum prepared pristine Si(100)-2x1 reconstructed surface. Programmed voltage pulses between the STM tip and hydrogen resist layer remove hydrogen atoms from selected sites, leaving the rest of the surface protected by the remaining hydrogen covering.^{19,20} An example of STM patterning is shown in Fig. 2. The features in the STM constant current map are proportional to the local density of states (LDOS). The bright spots in the STM filled-states image, although actually the sites of missing hydrogen atoms, are due to the larger contribution that the silicon dangling bonds make to the DOS relative to the surrounding hydrogen atoms. After patterning, the surface is then exposed to molecules of a phosphorus bearing compound such as phosphine, PH_3 , that attach to the highly reactive phosphorus dangling bonds exposed on removal of the

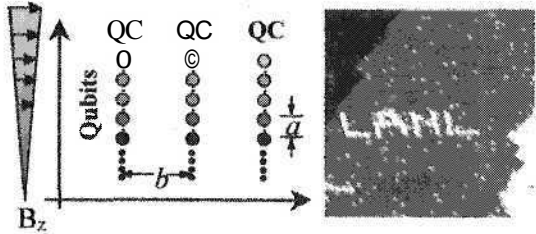


Fig. 1. Left: Layout of the SSQC device; each vertical line of qubits represents one copy of the QC; the readout will be a measurement of all copies. Right: Demonstration of STM patterning on silicon surface (the image is 800 Å across).

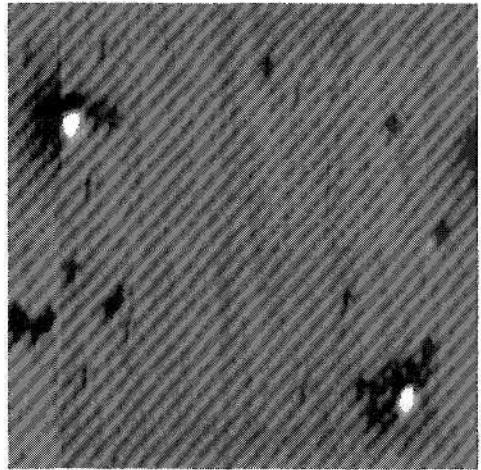


Fig. 2. Image showing STM patterned array of a pair of phosphorus dangling bonds after removal of two hydrogen from H-covered Si(100)-2x1 surface.

hydrogen atoms. The hydrogen is removed and phosphorus atoms exchange with surface silicon atoms during an annealing step. The surface is then buried under an isotopically pure ^{28}Si silicon epitaxial layer. The phosphorus dopants are activated upon encapsulation. The isotopically pure ^{28}Si silicon is necessary because the natural abundance of ^{29}Si , which has a spin $1/2$, in silicon is 4.7 %. If present, ^{29}Si would represent a natural source of decoherence. Finally, no barrier layer and no surface electrodes are needed in the scheme, which relies on optical methods for readout. The STM doubles as an electronic characterization probe of the local electronic environment of the qubits even after they are buried below the surface.¹³

A second requirement to implementing a quantum-computing device is to identify a means to address individual of qubits. In our spin-based scheme we use a 3 - 4 T external magnetic field, which splits the degeneracy of the spins, and a $10^4 - 10^6 \text{ Tm}^{-1}$ field gradient to shift the Larmor frequency of each qubit in the linear chain to a successively higher value, allowing individual spins to be addressed. Electron spins are fully polarized in this field and at low temperatures. In our case, the cryostat operates at temperatures below 1 K. The lone phosphorus electron spin is used to control the nuclear spin orientation via their hyperfine interaction. A combination of microwave and RF pulses devised by the modeling effort of our team are based on experimental parameters (Fig. 3). The spacing between adjacent phosphorus nuclei guarantees that the hyperfine interaction is much larger than the exchange interaction, which can be treated as a perturbation. Figure 4 shows a schematic of the electronic energy levels for a one nuclear spin - one electron spin in an external field including the electronic (dashed lines) and

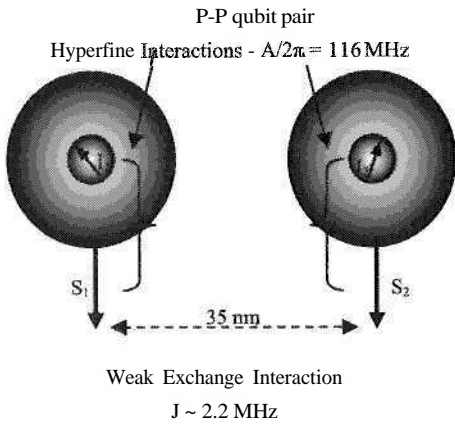


Fig. 3. Hyperfine and exchange interactions calculated for 35 nm spacing.

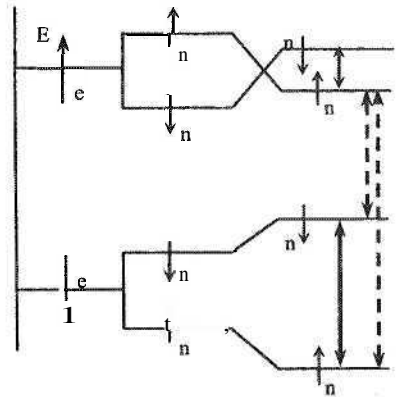


Fig. 4. Energy levels for one nuclear spin and one electron spin in an external magnetic field.

nuclear spin transitions (solid lines). The splitting in the diagram is exaggerated to show the transitions. In a two qubit system, to control the nuclear spins, for example, a microwave magnetic pulse is applied tuned to the resonance frequency of qubit one's electron, leaving qubit two's electron unaffected. RF pulses tuned to the transition of the

qubit one will flip its spin or not depending on the state of its electron. Initialization is accomplished by swapping the electron and nuclear spins then allowing the electron spins to relax back to their lowest energy state. Initially the electron spins are fully polarized in the magnetic field while the nuclear spins are in unknown states. The swap operation between electron 2 and nuclear spin 2 flips the nuclear spin into spin down state. Similarly, a swap operation on qubit 1 will rotate it into the down state. The swap operations are followed by a wait in which the electron spins relax back to their ground states. These operations result in the nuclear qubit spins initialized in the $|00\rangle$ state:

$$|\uparrow_{n2} \downarrow_{e2} \downarrow_{e1} \uparrow_{n1}\rangle \xrightarrow{\text{2 electron-nuclear swap-operations}} |\downarrow_{n2} \uparrow_{e2} \uparrow_{e1} \downarrow_{n1}\rangle \xrightarrow{\text{electron-relaxation}} |\downarrow_{n2} \downarrow_{e2} \downarrow_{e1} \downarrow_{n1}\rangle \equiv |00\rangle. \quad (3)$$

Similar procedures are used perform gate operations. In each case the state of the nuclear spin is controlled by operations using the state of its lone electron spin.

Readout is accomplished through optical spectroscopic methods. Excitons, generated in the substrate by a laser beam focused on the sample, diffuse and bind to the dopant sites where they act as probes of the spin states. Their recombination energies are sensitive to the local electronic spin states.^{14,15} The simplified diagram of the energy levels and allowed transitions shown in Fig. 5 modified from Ref. 15. For phosphorus, the highest intensity peak is for transition five,¹⁶ which can be monitored to determine the spin up/ spin down ratio. Significant improvement in peak sharpness can be obtained by using isotopically pure ²⁸Si.¹⁷

We have developed a new cryogenic based system for these experiments. A 7 T magnet, mm waveguide, and RF excitation coils are housed inside the new custom designed ultra-high vacuum ³He cryostat, which is equipped with an optical access for a laser beam and collection of the photons emitted during the exciton recombination, which are analyzed using a single element detector and photoluminescence spectrometer external to the cryostat.

5. Summary

We have developed a new concept for a QC device that will allow us to make simple quantum operations in a silicon-based solid-state system. To enable these demonstrations, we have developed a new low temperature experimental capability designed to perform demonstrations of simple quantum operations in the simplified prototype silicon-based spin device. As in the Kane scheme, STM is used to fabricate qubit spin arrays. Our method to achieve individual qubit addressability is to use a magnetic field and field gradient to shift the Larmor frequency of each qubit in each linear chain of phosphorus

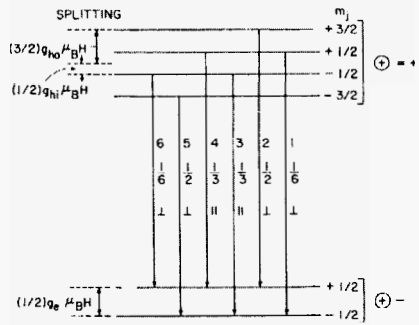


Fig. 5. Simplified energy diagram showing allowed exciton transitions (modified from Ref. 15).

atoms in the QC to successively higher value along the chain. We use duplicate QCs to gain signal intensity. The phosphorus nuclei and the lone electrons are manipulated using ESR and NMR techniques to carry out the quantum operations. Excitons, generated in the sample by a laser, are used to probe the state of the qubits. The intensity of a specific spectral line due to photons emitted during exciton recombinations will be used to determine the spin up-spin down ratios to readout the results of the quantum operations.

Acknowledgments

This work was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396.

