Cutting Edge Nanotechnology

Cutting Edge Nanotechnology

Edited by
Dragica Vasileska

In-Tech
intechweb.org

Published by In-Teh

In-Teh
Olajnica 19/2, 32000 Vukovar, Croatia

Abstracting and non-profit use of the material is permitted with credit to the source. Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published articles. Publisher assumes no responsibility liability for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained inside. After this work has been published by the In-Teh, authors have the right to republish it, in whole or part, in any publication of which they are an author or editor, and the make other personal use of the work.

© 2010 In-teh
www.intechweb.org
Additional copies can be obtained from:
publication@intechweb.org

First published March 2010
Printed in India

Technical Editor: Maja Jakobovic
Cover designed by Dino Smrekar

Cutting Edge Nanotechnology,
Edited by Dragica Vasileska

p. cm.
ISBN 978-953-7619-93-0

Preface

Semiconductor nanoelectronics has developed with unprecedented pace in the last twenty years. Year after year transistors have been continually scaled from micrometer to deep into nanometer regime (therefore the name Nanoelectronics). Up until this point it has been possible to scale conventional transistors but as the scaling process continues it becomes more difficult o follow Moore's law as challenges in photolithography become too difficult to be overcome and to produce devices with yields that will ensure components cost that provides profit to the manufacturers. What will happen at the end of the technology roadmap is still not clear even to Intel Fellows. To ensure further performance improvement in transistors operation two alternative strategies are typically followed. The first one is use of alternative materials, like strained Si and strained SiGe, for example, that offer higher mobilities for electrons and holes, respectively. The second avenue is to utilize alternative device structures. These alternative device structures come in different shapes: fully depleted silicon on insulator (SOI) devices, dual gate device structures, tri gate devices that are also known as FinFET device structures, MugFETs (multiple gate device structures), etc. A flow chart that explains the advantages and the disadvantages of both devices employing different materials and devices using different device designs is shown in Table 1 below.

Table 1. Non-classical metal oxide semiconductor field effect transistors designs.

Device	Ultrathin Body (UTB) SOI	Band-Engineered Transistor	Vertical Transistor	FinFET	Double-Gate
Concept	Fully-depleted SOI	SiGe or Strained Si; bulk Si or SOI	Double-gate or surround-gate structure		
Application/ D river	Higher performance, higher transistor density, lower power dissipation				

Advantages	Improved subthreshold slope; VT controllability	Higher drive current; compatible with bulk Si and SOI	Higher drive current; lithography independent gate length	Higher drive current; Improved subthreshold slope; improved short channel effect (SCE)	
Scaling Issues	Si film thickness, gate stack; worse SCE than bulk CMOS	High mobility film thickness (SOI); gate stack; integratability	Si film thickness; gate stack; integratability; process complexity; accurate TCAD	Gate alignment; Si film thickness; gate stack; integratability; process complexity; accurate TCAD	
Design	Device	Device	Device characterization; PD versus FD; compact		
Challenges	characterization; compact model and parameter extraction	characterization	model and parameter extraction; applicability to mixed signal applications		

Eventually, even the transistor structures listed in Table 1 will not be small enough and further miniaturization of devices will be required. Next in lane are nanowire transistors, single electron transistors and quantum dot structures. The nanowire devices show more attractive scaling in performance as they are made smaller. Specific applications for nanowire transistors in future include use in building high performance logic circuits as well as host of electronics applications on unconventional substrates, such as plastics, where such high performance devices have not been possible.

A conventional field effect transistor, the kind that makes all modern electronics work, is a switch that turns on when electrons are added to a semiconductor and turns off when they are removed. These on and off states give the ones and zeros that digital computers need for calculation. Interestingly, these transistors are almost completely classical in their physics. Only a few numbers that characterize their behavior are affected by quantum mechanics. However, if one makes a new kind of transistor, in which the electrons are confined within a small volume and communicate with the electrical leads by tunneling, all this changes. One then has a transistor that turns on and off again every time one electron is added to it; we call it a single electron transistor (SET). Furthermore, the behavior of the device is entirely quantum mechanical. The problem with SETs is that the switching behavior is most prominent at low temperatures. Nevertheless, cellular neural networks have been proposed by the University of Notre Dame team led by Craig Lent and co workers that utilize SETs.

Quantum dots, also known as nanocrystals, are a special class of materials known as semiconductors, which are crystals composed of periodic groups of II-VI, III-V, or IV-VI materials. Semiconductors are a cornerstone of the modern electronics industry and make possible applications such as the Light Emitting Diode and personal computer. Semiconductors derive their great importance from the fact that their electrical conductivity can be greatly altered via an external stimulus (voltage, photon flux, etc), making semiconductors critical parts of many different kinds of electrical circuits and optical applications. Quantum dots are unique class of semiconductor because they are so small, ranging from 2-10 nanometers (10-50 atoms) in diameter. At these small sizes materials behave differently, giving quantum dots unprecedented tunability and enabling never before seen applications to science and technology.

Finally, at the bottom of the hierarchy of scaling we have the direct manipulation of molecules and atoms. A new field has emerged, called molecular electronics. Molecular electronics (sometimes called moletronics) is an interdisciplinary theme that spans physics, chemistry, and materials science. The unifying feature of this area is the use of molecular building blocks for the fabrication of electronic components, both passive (e.g. resistive wires) and active (e.g. transistors). The concept of molecular electronics has aroused much excitement both in science fiction and among scientists due to the prospect of size reduction in electronics offered by molecular level control of properties. Molecular electronics provides means to extend Moore's Law beyond the foreseen limits of small-scale conventional silicon integrated circuits.

The main purpose of this book is to describe important issues in various types of devices ranging from conventional transistors (beginning chapters of the book) to molecular electronic devices whose fabrication and operation is discussed in the last few chapters of the book. As such this book can serve as a guide for identifications of important areas of research in micro, nano and molecular electronics. We deeply acknowledge the valuable contributions that each of the authors made in writing these excellent chapters that this book consists of.

Dragica Vasileska

Contents

Preface V

1. The need for practical regulation of developing commercial nanotechnology 001
Charles R. McConachie and J.D.

2. Improving Performance and Reliability of MOS Devices using Deuterium Implantation 015
Jae-Sung Lee

3. Heating Effects in Nanoscale Devices 033
Dragica Vasileska, Katerina Raleva and Stephen M. Goodnick

4. Low-Dimensional Group III-V Compound Semiconductor Structures 061
Nobuhiko P. Kobayashi

5. Analytic Approach to the Operation of RTD Ternary Inverters Based on MML 097
Juan Núñez, José M. Quintana and María J. Avedillo

6. Development of Superlattice Infrared Photodetectors 113
Shih-Hung Lin, Ying-Hsiang Wang, Che-Wei Chang, Jen-Hsiang Lu, Chun Chi Chen and Chieh-Hsiung Kuan,

7. "Bottom-up" Approaches for Nanoelectronics 137
Mrunal A. Khaderbad, Arindam Kushagra, M. Ravikanth and V. Ramgopal Rao

8. Templated Si-based nanowires via solid-liquid-solid (SLS) and vapor-liquid-solid (VLS) growth: Novel growth mode, synthesis, morphology control, characteristics, and electrical transport 167
Jae Ho Lee and Robert E. Geer

9. Probe Technologies for Micro/Nano Measurements 193
Kuang-Chao Fan,Fang Cheng, Yejin Chen, Bor-Shen Chen

10. Tensile and Flexural Deformation of Nickel Nanowires via Molecular Dynamics Simulations 207
Ram Mohan and Yu Liang

11. Advances in Nanowire-Based Computing Architectures 225
Jun Wu, Sriram Venkateswaran and Minsu Choi

12. New Class of Coolants: Nanofluids 253
Sadollah Ebrahimi, Anwar Gavili, Maryamalsadat Lajevardi, Taghi Dallali Isfahani, Iraj Hadi and Jamshid Sabbaghzadeh

13. Intersubband Transitions in the Quantum Dot Layers for Quantum Confined Photodetector 281
Shiang-Feng Tang, Tzu-Chiang Chen, Shih-Yen Lin and Hsing-Yuan Tu,

14. Intersublevel Relaxation Properties of Self-Assembled InAs/GaAs Quantum Dot Heterostructures 307
1Jiunn-Chy Lee and Ya-Fen Wu

15. Arithmetic Circuits Realized by Transferring Single Electrons 325
Wan-cheng Zhang and Nan-jian Wu

16. Single-Electron Circuits for Sigma-Delta Domain Signal Processing 349
Hisato Fujisaka

17. Carbon Nanotube Capacitors 375
Mark M. Budnik and Eric W. Johnson

18. Electrochemical Co-deposition of Metal-Nanoparticle Composites for Microsystem Applications 423
Kyle Jiang

19. Single Cell Analysis inside Environmental Scanning Electron Microscope (ESEM)-Nanomanipulator System 445
Mohd Ridzuan Ahmad, Masahiro Nakajima, Seiji Kojima, Michio Homma and Toshio Fukuda

20. Anti silver nanoparticle bacteria 471
Chi-Yea Yang

1

The need for practical regulation of developing commercial nanotechnology

Charles R. McConachie and J.D.
McConachie Law
Dallas, Texas, USA

1. Introduction

Nanotechnology began as a theoretical concept in 1959 in a talk by Nobel physicist Richard Feynman. By the 1980s the theory of nanotechnology became more of a fact when new microscopes were developed allowing scientists to see nanometers, down to one-billionth of a meter (Brown, 2008).

Commercial development of nanotechnology has expanded significantly as can be seen by the fact that between March 2006 and August 2008 the total number of consumer nanotechnology based products manufactured in the US rose from 125 to 426. In Asia the increase has been from less than 40 products to 227 in the same time period (Project on Emerging Nanotechnologies, 2009). A trip through Google with the search term "nanotechnology development" reveals approximately 5,320,000 different web sites (Google.com, 2009).

It is submitted that nanotechnology is a rapidly growing phenomena that has had and will have profound impact on man and the environment. Some of the impact will be good, especially in the new consumer products becoming available in all kinds of areas from new roofing insulation materials to new, incredible medical devices (McConachie, 2008). It is anticipated and predicted that this same nanotechnology development without regulation to protect the environment, health and safety (EHS) will result in profound and disturbing harm to man and the environment (Renn & Roco, 2006). The purpose of this chapter is to identify nanotechnology regulation that exists, present the rationale for maintaining *status quo ante* as well as for the promulgation of regulation promulgation of further regulation and, with an understanding of what the risks are likely to be, suggest that because there is now no binding regulation of nanotechnology mankind needs to take appropriate action before the EHS goes through its 9/11 event.

2. The State of Nanotechnology Regulation

On October 1, 2007, Dr. Patrick Lin, director of the Nanoethics Group in an article posted on the Nanoethics Group web site compared the development of nanotechnology with playing with fire; this because there is inadequate information and knowledge on the proper control of nanoparticles and what the dangers might be if there is a release of nanoparticles into the

atmosphere. Dr. Lin proposed that sufficient evidence exists to predict the existence of toxicological risks from nanotechnological exposure. As a result in his view nanotechnological particles should be regulated (Lin, 2007).

Ironically, there is at the time this chapter is being prepared, in November 2009, an almost total dearth of governmental regulation of nanotechnology and nanoparticles. Indeed, it was not until December 2006 that any government in the world enacted binding law to regulate nanotechnology, and that government is the Berkeley, California City Council in the US (Phillips, 2008). The City Council promulgated new law amending its hazardous materials law to include nanoparticles (Elvin, 2008). This local ordinance required researchers and manufacturers to report to the City of Berkeley what nanotechnology materials are being worked with and how the articles are handled to maintain safety (Elvin, 2008). Another US city, Cambridge, Massachusetts, considered the same kind of local ordinance, but as of July, 2008, had only gone as far as voting to accept recommendations of an advisory committee to track developments and changes and report back to the council (Bergenson, 2008). Whether it is coincidence or foresight that the only two cities to have preceded this far in nanotechnology regulation happen to be home to two of America's outstanding universities, Harvard and the University of California at Berkeley, is unknown.

As of February 2009 twenty-two states in the US had passed nanotechnology legislation. The various states legislation encompasses grants for research, business development and the like. Not one of these state statutes addresses any regulatory aspect of nanotechnology (Nanotechnology Statutes, 2009).

In the U.S., President Bush in 2004 signed into law the 21st Century Nanotechnology Research and Development Act (21-NRDA, 2004). While 21-NRDA contains important provisions for research and development, again, the Act does nothing to regulate by law nanoparticles. In 2007, and again in 2009 the US House of Representatives passed HR 554, the National Nanotechnology Initiative Amendments Act of 2009. The House passage of HR 554 in 2009 was a part of the February 2009 stimulus package. In both 2007 and 2009 the House without amendment passed the NNIAA. Ironically, the NNIAA has not been reported out of Committee in the Senate as of late August 2009. There are no hearings scheduled for HR 554 by the Senate Committee on Commerce, Science, and Transportation (HR 554, 2009). Even if the US Senate does take action with the NNIAA, the interesting aspect of the 2009 Amendments is that the bill contains any number of provisions for reporting, encouraging, studying, and advancing nanotechnology, while at the same time recognizing there are safety issues in nanotechnology development, and yet there is no new regulation of nanotechnology development or use in the 2009 Amendments.

The perceived need for nanotechnology regulation in the United States is not great while in Europe the official view of the European Commission is that no new regulations in the EU are needed because existing regulation leaves no regulatory void. According to the official responsible for regulatory aspects of nanotechnology at the European Commission, Cornelis Brekelmans, "[w]e are not in a regulatory void." At the Second Annual Nanotechnology Safety for Success Dialogue Workshop in October, 2008, Brekelmans stated that "We may decide to not authorize a product," and later the Commission might review, modify, or cancel an authorization (EurActiv, 2008).

Mr. Brekelman's perspective was challenged at the same Workshop by the leader of Greens/EFA, Axel Singhofen, who argued that "the reality is not quite how you [Brekelmans] present it." Contrary to Mr. Brekelmans stated views, Mr. Singhofen

advocated that developers of nanotechnology products should have to prove their safety before being allowed to enter the market (Azonano, 2008).
In both the US and Europe the prevailing government view either evidenced by word or lack of activity/interest is that the case for nanotechnology regulation of products being developed has yet to be made. On the other hand there are a number of non-governmental organizations (NGOs) such as Greens/EFA, Greenpeace and the International Risk Governance Council (IRGC) that hold to a different line. In a 2006 article published in the *Journal of Nanoparticle Research* entitled "Nanotechnology and the Need for Risk Governance," Renn and Roco held that the novel attributes of nanotechnology require the development of different routes to determine benefit-risk since regulation has not kept up with the development of new nanotechnology products (Renn & Roco, 2006).

3. A Look at the Risks from Nanotechnology

The lack of safety regulation of nanoparticles persists despite considerable work and research. In 2006 the International Risk Governance Council (IRGC) hosted a workshop in Switzerland concerning the "Conceptual Risk Governance Framework for Nanotechnology." The participants agreed that nanotechnology is divided into four broad generations of technology products and processes (Renn & Roco, 2006). With each successive generation the risks increase because the nanoproducts become more active and complicated.
The first generation, post 2000, consists of passive nanostructures. These steady function, or passive, nanoproducts consist for example of coatings, ultra precision engineering, polymers and ceramics. On March 5, 2008 Industrial Nanotech, Inc., announced that it was entering the commercial roof insulation market with lightweight thermal insulation based on its patented product line, Nansulate®, a passive nanoproduct (McConachie, 2008).
The second generation of nanoproducts, in the 2005 time frame, consists of active nanotechnology, which might include transistors, amplifiers, targeted drugs and chemicals, nanoscale fluids and laser-emitting devices. An active nanostructure product changes its state during operation. By way of example, a drug delivery nanoparticle changes its morphology and chemical composition. The new resultant state may also be subject to change from other changes in the biological, electronic, mechanical and magnetic properties (Renn & Roco, 2006).
The third generation or stage to begin next year, 2010, will be a system of nanosystems made up of various syntheses and assembling techniques. The third generation in medicine would include the production of an artificial organ made up of nanoscale cell tissues and scaffolds for cell engineering. In the area of nanoelectronics possible new devices would be based upon variables other than electrical charge. Third generation products with potential high risk include the behavior of engineered robotics, evolutionary artificial organs and modified viruses and brain cells (Renn & Roco, 2006).
The fourth generation, projected to begin in 2015, is where a heterogeneous molecular nanosytem has a specific structure and yet plays a different role. It is envisioned that molecules in devices will be used in new functions with new functions and structures. Nanomedicine products of the fourth generation would include cell aging therapies, stem cell nanocell therapy new genetic therapies (Renn & Roco, 2006).

Nanotechnology is about the creation of new products made up of new parts or ingredients to be used in new ways. In determining whether going forward nanotechnology presents sufficient risk to EHS so as to either regulate or limit it's admission to the marketplace, knowledge of what products based upon nanotechnology are being distributed in commerce and what products are being developed for use in commerce is a critical must. A great deal of the problem as pointed out by Renn and Roco is the ". . . uncertain/unknown evolution of the technology and human effects (for example, health changes at birth, brain understanding and cognitive issues and human evolution), as well as a framework through which organizations and policies can address such uncertainties" (Renn & Roco, 2006).

Put another way, the extent of the dangers from nanotechnology development have not been fully appreciated because of the fact that the properties of nanomaterials are not predictable based upon known laws of chemistry and physics. What one thinks *should happen* may very well have a completely different result in a nanotechnological base product. Part of the reason for the quite possible different distinctions, and thus the risk, is the fact that structure in a nanotechnology product is quite important in how both biological and physical behavior play out (Davies, 2006).

Citing Oberdorster and Maynard, Davies states:

> "We do not know enough about the toxicity and environmental effects to know whether . . . [nanotechnology] materials are also different in these respects, but it is likely, for example, that the toxicity of . . . [nanotechnology] materials is more related to their surface area than to their weight" (Davies, 2006).

Another perspective of the EHS risks that come from nanotechnology development are concerns about how penetration of human skin by nanoparticles, inhalation of nanoparticles effecting the lungs and respiratory system, the breach of the blood-brain barrier by nanoparticles in the bloodstream may all cause harm to man. As noted by Brown, a recent experiment reported in *Science Daily* that showed men's socks with an "odor fighting" feature when washed normally released ionic silver which after traveling through the wastewater process and entering natural waterways could very well harm the water ecosystems. This example shows that the law of unintended consequences clearly applies in any evaluation of EHS risks from nanotechnology (Brown, 2008).

4. A Worst Case Scenario?

There has not been a recorded serious EHS event caused by nanoparticles. The technology is new and commercial development is only now becoming common. There has been research into what in the real world might be viewed as a worst case scenario. Research by NASA (Life Sciences), Wyle Laboratories and UT Medical School (Pathology and Laboratory Medicine) in Houston, Texas inquired into the toxicity of carbon nanotubes to the lungs of mice. Five mice treated (under anesthesia) died within one week. All of the nanotubes introduced epitheloid granulomas, or tumor-like nodules, in the lungs. In some instances this resulted in inflammation of the lungs within 7 days. The mice that survived were sacrificed at 90 days and subsequent examination showed pronounced nodules and extensive necrosis (Lam et al, 2004). In the real world such unprocessed nanotubes are quite light. They could become airborne if released and potentially reach the lungs.

The researchers here concluded that carbon nanotubes are "more toxic than carbon black and can be more toxic than quartz" (Lam et al, 2004). The nanotubes used in the test were processed under different conditions with different heavy metals, such as nickel, iron and yttrium.

A nanoparticle that is popular in medical applications consists of metal nanoshells, nanoparticles that are tunable to electromagnetic radiation. The typical metal nanoshell is spherical core, i.e. silica, that is surrounded by a thin – often gold - shell. Such nanoshells are thought to be very beneficial in reducing carcinoma of the breast. Cancerous cells incubated and exposed to infrared light died while cells with no nanoshells were unharmed (Hirsch et al, 2003).

No one knows whether such nanoshells are safe. No one knows what happens to the nanoshells when cleared from the patient's dead cells by the immune system, or when the nanoshells are discharged or released. Indeed, no one knows what happens to the patient over the long term.

In 2003 the specter of nanotechnology disaster took a new turn when Prince Charles of Great Britain asked the Royal Society, the world's oldest scientific club to have a dialogue concerning the enormous risks when faced with self-replicating. This examination of the "grey goo" problem that commenced in 1986 when Dr. Eric Hexler first began describing the danger of the grey goo in the context of nanotechnology nanotechnology (Radford, 2003). By 2004 The Prince and Dr. Hexler both recanted on the idea that there is some valid science suggesting that grey goo will likely or even ever be close to rescue. Prince Charles reduced his criticism of nanotechnology from grey goo, acknowledging that it was quite likely such would not take place (Sheriff, 2004). Dr. Drexler, who is regarded as a leading early nanotechnology expert, lost considerable reputation when Richard Smalley, the Rice University chemist who shared the 1996 Nobel Prize for discovering Buckminsterfullerene, called Drexler out in late 2004 by saying Drexler was terribly wrong in predicting grey goo, and this just two days before President Bush signed into law the 21-NRDA in which nanotechnology was recognized as an important link to the future (Regis, 2004).

Even without gray goo being a realistic and serious EHS risk, there are sufficient unknowns to the safe use of nanotechnology so as to make credulous the concerns that developing nanotechnology, especially the third an fourth generations must be considered to contain risks that are not fully appreciated by man.

5. Nanotechnology Products Today

A recent Internet posting contained the first widely available inventory of nanotechnology consumer products (Project on Emerging Nanotechnologies, 2009).

There were more than 1,000 products in the Consumer Products Laboratory in August of 2009. The total number of nanotechnology based consumer products has increased 376 percent since 2006.. A total of 483 companies produced nanotechnology products located in 24 countries. By product category the most prevalent nanotechnology consumer product is in health and fitness. The growth of health and fitness products between 2006. 2009 was from slightly less than 150 to more than 605 of the total 1,015 products. By contrast only one other consumer product category, home and garden, had more than 150 products last year. Within the eight major product categories are found sub-categories. One sub-category of Home and Garden is Paint. Multi-functional products are categorized as "Cross Cutting."

"Coatings" is the sub-category of Cross Cutting, which means that a Coating consumer product based upon nanotechnology will have more than one purpose (Project on Emerging Nanotechnologies, 2009).
The regions of origin are reported in 2009 to be 540 of the 1,015 total from the US, 240 of 1,015 from East Asia, 154 products come from Europe and 66 products come from the rest of the world (Project on Emerging Nanotechnologies, 2009).

6. Existing Laws That Might Regulate Nanotechnology

With this kind of worldwide breakdown based upon region/country, it is not surprising that in determining what new regulation is necessary to protect man and the environment from the risks commonly recognized in new nanotechnology it is first necessary to have an understanding of what regulatory structures exist at the present, and if such structures are effective. An examination of US federal law that exists today provides a foundation.
The US Food and Drug Administration (FDA) is one of the oldest US consumer protection agencies. To market drugs or biologics in commerce the FDA must first approve an application and determine the product is both safe and effective (21 USC 355, 21 USC 360). Part of the approval process is that the drug or biologic will be manufactured in compliance with good manufacturing practices (GMPs) which include requirements concerning building facilities, such as design, lighting, ventilation, filtration, HVAC, plumbing, equipment and controls as well as controls of production and process (21 CFR 210, 21CFR211). FDA also is responsible for medical devices (21 USC 360). The approval process for medical devices is two-stepped. New, never before used devices must go through the full FDA review in what is described as a Premarket Application (PMA), while a medical device sold before October 1976 or that is substantially equivalent to a device lawfully on the market is submitted to FDA for clearance under what is known as a 510(K) notice (21 USC 360 1(k). The GMPs for devices, Quality Systems Regulation) mandate, as do the drug GMPs, that the production and process controls include environmental and contamination controls (21CFR820.70). There is support for the conclusion that as to drugs, biologics and medical devices the present US food and drug law is sufficient for purposes of regulating nanotechnology (Davies, 2006).
Unfortunately, the same may not be as true with other existing US regulatory schemes. For example, the Toxic Substances Control Act (TSCA) administered by the US Environmental Protection Agency (EPA) has been described as the primary vehicle to regulate nanotechnology because of its broad scope. One important question yet to be finally decided is whether nanoparticles may under this regulatory scheme be considered "new chemical substances." Both the National Resource Defense Council (NRDC) and Greenpeace argue that under the TSCA "all engineered" nanoparticles are "new chemical substances." Because of the divergence of views in the US and the way the US political system operates it is by no means certain that the courts will ultimately agree with NRDC and Greenpeace (Davies, 2006).
If the conclusion is nanoparticles are not "new chemical substances," Davies argues that the TSCA's "significant new use" rules (SNUR) could perhaps be utilized. In other words, the Administrator of EPA could conclude an existing chemical is to be regulated as though it were a new one. Whether this approach is practical is quite open to question. TSCA rulemaking is almost always a lengthy administrative process in which one chemical or

chemical group is considered at a time by the Administrator. Besides publishing required notices in the *Federal Register* any affected person could challenge the EPA by filing objections to the proposed rule. The upshot of such an objection may well result in an administrative hearing that is appealed first to the Administrator and then reviewed by the court of appeals. Going through this process one chemical group or one chemical entity at a time is not feasible in a developing new industry where changes, new developments or uses come with lightning rapidity. It should be remembered that of the existing US laws with broad coverage TSCA is considered to be the primary vehicle for regulating nanotechnology (Davies, 2006). This is not a bright prospect.

A second challenge to putting nanotechnology under the TSCA regulatory umbrella is when in the process TSCA should apply. If one assumes nanotechnology products fall under the EPA's jurisdiction by virtue of TSCA, toxicity downstream at the time the final formulation occurs cannot be assumed or predicted. While regulation of nanotechnology would be focused on final products TSCA would look to manufacturers of the basic forms of nanotechnology and expect these entities to anticipate, track and trace all possible final uses of the basic products. Should two or more basic nanoparticles be combined or joined to make a final product the new identity would remove the product from TSCA's present jurisdiction (Davies, 2006).

An additional problem with TSCA being effective to regulate nanotechnology is the requirement that the EPA must first meet a number of requirements prior to taking any regulatory action. This is seen quite clearly in *Corrosion Proof, et al. v. EPA* (Corrosion Proof, 1991). *Corrosion Proof* concerned the EPA's twelve-year proceeding to use the TSCA "to reduce the risk to human health posed by exposure to asbestos" (54 FR 29,460, 1989). The EPA was not the first US regulatory agency concerned over asbestos. In 1971, the Occupational Safety and Health Administration (OSHA) began limiting the exposure limit of asbestos, then @ 12 fibers per cubic centimeter (Corrosion Proof, at 1207, note 1, 1991).

Between 1979 and 1989 the EPA conducted its administrative proceeding leading up to the issuance of a final rule in 1989 that prohibited "the manufacture, importation, processing, and distribution in commerce of most asbestos-containing products" (Corrosion Proof Fittings at 1207-1208, 1991). The Final Asbestos Rule was to be implemented in stages over a six-year period. A number of domestic and foreign parties challenged the Final Rule claiming among other things that the rule-making process was fatally flawed because of a lack of due process and the lack of substantial evidence necessary to support the EPA's decision.

In American administrative law the doctrine of "substantial evidence" as a foundation for regulatory agencies reaching substantive decisions is well established. One seminal case that sets out the basic framework for court's to review EPA rule making for substantial evidence is *Chemical Manufacturers Association v. EPA*, where the appellate court held that determining substantial evidence meant whether (1) the regulated chemical in the environment was substantial in quantity and (2) whether exposure by humans to the chemical was significant/substantial (Chemical Manufacturers Association, 1990). In this context if the agency reaches a decision in exercising its judgment without reliance on set quantifiable risks, etc., it must alternatively "cogently explain why it has exercised its jurisdiction in a given manner" and provide a rational basis for what it did (Motor Vehicles Manufacturers Association, 1983).

In *Corrosion Proof* this is precisely what the court concluded the EPA had failed to do:

> We conclude that the EPA has presented insufficient evidence to justify its asbestos ban. We base this conclusion upon two grounds; the failure of the EPA to consider all necessary evidence and its failure to give adequate weight to statutory language requiring it to promulgate the least burdensome regulation required to protect the environment adequately (Corrosion Proof, at 1215, 1991).

The courts have also found fault in the process of the rule making itself. The EPA failed to allow cross-examination of all witnesses and failed to notify the parties until after the close of the hearings that it intended to use analogous exposure estimates to support the final rule. By not giving such notice, the petitioners were not able to challenge these estimates and make a record during the hearing. The court found fault with denying cross-examination, but held that defect, alone, was not sufficient to overturn the Final Rule (Corrosion Proof, at 1212, 1991).

EPA's failure to give any timely notice of its intent to use analogous exposure data to calculate its benefit risk methodology, however, did not fare as well. The court held that the EPA's analogous exposure data should have been available to the public's scrutiny before the record closed (Corrosion Proof, at 1212, 1991). The precedent for this conclusion was a similar instance where the Consumer Product Safety Commission failed to allow the public to comment on a conclusion it made about how its rule would impact the swimming pool slide market in an earlier case, *Aqua Slide 'N' Dive v. CPSC* (Aqua Slide, 1978).

What is seen from this examination of how the court has viewed EPA's efforts to regulate asbestos is the very real chance that EPA would take years to develop a rule under TSCA to control/ban/phase out specific nanoparticles because of risks only to have the reviewing court invalidate the rule due to the failure of the government to follow the law. Given the fact that the pace of nanotechnology technology development is ever increasing such delays in regulatory oversight are simply not acceptable.

OSHA was created in 1970 when the US Congress combined two existing occupational safety programs then located in the Department of Labor and what is now the Department of Health and Human Services (21 USC 651). The lead responsibility for enforcement of the OHSA Act was the Office of Safety and Health Administration located in the Department of Labor.

The key to OSHA regulation is the "occupational safety and health standard," which is "a standard which requires conditions, or the adoption of use of one or more practices, means, methods, operations or processes, reasonably necessary or appropriate to provide safe or healthful employment and places of employment" (21 USC 651).

While the regulation of nanotechnology could occur under OSHA, it is submitted that using OSHA to protect employees would not be effective. To know which products have a nanotechnology basis is not easy. Fairly sophisticated equipment would be in order and OSHA would have to first determine, for example, the relevant parameters from which to measure toxicity emanating from a factory or the environment around it. Additionally, OSHA lacks the breadth of resources needed to effectively regulate nanotechnology in a growing workplace.

Besides TSCA, the EPA is responsible for the enforcement of a number of other environmental protection laws such as The Clean Air Act, Clean Water Act, and the Resource Conservation and Recovery Act (Davies, 2006). These environmental laws generally authorize EPA to establish standards for acceptable pollution and then issue

permits to applicants that meet the standards. By definition, a firm that emanates waste that does not meet the established standard cannot obtain a permit, which is necessary to discharge the waste at issue (League of Wilderness Defenders, 2002).

All of these environmental laws suffer the same impediment to effective enforcement. Without sophisticated laboratory equipment and well-trained technicians locating nanotechnology products is quite challenging (Davies, 2006). In situations where the presence of nanoparticles is determined the issue then becomes the remedy. The EPA laws are not product specific and a complete ban of one or more nanoparticles from the environment may be fairly considered to be regulatory overkill. A possible exception to this statement is where one or more manufacturing facilities suffer leaks into the general environment of a nanomaterial that presents a substantial risk to the environment.

New industrial and commercial applications of nanotechnology are ever increasing. The estimated 2015 annual nanotechnology market, i.e. the fourth generation discussed above, is estimated at $1T dollars. Even with or perhaps because of such growth toxicity concerns from nanotechnology products continue to persist. Going back to 2001 safety problems with nanomaterials have been well known (Chenggang Li et al, 2009). Donaldson, Stone et al, of Napier University's Biomedicine Research Group reported in 2001 the very real health risks presented by ultra fine particles to the lungs (Donaldson et al, 2001). It is true, as Donaldson points out, that diseases in the lungs caused by inhaled particles are known as far back as the 14th Century. And while by the close of the 20th Century the significant death toll from asbestos and silica are coming to an end, a new particle-ultrafine is the subject of new concern.

American regulatory agencies have no worldwide monopoly on pre-market review. In the EU the critical document necessary to have a medicinal product distributed and sold commercially is the Marketing Authorization Application (MAA). Without a MAA for a drug, biologic or device the product may not be lawfully sold in the EU. A sponsor of a medicinal product files its MAA with the appropriate authority of a member state or to the European Medicines Agency (Marketing-Authorisation-Applications). Starting in 2005 submissions for oncology, diabetes, HIV and genital diseases must be submitted to the European Medicines Agency (EMEA). By virtue of this devolved system there are two approval procedures followed by the EU. The dual application process permits a sponsor of a new drug to apply for marketing authority (MA) in one member state and when approved to then request recognition of the MA by the remainder of the EU states (European Commission, at 28, 2006).

A question raised by some is whether the EU agencies are fully equipped and capable to make decisions that adequately protect EHS. The EMEA has conducted meetings among specialists throughout the EU to build expertise, establish professional relationships among different EU experts and to identify and satisfy needs (European Commission, at 34, 2006). In early 2008 the EMEA published a paper on regulation of nanotechnology (MHRA, 2008). Medicines for humans other than homoeopathic drugs require pre-market approval based in broad bush upon the US concept of safety and efficacy. The regulation includes authority for inspection by governmental officials, enforcing good clinical practices, good manufacturing and distribution practices and good laboratory practices. Should a regulated entity fail to meet required standards and procedures or produce adulterated mislabeled medicines regulatory officials have authority to inspect the premises and books and documents, to undertake prosecutions for consumer safety and punish wrongdoers

criminally and with confiscation orders. In sum other than the devolved system for medicine approval in the EU the differences between the US and the EU in the area of products requiring pre-market approval are not so different that there are sound reasons for concern.

The same is not necessarily true in the EU for other nanotechnology products. Specifically as to nanotechnology a 2009 "Safety for Success" dialogue took place in Brussels to discuss among other topics regulation (Nanowerk, 2009). In the Safety for Success meeting there was general agreement that in three areas coordinated effort was required:

1. "Developing trustworthy information on products containing nanomaterials that are on or near the market"
2. "Meaningful public engagement on the basis of shared definitions of nanotechnology."
3. "Ongoing regulatory reviews to provide clear guidance to industry on how to interpret regulatory frameworks . . . ."

Additionally, more research on nanotechnology risks was considered a priority, including gaining more knowledge about nanomaterials in the environment to make further clarification regarding existing regulations given the uncertainties of biological properties with nanomaterials. Finally, the stakeholders of the Safety for Success called for the introduction of post-marketing monitoring systems for nanoproducts in commerce.

From the record of the Safety for Success Dialogue it is submitted that the EU is certainly not as far along in the implementation of regulatory safety control of nanoproducts (with the possible exception of medicines and similar products) as the US. The reason may well be that the US has reached the conclusion that more regulation is necessary, but not yet implemented, while in the EU there is not yet general recognition that more regulation of nanotechnology development to protect EHS is indeed necessary. Recall that the official responsible for regulatory aspects of nanotechnology at the European Commission, Cornelis Brekelmans, has stated further regulation is not necessary as "[w]e [EU] are not in a regulatory void" (EurActiv, 2008).

With regard to devices, the EU follows Directive 98/79/EC for in vitro diagnostic devices that took effect in 1998 (In Vitro Directive, 1998). This was the first time that requirements on safety, quality, and performance bringing in vitro devices under regulation have been put in place.

8. How the 9/11 of Nanotechology Will Occur

The web page ***Responsible Nanotechnology*** sets out what many consider to be the most likely potential disasters from nanotechnology (CRNANO, 2004). War, economic meltdown, environmental meltdown from overproduction or leakage is the most obvious potential candidates. Without adequate regulation it is impossible to conclude that these risks are not real or cannot occur.

Another view comes from a European team that comprises *Nanologue* (Nanologue.net). Nanologue takes a new look at the potential future, both positive and negative. causing hundreds, perhaps thousands of injuries/deaths. In a time line format going forward advances in nanotechnology as well as disasters are set out. The future events in the time line are, of course, not real, but they do demonstrate how in a real sense the dark side of nanotechnology may impact on EHS. For example,

2010 The UK Government publically criticized the Global Framework on Emerging Technologies for moving too slowly and introduced its own, watered down, guidelines. These are voluntary.

2011 Workers at a factory in Toulouse went on strike, refusing to work with nanoparticles following a number of medical complaints. Demonstrations spread across Europe. The number of occupational health court cases increased.

A campaign by a major NGO was launched, calling for a moratorium on nanoscience and technologies until more is known about the health and environmental effects.

2012 In April, the process for delivering the Global Framework on Emerging technologies broke down and efforts to create a level playing field internationally were abandoned.

A major explosion occurred at a plant on the outskirts of Seoul, which releases several tons of nanoparticles into the environment (Nanalogue.net).

Under this scenario it does not get any better, with the result that the development of nanotechnology slows significantly.

9. Conclusion

Nanotechnology offers great potential in improving the quality of life for man as well as the environment. If this potential is to be achieved nanotechnology must be both fostered and controlled. Government and business realize that the fostering of nanotechnology is best served with the infusion of capital for research, capitalization, manufacturing and distribution. Regulation is not a word normally favored by business and is viewed positively by government only when government is pro-regulation. Of course, not all governments have the same views on regulation at the same time. The US government during President Bush's two terms was as a general rule more inclined to regulate business less than was government the proceeding eight years of President Clinton. Great Britain in the same way viewed regulation with less friendly eyes during the time Margaret Thatcher served as Prime Minister than when Labor and Tony Blair took over control of the Commons.

Nanotechnology, of course, is not political and does not recognize the borders of countries. If a spill of nanoparticles were to occur in Korea and create environmental havoc as postured above, governments and borders mean nothing. To keep the spill in Korea from doing harm to EHS potentially anywhere in the world, governments of countries where nanotechnology is being developed must come together and put into place common regulation that, in sum, will prevent the potential Korean spill from ever taking place. Such international cooperation is quite unusual, but not impossible. For nanotechnology to prosper over the long term, there is no other choice.

10. References

Azonano. (2008). No Regulatory Void on Nanotechnology, Says European Commission, http://www.azonano.com/News.asp?NewsID=8011, October 8, 2008.

CRNANO. (2004). Civil Rights in the Nano Era: Disater Scenarios, July 19, 2004. http://crnano.typepad.com/crnblog/2004/07/disaster_scenar.htmlEurActiv. (2008). No Regulatory Void on Nanotech, Says Commission, http://www.euractiv.com/en/science/regulatory-void-nanotech-commission/ article-176050, October 7, 2008. European Commission. (2006). Regulatory and Quality Assurance Frameworks for PGX: A Comparative Study of the US, EU and Four Member States Part 3, European Commission, EUR 22214 EN. 2006, p. 37. Google Search. (2009). http://www.google.com/search?hl=en&q=nanotechnology + development&aq =f&oq= &aqi =g1. [Search conducted June 10, 2009 by the author.] In Vitro Diagnostic Medical Devices Directive (98/79/EC) adopted, October 1998published, Official Journal of European Communities, December 7, 1998 (OJ No. L3317.12.98 p.1).Marketing-Authorisation-Applications. http:// www.eudrac.com/glossary/ Marketing-Authorisation-Applications.html MHRA. (2008). About Us, Medicines and Healthcare products Regulatory Agency, http://www.mhra.gov.uk/Aboutus/index.htmNanalogue.net. End of Project. Nananogue.net. http://www.nanologue.net/ Nanotechnology Statutes. (2009). http://www.ncsl.org/default.aspx?tabid=16835Nanowerk. (2009). Nanotechnology Safety for Success Dialogue, Nanowerk, http://www.nanowerk.com/spotlight/spotid=8830.phpProject on Emerging Nanotechnologies (2009). http://www.nanotechproject.org/ inventories/ consumer/

21 CFR 210. Title 21 Code of Federal Regulations, Part 210. Current Good Manufacturing Practice in Manufacturing, Processing, Packing, or Holding of Drugs; General. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm

21 CFR 211. Title 21 Code of Federal Regulations, Part 211. Current Good Manufacturing Practice for Finished Pharmaceuticals. http://www.accessdata.fda.gov/scripts/ cdrh/cfdocs/cfcfr/cfrsearch.cfm

21 CFR 820.70. Title 21 Code of Federal Regulations, Part 820.70. Quality System Regulation. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm

21 NRDA. (2004). 21st Century Nanotechnology Research and Development Act. Citation: 15 USC 7501 note, 117 STAT. 1924.

21 USC 355. Title 21 U.S. Code, Section 355. Federal Food, Drug, and Cosmetic Act: New Drugs. http://www.gpoaccess.gov/uscode/index.html

21 USC 360. Title 21 U.S. Code, Section 360. Federal Food, Drug, and Cosmetic Act: Registration of Producers of Drugs or Devices. http://www.gpoaccess.gov/ uscode/index.html

21 USC 651. Title 21 U.S. Code, Section 651. Occupational Safety Health Act et. seq.

54 FR.29,460 (1989). Volume 54 Federal Register, number 29, page 460. 1989. http://www.gpoaccess.gov/fr/index.html

HR 554. (2009). National Nanotechnology Initiative Amendments Act (NNIAA), http://www.govtrack.us/congress/bill.xpd?bill=h111-554 [The same bill was also passed by the House in the 110th Congress. See H.R. 5940, which passed the 110 Congress by 407 to 6, science.house.gov/Press/PRArticle.aspx?NewsID=2338 - 41k.]

Aqua Slide 'N' Dive v. CPSC. Vol. 569 Federal Reporter Second, p. 831. 5th Circuit US Court of Appeals, 1978)

Bergenson, Lynn L. (2008). City of Cambridge Adopts Recommendations for a Municipal Regulation of Health and Safety Policy on Nanomaterials, http://nanotech.lawbc.com/ 2008/08/articles/united-states/local/city-of-cambridge-adopts-recommendations-for-a-municipal-health-and-safety-policy-on-nanomaterials/

Brown, Nancy J. (2008). Is New Regulation Needed, and if so By Whom?, Legal Backgrounder, Vol. 23 no. 33, 2008. Washington Legal Foundation.

Chenggang Li, Haolin Liu, Yang Sun, Hongliang Wang, Feng Guo, Shuan Rao, Jiejie Deng, Yanli Zhang, Yufa Miao, Chenying Guo, Jie Meng, Xiping Chen, Limin Li, Dangsheng Li, Haiyan Xu, Heng Wang, Bo Li and Chengyu Jiang. (2009). Paman Nanoparticles Promote Acute Lung Injury by Inroducing Autophagic Cell Death through the Akt-TSC2-mTor Signaling Pathway, Journal of Molecular Cell Biology, June 2009.

Chemical Manufacturers Association v. EPA. Vol. 899 Federal Reporter Second, p. 344. 5th Circuit US Court of Appeals, 1990.

Corrosion Proof Fittings, et al v. EPA. Vol. 947 Federal Reporter Second, p. 1201. 9th Circuit US Court of Appeals, 1991.

Davies, J. Clarence. (2006). Managing the Effects of Nanotechnology, Woodrow Wilson International Center for Scholars, p.7. http://www.wilsoncenter.org/events/docs/Effectsnanotechfinal.pdf

Donaldson, K.; Stone, V., Clouter, A; Renwick, L.; MacNee, W. (2001), Ultrafine Particles, Occupational and Environmental Medicine 2001;58:211-216; doi:10.1136/oem.58.3.211, 2001

Elvin, George. (2006). Berkeley Nanotechnology Regulations Take Effect, http://www.nanotechbuzz.com/50226711/berkeley_nanotechnology_regulations_take_effect.php, December 2006.

Hirsch, L.R.; Stafford, R.J.; Bankston, J.A.; Sershen, S.R.; Rivera, B.; Pierce, R.E.; Hazle, J.D.; Halas, N.J.; and West, J.L. (). Nanoshell-Mediated Near-infrared Thermal Therapy of Tumors Under Magnetic Resonance Guidance, Proceedings of the National Academy of Sciences, Vol. 100, no. 23, November 11, 2003, pp. 13549–13554.

Lam Chiu-Wing; James John T.; McCluskey Richard; and Hunter Robert L. (2004). Pulmonary Toxicity of Single-wall Carbon Nanotubes in Mice 7 and 90 Days After Intratracheal Instillation, Toxicological Sciences, 77, 2004, pp.126-34. http://toxsci.oxfordjournals.org/cgi/content/full/77/1/126

League Of Wilderness Defenders/Blue Mountains Biodivdersity Project v. Forsgren. Volume 309, Federal Reporter Third, p. 1181, 9th Circuit US Court of Appeals, 2002.

Lin, Dr. Patrick. (2007). Nanotechnology Bound: the Case for More Regulation. The Springer Journal, No. 2, August 2007, pp. 105-122.

McConachie, Charles R. (2008). Practical Issues In Commercial And Regulatory Development Of Nanotechnology, Proceedings of NANO '08. 8th IEEE Conference, pp. 870-873, Arlington, Texas, August 2008.

Motor Vehicles Manufacturers Association v. State Farm Mutual Insurance, 463 United States 29, 103 Supreme Court 2856, 77 Lawyer's Edition 2d 443, 1983.

Phillips, Theresa. (2008). Nanotechnology Testing and Regulations Inadequate, http://biotech.about.com/b/2008/01/16/nanotechnology-testing-and-regulations-inadequate.htm, January 2008. [This is not to say there is no regulation al all. In the USA, certain products are regulated and if produced through nanotechnology, the existing regulation would apply. The best example is the US Food and Drug Administration must approve or clear biologics, drugs, medical devices, or food additives before these products may enter commerce, See 21 USC 321 et seq. Cosmetics and foods require no such premarket review.]

Radford, T. (2003). Brave New World or Miniature Menace, The Guardian, April 29, 2003.

Regis, E. (2004). The Incredible Shrinking Man, Wired, October 2004.

Renn, O. & Roco, M.C. (2006). Nanotechnology and the Need for Risk Governance, Journal of Nanoparticle Research, Vol. 8, 2006, pp. 153-191.

Sheriff, L. (2004). Prince Charles Gives Forth on Nanotech, The Register, July 12, 2004.

2

Improving Performance and Reliability of MOS Devices using Deuterium Implantation

Jae-Sung Lee
Uiduk University
South Korea

1. Introduction

Our advancing information network society is based on various kinds of digital communication systems in which versatile silicon integrated circuits (ICs) are indispensable key components. For the next-generation networking systems, there is need for higher packing density, higher quality and higher speed of ICs. Much effort is needed to improve the quality of thin silicon dioxide (SiO_2) used in submicron metal-oxide-semiconductor (MOS) field-effect transistors (FETs) in ICs. Since MOSFETs with thermal SiO_2 were developed in 1960, SiO_2 has been widely used as a gate insulator in MOSFETs and has played an important role in rapidly advancing IC development.

The quality of the gate insulator or gate oxide is closely related to the control of driving current, which is one of the most important factors for the MOS device. From this point of view, the practical minimum value of gate oxide thickness is below 2 nm at present, and at the same time, defect-free oxide must be prepared for both gate oxide and inter-gate oxide. Many kinds of contaminations from various processes and electrical-stressed-induced-damage on SiO_2 and the Si-SiO_2 interface must be extensively examined and removed.

Progressive silicon complementary MOS (CMOS) IC technologies are also based on reductions in the channel lengths of MOSFETs as well as on reductions in the gate oxide thickness. The reduction in the channel lengths leads to an increase in hot electron generation (for n-channel MOSFETs) owing to an increase in the electric field applied to the channel. The reduction in the gate-oxide film thickness results in an increase in the electric field applied to the thin gate-oxide films. Hot electron generation in short channel MOSFETs is followed by electron trapping in gate oxides on the silicon substrates. The electron trapping in gate oxides is the principal cause of instability for short channel MOSFETs, and electron traps in as-grown oxides need to be studied extensively. At any rate, improvement in the quality and reliability of thin SiO_2 is one of the most important concerns for next-generation MOS IC development.

The central theme of this work is to characterize the electrical reliability of MOS devices related to the nature of thin SiO_2 film. This monograph deals with hydrogen or deuterium process employed in fabrication to improve MOS device's reliability. The process of post-metallization anneal of the wafers at a low temperatures in hydrogen ambient is critical to CMOS fabrication technology to improve MOS device function by passivating the otherwise

electrically active interface traps, but it sets the stage for subsequent hydrogen-related degradation. Recently, an alternative process during which the interface traps are passivated by deuterium instead of hydrogen was demonstrated. This phenomenon can be understood as a kinetic isotope effect. The chemical reaction rates involving the heavier isotopes are reduced, and consequently, under the electrical stress, bonds to deuterium are more difficult to break than bonds to hydrogen. However, it is difficult to identify whether the deuterium could passivate the whole silicon dangling bonds along the gate length. Deuterium diffusion takes place primarily through the silicon oxide (SiO_2) in the MOS system because of the limited permeability of bulk Si, metal, and even poly silicon to deuterium. In the case of large scale ICs, therefore, the ability of deuterium to diffuse within the very thin gate oxide layer may be severely impeded, impacting the large-area devices.

2. Gate Oxide Reliability in MOS System

The microelectronics industry, including the internet and the telecommunications revolutions, owes its success largely to the existence of the thermal oxide of silicon, i.e., silicon dioxide(SiO_2). A thin layer of SiO_2 forms the insulating layer between the control "gate" and conducting "channel" of the transistors used in modern integrated circuits. As circuits are made more dense, all the dimensions of the transistors are reduced ("scaled") correspondingly, so that nowadays the SiO_2 layer thickness is 2 nm or less, and the reliability of such ultrathin oxide layers has become a major concern for continued scaling.
The reliability of SiO_2, i.e., the ability of thin film of this material to retain its insulating properties while subjected to high electric fields for many years, has always been an important issue and has been the subject of numerous publications over the last 43 years, since the realization that SiO_2 could be used as an insulating and passivating layer in silicon-based transistors. Oxide reliability and the experimental methods for accelerated testing have been the subject of earlier review papers.
For the relatively thick (> 10 nm) oxides used in earlier technologies, the breakdown mechanisms are actually fairly complex, and the detailed understanding of the intrinsic reliability has only come about in the more recent past. When a voltage is applied across the gate oxide, an electron current will flow if the gate voltage (V_g) is high enough and/or the oxide is thin enough. For thick oxides, the current is controlled by Fowler-Nordheim tunneling, while for thin oxides($\leq$ 3 nm) at voltages below about 3 V (corresponding to the barrier height between n-type silicon and the SiO_2) the current is due to direct quantum-mechanical tunneling. The electrons flowing across the oxide will trigger several processes depending on their energy. The defect generation mechanism at the lower energy process, which dominates at the voltages where present MOSFETs operation, is attributed to hydrogen release from the anode with a threshold gate voltage of about 5 V. As a consequence of the reaction of the released mobile hydrogen, a variety of defects such as electron traps, interface states, positively charged donor-like sates, etc., gradually build up to the point where the oxide breaks down destructively.
Low temperature post-metal anneal in hydrogen ambient is imperative from a fabrication standpoint since silicon dangling bonds at the Si/SiO_2 interface are electrically active and lead to the reduction of channel conductance. Electron spin resonance (ESR) on deep-submicron transistor Si/SiO_2 interface has in fact identified in the stress-induced P_b defects a spread in the distance between the silicon atom at which the defect is localized and its

nearest neighbors, which would correspond to a spread in the Si-H bond energy. Fig. 1. represents P_b defects at the (100) interface. Such defects appear as the source of interface trapped charges. Thermal activation of hydrogen from the P_b defect at the (111) interface has confirmed a spread in the bond energies of 0.08eV, and the spread of the energies of P_b defects at the (100) interface deems to be twice of that. The passivation process of P_b defects is described by the equation

$$P_b + H_2 \rightarrow P_bH + H \quad (1)$$

where P_bH is the passivated dangling bond.
In the case of bulk oxide defects generation, a model considers the interaction of the applied electric field E with the dipole moments associated with oxygen vacancies (weak Si-Si bonds) in SiO_2. The oxygen vacancies, known as E′ centers, generate dominant hole traps during the electrical operation. Fig. 2 shows the E′ defect in SiO_2 bulk system. The activation energy required for bond breakage is lowered by the dipolar energy, leading to a quantitative prediction for the field dependence of the activation energy for dielectric breakdown. Allowing for a distribution of energies of the weak bonds could account for a wide range of observations of the temperature- and field-dependence of SiO_2 breakdown times, since the defect which dominates the breakdown process may change depending on stress conditions.

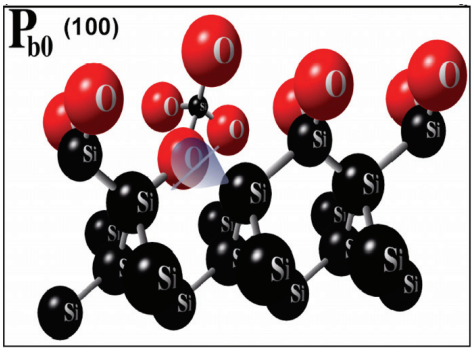


Fig. 1. Oxygen vacancy, P_b canter, at the Si/SiO_2 interface.

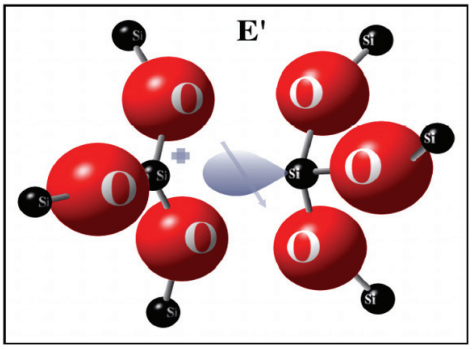


Fig. 2. Oxygen vacancy, E′ center, in SiO_2 bulk.

2.1 Hydrogen in Silicon and Silicon Dioxide

The introduction of hydrogen into silicon-based materials is an important step for the fabrication of many electronic devices. Besides hydrogen's ability to relieve network strain and passivate shallow donor states, hydrogen can also passivate electrically active midgap states. The latter are commonly associated with silicon dangling bonds and are found at surfaces, grain boundaries, interfaces, and in bulk silicon. Incorporation of hydrogen during the growth of amorphous silicon (a-Si) films is essential for producing devices such as solar cells. Also, device quality silicon-based transistors are annealed in a hydrogen-rich environment on order to passivate defects at the Si-SiO_2 interface.

A fundamental understanding of the Si-H dissociation process is essential for analyzing and controlling these phenomena. The present calculations build on earlier theoretical work which found that hydrogen interacts strongly with impurities as well as with defects in bulk crystalline and the dangling bond, where Si-H bonds are formed with bond strengths of up to 3.6 eV, similar to those found in silane. Although the energy to take a neutral hydrogen atom from an isolated dangling-bond site to free space is 3.6 eV, the energy to move the hydrogen into a bulk interstitial site is only about 2.5 eV.

Hydrogen exists in abundant quantities in bound forms in the oxide, the polysilicon gate, and the metal interconnects as a result of manufacturing process. Intrinsic defects generate electron-hole pairs and holes react with the bound forms to release mobile hydrogen. Hydrogen can migrate through the oxide as a neutral atom, as a positive ion (proton), or in other form, and reaches at the Si-SiO_2 interface where it reacts and new defects are generated. If the interface is first dry annealed, a process that is known to depassivate defects such as dangling bonds, introduction of H passivates the defects via the reaction

$$X + H \rightarrow SiH \tag{2}$$

where X stands for a dangling bond and SiH stands for a Si-H bond. If the interface contains H-passivated dangling bonds (SiH), introduction of H may depassivate some of them through the reaction

$$SiH + H \rightarrow X + H_2. \tag{3}$$

During device operation electronic defects are created that limit device lifetimes, and hydrogen has been observed to be involved in this degradation process. For instance, hydrogen is known to play a role during hot-electron degradation in silicon-based transistors, as well as during light-induced degradation in a-Si:H solar cells. The created defects are isolated and immobile. Hydrogen desorption from silicon dangling bonds is usually considered to be the dominant mechanism by which interface or bulk defects are created.

There is compelling evidence, however, that introduction of H induces additional defects at or near the interface. There defects can function as oxide traps, interface straps or border traps. Theoretical calculations so far focused on the behavior of hydrogen in bulk SiO_2. A great deal has been learned about the bonding of H in nominally perfect SiO_2 in different charge states and about cracking of H_2 molecules. Theoretical investigations of H at the Si-SiO_2 interface have been lacking, however, because of the enormous complexity of the problem.

2.2. Deuterium Effect in Degradation of MOS Devices

A large body of literature exists on hot electron and hydrogen related degradation of MOS devices. Degradation has been identified to be due to trap generation in the oxide as well as at the Si-SiO_2interface and the interface to polysilicon gates. While numerous experimental facts of the degradation have found a large variety of explanations, the actual mechanism of the damage has only been cursorily addressed. For damage within the oxide, electron hole defect recombination and the corresponding energy release have been identified as the likely cause. At the Si-SiO_2interface and the interface with polysilicon it is the release of hydrogen and the creation of dangling bonds that have been identified as causes of MOS device's degradation. However, the actual mechanism as to how the energy of electron-hole recombination or the energy of hot electrons (or holes) creates the defect has not been fully explained. A new large isotope effect of hot electron degradation by using deuterium instead of hydrogen for interface passivation was found.

The isotope effect can be used to distinguish hydrogen related hot electron damage from other mechanisms. It was initially discovered during scanning tunneling microscope (STM) experiments dealing with passivation and de-passivation of silicon surfaces in ultrahigh vacuum (UHV). These experiments showed that it takes a certain number of electrons (typically of the order of 10^6-10^8) having a certain energy to remove hydrogen from the (100) silicon surface. The same experiments performed with the isotope deuterium instead of hydrogen required roughly a factor or one hundred more electrons to remove deuterium for electron energies above ~ 4 eV. Recent STM experiments now show that this isotope effect increases dramatically for electron below 4 eV.

Silicon surface is passivated with hydrogen and then selectively depassivated by STM to form silicon nanostructure patterns that could be used for further chemical processing. In the course of these investigations, it is found that passivations with deuterium are significantly more resistant to STM depassivation. It takes higher voltages or significantly higher STM current densities to remove a given deuterium atom from the surface than necessary for hydrogen. The isotope effect is of the order of a factor of 100 at high STM voltages and much higher still at lower voltages. Typical measurement of the desorption yield for H and D is shown in Fig. 3. The strong dependence on the STM current is a signature of a process requiring multiple scattering events.

These basic STM experiments led to investigations of hot electron degradation of CMOS devices that were annealed in a deuterium atmosphere. Again a large isotope effect was found with transistor lifetimes being extended by factors of 10-50. Smaller improvements were observed under circumstances of large background hydrogen or reduced deuterium diffusion (e.g., nitride spacers).
The basic desorption mechanism toward which the isotope effect points is the creation (by hot electrons) of vibrational excitations of hydrogen bound to silicon (or polysilicon) at an interface. These vibrations and collisions with electrons having a few electron volts of energy can lead to desorption of the hydrogen, creating atomic hydrogen and a dangling bond. The freed atomic hydrogen subsequently can create further damage. The desorption mechanism itself determines critical energies and current densities and is therefore important for understanding and controlling degradation.

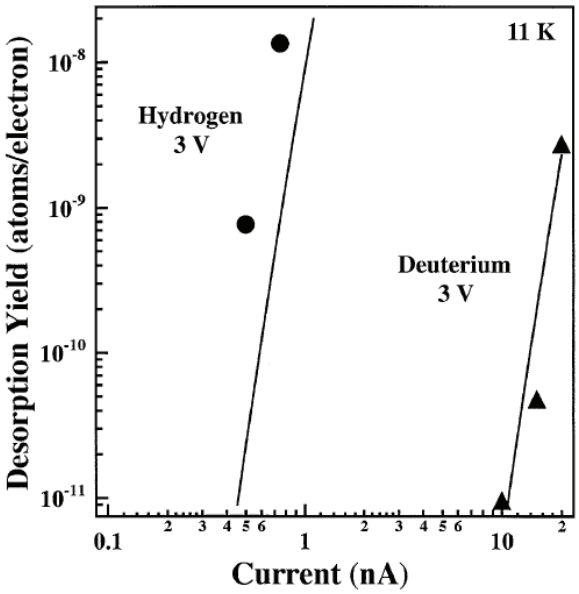


Fig. 3. Comparison of hydrogen and deuterium desorption yields at 3 V and 11 °K as a function of STM current showing an isotope effect and current dependence.

3. Deuterium Implantation in MOS Devices

The passivation with the annealing process is thought to be due to deuterium (D)-terminated, dangling bonds at the silicon surface, reducing interface trap density. This process relies on the diffusion of deuterium to the interface in the entire device area. Deuterium diffusion takes place primarily through the gate oxide, as depicted in Fig. 4, because of the limited permeability of bulk Si, metal, and even polysilicon to hydrogen or deuterium. Room temperature diffusion coefficient through Si is measured to be $\approx 10^{-15}$ cm^2/s, compared with $\approx 10^{-11}$ cm^2/s in SiO_2 . The hydrogen diffusion through the

polysilicon is further retarded at the grain boundaries, as has been demonstrated in a study of thin-film transistors (TFT). In the case of large scale ICs, therefore, the ability of hydrogen to diffuse within the very thin SiO_2 layer may be severely impeded, impacting the large-area devices. This is particularly alarming since it has, in addition, been reported that H_2 permeability in SiO_2 is reduced as the oxide thickness decreases.

Lyding et al. delivered the deuterium to the region of the gate oxide in an oven through thermal diffusion. This causes most of the deuterium to be wasted. In addition, during the sintering process, the deuterium may experience difficulty diffusing through some materials to reach the Si/SiO_2 interface, especially in those cases where several layers of metallization are located between the deuterium gas and the Si/SiO_2 interface.

Deuterium is introduced into the semiconductor devices by implantation in our study, instead of by thermal diffusion as was done by Lyding et al. The implantation may be accomplished at any step of the semiconductor process flow. In general, deuterium implantation is provided so that, during subsequent thermal cycles, the deuterium will diffuse to the gate oxide/silicon interface and become chemically attached to the dangling bonds at the interface, this generally being the Si/SiO_2 or polysilicon/SiO_2 interface. The energy, dose and defects of the implant are optimized to affect this.

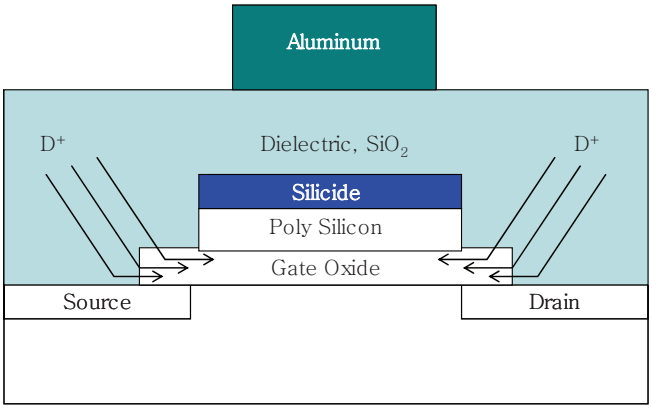


Fig. 4. Schematic of the device structure, indicating the path of D^+ diffusion through gate oxide from the edges of poly-silicon gate.

3.1. Calculation for Ion-Implant Process

There is need for the mathmetical calculation for the ion implantation. TRIM(Transport of ions in matter) is Monte Carlo computer program that calculates the interactions of energetic ions with amorphous targets. TRIM is a group of programs which calculate the stopping and range of ions (10 eV ~ 2 GeV/amu) into matter using a quantum mechanical treatment of ion-atom collisions. This calculation is made very efficient by the use of statistical algorithms which allow the ion to make jumps between calculated collisions and then averaging collision results over the intervening gap. During the collisions, the ion and atom have a screened Coulomb collision, including exchange and correlation interactions between

the overlapping electron shells. The ion has long range interactions creating electron excitations and plasmons within the target.
TRIM accepts complex targets made of compound materials with up to eight layers, each of different materials. It caculates both the final 3D distribution of the ions and also all kinetic phenomenons associated with the ion's energy loss: target damage, sputtering, ionization, and phonon production. Fig. 5 is the Setup Window for TRIM execution. The widow is used to input the data on the ion, target, and the type of TRIM calculation that is wanted. Almost all inputs have online explanations.

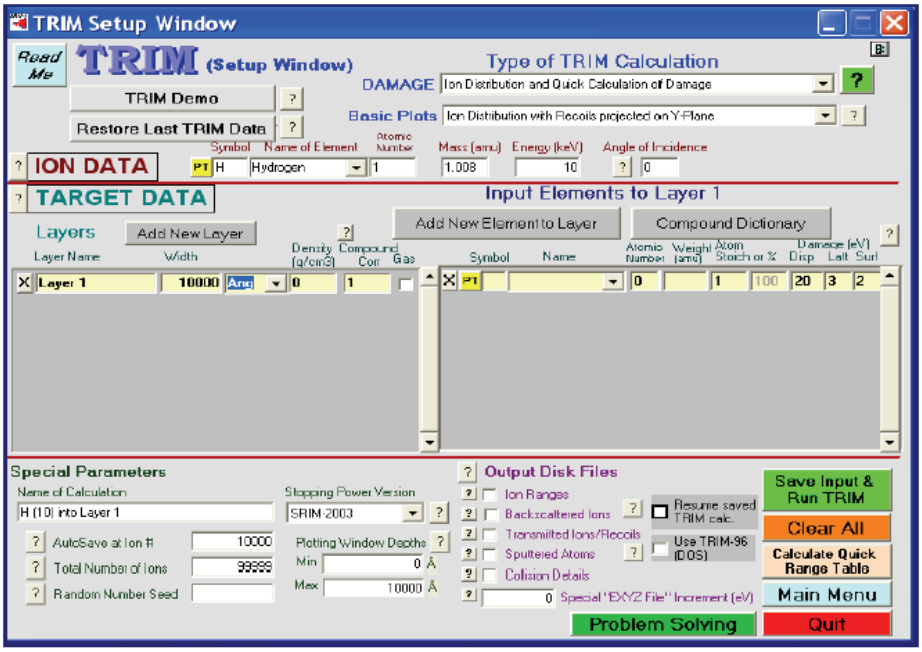


Fig. 5. TRIM Setup Window to calculate ion-implant process.

4. Experimental Results

4.1. Device Fabrication

In our study, both p- and n-MOSFETs were fabricated using standard CMOS processes for various channel lengths and widths down to 0.15 μm. Fig. 6 shows schematic cross section of our n-MOSFET device and experimental set up for the voltage stress measurements. The effective oxide thickness of our devices has a range of 3 ~ 7 nm. The gate oxide films were produced with a conventional furnace in H_2-O_2 ambient. The hydrogen (H) or deuterium (D) implantation was performed at the back end of the process line (after first metallization) to passivate the defects which spreaded in gate oxide area.

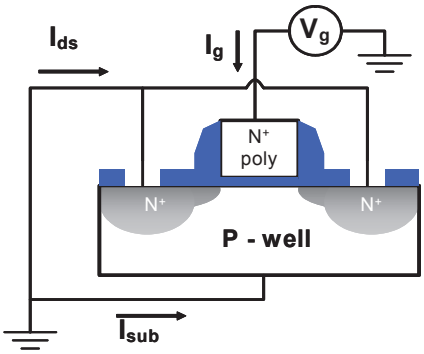


Fig. 6. Schematic cross section of n-MOSFET and experimental setup for the voltage stress measurements.

Transistors from a given wafer were divided into two groups. One group was implanted by H^+ ion with ≤ 60 keV and the other group was implanted by D^+ ion with ≤ 80 keV. Ion dose was fixed at $1X10^{14}/cm^2$ only for hydrogen implant as referential sample, while in case of deuterium implantation the ion dose have a range from $1X10^{10}/cm^2$ to $1X10^{16}/cm^2$ to find optimum process condition. Post-annealing was achieved at 400 ℃ for 30 minutes at N_2 ambient for the whole devices to activate the injected ions and to annihilate damages due to the implant process.

The implantation conditions for each ion were extracted through the computer simulation (TRIM tool). The total thickness from the top of first metal to the bottom of gate oxide was about 700 nm, including aluminum, silicon dioxide, and polysilicon layers, as shown Fig. 7. The control device was also prepared without our implantation processes to compare its electrical properties with those of our processed devices.

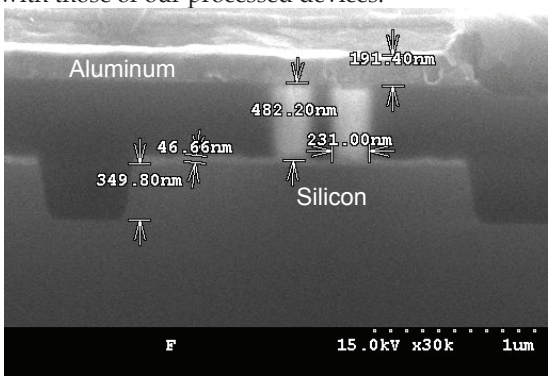


Fig. 7. Cross-section picture for our processed wafer. Two via contacts are shown between silicon substrate and aluminum metal layer.

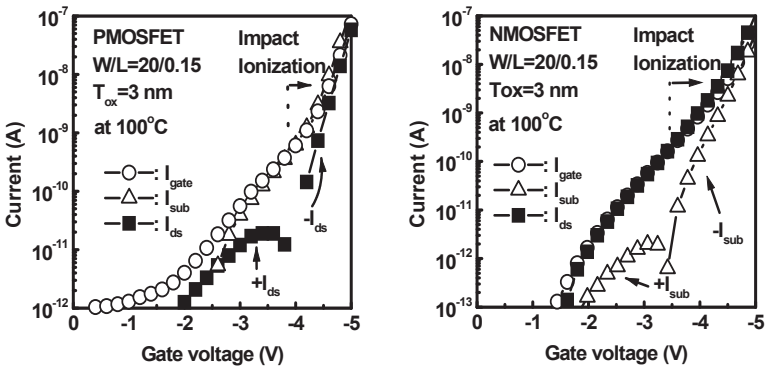
To investigate the reliability of our devices, the degradation phenomenon such as channel hot carrier injection (HCI), negative-bias temperature instability (NBTI), and stress-induced leakage current have been studied.

A voltage of V_g= V_d = ±3.0 V was applied to the 3 nm-thick-MOSFET gate at the room temperature to accelerate HCI gate oxide degradation. A source terminal is connected to substrate and grounded. Stress voltages of V_g=-2.8 V for NBTI was applied to the p-MOSFET gate at the temperature of 50 ~ 100 ℃. The source and drain terminals were connected to the substrate and grounded. For gate oxide leakage current measurements, large area MOSFETs (W/L=500 μm/500 μm) were used to avoid the edge effect. While the constant voltage, V_g= ±3.5 V, was applied to gate terminal, the gate current was measured simultaneously. The percent shifts (%) of saturation drain current (I_d) and the shift of threshold voltage (ΔV_{TH}) were measured to determine device parameter degradation. The percent shifts (%) of the gate current (I_g) were also monitored to assess gate oxide wear-out.

4.2. Result

The carrier separation experiment was conducted to measure the gate current I_g, the sum of source and drain currents I_{ds}, and the substrate current I_{sub} separately before stress, applying negative polarity of V_g. All currents flowing into the device are taken as positive.

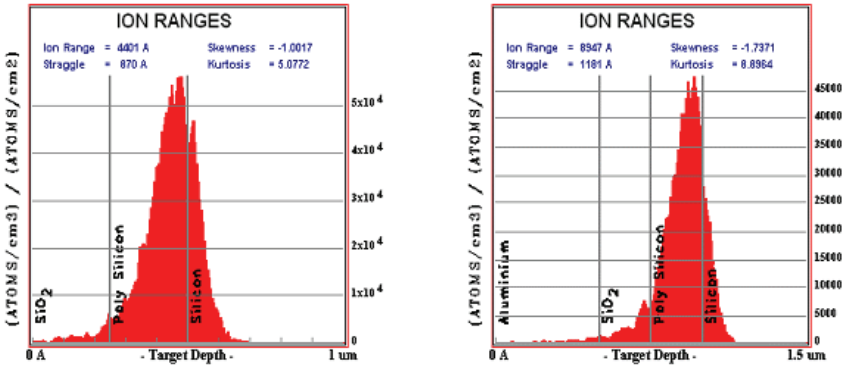
Fig. 8 shows I_g, I_{sub}, and I_{ds} versus sweeping V_g in our p- (a) and n-MOSFET (b) at 100 °C, respectively. In p-MOSFET, the conduction mechanism for three current components indicates that the electron current, when tunneled to the substrate, produces electron-hole pairs by impact ionization from near V_g=-4.0 V. The impacted ionized holes flow out through the source/drain; hence, those "hot" holes generate negative I_{ds}. Below V_g=-4.0 V, "cold" hole injection from silicon valence band and electron injection from polysilicon valence band become allowed, simultaneously. In n-MOSFET, the trend is quite similar to the case of p-MOSFET. However, I_{ds} measured the electron current and I_{sub} measured the hole current. Near V_g=-3.5 V, the increase of I_{sub} tends to slow down, and changes the sign from positive to negative that means the impact ionization could be dominant.



(a) (b)

Fig. 8. Carrier separation I-V curves for 3 nm-thick gate oxide devices (W/L : 20 μm/0.15 μm) measured in negative bias : (a) p-MOSFET; (b) n-MOSFET.

Before actual ion implantation its optimum condition was estimated through the computer simulation with TRIM. The material stack between aluminum and silicon substrate in our MOS device was composed of SiO_2 [300 nm] / poly-Si [250 nm] / gate oxide [3~7 nm] / Si-substrate. Fig. 9 shows the distribution of ion ranges for deuterium when the incident energy was 30 KeV for SiO_2/poly-Si/Si structure and 85 KeV for Al/SiO_2/poly-Si/Si structure. The each peak was close to the bottom of poly-Si gate, and then deuterium ion would be diffused into the gate oxide area through post-annealing process.



(a) (b)

Fig. 9. Ion range calculation with the computer simulation for (a) SiO_2/poly-Si/Si structure and (b) Al/SiO_2/poly-Si/Si structure.

Secondary ion mass spectroscopy (SIMS) measurement of Al/SiO_2/poly-Si/Si structure which was implanted by deuterium on the top Al with 85 KeV energy and 10^{16}/cm^2 dose is shown in Fig. 10. The SIMS analysis was done after 400 ℃ post-annealing process. The deuterium concentrations at the two SiO_2 interfaces are higher than that for aluminium area, indicating also deuterium concentration at the gate oxide region. Consequently, deuterium incorporation in the thin gate oxide (3 ~ 7 nm) was achieved at lower temperature through our implant process.

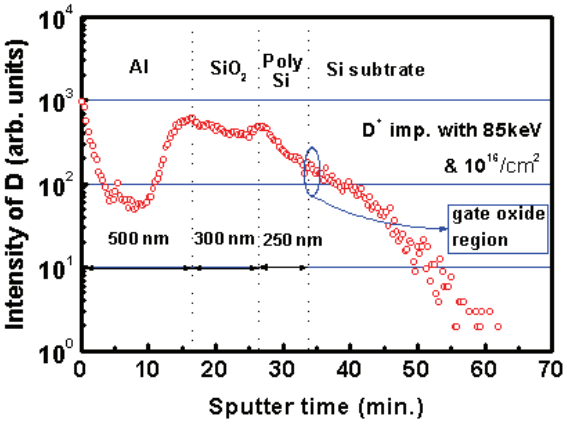


Fig. 10. SIMS analysis for Al/SiO_2/poly-Si/Si structure that was implanted with deuterium ion.

Fig. 11 represents the variation of gate current for both H^+ and D^+ implanted p-MOSFETs. The current was measured at the gate terminal while the gate voltage was being swept from 0 to 3 V. The device with deuterium implantation showed a lower gate leakage current for the entire range of sweeping voltages. Because the generation of stress-related bulk-oxide traps is suppressed by deuterium incorporation, we can infer that the isotope effect is valid in our new implant process. Therefore, deuterium implantation might be used to improve the reliability, if the suitable implant condition is obtained.

The normalized gate current is plotted as a function of stress time for p-MOSFETs with 3 nm-thick gate oxides in Fig. 12. The conventional, the hydrogen, and the deuterium-processed devices were stressed at a constant V_g=3.5 V voltage. The stress-induced leakage current is considered to be monitor for the defect generation in the gate oxide. From these curves, deuterium process shows fewer defects than hydrogen process, and the gate oxide implanted with deuterium, $1X10^{12}$/ cm^2 dose, generates almost the same number of defects as the gate oxide annealed in H_2. Trap generation rate increases with the deuterium (or hydrogen) concentration in the gate oxide, which is the same result as in the hydrogen or deuterium annealing process.

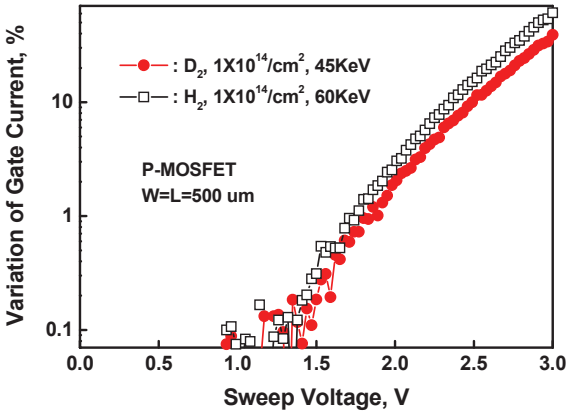


Fig. 11. Change of gate leakage current, after a constant voltage stress (V_g=3.5 V), of large-area p-MOSFETs implanted with H_2 and D_2, respectively.

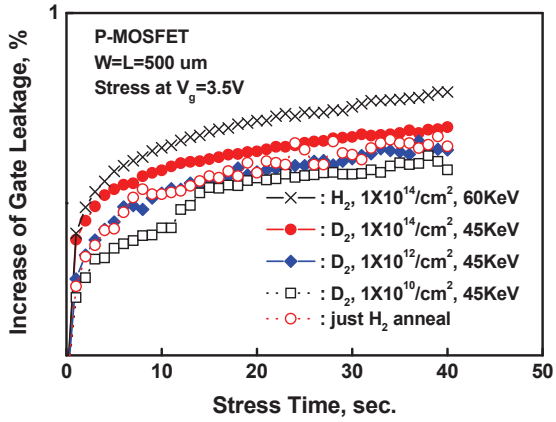
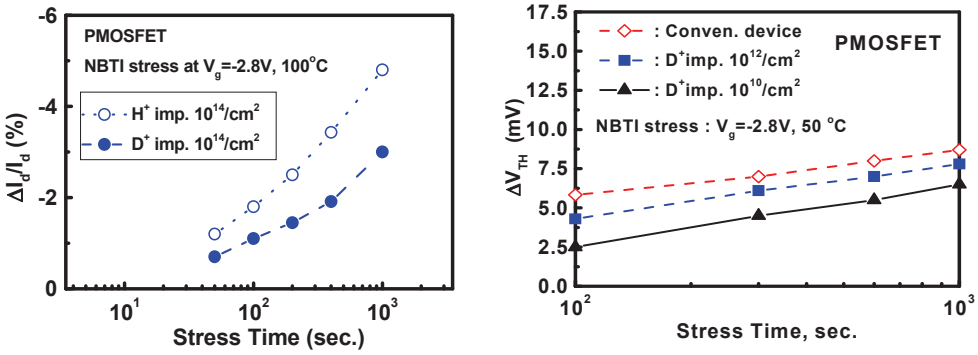


Fig. 12. Gate current transients during constant gate voltage, V_g = 3.5 V, for deuterium-implanted p-MOSFETs. These curves represent the evolution of defect generation.

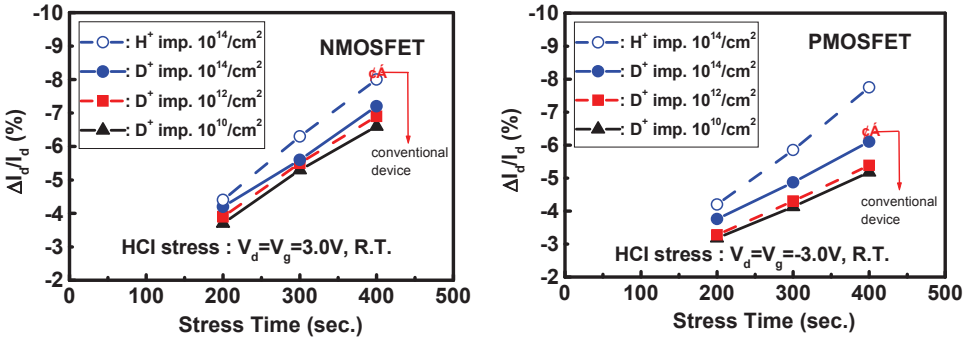
NBTI of p^+-gate p-MOSFETs has been the most serious issue for the reliability of ultra-thin gate oxide for continuous scaling down of devices. The isotope effect of the NBTI phenomenon was evaluated in our devices. Fig. 13 shows the saturation current and the threshold voltage dependence on the stress-time, respectively, for hydrogen- and deuterium-implanted p-MOSFETs. The degradation of deuterium-incorporated gate oxide is less remarkable than that of conventional gate oxide. Instead of hydrogen, using a deuterium can suppress the hydrogen-related precursors because the bonds to deuterium are more difficult to break than bonds to hydrogen.



(a) (b)

Fig. 13. (a) Decrease of saturation current and (b) variation of threshold voltage depending on NBTI stress time for both hydrogen-and deuterium-implanted p-MOSFETs.

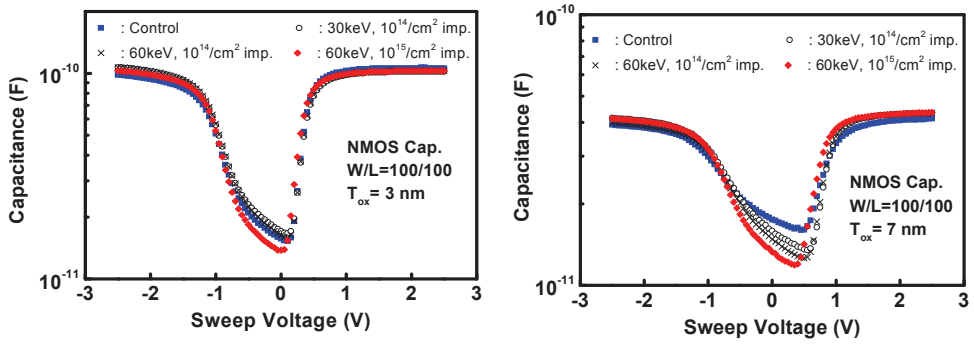
In the channel hot carrier stress test, the Si/SiO_2 interface is degraded by hot carriers that are travelling the device from source to drain. The accelerated stress tests performed on MOSFET typically results in localized oxide damage, which has been correlated to Si/SiO_2 interface trap states. The generation of the interface trap states is due to hot carrier stimulated hydrogen desorption and de-passivation of the silicon dangling bonds. Fig. 14 shows the decrease of the saturation current by HCI stress for hydrogen- and deuterium-implanted MOSFETs. The deuterium-implanted device shows the slight enhancement for both n- and p- types compare to the conventional and the hydrogen-implanted devices. We believe that the incident deuterium atoms were not fully replaced with as-bonded hydrogen atoms at the Si/SiO_2 interface, and the damage area was restricted to the narrow channel area near drain.



(a) (b)

Fig. 14. Decrease of saturation current depending on HCI stress time for both hydrogen- and deuterium-implanted (a) n-MOSFETs and (b) p-MOSFETs.

In case of a large-area capacitor (100 X 100 μm²), the capacitance-voltage (CV) curve was measured at 1 MHz frequency, as shown in Fig. 15, for the deuterium-implanted n-MOSFETs. A shift of threshold voltage was observed in the deuterium-implanted device, which is related to the interface states density. In addition, there is a reduction in the minimum capacitance, indicating that the net doping in the Si substrate has been decreased. We believe that the difference of threshold voltage between before and after D^+ implantation is due to two different mechanisms. One is the elimination of the interface states by D^+ implantation, while the other is the change in the net doping in the substrate, due to deactivation of channel dopants.



(a) (b)

Fig. 15. Variation of CV characteristics for both the control and the deuterium-implanted n-MOSFETs with (a) 3 nm-thick gate oxide and (b) 7 nm-thick gate oxide.

4.3. Discussion and Summary

The generation of charge trapping in the existing precursor is explained by the strained bond model. The strained bonds exist not only at the SiO_2/Si interface but also in the oxide bulk because strain extending 1-3 nm into oxide from SiO_2/Si interface has been observed. When a strained Si-O bond in SiO_2 is broken by a hydrogen ion under the high-pressure and high-temperature hydrogen anneal, structure is rearranged locally to relax the strain, generating a trivalent silicon and a nonbridging oxygen simultaneously. The trivalent silicon acts as a positively charged center after it traps an injected hole. The nonbridging oxygen may act as a neutral electron trap center.

In the annealing process hydrogen or deuterium atom reaches at the gate oxide layer through upper layers, like aluminum, silicon oxide, silicon nitride, and polysilicon, by diffusion mechanism. In case of deuterium annealing, low levels of deuterium could be expected because the silicon nitride or polysilicon layer acts as a barrier to deuterium. Therefore, deuterium bond may not distribute uniformly along the channel area. Non-uniform distribution of deuterium bond tends to lack an isotope effect during electrical stress. By means of our suggested implantation, the deuterium bond could be distributed intentionally in the gate oxide layer.

Fig. 16 illustrates the possible reaction for the generation of oxide traps in the deuterium-implanted n-MOSFETs gate oxide with the energy band diagram. When the deuterium bond

distributes through the gate oxide, two kinds of reactions, the interface and the bulk reaction, may occur independently. The interface reaction involves deuterium release that could produce positive deuterium ions. The deuterium ions bonded with non-bridging atoms in gate oxide bulk react with the energetic electrons or holes. The bond breakage is not accelerated as rapidly, because deuterium is twice as heavy compare to hydrogen, and the dissociation by injected electrons or holes is suppressed. In the process of dissociation the mass of the atom plays a significant role and a large kinetic isotope effect is the consequence.

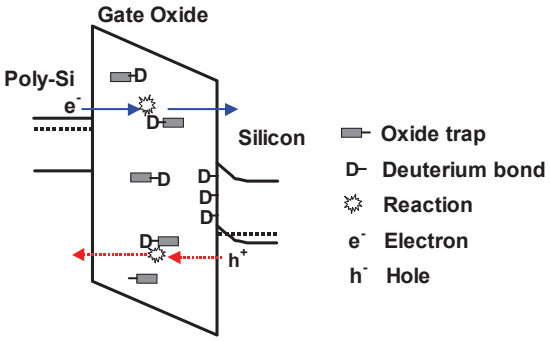


Fig. 16. Illustration for the reaction of injection carriers with the deuterium bonds existing at Si/SiO_2 interface and/or in SiO_2 bulk.

5. Conclusion

In the large-scale transistor structures, where gate oxide thickness of 7 nm and below is used, the post-metallization annealing (PMA) leaves a large number of Si/SiO_2 interface states unpassivated. Furthermore, the un-bonded hydrogen atoms in the SiO_2 is known as the main cause of the devcie degradation, such as NBTI and HCI. By means of our new method of deuterium implantation at the back-end process, we have improved the gate oxide reliability of MOSFETs, as compared with that of the conventional device. It was found that Si-D bonds (instead of Si-H) play a major role in suppressing the generation of oxide traps in our process. However, when the concentration of deuterium is redundant in gate oxide, excess traps are generated and degrades further the performance. We have suggested the new deuterium implantation process for the reliability of MOSFET, which is compatible with the conventional hydrogen annealing process.

6. Acknowledgment

This work was supported by the Proton Engineering Frontier Project (PEFP) funded by the Korean Government.

7. References

Blochl, P. E. & Stathis, J. H. (1999) Hydrogen electrochemistry and stress-induced leakage current in silica. *Phys. Rev., Lett.*, vol. 83, pp.372-375.

DiMaria, D. J. & Cartier, E. (1995) Mechanism for stress-induced leakage currents in thin silicon dioxide films. *J. Appl. Phys.*, vol. 78, pp. 3883-3894.

Edwards, A. H., Pickard, J. A. & Stahlbush, R. E. (1994) Interaction of hydrogen with defects in a-SiO_2. *J. Non-Cryst. Solids*, vol. 179, pp.148-161.

Guan, H., Li, M. F., He, Y., Cho, B. J. & Dong, Z. (2000) A thorough study of quasi-breakdown phenomenon of thin gate oxide in dual-gate CMOSFET's. *IEEE Trans. Electron Devices*, vol. 47, pp. 1608-1616.

Hess, K., Kizilyalli, I. C. & Lyding, J. W. (1998) Giant isotope effect in hot electron degradation of metal oxide silicon devices IEEE Trans. Electron Devices, vol. 45, pp. 406-416.

Jeppson, K. O. & Svensson, C. M. (1977) Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices. *J. Appl. Phys.*, vol. 48, pp. 2004-2014.

Kundu, T. & Misra, D. (2006) Enhanced SiO_2 reliability on deuterium-implanted silicon. *IEEE Trans. Device and Material Reliability*, vol. 6, no. 2, pp. 288-291.

Lee, J.-S. (2008) Modeling of time-dependent defect generation during constant voltage stress for thin gate oxide of sub-micron MOSFET. *Jpn. J. Appl. Phys.*, vol. 47, no. 1, pp.19-22.

Lee, M. H., Lin, C. H., & Liu, C. W. (2001) Novel methods to incorporate deuterium in the MOS structures. *IEEE Electron Device Lett.*, vol. 22, pp.519-521.

Lyding, J. W., Hess, K. & Kizilyalli, I. C. (1996) Reduction of hot electron degradation in MOS transistors by deuterium processing. *Appl. Phys. Lett., vol. 68, pp. 2526-2528.*

Pantelides, S. T., Rashkeev, S. N., Buczko, R., Fleetwood, D. M. & Schrimpf, R. D. (2000). Reactions of hydrogen with Si-SiO_2 interfaces. *IEEE Trans. Nucl. Sci.*, vol. 47, no. 6, pp. 2262-2268.

Priyanka, M., Lai, M., Kumar, R. & . Singh, S. N. (2005) Optimum hydrogen passivation by PECVD Si_3N_4 deposited crystalline silicon solar cells. *Photovoltaic Specialist Conference*, p. 1822, Jan. 2005.

Rost, T. A. & Wallace, R. M. (2001) Method for improving performance and reliability of MOS technologies and data retention characteristics of flash memory cells. U.S. Patent 6221705, Apr. 24.

Sakura, T., Utsunomiya, H., Kamakua, Y. and Taniguchi, K. (1998) A detailed study of soft- and pre-soft-breakdowns in small geometry MOS structures. *IEDM Tech. Dig.*, pp. 183-186, CA. 1998.

Schuegraf, K. F. & Hu, C. (1994) Hole injection SiO_2 breakdown model for very low voltage lifetime extrapolation. *IEEE Trans. Electron Devices*, vol. 41, no. 5, pp. 761-766.

Surma, B., Misiuk, A., Jun, J., M. Rozental, Wnuk, A., Ulyashin, A. G., Antonova, I. V., Popov, V. P. & Job, R. (1998) Effect of pressure treatment of electrical properties of hydrogen-doped silicon. *ASDAM' 98* pp. 47-48, Oct. 1998.

Tuttle, B. (1999) Structure, energetic, and vibrational properties of Si-H bond dissociation on silicon. *Physical Review B*, vol. 59, pp.12, 884-12,889.

Uchida, H., Inomata, S. & Ajioka, T. (1989) Effect of interface traps and bulk traps in SiO_2 on hot-carrier-induced degradation. *IEEE Int. Conference on Microelectronics Test Structures*, pp. 103-108, 1989.

Wu, E., Nowak, E., Aitken, J., Abadeer, W., Han, L. K. & Lo, S. (1998) Structural dependence of dielectric breakdown in ultra-thin gate oxides and its relationship to soft breakdown modes and device failure. *IEDM Tech. Dig.*, pp. 187-190, CA, 1998.

Wu, J., Rosenbaum, E., MacDonald, E., Li, B. E., Tracy, B. & Fang, P. (2000) Anode hole injection versus hydrogen release: The mechanism for gate oxide breakdown. *IEEE Int. Reliability Physics Symp.*, San Jose, CA, pp. 27-32, 2000.

Yamamoto, T., Uwasawa, K. & Mogami, T. (1999) Bias temperature instability in scaled p+ polysilicon gate p-MOSFETs. *IEEE Trans. Electron Devices*, vol. 46, pp. 921-926.

3

Heating Effects in Nanoscale Devices

Dragica Vasileska[1], Katerina Raleva[2] and Stephen M. Goodnick[1]
[1]*Arizona State University, Tempe, AZ, 85287-5706, USA*
[2]*University Sts Cyril and Methodius, Skopje, Macedonia*

1. Introduction

The ever increasing demand for faster microprocessors and the continuous trend to pack more transistors on a single chip has resulted in an unprecedented level of power dissipation, and therefore higher temperatures at the chip level. Thermal phenomena are not directly responsible for the electrical functionality and performance of semiconductor devices, but adversely affect their reliability. Four major thermally-induced reliability concerns for transistors are: (1) degradation of device thermal characteristics due to heating effects, (2) failure due to the electrostatic discharge phenomenon, (3) stresses due to different rates of thermal expansion of transistor constituents, and (4) failure of metallic interconnects due to diffusion or flow of atoms along a metal interconnect in the presence of a bias current, known as the electromigration phenomenon. Self-heating of the device and interconnects reduces electron mobility and results in a poor or, at best, non-optimal, performance of these devices and structures. Fig. 1 shows the trend of the average power density for high-performance microprocessors according to the ITRS. No flattening or slower decelerated increase will occur after the introduction of the Silicon on Insulator (SOI) technology. It should be noted that the power density shown in Fig. 1 is the average power density, i.e. the total chip power divided by the chip area. In logic circuits, such as microprocessors, the power is non-uniformly distributed. There are portions of the chip of quite low power dissipation (memory blocks) and, on the other hand, portions running at full speed with high activity factors where the power density can easily be more than a magnitude higher than the average chip power density from Fig. 1. The latter portions will create hot spots with quite high local temperature. The power density in the active transistor region (essentially the channel region underneath the gate) is again much higher than the average power density in a hot spot when the transistor is in the on-state. Thus, the treatment of self-heating and the realistic estimation of the power density is quite a complex problem.

Sometime within the next five years, traditional CMOS technology is expected to reach limits of scaling. As channel lengths shrink below 22 nm, complex channel profiles are required to achieve desired threshold voltages and to alleviate the short-channel effects. To fabricate devices beyond current scaling limits, Integrated Circuits (IC) companies are simultaneously pushing planar, bulk silicon CMOS design while exploring alternative gate stack materials (high-k dielectrics and metal gates), band engineering methods (using

strained Si or SiGe), and alternative transistor structures, such as fully-depleted (FD) SOI devices, dual gate (DG) structures, and FinFETs.

The problem with the SOI devices is that they exhibit self-heating effects (Majumdar, 1993; Pop, 2006). The buried oxide layer underneath the thin active silicon layer, having a thermal conductivity 100 times smaller than the value of bulk Si thermal conductivity, constitutes a large barrier for the heat removal from the active devices to the heat sink. Thus, it will become much more difficult to remove the heat generated in the active devices and considerable self-heating is expected to occur in the SOI transistors. This, in turn, causes a decrease of the on-currents and an increase of the off-current leading to a dramatic deterioration of transistor switching performance with consequences on the overall circuit performance. Also, the thermal conductivity of the thin silicon film decreases due to phonon boundary scattering.

Note that the removal of heat from deep inside the structure becomes very important issue in Today's nano-electronics Industry. Various novel semiconductor thermoelectric coolers and structures, such as thermionic (Shakouri *et al*.,1998) and nanowire (Chen & Shakouri, 2002) coolers have been proposed, developed, and investigated due to the advance of nanotechnology. For example, thermionic emission current in heterostructures can be used to achieve evaporative cooling by selective emission of hot electrons over a barrier layer from cathode to anode. Such structure can effectively build a temperature gradient within the range of electron mean free path (a few hundred nanometers), which can be used to remove the heat from a CMOS hotspot.

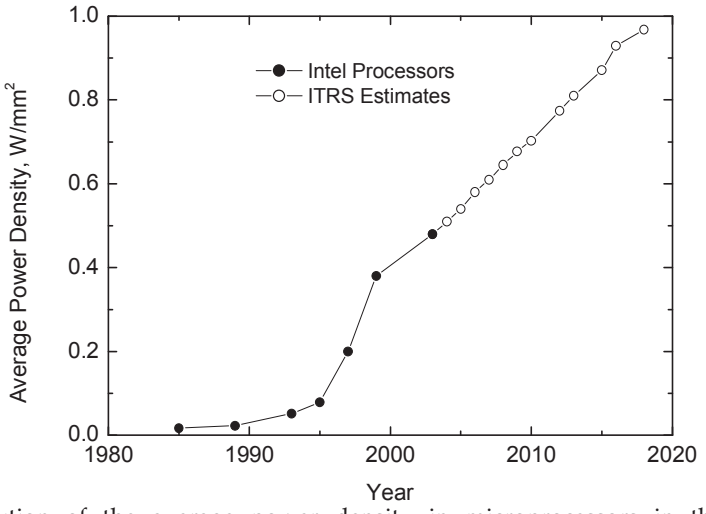


Fig. 1. Evolution of the average power density in microprocessors in the past and expectations of the ITRS until 2018.

In summary, the self-heating effect is particularly important for transistors in SOI circuitry where the device is separated from the substrate by a low thermal conductivity buried silicon dioxide layer as well as copper interconnects that are surrounded by low thermal conductivity dielectric materials (Borkar, 1999). Accurate thermal modeling and design of

microelectronic devices and thin film structures at micro- and nanoscales poses a challenge to the thermal engineers who are less familiar with the basic concepts and ideas in sub-continuum heat transport.

2. Some General Aspects of Heat Conduction

The energy given up by the constituent particles such as atoms, molecules, or free electrons of the hotter regions of a body to those in cooler regions is called heat. Conduction is the mode of heat transfer in which energy exchange takes place in solids or in fluids in rest (i.e. no convection motion resulting from the displacement of the macroscopic portion of the medium) from the region of high temperature to the region of low temperature due to the presence of temperature gradient in the body. The heat flow can not be measured directly, but the concept has physical meaning because it is related to the measurable scalar quantity called temperature. Therefore, once the temperature distribution T(r,t) within a body is determined as a function of position and time, then the heat flow in the body is readily computed from the laws relating heat flow to the temperature gradient. The science of heat conduction is principally concerned with the determination of temperature distribution within solids.

The basic law that gives the relationship between the heat flow and the temperature gradient, based on experimental observations, is generally named after the French mathematical physicist Joseph Fourier, who used it in his analytic theory of heat. For a homogeneous, isotropic solid (i.e. material in which thermal conductivity is independent of direction) the Fourier law is given in the form

$$\mathrm{q}(\mathrm{r},t) = -\kappa \nabla T(\mathrm{r},t) \quad W/m^2 \tag{1}$$

where the temperature gradient is a vector normal to the isothermal surface, the heat flux vector q(r,t) represents heat flow per unit time, per unit area of the isothermal surface in the direction of the decreasing temperature, and κ is called the thermal conductivity of the material which is a positive, scalar quantity. Since the heat flux vector q(r,t) points in the direction of decreasing temperature, a minus sign is included in Eq. (1) to make the heat flow a positive quantity. When the heat flux is in W/m^2 and the temperature gradient is in °C/m, the thermal conductivity κ has the units W/(m °C).

Clearly, the heat flow rate for a given temperature gradient is directly proportional to the thermal conductivity κ of the material. Therefore, in the analysis of heat conduction, the thermal conductivity of the material is an important property, which controls the rate of heat flow in the medium. There is a wide difference in the thermal conductivities of various engineering materials. The highest value is given by pure metals and the lowest value by gases and vapors; the amorphous insulating materials and inorganic liquids have thermal conductivities that lie in between. Thermal conductivity also varies with temperature. For most pure metals it decreases with temperature, whereas for gases it increases with increasing temperature.

At nanometer length scales, the familiar continuum Fourier law for heat conduction is expected to fail due to both classical and quantum size effects (Geppert, 1999; Zeng *et al.*, 2003; Majumdar, 1993). The past two decades have seen increasing attention to thermal conductivity and heat conduction in nanostructures. Experimental methods for characterizing the thermal conductivity of thin films and nanowires have been developed

and are still evolving. Experimental data have been reported on various nanostructures: thin films, superlattices, nanowires, and nanotubes. Along the way, models and simulations have been developed to explain the experimental data. This section summarizes some past work and the current understanding of heat conduction in nanostructures. We first give a brief overview on the fundamental physics that distinguishes phonon heat conduction in nanostructures from that in macrostructures. Then we discuss a few size effects in nanostructures that impact their thermal conductivity.

Heat conduction in dielectric materials and most semiconductors is dominated by lattice vibrational waves. The basic energy quantum of lattice vibration is called a phonon, analogous to a photon which is the basic energy quantum of an electromagnetic wave. Similar to photons, phonons can be treated as both waves and particles. Size effects appear if the structure characteristic length is comparable to or smaller than the phonon characteristic length. Two kinds of size effects can exist: the classical size effect, when phonons can be treated as particles, and the wave effect, when the wave phase information of phonons becomes important. Distinction between these two regimes depends on several characteristic lengths (Chen, 2001). The important characteristic lengths of phonon heat conduction are the mean free path, the wavelength, and the phase coherence length (Reif, 1985). The mean free path is the average distance that phonons travel between successive collisions. The mean free path Λ is often estimated from kinetic theory as

$$\kappa = \tfrac{1}{3} C_V v_P \Lambda \tag{2}$$

where C_V and v_P are the volumetric specific heat capacity of a phonon, and the phonon velocity inherent in a material. In silicon, for example, the phonon mean free path is on the order of ~300 nm (Ashegi *et al.*, 1998). The phase of a wave can be destroyed during collision, which is typically the case in inelastic scattering processes, such as the phonon-phonon collision.

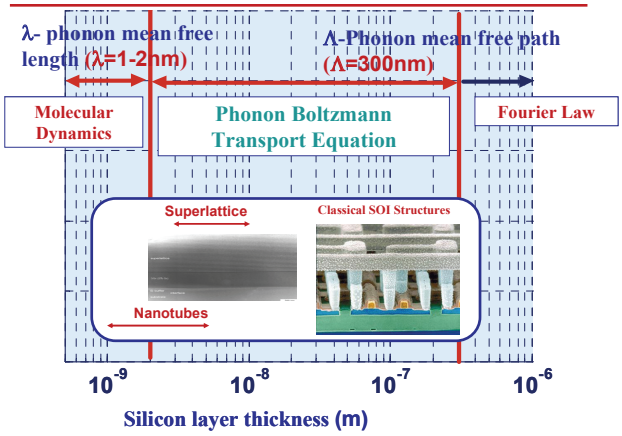


Fig. 2. Regime map for phonon transport in ultra-thin silicon layers. Mean free path Λ is a distance that phonons travel on average before being scattered by other phonons. If the dimension of the silicon layer is smaller than Λ, the Boltzmann Transport Equation (BTE) should be used for heat transfer analysis of the thin film. The dominant phonon wavelength, λ, at room temperature, is on the order of 2-3 nm. Analogous phonon wave simulations should be performed for devices with thicknesses comparable to λ.

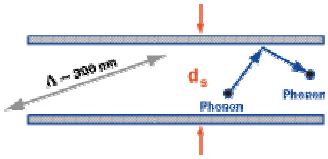


Fig. 3. Phonon-boundary scattering is responsible for a large reduction in the thermal conductivity of a thin silicon layer where the thickness of the film, d_s, is comparable to or smaller than the phonon mean free path, Λ.

Fig. 2 compares the dimensions of several nanostructures (e.g., an SOI device and a superlattice structure) with the dominant phonon mean free path (MFP) and wavelength at room temperature. This graph also provides a general guideline for the appropriate treatment of phonon transport in nanostructures. Phonon transport can be predicted using the Boltzmann particle transport equations (BTE), which is required only when the scattering rates of electrons or phonons vary significantly within a distance comparable to their respective mean free paths.

The BTE simply takes care of the bin counting of the energy carrier particles of a given velocity and momentum, scattering in and out of a control volume at a point space and time. Analysis of the heat transfer in microelectronic devices, interconnects and nanostructures using the BTE is very cumbersome and complicated, even for simple geometries, and has been the topic of research and development in the field of micro- and nanoscale heat transfer for the past two decades (Choi & Maruyama, 2003). Eq. (2) can only be considered as providing the qualitative behavior of a thermal conductivity from which the thermal conductivity is found to be proportional to the phonon mean free path (MFP). The phonon MFP is well known to become shorter as the system is hotter because the phonon population is increased, which causes the collision frequency among phonons to be high. Increased phonon collisions prevent the phonons with high energy in the hot region from moving to the cold region and vice versa. This means that the energy transport is low; consequently the thermal conductivity is low. Therefore, it can be inferred that phonon scattering governs the thermal conductivity.

Detailed descriptions and analyses of the ballistic heat transfer in a semiconductor/ metallic layer are beyond the scope of this book chapter. However, the most prominent manifestation of ballistic heat transport in thin films would occur in the form of large reductions in thermal conductivity compared to the bulk values. Ballistic phonon transport in silicon films, or phonon-boundary scattering (see Fig. 3), has been investigated through large measured reductions in the lateral thermal conductivity compared to the bulk value near room temperature (Ju & Goodson, 1995; Liu & Ashegi, 2004; [a]Liu & Ashegi, 2005).

The lateral thermal conductivity of the thin silicon layer decreases as the thickness of the film is reduced. Deviation of the thermal conductivity from the bulk value takes a sharp dive as the thickness of the film is reduced beyond 300 nm, which is the order of magnitude for the phonon mean free path in silicon at room temperature. For example, the thermal conductivity of the 20 nm thick silicon layer is nearly an order of magnitude smaller than the bulk value. The impact of phonon-boundary scattering on the thermal conductivity of a thin silicon layer can be predicted using the BTE solution and the theory described by Asheghi *et al*., such that it agrees very well with the experimental data.

One way to estimate the impact of the micro/nanoscale effect is to use the modified thermal conductivity values for thin silicon and copper layers in conventional thermal simulation tools that use the continuum theory or diffusion equation. In order to perform more realistic estimates of the current degradation and the hot spot temperature we have followed the approach of Sondheimer (Sondheimer, 2001) that takes into account phonon boundary scattering by assuming it to be purely diffusive. Namely, the thermal conductivity of a semiconductor film of a thickness a, under the assumption that the z-axis is perpendicular to the plane of the film, the surfaces of the film being at z=0 and z=a, is given by:

$$\kappa(z)=\kappa_0(T)\int_0^{\pi/2}\sin^3\theta\left\{1-\exp\left(-\frac{a}{2\lambda(T)\cos\theta}\right)\cosh\left(\frac{a-2z}{2\lambda(T)\cos\theta}\right)\right\}d\theta \tag{3}$$

where $\lambda(T)$ is the mean free path expressed as $\lambda(T)=\lambda_0(300/T)$ nm where room temperature mean free path of bulk phonons is taken to be $\lambda_0=290$ nm. Selberherr (Palankovski & Selberherr, 2001; Sivaco Manuals) has parameterized the temperature dependence (Asheghi *et al*., 1998) of the bulk thermal conductivity in the temperature range between 250K and 1000K. In our case we find that the appropriate expression is:

$$\kappa_0(T)=\frac{135}{a+bT+cT^2}\quad \text{W/m/K} \tag{4}$$

where a=0.03, b=1.56×10^{-3}, and c=1.65×10^{-6}. Eqs. (3) and (4) give excellent agreement with the experimental and the theoretical data reported in a later Asheghi paper (Asheghi *et al*., 2005). Our group has utilized Eqs. (3) and (4) to obtain both the temperature and thickness dependence of the thermal conductivity for different temperatures, as shown in Fig. 4, compared to experimental data of Asheghi and co-workers (Asheghi *et al*., 2005) at 300K. This model for the thermal conductivity is then implemented into the energy balance equation for acoustic phonons in our electro-thermal device simulator, and different device technology nodes have been examined.

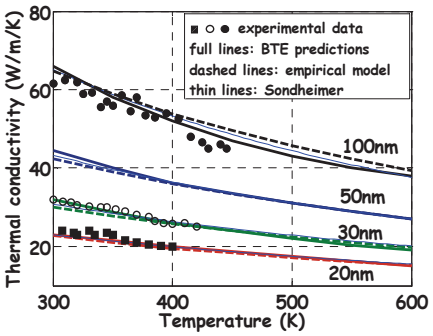


Fig. 4. Silicon film thickness dependence of the average thermal conductivity at T=300 K vs. active silicon layer thickness. Experimental data are taken from the work of Asheghi and co-workers (Asheghi *et al*., 2005).

In the ultra-fast laser heating processes at time scales of 10^{-15} to 10^{-12} seconds, as well as high speed transistors switching at timescales in the order of 10^{-11} seconds, the temperatures of the electron and phonon systems are not in equilibrium and may differ by orders of magnitude. Even after the phonon and electron reach equilibrium, the energy carried away by phonons can travel only to 10-100 nanometers; therefore, the temperature of the transistor can easily rise to several times its designed reliability limit. Under these circumstances, regardless of the cooling solution at the packaging level, a catastrophic failure at the device level can occur, because the impact of the rapid temperature rise is limited to the device and its vicinity. As a result, while the package level cooling solutions can reduce the quasi steady-state/average temperature across a microprocessor or at length scales in the order of one millimeter, it has very little impact at micro/nanoscales. Basically, there is no practical way to reduce the temperature at the device and interconnect level by means of a cooling device or solution; therefore, the options for thermal engineering of these devices are very limited. However, intelligent electro-thermal design along with careful floor planning at the device level can largely reduce the temperature rise within a device. This means that the role of the thermal engineer is to properly anticipate – perhaps in full collaboration with electrical engineers – and prevent the problem at the early stages and at the device level, rather than to pass the problem to the package-level thermal engineers.

3. Early attempts to Modeling Heating Effects in State of the Art Devices

In order to understand the modeling of heating effects in commercial device simulators, we first derive the differential equation of heat conduction for a stationary, homogeneous, isotropic solid with heat generation within the body. Heat generation in general may be due to nuclear, electrical, chemical, or other sources that may be a function of time and/or position. The heat generation rate in the medium, generally specified as heat generation per unit time, per unit volume, is denoted H(r,t) and is given in W/m^3.
We consider the energy balance equation for a small control volume stated as: rate of heat entering through the bounding surfaces of V + rate of energy generation in V = rate of storage of energy in V. In other words the rate of heat entering through the bounding surfaces of V (term 1) is:

$$Term1 = -\int_A q \cdot n dA = -\int_V \nabla \cdot q dV \tag{5}$$

where A is the surface area of the volume element V, n is the outward-drawn normal unit vector to the surface element dA, q is the heat flux vector at dA; here the minus sign is included to ensure that the heat flow is into the volume element V, and the divergence theorem is used to convert the surface integral to volume integral. The remaining two terms are evaluated as:

$$Term2 = \text{Rate of energy generation in } V = \int_V H(r,t) dV \tag{6}$$

$$Term3 = \text{ rate of energy storage in } V = \int_V \rho C_V \frac{\partial T(r,t)}{\partial t} dV \tag{7}$$

Combining Eqs. (5) – (7) yields:

$$\int_V \left[-\nabla \cdot q(r,t) + H(r,t) - \rho C_V \frac{\partial T(r,t)}{\partial t} \right] dV = 0 \tag{8}$$

The last equation is derived for an arbitrary small volume element V within the solid, hence the volume V may be chosen so small as to remove the integral and one obtains:

$$-\nabla \cdot q(r,t) + H(r,t) = \rho C_V \frac{\partial T(r,t)}{\partial t} \tag{9}$$

Substituting q(r,t) from Eq. (1) into Eq. (9) finally yields the differential equation of heat conduction for a stationary, homogeneous, isotropic solid with heat generation within the body as:

$$-\nabla \cdot \left[\kappa \nabla T(r,t) \right] + H(r,t) = \rho C_V \frac{\partial T(r,t)}{\partial t} = c \frac{\partial T(r,t)}{\partial t} \tag{10}$$

The energy generation term in Eq. (10) is discussed in more details in Section 3.1 below.

3.1 Form of the Heat Source Term

Lai and Majumdar (Lai & Majumdar, 1996) developed a coupled electro-thermal model for studying thermal non-equilibrium in submicron silicon MOSFETs. Their results showed that the highest electron and lattice temperatures occur under the drain side of the gate electrode, which, also corresponds to the region where non-equilibrium effects such as impact ionization and velocity overshoot are maximum. Majumdar *et al*. (Majumdar *et al*., 1995) have analyzed the variation of hot electrons and associated hot phonon effects in GaAs MESFETs. These hot carriers were observed to decrease the output drain current by as much as 15%. Thus, they concluded that both electron and lattice heating should be included in the electrical behavior of devices.

As it has been recognized that the simulation of devices operated under non-isothermal conditions was of growing importance, the heat flow equation given in Eq. (10) has been added to conventional drift-diffusion and/or hydrodynamic models to account for the mobility degradation due to lattice heating. There has been a discussion on the form of the heat generation term, details of which can be found in an excellent paper by Wachutka (Wachutka, 1990). Briefly, three different models are most commonly used and these include: (1) Joule Heating, (2) electron-lattice scattering and (3) the phonon model. Although these three models yield identical results in equilibrium, under non-equilibrium conditions the results of the three models can vary significantly.

Case 1: Within the Joule heating model, the thermal model consists of the heat diffusion equation using a Joule heating term as the source. The source term is computed from the electrical solution as the product of the local field and the current density (Gaur &,Navon, 1976)

$$H = J \bullet E \tag{11}$$

This source term is similar to the one used by Leung and co-workers (Leung *et al*., 1997) and assumes that recombination heating is negligible. In this case the "hot spot" will occur near the location where the dot product of the field and of the current density is the largest. Simulations that have used this expression as a heating term suggest that the bulk of the heating will occur directly under the gate region where most of the voltage drop occurs and where the current density is the largest because of the restricted electron flow path due to the depletion regions. A study by Raman and co-workers on lightly doped drain (LDD) devices suggested that the location of the hot-spot occurs at the drain side of the gate. The complete Leung and co-workers expression used for the source term is

$$H = J \bullet E + (R - G)(E_G + 3k_B T) \tag{12}$$

where the second term represents the heating rate due to non-radiative generation (G) and recombination (R) of electron-hole pairs. E_G is the semiconductor band-gap, k_B is the Boltzmann constant and T is the lattice temperature.

Case 2: Within the electron-lattice scattering model, the thermal system is represented as a single lattice temperature and is considered to be in thermal equilibrium. However, since the heat generation is due to non-equilibrium electron temperatures, the source term is then taken as a scattering term obtained from the relaxation time approximation and moments of the BTE. In essence, the transport is similar to case 1 in that the heat diffusion equation governs transport in the solid, except for the fact that the source term is now given as a moment of the relaxation time approximation, i.e.

$$H = \frac{3\rho k_B}{2}\left(\frac{T_e - T_L}{\tau_{e-L}}\right) \tag{13}$$

In the above expression T_e is the electron temperature, T_L is the lattice temperature and $\tau_{e\text{-}L}$ is the electron lattice scattering time constant.

Case 3: Phonon-model. Under thermal non-equilibrium conditions a system of two phonons is used as represented later in the text. In this case, the 'lattice' temperature is taken to be the acoustic phonon temperature T_A, because this is the mode responsible for diffusion. The energy balance equations for the acoustic and optical modes were for the first time derived by Majumdar and co-workers starting from the phonons Boltzmann transport equation. In all our investigations we have pursued this approach for the description of the phonon bath. A variant of this approach, that has been pursued by the Leeds group (Sadi *et al*., 2007) and by Eric Pop and co-workers [Pop, 2006], counts the number of generated acoustic and optical phonons in a given branch and mode. Then, the total heat generation rate per unit volume is computed as

$$H = \frac{n}{N_{sim}\Delta t}\sum\left(\hbar\omega_{ems} - \hbar\omega_{abs}\right) \tag{14}$$

where n is the electron density, N_{sim} is the number of simulated particles and Δt is the simulation time.
In our opinion, the best solution to the submicron heat transport problem only has been provided by Narumanchi and co-workers (Naramunchi, 2004). In their recent study they propose a model based on the solution of the BTE, accounting for the transverse acoustic and longitudinal acoustic as well as optical phonons. Their model incorporates realistic phonon dispersion curves for silicon. The interactions among the different phonon branches and different phonon frequencies are considered and the proposed model satisfies energy conservation. Frequency-dependent relaxation times, obtained from perturbation theory, and accounting for phonon interaction rules, are used. In the calculation of the relaxation rates, they have included impurity scattering and the three-phonon interactions (the normall (N) process and the Umklapp (U) process). U processes pose direct thermal resistance while N processes influence the thermal resistance by altering the phonon distribution function. In this study, the BTE is numerically solved using a structured finite volume approach. Using this model, experimental in-plane thermal conductivity data for silicon thin films over a wide range of temperatures are matched satisfactorily.

4. The ASU Approach to Solving Lattice Heating Problem in Nanoscale Devices

To properly treat heating without any approximations made in the problem at hand, one in principle has to solve the coupled Boltzmann transport equations for the electron and phonon systems together. More precisely, *one has to solve the coupled electron − optical phonons – acoustic phonons – heat bath problem*, where each sub-process involves different time scales and has to be addressed in a somewhat individual manner and included in the global picture via a self-consistent loop. Let us consider the coupled system of semi-classical Boltzmann transport equations for the distribution functions of electrons $f(\mathbf{k},\mathbf{r},t)$ and phonons $g(\mathbf{k},\mathbf{r},t)$:

$$\begin{aligned}\left(\frac{\partial}{\partial t}+v_e(\mathrm{k})\cdot\nabla_r+\frac{e}{\hbar}E(\mathrm{r})\cdot\nabla_k\right)f&=\sum_{\mathrm{q}}\left\{W_{e,\mathrm{q}}^{\mathrm{k+q}\to\mathrm{k}}+W_{a,-\mathrm{q}}^{\mathrm{k+q}\to\mathrm{k}}-W_{e,-\mathrm{q}}^{\mathrm{k}\to\mathrm{k+q}}-W_{a,\mathrm{q}}^{\mathrm{k}\to\mathrm{k+q}}\right\}\\ \left(\frac{\partial}{\partial t}+v_p(q)\cdot\nabla_r\right)g&=\sum_{\mathrm{k}}\left\{W_{e,\mathrm{q}}^{\mathrm{k+q}\to\mathrm{k}}-W_{a,\mathrm{q}}^{\mathrm{k}\to\mathrm{k+q}}\right\}+\left(\frac{\partial g}{\partial t}\right)_{p-p}\end{aligned} \tag{16}$$

Here $W_{e,\mathrm{q}}^{\mathrm{k+q}\to\mathrm{k}}$ is the probability for electron transition from k+q to k due to emission of phonon q. Similarly $W_{a,\mathrm{q}}^{\mathrm{k+q}\to\mathrm{k}}$ refers to processes of absorption. The system is nonlinear, as the probabilities W depend on the product $f \bullet g$ of the electron and phonon distribution functions. The transfer of energy between electrons and phonons is due to the terms W, with a timescale of the order of 0.1 ps (see Fig. 5 below for more details). This equation set poses a multi-scale problem since the left hand sides involve different time scales: the velocity v_p of the phonons is two orders of magnitude lower than the velocity v_e of the electrons. Accordingly, the heat transfer by the lattice is much slower process than the charge transfer.
Note that, when considering the electron lattice coupling, the energy transfer from electrons to the high-energy optical phonons is very efficient. However, optical phonons possess negligible group velocity and, thus, do not participate significantly in the heat diffusion.

They instead must transfer their energy to acoustic phonons, which diffuse heat. The energy transfer between phonons is relatively slow compared to electron-optical phonon transport and, thus, thermal non-equilibrium may also exist between optical and acoustic phonons. Fig. 5 shows the primary path of thermal energy transport and the associated time constants.

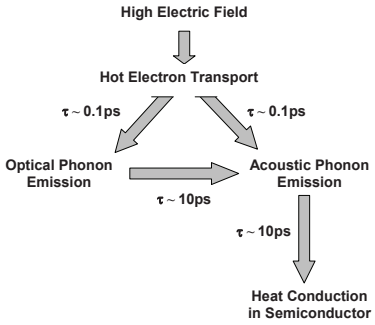


Fig. 5. The most likely path between energy carrying particles in a semiconductor device is shown together with the corresponding scattering time constants.

According to Fig. 5, the primary path of energy transport is represented first by scattering between electrons and optical phonons (T_{LO}) and then optical phonons to the lattice (T_A) (Tien *et al*., 1998). The BTE for the two kinds of phonons is then used to provide the energy balance of the process. As we already mentioned, the direct solution of the phonon Boltzmann equation itself is a very difficult task as it is difficult to mathematically express the anharmonic phonon decay process, and in addition to this one has to solve separate phonon Boltzmann equation for each mode of the acoustic and optical branches. As already noted in Section 2 above, some attempts to solve the problem within the relaxation time approximation have been made by Narumanchi and co-workers (Naramunchi *et al*., 2004). If we now include the electrons and holes into the picture with their corresponding Boltzmann transport equations, then the solution of the electron-hole-phonon coupled set of equations becomes a formidable task even for today's high performance computing systems. Therefore, some simplifications of this global problem are needed. Since for device simulation we are mainly focused on accurately calculating the *IV*-characteristics of a device, the self-heating being a by-product of the current flow through the device can be treated in a more approximate manner, but which is still more accurate than the local heat conduction model. Majumder and co-workers (Lai & Majumdar, 1996; Majumdar *et al*., 1995), starting from the phonon Boltzmann equations derived energy balance equations separately for the optical phonon and the acoustic phonon bath. One can also derive these energy balance equations starting from the energy conservation principle. For electric field, $E \geq 10^6$ V/m, electrons lose energy to optical phonons and optical phonons decay to acoustic phonons. Using the first law of thermodynamics, the energy conservation equations for optical and acoustic phonons are:

$$\frac{\partial W_{LO}}{\partial t} = \left(\frac{\partial W_e}{\partial t}\right)_{coll} - \left(\frac{\partial W_{LO}}{\partial t}\right)_{coll}, \tag{17}$$

$$\frac{\partial W_A}{\partial t} = \nabla(\kappa_A \nabla T_A) + \left(\frac{\partial W_{LO}}{\partial t}\right)_{coll}, \tag{18}$$

W_e, W_{LO} and W_A are electron, optical phonon and acoustic phonon energy densities, respectively. We next use

$$dW_{LO} = C_{LO} dT_{LO} \text{ and } dW_A = C_A dT_A \tag{19}$$

where C_{LO} (specific heat capacity for optical phonons) can be estimated from Einstein model, while C_A (specific heat capacity for acoustic phonons) from Debye model. Next, the collision terms are expressed using the relaxation time approximation (RTA):

$$\left(\frac{\partial W_e}{\partial t}\right)_{coll} = n \cdot \frac{\frac{3}{2}k_B T_e + \frac{1}{2}m^* {v_d}^2 - \frac{3}{2}k_B T_{ph}}{\tau_{e-ph}}, \tag{20}$$

$$\left(\frac{\partial W_{LO}}{\partial t}\right)_{coll} = C_{LO}\frac{T_{LO} - T_A}{\tau_{LO-A}}, \tag{21}$$

where T_e is the electron temperature, v_d is the electron drift velocity and T_{ph} can be optical or acoustic phonon temperature depending on which kind of phonons electrons interact with. Combining Eqs. (17)-(21), one gets:

$$C_{LO}\frac{\partial T_{LO}}{\partial t} = \frac{3}{2}nk_B\left(\frac{T_e - T_{LO}}{\tau_{e-LO}}\right) + \frac{nm^*{v_d}^2}{2\tau_{e-LO}} - C_{LO}\left(\frac{T_{LO} - T_A}{\tau_{LO-A}}\right) \tag{22a}$$

$$C_A\frac{\partial T_A}{\partial t} = \nabla(\kappa_A \nabla T_A) + C_{LO}\left(\frac{T_{LO} - T_A}{\tau_{LO-A}}\right). \tag{22b}$$

For electric field, E<10^6 V/m, electrons lose energy directly to acoustic phonons and in that case:

$$C_A\frac{\partial T_A}{\partial t} = \nabla(\kappa_A \nabla T_A) + \left(\frac{\partial W_e}{\partial t}\right)_{coll} \tag{23a}$$

$$C_A\frac{\partial T_A}{\partial t} = \nabla(\kappa_A \nabla T_A) - \frac{3}{2}\frac{nk_B}{\tau_{e-A}}T_A + n \cdot \frac{\frac{3}{2}k_B T_e + \frac{1}{2}m^*{v_d}^2}{\tau_{e-A}} \tag{23b}$$

Under the assumption of very low electric fields, the electron temperature and acoustic phonon temperature equal the lattice temperature and terms 2 and 3 in Eq. (23) cancel. Using the low field conductivity and the mobility expressions we get that the heat source term reduces to the last term of Eq. (23b). Namely

$$q_{gen} = J \bullet E = \sigma E^2 = \frac{\sigma v_d^2}{\mu^2} = \frac{nm^* v_d^2}{\tau} \tag{24}$$

where it is assumed that for low doping concentrations the relaxation time τ that appears in Eq. (24) is the acoustic phonon relaxation time since acoustic phonon scattering, being isotropic scattering process, is most effective in randomizing the carrier momentum when the carrier energy is very low (low applied electric fields). In our electro-thermal simulator, we solve steady-state versions of the Eqs. (22a) and (22b) for the optical and acoustic phonon temperatures self consistently with the Monte Carlo simulation of the electron Boltzmann transport equation for modeling *n*-channel fully-depleted (FD) silicon on insulator (SOI) devices.

In other words, to simplify the calculations to a tractable form in the present work, we have used a Chapman-Enskog type expansion (Cercignani, 1988) to replace the microscopic phonon transport equation by a diffusion problem for the local density and energy of the phonons, where the diffusion coefficients are dependent on the state of the electrons. This method involves computation of the phonon energy dependent scattering tables in the Ensemble Monte Carlo (EMC) code for the electron transport, which already represents a big improvement over the current state of the art, where, as already discussed, the coupling of thermal and charge effects is strictly one way. This approach also takes care of the multi-scale nature of the problem, assuming a quasi steady state: the phonons are in steady state, albeit with a spatially dependent temperature distribution. In the present work we neglect the effect of the optical phonon dispersion on the tabulated scattering rates.

Using our electro-thermal simulator we demonstrate that, contrary to earlier predictions, the non-locality of the electron transport and velocity overshoot effects significantly reduce current degradation in nanoscale devices due to the fact that carriers traveling ballistically do not dissipate energy in the active region of the device, rather in the drain contact. We also show that in nanoscale devices, the hotspot corresponding to the peak temperature moves towards the drain end of the channel where removal of heat is more effective, and where less effect on the transport dynamics occurs underneath the gate.

4.1 Electro-Thermal Particle-Based Device Simulator Description

As illustrated in Fig. 6, and discussed in details in Ref. (Vasileska *et al*., 2009), we self-consistently couple the Monte Carlo solution of the electron Boltzmann transport equation with the energy balance equations for both optical and acoustic phonons. This itself is a difficult problem as we are coupling a particle picture for electrons (which is inherently noisy) with a continuum model for the phonons. To achieve convergence of the coupled scheme, both temporal and spatial averaging of the variables extracted from the Monte Carlo (MC) solver (e.g. the electron density, drift velocity and temperature) must be performed. The number of simulated particles in the model contributes significantly to the smoothness of the variables being transferred to the energy balance solver. Within each "outer iteration" we solve the Boltzmann Transport equation for the electrons using the Ensemble Monte Carlo (EMC) method for a time period of 10 ps to ensure that steady state conditions have been achieved. The required variables are then passed to the thermal solver which gives the updated optical and acoustic phonon temperatures. This constitutes one Gummel cycle or "one outer iteration" (Raleva et al., 2008; Sridhaharan et al., 2008).

In our research effort, the EMC code for the carrier BTE solution (He, 1999; Ahmed, 2005) has been modified as well. As we have variable lattice temperature in the hot-spot regions, we have introduced the concept of *temperature dependent scattering tables*. For each combination of acoustic and optical phonon temperature, one energy dependent scattering

table is created. These scattering tables involve additional steps in the MC phase (Fig. 7 – right panel), because to choose randomly a scattering mechanism for a given electron energy, it is necessary to find the corresponding scattering table. To do that, first, the electron position on the grid needs to be found, in order to know the acoustic and optical phonon temperatures in that grid point, and then the scattering table with "coordinates" (T_L,T_{LO}) is selected. Using current state of the art computers, the pre-calculation of these scattering tables does not require much CPU time or memory resources and is done once in the initialization stages of the simulation for a range of temperatures. An interpolation scheme is then adopted afterwards for temperatures for which we do not have the appropriate scattering table.

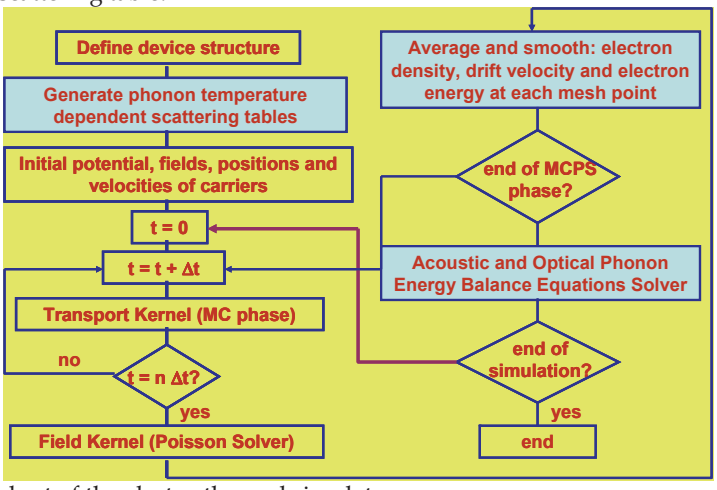


Fig. 6. Flow-chart of the electro-thermal simulator.

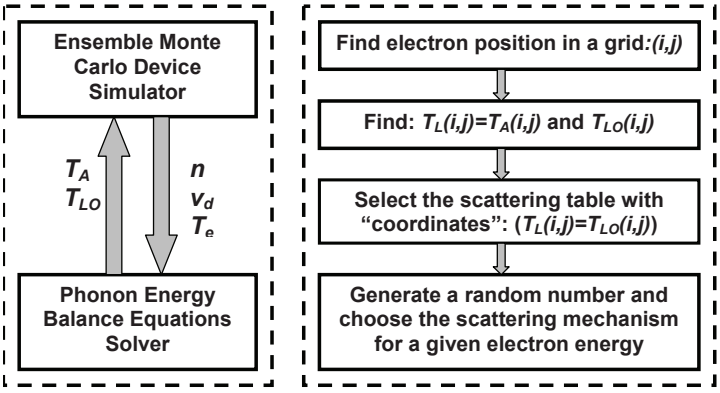


Fig. 7. Left panel – Exchange of variables between the two kernels. Right panel: Choice of the proper scattering table.

To properly connect the particle-based picture of electron transport with continuous, "fluid-like" phonon energy balance equations, a space-time averaging and smoothing of electron density, drift velocity and electron energy are included. At the end of each MC time step, the electrons are assigned to the nearest grid point. Then, the drift velocities and thermal energies are averaged with the number of electrons at the corresponding grid points. After the MC phase, a time averaging of electron density, drift velocity and thermal energy is done and the electron temperature distribution is calculated. It is assumed that the drift energy is much smaller than the thermal energy. The smoothing of these variables is necessary, because most of the grid points, especially at the interfaces, are rarely populated with electrons. This leads to very low lattice temperatures in those points. The exchange of variables between electron and phonon solvers is shown on the left panel of Fig. 7.

Because SOI devices consist of two distinct regions, the silicon device layer and the buried oxide layer (in which the phonons have significantly smaller mean-free paths), the phonon BTE is solved in the silicon layer to accurately model heat transport, but the simpler heat diffusion equation is used in the amorphous BOX because the characteristic length-scale of conduction is much smaller than the film thickness. The two distinct computational regions are coupled through interface conditions that accounts for differences in material properties. For the coupling of the silicon and oxide solution domains, it is necessary to calculate the flux of energy through the interface between the two materials at each point along the interface for every time step.

To simulate the steady-state state behavior of a device, the system is started in some initial condition, with the desired potential applied to the contacts, and then the simulation proceeds in a time stepping manner until steady-state is reached. A common starting point for the initial guess is to start out with charge neutrality, i.e., to assign particles randomly according to the doping profile in the device, so that initially the system is charge neutral on the average. After assigning charges randomly in the device structure, charge is then assigned to each mesh point using an adequate particle mesh (PM) coupling method, and Poisson's equation is solved. The forces are then interpolated on the grid, and particles are accelerated over the next time step. When the system is driven into a steady-state regime and MC simulation time has elapsed, we calculate the steady-state current through a specified terminal.

To continue with the thermal part of the simulation, the average electron density, drift velocity and electron temperature must be calculated on a grid. For the given bias conditions, the average electron density in the channel (Fig. 8) is very high at the Si-SiO_2 interface near the source injection barrier, while pinch off region exists near the drain.

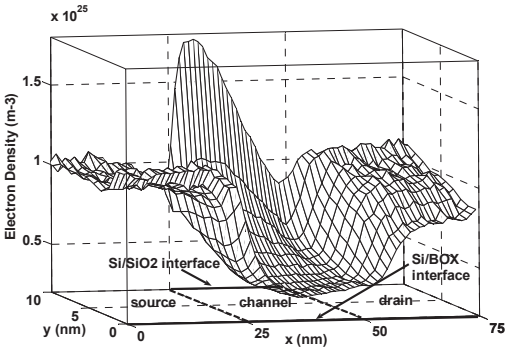


Fig. 8. Sample electron density.

The two components of the electron kinetic energy are presented in Fig. 9. They show that the thermal part of electron kinetic energy is much larger than the drift part, so electron temperature can be calculated from the thermal energy.

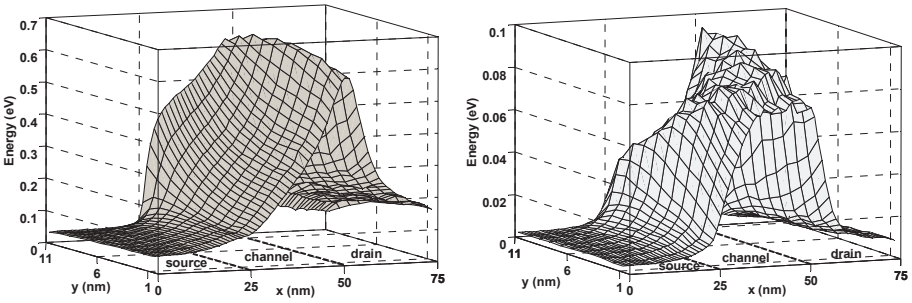


Fig. 9. Left panel: electron thermal energy. Right panel: electron drift energy.

Finally the acoustic and optical phonon temperatures are calculated with the phonon energy balance equation solver. During the simulation, the gate contact and the bottom of the BOX are set to 300 K, while Neumann boundary conditions for the heat transfer are used in all other outer surfaces. Also, the tolerance used in the "thermal" successive over-relaxation (SOR) algorithm equals 0.001, which leads to very fast convergence.

When the simulation starts, all variables obtained from the first iteration of the EMC solver, are calculated using uniform distribution for the acoustic and optical phonon temperatures. This means that only one scattering table is used for all electrons, no matter where they are located in the device. When the phonon temperatures are computed from the phonon energy balance equations, they are "returned" at the beginning of the MC free-flight – scattering phase. Now, for each mesh point, we have a scattering table which corresponds to the acoustic and optical phonon temperatures at that point. In this case, the electron position defines which scattering table is "valid "and then, by generating a random number, the scattering mechanism is chosen for the given electron energy. The impact of the phonon temperature dependent scattering tables can be demonstrated by counting the number of

energy-exchange electron-phonon scattering events. From the results obtained it could be concluded that the inclusion of the phonon temperature dependent scattering table increases the number of electron-phonon interactions. We have also found that electrons with energies below 50 meV scatter mainly with acoustic phonons in silicon, while those with higher energy scatter strongly with the optical modes. The optical phonon modes have low group velocity (on the order of 1000 m/s) and their occupation is also relatively low, hence they contribute very little to the heat transport. The primary heat carriers in silicon are the faster acoustic phonon modes, which are significantly populated and have group velocities from 5000 m/s for transverse modes to 9000 m/s for longitudinal acoustic modes. Optical phonons decay into acoustic modes, but over relatively long time scales, i.e. picoseconds, compared to the order of tenths of picoseconds (Ferry, 2000). Under high field conditions, this can lead to the creation of a phonon energy bottleneck which can cause the density of optical phonon modes to build up over time, leading to more scattering events and impeding electron transport (Artaki & Price, 1989).

To test the overall convergence of the coupled EMC and thermal codes, we registered the variations of the drain current with the number of thermal iterations for a given bias condition. We have simulated 25nm fully-depleted SOI structure (Fig. 10 left panel). The device geometrical dimensions are: channel length = 25 nm, source/drain length=25 nm, Si layer thickness = 10 nm and BOX thickness = 50 nm. The results of simulations given in the right panel of Fig. 10, and show that only 3-5 iterations are necessary to obtain the steady-state solutions of the current.

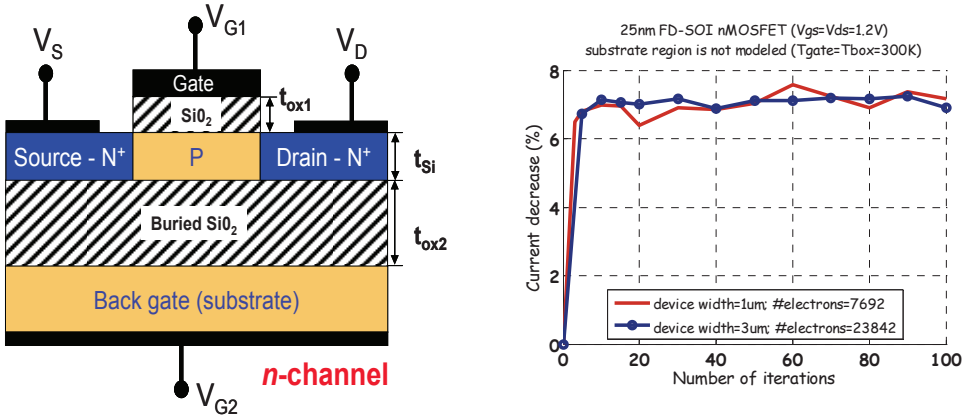


Fig. 10. Left panel - Cross-section of the simulated devices. Right panel - Current decrease variations with the number of thermal iterations for different number of simulated electrons for 25nm fully-depleted SOI device. For the thermal part of simulations, the bottom of the BOX and the gate electrode are assumed to be isothermal boundaries, set to 300K. Substrate region is not modeled.

To further investigate the accuracy of the results and the convergence of the outer current iterative scheme, we increase the device width in order to increase the number of simulated electrons. Namely, the statistical error of the MC method shrinks as the number of simulated carriers, N_{sim}, increases, and it drops as $1/(N_{sim})^{1/2}$ (Ferry, 2000). The statistical

uncertainty of the results is 1.14% when device width is 1 Mm and 0.65% for 3 Mm device width, so the smoother convergence of the results is obtained with the larger number of simulated carriers (see Fig. 10 - right panel).

Simulation results for the isothermal and non-isothermal current are summarized in Table 1. It is obvious that SOI structure has larger current degradation, since SiO_2 has very poor thermal conductivity compared to diamond and AlN. But, even though diamond has over 6 times larger thermal conductivity than AlN, due to the differences in the dielectric constants, the current degradation is slightly higher in Si on AlN than in Si on Diamond. This is explained in more details in Ref. (Vasileska, et al., 2009) where we discuss the interplay of the dielectric constant of the BOX, the thickness of the BOX and the thermal conductivity of the BOX on the overall magnitude of the current and the current degradation for 25 nm channel length devices.

The left panel of Fig. 11 shows the effect of lattice heating on the *I-V* characteristics of a 25 nm gate length fully depleted SOI structure. The velocity overshoot is clearly seen on the right panel of Fig. 11. For this particular device structure, the corresponding degradation of device characteristics due to thermal effects is relatively small, less than 10%. As seen from the temperature maps of acoustic and optical phonons shown in Fig. 12, the maximum rise in the lattice temperature is on the order of 100 °C on the drain side of the gate as expected. For comparison, we also compare in Fig. 12 the effect of an elevated lattice temperature of 400K assuming an isothermal model compared to the non-uniform model, which shows that most of the effect observed in the *I-V* characteristic is due to lattice heating.

Device	Device width = 1μm			Device width = 3μm		
	Current (mA/um) isothermal	Average current (mA/um) thermal	Current decrease (%)	Current (mA/um) isothermal	Average current (mA/um) thermal	Current decrease (%)
FD-SOI	1.7706	1.6464	7.01	1.8238	1.6952	7.05
FD-SOAIN	1.7986	1.7669	1.76	1.8494	1.8201	1.58
FD-SOD	1.7884	1.7646	1.33	1.8398	1.8139	1.41

Table 1. Comparison of current degradation for different device technologies.

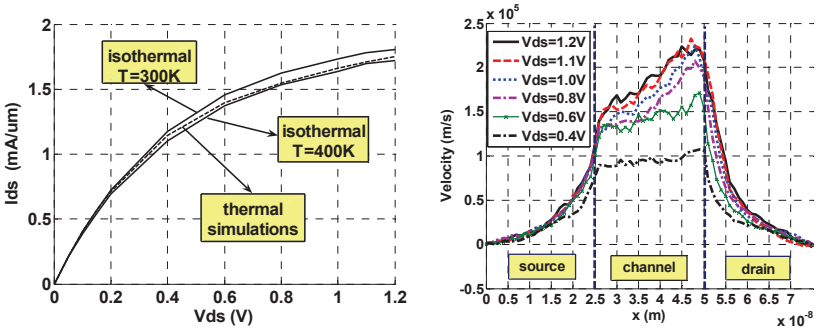


Fig. 11. Left panel: Output characteristics for V_G=1.2 V. Top/middle curve correspond to the case of excluded/included lattice heating model. Bottom curve is an isothermal simulation but for lattice temperature T =400K throughout the whole simulation domain. Right panel:

Velocity along the channel for V_G=1.2 V and different values of V_D. Note that for V_D>0.4 V electrons are in the velocity overshoot regime which suggests that lattice heating does not significantly degrade the device characteristics.

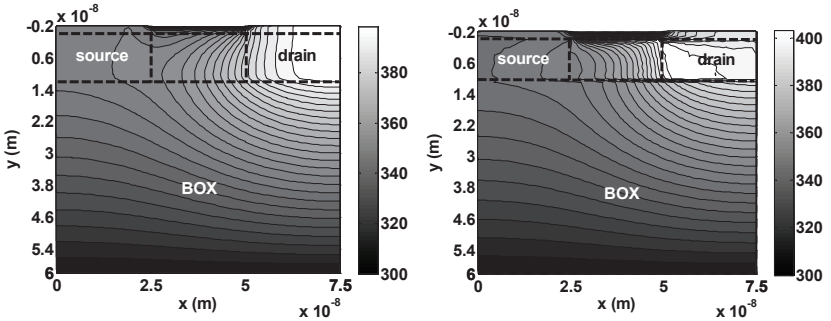


Fig. 12. Left panel - Acoustic phonon temperature for V_G=1.2 and V_D=1 V. Notice the significant heating of the lattice that equals the acoustic phonon temperature in our model. Right panel - Optical phonon temperature for V_G=1.2 V and V_D=1 V. Notice the region near the drain with higher optical phonon temperature with respect to the acoustic phonon temperature.

4.2 Thermal degradation with scaling of device geometry

In addition to the previously noted observation regarding the influence of the velocity overshoot, we modeled larger fully-depleted SOI device structures and we also investigated the influence of the temperature boundary condition on the gate electrode on the current degradation due to heating effects. The geometrical dimensions of the simulated fully-depleted SOI MOSFETs are given in Table 2, while Table 3 gives the percentage of the current decrease due to heating effects with the variation of the gate electrode temperature for the device structures being considered. The calculated results show that the current degradation is more prominent for larger devices and for higher gate temperatures. For 80 nm and larger devices, simulated carriers are not in the velocity overshoot regime in the larger portion of the channel (especially near the source end of the channel). Snapshots of the lattice temperature profiles in the silicon layer for these devices when the gate temperature is set to 300K and 400K, are given in Fig. 13 and 14, respectively. From these snapshots one can observe that: (a) the temperature in the channel is increasing with the increase of the channel length, (b) the maximum lattice temperature region (hot spot) is in the drain and it shifts towards the channel for larger devices. This behavior is more drastic for higher gate temperatures (see Fig. 14).

Gate Length (nm)	Gate Oxide Thickness (nm)	Active Si-Layer Thickness (nm)	BOX Thickness (nm)	Channel Doping Concentration (cm^{-3})	Bias conditions $V_{GS}=V_{DS}$ (V)	Current (mA/um) Isothermal value (300K)
25	2	10	50	1×10^{18}	1.2	1.82
45	2	18	60	1×10^{18}	1.2	1.41
60	2	24	80	1×10^{18}	1.2	1.14
80	2	32	100	1×10^{17}	1.5	1.78
90	2	36	120	1×10^{17}	1.5	1.67
100	2	40	140	1×10^{17}	1.5	1.57
120	3	48	160	1×10^{17}	1.8	1.37
140	3	56	180	1×10^{17}	1.8	1.23
180	3	72	200	1×10^{17}	1.8	1.03

Table 2. Geometrical dimensions of the simulated fully-depleted SOI MOSFETs

Fig. 15 gives the ensemble averaged lattice and optical phonon temperatures along the channel in the silicon layer only for three technologies of devices being considered (25 nm, 80 nm and 180 nm). Notice that there is a bottleneck between the lattice and the optical phonon temperature in the channel which is more pronounced for shorter devices, due to the fact that the energy transfer between optical and acoustic phonons is relatively slow compared to the electron-optical phonon processes and the fact that the electrons are in the velocity overshoot (and since the channel is very short, they spent little time in the channel). To better understand the phonon temperature bottleneck, different cross-sections (at Si/SiO_2 interface, at half Si-layer width, at Si/BOX interface) of the lattice and the optical phonon temperature profiles in the channel direction were investigated as well. We find that the bottleneck is decreasing from Si/SiO_2 interface to Si/BOX interface. For shorter devices, it exists in the whole channel region, which is not a case for longer devices (thicker Si-layer and longer channel length). From the results we have presented here one can conclude that the higher the temperature in the channel and/or the longer the electrons are in the channel, the larger the degradation of the device electrical characteristics is due to the heating effects.

Type of simulation	Gate Electrode Temperature (K)	25 nm FD SOI nMOSFET Vgs=Vds=1.2V		45 nm FD SOI nMOSFET Vgs=Vds=1.2V		60nm FD SOI nMOSFET Vgs=Vds=1.2V	
		Current (mA/um)	Current Decrease (%)	Current (mA/um)	Current Decrease (%)	Current (mA/um)	Current Decrease (%)
isothermal	300	1.929	\	1.5645	\	1.2457	\
thermal	300	1.8176	5.78	1.4717	5.93	1.1885	4.73
thermal	400	1.7467	9.45	1.3957	10.79	1.1124	10.83
thermal	600	1.5997	17.1	1.2612	19.39	0.9882	20.79

Type of simulation	Gate Electrode Temperature (K)	80 nm FD SOI nMOSFET Vgs=Vds=1.5V		90 nm FD SOI nMOSFET Vgs=Vds=1.5V		100 nm FD SOI nMOSFET Vgs=Vds=1.5V	
		Current (mA/um)	Current Decrease (%)	Current (mA/um)	Current Decrease (%)	Current (mA/um)	Current Decrease (%)
isothermal	300	1.7810	\	1.6671	\	1.5743	\
thermal	300	1.5850	11	1.4805	11.19	1.4004	11.05
thermal	400	1.4612	17.96	1.3678	17.95	1.2833	18.48
thermal	600	1.2571	29.42	1.1784	29.31	1.1158	29.12

Type of simulation	Gate Electrode Temperature (K)	120 nm FD SOI nMOSFET Vgs=Vds=1.8V		140 nm FD SOI nMOSFET Vgs=Vds=1.8V		**180 nm FD SOI nMOSFET Vgs=Vds=1.8V**	
		Current (mA/um)	Current Decrease (%)	Current (mA/um)	Current Decrease (%)	Current (mA/um)	Current Decrease (%)
isothermal	300	1.3657	\	1.2341	\	1.0321	\
thermal	300	1.1439	16.34	1.0416	15.6	0.8570	15.22
thermal	400	1.0630	22.16	0.9654	21.85	0.8100	21.52
thermal	600	0.9256	32.22	0.8311	32.27	0.6940	32.76

Table 3. Current decrease with gate temperature for different FD SOI technologies.

5. Open Problems

From the discussion presented so far, we might conclude that device simulation without lattice heating has reached high levels of sophistication even in the quantum domain area. The inclusion of lattice heating is done using the energy balance picture by the group from Arizona State University or by including the heat flux term that accounts for the number of generated acoustic and optical phonons in a given mode.

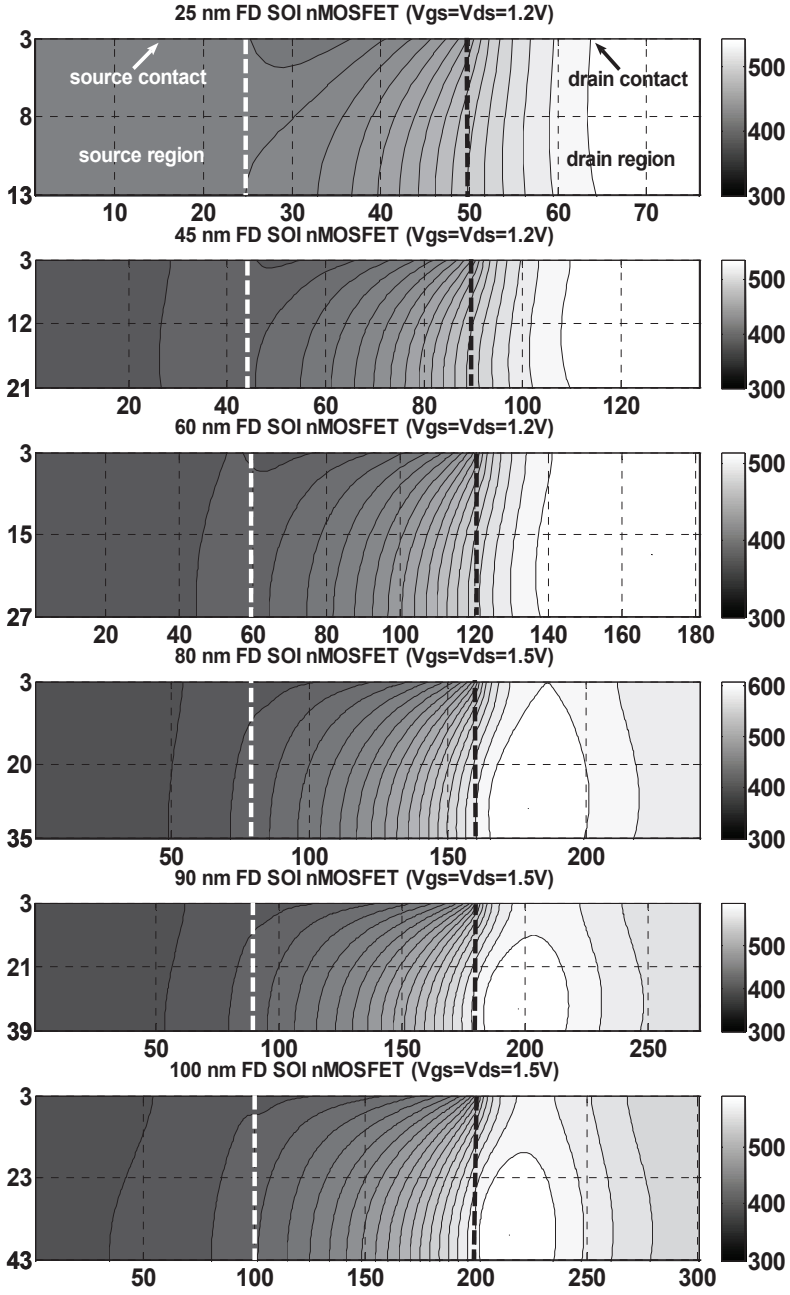


Fig. 13. Lattice temperature profiles in the silicon layer for FD SOI MOSFETs from Table 2 with gate temperature set to 300K. (25 nm –top, 100 nm-bottom).

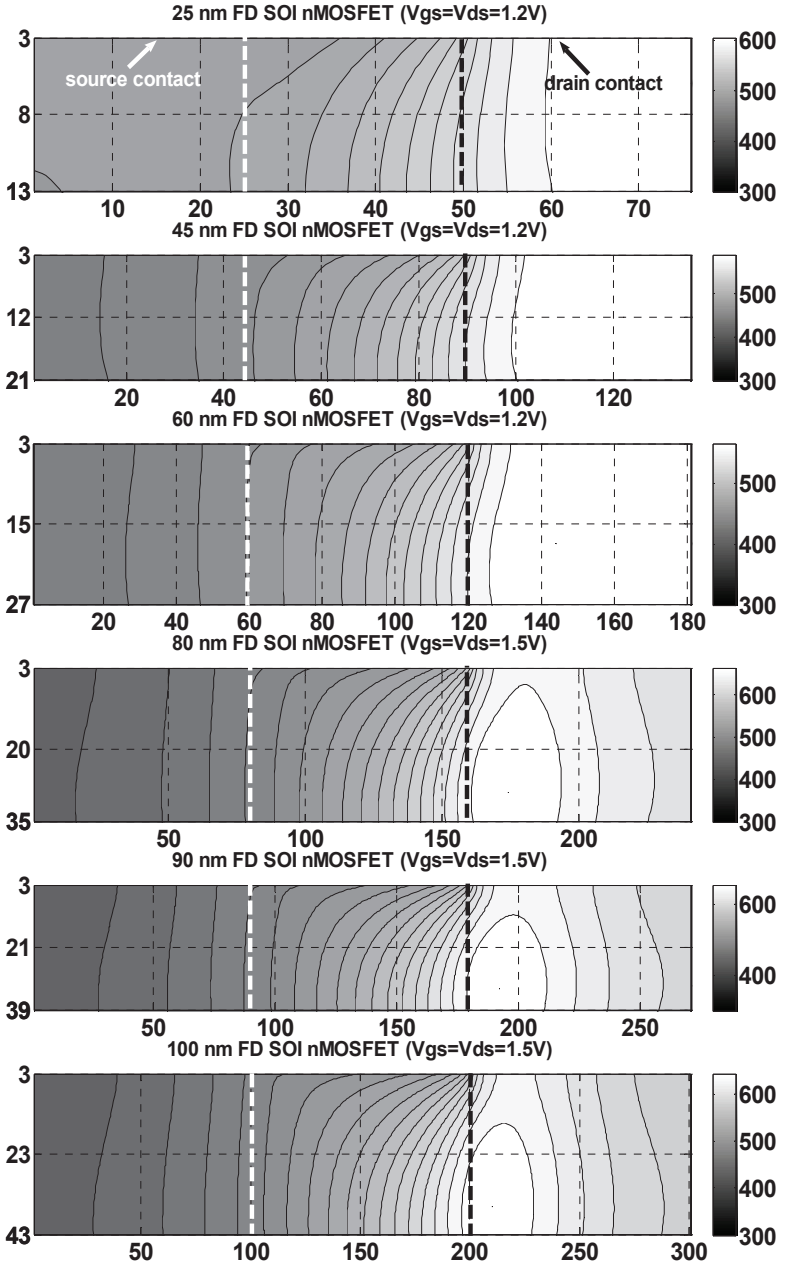


Fig. 14. Lattice temperature profile in silicon layer for different device geometries of fully-depleted SOI MOSFET when gate temperature is set to 400K.

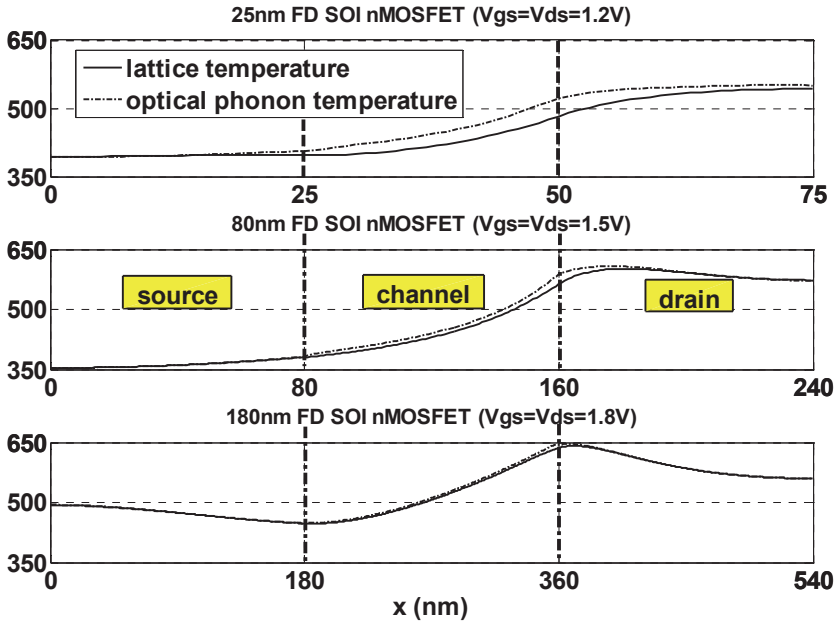


Fig. 15. Averaged lattice(full) and optical phonon(dashed) temperature profiles in the channel direction in the silicon layer for 25 nm (top), 80nm (middle) and 180 nm (bottom) FD SOI n-channel MOSFETs with gate temperature set to 300K.

Both of these approaches have their advantages and disadvantages. Important point is that understanding heat conduction surrounding nanostructures is just at the beginning. Many questions remain to be answered and new applications need to be explored. Some of these are discussed below.

1. The experimental studies carried out by Goodson's group provide the first quantitative evidence of non-local phonon heat conduction effects surrounding nanostructures. The effects are more pronounced at low temperatures, as expected (Sverdrup et al. 2001). In their experiments, the heater (also serving as a temperature sensor) is defined by a *pn*- junction. Because the calibration is through measuring the resistance of the heater when the sample temperature is uniformly raised, there are lingering questions on how accurate it is to use such calibration data to infer the heater temperature due to the spatial temperature gradient that may overlap with the space charge region of the *pn*-junction. More experimental studies are needed to provide further experimental data for analysis and to push to even smaller structures.

2. When the heat conduction is nonlocal, the transport is highly non-equilibrium, the temperature used to represent the modeling results is at best a measure of the local energy density, rather than their typical thermodynamic meaning. On the other hand, in microelectronics, the device reliability is often associated with the temperature through the Arrehnius law, which is a manifestation of the Boltzmann distribution and is a result obtained under the assumption of local equilibrium. A valid question is what does this temperature means for a device. If, for example, the calculated temperature corresponding

to the local energy density reaches the melting point, does it mean that local melting occur? The molecular dynamics simulations may shed some light on this question, as is discussed in Goodson et al.'s paper [Asheghi *et al*., 1998].

3. The simulations so far are based on either Monte Carlo methods or the Boltzmann equation and take the various relaxation times as input parameters. These parameters are subject to a wide range of uncertainties. The phonon relaxation time, for example, were mostly derived from the modeling work in the 1950s and 1960s on the thermal conductivity of bulk materials. The relaxation time in these models is often obtained under various approximations, such as the Debye model for the phonon dispersion. There are many variables in these past models that can lead to different values of the relaxation time. There is a clear need for more accurate information on the relaxation times. Molecular dynamics simulation may be one way to obtain them. Similarly, electron-phonon scattering processes also need further consideration, particularly when electrons have very different temperatures from that of phonons.

4. In addition to the different application problems in microelectronics and in data storage, there are also new fundamental problems associated with nonlocal heat conduction. Past studies have considered phonons only. The concurrent electron and phonon nonlocal transport and nonequilibrium transport is an example. A recent proposal is to use the departure from equilibrium between electrons and phonons at a small contact point as a means to increase the thermoelectric energy conversion efficiency (Goshal et al., 2002). The transport in this case can well become nonlocal for both electrons and phonons. This possibility was raised, but has never been studied (Chen, 1996).

5. The above examples emphasize the small length scales involved in nano-devices and nanomaterials. Short time scales are also becoming increasingly important. Similar questions can be raised for transport at short time scales as for the small length scales. Lasers can deliver a pulse as short as a few femtoseconds (1 fs = 10^{-15} s). Energy transduction mechanisms at such short time scales can differ significantly from that at macroscale. Microelectronic devices are pushing to the tens of gigaherts clock frequency with a much shorter transient time. The device temperature rise in such short time scales can be very different from predictions of the Fourier law.

6. References

Ahmed, S. S. (2005) . Ph.D. Thesis, Arizona State University, Advisor: Prof. D. Vasileska.

Artaki, M. & Price, P. J. (1989). Hot phonon effects in silicon field-effect-transistors, *J. Appl. Phys*. Vol. 65, 1989, pp. 1317-1320.

Asheghi, M.; Touzelbaev, M.N.; Goodson, K.E.; Leung, Y.K. & Wong, S.S. (1998). Temperature Dependent Thermal Conductivity of Single-Crystal Silicon Layers in SOI Substrates, *ASME Journal of Heat Transfer*, Vol.120, 1998, pp. 30-33.

Asheghi, M.; Touzelbaev, M. N.; Goodson, K. E.; Leung, Y. K.; & Wong, S. S. (1998). Temperature Dependent Thermal Conductivity of Single-Crystal Silicon Layers in SOI Substrates, *ASME Journal of Heat Transfer*, Vol.120, 1998, pp. 30-33.

Borkar, S. (1999). Design Challenges of Technology Scaling, *IEEE Micro*, Vol. 19, 1999, pp. 23-29.

Cercignani, C. (1998). *The Boltzmann Equation and its Applications*, Vol. 67 of *Applied Mathematical Sciences*, Springer-Verlag, 1988.

Chen, G. (1996). J.Heat Transf., Vol. 118, 1996, pp. 539- .

Chen, G. (2001). Ballistic-Diffusive Heat-Conduction Equations, *Physical Review Letters*, Vol. 86, 2001, pp. 2297-2300.

Chen, G. & Shakouri, A. (2002). Nanoengineered Structures for Solid-State Energy Conversion, *Transactions of the ASME*, Vol. 124, 2002, pp. 242-252.

Choi, S.-H.; Maruyama, S. (2003). Evaluation of the phonon mean free path in thin films by using classical molecular dynamics, *Journal of the Korean Physical Society*, Vol. 43, 2003, pp. 747-753.

Ferry, D. K. (2000). *Semiconductor Transport*, New York: Taylor & Francis, 2000.

Gaur, S. P. & Navon, D. H. (1976). Two-dimensional carrier flow in a transistor structure under non-isothermal conditions, *IEEE Trans. Electron Devices*, Vol. 23, 1976, pp. 50-57.

Geppert, L. (1999). Solid state [semiconductors. 1999 technology analysis and forecast], *Spectrum, IEEE*, Vol. 36, 1999, pp. 52-56.

Goshal, U. et al. (2002). , *Appl. Phys. Lett.*,Vol. 80, 1976,pp. 3006-

He, X. (1999). MS Thesis, Arizona State University, 1999, Advisor: Prof. D. Vasileska.

Ju, Y. S. & Goodson, K. E. (1999). Phonon Scattering in Silicon Films with Thickness of Order 100 nm, *Applied Physics Letters*, Vol. 74, 1999, pp. 3005-3007.

Lai, J. & Majumdar, A. (1996). Concurent thermal and electrical modeling of submicrometer silicon devices, *J. Appl. Phys.* , Vol. 79, 1996, pp. 7353- .

Leung, Y. K.; Paul, A. K.; Goodson,K. E.; Plummer, J. D. & Wong, S.S. (1997). Heating mechanisms of LDMOS and LIGBT in ultrathin SOI, *IEEE Electron Device Lett.*, Vol. 18, 1997, pp. 414-

Liu, W. & Asheghi, M. (2004). Phonon-Boundary Scattering in Ultra Thin Single-Crystal Silicon Layers, *Applied Physics Letters*, Vol. 84, 2004, pp. 3819-3821.

[a] Liu, W. & Asheghi, M. (2005). Thermal Conductivity of Ultra-Thin Single Crystal Silicon Layers, *Journal of Heat Transfer,* Vol. 128, 2005, pp. 75-83.

[b] Liu, W.; & Asheghi, M. (2005). *J. Appl. Phys.*, Vol. 98, 2005, 123523-1.

Majumdar, A. (1993). Microscale Heat Conduction in Dielectric Thin Films, *Journal of Heat Transfer*, Vol. 115, 1993, pp.7-16.

Majumdar, A.; Fushinobu, K. & Hijikata, K. (1995). Effect of hate voltage on hot electron and hot phonon interaction and transport in a submicrometer transistor, *J. Appl. Phys.* Vol. 77, 1995, pp. 6686- .

Narumanchi, S. V. J.; Murthy, J. Y. & Amon, C. H. (2004). Submicron heat transport model in silicon accounting for phonon dispersion and polarization, *Transactions of the ASME,* Vol. 126, 2004, pp. 946-955.

Palankovski, V. & Selberherr, S. (2001). Micro materials modeling in MINIMOS-NT, *Journal Microsystem Technologies*, Vol. 7, November 2001, pp. 183-187.

Pop, E.; Sinha, S. & Goodson, K. E. (2006) "Heat genareation and transport in nanometer-scale transistors", *Proceedings of the IEEE*, Vol. 94, 2006, pp. 1587-1601.

Raleva, K.; Vasileska, D.; Goodnick, S. M. & Dzekov T. (2008). Modeling thermal effects in nano-devices, *Journal of Computational Electronics, DOI 10.1007/s10825-008-0189-3 © Springer Science+Business Media LLC 2008, J. Computational Electronics,* Vol. 7, 2008, pp. 226-230.

Reif, F. (1985). *Fundamentals of Statistical and Thermal Physics* , McGraw-Hill, London, 1985.

Sadi, T. Kelsall, R.W. & Pilgrim, N. J. (2007) "Electrothermal Monte Carlo Simulation of Submicrometer Si/SiGe MODFETs", IEEE Trans. on Electron Devices, Vol.54, No.2, February 2007, pp. -

Silvaco Manual (www.silvaco.com).

Shakouri, A.; Lee, E. Y.; Smith, D. L.; Narayanamurti, V. & Bowers, J. E. (1998). Thermoelectric Effects in Submicron Heterostructure Barriers, *Microscale Thermophysical Engineering*, Vol. 2, 1998, pp. 37-47.

Sondheimer, E. H. (2001). The Mean Free Path of Electrons in Metals, Advances in Physics, Vol. 1, no. 1, Jan. 1952, reprinted in Advances in Physics, Vol. 50, 2001, pp. 499-537.

Sridharan, S.; Venkatachalam, A. & Yoder P. D. (2008). Electrothermal analysis of AlGaN/GaN high electron mobility transistors, *Journ. of Compt. Electronics* Vol. 7, 2008, 236–239.

Sverdrup, P. G.; Sinha, S.; Uma, S.; Asheghi, M. & . Goodson, K. E. (2001). *App. Phys. Let*. Vol. 78, 2001, pp. 3331 -.

Tien, C. L.; Majumdar, A. & Gerner, F. M. (1998). *Microscale Energy Transport*, Taylor&Francis, 1998.

Vasileska,D.; Raleva, K. &. Goodnick, S. M. (2008). Modeling heating effects in nanoscale devices: the present and the future, *Journal of Comp. Electronics*, DOI 10.1007/s10825-008-0254-y, 2008.

Vasileska, Raleva and Goodnick, S.M. Self-Heating Effects in Nano-Scale FD SOI Devices: The Role of the Substrate, Boundary Conditions at Various Interfaces and the Dielectric Material Type for the BOX, IEEE Trans. Electron Devices, Vol. 56, No. 12, pp. 3064-3071 (2009).

Wachutka, G.K. (1990). Rigorous Thermodynamic Treatment of Heat Generation and Conduction in Semiconductor Device Modeling, *IEEE Trans. Comp. Aided Design*, Vol 11, 1990, pp 1141-1149.

Zeng, G.; Fan, X.; LaBounty, C.; Croke, E.; Zhang, Y.; Christofferson, J.; Vashaee, D.; Shakouri, A. & Bowers, J.E. (2003). Cooling Power Density of SiGe/Si Superlattice Micro Refrigerators, *Materials Research Society Fall Meeting 2003, Proceedings* Vol. 793, paper S2.2, December 2003, Boston, MA.

4

Low-Dimensional Group III-V Compound Semiconductor Structures

Nobuhiko P. Kobayashi
University of California Santa Cruz, Santa Cruz, California, U.S.A.
Advanced Studies Laboratories, University of California Santa Cruz and NASA Ames Research Center, Moffett Field, California, U.S.A.
Bio-Info-Nano Research and Development Institute, University of California Santa Cruz and NASA Ames Research Center, Moffett Field, California, U.S.A.

1. Introduction

1.1 Epitaxial Growth

Epitaxial growth of inorganic materials in the form of single-crystal thin films can be viewed as a special case of crystal growth that is essentially a first-order phase transition exhibited by a wide range of single chemical elements and a variety of compounds including insulator, metals, and semiconductors. Epitaxial growth of thin films can also be viewed as a unique case of those among various deposition processes of thin films. Epitaxial growth of thin films requires a combination of stringent crystal growth parameters that need to be dynamically tuned to establish appropriate growth environment. Everlasting quest to raise the level of perfection of single-crystal thin films by understanding and improving epitaxial growth processes has been a scientific subject that has inspired many scientists in the past. An early attempt of constructing a theory of crystal growth from vapor phase, including many essential aspects of crystal growth, treated surface features (e.g., steps and kinks) on a crystal surface in equilibrium with vapor by obtaining the rate of advancement of monomolecular steps as a function of supersaturation in the vapor and mean concentration of kinks in the steps.[1] Influence of dislocations was also analyzed and growth rate of a surface containing dislocations was shown to be proportional to the square of the supersaturation for low supersaturation and to the first power for high supersaturation. Equilibrium structures of steps (e.g., the statistics of kinks in steps) were also studied in terms of surface temperature, binding energy parameters, and crystallographic orientation. Shape and size of a two-dimensional nucleus in unstable equilibrium with a given supersaturation at a given temperature were obtained. Analyses on the temperature dependence of the structure of perfect surfaces (i.e., surfaces free from steps at absolute zero temperature) showed that a perfect surface remains flat until the surface reaches a roughening transition temperature at which the surface undergoes a morphological transition by which the surface becomes rough dramatically.[2] With the view more practical, a unified theory was developed to understand the physics of epitaxial thin films and applied to tailor epitaxial growth conditions to obtain perfect epitaxial thin films of uniform

thickness.[3] The pursuit of superior epitaxial growth will continues to motivate scientists and engineers in a variety of technical fields for the years to come as the growing demand for high-quality single-crystal thin films needs to be met for a wide range of advanced solid-state devices.

Apparently in epitaxial growth, one of the prerequisites that drive ordinary thin film deposition into epitaxial growth is the use of a single-crystal substrate. A single-crystal substrate is characterized by its constituent atoms chemically bonded each other to form a three-dimensional network that can be described by specific translational and rotational crystallographic symmetries with the presence of long-range atomic ordering. When a thin film is epitaxially grown on a single-crystal substrate, the thin film grows as a single-crystal and a common interface is created between the single-crystal epitaxial thin film and the single-crystal substrate. A single-crystal epitaxial thin film exhibits a crystal lattice having a definite crystallographic orientation with respect to that of a single-crystal substrate, in other words, a single-crystal epitaxial thin film has a specific crystallographic registry with a single-crystal substrate. The surface of a single-crystal substrate, a part critically relevant to epitaxial growth, plays a major role in epitaxial growth, in particular, in the early stage of epitaxial growth. Surfaces of semiconductor single-crystal substrates are known to exhibit a variety of superstructures (surface reconstructions) that have translational and rotational symmetry different from those present in the bulk part of a semiconductor single-crystal substrate. Although various types of surface reconstructions are observed on surfaces of semiconductor single-crystal substrates under different epitaxial growth conditions, a semiconductor thin film grown on a semiconductor single-crystal substrate is eventually connected coherently to the single-crystal substrate, forming an atomically seamless interface between the grown single-crystal semiconductor epitaxial thin film and the single-crystal semiconductor substrate.

Epitaxial growth has been reviewed by numerous times and various theories intended to describe different aspects of epitaxial growth were developed. In particular, nucleation processes in the early stage of epitaxial thin film growth were extensively studied.[4,5,6,7,8,9,10,11] Nucleation processes in epitaxial growth of thin films have been viewed as one of the scientific fields where experimental demonstrations and theoretical analyses progress side-by-side because the nucleation process in the early stage of thin film growth is an ideal case where experimental variable can be fairly well controlled. The study of the nucleation processes led us to recognize the importance of surfaces in the context of epitaxial growth of thin films. Based on thermodynamics, physical properties of, for instance, equilibrium surfaces can be deduced from the knowledge of the free energy associate with equilibrium surfaces. While a surface viewed as the interface between the surface of an epitaxial thin film and surrounding (e.g., vapor phase established in given epitaxial growth environment) is important for the nucleation in epitaxial growth of thin films, the interface between an epitaxial thin film and a substrate or between an epitaxial thin film and another epitaxial thin film is equally critical. The energy associated with the interface between an epitaxial thin film (or multiple epitaxial thin films) and a substrate is of great importance, in particular, when the thin film is made of a material different from that of the substrate. Further understanding of epitaxial growth of thin films in terms of morphological and structural evolution is clearly the goal that can only be attained by continuous efforts in both experimental and theoretical advancement.

1.2 Heteroepitaxial Growth

Among a wide range of material systems that can be obtained by epitaxial growth (i.e., a single-crystal thin film on a single-crystal substrate), several unrivaled characteristics exhibited by such material systems as those made of group III-V compound semiconductors are recognized as their flexibility of being assembled into hetero-structures where dissimilar group III-V compound semiconductors with various chemical compositions are coupled in the form of multiple thin films. Sophisticated epitaxial growth techniques such as molecular beam epitaxy (MBE) and metal organic vapor chemical vapor deposition (MOCVD) have been successfully implemented for the growth of group III-V compound semiconductor multiple thin films both at the level of high-volume manufacturing and cutting-edge research environment and widely used for a variety of electronic and optoelectronic devices including hetero-bipolar transistors, light-emitting-diodes, laser diodes, and multi-junction solar cells.

Epitaxial growth of a thin film (or multiple thin films) of various group III-V compound semiconductors or related alloys on a substrate dissimilar to the thin films to be grown (i.e., heteroepitaxial growth), however can be exceptionally complex in contrast to simple epitaxial growth of similar materials (i.e., homoepitaxial growth such as a silicon thin film on a silicon substrate, a germanium thin film on a germanium substrate, and a gallium arsenide thin film on a gallium arsenide substrate). For instance, simple pictures of nucleation processes and evolution processes that work for homoepitaxial growth may no longer be valid for heteroepitaixl growth because new constraints, such as misfit strain associated with two dissimilar materials having different lattice constants or chemical incompatibility, are present in heteroepitaxial growth. When two materials having different lattice constants are connected via a two-dimensional interface, misfit strain needs to be included in the description of heteroepitaxial growth. Epitaxial growth of a single-crystal thin film on a single-crystal substrate with misfit strain would be recognized as one of the most challenging, yet most appealing, in both scientific and engineering subject that needs to be addressed to gain utmost benefits out of this concept technologically very important.

Within thermodynamic framework, the key topic of heteroepitaxial growth of a single-crystal thin film on a single-crystal substrate with mismatch is free energy associated with the interface (i.e., interface free energy) between the thin film and the substrate. A series of early investigations employed an array of dislocations that connected two infinite single-crystals, giving a clear illustration of a very simple model for a heteroepitaxial structure analyzed semi-quantitatively.[12,13] This simple model was further advanced by the establishment of models that dealt with a single dislocation to calculate three types of inter-crystalline boundaries; a boundary due to a difference of atomic spacing, a twist boundary, and a symmetrical tilt boundary[14], and later the model for two crystals having infinite thickness was refined by introducing mismatch as an independent parameter and treating the thin film as a film with finite thickness in contrast to the substrate having infinite thickness, putting an emphasis on epitaxial growth in which a thin film was grown on a thick substrate.[15,16]

From technological perspective, the goal is to find a route by which, for given misfit, an epitaxial thin film can be grown without generating crystallographic defects (e.g., misfit dislocations) detrimental to a targeting application, therefore it is essential to correlate misfit to the generation of misfit dislocations in terms of available growth parameters that can be experimentally tuned. The various concepts and forms of what is called critical thickness for

heteroepitaxial growth with lattice mismatch were derived. Critical thickness for given misfit provides a useful guide in designing an epitaxial growth process as it gives a thickness beyond which the generation of misfit dislocations is substantially accelerated. For instance, the critical thickness for the growth of a Ge_xSi_{1-x} thin film on a Si substrates was evaluated by assuming that the generation of misfit dislocation sets in when the areal strain energy density of the film exceeds the energy density associated with the formation of a screw dislocation (i.e., energy balance) at a distance from the free surface equal to the film thickness.[17] In addition to the energy balance, force balance was also used to predict a critical thickness for $GaAs/GaAs_{0.5}P_{0.5}$ multiple thin films with the motivation to reduce misfit dislocation density by (a) using film thicknesses below those at which misfit dislocations are formed between layers, (b) matching lattice parameters of the substrate to those of the multilayer taken as a whole; and (c) using misfit strain to drive threading dislocations out of samples.[18]

The concept of critical thickness defined for the transition at which a coherent thin film (a film fully strained and without dislocations) turns into an incoherent thin film (i.e., a film with dislocations) was analyzed within the framework of two models, Frenkel-Kontorowa and Volterra models. [19] It was recognized that the epitaxial growth is basically a dynamical process (i.e., a true equilibrium state is not reached during epitaxial growth), thus coherent-incoherent transition essentially driven by free energy gradients and reaching equilibrium is hindered by energy barriers associated with the generation of misfit dislocations. Inherent nature of epitaxial growth occurring on a two-dimensional plane was explicitly analyzed.[20] Dependence of equilibrium configurations of a single-crystal thin film on the strengths of substrate-thin film, thin film-thin film bonds, and the interfacial misfit was examined in two-dimension,[20] in contrast to early analyses based on a one-dimensional model in which misfit was treated along only one crystallographic direction. The analysis postulated that misfits along two different crystallographic orientations may differ from each other and could be accommodated by cross grids of dislocations. A variety of practical cases have been analyzed to predict detrimental influences of misfit on epitaxial growth and a range of practical approaches have been proposed to minimize damaging influences of misfit and to take full advantage of heteroepitaxial growth.[21,22,23,24,25,26]

In addition to numerous attempts to describe epitaxial growth of thin films within the framework of equilibrium structures, approaches with atomistic view (i.e., *ab initio* calculations) have been also comprehensively investigated. Phenomenological models such as Frenkel-Kontorowa and Volterra models gave intuitive physical pictures, which was the main advantage; while the *ab initio* calculations were developed with the goal of obtaining, for instance interface energy more quantitatively accurately. Early atomistic approaches were used to directly calculate interface energy. [27,28,29] For group III-V compound semiconductors characterized by fairly strong covalent chemical bonds, however substantial discrepancies between outcomes of equilibrium approaches and those obtained from experiments were often noticed due partly to relatively high Peierls barriers. The important role of thermal fluctuations and the influence of free energy gradients were suggested[30] and the dynamics of misfit dislocations relieved by continuous plastic deformation was described.[31] The model, referred to as configuration-dependent reactive incorporation (CDRI) model,[32.33.34] with an atomistic view for the epitaxial growth of group III-V compound semiconductor thin films was developed [37]through kinetic Monte-Carlo simulations and introduced with inspiration of several pioneering experimental works[35,36]

on the nature of the adsorption, dissociative reaction, and incorporation of arsenic molecules in homoepitaxial growth of gallium arsenide and related alloys. The CDRI model accounts for a range of surface kinetic processes such as[37] sticking coefficients of the group III atoms, adsorption coefficients for the group V molecular species, intra-planar migration of group III atoms, inter-planar migration of the group III atoms, the group III local configuration dependent reaction rates for the dissociative molecular reaction of physisorbed group V diatomic molecules, the associative reaction of chemisorbed group V atoms to form diatomic molecules and its subsequent desorption.

Epitaxial growth of thin films (quasi two-dimensional), serving as a basic building block, predominate most of electronic and optoelectronic devices commercially available. In contrast to epitaxial thin films, low-dimensional epitaxial structures, such as semiconductor three-dimensional (3D) islands (quasi zero-dimensional) and semiconductor nanowires (quasi one-dimensional), have opened a new paradigm for group III-V compound semiconductors, emerging as nanometer-scale low-dimensional structures. A range of coherent epitaxial low-dimensional semiconductor structures have been demonstrated on various substrates/surfaces physically incompatible. In this chapter, the first part will describe indium arsenide (InAs), one of group III-V compound semiconductor binary alloys, grown, by molecular beam epitaxy, into the form of three-dimensional (3D) islands (also referred to as self-assembled quantum dots) on gallium arsenide surfaces. Two-dimensional (i.e., nucleating clusters) to three-dimensional (i.e., 3D islands) morphological transition and size evolution/vertical alignment of InAs 3D islands in a single and multiple stacks will be illustrated. The lateral size and size dispersion are found to first increase drastically with a small amount of additional InAs deposition and then decrease and saturate, indicating the onset of a natural tendency for lateral size equalization. The vertical alignment of multiple stacks of 3D islands is a result of strain field generated by InAs 3D islands embedded within a GaAs matrix.

In the second part, indium phosphide (InP), another group III-V compound semiconductor binary ally, nanowires grown by metal organic chemical vapor deposition on non-single-crystal surfaces are described. Unlike conventional epitaxial growth of thin films, the proposed route for growing nanowires requires no single-crystal surface. In principle, only short-range atomic order, in contrast to long-range atomic order required for epitaxial growth of thin film, is necessary for nanowires. A template layer that possesses short-range atomic order prepared on a non-single-crystal surface is employed. Ensembles of InP nanowires grown on a template prepared on a non-single-crystal surface are found to be single-crystal and electrically/optically active, which unlocks attractive applications where III–V compound semiconductors are functionally integrated onto various incompatible material platforms.

2. Indium Arsenide Three-dimensional Islands

2.1 Background

In a wide range of epitaxial growth of thin films, a thin film evolves by undergoing three major stages; nucleation of small nuclei (monolayer and double-monolayer islands, etc.), growth of nuclei, and coalescence of nuclei.[38] Presence of misalignment of nuclei caused by the rotation of nuclei around their axis perpendicular to the surface on which nucleation progressed was a popular subject examined and discussed comprehensively within the

scope of epitaxial growth of thin films, which suggested that misfit dislocations generated in a nucleus change misfit strain within the nucleus.[39] Stability of small nuclei seen in the nucleation stage was studied in terms of their size, misfit, stiffness, and strength of film-substrate interaction, which led conclusions that strain energy associated with both crystallographic registry and misfit promotes three-dimensional growth (i.e., the formation of three-dimensional islands) when the misfit is large.[40] Analysis of equilibrium configurations of growing epitaxial islands was also carried out by considering discrete characteristics of consecutive layers of a growing epitaxial three-dimensional (3D) island and interaction between them.[41] The aim of studying epitaxial growth of 3D islands under equilibrium is to find mutual dependence among various physical aspects involved in the epitaxial growth of 3D islands resulted from layered growth within the islands. Such physical aspects includes, for instance, the number of atomic layers consisting of an island, the number of atoms contained in each layer, chemical bonding within and between layers, substrate interfacial bonding, natural misfit and number of misfit dislocations, over-all strain in an 3D island, the conditions required for stable coherent configurations, etc. For given misfit between an overgrowth and a substrate, the misfit may not be entirely accommodated by an integral number of identical misfit dislocations. Thus, the remaining misfit is accommodated by residual elastic strain within 3D islands. Ultimate goal is to form coherent 3D islands by engineering misfit strain with accurate and reproducible growth conditions.

With highly-sophisticated controllability of every aspect of epitaxial growth conditions provided by, for instance, molecular beam epitaxy and metal organic chemical vapor deposition, epitaxial growth of group IV elemental and group III-V compound semiconductors on lattice-mismatched substrates (i.e., strained epitaxial growth) has been employed as a technique to demonstrate semiconductor 3D islands (also referred to as quantum dots) with nanometer-scale size. Such 3D islands have been extensively studied with the hope of developing further advanced electronic and optoelectronic solid-state devices. The formation of 3D islands is in principal driven by a specific mechanism by which given misfit strain is accommodated by morphological change accompanying with or without plastic deformation. The level of understanding physical pictures of epitaxial growth of 3D islands does not seem to have reached its completion even though enormous efforts have been made in both experimental and theoretical studies.

Epitaxial growth of thin films, in contrast to epitaxial growth of 3D islands, implies that relevant growth processes occur over large area (i.e., size of the area, on a substrate, over which epitaxial growth proceeds does not have direct influence on the way epitaxial growth of a thin film progresses, in other words, size of the area over which epitaxial growth of a thin film progresses is considered to be infinite). Unlike thin films, 3D islands are characterized as small (lateral size ~several tenths of nanometers, and height ~several nanometers) isolated three-dimensional structures, each of which is connected to the surface of a substrate through very small area comparable to its lateral size as in Fig. 1.[42] Semiconductor 3D islands, at least those drawing significant attention with the hope of superior optoelectronic devices in the last decade, are expected to exhibit unique physical characteristics associated with their unique features such as shape, size and electronic structures. Among various semiconductor 3D islands, two material systems; silicon/germanium grown on silicon surfaces and indium arsenide (InAs) grown on gallium

arsenide (GaAs) surfaces have been used as convenient vehicles for extensive study because of their inherent simplicity (e.g., chemical composition).

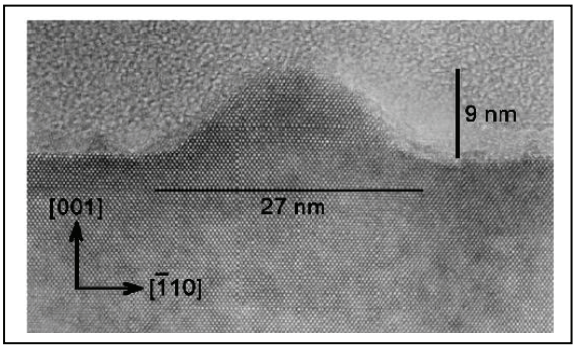


Fig. 1. Lattice image of InAs 3D island obtained by [110] cross-sectional transmission electron microscope lattice imaging. The width and height of the island are 27 and 9 nm, respectively.[42]

Comprehensive reviews in the field of 3D islands (also referred to as quantum dots) are available.[43] In the following section, a series of studies on InAs 3D islands grow on GaAs substrates are described by focusing on three topics; two-dimensional to three-dimensional morphological transition, lateral size equalization, and vertical alignment of an ensemble of 3D islands, all of which are crucially important issues that need to be addressed in the course of implementation of ensembles of 3D islands as a part of solid-state devices.

2.2 Two-dimensional to Three-dimensional Morphological Transition

Formation of coherent 3D islands driven by lattice mismatch between an overgrowth and a substrate has been a fundamental area of research both experimental and theoretical in the field of epitaxial growth in nanometer-scale represented by numerous examples, in particular, with semiconductors such as Ge/Si[44,45] and In(Ga)As/GaAs material systems.[46,47,48,49,50,51,52,53,54,55,56,57] The main focus is to understand the nature of surfaces changing from two-dimensional morphology (i.e., continuous thin films or low-profile, one- or two-monolayer height, two-dimensional clusters appearing in the early stage of highly strained epitaxial growth) to three-dimensional morphology (i.e., surfaces with 3D islands). From solid-state device perspective, early reports on laser oscillation[58,59] from devices that employed InGaAs 3D islands embedded in a GaAs matrix led further studies to clearly correlate optical characteristics of ensembles of 3D islands and lasing characteristics. InAs/GaAs(001) material system has been a vehicle to further understand the atomistic nature of two-dimensional cluster (1ML high) to three-dimensional island (typically 2~4 nm high) transition.

The level of understanding atomistic nature of the two-dimensional to three-dimensional (2D-3D) morphological transition observed in highly-mismatched heteroepitaxial growth has certainly improved dramatically in the past ten years. For many years, the 2D-3D morphological transition was believed to be associated with the formation of defects such as dislocations. However, experimental demonstrations with two semiconductor systems; InGaAs grown on GaAs and Ge on Si, clearly showed that the formation of coherent 3D

islands (i.e., 3D islands without structural defects) is possible in the 2D-3D morphological transition, which has led extensive studies aiming at better atomistic and kinetic understanding of the 2D-3D morphological transition.[33,49,51,52,53,60,61] Therefore, reviewing some of early studies on the 2D-3D morphological transition is worth to acknowledge how much the understanding has advanced. The growth mode characterized by the 2D-3D morphological transition is classically referred to as the Stranski-Krastanow growth mode; however there is clear evidence that the 2D-3D morphological transition in highly-mismatched epitaxial growth is much more complex than the spontaneous change from 2D to 3D morphology described in the framework of the Stranski-Krastanow growth mode.