References

1. B. E. Kane, *Nature* **393**, 133 (1998).
2. J. Chiaverini, J. Britton, D. Leibfried, E. Krill, M. D. Barrett, R. B. Blakestad, W. P. Itano, J. D. Jost, C. Langer, R. Ozeri, T. Schaetz, and D. J. Wineland, *Science* **308**(5724), 997 (2005).
3. G. Birkel, F. B. J. Buchkremer, R. Dumke, and W. Ertmer, *Optics Commun.* **191**(1-2), 67 (2001).
4. J. Wrachtrup and F. Jelezko, *J. Physics: Condensed Matter* **18**(21), S807 (2006).
5. I. Chiorescu, Y. Nakamura, C. J. P. M. Harmans, and J. E. Mooij, *Science* **299**(5614), 1869 (2003).
6. L. M. K. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chuang, *Nature* **414**, 883 (2001).
7. A. A. Clerk, S. M. Girvin, A. K. Nguyen, and A. D. Stone, *Phys. Rev. Lett.* **89**(17), 176804-1-4 (2002).
8. D. P. DiVincenzo, in *Mesoscopic Electron Transport*, eds. L. P. Sohn, G. Kowenhoven, L. L. Schoen, 657 (Kluwer Academic Publications, Dordrecht, 1997).
9. D. Rugar, R. Budakian, H. J. Mamin, and B. W. Chui, *Nature* **430**(6997), 329 (2004).
10. G. C. Abeln, T.-C. Shen, J. R. Tucker, and J. W. Lyding, *Microelectronic Eng.* **27**(1-4), 23 (1995).
11. T. Hitosugi, T. Hashizume, S. Heike, H. Kajiyama, Y. Wada, S. Watanabe, T. Hasegawa, and K. Kitazawa, *Appl. Surf. Sci.* **130-132**, 340 (1998).
12. R. J. N. Coope, T. Tiedje, S. L. Konsek, and T. P. Pearsall, *Ultramicroscopy* **68**, 257 (1997).
13. G. W. Brown, B. P. Uberuaga, H. Brube, and M. E. Hawley, *Phys. Rev. B* **72**, 195323 (2005).
14. A. S. Kaminskii, V. A. Karasyuk, and Ya. E. Polkrovskii, *Zhurnal Eksperimental'noi I Teoreticheskoi Fiziki* **79**(2), 422 (1980).
15. D. G. Thomas and M. Gershenson, *Phys. Rev.* **131**(6), 2397 (1963).
16. R. Sauer and J. Weber, *Phys. Rev. Lett.* **36**(1), 48 (1976).
17. D. Karaiskaj, M. L. W. Thewalt, T. Ruf, M. Cardona, H.-J. Pohl, G. G. Deviatykh, P. G. Sennikov, and H. Riemann, *Phys. Rev. Lett.* **86**(26), 6010 (2001).

This page intentionally left blank

QUANTUM ENGINEERING FOR THREAT REDUCTION AND HOMELAND SECURITY

GENNADY P. BERMAN

Theoretical Division and CNLS, Los Alamos National Laboratory, MS B213, Los Alamos, NM 87545, USA
gpb@lanl.gov

ALAN R. BISHOP

Theory, Simulation & Computation Directorate, Los Alamos National Laboratory,
MS B213, Los Alamos, NM 87545, USA
arb@lanl.gov

BORIS M. CHERNOBROD

Theoretical Division and CNLS, Los Alamos National Laboratory, MS B213, Los Alamos, NM 87545, USA
boris@lanl.gov

We review the results of our current research on quantum engineering which include the theory, modeling and simulations of quantum devices for potential applications to threat reduction and homeland security. In particular, we discuss: (i) scalable solid-state quantum computation with qubits based on (a) nuclear spins of impurity atoms in solids, (b) superconducting junctions, and (c) unpaired electron spins of spin radicals in self-assembled organic materials; (ii) quantum neural devices; (iii) quantum annealing; (iv) novel magnetic memory devices based on magnetic tunneling junctions with large tunneling magnetoresistance; (v) terahertz detectors based on microcantilever as a light pressure sensor; (vi) BEC based interferometers; (vii) quantum microscopes with a single-spin resolution based on (a) a magnetic resonant force microscopy and (b) an optically detected magnetic resonance; and (viii) novel approach for suppression of fluctuations in free space high-speed optical communication. Finally, we describe the similarities between the behavior of cross sections in reactions with heavy nuclei in the regions of strongly overlapped resonances and electron conductivity in semiconductor heterostructures.

Keywords: Quantum computer; spin; magnetic memory; cantilever; terahertz detector.

1. Introduction

Recent breakthroughs have led to a dramatic increase in our ability to control the quantum world; scientists can manipulate the microscopic quantum realm to an extent that would have been unimaginable two decades ago. Simultaneously, engineering developments are leading us into the world of quantum phenomena. The resulting inter-dependence of quantum science and engineering is giving birth to the discipline of “quantum

engineering". This recognizes that fundamental quantum phenomena are playing a crucial role at the frontiers of engineering. As examples: quantum calculations are directly impacting the design of the next generation of sensors and detectors; quantum-based protocols point the way to developing strategies for eavesdropping-proof communications and for ultra-efficient computation; and ultra-cold matter wave optics promises higher resolution detection schemes and microscopic imaging.

The far-reaching prospects offered by the emerging technology of quantum engineering has already been recognized in academia, where new centers that bridge physics, chemistry, mathematics and engineering departments have been created. The importance of quantum engineering for threat reduction and homeland security is being recognized by many funding agencies, including NSF, NSA, CIA, DARPA, NRO, ARO, DOE, and ONR. Their interests include applications such as secure communication techniques, high speed information processing, efficient energy conversion, and high resolution sensing and detection.

In this paper, we review the results of our research which connect fundamental quantum science with application-oriented engineering. We focus on the new opportunities created by breakthroughs in fundamental science driven applications in quantum information, spintronics, cold atom physics, quantum optics, nuclear physics, and condensed matter. These areas promise revolutionary advances in communication, computation, detection and sensing applications relevant to threat reduction and homeland security applications.

2. Dynamical Perturbation Theory for Scalable Solid-State Quantum Computation

One of the future quantum technologies is quantum computation.¹ Successful implementation of quantum computation requires combining significant efforts in fundamental and applied fields at the intersections of physics, chemistry, computer science, nano-technology, and material science. Crucial mathematical and physical problems must be solved in order to understand the dynamical aspects of quantum computation. One of these problems is the creation of a dynamical theory for scalable solid-state quantum computation – one of the main subjects of our research.²⁻⁷ The processes of the creation of quantum data bases, the storage and searching of quantum information, the implementation of quantum logic gates, and all the steps involved in quantum computation are dynamical processes. When qubits representing a register in a quantum computer are in superpositional states, they are not eigenstates of the Hamiltonian describing the quantum computer. These superpositions are time-dependent. Understanding and optimizing their dynamics is very important. To design a working quantum computer, modeling, simulations, optimizing and benchmarking the performance of quantum logic operations and fragments of quantum computation are essential. There are two main obstacles to simulating useful quantum logic operations with a large number of qubits on a digital (classical) computer: (1) the related Hilbert space is extremely large, $D_L = 2^L$, where L is the number of qubits (spins $\frac{1}{2}$), and (2) even if the initial state of a quantum computer does not involve many basic states (eigenstates), the number of excited eigenstates can rapidly grow during the process of performing quantum logic operations. Generally, both of these

obstacles exist. The expectation is that future quantum computers will overcome these (and others) obstacles. At the same time, many useful quantum operations can be modeled, simulated and benchmarked on a digital computer even if the number of qubits is quite large, say $L = 1000$. How can one do this? One way is to create (and use) a *perturbation theory of quantum computation*. To this end, one should consider a quantum computer as a many-particle quantum system and utilize small parameters. Usually, almost in *all* physical problems there exist small parameters which allow one to simplify the problem to find approximate solutions. The objective in this approach is to build a solution with bounded errors.

In Ref. 2 (see also references therein) we developed a perturbation theory for an L using spin scalable quantum computer. All spins are arranged in a 1D chain, and are placed in the permanent magnetic field $B_0^z(x)$, which has a gradient in the x -direction. In this case, all spins (qubits) have unique Larmor frequencies, ω_l . The Hamiltonian of the quantum register is:

$$H_0 = -\sum_{l=0}^{L-1} \omega_l S_l^z - 2J \sum_{l=0}^{L-2} S_l^z S_{l+1}^z, \tag{1}$$

where J is the constant of interaction between neighboring spins. The quantum protocol is implemented by a sequence of *rf*-pulses. The Hamiltonian of the n -th pulse is:

$$V^{(n)} = -\Theta^{(n)}(t)(\Omega^{(n)}/2) \sum_{l=0}^{L-1} \left(S_l^- e^{-i(\nu^{(n)l} + \varphi^{(n)})} + C.C. \right), \tag{2}$$

where $S_l^{z,\pm}$ are spin $1/2$ operators; the function $\Theta^{(n)}(t)$ equals 1 during the n -th pulse and 0 otherwise; $\Omega^{(n)}$, $\nu^{(n)}$, and $\varphi^{(n)}$ are the Rabi frequency, frequency, and the phase of the n -th pulse. In order to simulate quantum logic operations with many qubits, we developed a perturbation approach which allowed us to minimize, in a controlled way, systematic errors produced by non-resonant interactions of spins with *rf*-pulses. Our approach is based on two small parameters:

$$\varepsilon = \left[\frac{\Omega}{2J} \sin(\lambda \tau) \right]^2 \ll 1, \quad \mu \approx \left(\frac{\Omega}{\delta\omega} \right)^2 \ll 1, \tag{3}$$

which naturally appear in this quantum computer model. In Eq. (3), $\lambda^2 = \Omega^2 + 4J^2$, $\delta\omega$ is the difference of Larmor frequencies between neighboring qubits, and τ is the time of the pulse. Our perturbation approach allowed us to simulate, in a controlled way, many useful quantum gates and algorithms in a quantum computer with a large number of qubits, including: a quantum Controlled Not gate between remote qubits, quantum teleportation, quantum full adder, and others. In Ref. 5 we applied our approach to superconducting scalable quantum computers based on Josephson junctions. In Refs. 6 and 7 we considered a quantum computer based on self-assembled organic materials with electron spin radicals serving as qubits. In Ref. 8 we demonstrated the influence of collective decoherence on the performance of the Shor algorithm in a scalable spin based quantum computer.^{1,2}

3. Quantum Neural Devices

Nowadays, high speed pattern processing (or wide-band signals) is required by many applications. If we want to process patterns as a whole we need computers able to operate on registers containing thousands of bits or even more. There are at least two severe obstacles for the development of pattern processing computers: (i) the number of gates which should be manufactured on a chip grows exponentially with the size of the register; (ii) the code length grows exponentially with the size of the register. Both these obstacles led researches to the conclusion that it is necessary to find a new approach to the development of pattern processing computers, which must differ essentially from classical computers. Neurocomputers are one of the possible solutions which will allow us to overcome two mentioned difficulties of patterns processing. These computers will use new types of processing elements and also reject the algorithmic approach. Indeed, neurocomputers are built from nonlinear elements called neurons and use learning instead of programming. The power of a quantum neurocomputer can be demonstrated on a simple example of pattern associations. Suppose that a quantum register consists of $d + n$ qubits. Suppose also that the first d qubits are used to represent a binary weight of input connections of a single-pattern "classical" neuron, while the other n qubits are used to represent additional binary weights, whose observed values correspond to the components of an associated pattern. Using quantum entanglement and a superposition principle one can build the wave function of the register

$$|\Psi\rangle = 2^{-d/2} \sum_{i=0}^{2^d-1} |d_i\rangle |n_i\rangle, \quad (4)$$

which simultaneously combines the exponentially large number of all possible "classical" input patterns $|d_i\rangle$ (with binary weights, d_i) and their associated output patterns $|n_i\rangle$ (with binary weights, n_i). A signal which "measures" the state $|n_i\rangle$ will immediately collapse the wave function in Eq. (4) into the state $|d_i\rangle$. Discussions on quantum neural technologies can be found in Ref. 9 and references therein.

4. Quantum Annealing

Many physical and combinatorial problems associated with complex networks of interacting degrees of freedom can be formulated as a problem of determining the ground state or minimum cost state of the corresponding Ising (or Heisenberg) type spin Hamiltonian H_s with complicated frustrated interactions between spins (not necessarily the nearest-neighbor interactions). One of the approaches to solve this problem is quantum annealing which can be mathematically formulated as follows (different formulations are also possible¹⁰). Consider the time-dependent Hamiltonian:

$$H(t) = H_s \frac{t}{\tau} + H_{eff} \left(1 - \frac{t}{\tau}\right). \quad (5)$$

In Eq. (5), H_s is the Hamiltonian of the corresponding spin Ising (or Heisenberg) system; τ is a given time-interval of numerical simulations; time t varies from 0 to τ . Suppose that at $t = 0$ the ground state of the Hamiltonian H_{eff} can be easily calculated. Then, if to increase time t slowly enough, one can hope to approach the ground state of the Hamiltonian H_s at time $t = \tau$. As was demonstrated for Ising Hamiltonians, in many cases the process of quantum annealing is much faster than the corresponding process of classical annealing.¹⁰ Our preliminary results with Heisenberg Hamiltonians¹¹ demonstrate the advantages of a quantum annealing approach in finding the ground state of H_s . Quantum annealing can also be used for creation of exotic materials, including those required for quantum computation.

5. Magnetic Memory Devices Based on Magnetic Tunneling Junctions

At present, “ferromagnetic metal/MgO/ferromagnetic metal” magnetic tunnel junctions (MTJs) based on iron (Fe) and its compounds CoFe and CoFeB have demonstrated high values of tunneling magnetoresistance (TMR) at room temperature.^{12,13} Such MTJs are characterized by low potential barriers of the order of 0.34-0.50 eV. This is surprising, since large values of TMR are observed also for high potential barriers in MTJs based on Al₂O₃. Moreover, the TMR of low-barrier MTJs increases with an increase in barrier width. At the same time, the inverse picture is observed in Al₂O₃-based MTJs. The low-barrier MTJs have both a low resistance-area product and a high TMR. They are very favorable for developing magnetoresistive random-access-memory (MRAM).^{12,13} We demonstrated in Refs. 14 and 15 that both low and high barrier MTJs can be described by a two-band model of free electrons in ferromagnetic electrodes. Our theory works at finite values of the bias voltage V_a . Additionally, in Ref. 15 the influence of image forces on the TMR and on the electron current density in MTJs is taken into account. In particular, we show in Ref. 15 that the electron current density in low barrier MTJs is sufficient for switching the current-driven magnetization of the ferromagnetic electrodes.^{12,13} The results obtained in Ref. 15 considerably extend our knowledge about the physics of tunnel processes in magnetic nanostructures and can be used for developing MRAM with desired parameters.

6. Terahertz Detector Based on Microcantilever as a Light Pressure Sensor

Terahertz waves (1-25 THz) lie between the infrared and microwave regions of the electromagnetic spectrum. Recent increased interest to this spectral region is stimulated by applications of great importance for threat reduction and homeland security, including remote sensing of explosive materials, chemical and biological agents, security inspection and THz imaging. The sensitivity of uncooled detectors, which are available presently, is not enough for these applications. Recent progress in the performance of terahertz detectors is partly related to opto-mechanical uncooled detectors based on microcantilevers. These detectors exploit the thermomechanical effect related to the bending of a biomaterial microcantilever due to temperature changes. We propose to use the light pressure effect instead of the thermomechanical effect. We demonstrated recently^{16,17} that using the light

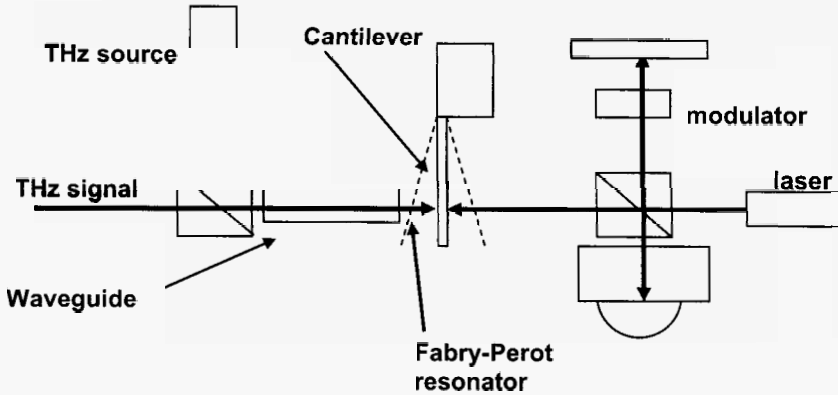


Fig. 1. Microcantilever based terahertz detector.

pressure microcantilever sensor as a detector significantly improves the sensitivity of the frequency modulation spectroscopy. To measure the signal from a thermal emitting object, we propose to use a local terahertz oscillator as a heterodyne source (see Fig. 1). The detector senses the spectral components of the signal shift relative to the heterodyne frequency by a value equal to the cantilever resonance frequency. The spectrum of the signal could be scanned by tuning the heterodyne frequency. To enhance the sensitivity of the detector, we use a Fabry-Perot optical resonator (see Fig. 1). Our theoretical analysis shows that the sensitivity is improved by, at least, three orders of magnitude as compared with conventional terahertz detectors, which are now available. An optical cavity scheme similar to one shown in Fig. 1 was realized for an efficient laser cooling of a microcantilever,^{18,19} which has a potential to obtain quantum mechanical ground state of the microcantilever. The utilization of quantum mechanical regime for terahertz detection could lead to unprecedented high sensitivity of a microcantilever based sensor.

7. Bose-Einstein Condensate Based Interferometers

Bose-Einstein condensates (BECs), with their coherence properties, have attracted wide interest for their possible application to ultraprecise interferometry and ultraweak force sensors. Since condensates, unlike photons, consist of interacting particles, they may permit the realization of specific quantum states needed as input to an interferometer to approach the Heisenberg limit, the supposed lower bound to precision phase measurements.²⁰ To this end, we study in Ref. 21 the sensitivity to external weak perturbations of a representative matter-wave Mach-Zehnder interferometer whose inputs are two Bose-Einstein condensates created by splitting a single condensate into two parts. The interferometric phase sensitivity depends on the specific quantum state created with the two condensates, and, therefore, on the time scale of the splitting process. We identify in Ref. 21 three different regimes, characterized by a phase sensitivity $\Delta\theta$ scaling with the total number of condensate

particles N as (i) the standard quantum limit $\Delta\theta \sim 1/N^{1/2}$, (ii) the sub shot-noise $\Delta\theta \sim 1/N^{3/4}$, and (iii) the Heisenberg limit $\Delta\theta \sim 1/N$. However, in a realistic dynamical BEC splitting, the $1/N$ limit requires a long adiabaticity time-scale, which is hardly practical experimentally. On the other hand, the sub-shot-noise sensitivity $\Delta\theta \sim 1/N^{3/4}$, can be reached in a realistic experimental setting. We also demonstrated²¹ that the $\Delta\theta \sim 1/N^{3/4}$ scaling is a rigorous upper bound in the limit $N \rightarrow \infty$, while keeping constant all parameters of the bosonic Mach-Zehnder interferometer.