In the following section, a series of systematic studies on InAs three-dimensional (3D) islands grown on GaAs(001) surfaces by molecular beam epitaxy (MBE) are described. *In situ* ultrahigh vacuum (UHV) scanning tunneling microscope combined with atomic force microscope (STM/AFM) and *ex situ* photoluminescence (PL) and photoluminescence excitation (PLE) were used to study the growth of InAs 3D islands. The amount of InAs delivered onto the GaAs(001) surfaces was varied within the range that covered sub-monolayer to well-formed 3D islands prior to the onset of island coalescence to investigate the initial formation and the subsequent evolution of InAs 3D islands. The details of the sample preparation are as follows and also described elsewhere. [51,62] All samples were grown by MBE on silicon doped n+-GaAs substrates with orientation of (001) ± 0.1 degrees. The growth was monitored by reflection high energy electron diffraction (RHEED). The deposition rate of InAs was calibrated by RHEED on a InAs(001) substrate and determined to be 0.22 ML/s at the substrate temperature of 500 °C and with arsenic beam equivalent pressure of $6x10^{-6}$ Torr. InAs was deposited at 500 °C on a 500 nm GaAs buffer layer that showed clear *c*(4x4) surface reconstruction. The amount of InAs deposition given with respect to the GaAs(001) surface atomic density was confirmed to be reproducible within 0.022 ML. On *in situ* RHEED observation, weak spots superposed on faint streaks appeared at 1.57 ML of InAs coverage, indicating the onset of InAs 3D island formation. All samples were cooled down immediately after the deposition of InAs by shutting off electrical power to the substrate heater in order to minimize post-growth evolution of the grown surfaces so that the nature of InAs 3D island evolution as close as possible to the growth conditions was examined. No intentional annealing or growth interruptions was conducted as these post-growth thermal processes are known to promote a variety of post-growth evolution on grown surfaces.[63,64] The morphological instabilities such as mounds were reported on samples grown with growth interruption[65], which needed to be avoided as much as possible because the formation of 3D islands is extremely sensitive to underlying surface morphology. For instance, diffusion of species on a rough surface can be highly anisotropic.[66] Samples were transferred to the UHV STM/AFM chamber through the UHV interconnect for *in situ* surface characterization. Samples used for *ex situ* PL and PLE were grown on semi-insulating GaAs substrates by following exactly the same procedure described above for the STM/AFM samples with one exception that they were capped with a GaAs thin film grown by migration enhance epitaxy at the substrate temperature of 400 °C.[67]

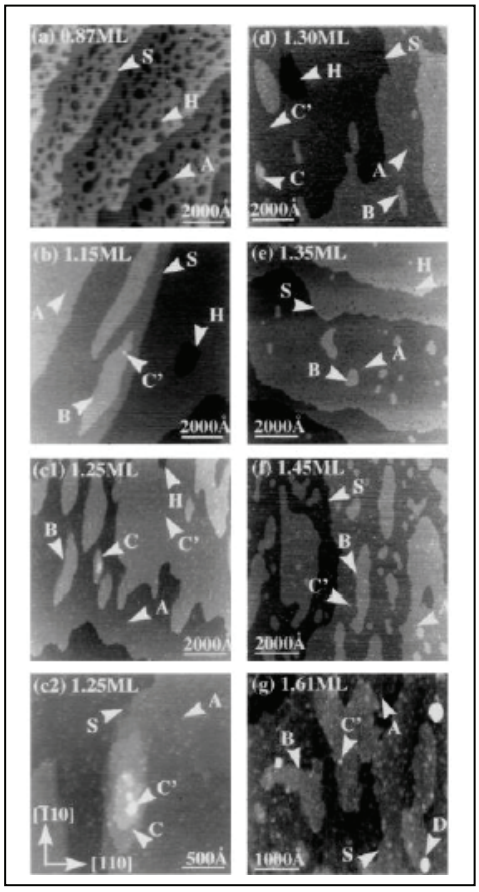


Fig. 2. STM images showing the evolution of InAs on GaAs(001) for depositions of (a) 0.87, (b) 1.15, (c1, c2) 1.25, (d) 1.30, (e) 1.35, (f) 1.45, and (g) 1.61 ML. The labels in the figure denote small 2D clusters (A), large 2D clusters (B), small Q3D clusters (C□), large Q3D clusters (C), 3D islands (D), 1 ML high steps (S), and 1 ML deep holes (H).[68]

Fig. 2 shows a series of STM images of samples grown with the amount of InAs deliveries in the range from 0.87 to 1.61 ML.[68] A varieties of surface features are clearly seen on the samples at different amount of InAs deliveries. 1 ML high steps (labeled *S*) with 200~400 nm wide terraces are seen, which is consistent with the GaAs substrates with tilt of approximately 0.1° used for the growth. The first InAs layer seen in panel (a), generally referred to as a wetting layer, appears to be incomplete up to 1.35 ML InAs delivery in panel (e), evidenced by 1 ML deep holes (labeled *H*). At 0.87 ML in panel (a), the holes cover approximately 20% of the surface. Clusters of up to few nm wide, lateral size <20 nm, and 1 ML high with a high density (>10^{11} cm^2), referred to as *small 2D clusters* (labeled *A*) are seen. Additional InAs delivery, as seen in panel (b), leads to the formation of 1 ML high clusters with width ≥50 nm and lateral sizes up to hundreds of nm (referred to as *large 2D clusters*

and labeled *B*) on top of the still incomplete wetting layer. These large 2D clusters are elongated in the [1-10] direction as seen in panel (b). At the deposition of 1.15 ML in panel (b), features that are 2-4 ML high with respect to the flat InAs surface and up to 20 nm wide (labeled *C'*), referred to as *small quasi-3D (Q3D) clusters*, start to appear. At 1.25 and 1.30 ML InAs delivery in panels (c1) and (d), in addition to the small Q3D clusters, clusters of height of 2-4 ML with lateral extension ≥50 nm, referred to as *large Q3D clusters* are present (labeled *C*). Furthermore, features consisting of a large Q3D cluster (*C*) topped by a small one (*C'*) up to 5 ML high (i.e., features with an extended base and a narrow top) are seen in panel (c2). A notable finding was that the Q3D clusters (small and large) *disappear* once at 1.35 ML, exhibiting a 2D-like surface. Then, as the InAs delivery is further increased to 1.45 ML, only small Q3D clusters *reappear* at a density 2 orders of magnitude higher well before the first 3D islands (7-14 ML high with a lateral size <25 nm) labeled *D* in panel (g) form at 1.57 ML (the critical delivery denoted Q_c). Between 1.57 and 1.74 ML delivery, 2D clusters, small Q3D clusters, and 3D islands co-exist as shown for 1.61 ML deposition in panel (g). Evolution of the 2D clusters, Q3D clusters, and the 3D islands in terms of their area density as a function of InAs delivery is shown in Fig. 3.[68] The Q3D cluster density evidently indicates the appearance, disappearance, and reappearance of 3D morphological features well in advance of the regime in which well-formed 3D islands form. The presence and behavior of the Q3D clusters play a major role in analyzing PL and PLE studies in the following section.

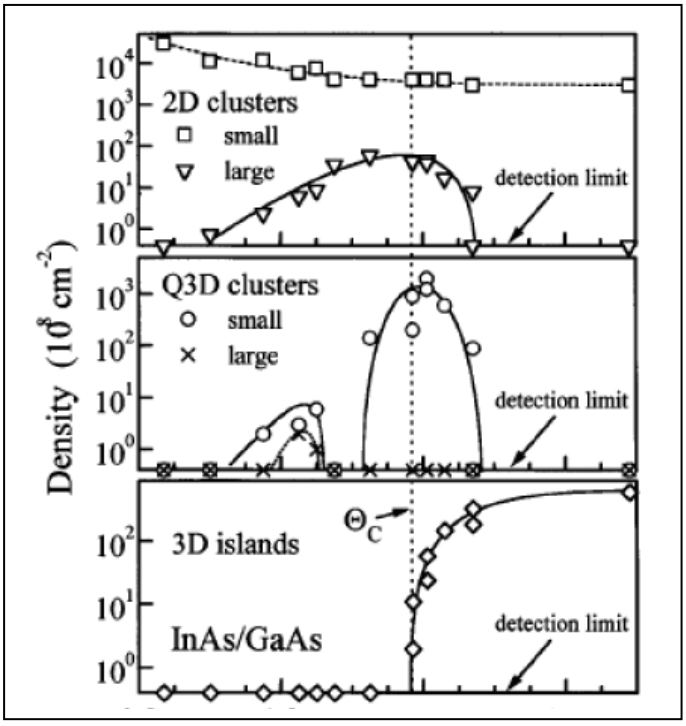


Fig. 3. Area density of 2D and Q3D clusters as well as 3D InAs islands on GaAs(001) as a function of InAs delivery.[68]

Fig. 4 shows the evolution of the PL spectra from 1.00 to 2.00 ML InAs delivery. The narrow peak, attributed to recombination in the wetting-layer, near 1.45 eV evolves with increasing InAs delivery and vanishes just beyond Q_c. The peak at 1.215 eV for the 2.00 ML sample is attributed to recombination in 3D islands as reported earlier.[52,57,69] The PL spectra of the 1.15 and 1.25 ML samples reveal weak, yet distinct, peaks at 1.322 and 1.274 eV, respectively, however, no peak is resolved in the range between 1.20 and 1.35 eV for the 1.35 and 1.45 ML samples. At 1.55 ML delivery (just below Q_c), emission in the range between 1.20 and 1.35 eV reappears, indicating a re-entrant PL behavior showing similarity to that observed in Figs. 2 and 3. The peaks at 1.322 and 1.274 eV in the 1.15 and 1.25 ML samples are attributed to the large 2D and large Q3D clusters serving as parts where the wetting-layer is locally thicker. A comparison with the STM information on the morphology of the InAs layer suggests that *small* Q3D clusters (i.e., feature *C'* in Fig. 2) on top of large 2D and large Q3D clusters act as optically active quantum dots. The observation of emission from the coupled structure of small Q3D clusters on large 2D clusters also explains the disappearance of the PL in the 1.20 to 1.35 eV range for the 1.35 ML sample with no small Q3D clusters, indicating that dominant nonradiative recombination takes place in the large 2D clusters.

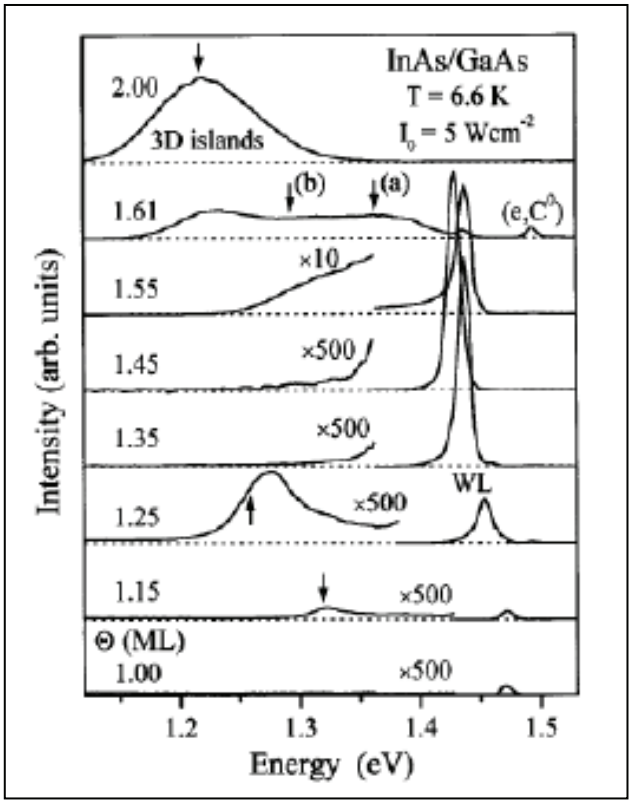


Fig. 4. PL spectra for InAs on GaAs(001) at various depositions Q, excited at 514 nm with a density of 5 Wcm^{-2}. The peak at 1.493 eV (1.61 ML sample) is attributed to carbon-related recombination in the GaAs barrier.[68]

The structural and optical studies on the comparable samples provide a detailed picture of the kinetically controlled evolution of heteroepitaxially grown InAs surfaces. InAs is incorporated in higher layers even before the first InAs layer is completed. In addition to 2D clusters, Q3D clusters form for depositions as low as 1.15 ML due mainly to the strain-driven surface kinetics near the edges of 2D clusters. The strain fields at the edges of 2D clusters lead to an asymmetry in inter-planar In diffusion,[33] promoting the formation of Q3D clusters possible. Therefore, the Q3D features disappear above 1.30 ML because of the change in the surface strain fields when the first InAs ML reaches nearly its complete coverage. The reappearance of the small Q3D clusters with a further increase in In delivery leads to a low energy tail in the wetting-layer peak in Fig. 4 for 1.55 ML. The high density of small Q3D clusters at 1.61 ML, corresponding to an average separation of 22 nm indicates that they act as precursors of 3D islands. It must be emphasized however, that surface morphology of InAs on GaAs(001) material systems near the 2D-3D morphological transition shows a variety of surface features,[53,61,70.71.72] thus a unique picture would need to be established for a series of samples grown under specific growth conditions and schemes. For instance, InAs/GaAs(001) system grown with a growth interruption and studied by atomic force microscope showed that the evolution of 3D islands exhibited two transition onsets at 1.45 and 1.59 ML of InAs coverage, corresponding to the formation of two distinguishable families, small and large islands.[73] The transition between the two families of islands and the explosive nucleation of the large islands seem to be associated with the erosion of step edges around the islands. Furthermore, *in situ* RHEED measurements of the surface evolution during the growth of InAs 3D islands on GaAs showed that InAs 3D islands desorbed when arsenic supply was interrupted.[74]

2.3 Lateral Size Equalization

The formation of three-dimensional (3D) coherent islands in the process of heteroepitaxial growth of material systems with large lattice mismatch and their unique optical properties, demonstrated for InAs 3D islands grown on GaAs,[67,75] have motivated us to find a way to describe mechanisms that govern the formation and evolution of semiconductor 3D islands in the framework of Stranski–Krastanow growth mode. Growth conditions (e.g., the influence of group V pressure, substrate temperature, and growth procedure without or with growth interruption during InAs delivery) employed for the formation of InAs 3D islands have been carefully examined.[52,53,67,75,76,77,78,79] The kinetic processes impacting the evolution of semiconductor 3D islands and their size distribution are two subjects highly relevant to envisioning practical implementation of ensembles of semiconductor 3D islands as an active or a passive part in solid-state devices Therefore, systematic study, using *in situ* ultrahigh vacuum (UHV) atomic force microscope (AFM) and scanning tunneling microscope (STM), on the growth of InAs 3D islands on GaAs(001) surfaces is essential. The unique evolution of InAs 3D islands at sequential growth stages characterized by specific InAs coverage is described in the following section with the aim at minimizing size distribution of an ensemble of InAs 3D islands.

All experiments were carried out in the integrated, ultra-high-vacuum-interconnected, molecular beam epitaxy (MBE) growth-processing-characterization system to conduct growth and analysis without having samples exposed to air. InAs deposition was done at the substrate temperature (T_s) 500 °C with the arsenic beam equivalent pressure (P_{As4}) $6x10^{-6}$ Torr on GaAs(001) surfaces showing *c*(4x4) surface reconstruction ensured by reflection

high-energy electron diffraction (RHEED). The desired amount of InAs (0.87 to ~2.18 ML with respect to the GaAs(001) surface atomic density) was delivered at a growth rate of 0.22 ML/s with an accuracy and precision of ±0.07 ML and ±0.022 ML, respectively. As the In incorporation coefficient is expected to be unity under the growth conditions employed in the experiment, all the delivered InAs is incorporated into the solid. At ~1.57 ML InAs delivery, the RHEED pattern showed weak spots along with faint streaks, signifying the onset of 3D island-like features. The power to the substrate heater was turned off immediately after the deposition of InAs to let T_s drop rapidly. At T_s of 450 °C, P_{As4} was decreased to 1x10^{-6} Torr and at T_s of 400 °C the As_4 flux was shut off completely. The above-mentioned growth and cooling procedures were strictly followed for each growth run in order to maintain consistency among all the samples. The unique cooling procedure minimizes and, more importantly, keeps post-growth evolution of the grown surfaces similar from sample to sample. Contact-mode AFM was used to measure density, lateral width and height of InAs 3D islands. To avoid ambiguities in determining the island size dispersion, all island size measurements on a given sample were completed with the same AFM tip. Two different AFM tips were used for the study of our samples and the data obtained using these two tips were found to be consistent with each other. Therefore, while the absolute values of the measured lateral sizes, as is well known in scanning probe microscopy, are an upper bound[80], the relative values within a sample and between samples can be relied upon.

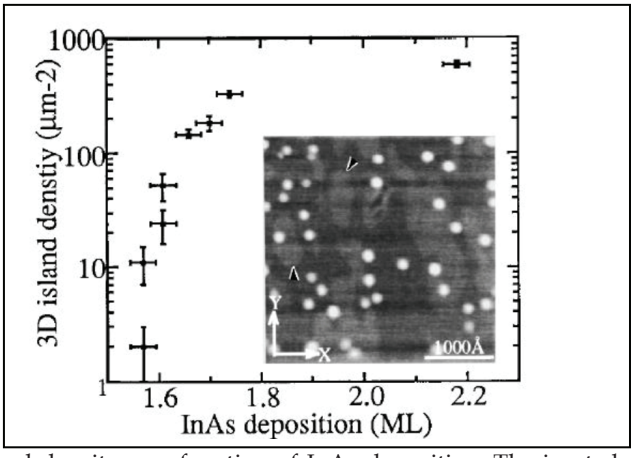


Fig. 5. InAs 3D island density as a function of InAs deposition. The inset shows an *in situ* UHV AFM image for ~1.70 ML InAs deposition on GaAs(001). The bright "spots" are 3D islands while the "patches" in the background (some of these are indicated by arrows) are 2D clusters. The notations X and Y refer to the crystallographic directions [110] and [1 $\bar{}$10], respectively.[51]

The AFM image shown as an inset in Fig. 5 is for ~1.70 ML InAs delivery and represents the AFM images obtained for InAs delivery >1.57 ML. The bright spots are the InAs 3D islands. The ''patches'' (marked by two arrow-heads) are 2D clusters with one monolayer high with respect to their surroundings. The areal density of the InAs 3D islands is plotted in Fig. 5, which indicates that the areal density increases rapidly for InAs delivery from 1.57 to 1.74

ML and then slows down from 1.74 to 2.18 ML, reaching a maximum value of ~6x10^2 μm^{-2} before island-coalescence begins to set in. Correspondingly, the STM determined that the areal density of the quasi-3D clusters rises to a maximum of ~2x10^3 μm^{-2} at 1.61 ML and then drops, reaching near zero for >1.75 ML InAs delivery. The InAs 3D island lateral size and height distributions are plotted in Fig. 6(a) and (b) for different InAs 3D island densities. The lateral size of a 3D island is defined as the full width at half maximum of the island height. As seen in Fig. 6(a), initially when the 3D island density is ~11 μm^{-2} the islands have a narrow lateral size distribution, however as the island density increases to around 24 μm^{-2}, the island lateral size distribution broadens significantly without an evident peak. Concomitantly, the average lateral size seems to increase as well. Remarkably, further increase in island density (~58 μm^{-2}) makes the island lateral size distribution narrow again and the average lateral size also becomes smaller. Thereafter, with increasing island density, the lateral size distribution becomes narrower, while maintaining an essentially constant average lateral size, and then practically invariant as a function of coverage prior to the onset of island coalescence. The evolution of the InAs 3D island height distribution as seen in Fig. 6(b) appears to exhibit the similar but not identical trend to that of the lateral size distribution. Certain variation in the heights is observed at any given island density, which clearly indicates that the islands coexisting at a given evolution stage do not have definite shape (e.g., equilibrium shape), thus their formation is kinetically controlled. It is important to note that the transition from a narrow to a broad and back to a narrow island lateral size distribution under the growth conditions employed in the experiment occurs over a very small increase (<0.08 ML), thus extreme care in the control of deposition amount is required for this notable observation. An initial broadening and subsequent narrowing reported[53,76,77] for near equilibrium islands further suggests that this nature of island evolution is a general phenomenon occurring over the entire range from highly kinetically controlled regime to near thermodynamic equilibrium.

The above remarkable findings in this carefully controlled series of systematic experiments thus not only provide stringent tests for models[33,60,78,81] of strained epitaxy and island formation but also provide insights into the nature of atomistic kinetic processes[33,78] consistent with the observations. The observed initial broadening in the island lateral size distribution and increase in average size up to island density of ~24 μm^{-2} can be understood if, at the earliest stages, the 3D island initiation rate is constant and their growth rates are independent of each other. In this case the size of an island at this stage is simply determined by the time that passes since its initiation provided surface diffusion and incorporation of atoms at the island periphery are not rate limiting steps. The narrowing of the lateral size distribution and the reduction in the average lateral size beyond ~24 μm^{-2} island density in Fig. 6(a) indicate that the above mechanism of island formation at the earliest stage is subsequently replaced by some other evolution mechanism(s) as deposition proceeds. Two important kinetic processes contributing to island growth provide mechanisms consistent with this narrowing. The first is the suggestion based upon thermodynamic energy considerations[33,78,81] that the growth rate of 3D islands would diminish when an island grows to a certain size due to the accumulation of elastic strain energy. From the kinetic viewpoint underlying strained growth simulations,[33] this has been modeled as an attendant increase in an energy barrier for atom incorporation at island edges. As a consequence, at this stage, smaller islands will grow more rapidly, along with the possible formation of new islands. The second contribution is suggested[78] to arise when,

with increasing island density in response to continued InAs delivery, the island-induced strain fields in the substrate (that help to stabilize coherent larger islands in the first place) begin to interact.

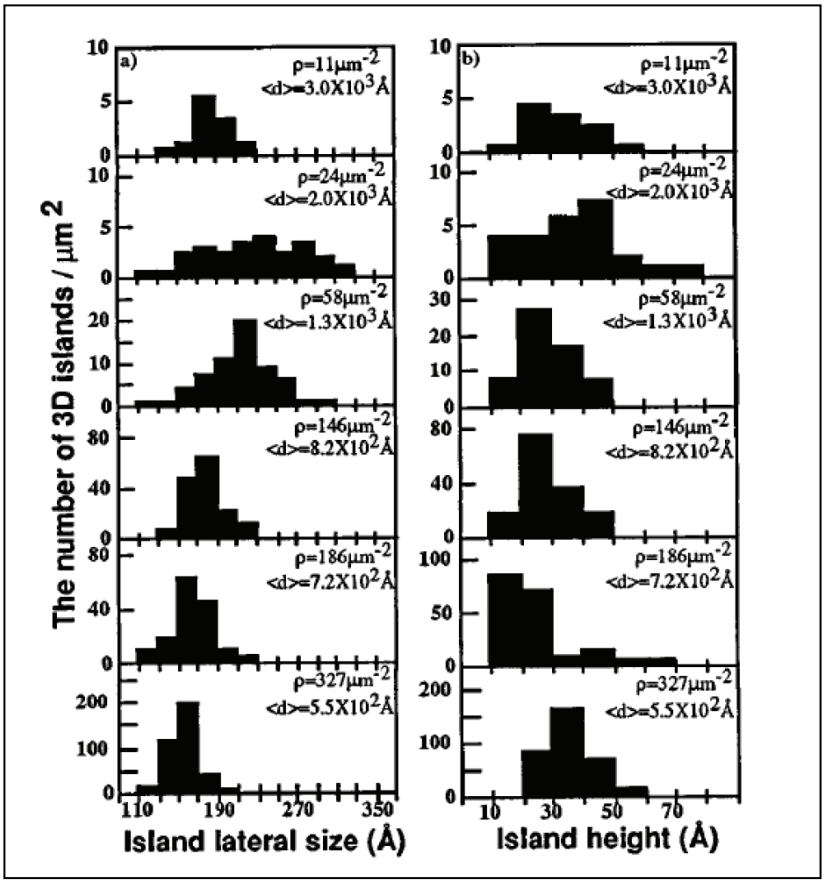


Fig. 6. InAs 3D island (a) lateral size and (b) height distribution for different total InAs 3D island densities, ρ. Also indicated is the mean inter-island separation, $\langle d \rangle = \rho\text{-}1/2$.[51]

At this stage a *driving force* for preferential migration of atoms towards smaller islands can arise due to a downward tilt in the surface potential towards the smaller islands caused by the presence of the interacting island-induced strain fields in the substrate. The interacting strain fields in the substrate and the wetting layer also provide a means to destabilize the initially largest islands by not accommodating as much strain relief in the substrate and thus, from a kinetic viewpoint, decrease the barrier for detachment of atoms from the larger islands. The detachment of atoms from the larger islands can give rise to the loss of material from the initially largest islands, consistent with the AFM results shown in Fig. 6(a) and contributes to island size equalization as well as narrowing of the average size and potentially a change in the island shape. Such *kinetic* processes are likely to begin to impact island growth when the inter-island distance becomes comparable to the range of the

individual island-induced strain fields and a characteristic length, such as the surface diffusion length of the adatoms. The observed value of island density (~24 μm^{-2}) in Fig. 6(a) can be converted to a mean inter-island separation, which results in ~200 nm, the critical inter-island distance, for which the manifestation of island–island interaction is expected to rise rapidly. Indeed, the diffusion length of In during InAs growth on GaAs under the same growth conditions as used in this study has been independently estimated[56] to be ~280 nm, which is comparable to the critical inter-island distance indicated by the AFM studies of the island size evolution behavior described above. The similar trend of narrowing lateral size distribution was reported on double-stacked InAs 3D island on GaAs at island density much higher (~1800 μm^{-2}) than those in our studies[82], indicating that the proposed mechanisms that control overall island lateral size and lateral size distribution are fairly applicable to various InAs 3D islands formed under substantially different growth conditions and/or growth schemes. Various mechanisms that would result in narrow island size distribution have been proposed,[60,83,84] and this subject will continuously be a critical and challenging in the epitaxial growth of semiconductor 3D islands.

2.4 Vertical Alignment

"Stress Engineering" can be referred to as, within the scope of epitaxial growth, the deliberate use of elastic strain naturally or artificially generated within a medium to control kinetics in the process of epitaxial growth. Misfit strain, for instance, can provide a natural driving force for 3D islands in lattice mismatched growth such as InAs on GaAs. Coherent InAs 3D island formation on GaAs(100) offers an bottom-up approach to fabrication of optically efficient 3D islands.[75,85] Given the random initiation of islands on a starting surface of GaAs(100), a certain degree of spatial randomness always sets, resulting in spatial randomness of the locations of 3D islands. In the following section, strain fields induced by 3D islands embedded in a protective cap layer are used, with the spirit of "Stress Engineering", to demonstrate self-organization of InAs 3D islands along the vertical (i.e., growth) direction to examine the potential role of the evolving and interacting strain fields generated within a cap layer by multiple InAs 3D islands.

Molecular beam epitaxy (MBE) was used to grow InAs 3D islands on GaAs(100) substrates.[51] A GaAs buffer layer was grown on a GaAs(100) substrate first, then, the first set of InAs 3D islands were formed at the substrate temperature (T_s) 500 °C with the arsenic beam equivalent pressure (P_{As4}) $6x10^{-6}$ Torr on the GaAs buffer layer showing c(4x4) surface reconstruction. Subsequently, the first set of InAs 3D islands were overgrown by a GaAs spacer layer. The thickness of the GaAs spacer layer was varied in the range from 30 ML to 200 ML. The GaAs spacer layer was further followed by the growth of the second set of InAs 3D islands. This combination of the growth of InAs and GaAs spacer was repeated one or more times to obtain two or more sets of islands separated by GaAs spacer layers. Based on RHEED analysis, the growth of InAs on a GaAs spacer layer did not show any apparent differences from the growth of the first set of InAs 3D islands. Transmission electron microscope (TEM) was used to investigate the samples having multiple sets of InAs 3D islands separated by multiple GaAs spacer layers having various thicknesses.

Fig. 7(a) shows a representative [011] cross-sectional TEM (XTEM) picture of a sample with two sets of InAs 3D islands separated by a GaAs spacer of 46 ML. From the strain contrast of the islands in the two sets, each island in the second set is seen to be located on top of another in the first set, which indicates that a strong one-to-one vertical correlation between

the two sets is present. A detailed statistical analysis on the sample reveals that the paring probability of islands in the two sets, measured from XTEM projected images, is 0.885 ± 0.032. The pairing probability drops to 0.492 ± 0.04 when the GaAs spacer thickness increases to 92 ML as in Fig. 7(b).

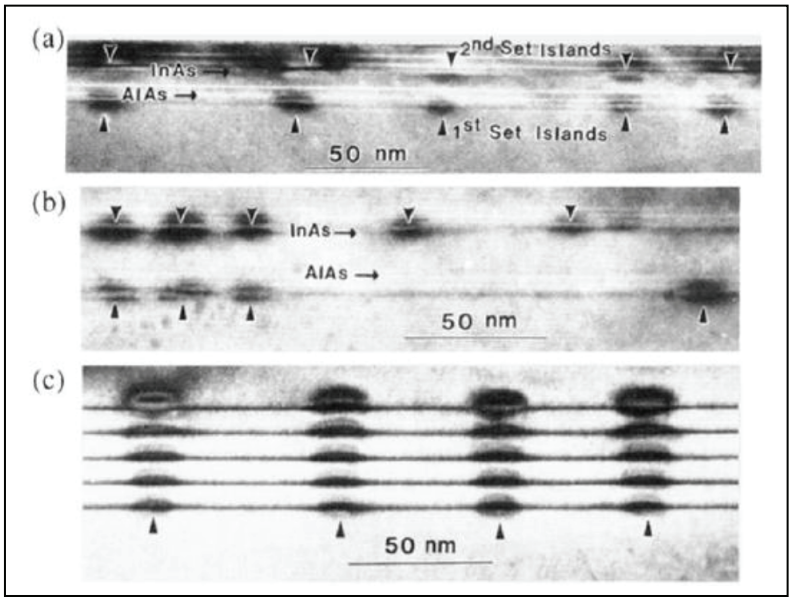


Fig. 7. Typical *g* = (400) bright field TEM pictures taken along [011] azimuth for the samples with two sets of islands separated by (a) 46 and (b) 92 ML spacer layers, respectively. Arrows point to the island positions indicated by the strain contrast. (c) A typical *g*= (200) dark field TEM picture of a sample with five sets of islands separated by 36 ML spacer layers.[56]

For a sample with five sets of islands separated by 36 ML thick spacers grown by migration enhanced epitaxy in Fig. 7(c), the probability of island pairing to the set just below is maintained at approximately ~0.95 for all adjacent island sets. High resolution lattice images confirmed that the spacer layer is atomically flat and the islands are coherent. Possible intermixing between InAs 3D islands and a GaAs spacer layer is an issue that needs to be further addressed since the evidence of intermixing between InAs 3D islands and a GaAs cap layer during the growth of the cap layer and resulting change in shape and composition of InAs 3D islands were reported.[86] Fig. 8 summarizes the pairing probabilities obtained as a function of the spacer thickness using both [011] and [011-] XTEM azimuths. Three distinguishable regimes can be seen; (1) for small spacer thickness <36 ML, the probability is greater than 95%, indicating a nearly completely correlated behavior, (2) a regime of gradual decrease in the probability, and (3) for larger thickness, the probability saturation at a value corresponding to random overlapping of islands. A model analysis accounting for the mechanochemical surface diffusion gives an island average size and average separation dependent characteristic spacer layer thickness below which a vertically self-organized growth occurs.[56] Vertical alignment of InAs 3D islands are also investigated as a means to

improve size distribution. The thickness of a GaAs spacer layer was varied to study its influence on the structural and optical properties of InAs 3D islands. The structural and optical properties of multi-stacked InAs 3D islands grown on GaAs surfaces were studied by photoluminescence measurements. The PL full width at half maximum, reflecting the size distribution of the islands, was found to reach a minimum for an GaAs spacer layer with a thickness of 30 ML.[87] Similar vertical self-organization of InAs 3D islands with GaAs spacer layers was observed in samples grown by metal organic chemical vapor deposition,[88] suggesting this particular "Stress Engineering" is applicable for various epitaxial growth methods.

3. Indium Phosphide Nanowires on Non-single-crystal Surfaces

3.1 Background

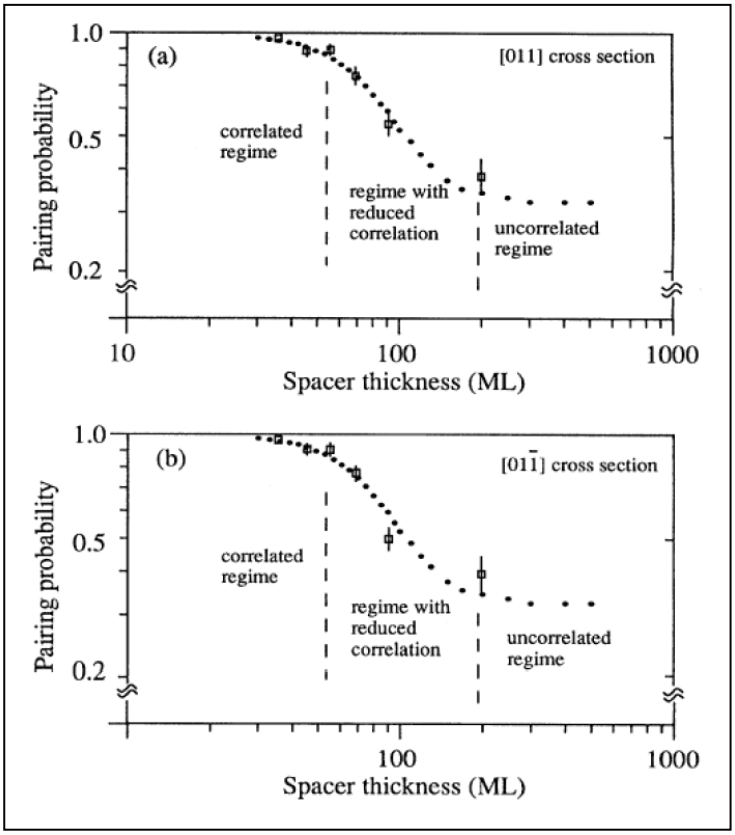


Fig. 8. Experimentally observed paring probabilities (open squares) as a function of the spacer thickness shown on a log-log plot of samples with two sets of islands (a) [011] cross section, and (b) [011-] cross section.[56]

As a natural extension from epitaxial growth of thin films (two-dimensional epitaxial growth), epitaxial growth of three-dimensional (3D) islands (zero-dimensional epitaxial growth) described in the previous section appears to fall into a category of a special case of the formation of thin films accompanying significant morphological transition because of large misfit between a thin film and a substrate. Completely apart from thin films and 3D islands can be seen in the recent surge in the field of one-dimensional epitaxial growth (e.g., semiconductor nanowires). In particular, nanowires of group IV elemental and group III-V compound semiconductors grown with the aid of metal catalyst in the form of nanometer-scale particles are often referred to as vapor-liquid-solid (VLS) growth proposed over forty years ago.[89] Nanowires can be viewed as a structure highly anisotropic in shape with the first and the second dimensions (i.e., diameter in the range of tens of nanometers) much smaller than the third dimension (i.e., length up to several micrometers). In a nanowire, chemical composition can be modulated along its long axis as well as in a concentric manner as seen in Fig. 9(a) and (b), respectively.[90,91] A wide variation in nanowire growth mechanisms that appears to be different from the VLS growth, for instance vapor-solid-solid (VSS) and nanowire growth without metal catalyst, has also been reported.[92,93] The growing number of literatures in this field clearly indicates that various techniques for the growth of nanowires will be continuously developed. Nanowires are expected to exhibit, as seen for semiconductor 3D islands, intriguing physical characteristics strongly associated with their unique electronic structures substantially different from those of bulk and with a large surface to volume ratio, a characteristic totally ignored in bulk. Lateral size (i.e., diameter) of semiconductor nanowires can be comparable to a variety of characteristic lengths such as exciton Bohr radius, electron/hole mean free paths, and electron phase coherence length, etc., thus in nanowires, physical properties related to these characteristics lengths are expected to be strongly influenced by the diameter of nanowires.[94] A nanowire having a large surface to volume ratio implies that the number of atoms existing on the surface of a nanowire is no longer negligible compared to the number of atoms inside of a nanowire, thus the surface of a nanowire is expected to contribute considerably to the ways physical properties are exhibited by a nanowire. For instance, such a bulk property as electrical resistivity cannot be simply defined for a material in the form of nanowire because of the presence of significant contribution from its surface to electrical transport properties. Several comprehensive reviews in the field of nanowires are available.[95] Apparently large surface to volume ratio can be either advantage or disadvantage. Large surface to volume ratio would be very beneficial when solid-state sensors that employ nanowires are envisioned. In contrast, proper handling is required to minimize undesirable physical contributions from large surface area. For instance, nonradiative carrier recombination associated with surface states that scale with total surface area would adversely affect performance of solid-state devices in which photons need to be generated by radiative carrier recombination or photons need to be converted into electrons and holes.

When nanowires are implemented as an active or a passive part of solid-state devices, advantage of using nanowires is recognized as that the volume of a single nanowire is much larger than that of, for instance, a single 3D island. Given such an anisotropic shape, nanowires are of great interest as they are expected to show strong quantum confinement for charged carriers within their cross-section (i.e., within the area defined by their diameter) while charged carriers would move fairly freely along their length. Within the scope of epitaxial growth, in particular heteroepitaxial growth, epitaxial growth of single-crystal

nanowires with diameter significantly smaller than those length-scales relevant to heteroepitaxial growth of thin films offers an opportunity of setting environment where a significant degree of elastic relaxation of misfit strain is expected,[90,96,97,98] which further offers significant advantage when, for instance, heterogeneous integration of nanowires as an optically active part on integrated electronic circuits is envisioned. (e.g., optoelecronic functionalities of group III-V compound semiconductor nanowires integrated on silicon-based electronics). It should be noted, however, the implementation of nanowires as an active part of solid-state devices would not be as straightforward as it appears. Ensembles of quantum dots have been employed as an active part of solid-state devices such as lasers, light-emitting diodes, and photodetectors by simply replacing an active semiconductor thin film by an ensemble of semiconductor quantum dots embedded in a thin film. Nanowires, however, owing to their highly anisotropic shape, non-conventional device designs need to be developed to accommodate one-dimensional shape of nanowires without hampering their unique characteristics. In the following section, our research on group III-V compound semiconductor nanowires is described with strong emphasis on developing a practical yet flexible route to integrate ensembles of nanowires onto various material platforms for a wide range of solid-state devices.

3.2 Ensembles of Nanowires on Non-single Crystal Surfaces

Employing high-quality single-crystal substrates for epitaxial growth is a prevalent practice when high-quality epitaxial films over large-area are required, as a result, significant efforts have been dedicated in developing epitaxial growth processes that produce epitaxial films that are homogeneous in composition and uniform in thicknesses over required areas.

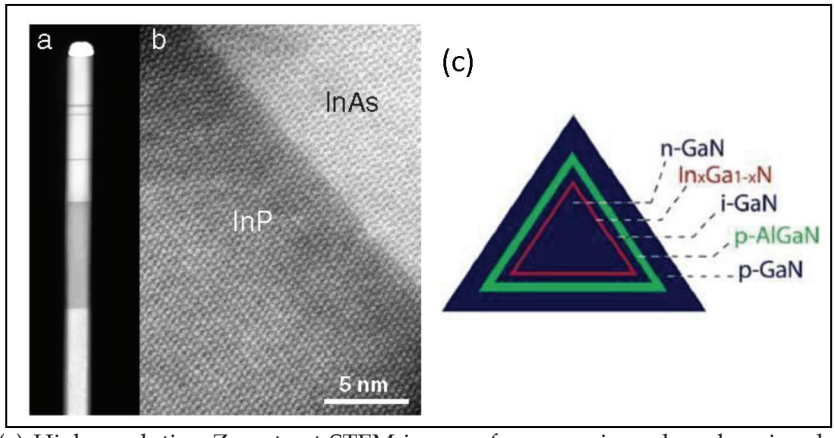


Fig. 9. (a) High resolution Z-contrast STEM image of a nanowire, where heavier elements, such as the gold particle at the top of the wire, show up brighter. The bright areas of the wire are InAs and the darker area is a 150 nm InP segment. (b) A high-resolution STEM image shows the atomically sharp interface, free from dislocations, between the InAs and the InP segments.[90] (c) Cross-sectional view of a core/multi-shell nanowire structure.[91]

Apart from large-area epitaxial films, even in the synthesis of nanometer-scale semiconductor structures, the use of single-crystal substrates as a starting platform is very common. For instance, epitaxial growth of semiconductor nanowires generally makes use of single-crystal substrates. In fact, the growth of group III-V compound semiconductor nanowires has been demonstrated almost exclusively on single-crystal substrates, including GaAs or InP nanowires grown either on single-crystal group III-V compound semiconductor substrates[99] or on single-crystal Si substrates.[99,100]

In the growth of semiconductor nanowires on single-crystal substrates, surfaces of single-crystal substrates provide epitaxial information (e.g., crystallographic symmetries and lattice constants, etc.) to "seeds" of nanowires in the early stage of the growth. A variety of growth mechanisms have been suggested for the growth of various types of nanowires.[89,101] In contrast to the growth of 3D islands (e.g., InAs on GaAs (100) surfaces), the growth of nanowires can be viewed as a significantly complicated process due partly to the mechanism that leads to quasi one-dimensional shapes resulted from the anisotropy in growth rate in two different directions (i.e., directions perpendicular and parallel to the long axis of a nanowire). Another feature contributing to various complications involved in the growth of nanowires is the presence of metal catalysts (e.g., metal-catalyzed nanowire growth with the presence of transition metal or noble metal particles).[102,103,104] Catalyst-free nanowire growths were also suggested,[105,106,107] however the level of complexities that would have to be handled in analyzing the growth of nanowires without metal-catalysts appears to be even much higher than that in metal-catalyzed nanowire growths. Therefore, the discussion on the role played by single-crystal substrates in the growth of nanowires, with or without metal catalysts, appears to be still in its infancy. However, it is conceivable that if an individual nanowire "sees" short-range atomic order within an area on the scale that is comparable to or moderately larger than the size of a nanowire itself in its early stage of evolution, an individual nanowire would not be able to tell whether it sits on a single-crystal substrate or a small crystallite present in a non-single-crystal substrate. Apparently the major difference between the two cases; nanowires grown on a single-crystal substrate and those grown on small crystallites in a non-single-crystal substrate, is that two different locations on the surface of a single-crystal substrate are correlated with its specific crystallographic translational operation whereas two different small crystallites that exist in a non-single-crystal substrate are not geometrically correlated, implying that, as long as geometrical synchronization or organization among a group of nanowires is not required, the use of a single-crystal substrate should not a required condition for the growth of nanowires.

In the following section, the idea of growing single-crystal group III-V compound semiconductor nanowires on various non-single-crystal substrates is reviewed. The structural concept of what is proposed is rather simple as schematically illustrated in Figs. 10(a)~(c). As postulated in Introduction, a starting substrate is non-single-crystal including amorphous and poly-crystal (note that there are a wide range of variations in this class of material, e.g., microcrystalline and nanocrystalline, etc.). Not constrained by the availability of appropriate single-crystal substrates, the freedom in choosing a starting substrate from a variety of materials will maximize flexibility in designing solid-state devices that utilize semiconductor nanowires. Although a non-single-crystal substrate can be chosen from such materials as glass or ceramics, one of the requirements that need to be carefully considered when a non-single-crystal substrate is selected is related to the fact that a non-single-crystal

substrate needs to withstand all succeeding process steps, in other words, a selected non-single-crystal substrate has to be physically and chemically stable at every process step towards the end of the fabrication process sequence necessary for solid-state devices that employ nanowires.

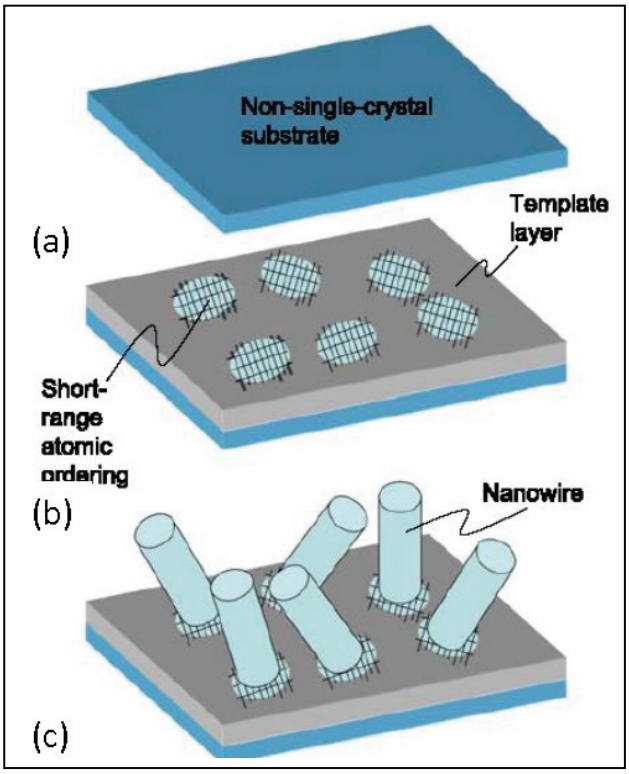


Fig. 10. Structural concept of the proposed route to grow semiconductor nanowires on a non-single-crystal substrate is schematically shown. A starting substrate in (a) can be chosen from various types of non-single-crystal including amorphous and poly-crystal. Once an appropriate non-single-crystal substrate is obtained, a template layer is prepared on the non-single-crystal substrate as in (b). The template layer can be made of a variety of materials as long as the template layer "locally" exhibits short-range atomic ordering on its surface. After nanowires are grown in (c), an ensemble of nanowires can be electrically accessible through the template layer.[109]

Amorphous silicon oxide (a-SiO_2) surfaces were used as a starting non-single-crystal substrate as shown in Fig. 10(a).[108] Subsequently, a template layer is prepared on the a-SiO_2 surface as in Fig. 10(b). The template layer can be made of a variety of materials as long as the template layer "locally" exhibits short-range atomic order on its surface. It is this short-range atomic order that is transferred to nanowires in the same sense as that in conventional epitaxial growth of a single-crystal thin film on a single-crystal substrate. As postulated earlier, it is short-range atomic order with a linear-scale comparable to or several times the diameter of nanowires that is critical for the formation of epitaxial single-crystal nanowires.

Although semiconductor single-crystal nanowires have been formed on non-single-crystal surfaces that were electrically insulating,[109,110,111] it is more advantageous, in specific applications as introduced later, to form an ensemble of single-crystal semiconductor nanowires with their one end electrically connected onto an electrically conductive template layer.

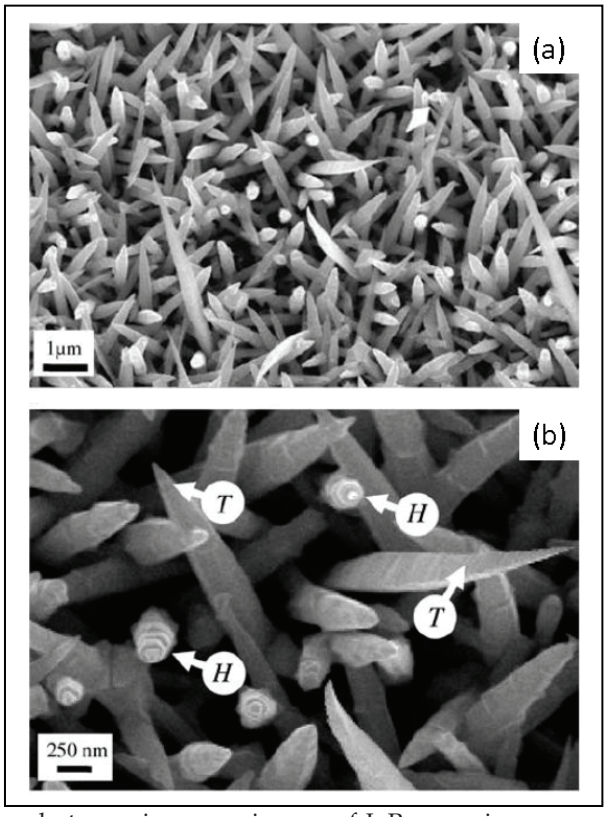


Fig. 11. Scanning electron microscope images of InP nanowires grown on a hydrogenated silicon surface. (a) top view at a low magnification, InP nanowires are randomly oriented with respect to the surface normal. (b) top view at a high magnification, the two types crystal habits are clearly seen.[112]

It is further advantageous if the electrical properties of a template layer are explicitly controlled so that an ensemble of nanowires is electrically accessed through an electrically conductive template layer as depicted in Fig. 10(c). This feature is, in fact, the one of two intriguing characteristics of our idea proposed here, that are; (1) epitaxial growth of semiconductor nanowires can be done on a non-single-crystal substrate by employing a template layer that possesses short-range atomic order and (2) an electrically conductive template layer establishes access to an ensemble of nanowires from external micrometer electrical circuits. In the selection of a material for the template layer, hydrogenated silicon (Si:H) was chosen for the demonstration of our idea because Si:H is essentially a

semiconductor that can be doped with a controlled manner to intentionally tune its electrical transport properties, which is significantly advantageous when solid-state devices that require *pn* junctions or Ohmic contacts are designed.

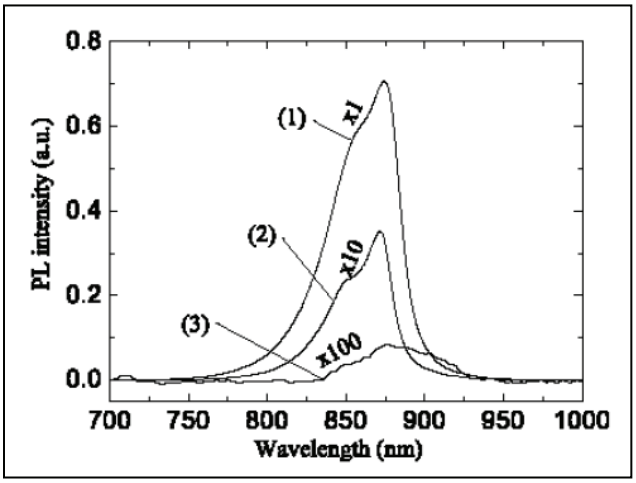


Fig. 12. A set of micro-photoluminescence spectra collected from the ensemble of InP nanowires at room temperature. The spectra (1)–(3) represent spectra collected from the InP nanowires on hydrogenated silicon at three different excitation power densities 6×103, 6×102and 6×101Wcm−2, respectively.[112]

Indium phosphide (InP) deposited with the presence of Au nanoparticles on the Si:H template layer prepared, as shown in Figs. 10, on the *a*-SiO_2 substrate was found to form nanowires. Fig. 11 is a scanning electron microscopy (SEM) image of the resulting InP nanowires. All nanowires appear to be randomly oriented as expected from the fact that the surface of the Si:H template layer on which the InP nanowires form is non-single-crystal. As seen in Fig. 11, two different crystal habits, some of which are marked with "*H*" and "*T*", were found in the ensemble of the nanowires. The nanowires denoted with *T* have a triangular cross section while the ones marked as *H* exhibit a hexagonal cross section. Since a specific crystal habit often reflects a unique crystallographic symmetry originated to that exists inside a single crystal, two different crystal habits seen in the SEM image suggest that there are two types of crystal structures co-existing in the ensemble of nanowires. The tapered feature of the both types of nanowires is clearly seen. Nominal geometrical dimensions of the two types of nanowires are summarized as; length is approximately 1.5~2 μm, lateral size at its bottom approximately 300~500 nm. The tapering angle of the nanowires (i.e., the angle between the long axis and side wall) is 33~39°. Analysis with transmission electron microscopy (TEM) revealed that the nanowires are single-crystal and there are two-types of crystal structures; zinc-blende (*ZB*) and wurtzite (*WZ*) co-existing within the ensemble of nanowires.[112] Photoluminescence studies shown in Fig. 12 suggest that the ensembles of InP nanowires are optically active.[113]

3.3 Building Blocks for Nanowire Solid-State Devices

Fabrication of robust electrical connections across nanowires is an essential step for implementing nanowires as an active part of solid-state devices. Electrical connections onto nanowires are also required to measure electrical transport properties of nanowires. Various techniques of contacting nanowires have been demonstrated by placing an individual nanowire across pre-fabricated metal electrodes on an insulating surface, or alternatively, placing an individual nanowire on an insulating surface and then, putting down a pair of electrodes onto the nanowire.[114,115] Apparently, these methods are not favorable to establishing reliable electrical contacts at large wafer scale. Nanowires bridging across a gap between semiconductor electrodes have also been demonstrated for silicon and III-V compound semiconductors,[116,117,118] in which single-crystal substrates were used. For certain applications, however the use of single-crystal substrates is not always desirable, which motivated us to develop the growth of group III-V compound semiconductor nanowires on non-single-crystal surfaces prepared on various substrates.

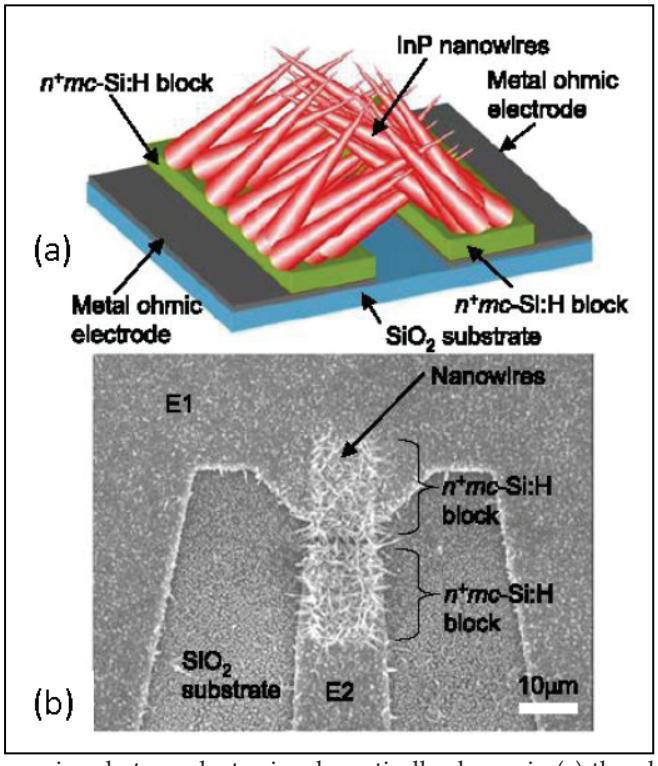


Fig. 13. InP nanowire photoconductor is schematically shown in (a) the planar structure in which InP nanowires can be selectively formed on *n+mc-Si:H surface offers great flexibility and simplification in building* devices that employ an ensemble of InP nanowires. Shown in (b) is an SEM top-view image collected on a representative InP nanowire photoconductor consisting of a pair of metal Ohmic electrodes (labeled E1 and E2) in a coplanar configuration that allows us to perform high speed testing.[109]

112 Even for nanowires on non-single-crystal surfaces, electrical contacts still need to be made for device applications. In the following section, an ensemble of InP nanowires (InP NWs) that connect two planar hydrogenated silicon (Si:H) segments that serve as templates is described. The two Si:H segments are separated by a gap and bridged by having the InP NWs either be fused and/or grow directly across the gap. Photoelectric characteristics of an ensemble of InP NWs are presented with the view toward a wide range of optoelectronic solid-state devices. Two important features obtained by combining InP NWs with Si:H are (1) InP NWs maintain all the properties of a direct bandgap semiconductor and (2) thin film characteristics of Si:H provide not only a template for the catalytic growth of InP NWs but also electrical contacts to an ensemble of InP NWs. As shown in Fig. 13(a), an amorphous surface was prepared by growing a 2 μm thick silicon dioxide (SiO_2) on a 4 inch Si(100) substrate. After forming Pt/Ti electrodes onto the SiO_2 surface, a heavily doped *n*-type Si:H film (*n*+-Si:H) was deposited onto the patterned Pt/Ti electrodes. The *n*+-Si:H film was then patterned into two segments spatially separated and electrically isolated from each other. The processed wafer was then coated with Au nanoparticles. Finally, InP was deposited by low-pressure metal organic chemical vapor deposition on the processed wafer. InP grew into randomly oriented InP NWs formed selectively onto the *n*+-Si:H segments. The NWs bridged the gap by either fusing themselves or going across the gap. A planar structure in which InP NWs are selectively formed on *n*+-Si:H surface offers great flexibility and simplification in building solid-state devices that employ an ensemble of InP NWs. This planar structure (InP NW photoconducotor) was used to evaluate photoconductive properties of an ensemble of InP NWs. Shown in Fig. 13(b) is a scanning electron microscope image (top view) of the InP NW photoconductor on which a pair of electrodes (labeled E1 and E2) in a coplanar configuration is seen. A large number of randomly oriented nanowires are clearly seen on the two *n*+-Si:H segments. Detailed inspections by viewing the photoconductor from the side revealed that nanowires approaching from the two facing *n*+-Si:H segments collided and "fused" together in free space.

DC current-voltage (*IV*) characteristics of the InP NWs photoconductor were measured at room temperature.[119] Photoresponse was also obtained by illuminating the device with a laser at 633, 780, or 1550 nm. A 633 nm laser with various optical power densities in the range of 3~25 W/cm^2 (corresponding optical power of $6x10^{-7}$~$5x10^{-6}$ W) was used to illuminate the gap between the two *n*+-Si:H segments with InP NWs. As a control, the device prior to the InP NW growth was also evaluated. Prior to the InP NW growth, the measured photocurrent was <0.5 nA at all bias voltages with the highest illumination level, which was roughly three orders of magnitude smaller than those obtained from the device with InP NWs, suggesting the presence of negligible leak current in the reference sample without InP NWs. A semi-log plot in the inset of Fig. 14 shows the DC *IV* characteristics of the InP NW photoconductor obtained under different optical powers ($5x10^{-6}$ W for the top curve in black and no illumination for the bottom curve in dark green). The *IV* characteristics were remarkably symmetric with respect to 0 V and essentially Ohmic, revealing a clear contrast to nonlinear characteristics observed for nanowires in other configurations.[120] More than an order of magnitude increase in current at ±5 V, compared to that without illumination, was observed when the device was illuminated at $5x10^{-6}$ W. In the main plot of Fig. 14, a set of data points in different colors at a specific incident optical power represents resistivities of the InP NWs obtained at various bias voltages. As in a metal-semiconductor-metal photoconductor, the resistivity at a specific bias voltage

decreases as the incident optical power increases. The resistivity under the illumination at incident optical power >5x10^{-6} W seems to saturate, suggesting the resistivity is limited by the finite total volume provided by a limited number of nanowires available for the generation of excess electrons and holes. Unlike other nanowire-based photodetectors,[121] the photoconductor response presented above is insensitive to the polarization of the incident light, which is most likely due to the random orientation of the InP NWs. We believe that the arrangement of incorporating an ensemble of nanowires as an active part in solid-state devices presented in this study could be a new route to design solid-state devices that employ nanowires with highly anisotropic shapes.

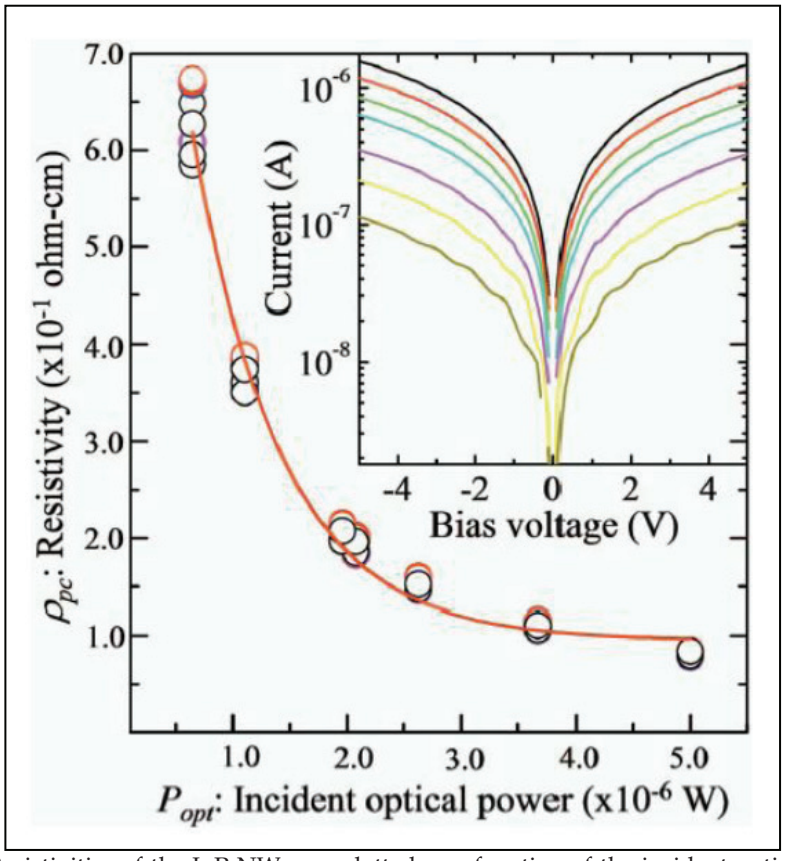


Fig. 14. Resistivities of the InP NWs are plotted as a function of the incident optical power by which the InP NW active region was illuminated. A set of data points in different colors at a specific incident optical power represents the resistivities obtained at different bias voltages ranging from −5 to 5 V with an increment of 1 V. Plotted in the inset are current-voltage characteristics collected under different levels of the incident optical power 5x10−6W for the top curve and 0 W for the bottom curve.119

4. Summary

InAs three-dimensional (3D) islands grown by molecular beam epitaxy on GaAs are described. The two-dimensional to three-dimensional morphological transition was studied with two complementary analytical techniques; scanning tunneling microscope and photoluminescence/photoluminescence excitation spectroscopy, to identify key features that exist at different stages of morphological evolution. The re-entrant behavior of small quasi-3D clusters existing prior to the appearance of 3D islands suggests that the quasi-3D clusters act as precursors of the 3D islands. Size evolution and vertical alignment of InAs 3D islands in a single and multiple stacks are illustrated. The lateral size and size dispersion were found to first increase drastically with a small amount of additional InAs deposition and then decrease and saturate, indicating the onset of a natural tendency for size equalization. The vertical alignment of multiple stacks of InAs 3D islands is a result of strain field generated by the 3D islands embedded within a GaAs matrix. All the insights we learned in the series of experiments would help us find a way to implement a group of 3D islands as a part of solid-state devices.

InP nanowires grown by metal organic chemical vapor deposition on non-single-crystal surfaces are described. Unlike conventional epitaxial thin films, the proposed route for growing nanowires requires no single-crystal surfaces. In principle, only short-range atomic order, in contrast to long-range atomic order required for the growth of epitaxial thin films, is necessary for nanowires. A template layer that possesses short-range atomic order prepared on a non-single-crystal surface is employed. Ensembles of InP nanowires were found to be single-crystal and electrically/optically active. DC electrical properties of simple photoconductors that employs an ensemble of InP nanowires were fabricated and characterized under light illumination. Our proposed path to implement semiconductor nanowires would unlock a path towards disruptive applications that require III–V compound semiconductors functionally integrated onto various material platforms.

5. Acknowledgement

I am grateful to many colleagues and peers in the course of gathering elemental building blocks for this chapter. Special thanks are due to Prof. A. Madhukar, Dr. Q. Xie, Dr. T. R. Ramachandran, Dr. R. Heitz, Dr. R. S. Williams, Dr. S.-Y. Wang, Prof. M. S. Islam, and Dr. Logeeswaran, V. J.

6. References

[1]Burton, W. K., Cabrera, N., Frank, F. C. (1951). The growth of crystals and the equilibrium structure of their surfaces, *Phil. Trans. Roy. Soc.* Vol. A234, pp. 299-358

[2]Gallet, F., Balibar, S., Rolley, E., Nozieres, P. (1987). The dynamic broadening of the roughening transition, *Japanese Journal of Applied Physics*, Sup. Vol. 26, pp. 389-90

[3]van der Merwe, J. H. (1993). Theoretical Considerations in Growing Uniform Epilayers, *Interface Science*, Vol. 1, pp. 77-86

[4] Venables, J. A. (1973). Rate equation approaches to thin film nucleation kinetics, *Philosophical Magazine,* Vol. 27, pp. 697-738

[5] Frankl, D. R., Venables, J. A. (1970). Nucleation on substrates from the vapour phase, *Advances in Physics*, Vol. 19, pp. 409-456

[6] Halpern, V. (1969). Cluster growth and saturation island densities in thin-film growth, *Journal of Applied Physics*, Vol. 40, pp. 4627-4636

[7] Joyce, B. A., Bradley, R. R., Booker, G. R. (1967), A study of nucleation in chemically grown epitaxial silicon films using molecular beam techniques. III. Nucleation rate measurements and the effect of oxygen on initial growth behavior, *Philosophical Magazine*, Vol. 15, pp. 1167-1187

[8]Lewis, B. (1971). Physical processes in epitaxial growth, *Thin Solid Films*, Vol. 7, pp. 179-217

[9]Stowell, M. J. (1972). Direct observations of epitaxial growth, *Thin Solid Films*, Vol. 12, pp. 341-54

[10]Stowell, M. J. (1970). Capture numbers in thin film nucleation theories, *Philosophical Magazine*, Vol. 26, pp. 349-60

[11]Zinsmeister, G. (1971). Theory of thin film condensation. IV. Influence of a variable collision factor, *Thin Solid Films*, Vol. 7, pp. 51-75

[12]Shockley, W., Read, W. T. (1949). Quantitative predictions from dislocation models of crystal grain boundaries, *Physical Review*, Vol. 75, pp. 692

[13]Read, W. T., Shockley, W. (1950). Dislocation models of crystal grain boundaries, *Physical Review*, Vol. 78, pp. 275-289

[14]van der Merwe, J. H. (1950). On the stresses and energies associated with inter-crystalline boundaries, *Proceedings of the Physical Society*, Vol. A63, pp. 616-637

[15]van der Merwe, J. H. (1962). Crystal interfaces. Part I. Semi-Infinite Crystals, *Journal of Applied Physics*, Vol. 34, pp. 117-122

[16]van der Merwe, J. H. (1963). Crystal interfaces. Part II. Finite overgrowths, *Journal of Applied Physics*, Vol. 34, pp. 123-127

[17]People, R., Bean, J. E. (1985). Calculation of critical layer thickness versus lattice mismatch for Ge_xSi_{1-x}/Si strained-layer heterostructures, *Applied Physics Letters*, Vol. 47, 322-324

[18] Matthews, J. W., Blakeslee, A. F. (1976). Defects in epitaxial multilayers III. Preparation of almost perfect multilayers, *Journal of Crystal Growth*, Vol. 32, pp. 265-273

[19]van der Merwe, J. H., Jesser, W.A. (1988). The prediction and confirmation of critical epitaxial parameters, *Journal of Applied Physics*, Vol. 64, pp. 4968-4974

[20]van der Merwe, J. H. (1970). Equilibrium structure of a thin epitaxial film, *Journal of Applied Physics*, Vol. 41, pp. 4725-4731

[21]Matthews, J. W. (1972). Influence of misfit dislocations on the alignment of epitaxial thin films, *Thin Solid Films*, Vol. 12, pp. 243-246

[22]Matthews, J. W., Blakeslee, A. E. (1974). Defects in epitaxial multilayers I. Misfit Dislocations, *Journal of Crystal Growth*, Vol. 27, pp. 118-125

[23]Matthews, J. W., Blakeslee, A. E. (1975). Defects in epitaxial multilatyers II. Dislocation pile-ups, threading dislocations, slip lines and cracks, *Journal of Crystal Growth*, Vol. 29, pp. 273-280

[24]Matthews, J. W., Blakeslee, A. E., Mader, S. (1976). Use of misfit strain to remove dislocations from epitaxial thin films, *Thin Solid Films*, Vol. 33, pp. 253-266

[25]Matthews, J. W., Jackson, D. C., Chambers, A. (1975). Effect of coherency strain and misfit dislocations on the mode of growth of thin films, *Thin Solid Films*, Vol. 26, pp. 129-134

[26]van der Merwe, J. H. (2001). Interfacial energy: bicrystals of semi-infinite crystals, *Progress in Surface Science*, Vol. 67, pp. 365-381

[27]Fletcher, N. H. (1964). Crystal interfaces, *Journal of Applied Physics*, Vol. 35, pp. 234-240

[28]Fletcher, N. H. (1967). Structure and energy of crystal interfaces II. A simple explicit calculation, *Philosophical Magazine*, Vol. 16, pp. 159-164

[29]Fletcher, N. H., Adamson, P. L. (1966). Structure and energy of crystal interfaces. I. Formal development, *Philosophical Magazine*, Vol. 14, pp. 99-110

[30]Tsao, J. Y., Dodson, B. W., Pieraux, S. T., Cornelson, D. M. (1987). Critical stresses for Si_xGe_{1-x} strained-layer plasticity, *Physical Review Letters*, Vol. 59, pp. 2455-2458

[31]Dodson, B. W., Tsao, J. Y. (1987). Relaxation of strained-layer semiconductor structures via plastic flow, *Applied Physics Letters*, Vol. 51, pp. 1325-1327

[32]Ghaisas, S. V., Madhukar, A. (1986). Role of Surface Molecular Reactions in Influencing the Growth Mechanism and the Nature of Non-Equilibrium Surfaces: A Monte-Carlo Study of Molecular Beam Epitaxy, *Physical Review Letters*, Vol. 56, pp.1066-1069

[33]Ghaisas, S. V., Madhukar, A. (1989). Kinetic Aspects of Growth Front Surface Morphology and Defect Formation During MBE Growth of Strained Thin Films, *Journal of Vacuum Science and Technology*, Vol. B7, pp. 264-268

[34]Ghaisas, S. V., Madhukar, A. (1988). Influence of compressive and tensile strain on growth mode during epitaxical growth: a computer simulation study, *Applied Physics Letters*, Vol. 53, pp. 1599-1601

[35]Arthur, J. R. (1974). Surface stoichiometry and structure of GaAs, *Surface Science*, Vol. 43, pp. 449-461

[36]Foxon, C. T., Joyce, B. A. (1975). Interaction kinetics of As_4 and Ga on {100} GaAs surfaces using a modulated molecular beam technique, *Surface Science*, Vol. 50, pp. 434-450

[37]Madhukar, A., Ghaisas, S. V. (1988). The nature of molecular beam epitaxial growth examined via computer simulations, *CRC Critical Reviews in Solid State and Materials Sciences*, Vol. 14, pp. 1-130

[38]Pashley, D. W. (1965). The nucleation, growth, structure and epitaxy of thin surface films, *Advances in Physics*, Vol. 14, pp. 327-410

[39]Matthews, J. W. (1972). Misfit dislocations and the alignment of epitaxial islands, *Surface Science*, Vol. 31, pp. 241-256

[40]van der Merwe, J. H., Bauer, E., (1989). Influence of misfit and bonding on the mode of growth in epitaxy, *Physical Review*, Vol. B39, pp. 3632-3641

[41]Stoop, L. C. A., van der Merwe, J. H. (1973). A simple model for layered growth in small epitaxial islands, *Thin Solid Films*, Vol. 17, pp. 291-309

[42]Shiramine, K., Muto, S., Shibayama, T., Sakaguchi, N., Ichinose, H., Kozaki, T., Sato, S., Nakata, Y., Yokoyama, N., Taniwaki, M. (2007). Tip artifact in atomic force microscopy observations of InAs quantum dots grown in Stranski–Krastanow mode, *Journal of Applied Physics*, Vol. 101, pp. 033527-5

[43]Bimberg, D., Grundmann, M., Ledentsov, N. N. (1988). Quantum Dot Heterostructures, John Willey & Sons, West Sussex, England

[44]Eaglesham, D. J. , Cerullo, M. (1990). Dislocation-free Stranski-Krastanow growth of Ge on Si(100), *Physical Review Letters*, Vol. 64, pp. 1943-1946

[45]Mo, Y.-W., Savage, D. E., Swartzentruber, B. S., Lagally, M. G., Kinetic pathway in Stranski-Krastanov growth of Ge on Si(001), *Modern Physics Letters B*, Vol. 4, pp. 1379-1384

[46]Guha, S., Madhukar, A., Rjkumar, K. C. (1990). Onset of incoherency and defect introduction in the initial stages of molecular beam epitaxial growth of highly strained $In_xGa_{1-x}As$ on GaAs(100), *Applied Physics Letters*, Vol. 57, pp. 2110-2112

[47]Houzay, F., Guille, C., Moison, J. M., Henoe, P., Barthe, F., (1987). First stages of the MBE growth of InAs on (001) GaAs, *Journal of Crystal Growth*, Vol. 81, pp. 67-72

[48]Lewis, B.F., Lee, T.C., Grunthaner, F.J., Madhukar, A., Fernandez, R., Maserjian, J. (1984). RHEED oscillation studies of MBE growth kinetics and lattice mismatch strain-induced effects during InGaAs growth on GaAs(100), *Journal of Vacuum Science & Technology B*, Vol. 2, pp. 419-424

[49]Snyder, C.W., Orr, B.G., Kessler, D., Sander, L. M. (1991). Effect of strain on surface morphology in highly strained InGaAs films, *Physical Review Letters*, Vol. 66, pp. 3032-3035

[50]Madhukar, A., Xie, Q., Chen, P., Konkar, A. (1994). Nature of strained InAs three-dimensional island formation and distribution on GaAs(100), *Applied Physics Letters*, Vol. 64, pp. 2727-2729

[51]Kobayashi, N. P., Ramachandran, T. R., Chen, P., Madhukar, A. (1996). In Situ atomic force microscope studies of the evolution of InAs three-dimensional islands on GaAs(001), *Applied Physics Letter*, Vol. 68, pp. 3299-3301

[52]Gerard, J.M., Genin, J.B., Lefebvre, J., Moison, J. M., Lebouche, N., Barthe, F. (1994). Optical investigation of the self-organized growth of InGaAs/GaAs quantum boxes, *Journal of Crystal Growth*, Vol. 150, pp. 351-356

[53]Leonard, D., Pond, K., Petroff, P. M. (1994). Critical layer thickness for self-assembled InAs islands on GaAs, *Physical Review B*, Vol. 50, pp. 11687-11692

[54]Polimeni, A., Patane, A., Capizzi, M., Martelli, F., Nasi, L., Salviati, G. (1996). Self-aggregation of quantum dots for very thin InAs layers grown on GaAs, *Physical Review B*, Vol. 53, pp. R4213-16

[55]Bressler-Hill, V., Lorke, A., Varma, S., Petroff, P. M., Pond, K., Weinberg, W. H. (1994). Initial stages of InAs epitaxy on vicinal GaAs(001)-(2*4), *Physical Review B*, Vol. 50, pp. 8479-8487

[56]Xie, Q., Madhukar, A., Chen, P., Kobayashi, N. P. (1995). Vertically self-organized InAs quantum box islands on GaAs(100), *Physical Review Letters*, Vol. 75, pp. 2542-2545

[57]Grundmann, M., Christen, J., Ledentsov, N. N., Bohrer, J., Bimberg, D., Ruvimov, S. S., Werner, P., Richter, U., Gosele, U., Heydenreich, J., Ustinov, V. M., Egorov, A. Yu., Zhukov, A. E., Kopev, P. S., Alferov, Zh. L. (1995). Ultranarrow luminescence lines from single quantum dots, *Physical Review Letters*, Vol. 74, pp. 4043-4046

[58]Xie, Q., Kalburge, A., Chen, P., Madhukar, A. (1996). Observation of lasing from vertically self-organized InAs three-dimensional island quantum boxes on GaAs (001), *IEEE Photonics Technology Letters*, Vol. 8, pp. 965-967

[59]Bimberg, D., Ledentsov, N. N., Grundmann, M., Kirstaedter, N., Schmidt, O. G., Mao, M. G., Ustinov, V. M., Egorov, A. Y., Zhukov, A. E., Kopev, P. S., Alferov, Zh. L. Ruvimov, S. Gosele, U., Heydenreich, J. (1996). InAs-GaAs quantum pyramid lasers: in situ growth, radiative lifetimes and polarization properties, *Japanese Journal of Applied Physics*, Part 1,Vol. 35, pp. 1311-1319

[60]Priester, C., Lannoo, M (1995). Origin of self-assembled quantum dots in highly mismatched heteroepitaxy, *Physical Review Letters*, Vol. 75, pp. 93-96

[61]Ramachandran, T. R., Heitz, R., Chen, P., Madhukar, A. (1996). Mass transfer in Stranski-Krastanow growth of InAs on GaAs, *Applied Physics Letters*, Vol. 70, pp. 640-642

[62]Ramachandran, T.R., Heitz, R., Kobyaahsi, N. P., Kalurge, A., Yu, W., Chen, P., Madhukar, A. (1997). Re-entrant behavior of 2D to 3D morphology change and 3D island lateral size equalization via mass exchange in Stranski-Krastanow growth: InAs on GaAs(001), *Journal of Crystal Growth*, Vol. 175/176, pp. 216-223

[63]Suekane, O., Hasegawa, S., Tanaka, M., Okui, T., Nakashima, H. (2002). Annealing effect on InAs islands on GaAs(001) substrates studied by scanning tunneling microscopy, *Applied Surface Science*, Vol. 190, pp. 218-221

[64] Lu, Z. D., Xu, J. Z., Zheng, B. Z., Xu, Z. Y., Ge, W. K. (1999). Effect of growth interruption on the optical properties of InAs/GaAs quantum dots, *Solid State Communications*, Vol. 109, pp. 649-653

[65]Patella, F., Arciprete, F., Placidi, E., Nufris, S., Fanfoni, M., Sgarlata, A., Schiumarini, D., Balzarotti, A., (2002). Morphological instabilities of the InAs/GaAs(001) interface and their effect on the self-assembling of InAs quantum-dot arrays, *Applied Physics Letters*, Vol. 81, pp. 2270-2272

[66]Schwoebel, R. L., Shipsey, E. J. (1966). Step Motion on Crystal Surfaces, *Journal of Applied Physics*, Vol. **37**, pp. 3682-3686

[67]Xie, Q., Chen, P., Kalburge, A., Ramachandran, T. R., Nayfonov, A., Konkar, A., Madhukar, A. (1995). Realization of optically active strained InAs island quantum boxes on GaAs(100) via molecular beam epitaxy and the role of island induced strain fields, *Journal of Crystal Growth*, Vol. 150, pp. 357-363

[68]Heitz, R., Ramachandran, T. R., Kalburge, A., Xie, Q., Mukhametzhanov, I., Chen, P., Madhukar, A. (1997). Obaservation of reentrant 2D to 3D morphology transition in highly strained epitaxy: InAs on GaAs, *Physical Review Letters*, Vol. 78, pp. 4071-4074

[69]Heitz, R., Grundmann, M., Ledentsov, N. N., Eckey, L., Veit, M., Bimberg, D., Ustinov, V. M., Egorov, A. Yu., Zhukov, A. E., Kop'ev, P. S., Alferov, Zh. I. (1996). Multiphonon-relaxation processes in self-organized InAs/GaAs quantum dots, *Applied Physics Letters*, Vol. 68, pp. 361-363

[70]Krzyzewski, T. J., Joyce, P. B., Bell, G. R., Jones, T. S. (2002). Role of two- and three-dimensional surface structures in InAs-GaAs(001) quantum dot nucleation, *Physical Review B*, Vol. 66, pp. 121307-1-4

[71]Costantini, G., Rastelli, A., Manzano, C., Acosta-Diaz, T. P., Katsaros, G., Songmuang, R., Schmidt, O. G., Knewl, O. V., Kern, K. (2005). Pyramids and domes in the InAs/GaAs(001) and Ge/Si(001) systems, *Journal of Crystal Growth*, Vol. 278, pp. 38-45

[72] Patella, Fl, Nufris, S., Arciprete, F., Fanfoni, M., Placidi, E., Sgarlata, Al., Blzarotti, A. (2003). Tracing the two- to three-dimensional transition in the InAs/GaAs(001) heteroepitaxial growth, *Physical Review B*, Vol. 67, pp. 205308-1-5

[73]Arciprete, F., Placidi, E., Sessi, V., Fanfoni, M., Patella, F., Balzarotti, A., Sessi, M., Fanfoni, F., Patella, Balzarotti, A. F., Arciprete, E., Placidi, V. (2006). How kinetics drives the two- to three-dimensional transition in semiconductor strained heterostructures: The case of InAs/GaAs(001), *Applied Physics Letters*, Vol. 89, pp. 041904 -1-3

[74]Heyn, Ch., Endler, D., Zhang, K., Hansen, W. (2000). Formation and dissolution of InAs quantum dots on GaAs, *Journal of Crystal Growth*, Vol. 210, pp. 421-428

[75]Drexler, H., Leonard, D., Hansen, W., Kotthaus, J. P., Petroff, P. M. (1994). Spectroscopy of quantum levels in charge-tunable InGaAs quantum dots, *Physical Review Letters*, Vol. 73, pp. 2252-2255

[76]Ponchet, A., LeCorre, A., L'Haridon, H., Lambert, B., Salaun, S. (1995). Relationship between self-organization and size of InAs islands on InP(001) grown by gas-source molecular beam epitaxy, *Applied Physics Letters*, Vol. 67, pp. 1850-1852

[77]Moison, J. M., Houzay, F., Barthe, F., Leprince, L., Andre, E., Vatel, O. (1994). Self-organized growth of regular nanometer-scale InAs dots on GaAs, *Applied Physics Letters*, Vol. 64, pp. 196-198

[78]Madhukar, A., Chen, P., Xie, Q., Konkar, A., Ramachandran, T. R., Kobayashi, N. P., Viswanathan, R. (1995). Low Dimensional Structures prepared by Epitaxial Growth or Regrowth on Patterned Substrates , *Proceedings of the NATO Advanced Workshop*, pp. 19-33, Ringberg Castle, Germany, February, 1995, The Netherlands

[79]Chen, P., Xie, Q., Madhukar, A., Chen, L., Konkar, A. (1994). Mechanisms of strained island formation in molecular-beam epitaxy of InAs on GaAs(100), *Journal of Vacuum Science & Technology B,* Vol. 12, pp. 2568-2573

[80]Gong, Q., Liang, J. B., Xu, B., Ding, D., Li, H. X., Jiang, C., Zhou, W., Liu, F. Q., Wang, Z. G., Qiu, X. H., Shang, G. Y., Bai, C. L. (1998). Analysis of atomic force microscopic results of InAs islands, *Journal of Crystal Growth*, Vol. 192, pp. 376-380

[81]Drucker, J. (1993). Coherent islands and microstructural evolution, *Physical Review B*, Vol. 48, pp. 18203-18206

[82]da Silvaa, M. J., Quivya, A. A., Gonz!alez-Borrerob, P. P., Marega Jr., E., Leitea, J. R. (2002). Atomic-force microscopy study of self-assembled InAs quantum dots along their complete evolution cycle, *Journal of Crystal Growth*, Vol. 241, pp. 19–30

[83]Jesson, D. E., Chen, G., Chen, K. M., Pennycook, S. J. (1998). Self-limiting growth of strained faceted islands, *Physical Review Letters*, Vol. 80, pp. 5156-5459

[84]Wang, L. G., Kratzer, P., Moll, N., Scheffler, M. (2000). Size, shape, and stability of InAs quantum dots on the GaAs(001) substrate, *Physical Review B*, Vol. 62, pp. 1897-1904

[85]Fafard, S., Leonard, D., Merz, J. L., Petroff, P. M. (1994). Selective excitation of the photoluminescence and the energy levels of ultrasmall InGaAs/GaAs quantum dots, *Applied Physics Letters*, Vol. 65, pp. 1388-1390

[86]Eisele, H., Lenz, A., Heitz, R., Timm, R., Dähne, M., Temko, Y., Suzuki, T., Jacobi, K. (2008). Change of InAs/GaAs quantum dot shape and composition during capping, *Journal of Applied Physics*, Vol. 104, pp. 124301-5

[87]Bouzaiene, L., Ilahi, B., Sfaxi, L., Hassen, F., Maaref, H., Marty, O., Dazord, J. (2004). Tuning vertically stacked InAs/GaAs quantum dot properties under spacer thickness effects for 1.3 μm emission, *Appl. Physics*, Vol., A79, pp. 587–591

[88]Heinrichsdorff, F., Mao, M. -H., Kirstaedter, N., Krost, A., Bimberg, D., Kosogovb, A. O., Werner, P. (1997). Room-temperature continuous-wave lasing from stacked InAs/GaAs quantum dots grown by metalorganic chemical vapor deposition, *Applied Physics Letters*, Vol. 71, pp. 22-24

[89]Wagner, R. S., Ellis, W. C. (1964), Vapor-liquid-solid mechanism of single crystal growth, *Applied Physics Letters*, Vol. 4, pp. 89-90

[90]Larsson, M. W., Wagner, J. B., Wallin, M., Hakansson, P., Froberg, L. E., Samuelson, L., Wallenberg, L. R. (2007). Strain mapping in free-standing heterostructured wurtzite InAs/InP nanowires, *Nanotechnology*, Vol. 18, pp. 015504-1-8

[91]Qian, F., Gradecak, S., Li, Y., Wen, C. –Y., Lieber C. M., (2005). Core/Multishell Nanowire Heterostructures as Multicolor, High-Efficiency Light-Emitting Diodes, *Nano Letters*, Vol. 5, pp. 11-2287-2291

[92]Persson, A. I., Larsson, M. W., Stenstrom, S., Ohlsson, B. J., Samuelson, L., Wallenberg, L. R. (2004). Solid-phase diffusion mechanism for GaAs nanowire growth, *Nature Materials*, Vol. 3, pp. 677-681

[93]Novotny, C. J., Yu, P. K. L. (2005). Vertically aligned, catalyst-free InP nanowires grown by metalorganic chemical vapor deposition, *Applied Physics Letters*, Vol. 87, pp. 203111-1-3

[94]Guichard, A. R., Barsic, D. N., Sharma, S., Kamins, T. I., Brongersma, M. L. (2006). Title: Tunable light emission from quantum-confined excitons in $TiSi_2$ -catalyzed silicon nanowires, *Nano Letters*, Vol. 6, pp. 2140-2144

[95]Dick, K. A. (2008). A review of nanowire growth promoted by alloys and non-alloying elements with emphasis on Au-assisted III-V nanowires, *Progress in Crystal Growth and Characterization of Materials*, Vol. 54, pp. 138-73 and references herein.

[96]Thelander, C., Agarwal, P., Brongersma, S. Thelander, C. Agarwal, P., Brongersma, S., Eymery, J., Feiner, L. F., Forchel, A., Scheffler, M., Riess, W., Ohlsson, B. J., Gösele, U., Samuelson, L. (2006). Nanowire-based one-dimensional electronics, *Materials Today*, Vol. 9, pp. 28-35

[97]Ertekin, E., Greaney, P.A., Chrzan, D.C., Sands, T. D. (2005). Equilibrium limits of coherency in strained nanowire heterostructures, *Journal of Applied Physics*, Vol. 97, pp. 114325-1-10

[98]Chuang, L. C., Moewe, M., Chase, C., Kobayashi, N. P., Chang-Hasnain, C., Crankshaw, S. (2007). Critical diameter for IIIVnanowires grown on lattice-mismatched substrates, *Applied Physics Letters*, Vol. 90, pp. 043115-1-3

[99]Yazawa, M., Koguchi, M., Hirutna, K. (1991). Heteroepitaxial ultrafine wire-like growth of InAs am GaAs substrates, *Applied Physics Letters*, Vol. 58, pp. 1080-1082

[100]Mårtensson, T., Patrik C., Svensson, T., Wacaser, B. A., Larsson, M. W., Seifert, W., Deppert, K., Gustafsson, A., Wallenberg, L. R., Samuelson, L. (2004). Epitaxial III-V Nanowires on Silicon, *Nano Letters*, Vol. 4, pp. 1987-1990

[101] Persson, A. I., Ohlsson, B. J., Jeppesen, S., Samuelson, L. (2004). Growth mechanisms for GaAs nanowires grown in CBE, Journal of Crystal Growth, Vol. 272, pp. 167-174

[102]Paiano, P., Prete, P., Loverginea, N., Mancini, A. M. (2006). Size and shape control of GaAs nanowires grown by metalorganic vapor phase epitaxy using tertiarybutylarsine, *Journal of Applied Physics*, Vol. 100, pp. 094305-1-4

[103]Kamins, T. I., Sharma, S., Yasseri, A. A., Li, Z., Straznicky, J. (2006). Metal-catalysed, bridging nanowires as vapor sensors and concept for their use in a sensor system, *Nanotechnology*, Vol. 17, pp. s291-297

[104]Sko1ld, N., Karlsson, L. S., Larsson, M. W., Pistol, M. –E., Seifert, W., Trägårdh, J., Samuelson, L. (2005). Growth and optical properties of strained GaAs-GaxIn1-xP core-shell nanowires, *Nano Letters*, Vol. 5, pp. 1943-1947

[105]Novotny, C. J., Yu, P. K. L. (2005). Vertically aligned, catalyst-free InP nanowires grown by metalorganic chemical vapor deposition, *Applied Physics Letters*, Vol. 87, pp. 203111-1-3

[106]Mattila, M., Hakkarainen, T., Lipsanen, H., Jiang, H., Kauppinen, E. I. (2006). Catalyst-free growth of In(As)P nanowires on silicon, *Applied Physics Letters*, Vol. 89, pp. 063119-1-3

[107]Noborisaka, J., Motohisa, J., Fukui, T. (2005). Catalyst-free growth of GaAs nanowires by selective-area metalorganic vapor-phase epitaxy, *Applied Physics Letters*, Vol. 86, pp. 213102-1-3

[108]Kobayashi, N. P., Mathai, S., Li, X., Logeeswaran, V. J., Islam, M. S., Lohn, A., Onishi, T., Stranznicky, J., Wang, S. –Y., Williams, R. S. (2009). Ensembles of indium phosphide nanowires: physical properties and functional devices integrated on non-single crystal platforms, *Applied Physics A*, Vol. 95, pp. 1005-1013.

[109]Gudiksen, M. S., Wang, J., Lieber, C. M. (2002). Size-Dependent Photoluminescence from Single Indium Phosphide Nanowires, *Journal of Physical Chemistry B*, Vol. 106, pp. 4036-9

[110]Vaddiraju, S., Mohite, A., Chin, A., Meyyappan, M., Sumanasekera, G., Alphenaar, B. W., Sunkara, M. K. (2005). Mechanisms of 1D Crystal Growth in Reactive Vapor Transport: Indium Nitride Nanowires, *Nano Letters*, Vol. 5, pp. 1625-1631

[111]Gao, P. X., Lao, C. S., Hughes, W. L., Wang, Z. L. (2005). Three-dimensional interconnected nanowire networks of ZnO, *Chemical Physics Letters*, Vol. 408, pp. 174-178

[112]Kobayashi, N.P., Wang, S.-Y., Santori, C., Williams, R. S. (2006) Growth and characterization of indium phosphide single-crystal nanoneedles on microcrystalline silicon surfaces, *Applied Physics A*, Vol. A85, pp. 1-6

[113]Kobayashi, N.P., Shih-Yuan Wang, Santori, C., Williams, R. S. (2007). Indium phosphide nanoneedles on non-single crystalline semiconductor surfaces, *Japanese Journal of Applied Physics*, Part 1, Vol. 46, pp. 6346-6351

[114]Pettersson, H., Tragardh, J., Persson, A.I., Landin, L., Hessman, D., Samuelson, L. (2006). Infrared photodetectors in heterostructure nanowires , *Nano Letters*, Vol. 6, pp. 229-232

[115]Salfi, J., Philipose, U., de Sousa, C. F., Aouba, S., Ruda, H. E. (2006). Electrical properties of Ohmic contacts to ZnSe nanowires and their application to nanowire-based photodetection, *Applied Physics Letters*, Vol. 89, pp. 261112-1-3

[116]Haraguchi, K., Hiruma, K., Katsuyama, T., Tominaga, K., Shirai, M., Shimada, T. (1996). Self-organized fabrication of planar GaAs nanowhisker arrays, *Applied Physics Letters*, Vol. 69, pp. 386-387

[117]Islam, M.S., Sharma, S. , Kamins, T. I., Williams, R. S. (2004). Ultrahigh-density silicon nanobridges formed between two vertical silicon surfaces, *Nanotechnology*, Vol. 15, pp. L5-8

[118]Yi, S. S., Girolami, G., Amano, J., Islam, M. S., Sharma, S., Kamins, T. I. (2006). InP nanobridges epitaxially formed between two vertical Si surfaces by metal-catalyzed chemical vapor deposition, *Applied Physics Letters*, Vol. 89, pp. 133121-1-3

[119]Kobayashi, N. P., Logeeswaran, V. J., Islam, M. S., Li, X., Straznicky, J. ,Wang, S. –Y., Williams, R. S., Chen, Y. (2007). Hydrogenated microcrystalline silicon electrodes connected by indium phosphide nanowires, *Applied Physics Letters*, Vol. 91, pp. 113116-1-3

[120]Law, J. B. K., Thong, J. T. L. (2006). Simple fabrication of a ZnO nanowire photodetector with a fast photoresponse time, *Applied Physics Letters*, Vol. 88, pp. 133114-1-3

[121]Hayden, O., Agarwal, R., Lieber, C. M. (2006). Nanoscale avalanche photodiodes for highly sensitive and spatially resolved photon detection, *Nature Materials*, Vol. 5, pp. 352-356

5

Analytic Approach to the Operation of RTD Ternary Inverters Based on MML

Juan Núñez, José M. Quintana & María J. Avedillo
Microelectronics Institute of Seville (IMSE-CNM-CSIC) and University de Seville
Spain

1. Introduction

Resonant tunnelling diodes (RTDs) are very fast non linear circuit elements which have been integrated with transistors to create novel quantum devices and circuits. They are today considered the most mature type of quantum-effect devices, already operating at room temperature, and being promising candidates for future nanoscale integration. The incorporation of tunnel diodes into transistor technologies has demonstrated improved circuit performance: higher circuit speed, reduced component count, and/or lowered power consumption (Mazumder et al., 1998), (Broekaert et al., 1998; Sano et al., 2001; Kawano et al., 2003). Thus, RTD based circuits has been receiving a great amount of interest in the last years. Most of the reported working circuits have been fabricated in III/V materials while Si-based tunnelling diodes compatible to standard CMOS fabs are currently an area of active research (Sudirgo et al., 2004). In fact, it has been claimed that augmenting CMOS with RTDs could be the way to extend the lifetime of CMOS and fully exploiting its huge economical investments (Sudirgo et al., 2004). Recent advances in the development of those Si-based RTDs have raised a renewed interest on circuit design using RTDs and transistors.