8. Quantum Microscopes with a Single-Spin Resolution

In Ref. 22 (see also references therein) a theoretical approach was developed for a quantum microscope with a single-spin resolution based on magnetic resonance force microscopy (MRFM). This direction was initially developed in Ref. 23. The first experimental measurement of a signal from a single spin using MRFM was demonstrated in Ref. 24. In Refs. 25 and 26 we developed a novel approach for measurement of single electron and nuclear spin states based on optically detected magnetic resonance (ODMR) in a nano-probe located at the apex of an AFM tip. The method provides single electron spin sensitivity with nano-scale spatial resolution. The novel aspects of our approach are: (1) use of a nano-probe for the ODMR, (2) non-destruction of the spin state being measured, (3) nano-scale spatial resolution, and (4) high sensitivity of ODMR to a single electron or nuclear spin orientation relative to an external magnetic field. Single electron spin measurements in molecules were independently performed in 1993 by two groups led by Moerner *et al.*²⁷ and Orrit *et al.*²⁸ Today, the principles of detection of a single electron spin based on ODMR are well established. The limitation of the lateral resolution of ODMR is related to the size of the light spot. The highest resolution is obtained by a near-field scanning optical microscope, which has light spot size of about 30-50 nm. Another limitation of the ODMR technique is that the unpaired electron has to be a part of a molecule, which absorbs or emits light. Our modified ODMR approach consists of an AFM with a photoluminescent material (nanoparticle) located at the apex of the AFM tip (see Fig. 1 in Ref. 26), which exhibits ODMR in the vicinity of an unpaired electron or nuclear spin in the sample. This approach transfers the detection of electron magnetic resonance from a microwave frequency domain to an optical domain that significantly increases measurement sensitivity. In a semiconductor, quantum-dot confinement leads to a replacement of continuous bands of energy by molecular-like energy level structures. The theory of the spectrum of quantum dots shows that the ODMR spectrum depends on the mutual hole-electron interaction in the exciton. In the ODMR method, the nanoparticle is excited from its ground state 1 into the first excited state 2 by absorption from a laser field, and subsequently decays through non-radiative transitions to magnetic sublevel states 3 (see Fig. 2 in Ref. 26). An external magnetic field lifts the degeneracy of the triplet states and changes the energy splitting between these magnetic sublevels as well as their populations and lifetimes. Resonance is achieved by scanning the magnetic field from a nearby radio-frequency coil to induce transitions between these magnetic sublevels to change their relative populations, thus increasing or decreasing the intensity of specific peaks in the photoluminescence spectrum

resulting from decay back to the ground state. The proposed optical illumination scheme, known as an apertureless near-field scanning microscopy probe, leads to significant enhancement of the optical field in the vicinity of a sharp tip. The intensity enhancement near a silicon tip can be of several orders of magnitude. This enhancement arises from the antenna effect of the elongated probe. Thus nanoprobe absorption in an evanescent laser field could be significantly enhanced when the probe is placed at the apex of a sharp silicon tip. The potential resolution of this method is related to the size of the photoluminescent probe, typically 1 - 10 nm. There are two possible types of measurements. In the first type of measurement, the frequency of the oscillating magnetic field is fixed and the external permanent magnetic field is varied. In the second type of measurements, the permanent magnetic field is fixed and the frequency of the oscillating magnetic field is varied in the vicinity of the resonance frequency. One of the most promising applications for this method is the non-destructive measurement of a quantum bit (qubit) single spin state in a quantum computer (QC).

9. Suppression of Fluctuations in Free Space Optical Communication

Free space optical communications (FSOC) is well recognized as a solution for rapidly deploying wireless networks with gigabit/s data rates. One application, of great importance for threat reduction and homeland security, is disaster recovery after a terrorist attack or natural catastrophe. The effectiveness of FSOC was proven after the tragic events of September 11, 2001 which severely damaged segments of the Manhattan communications system. Service was quickly restored using FSOC as a temporary solution. Microwave systems cannot meet the bandwidth requirements for mobile adaptive networks capable of providing data, voice and video information. FSOC does provide the required bandwidth, but optical communication systems have limitations associated with atmospheric conditions such as smog, snowfall, and turbulence. In the case of heavy smog or snowfall the solution is the utilization of microwave techniques as a complementary solution. The temporary reduction in data rate is considered acceptable for many applications, and these binary systems are commercially available now as a solution of the "last mile bottleneck" problem. In the case of atmospheric turbulence the situation is more complex. Laser beams experience three major effects. First, the beam phase-front is distorted by fluctuations of the air refractive index, causing intensity fluctuations (scintillations). Second, eddies with sizes larger than the beam diameter randomly deflect the whole beam, a phenomenon called 'beam wandering'. Third, a laser beam propagating through the turbulent atmosphere spreads more than diffraction theory predicts, because for long distances the angular spreading due to refraction by atmospheric turbulence dominates. Of these three problems, scintillation is the most severe since it causes a significant increase in bit error rate (BER) and a resulting degradation in communications performance. For example, even for moderate turbulence, high quality laser communications, with $BER < 10^{-9}$, is limited to distances less than 2.5 km. However, laser beams in communication systems need to travel more than 10 km in the regime of strong turbulence. Another example is the quantum key distribution (QKD) system whose transmission distances are severely limited by

turbulence.²⁹ Several approaches have been developed to mitigate the effects of turbulence on FSO. They include the use of aperture averaging, phase diffusers, adaptive optics, and special data encoding (see the detailed review³⁰). However, none of these methods solves the problems described above, and a new approach is needed.

Our research^{31,32} is oriented to high data rate secure communications in the turbulent atmosphere for areas without network access, and two-way communications between earth stations and satellites. The proposed systems will increase bandwidth by orders of magnitude beyond microwave capabilities. In Ref. 31 we developed a quantum perturbation theory for free space optical communications based on path integrals. This approach allows us to obtain the scintillation index for the case of a partially coherent beam (PCB), including the limit of strong turbulence and long-distance propagation (up to 10km). Our approach uses the distribution function of photons (photon density in phase space), $f(r, q, t)$, where r and q are the coordinate and the momentum of a photon, to describe the beam characteristics at an arbitrary instant t . The first and second moments of the distribution function ($\langle f \rangle$ and $\langle ff \rangle$) are used to obtain the beam size and intensity fluctuations, respectively. Our approach can be used for both stationary beams as well as for beams with varying intensity. In Ref. 32 this approach is applied in combination with spectral encoding which allows a high-speed free space optical communication.

10. Quantum Engineering with Heavy Nuclei: Connection with Electron Transport in Semiconductor Heterostructures

In Refs. 33-36 a new approach was developed for studying collective phenomena both in nuclear physics and in condensed matter physics. For mesoscopic solid state systems this approach allows one to describe simultaneously the intrinsic structure of the system and the interaction with the environment. In nuclear physics, this approach takes into account *both* strong interactions between nucleons and coupling to the continuum. In this problem, the standard shell model techniques, adjusted for the bound states, cannot work, and continuum effects must be properly accounted for. At some certain strength of the continuum coupling the system of strongly interacting nucleons undergoes a restructuring that reveals new collective phenomena. This is an area in which the conventional division of nuclear physics into “structure” and “reactions” becomes inappropriate and two views of the process, from the inside (structure and properties of bound states) and from the outside (cross sections of reactions), have to be unified. The effective Hamiltonian of a compound nucleus with open decay channels has the form:³³⁻³⁶ $H(E) = H_0 + V(E) - iW(E)/2$. Here H_0 stands for the mean-field Hamiltonian of N particles (quasi-particles) occupying single-particle levels of the standard nuclear shell model. The term $V(E)$ describes a strong residual interaction between the particles; its energy dependence is due to the virtual continuum coupling that renormalizes the conventional shell model interaction. In contrast to the standard shell model, the coupling to open continuum channels is taken into account explicitly by the term $W(E)$. This term contains the amplitudes for the transition from the intrinsic many-body state to the open continuum channel. The goal of our investigations is to reveal how statistical properties of cross sections depend on the strength of the inter-particle interaction

and on the coupling to the continuum. One of the effects studied within this approach is the so-called *width redistribution* (WR) that occurs when the unstable states overlap their widths. Specifically, for many unstable states, $N \gg 1$, when the coupling, $W(E)$, exceeds a critical value, the WR of the decay widths results in the formation of M short-lived (*wide* or *super-radiant*) states and $N-M$ relatively stable (*narrow* or *trapped*) states. The universality of this effect has been demonstrated in numerous applications in nuclear and atomic physics, chemical reactions and condensed matter on the mesoscopic scale.

Acknowledgments

We are thankful to M. G. Boshier, G. W. Brown, G. L. Celardo, D. A. Dalvit, M. E. Hawley, F. M. Izrailev, D. H. Sharp, D. L. Smith, E. M. Timmermans, V. G. Zelevinsky, and W. H. Zurek for useful discussions. This work was carried out under the auspices of the National Nuclear Security Administration of the U. S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396. Our research was partly supported by the Laboratory Directed Research & Development (LDRD) Program at LANL, by the Defense Advanced Research Projects Agency (DARPA), by the National Security Agency (NSA) and Advanced Research and Development Activity (ARDA), and by the Army Research Office (ARO).

References

1. G. P. Berman, G. D. Doolen, R. Mainieri, and V. I. Tsifrinovich, *Introduction to Quantum Computers* (World Scientific Publishing Company, 1998).
2. G. P. Berman, D. I. Kamenev, and V. I. Tsifrinovich, *Quantum Computation with Many Quantum Bits* (Rinton Press, 2005).
3. G. P. Berman, G. W. Brown, M. E. Hawley, D. I. Kamenev, and V. I. Tsifrinovich, Implementation of quantum logic operations and creation of entanglement in a silicon-based quantum computer with constant interaction, *quant-ph/0512174*.
4. G. P. Berman, D. I. Kamenev, and V. I. Tsifrinovich, Creation of entanglement in a scalable spin quantum computer with long-range dipole-dipole interaction between qubits, *quant-ph/0603227*.
5. G. P. Berman, A. R. Bishop, D. I. Kamenev, and A. Trombettoni, Quantum logic operations and creation of entanglement in a scalable superconducting quantum computer with long-range constant interaction between qubits, *Phys. Rev. B* **71**(1), 014523-1-11 (2005).
6. G. P. Berman, D. L. Allara, and V. I. Tsifrinovich, Entangled spin states in self-assembled monolayer systems, *Phys. Rev. B* **66**(19), 193406-1-3 (2002).
7. Z. Rinkevicius, G. P. Berman, D. L. Allara, and V. I. Tsifrinovich, S. Tretiak, Characteristic parameters and dynamics of two-qubit system in self-assembled monolayers, *quant-ph/0411202*.
8. G. P. Berman, D. I. Kamenev, and V. I. Tsifrinovich, Collective decoherence of the superpositional entangled states in the quantum Shor algorithm, *Phys. Rev. B* **71**(3), 32346-1-5 (2005).
9. A. A. Ezhov and G. P. Berman, *Introduction to Quantum Neural Technologies* (Rinton Press, 2003).
10. A. Das and B.K. Chakrabarti (Eds), *Quantum Annealing and Related Optimization Methods* (Springer-Verlag, 2005).

11. G. P. Berman, V. N. Gorshkov, and V. I. Tsifrinovich, Quantum annealing with Heisenberg systems, in preparation.
12. S. Parkin, Spin-polarized current in spin valves and magnetic tunnel junctions, *MRS Bulletin* **31**, 389-94 (2006).
13. S. A. Wolf, D. Treger, and A. Chtchelkanova, Spintronics: The Future of Data Storage? *MRS Bulletin* **31**, 400-3 (2006).
14. N. N. Beletskii, G. P. Berman, A. R. Bishop, S. A. Borysenko, and V. M. Yakovenko, Magnetoresistance and spin polarization of electron current in magnetic tunneling junctions, *LANL Preprint*, LA-UR-06-5008.
15. N. N. Beletskii, G. P. Berman, S. A. Borysenko, S. A. Wolf, and V. M. Yakovenko, Magnetoresistance of magnetic tunnel junctions without low barrier height, *LANL Preprint*, LAUR-06-7204.
16. B. M. Chernobrod, G. P. Berman, and P. W. Milonni, Improving the sensitivity of frequency modulation spectroscopy using nanomechanical cantilever, *Appl. Phys. Lett.* **85**, 3896-8 (2004).
17. G. P. Berman, A. R. Bishop, B. M. Chernobrod, M. E. Hawley, and G. W. Brown, Sensitivity analysis of a proposed novel opto-nano-mechanical photodetector for improving the performance of LIDAR and local sensor, *J. Phys., Conference Series* **38**, 171-3 (2006).
18. C. H. Metzger and K. Karrai, Cavity cooling of a microlever, *Nature* **432**, 1002-5 (2004).
19. S. Gigan, *et al.*, Self cooling of a micro-mirror by radiation pressure, <http://arxiv.org/abs/quant-ph/0607068>.
20. V. Giovannetti, S. Lloyd, and L. Maccone, Quantum-enhanced measurements: Beating the standard quantum limit, *Science* **306**, 1330-6 (2004).
21. L. Pezzé, L. A. Collins, A. Smerzi, G. P. Berman, and A. R. Bishop, Sub-shot-noise phase sensitivity with a Bose-Einstein condensate Mach-Zehnder interferometer, *Phys. Rev. A* **72**, 043612-8 (2005).
22. G. P. Berman, V. N. Gorshkov, F. Borgonovi, and V. I. Tsifrinovich, *Magnetic Resonance Force Microscopy and a Single-Spin Measurement* (World Scientific Publishing Company, 2006).
23. J. A. Sidles, Noninductive detection of single-proton magnetic resonance, *Appl. Phys. Lett.* **58**, 2854-6 (1991).
24. D. Rugar, R. Budakian, H. J. Mamin, B. W. Chui, Single spin detection by magnetic resonance force microscopy, *Nature* **430**(6997), 329-32 (2004).
25. B. M. Chernobrod and G. P. Berman, Spin microscope based on optically detected magnetic resonance, *J. App. Phys.* **97**(1), 014903-1-3 (2005).
26. G. P. Berman, A. R. Bishop, B. M. Chernobrod, M. E. Hawley, G. W. Brown, and V. I. Tsifrinovich, Measurement of single electron and nuclear spin states based on optically detected magnetic resonance, *J. Phys.: Conference Series* **38**, 167-170 (2006).
27. J. Köhler, J. A. J. M. Disselhorst, M. C. J. M. Donckers, E. J. J. Groenen, J. Schmidt and W. E. Moerner, Magnetic resonance of a single molecular spin, *Nature* **363**, 242-4 (1993).
28. J. Wrachtrup, N. Borczykowski, J. Bernard, M. Orrit, and R. Brown, Optical detection of magnetic resonance in a single molecule, *Nature* **363**, 244-5 (1993).
29. P. W. Milonni, J. H. Carter, C. G. Peterson, and R. J. Hughes, Effects of propagation through atmospheric turbulence on photon statistics, *J. Opt. B: Quantum and Semiclass. Opt* **6**, S742-5 (2004).
30. L. C. Andrews, R. L. Phillips, C. Y. Young, *Laser Beam Propagation in the Random Media* (SPIE – The International Society for Optical Engineering, 2001).
31. G. P. Berman and A. A. Chumak, Photon distribution function for long-distance propagation of partially coherent beams through the turbulent atmosphere, *Phys. Rev. A* **74**, 13805-11 (2006).

32. G. P. Berman, A. R. Bishop, B. M. Chernobrod, D. C. Nguyen, and V. N. Gorshkov, Suppression of intensity fluctuations in free space high-speed optical communication based on spectral encoding of a partially coherent beam, *LANL Preprint*, LA-UR-06-4251.
33. A. Volya and V. Zelevinsky, Exploring quantum dynamics in an open many-body system: transition to superradiance, *J. of Optics B: Quantum and Semiclassical Optics* **5**(3), S450-S456 (2003).
34. A. Volya and V. Zelevinsky, Super-radiance and open quantum systems, Nuclei and Mesoscopic Physics, *Workshop on Nuclei and Mesoscopic Physics (WNMP 2004)*, East Lansing, MI, USA (2004); *AIP Conf. Proc.* **777**, 229-249 (2005).
35. G. P. Berman, G. L. Celardo, F. M. Izrailev, and V. G. Zelevinsky, Novel approach for fluctuating cross sections in heavy nuclei in the region of strongly overlapped resonances, *LANL Preprint*, LA-UR-06-4067.
36. G. L. Celardo, F. M. Izrailev, V. G. Zelevinsky, and G. P. Berman, Resonance widths and spacings at the transition from isolated to overlapping resonances, (in preparation).

STRONG PHASE SHIFT MASK MANUFACTURING ERROR IMPACT ON THE 65 NM POLY LINE PRINTABILITY

NADYA BELOVA*

LSI Logic Corporation, 1621 Barber Lane, Milpitas, CA 95035, USA

A methodology and Monte Carlo simulation flow with the integrated LSI Logic package, Molotof, was applied to the 65 nm poly line sensitivity analysis. Strong phase shift mask (sPSM) manufacturing process specification was optimized to obtain image critical dimensions (CDs) and image placement errors (IPes) complying with technology process tolerances. Reticle manufacturing statistical variations of phase depth, phase width, and phase intensity imbalance were used to generate a virtual sPSM for imaging poly lines on a wafer. A criterion for qualifying reticle specification is to obtain all latent image CDs and IPes within a photo process module allowed range for a given mask specification. The approach allows for computing reticle and litho budgets into CD imaging performance. Sensitivity to a single parameter variation and full statistical analysis of the 65 nm poly line imaging performance affected by reticle manufacturing errors is presented. The optimum reticle specification, yielded 100% of critical dimensions and image placement errors, was found in simulation and confirmed by empirical data.

Keywords: Strong PSM; critical dimension; image placement error; Monte Carlo simulation

1. Introduction and Motivation

Manufacturing a strong phase shift mask (sPSM) for printing poly gates down to 65 nm and below requires a new specification for qualifying reticle performance. Three parameters, phase depth (PD), phase width (PW) and phase intensity imbalance (PI) in addition to the regular mask performance control parameters (reticle CD, registration, defect sizes) are taken into consideration for the alternating phase shift reticle qualification. In Fig. 1 the final sPSM is schematically shown. Figure 1(a) presents phase depth created by etching trenches into quartz. Figure 1(b) shows clear opening width of a zero-phase (PW₀) and a 180°-phase (PW₁₈₀) at the mask top view. Figure 1(c) illustrates the relative intensity profile of the light passing through the mask. The sPSM topography results in the intensity imbalance [Fig. 1(c)] due to light scattering at sidewalls and edges of the 180°-phase openings.¹ Rigorous 3-D electromagnetic simulation of the sPSM shows up to 25% of the relative intensity loss between zero- and 180°-phases.^{2,3} By assigning different transmission levels to the zero-phase (equal to one)

* Currently with ON Semiconductor Corporation
23400 NE Glisan St., Gresham, OR 97030, USA Phone (503) 618-5590 e-mail:
Nadya.Strelkova@onsemi.com