RTDs exhibit a negative differential resistance (NDR) region in their current-voltage characteristics. Figure 1*a* shows it enhancing key parameters for circuit design: peak current and voltage, I_p and V_p, and valley current and voltage, I_v and V_v. Three regions are defined according to Figure 1*a*: two regions of positive (I and III) and one of negative (II) differential resistance. Circuit applications of RTDs are mainly based on the MOnostable-BIstable Logic Element (MOBILE) (Maezawa & Mizutani, 1993; Akeyoshi et al., 1993; Chen et al., 1995; Pacha et al., 2000; Avedillo et al., 2006a). The basic MOBILE is a rising edge triggered current controlled gate which consists of two RTDs (the load and driver RTDs) connected in series and driven by a switching bias voltage (V_{bias}). When connected in series, RTDs provide multiple-peak structures in their *I-V* characteristics, which make it attractive for multiple-valued logic (MVL) (Waho et al. 1996; Soderstrom, Andersson, 1998; Seabaugh et al. 1992). MVL circuit applications are based on the Monostable-to-Multistable transition Logic (MML) (Waho et al., 1998), an extension of the binary MOBILE. Logic operation is based on the sequential switching (in increasing order of peak current values) of the RTDs connected in series, which is produced when the bias voltage rises to an appropriate value. Logic functionality is achieved by embedding an input stage (compound-semiconductor transistors, HEMT or HBT) which modifies, according the applied input signal, the peak current of some of the RTDs. MML

circuits have been mainly applied in communication system, mainly analog to digital converters and multiple-valued quantizers embedded on sigma-delta modulators, where their high-speed performance is expected to be of the most importance for future commercial applications, (Chibashi et al., 2004; Maezawa, 2005; Eguchi, 2005).

Several works have been dedicated to the performance modelling of MOBILE gates (Quintana & Avedillo, 2005; Aoyama et al., 2002; Matsuzaki et al., 2004; Uemura & Mazumder, 2002; Avedillo et al., 2006b; Quintana et al., 2006) from both a simulation and an analytical point of view. However, an analytical approach allows technology independence and reuse. Some of these papers (Aoyama et al., 2002; Matsuzaki et al., 2004; Uemura & Mazumder, 2002) have studied the maximum operating speed of MOBILE gates. In one of them (Aoyama et al., 2002), it is shown that these gates operate properly in a certain frequency range; that is, they exhibit both a minimum operating frequency and a maximum one. The frequency range depends on the gate fan-out. From the design point of view it should be desirable gates without the minimum limit (correct operation from DC up to a maximum frequency). Through extensive simulations, a relationship between RTD areas and transistor size that must be satisfied for a given MOBILE gate, in a specific technology, to operate properly at very low frequencies can be derived. Analytical design constraints for a DC correct operation have been studied in (Quintana & Avedillo, 2005) which allow designing MOBILE gates without minimum operating frequency problems. In addition, one of the most attractive features of MOBILE-based circuits, as it is their self-latching operation (which allows pipelining at the gate level, and thus very high through-output, without any area overhead associated to the addition of the latches), has been shown to be not inherent to the practical circuit topologies employed to implement MOBILE circuits (Avedillo et al., 2006b), and some analytical design guidelines relating circuit parameters have been developed to avoid the problem (Quintana et al., 2006).

Successful operation of MML circuits has been demonstrated in (Waho et al., 1998), where design of MML ternary inverters and literal gates is addressed. However, an analysis of the self-latching capabilities of the MML structure depending on technological parameters has not

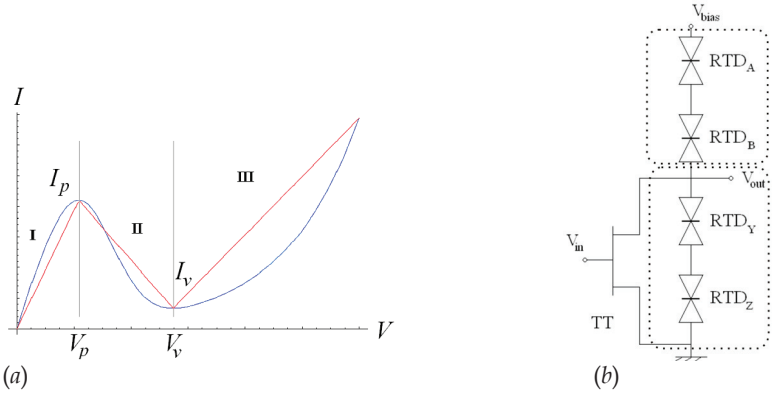


Fig. 1. (*a*) I-V characteristic for a LOCOM RTD (blue) and its linear approach (red) and (*b*) typical structure of a MML ternary inverter.

been yet performed. This analysis is crucial to implement quantizers in which the output level must be held even if the input changes (Gan & Su, 1997). To analytically study this problem we have selected a ternary valued inverter and resorted to piecewise linear descriptions for the RTD driving point characteristics, which allow us to obtain relations between RTD and transistor parameters that ensure a correct behaviour of the structure. The chapter is organized as follows: Section II deals with the operation principle of RTD-based ternary inverters. The importance of a correct sizing of the parameters of the structure is analyzed in Section III. The relationships between parameters to ensure a proper behaviour are derived in Section IV, and simulations results using the LOCOM (Prost et al., 2000) technology are shown in Section V. Finally, some conclusions are given in Section VI.

2. Operation Principle

Figure 1b depicts the typical structure of a ternary inverter based on MML composed of four series connected RTDs, one transistor, and driven by a switching bias voltage, V_{bias}). When V_{bias}) rises, RTDs are sequentially switched in increasing order of their peak current values. The high value for the bias voltage is selected in order to have two RTDs switched when it is applied so that three different output voltages can be observed: a low value when the two upper RTDs are switched; a high value when the two RTDs in the bottom are switched and, finally, a medium output value when one of the RTDs in the load and another in the driver are switched. The HFET provides the logic functionality as its input modulates the drain to source current of the transistor, and consequently the effective peak current of some of the RTDs, modifying their switching sequence. In our analysis, we have considered three specific, feasible voltage values of V_{in}, the high (V_{in}^H), medium (V_{in}^M) and low (V_{in}^L) voltages, associated with the logical "2", "1" and "0", respectively.

A first approach to the operation of the structure by analyzing the switching sequence of the RTDs is not suitable to derive relationships among technological and circuit parameters for correct operation, as it is our target, since it is difficult to add the effect of the transistor. This problem can be efficiently solved by considering that two NDR devices can be identified: the load and the driver NDRs (NDR_L and NDR_D, respectively). The first one, the load, is composed of two RTDs (RTD_A and RTD_B in Figure 1b); the second one, the driver NDR, is composed of two RTDs (RTD_Y and RTD_Z) and the transistor TT. The MOBILE operation principle can be applied to this structure consisting of two NDR devices. The order in which transitions are carried out depends on the relationship between the peak currents of each NDR device (Waho et al., 1998).

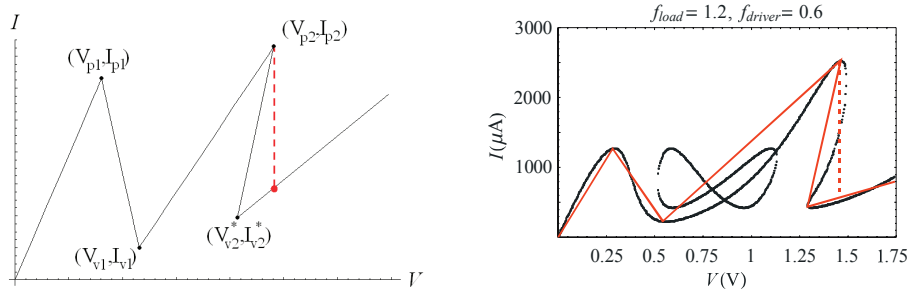
The joint I-V driving point characteristic for two non-linear series-connected RTDs is very complex and some kind of simplification is needed. To perform an analytical study, the I-V characteristic of an RTD has been approximated by a piecewise linear characteristic. Thus, simple geometrical considerations allow obtaining their joint I-V representation. Figure 1a shows this piecewise linear for the RTD obtained from a nonlinear RTD driving point characteristic. Thus, current through the RTD is given by

$$\begin{cases} f\, m_{\mathrm{I}}\, V & V \le V_p \\ f(m_{\mathrm{II}}(V - V_p) + I_p) & V_p \le V \le V_v \\ f(m_{\mathrm{III}}(V - V_v) + I_v) & V \ge V_v \end{cases} \quad (1)$$

where V is the voltage applied to the RTD and f [1] the area factor. If two generic piecewise linear series-connected RTDs, RTD_1 and RTD_2, with area factors f_1 and f_2 respectively ($f_1 > f_2$) are considered, the new joint I-V characteristic has two peak and two valley voltages (V_{p1}, V_{p2}, V_{v1} and V_{v2}, as shown in Figure 2a) which can be easily calculated depending on the values of f_1, f_2, I_p, I_v, V_p and V_v (Gan & Su., 1997). For example, and concerning the peak currents of the new joint IV characteristic, the first one, I_{p1}, is given by the smallest of the individual peak currents ($f_2 I_p$) and the second one, I_{p2}, by the largest one ($f_1 I_p$). The approach is very good as shown in Figure 2b where the driving point characteristic of two non-linear series-connected LOCOM RTDs (with areas $Area_1$=12μm² and $Area_2$=6μm²) and the corresponding joint I-V characteristic from two piecewise linear RTDs which has been theoretically obtained, are depicted. The correspondence between peaks and valleys of both representations becomes apparent.

It is easy to prove that for the representation in Figure 2a, the second valley voltage is always below the second peak voltage. In order to handle well-defined functions, we have made both voltages to be equal and, consequently, the second valley current has been changed as the red dotted line in Figure 2a indicates. The I-V characteristic so obtained does not include some regions that would appear in the complete representation corresponding to two linear series-connected RTDs, but they do not modify the normal operation of the inverter.

The I-V characteristic of both NDR_D and NDR_L shows two peaks and two valleys. Peak and valley voltages and currents in the NDR_L are directly obtained from RTD_A and RTD_B. For



(*a*) (*b*)

Fig. 2. (*a*) I-V characteristic of two series connected RTDs and (*b*) joint I-V representation of two non-linear (in black) and piecewise linear (in red) series-connected RTDs.

NDR_D, peak and valley voltages are directly obtained from RTD_Y and RTD_Z, but peak and valley currents are modified by the transistor, and calculated by adding to the original values (obtained from RTD_Y and RTD_Z) the corresponding amount due to the transistor current for the applied input and a drain to source voltage equal to the peak and valley voltages of the I-V characteristic obtained from RTD_Y and RTD_Z.

Due to the existence of two peaks in the I-V characteristic of both NDR_D and NDR_L, the operation of the ternary inverter shows two "transitions". The relation between the peak currents of the driver and the load during the evaluation phase, depending on the value of the input voltage, will determine the final value of the output voltage (usually RTD areas fulfill the relation $Area_A > Area_B > Area_Y > Area_Z$ (Waho et al., 1998)). For example, if $V_{in}=V_{in}^L$, the relation between peak currents in the NDRs must be $(I_{p1})_D < (I_{p2})_D < (I_{p1})_L < (I_{p2})_L$, where subindex D or L corresponds, respectively, to the driver or the load NDR. In this case, both transitions are due to that RTDs in NDR_D, RTD_Z and RTD_Y are the first to commute and the output is a 'high' level. When $V_{in}=V_{in}^M$, two possibilities can be found to reach a 'medium' logical level of the output: $(I_{p1})_D < (I_{p1})_L < (I_{p2})_D < (I_{p2})_L$ (RTD_Z commutes before RTD_B), or $(I_{p1})_L < (I_{p1})_D < (I_{p2})_L < (I_{p2})_D$ (RTD_B commutes firstly). Finally, for $V_{in}=V_{in}^M$, the relation between peak currents in the NDRs must be $(I_{p1})_L < (I_{p2})_L < (I_{p1})_D < (I_{p2})_D$, i.e., the first and the second transition are due to the RTDs in NDR_L (RTD_B and RTD_A).

3. The Multistability Disappearance Problem

In order to show concrete examples we have used a piecewise linear model for the RTDs which has been derived from the nonlinear I-V characteristic of RTDs fabricated in the LOCOM (Prost et al., 2000) fabrication process. Figures 3a and 3b show the output response to input voltage changes from two inverters with different sized transistor. Both of them evaluate correctly but, they behave in a different way when the input changes for $V_{bias}=V_{bias}^H$. In the first case (Figure 3a), a variation of the input has no effect on the output, thus, the MML inverter has a correct behaviour. However, for the second inverter (Figure 3b), with a different sized transistor, a change of the input node voltage from V_{in}^M to V_{in}^H forces the output to change from a medium to a low level (multistability disappearance).

A good criterion for predicting this problem can be derived through an analysis of the DC operation of the ternary inverter (Núñez et al., 2006). By applying the Kirchoff Laws to the circuit in Figure 1b, we obtain

$$I_{RTD,L}^{series}[V_{bias} - V_{out}] = I_{RTD,D}^{series}[V_{out}] + I_{DS}[V_{in}, V_{out}] \tag{2}$$

where $I_{RTD}^{series}[v]$ and $I_{DS}[V_{GS},V_{DS}]$ represent the mathematical description of the driving point characteristic of two series connected piecewise linear RTDs ($I_{RTD,D}^{series}[v]$ for the driver and $I_{RTD,L}^{series}[v]$ for the load) and the transistor, respectively. The set of solutions plotted in the V_{in}-V_{out} plane, gives information about what happens when the input voltage changes its value

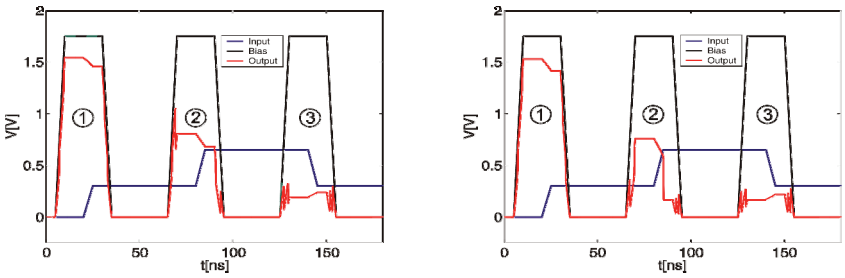


Fig. 3. Input (blue), bias (black) and output (red) voltages for an MML ternary inverter with f_Z=0.6, f_Y=0.7, f_B=1.1, f_A=1.2 and (*a*) W=6μm, L=1μm (does not exhibit multistability problems) and (*b*) W=10, L=1μm (multistability disappearance when the input node changes from $V_{in}{}^M$ to $V_{in}{}^H$).

for a fixed bias voltage. The problem in Figure 3 comes from the disappearance of one (or more) of the stable states in the DC solution representation for V_{bias}=$V_{bias}{}^H$, as it will be explained in the following.

Let us consider the disappearance of the highest output level when the input voltage increases its value to V_{in}=$V_{in}{}^H$ (transition marked as "2" in Figure 3b). This malfunction appears when NDR_L is biased about its first peak voltage (RTD_B is in its peak and commutes, forcing a change in the output node), thus, V_{out} is approximately $V_{bias}{}^H$-$(V_{p1})_L$. Figure 4a depicts the load curve for a properly-sized MML inverter, in which the five solutions found have been marked in red and represented in the V_{in}-V_{out} plane of Figure 4b. Let us consider the case in which the input is fixed at a low value. A malfunction is found around V_{out}≈$(V_{p1})_D$ when V_{in}=$V_{in}{}^L$, which happens when RTD_Z reaches its peak voltage, increasing the value of the output node. In order to find solutions to Eq. (2) around V_{out} ≈ $(V_{p1})_D$ it is mandatory that the first peak current of NDR_D is above the current of NDR_L when it is biased with a voltage equal to $V_{bias}{}^H$-$(V_{p1})_D$ (see Figure 4c and 4d). Finally, if V_{in}=$V_{in}{}^M$, two multistability problems must be analyzed. The first problem, occurs around the maximum value of the output, that is, $V_{bias}{}^H$-$(V_{p1})_L$. To avoid this malfunction, the first peak current of NDR_L has to be above the current of NDR_D when V_{out}= $V_{bias}{}^H$-$(V_{p1})_L$. A similar reasoning is carried out to deal with the malfunction that appears when it is biased with $V_{bias}{}^H$-$(V_{p1})_D$ in which the current through NDR_L has to be under the first peak current of NDR_D.

Let us consider the previously described structures of the MML inverter in figures 3a and 3b. Figure 5a depicts the V_{in}-V_{out} plot corresponding to the one which has the correct behaviour (Figure 3a), where the red dotted line marked with "1" represents the output evolution as V_{in} changes from V_{in}=$V_{in}{}^L$ to V_{in}=$V_{in}{}^M$). The vertical line corresponds to the input medium voltage level and the green point in Figure 5a is the final output value for this transition. The second red dotted line (marked with "2") depicts the output evolution when the input varies from the medium (output marked as a blue point) to the highest level (output as a red point). On the other hand, Figure 5b depicts the V_{in}-V_{out} plot corresponding

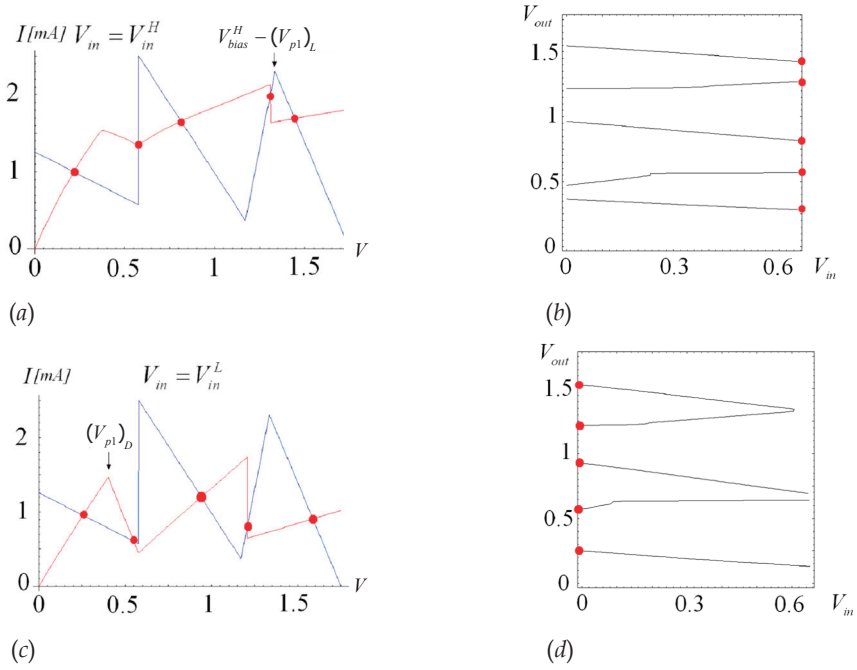


Fig 4. Load curves and V_{in}-V_{out} plots and solutions to Eq. (2) marked in red, for (*a*), (*b*) $V_{in} = V_{in}^H$; and (*c*), (*d*) $V_{in} = V_{in}^L$.

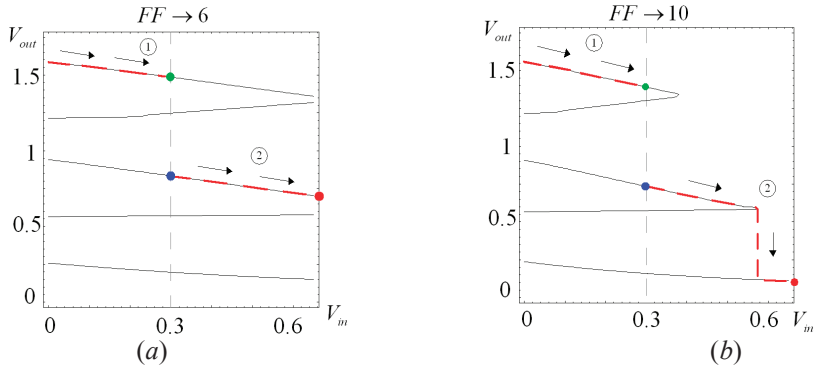


Fig 5. V_{in}-V_{out} plots for the structures of (*a*) Fig. 3*a* and (*b*) Fig. 3*b*, where the red dashed line depicts the evolution of the output as indicated in these figures.

red dotted line marked with "2" (input increasing from V_{in}=V_{in}^{M} to V_{in}=V_{in}^{H}), we can check that when the input voltage is close to 0.6V, the output falls down to the lowest level. The output does not maintain its value and the multistability property is not verified.

4. Critical Dependencies

To guarantee a correct DC operation, we need consider, first, the relations between the area factors of the RTDs and the form factor ($FF = W/L$) of the transistor for a correct evaluation of the structure for fixed input voltages. Case of the transconductance is not proportional to $1/L$ (as in the HFET), this form factor corresponds to W for a fixed L. If V_{in}=V_{in}^{L}, the first transition is due to RTD_Z, thus we must ensure that the first peak current of the driver is less than the first peak current of the load. The critical situation appears when the first peak current of NDR_D is equal to the corresponding one of NDR_L. Thus, the first relation between parameters comes from:

$$f_Z I_p + FF \cdot I_{DS}[V_{in}^{L}, (V_{p_1})_D] < f_B I_p \tag{3}$$

The second RTD to commute is RTD_Y, so that second peak current of the driver must be less than the first peak current of the load (which is originated by the smallest RTD of the load, that is, RTD_B). The limit case is obtained when the first peak current of NDR_L coincides with the second peak current of NDR_D. According to this, a new expression for a maximum value of FF is derived,

$$f_Y I_p + FF \cdot I_{DS}[V_{in}^{L}, (V_{p_2})_D] < f_B I_p \tag{4}$$

Now, if V_{in}=V_{in}^{M}, the sequence of commutation of RTD_Z and RTD_B determines the pair of inequalities to be considered.. Only two of these four inequalities need to be used because they are the most restrictive constraints,

$$f_Y I_p + FF \cdot I_{DS}[V_{in}^{M}, (V_{p_2})_D] > f_B I_p \tag{5}$$

$$f_Z I_p + FF \cdot I_{DS}[V_{in}^{M}, (V_{p_1})_D] < f_A I_p \tag{6}$$

Finally, for the highest input voltage, the RTD switching sequence is RTD_B and then RTD_A, and the following conditions can be derived,

$$f_Z I_p + FF \cdot I_{DS}[V_{in}^{H}, (V_{p_1})_D] > f_B I_p \tag{7}$$

$$f_Z I_p + FF \cdot I_{DS}[V_{in}^H, (V_{p_1})_D] > f_A I_p \tag{8}$$

Four restrictions to the feasible set of values of the circuit parameters can be derived through the analysis performed in Section III. Figure 6a depicts a V_{in}-V_{out} plot for V_{in} increasing from $V_{in}{}^L$ to V_{in}=$V_{in}{}^H$, where the intersections with the right ordinate axis correspond to feasible solutions to Eq. (2) for the high value of the input. Upper point (marked in red) corresponds to a double solution to Eq. (2) and indicates the critical situation for which higher values of *FF* would entail an incorrect behaviour. Thus, an expression concerning to a maximum *FF* is obtained,

$$I_{RTD,D}^{series}[V_{bias}^H - (V_{p_1})_L] + FF \cdot I_{DS}[V_{in}^H, V_{bias}^H - (V_{p_1})_L] < f_B I_p \tag{9}$$

When V_{in} decreases from $V_{in}{}^H$ to $V_{in}{}^L$, the decision is taken around $V_{out} \approx (V_{p1})_D$ (the red point in Figure 6b) and must guarantee that the peak current of the driver is under the current of the load. Thus,

$$f_Z I_p + FF \cdot I_{DS}[V_{in}^L, (V_{p_1})_D] > I_{RTD,L}^{series}[V_{bias}^H - (V_{p_1})_D] \tag{10}$$

For the medium input voltage, V_{in}=$V_{in}{}^M$, two conditions can be derived. A maximum value of FF is obtained by considering that the first peak current of the load must be above the current through the driver (the critical situation is depicted by means of the red point Figure 6c),

$$I_{RTD,D}^{series}[V_{bias}^H - (V_{p_1})_L] + FF \cdot I_{DS}[V_{in}^M, V_{bias}^H - (V_{p_1})_L] < f_B I_p \tag{11}$$

Finally, the last relationship between parameters is derived by considering that the first peak current of NDR_D is equal to the current through NDR_L (as shown in the red point in Figure 6d).

$$f_Z I_p + FF \cdot I_{DS}[V_{in}^M, (V_{p_1})_D] > I_{RTD,L}^{series}[V_{bias}^H - (V_{p_1})_D] \tag{12}$$

5. Simulation Results

In order to check our approach we have performed a comparison between our piecewise linear theoretical approach and HSPICE simulations using the nonlinear RTD model from LOCOM. Eq. 3 to Eq.12 have been employed to analyze how the DC operation of a ternary inverter is modified when some key parameters are changed. HFET and RTDs from the LOCOM technology have been used. For this technology, the RTD has a peak voltage, V_p, of 0.21V, the peak current density is 21 KA/cm² (giving an I_p of 2.1m*A* for an RTD of area factor of 1), and the peak current ratio is about 6.25 at room temperature. The transistor is a depletion HFET with threshold voltage -0.2V. Bias and input levels are $V_{bias}{}^H$=0.175*V*,

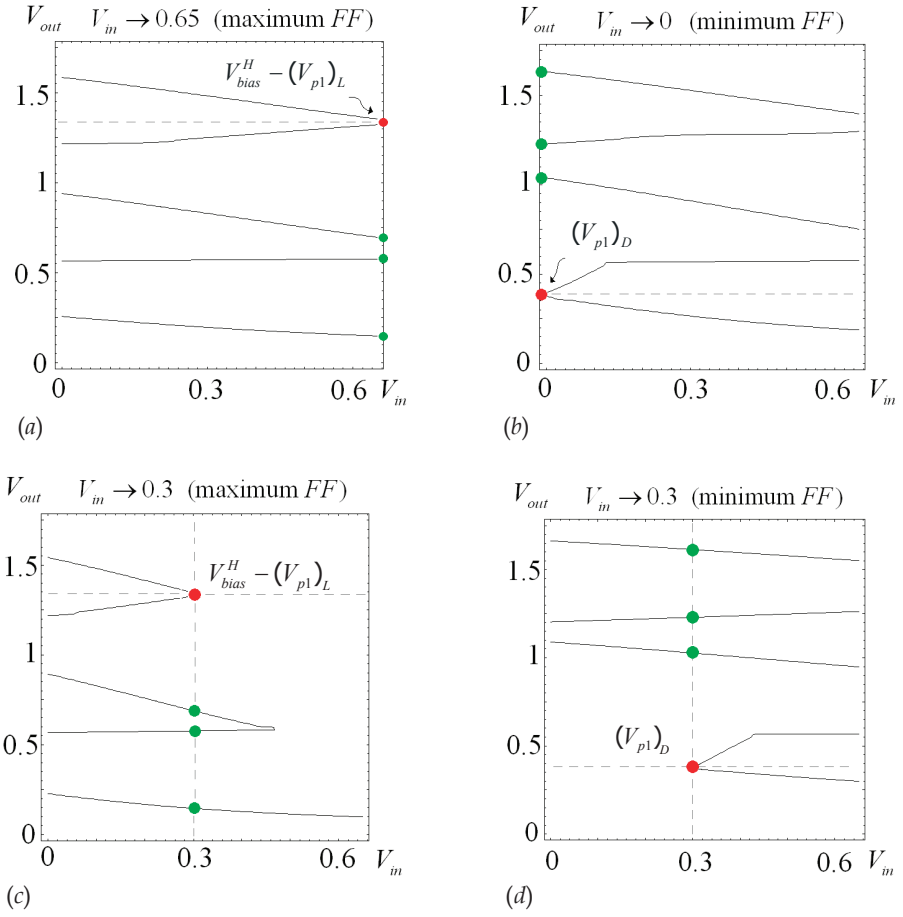


Fig. 6. V_{in}-V_{out} plots pointing out the critical (in red) solutions to Eq. (2) when (*a*) $V_{in} = V_{in}^{H}$, (*b*) $V_{in} = V_{in}^{L}$ and (*c*), (*d*) $V_{in} = V_{in}^{M}$.

$V_{bias}^{L}=0V$, $V_{bias}^{M}=0.3V$, and $V_{bias}^{H}=0.65V$. HSPICE simulations have been performed by using the nonlinear I-V characteristic of individual RTDs.

To reduce the set of parameters which can be modified, we have defined Δ_1 as the difference between the area factors of the RTDs from the same NDR and Δ_2 as the difference between the smallest are factor of the load and the highest of the driver, that is, $\Delta_1=f_Y-f_Z=f_A-f_B$ and $\Delta_2=f_Y-f_B$. Figure 7a shows the set of constraints from which the theoretically correct operation region (the one that guarantees a correct evaluation and multistability preserving operation) can be derived for an inverter with f_Z=0.6 and Δ_1=0.08. Feasible pairs (Δ_2,*FF*) have been calculated by varying Δ_2, as the shady region indicates. Multistability constraint which

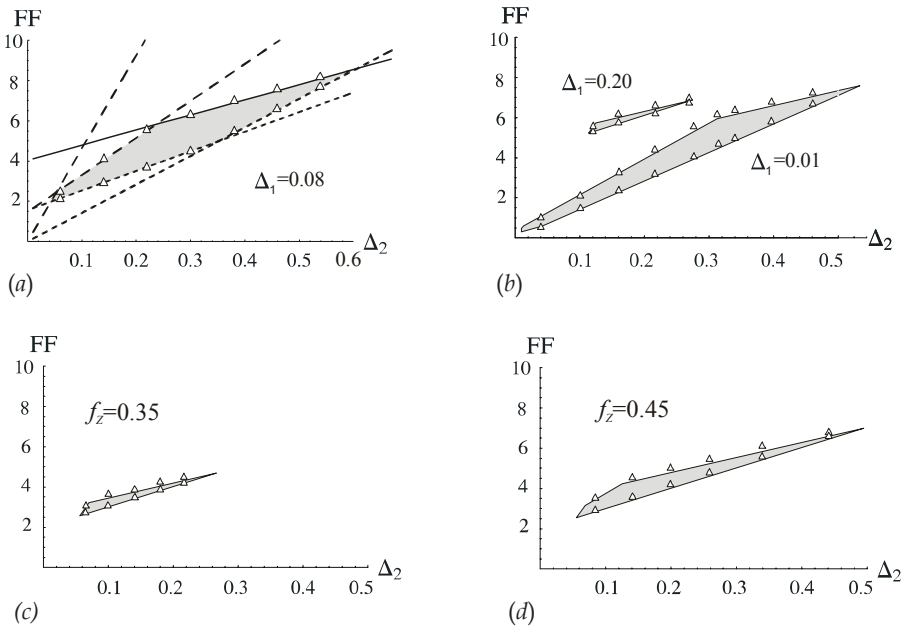


Fig. 7. (*a*) Set of constraints from which the correct operation region is derived. Grey area depicts feasible values of *FF vs.* $\Delta_2 = f_B - f_Y$ for $\Delta_1 = f_A - f_B = f_Y - f_Z = 0.08$ and $f_Z = 0.6$. Triangles delimit the region given by HSPICE simulations with nonlinear RTD models. Feasible (Δ_2,*FF*) pairs for (*b*) $\Delta_1 = 0.01$ and $\Delta_1 = 0.20$, $f_Z = 0.6$. (*c*) $\Delta_1 = 0.1$, $f_Z = 0.35$. (*d*) $\Delta_1 = 0.1$, $f_Z = 0.45$.

provides a maximum FF has been represented by a solid line and evaluation constraints by different broken lines (providing maximum and minimum FF values). From the figure it is apparent the shrinkage of the feasible operation region due to the inclusion of multistability constraints. Figure 7b shows only the feasible sets (the feasible region) for two values of the difference between f_y and f_z (Δ_1), 0.01 and 0.20, respectively. Triangles are operation limit conditions obtained with HSPICE and the nonlinear RTD models from LOCOM. It can be observed the progressive shrinking of the feasible region and the agreement between theoretical and simulation results. Practically, for $\Delta_1 > 0.2$ there are no inverters with a proper behaviour for this value of f_Z. In Figures 7c and 7d the effect of a modification of f_Z is analyzed. A very close correspondence between both theoretical and simulation results can be also observed.

Figure 8a depicts feasible areas of correct operation when $\Delta_1 = \Delta_2 = \Delta$ for different values of f_Z, where the feasible region is widen when f_Z increases its value. Finally Figures 8b and 8c depict pairs (Δ_{1D},*FF*) and (Δ_{1L},*FF*), that allow a correct DC operation, where $\Delta_{1D} = f_Y - f_Z$ and $\Delta_{1L} = f_A - f_B$, (for $\Delta_2 = 0.1$). Small values of Δ_{1L} give narrow regions of correct behaviour, whereas an increase of Δ_{1D} shifts up the feasible area of correct DC operation.

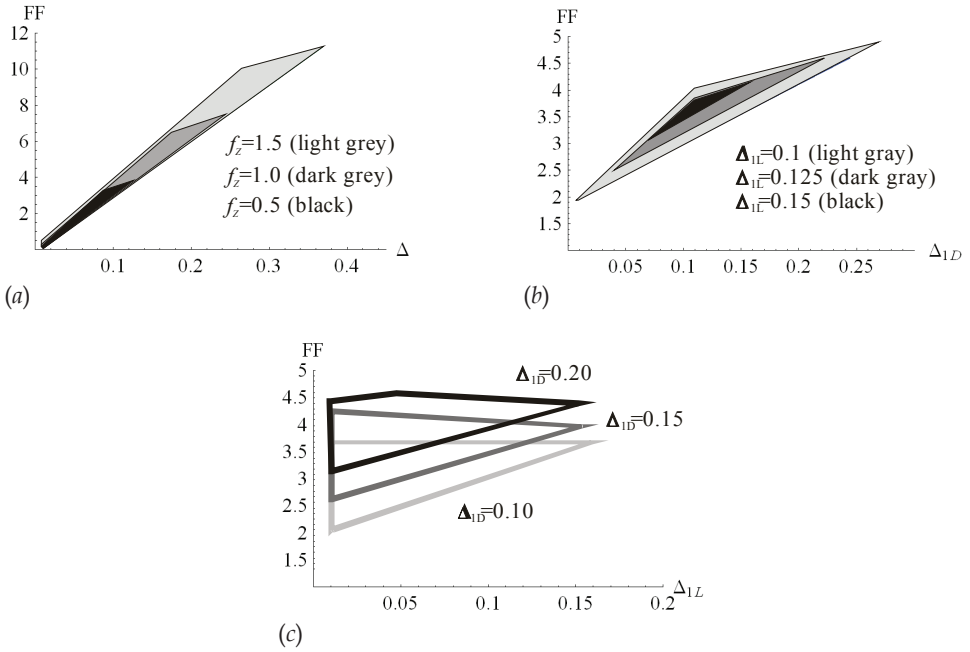


Fig. 8. (*a*) Region of correct DC behaviour *FF* vs $\Delta = \Delta_1 = \Delta_2$ for f_Z={0.5, 1.0, 1.5}. (*b*) *FF vs.* $\Delta_{1D} = f_Y - f_Z$ for $\Delta_{1L} = f_A - f_B$ ={0.1,0.125,0.15} and (*c*) *FF vs.* Δ_{1L} for Δ_{1D}={0.1,0.15,0.20}, with Δ_2=0.1.

The effect of the variation of the peak current density, j_p, is analyzed in Figure 9a by depicting feasible pairs (Δ_2,*FF*) for j_p=18, 42 and 66 KA/cm², and Δ_1=0.01. By increasing j_p, the feasible region is widened and shifts counterclockwise. Finally, Figures 9b and 9c depict the correct operation region for two different values of the peak voltage, V_p=0.24 and V_p=0.15, respectively, and keeping constant the values of the other parameters. A reduction of the final set of feasible pairs (Δ_2,*FF*) is observed while decreasing V_p. HSPICE simulation results have been also included.

6. Conclusions

One additional problem to size the MML inverter used in quantizers has been pointed out. It is the multistability fault, which can prevent the correct operation of circuits, such as quantizers, in which this property is essential. To analytically study the problem, a piecewise linear description for the RTD driving point characteristics has been used. A procedure to calculate the relationships between circuit parameters in order to obtain correct

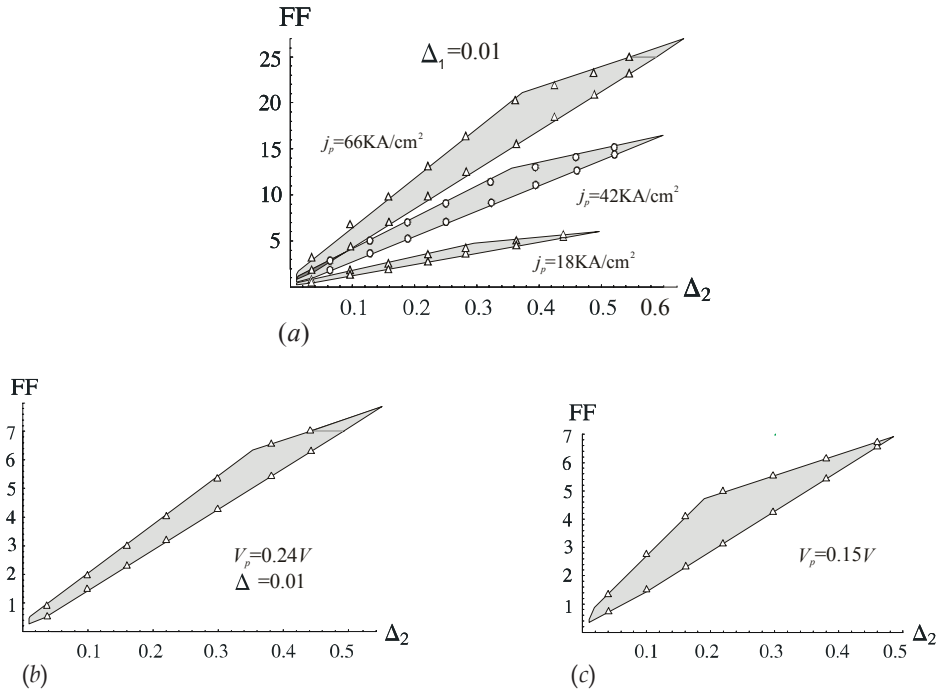


Fig. 9. Set of feasible of values of *FF vs.* $\Delta_2 = f_B - f_Y$ for $\Delta_1 = f_A - f_B = f_Y - f_Z = 0.01$. (*a*) Effect of the variation of j_p. Correct DC operation region for j_p={18, 42, 66}KA/cm². Feasible pairs (Δ_2,*FF*) for different values of the peak voltage, (*b*) V_p=0.24*V* and (*c*) V_p=0.15*V*.

DC operating regions has been derived. HSPICE simulations using a nonlinear RTD model agree very well with our piecewise approach.

7. Acknowledgment

This work has been funded by the Spanish Government under project NDR, TEC2007-67245/MIC, and the Junta de Andalucía through the Proyecto de Excelencia TIC-2961.

8. References

Akeyoshi, T.; Maezawa, K. & Mizutani, T.. (1993). Weighted Sum Threshold Logic Operation of MOBILE's (Monostable-bistable Transition Logic Element) using Resonant-tunnelling Transistors. *IEEE Electron Device Lett.*, Vol. ED-14, pp. 475-477.

Aoyama, et al. (2002). Effects of the HEMT Parameters on the Operation Frequency of Resonant Tunneling Logic Gate MOBILE. *Electronics and Communications in Japan*, Part 2, Vol. 85.

Avedillo, M.J.; Quintana, J.M. & Pettenghi, H. (2006a). Increased Logic Functionality of Clocked Series-Connected RTDs. *IEEE Trans. on Nanotechnology*, Vol. 5, No. 5, pp. 606 - 611.

Avedillo, M.J.; Quintana, J.M. & Pettenghi, H. (2006b). Self-latching Operation of MOBILE Circuits using Series-Connection of RTDs and Transitors. *IEEE Trans. on CAS*, Vol. 53, No. 5.

Broekaert, T., et al. (1998). A Monolithic 4-bit 2-gsps Resonant Tunneling Analog-to-Digital Converter. *IEEE J. of Solid-State Circuits*, Vol. 33, pp. 1342-1349.

Chen, K.J.; Akeyoshi, T. & Maezawa, K. (1995). Monolithic Integration of Resonant Tunnelling Diodes and FETs for Monostable–bistable Transition Logic Elements (MOBILEs). *IEEE Electronic Device Letters*, Vol. 16, pp. 70-73.

Chibashi, M.; Eguchi, K. & Waho, T. (2004). A novel delta-sigma modulator using resonant tunnelling quantizers. *Proc. IEEE Int. Symposium on Circuits and Systems (ISCAS'04)*, Vol. 1, No. 1, pp 533-536.

Eguchi, K.; Chibashi, M. & Waho, T. (2005). A Design of 10-GHz Delta-Sigma Modulator using a 4-level Differential Resonant-tunneling Quantizer. *IEEE Proc. on Multiple-Value Logic (ISMVL'05)*, pp 43-47.

Gan, K.-J. & Su Y.-K (1997). Modeling Current-Voltage and Hysteretic Current-Voltage Characteristics with Two Resonant Tunneling Diodes Connected in Series. *Solid State Electronics*, Vol. 41, No. 12, pp. 1917-1922.

Kawano, Y., et al. (2003). 88GHz Dynamic 2:1 Frequency Divider using Resonant Tunnelling Chaos Circuit. *IEEE Electronics Letters*, Vol. 39, no. 21, pp. 1546-1548.

Maezawa K. & Mizutani, T. (1993). A New Resonant Tunneling Logic Gate Employing Monostable-Bistable Transition. *Jpn. J. Appl Phys.*, vol. 32, no. 1A-B, pp. L42-L44.

Maezawa, K. (2005). Resonant Tunneling Diodes and Their Application to High-Speed Circuits. *IEEE Compound Semiconductor Integrated Circuit Symposium (CSIC'05)*, pp. 97-100.

Matsuzaki, H.; Fukuyama, H. & Enoki, T. (2004). Analysis of Transient Response and Operating Speed of MOBILE. *IEEE Trans. on Electron Devices*, Vol. 51, pp. 616-622.

Mazumder, P., et al. (1998). Digital Circuit Applications of Resonant Tunneling Devices", *Proc. of the IEEE*, Vol. 86, no. 4, pp. 664-686.

Núñez, J.; Quintana, J.M. & Avedillo, M.J. (2006). Holding Disappearance in RTD-based Quantizers. *European Nano Systems 2006*, pp. 120-124.

Pacha, C. et al (2000). Threshold Logic Circuit Design of Parallel Adders Using Resonant Tunneling Devices. *IEEE Trans. on Very Large Scale Integration Systems*, Vol. 8, no. 5, pp. 558-572.

Prost, W., et al. (2000). EU IST Report LOCOM no. 28 844 Dec. 2000.

Quintana, J.M. & Avedillo, M.J. (2005). Transistor Critical Sizing in a MOBILE Follower. *Electronics Letters*, Vol. 41, No. 10, pp. 583-584.

Quintana, J.M.; Avedillo, M.J. & Pettenghi, H. (2006). Self-latching operation limits for MOBILE circuits. *Proc. IEEE Int. Symposium on Circuits and Systems (ISCAS'06)*, pp 4579-4582.

Sano, K., et al. (2001). An 80-Gb/s Optolectronic Delayed Flip-Flop IC using Resonant Tunneling Diodes and Uni-Traveling- Carrier Photodiode. *IEEE J. of Solid State Circuits*, Vol. 36, no. 2, pp. 281-289.

Seabaugh, A. C.; Kao, Y.-C. & Yuan H.-T. (1992). Nine-State Resonant Tunnelling Diode Memory. *IEEE Electron Devices Letters*, Vol. 13, pp. 479-481.

Soderstrom, J. & Andersson, T.G. (1998). A multiple-state memory cell based on the resonant tunneling diode. *IEEE Electron Device Letters*, Vol. 9, Issue 5, pp. 200-202.

Sudirgo S., et al. (2004). Monolithically Integrated Si/SiGe Resonant Interband Tunnel Diode/CMOS Demonstrating Low Voltage MOBILE Operation. *J. of Solid-State Electronics*, Vol. 48, pp. 1907-1910.

Uemura, T. & Mazumder, P. (2002). Rise Time Analysis of MOBILE Circuit. *Proc. IEEE Int. Symposium on Circuits and Systems (ISCAS'02)*, Vol. 5, pp. 26-29.

Waho, T.; Chen, K.J. & Yamamoto, M. (1996). A novel multiple-valued logic gate using resonant tunneling devices. *IEEE Electron Device Letters*, Vol. 17, no. 5, pp. 223–225.

Waho, T.; Chen, K.J. & Yamamoto, M. (1998): "Resonant-Tunneling Diode and HEMT Logic Circuits with Multiple Thresholds and Multilevel Output. *IEEE J. of Solid-state Circuits*, Vol. 33, No. 2, pp. 268-274.

6

Development of Superlattice Infrared Photodetectors

Shih-Hung Lin, Ying-Hsiang Wang, Che-Wei Chang,
Jen-Hsiang Lu, Chun Chi Chen and Chieh-Hsiung Kuan,
Department of Electrical Engineering, National Taiwan University
Taipei, 10617 Taiwan, Republic of China

1. Overview and Perspective

Infrared (IR) detectors play a critical role in both military and civilian applications and have been widely researched in recent decade. Because the atmospheric transparent windows for the IR radiation exist within the spectral ranges of 3−5 and 8−12 μm, and the spectrum of the black-body radiation at room temperature has a peak at 10 μm, the detectors with the 8-12 μm detection spectra are helpful to identify the heat radiation from a target at room temperature. Such detectors are the research focus in this chapter.

The employment of intersubband transitions for the infrared radiation detection has drawn much attention. The transition is completed by the electrons which absorb photons with the appropriate energy equal to the suband energy difference to transit from the low subband to the high one. The intersubband photodetectors can be made from semiconductor heterostructures of multiple quantum wells (West & Eglash, 1985) (Harwit & Harris Jr., 1987) (Levine et al., 1987) or superlattices as shown in Fig. 1. Infrared detection will be done by the intersubband transitions between two quantum states in the multiple quantum wells or two minibands in the superlattices. The wells are sandwiched by thick barriers in the multiple quantum well structure. Therefore electron wavefunctions in the wells would not interact with each others and discrete quantum states are formed. Contrarily, the adjacent wells in the superlattice are separated by thin barriers. Minibands are formed in the superlattice region by the coupling of electron wavefunctions.

As shown in Fig. 1, in comparison with the quantum well infrared photodetectors (QWIPs), superlattice infrared photodetectors (SLIPs) have three different characteristics. The first one is the low operational bias. The electrons in the miniband of the superlattice (SL) are conductive while those in the quantum states of the multiple quantum wells (MQWs) are confined. The SL hence becomes a low resistance structure and thus no externally applied bias drops on the SL under low bias range. Therefore, the current blocking layer is needed to decrease the dark current in SLIPs and can determine the operational bias range.

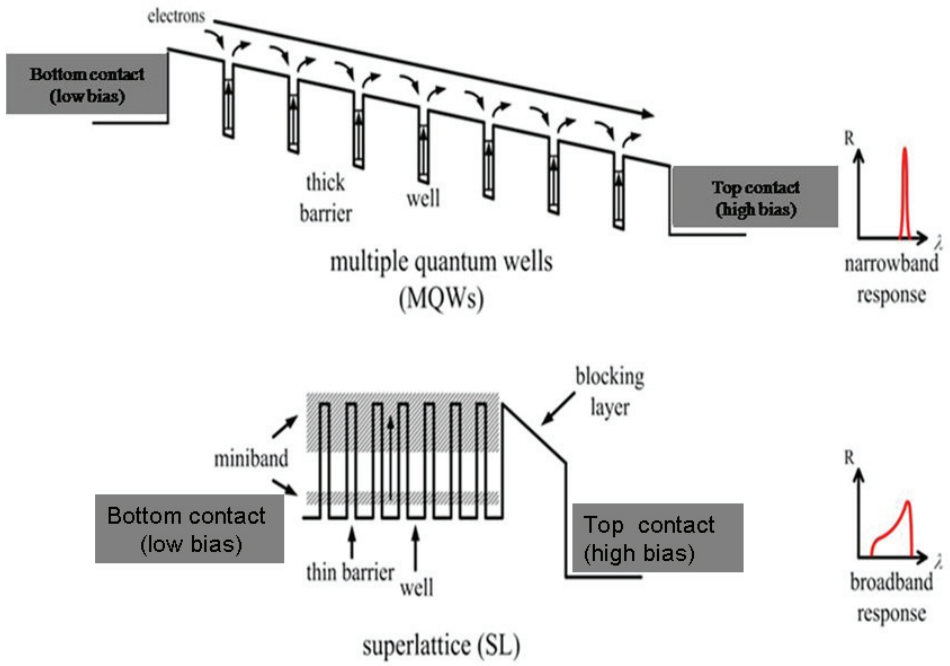


Fig. 1. The schematic band diagrams of multiple

We can design various structures of the blocking layers in order to reach the lower bias and less power consumption for operation. The second characteristic is broadband response of the SL. The transition type of SL is for a miniband to another miniband, so the absorption wavelength range is broader in comparison with QWIPs. Hence, the broadband detection can be done easily by using SL.

The third characteristic is voltage-tunable. We can design a proper current blocking layer to tune the absorption wavelength range of SLIPs with externally applied bias. Consequently, the functions of the blocking layer in SLIPs are not only to reduce the dark current but also to act as a bias-tuned energy filter. The above SLIP performances have been proved by our work (Chen et al., 2002) (Chen et al., 2003) (Lu et al., 2003). In particular, the voltage tunability will be explained in details in Sec 2.

In the detection of the thermal image, a focal plane array which is composed of the elements containing a detector and associated readout integrated circuit (ROIC). In order to avoid saturation of capacitance in ROIC, the low total current of the detector is necessary. In the MQW case, the photoexited electrons need a high bias to transport through whole structure and contribute to the photocurrent, but the dark current increases dramatically under high bias. However, no externally applied bias drops on the SL under low bias range. The SL structure does not need a high bias to transport electrons. Lower biases and less power consumption can be achieved if the SLIP is chosen and the saturation of capacitance in ROIC can also be avoided.

Although detector performance is better when the temperature is lowered (Gumbs et al., 1994) (Jiang et al., 1999) (Majumdar et al., 2003), the cost for the liquid nitrogen cooling is

much cheaper than that for the liquid helium one. Therefore the detector has to be operated at the temperature higher than 77K. To achieve this goal for the SLIP, different blocking layers are needed to improve the operation temperature.

No. of detectors	structure descriptions	note
Detector 1	double SLs separating by a barrier	
Detector 2	one SLs sandwiched by double barriers(one 500nm thick and one 50nm thin barriers)	etched to emitter
Detector 3	one SLs sandwiched by double barriers (one 500nm thick and one 50nm thin barriers)	etched to SLs
Detector 4	one SLs with 50nm barrier	
Detector 5	one SLs and graded barrier with MQWs inserting	
Detector 6	one SLs with 300nm barrier	

Table 1.1. The descriptions of all detectors in this chapter

The principal goals of this chapter are the improvement of the responsivity and decrement of noise under low bias in order to increase the final detectivity and operation temperature. The demonstration of the flexibility with SL structure is also included in our principal goal. In the following sections, various SL structures and blocking layers are composed to become a detector with a specific principal goal. At first, a graded blocking layer is sandwiched between two different SLs to reach the multi-color detection which demonstrate the flexibility of the SL structure. Then double barriers with SL inserting are utilized to improve the responsivity. Next is the MQWs are integrated with the SLIP to act as a noise filter. Finally, the conclusion and future work will be discussed. Table 1.1 lists all the detectors shown in this chapter.

2 Multi-color Superlattices Infrared Photodetector

2.1 Introduction

In various applications such as aerospace observation, target discrimination and temperature sensing, an infrared photodetector for multi-color detection is essential.Because of the maturity and flexibility of band engineering, many works of intersubband multi-color infrared photodetectors have been devoted. The structures with quantum wells (Gravé et al., 1992) (Köck et al., 1992) (Martinet et al, 1992) (Tsai et al., 1993) (Liu et al., 1993) or superlattices (Hsu et al., 2000) (Chen et al., 2002) were reported to realize multi-color photodetectors.

In this section, we adopt two distinct superlattices separated by a blocking layer to achieve multi-color detection. The spectral responsivity of our photodetector is switchable between two wavelength regimes (6~8.5 μm and 7.5~12 μm) by the bias polarity, and is also voltage-tunable in each wavelength regimes.

2.2 Sample Structure

Fig. 2.1 shows the band diagram of the double superlattice infrared photodetector and the corresponding energy levels. We denote this detector as Detector 1. The barrier height of the blocking layer is higher than the energy level of the bottom edge in the second mini-band of the superlattice. It is designed for the operation of the voltage tunability. In particular, the transitions of the electrons between the two miniband of the superlattice can be separately into the long-wavelength and the short-wavelength absorptions which correspond to the respective transitions from the top state of the first miniband into the bottom state of the second miniband and from the bottom state of the first miniband into the top state of the second miniband.

The stucture contains sequentially a 400 nm top contact layer, a 14-period top superlattice, a blocking barrier, 14-period bottom superlattice, and a 500 nm bottom contact layer. Top superlattice is composed of 6 nm $Al_{0.31}Ga_{0.69}As$ barrier and 4.5 nm GaAs well, and bottom one is composed 4 nm $Al_{0.31}Ga_{0.69}As$ barrier and 6 nm GaAs well. The blocking barrier consists of a 60 nm $Al_{0.27}Ga_{0.73}As$ layer, a 50 nm graded $Al_xGa_{1-x}As$ layer with x increasing from 0.27 to 0.31, and a 60 nm $Al_{0.31}Ga_{0.69}As$ layer. Both supelattice wells and the contact layers are doped with 1×10^{18} cm-3 of Si. The blocking barrier and the superlattice barriers are left undoped.

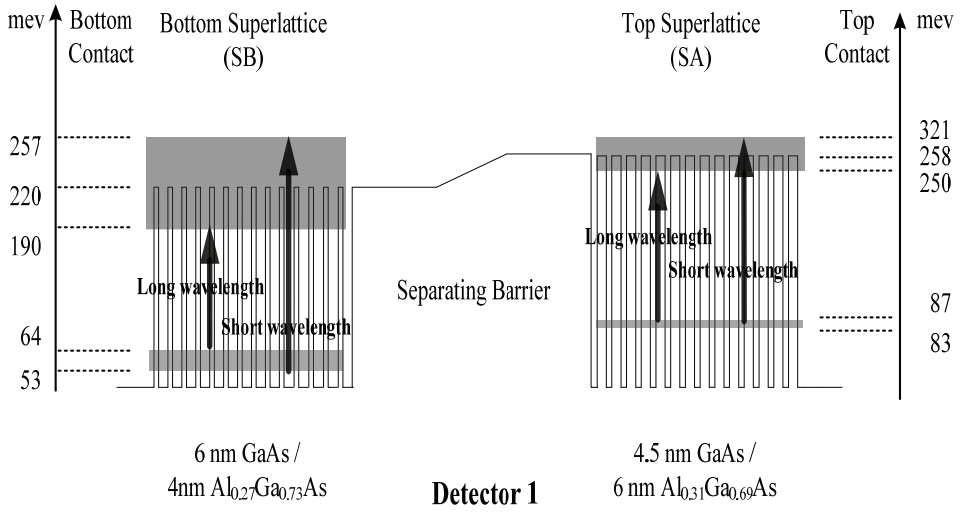


Fig. 2.1. The band diagram of Detector 1. It contains two superlattices and a separating barrier.

2.3 Operation Principle

To show the operation principles, the band structure under a positive bias is shown in Fig. 2.2. The voltage polarity is taken as positive if high potential is applied on the top contact. Because superlattices act as a low-resistance region, the applied voltage is almost totally dropped on the separating barrier. For convenience, we labeled top superlattices as superlattices A (SA), and the bottom one as superlattices B (SB). Under strong electric field applied on the barrier, the photoelectron in the second miniband of SB can tunnel through

the separating barrier. The escaped photoelectrons result in positive charges to attract electrons from the bottom contact and cause photocurrent in the external circuit.
On the contrary, the escaped photoelectrons in SA attract electrons from the top contact and result in internal current circulation as shown in Fig. 2.2. Therefore, only SB is active under positive bias. In the same way, only SA is active under negative bias. This characteristic makes the spectral responsivity switchable by the bias polarity between the two wavelength regimes corresponding to the respective miniband transition of SA and SB.

As shown in Fig 2.1, the barrier height of the blocking layer is actually a little bit higher than the energy level of the bottom edge in the second mini-band of the superlattice. This structure is designed for the operation of the voltage tunability. The operation principle for the voltage tunability in the superlattice with a single barrier is shown in Fig 2.3. The flat barrier of the blocking layer is assumed for the convenience to understand.

When the electron absorbs a photon to transit from the first miniband to the second miniband and become a photoelectron. The photoelectrons distributions corresponding to the second miniband energy are shown in the left and right insets of Fig 2.3 for long-wavelength and short-wavelength absorptions respectively. The insets under the same column are the same. For the long-wavelength absorption, the photoelectrons are always accumulated in the bottom state of the second miniband or relax into the first miniband as shown in the left inset. While for the short-wavelength absorption, most of the photoelectrons would relax from the top state of the second miniband into the bottom state or into the first miniband.

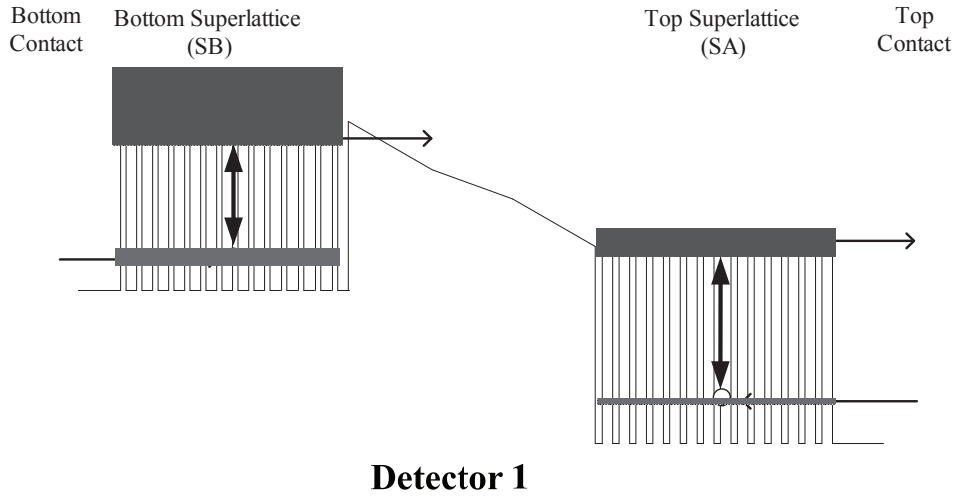


Fig. 2.2. The band diagram of Detector 1 under positive bias

Therefore, there are more photoelectrons in the bottom state than those in the top state as shown in the right inset. Fig 2.3 (a) and (b) demonstrate the tunneling behavior of the photoelectrons for the long-wavelength absorption under zero and high bias respectively, while Fig 2.3 (c) and (d) show the behavior of the photoelectrons for the short-wavelength absorption. For the long-wavelength absorption, especially under low bias, those photoelectrons which are accumulated in the bottom state of the second miniband can not

tunnel through the blocking layer as shown in Fig. 2.3 (a). When bias increases, for those photoelectrons, the tunneling path decreases and the associated tunneling photocurrent increases dramatically with tunneling probability increasing.

On the other hand for the short-wavelength absorption, photoelectrons with energy higher than barrier height can pass through the blocking layer even at zero bias as shown in Fig. 2.3 (c). As the applied bias increases, the number of photoelectrons, which can pass through the blocking layer, increases because of the increasing tunneling photoelectrons as shown in Fig. 2.3 (d).

In brief, under low bias, the photoelectrons generated by short wavelength radiation have higher energy and tunneling probability to contribute to the photoresponse. Short wavelength radiation dominates the spectral responsivity under low applied voltage. However, under high bias, the tunneling probability of the photoelectrons in the bottom state of the second miniband increases, and the long wavelength responsivity dominates the spectral responsivity. Fig 2.4 is the schematic illustration of the responsivity under low bias, medium bias, and high bias. Under low bias, the reponsitivety is dominated by the short-wavelength absorption. Under medium bias, the responsivity for the short-wavelength and long-wavelength absorption are compatible. Under high bias, the responsivity is dominated by the long-wavelength absorption.

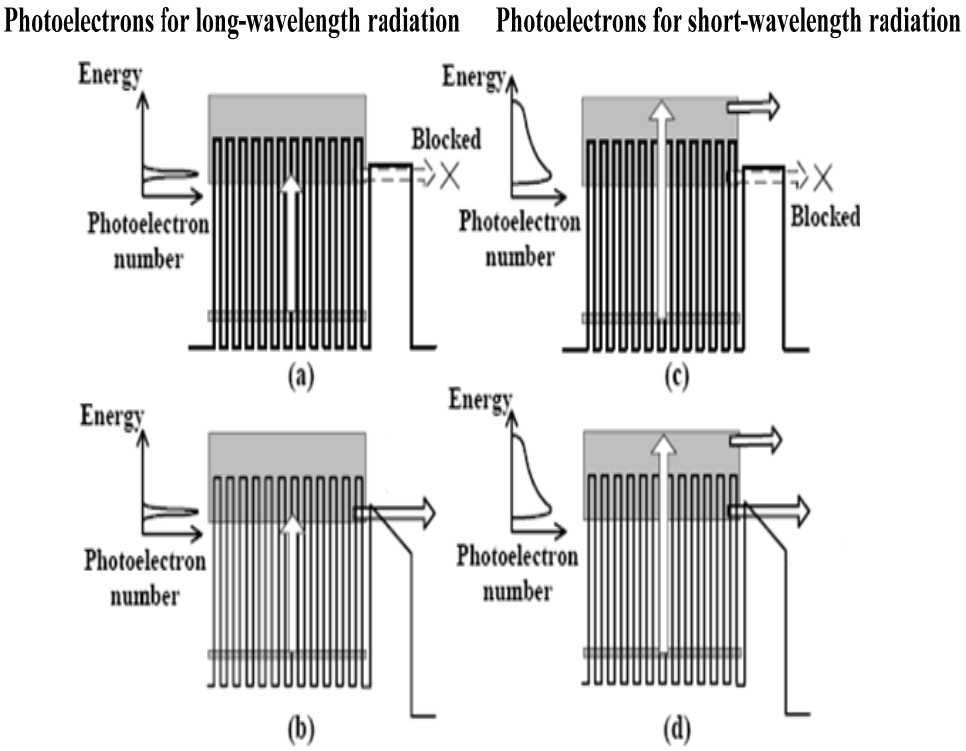


Fig. 2.3. Schematic illustration of photoelectron transport in a superlattice combined with a

blocking layer (a) under zero bias and illuminated by long-wavelength light (b) under high bias and illuminated by long-wavelength light. (c) under zero bias and illuminated by short-wavelength light (d) under high bias and illuminated by short-wavelength light. The distinct insets are the photoelectrons distribution corresponding to the second miniband energy for the long-wavelength and the short-wavelength absorption.

2.5 Responsivity

The measured spectral responsivity at 30 K is shown in Fig. 2.5. Under positive bias, only SB is active and dominates spectral responsivity, which is tunable in 7.5~12 μm through the magnitude of the applied bias. Under negative bias, the spectral responsivity is dominated by the SA, and is also tunable in 6~8.5 μm.

The theoretical calculation of the miniband as shown in Fig 2.1 demonstrates the short-wavelength and the long-wavelength absorptions for the SA happen at 5.2 μm and 7.6 μm and those for SB occur at 6 μm and 9.8 μm. The measured ones as show in Fig 2.5 are 6.8 μm and 7.6 μm for the short-wavelength and the long-wavelength absorptions of SA and 8.5 μm and 9.8 μm for the SB. The long-wavelength agrees very well but the short-wavelength does not. It is attributed to the relaxation of the most photoelectrons in the top states of the second miniband.

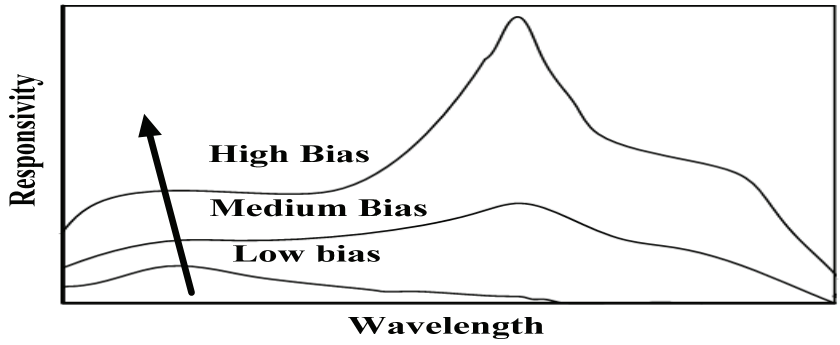


Fig. 2.4. Schematic illustration of the responsivity under low bias, medium bias, and high bias

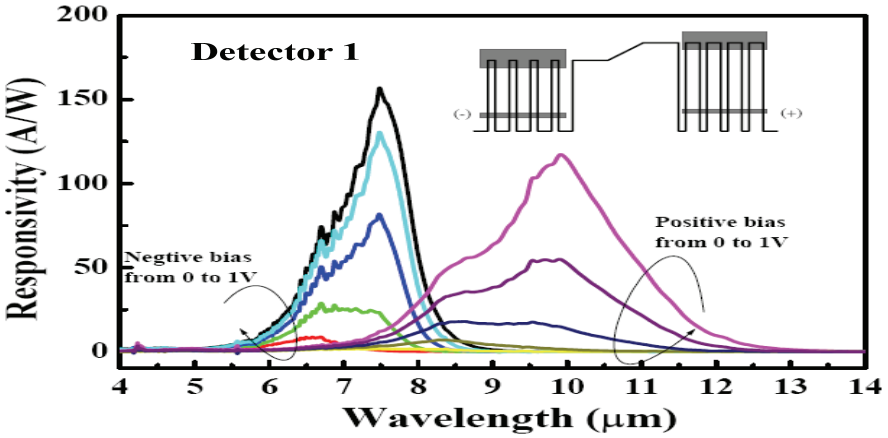


Fig. 2.5. The measured spectral responsivity at 30 K under several bias voltages

2.6 Detectivity

To evaluate the detectivity, the noise performance was measured from 1 to 8 kHz, which is limited by the noise and bandwidth of our amplifier. Under careful electrical and optical isolation, the current noise power spectral density at 77 K was measured with Detector 1 immersed in a 77 K liquid nitrogen dewar. With sophisticated calibration of the system noise, the detector noise under positive bias was extracted from the total measured noise, and is shown in Fig. 2.6. The noise of Detector 1 is white noise in the frequency range of our measurement system. Also shown in the inset of Fig. 2.6 are the measured noise data and the shot noise $2eI_d$ calculated with the dark current at 77 K. The dashed line in the inset represents the estimated minimum resolvable noise of our noise measurement system.

It is observed that the measured noise PSD agrees with the estimated shot noise at 77 K for voltage larger that 0.5 V. Therefore, it is concluded that the noise source comes from the shot noise of the electrons tunneling through the blocking barrier. Since the dark current at 77 K under negative bias is much smaller than that under positive bias, it is not resolvable in our noise measurement system. In the following evaluation of detectivity, we assume the noise of our detector is also shot noise of the dark current. The zero background peak detectivity calculated with the shot noise of our detector is 2.3×10^{10} cmHz$^{0.5}$/W at 50 K under 0.7 V with wavelength being 9.8 μm, and is 8.7×10^{10} cmHz$^{0.5}$/W at 70 K under –0.7 V with wavelength being 7.4 μm.

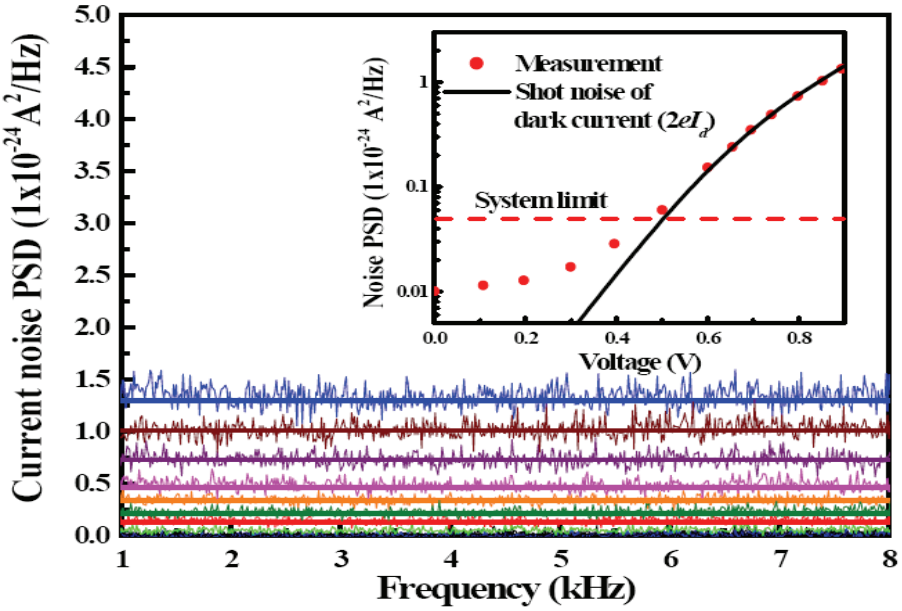


Fig. 2.6. The measured current PSD with the Detector 1 immersed in 77 K liquid nitrogen. Each line corresponds to each point in the inset. The solid curve represents the estimated shot noise calculated with the dark current at 77 K. The dashed line indicated the minimum resolvable noise of our measurement system.

2.7 Summary

In summary, we have designed and fabricated an infrared photodetector with two superlattices separated by a blocking barrier. The spectral responsivity of Detector 1 is tunable by the bias magnitude and is switchable by the bias polarities. These show that the structure is very appropriate to realize the multi-color infrared photodetector.

3. Double-Barrier Superlattice Infrared Photodetector

3.1 Introduction

An ordinary SLIP is realized by a SL structure with a single barrier. Because the SL is a low resistance structure, it is considered there is no applied bias dropped on the SL under the low bias. The photo-electrons in the second miniband of the SL can move either forward or backward(Chen et al., 2002) (Chen et al., 2003) (Lu et al., 2003) . Some backward and escaping photo-electrons have no contribution to photocurrent. For the forward photo-electrons, the probability for those electrons to tunnel through the single barrier is assumed to be P. Those photo-electrons which move forward but can not tunnel through the barrier will finally relax into the first miniband. The total loss of photo-electrons is calculated to be (1- P/2) x100% and must be higher than 50%.

In order to solve this problem, we add a thick barrier into the left of this structure to increase the probability of bounce for photo-electrons in the second miniband.Then photo-electrons can have higher opportunity to tunnel through the thin barrier and contribute to photoresponse. This detector denoted by double-barrier superlattice infrared photodetector.

3.2 Sample structure

The sample structure of this detector from bottom to top is as following: a bottom contact layer, a 500 nm $Al_{0.28}Ga_{0.72}As$ thick barrier, a 15-period SL, a 50 nm $Al_{0.28}Ga_{0.72}As$ thin barrier and a top contact layer. Each period of the SL is composed of 6.5 nm GaAs well ($N_d=2\times10^{17}$ cm^{-3} Si) and 3.5 nm $Al_{0.32}Ga_{0.68}As$ barrier (undoped). The sample was fabricated with the 100 μm × 100 μm square mesas then etched down to the bottom contact layer and evaporate metal onto the top of each mesa and the bottom contact. Hence, electrons must traverse the thick and thin barriers to generate the current. A 45° facet on the GsAs substrate is made to allow the TM polarized infrared light radiate on the photodetector to measure FTIR spectrum. The absorption wavelength range of the SL is about 6 to 10 μm. The positive bias polarity is defined if high potential is applied on the top contact.Above device is denoted by Detector 2.

3.3 IV characteristic and excess electrons in the wells

The temperature-dependent current-voltage (I-V) curves of Detector 2 are shown in Fig. 3.1. The solid curves and the dashed curve are dark current at 50-100 K and photocurrent measured at 20 K under 300 K background radiations, respectively. A very asymmetric I-V curve is observed. The current magnitude of positive bias increases more dramatically than that of negative bias.

A possible mechanism qualitative to explain the dramatic asymmetry of the IV curve is shown in Fig. 3.2 (a) (b) and(c) which correspond to the negative, low positive, and high positive biases respectively. As bias magnitude increases, most of bias will drop on the thick barrier. Under the negative bias as shown in Fig. 3.2 (a), a small amount of positive ion charges appear in the left of the thick barrier due to the high doping density, a small amount of voltage drop is on the thin barrier. Electrons can be injected from the top contact through the thin barrier then relax onto SL. At the same time, the same amount of electrons is injected from SL through the thick barrier onto the bottom contact. No current component injected directly from the top contact through the SL into the bottom contact is expected to appear.

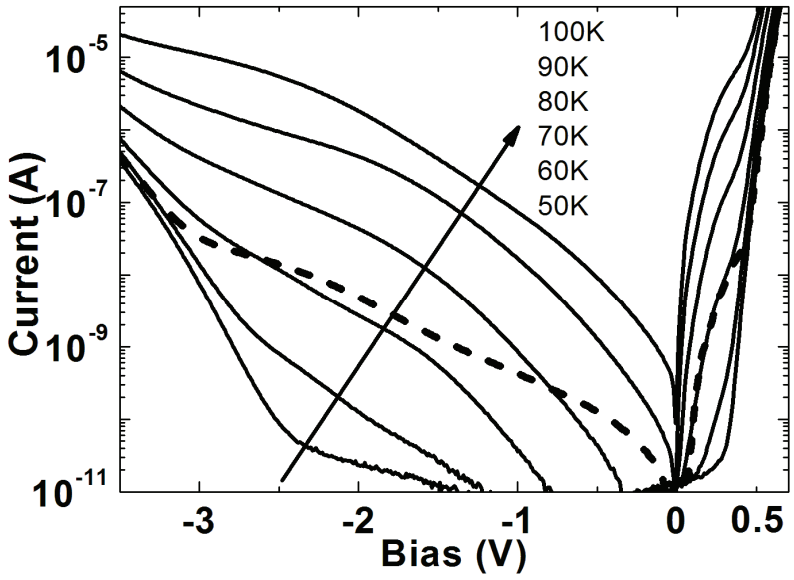


Fig. 3.1. (a) The dark current (solid lines) at different temperatures and the photocurrent (dashed line) versus the bias voltage of the Detector2.

On the other hand, under the positive bias as shown in Fig. 3.2 (b) and (c), the electrons in the SL near the thick barrier are repelled to move toward the thin barrier and a large amount of positive ion charges appear in this region due to the low doping density 4×10^{17} cm^{-3} in SL compared with the doping density, 1×10^{18} cm^{-3} in the contacts. A part of repelled electrons is expected to tunnel through the thin barrier and accumulate near the left side of the thick barrier through the external circuit as shown in Fig. 3.2 (b) and(c). The other part of repelled electrons will be trapped in the wells due to the breakdown of SL miniband. It is noted the SL still have miniband near the thin barrier. The trapped electrons are denoted as excess electrons in the wells. As bias increasing, the excess electrons in the wells tend to tunnel through the thin barrier. As bias is high enough, Fig. 3.2 (c) demonstrate the band diagram after the excess electrons all tunnel through the thin barrier.

In Fig. 3.2 (b), due to the misalign of the electron states in the wells under the region of positive electric field ,the excess and doping electrons can not move toward the positive ion charges to neutralize them. A large negative field on the thick barrier produced by the positive ion charges and accumulating electrons at the bottom contact near the thick barrier renders the dark current to increase dramatically. Most of electrons injected from bottom contact through the thick barrier can be accelerated by positive ion charges and pass through the SL and thin barrier directly into the collector contact .Some electrons will relax into the SL and the same amount of electrons are injected through the thin barrier into the top contact. The amount of the electrons injected directly from the bottom contact into the top contact is expected to increase as the positive bias increases.

In addition in Fig. 3.2 (b), due to the positive electric field formed by the positive ion charges and excess electrons in the wells, the potential energy rise from the minimum point.

Therefore the total bias between the bottom and top contacts is about one tenth biases on the thick barrier. This can explain the asymmetry and the positive bias range is about ten times larger than negative bias range for the same current level. In other words, for the same magnitude of negative and positive biases, the dark current in positive bias is much larger than that under negative bias.
In order to solve this problem, we etch the sample down to SL to fabricate the metal contact on it directly instead of the bottom contact layer. We denoted this device as Detector 3 to differentiate from the above bottom contact sample Detector 2. Fig. 3.3 (a) and (b) show the band diagrams and the related cross-sectional plots of Detector 2 and Detector 3 respectively. For Detector 3, electrons escaped form the SL can be supplied from the metal contact immediately to overcome this problem and to improve the performance.

The I-V curves of Detector 3 at different temperatures are shown in Fig. 3.4. In comparison with the curves shown in Fig. 3.1, the operating bias range under both biases is symmetric. The inset shows the current comparison of Detector 2 and the Detector 3, it is observed that the Detector 3 has the lower dark current. Similar to the comparison of IV curve under positive and negative biases in Detector 2, it is attributed that no high injected current occur directly from the bottom contact to the top contact in Detector 3.

3.4 Responsivity and multiple bounce of the photoelectrons

The spectral responsivity under negative and positive biases of Detector 2 is shown in Fig. 3.5. (a),(b) respectively. For this spectra responsivity, the operational voltage ranges under positive and negative biases are consistent with the voltage range of I-V curve as shown in Fig. 3.1. The negative applied bias has to be ten times larger than that under positive bias for the same magnitude of responsivity. The spectral responsivities under positive bias of Detector 3 and 4 are shown in Fig. 3.5 (c) (d) respectively, where Detector 4 is a SLIP which has the same SL structure and single thin barrier. The applied bias ranges from 0.05 to 0.35V in both detectors. It is noted that the spectral responsivities of these four figures have the similar dependence on the bias with that as shown in Fig. 2.4. that is the short wavelength response appears under the low bias, the long wavelength response dominates under the high bias, and they are comparable under medium bias.

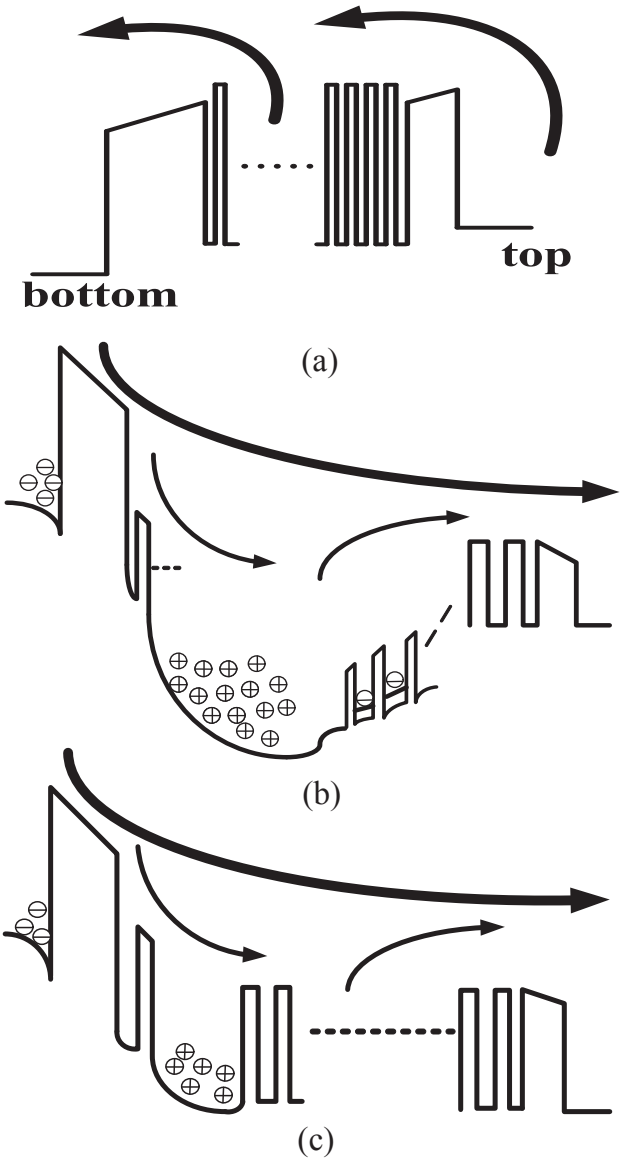


Fig. 3.2. Schematic illustration of band diagram of Detector 2 under (a) negative bias (b) low positive bias (c) high positive bias

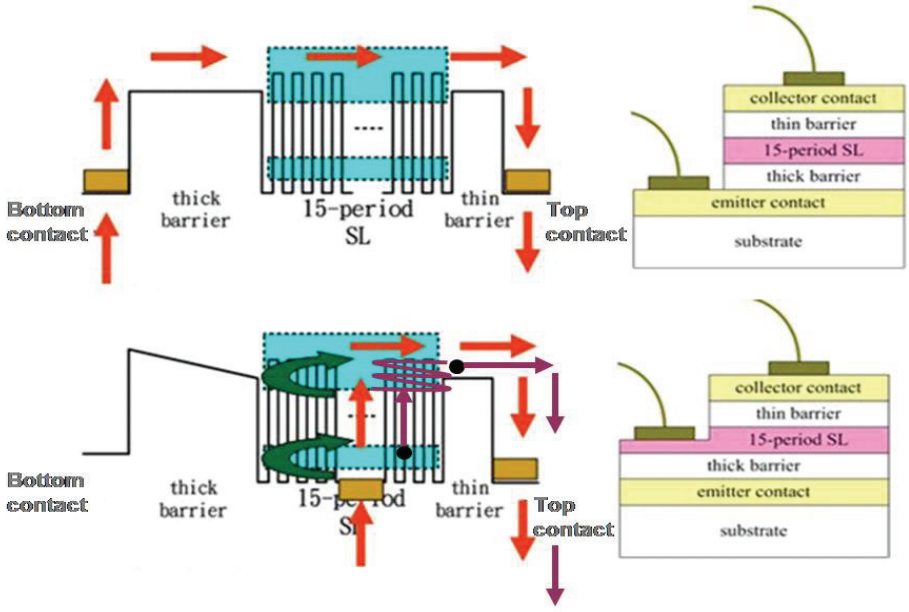


Fig. 3.3. The schematic band diagrams and the cross-sectional plots of (a) Detector 2 and (b) Detector 3. The arrows show the electron transport in these samples. For Detector 2, electrons have to transverse both of the barriers. However, in Detector 3, electrons have to transverse only the thin barrier.

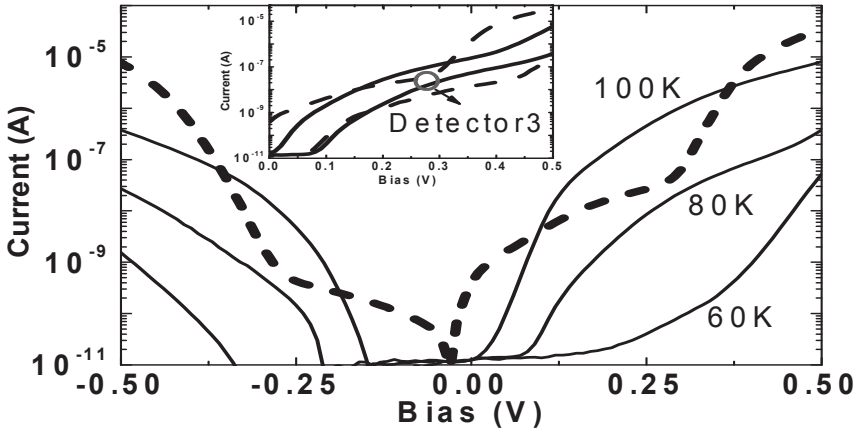


Fig. 3.4.(a) The dark current (solid lines) at different temperatures and the photocurrent (dashed line) versus the bias voltage of Detector 3. Inset shows the current comparison of Detector 2 and Detector 3. The solid lines and dashed line present the photocurrent and the dark current at 80K, respectively.

To further demonstrate the photocurrent improvement of Detector 3, the peak responsivities of Detectors 2, 3 and 4 at 80 K under several positive biases are shown in Fig. 3.6. Detector 3 has the higher responsivity than Detectors 2 and 4 under all applied biases. In particular at very low bias such as 0.15 V, the responsivity at 9.2 μm of Detector 2, at 9.4 μm of Detector 3 and at 9.2 μm of Detector 4 .It is noted that the responsivity of Detector 3 increases smoothly under all biases, but there are threshold voltages at 0.15V and 0.20V for the Detector 2 and 4 respectively. Especially under 0.15V, the responsivity of Detector 3 is at least ten times higher than those of Detectors 2 and 4. This is consistent with the comparison of the resulted background photocurrents of Detectors 2 and 3 shown in the inset of Fig. 3.4.
Under low positive bias, photoexcited electrons of Detector 3 can bounce several times in second miniband and enhance probability of tunneling through the thin barrier meanwhile the supply of electrons from the SL contact is no longer difficult as shown in Fig. 3.3(b).The similar hyperbolic shape of Detector 2 increases the threshold voltage of Detector 2 even to the higher bias than that of Detector 3. Under high positive bias, photoexcited electrons in second miniband of Detectors 3 and 4 can tunnel through the thin barrier at one time. So we can observe the responsivity difference between Detector 3 and 4 gradually decreased with bias increasing.

3.5 Detectivity

Because Detector 3 has the much higher responsivity about ten times than Detector 2 and Detector 4 at very low bias, its detectivity is also better in that range. Table3.1 shows the comparison of detectivity at 80 K of the three detectors. It is observed that Detector 3 has the highest detectivity than the other detectors. The maximum detectivity value of Detector 3 is 1.2×10^{10} cmHz$^{1/2}$/W for 9.4um wavelength, which is higher than all of our previous works at 80K(Chen et al., 2002) (Chen et al., 2003). Therefore, the above analysis shows that the thick barrier is helpful to enhance the photoresponse, especially under low bias.

Detector number @ 80K	Wavelength	Detectivity (cmHz$^{1/2}$/W)
Detector 2	9.22 μm	1.2×10^{9}
Detector 3	9.4 μm	1.2×10^{10}
Detector 4	9.7 μm	3.5×10^{9}

Table 3.1. Shows the comparison of detectivity at 80 K of the three detectors.

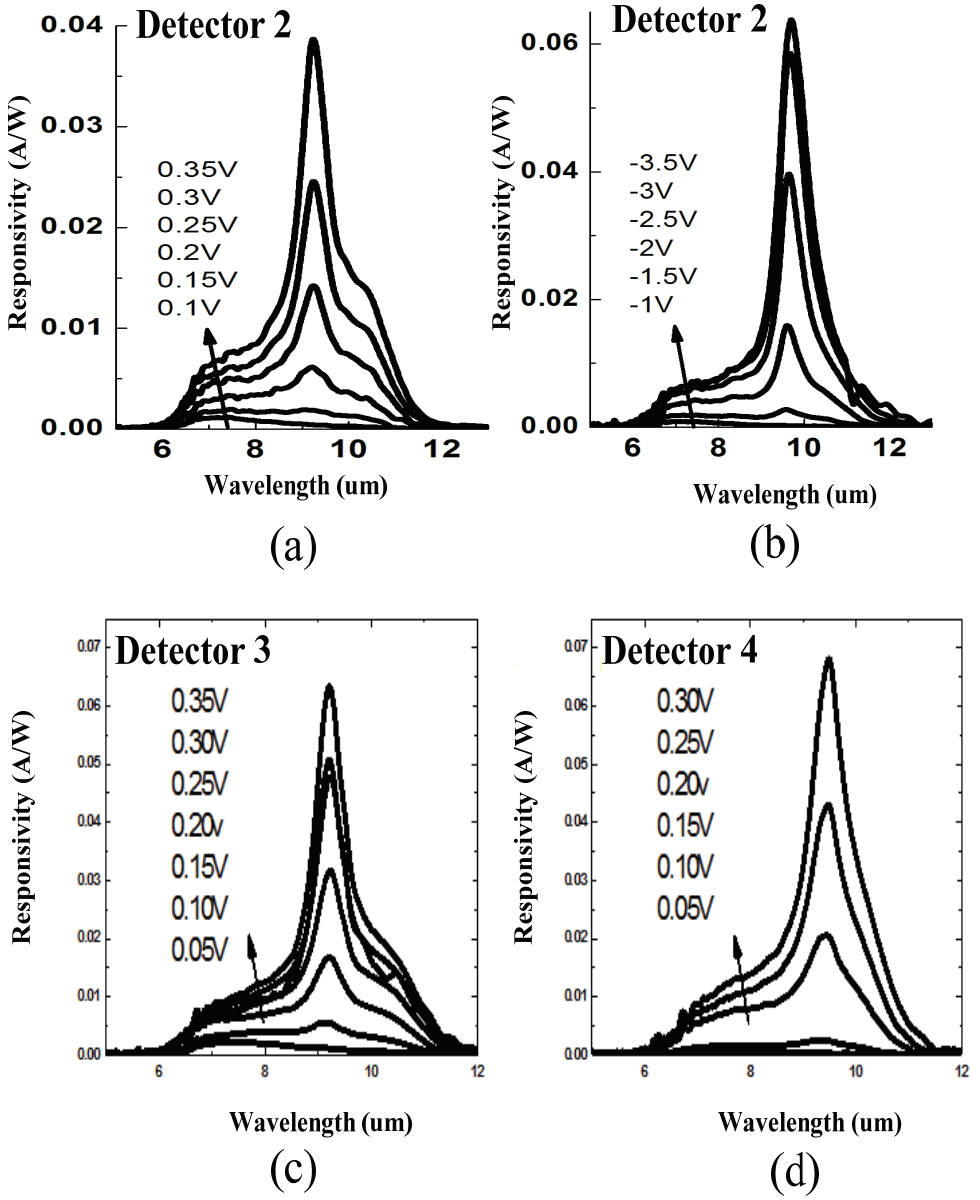


Fig. 3.5. (a) (b) The spectral responsivity of Detector 2 at 80 K under positive and negative bias respectively. (c) (d) The spectral responsivity of Detector 3 and 4 at 80 K under positive respectively.

3.6 Summary

We have investigated a double-barrier SLIP whose structure is a SL sandwiched between two asymmetric barriers. The thick and thin barriers can let the photoelectrons bounce several times in the second miniband of the SL to improve the responsivity dramatically at very low bias. However, the excess electrons trapped in the SL,the accumulating electrons at the bottom contact near the thick barrier and positive ion charges inside the SL result in high dark current under a small bias. A large electric field across the thick barrier makes the performance of the detector worse. The metal contact is fabricated directly onto the SL instead of the bottom layer to solve this problem. Detector 3 demonstrates higher responsivity under very low bias, although we demonstrate the responsivity at 80K yet it can be operated at 100K. Detector 3 also shows the higher responsivity at 80k than our previous works for the Detector 4.

4. The SLIP integrated with MQWs

4.1 Inroduction

In order to reduce the dark current and noise power under low bias, an ordinary SLIP with a thin barrier followed by additional MQWs has been developed as shown in Fig. 4.1. The whole structure can be divided into two parts. The SL is used as an active region for the primary photoresponse while the MQWs are considered as a noise filter for noise reduction. Because of the high capture probability of the wells for the electrons under low bias, almost all electrons traversing a barrier are recaptured into the neighboring well. The photoelectrons and darkelectrons from the SL are hence recaptured in the first well of the MQWs and produce the capture current. This current must be balanced by the emission current from the well under steady state due to the charge neutrality. It assures that the total current, especially the photocurrent generated from SL can also flow through the whole structure without loss under low bias as indicated by the arrows in Fig. 4.1.

Furthermore, MQWs are capable of providing additional photocurrent to improve photoresponse under low-voltage operation.

Because of thermally assisted tunneling of the electrons through the barrier, the noise of the SLIP with a single barrier is proved in Fig 2.6 as shot noise:

$$i_{SB}^2 = 2eI_D$$

where i_{SB} is the noise current power spectral density (PSD) of the SLIP and I_D is the dark current. On the other hand, the noise of MQWs is the generation-recombination noise.

$i_{QW}^2 = 4eg_nI_D$ and $g_n = 1/(N\,p_c)$

where i_{QW} is the noise current PSD of the MQWs, g_n is the noise gain of MQWs, I_D is the dark current, and N is the number of wells. The p_c is the capture probability of the single well and is approximately 1 under low bias when the noise gain is the smallest as indicated by the downward arrows in Fig. 4.1. The total noise current PSD of the combination of SLIP with MQWs is still the generation and recombination noise since their current always passes through the MQWs. Without the loss of the photocurrent from the SL, the MQWs can act as a noise filter under the low bias.

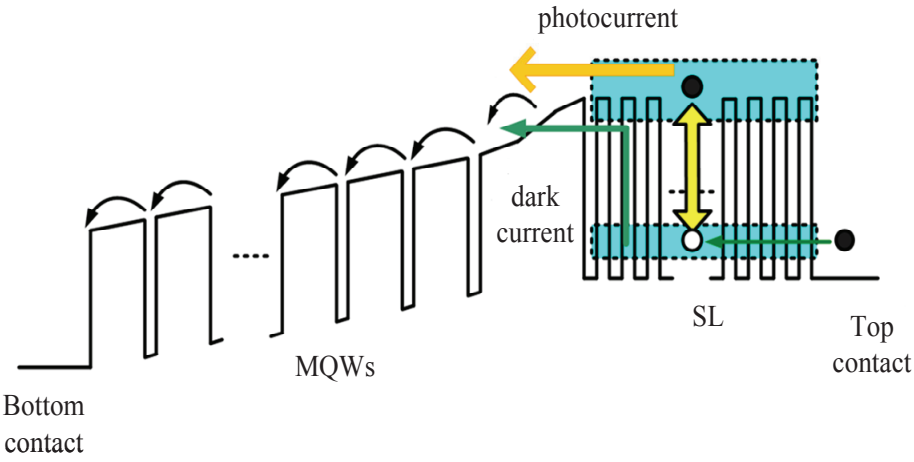


Fig. 4.1. The band diagram of SL integrated with MQWs.

4.2 Sample Structure

For the comparison of the differences between SLIP and SLIP with MQWs, Detector 5 and Detector 6 are designed. The structure of Detector 5 from top to bottom is as follows: a contact layer, a SL, a 60 nm $Al_xGa_{1-x}As$ graded barrier, a 50-period MQWs and a contact layer. Detector 6 is a SL with a single barrier sandwiched between two contact layers. The single barrier is a 300 nm $Al_{0.25}Ga_{0.75}As$ layer. For the MQWs, each period consists of 6nm GaAs well (n = 4 x 10^{17} cm^{-3}) and 50nm $Al_{0.21}Ga_{0.79}$ As barrier (undoped). And each period of the SL consists of 6nm GaAs well (n = 4 x 10^{17} cm^{-3}) and 4nm $Al_{0.29}Ga_{0.71}As$ barrier (undoped).

4.3 Noise gain

It is clearly that if gn is less than 0.5, the noise of SLIP with MQWs is lower in comparison with an ordinary SLIP. Fig. 4.2 shows the measured noise gain of SLIP with 50-period MQWs. It is noted that the noise gain is very small under the low bias range. For example, i_{QW}^2 can be 50 times less than i_{SB}^2 at 0.01V. It is the reason why the SLIP with MQWs is appropriate for the operation at low bias as mentioned above.