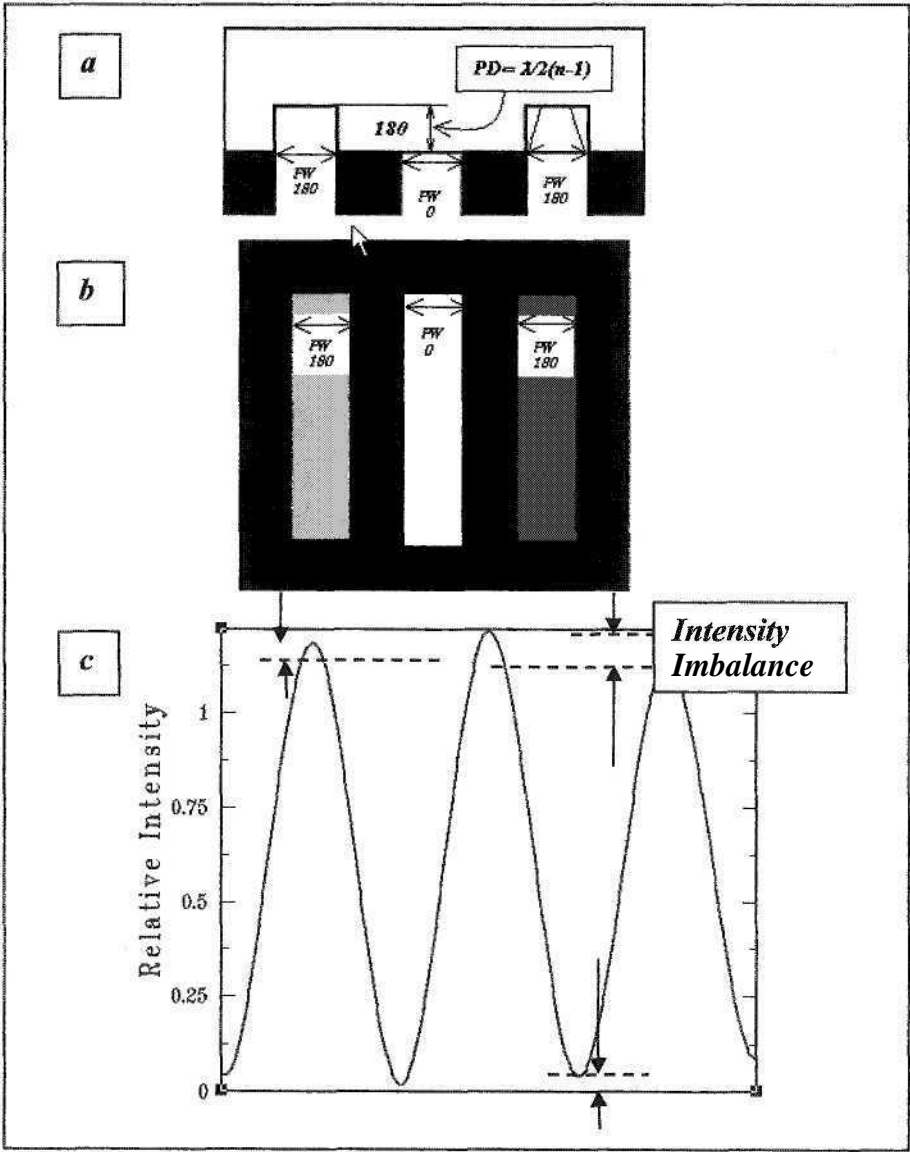


Fig.1. Strong phase shift mask manufactured by a single etch quartz process (a). The phase depth (PD) and phase width (PW) parameters are illustrated in (a), (b). Different coloring of zero- and 180° -phase openings on the reticle top view represents different transmission levels translated into an unbalanced intensity profile (c).

and to the 180° -phase (equal to π) the sPSM intensity imbalance can be simulated. This assumption is presented in Fig. 1(b) by different shades of the zero- and 180° -phases.

The depth of etched quartz trenches of an ideal alternating phase shift mask is equal to $\lambda/2(n-1) = 180^\circ$, where λ is the wavelength of an illumination source, and $n(\lambda)$ is the refractive index. The widths of the zero- and 180° -openings, in general, not equal to each

other ($PW_0 \neq PW_{180}$) are adjusted to eliminate an intensity imbalance shown in Fig. 1(c).

Systematic mask manufacturing errors, presented by mean values of $\langle PD \rangle$, $\langle PW_0 \rangle$, $\langle PW_{180} \rangle$, $\langle PI \rangle$, and statistical variations (ΔPD , ΔPW_0 , ΔPW_{180} , ΔPI) translate into critical dimension and image placement errors on wafer printed images. The overall impact of the sPSM manufacturing errors on imaging performance can be measured by collecting and analyzing CD/IFE statistics of printed images.

The motivation for this research was (1) to justify the strong phase shift reticle specification and associated mask making cost and (2) to understand which parameter of sPSM specification can be relaxed without significant impact on imaging performance.

In Sec. 2 the method and metrology for quantitative characterization of the sPSM error impact on line printability is discussed. Section 3 presents results of applying the methodology for sensitivity and Monte Carlo analysis of the 65 nm poly line printability. The article is concluded by a summary.

2. Method and Analysis Flow

The input variables used for simulating the impact of mask manufacturing errors on the poly line printability are presented in Fig. 2. The ideal sPSM gds mask is presented by mask cross-section in Fig. 3(a). Zero-phase opening is designed on a zero-phase gds layer with parameters $PD = 0$, $PW_0 = \text{const}$, $PI = 1$. The 180°-phase opening is designed on the 180°-phase gds layer with parameters $PD = \langle PD \rangle$, $PW_{180} = \text{const}$, $PI = \langle PI \rangle$. The mask manufacturing errors are reproduced by distribution functions for variables (PD , PW_0 , PW_{180} , PI) with correspondent ΔPD , ΔPW_0 , ΔPW_{180} , ΔPI . The normal distributions are assumed in this work for all mask parameters in consideration. Randomly generated 180°-phase opening parameters are used to transform an ideal 180°-phase gds mask into a virtual statistical mask [Fig. 2(b)], that simulates the systematic and statistical reticle fabrication process errors.

For quantitative characterization of printability, two response functions of mask parameters are calculated from an intensity distribution profile obtained from an aerial image: $CD(PD, PW_0, PW_{180}, PI)$ and $IFE(PD, PW_0, PW_{180}, PI)$ [Fig. 3(c)]. Dark line CD is measured as a distance between two coordinates at a specific intensity threshold representing an exposure dose in a lithography process. IFE is calculated as a displacement of a printed line center from the ideal mask geometrical center.

The self-explanatory Fig. 3 shows the simulation flow for studying the impact of the mask errors on the CD and IFE responses. The flow includes steps of defining input parameter distributions; importing input variables into design layer for generating a virtual statistical gds mask; simulating an aerial image of a gds mask; calculating CD and IFE from an intensity distribution profile extracted from the aerial image; and, finally, calculating response function CD/IFE statistics.

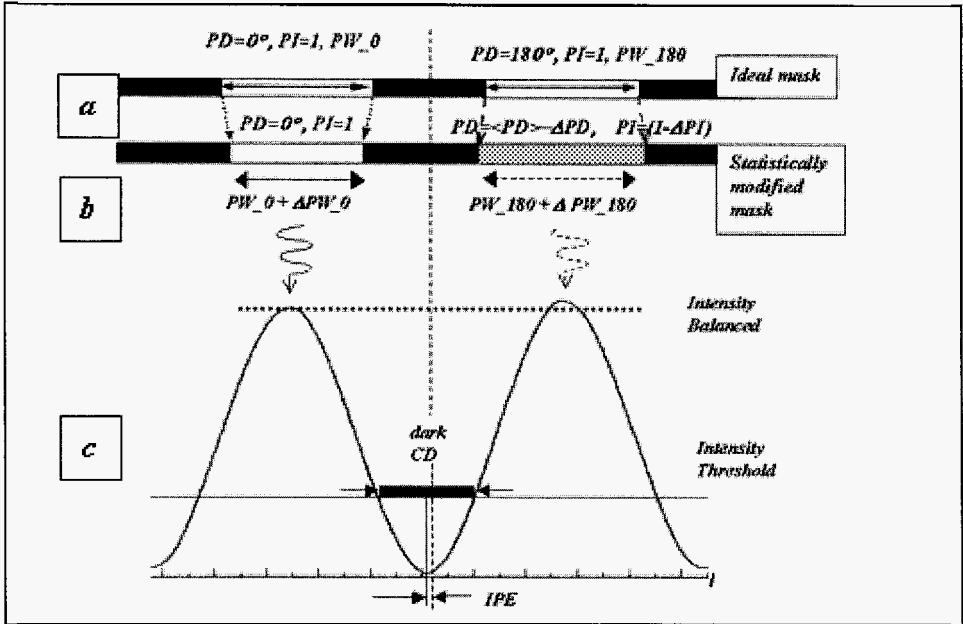


Fig. 2. Phase depth (PD), zero- and 180° -phase opening widths (PW_0, PW_{180}) and phase imbalance (PI) errors modify an ideal mask (a) into a virtual random mask (b). Critical dimension (CD) and image placement error (IPE) metrology is shown in (c).

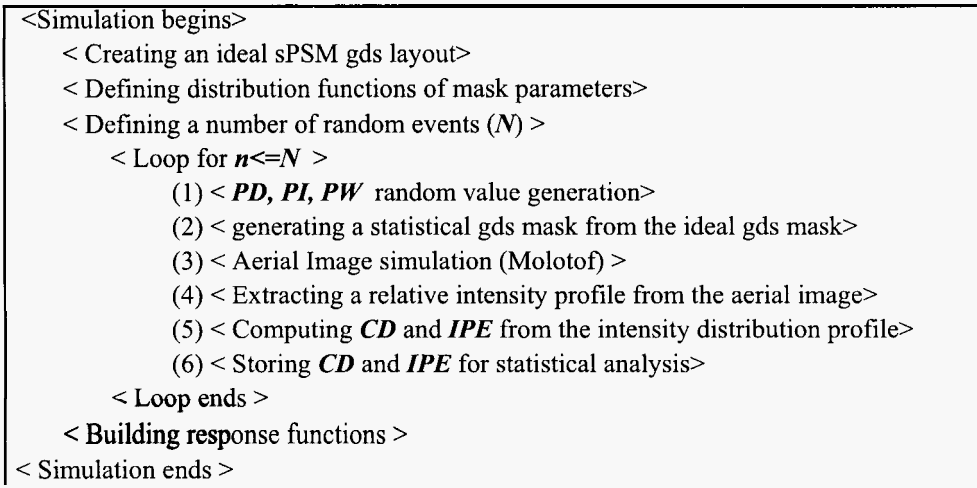


Fig. 3. Simulation flow to perform mask imaging simulation and analysis of response functions to the mask manufacturing process variations. The key module of the flow is the Molotof image simulator.

The central place of the flow belongs to Molotof image simulator developed at LSI Logic. Molotof image simulator easily integrates into any Resolution Enhancement Technique (RET) flow. Molotof has better CPU-time performance than any other commercially available software. The CPU-time per one simulation loop of mask imaging is critical to perform Monte Carlo statistical analysis.

3. Results and Discussion

This section presents simulation results of the 65 nm poly line printability affected by sPSM manufacturing process variations. The dense pattern of straight lines with the constant width = 65 nm in the 160 nm and 180 nm pitches was used for simulation (see Fig. 4).

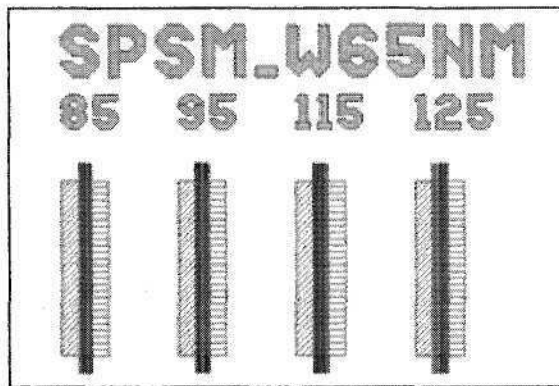


Fig. 4. Strong PSM layout of dense lines with constant width equals to 65 nm and different spacing ($s = 85$ nm, 95 nm, 115 nm, etc.) used for printability statistical analysis. Different shades on the left and the right sides of the pattern represent coloring of the zero- and 180°-phases.

The simulation was carried out using 193 nm exposure wavelength. A circular illumination source with numerical aperture $NA = 0.7$ and the partial coherence factor $\sigma_p = 0.35$ was used for mask exposure.

An intensity threshold (dose) for *CDIPE* measurements was calibrated to print $CD = 65$ nm in the target pitch ($p = w + s = 160$ nm) of an ideal sPSM with NO intensity imbalance between clear openings. An aerial image model was applied for the mask simulation. Single exposure of sPSM was simulated. No trim mask was taken into consideration. It was assumed, an input mask had no imbalance between zero-phase and 180°-phase, i.e. width openings equal to $\langle PW_0 \rangle = \langle PW_{180} \rangle$, and transmission factor equal to $\langle PI \rangle = 1$. The 180°-phase systematic depth error ($\langle PD \rangle$), and random errors (ΔPD , ΔPW_0 , ΔPW_{180} , ΔPI) with normal distributions were taken into consideration. Each statistical experiment simulated 250 random masks, each virtual mask affected by four variables.

Simulation and empirical results are presented in Tables 1 and 2 and in Figs. 5-7. Tables 1 and 2 show simulated statistics for the 65 nm line in the 160 nm pitch. Table 1

presents simulated *CD* and *IPE* mean values, standard deviations, range, and combined *CD/IPE* yield in response to different reticle *CD* control specifications. Table 2 graphically illustrates results from the Table 1. Table 2 shows *CD* distribution functions and *CD* vs. *IPE* scattering plots used for *CD/IPE* yield calculation. Figure 5 presents empirical and simulated *CD* distributions for the 65 nm line in the 160 nm pitch. Figure 6 shows wafer *CD* distributions and *CD* vs. *IPE* scattering plot in different pitches (160 nm and 180 nm). Figure 7 illustrates strong phase shift reticle and photolithography process budgets in the line printability performance.

Sensitivity simulation of the poly line *CD* and *IPE* in the 160 nm pitch to a single parameter variation shows very little impact on *CD* and no impact on *IPE* of the phase depth errors. The 20% range in *PI* translates into approximately ~4.2 nm of *CD* and ~ 4 nm of *IPE* range. The 180°-phase opening width variation within (+/- 10 nm/per side) results in approximately 10.5 nm range of *CD* and 10 nm of *IPE*. The conclusion from the sensitivity analysis is that *CD/IPE* performance at the best focus is dramatically affected by the mask *CD* variations. The second factor is the transmission loss at 180°-phase openings, and the third factor is the phase depth.

Table 1. Input parameter statistics and *CD/IPE* response for the 65 nm line in the 160 nm pitch. The *CD/IPE* yield is calculated as a ratio of the number of events resulting in responses violating *CD* = (65 +/- 5) nm and *IPE* = +/-5 nm tolerance, to the total number of simulated Monte Carlo samplings.

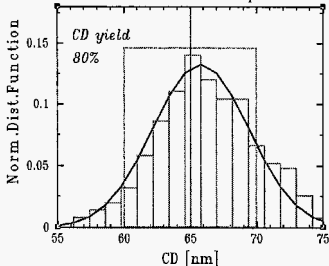
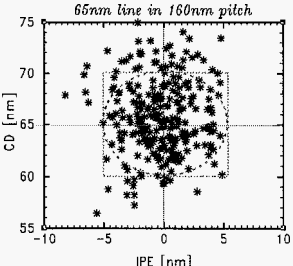
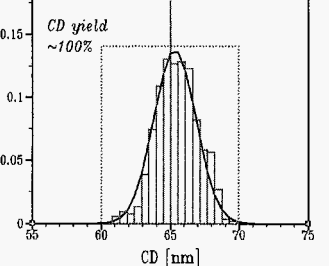
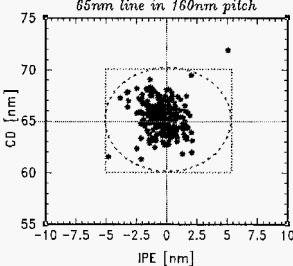
#	PD [degrees]		PI [au]	ΔPW [nm] / per side	CD [nm]			IPE [nm]		CD/IPE yield for CD=(65 +/- 5) nm IPE= +/- 5nm
	mean	3 σ	3 σ	3 σ	mean	range	3 σ	range	3 σ	
1	179	5°	10%	9	65.8	17.5	10.8	13	7.8	78%
2	179	1°	1%	9	65.5	14	8.7	12	6.7	89%
3	179	1°	1%	6	65.4	11	6	10.1	4.6	97%
4	179	1°	1%	4.5	65.3	10.6	4.6	9.8	3.6	100%
5	179	5°	5%	5	65.4	11	6.3	6.8	4.3	94%

Statistical analysis of poly line printability due to mask fabrication errors with $\langle PD \rangle = 179^\circ$, $3\sigma PD = 5^\circ$, transmission error $3\sigma PI = 10\%$, and $3\sigma PW = 9$ nm/per side gave 17.5 nm *CD* range, and 13 nm *IPE* range. The combined *CD/IPE* yield was ~ 78% for the *CD/IPE* range +/-5 nm (Tables 1 and 2, row 1). *CD*-yield is calculated as an integral of a normalized distribution function limited by +/-5 nm range of the nominal 65nm *CD* presented by dotted boxes in Table 2, column 2. The *CD/IPE* combined yield is calculated as a ratio of the number of events falling out of the +/- 5 nm range shown by dotted boxes in figures of Table 2, to the total number of Monte Carlo samplings. The *CD* yield is slightly higher than the combined *CD/IPE* yield, (compare columns 2 and 3 of the Table 2). Yield definitions should be determined prior to yield estimation:

The yield computed within a *CD/IPE* ellipse is lower than the yield calculated within a *CD/IPE* rectangle as can be seen in figures of Table 2, column 3.

Several reticle specifications were simulated, see Table 1. The reticle specification was optimized to obtain 100% *CD/IPE* yield. The reticle with specification #4 with $3\sigma PW = 4.5$ nm/per side, $3\sigma PD = 1^\circ$ and $3\sigma P = 1\%$ will also provide good yield margins presented in Table 2.

Table 2. Normalized distribution functions and *CD* versus *IPE* scattering plots used for the *CD*-yield and combined *CD/IPE* yield calculation illustrate statistical simulation results presented in Table 1.