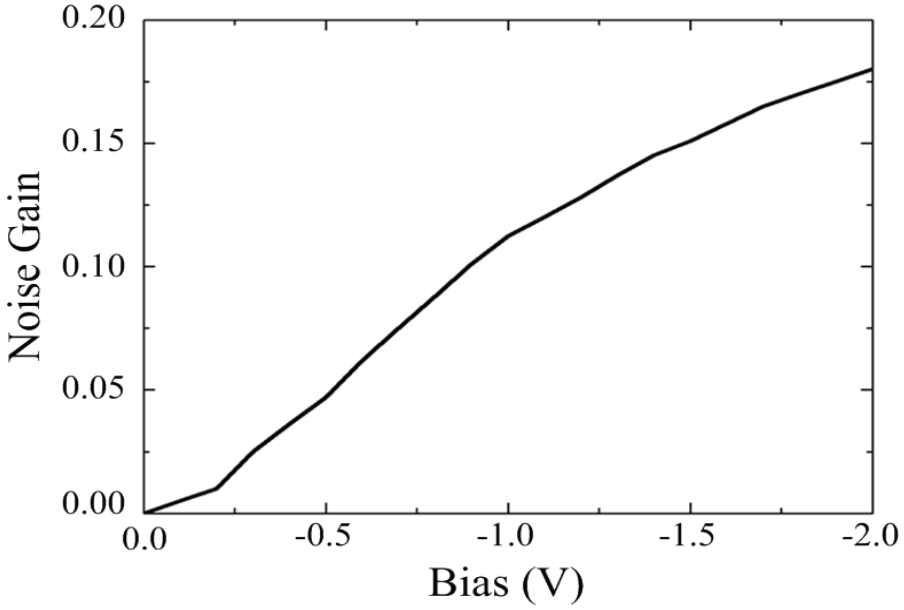


Fig. 4.2. The noise gain of SL integrated with MQWs under negative bias at 80 K.

4.4 Responsivity

Fig. 4.3 shows the photoresponse of Detector 5 at 80K with different biases. The line without solid squares is for the positive bias on the top contact and the other one is for the negative bias. Based on the band diagram in Fig. 4.1, only the MQWs are active under the positive bias while both the SL and MQWs are active under the negative bias.

As shown in the Fig. 4.3, it is obvious that the responsivity magnitude increases with the bias magnitude for both bias polarities. By comparing the spectral range for both bias polarities, it is concluded that the SL has a responsivity peak at 7.4μm which correspond to short wavelength absorption in the SL structure and the MQWs have a peak at 11.4μm. Especially the responsivity for the MQWs increases quickly under the positive bias range but less than or compatibly with the SL one under the negative bias range. Therefore it is expected that the SLIP with the MQWs can be operated under 0V to -0.7V. In particular, the SL responsivity is dominant for 0V to -0.3V which the noise gain is also very small.

Fig. 4.4 shows the responsivity of Detector 5 under -0.2V bias at 40K, 60K and 80K. The responsivity spectral range is actually compatible with 300K blackbody radiation spectrum which is sketched as the dotted curve. For the responsivity, the functions of the MQWs are not only to act as a noise filter but also to enhance the long wavelength response for broadband detection.

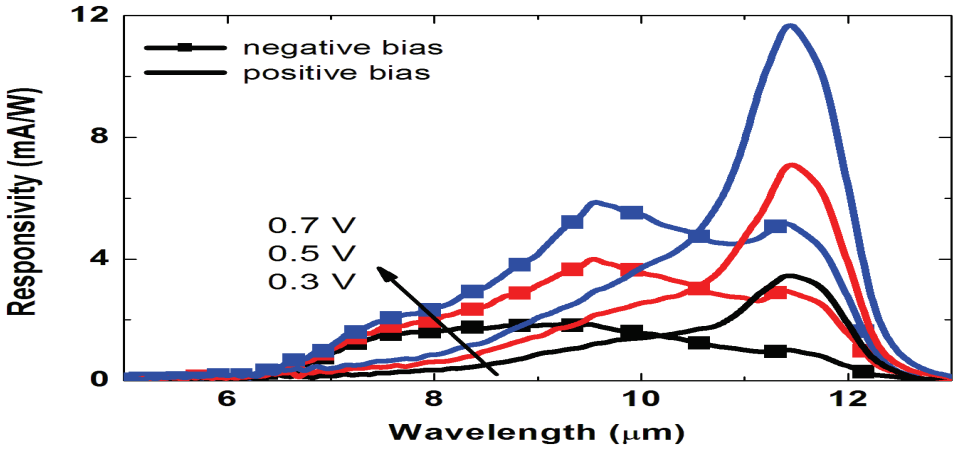


Fig. 4.3. The photoresponse comparison of Detector 5 under different positive and negative biases at 80 K. It is observed that the response spectra under negative bias are more broadband and flatband than those under positive bias.

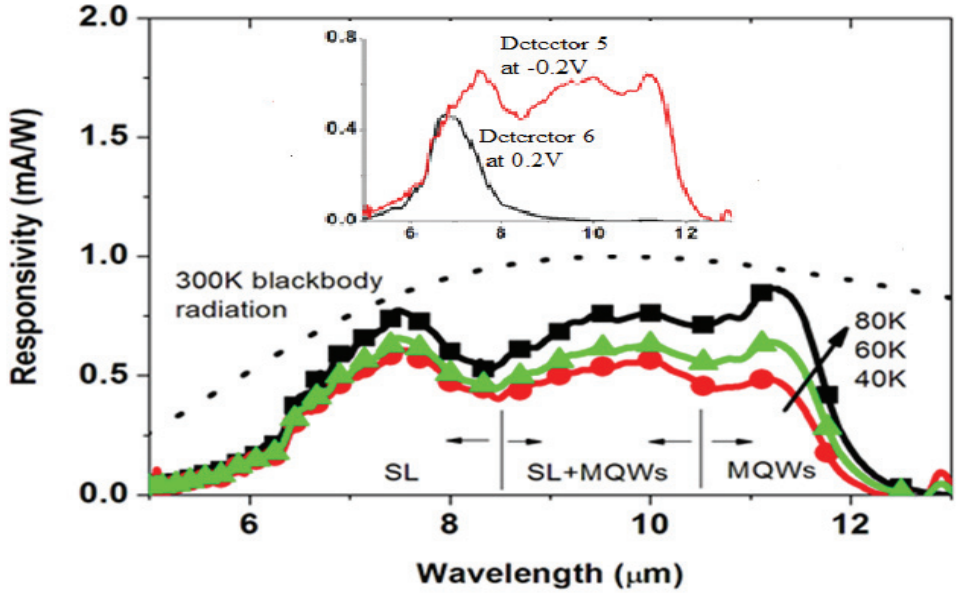


Fig. 4.4. The photoresponse of Detector 5 under -0.2 V at different temperatures. Its detection range is compatible with the intensity distribution of 300 K blackbody radiation which is sketched as the dotted line (not in scale). The inset shows the response of Detector 5 at -0.2 V and that of Detector 6 at 0.2 V

The reason why the responsivity increases a little bit with the temperature rising is attributed to the low doping density of 4 x 10^{17} cm-3 in the wells of the SL and MQWs. Some

of the electrons are frozen in the impurity states in the wells. As the temperature rises, they can escape from the impurity states into the well to improve the responsivity. This freezing effect actually becomes the less at the higher temperature .Comparing Detector 1 in sec2 and Detector 5 in this sec, the responsivity of Detector 1 is temperature independent because of high doping density in the wells and no freezing effect occur . In additions, the advantage for the low doping density is to decrease the dark current in order to improve the detector operation at the high temperature.
In the inset of Fig. 4.4, the responsivity comparison between Detector 5 and Detector 6 is demonstrated. The bias is given under the same electric field on the blocking layer in the two samples. It is obvious that Detector 5 has a broader spectrum than Detector 6 due to the additional contribution of the MQWs.

4.5 Detectivity

Fig. 4.5(a) shows the detectivity (D*) of Detector 5 under positive and negative biases at 50K, 60K, 80K. Detector 5 consists of SL and MQWs, the responsivities for the wavelength 11.7 μm(MQW peak) and 7.4 μm(SL peak) were chosen respectively to calculate the D* under positive and negative biases. Under negative bias, as bias magnitude decrease from -0.7V the D* increase quickly. On the other hand, under the positive bias, the D* decrease when the bias magnitude decrease from 0.7V. It is attributed to the additional contribution of the SL responsivity under the low negative bias except the contribution of MQWs.
Despite the small SL responsivity under the low negative bias, both the corresponding dark current and the noise gain are low enough to make D* increase as the bias magnitude decreases. Hence the maximum D* under positive bias (MQWs) is still less than maximum D* under negative bias (SL+MQWs) at the same temperature. Fig. 4.5(b) shows the comparison of D* between Detectors 5 and 6 for 7.4 μm. The D* under negative and positive biases represents the respective performance for the Detectors 5 and 6. The D* of Detector 5 is better than that of Detector 6 under the same bias magnitude for any temperature although they have similar responsivity. It is attributed to the different noise performance i.e., generation-recombination noise noise and shot noise for Detectors 5 and 6 respectively.

4.6 summary

The SL integrated with the MQWs shows low noise, and broadband response under the low bias. In comparison with MQWs, the additional contribution of SL responsivity renders the better performance in our detector. In comparison with ordinary SLIP, the MQW in our detector shows better noise performance. Therefore this detector is appropriate for operation under the low biases and high temperature (≥80K).

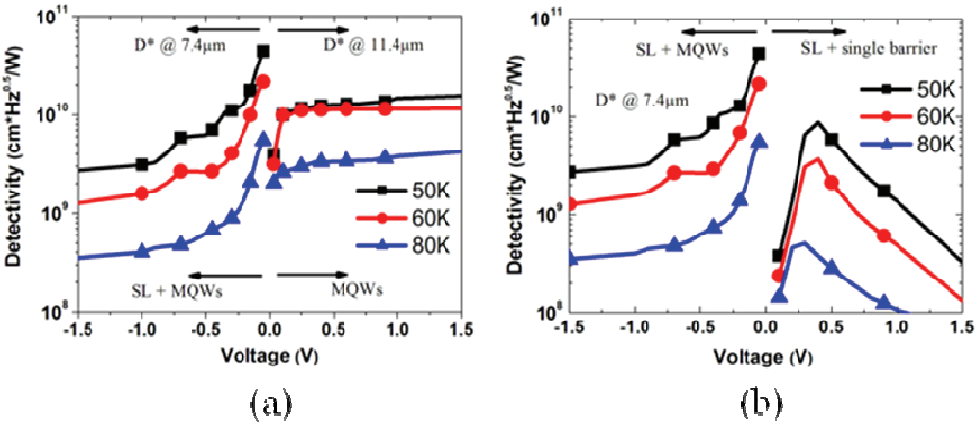


Fig. 4.5. (a) The detectivity (D*) of Detector 5 versus voltage at 50 K, 60 K and 80 K under 7.4μm and 11.4μm. Because the photoresponse comes from the different structures, the corresponding D* wavelengths under positive and negative bias are 11.4 and 7.4 μm, respectively. (b) The comparison of D* at 7.4 μm between Detector 5 and Detector 6.

5. The conclusion and future work

Various SL structures and blocking layers are combined to form a detector with a specific application. At first, a blocking barrier separated two distinct superlattices was made to obtain the multi-color detection which demonstrate the flexibility of the SL structure. In comparison with the conventional multicolor photodetectors, the double-superlattice photodetectors show advantages including temperature invariance of the spectral responsivity, the flexibility for the design of a multicolor photodetector and the same order of responsivities in the two different wavelength regions. But the main disadvantage is its low operational temperature. It is also noted the noise performance of such SLIPs belongs to shot noise compared with generation-recombination noise of the QWIPs the shot noise actually higher than generation-recombination noise under the low bias. For a detector operated under the low bias, the noise performance should be improved.

In order to increase the operation temperature, we design the double barriers with SL inserting. The additional thick barrier compared to ordinary SLIP is used to bounce electrons backward to enhance the photoresponse. However, the dark current increase dramatically under low bias and threshold voltage of photocurrent is postponed. The possible model is given to explain this phenomenon inside the SL the excess electrons trapped in the wells and positive ion charges near the thick barrier form a large electric field to reduce the potential bias. At the same time, the accumulating electrons in the contact near the thick barrier and positive ion charges inside the SL form a large electric field to increase dark current dramatically. To solve this problem, the metallic contact is made on the SL instead of the bottom contact. For this detector, we observe a higher responsivity than the ordinary SLIP which has only one thin barrier. The thick barrier can really enhance the photoresponse of SL, especially under low bias. The resulted operation temperature for this detector is higher than 80K.

To improve noise performance under low bias, we further design a SL followed by a

multiple quantum wells (MQWs). The MQWs are used to reduce the noise current power and add the response range. The noise gain of generation and combination noise is very small under low bias compared to the shot noise. On the other hand, the SL is utilized to inject the photocurrent into the QWIPs under low bias. In fact, the photocurrent of the SL is not reduced by the MQWs. As the result, this detector has the higher D* than the ordinary SLIP and QWIP, at the temperature of 80K.

The double-superlattice infrared photodetectors show multicolour detection and the flexibility of SL structure while the double-barrier superlattice infrared photodetectors and the SLIP integrated with MQWs can be operated at high-temperatur with high detectivity. The combination of quantum well infrared photodetector (QWIP) and double-barrier SLIP will be completed in the future. Even the operating temperature is expected to be above 100K with high detectivity .Such kind of SLIP is promising candidate of a pixel in the focal plane array especially such detector would not saturate the capacitor in ROIC because of the low dark current under low bias.

6. References

A. S. Sedra, and K. C. Smith, "Microelectronic Circuits", Oxford University Press, USA, 4 edition, ch. 4

B. F. Levine (1993). Restore Desktop View

C. C. Chen, H. C. Chen, C. H. Kuan, S. D. Lin, and C. P. Lee (2002). Multicolor infrared detection realized with two distinct superlattices separated by a blocking barrier. Appl. Phys. Lett. 80, 2251.

C. C. Chen, H. C. Chen, M. C. Hsu, W. H. Hsieh, C. H. Kuan, S. Y. Wang, and C. P. Lee (2002). Performance and application of a superlattice infrared photodetector with a blocking barrier. J. Appl. Phys. 91, 943.

Chen, C. Y. & Kuan C. H. (2000). Design and calibration of a noise measurement system. IEEE T.I.M. , 77, 49

Gravé, I. ; Shakouri, A. ; Kuze, N. & Yariv, A. (1992). Voltage-controlled tunable GaAs/AlGaAs multistack quantum well infrared detector. Appl. Phys. Lett. , 60, 2362

Gumbs, G. ; Huang, D. & Fessatidis, V. (1994). Many-body effects on temperature dependence of the interband absorption in quantum wells. J. Appl. Phys. , 75, 7942

Gunapala S. D. et al. (2000). 640×486 long-wavelength two-color GaAs/AlGaAs quantum well infrared photodetector (QWIP) focal plane array camera. IEEE T-ED, 47, 963

Harwit, A. & Harris Jr., J. S. (1987). Observation of Stark shifts in quantum well intersubband transitions. Appl. Phys. Lett. , 50, 685

Hsu, M.C. ; Kuan, C.H. & Wang, S. Y. (2000). Multicolor infrared detection using two stacks of superlattice structures in a back-to-back configuration. Appl. Phys. Lett. , 77, 2240

J. H. Lu, Y. Y. Yang, C. C. Chen, C. H. Kuan, H. T. Chen, and S. C. Lee (2003). Study of period number effect in the superlattice infrared photodetector. Infra. Phys. & Technol. 44, 399.

Jiang, X. D. ; Li, S. S. & Tidrow, M. Z. (1999). Investigation of a multistack voltage-tunable four-color quantum-well infrared photodetector for mid- and long-wavelength infrared detection. IEEE J. Quant. Electron. , 35, 1685

K. K. Choi, "The Physics of Quantum Well Infrared Photodetectors," World Scientific, River Edge, NJ, 1997, pp.294

Köck, A. ; Gornik, E. ; Abstreiter, G. ; Böhm, G. ; Walther, M. & Weimann, G. (1992). Double wavelength selective GaAs/AlGaAs infrared detector device. Appl. Phys. Lett. , 60, 2011

Levine, B. F. ; Choi, K. K. ; Bethea, C. G. ; Walker, J. & Malik, R. J. (1987). New 10 μm infrared detector using intersubband absorption in resonant tunneling GaAlAs superlattices. Appl. Phys. Lett. , 50, 1092

Liu, H. C. ; Li, J. ; Thompson, J. R. ; Wasilewski, Z. R. ; Buchanan, M. & Simmons, J. G. (1993). Multicolor voltage-tunable quantum-well infrared photodetector. IEEE Elect. Dev. Lett. , 14, 566

Lu, J. H. ; Yang, Y. Y. ; Chen, C. C. ; Kuan, C. H. ; Chen H. T. & Lee, S. C. (2003). Study of period number effect in the superlattice infrared photodetector. Infra. Phys. & Technol. , 44, 399

Majumdar, A. ; Choi, K. K. ; Rokhinson, L. P. & Tsui, D. C. (2003). Temperature dependence of electron transfer in coupled quantum wells. Appl. Phys. Lett. , 82, 686

Martinet, E. ; Rosencher, E. ; Luc, F. ; Bois, Ph. ; Costard, E. & Delaître, S. (1992). Switchable bicolor (5.5–9.0 μm) infrared detector using asymmetric GaAs/AlGaAs multiquantum well. Appl. Phys. Lett. , 61, 246

Tsai, K. L. ; Chang, K. H.; Lee,C. P. ; Huang, K. F. ; Tsang, J. S. & Chen, H. R. (1993). Two-color infrared photodetector using GaAs/AlGaAs and strained InGaAs/AlGaAs multiquantum wells. Appl. Phys. Lett. , 62, 3504

West, L. C. & Eglash,S. J. (1985). First observation of an extremely large-dipole infrared transition within the conduction band of a GaAs quantum well. Appl. Phys. Lett. , 46, 1156

7

"Bottom-up" Approaches for Nanoelectronics

Mrunal A. Khaderbad[1], Arindam Kushagra[1],
M. Ravikanth[2] and V. Ramgopal Rao[1]
[1]*Centre for Excellence in Nanoelectronics, EE Department, IIT Bombay,*
[2]*Department of Chemistry, IIT Bombay,*
Powai, Mumbai-400076
India

1. Introduction

Over the last 40 years, feature sizes in complementary metal oxide semiconductor (CMOS) technologies have been scaled from 3 µm to the current sub-50 nm using the "top-down" scaling techniques (Nowak, 2002). This "scaling" has resulted in an increased processing power and transistor density while reducing the cost per transistor (Chao Li et al., 2007). These classical methods employ a sequence of deposition, pattern definition, doping, lithographic and etching steps to build solid-state semiconductor devices and integrated circuits. As the process technologies scale beyond the sub-10nm feature sizes, above fabrication methods result in increased process costs, variability and longer fabrication turnaround times. To push the CMOS technology to its limits and to reap the benefits of scaling, non-traditional alternatives are needed while fabricating devices. One such approach is the use of "bottom-up" nanotechnologies or even a combination of the bottom-up and the "top-down" fabrication methodologies (Wei et al., 2007). In the bottom-up approach, analogous to the biological systems, atoms or organic molecules are self-assembled to build electronic structures with novel electronic, optical, or magnetic properties. Integrated circuits obtained with this approach of molecular-level control of material composition and structure may lead to devices and fabrication strategies not possible with top-down methods (Amarchand Satyapalan et al., 2005).

The field of self-assembled monolayers (SAMs), especially mono and multilayer assemblies of organic materials on various substrates has been most extensively studied in recent years. The compelling force for this research is the importance of such formations to nanoelectronic device fabrication, for modifying the surface wetting/adhesion properties, for sensor applications and for corrosion resistance and molecular electronics (Kaushik Nayak et al. 2007). With the limitations in lithography techniques, to get features smaller than 10 nm, the molecular self-assembly provides the route to both smaller features and lower costs. Because of the ability to form layers with atomic resolution thickness and spacing, they are used as ultra thin resists and passivating layers. CNTs, polyphenylenes, porphyrins and DNA strands are some of the molecules that are being actively researched for the above applications (Reimers et al., 1996).

Porphyrin and metalloporphyrin systems are excellent materials for molecular electronics because of their diverse structural motifs and associated electrical, optical and chemical properties. Porphyrin nanostructures like tubes, spheres, wires, rods and structures with more complex morphologies, present an opportunity for integration of these functional molecules into electronic and optoelectronic devices (Botti et al., 2002). DNA provides basic building blocks for constructing functionalized nanostructures with four major features: *molecular recognition, self-assembly, programmability, and predictable nanoscale structure* (Braun et al., 1998; Mrunal Khaderbad et al., 2008).
In this chapter, we review the formation of 5-(4-Hydroxyphenyl)-10, 15, 20-tri (p-tolyl) Zn (II) porphyrin SAM on silicon dioxide (SiO_2) and hydrogen silesquioxane (HSQ) and its application as Cu diffusion barriers for ULSI metallization and in microfluidics. We discuss the use of meso-pyridyl pophyrin SAM on gold and its interesting properties for molecular electronic applications. We present some of the approaches and current research status in DNA templated nanowire fabrication and the potential use of DNAs in transistor realization at the molecular level.

2. Challenges in Nanoscale Technologies

For the past twenty years, CMOS technology is the leading technology used in microprocessors, static/dynamic memory, microcontrollers and other logic circuits. As we approach the scaling limits, new ways of building logic and memory structures will have to be considered. One such approach is using bottom-up methods with the conventional fabrication metods.
In the sub-50 nm CMOS processes, multilevel interconnects with copper (Cu) are currently being used to minimize interconnect delay, coupling, and power dissipation. Copper is one of the best known electrical conductors having a very low resistivity and high electromigration resistance. However, copper diffusion through the dielectric is a serious reliability issue (Dallaporta, 1990). Diffusion of copper in Si, SiO_2 and low-k inter layer dielectrics (ILDs) at higher temperatures increases the device leakage currents, thus degrading the device performance and life time (Shacham-Diamond et al., 1991). Because of these issues Cu needs a suitable drift/diffusion barrier whose thickness is scaleable along with the other technology parameters.
Next section discusses the application of hydroxyl-phenyl porphyrin SAMs in nano-scale technologies. First, we discuss, how 5-(4-Hydroxyphenyl)-10, 15, 20-tri (p-tolyl) Zn(II) porphyrin SAM is useful in preventing copper diffusion in SiO_2 and HSQ. Surface modification of substrates like SiO_2 using porphyrin SAMs has a tremendous value in nanofluidics. The porphyrins with meso-pyridyl groups are useful to prepare water-soluble porphyrins, which can bind with biological molecules such as DNA and other proteins. We also review the formation of and characterization of SAM of meso-pyridyl porphyrin having a thiol linker [such as 5-(4-(2- (4-(S-Acetylthiomethyl)phenyl)ethynyl)phenyl)- 10,15,20-tris (4-pyridyl) porphyrin] on gold in view of its application in molecular electronics. In biology, it is well known that the bottom-up approach can be used to synthesize complex and sophisticated DNA networks. Section 4 describes how the DNA is used as a template for molecular wiring and in quantum computing.

3. Porphyrin Self-Assembled Monolayers for Nanoelectronic Application

3.1 Self-Assembled Monolayers

SAMs form by chemisorption and self-organization of functionalized organic/bio molecules on to the surfaces of different substrates. The molecules that form SAMs are called *surfactants*. Surfactants comprise of a *head-group* which binds to the substrate, an *end-group* that constitutes the outer surface of the film, and a *backbone* that connects head-group and end-group and affects the intermolecular separation and molecular orientation. A SAM has an interesting property of inherent surface reconstruction, a phenomenon that often causes rapid randomization of surface atoms. Since self-assembly is a process in which it reaches a thermodynamically stable system, it tends to eliminate faulty structures during the assembling process. There are numerous surfaces that have been employed in SAM construction including conductors such as copper, silver and palladium; semiconductors such as silicon, gallium arsenide and cadmium sulfide; and insulators such as silicon oxide (Abraham Ulman, 1996).

An important monolayer system for its many technological applications is the structures of silanes on hydroxylated surfaces. For monolayer self-assembly; alkylchlorosilanes, alkylalkoxysilanes, and alkylaminosilanes need hydroxylated surfaces as substrates. Self-assembly of these molecules takes place through the formation of polysiloxane, which is connected to surface silanol groups (-SiOH) via Si-O-Si bonds. Fig. 1a shows the process for the formation of SAM on silicon dioxide using silanes, organometallics and alcohols (Aswal et al., 2006).

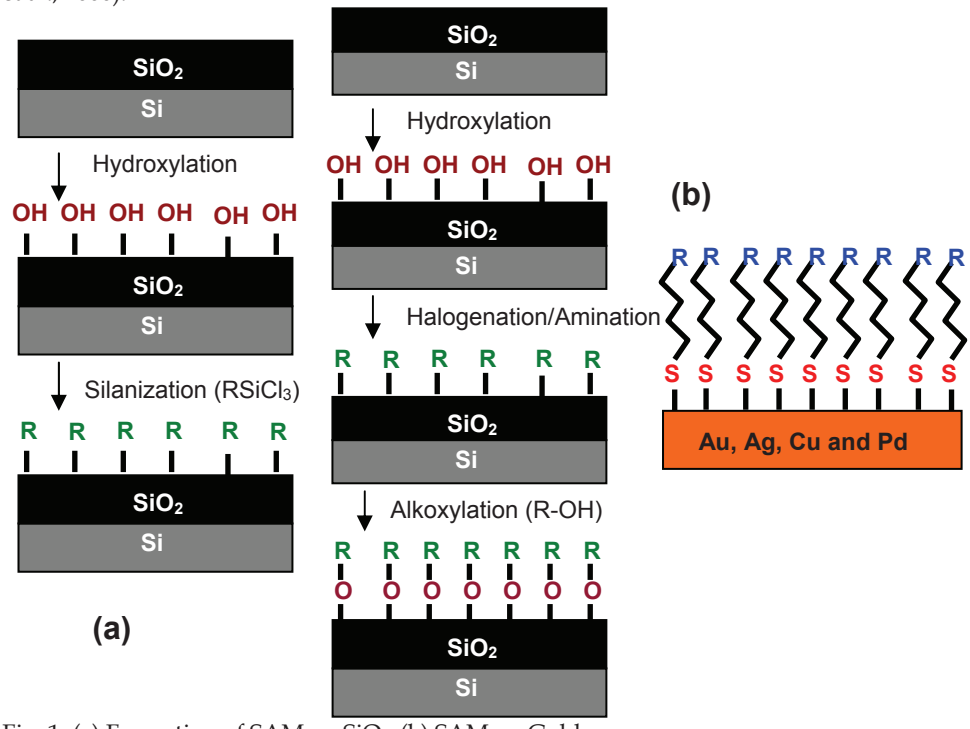


Fig. 1. (a) Formation of SAM on SiO_2 (b) SAM on Gold

Substrates on which these monolayers have been prepared using the above process include silicon oxide, aluminum oxide, quartz, glass, mica, zinc selenide, germanium oxide, and gold.

The widely studied systems of SAMs are alkanethiolates $CH_3(CH_2)_nS^-$ on gold. Well defined arrangement of organic surface phases was first observed in 1983 by the immersion of a gold substrate in the dialkylsulfide solution. Besides gold, thiols bind very strongly to silver, palladium and copper (Abraham Ulman, 1998). Fig. 1b. shows the formation of SAM on gold.

It is well known that there are a number of head groups that bind to various dielectrics, metals and semiconductors. Table. 1. shows the ligands that bind to the various substrates (Christopher Love et al., 2005). The binding mechanism is given in the third column.

Substrate	**Ligand**	**Binding**
M (M = Au, Ag, Cu, Pd)	RSH, ArSH (thiols)	RS-M
Au	RSSR' (disulfide) RSR'(sulfide) RSO_2H R_3P	RS-Au RS-Au RSO_2-Au R_3P-Au
Pt	RNC	RNC-Pt
GaAs InP	RSH	RS-GaAs RS-GaAs
SiO_2, glass	$RSiCl_3$,$RSi(OR')_3$	Siloxane
Si/Si-H	$(RCOO)_2$ $RCH=CH_2$	R-Si RCH_2 CH_2Si
Si/Si-Cl	RLi, RMgX	R-Si
Metal Oxides	RCOOH RCONHOH	$RCOO^-$---MO_n RCONHOH--- MO_n
ZrO_2 In_2O_3/SnO_2	RPO_3H_2	RPO_3^{2-}---Zr^4 RPO_3^{2-}---M^{n+}

Table. 1. Substrates and ligands that form SAMs

In the following sections, importance of porphyrins and porphyrin derivatives, formation of porphyrin SAMs on various substrates and structural/material characterization of these SAMs are discussed.

3.2 Porphyrins

Porphyrins are nitrogen containing compounds derived from the tetrapyrrole porphin molecule. The basic structure of the porphyrin macrocycle consists of four pyrrolic subunits linked by four methine bridges.

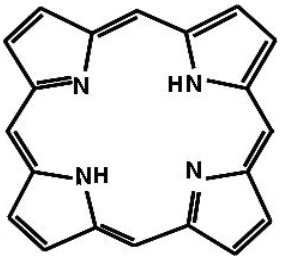


Fig. 2. Porphin molecule

Figure 2 shows the structure of a porphin molecule. Porphyrins bind metals to form complexes, usually with a charge of 2^+ or 3^+, which resides in the central N_4 cavity formed by the loss of two protons. These metallo-porphyrins play an important role in biology. Chlorophyll is a Mg-porphyrin and the Fe(II) porphyrin complex is the part of hemoglobins and myoglobins, which are responsible for oxygen transport and storage in living tissues (David Dolphin, 1978).

A wide variety of porphyrin arrays of ever increasing size have been constructed by the traditional methodology of covalently linking porphyrins. The importance of porphyrins, porphyrazines, and phthalocyanines in various fields derives from their photophysical and electrochemical properties, stability, and highly predictable and robust structure. The research developments in the formation and characterization of functional, porphyrinic materials and devices are self-assembled porphyrin arrays into phototransistors and photonic devices; SAMs of porphyrin molecules for sensors and nanotechnology applications; metalloporphyrins as stochastic sensors; and covalently bound arrays of pyrrole units as photonic materials.

Potential use of these multi-functional nanostructures is in nanoelectronics and nanophotonics, as previous research on porphyrin crystals and aggregates have demonstrated conducting and semiconducting properties (Kenneth S. Suslick et al., 2000). Photoconductivity and non-linear optical properties with visible light have also been demonstrated in porphyrinic materials (Schwab et al., 2004). With porphyrins forming nanostructures such as tubes, spheres, wires, rods, and structures consisting of complex morphologies, an opportunity presents itself for integration of these functional porphyrin-based nanostructures into electronic and optoelectronic devices (Anthony et al., 2000). Porphyins and porphyrin derivatives are used in numerous applications like in pressure-sensitive paints (Grenoble et al., 2005), organic field effect transistors (OFETs) (Berliocchi. M et al., 2004), bio-sensors (Papkovsky et al., 2000), explosive detectors (Shengyang Tao et al., 2007, Dudhe et al., 2008, 2009) and in TFTs. Wende et.al, 2007, demonstrated the substrate-induced magnetic ordering and switching of iron porphyrin molecules. Above studies open up an avenue for spin-dependent molecular electronics. Chao Li et al., investigated the potential applications of porphyrins in memory storage devices. Masahiro Kawao et al. prepared conducting oligo-diethynyl-porphyrin wires with length exceeding 600 nm. Their

applications include nonlinear optics, catalysts, sensors, actuators, memory devices, organic FETs and therapeutics.

3.3 Hydroxy-Phenyl Porphyrin SAM formation on SiO_2 and HSQ

The hydroxy-phenyl porphyrin SAM on SiO_2 was prepared following a chemisorption technique (Onclin, 2005). The silicon dioxide substrate used to prepare porphyrin SAM was prepared by thermally growing SiO_2 on a RCA (Radio Corporation of America) cleaned p-type (100) Si wafer. The substrate was then cut into the required size and cleaned by ultrasonic rinsing in Isopropyl alcohol. Then the SiO_2 substrate was dipped in sulphochromic acid (mixture of 1 ml DI water, 0.5 mg $K_2Cr_2O_7$ and 100 ml H_2SO_4) for 10 minutes. This removes any native carbon impurities and creates OH groups on the SiO_2 surface by opening siloxane bonds and forming silanol groups (SiOH) on the surface. After the sulphochromic acid treatment, the SiO_2 substrate was rinsed in DI water and dried under Ar gas flow.

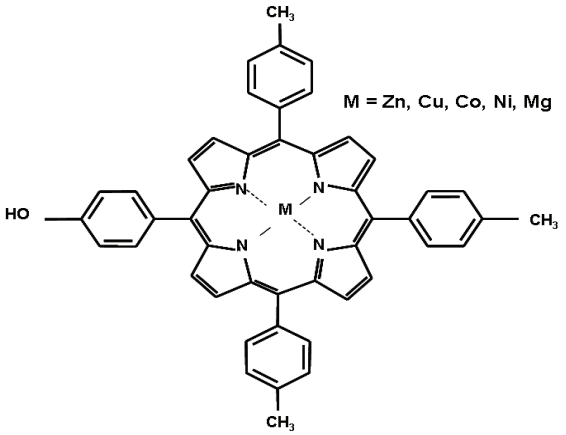


Fig. 3. 5-(4-Hydroxyphenyl)-10, 15, 20-tri (p-tolyl) porphyrin. This molecule 5-(4-Hydroxyphenyl)-10, 15, 20-tri (p-tolyl) porphyrin is used to prepare a self-assembled monolayer on silicon dioxide surface and on HSQ.

This was followed by vacuum heating (vacuum pressure ~ 10^{-2}mbar) the substrate at 110^0 C for 1 hour. 4 mg of porphyrin was dissolved in 20 ml of toluene to prepare 10^{-4} M solution. The SiO_2 substrate was immersed in the above solution for 30 minutes. During immersion, the head groups of the porphyrin molecule chemically bond with the silanol groups on SiO_2 surface forming a self-assembled monolayer (Fig. 4).

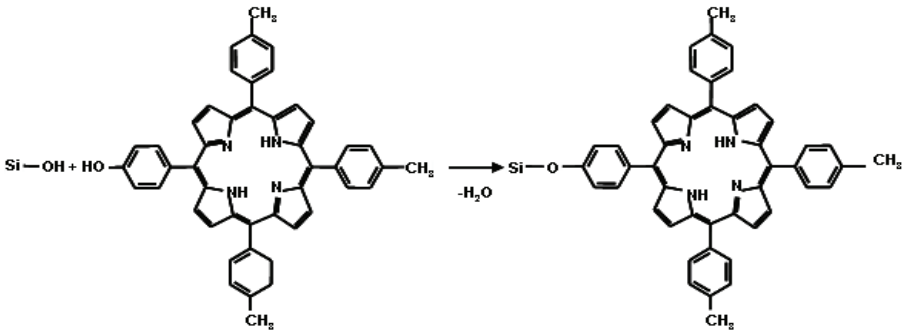


Fig. 4. Formation of hydroxyl phenyl porphyrin SAM on SiO_2

This was followed by rinsing the substrate with toluene and drying under Ar gas flow. Finally the substrate was heated at 120^0 C for 45 minutes to remove the water molecules and then stored for characterization. Above method can be used to prepare SAM on HSQ.
The formation of SAMs is assessed using various methods, which include surface probe microscopies (such as AFM and STM), Fourier transform Infrared spectroscopy (FTIR), UV-vis spectroscopy, tunneling electron microscopy (TEM), sum frequency generation (SFG), helium diffraction, electron diffraction, contact angle, ellipsometry, and NEXAFS. Ultraviolet-visible spectropscopy (uv = 200-400 nm, visible = 400-800 nm) corresponds to electronic excitations between the energy levels that correspond to the structure and orbitals of the molecular systems. The following electronic transitions can occur by the absorption of ultraviolet and visible light: σ to σ*, n to σ*, n to π* and π to π*.

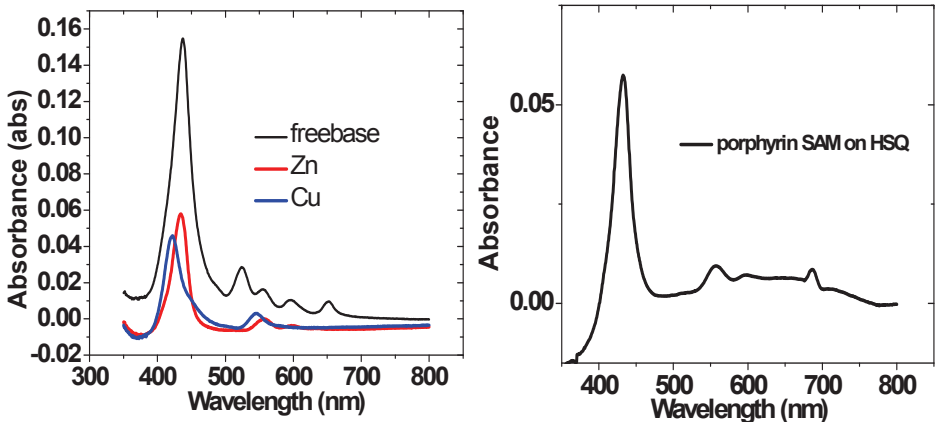


Fig. 5. (a) Ground state UV absorption spectra of hydroxy-phenyl porphyrin SAM on SiO_2 (b) UV-Visible spectrum of Zn-TPP-OH SAM on HSQ

In porphyrin visible absorption spectra, the highly conjugated aromatic macrocycle shows an intense absorption (extinction coefficient > 200,000) in the neighborhood of 400 nm; this absorption maximum is referred to as the "Soret Band". Visible spectra of porphyrins also show several weaker absorptions (Q Bands) at longer wavelengths (450 to 700 nm).

Variations of the peripheral substituents on the porphyrin ring cause minor changes to the intensity and wavelength of the absorption features. The protonation of two of the inner nitrogen atoms or the insertion/removal of metal atoms into the macrocycle strongly change the visible absorption spectrum. Fig 5 (a) illustrates the ground state UV absorption spectra of the hydroxy-phenyl porphyrin SAM on SiO_2. Spectra were recorded using Perkin-Elmer Lambda 35 spectrophotometer at room temperature in the wavelength range of 350 to 800 nm). For porphyrin in toluene, the Soret band was observed at 418 nm. For the porphyrin on SiO_2 substrate, relatively low absorbance was observed. The Soret band of porphyrin on SiO_2 was broadened and red shifted to 426 nm compared to the Soret band of porphyrin in toluene. This red shift indicates that the porphyrin molecules are arranged in a side-by-side orientation in the molecular self-assembly. Fig. 5 (b) illustrates the ground state UV absorption spectra of the hydroxy-phenyl porphyrin SAM on HSQ. Soret band shift and broadening of peak in the spectrum confirm ZnTPP-OH SAM formation on HSQ.

3.3.1 Preparation of meso-pyridyl Porphyrin SAM on Gold

The preparation of meso-pyridyl porphyrin Self-Assembled Monolayer (SAM) on gold surface was explained in the work done by Amarchand Sathyapalan et al., 2005. In this work, a meso-pyridyl porphyrin having a thiol linker such as 5-(4-(2-(4-(S-acetylthiomethyl) phenyl) ethynyl) phenyl) porphyrin shown in Fig. 6 was synthesized and used for the formation of self-assembled monolayers on a gold substrate. The meso-pyridyl porphyrin SAM was prepared by the base-promoted method described elsewhere (Chen et al., 2000). The formation of self-assembled monolayers on gold surface is a spontaneous process. The specificity of the gold-sulfur interaction has provided an extremely convenient route to the formation of chemisorbed molecular films. The procedure of SAM formation that is followed is simple, and flexible enough to change it to suit different compounds. The SAMs formed by this method are very stable due to the nature of the adsorption which is via a chemical bond.

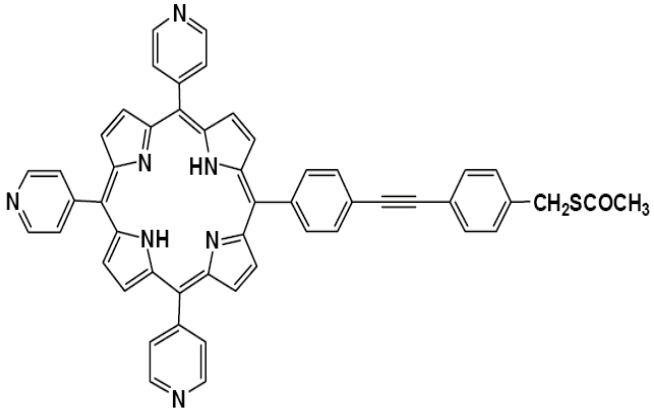


Fig. 6 Chemical structure of 5-(4-(2-(4-(*S*-acetylthiomethyl) phenyl) ethynyl) phenyl)-10, 15, 20-tris (4-pyridyl) porphyrin.

Fourier transform infrared (FTIR) spectroscopy gives the molecular orientation and ordering in a self-assembled monolayer. For this study, grazing incidence (80° to the surface normal)

reflection absorption FTIR spectroscopy was used. Fig. 7 shows the FTIR spectrum of meso-pyridyl porphyrin SAM on gold.

A broad and strong band at 3433 cm^{-1} arises mainly from the O–H stretching mode of water molecules bound to the cationic porphyrin. A band due to the stretching mode of the N–H group of the porphyrin is barely resolved at about 3315 cm^{-1} as a shoulder on the low wavenumber side of the broad band at 3433 cm^{-1}. Besides this high wavenumber band, two bands at 970 and 723 cm^{-1} due to N–H in-plane and out-of-plane bending modes, respectively, also appear in the low wavenumber region. A strong band at 2924 cm^{-1} and a medium band at 2357 cm^{-1} are ascribed to CH_2 antisymmetric and symmetric stretching modes, respectively. A strong band located at 1637 cm^{-1} is ascribed to the C==N stretching mode. Some weak and medium bands due to the vibrational modes of the porphyrin ring appear in the region 1600–680 cm^{-1} of the solid spectrum.

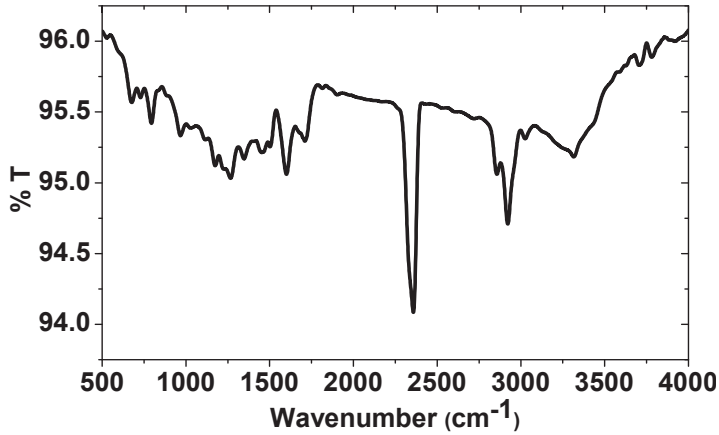


Fig. 7. FTIR spectrum of meso-pyridyl porphyrin SAM on gold.

The surface coverage of a monolayer is examined by measuring the surface morphology using the atomic force microscopy (AFM) (Fig. 8). The 5-nm resolution images demonstrated the formation of SAMs meso-thiol porphyrins on gold surfaces. Basic hexagonal $\sqrt{3} \times \sqrt{3}$ R 30^0 arrangement with highly ordered monolayers was observed.

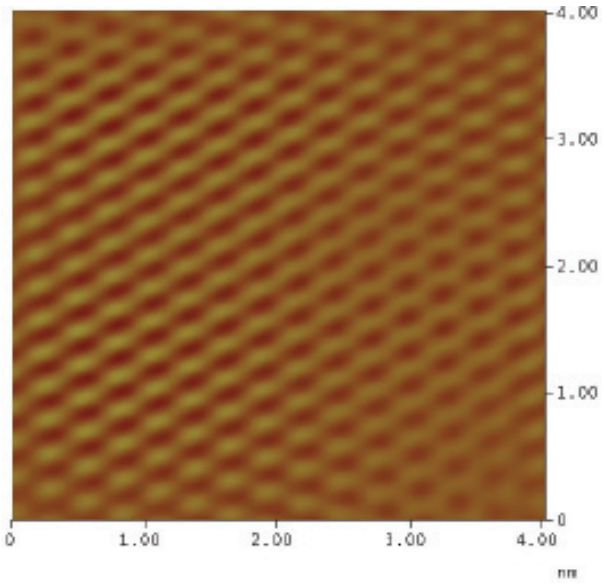


Fig. 8. AFM image of porphyrin SAM on gold

3.4. Porphyrin SAMs as Cu Diffusion Barriers in ULSI metallization

3.4.1 Cu Diffusion Barriers in ULSI metallization

For the sub-nm CMOS technology, ultrathin diffusion barriers (1-3nm) are needed for the copper interconnects to suppress the diffusion of Cu into silicon and into inter layer dielectrics (ILDs) (Awaya et al., 1996). In this regard, refractory metal binary and ternary nitrides have been investigated for their copper diffusion barrier properties. Reactive sputtering is used to deposit ternary nitride alloys, such as W-Ge-N, Ta–Si–N, W–Si–N, W–B–N, and Ta–W–N. Since the resistance of interconnects is affected by the thickness of the barrier layer (Koike et. al., 2005), thinner barrier layers (less than 10 nm) with TiSiN and WN have been deposited by chemical vapor deposition (CVD). Many research groups tried to address the issue of barrier layer thickness by depositing ultra thin TiN or WNC layers by atomic layer deposition (ALD, ALCVD) (Eleres et. al., 2002). ALD is known to be effective in depositing conformal and thin barrier layers which are in amorphous or polycrystalline phase (Jill S. Becker and Roy G. Gordon, 2003). The major failure mechanism of diffusion barriers is the grain boundary diffusion and the barrier layers deposited by the above discussed methods tend to be ineffective due to their high defect densities and fast diffusion paths such as nano-pipes or grain boundaries (Rosenberg et. al., 2000). Thus, deposition of continuous and uniform ultrathin layers is difficult by conventional physical and chemical vapor deposition methods.

An alternative and viable approach to this problem is to use self-assembled monolayers (SAMs) of molecules, which can withstand back-end of line (BEOL) processing conditions and meet the diffusion barrier requirements. These methods are also extremely cost-effective compared to other deposition techniques as they do not require ultrahigh vacuum

(UHV) or other specialized equipment. Krishnamoorthy et. al., in 2001 have reported the use of monolayers of self-assembled amphiphilic organosilanes, and polyelectrolytes, as diffusion barriers at the Cu–SiO_2 interface. Molecules with different side chains and different terminal groups have been studied to immobilize Cu through strong local interfacial bonding and to improve interfacial adhesion (Ganesan et. al., 2004). In the following section, the potential application of hydroxy-phenyl Zn(II) porphyrin SAM as a Cu diffusion barrier for ULSI metallization has been explained.

3.4.2 Porphyrin SAMs as Cu Diffusion Barriers in ULSI metallization

Using the porphyrin SAM as a diffusion barrier has multiple advantages like excellent thickness control (molecular monolayer), conformal layer formation and uniform coverage. Presence of aromatic rings in porphyrins sterically hinders the Cu ion diffusion through the SAM layer. Also, the steric effect may get enhanced by the presence of Zn(II) ions in the center of the porphyrin macrocycle. Zn in hydroxy-phenyl Zn(II) porphyrin molecule prevents Cu ion diffusion into SiO_2 due to its electronegativity and strong binding to the porphyrin molecule. Adding to the above effects, the pyrrole subunits within the porphyrin macrocycle containing nitrogen may play a key role in the prevention of Cu diffusion, because Cu-N and Cu-P bonds have been known for their involvement in diffusion barrier mechanisms (Mrunal Khaderbad et al., 2008; Urmimala Roy et al., 2009).

In the following sections, the results of bias-temperature-stress (BTS) CV analysis on Cu/SiO_2/p-Si, Cu/SAM/SiO_2/p-Si, Cu/HSQ/p-Si and Cu/SAM/HSQ/p-Si have been presented. It shows that the hydroxy-phenyl Zn(II) porphyrin SAM is effective in preventing the diffusion of mobile Cu ions into SiO_2 as well as HSQ.

Bias Stress Temperature effects on Cu/SiO_2/p-Si and Cu/SAM/SiO_2/p-Si MOS capacitors (MOSCAPs):

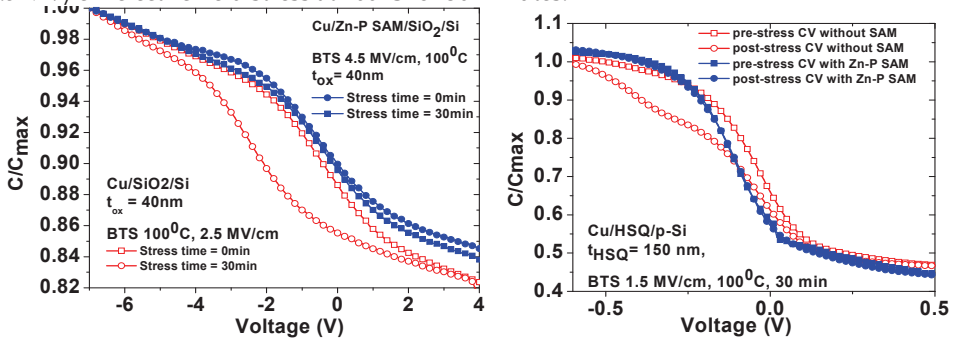
Copper can diffuse through SiO_2, Si or ILDs under high bias-temperature stress (BTS) conditions. Previous research showed that in atmospheric nitrogen ambient, copper drift in oxide will occur at 350^0C or higher. In the presence of an electric field, at temperatures as low as 100^0C, positive Cu ions (Cu^+ or Cu^{2+}) drift rapidly through inter-layer dielectrics (ILD) (Cluzel et al., 2002; Loke et al., 1998). The copper ion diffusion under BTS conditions results in the shift of MOSCAP C-V (capacitance-voltage) characteristics. This shift can be calculated using the following equation:

$$V_{FB} = \Phi_{ms} - 1/C_{ox}(Q_f + Q_m\gamma_m + Q_{it}) \quad (1)$$

where *Cox* is the oxide capacitance; Φ_{ms} is the difference in the work functions of the metal and semiconductor; Q_f, Q_m, Q_{it} are fixed, mobile and interface-trap charges respectively and γ_m is the centroid of the mobile charge.

BTS studies were carried out on Cu/SiO_2(HSQ)/p-Si and Cu/SAM/SiO_2(HSQ)/p-Si MOSCAP test structures to characterize the Cu diffusion. SAM formation in these MOSCAP structures was done using the recipe explained in section 3.3. Fig. 9 (a) describes the pre-stress and post-stress C-V (normalized with respect to C_{max}) characteristics for the Cu MOS capacitors (t_{ox} = 40nm) with and without the porphyrin SAM, obtained at 50 kHz frequency, using Agilent 4284-A precision LCR meter. Fig. 9 (b) shows the C-V characteristics for the Cu MIS capacitors (t_{HSQ} = 150nm) with and without the porphyrin SAM, obtained at 1MHz frequency. The Cu/SiO_2/p-Si MOS capacitor was subjected to 2.5 MV/cm electric field

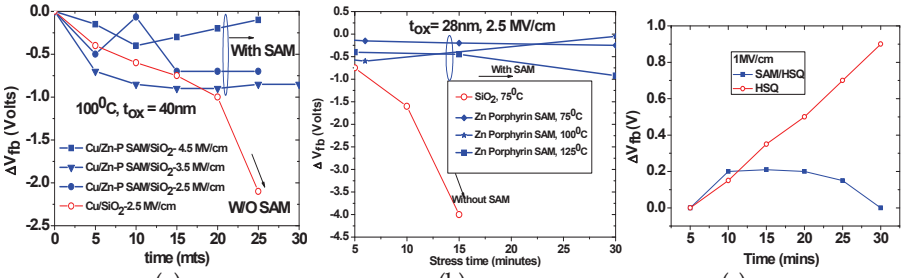
stress at 100^{0}C for 30 minutes, where as, the Cu/HSQ/p-Si MIS capacitor was subjected to 1.5 MV/cm electric field stress at 100^{0}C for 30 minutes.



(a) (b)

Fig. 9. (a) The pre-stress and post-stress C-V characteristics for the Cu MOS capacitors with and without the porphyrin SAM (b) Pre-stress and post-stress HFCV plots of Cu/SAM/HSQ/p-Si MIS capacitor after BTS of 30 min (t_{HSQ} = 150nm)

Comparing C-V plots in figs. 9(a) and 9 (b), it is clear that C-V curve shift is less in the case of Cu/SAM/SiO_2/Si MOS structure compared to that of MOS structure without SAM. Fig. 10 shows ΔV_{fb} (V_{fb} shift) versus stress time for MOS (MIS) structures with and without SAM. The Cu MOS capacitors with SAM were subjected to higher fields (4.5 MV/cm) at the same temperature (100^{0}C) and stress time (30 mins), and still show superior properties (Fig. 10a).



(a) (b) (c)

Fig. 10. Flatband Voltage (ΔVfb) versus stress time. (a) Cu/SiO_2/Si MOSCAPs were stressed at 100^{0}C and 2.5 MV/cm; Cu/SAM/SiO_2/Si capacitors were subjected to higher stress fields stress at 100^{0}C. (b) MOSCAPs without SAM were subjected to a stress at different temperatures. (c) Flatband Voltage (ΔV_{fb}) versus stress time curve of Cu/HSQ/Si and Cu/SAM/HSQ/Si MIS capacitors.

3.5 Hydroxy-Phenyl Porphyrin SAMs for micro/nanofluidic applications

Controlling liquid flow within networks of nanochannels is crucial for the design and fabrication of nanofluidic devices. The hydrophobic and hydrophilic characteristics of the surface have been exploited to handle and control liquid flows in the above systems. Hydrophobic and hydrophilic regions inside microchannels can be patterned through surface modification or microcontact printing of SAMs (Kaushik Nayak et al., 2007).

Porphyrin self-assembled monolayer chemistry can be used to modify surface wetting properties of a variety of materials.
The contact angle measurements are known to be effective in characterizing the surface wetting properties as the contact angle depends on the free energies of the liquid and the surface. Fig. 11 shows the water contact angle measurements of 50 μL of sessile DI water (Resistivity, ρ ~ 18.2 MΩ-cm) drop on SiO_2 surface and on hydroxy-phenyl porphyrin SAM.

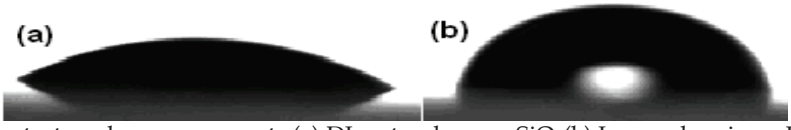


Fig. 11. Contact angle measurements (a) DI water drop on SiO_2(b) Image showing a DI water drop on hydroxy-phenyl porphyrin SAM on SiO_2

The water drop on SiO_2 surface exhibited contact angles of 30 ± 2^0 showing hydrophilic nature of the surface (Fig. 11a). In the case of hydroxy-phenyl porphyrin SAM on SiO_2, the contact angles were found to increase up to 78 ± 3^0, as shown in Fig. 11 (b). Above property shows that the patterning of hydrophilic substrates with porphyrin SAMs has tremendous applications in micro/nanofluidics and that the SAMs are effective in modifying the surface properties.

3.6 Meso-pyridil Porphyrin SAM on Gold for molecular electronics applications

Molecular electronics, electronics that uses molecules as building blocks for making electronic devices such as transistors, has the potential to extend Moore's Law beyond the foreseen limits of existing silicon technology. It uses simple chemistry to promote molecules with electronic or optical functionality and their self assembly to build active electronic devices. CNTs, polyphenylenes, porphyrins and DNA strands are some of the molecules that are being actively researched upon for the above purposes. Many techniques have been proposed to probe the conductance of single molecules, either using a fixed gap between two electrodes (fabricated by e-beam lithography or as a mechanically controllable break junction or break-junction using electromigration) or using conductive atomic force microscopy (AFM) and scanning tunneling microscopy (STM) techniques (Chen et al., 2007; Akkerman et al., 2008). A meso-pyridil SAM on gold with a thiol linker can be formed as explained in section 3.3.1. Amarchand Satyapalan et at. in 2005 reported its structural and electrical characteristics in view of their applications in molecular electronics. It was observed that the electronic characteristics measured by scanning tunneling spectroscopy (STS) showed that this monolayer has a non-linear IV behavior, similar to a semiconductor junction with a barrier potential (Reed et al., 1997). This barrier behavior can be explained with the help of alignment of molecular orbital levels (HOMO/LUMO) with that of metal's Fermi energy level. Depending on the substrate/bias voltages V, the molecule conducts strongly when

$$eV > E_L - E_f \quad \text{(positive substrate bias)} \quad (2)$$

and

$$-eV < E_f - E_H \quad \text{(negative substrate bias)} \quad (3)$$

where E_L and E_H are LUMO and HOMO energy levels of the molecule respectively and E_f is the metal Fermi level. The threshold for conduction for such type of a molecular junction is given by (Datta et al., 1997):

$$eV > \min [(E_f - E_H/\eta), (E_L - E_f/1-\eta)] \quad \text{(positive substrate bias) (4)}$$

$$-eV < \min [(E_f - E_H/1-\eta), (E_L - E_f//\eta)] \quad \text{(negative substrate bias)(5)}$$

where η is the ratio of distance of substrate from the centre of the molecule to the distance of substrate from the tip.

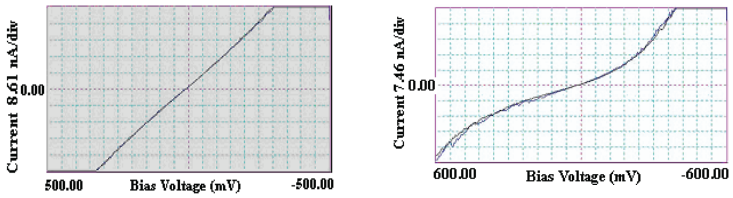


Fig. 12. (a) I-V characteristics of bare gold substrate (b) I-V characteristics of the SAM on gold

Figure 12 shows the STS plots of the gold substrate and of the 5-(4-(2-(4-(SAcetylthiomethyl) phenyl)ethynyl)phenyl)-10,15,20-tris(4-pyridyl) porphyrin SAM in ambient laboratory conditions. It is evident that the tunneling current is almost negligible before a certain cut-in voltage and rises sharply afterwards in case of porphyrin SAM on gold.

4. DNA for Nanoelectronics' Applications

DNA is fast becoming a material of choice for the bottom-up approach in the fabrication of nanometer-scale electronic devices. Eley and Spivey in 1961 first predicted the tentative use of DNA as molecular wires (Eley et al., 1961). Easy availability (second-most abundant class of biomolecules, next to proteins), self-assembly property and its ability to be manipulated *in vitro* has put DNA into one of the top-priority alternatives for the bottom-up approach. If biomolecules are to be chosen for the bottom-up self-assembly oriented approach, DNA should be given higher priority as compared to proteins because proteins are not robust to extreme physical and chemical conditions. Inherent programmability of DNA through variation of its base sequence is another attractive feature of the DNA. In addition to the above points, DNA also offers the possibility of *in vitro* precise manipulation which makes possible interesting device applications for nanoelectronics applications.

4.1 Relevant properties of DNA

DNA is a duplex (double-stranded) polymeric molecule. Each strand is itself a polymer, consisting of nitrogenous bases namely purines (adenine and guanine) and pyrimidines (thymine, cytosine and uracil). The only exception that contains uracil molecule in the DNA is a bacteriophage *PBS1* (Savva, 1995). Fig. 13 shows the chemical structure of the five bases.

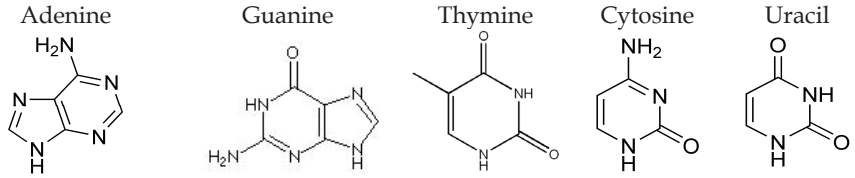


Fig. 13. Different bases in a DNA molecule (Neidle, 2008)

These five bases form hydrogen-bonds between each other so as to stabilize the duplex structure of the molecule. Adenine forms double H-bonds with thymine or uracil whereas guanine forms a triple H-bond with cytosine (as shown in fig. 14).

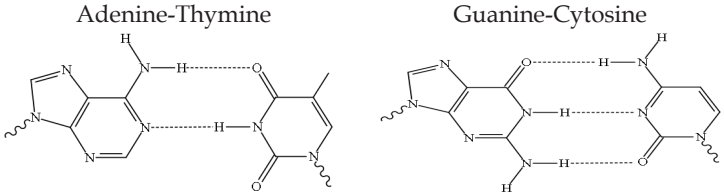


Fig. 14. Hydrogen bond formation between different purine-pyrimidine base pais (Neidle, 2008) The DNA molecule is negatively-*supercoiled* like a rope, twisted opposite to the direction of helix thus pushing the bases away from each other. This structural property helps in easy unwinding of the double-helix during the process of DNA replication. In the ambient physical conditions, three types of DNA conformations exist: A-DNA, B-DNA and Z-DNA. Out of these, B-DNA is the most common form of DNA available (hence most widely studied scientifically) and Z-DNA is the rarest form of DNA (forms under extreme stringent conditions). Different conformations are formed due to the change in the sequence of DNA, the amount and direction of supercoiling, chemical modifications of the bases and the solution conditions (like the concentration of heavy metal ions or polyamines). There are several geometrical attributes of different forms of DNA which play major roles in the formation of electronic templates. These attributes are listed in table-2 as follows:

Geometry attribute	**A-form**	**B-form**	**Z-form**
Helix sense	right-handed	right-handed	left-handed
Repeating unit	1 bp	1 bp	2 bp
Rotation/bp	33.6°	35.9°	60°/2
Mean bp/turn	11	10.5	12
Inclination of bp to axis	+19°	−1.2°	−9°
Rise/bp along axis	2.4 Å (0.26 nm	3.4 Å (0.34 nm)	3.7 Å (0.37 nm)
Rise/turn of helix	24.6 Å (2.46 nm)	33.2 Å (3.32 nm)	45.6 Å (4.56 nm)
Mean propeller twist	+18°	+16°	0°
Diameter	26 Å (2.6 nm)	20 Å (2.0 nm)	18 Å (1.8 nm)

Table 2. Geometrical attributes of A-, B- and Z-forms of DNA (Neidle, 2008; Ghosh et al., 2003)

Another important physical feature that needs a mention is the presence of major and minor grooves in the DNA molecule (fig. 15).

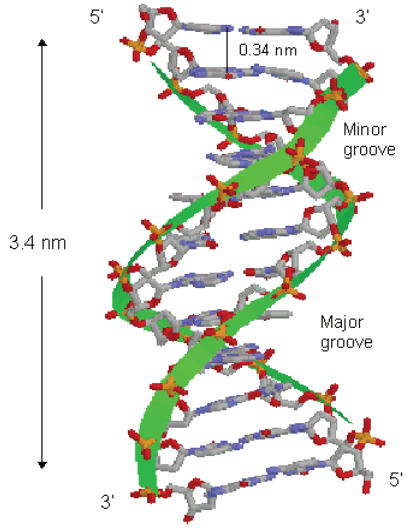


Fig. 15. Schematic of DNA molecule showing major and minor grooves (Molecular Biology Web Book, 2009)

The width of a major groove is ~ 2 Å and a minor groove is ~ 1.2 Å. Different template materials like porphyrins and others that bind to DNA usually prefer major grooves for this purpose. This is because of the greater exposure of bases through the respective groove (Molecular Biology Web Book, 2009). DNA being a biological molecule could be manipulated to the needs of electronics by biological means. Selective manipulation of the molecule is possible by the use of different enzymes (Molecular Biology Web Book, 2009) as listed below:

- Nucleases: Enzymes that degrade DNA by hydrolysis of the phosphodiester bonds in the strands.
    - Exonucleases hydrolyze phosphodiester bonds from the ends of DNA strands.
    - Endonucleases hydrolyze phosphodiester bonds within the DNA. Most frequently used nucleases are restriction endonucleases, which cleave DNA strands at a specific sequence location.
- Ligases: Enzymes that close the nicks; recover the broken phosphodiester bonds in a double-stranded (ds) DNA
- Polymerases: Synthesize polynucleotide chains from nucleoside triphosphates. Sequence of their products match with the pre-determined polynucleotide chains known as *templates*. All polymerases work in 5′→3′ direction (by adding nucleotides to 3′-OH group of preceding nucleotide)

After the manipulation process is done, DNA molecule could be separated on the basis of number of base-pairs in the molecule by *gel-electrophoresis*. The amplification / selective duplication of DNA molecules could be done using *polymerase-chain reaction* colloquially known as *PCR*. In this process, a RNA primer is required to start the reaction. Then DNA

polymerase takes over, adding individual dNTPs present in the reaction mixture according to the sequence as desired. This process repeated n times give 2^n copies of the desired DNA molecule (Saiki et al., 1985; Saiki et al., 1988).

4.2 Transfer of charge through DNA

The much debated topic in the scientific community is the conductivity of bare DNA molecule. Scientists have reported DNA as superconducting (Kasumov et al., 2001), metallic (Fink et al., 1999; Cai et al., 2000; Tran et al., 2000; Yoo et al., 2001), semiconducting (Porath et al., 2000; Rakitin et al., 2001) and insulating (Braun E. et al., 1998; dePablo et al., 2000; Storm et al., 2001) as well. There could be a plethora of reasons for the ambiguity in the observations made, resulting in different conductivity profiles of bare DNA. Different lengths of DNA molecules could account for different conductivity observations. Some other reasons might include different base-pair sequences, whether DNA is in form of ropes or single molecules during experiment, effects of ions or counterions in the environment, due to deformation of DNA molecules (e.g. stretching changes the stacking of p-orbitals between base pairs), presence of free-standing or surface-bound DNA molecules, variability in sample preparation, variability in measurement conditions (humidity, thermal fluctuations in solution) and variability in detection protocols .

With the observed conductivities of bare DNA, it is safe to infer that bare DNA is not useful for electronic applications. Hence to use DNA molecules in electronics, their conductivity need to be brought to the levels of semiconductors or metals.

4.3 Probable mechanisms of charge transfer in DNA

The most studied form of charge transfer in DNA is the hole transfer process. The studies on electron transfer process are still in their infancy (Wagenknecht, 2005). There are three mechanisms of hole transfer that have been studied widely *viz.* molecular-wire mechanism, polaron-like mechanism, superexchange mechanism and hopping mechanism. This transfer process could go over few microns which renders the molecule suitable for nanoscale electronic applications.

The hole transfer is an oxidative highest occupied molecular orbital (HOMO) controlled process. This implies that the Fermi level of the DNA molecule (as a bridge) is approximately at the level of the valence band of the metallic source and drain electrodes. In contrast, the electron transfer is a reductive lowest unoccupied molecular orbital (LUMO) controlled process. Its implication is that the Fermi level of the DNA molecule is approximately at the level of the conduction band of the metallic source and drain electrodes.

4.4 Making DNA useful for electronic applications

To change the conductivity levels of DNA, one could employ different chemical, physical and biological measures. In the following paragraphs, these measures will be discussed in some detail.

The measures that might classify into chemical ones might include incorporation of transition-metal (Zn, Ni or Co) ion into DNA and introduction of metal-ligating intercalating planar chelators. The incorporation of transition metal ions into DNA has been widely studied by Wood *et al* (Wood, 2002). They have shown that M-DNA is a complex

between transition-metal ions Zn^{2+}, Ni^{2+} and Co^{2+} and duplex DNA which forms at pH ~ 8.5 (Aich et al., 1999).

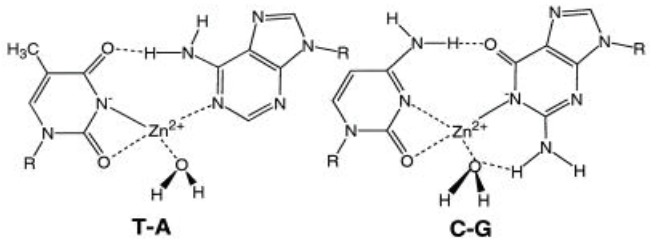


Fig. 16. Zn^{2+} ions forming hydrogen bonds with neighbouring bases in M-DNA (Wood et al., 2002)

The formation of the M-DNA depends on the nature and concentration of the metal ion as well as the pH and DNA concentration (when metal ions are present in excess amount). Figure 16 (Wood, 2002) shows how a metal ion trapped between a base pair in M-DNA would look like. Rakitin *et al* (Rakitin et al., 2001) demonstrated the change of semi-conductive behavior of λ-DNA in its B-form to metallic behavior by incorporating zinc in every base pair i.e. M-DNA.

The intercalating molecules that bind with the DNA molecules, preferably bind that portion of the DNA which has extrahelical structural elements. Typically, these molecules include metal-porphyrins and other coordination compounds like $[PtCl_4]^-$, cisplatin etc. Porphyrins usually bind to the major groove of DNA double helix to the extrahelical flip-out bases. The binding constant of porphyrins to DNA is ~ 10^6 M^{-1} (Jia et al., 2006). In the DNA-porphyrin complex, assuming the porphyrin molecules are placed at statistically equal intervals, the porphyrin-molecule-trapped metal ions serve as electron carriers and hence act as metallic nanowire.

Physical measures include templating DNA with metals (Mertig et al., 2002; Rajwade, 2007), templating DNA with metallic nanoparticles (Braun G. et al., 2005) / semiconductor quantum dots (proposed), templating DNA with conducting / semiconducting CNTs (Keren et al., 2003; Xin, 2006) and templating DNA with conducting polymers like polyaniline (PANI) (Ma et al., 2004; Rajwade, 2007; Khaderbad et al., 2008).

Covering DNA with metals require either continuously coating the molecule with metal or by the use of metallic nanoparticles. For continuous coating, the most common method used is electroless plating. It proceeds first with the formation of metal complexes or ions reacted with DNA to (specific) bonding sites on the molecule. Then, active sites are transformed to seeds by adding a reducing agent or by a reducing agent attached directly to the DNA. At last the seeds serve as catalysts for further reduction of metal in the third step and subsequently get enlarged (Fig. 17). This process is known as electroless plating. This method has been effectively used by E. Braun *et al* (E. Braun, 1998) for the synthesis of DNA-templated silver nanowires. The process is described later in the chapter.

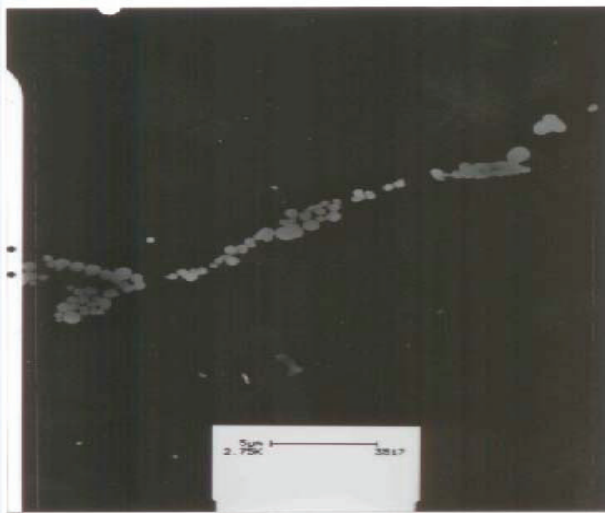


Fig. 17. TEM image of Metallization on DNA: TEM image (2750X) of metallized DNA strands in solution phase. The metal clusters formed along the length of the DNA are approximately 50 nm thick. The thickness of this wire formed was never observed to be uniform and varied in the range of 10-100 µm (Rajwade, 2007)

Polyaniline (PANI) deposition on DNA is of greater concern to researchers since PANI is a conducting polymer. The continuous conductive PANI nanowires are achieved at pH around 4.0. Ma *et* al have demonstrated that PANI can be reversibly doped and undoped on the basis of simple acid / base chemistry (by the addition of hydorchloric acid / ammonium hydroxide, respectively). The formation of continuous PANI nanowires over DNA has been acheived by Ma *et al* (Ma, 2004). Individual aniline molecules sit over DNA, which upon addition of horse radish peroxidase (HRP) and hydrogen peroxide (pH ~ 4) form continuous polyaniline nanowires.

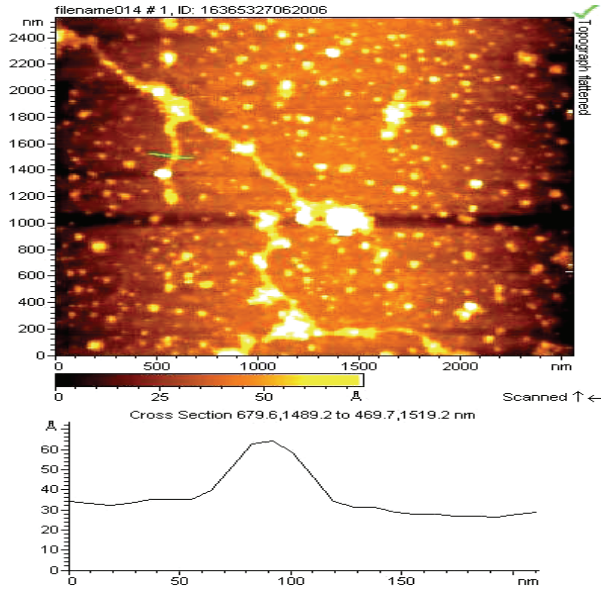


Fig. 18. AFM image of DNA/PANI complexes incubated in 100 µM aniline: 2.5 µm X 2.5 µm tapping mode image of DNA/PANI complex incubated in 100 µM aniline deposited on mica surface. The sectional analysis along the green line in the figure establishes the height of the DNA/PANI complex to be ~3.5 nm (Rajwade, 2007)

Biologically, DNA metallization has also been suggested using telomerase (Weizmann et al., 2004). The first case that has been reported indicates the use of gold nanoparticle labeled nucleic acid complementary to the telomere sequence. The second case reports of covalent attachment of ester-functionalized gold nanoparticles to polyamino-functionalized telomere templates.

4.5 Realization of DNA-based electronic devices

To make a working DNA-based electronic device, one should have a suitable DNA sample, a suitable substrate for immobilization of DNA. After the sample and substrate are available, one needs to take care of the alignment of DNA on the substrate. Next, the electrical characterization of DNA-based materials has to be done to determine whether the fabricated devices (interconnects and FETs) could be used for electronic purposes or not.

4.5.1 Suitable DNA samples and substrates

The DNA samples that are used for electronics related applications are λ-DNA, salmon-sperm DNA, calf-thymus DNA, poly(dG)–poly(dC) and poly(dA)–poly(dT) salts (Garcia, 2007). The substrate that one chooses must have a specific affinity towards the DNA molecules. For that it is necessary to modify the substrate surface or the DNA molecules or both. Garcia has proposed several substrates that could be used for the synthesis of DNA-templated nanodevices in his doctoral thesis. The suitable substrates include muscovite mica, highly oriented pyrolytic graphite (HOPG), single-crystal silicon, thermally grown

SiO_2, alkylsilane SAMs and films, glass, thin gold films, synthetic polymer films, anionic surfaces and gold (Garcia, 2007). The attachment of DNA to the surface is ensured by suitable surface modifications as well as the modifications of DNA molecule itself. DNA modifications include thiolation, biotinylation etc of the molecule; henceforth it could bind to gold surface and streptavidin-coated surface respectively (Seidel, 2003).

4.5.2 Alignment of DNA on substrates

The alignment of DNA molecules on the modified substrates can be done using either the molecular combing, hydrodynamic stretching or stretching by nitrogen blowing or spin coating or by PDMS stamps (Dewaratt, 2002; Garcia, 2007). In molecular combing, a minuscule drop of a dilute aqueous DNA solution is translated over a conditioned surface, followed by rinsing with purified water. DNA molecules at the interface facilitate surface-attractive interactions. As the air blows and the drop moves, the DNA molecules at the interface get aligned on the surface after leaving the solution phase sticking with the substrate.

In hydrodynamic stretching, a cylindrical chamber is filled with buffer solution containing DNA in it. The substrate surface is dipped vertically into the chamber. Then, the solution in the chamber is evacuated using a micro-liter pump. The DNA is stretched onto the surface by the lowering meniscus of the liquid. This process gives a better control over the stretching process of DNA on the substrate. In case of stretching by blowing nitrogen over the sample, random stretching takes place since the speed of stretching is in the order of few mm/s. There is less control over the stretching of the DNA over the modified surface in N_2 blowing, as compared to hydrodynamic stretching (Dewarrat, 2002).

Spin coating can be used effectively to align very small amounts of DNA sample on the substrate. However, this process suffers from a generic drawback of greater deposition of material at the edge of the substrate surface. Microfabricated PDMS stamps can also be utilized to align DNA on surfaces and this method provides a good control over DNA deposition (Garcia, 2007).

Diez *et al* have reported that DNA could also be aligned on a plain substrate by the aid of other biomolecules like tubulin (Diez et al., 2003). Biotin coated DNA bound to streptavidin coated tubulin was stretched over the glass surface using a motor protein, kinesin. As the tubulin-DNA system progresses aided by kinesin, DNA gets attached to the glass surface by either pH dependent mechanism or due to the fact that glass has been coated with streptavidin.

4.5.3 Electrical characterization of DNA and DNA-templated materials

To electrically characterize any molecule, it needs to be probed with the electrical contacts at its two ends to make contact to a single-molecule. There has been a substantial amount of work done to study the electronic transport through different molecules. Different methods that will be discussed in this section that could measure the electrical conductivity of DNA and related materials are mechanically controlled break junctions (MCBJs) (Kang et al., 2008), scanning tunneling microscopy (STM) (Zareie et al., 2003), conductive or current-sensing AFM (Inoue et al., 2008), two- and four-probe methods (Hartzell, 2004).

MCBJ technique is based on the premise of using single-atom contacts to measure the conductivity of single molecules. Kang *et al* (Kang et al., 2008) have done the

characterization of DNA using this method. To fabricate the MCBJ device, a thin film of gold is patterned on an elastic substrate using e-beam lithography. The elastic substrate-gold system is clamped on the two sides by counter supports along the face of gold deposition. The substrate is then pushed from the bottom, to the point that gold film gradually gets thinner and breaks, leaving an atom-thick contact. The distance between the nano-contacts could be engineered by the movement of the pushing rod.
In scanning probe microscopy (SPM) methods, one end of the DNA is bound to a metallic base (like gold) which has a bias with respect to the tip, which will measure the *I-V* characteristic of the DNA (Felice et al., 2007). Figure 19 depicts the way in which SPM techniques could be used to electrically characterize DNA.

Fig. 19. The substrate is insulating in the left case; the figures depict the measurement of *I-V* characteristics of DNA using SPM techniques (Felice et al., 2007)

The simplest schematic of a two probe method is shown in figure 20. DNA is probed by two electrical contacts and a potential difference is applied across them. Thus, *I-V* characteristics of DNA are measured using this method.

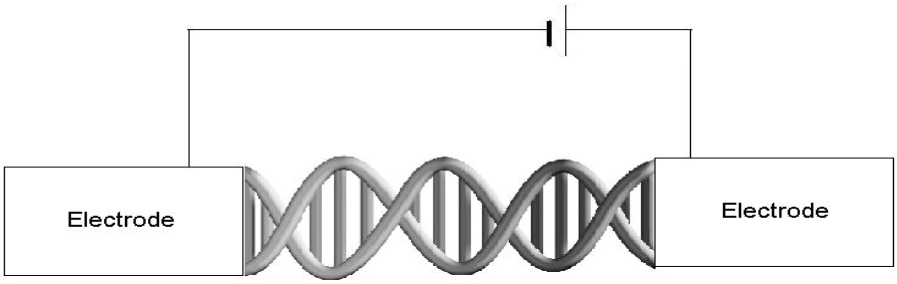


Fig. 20. Two probe method of characterizing DNA (Felice et al., 2007; Hartzell, 2004)

The two and four probe methods are used to calculate the sheet-resistance (resistivity per unit thickness) of different materials. Single DNA is to be spread across the pins of the four probe instrument for molecular characterization by either physical or chemical means, as discussed above. Figure 21 shows the working schematic of the four probe measurement method. Figures 22(a) and 22(b) show one of the possible configurations of the four-probe device, realized by Rastogi for his master's thesis.

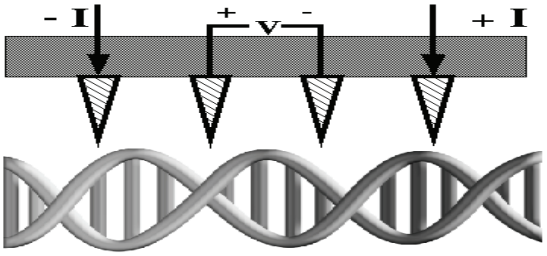


Fig. 21. Four probe device configuration

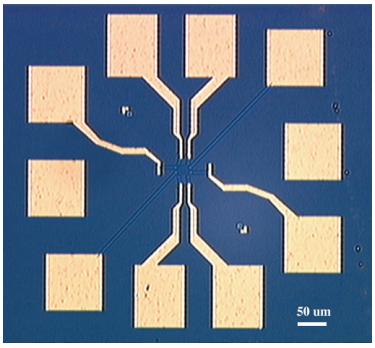


Fig. 22(a). Optical microscope image of a single device containing the nano-electrodes and the connecting-lines pattern (Rastogi, 2002)

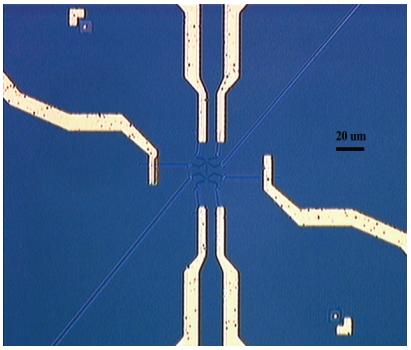


Fig. 22(b). A magnified optical microscope view showing the details of the connecting lines (Ti/Au) about 20 nm thick, which connect the inner (inside yellow ring) electrode pattern (Pt - 4 nm thick) to the contact pads (130000 nm thick). They were written in the first step of E-Beam Lithography (Rastogi, 2002)

The principal challenge, after listing out the possible experimental techniques to characterize DNA molecule, is to ensure that DNA is stretched out linearly across the measurement system. It could be done either by physical means (as discussed in previous section of alignment of DNA) or by chemically modifying either DNA or the electrical contacts or both. Chemical modification can be done by either silanizing the contact surface so that DNA sticks to it or by labeling the DNA molecule with biotin and contact surface with streptavidin. Another method is to label the DNA with thiol-groups and make contacts with gold. Another approach is the combination of the above two methods to ensure that DNA is properly stretched between two electrodes or attaching complementary ss-DNA to the electrodes such that overhanging fragments of enzyme-digested ds-DNA could bind to them.

4.5.4 Realization of DNA-based electrical interconnects and field effect transistors

E. Braun *et al* made first ever DNA- templated metal (silver) nanowires (Braun E. et al., 1998). Oligonucleotides with overhanging bases with two different sequences were attached

to the 50 μm parellel electrodes. Linear λ-DNA has the overhanging sequences, complementary to the overhangs attached to the electrodes. When λ-DNA was flown perpendicular to the faces of the electrodes, it got stuck with the complementary sequences on the electrodes. Thus, a λ-DNA bridge was formed between two electrodes. DNA metallization was done by deposition of silver ions on the negatively charged DNA molecule. The resultant the silver-DNA complex was reduced using a basic hydroquinone solution (pH ~ 10.5) forming small silver aggregates on the λ-DNA template. The silver template is further *developed* using an acidic hyroquinone solution (pH ~ 3.5) and silver ions under low light conditions forming a silver nanowire. The nanowire thus formed is an excellent alternative for DNA-templated electronic interconnects. Many different groups have demonstrated the formation of DNA-templated copper, aluminum, silver, gold as well as platinum nanowires (Monson et al., 2003; Knez et al., 2003; Kryachko et al., 2005; Mertig et al., 2002; Ongaro et al., 2005; Braun G. et al., 2005) that could be used as interconnects.

Keren *et al* reported the selective metalization of the DNA (Keren et al., 2002). It used *E. coli* RecA protein bound to the homologous ss-DNA sequence. RecA is an ATPase that mediates homologous recombination process in bacterial systems. The RecA-DNA complex would bind to the ds-DNA on which selective metallization is to be done (as shown in figure 23). This complex would act as a mask during the electroless deposition of silver and finally gold on exposed ds-DNA.

Moving on to the fabrication of DNA-mediated field effect transistors, synthesis of DNA-templated carbon nanotube FET has been reported (Keren et al., 2003). This includes the selective-metallization-of-DNA concept and semiconducting carbon nanotubes are templated over RecA using anti-RecA antibodies.

Other FETs include M-DNA based transistor (Nokhrin et al., 2007) and altered base pair mediated transistor (Maruccio et al., 2003). The M-DNA transistor works when a voltage applied perpendicular to its helix displaces metal ion and hence, differences in the site energies are created. This displacement can be controlled by changing the voltage applied. This could function as a gate, thus controlling the gate current. The altered base pair FET works with a modified base system: a deoxyguanosine (dG) derivative. In this case, the device behaves like a p-channel MOSFET with a maximum V_{out}/V_{in} ratio of 0.76.

5. Conclusion

In this chapter, we have shown that a combination of top-down & bottom-up approaches offers an effective way to address the scaling challenges faced by the CMOS technologies. For copper interconnect technologies, it has been shown that porphyrin SAMs can be used as effective diffusion barriers, thus improving the device performance and life-time. Zn-P SAM preparation can be done using wet/vapor chemical method and can withstand BEOL process conditions. BTS studies on MOS structures confirm the effectiveness of zinc porphyrin SAM as a good copper diffusion barrier. Patterning the SiO_2 surface using 5-(4-Hydroxyphenyl)-10, 15, 20-tetra (p-tolyl) porphyrin SAM in order to achieve alternate hydrophobic and hydrophilic surfaces has tremendous potential in micro/nanofluidics. Barrier properties of meso-pyridil porphyrin SAM on gold shows that the porphyrins have interesting applications for realization of molecular devices.

Bare DNA is not useful for electronics related applications. It has to be templated by some or the other material that takes the DNA conductivity to the level of semiconductors or

metals. Also, it is seen that different manipulations on the DNA could be done using nucleases or DNA-binding proteins leading to the DNA-mediated electronic interconnects or DNA-mediated field effect transistors. This opens up an entirely broad spectrum of research possibilities in the possible use of other biomolecules for electronic applications. This study also opens up the field of electronic transport studies in biomolecules that will give an electronic perspective to these materials.

6. References

Aich, P.; Labiuk, S. L.; Tari, L. W.; Delbaere, L. J. T; Roesler, W. J.; Falk, K. J.; Steer, R. P. & Lee, J. S. (1999). M-DNA: A Complex Between Divalent Metal Ions and DNA which Behaves as a Molecular Wire. *J. Mol. Biol.* 294, 477-485

Aswal, D. K. ; Lenfant, S. ; Guerin, D. ; Yakhmi, J. V. & Vuillaume, D. (2006). "Self assembled monolayers on silicon for molecular electronics", Analytica Chimica Acta 568, 84–108.

Awaya, N.; Inokawa, H.; Yamamoto, E.; Okazaki, Y.; Miyake, M.; Arita, Y. & Kobayashi, T. (1996). "Evaluation of a Copper Metallization Process and the Electrical Characteristics of Copper-Interconnected Quarter-Micron CMOS," IEEE Transactions on Electron Devices, Aug, Volume: 43, Issue 8, page(s): 1206-1212.

Becker, J. S. & Gordon, R. G. (2003). "Diffusion barrier properties of tungsten nitride films grown by atomic layer deposition from bis(*tert*-butylimido)bis(dimethylamido)tungsten and ammonia", Appl. Phys. Lett. 82, 2239-2241.

Berliocchi, M.; Manenti, M.; Bolognesi, A.; Di Carlo, A.;Lugli, P.; Paolesse, R.; Mandoy, F.; Di Natale, D.; Proietti, E.; Petrocco, G. & D'Amico,A. (2004). "Charge transport in pentacene and porphyrin-based organic thin film transistors", *Semicond. Sci. Technol*. 19 S354-S356.

Botti, S.; Ciardi, R.; Terranova, M. L.; Piccirillo, S.; Sessa, V. & Rossi, M. (2002) Carbon nanotubes and nanowires grown from spherical carbon nano-particles. Chemical Physics Letters, 355: 395–399.

Braun, E.; Eichen, Y.; Sivan, U. & Ben-Yoseph, G. (1998). DNA-templated assembly and electrode attachment of a conducting silver wire. *Nature.* 391, 775-778

Braun, G.; Ingaki, K.; Estabrook, R.A.; Wood, D.K.; Levy, E.; Cleland, A.N.; Strouse, G.F. & Reich, N.O. (2005). Gold Nanoparticle Decoration of DNA on Silicon. Langmuir. 21(23), 10699-701

Burrell, A. K.; Michael & Wasielewski, R. (2000). "Porphyrin-based nanostructures: routes to molecular Electronics", Journal of Porphyrins and Phthalocyanines *J. Porphyrins Phthalocyanines* 4, 401–406.

Cai, L.; Tabata, H. & Kawai T. (2000). Self-assembled DNA networks and their electrical conductivity. *Appl. Phys. Lett.* 85, 44992-44995

Chen, J.; Wang, W.; Reed, M. A.; Rawlett, A. M.; Price, D. W. & Tour, J. M. (2000). "Room-temperature negative differential resistance in nanoscale molecular junction," Appl. Phys. Lett., vol. 77, pp. 1224-1226.

Cluzel, J.; Mondon, F.; Blachier, D.; Morand, Y.; Martel, L. & Reimbold, G. (2002). "Electrical characterization of copper penetration effects in silicon dioxide," Annual lntemational Reliability Physics Symposium, IEEE, p431-432.

Dallaporta, H.; Liehr, M. & Lewis, J. E. (1990). "Silicon dioxide defects induced by metal impurities," Phys. Rev. B 41, 5075 - 5083.

Datta, S.; Tian, W.; Hong, S.; Reifenberger, R.; Henderson, J. & Kubiak, C. P. (1997). „Current-Voltage Characteristics of Self-Assembled Monolayers by Scanning Tunneling Microscopy", Phys. Rev. Lett. 79, 2530.

David Dolphin. (1978). "The Porphyrins", Physical Chemistry, Part C, Academic Press, New York.

dePablo, P. J.; Moreno-Herrero, F.; Colchero, J.; Gomez Herrero, J.; Herrero, P.; Baro, A.; Ordejon, P.; Soler, J.M. & Artacho, E. (2000). Absence of dc-Conductivity in λ-DNA. *Phys. Rev. Lett.* 85, 4992-4995

Dewarrat F. (2002). *"Electric Characterization of DNA"*. Doctoral thesis. University of Basel, Switzerland

Diez, S.; Reuther, C.; Dinu, C.; Seidel, R.; Mertig, M.; Pompe, W. & Howard, J. (2003). Stretching and Transporting DNA Molecules Using Motor Proteins. *Nano Letters.* 3(9), 1251-1254

Dudhe, R. S.; Tiwari, S. P.; Raval, H. N.; Khaderbad, M. A; Sinha, J.; Yedukondalu, M.; Ravikanth, M.; Kumar, A.; & Rao, V. R. (2008). "Explosive vapour sensor using poly (3-hexylthiophene) and Cu-tetraphenyl-porphyrin composite based organic field effect transistors", Applied Physics Letters, Vol. 93, 263306.

Dudhe, R. S.; Mishra, P.; Sinha, J.; Kumar, A. & Rao, V. R. (2009) "A MWCNT/P3HT Based chemiresistor for RDX Detection", Proceedings of the International Conference on MEMS, Chennai, India January 3-5.

Elers, K. E.; Saanila, V.; Soininen, P. J.; Li, W. M.; Kostamo, J. T.; Haukka, S.; Juhanoja, J. & Besling, W. F. A. (2002). "Diffusion Barrier Deposition on a Copper Surface by Atomic Layer Deposition", Chemical Vapor Deposition, 8(4), 149-153.

Eley, D. D. & Spivey, D. I. (1961). Semiconductivity of organic substances. Part 7. The polyamides. *Trans. Faraday Soc.* 57, 2280 – 2287

Felice, R. D. & Porath, D. (2007). DNA based Nanoelectronics, In: NanoBioTechnology: BioInspired Devices and Materials of the Future, Shoseyov O. (Ed.), Levy I. (Ed.), 141-185, Humana Press. 1 edition. ISBN-13: 978-1588298942, New Jersey, USA

Fink, H.W. & Schoenenberger, C. (1999). Electrical conduction through DNA molecules. *Nature*. 398, 407-410

Ganesan, P. G.; Singh, A. P.; & Ramanath, G. (2004). "Diffusion barrier properties of carboxyl- and amine-terminated molecular nanolayers," Appl. Phys. Lett., 85, 579-581.

Garcia, H. A. B. (2007). DNA-templated nanomaterials. Doctoral Thesis. Brigham Young University, U.S.A.

Ghosh, A. & Bansal, M. (2003). A glossary of DNA structures from A to Z. Acta Crystallogr D Biol Crystallogr 59(4), 620–626

Grenoble, S; Gouterman, M.; Khalil, G.; Callis, J & Dalton, L. (2005). "Pressure-sensitive paint (PSP): concentration quenching of platinum and magnesium porphyrin dyes in polymeric films", Journal of Luminescence 113 , 33–44.

Hartzell, B. M. (2004). DNA manipulation and characterization for nanoscale electronics. Doctoral Thesis. Ohio University, U.S.A.

Inoue, T.; Qian, J.; Liao, S.; Li, Y.; Narayanamurty, V.; Vasudev, M.; Dutta, M. & Stroscio, M. A. (2008). Comparison of calculated and measured I – V curves for DNA. Journal of Computational Electronics. 7(1), 43-45.

Jia, T.; Jiang, Z; Wang, K. & Li, Z. (2006). Binding and photocleavage of cationic porphyrin–phenylpiperazine hybrids to DNA. Biophysical Chemistry. 119 (3), 295-302

Kang, N.; Erbe, A. & Scheer, E. (2008). Electrical characterization of DNA in mechanically controlled break-junctions. New J. Phys. 10, 023030-023038

Kasumov, A. Yu.; Kociak, M.; Guéron, S.; Reulet, B.; Volkov, V. T.; Klinov, D. V. & Bouchiat, H. (2001). Proximity-Induced Superconductivity in DNA. Science. 291(5502), 280-282

Keren, K.; Krueger, M.; Gilad, R.; Ben-Yoseph, G.; Sivan, U. & Braun, E. (2002). Sequence-specific molecular lithography on single DNA molecules. Science. 297, 72-75

Keren, K.; Berman, R.S.; Buchstab, E.; Sivan, U. & Braun, E. (2003). DNA-templated carbon nanotube field-effect transistor. Science. 302, 1380-1382

Khaderbad, M. A.; Nayak, K.; Yedukondalu, M.; Ravikanth, M.; Mukherji, S. & Rao, V. R. (2008). "Metallated Porphyrin Self Assembled Monolayers as Cu Diffusion Barriers for the Nano-scale CMOS Technologies", Proceedings of the 8th IEEE Conference on Nanotechnology, August 18-21, 2008, Arlington, Texas USA.