Variables presented in Table 1	Normalized <i>CD</i> distribution functions. <i>CD</i> yield is calculated as an integral of a distribution function within $CD = (65\pm 5)$ nm shown by a dotted box.	<i>CD</i> . vs. <i>IPE</i> scattering plots. <i>CD/IPE</i> yield presented in Table 2 is calculated as a ratio of the failing events (outside a dotted box) to the total number of samplings.
<p>Spec #1:</p> <p>$\langle PD \rangle = 179^\circ$ $3\sigma PD = 5^\circ$</p> <p>$3\sigma PI = 10\%$</p> <p>$3\sigma PW_0 = 3\sigma PW_{180} = 9$ nm / per side</p>	<p><i>CD</i>-yield ~ 80%</p> 	<p><i>CD/IPE</i>-yield ~ 78%</p> 
<p>Spec #4</p> <p>$\langle PD \rangle = 179^\circ$ $3\sigma PD = 1^\circ$</p> <p>$3\sigma PI = 1\%$</p> <p>$3\sigma PW_0 = 3\sigma PW_{180} = 4.5$ nm / per side</p>	<p><i>CD</i>-yield ~ 100%</p> 	<p><i>CD/IPE</i>-yield ~ 99.6%</p> 

The realistic reticle may not produce 100% wafer *CD/IPE* yield within the photo process *CD* tolerance. The calibration of reticle control parameters and illumination conditions have to be done before doing the imaging simulation. It was found that the match between simulated and empirical *CD*s is achieved with the reticle specification fabricated with the specification #5 from the Table 1 ($3\sigma PW = 5$ nm/per side, $3\sigma PD = 5^\circ$ and $3\sigma PI = 5\%$). Figure 5 compares the cumulative probabilities of the wafer *CD* distribution (circles) with simulated *CD* distributions (curves) obtained with two virtual

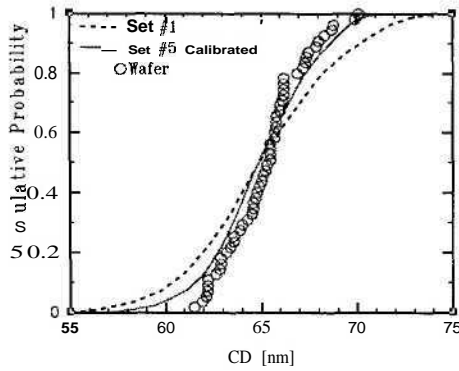


Fig .5. Comparison of wafer *CD* and simulated *CD* cumulative probabilities in the 160 nm pitch. Dashed curve presents the non-optimized reticle specification #1 from the Table 1. Solid curve presents the simulated *CD* distribution with the calibrated mask specification #5: $3\sigma PW = 5$ nm/per side, $3\sigma PD = 5^\circ$ and $3\sigma PI = 5\%$. Wafer *CD* statistics presented by 55 SEM measurements.

reticle specifications (#1 and #5 from the Table 1). Simulation done with calibrated reticle specification #5 shows a very good agreement with the empirical distribution. In the 160 nm pitch the empirical mean $\langle CD \rangle = 66.8$ nm, standard deviation $3\sigma = 6.15$ nm, with *CD* range equal to 8.6 nm, while simulated mean $\langle CD \rangle = 65.5$ nm, standard deviation $3\sigma = 7.8$ nm, with *CD* range equal to 13 nm. In the 180 nm pitch the empirical mean $\langle CD \rangle = 64.0$ nm, standard deviation $3\sigma = 4.8$ nm, with *CD* range equal to 8.3 nm, while simulated mean $\langle CD \rangle = 65.4$ nm, standard deviation $3\sigma = 4.4$ nm with *CD* range of 7.6 nm.

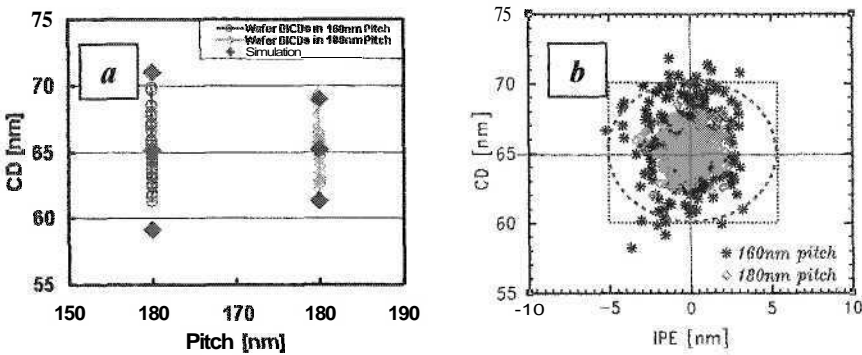


Fig. 6. Empirical *CD* distributions vs. simulated *CD* statistics in the 160 nm and 180 nm pitches (a). The *CDIPE* distributions used for estimating *CDIPE* combined yield (b): *CDIPE* yields 94% in the 160 nm pitch and *CDIPE* yields 100% with wide margins in the 180 nm pitch.

In the larger pitches the wafer *CD* range decreases. Figure 6(a) shows the empirical *CD*s compared to simulated *CD* distributions for 65 nm line in the 160 nm and 180 nm pitches. The empirical *CD* distributions in Fig. 6(a) fit inside of simulated $CD = (65 \pm 3\sigma)$ nm range shown by solid diamonds.

The advantage of simulation is that with calibrated reticle and imaging parameters, the image placement error can be computed. Figure 6(b) shows the $CD/1PE$ distributions used for estimating $CD/1PE$ yield in the 160 nm (stars) and 180 nm (diamonds) pitches: the $CD/1PE$ margins in the 180 nm pitch are at least twice better than $CD/1PE$ performance in the 160 nm pitch. Since Molotof simulator predicted CD distributions with a good accuracy, the simulated image placement error and $CD/1PE$ combined yield are most likely also estimated accurately.

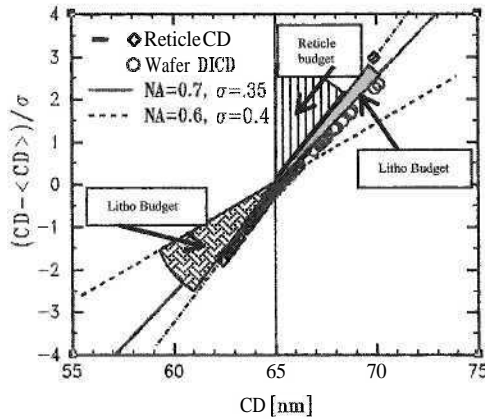


Fig. 7. Segregating reticle and litho process contributions into CD imaging budget. Image simulation was in 160 nm pitch at two different illumination conditions ($0.6NA/0.4\sigma_p$ and $0.7NA/0.35\sigma_p$) with the calibrated reticle specification $3\sigma PW = 5$ nm/per side, $3\sigma PD = 5^\circ$ and $3\sigma PI = 5\%$. Litho budget increases under not-optimized illumination conditions.

The simulation also allows segregating reticle and litho process budgets contributing into overall CD imaging performance presented in Fig. 7. Reticle CD normalized deviations, presented by 72 reticle CD s (diamonds) done at a mask shop and virtual mask gds CD s (dot-dashed line) are compared to wafer (circles) and simulated CD s under two different illumination conditions ($0.6NA/0.4\sigma_p$ and $0.7NA/0.35\sigma_p$). As to be expected, the litho process adds extra factors into CD performance. Under not-optimized illumination conditions (ex: $0.6NA/0.4\sigma_p$) the litho budget in CD performance can be significant.

Summarizing simulation and empirical results, we can say that the strong phase shift reticle used for exposure and CD collection has the best mask industry standard CD control (MTT and range are ~ 20 - 30 nm, $4X$) and PD control (MTT $< 5^\circ$, and range $< 5^\circ$).⁶ The results are consistent with the results of the 90 nm poly line printability statistical analysis we published in Ref. 4. By this analysis, one can estimate different reticle manufacturing factors and photolithography factors both affecting wafer CD performance.

The phase depth error impact requires additional study since wafer *CDs* and image placement errors vary considerably through focus/dose with the mean-to-target phase depth error.

5. Conclusions

The methodology for evaluating a reticle manufacturing budget in sPSM imaging performance was applied to the 65 nm poly line printability statistical analysis. The method and analysis flow was used for optimizing and justifying a reticle specification. Simulated results showed an excellent agreement with empirical data. The method provides a metrology for strong phase shift reticle quality control independent from a mask shop.

The sensitivity analysis showed the dominating factor in the imaging budget is the *CD* control of the 180°-phase openings (*PW*₁₈₀). The combined effect of *PW*, *PI* and *PD* variations is significantly stronger than a single parameter variation.

The phase depth error impact on imaging performance requires an additional study through statistical dose and focus variations, which can be carried out by the proposed Monte Carlo simulation flow.

Molotof image simulator can be used for image placement measurements for provided reticle specifications and process conditions.

One should take reasonable doubts in the simulation results and treat them with discretion.

Acknowledgements

Sincere thanks to scientists and engineers of the Advanced Development Labs at LSI Logic Corporation for assistance with Molotof package. I also thank process engineers for providing reticle *CD* and wafer *CD* data.

References

1. A. Kwock-Kit Wong, *Resolution Enhancement Techniques in Optical Lithography* (SPIE Press, Bellingham, Washington, 2001).
2. A. Erdman, Topography effects and wave aberrations in advanced PSM-technology, *Proc. SPIE, Santa Clara*, 345–355 (2001).
3. C. E. Tabery and C. A. Spence, Evaluation of 3D alternating PSM structures using mask topography simulation and AIMS at $\lambda=193$ nm, *Proc. SPIE, Santa Clara*, 429–439 (2001).
4. N. Belova, J. Jonson, S. Khodabandeh, and E. Croffie, Statistical analysis of poly line printability affected by sPSM manufacturing errors, *Proc. SPIE, Santa Clara*, **5379**, 268–278 (2004).
5. K. Taravade, E. Croffie, and A. M. Jost, Two Dimension Image-based Model Calibration for OPC Application, *Proc. SPIE, Santa Clara* (2004).
6. W. D. Kim, S. Jessen, C. Aquino, M. D. Eickhoff, P. Lim, N. Fukuhara, Y. Kikuchi, J. Tanzawa, *Mask Challenges and Capability Development for the 65-nm Device Technology Node: First Status Report*, Photomask, BACUS News, February 2004, **20**(2) (SPIE, Bellingham, USA, 2004).

PHYSICS AND MODELING OF TERA- AND NANO-DEVICES

Physics and Modeling of Tera- and Nano-Devices is a compilation of papers by well-respected researchers working in the field of physics and modeling of novel electronic and optoelectronic devices. The topics covered include devices based on carbon nanotubes, generation and detection of terahertz radiation in semiconductor structures including terahertz plasma oscillations and instabilities, terahertz photomixing in semiconductor heterostructures, spin and microwave-induced phenomena in low-dimensional systems, and various computational aspects of device modeling. Researchers as well as graduate and postgraduate students working in this field will benefit from reading this book.