Khaderbad, M. A.; Mukherji, S. & Rao, V. R. (2008). "DNA Based Nanoelectronics", Recent Patents on Electrical Engineering, Volume 1, Number 2, June 2008, pp. 115-120 (Bentham Sciece Publishers) (Invited review article)(ISSN: 1874-4761 - Volume 1).

Koike, J. & Wada, M. (2005). "Self-forming diffusion barrier layer in Cu–Mn alloy metallization", Appl. Phys. Lett. 87, 041911

Krishnamoorthy, A.; Chanda, K.; Murarka, S. P.; Ramanath, G. & Ryan, J. G. (2001). "Self-assembled near-zero-thickness molecular layers as diffusion barriers for Cu metallization," A ppl. Phys. Lett., vol.78, pp. 2467-2469.

Li, C.; Lei, B.; Fan, W..; Zhang, D.; Meyyappan, M. & Zhou, C. (2007). "Molecular Memory Based on Nanowire–Molecular Wire Heterostructures", Journal of Nanoscience and Nanotechnology Vol.7, 138–150.

Loke, A. L. S.; Wetzel, J. T.; Changsup Ryu; Won-Jun Lee & Wong, S. (1998). "Copper drift in low-K polymer dielectrics for ULSI metallization," VLSI Technology syposium, IEEE Jun, page(s): 26-27.

Love, J. C.; Lara, A.; Estroff; Kriebel, J. K.; Nuzzo, R. G. & Whitesides, G. M. (2005). "Self-Assembled Monolayers of Thiolates on Metals as a Form of Nanotechnology", Chem. Rev., 105, 1103-1169

Ma, Y.; Zhang, J.; Zhang, G. & He, H. (2004). Polyaniline Nanowires on Si Surfaces Fabricated with DNA Templates. J. Am. Chem. Soc. 126(22), 7097-7101

Marrucio, G.; Visconti, P.; Arima, V.; D'Amico, S.; Biasco, A.; D'Amone, E.; Cingolani, R.; Rinaldi, R.. Masiero, S.; Giorgi, T. & Gottarelli, G. (2003). Field Effect Transistor Based on a Modified DNA Base. Nano Letters. 3(4), 479-483

Mertig, M.; Ciacchi, L. C.; Seidel, R. & De Vita, W. (2002). DNA as selective metallization template. Nano Letters. 2(8), 841-844

Molecular Biology Web Book (2009). Web Books Publishing

Nayak, K.; Kulkarni, P; Deepu, A.; Sitaraman, V.; Punidha, S.; Saha, A.; Ravikanth, M.; Mitra, S; Mukherji, S. & Rao, V. R. (2007). "Patterned Microfluidic Channels using

Self-assembled Hydroxy-phenyl Porphyrin Monolayer", Proceedings of the 7th IEEE International Conference on Nanotechnology, August 2-5, Hong Kong.

Neidle S. (2008). Principles of Nucleic Acid Structure, Academic Press (Elsevier), ISBN-13: 978-0-12-369507-9/ISBN-10: 0123695074, Oxford, U.K.

Nokhrin, S.; Baru, M. & Lee, J.S. (2007). A field-effect transistor from M-DNA. Nanotechnology. 18(9), 095205-095209

Nowak, E. J. (2002). "Maintaining the benefits of Scaling", IBM J. RES. & DEV. VOL. 46 NO. 2/3 MARCH/MAY, 169-180.

Onclin, S.; Ravoo, B. J. & Reinhoudt, D. N. (2005). "Engineering silicon dioxide surfaces using self-assembled monolayers", Angew. Chem. Int. Ed., 44, 6282-6304.

Ongaro, A.; Griffin, F.; Beecher, P.; Nagle, L.; Iacopino, D.; Quinn, A.; Redmond, G. & Fitzmaurice, D. (2005). DNA-Templated Assembly of a Protein-Functionalized Nanogap Electrode. *Chem. Mater*. 17(8), 1959-1964

Papkovsky, D. B.; O'Riordan, T. & Soini, A. (2000). "Phosphorescent porphyrin probes in biosensors and sensitive bioassays", Biochem. Soc. Trans. 28, (74–77).

Porath, D.; Bezryadin, A.; deVries, S. & Dekker, C. (2000). Direct measurements of electrical transport through DNA molecules. *Nature*. 403, 635–338

Rajwade, S. (2007). Polyaniline Nanowires on DNA Templates for Nanoelectronics Applications. Master of Technology thesis. Indian Institute of Technology Bombay, India

Rakitin, A.; Aich, P.; Papadopoulos, C.; Kobzar, Y.; Vedeneev, A.S.; Lee, J.S. & Xu, J.M. (2001). Metallic Conduction through Engineered DNA: DNA Nanoelectronic Building Blocks. *Phys. Rev. Lett*. 86, 3670-73

Rastogi, S. (2002). Novel approach to DNA-based Molecular Electronics & Towards Four-probe I-V measurements of DNA. Master of Technology thesis. Indian Institute of Technology Bombay, India

Reimers, J. R.; Lü, T. X.; Crossley, M. J. & Hush, N. S. (1996). "Molecular electronic properties of fused rigid porphyrin-oligomer molecular wires", *Nanotechnology* **7** 424-429.

Reed, M. A.; Zhou, C.; Muller, C. J.; Burgin, T. P. & Tour, J. M. (1997). "Conductance of a molecular junction", *Science* 278 (5336), 252

Richter, J. (2003). *Metallization **of** DNA. Physica E*. 16, 157–173

Rosenberg, R.; Edelstein, D. C.; Hu, C. K. & Rodbell, K.P. (2000). *Annu. Rev. Matter. Sci.* **30**, pp. 229–262.

Saiki, R. K.; Scharf S.; Faloona F.; Mullis K. B.; Horn G. T.; Erlich H. A. & Arnheim N. (1985). "Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia". *Science* 230, 1350–1354

Saiki, R. K.; Gelfand, D. H.; Stoffel, S.; Scharf, S. J.; Higuchi, R.; Horn, G. T.; Mullis, K. B. & Erlich, H. A. (1988). "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase". Science 239, 487–491

Sathyapalan, A.; Lohani, A.; Santra, S.; Goyal, S.; Ravikanth, M.; Mukherj, S. & Rao, V. R. (2005). "Preparation, characterization and electrical properties of a novel self-assembled meso-pyridyl pophyrin monolayer on gold surfaces", Australian Journal of Chemistry, Vol.58, pp. 810-816.

Savva, R. & Pearl, L.H. (1995). Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase–uracil glycosylase inhibitor protein complex. Nature Structural Biology 2, 752–757

Schwab, A. D. (2004). "Photoconductivity of Self-Assembled Porphyrin Nanorods", Nano Letters, Vol. 4, No. 7, 1261-1265.

Seidel, R. (2003). Methods for the development of a DNA based nanoelectronics. Doctoral Thesis. Technical University, Dresden, Germany

Shacham-Diamond, Y.; Dedhia, A.; Hoffstetter, D. & Oldham, W. G. (1991). "Reliability of copper metallization on silicon-dioxide," VLSI Multilevel Interconnection Conference, IEEE, Santa Clara, USA.

Shengyang Tao & Guangtao Li. (2007). "Porphyrin-doped mesoporous silica films for rapid TNT detection", Colloid Polym Sci 285:721–728

Storm, A.J.; vanNoort, J.; deVries, S. & Dekker, C. (2001). Insulating behavior for DNA molecules between nanoelectrodes at the 100 nm length scale. *Appl. Phys. Lett.* 79, 3881-3883

Suslick, K. S.; Rakow, N. A.; Kosal, M. E.; & Chou J-H. (2000). "The materials chemistry of porphyrins and metalloporphyrins", Journal of Porphyrins and Phthalocyanines Volume 4, Issue 4, Pages 407 – 413.

Tran, P.; Alavi, B. & Gruner, G. (2000). Charge transport along the lambda-DNA double helix. *Phys. Rev. Lett*. 85, 1564-1567

Ulman, A. (1996). "Formation and Structure of Self-Assembled Monolayers", Chem. Rev. 96, 1533-1554.

Ulman, A. (1998). "Self-Assembled Monolayers of Thiols", Thin Films, Volume 24, Academic Press, San Diego.

Roy, U.; Khaderbad, M. A.; Yedukondalu, M.; Ravikanth, M.; Mukherji, S. & Rao, V. R. (2009). "Hydroxy-phenyl Zn(II) Porphyrin Self-Assembled Monolayer as a Diffusion Barrier for Copper-Low k Interconnect Technology", Proceedings of the IEEE International Workshop on Electron Devices & Semiconductor Technology, Mumbai, India.

Wagenknecht, H. (2005). Principles and Mechanisms of Photoinduced Charge Injection, Transport and Trapping in DNA, In: *Charge Transfer in DNA: From Mechanism to Application,* Wagenknecht H. (Ed.), 1-26, Wiley-VCH. ISBN-13: 978-3527310852, Weinheim, Germany

Wei, L. & Charles Lieber, M. (2007). "Nanoelectronics from the bottom-up", nature materials, VOL 6, 841-850.

Weizmann, Y.; Patolsky, F.; Popov, I. & Willner, I. (2004). Telomerase-generated templates for the growing of metal nanowires. *Nano Letters* 4(5), 787-792

Wende, H.; Bernien, M.; Luo, J.; Sorg, C.; Ponpandian, N.; Kurde, J.; Miguel, J.; Piantek, M.; Xu, X. & Eckhold, (2007). "Substrate-induced magnetic ordering and switching of iron porphyrin molecules", P. Nat. Mater. 6, 516-520.

Wood, D. O.; Dinsmore, M. J.; Bare, G. A.; & Lee, J. S. (2002). M-DNA is stabilized in G-C tracts or by incorporation of 5-fluorouracil. *Nucleic Acids Res*. 30(10), 2244–2250

Xin, H. (2006). DNA-templated surface alignment and characterization of carbon nanotubes. Doctoral Thesis. Brigham Young University, U.S.A.

Yoo, K.H.; Ha, D.; Lee, J.O.; Park, J.W.; Kim, J.; Kim, J.J.; Lee, H.Y.; Kawai, T. & Choi, H.Y. (2001). Evidence of phonon-assisted tunnelling in electrical conduction through DNA molecules. *Phys. Rev. Lett*. 87, 19102-19104

Zareie, M. H. & Lukins P. B. (2003). Atomic-resolution STM structure of DNA and localization of the retinoic acid binding site. *Biochemical and Biophysical Research Communications*. 303(1),153-159.

8

Templated Si-based nanowires via solid-liquid-solid (SLS) and vapor-liquid-solid (VLS) growth: Novel growth mode, synthesis, morphology control, characteristics, and electrical transport

Jae Ho Lee and Robert E. Geer
College of Nanoscale Science and Engineering
University at Albany, SUNY
USA

1. Introduction

The 'scaling down' of device and interconnect features in CMOS technology to the deep sub-100 nm regime has motivated substantial research focusing on the development of nanoscale building blocks formed via self-assembled or 'bottom-up' synthesis to complement prevailing lithographic or 'top-down' CMOS processing. Examples of such nanoscale building blocks include metallic and semiconducting nanowires or nanobelts, carbon nanotubes and, most recently, n-layer graphenes. In particular, one-dimensional (1-D) nanostructures have attracted much attention because of their unique electrical, optical, mechanical and thermal properties and their potential utility in a wide variety of nanoelectronic and optoelectronic applications. Non-carbon 1-D nanomaterials fabricated from bottom-up synthesis can be used as fundamental building blocks for nanoscale devices and circuits and may have the potential to replace certain, conventional top-down processes. In particular, silicon nanowires (SiNWs) may be an attractive alternative to conventionally processed Si transistors if their intrinsic self-assembly can be harnessed to obviate the need for complex lithographic techniques for device fabrication. In addition, SiNWs can potentially function as both the switch (i.e. transistor) and local interconnect (e.g. metal silicide nanowire) to form an inherently integrated nanoelectronic system – potentially on the same self-assembled nanostructure (Morales *et al.* 1998; Lu *et al.* 2007; Colli *et al.* 2007; Wu *et al.* 2004). Recent research has demonstrated that Si-based silicide nanowires may yield performance superior to conventional Cu interconnects at sub 10-nm wire widths which highlights the excellent potential for SiNW-based systems (Wu *et al.* 2004; Zhang *et al.* 2000; Kim *et al.* 2005; Kim *et al.* 2003). And since NiSi has been shown to be a good electrical contact material for gate, source and drain in Si CMOS (Lavoie *et al.* 2003; Kittl *et al.* 2003; Morimoto *et al.* 1995), NiSi nanowires, in particular, comprise an attractive nanoscale building block material (Wu *et al.* 2004).

Currently, several methods are available to synthesize silicon nanowires including laser ablation (Morales *et al.* 1998; Zhang *et al.* 1998), physical vapor deposition (Zakharov *et al.*

2006), thermal evaporation (Yu *et al.* 1998), and chemical vapor deposition (Westwater *et al.* 1997; Cui *et al.* 2001; Hochbaum *et al.* 2005). In the literature these methods have been roughly categorized according to the underlying growth modes described, respectively, as solid-liquid-solid (SLS) growth (Paulose *et al.* 2003; Yan *et al.* 2000), vapor-liquid-solid (VLS) growth (Wagner *et al.* 1964; Givargizov *et al.* 1975), and oxide assisted growth (OAG) (Zhang *et al.* 2001a; 2001b; Zhang *et al.* 2003).

For eventual technological impact it is essential to understand the fundamental processes by which SiNWs are synthesized as well as develop approaches to tailor the functionality of SiNWs through post-growth processing. *This chapter reviews recent novel research results in both areas regarding SiNW-based materials*. Section 2 presents a brief overview of the fundamental growth mechanisms of SiNWs. Sections 3 and 4 describe the experimental methodology for SiNW growth and the details of SLS and VLS SiNW synthesis, respectively, used for the research presented here. Section 5 details the utilization of post-growth processing SiNWs in the formation of conductive core-shell nanostructures. This includes detailed electrical transport testing and modeling. Section 6 reviews the recent observation of a novel SLS-based SiNW growth mode in which nanowire growth is preferentially initiated at oxide-meditated etch pits in the crystalline Si wafer substrate. Lastly, Section 7 presents a brief summary of the work described in this chapter and highlights opportunities for future research and development.

2. Growth Mechanisms

The solid-liquid-solid (SLS) growth approach is a relatively straightforward technique to synthesize silicon nanowires because it does not require a gas phase precursor such as SiH_4 or $SiCl_4$. Via the SLS process, as shown in Fig. 1, silicon nanowires can be directly grown on a silicon substrate which acts as the silicon source (Paulose *et al.* 2003). In the SLS process a catalyst is deposited on a single-crystal silicon substrate. The annealing of the catalyst-deposited Si substrate results in metal-silicon alloy nano-droplet formation. Continuous diffusion of silicon atoms from the substrate to the droplet at elevated temperatures causes saturation of silicon atoms inside the alloy droplet which leads to precipitation of silicon at the surface of the droplet. The surface Si precipitate forms a Si growth front resulting in nanowire formation from the catalyst because of a negative temperature gradient at the droplet surface (e.g. due to a gas flow) (Paulose *et al.* 2003).

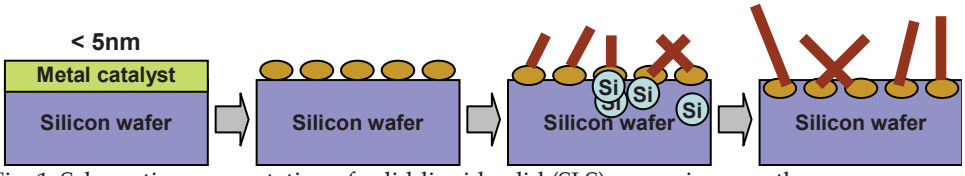


Fig. 1. Schematic representation of solid-liquid-solid (SLS) nanowire growth.

The vapor-liquid-solid (VLS) growth mechanism is widely cited method for nanowire synthesis. It utilizes silane (SiH_4) or $SiCl_4$ instead of a bare silicon wafer for the silicon source, or a laser ablated Si-metal catalyst target (Morales *et al.* 1998; Chen *et al.* 2002). The VLS growth mode was first described in detail decades ago (Wagner *et al.* 1964). A metal catalyst (Au in our experiments) is used in VLS growth as favored sites for absorption of the gas phase precursor. As temperature increases, the thin Au catalyst films break into nano-

scale droplets above the eutectic temperature and Au-Si alloy nanodroplets are formed that supersaturate upon continued exposure to the precursor gas, resulting in the precipitation of the solid nanowire, shown in Fig 2. In contrast to SLS growth, the alloy clusters can generally be observed at the tips of the wires.

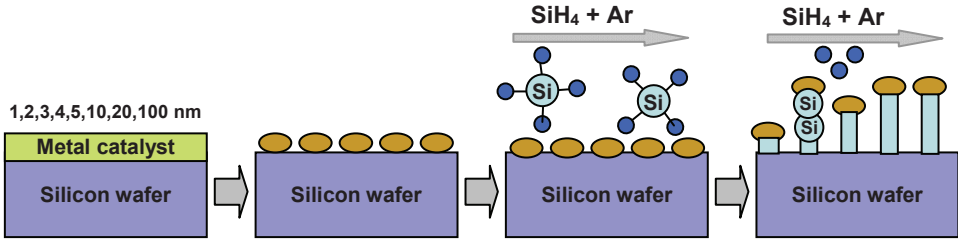


Fig. 2. Schematic representation of vapor-liquid-solid (VLS) nanowire growth.

In contrast to the SLS and VLS SiNW growth modes, SiNW synthesis via oxide assisted growth (OAG) is induced by the presence of silicon oxide instead of a metal catalyst. During the process Si-rich oxide clusters prefer to form Si-Si bonds among one another while oxygen-rich silicon oxide cluster prefers to form Si-O bonds with other oxygen rich SiO_x clusters (Zhang *et al.* 2001a; 2001b; Zhang *et al.* 2003). Highly reactive silicon atoms in the deposited silicon-rich oxide clusters are strongly bonded to the silicon substrate. In the same cluster, exposed non-bonded, reactive silicon atoms exposed to the vapor phase act as nuclei for adsorption of additional reactive silicon oxide clusters (Zhang *et al.* 2003). The result is the formation of a self-assembled Si-rich oxide growth front exhibiting a wire-like morphology.

3. Experimental Methods

For the SiNW synthesis described in this chapter p-type (100) and (111) silicon wafers with electrical resistivity in the 1-20 Ω•cm range were used. Wafers were cleaned with diluted hydrofluoric acid (5%) to remove the native oxide layer and ultrasonicated in acetone to remove organic contamination. The cleaned samples were immediately loaded into a physical vapor deposition (PVD) chamber (evacuated to 5×10^{-7} torr) for sputtering of a Au catalyst film with a thickness of 4 nm for SLS growth. For VLS nanowire growth, a Au catalyst film was deposited on the Si substrate with a thickness that varied from 1-100 nm.
Following Au deposition the silicon samples were placed inside an annealing chamber. The annealing chamber was evacuated to a base pressure of approximately 5 Torr and backfilled with high purity (99.999%) Ar gas. The total pressure of the system was then raised to atmospheric pressure through an Ar gas flow. For SLS nanowire growth, the Au-deposited Si samples were annealed at 1000 °C (30 minute ramping time) under Ar (also 99.999 % purity) gas flow at 2,000 sccm. The annealing duration, as shown in Fig. 3, was varied from 10 minutes to 120 minutes after the temperature ramp to investigate the effect of annealing time on the Si nanowire diameter, length, and overall morphology.

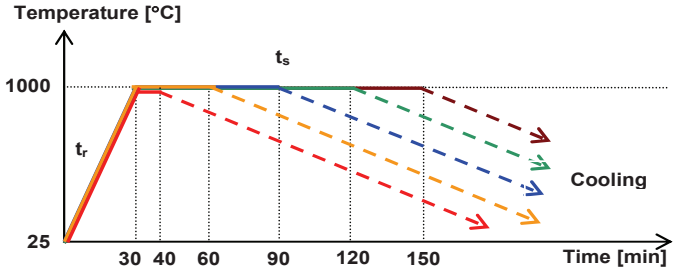


Fig. 3. Annealing duration conditions for SLS growth.

To investigate the SLS SiNW synthesis process, various experimental parameters including Ar gas flow rate and sample position within the chamber were varied. Au-deposited samples were positioned at various points from a lower temperature region within the chamber near the inlet of Ar gas (as compared to the center of the annealing chamber). Early stages of SLS SiNW growth were observed with a shorter annealing duration (typically, below 15 minutes – see Section 6). The temperature inside the oven could be reduced by utilizing an Ar flow with 2,000 sccm when the oven heater is turned off; however, there was a latent annealing duration of approximately 8 min while temperature cooled to ~ 800 °C. To eliminate this unnecessary duration and to enable investigation of the early stages of nanowire growth, 20,000 sccm of Ar was flown inside the chamber, which quenched the growth relatively quickly. Vapor-liquid-solid (VLS) SiNW growth was also investigated to optimize growth rate and diameter distribution. The Au catalyst film thickness was varied from 1 to 100 nm in thickness. Likewise, an optimized VLS SiNW growth rate in the reactor was achieved at a temperature of 500 °C.

For core-shell metal and metal-silicide wire structures surface metallization of SiNWs utilized nickel and the tungsten. Metal deposition and subsequent thermal processing was carried out for as-deposited SiNWs to investigate surface silicide formation. Nickel was deposited on the nanowires via e-beam evaporation (calibrated for an effective blanket film thickness of 150 nm) on SLS and VLS SiNWs. Tungsten was deposited on VLS SiNWs with 2 nm and 4 nm thickness likewise using an atomic layer deposition (ALD) process. Nickel deposited SiNW samples were annealed using a rapid thermal annealing system at 550 °C for 5 min (for SLS, VLS nanowire samples) and at 600 °C for 5 min (for SLS nanowire samples) with a 10 min ramping time to compare resultant SiNW morphologies (subsequently investigated by scanning electron microscopy (SEM)).

Metal-coated SiNWs and as-grown SiNWs were dispensed on metal-patterned Si wafers to carry out two-point and four-point electrical conductivity measurements. Electrical contacts were formed using direct-write Pt electrode deposition within a dual-beam focused ion beam SEM (FIB-SEM). Structural and compositional properties of these wires were analyzed using scanning electron microscopy (SEM), energy dispersive x-ray spectroscopy (EDS), and transmission electron microscopy (TEM).

4. Synthesis of Silicon Nanowires

4.1 Solid-liquid-solid (SLS) growth

Scanning electron microscopy (SEM) micrographs of typical SLS SiNWs synthesized for various-length anneals at 1000 °C are shown in Fig. 4. The furnace temperature was ramped to 1,000 °C for 30 min under 2,000 sccm Ar flow. After the ramp the furnace temperature was held constant for durations ranging from 10 minutes to 120 minutes. The effects of SLS-growth annealing duration have been reported (Lee *et al*. 2008). An approximately bimodal diameter distribution was observed. In the previous work (Lee *et al*. 2008), variation of the anneal duration and Ar gas flow rate served to quench the SiNW growth and also served to coarsely regulate nanowire diameter. This preliminary conclusion follows from the basic SLS growth mechanism if high Ar flow rates (~ 20,000 sccm) are sufficient to reduce the temperature at the alloy droplet below that necessary for SiNW growth.

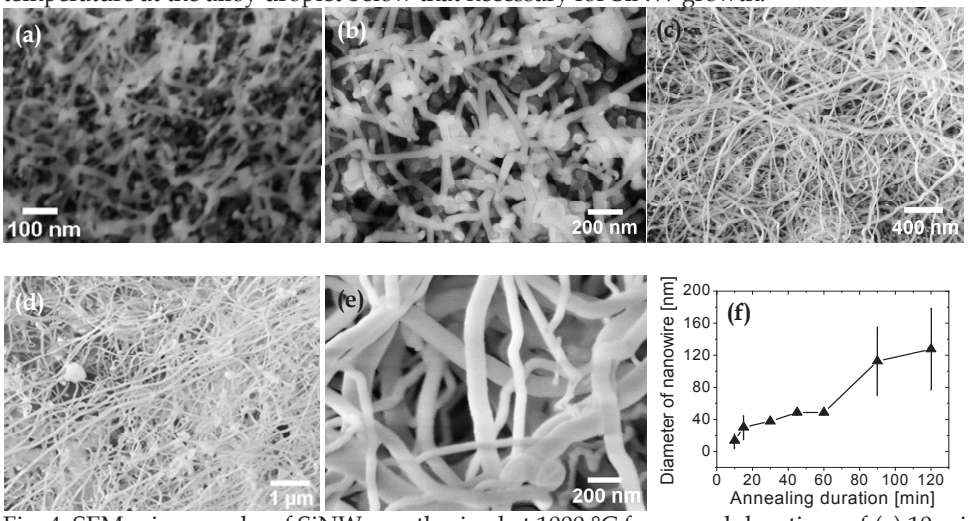


Fig. 4. SEM micrographs of SiNWs synthesized at 1000 °C for anneal durations of (a) 10 min, (b) 15 min, (c) 45 min, (d) 90 min, (e) 120 min. Diameters ranged from 5 nm to 178 nm. (f) Dependence of annealing duration on SLS nanowire growth (Lee *et al*. 2008).

Transmission electron microscopy (TEM) was used to carry out compositional and structural analyses of SLS nanowires (Fig. 5(a)). Nanowires were detached from the initial silicon wafer using ultrasonication in a methanol solution as described above. A selected-area electron diffraction (SAED) pattern (Fig. 5(b)) was acquired from an individual SiNW and indicated a locally amorphous structure. The EDS spectra of as-deposited SiNWs on highly oriented pyrolytic graphite (HOPG) substrates indicated the presence of oxygen (Fig. 5(c)) and implied that the nanowires are likely SiO_x in nature, consistent with the observation of an amorphous microstructure. It should be noted that nanowires grown via the SLS method are generally in an amorphous state, while crystalline nanowires have been typically observed via VLS growth. This may result from the high temperature of the SLS growth process and the relatively high growth rates of those nanowires, which lead to amorphous rather than crystalline microstructure.

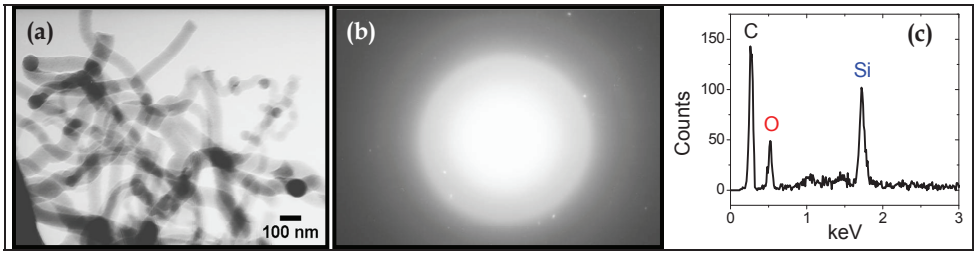


Fig. 5. (a) TEM micrograph of SLS-grown SiNW, (b) SAED pattern. (c) EDS spectra of SiNW. Si and O are the primary elements detected. The SEM-EDS was performed on an HOPG substrate. The carbon peak is due to the substrate.

4.2 Vapour-liquid-solid (VLS) growth

SEM micrographs of VLS SiNWs synthesized for various catalyst Au thicknesses at 500 °C are shown in Fig. 6. It is clearly apparent that the VLS-grown nanowires show a linear morphology instead of the entangled structure exhibited by SLS grown nanowires. The tilted SEM micrograph (Fig. 6(e)) confirms the vertical growth of nanowires from the Si wafer. In Fig. 6(f) a silicon pillar structure was observed.

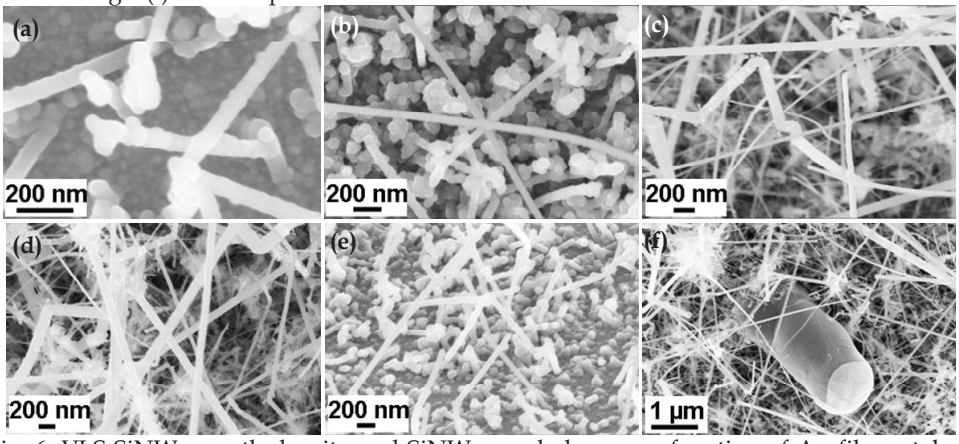


Fig. 6. VLS SiNW growth density and SiNW morphology as a function of Au film catalyst thickness. SEM images of VLS SiNWs for Au catalyst thickness of (a) 1 nm, (b) 5 nm, (c) 20 nm and (d) 100nm. VLS SiNW diameters ranged from 5 nm to 120 nm, (mean ~ 60 nm). (e) Tilted SEM image to show vertical SiNW growth, (f) SEM imaging showing a Si pillar structure formed during SiNW growth.

If sufficient processing control could be established, such large self-assembled Si structures may have potential for use in 3D IC interconnections upon metallization or metal silicidation. Ni was also used as a catalyst for VLS growth (Fig 7.). Experimental results show that the SiNW morphology from a Ni catalyst is similar to that of Au-catalyzed VLS SiNWs. Bi-directional growth was observed at the edge of one Ni-catalyzed VLS SiNW sample (Fig 7(b)).

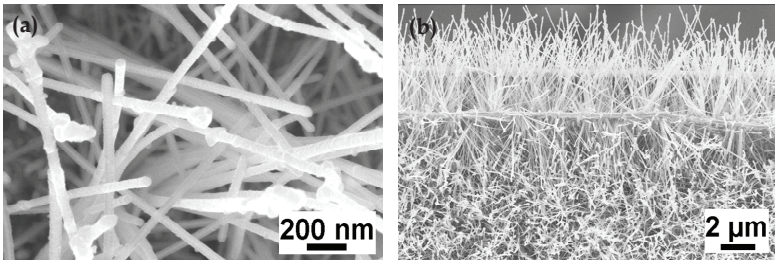


Fig. 7. SEM micrographs, (a) VLS SiNW growth using Ni catalyst, (b) bi-directional growth.

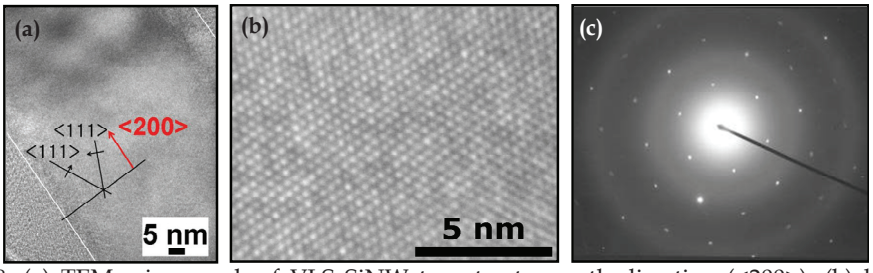


Fig. 8. (a) TEM micrograph of VLS SiNW to extract growth direction (<200>), (b) lattice structure of VLS SiNW, (c) selected-area electron diffraction (SAED) patterns from a SiNW. The SEAD pattern shows the perfect crystalline structure of an as-grown VLS SiNW.

Transmission electron microscopy (TEM) of VLS SiNWs was also undertaken for compositional and structural analyses (Fig. 8) and for comparison with SLS-grown nanowires. VLS Si nanowires were detached from the initial silicon wafer using ultrasonication in a methanol solution as described above. Selected-area electron diffraction (SAED) patterns (Fig. 8(c)) indicated a crystalline structure for all VLS nanowires investigated. It should be noted that nanowires grown via VLS are typically crystalline, while amorphous nanowires have been typically observed via SLS growth. Both trends agreed with the work presented here.

The physical properties of both VLS and SLS SiNWs were further investigated with Raman scattering. Figure 9(a)-(c) shows Raman spectra of an as-grown SLS and VLS SiNW compared to that of the Si wafer substrate. The SLS SiNW Raman peak is located at 509 cm^{-1}, and the VLS SiNW Raman peak is located at 517 cm^{-1} while the Si substrate Raman peak is located at 520 cm^{-1}. All peaks correspond to the first-order transverse optical phonon mode (TO) in locally ordered Si. The substantial SiNW TO peak shift and broadening evident in the SLS SiNW results from the substantial local Si-Si disorder. This observation is consistent with prior Raman measurements of SiO_x systems (Yu *et al.* 1998; Li *et al.* 1999; Li *et al.* 2005). Figure 9(d)-(f) compares the Raman spectra for the same phonon mode from an as-grown VLS SiNW, a similar VLS SiNW following 56 days of ambient exposure, and an as-grown SLS SiNW. Following substantial ambient exposure the Si Raman peak is broadened and shifted to 498 cm^{-1} indicative of significant oxidation and possible amorphization.

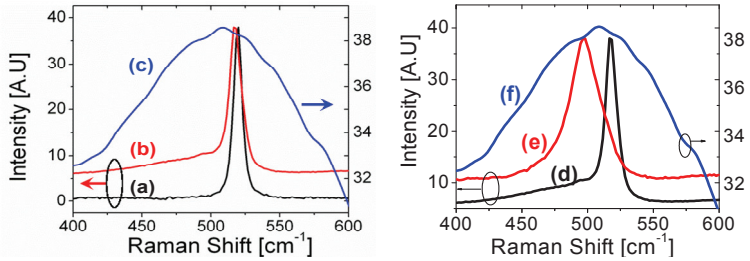


Fig. 9. (Left) Si-Si Raman band from (a) silicon wafer, (b) as-grown VLS SiNW, and (c) as-grown SLS SiNW. (Right) Si-Si Raman band from (d) as-grown VLS SiNW, (e) VLS SiNW after 56 days ambient exposure, and (f) as-grown SLS SiNW.

5. Core-shell Approaches for Metallic Channel Formation

5.1. Synthesis, metallization, structural and compositional characterization of SLS SiNWs

To investigate the formation of conductive shells on SiNWs, nickel was deposited on as-grown SLS SiNWs via e-beam evaporation calibrated for an effective blanket film thickness of 150 nm. Post-deposition thermal processing was carried out for nickel silicide formation (550 °C for 5 minutes via rapid thermal annealing (RTA) and 600 °C for 5 minutes with a 10 minute ramp). The post-anneal nanowire surface morphology was sensitive to the anneal temperature and ramp rate. Rapid ramps resulted in an atomically-smooth Ni-SiNW surface morphology. Slow annealing resulted in a rough Ni-SiNW surface morphology indicative of non-uniform silicide domain formation, as shown in Fig. 10(a), (b), and (c).

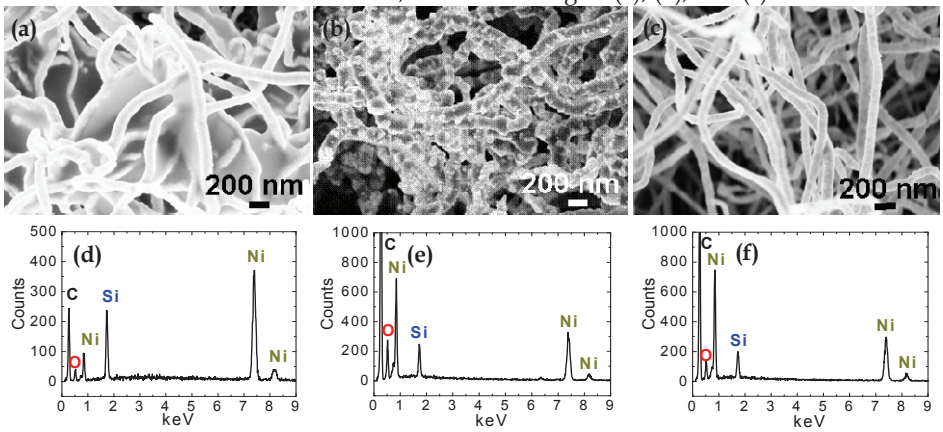


Fig. 10. (a), (d) SEM image and EDS spectrum of as-grown SiNWs after Ni deposition. (b), (e) SEM image and EDS spectrum of Ni-deposited SiNWs after slow annealing at 600 °C for 5 min in Ar ambient with oxygen filter. Note the rough NW surface morphology. (c), (f) SEM image and EDS spectrum of Ni-deposited SiNWs after rapid thermal annealing (RTA) at 550 °C for 5 min in N_2 ambient, Note the smooth NW surface morphology.

Energy dispersive x-ray spectroscopy (EDS) was performed for post-annealed Ni-deposited SiNWs on a HOPG substrate for compositional analysis. Figure 10(d), (e), and (f) shows EDS

spectra of SiNWs after Ni deposition, after slow annealing, and after RTA, respectively. These results confirm the presence of nickel on individual SiNWs after thermal processing. Note the substantial increase of the Lα peak (0.85 keV) relative to the Kα peak (7.47 keV) for the annealed Ni-SiNWs. This implies nickel silicide formation (Kim *et al.* 2005; Song *et al.* 2007 ; Lee *et al.* 2004). The EDS data in Fig. 10 also confirm the presence of oxygen in the SiNWs. The Ni decoration on the surface of a nanowire after RTA is also shown in the SEM and TEM images of Fig. 11(a) and (b).

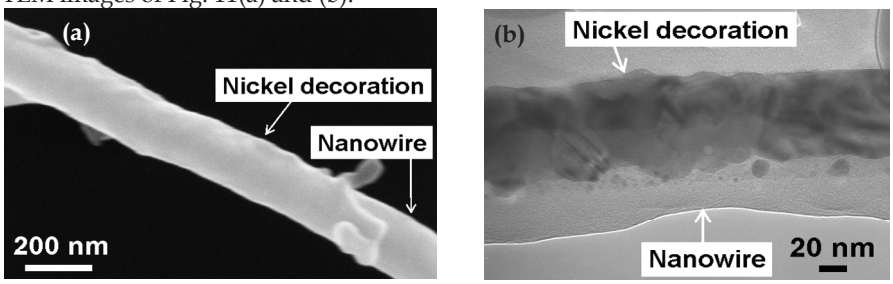


Fig. 11. (a) SEM image of Ni decorated SLS SiNW after RTA on HOPG substrate, (b) TEM image of Ni decorated SLS SiNW after RTA.

5.2. Synthesis, metallization, structural and compositional characterization of VLS SiNWs

Similar to the SLS SiNWs, nickel was deposited on as-grown VLS SiNWs by e-beam evaporation calibrated for an effective blanket film thickness of 150 nm. Post-deposition thermal processing was carried out for nickel silicide formation at 550 °C for 5 minutes via RTA. SEM micrographs in Figs. 12(c) and (d) show that deposited Ni does not coat VLS SiNWs conformally. The asymmetric coating induces mechanical deformation during RTA processing due to thermal expansion differences (Fig. 12(g)). To investigate nickel silicide formation during RTA processing, energy dispersive x-ray spectroscopy (EDS) was performed for as-grown SiNWs (Fig. 12(b)), post-annealed Ni-deposited SiNWs (Fig. 12(f)), Ni-etched nanowires after RTA (Fig. 12(i)), and Ni-etched nanowires without RTA (Fig. 12(k)). Samples for EDS analysis were transferred to a HOPG substrate for compositional analysis. EDS data confirmed that Ni was completely removed during the etching process in the absence of RTA. In contrast Ni was present for samples that underwent RTA. This implies Ni silicide formation at the Ni/SiNW interface since the Ni etchant used does not attack stoichiometric nickel silicide phases. Fast Fourier transform (FFT) diffraction patterns and TEM micrographs of Ni silicided nanowires suggest that they are crystalline (Fig. 13).

VLS SiNWs were also coated with tungsten (2-4 nm thick layer) via an atomic layer deposition (ALD) process. This process was chosen for its ability to provide a conformal coating of W on the high surface area SiNWs. Following W ALD, the samples were annealed in the RTA furnace at 800 °C under a N_2 flow to form W silicide. TEM-EDS was performed on the resulting structures to confirm the presence of W (Fig. 14) although it is not, in principle, possible to distinguish WO_x from WSi_x from this data.

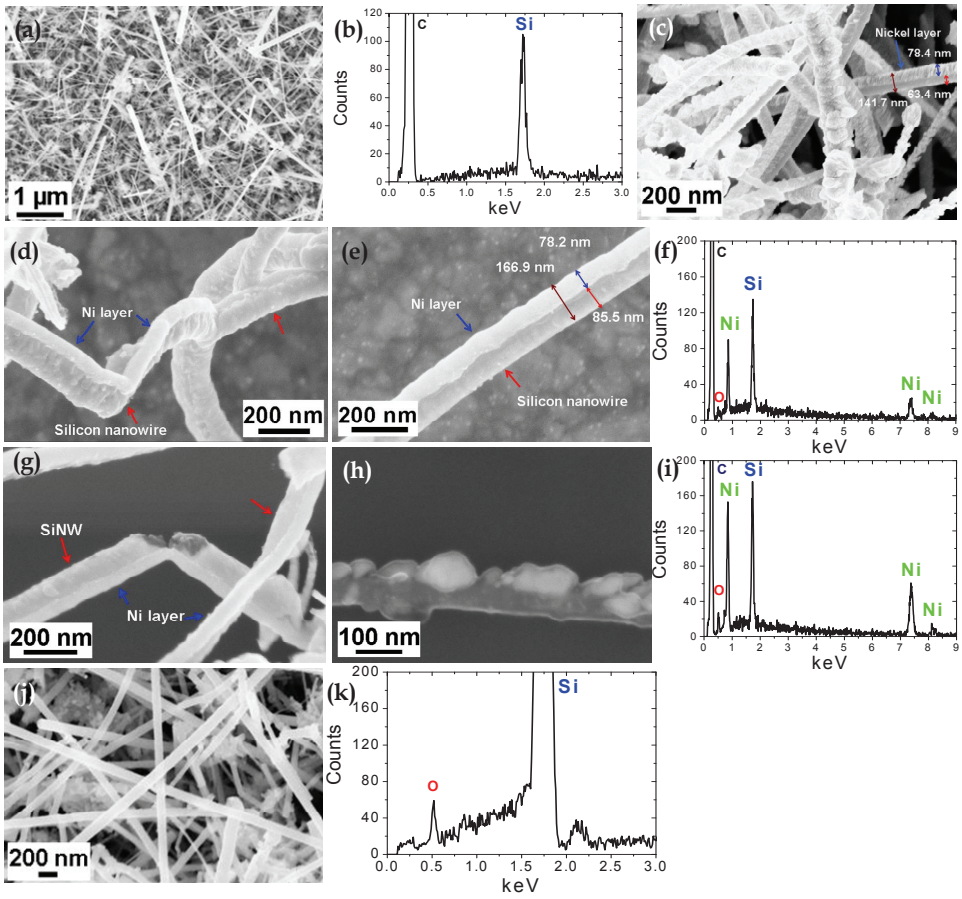


Fig. 12. (a) SEM micrograph of as-grown VLS SiNWs; (b) EDS of as-grown VLS SiNWs; (c) SEM micrograph after Ni coating; (d) (e) and (f) SEM images and corresponding EDS data for VLS Ni-SiNWs after RTA; (g) (h) and (i) SEM images and corresponding EDS data for post-RTA Ni-SiNWs followed by Ni etching; (j) and (k) SEM and corresponding EDS data for Ni-SiNWs after Ni etch (no RTA); (i) and (k) imply the presence of $NiSi_x$.

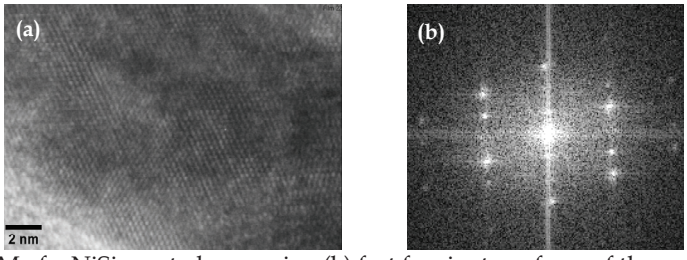


Fig. 13. (a) TEM of a $NiSi_x$-coated nanowire, (b) fast fourier transform of the same nanowire

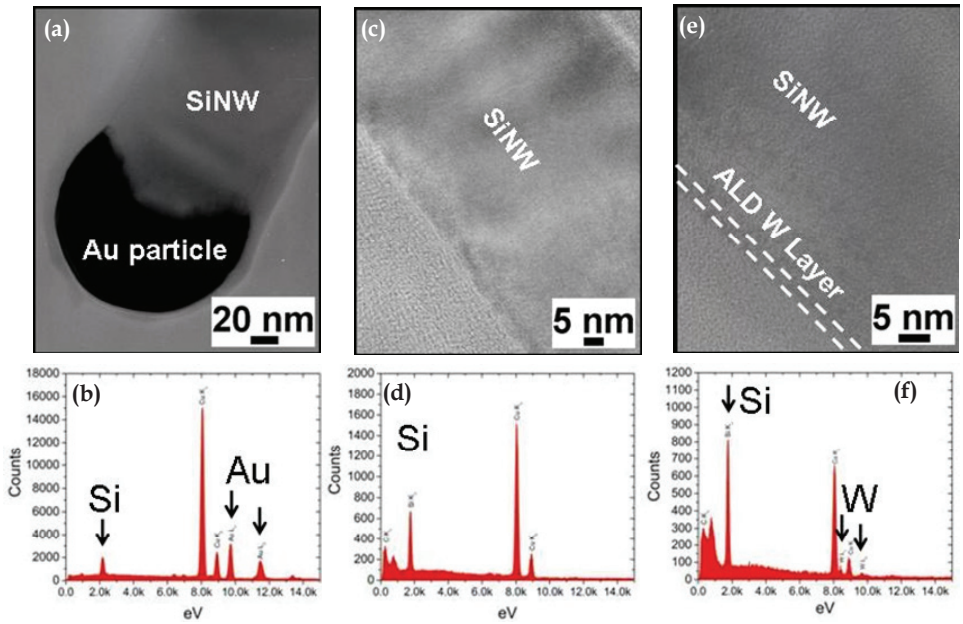


Fig. 14. (a) SEM image of Au catalyst particle at the head of a VLS SiNW, (b) corresponding EDS spectrum (c) SEM image at the center of the nanowire, (d) corresponding EDS spectrum, (e) SEM image of the edge of an ALD-W coated VLS SiNW (nominal W thickness ~ 2-4 nm), (f) corresponding EDS spectrum confirming presence of W.

5.3. Electrical transport of as-grown and metal-coated SiNWs

As-grown silicon nanowires via SLS and VLS and metal/metal-silicide coated SiNWs (both SLS and VLS) were dispensed on SiO_2-coated Si wafers patterned with a metal (Au) electrode pattern to carry out electrical conductivity measurements. Electrical contacts between the dispensed SiNWs and the Au electrodes were formed via direct-write FIB-based Pt deposition. Figures 15(a) and 15(c) display a SEM micrograph and associated current-voltage characteristic, respectively, of an as-grown SLS SiNW. The I-V response of the SLS SiNW is linear and the two-point resistivity of similar SiNWs ranged from 20 Ωcm to 2×10^5 Ωcm. The observed electrical resistivity range for as-deposited SiNWs rules out a dominant SiO_2 stoichiometry, although it is not inconsistent with local SiO_x compositions within the nanowire. (No current measurement was possible for open Pt electrodes confirming the insulative properties of the SiO_2 dielectric in the electrical test structure).

The resistivity of thermally processed Ni-SLS SiNWs (Section 5.1) was measured by two- and four-point electrical resistivity measurements (Fig. 15 (b), (d), (e)). Two- and four-point measurements yield resistances of 2 MΩ and 55 kΩ respectively, for the 9.2 μm long Ni-SLS SiNW with a diameter of 165 nm. Assuming, for the sake of simplicity, that the entire SiNW has been converted to a metal silicide phase the latter value implies an equivalent resistivity of 0.013 Ωcm. This data point is represented by a black circle in Fig. 15 (f). Figure 15 (f) also shows the electrical resistivity distribution of other as-grown SiNWs (blue squares) and Ni-SiNWs (red circles). The lowest measured effective electrical resistivity was 0.02 Ωcm

assuming complete silicidation. (This assumption is discussed in more detail in the following section). After thermal processing of evaporated Ni, the SLS SiNW conductivity is increased up to 7 orders of magnitude.

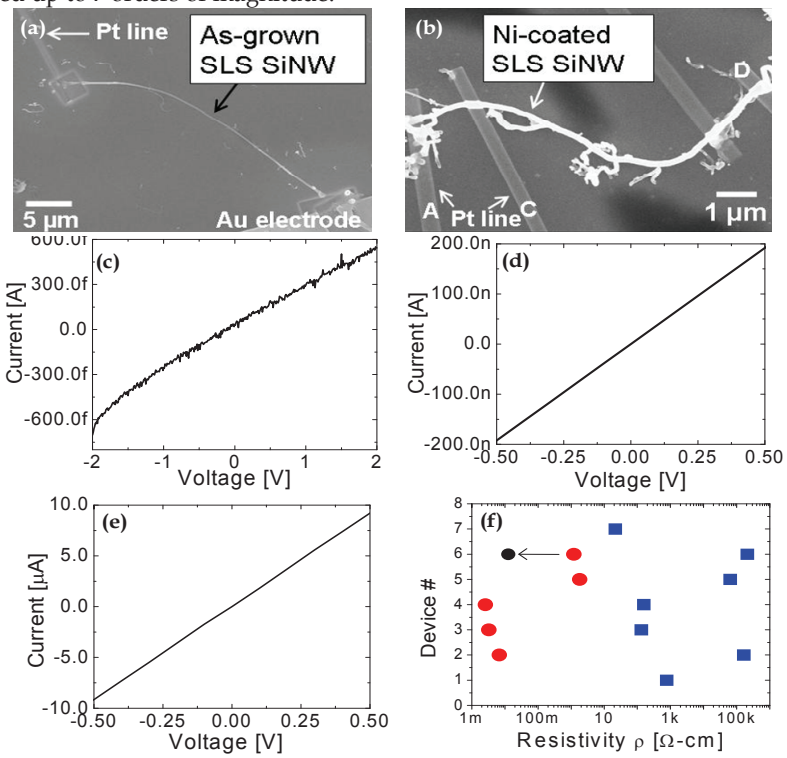


Fig. 15. SEM micrographs of (a) as-grown SLS SiNW in 2-point test structure and (b) annealed Ni-SLS SiNW in 4-point test structure. (c) 2-point I-V data of as-grown SLS SiNW (29.2 µm length, 168 nm diam.), (d) 2-point I-V data of annealed Ni-SLS SiNW (9.2 µm length, 165 nm diam.), (e) 4-point I-V data of annealed Ni-SLS SiNW (9.2 µm length, 165 nm diam.), (f) electrical resistivity from as-grown SiNWs (blue squares) and annealed Ni-SiNWs (red circles). Arrow indicates resistivity drop using 4-point in place of 2-point I-Vs.

Figure 16 shows representative SEM micrographs and associated current-voltage characteristic of an as-grown VLS SiNW, a W-coated VLS SiNW, and a $NiSi_x$ VLS SiNW, respectively. The as-grown VLS SiNW exhibits the expected I-V response for undoped Si. W-coated and $NiSi_x$ nanowires exhibit ohmic conductance (surface). Four-point electrical resistivity measurements were performed on 6 groups of nanowire devices: as-grown VLS SiNWs, $NiSi_x$ VLS nanowires (Ni by e-beam evaporation), ALD W-coated nanowires (2nm W, non-annealed), ALD W-deposited nanowires (2nm W, annealed), ALD W-deposited nanowires (4nm W, annealed), and Ni catalyst SiNWs. Figure 17 shows the electrical resistivity distribution of all 6 groups of these nanowires including as-grown SLS SiNWs and as-deposited Ni-SiNWs as well as as-grown VLS SiNWs. For this plot it was assumed that all Ni-metallized SiNWs were completely converted to metal silicide (i.e. the entirety of

the SiNW cross-section is metallic) and that electrical conduction was confined to the W-region of the W-coated SiNWs. As expected, metallization increased nanowire electrical conductivity by as much as 8 orders of magnitude.

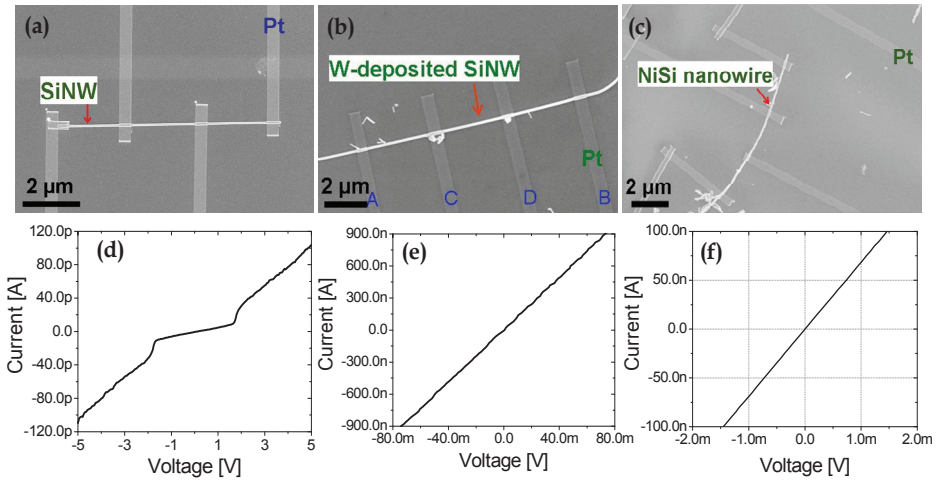


Fig. 16. (a) As-grown VLS SiNW, (b) ALD-W coated SiNW, and (c) $NiSi_x$ SiNW on 4 point electrical test structures. (d) I-V plot of the SiNW shown in (a), (e) I-V plot of the ALD-W coated SiNW shown in (b), (f) I-V plot of $NiSi_x$ nanowire shown in (c)

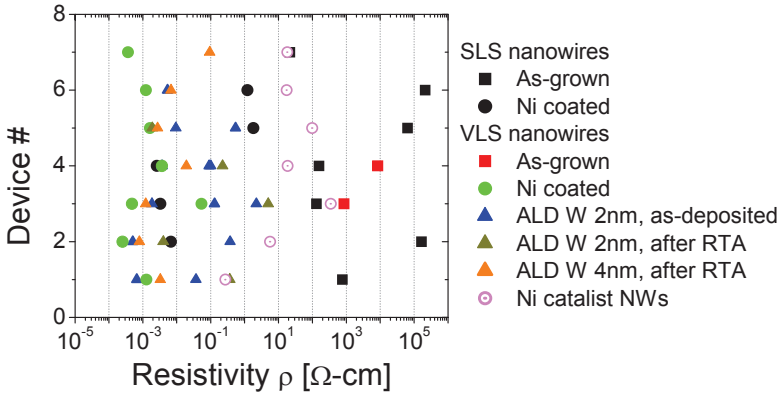


Fig. 17. Comparison of measured electrical resistivity from SiNWs. Device groups include as-grown SLS and VLS SiNWs, Ni-coated SiNWs, $NiSi_x$ nanowires, ALD W-coated VLS SiNWs (2nm W-non-annealed, 2nm W-annealed, and 4nm W-annealed), and Ni catalyst SiNWs.

5.4. Modeling and simulation of electrical transport

To further elucidate the silicidation and metallic conduction of metallized SiNWs analytical and finite-element modeling was carried out. The modeling was restricted to ALD W-coated SiNWs due to the relative uniformity of the W film. Firstly, it was necessary to estimate the

thickness of the post-RTA WSi_2 layer relative to the thickness of a remnant W layer – in the event not all W was consumed during silicidation. Figure 18 schematically illustrates this situation wherein the post-RTA, W-coated SiNW is divided into three distinct radial regions: 1) a single crystalline Si core; 2) a WSi_2 annular region; and 3) an outer W region. Based on simple stoichiometric arguments the annular width of the WSi_2 region depends on the amount of Si and W consumed during the silicidation reaction. Based on bulk Si, W, and WSi_2 unit cell sizes, the thickness of the WSi_2 layer as depicted in Fig. 18 can be calculated as a function of consumed W fraction. These calculations are shown in Fig. 19(a). The overall electron transport across a WSi_2-coated SiNW was simulated using COMSOL Multiphysics finite element analysis software. Figure 19(b) shows current flow through a modeled W/WSi_2-coated SiNW. The current is confined primarily to the WSi_2 region.

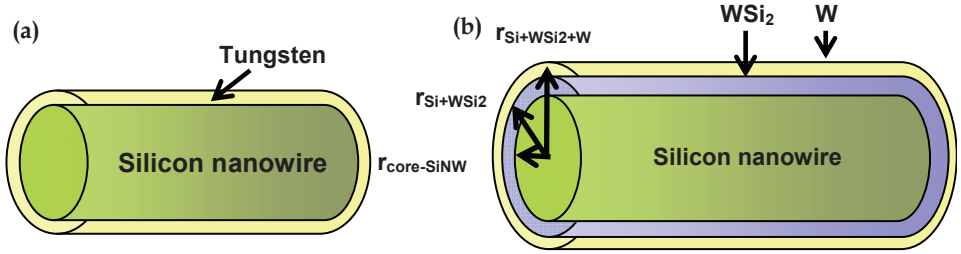


Fig. 18. Schematic of tungsten deposited nanowire (a) before silicidation, (b) after silicidation

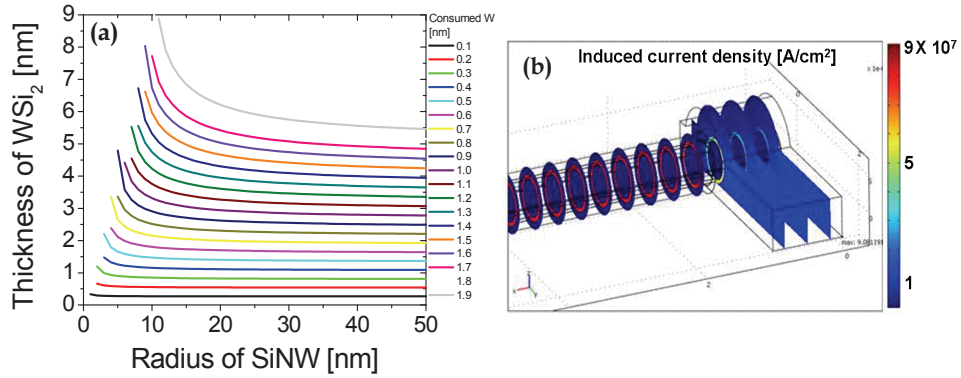


Fig. 19. (a) Thickness (radius) change of the metallic WSi_2 annular region during silicidation for a 2 nm tungsten film as a function of W fraction, (b) representative result of COMSOL Multiphysics simulation. The Si-WSi_2-W structure was used under voltage bias with a metal electrode. Current density was calculated along the nanowire axis. The red circular region denotes the current-carrying annulus.

The resistivity modeling utilized a simple expression to incorporate finite size effects in the W and WSi_2 annular films. Shown in Equation 1, this expression for thin film electrical resistivity is expressly for the case where surface scattering is dominant (Hauder *et al.* 2001; Wissmann *et al.* 2007; Lee *et al.* 2004):

$$\rho = \rho_o + \rho_{GB} + \rho_{SS} + \rho_{SR} \approx \rho_o + \rho_{SS} \approx \rho_o \{1+ (3/16) \bullet l_o/d\} \quad (1)$$

Here, ρ_o is the bulk resistivity, ρ_{GB} is the grain boundary scattering contribution, ρ_{SS} is the surface scattering contribution, ρ_{SR} is the roughness contribution, l_o is the electron mean free path, and d is thickness of the film.

For W, ρ_o and l_o are approximately equal to 8.7 μΩ-cm and 33 nm, respectively (Steinhogl *et al.* 2005). For WSi_2 these values are 46 μΩ-cm and 10nm, respectively (Santucci *et al.* 1998). Therefore, an ideal annular resistivity for the 2 nm W-coated VLS SiNW would be 34. 8 μΩ-cm. The best experimental resistivity of the 2 nm ALD W coating on the VLS SiNW in this measured for this work was 506 μΩ-cm. The order of magnitude difference implies that surface roughness, thickness variation, and, possibly, local compositional variations (e.g. oxidation) have a large effect on the effective W-coating electrical resistivity. For the case of the WSi_2 coatings the best measured experimental resistivity was 910 μΩ-cm for an RTA-process 4 nm ALD W coating on a VLS SiNW. Assuming full silicidation the thickness of WSi_2 was 11 nm which translated to an ideal resistivity (using Eq. 1 via the COMSOL package) of 58 μΩ-cm. As for the case of the W film the order of magnitude difference between idealized model and experiment is attributed to disorder and compositional variation in the WSi_2 coating. Modeling results show most current flows through the WSi_2 shell layer as expected (Fig. 19(b)).

6. Novel Growth Mode of Solid-liquid-solid (SLS) Silicon Nanowires

The growth processes for the SLS Si nanowires described in Sections 4 and 5 were optimized with respect to SLS SiNW diameter control and overall nanowire length for evaluation of Ni-based and W-based metallization, electrical transport, and comparison with VLS SiNWs. Those optimization studies also focused on the very early and intermediate stages of SLS SiNW growth and yielded critical insights regarding the initial nucleation of SLS nanowires. Specifically, a novel and previously unreported, high temperature solid-liquid-solid (SLS) silicon nanowire growth mode has been observed and investigated. In this novel mode SLS nanowire nucleation and subsequent growth was uniquely promoted by – and coupled to – the formation of thermally-etched pyramidal pits in the Si substrate which formed during the high-temperature anneal phase BEFORE the onset of SLS SiNW formation. The silicon-oxide-mediated thermal pit formation process enhanced Si transport to Au-Si alloy droplets directly adjacent to the pyramidal pits. Consequently, SLS nanowire nucleation and growth was preferentially promoted at the pit edges. The catalytic nature of the pyramidal pits resulted in the observation of SLS nanowire 'blooms' at the pit locations. Subsequent nanowire growth – occurring both at the pit sites and from Au-Si alloy droplets distributed across the planar surfaces of the Si wafer – eventually occluded the pits complicating experimental observation of this mode. A rapid ramp of the Ar gas flow introduced to the SLS growth chamber at selected points in the annealing cycle was sufficient to quench nanowire growth and permit experimental observation of this novel mode. This newly observed process is termed 'thermal pit-assisted growth.' Section 6.1 describes the experimental configuration for growth mode analysis of SLS Si nanowires. Section 6.2 documents the experimental observations of the stages of thermal pit-assisted SLS SiNW growth and presents compositional analysis for determination of Au-Si alloy droplet disposition with respect to the pyramidal pits. Section 6.3 presents and discusses a schematic outline of the thermal pit-assisted growth mode.

6.1. Experimental configuration for growth mode analysis of SLS Si nanowires

Central to the observation of thermal pit-assisted SLS SiNW growth was the ability to differentially quench SiNW growth across the sample for comparative SEM analysis. As noted in Section 3 quenching was achieved by positioning Au-deposited samples near the chamber's Ar gas inlet (as compared to the center of the annealing chamber) to exploit the large spatial and temporal temperature gradients. For short annealing durations (typically, below 15 minutes) the temperature near the Ar inlet was rapidly modulated via gas flow. Low Ar flows (~2,000 sccm) introduced upon deactivating the heater yielded a slow cool down (approximately 8 min to reduce chamber temperature below 800 ^{o}C and cease SLS SiNW growth – i.e. latent annealing). High Ar flows (~20,000 sccm) resulted in a rapid quench over tens of seconds and was effective at preserving early stage growth configurations and substantially reducing the duration of latent annealing.

Figure 20 shows a SEM micrograph of a sample adjacent to the Ar inlet within the annealing chamber (low temperature/fast quench region). This single sample exhibited various regions reflecting different stages of SLS Si nanowire growth.

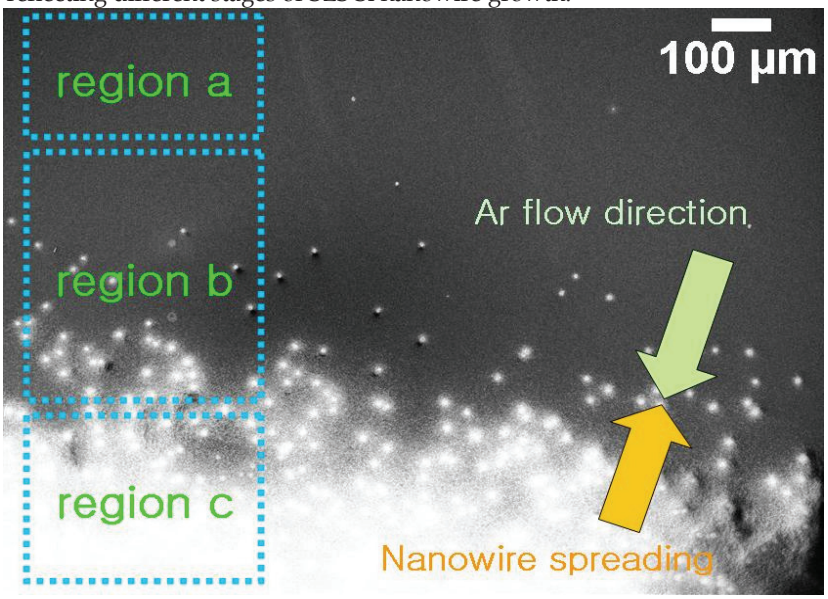


Fig. 20. SEM image of a sample placed close to the Ar inlet inside the oven (low temperature/fast quench region). This sample has three distinct regions in terms of differentiable SLS Si nanowire growth stages.

Three regions were identified. In region (a) Au-Si alloy nanoparticles reside on the surface although SiNW nucleation had not yet occurred (the Au-Si particles are not visible in the magnification shown in Fig. 20). In region (b) so-called 'germinated' Si nanowires could be observed. In region (c) fully-formed Si nanowires were clearly present. Figure 21(a) shows Au-Si alloy particles from the region 'a' depicted in Fig. 20. Little if any SiNW nucleation is evident. The SEM micrographs shown in Figs. 21(b) and (c) were acquired in the 'b' region and clearly capture the initial formation of Si nanowires from droplets wherein the amorphous Si nanowire can be seen emerging or 'germinating' from the droplets. Flower-

like structures were observed (Figs. 21(d), (e)) from the 'c' region in which dense SLS SiNW nucleation was localized. SEM micrographs from deeper within the 'c' region also show dense mats of nanowires from the flower-like structures that had spread out and locally covered the substrate (Fig. 21(f)). The morphology of this latter feature is more consistent with SLS SiNW morphologies observed from regions of the substrate located at the center of the chamber. SLS nanowires originating from the flower-like structures were substantially greater in length and density than more isolated SLS Si nanowires nucleating nearby. *On closer inspection, the flower-like nanowire clusters were seen to originate from pyramidal pits in the Si substrate. This growth mode has not been heretofore observed for SLS Si nanowires.*

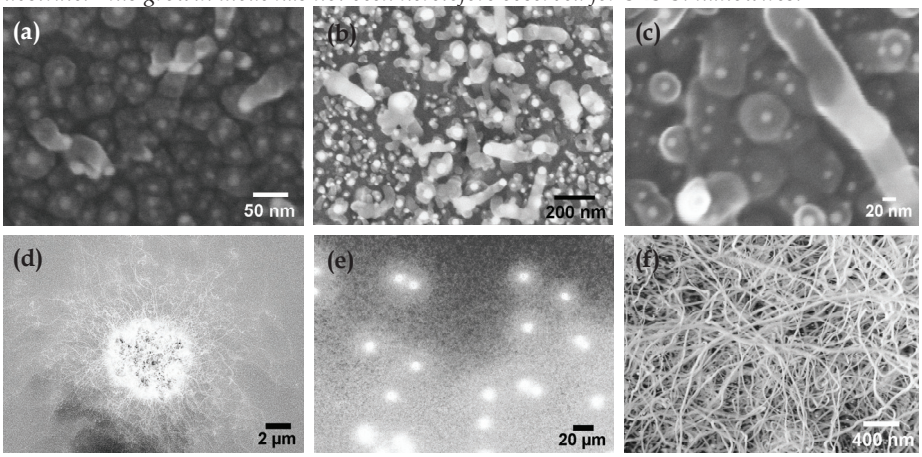


Fig. 21. SEM images of SiNWs at various regions on a single sample in growth. (a) Au-Si alloy particles, (b) and (c) germination of SiNWs, (d) and (e) flower-like nanowire nucleation regions, (f) normal morphology of nanowires.

6.2. Experimental observation of a novel SLS Si nanowire growth sequence

Before elaborating on the observation of the 'thermal pit assisted growth mode' it is important to differentiate it from previously reported SLS growth processes. In SLS SiNW growth the Au-Si alloy droplet is generally observed at the bottom of the nanowire on the surface of the substrate. It has been previously observed that SiNWs can nucleate from alloy droplets inside pits on the silicon surface – pits that had been created by the inflow of the silicon to the alloy droplet (Paulose *et al.* 2003; Sekharet *et al.* 2008; Kim *et al.* 2008). This type of growth was also observed for the work reported here. An example is shown in Fig 22(a). SLS Si nanowire growth can also proceed without necessitating pit formation (Wang *et al.* 2006; Elechiguerra *et al.* 2004). Such a process was also observed for the work reported here (Figs. 22(b), and (c)). These conventional processes (growth modes) are typically used to explain the growth mechanism of SLS nanowires.

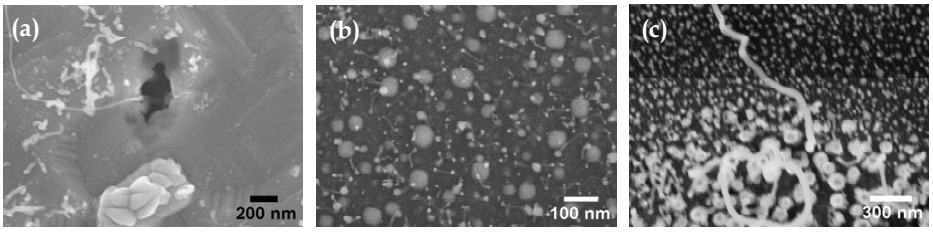


Fig. 22. SEM images of: (a) nanowire and pit created by the inflow of the silicon to the alloy droplet, (b) nanowires directly grown from the substrate and nanoparticles, (c) nanowire germination and growth.

However, these growth modes are distinct from the sequence depicted in Fig. 21. Due to the spatial and temporal thermal gradients afforded by sample placement in the chamber and the rapid Ar gas flow, a spatially-resolved 'timeline' for this novel growth process was documented via SEM imaging at appropriate positions on the substrate. A clear and more detailed series of micrographs documenting this thermally-quenched sequence of SLS Si nanowire growth steps is shown in Fig. 23.

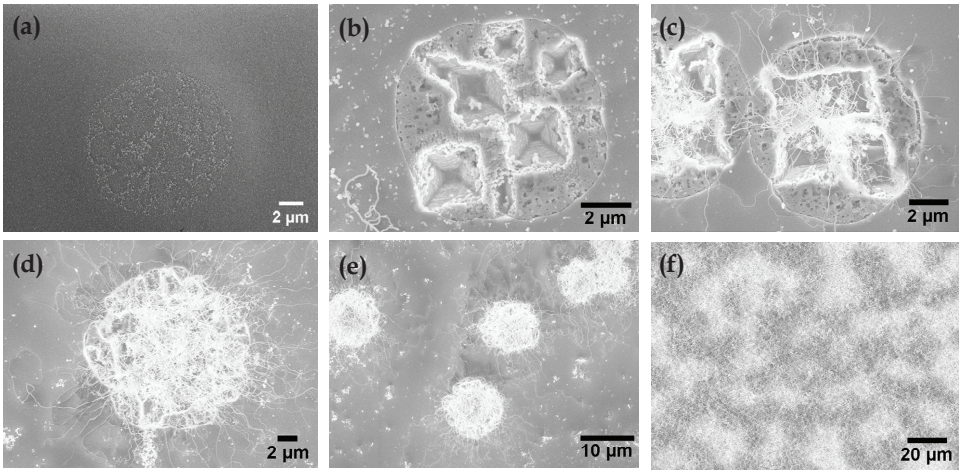


Fig. 23. SEM images of nanowire synthesis steps via thermal pit assisted growth, (a) formation of 'embryos' on Si substrate, (b) thermal pit formation - the pyramidal pits are created by a thermal etching process, (c) initial nanowire growth between the pits, (d) coverage of pyramidal pit by SLS SiNWs (e) nanowires covering multiple pits, (f) nanowires coalescing from neighboring pyramidal pits to cover the substrate.

Figure 23(a) shows the first step in thermal pit assisted SLS SiNW growth – a circular region approximately 10 µm in diameter of local volume discontinuity (swelling) developed in the Si wafer substrate. Regions of the substrate which saw lower temperatures did not exhibit these features. The volume discontinuity is attributed to local oxidation of the Si substrate. Similar localized volume increases have long been observed during thermal oxidation processing of Si (section 6.3). Figure 23(b) shows the next step in this novel SLS Si growth

mode – thermally-induced formation of pits in circular oxidation regions. These pits were pyramidal in shape with sidewalls comprised of intersecting {111} planes in the (100) silicon wafer. It was observed for multiple samples and experimental runs that the pyramidal pits in the Si substrate were formed *prior* to SLS Si nanowire growth. Figures 23(c) and 23(d) show SiNW nucleation and growth which occurred preferentially at the pyramidal pits. Figures 23(e) and 23(f) shows coalescence of SLS SiNWs from the pyramidal pit regions.
Figure 24 affords closer inspection of the pyramidal pit area at the onset of SLS SiNW nucleation. Firstly, Au-Si alloy droplet concentration is localized at the edge of the pyramidal pits (Fig. 24(a)). Secondly, the silicon oxide growth front defining the circular region encompassing the pyramidal pit is clearly visible (black arrows in Fig. 24(b)). The silicon oxide exhibits a granular or porous structure consistent with the local oxidation-induced volume change.

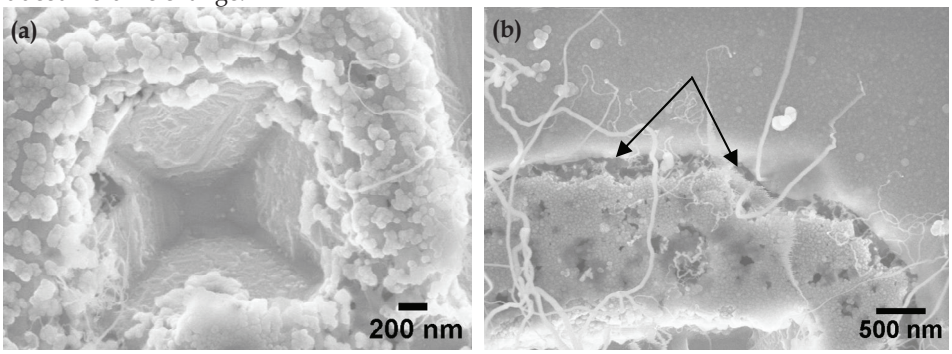


Fig. 24. SEM images: (a) the center of a pit showing Au-Si alloy particles on the edge of the pit; (b) the border of the pit area. In (b) The growth front separating the crack oxidized region from the Si surface is clearly visible (black arrows).

The material disposition associated with the regions in and around the pyramidal pits shown in Figs. 23 and 24 was investigated via SEM-based energy dispersion spectroscopy (SEM-EDS). Representative SEM-EDS spectra acquired in and near a cluster of pyramidal pits are shown in Fig. 25. Figure 25(a) shows a SEM-micrograph of the pit cluster and positions where EDS spectra were acquired. An EDS spectrum acquired from the bright region at the edge of the pits is shown in Fig. 25(b) and confirms the presence of Au, Si, and O, as expected for the presence of Au-Si alloy droplets (see Fig. 24(a)) and SiO_x. The lower contrast region between the pits (but within the circular oxidized region) reveals no Au (Fig. 25(c)) and implies that Au-Si alloy droplets congregate preferentially at the pit edges (although the spectrum confirms the presence of oxidized Si). Away from the circular region confining the pits, EDS of the bare Si substrate does show smaller amounts of oxygen (Fig. 25(d)) which is attributed to the native oxide. Interestingly, EDS data from the very center of a pit (Fig. 25(e)) shows no appreciable Au or O. Trace amounts of O are observed further up the pit walls, closer to the pit edge (Fig. 25(f)). Combined with the SEM image data of Figs. 21, 23 and 24, the EDS data of Fig. 25 enabled the formulation of a SLS SiNW growth mechanism promoted by the pyramidal pits. An elaboration of this mechanism is provided in the following section.

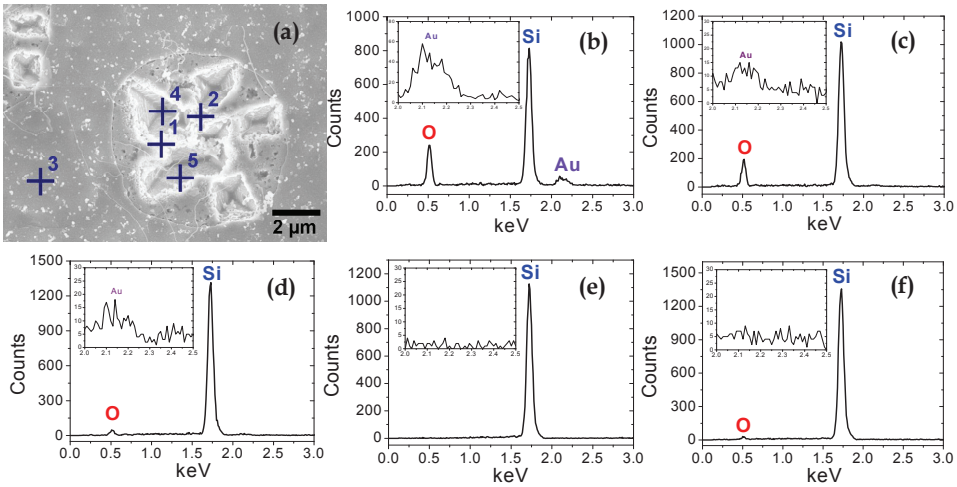


Fig. 25. SEM-EDS spectra at a thermal pit cluster. (a) SEM image of the pit cluster; (b) EDS spectrum at a high-contrast region near a pit-edge (position 1); (c) EDS spectrum at a low-contrast region between pits (position 2); (d) EDS spectrum at the silicon substrate away from the pit cluster (position 3); (e) EDS spectrum at the center of a pyramidal pit (position 4); (f) EDS position at the sidewall of a pyramidal pit near the edge (position 5).

6.3. Mechanistic description of thermal-pit-assisted SLS Si nanowire growth

Based on extensive SEM micrograph and SEM-EDS analysis (consistent with data shown in Figs. 21, 23, 24, and 25) of thermal pit assisted SLS SiNW growth it is concluded that a four-step sequence typifies this growth mode: (1) Nucleation/growth of circular silicon oxide regions; (2) thermal-driven pyramidal pit formation; (3) enhanced Au-Si alloy droplet formation at pyramidal pit edges; and (4) onset of SLS SiNW growth from enhanced Au-Si alloy droplets.

The first step in the aforementioned sequence is consistent with previous observations of the decomposition and removal of SiO at high temperatures (1100 °C) from Si wafers, leaving nearly circular regions several microns in diameter (Rubloff *et al.* 1986; Tromp *et al.* 1985). In this process, solid state SiO_2 reacts with solid Si atoms to form volatile SiO groups. This reaction and the subsequent release of SiO are consistent with the circular regions shown in Figs. 23 and 25 and the porous microstructure within these regions seen in Fig. 24(b).

The oxidation kinetics in the circular regions are also consistent with pyramidal pit growth at high temperatures (> 1000 ^{o}C). As noted in Section 6.2, Fig. 23(b) indicates that the thermally-etched pyramidal pits are formed along (111) planes from the (100) silicon wafer in agreement with observations of other research groups (Ueda *et al.* 2004; Reisman *et al.* 1990; Suzuki 2000a; 2000b). Trace amounts of oxygen enhanced this thermal etching process at high temperatures (Reisman *et al.* 1988, Reisman *et al.* 1990) with pits originating at Si substrate dislocation sites (Yazdi *et al.* 2007). This supports our observation of pyramidal pit growth in the circular oxidized regions of the Si substrate (Figs. 23(a) and 23(b)). The source of the oxygen is attributed to solid state SiO_2 (Reisman *et al.* 1988), not gas-phase oxygen from the chamber. Indeed, research has shown that pyramidal pit formation occurred preferentially under low oxygen partial pressure (below $3x10^{-5}$ atm) during annealing in Ar

(Suzuki 2000a; 2000b; Suzuki 2001). Those researchers observed no pit formation at higher partial oxygen pressures. This is consistent with our experimental configuration. High purity industrial grade Ar (99.999%) was utilized in our experiment with an oxygen filter loop. The estimated partial pressure of oxygen in our system is at least 10^{-6} atm from the Ar gas and does not approach levels close to $3x10^{-5}$ atm even considering other oxygen sources (e.g. quartz tube furnace, sample holder, vacuum excursions, etc…).

The formation of volatile SiO groups as a driving force for the pyramidal etch front is confirmed by the EDS spectrum in Fig. 25(e). No oxygen or Au is evident at the very center of the pit while small trace amounts of oxygen (but no Au) are revealed on the pyramidal pit sidewall nearer the pit edge. The relatively high SiO partial pressure at the pit results in adsorption on Au nanoparticles near the pit edges. This has two primary results. Firstly, Si saturation occurs in the Au-Si alloy droplets at the pit edges comparatively faster than for Au-Si alloy droplets formed on the planar Si wafer surface away from the pyramidal pits. Secondly, oxygen from the SiO species adsorbed on the Au nanoparticles is available to react with the Si substrate thus driving the circular SiO_x growth front and promoting additional pyramidal pits. As the pits cluster, their etch fronts converge, enhancing local concentration of Au-Si alloy droplets leading to a density of SLS SiNW nucleation sites. This speculation is wholly consistent with the observation of a relatively high density of Au-Si alloy nanoparticles adjacent to the thermally generated pits (Figs. 23(c) and 24(a)). Also, Au-Si alloy droplets at the pit edges exhibited, on average, larger diameters as compared to alloy nanoparticles formed at the planar Si wafer surface away from the pits. As a natural consequence, SLS Si nanowires nucleate and grow more rapidly from the Au-Si alloy droplets at the thermally-generated pit edges as compared with Au-Si alloy droplets at the planar Si wafer surface. This speculation is strongly supported by SEM micrographs shown in Figs. 23(e), (f).

It is essential to differentiate the mechanism described above with so-called oxide-assisted growth (OAG). In OAG thermal evaporation or laser ablation is used to evaporate SiO or SiO_2-Si. The resulting SiO vapor decomposes into Si and SiO_2 resulting in a nanowire-like growth front as described in the literature (Zhang *et al.* 2001a; 2001b; Zhang *et al.* 2003). In our experiment, the volatile SiO results specifically from pit formation. Nanowire growth is specifically catalyzed by Au nanoparticles at the pit edges which serve as SiO adsorption sites. The Au nanoparticles simultaneously promote Si nanowire formation through the SLS process *and* maintain local SiO_x growth further promoting pit formation.

This process is described, schematically in Fig. 26. An Au-coated Si wafer is loaded into the chamber. As the sample is heated the Au film dewets to form Au nanoparticles, resulting in Au-Si alloy droplets on the substrate. At selected regions, larger Au-Si alloy droplets are formed via ripening effects. Dangling bonds associated with substrate silicon atoms promote local oxidation (from both low partial pressure gas phase oxygen and solid oxide sources) as shown in Fig. 26(a). Continued oxidation and SiO volatilization lead to the formation of circular SiO_x regions (Fig. 26(b)) and eventually to pyramidal pit etching (Fig. 26(c)). Continued pit etching results in an increase in volatile SiO concentration which promotes relatively faster Si saturation of Au-Si alloy droplets (Fig. 26(d)). This promotes more rapid SLS SiNW growth compared with nucleation on the planar Si substrate. If pyramidal pit density is sufficient a mat-like morphology of thermal pit assisted SiNWs coalesces and forms a dense network or SiNWs.

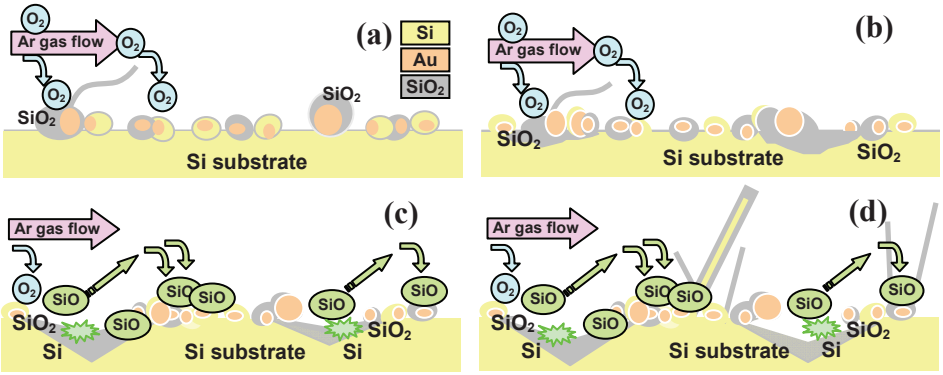


Fig. 26. Schematic illustration of mechanism of thermal pit assisted growth mode, (a) oxidation of germinated silicon, (b) oxidation of silicon substrate, (c) reaction between solid state SiO_2 and Si, resulting in volatile SiO and etched pit. Volatile SiO groups can adsorb to Au-Si alloy droplets at the pit edge. (d) nanowire growth from accumulated Au-Si alloy droplets.

7. Conclusion

Self-assembled Si nanowires were synthesized via SLS and VLS growth processes and characterized as a template for surface metal silicide formation for electron transport. Also, the fundamental kinetics and growth mechanisms of SLS Si nanowires were investigated. The diameter of all nanowires investigated ranged from 5 nm to 180 nm. The diameter of as-grown SLS SiNWs was controlled, to an extent, through the annealing time.

Post-growth Ni deposition and thermal processing was carried out for nickel silicide formation at the SLS SiNW surface. The surface morphology of post-annealed SLS Ni-SiNWs was sensitive to anneal temperature and ramp rate. Rapid ramps resulted in an atomically smooth Ni-SiNW surface morphology. Slow annealing resulted in a rough Ni-SiNW morphology indicative of nonuniform domain formation. The presence of Ni on the surface of SLS SiNWs after RTA processing was confirmed via SEM and TEM imaging. EDS measurements also confirmed that the nickel remains on the SLS SiNW surface after thermal processing. Moreover, the substantial increase of the Lα peak (0.85keV) relative to the Kα peak (7.47 keV) in the EDS spectra for the annealed Ni-SiNWs strongly implies nickel silicide formation. Electrical conductivity measurements were performed on as-grown SLS Si nanowires and metal-silicide coated SLS Si nanowires by two-point and four-point methods. NiSi-coated SLS SiNWs exhibited an improvement in electrical conductivity of several orders of magnitude compared with that of as-grown SLS silicon nanowires.

VLS Si nanowire synthesis was carried out to study formation of surface metal silicide layers. That process utilized a SiH_4/Ar mixture at 500 °C. The diameters of VLS SiNWs were modulated by catalyst thickness and annealing duration. TEM analysis confirmed grown VLS SiNWs were crystalline with a <200> growth direction. TEM-EDS results showed ALD W was successfully deposited on the surface of SiNWs (~ 2 nm, 4 nm thickness). E-beam evaporated Ni was deposited on VLS SiNWs. SEM-EDS confirmed Ni remained on the VLS nanowire surface after etching the excess Ni. This observation strongly implied NiSi formation. Fourier transform analysis of TEM selected-area-diffraction-patterns likewise

implied NiSi formation. Simulation and modeling were performed to investigate electrical transport on W silicide layer. I-V characteristics of W-coated VLS SiNWs and Ni-decorated SiNWs showed an improvement in electrical conductivity (~ 8 orders of magnitude) compared with that of as-grown VLS silicon nanowires. Electrical resistivity measurements confirm ohmic behaviour for electron transport. ALD Ni will be a prospective candidate for silicide formation on SiNW surfaces in the future due to the conformal nature of its deposition and growth.

A previously unreported SLS growth process was observed. For that growth process it is concluded that the high temperature formation of pyramidal etch pits in the Si wafer substrate resulted in local enhancement of Au-Si alloy droplet formation through production of volatile SiO groups from a solid state SiO_2-Si reaction. Saturation of the Au-Si alloy droplets at the pyramidal pit edges due to adsorption of SiO groups occurred more rapidly than Si saturation of Au-Si alloy droplets at the planar Si wafer surface. This resulted in relatively faster SLS SiNW nucleation and growth. This growth mode is termed 'thermal pit assisted growth.'

The authors would like to acknowledge the Semiconductor Research Corporation, DARPA, and the New York State Office of Science, Technology and Research, for their support of this work. The authors would also like to acknowledge Eric T. Eisenbraun, Kun Ho, Rama Kambhampati, Juntao Li, Miguel A. Rodriguez, Daniel Greenslit and Isaac Lund for experimental support and technical discussions.

8. References

Chang, C. & Shen, P. (2000) Thermal-etching development of a-Zn_2SiO_4 polycrystals: effects of lattice imperfections, Mn-dopant and capillary force. *Materials Science and Engineering*, A288, 42–46, 0921-5093

Chen, Y. Q.; Zhang, K.; Miao, B.; Wang, B. & Hou, J. G. (2002). Temperature dependence of morphology and diameter of silicon nanowires synthesized by laser ablation. *Chemical Physics Letters*, Vol. 358, 396-400, 0009-2614

Colli, A.; Fasoli, A.; Beecher, P.; Servati, P.; Pisana, S.; Fu, Y.; Flewitt, A. J.; Milne, W. I.; Robertson, J.; Ducati, C.; De Franceschi, S. ; Hofmann, S. & Ferrari, A. C. (2007). Thermal and chemical vapor deposition of Si nanowires: Shape control, dispersion, and electrical properties. *Journal of Applied Physics*, Vol. 102, 034302, 1089-7550

Cui, Y.; Lauhon, L. J.; Gudiksen, M. S.; Wang, J. & Lieber, C. M. (2001). Diameter-controlled synthesis of single-crystal silicon nanowires. *Applied Physics Letters*, Vol. 78, No. 15, 2214-2216, 1077-3118

Elechiguerra, J. L.; Manriquez, J. A. & Yacaman, M. J. (2004) Growth of amorphous SiO2 nanowires on Si using a Pd/Au thin film as a catalyst. *Applied Physics A*, Vol. 79, 461–467, 1432-0630

Filonov, A. B.; Tralle, I. E.; Dorozhkin, N. N.; Migas, D. B.; Shaposhnikov, V. L.; Petrov, G. V.; Anishchik, V. M. & Borisenko, V. E. (1994). Semiconducting properties of hexagonal chromium, molybdenum, and tungsten disilicides. *Physica Status Solidi (b)*, Vol. 186, No. 1, 209-215, 1521-3951

Futagami, M. & Hamazaki, M. (1982). Thermal etching of 100 silicon surface. *Japanese Journal of Applied Physics*, Vol. 21, No. 12, 1782, 1347-4065

Givargizov, E. I. (1975). Fundamental aspects of VLS growth. *Journal of Crystal Growth*, Vol. 31, 20-30, 0022-0248

Hauder, M.; Gstöttner, J.; Hansch, W. & Schmitt-Landsiedel, D. (2001). Scaling properties and electromigration resistance of sputtered Ag metallization lines. *Applied Physics Letters*, Vol. 78, No. 6, 838-840, 1077-3118

Hochbaum, A. I.; Fan, R.; He, R. & Yang, P. (2005). Controlled growth of Si nanowire arrays for device integration. *Nano Letters*, Vol. 5, No. 3, 457-460, 1530-6992

Kim, G. B.; Yoo, D. J.; Baik, H. K.; Myoung, J. M.; Lee, S. M.; Oh , S. H. & Park, C. G. (2003). Improved thermal stability of Ni silicide on Si 100 through reactive deposition of Ni. *The Journal of Vacuum Science and Technology B*, Vol. 21, 319-322, 1520-8567

Kim, J. & Anderson, W. A. (2005). Spontaneous nickel monosilicide nanowire formation by metal induced growth. *Thin Solid Films*, Vol. 483, 60-65, 0040-6090

Kim, J. H.; An, H. H.; Woo, H. J. & Yoon, C. S. (2008). The growth mechanism for silicon oxide nanowires synthesized from an Au nanoparticle/ polyimide/ Si thin film stack. *Nanotechnology*, Vol. 19, 125604, 1361-6528

Kittl, J. A.; Lauwers, A.; Chamirian, O.; Van Dal, M.; Akheyar, A.; De Potter, M.; Lindsay, R. & Maex, K. (2003). Ni- and Co-based silicides for advanced CMOS applications. *Microelectronic Engineering*. Vol. 70, 158-165, 0167-9317

Lavoie, C.; d'Heurle, F. M.; Detavernier, C. & Cabral, C. (2003). Towards implementation of a nickel silicide process for CMOS technologies. *Microelectronic Engineering*, Vol. 70, 144-157, 0167-9317

Lee, J. H.; Rogers, P.; Carpenter, M. A.; Eisenbraun, E. T.; Xue, Y. & Geer, R. E. (2008). Synthesis and properties of templated Si-based nanowires for electrical transport. *Proceedings of 8th IEEE Conference on Nanotechnology*, 584-587, Dallas, TX, USA, Aug. 2008

Lee, J. W.; Kim, J. K.; Kim, S. H.; Sun, H. J.; Yang, H. S.; Shon, H. C. & Kim, J. W. (2004). Physical and electrical characteristics of physical vapor-deposited tungsten for bit line process. *Japanese Journal of Applied Physics*, Vol. 43, No. 12, 8007-8012, 1347-4065

Lee, K. S.; Mo, Y. H.; Nahm, K. S.; Shim, H. W.; Suh, E. K.; Kim, J. R. & Kim, J. J. (2004). Anomalous growth and characterization of carbon-coated nickel silicide nanowires. *Chemical Physics Letters*, Vol. 384, 215-218, 0009-2614

Li, B. Z & Aitken, R. G. (1985). Electrical transport properties of tungsten silicide thin films. *Applied Physics Letters*, Vol. 46, No. 4, 401-403, 1077-3118

Li, B.; Yu, D. & Zhang, S. L. (1999). Raman spectral study of silicon nanowires. *Physical Review B*, Vol. 59, No. 3, 1645-1648, 1538-4489

Li, C.; Fang, G.; Sheng, S.; Chen, Z.; Wang, J.; Ma, S. & Zhao, X. (2005). Raman spectroscopy and field electron emission properties of aligned silicon nanowire arrays. *Physica E*, Vol. 30, 169-173, 1386-9477

Lu, W & Lieber, C. M. (2007). Nanoelectronics from the bottom up. *Nature Materials*, Vol. 6, 841-850, 1476-4660

Morales, A. M. & Lieber, C. M. (1998). A laser ablation method for the synthesis of crystalline semiconductor nanowires. *Science*, Vol. 279, No 9, 208-211, 1095-9203

Morimoto, T.; Ohguro, T.; Momose, S.; Iinuma, T.; Kunishima, I.; Suguro, K.; Katakabe, I.; Nakajima, H.; Tsuchiaki, M.; Ono, M.; Katsumata, Y. & Iwai, H. (1995). Self-aligned nickel-mono-silicide technology for high-speed deepsubmicrometer logic CMOS ULSI. *IEEE Transactions on Electron Devices*, Vol. 42, No. 5, 915-922, 0018-9383

Murarka, S. P. (1983). *Silicide for VLSI Applications*, Academic Press Inc., New York, 0125112203

Pan, Z. W.; Dai, Z. R.; Ma, C. & Wang, Z. L. (2002). Molten gallium as a catalyst for the large-scale growth of highly aligned silica nanowires. *Journal of the American Chemical Society*, Vol. 124, No. 8, 1817-1822, 1520-5126

Paulose, M.; Varghese, O. K. & Grimes, C. A. (2003). Synthesis of gold-silica composite nanowires through solid-liquid-solid phase growth. *Journal of Nanoscience and Nanotechnology*, Vol. 3, 341-346, 1361-6528

Reisman, A.; Edwards, S. T. & Smith, P. L. (1988). On the thermal etching of silicon. *Journal of The Electrochemical Society*, Vol. 135, No. 11, 2848-2858, 1945-7111

Reisman, A.; Temple, D. & Smith, P. L. (1990). Further comments on the thermal etching of silicon: The surface morphology of (100), (111) and (110) wafers in the temperature range 900°C-1150°C. *Journal of The Electrochemical Society*, Vol. 137, No. 1, 284-290, 1945-7111

Rubloff, G. W.; Tromp, R.M.; Van Loenen, E. J.; Balk, P. & LeGoues, F.K. (1986). Summary abstract: high temperature decomposition of SiO2 at the Si/SiO2 interface. *The Journal of Vacuum Science and Technology A*, Vol. 4, No, 3, 1024-1025, 1520-8559

Santucci, S.; Lozzi, L.; Passacantando, M.; Picozzi, P.; Petricola, P.; Moccia, G.; Alfonsetti, G.; & Diamanti, R. (1998). Studies on structural, electrical, compositional, and mechanical properties of WSi_x thin films produced by low-pressure chemical vapor deposition. *Journal of Vacuum Science and Technology A*, Vol. 16, 1208-1212, 0734-2101

Sekhar, P. K.; Ramgir, N. S.; Joshi R. K. & Bhansali, S. (2008). Selective growth of silica nanowires using an Au catalyst for optical recognition of interleukin-10. *Nanotechnology*, Vol. 19, 245502, 1361-6528

Song, Y. & Jin, S. (2007). Synthesis and properties of single-crystal β3-Ni3Si nanowires. *Applied Physics Letters*, Vol. 90, No. 17, 173122, 1077-3118

Steinhogl, W.; Steinlesberger, G.; Perrin, M.; Sheinbacher, G.; Schindler, G.; Traving, M.; & Engelhardt, M. (2005) Tungsten interconnecs in the nano-scale regime. *Microelectronic Engineering*, Vol. 82, 266-272 0167-9317

Suzuki, T. (2000a). Effect of annealing a silicon wafer in argon with a very low oxygen partial pressure. *Journal of Applied Physics*, Vol. 88, No. 11, 6881-6886, 1089-7550

Suzuki, T. (2000b). Relation between the suppression of the generation of stacking faults and the mechanism of silicon oxidation during annealing under argon containing oxygen. *Journal of Applied Physics*. Vol. 88, No. 2, 1141-1148, 1089-7550

Suzuki, T. (2001). Oxygen partial pressure dependence of suppressing oxidation-induced stacking fault generation in argon ambient annealing including oxygen and HCl. *Applied Surface Science*, 180, 168-172, 0169-4332

Tromp, R.; Rubloff, G. W.; Balk, P. & LeGoues, F. K. (1985). High-temperature SiO2 decomposition at the SiO2/Si interface. *Physical Review Letters*, Vol. 55, No. 21, 2332-2335, 1092-0145

Ueda, K. & Yoshimura, M. (2004). Formation of micromeshes by nickel silicide. *Thin Solid Films*, Vol. 464-465, 208-210, 0040-6090

Wagner, R. S. & Ellis, W. C. (1964). Vapor-liquid-solid mechanism of single crystal growth. *Applied Physics Letters*, Vol. 4, No. 5, 89-90, 1077-3118

Wang, C. Y.; Chan, L. H.; Xiao, D. Q.; Lin, T. C. & Shiha, H. C. (2006). Mechanism of solid-liquid-solid on the silicon oxide nanowire growth. *The Journal of Vacuum Science and Technology B*, Vol. 24, No. 2, 613-617, 1520-8567

Westwater, J.; Gosain, D. P.; Tomiya, S.; Usui, S. & Ruda, H. (1997). Growth of silicon nanowires via gold/ silane vapor–liquid–solid reaction. *The Journal of Vacuum Science and Technology B*, Vol. 15, No. 3, 554-557, 1520-8567

Wissmann, P. & Finzel, H. (2007). *Electrical resistivity of thin metal films*, Springer, Berlin, New York, 3540484884

Wu, Y.; Xiang, J.; Yang, C.; Lu, W. & Lieber, C. M. (2004). Single-crystal metallic nanowires and metal/ semiconductor nanowire heterostructures. *Nature*, Vol. 430, 61-65, 1476-4687

Yan, H. F.; Xing, Y. J.; Hang, Q. L.; Yu, D. P.; Wang, Y. P.; Xu, J.; Xi, Z. H. & Feng, S. Q. (2000). Growth of amorphous silicon nanowires via a solid–liquid–solid mechanism. *Chemical Physics Letters*, Vol. 323, 224-228, 0009-2614

Yazdi, G. R.; Syvajarvi, M. & Yakimova, R. (2007). Formation of needle-like and columnar structures of AlN. *Journal of Crystal Growth*, Vol. 300, 130-135, 0022-0248

Yu, D. P.; Bai, Z. G.; Ding, Y.; Hang, Q. L.; Zhang, H. Z.; Wang, J. J.; Zou, Y. H.; Qian, W.; Xiong, G. C.; Zhou, H. T. & Feng, S. Q. (1998). Nanoscale silicon wires synthesized using simple physical evaporation. *Applied Physics Letters*, Vol. 72, No. 26, 3458-3460, 1077-3118

Zakharov, N.D.; Werner, P.; Gerth, G.; Schubert, L.; Sokolov, L. & Gösele, U. (2006). Growth phenomena of Si and Si/Ge nanowires on Si (111) by molecular beam epitaxy. *Journal of Crystal Growth*, Vol. 290, No. 1, 6-10, 0022-0248

Zhang, R. Q.; Chu, T. S.; Cheung, H. F.; Wang, N. & Lee, S. T. (2001a). High reactivity of silicon suboxide clusters. *Physical Review B*, Vol. 64, No. 11, 113304, 1538-4489

Zhang, R. Q.; Chu, T. S.; Cheung, H. F.; Wang, N. & Lee, S. T. (2001b). Mechanism of oxide-assisted nucleation and growth of silicon nanostructures. *Materials Science and Engineering*, C16, 31–35, 0928-4931

Zhang, R. Q.; Lifshitz, Y. & Lee, S. T. (2003). Oxide-assisted growth of semiconducting nanowires. *Advanced Materials*, Vol. 15, 635-640, 1521-4095

Zhang, X. W.; Wong, S. P.; Cheung, W. Y. & Zhang, F. (2000). Formation and electrical transport properties of nickel silicide synthesized by metal vapor vacuum arc ion implantation. *Material Research Society Symposium*, Vol. 611, C6.5.1

Zhang, Y. F.; Tang, Y. H.; Wang, N.; Yu, D. P.; Lee, C. S.; Bello, I. & Lee, S. T. (1998). Silicon nanowires prepared by laser ablation at high temperature. *Applied Physics Letters*, Vol. 72. No. 15,1835-1

9

Probe Technologies for Micro/Nano Measurements

Kuang-Chao Fan[1,2], Fang Cheng[1], Yejin Chen[1] and Bor-Shen Chen[2]
[1]Hefei University of Technology, [2]National Taiwan University
[1]China, [2]Taiwan

1. Introduction

By definition, the nano-scale ranges from 100nm to 0.1nm. Any new development with size or function falls into this range is named "nanotechnology". The approach of nano-scale research can be in two ways: the bottom-up and the top-down. The bottom-up structure is built up with molecular particles and is the focus of science researches. The top-down approach is to miniaturize visible components from macro- to micro-size and finally downscale to nano-size. This is the way that the engineering technology can discover.

The mission of an engineer is to develop some technologies that can benefit to the industries. Most of the current industrial technologies still remain in the meso- to macro-scale level of the feature size. During the past decade, the MEMS (Micro-Electro-Mechanical System) and energy beam lithography technologies have attracted many researchers to the material processing in sizes from micro to high nano scales (100nm to 10nm). In order to follow the fashion, MEMS has been renamed to NEMS (Nano-Electro-Mechanical System) in recent years. The NEMS, however, can only provide manufacturing process in two and half dimensions (2.5D) through layer deposition, etching, etc. For the true 3D micro parts fabrication the concept of micro machine tools has to be realized. This is the new focus after the nanotechnology in the world, namely the "micro/nano system" (Ni, 2004; Sawada, 2007). A typical example can be found from Sasaki (2007) who successfully used an ultra-precision 5-axis micro-milling machine to cut a complex sculpture BOSATSU model in 2mm size.

It is understood that the metrology should always follow the step of manufacturing to ensure the quality of the products. For any 3D profile the 3D measurement must be employed. Conventional probes for dimensional measurement of parts in macro scale are no more capable for the meso- to micro- sized parts that require accuracy to the degree of 100nm to 10nm. The technology of micro/nano-scale 3D measurement is still a bottleneck for the industry.

This Chapter will address probe technologies for micro/nano measurements. Both of the non-contact and contact types of probes will be reviewed. For the non-contact probe, the principles and applications of focus probe and confocal microscope will be addressed. For

the contact type, the fabrication of micro probe with quality spherical tip will be introduced. A newly developed 3D touch probe for a micro-CMM (Fan, 2006) will be described.

2. Non-contact probes for micro/nano measurements

With the rapid development of micro system technologies and the increasing market need of various micro objects, the measurement system of 3D micro/nano profiles with nanometer accuracy is urgently demanded. Currently, 3D micro/nano structures can be measured by some non-contact techniques to the nanometer accuracy, such as the hologram diffraction method, optical focus probe method, and SPM (Scanning Probe Microscope) methods. They are all sensitive to the reflectivity of the measured surface and thus only limited to a single material at a time. In this section, two types of non-contact probes are introduced. One is the focus probe, which is insensitive to surface material but only allows point measurement. The second is the confocal microscope for area measurement.

2.1 Focus probe

Laser focus probe is based on the astigmatic principle. Its optical path can be expressed by Fig. 1. A light from a laser diode is primarily polarized by a grating plate. Having passed through a beam splitter and a quarter wave plate (mounted on the beam splitter), it is focused by an objective lens onto the object surface as a spot approximately 1 μm in diameter, about 2 mm from the sensor. The reflected beam signal is imaged onto a four-quadrant photo detector within the sensor by means of the quarter wave plate. The photodiode outputs are combined to give a focus error signal (FES) which is used to respond to the surface variation. At the focal plane the spot is a pure circle. When the object moves up or down away from the focal plane, the spot appears an elliptical shape in different orientations. The corresponding focus error signal (FES) provides an S-curve signal proportional to the object movement, as shown in Fig. 2. Previous studies showed that on a single material, such as DVD disk, this probe can accurately measure 3D surface to nanometer accuracy (Fan, 2000; Mastylo, 2004). Products fabricated by NEMS process are, however, mostly built up different materials on the silicon substrate. Like other types of scanning probes, such as the hologram diffraction method or the SPM (Scanning Probe Microscope) methods, non-contact probes always have different characteristics curves with respect to different materials. A solution that utilizes the normalized FES (NFES) technique is proposed to cope with this problem.

When the probe measures a sample surface with high reflectivity, the NFES signal is:

$$S_H = [(A+C)-(B+D)]/(A+B+C+D) \quad (1)$$

If the probe measures a sample surface with low reflectivity, the signals of the four quadrant sensors reduce K times at the same time, and the NFES is:

$$\begin{aligned} S_L &= [(A/K+C/K)-(B/K+D/K)]/(A/K+B/K+C/K+D/K) \\ &= [(A+C)-(B+D)]/(A+B+C+D) \end{aligned} \quad (2)$$

As shown in Eq. (2), the S-curve of sample surface with low reflectivity is the same as the one of high reflectivity. It proves that the probe could measure the 3D profile successfully no matter the surface contains any kind of material.

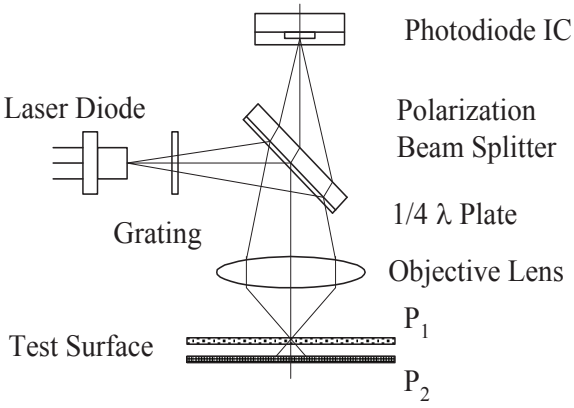


Fig. 1. Optical system of the focus probe

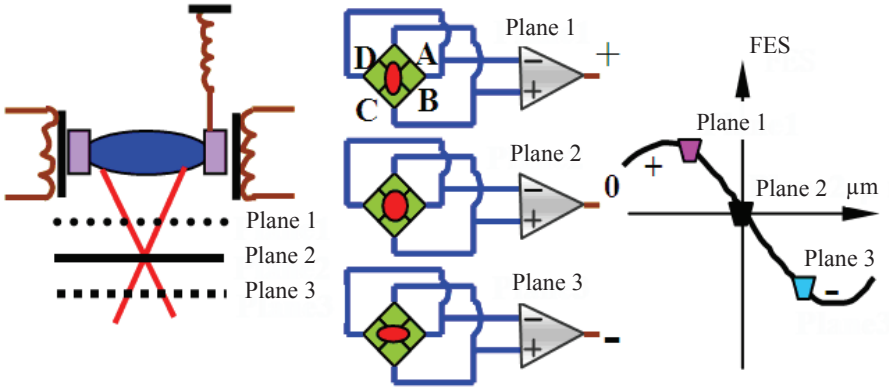


Fig. 2. Focus error signal (FES) with respect to the distance

Calibration tests were conducted on different materials with the focus probe mounted onto a nanopositioning machine, made by SIOS Co. model NMM1 (Fan, 2007). Fig. 3 plots the calibrated FES and NFES curves of different materials. Due to the different reflective ratios each material has independent FES curve. After employing the normalization technique, all the NFES curves are almost the same. Measurement range is about 6μm, resolution is 1nm, and accuracy is about a few nanometers depending on the matrerial. Fig. 4 presents two measured samples, one is a MEMS part and the other one is a PTB step height of 69nm. The first case demonstrates the capability of measuring composite material. The second case verifies the accuracy of this probe.

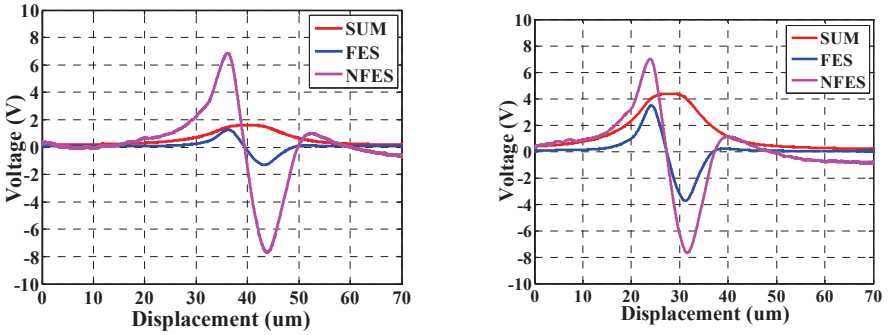


Fig. 3. Calibrated FES and NFES curves for silicon surface (left) and mirror surface (right)

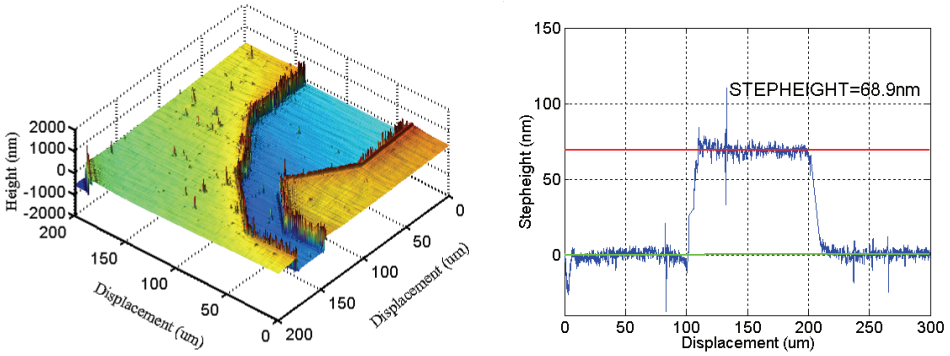


Fig. 4. Measured samples: MEMS Nickel on ALGaN (left) and PTB 69nm step height (right)

2.2 Confocal microscope

In recent years, the technique of confocal microscopy, first described by Minski (1957), has become a more and more powerful tool for surface characterization, in parallel with the development of computer-based image processing systems. The basic principle of confocal microscopy (Minski named it double focusing microscopy) is shown in Fig. 5. Light emitted from a point light source (for example a laser beam focused onto an illumination pinhole) is imaged onto the object focal plane of a microscope objective (the first focusing). A specimen location in focus leads to a maximum flux of light through the detector pinhole (the second focusing), whereas light from defocused object regions is partly suppressed. The variation of the detected intensity is a function of distance change. It, however, can only measure point by point. The white light confocal microscope was then proposed to measure one area profile at a time in conjunction with digital image processing technique. However, in order to detect the whole surface profile one more scanning principle has to be added, such as the point scan by rotating a Nipkow disk (Jordany, 1998) or the structure light projection method with DMD (Bitte, 2001).

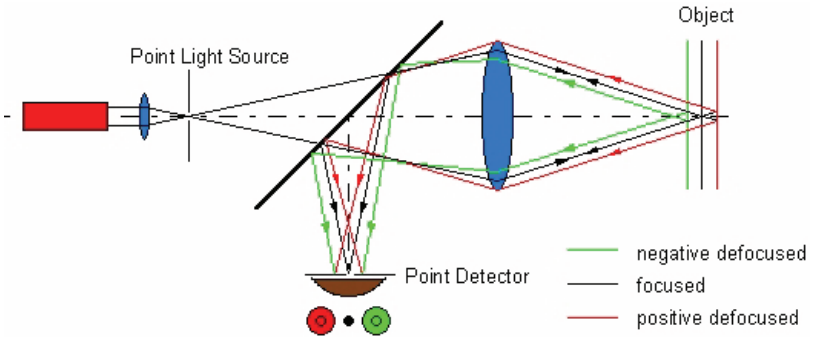


Fig. 5. Principle of laser confocal microscopy

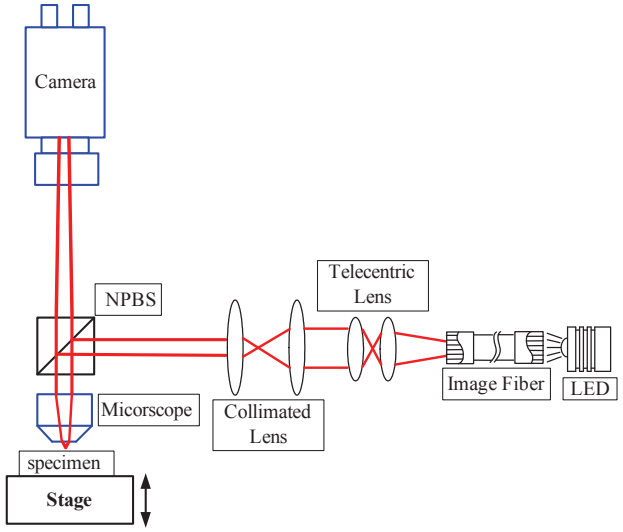


Fig. 6. Image fiber LED confocal microscope

A low cost confocal microscope probe for area detection has been developed (Fan, 2007). The system principle is shown in Fig. 6. It uses a high intensity LED to project light through an image fiber bundle onto the work surface. The fiber bundle consists of 80000 fibers with about 150μm in diameter each. The hexagonal grid pattern of the fiber bundle projected onto the work surface can be treated as a structural light. In the collected image frame the grey level change of the pixels is proportional to the distance out of focal plane of the probe. Fig. 7 shows an example of the photo spacer profile measurement on the LCD color filter plate. The result is the same as the one measured by a commercial white light interferometer. The thickness of the color filter's RGB film can also be obtained in the same graph. The accuracy of this fiber bundle confocal microscope relies on the accuracies of the positioning stage and the characteristic curve of the grey level vs. focal distance. Current system accuracy is about 100nm. It is low cost and suitable for micro level measurements.

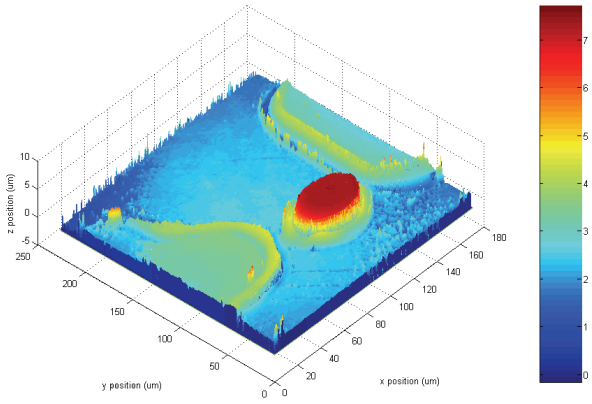


Fig. 7. Measured color filter film and the spacer

3. Contact probe for micro-CMM

3.1 The needs for contact probes

Ultrahigh precision 3D surface measurement technologies have been paid much attention in research during the last ten years. Although many non-contact measurement systems have been developed and commercialized successfully for meso- to micron- or micron- to nano-scaled 3D geometric measurement, such devices cannot cope with the side wall geometry measurement of high aspect ratio micro holes, grooves, and edges. The design and manufacturing of the contact probe becomes one of the critical factors to achieve the measurement capability. In addition, the probe only provides information at the moment of contact to the object being measured. To measure the full scale of any micro part, the object has to be precisely moved by an ultraprecision 3-axis positioner. Combining the contact probe system with the positioning system is generally called the coordinate measuring machine (CMM). Conventional CMM, as a versatile dimensional metrology tool, can only measure macro- to meso-scaled parts because of the size limitation of the probing system. The system design and integration of a contact type micro/nano-scaled three dimensional coordinate measuring machines (3D CMM) (Fan, 2006) has become increasingly important. This kind of micro-CMM requires higher measurement accuracy and resolution than conventional macro-scale 3D CMM.

Several researches have developed micro- or nano-CMM that can measure meso- to micro-scaled parts in nanometer resolution, mostly with specially designed probe systems. The principle of touch-sensing mechanism can be categorized to two types, namely the touch-triggered probe and the touch-analog probe. The triggering probe only detects the moment of contact and outputs a triggered signal to lock the current position displayed by the CMM. A typical example of this type can be found in the Mitutoyo Corporation (UMAP 130). It is based on the detection principle using a vibrating element whose amplitude is reduced on contact with a surface (Masuzawa, 1993). This effect is independent of the approach direction, so true 3D measurements are possible. The touch-analog type is designed with moveable suspension plate or frame in the mechanism and the motion can be detected by built-in sensors. A variety of such probe systems have been designed, such as the silicon-

based (Haitjema, 2001), flexture structure-based (Kung, 2007), and NPL probe (Peggs, 1999). More details can be found in (Weckenmann, 2006).
The probing ball has to be made as small as possible. Most of the above-mentioned probes produce the stylus by gluing a commercially available micro sphere onto a metal stem. Such a process will inevitably incur the offset error of the ball to the center of the stylus. Even though many probe systems announce the measuring uncertainty from 50nm to 20nm, the technique of probe radius compensation has never been mentioned.
A new analog contact probe is designed and a ptototype is made in this study. This contact probe is composed of a monolithic fiber stylus with a ball tip, a floating plate and focus sensors. The ball tip is fabricated using optical fiber with melting and solidification processes, which can ensure the required diameter, roundness and center offset to the tolerance less than 1μm. The design principle and calibration of the innovative probe are explained in the follwing sections.

3.2 Fabrication of stylus

A technique of fabricating monolithic probe stylus with melting and solidification processes of a thin glass fiber to form a micro-sphere tip has been developed by the authors (Fan, 2006). The single mode (SM) glass fiber with 125μm diameter is selected to manufacture the ball tip. Glass fiber has sufficient mechanical strength to perform contact measurements. It can be easily fused within an appropriate heating field. The fiber tip absorbs the arc discharging power and melts instantaneously. The is the principle of fiber splicing, as shown in Fig. 8. Due to the surface tension, the melting part of the fiber starts to form a spherical tip gradually during solidification. The FITEL S199S model single mode fiber fusion splicer is the basic apparatus employed in this experiment, as shown in Fig. 9. The principle of the fiber splicerA fabricated fiber is mounted onto a V-groove fiber holder and the fiber fixture is precisely positioned by an XY stage. With proper selection of the process parameters in the splicer the end face of the glass fiber can be melted and then solidified to a micro-sphere due to the surface tension phenomenon. However, the sphere will droop due to the gravity effect. Therefroe, a rotary stage is added to keep rotating the melt fiber while it is solidified in order to compensate the gravity effect.
The quality of the fiber probe can be controlled by process four parameters, including arc power, fusing times, cleaning arc power offset and cleaning time. The optimal fabricating parameters are obtained for getting precise probe diameter, best probe roundness and smallest center offset using the Taguchi method (Fan, 2008). The image of one of the probes viewed at four angular positions is shown in Fig. 10. The corresponding measured results are summarized in Table. 1. All errors can be controlled to within 1 μm.
In practice, the fiber stylus may produce elastic deformation under contact force, which will yield complex mechanics. The stylus has to be hardened. One solution is to insert the fiber stem into an medical injection needle so that the stem can be very rigid. Besides, the tip ball is coated with a chromatic film by deposition process with rotating technique so that the ball can resist the wear.

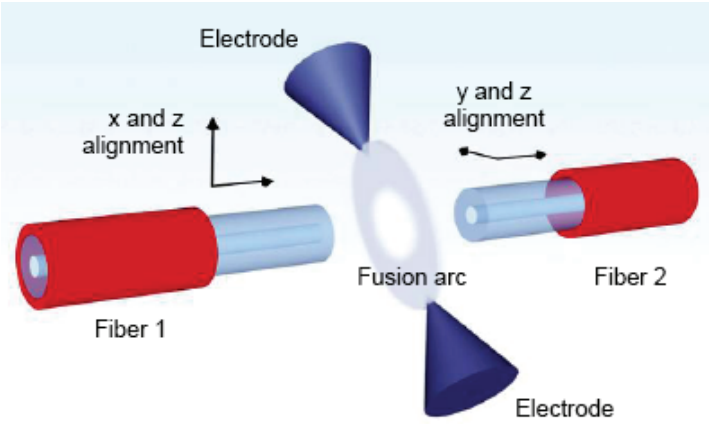


Fig. 8. Principle of fiber fusion splicing

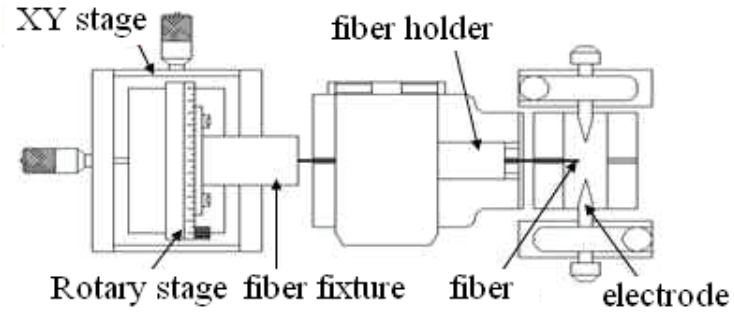


Fig. 9. Experiment setup for fiber tip fabricating

Angle of view	0°	90°	180°	270°
Diameter	314.07	313.84	314.05	313.89
Roundness	0.54	0.82	0.37	0.61
Center offset	0.37	0.96	0.28	0.72

Table 1. Measurement results of the tip at four angles (in μm)

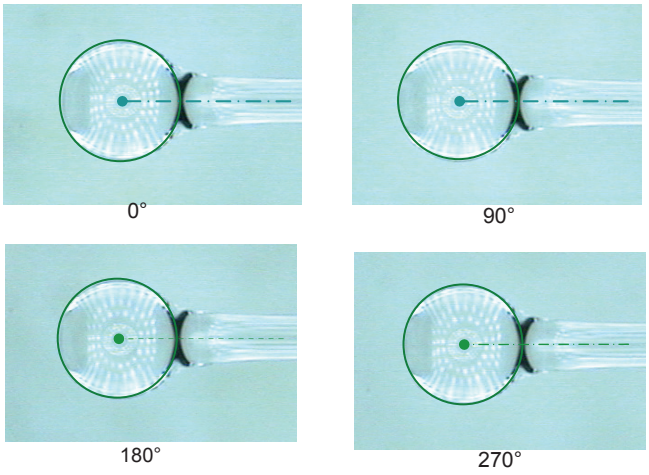


Fig. 10. Image of one fiber probe at different views

3.3 The suspension mechanism

The main task of the suspension mechanism is to give the probe a stable rest position and a tilt angle relative to the contact force in three orthogonal directions. Fig. 11 shows the proposed mechanism of a contact probe. The glass fiber probe stylus is fixed to a floating plate which is suspended by four evenly distributed wires connected to the probe case. The contact force will cause tilt motion of the plate and four wires will perform elastic deformation. Four mirrors mounted onto respective extended arms will amplify the up/down displacement at each mirror position. As shown in Fig. 12, these displacements can be detected by four corresponding focus probes developed in this work, as described in section 2.1. The dimension of the mechanism can be simulated by finite element method to obtain optimum design. Because of the symmetrical geometry, the force-motion sensitivity should be symmetrical in X-Y plane.

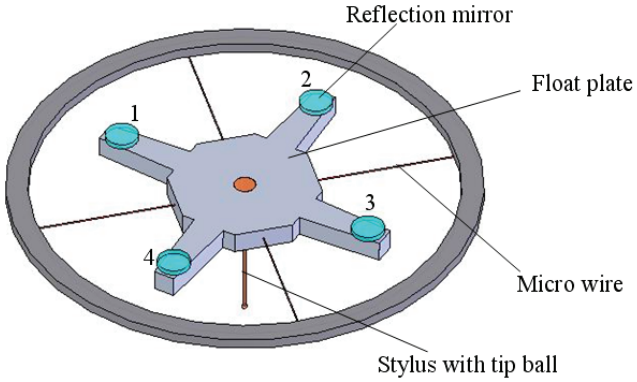


Fig. 11. The suspension mechanism of the contact probe

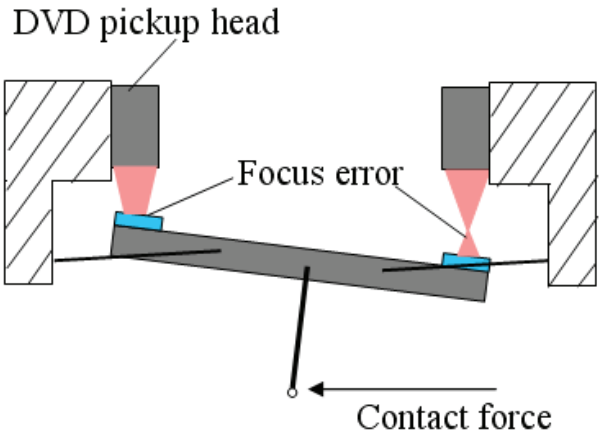


Fig. 12. Structure and motion of a touch-analog probe

3.4 Sensor integration

The sensor used in this touch analog probe is based on the laser focus probe which is reconfigured from the DVD pickup head. Being a mass produced device, the pickup head is cheap and very accurate. The laser focus probe adopts astigmatic principle, as described in section 2.1. The displacement of the reflection mirror will output a focus error signal (FES) from the embedded quadrant photodetector. This FES performs extremely linear curve in the focusing range with a resolution to 1nm. The pickup head is reconfigured with a special design, as shown in Fig. 13. Fig. 14 shows the integration of four sensors in the probe head without the floating mechanism. Since the FES signal is analog to the tip displacement after contact, the combination of four FES signals can reveal the angle and displacement of the probe ball, being a contact scanning mode.

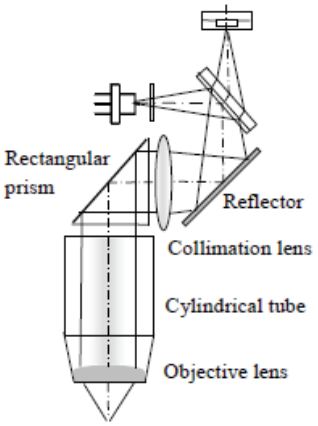


Fig. 13. Reconfigured DVD pick-up head

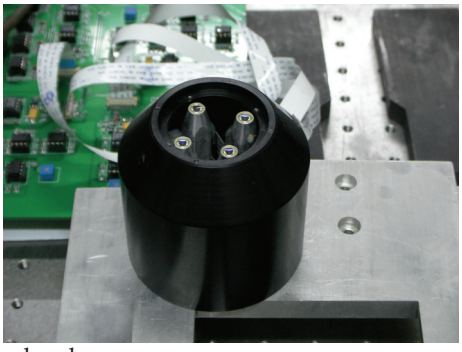


Fig. 14. Photo of the probe head

3.5 Calibration of the contact probe

(1) Contact force calibration

A highly sensitive force measurement appratus has been designed and its schematic diagram is shown in Fig. 15a. A thin leaf plate is built in the base. When a small force is applied to its end, a bending angle can be detected by an angular sensor, which is made from a DVD pickup head based on the principle of autocillimator, as shown in Fig. 15b. This force sensor has been calibrated by some known weights and its linear range of force to voltage is appraent in the range of 1mN. In the experiment, a microscope CCD is used to control the velocity of probe contact, as shown in Fig. 16a. Changing the contact point of the probe ball to its normal direction, the corresponding contact force can be measured. Calibrated results are shown in Fig. 16b from which it can be seen that the forces are nearly balanced in horizontal XY plane. The average contact force calibrated is 109μN.

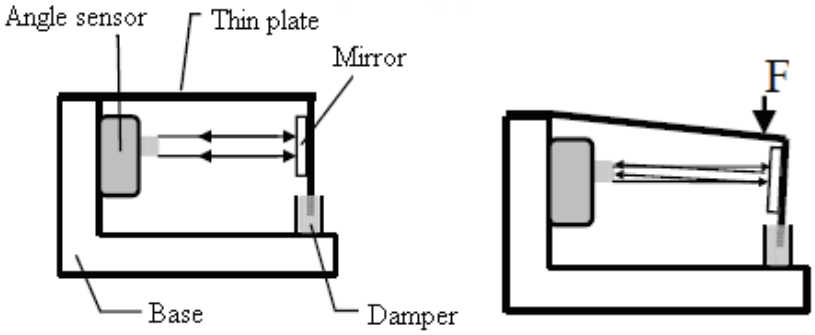


Fig. 15. Force sensor, (a) Schematic diagram, (b) Under load

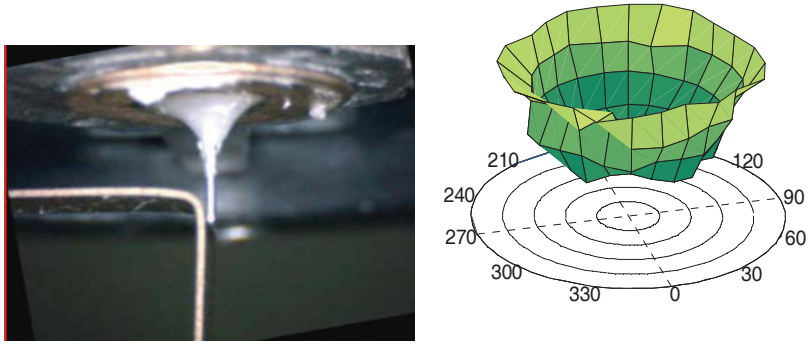


Fig. 16. Contact force calibration, (a) experimental photo, (b) 3D force plot

(2) Contact displacement calibration
The moments of touch, trigger and analog motion of the probe relative to the measured part are different, as shown in Fig. 17. Before contact, the probe is at rest with no output signals from the focus sensors. The trigger point is at the time the sensors detect the tilt motion of the floating plate. The distance between the touch and the trigger points is called the pretravel distance, and distance after the trigger point is called the overtravel or the analog distance. Within the analog range, the sensor output is linear to the displacement. The trigger point is the intersection of two linear lines. The actual displacement can be calibrated by a linear stage with a laser interferometer as the feedback sensor. Repeating this program for ten times, the standard deviation is calculated and the result is 8.9nm.
For the directional tests of contact, 24 directions with 15 degrees separation on the horizontal plane are tested. Fig. 18 shows quite symmetrical result with only 20nm variation at trigger points. Compared to normal 3-point support CMM probe that will appear 3 lobes characteristics, this 4-point support probe exhibits insignificant lobing effect, being insensitive to the contact direction. This contact probe can be equipped to a micro-CMM, which is being developed by the author's group.

4. Conclusion

This chapter addresses the needs of probe technology for micro/nano measurements of meso/micro parts. Three kinds of probes have been developed, namely the focus probe, the confocal microscope, and the contact probe. The focus probe is for pint-to-point measurement of 3D profile featuring high resolution (1nm) and high accuracy (a few nanometers) and insensitive to material property by normalization technique. The confocal microscope using fiber image bundle technique can measure area profile large vertical range but the accuracy is limited to 0.1μm at the present. The contact probe can measure side walls with minimum contact force of 50μN. It has to be equipped into a high precision micro-CMM to achieve 3D measurement with either touch trigger mode or touch analog mode. For different measurement requirements we shall select proper probes.

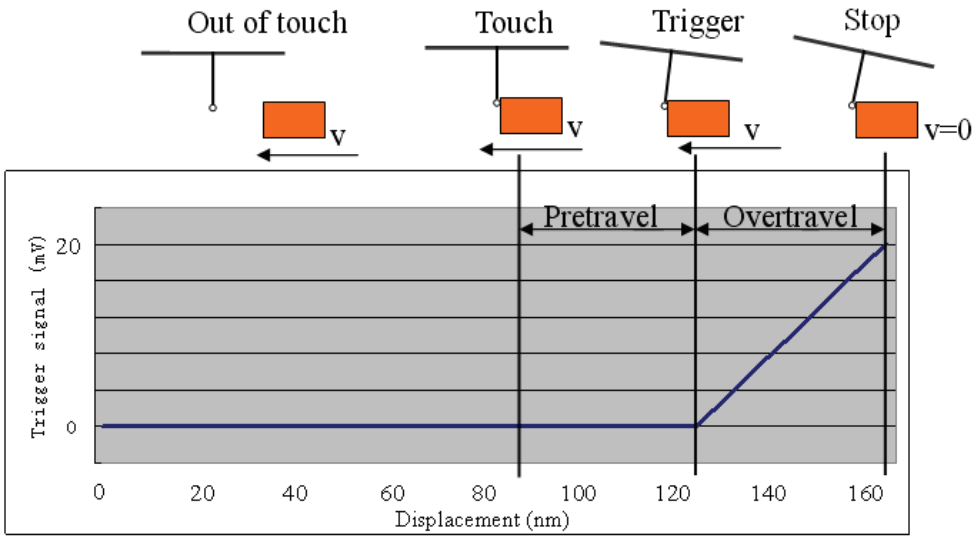


Fig. 17. The output signals before and after contact

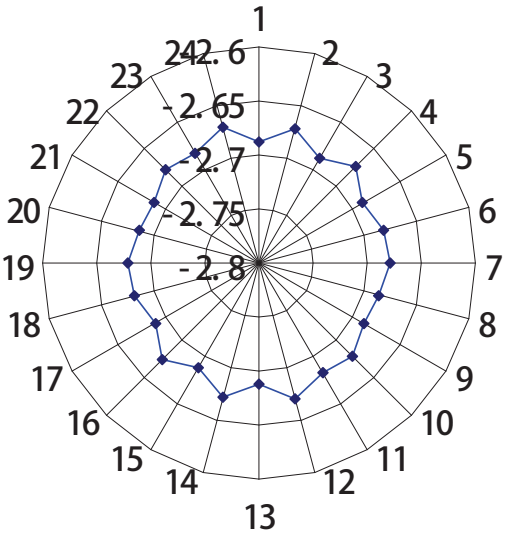


Fig. 18. Probe response to different contact angles

5. References

Minsky, M. (1961). Microscopy apparatus, USA patent 3013467.

Masuzawa, T.; Hamasaki, Y. & Fujino, T. (1993). Vibroscanning method for nondestructive measurement of small holes, *Ann. CIRP*, Vol. 42, 589–92.

Jordany, H.; Wegner, M. & Tiziani, H. (1998). Highly accurate non-contact characterization of engineering surfaces using confocal microscopy, *Measurement Science and Technology*, Vol. 9, 1142–1151.

Peggs, G. N., Lewis, A.J. (1999). Design for a compact high-accuracy CMM, *Annals CIRP*, Vol. 48, 417-420.

Fan, K. C.; Lin, C. Y. & Shyu, L. H. (2000). Development of a Low-cost Focusing Probe for Profile Measurement, *Measurement Science and Technology*, Vol. 11, No. 1, 1-7.

Bitte, F.; Dussler, G. & Pfeifer, T. (2001). 3D micro-inspection goes DMD, *Optics and Lasers in Engineering*, Vol. 36, 155-167.

Haitjema, H.; Pril, W.O. & Schellekens, P. (2001). Development of a Silicon-based Nanoprobe System for 3-D Measurements, *Annals CIRP*, Vol. 50, 365-368.

Mastylo, R.; Manske, E. & Jäger, G. (2004). Development of a focus sensor and its integration into the nanopositioning and nonomeasuring machine, *TM-Technisches Messen*, Vol. 71, No. 11, 596-602.

Ni, J. (2004). Future direction of micro/meso-scale manufacturing, invited speech, Proceedings of the 6th ICFDM'2004, Xi'an, China, June 21-23.

Sawada, K. (2004), Explanation of ultra precision (nano) machine tools, keynote speech, Proceedings *of the 1st ICPT*, Hamamatsu, Japan, June 9-11.

Fan, K. C.; Fei, Y. T.; Yu, X. F.; Chen, Y. J.; Wang, W. L.; Chen, F. & Liu, Y. S. (2006) Development of a low cost micro-CMM for 3D micro/nano measurements, *Measurement Science and Technology*, Vol. 17, 524-532.

Fan, K.C.; Hsu, H.Y.; Hung, P.Y.& Wang, W.L. (2006). Experimental study of fabricating a micro ball tip on the optical fiber, *J. of Optics A: Pure and Applied Optics*, Vol. 8, 782-787.

Weckenmann, A.; Peggs, G. & Hoffmann, J. (2006). Probing systems for dimensional micro- and nano-metrology, *Measurement Science and Technology*, Vol. 17, 504-509.

Fan, K.C., Chen, Y.J. and Wang, W.L. (2007) Probe technologies for micro/nano measurement, Invited paper, *Proceedings of IEEE-Nano*, 989-993, August 2-5, Hong Kong.

Fan, K.C.; Jager, G.; Chen, Y. J. & Mastylo, R. (2007) Development of a non-contact optical focus probe with nanometer accuracy, *Proceedings of the 7th EUSPEN conference*, 278-281, May 20-24, Bremen, Germany.

Küng, A.; Meli, F. & Thalmann, R. (2007). Ultraprecision micro-CMM using a low force 3D touch probe, *Measurement Science and Technology*, Vol. 18, 319-327.

Sasaki, T.; Ishida, T.; Teramoto, K.; Kawai, T. & Takeuchi, Y. (2007), Ultraprecision micromilling of a small 3-D parts with complicated shape, *Proceedings of the 7th EUSPEN conference,* 388-391, May 20-24, Bremen, Germany.

Fan, K.C.; Wang, W.L. & Chiou, H.S. (2008). Fabrication optimization of a micro-spherical fiber probe with the Taguchi method, *J of Micromechanics and Microengineering*, Vol. 18, 0150111, 1-8.

10

Tensile and Flexural Deformation of Nickel Nanowires via Molecular Dynamics Simulations

Ram Mohan and Yu Liang
North Carolina A & T State University
Greensboro, NC, USA

1. Introduction

Metallic nanowires show great potential for applications in minimization of electronic devices due to their extraordinary mechanical strength and electrical properties. Their desirable property characteristics with the smallest dimensions for efficient transport of electrons show potential for use as interconnects and critical devices in nanoelectronics and nano-optoelectronics (Chen, et al., 2006). These metallic nanowires also show potential for applications in electronic packaging, nanoelectronic and nano mechanical devices. A significant issue in the application of these metallic nanowires is their structural strength and stability under mechanical and thermal loading conditions. The deformation behavior of these nanowires under different mechanical loads (for e.g., tensile, bending) is poorly known. Experimental investigations of these behaviors are impractical due to their size and the complications of applying these loading conditions via nano load cells within high resolution microscope systems. Computational techniques based on molecular dynamics (MD) simulations of the representative atomistic configuration of the metallic nanowires provide an effective means of understanding the mechanical deformation behavior of these nanowires.

In this chapter, we discuss the tensile and flexural dynamic deformation behavior of the Nickel (Ni) nanowires due to tensile loading and flexural bending via molecular dynamics simulations. The tensile and flexural deformation behaviors based on the atomistic model configurations of Nickel nanowires are analyzed. The stress-strain constitutive behavior, tensile strength and the Young's modulus for various Ni nanowire configurations are investigated and presented. The natural frequency of the flexural deformation of these nanowires via molecular dynamics simulations is obtained and analyzed. The simulation of the deformation behavior in metallic nanowires modeled as atomistic systems at finite temperatures is a dynamic process and is conducted using classical molecular dynamics.

Focusing on the mechanical behavior of nanowires, it is known that the properties of material configurations at nanometer dimensions can be rather different from those of the bulk material. In the past decades, the rapid progress of miniaturization of electronic devices and nanoscale measurement systems has aroused an interest in nanometer scale materials such as nanowires (Ju, 2004) (Liang, 2003) (Park, 2005) (Silva et. al., 2004), (Silva,

2004) and nanotubes (Liew et. al., 2004). These are not only known for their applications in nanometer scale integrated circuits but also for nanoscale tips of scanning tunneling microscope (STM) (Young, 1966) and atomic force microscope (AFM) (Binnig et.al., 1986). This chapter investigates the tensile and flexural deformation of Nickel (Ni) nanowires using the classical molecular dynamics (Frenkel & Smit, 2001) (Haile, 1997) (Plimpton, 1995), (Rapaport, 1997) (Schlick, 2002) simulation method. Molecular dynamics simulations provide a means to predict mechanical behavior of nano material systems. Experimental techniques for such mechanical behavior of the nano dimensional nanowire configurations, as discussed in this chapter are not only practical but are also difficult.

Prior work in the literature exists on the deformation of the Copper and Gold nanowires (Chang & Fang, 2003) (Ikeda et. al., 1999) (Ju et. al., 2004) (Liang & Zhou, 2003) (Liew et. al., 2003) (Park & Zimmerman, 2005) (Silva et. al., 2004) (Silva, 2004) via molecular dynamics (MD) methods. In this chapter, a nanowire configuration formed from the single crystals of Nickel in the <001> (longitudinal), <100> and <010> (transverse) directions is employed. In particular, this chapter focuses on the mechanical properties of Nickel nanowires by means of molecular dynamics methods. The yielding and the fracture mechanisms (e.g., stress-strain relation of Ni nanowires, which are elongated by a tensile deformation are investigated. The effects of strain-rate, specimen size and ensemble temperature (Chang & Fang, 2003) (Cheung & Yip, 1994) on the tensile deformation behavior are addressed. Most of the current literature is focused on the tensile deformation of nanowire configurations of different materials with very limited and non-existent work on their bending deformation behavior. In addition to this tensile deformation, this chapter also focuses on the flexural deformation of the Nickel nanowire beams under flexural bending conditions via molecular dynamics modeling and simulations.

This chapter is organized as follows. Brief discussions on the theoretical basis of molecular dynamics (MD) simulation methods which in essence are numerically similar to the solution of the multi-body dynamics problems in mechanics are presented. This is followed by a brief discussion on the computation of nano level stresses from the atomic displacement and velocity fields in such MD simulations. These numerical methodologies briefly described in this chapter form the basis of the general purpose molecular dynamics solver (LAMMPS: Large-Scale Atomic/Molecular Massively Parallel Simulator) from Sandia National Laboratory (Plimpton, 2005). The force field interactions between the Nickel atoms are modeled based on the embedded atom potential (EAM). This is followed by the discussions on the analysis of the tensile and flexural bending deformation behavior of Nickel nanowires under various boundary, loading and temperature conditions.

2. Overview of Molecular Dynamics Simulations

Molecular dynamics techniques are generally employed in the simulations of atomistic or molecular level systems. The solution approaches are similar to multi-body problems of mechanics. Computationally, a MD simulation consists of step-by-step numerical solution of the classical Newton equations of motion. The general form of the equations of motion used in MD simulations is given by

$$F_i \equiv -\nabla_i \Pi(X) \equiv -\frac{\partial \Pi(X)}{\partial X_i} = m_i \frac{d^2 X_i}{dt^2} \equiv m_i \frac{dv_i}{dt} \equiv m_i a_i \qquad (1)$$

In this chapter N is the total number of atoms, Π is the potential energy function, and X denotes the positions of atoms; m_i, v_i, and a_i are the mass, velocity and acceleration of atom i. Any molecular level system can be completely formulated by the positions X and velocity V (or momentum P) of atoms. The above equation is similar to the Newton equations of motion employed in continuum mechanics but applied at the atom level. The dynamic behavior of the time-dependent atom motion is computed using an integrator such as the Verlet integrator (Cheung & Yip, 1994) to calculate the trajectories of the atoms. The time-scale involved in the MD simulations is of the order of O (10^{-12} – 10^{-15} sec) and the length-scale is of order O (10^{-10} – 10^{-7} m). Figure 1 shows the general solution flow of the molecular dynamics simulations.

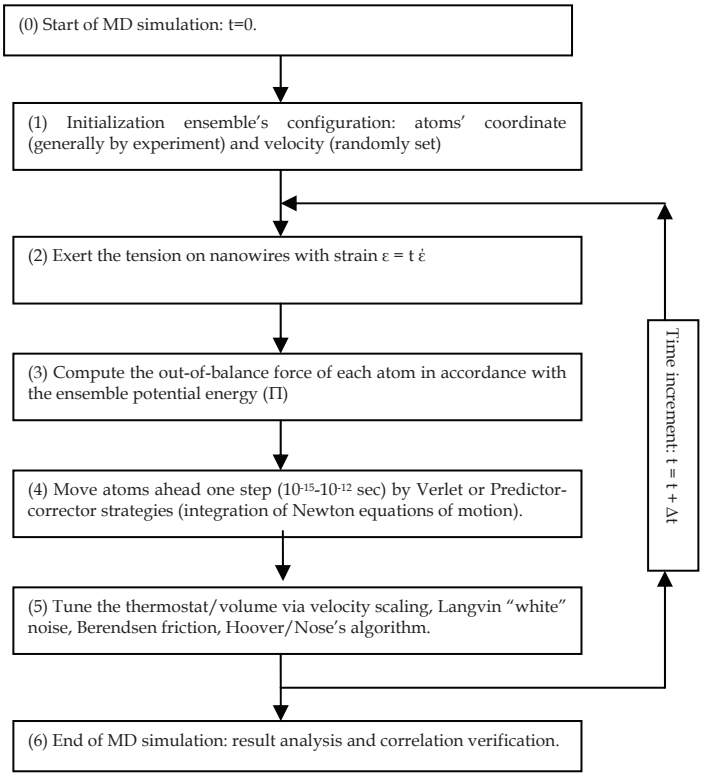


Fig. 1. Solution Flow of Molecular Dynamics Simulation

As illustrated in Figure 1, the numerical solution of Newton's equations of motion (Eq. 1) involved in the MD simulations consists of the following steps:

(1) An initial configuration (the location coordinates and velocity of atoms at time *t=0*). The initial coordinates of the atoms are generally obtained from experiments (Plimpton, 2005). In this work, the atoms of Nickel nanowires are initially configured according to the structure of Nickel crystal (FCC with lattice constant of ג=3.52Å) (Plimpton, 2005). The initial velocity of atoms is usually determined from a random distribution such that the molecular system has a zero initial momentum, as defined by Eq. 2.

$$P = \sum_{i=1}^{N} P_i = \sum_{i=1}^{N} m_i v_i = 0 \tag{2}$$

The velocities of atoms are also chosen randomly using Maxwell-Boltzmann and Gaussian distribution. For a given temperature T, the probability of an atom at velocity v is given by (Frenkel & Smit, 2001) (Haile, 1997) (Schlick, 2002)

$$\varphi(v_i) = \left(\frac{m_i}{2\pi k_B T} \right)^{1/2} \exp\left(-\frac{1}{2} \frac{m_i v_i^2}{k_B T} \right), \tag{3}$$

where k_B is Boltzmann's constant and T is absolute temperature. One of the difficulties in molecular dynamics simulation is calculating the temperature of the system, because temperature is a statistical quantity. However, if there are a large enough number of atoms, a statistical average temperature can be estimated from the *instantaneous temperature* (Frenkel & Smit, 2001) (Haile, 1997) (Schlick, 2002), which can be found from the kinetic energy of the system. This instantaneous kinetic temperature is given by

$$T = \frac{1}{3Nk_B} \sum_{i=1}^{N} \frac{P_i^2}{m_i} \tag{4}$$

This instantaneous kinetic temperature is an indication of the mechanical activity at the atomistic level. This temperature is employed in the simulation studies discussed in this chapter to understand the effect of temperature on the tensile deformation presented in a later section behavior of Nickel nanowires.

(2) Apply an external load/displacement/strain to the initial configuration. In this work, an external axial strain $\varepsilon=\dot{\varepsilon}t$ where $\dot{\varepsilon}$ is a constant strain-rate has been used for the tensile deformation and an external bending loading force is used for the flexural deformation.

(3) The deformation of the system under the action of an external force is computed using the potential energy function. Commonly used potential energy functions are empirical potentials, semi-empirical potentials, and ab-initio (first principles) potentials (Frankel & Smit, 2001) (Haile, 1997). Only empirical potentials are applicable for practical large-scale problems. Embedded atom method (EAM) (Frankel & Smit, 2001) (Haile, 1997) (Plimpton, 2005) (Schlick, 2002) is one such empirical potential that provides a precise and efficient description for metals and metal alloys. This potential is used in this study.

(4) The dynamical behavior of the atomistic system is computed using the potential Π and its gradient $\Delta\Pi$. The time integration gives a snapshot of the atomic configuration (called trajectory) that changes at each discrete time level. No analytical method is applicable for the solution of dynamic system (Eq.1). Commonly used numerical techniques include Verlet, leap-frog, velocity-corrected Verlet, Beeman and predictor-corrector methods (Frankel & Smit, 2001) (Haile, 1997) (Plimpton, 2005) (Schlick, 2002), etc. The MD simulation code (LAMMPS (Plimpton, 2005)) used in this study employs the velocity verlet scheme (described by Eqs. 5-8 below) to simulate the tensile and flexural deformation behavior of nanowires.

$$X(t+\delta t) = X(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \mathrm{O}(\delta t^3) \tag{5}$$

$$v(t+\frac{\delta t}{2}) = v(t) + \frac{1}{2}a(t)\delta t + \mathrm{O}(\delta t^2) \tag{6}$$

$$a(t+\delta t) = -\frac{\nabla \Pi(X(t+\delta t))}{m} \tag{7}$$

$$v(t+\delta t) = v(t+\frac{\delta t}{2}) + \frac{\delta t}{2}a(t+\delta t) \tag{8}$$

(5) During the simulation of dynamic systems, the velocities of atoms need to be updated constantly to make the temperature of ensemble stable. The commonly used consistency strategies are velocity scaling, Langvin "white" noise, Berendsen friction, and Hoover/Nose's (Frankel & Smit, 2001) (Haile, 1997) (Schlick, 2002). Velocity scaling is used to maintain the temperature in this work on the tensile and flexural deformation of Nickel nanowires.

Molecular dynamics is a great tool, but there are limitations. Due to its high computational cost for large time-scale and length-scale problems, molecular dynamics method is only applicable for small scale problems. With the availability of parallel computing systems, LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) (Plimpton, 2005), an efficient parallel MD simulation code developed by Sandia National Lab is employed. A brief description of LAMMPS is presented next.

2.1 Major Features of LAMMPS

The MD simulations require keeping track of the nearby particles of each atom. For computational efficiency, LAMMPS uses Verlet neighbor lists (Chialvo & Debenedetti, 1990) (Frenkel & Smit, 2001) (Rapaport, 1997) to keep track of nearby particles. As an extension of cutoff-radius methods, Verlet neighbor methods set neighbor cutoff radius as potential cutoff plus a "skin". The updating will not be triggered till a molecule travels a distance greater than the skin thickness. In this way, the updating of neighbor list (Chialvo & Debenedetti, 1990) becomes much less frequent. In this work, the skin size is set to be 0.03 (Plimpton, 2005).

On parallel, multi-processor computing systems, LAMMPS uses spatial-decomposition techniques (Plimpton, 2005) to partition the simulation domain into smaller 3D sub-domains. Each sub-domain is assigned to each processor. Processors communicate and store "ghost" atom information (Plimpton, 2005) for atoms that border their sub-domain.

A single macroscopic state that consists of different microscopic states is defined by an ensemble (Plimpton, 2005). Ensembles with specific characteristics that can be defined in LAMMPS include:

- Microcanonical ensemble (NVE): The thermodynamic state is characterized by a fixed number of atoms, *N*, a fixed volume, *V*, and a fixed energy; *E*. NVE corresponds to an isolated system.

- Canonical Ensemble (NVT): This is a collection of all systems whose thermodynamic state is characterized by a fixed number of atoms, *N*, a fixed volume, *V*, and a fixed temperature, *T*.
- Isobaric-Isothermal Ensemble (NPT): This ensemble is characterized by a fixed number of atoms, *N*, a fixed pressure, P, and a fixed temperature, *T*.
- Grand canonical Ensemble (mVT): The thermodynamic state for this ensemble is characterized by a fixed chemical potential, m, a fixed volume, *V*, and a fixed temperature, *T*.

In this work, NVE ensemble is employed for the tensile and flexural deformation behavior of Nickel nanowires.

In the molecular dynamics simulations, periodic boundary conditions are commonly used to enable representing infinite size material configurations with only a small number of atoms. The particles do not interact across the boundary and do not move out of their enclosed periodic box. Such periodic boundary conditions are generally used with LAMMPS. However, to simulate the tensile and flexural behavior of Nickel Nanowires in this work, the molecular system is set to be non-periodic.

2.2 Stress Definition in MD (Virial Stress)

Cauchy stress, which represents force per unit area of the deformed solid, is commonly used in classical mechanics problems (Frankel & Smit, 2001) (Haile, 1997). However, this is not applicable for discrete systems such as those in molecular dynamics problems. LAMMPS uses virial stress (Zhou, 2003) (Zimmerman et. al., 2003) to describe the macroscopic (continuum) stress in accordance with the microscopic quantities. Given the phase status of atoms, the macroscopic stress tensor in a macroscopically small, but microscopically large volume Ω is given by (Zhou, 2003) (Zimmerman et. al., 2003):

$$\sigma_{\alpha\beta} = \frac{1}{\Omega}\sum_{i\in\Omega}\{-m_i(v_{i,\alpha}-\bar{v}_\alpha)(v_{i,\beta}-\bar{v}_\beta)+\frac{1}{2}\sum_j(X_{j,\alpha}-X_{i,\alpha})f_{ij,\beta}\} \quad (9)$$

where

$$f_{ij,\beta} = \frac{\partial \Pi_i}{\partial X_j} \quad (10)$$

Here m_i is the mass of the i-th molecule in Ω, X_i is its position (α and β indicates Cartesian components), v_i its velocity, ū the local average velocity, and f_{ij} is the force on molecule i exerted by another molecule j. This virial stress is used to compute the stress values for the tensile and flexural deformation behavior in the present simulations.

3. Tensile Deformation of Nickel Nanowires

The strain-stress relationship and the tensile deformation behavior of Nickel nanowires investigated via MD simulations are presented next. The effect of temperature, strain rate and specimen size on the mechanical properties (e.g., the maximum yielding stress and Young's modulus) are studied and discussed.

3.1 Computational Model Configuration of Tensile Nickel Nanowires

Figure 2 shows the computational model configuration of Nickel nanowires used in this work. The Nickel nanowires are made of Nickel FCC crystals with initial surface orientation of <100>, <010> and <001> (Liang & Zhou, 2003). The lattice constant of faced-centered cubic (FCC) (Frankel & Smit, 2001) (Haile, 1997) (Schlick, 2002) Nickel crystal is ג=3.52 Å.

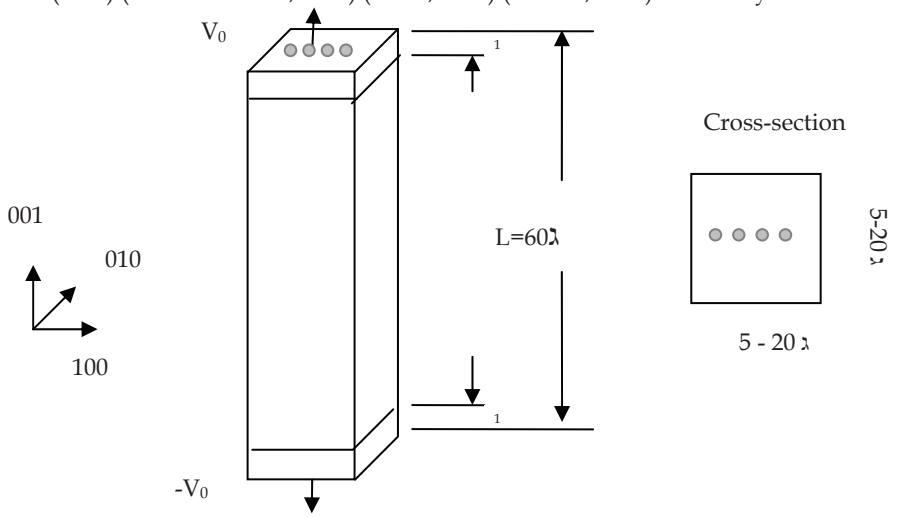


Fig. 2. Configuration of Tensile Nickel Nanowires for Tensile Deformation

In this work, the length (L) of Ni nanowires was set to be 60ג in the <001> directions for the MD simulations of the tensile deformation behavior. Different cross-sectional sizes that range from 5 to 20 ג for a side formed the different Nickel nanowire configurations. Constant velocities $\pm V_0$ are enforced on the top and bottom layers of the nanowires. These top and bottom layers define the boundary layer and have identical size of 1 lattice constant, along the <001> crystalline direction.

Based on the velocities over boundary layers ($\pm V_0$), the nanowire model system deform under strain rate given by:

$$\varepsilon' = \frac{2V_0}{L} \quad . \tag{11}$$

Here L is the length of Nickel Nanowires (Liang & Zhou, 2003).

Different velocities were employed in the simulations to give different strain rate conditions. Table 1 shows the velocity of boundary layers (V_0) and the resulting strain-rate. It should be noted that in LAMPPS the velocity is always expressed as lattice-constant (ג) per pico-second.

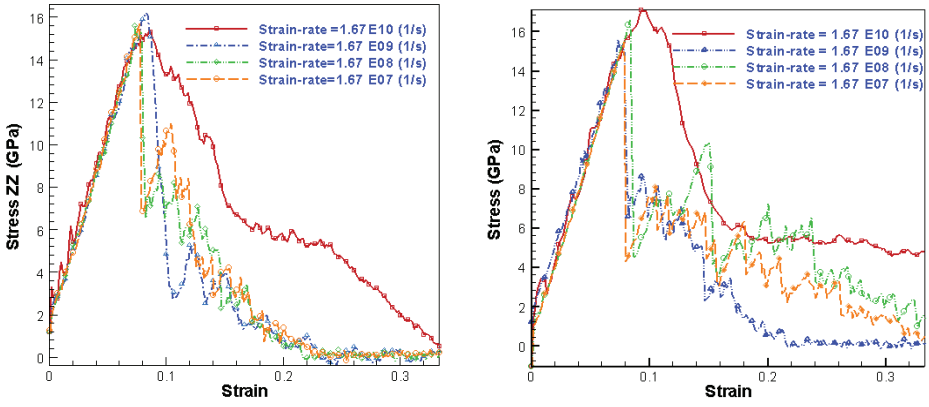
$\dot{\varepsilon}$(1/s)	V_0 (Å/ps)
1.67×10^{7}	5×10^{-4}
1.67×10^{8}	5×10^{-3}
1.67×10^{9}	5×10^{-2}
1.67×10^{10}	5×10^{-1}

Table 1. Tension Velocity of Nickel Nanowires and the Resulting Strain Rate

The tensile behavior of Ni nanowires and the influencing factors investigated through the MD simulations are presented next.

3.2 Effect of Strain Rate

Figure 3 presents the tensile stress-strain behavior obtained from 5×5×60 and 10×10×60 Nickel nanowires at various strain rates (1.67×10^{7} – 1.67×10^{10} s^{-1}). These simulations are based on an NVE ensemble at temperature T=*300K*. It is observed that a higher strain-rate led to higher oscillations in the stress-strain curve.



(A) 5×5×60; (B) 10×10×60

Fig. 3. Stress-Strain Curve of Nickel Nanowires under Various Strain-rates at T=300K.

The Young's modulus is determined from the results of tensile stress – strain curve for the strain $\varepsilon < 0.08$ using a linear regression. Table 2 and 3 shows the Young's modulus and maximum yielding stress of Nickel nanowires respectively under tensile loading.

Strain-rate($\dot{\varepsilon}$)	5×5×60	10×10×60
1.67×10^{7}	191.27	184.2127
1.67×10^{8}	189.82	192.0556
1.67×10^{9}	184.33	190.3636
1.67×10^{10}	182.71	187.2973

Table 2. Young's Modulus (GPa) of Nickel Nanowires for Various Strain-rates

This data indicates that strain rate does not significantly influence the Young's modulus and the maximum yield stress.

Max. Yielding Stress (GPa)	5×5×60	10×10×60
1.67×10^{7}	15.7178	15.4835
1.67×10^{8}	15.6100	16.6046
1.67×10^{9}	16.2640	15.6100
1.67×10^{10}	15.3314	17.0836

Table 3. Maximum Yielding Stress (GPa) of Nickel Nanowires for Various Strain-rates

Figures 4 and 5 present the progressive deformation and failure of 5×5×60 Ni nanowires for the strain rates $1.67x10^{7}$ and $1.67x10^{10}$ (s^{-1}) respectively. As seen from these results, it is observed that the yielding slip planes, cross slip and the breaking neck (Chang & Fang, 2003) (Schlick, 2002) (Silva, et. al., 2004) of Nickel Nanowires are influenced by the strain-rate. The two deformation configurations presented in figures 4 and 5 are an intermediate configuration during the deformation and the final yielding configuration. These figures clearly show that the strain rate influences the yielding slip planes, cross slip, and the breaking neck of the Ni nanowires.

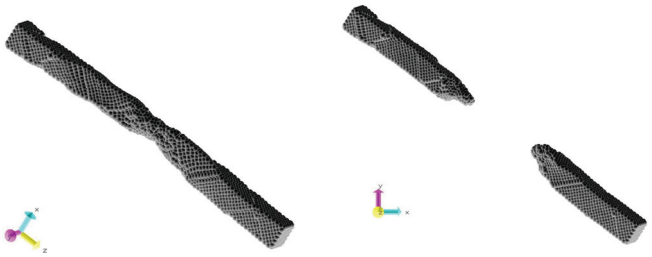


Fig. 4. Tensile Deformation and Failure of 5X5X60 Nickel Nanowire; Strain Rate = 1.67×10^{7} (s^{-1})

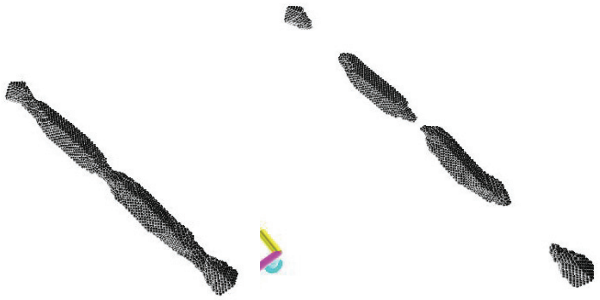


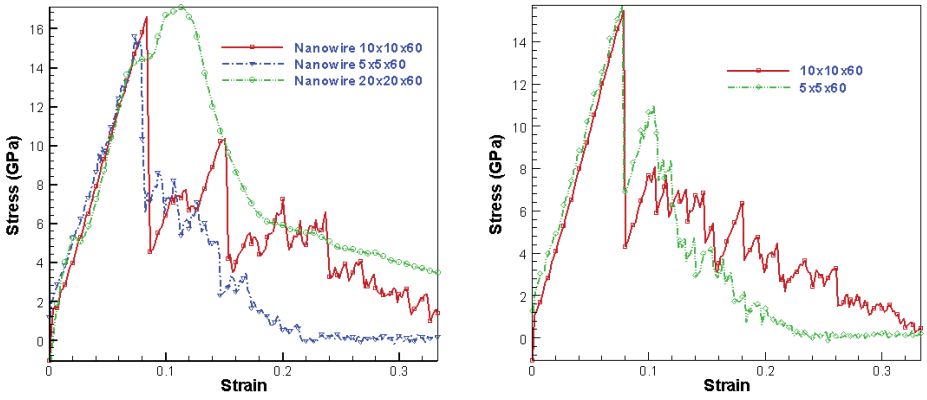
Fig. 5. Tensile deformation and failure of 5×5×60 Nickel nanowires; strain rate=1.67×10^{10} (s^{-1})

3.3 Effect of Nanowire Cross Section Size

For a specific strain-rate (1.67×10^{10} s^{-1} or 1.67×10^{7} s^{-1}), Figure 6 presents the comparison of the simulated strain-stress behavior of Nickel nanowires for varying cross-sectional dimensions (5-20 ג) (Liang & Zhou, 2003). The results indicate that cross-sectional dimension does not significantly affect the Young's modulus of Nickel Nanowires. The larger cross-sectional dimension leads to larger maximum yielding stress. Further, it is observed from Figure 6 that larger cross-sectional dimension nanowires have less numerical oscillations

amplitude compared to the smaller cross-sectional dimension nanowires. This phenomenon may be caused due to the dynamic wave effect or phonon drag that impedes the motion of dislocations (Liang & Zhou, 2003).

Table 4 and 5 shows the Young's modulus and maximum yielding stress of tensile Nickel nanowires, respectively for various cross-sectional dimensions. These results indicate that the cross-sectional size does not significantly affect the Young's modulus and the maximum stress obtained from the simulations.



(A) strain-rate = 1.67×10^{10} s^{-1} (B) strain-rate = 1.67×10^{7} s^{-1}

Fig. 6. Strain-stress curve of Nickel nanowires for Various cross-sectional dimensions

Strain-rate($\dot{\varepsilon}$)	5×5×60	10×10×60	20×20×60
1.67×10^{7}	191.27	184.2127	-
1.67×10^{10}	182.71	187.2973	184.0893

Table 4. Young's Modulus (GPa) of Nickel Nanowires for Various Cross-sectional Dimensions

Strain-rate($\dot{\varepsilon}$)	5×5×60	10×10×60	20×20×60
1.67×10^{7}	15.7178	15.4835	-
1.67×10^{10}	15.3314	17.0836	17.1271

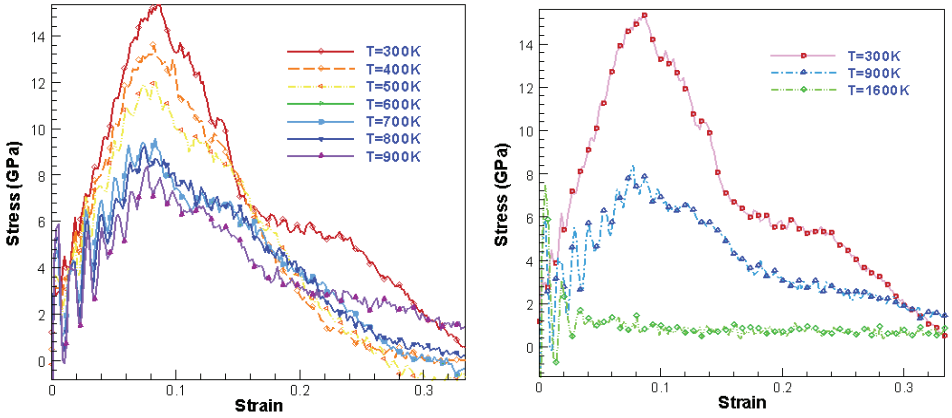
Table 5. Maximum Yielding Stress (GPa) of Nickel Nanowires for Various Cross-sectional Dimensions

3.4 Effect of Temperature

Temperature plays an important role in the deformation behavior of nanomaterial systems. Molecular dynamics simulations as presented and discussed in this chapter provide an effective predictive methodology to understand the deformation behavior of nanomaterial systems (such as the nanowire). This section focuses on the effect of varying temperature on the mechanical properties and deformation behavior of Nickel nanowires (Chang & Fang, 2003). For this analysis, molecular dynamics simulations are performed for various temperature conditions of the molecular ensemble. Figure 7(A) shows the stress-strain behavior at temperature ranging from 300K to 900K in steps of 100K. Figure 7(B) includes

the stress-strain behavior at a higher temperature of 1600K which is close to the melting point of Nickel (1728K) along with the behavior at room temperature (300K) and an elevated temperature (600K). The simulated stress – strain deformation behavior obtained from Figure 7(A) indicates that:

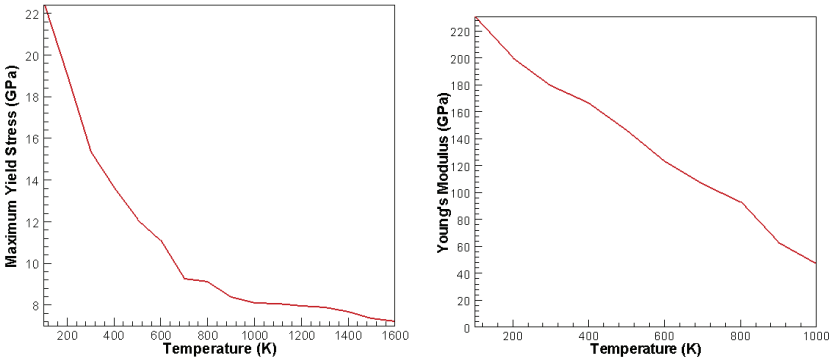
- The maximum yielding stress decreases as the temperature of the ensemble increases
- With the increase of ensemble temperature, the thermal oscillations due to numerical corrections and temperature scaling also become dominant.



(A) In the temperature range 300-900K (B) Near melting point.

Fig. 7. Stress-Strain behavior of 5×5×60 Nickel nanowires at various temperatures

Figure 8(A) shows the change in the maximum yield stress of Nickel nanowire at varying temperatures ranging 100K-1600K. As can be expected, the maximum yield stress decreases with increasing temperature. Figure 8(B) shows the associated Young's modulus of Nickel nanowires obtained from molecular dynamics simulations at various temperatures. As seen from this figure, the Young's modulus of Nickel nanowires decreases fairly linearly as the temperature increases.



(A) Maximum yield stress (B) Young's modulus

Fig. 8. Effect of Temperature on the Maximum Yield Stress and Young's Modulus of Nickel Nanowires (5×5×60)

Figure 9 illustrates the superior ductility of Nickel nanowires around melting temperature. As seen from figure 9 obtained from the molecular dynamics simulations, it is observed that when the ensemble is maintained at 1600K, the nanowires can be greatly extended without a ductile fracture. Further discussions of molecular dynamics simulations on the transition from brittle to ductile are presented in Cheung and Yip (Cheung & Yip, 1994).

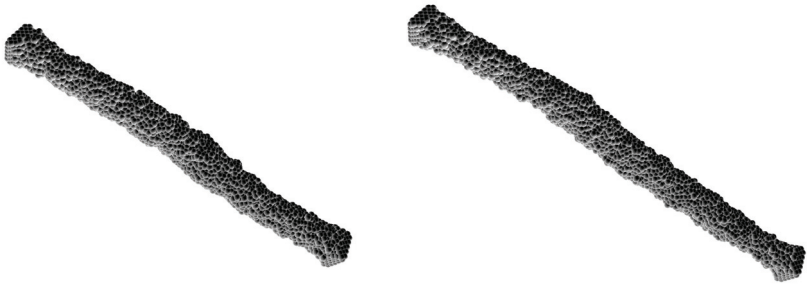


Fig. 9. Deformation configuration of 5×5×60 Nickel nanowire at T=1600K

4. Flexural Deformation of Nickel Nanowires

This section focuses on the flexural deformation behaviour of Nickel nanowires due to flexural bending based on their molecular configurations using molecular dynamics simulations. In particular, the deformation vibration frequencies obtained from the molecular dynamics simulations are compared with the natural frequencies based on classical beam theory for two different boundary conditions.

4.1 Computational Model Configuration and Analysis of Flexural Nickel Nanowires

Figure 10 presents the configuration of Nickel nanowire beams and the corresponding molecular model. The molecular model configuration is based on single crystals of Nickel in the <001> (longitudinal direction), <010> and <100> (transverse directions) directions and with a dimension of 120×10×1 cubic lattice constants (rectangular cross section with a longer span). Two types of boundary conditions are employed in the simulations: 1. Both ends pinned (i.e., simply supported), 2. Both ends clamped. The nanowire beam deflects under the action of applied loading and when the loading force is removed, the displaced beam would try to return to its original position. The inertia of the beam would cause the beam to vibrate. The transient flexural bending dynamic behavior of the molecular configuration of Nickel nanowire beams are investigated and analyzed.

The transient molecular dynamic simulations compute the new position of the atoms in the Nickel nanowire beam subjected to the flexural loading and the boundary constraints. The time increments are however significantly small in these dynamic simulations. A Mean Square Displacement (MSD (u(t))) is defined and used as a measure of the average distance an atom in the model travels over a certain time interval period. This is defined as:

$$msd(u(t)) = \frac{1}{N}\sum_{i=1}^{N} u_i^2(t) = \frac{1}{N}\sum_{i=1}^{N} (r_i(t) - r_i(0))^2 \tag{12}$$

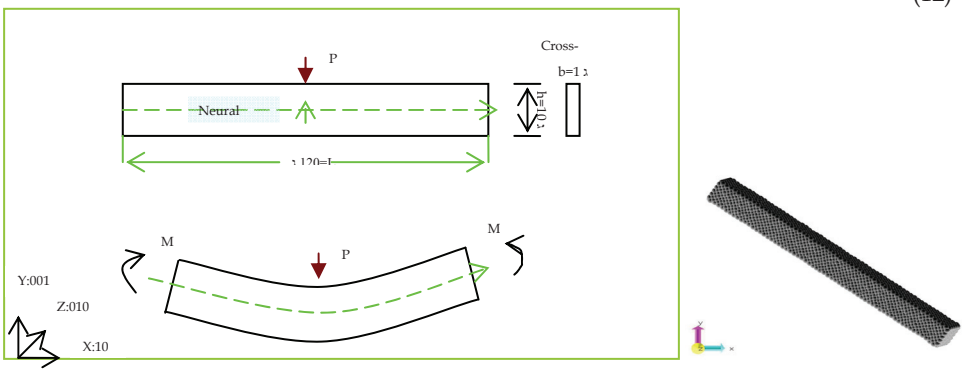


Fig. 10. Configuration of Nickel Nanowire Beams for Flexural Deformation and Molecular Model

The displacement $\boldsymbol{u}_i(t)=\boldsymbol{r}_i(t)-\boldsymbol{r}_i(0)$ is the distance traveled by molecule i over some time interval t, and the squared magnitude of this vector is averaged over many such time intervals. This MSD displacement value is used in the analysis of time dependent displacement response of the Ni nanowires. The dynamic displacement response under two different boundary conditions for flexural bending is studied.

4.2 Simply Supported Nickel Nanowire Beam

The Nickel nanowire beam configuration as shown in figure 10 is simply supported (rotations are possible at the ends) and is subjected to a dynamic concentrated point load at the center of the nanowire beam. Two different load values (F=0.01eV/Å and 0.03eV/Å) are analyzed.

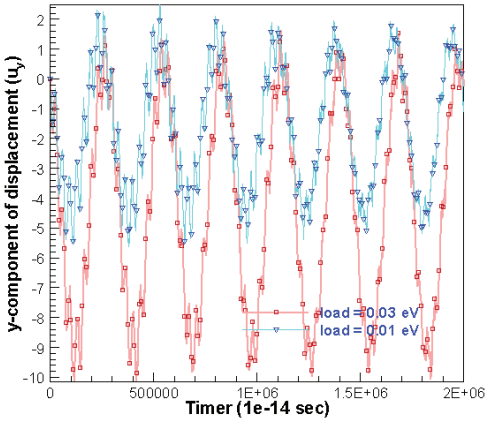


Fig. 11. Transient Dynamic Vertical Displacement at the Center (simply supported ends)

Figure 11 presents the dynamic mean square vertical displacement variation with time at the center of the nanowire beam under these two loading conditions. These are based on the averaged vertical displacement of the atoms in the cross-sectional plane at the center in all cases.

The dynamic vertical displacement at the center is proportional to the loading value and increases with a higher magnitude of external loading. The natural frequency of the dynamic vibration as computed from the displacement - time profile is however independent of the magnitude of external loading. The computed angular frequency from the predicted time dependent deflection of the molecular model of the Nickel nanowire beam shown in Figure 11 is 2.4166E+09 Hz.

The natural frequency for the case of a simply supported beam based on classical beam theory analysis is given by (Megson, 1996) (Voltera & Zackmanoglou, 1965) (Tedesco et. .al., 1999)

$$\omega_n = (n\pi)^2 \sqrt{\frac{EI}{\rho A L^4}} \tag{13}$$

Using a Young's Modulus value of E = 190 GPa (1.1859 eV/A^3) obtained from molecular dynamics modeling of the tensile stress strain deformation of Nickel nanowires discussed in the earlier section, the mode 1 frequency value obtained from the classical beam theory is 2.5244E+10 Hz. This frequency as obtained from the classical beam theory is at least one order higher than the frequency obtained from the time dependent deflection using molecular dynamics simulations. The classical elastic beam theory based on continuum mechanics principles also indicate that the natural frequency of vibration of a simple supported beam is independent of the magnitude of the external loading and depends only on the beam cross sectional moment of inertia, cross-sectional area, length and modulus of elasticity of the material. The natural frequency obtained from molecular dynamics simulations for the loading and simply supported boundary conditions as presented in figure 11 is also independent of the magnitude of the external loading.

4.3 Clamped Nickel Nanowire Beam

The previous section considered the case of a simply supported boundary condition configuration where the rotation at the ends of the flexural beam is permitted. The same nanowire beam is fixed at both ends (displacement and the rotation at the ends are zero) and is subjected to external loading force at the center. As before, two different loading values are investigated. Figure 12 presents the computed dynamic displacement response of the loaded center of the nanowire beam. As seen from figure 12, the dynamic displacement magnitude depends on the external loading value while the frequency of the dynamic displacement is independent of the external loading values. This is in direct correlation with the analytical results of natural frequency based on the classical beam theory.

The computed angular frequency obtained from the predicted time dependent deflection of the clamped Nickel molecular beam shown in Figure 12 is 2.3271E+09 Hz. The frequency of vibration based on the classical beam theory for this case of clamped ends is given by (Megson, 1996) (Voltera & Zackmanoglou, 1965) (Tedesco et. .al., 1999)

$$\omega_n = (K_1 L)^2 \sqrt{\frac{EI}{\rho A L^4}}; \quad K_1 = 4.73 \tag{14}$$

Using the same Young's Modulus for the Nickel nanowire as before, the mode 1 natural frequency as obtained based on the classical beam theory is 5.7225E+10 Hz. This frequency obtained from the classical beam theory for this clamped Nickel nanowire beam configuration is at least one order higher than the frequency obtained from the time dependent center point load deflection using the molecular dynamics simulations. The classical beam theory based on continuum mechanics principles also indicate that the natural frequency of a clamped beam is independent of the external loading and depends only on the beam cross sectional moment of inertia, cross-sectional area, length, and modulus of elasticity of the material. This was also the case in the frequency of the nanowire beams obtained from molecular dynamics simulations as presented in figure 12.

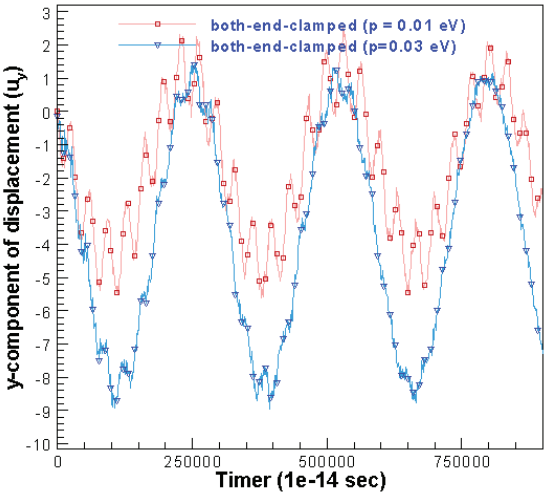
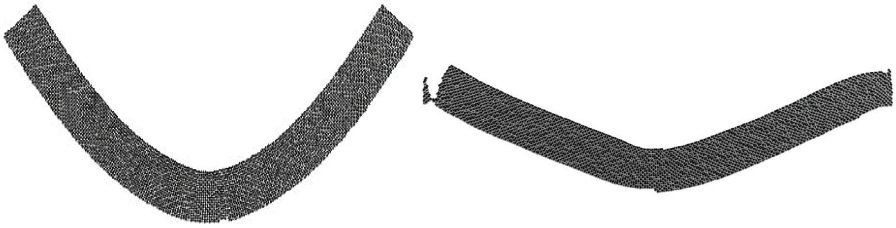


Fig. 12. Transient Dynamic Vertical Displacement at the Center (clamped ends)

4.4 Effect of Large Loading

The flexural, bending failure behavior of these Nickel nanowire beams with a larger loading is studied. As the loading is increased, the failure profile from the molecular dynamics simulations clearly shows the effect of boundary conditions. A shear slip failure along the mid-plane at the center of the nanowire beam where the shear forces are higher is seen in the case of nanowire beam with the clamped ends. The simply supported nanowire beam configuration presents a similar failure at the center of the nanowire beam but did not show predominant shear slip failure. Figure 13 shows the deformed shape at failure with a larger load for both the boundary condition cases discussed earlier. Molecular dynamics modeling provides effective means of understanding the deformation of Nickel nanowires using the molecular model configuration. These simulations are effective in understanding the deformation and failure behavior of these material systems and are especially useful as the

experimental techniques are not practical and conducive for such length scale material systems.



(A) Simply Supported Ends. (B) Clamped Ends

Fig. 13. Deformed Shape on Failure (left: simply supported; right: clamped)

5. Summary and Concluding Remarks

As the metallic nanowires are increasingly used in electronic devices, it is essential to understand their mechanical deformation behavior, structural and mechanical stability under external mechanical and thermal loading conditions. This mechanical strength and stability are important considerations during the life cycle of these electronic devices, as well as in other applications of such metallic nanowires. Computational modeling and simulations based on molecular dynamics modeling and the associated inter atomic potential provide an effective means to understand these mechanical behaviors. These computational modeling and simulations are based on the molecular model configurations of these metallic nanowires. In this chapter, molecular dynamics modeling of the two fundamental mechanical behaviors of Nickel Nanowires, namely the tensile deformation and failure of Nickel nanowires; and the flexural bending deformation behavior of Nickel nanowire beam configurations are presented and discussed. The tensile deformation behavior focused on the stress-strain behavior and the tensile failure of the nanowires. The flexural deformation behavior focused on the vibration characteristics of the nanowire beams subjected to external loading and its release. Nanowire inter-connectors in electronic devices often are subjected to such loading conditions and these enabling molecular computational modeling and simulations for nanotechnology applications provide an effective means for such analysis.

The molecular dynamics simulations for the tensile mechanical deformation behavior in Nickel nanowires employed the molecular configurations of Nickel (FCC crystal structure). The deformation behavior of these atomic molecular configurations is modeled based on the force field interactions between the Nickel atoms at a finite temperature. These simulations are conducted using a classical molecular dynamics code LAMMPS (Large Scale Atomic/Molecular Massively Parallel Simulator) from Sandia National Laboratory. The force field interactions between the Nickel atoms are modeled based on the embedded atom potential (EAM).

Molecular dynamics modeling of the tensile behavior of Nickel nanowires under different strain rates were conducted to predict the stress-strain behavior and Young's Modulus of these nanowires. The tensile deformation behavior indicated that the Young's Modulus was independent of the cross sectional area of the nanowire and the strain rate employed in the tensile deformation. Also, as the temperature increases, the Young's modulus of Nickel

nanowire is found to decrease with improved ductility. At a temperature close to the melting temperature, the molecular Nickel nanowire configuration behaves as a highly ductile, fluid material system.

The bending and flexural deformation and vibration behavior indicated that the frequency of the vibrations as compared from time displacement deformation behavior of the molecular configurations of the Nickel nanowire beams are independent of the magnitude of external loading. This was the case for both the simply supported and clamped end boundary configurations discussed in this chapter. This is consistent with the natural frequency based on the classical beam theory for such boundary and loading conditions. The magnitude of natural frequency was however higher when compared with the classical beam theory. The displacement magnitude was also significantly higher in the case of a larger loading. A higher loading however causes the failure of the simulated Nickel nanowire beam. A shear-slip type of failure along the mid plane at the mid-point of the beam was observed with a larger bending flexural loading. This shear-slip failure is predominant in the case of the clamped boundary condition.

Experimental investigations of the mechanical behavior of metallic nanowires are impractical and computational molecular modeling and simulations as analyzed and discussed in this chapter provide effective means for such analysis.

6. Acknowledgements

The support in part by the Office of Naval Research (Award Number: N00014-09-1-0842) is acknowledged by Mohan. Liang thanks W. Liang and M. Zhou at Georgia Tech for sharing their experiences on the application of molecular dynamics. Thanks to S. Plimpton for the assistance with LAMMPS. The molecular configuration diagrams presented in this chapter were generated using VMD (Humphrey et. al., 1996).

7. References

Binnig, G.; Quate, CF.; Gerber, CH. (1986). Atomic Force Microscope, *Phys. Rev. Lett*, 56.,930-933.

Chang, WJ.; Fang, TH. (2003). Influence of Temperature on Tensile and Fatigue Behavior of Nanoscale Copper Using Molecular Dynamics Simulation, *Journal of Physics and Chemistry of Solids*, 64. 1279-1283.

Chen, H.; Lin, C.; Chen, W.; Hsu, Y.; & Uang, R. (2006). Process Dependent Thermal-Mechanical Analysis and Design of Novel Nanowires Based Anisotropic Conductive Film Assembly, *Proceedings of International Conference on Microsystems, Packaging and Assembly*, IEEE Paper 1-4244-0735-4/06, Taiwan, IEEE.

Cheung, KS.; Yip, S. (1994). A Molecular Dynamics Simulation of Crack Tip Extension: The Brittle to Ductile Transaction, *Modeling Simul. Mater. Sci. Eng*., 2, 865-892.

Chialvo, AA.; Bebenedeti, PG. (1990). On the Use of the Verlet Neighbor List in Molecular Dynamics, *Computer Physics Communications*, 60, 2, 215-224.

Frankel, D.; Smit, B. (2001). *Understanding Molecular Simulations: From Algorithms to Applications*, Academic Press, ISBN 0-12-267370-0, San Diego.

Haile, J. (1997). Molecular Dynamics Simulations: Elementary Methods, *John Wiley & Sons, Inc*., ISBN 0-471-81966-2, New York.

Humphrey, W.; Dalke, A.; Schulten, K. (1996). VMD – Visual Molecular Dynamics, *J. Molec. Graphics*, 14, 33-38.

Ikeda, H.; Qi, Y.; Cagin, T.; Samwer, K.; Johnson, WL.; Goodard III, WA. (1999). Strain Rate Induced Amorphization in Metallic Nanowires, *Physical Review Letters*, 82, 2900-2903.

Ju, SP.; Lin, JS.; Lee, WJ. (2004). A Molecular Dynamics Study of the Tensile Behavior of Ultra-thin Gold Nanowires, Nanotechnology, 15, 1221-1225.

Liang, WW.; Zhou, M. (2003). Size and Strain Rate Effects in Tensile Deformation of Cu Nanowires, *Proceeding of the 44th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics, and Materials Conference*, Norfolk, VA.

Liew, KM.; He, XQ.; Wong, CH. (2004). On the Study of Elastic Plastic Properties of Multi-Walled Carbon Nanotubes under Axial Tension using Molecular Dynamics Simulation, *ACTA Material*, 52, 2521-2527.

Megson, THG. (1996). *Structural Stress Analysis*, John Wiley & Sons Inc., ISBN 0 470 23563 2, New York.

Park, HS.; Zimmerman, JA. (2005). Modeling Inelasticity and Failure in Gold Nanowires, *Physical Review B*, 72, 54-106.

Plimpton, SJ. (2005). LAMMPS, A Large Scale Atomic/Molecular Massively Parallel Simulator, *www.cs.sandia.gov/~sjplimp/lammps.html*

Plimpton, SJ. (1995). Fast Parallel Algorithms for Short Range Molecular Dynamics, *J. Comp. Physics*, 117., 1-19.

Rapaport, D. (2004). *The Art of Molecular Dynamics Simulation*, Cambridge University Press, ISBN 0 521 82568 7, Cambridge.

Scarola, VW.; Mahan, GD. (2002). Phonon Drag Effect in Single Walled Carbon Nanotubes, *Physical Review B*, 66, 205405.

Schlick, T. (2002), Molecular Modeling and Simulation – An Interdisciplinary Guide, *Springer-Verlag Inc.*, ISBN 0-387-95404-X, New York.

Silva, EZ.; Silva, AJR.; Fazzio, A. (2004). Breaking of Gold Nanowires, *Computational Materials Science*, 30, 73-76.

Silva, EZ. (2004). Theoretical Study of the Formation, Evolution and Breaking of Gold Nanowires, *Phys. Rev. B*, 69., 115411.

Tedesco, JW.; McDougal, WG.; Ross, CA. (1999). *Structural Dynamics – Theory and Applications*, Prentice Hall, ISBN 978-0673980526.

Volterra, E.; Zackmanaglou, EC. (1965). *Dynamics of Vibrations*, Charles E. Merrill Books Inc., ASIN B000BSN9QS, Columbus.

Young, RD. (1966). Scanning Tunneling Microscope, *Rev. Sci. Instrum.*, 37, 275.

Zhou, M. (2003). A New Look at the Atomic Level Virial Stress: On Continuum Molecular System Equivalence, *Proc. R. Soc. Lond. A*, 2347-2392.

Zimmerman, JA.; Webb, EB.; Hoyt, JJ.; Jones, RE.; Klein, PA.; Bammann, DJ. (2003). Calculation of Stress in Atomistic Simulation, *Modeling Simul. Mater. Sci. Eng.*, 12, 319-332.

11

Advances in Nanowire-Based Computing Architectures

Jun Wu, Sriram Venkateswaran and Minsu Choi
Missouri University of Science and Technology
United States

1. Introduction

The end of photolithography as the driver for Moore's Law is predicted within seven to twelve years and nanotechnologies are emerging that are expected to continue the technological revolution (Bourianoff et al., 2003). Recently, numerous nanoscale logic devices have been proposed based on nanoscale components such as CNTs and SiNWs; computing architectures are also being proposed using them as primitive building blocks.

One of the most promising nanotechnologies is the nanowire crossbar-based architecture, a two-dimensional array (i.e., nanoarray) formed by the intersection of two orthogonal sets of parallel and uniformly-spaced nanometer-sized wires (Dehon et al., 2003; Rueckes et al., 2000), such as carbon nanotubes (CNTs) and silicon nanowires (SiNWs). Experiments have shown that such wires can be aligned to construct an array with nanometer-scale spacing using a form of directed self-assembly and the formed crosspoints of nanoscale wires can be used as programmable diodes, memory cells or FETs (Field-Effect Transistors); therefore, nanoscale logic devices can be realized.

Nanowire crossbars offer both an opportunity and a challenge. The opportunity is to achieve ultra-high density which has never been achieved by photolithography (a density of 10^{11} crosspoints per cm^2 has been already achieved (Melosh et al., 2003)). The challenge is to make them simple enough to be manufactured and reliable enough to be used in everyday computing applications. Ultra-high density fabrication could potentially be very inexpensive if researchers can actualize a chemical self-assembly, but such a circuit would require laborious testing, repair and reconfiguration processes, implying significant overhead costs. All reconfigurable computing architectures based on nanowire crossbars are commonly envisioned to be used for synchronous circuits and systems. Thus, a complex clock distribution network should be fabricated along with nanowire crossbars and precise timing control should be practiced to avoid all timing-related faults induced by physical design parameter variations caused by nanoscale nondeterministic assembly.

2. Benefits and Challenges

Unlike the conventional clocked counterparts, a new asynchronous architecture for carbon nanotube (CNT) and silicon nanowire (SiNW)-based reconfigurable nanocomputing systems is proposed to address aforementioned issues in this chapter. The proposed

asynchronous nano-architecture is based on delay-insensitive data encoding and self-timed logic referred to as the Null Convention Logic (NCL) – making it totally clock-free. The potential benefits can be summarized as follows:
1. Manufacturability: Distributing the clock signal is not needed. All clocking-related hardware components can be removed from the overall hardware design. This not only saves on circuit area, but also facilitates the homogeneity of the system – an important issue for manufacturing based on molecular self-assembly.
2. Scalability: The overall circuit is self-timed – meaning that all timing information is embedded in the encoding. This unique aspect of the proposed asynchronous architecture provides better scalability than its clocked counterparts, since the timing is locally handled. The timing complexity remains the same even though the size of the circuit gets larger.
3. Robustness: Due to non-determinism of the directed self-assembly paradigm, nanowire crossbar circuits are anticipated to exhibit large variations in physical parameters. Since any physical variation in an electrical parameter may have its own negative effect on the timing behavior of the circuit, being able to design delay-insensitive circuits (i.e., correct operation of the circuit is independent of the timing) is a significant capability and it would greatly increase the robustness of the circuit to design parameter variations.
4. Modularity: Asynchronous circuits can be divided into modules that are designed without considering other modules, especially with respect to timing relationships. Modularity facilitates configurability of circuits.
5. Defect and Fault-Tolerance: In addition to the complete removal of many timing-related failure modes, testing complexity is reduced in that stuck-at-1 faults simply halt the circuit. Also, in case of dual-rail encoding, 11 is considered as an invalid code. So, any permanent or transient fault that results in this invalid codeword can be eventually detected. Only stuck-at-0 faults and some other transient faults need to be exercised with applied patterns. Design time and risk as well as circuit testing requirements are expected to be decreased because of elimination of complexity of clock and its associated critical timing issues.
In order to be a viable nanotechnology, the nanowire crossbar-base systems should be:
1. Structurally simple and scalable enough to be fabricated by bottom-up manufacturing technique.
2. Robust enough to tolerate extreme parametric variations.
3. Defect and fault-tolerant enough to overcome the extreme defect densities, aging factors and transient faults.
4. Able to support at-speed verification and reconfiguration.

3. Nanoscale Devices and Computing Architecture

3.1 CNT and SiNW Crossbar-Based Nanoscale Devices

Unlike CMOS, chemically-assembled nanoscale components (such as CNTs and SiNWs) are unlikely to be used to construct complex aperiodic structures (Goldstein et al,. 2001; Goldstein et al,. 2002; Mishra et al,. 2002). Thus, unconventional architectures are often desirable. One of the most promising computational nanotechnologies is the crossbar-based architecture (Kuekes et al,. 2000; Kuekes et al,. 2001; Chen et al,. 2003; Ziegler et al,. 2002; Stan et al,. 2001), a two-dimensional array (nanoarray) formed by the intersection of two orthogonal sets of parallel and uniformly-spaced nanometer-scale wires, such as carbon nanotubes (CNTs) and silicon nanowires (SiNWs). A typical nanoscale crossbar structure is

shown in Fig.1. Experiments have shown that such nanoscale wires can be aligned to construct an array with nanometer-scale spacing using a form of directed self-assembly (Nicewarner et al,. 2002; Huang et al,. 2001; Cui et al,. 2001; Mirkin et al,. 2000;)

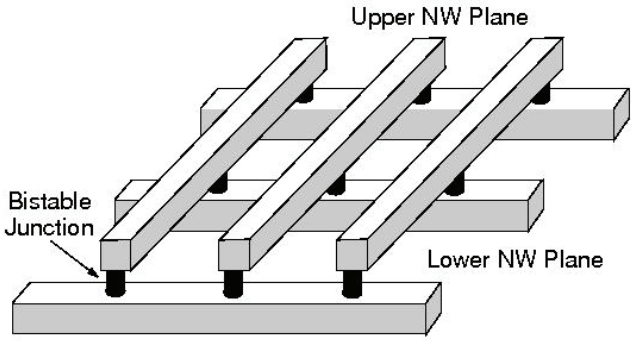


Fig. 1. Typical nanowire crossbar architecture

One or more crosspoints can be grouped together to form a memory or logic device. Lieber research group have shown electro-mechanical switching devices using suspended nanotubes (Rueckes et al,. 2000). The NT-NT junction is bistable with an energy barrier between the two states. SiNWs can be substituted for the lower wire, and these junctions can rectifying such that the connected state exhibits p-n-diode rectification behavior. Thus, diode-resistor logic can be realized in nanoscale. Doped SiNWs exhibit FET (Field Effect Transistor) behavior (Huang et al,. 2001). That is, oxide can be grown over the SiNW to prevent direct electrical contact of a crossed conductor. The electrical field of one wire can then be used to gate the other wire by locally evacuating a region of the doped SiNW of carriers to prevent conduction. In other words, as the gate voltage is changed, conductance increases or decreases between the source and drain. CNTs also demonstrate FET behavior (Bachtold et al,. 2001; Derycke et al,. 2001; Soh et al,. 1999) and can be also used as memory devices (Finkelstein et al,. 2003; Rueckes et al,. 2000) or interconnects (Dekker et al,. 1999).
Molecular resonant tunneling diodes, often called negative differential resistors (NDRs), have also been realized (Bhattacharya et al,. 2001; Chen,J. et al,. 1999; Chen ,J. Et al,. 2000; Yu et al,. 1999). Devices with useful peak-to-valley ratio have been measured at room temperature (Chen,J. et al,. 2000). These molecules can be used as the basis of logic families (Chen,J. et al,. 1999) or as the core of a molecular latch which also provides signal restoration and I/O isolation (Goldstein et al., 2002).

3.2 Nanoscale Universal Computing Architecture

DeHon et. al. h shown how to organize the CNTs, SiNWs and molecular-scale devices that are now being developed into an operational reconfigurable computing system (referred to as the NanoPLA) (Dehon et al,. 2003; Naeimi & Dehon et al. 2004). The molecular-scale wires can be arranged into interconnected, crossed arrays with switching devices at their crosspoints. Similar nanoscale reconfigurable homogeneous architecture, namely Nano-Fabric, has been also proposed by Goldstein et. al. ((Goldstein et al,. 2002). Lieber research group referred to such a nanoscale reconfigurable homogeneous structural paradigm for

computing as the Universal Computing Architecture (UCA). Such nanoscale computer architectures share common characteristics - they support unconventional nanoscale manufacturing paradigm via simple homogeneous periodic structures and reconfigurability for post-fabrication design mapping. In general, nanowire-based reconfigurable systems based on the UCA concept are referred to as the **Nanowire Reconfigurable Crossbar Architecture (NRCA)**.

There are considerable on-going research and development efforts to fabricated nanoscale crossbar-based circuits and systems using advanced lithography as well. For example, researchers at HP successfully fabricated 8*8 (i.e., 64 bits) crossbar memory arrays (Paulson et al,. 2005; Chen, Y et al,. 2003) using nano-imprint lithography (Mcalpine et al,. 2005; Hiroshima et al,. 2002; Xia et al,. 1999). Non-volatile bistable Rotaxane molecules are sandwiched between two orthogonal metal wires to form a non-volatile memory cell.

About 75% of the memory cells are tested functional and all the other cells are either stuck-open or stuck-closed: therefore, non functional. It is also easily predictable that even higher defect densities will be induced as the size of such devices scale down. Therefore, enhancement of lithographic resolution and defect tolerance are two key challenges that advanced lithographic fabrication methods, such as nano-imprint lithography, face.

3.3 Bottom-Up Paradigm for Nanowire Crossbar Assembly

CNTs and SiNWs are the most promising building blocks for nanoscale computing systems. Unfortunately, synthesis of such nanowires and high-density integration of devices and systems based on nanowires are fully different from conventional top-down lithographic fabrication techniques, because such nanowires must be synthesized first, then assembled into functional devices and systems in a bottom-up manner.

Langmuir-Blodgett (LB) method (Huang et al,. 2001) is an effective way to hierarchically assemble 1D nanostructures into integrated nanosystems based on fluidic flow. In this method, SiNWs or CNTs can be aligned by passing a suspension of NWs through microfluidic channel structures. Ordered monolayers are formed over a large area and transferred to substrates by the LB method. It has been demonstrated that virtually all of the NWs are aligned along the flow direction. Alternating the flow in orthogonal directions in a two-phase assembly process results in crossbar structures having high yield.

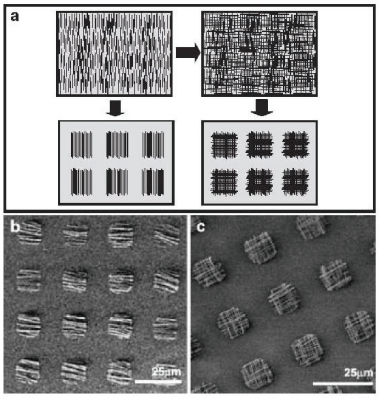


Fig. 2. Hierarchical patterning by fluidic flow and LB method (Whang et al. 2004)

The aligned, controlled spacing NW structures produced by the LB assembly method exhibit fluctuations in the average alignment direction and poor end-to-end registry. So, interconnected finite-size arrays of nanoscale devices are more desirable than monolithic structures for integrated nanosystems, because hierarchical organization reduces the probability that small numbers of defects will cause catastrophic failure in the whole system. Hence, by adjusting the array size to be less than the average NW length it is possible to minimize the number of NWs that fail to span the width of an array due to poor end-to-end registry. Following uniform transfer of NWs of a specified spacing onto a substrate, photolithography is used to define a pattern over the entire substrate surface, which sets the array dimensions and array pitch, and then the NWs outside the patterned array are removed by gentle sonication. SEM (Scanning Electron Microscope) images of both parallel NW arrays and crossed NW arrays are shown in Fig.2.((b) and (c)), respectively.

3.4 Overview of Asynchronous Design Methods - Null Convention Logic for Clock-Less Computing

Asynchronous circuits fall into two main categories: delay-insensitive and bounded-delay models. Paradigms, like NCL (Null Convention Logic), assume delays in both logic elements and interconnect to be unbounded, although they assume that wire forks are isochronic (Fant et al,. 1996). This implies the ability to operate in the presence of indefinite arrival times for the reception of inputs. Completion detection of the output signals allow for handshaking to control input wavefronts. On the other hand, bounded-delay models such as Huffman circuits (Unger, 1969), burst-mode circuits (Nowick & Dill, 1991), and micropipelines (Sutherland, 1989) assume that delays in both gates and wires are bounded. Delays are added based on worse-case scenarios to avoid hazard conditions. This leads to extensive timing analysis of worse-case behaviour to ensure correct circuit operation. Since NCL exhibits neither of these characteristics, it is well-suited for the proposed clock-free nanowire crossbar architecture and bounded delay models are not addressed further.

Null Convention Logic (NCL) is a delay-insensitive logic which is one of the Asynchronous circuits categories. NCL circuits utilize dual-rail or quad-rail logic to achieve this delay-insensitivity (Bandapati et al,. 2003). In dual-rail logic, a signal D consists of two wires D^0 and D^1; which may assume any value from the set DATA 0, DATA 1, and NULL. The DATA 0 state ($D^0 = 1$, $D^1 = 0$) corresponds to a Boolean logic 0, the DATA 1 state ($D^0 = 0$, $D^1 = 1$) corresponds to a Boolean logic 1, and the NULL state ($D^0 = 0$, $D^1 = 0$) corresponds to the empty set meaning that the value of D is not yet available. The two rails are mutually exclusive, so that both rails can never be asserted simultaneously; this state is defined as an illegal state (Bandapati et al,. 2003).

NCL is self-timed logic paradigm in which control is inherent. Whenever the circuit inputs become available, it does not need timing analysis for correct operation. NCL circuits switch between DATA (data representation) and NULL (control representation). According to the Seitz's weak switching conditions: outputs transfer from DATA to NULL only until all inputs of a combinational circuit transition from DATA to NULL; likewise, outputs transfer from NULL to DATA only until all inputs of a combinational circuit transition from NULL to DATA; most delay insensitive systems (Singh et al,. 2002) have at least two register stages, including NCL (Smith, 2006). We call them NCL registers; they are connected by request and acknowledge handshaking signals (Ki and Ko). NCL uses symbolic completeness of expression to achieve self-timed behaviour. A symbolically complete expression is defined

as an expression that only depends on the relationships of the symbols present in the expression without a reference to the time of evaluation. Traditional Boolean logic is not symbolically complete; the output of a Boolean gate is only valid when referenced with time. NCL uses threshold gates with hysteresis for its composable logic elements. There are 27 threshold gates in all, as Table 1 shows. One type of threshold gate is the THmn gate, where $1 \leq m < n$. A THmn gate corresponds to an operator with at least m signals asserted as its set condition and all signals de-asserted as its reset condition. THmn gates have n inputs. At least m of the n inputs must be asserted before the output will become asserted. Because threshold gates are designed with hysteresis, all asserted inputs must be de-asserted before the output will be de-asserted. Hysteresis is used to provide a means for monotonic transitions and a complete transition of multi-rail inputs back to a NULL state before asserting the output associated with the next wavefront of input data. Fig. 3 shows the basic symbol of THmn gate and TH34w2 gate.

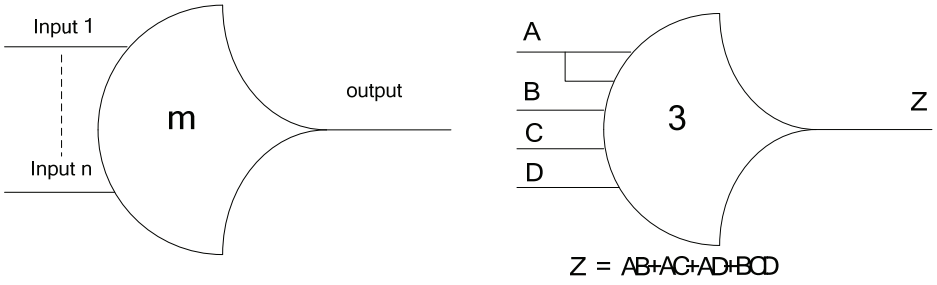


Fig. 3. THmn gate symbol

Another type of threshold gate is referred to as a weighted threshold gate. Weighted threshold gates are denoted as THmnWw_1..w_R. Weighted threshold gates have an integer value, $m = w_R > 1$, applied to inputR4. Here $1 = R < n$; where n is the number of inputs; m is the gate's threshold; and w_1, w_2,...w_R, are the integer weights of input1, input2, ...inputR, respectively. For example, consider a TH34W2 gate, whose n = 4 inputs are labeled A, B, C, and D. The weight of input A, W(A), is therefore 2. Since the gate's threshold, m, is 3, implies that in order for the output to be asserted, either inputs B, C, and D must all be asserted, or input A must be asserted and any other input, B, C, or D must also be asserted. NCL threshold gates may also include a reset input to initialize the gate's output.

NCL Gate	Function
TH12	A + B
TH22	AB
TH13	A + B + C
TH23	AB + AC + BC
TH33	ABC
TH23w2	A + BC
TH33w2	AB + AC
TH14	A + B + C + D
TH24	AB + AC + AD + BC + BD + CD
TH34	ABC + ABD + ACD + BCD
TH44	ABCD
TH24w2	A + BC + BD + CD
TH34w2	AB + AC + AD + BCD
TH44w2	ABC + ABD + ACD
TH34w3	A + BCD
TH44w3	AB + AC + AD
TH24w22	A + B + CD
TH34w22	AB + AC + AD + BC + BD
TH44w22	AB + ACD + BCD
TH54w22	ABC + ABD
TH34w32	A + BC + BD
TH54w32	AB + ACD
TH44w322	AB + AC + AD + BC
TH54w322	AB + AC + BCD
THxor0	AB + CD
THand0	AB + BC + AD
TH24comp	AC + BC + AD + BD

Table 1. 27 Fundamental NCL Gates (Smith, 2006).

By employing threshold gates for each logic rail, NCL is able to determine the output status without referencing time. NCL switches monotonically. All combinational logic clouds switch from a condition of all NULL to a condition where some of the wires are at DATA. Once a wire has switched to DATA, it remains there until the combinational logic cloud is returned to NULL state. This monotonic switching behaviour makes NCL designs glitch-free. Because every threshold element in NCL exhibits inherent storage behaviour due to hysteresis, storage cells such as latches or flip-flops are unnecessary. Data can be pipelined by creating registration elements from the gate primitives. Using these same registration elements, data can be fed back on itself to create sequencers in much the same way sequencers are built in traditional logic. Data Output: The proper conditions have been met to provide DATA at the output of the registration element.

4. Proposed Architecture-Clock Free Nanowire Crossbar Architecture

4.1 Programmable Gate Macro Block (PGMB)

The primitive unit of the proposed architecture is referred to as the Programmable Gate Macro Block (PGMB). Each block is made of one AND-plane and one OR-plane formed by the diode crossbars. Vertical nanowires with pull-up resistors form product terms and horizontal wires with pull-down resistor add them using OR-logic. So, each PGMB can be programmed to embed a NCL gate macro function in SOP (Sum-Of-Product) form. It also

has a feedback loop which drives the output back to an input wire. The maximum number of inputs to any threshold gate is 4. Along with this, it needs a feedback to implement any of the 27 threshold gate macros. Notably, there is no need to add complemented primary inputs as done in NanoFabric (Goldstein et al., 2001), since inversion can be easily done in NCL by crossing a dual-rail signal, D, consists of two wires, D^0 and D^1. Therefore, there is no need to use FET-based (Field Effect Transistor) inverters as done in NanoPLA (Dehon et al,. 2003). This unique encoding-level inversion capability is another significant merit of the proposed architecture. Fig.4 shows the programming of TH23 and TH34W2 gates on PGMB where small dots represent activated programmable crosspoints. The output of the TH23 gate is given by the logic Z = AB + BC + CA + (A+B+C)Z', where Z' is the previous output of the TH23 gate which is fed back to the input nanowire. Likewise, the output of the TH34W2 gate is given by the logic Z = AB + AC + AD + BCD + (A + B + C + D)Z'. Notably, these two gates embedded in PGMB are core building blocks in NCL full adder design, which will be discussed in the next subsection.

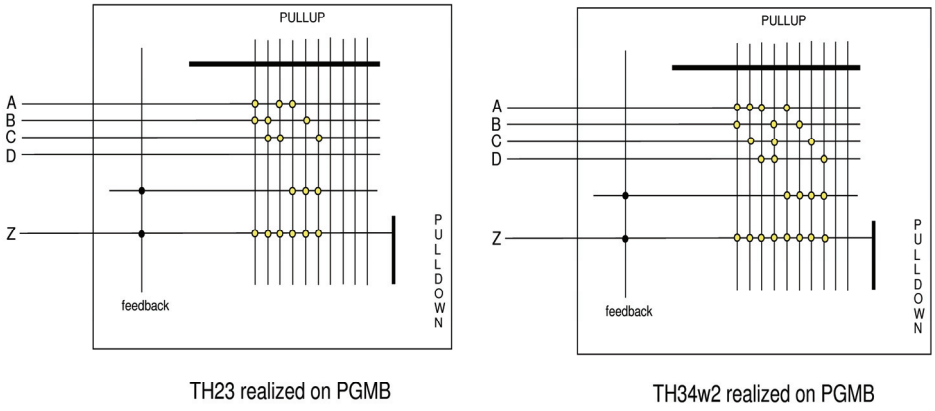


Fig. 4. TH23 and TH34w2 gates configured on PGMB. All 27 gate macros can be programmed on PGMB and additional row nanowires and column nanowires can be added as redundancy in case of high defect rate.

4.2 Addressing Decoding strategy

Snider et. al. (Snider et al., 2005) has show that the demultiplexers are expected to be the key components in interfacing submicrometer-scale and nano-scale electronic circuit. Demultiplexers are used for interfacing nanoelectronic circuits and outside conventional CMOS (Kuekes et al, 2001). We can through a small number of micro-scale wires to control a large number of nano-scale wires by forcing a 'selected' or 'unselected' voltage onto nanowires (Snider et al,. 2005). In our proposed clock-free nanowire crossbar architecture, we use it to control programmable junction(crosspoints) of different PGMBs. Fig.5 shows the implementation of a demultiplexer forming by a diode crossbar and resistors. The address lines of the demultiplexer are micro scale wires which could be implemented on a CMOS substrate, and the output lines are nanowires (Snider et al,. 2005). The relationship between the number of input microwires(M) and the number of output nanowires(N) is $N=2^M$; If we have eight outputs which need at least three inputs for constituting a

demultiplexer. For example, each PGMB's size is 6*10, for covering all crosspoints which can be programed, demultiplexers located on rows and columns are needed respectively. See Fig.6.

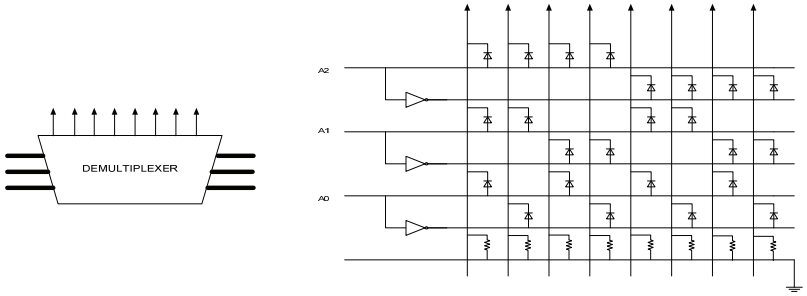


Fig. 5. A demultiplexer consist of diode crossbar and resisors,input address is A2,A1,A0 (Snider et al., 2005).

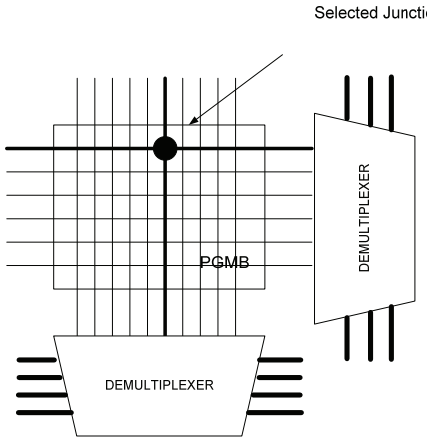


Fig. 6. Two demultiplexers control one PGMB for selecting different crosspoints.

4.3 Configurable Logic Block(CLB)

For the purpose of reconfiguring all of the 27 threshold gates arbitrarily, we need to design a reconfigurable architecture to make it work no matter what kind of logic those gates possess. We designed three different Configurable Logic Blocks (CLB) and named them CLB-1, CLB-2, CLB-3. Fig.7 shows that each CLB consists of four PGMBs which are surrounded by nanowires, and demultiplexers. Horizontal nanowires cross over vertical nanowires forming the crosspoints. In order to utilize the programmability of crosspoints for configuring logic functions, we use different demultiplexers to select them. Output signals from the top demultiplexer determines which nanowires are selected as the input signals of the PGMB. The demultiplexer along adjacent edges of nanowire/nanowire crossbar is used for selecting different PGMBs to receive the input signals. We program each crosspoint through controlling the up-side and right-side demultiplexer at the same time. For example, we program the crosspoints to ON state or OFF state by driving the up-side demultiplexer's

output nanowires with positive voltage and driving the right-side demultiplexer's output with negative voltage. In this case we assume those unselected outputs are driven with ground, then the intersection of driven nanowires have a voltage drop which is different from other unselected crosspoints in the crossbar. This allows us to configure crosspoints sensitive to voltage drops across them (Snider et al,. 2005)..

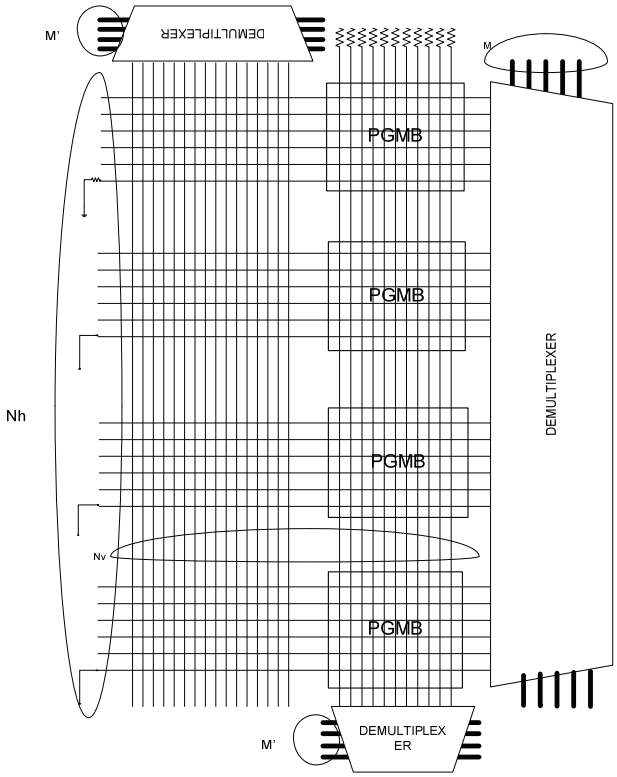


Fig. 7. The first version of Configurable Logic Block (CLB-1).

Fig.8 shows the other two versions of CLB. The main difference between them is the number of demultiplexers. For CLB-1, there is only one demultiplexer on the right-side, so only one crosspoint can be selected at one time. For CLB-2, there are two demultiplexers on the right side, so two crosspoints can be selected at the same time, however, the location of the two crosspoints have to be in the different group of PGMBs(assume the above two PGMBs are group1, another two PGMBs are group2). For CLB-3, there are four demultiplexers on the right side, so four crosspoints can be selected at the same time, the precondition is four crosspoints from the four different PGMBs respectively. Compare above mentioned three types of CLB architecture, it is possible to see the advantage of those who has less demultiplexers is easier to control, but it need more time to access crosspoints. We will analysis and compare the detail data of different CLBs in Mathematical Analysis section.

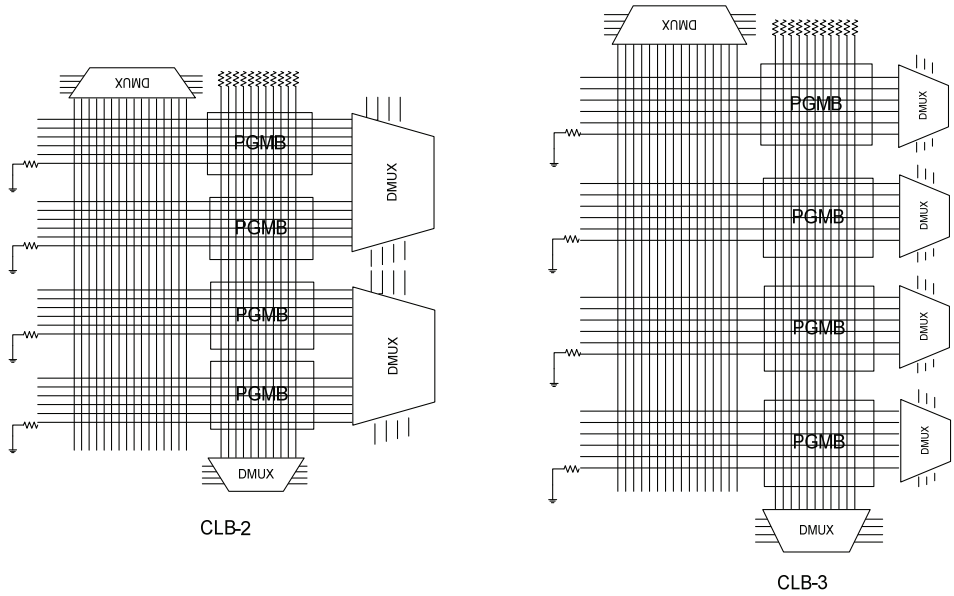


Fig. 8. The second and third version of Configurable Logic Block(CLB-2 and CLB-3).

4.4 Case Study: Asynchronous Full Adder

A full adder can be implemented by using threshold gates as shown in Fig.9.

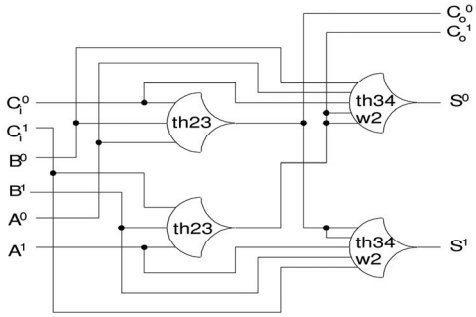


Fig. 9. 1-bit full adder implement by threshold gates (Smith, 2006).

As the figure shows, it consists of two TH23 gates and two TH34w2 gates. This requires three inputs which are X, Y for addition and Ci for carry in, two outputs are summation S and carry-out Co. These bits are represented by X^0, X^1, Y^0, Y^1, Ci^0, Ci^1, Co^0, Co^1, S^0, S^1 encoded in dual rail logic respectively; then we use our proposed architecture to implement 1-bit full adder.

As mentioned above, crosspoints can be programmed by demultiplexers, and the input signals can be routed to the corresponding PGMBs. Co^0 and Co^1 are the output signals of

TH23 gates; they are also the input signals of TH34w2 gates. Fig.10 shows these two signals are propagated to the TH34w2 PGMBs and located in the free nanowires.

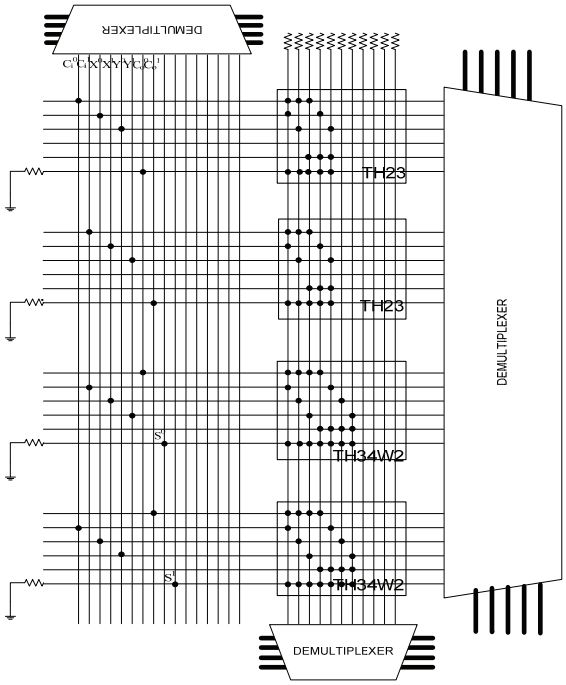


Fig. 10. 1-bit full adder implement by CLB-1.

4.5 NCL Register

NCL Register stage consists of three threshold gates. Fig.11 shows how the 1-bit NCL register is implemented by threshold gates. There are two TH22 gates are used to generate a handshaking signal which helps in synchronizing the circuit. There are two signals: 'REQUEST FOR DATA' and 'REQUEST FOR NULL' generated by the registers that are passed on the previous register. Ki (input of successive stage) and Ko (output of previous stage/input of next stage) are the handshaking signals, I^0, I^1, O^0, O^1 are input and output data rails respectively.

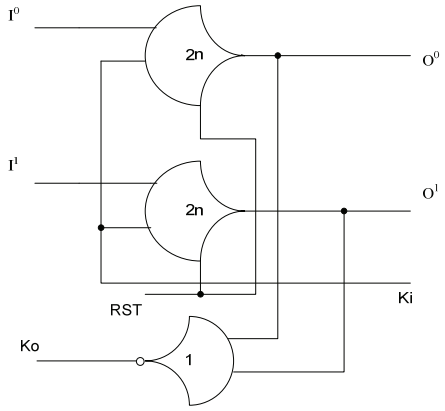


Fig. 11. 1-bit NCL register implement by threshold gates.

Fig.12 shows the 1-bit NCL register is implemented by our proposed NRCA. As the same routing method of 1-bit full adder, we route required signals into different crosspoints on the crossbar, and get the logic output. Since Ko is the invert value of output, the inversion can be easily done in NCL by crossing a dual-rail signal. For example, when the output is Boolean logic 0 ($Ko^0 = 1$, $Ko^1 = 0$), after crossing the dual-rail signal, it becomes ($Ko^0 = 0$, $Ko^1 = 1$), which corresponds the Boolean logic 1, output data successfully be inverted.

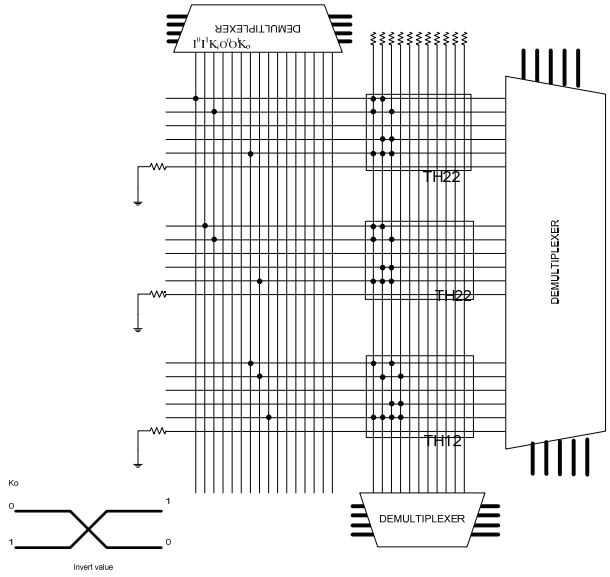


Fig. 12. 1-bit NCL register implement by CLB-1.

4.6 Connection between Combination Logic and NCL Registers

As NULL Convention Logic for Clock-Less computing section mentioned, we need handshaking signals to connect two delay-insensitive NCL registers with delay-insensitive combination logic circuit (Fig.13).

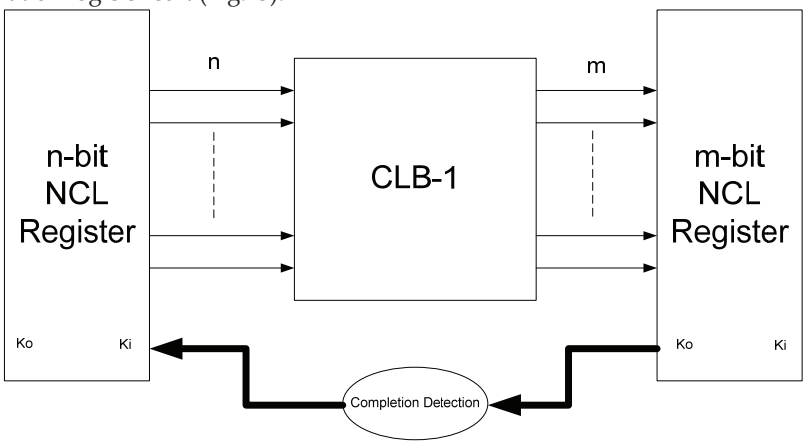


Fig. 13. NCL combinational logic block with input and output registrations.

The case of Asynchronous Full Adder can be used as an example to explain the details (Fig.14). For a 1-bit full adder, there are three inputs X^0, X^1, Y^0, Y^1, Ci^0, Ci^1 encoded in dual rail logic respectively, 3-bit NCL register is needed; also because the full adder has two outputs Co^0, Co^1, S^0, S^1 encoded in dual rail logic respectively, so another 2-bit NCL register is needed. Assume the initial value of Ki is zero, and then we get all zero for the outputs from 3-bit NCL register, now the system is on NULL state; at the same time, 2-bit NCL register which accepts outputs from full adder are all zero. Therefore the feedback handshaking signal Ko becomes 1. After feedback it to the previous 3-bit NCL register as an updated Ki, the system state makes a transition to the DATA state, finally we can get the correct output result of Full Adder. The input registration alternates NULL and DATA waves and those waves are initiated by the output registration's feedback signal.

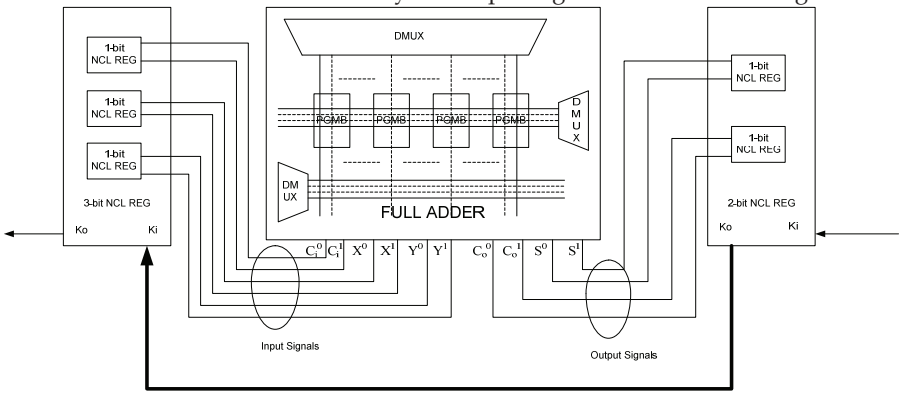


Fig. 14. Full Adder implementation with input and output registrations.

5. Hierarchical Architecture

Hierarchical interconnection architecture has been widely chosen by FPGA designs due to the easiness of programming and flexible routing. The proposed asynchronous NRCA is also based on hierarchical structure to interconnect PGMBs. The first level is Configurable Logic Blocks (CLB), which consists of 4 PGMBs, which is level 1. Level 2 consists of four CLBs using the same principle. Level 3 consists of four 4*CLBs and level 4 consists of four 16*CLBs and so on. The number of levels depends on what kind of logic circuit are processing. Level 1 has been introduced in the previous section. Hierarchical structures used at higher levels are described in this section.

5.1 Level 2 – 4*CLB

Each CLB has 16 distinct inputs and 4 outputs, the architecture shows on Fig.15. Interconnect switching block is used for interconnecting the input/output signals to the different blocks or up to the next level. A 4-bit NCL register would be a good design example for explaining this architecture. Fig.16 shows how to implement the logic circuit by using 4*CLB architecture. There are D^0~D^3 and Q^0~Q^3 are input and output data rails respectively, Ki and Ko are handshaking signals.

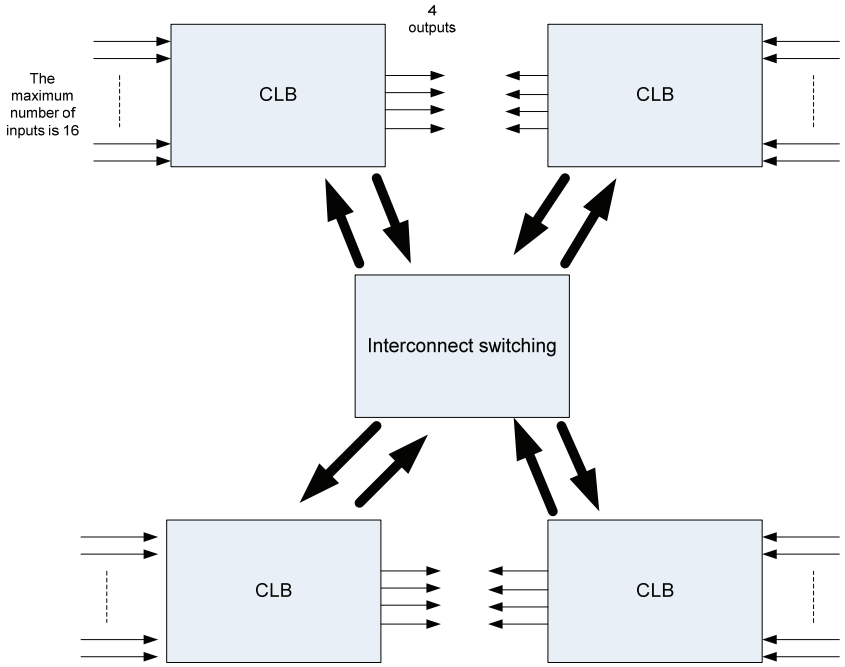


Fig. 15. level-2:4*CLB architecture.

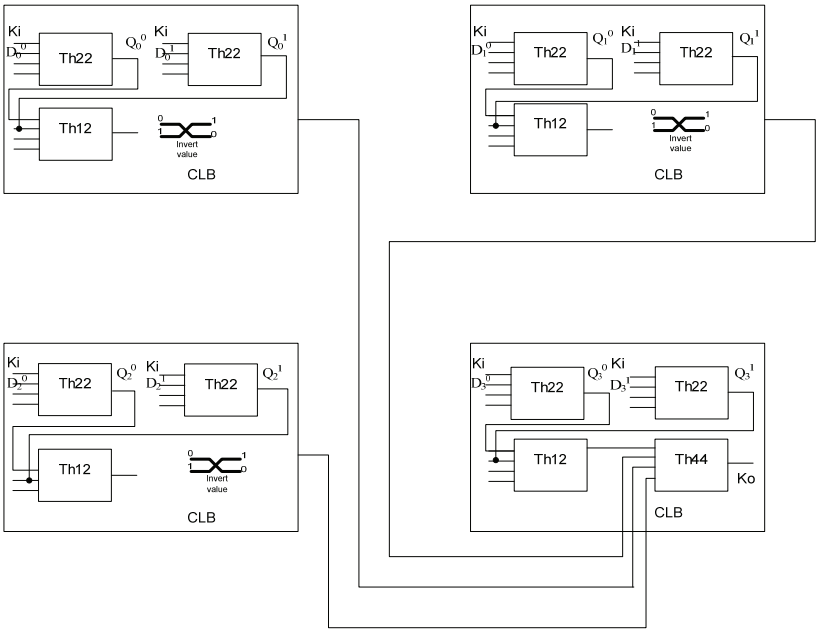


Fig. 16. 4-bit NCL register implemented at level-2 structure.

5.2 Hierarchical Interconnect Switching

The design of Hierarchical Interconnect Switching is one of key elements for the whole system, there are several programmable interconnect strategies used by various designers, one of very popular interconnect technique is pass transistor interconnect. We use this technique for our proposed architecture so that the interconnect switching system can be controlled flexibility through different arrangement of pass transistors. It constitutes by transistors and latches, the latch is used for storing the gate input of the transistor, we can also use SRAM to store the value for the switch input, and it can be programmed during the configuration phase (Singh et al., 2002).

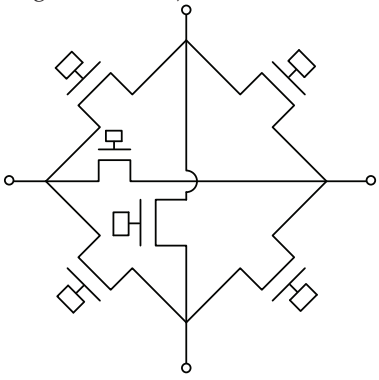


Fig. 17. Six Pass Transistors Interconnects (Lee et al., 1999).

Fig. 17 shows the different arrangements of various interconnect schemes using pass transistors. This scheme contains 6 transistors and for each interconnect point, it would has 3 different directions to be chosen. According to this property, we can build up our own interconnect switching matrix for routing a set of inputs to a set of outputs (Fig.18).

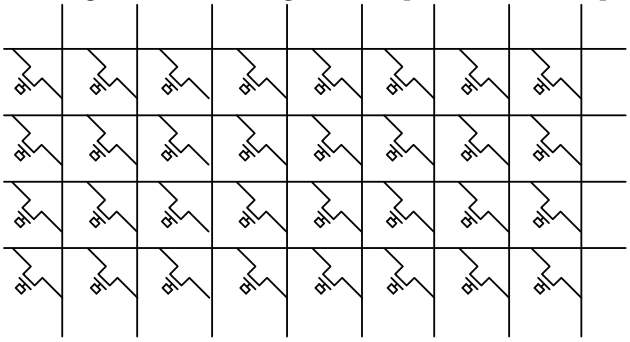


Fig. 18. 8*4 Switching Matrix

6. Parameters

Cost of the final manufactured product is become more and more important to a digital electronics. There are numerous factors such as area, manufacturability and so on would influence the cost. So analyzing them from mathematical angle is necessary. To show the design trade-offs of the three CLBs design, two difference benchmarks will be used; namely area performance and execution complexities. Section VI models the area, execution steps and timing issues of these structures.

6.1 Area

The following feature size parameters are used in this chapter:

- N_v: The number of nanowires of vertical.
- N_h: The number of nanowires of horizontal.
- M : The number of microwires of demultiplexers.
- N_d: The number of nanowires which outputs from demultiplexers.
- P_n: The nanowire pitch for nanowires which are inputs to diodes.
- P_m: The microwire pitch for microwires.
- D_n: Diameter(width) of each nanowires.
- D_m: Diameter(width) of each microwires.
- A_d: The area of each demultiplexers.

For the 45nm node, the lithographic interconnect pitch is 105nm [ITRS 2001]. Technology developments suggest we can build and assemble 10nm pitch nanowires with crosspoints at every nanowire-nanowire crossing (Dehon et al., 2003). So P_n = 10nm, P_m = 105nm. And nanowires with diameters down to 3nm have been demonstrated (Cui et al. 2001).

Compare the area from different design of CLBs from Fig.7.and Fig.8., we can see the basic area composition of each tile.

$$\text{Width} = Pn(Nv-1) + Pm(M-1) \quad (1)$$

$$\text{Height: } Pn(Nh-1) + Pm(M'-1)*2 \quad (2)$$

$$\text{Area} = \text{Width} * \text{Height} \quad (3)$$

The only difference between CLB-1,CLB-2,CLB-3 is the number of demultiplexers. So "M" (number of microwires of demultiplexers on the right side) is different fro m each other (M=5(CLB-1); M=4(CLB-2); M=3(CLB-3)). Fig.19 shows the area comp arison of different design of CLBs, the CLB-3 is the smallest since the four demultip lexers can be located on the same plate by using the assembly techniques.

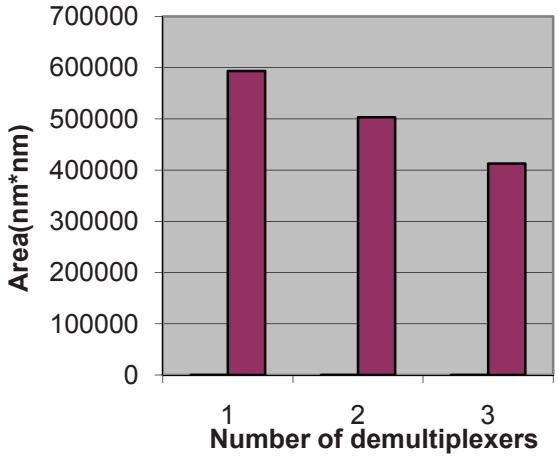


Fig. 19. Area comparison of three different designs of CLBs

6.2 Execute steps of each CLB

For improving the design, we not only take into account the area, we also consider about the complexity of the operation. One of factors is the number of steps for processing a logic circuit. Because it is directly impact the efficiency of design. The relationship between numbers of crosspoints needed to be programmed and the average steps for processing all of them is shown in Table 2. It shows that as increasing of number of points, the average steps are also increasing, but for different CLBs, the average steps are also different.

Numbers of Points	Average Steps		
	CLB-1	CLB-2	CLB-3
2	2	1.947	1.947
4	4	3.684	3.555
6	6	5.2105	4.926
8	8	6.5263	6.0625
10	10	7.6316	7.0013

Table 2. Execution steps comparison of different designs of CLBs

CLB-2 and CLB-3 have 2 and 4 demultiplexers on the right side respectively. We distinguish the first two PGMBs as a group, and the last two PGMBs as another group. For the CLB-2; as Fig.8 shows, it has capability of programming 2 points at the same time, only when the 2 points are distributed to the 2 different groups respectively, because each multiplexer controls 2 PGMBs at one time. Applying the same principle, the CLB-3 has the probability of programming 4 points at the same time, and the precondition is that 2 points have to be distributed to each PGMBs respectively because each demultiplexer can control only one PGMB.

7. Testing and Defect & Fault – Tolerance

Defects and faults are the worst problems with the nanoscale integration. Conventional fault-tolerant and reliable design techniques are primarily optimized for highly stable process technologies where reliability characterization of chip materials and circuit operation are well understood. However, conventional methods will become inadequate in the near future due to increased defect and failure rates, especially in case of nanoscale integration. New research in multi-scale fault-tolerance and reliability techniques is, therefore, critically needed for commercial adoption of emergent nanotechnologies (Heath et al., 1998). It is anticipated that NRCA will have significantly higher defect density than CMOS, as high as 10% (Huang et al., 2004; Jacome et al., 2004). High-density fabrication could potentially be very inexpensive if researchers can actualize a chemical self-assembly, but such a circuit would require laborious testing, diagnosis, repair and reconfiguration processes, implying a significant overhead cost, as well. Crossbar arrays and supporting nanoscale interfaces are also susceptible to transient faults that are usually caused by interferences such as crosstalk and radiation. Thus, reliability of NRCA is also an issue as the size of devices shrink and the integration density increases.

7.1 Defect-Unaware Testing and Reconfiguration

For the proposed clock-free nanowire crossbar architecture, we will take a different direction to overcome its defects. When we embed a large-scale system on the proposed clockless nanowire crossbar system, it is certainly time-consuming and impractical to scan through all programmable crosspoints of the given nanowire crossbars to locate defects. Instead, the given physical system design can be placed and routed without being aware of defects. Then, each PGMB can be functionally tested to significantly reduce the test space.

After testing, PGMBs and switching resources affected by those defects (i.e., stuck-OPEN crosspoints) can be further reconfigured to tolerate them. There are three possible ways to tolerate defects in PGMB. The first way is to reconfigure the order of primary input variables by reconfiguring the incoming switching block. In case of TH23, there are 3!=6 different ways to rearrange the order of inputs. Any order that can be used to avoid defective crosspoint(s) in the first three rows can be selected. The second way is to rearrange the order of columns of the given PGMB. Since all threshold gate macros have a simple Sum-of-Product (SoP) structure, by the commutative law, the order of these product terms can be rearranged without affecting the functionality of the gate. The third way is to include redundant rows and columns in each PGMB. Any row or column with an excessive number of defects that cannot be tolerated by the first two methods can be simply purged out by replacing it with a spare wire in this case. In case the PGMB under consideration is too

defective to repair, it should be discarded and another PGMB should be used to realize the given threshold gate function.

7.2 Defect - Aware Approach

Defect Aware Approach needs to generate a defect map for the given PGMB, and compare the pattern of defect free crosspoints with the required pattern. This approach makes use of the available redundant crosspoints to instead of the defective crosspoints on each PGMB. According to the architecture of PGMB, columns are used for the AND plane, it can be interchanged without affecting functionality, that means for each threshold gate, it can be represented in several different ways without affecting the logic functionality. The defect rate of this approach is lower than Defect Unaware Approach, but the drawback being it needs more time for testing and reconfiguration because each PGMB has to be tested to locate all defective crosspoints, and make sure avoid them when the netlist is actually placed and routed.

The proposed clock-free architecture uses NCL encoding. Stuck-at-1 faults simply halt the circuit, since the NCL circuit cannot make a transition from DATA (either 01 or 10) to NULL (00) - therefore, can be easily screened out. Also, in case of dual-rail encoding, 11 is considered as an invalid code. So, any permanent or transient fault that results in this invalid codeword can be eventually detected as faulty. Only stuck-at-0 faults and some other transient faults need to be exercised with applied patterns.

7.3 Parametric Simulation Results

The illustration of the behaviour of programmability with varying defect rates for both the techniques is illustrated in Fig.20. The defect-aware approach surely outperforms the defect-unaware approach, but the defect-unaware approach still shows good programmability when the defect rate is considerably lower.

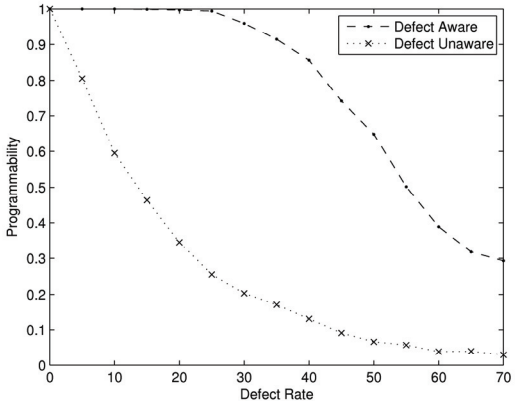


Fig. 20. Programmability comparison for defect-aware and -unaware approaches

It also shows almost all the gates are programmed even at 30% defect rate and then programmability (ratio of successfully programmed PGMB's to the total number of PGMB's) reduces, which is inevitable. This is much better in terms of programmability as compared

to the defect aware technique but it takes time to generate defect map and then perform mapping and placement, but is very good at utilizing the inherent redundancy. This helps when we have a highly defective fabrication process; the algorithm bypasses the defective crosspoints and places the crosspoints without affecting the functionality of the gate. For both the techniques 20% of the rows have been dedicated to the OR crossbar logic. When the defect rate is considerably lower than 5%, the defect-unaware approach still shows good programmability.

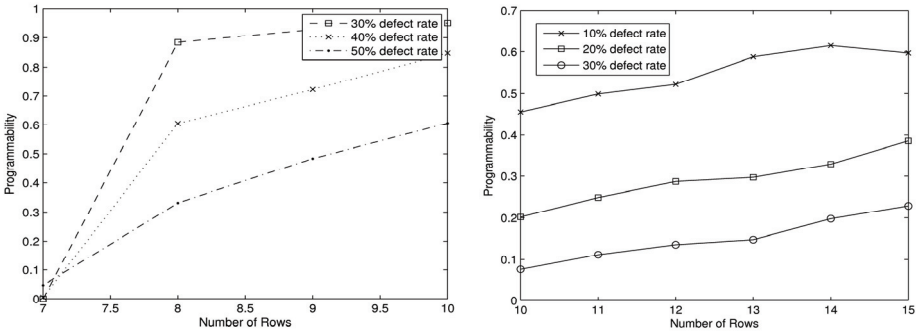


Fig. 21. Programmability VS Number of rows for defect-unaware and defect-aware.

Figure 21. provides useful analysis for the behaviour of programmability with change in dimension of the PGMB at different defect rates. These graphs also indicate the superiority of defect aware approach over the unaware approach.

7.4 Functional Test Algorithm (FTA)

The proposed functional test algorithm is a post configuration test scheme which makes use of the Boolean function of the threshold gate being implemented to test the programmable crosspoint locations on the PGMB. Each THmn gate has its own distinctive programmable co-ordinate locations. The proposed test scheme aims to test only those programmable ON crosspoints in the given programmable PGMB. This algorithm uses the functional expression that is unique to each THmn gate. The functional test scheme uses "test tuples" for the purpose of testing the programmable crosspoints. Test tuples are joint combinations of input bit patterns and previously asserted output. Table 3 can be used to clearly understand this concept. Consider the implementation of TH23 gate. Assume there is a fault at the coordinate location (1, 3). The fault at this ON point gives a faulty output F*=1 when input 001 is used. The desired output in case there is no fault at any crosspoint is F=F'(the previously asserted output). In case the previously asserted output is set to 0 and followed up with input pattern 001, it will be possible to stimulate the fault. By using a combination of F' and input bits, we can detect faults in the programmable crosspoints. These sets of inputs used to detect the faulty crosspoint locations are called "test tuples". Test tuples having one to one correspondence with the programmable cross- points are called lower order test tuples. Higher order test tuples can test for defects in more than a single crosspoint simultaneously.

ABC	F	F*(1,1)	F*(1,3)	F*(1,4)	F*(2,1)	F*(2,2)	F*(2,5)	F*(3,2)	F*(3,3)	F*(3,6)
000	0	0	0	F′=1	0	0	F′=1	0	0	F′=1
001	F′	F′	1	F′	F′	1	F′	F′	F′	F′
010	F′	1	F′	F′	F′	F′	F′	1	F′	F′
011	1	1	1	1	1	1	1	1	1	1
100	F′	F′	F′	F′	1	F′	F′	F′	F′	1
101	1	1	1	1	1	1	1	1	1	1
110	1	1	1	1	1	1	1	1	1	1
111	1	1	1	1	1	1	1	1	1	1
ABC	F	F*(5,4)	F*(5,5)	F*(5,6)	F*(6,1)	F*(6,2)	F*(6,3)	F*(6,4)	F*(6,5)	F*(6,6)
000	0	0	0	0	0	0	0	0	0	0
001	F′	F′	F′	1	F′	F′	F′	F′	F′	0
010	F′	F′	1	F′	F′	F′	F′	F′	0	F′
011	1	1	1	1	1	1	1	1	1	1
100	F′	F′	F′	F′	F′	F′	F′	0	F′	F′
101	1	1	1	1	1	1	F′=0	1	1	1
110	1	1	1	1	1	F′=0	1	1	1	1
111	1	1	1	1	1	1	1	1	1	1

Table 3. Truth Table for TH23 Gate and all faulty functions that can be resulted from single crosspoint defect

A pseudo code for the functional test algorithm is described in Table 4.

START
1. Define PGMB dimensions and threshold gate to be programmed.
2. Map THmn gate onto the PGMB and generate the corresponding truth table.
3. IF OR plane defect detection is the prime objective THEN
   Set test tuples with low priority tuples having direct correspondence with programmable crosspoints to be used ahead of the high priority ones.
   ELSE
   Set test Tuples In order of decreasing priority levels.
4. Use next test input tuple to check for programmable defect.
5. IF False output is observed THEN
   Increment fault count and note possible defect location.
   ELSE
   The tested location is defect free.
6. Go to Setp 4.
7. Summarize the results of the test.

STOP

Table 4. Pseudo code describing the Functional Test Algorithm

7.5 Fault-Tolerant Placement Schemes Using FTA

Table 5 shows the number of OR locations utilized to implement few of the commonly and importantly used THmn gates.

TH12	TH23	TH24	TH34	TH33w2	TH34w2	TH44w2
4	6	10	8	5	8	7

Table 5. The number of programmable OR locations for THmn gates

Having studied the mapping patterns of THmn gates and defect distributions, it has been noticed that the OR plane is vulnerable to have a physical defect overlap with a programmable ON location of any threshold gate macro. Since a majority of programmable ON crosspoints fall on a single OR plane, it is essential to ensure OR plane reliability. With the inclusion of a redundant OR wire, the reliability of the OR plane can be enhanced. With the inclusion of a redundant OR wire, in case an OR point is defective, the connection can be moved to the redundant wire without programming other crosspoints in the column which contribute to the product term. Another advantage of introducing the OR plane is that since the OR planes are ORed together, the realization is not altered in any manner. As far as testing overheads are concerned, with the addition of a redundant row, only single additional input test tuple needs to be used to test the single OR location. If redundant OR row is not introduced, in case of a defect at OR location, the entire column will have to be moved to another location and all the corresponding crosspoint locations will have to be tested using additional test tuples for defects. Not only will the number of programmable locations increase with this approach, but the testing space will also increase drastically. In case of some of the THmn gates such as TH12, TH23w2 where no more than 50% of the programmable OR locations are used, it would be better to rearrange the columns instead of using a redundant OR row. In this section, different modelling and placement schemes that could be used to address these mapping issues are presented.

8. Conclusion

In this chapter, we introduced our new proposed clock-free nanowire crossbar architecture based on delay-insensitive logic known as Null Convention Logic. The complex clock distribution network can be removed from the hardware and many clocking-related failure modes can be intrinsically eliminated by the proposed clock-free architecture. Based on this architecture, we can construct a hierarchical architecture for executing more complex logic function, such as the 4-bit NCL register and so on. For defect tolerance, we proposed two mapping and placement techniques, namely Defect Unaware Approach and Defect Aware Approach. Even though the defect-aware approach is better than the defect-unaware approach especially when the defect rate is higher, it requires much more laborious testing (i.e., each PGMB should be tested to locate all defective crosspoints) and reconfiguration (i.e., all defective crosspoints should be avoided when the netlist is actually placed and routed) tasks. However, the defect-unaware approach can be simpler since the netlist is directly mapped without considering any defects. After that, PGMBs and can be functionally tested to locate ones with faults. These faulty ones then can be tested and reconfigured to avoid defects. Our future direction is to develop automated design optimization tools, testing schemes, and maximize the utility of PGMBs in spite of the inherent fabrication defects.

9. References

Bachtold, A.; Hadley, P; Nakanishi, T & Dekker, C. (2001). Logic circuits with carbon nanotube transistors, *Science*, vol. 294, pp. 1317-1320.

Bourianoff, G. (2003). The future of nanocomputing, *IEEE Computer,* Vol. 38, No. 8, pp. 44-53.

Bhattacharya, M. & Mazumder, P. (2001). Augmentation of SPICE for simulation of circuits containing resonant tunneling diodes, *IEEE Trans. Computer-Aided Design,* vol. 20, pp. 39-50.

Bandapati, S.K.; Smith, S.C. & Choi, M. (2003). Design and Characterization of Null Convention Self-Timed Multipliers, *IEEE Design and Test of Computers,* vol. 20, no. 6, pp. 26-36.

Chen, J.; Reed, M.A.; Rawlett, A.M. & Tour, J.M. (1999). Large on-off ratios and negative differential resistance in a molecular electronic device, *Science*, vol. 286, pp. 1550-1552.

Chen, J.; Wang, W.; Klemic, J. Reed, M.A. & Axelrod, B.W. (2002). Molecular wires, switches and memories, *Annals Of The New York Academy Of Sciences,* Vol. 960, pp. 69-99.

Chen, Y.; Ohlberg, D. & Li, X. et. al. (2003). Nanoscale molecular-switch devices fabricated by imprint lithography, *Applied Physics Letters,* Vol.82, No. 10, pp. 1610-1612.

Chen, Y.; Jung, G.; Ohlberg, D. Li, X. et. al. (2003). Nanoscale molecular-switch crossbar circuits, *Nanotechnology*, no. 14, pp. 462-468.

Culbertson, W.B.; Amerson, R.; Carter, R.; Kuekes, P. & Snider, G. (1997). Defect tolerance on the TeraMAC custom computer, I*EEE Symp. FPGAs Custom Computing Machines,* pp. 116-124.

Cui, Y. & Lieber, C.M. (2001). Functional nanoscale electronic devices assembled using silicon nanowire building blocks, *Science*, vol. 291, pp. 851-853.

DeHon, A. (2003). Array-Based Architecture for FET-Based, Nanoscale Electronics, *IEEE Transactions on Nanotechnology*, Vol. 2, No. 1, pp. 23-32.

Dekker, C. (1999). Carbon nanotubes as molecular quantum wires, *Physics Today*, pp. 22-28.

Derycke, V.; Martel, R.; Appenzeller, J. & Avouris, Ph. (2001). Carbon nanotube inter- and intramolecular logic gates, *Nano Letter,* Vol. 1, No. 9, pp. 453-456.

Durbeck, L.J.K. & Macias, N.J. (2001). The cell matrix: an architecture for nanocomputing, *Nanotechnology*, Vol. 12, pp. 217-320.

Fant, K.M. & Brandt, S.A. (1996). NULL Convention Logic: a complete and consistent logic for asynchronous digital circuit synthesis, *International Conference on Application Specific Systems, Architectures and Processors,* pp. 261 – 273.

Finkelstein, H.; Asbeck, P.M. & Esener, S. (2003). Architecture and analysis of a self-assembled 3D array of carbon nanotubes and molecular memories, *IEEE Conference on Nanotechnology*, pp. 12-14.

Franzon, P. & Nackashi, D. (2000). Moletronics: a circuit design perspective, *Int. Conf. SPIE Smart Electronics and MEMS,* vol. 4236, pp.80-88.

Goldstein, S.C. & Budiu, M. (2001). Nanofabrics: spatial computing using molecular nanoelectronics, *Proc. 28th Int. Symp. Computer Architecture,* 2001, pp. 178-189.

Goldstein, S.C. & Rosewater, D. (2002). Digital logic using molecular electronics, *Int. Solid-State Circuits Conf.,* pp. 125.

Heath, J.R.; Kuekes, P.J.; Snider, G.S. & Williams, R.S. (1998). A defect- tolerant computer architecture: Opportunities for nanotechnology, *Science*, Vol. 280, pp. 1716-1721.

Hiroshima, H.; Komuro, M.; Kasahara, N. & Kurashima, Y. (2002). Pattern defects of nanoimprint in atmospheric conditions, *International Microprocesses and Nanotechnology Conference*, pp. 22-23.

Huang, Y.; Duan, X.; Wei, Q. & Lieber, C.M. (2001). Directed assemble of one-dimensional nanostructures into functional networks, *Science*, Vol. 291, pp. 630-633.

Huang, Y.; Duan, X.; Cui, Y.; Lauhon, L.J.; Kim, K. & Lieber, C.M. (2001). Logic gates and computation from assembled nanowire building blocks, *Science*, vol. 294, pp. 1313-1316.

Huang, J.; Tahoori, M.B. & Lombardi, F. (2004). On the defect tolerance of nano-scale two-dimensional crossbars, IEEE International Symposium on Defect and Fault Tolerance in VLSISystems, pp. 96-104.

Jacome, M.; He, C.; Veciana, G. & Bijansky, S. (2004). Defect tolerant probabilistic design paradigm for nanotechnologies, *IEEE/ACM Design Automation Conference (DAC),* pp. 1-6.

Kuekes, P.J.; Heath, J.R. & Williams, R.S. (2001). Molecular-wire crossbar interconnect (MWCI) for signal routing and communications, *U.S. Patent 6 314 019.*

Kuekes, P.J.; & Williams, R.S. (2001). Demultimplexer for a molecular wire crossbar network (MWCN DEMUX), *U.S. Patent 6 256 767.*

Lee,H. & Flynn, M.J.(1999). High-Speed Interconnect Schemes for a Pipelined FPGA, *Technical Report, Stanford University,* UMI Order Number: CSL-TR-99-786.

Mirkin, C. (2000). Programming the assembly of two- and three-dimensional architectures with DNA and nanoscale inorganic building blocks, *Inorganic Chemistry*, vol. 39, pp. 2258-2272.

Mishra, M. & Goldstein, S. (2002). Scalable defect tolerance for molecular electronics, *Workshop Non-Silicon Computation (NSC-1),* pp. 78.

Mcalpine, M.C.; Friedman, R.S. & Lieber, C.M. (2005). High-Performance Nanowire Electronics and Photonics and Nanoscale Patterning on Flexible Plastic Substrates, *Proceedings of the IEEE,* Vol. 93, No. 7, pp. 1357-1363.

Melosh, N.; Boukai, A.; Diana, F. & Gerardot, B. (2003). Ultrahigh-density nanowire lattices and circuits, *Science*, Vol. 300, pp. 112-115.

Naeimi, H. & Dehon, A. (2004). A greedy algorithm for tolerating defective crosspoints in nanoPLA design, *IEEE International Conference on Field-Programmable Technology*, pp. 49-56.

Nowick, S.M. & Dill, D.L. (1991). Synthesis of Asynchronous State Machines Using a Local Clock, *Proceedings of ICCAD,* pp.192-197.

Nicewarner-Peana, S.R.; Raina, S.; Goodrich, G.P.; Fedoroff, N.V. & Keating, C.D. (2002). Hybridization and extension of Au nanoparticle-bound oligonucleotides, J*ournal of American Chem. Soc.,* vol. 124, pp. 7314-7323.

Paulson, L.D. (2005). Researchers work on transistor successor, *IEEE Computer*, Vol. 38, No. 5, pp. 17.

Rueckes, T.; Kim, K.; Joselevich, E.; Tseng, G.; Cheung, C. & Lieber, C. (2000). Carbon nanotube based nonvolatile random access memory for molecular computing, *Science*, vol. 289, pp. 94-97.

Soh, C.; Quate, C. & Morpurgo, C. (1999). Integrated nanotube circuits: controlled growth and ohmic contacting of single-walled carbon nanotubes. *Appled Physics Letter,* vol. 75, no. 5, pp. 627-629.

Sutherland, Ivan E. (1989). Micropipelines, *Communications of the ACM,* Vol. 32, No. 6, pp. 720-738.

Stan, M. (2001). A scaling scenario for nanoelectronic technologies, Georgia Tech Conf. Nanoscience and Nanotechnology, pp.103.

Smith, S.C.; DeMara, R.F.; Yuan, J.S.; Ferguson, D. & Lamb, D. (2004). Optimization of NULL Convention Self-Timed Circuits, *Integration, The VLSI Journal,* Vol. 37, No. 3, pp. 135-165.

Smith, S.C.(2006). Speedup of NULL convention digital circuits using NULL cycle reduction, *Journal of system architecture*, pp. 411-422.

Seitz, C.L. (1980). System Timing, *Introduction to VLSI System*, pp. 218-162.

Singh, A. And Marek-Sadowska, M. (2002). FPGA interconnect planning, Proceedings of the 2002 international workshop on System-level interconnect prediction. pp 23-30.

Snider, G.S; Robinett, W.(2005). Crossbar demultiplexers for nanoelectronics based on n-hot codes, *Nanotechnology, IEEE Transactions on,* pp. 249-254.

Snider, G.S.; Kuekes, P.; Hogg, T.; Willams, R.S. (2005). Nanoelectronic Architectures, *Applied Physics,* pp 1183-1195.

Unger, S.H. (1969). Asynchronous Sequential Switching Circuits, *Wiley*, New York.

Whang, D.; Jin, S. & Lieber, C.M. (2004). Large-Scale Hierarchical Organization of anowires for Functional Nanosystems, *Japanese Journal of Applied Physics*, Vol. 43, No. 7B.

Xia, Y.; Rogers, J.; Paul, K. & Whitesides, G. (1999). Unconventional methods for fabricating and patterning nanostructures, *Chemistry Review,* vol. 99, pp. 1823-1848.

Yu, Y.S.; Jung, Y.I.; Park, J.H.; Hwang, S.W. & Ahn, D. (1999). Simulation of single-electron/CMOS hybrid circuits using SPICE macromodeling, *J. Korean Phys. Soc.*, vol. 20, no. 35, pp. S991-S994.

Ziegler, M.; Rose, G. & Stan, M. (2002). A universal device model for nanoelectronic circuit simulation, *IEEE Conf. Nanotechnology (IEEE-NANO),* pp. 83-88.

Ziegler, M. & Stan, M. (2002). Design and analysis of crossbar circuits for molecular nanoelectronics, *IEEE Conf. Nanotechnology (IEEE-NANO),* pp. 323 – 327.

12

New Class of Coolants: Nanofluids

Sadollah Ebrahimi[1], Anwar Gavili[1], Maryamalsadat Lajevardi[1],
Taghi Dallali Isfahani[1, 2], Iraj Hadi[1] and Jamshid Sabbaghzadeh[1]
[1] *National Laboratory Center of Laser Technology, Iran*
[2]*Iran University of Science and Technology, School of Metallurgical and Materials Science, Narmak, Tehran, Iran*

1. Introduction

Today more than ever, ultrahigh-performance cooling plays an important role in the development of energy-efficient heat transfer fluids which are required in many industries and commercial applications. However, conventional coolants are inherently poor heat transfer fluids. Nanofluid a term coined by Choi in 1995 is a new class of heat transfer fluids which is developed by suspending nanoparticles such as small amounts of metal, nonmetal or nanotubes in the fluids. The goal of nanofluids is to achieve the highest possible thermal properties at the smallest possible concentrations (preferably<1% by volume) by uniform dispersion and stable suspension of nanoparticles (preferably<10 nm) in host fluids. We have divided this chapter to four sections. Section 1 has focused on the two methods of synthesizing nanofluids and different methods for dispersing spherical and cylindrical nanoparticles such as Ag, Cu in a host fluid and also the common methods for measuring the thermal conductivity of nanofluids. Section2 has discussed on the thermal conductivity of nanofluids respect to pure fluids to explain the effective thermal conductivity of nanofluids. In this section various theoretical models on the effective thermal conductivity of nanofluids for spherical and cylindrical nanoparticles have been investigated. Section 3 represents the limited understanding of convective heat transfer in nanofluids containing carbon nanotubes that has been developed in recent years. For example the natural and forced convection of nanofluids in a heated cavity are investigated. In recent years there are theoretical and numerical researches for studying the heat transfer effect of nanofluids in microchannels because of their high thermal conductivity. In the last section we discussed about the theoretical and experimental researches for optimizing the microchannel heatsink performance.

2. Synthesis of Nanofluids

In this section we are concerned with the different methods of synthesizing nanofluids. Also the various methods of determining the thermal conductivity of nanofluids will be introduced.
The thermal conductivity of heating or cooling fluids is a very important property in the development of energy-efficient heat transfer systems. At the same time, in all processes

involving heat transfer, the thermal conductivity of the fluids is one of the basic properties taken account in designing and controlling the process. Conventional heat transfer fluids have inherently poor heat transfer properties compared to most solids which is due to the higher thermal conductivities of solids (in orders of magnitude) compared to traditional heat transfer fluids. To overcome the rising demands of modern technology and also to reduce the limitations there is a need to develop new types of fluids that will be more effective in terms of heat exchange performance (Assael etal.,2006). Long ago the basic concept of dispersing solid particles in fluids (mm or μm sized) to enhance the thermal conductivity was introduced by James Clerk Maxwell. Although these fluids showed improved thermal conductivities but due to problems such as: sedimentation, erosion, fouling, increased pressure drop of the flow channel and also clogging of microchannels and flow passages they have never been of interest for practical applications (Trisaksri & Wongwises,2007; Wang & Wei, 2009). However for this reason finding other ways to improve heat transfer properties of conventional fluids became a serious challenge for many researchers. By the introduction of nanotechnology which involved the synthesis of nanoparticles new ideas such as dispersing nanoparticles instead of mm or μm particles in conventional fluids arised. Nanoparticles (1-100nm) have unique properties such as: size dependent physical properties (color, conductivity), large surface area (the specific surface area of nanoparticles is 3 orders of magnitude greater than that of microparticles), large number density (for a given mass of material there are a greater number of particles as the size decreases) and surface structure (nanoparticles have ~20% of their atoms near the surface, allowing them to absorb and transfer heat more efficiently) and finally due to their small size they improve the stability of the suspensions. All of these properties made them very attractive and many researchers started investigating on them. In 1995 Choi, for the first time introduced the term "Nanofluids".

Nanofluids are a new class of heat transfer fluids. These fluids are developed by suspending small amounts of metal and/or non-metal nanoparticles or also nanotubes in conventional fluids (Trisaksri & Wongwises 2007). It should be noted that nanofluids consist of two constituent phases which are conventional fluids and nanoparticles although they cannot be separable. Until now most of the research has been focused on the production of nanofluids with particles such as aluminum oxide, copper, copper oxide, gold, silver, silica nanoparticles and carbon nanotubes in base fluids such as water, oil, acetone, decene and ethylene glycol (Trisaksri & Wongwises 2007). There are two different approaches for the synthesis of nanoparticles. The bottom-up approach which is the formation of nanoparticles from constituent atoms and the synthesis of nanostructures from bulk referred to the top-down approach. Also nanoparticles can be classified by the technique that has been used for their synthesis which are:

1- Physical methods
   1-1 Mechanical grinding
   1-2 Inert gas condensation
2- Chemical methods
   2-1 Chemical precipitation
   2-2 Chemical vapor deposition
   2-3 Micro-emulsions
   2-4 Spray pyrolysis
   2-5 Thermal spraying

Also the most common techniques for the production of carbon nanotubes are:

1- Arc discharge
2- Laser ablation
3- Chemical vapor deposition (CVD)

2.1. The goal and application of nanofluids

As shown in figure1 for the preparation of nanofluids the nanoparticles can be dispersed in different fluids.

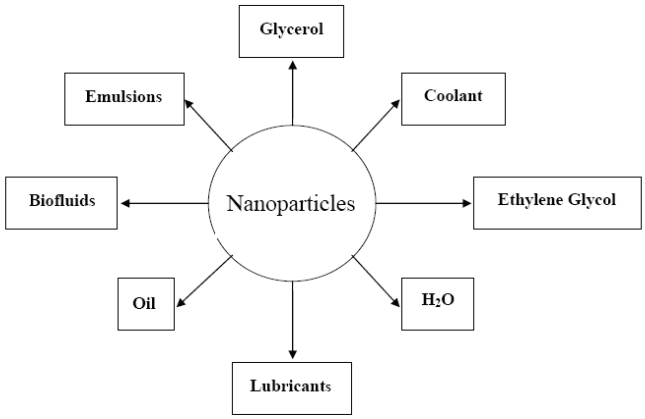


Fig. 1. A number of liquids (heat transfer fluids) that can host nanoparticles for the production of nanofluids.

The most important goal in nanofluid research is to create and develop a nanofluid with stability and ultra-high thermal conductivity for industrial applications. The use of these nanofluids can have a lot of benefits which are: the improvement of heat transfer, reduction in pumping power and lower operating costs, miniaturizing of smaller and lighter heat exchangers, reduction of emissions, suitable for small flow passages like microchannels and reduction in heat transfer fluid inventory.

There are many engineering and medical applications for nanofluids. They can be used for a wide variety of industries, ranging from transportation to energy production and supply to electronics, they can be used to cool car engines, welding equipment and high heat flux devices such as high-power microwave tubes and high-power laser diode arrays they can also flow through tiny passages in MEMS. Also magnetic nanoparticles in biofluids can be used as delivery vehicles for drugs or radiation, providing new cancer treatment techniques because magnetic nanoparticles absorb much more power than microparticles and are more adhesive to tumor cells than normal cells.

2.2. Methods for producing nanofluids

The delicate preparation of a nanofluid is important because nanofluids need special requirements such as an even suspension, durable suspension, stable suspension, low agglomeration of particles and no chemical change of the fluid. There are two fundamental methods to obtain nanofluids (Mamut 2004).

2.2.1. Two step process:

This technique is also known as Kool-Aid method which is usually used for oxide nanoparticles. In this technique nanoparticles are obtained by different methods (in form of powders) and then are dispersed into the base fluid. The main problem in this technique is the nanoparticle agglomeration due to attractive Van der Waals forces.

2.2.2. One step process:

In this process the dispersion of nanoparticles is obtained by direct evaporation of the nanoparticle metal and condensation of the nanoparticles in the base liquid and is the best technique for metallic nanofluids such as Cu nanofluids. The main problems in this technique are low production capacity, low concentration of nanoparticles and high costs. While the advantage of this technique is that nanoparticle agglomeration is minimized.

The suspensions obtained by either case should be well mixed, uniformly dispersed and stable in time. Also it should be noted that the heat transfer properties of nanofluids could be controlled by the concentration of the nanoparticle and also by the shape of nanoparticles.

2.3. Methods for dispersing particles

Due to the high surface energy of nanoparticles they tend to agglomerate to decrease their surface energy. The agglomeration of nanoparticles causes rapid settling which deteriorates the properties of nanofluids. To keep the nanoparticles from agglomeration they are coated with a surfactant (steric dispersion) or charged to repulse each other in a liquid (electrostatic dispersion).

Although the addition of the dispersant could influence the thermal conductivity of the base fluid itself, and thus, the real enhancement by using nanoparticles could be overshadowed.

There are other dispersion methods such as using a high-speed disperser or an ultrasonic probe/bath and also changing the pH value of the suspension (Chopkar etal., 2006). The selection of suitable dispersants depends mainly upon the properties of the solutions and particles and the use of these techniques depends on the required application of the nanofluid.

However metallic nanofluids due to their low thermal conductivity have limited interest but metallic nanofluids especially Cu nanofluids and Ag nanofluids due to there high thermal conductivity are the common nanofluids. More specifically we can say that all the metallic nanofluids compared to oxide nanofluids show much more enhancements so that metallic nanofluids and their volume percent is reduced by one order of magnitude at comparable K enhancements. There are a number of factors other than the thermal conductivity of the dispersed phase which should be considered such as the

average size of the nanoparticles, the method employed for the preparation of the nanofluids, the temperature of measurements and the concentration of the dispersed solid phase.
The thermal conductivity for some engineering liquids and bulk materials are presented in table1:

Material	Thermal Conductivity $\left[\frac{W}{mK}\right]$	Density $\left[\frac{kg}{m^3}\right]$
Water	0.613	997
Ethylene Glycol	0.252	1114
Engine Oil	0.145	884
Aluminum Oxide	40.0	3970
Silicon	148.0	2330
Aluminum	237.0	2702
Copper	401.0	8933
Gold	317.0	19300
Silver	429.0	10500

Table 1. Thermal conductivities of several liquids and solids (Mamut 2009)

The dispersion of copper (Cu) nanoparticles and alternatively carbon nanotubes (CNTs) has provided the most promising results so far, with reported thermal conductivity enhancements of up to 40% and 160% respectively in relation to the base fluid. Because of the high enhancement of the thermal conductivity of CNT nanofluids we will describe them more.

2.4. Carbon Nanotube Suspensions

Carbon nanotubes (CNT) are fascinating materials with microscale length and nanoscale diameter. They have a very high thermal conductivity (6600W/mK for single-walled CNTs) and a low density. Untreated CNTs are not miscible with water and for this reason to form CNT-nanofluids ultrasonic treatment has to be applied whether with the addition of surfactants (Anionic, Cationic or Non-ionic) or not. Usually CNTs are functionalized to easily disperse in the base fluid.
Until know the greatest increase in the enhancement of the thermal conductivity of nanofluids is reported by Choi et al. They dispersed 1 vol. % multiwalled carbon nanotubes (MW-CNTs) with mean diameters of about 25nm and lengths of 50μm (resulting in a length-to-diameter ratio of about 2000) in a synthetic poly (α-olefin) oil as the base fluid and observed an 160% increase (Keblinski etal.,2005).

2.5. Possible heat transfer mechanisms

Existence of agglomerates and close packing of the dispersed phase as well as the presence and type of the dispersants can negatively influence the increase of the thermal conductivity. On the other hand a lot of parameters are responsible for the unique thermal behavior of nanofluids which are: the particle size and shape, the length-to-diameter ratio, homogenization time, volume fraction of the dispersed phase, interfacial resistance, the ordered structure of the liquid at the solid-liquid interfaces and the Brownian motion of the

nanoparticles enabling the formation of loosely packed clusters and convection-like effects at the nanoscale (Assael etal.,2006).

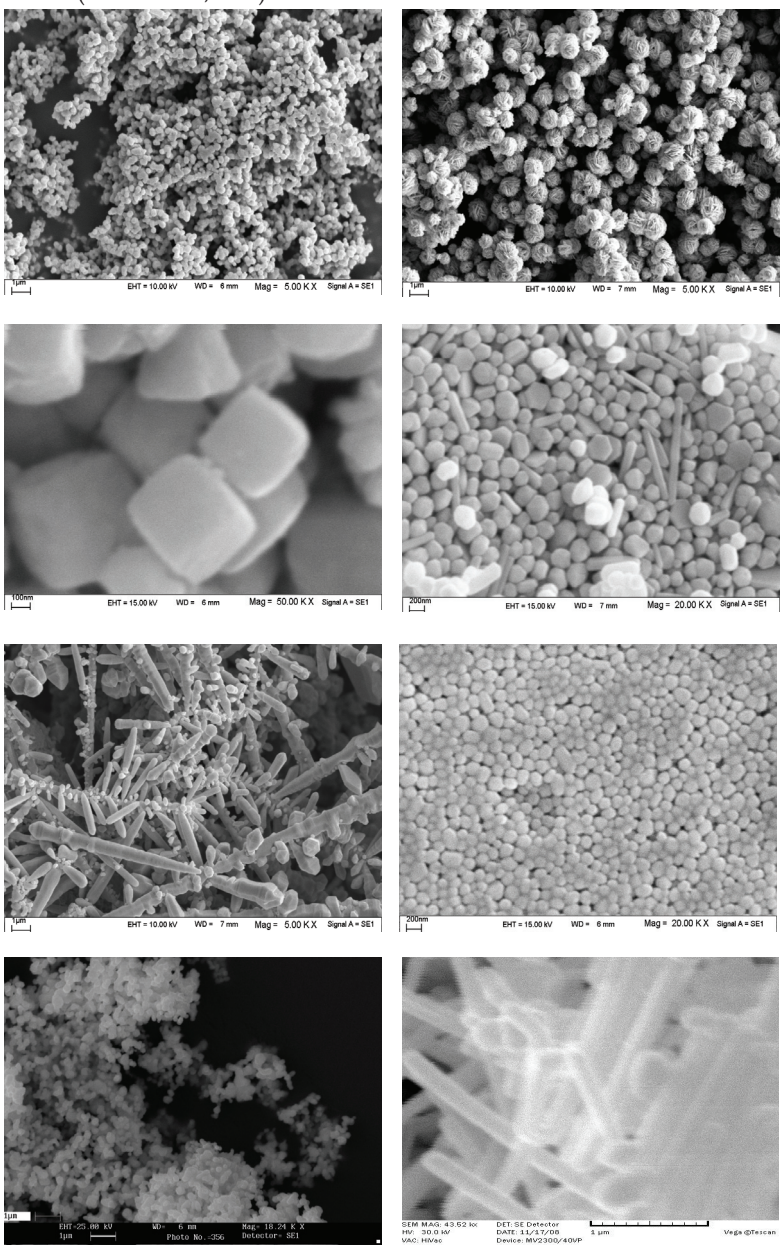


Fig. 2. SEM images of Silver micron particles, nanoparticles and nanowires with different morphologies (Dalalli Isfahani etal.,2009).

2.6. Measurement of the thermal conductivity of nanofluids

Many researchers have reported experimental studies on the thermal conductivity of nanofluids. The transient hot wire method (THW), temperature oscillation, transient planar source (TPS) and the steady –state parallel plate method have been employed to measure the thermal conductivity of nanofluids. However, THW method is one of the most accurate for determining thermal conductivities of materials and for this reason it has been extensively used. The THW technique works by measuring the temperature/time response of the wire to an abrupt electrical pulse. The wire is used as bath heater and thermometer. A derivation of Fourier's law and temperature data were used to calculate the thermal conductivity. The results from all of the available experimental studies indicated that nanofluids containing a small amount of nanoparticles have substantially higher thermal conductivity than those of base fluids. The major advantage of this method lies in its almost complete elimination of the effects of natural convection whose unwanted presence creates problems for measurements made with a steady-state apparatus. In addition the method is very fast relative to steady-state techniques.

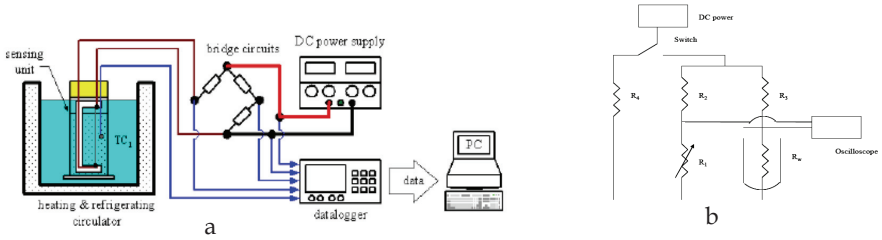


Fig. 3. (a): Experimental apparatus, (b) : schematic diagram of hot-wired apparatus(FOO , 2005).

2.7. The affecting parameters on the thermal conductivity

Overally we can say that the smaller the size the greater the stability of colloidal dispersion, the greater the stability of colloidal dispersion the greater the probability of interaction and collision among particles and collision among particles and fluid and the greater the effective heat energy transport inside the liquid (Xue 2003).

Thermal conductivity enhancement ratio ($K_{effective}= K_{nanofluid}/K_{basefluid}$) and the parameters that most affect the thermal conductivity of nanofluids are:

1- Particle volume fraction
2- Particle material
3- Base fluids
4- Particle size
5- Temperature

2.8 Important points about the heat transfer of nanofluids

Finally it should be noted that the most important reason of using nanofluids is to improve the heat transfer of fluids. Our experience in this field has showed that other properties besides the thermal conductivity such as the Cp, α , heat fusion and etc. of the nanoparticle have great impact on the heat transfer of nanofluids. This point has not been mentioned else

ware. Most of the researchers have mainly focused on the thermal conductivity of nanofluids without considering these properties. This is the main key in making useful nanofluids for practical applications. Some other key points for nanofluids are:
1- The thermal conductivity enhancement ratio increases with increasing particle volume fraction.
2- The sensitivity to volume fraction depends on particle material and base fluid (the sensitivity is higher for particle material with higher thermal conductivity and base fluid with lower thermal conductivity).
3- The thermal conductivity of nanofluids shows higher sensitivity to temperature then that of the base fluid, consequently the thermal conductivity enhancement ratio shows also high sensitivity to temperature.
4- The particle shape also affects the thermal conductivity. Elongated particles show higher thermal conductivity enhancement ratio than spherical particles.
5- The additives used to prevent particle agglomeration seem to increase the thermal conductivity enhancement ratio.
6- The thermal conductivity enhancement ratio increases with acidity.
7- For the suspensions containing the same base liquid and nanoparticles, the thermal conductivity enhancements were highly dependent on the specific surface area (SSA) of the nanoparticles.
8- For the suspensions using the same nanoparticles, the enhanced thermal conductivity ratio decreased with increasing thermal conductivity of the base fluid.

3. Dynamic models of thermal conductivity in nanofluids

In recent years nanofluids has proved the enhancement in thermal conductivity of pure fluids. At the 1995 annual winter meeting of the American Society of Mechanical Engineers, Choi (1995) presented the remarkable possibility of doubling the convection heat transfer coefficients using ultrahigh-conductivity nanofluids instead of increasing pumping power by a factor of 10. The high thermal conductivity enhancement of nanofluids cannot be predicted by any of the classical models. Typically, the classical theories (Maxwell 1881) consider the effect of particle concentration, particle and fluid conductivity and the particle shape effect (Hamilton and Crosser 1962). Traditional solid/liquid conductivity theories for large particles have also been studied. The traditional models are in lack of any predicted dependency of conductivity on spherical particle size. However, it is still difficult to find a suitable high thermal conductivity fluid for microchannel heat sinks. In fact, thermal conductivity of fluids plays an important role in the development of energy-efficient heat transfer fluids (Ryu et al., 2002; Ryu, et al., 2003; Jang et al., 2006; Chein et al., 2005; Min et al., 2004; Xie et al., 2002; Keblinski et al., 2002; Choi et al., 2001; Eastman, 2001; Xie et al., 2003). The resulting heat transfer nanofluids posses significant high thermal conductivity compared to unfilled liquids. It was demonstrated that thermal conductivities predicted by traditional theoretical models such as Hamilton and Crosser's (1962) are much lower than the measured data for metallic and oxide nanofluids. To develop such theories, Wang et al. (1999) were the first to propose new static and dynamic mechanisms, they suggested for the first time that nanoparticle size is important in enhancing the thermal conductivity of nanofluids. They also attributed enhanced conductivity to the chain structure of nanoparticle clusters. Xuan and Li (2000) suggested several possible mechanisms for

enhanced thermal conductivity of nanofluids, such as the increased surface area of nanoparticles, particle–particle collisions, and the dispersion of nanoparticles. Keblinski et al. (2002) proposed four possible microscopic mechanisms for the anomalous increase in the thermal conductivity of nanofluids: Brownian motion of the nanoparticles, molecular-level layering of the liquid at the liquid–particle interface, the ballistic rather than diffusive nature of heat conduction in the nanoparticles, and the effects of nanoparticle clustering. Xuan et al. (2003) developed a dynamic model into which the effects of Brownian motion of nanoparticles and the aggregation structure of nanoparticle clusters (i.e., fractals) are taken. Yejine (2005) attempted to investigate theoretically the mechanism of the effective thermal conductivity of nanofluids. They investigated a theoretical model which includes particle-fluid mixture; the liquid molecules close to a particle surface which are known to form layered structures and behave much like a solid (nanolayer), with the thickness of nanometer. Also the nanoparticle motion causes micro-convection of the suspending fluid, and thus enhances the overall heat transfer in nanofluids. Jang and Choi (2004) have constructed a theoretical model based on kinetics, Kapitza resistance, and convection. They have derived a general expression for thermal conductivity of nanofluids involving three modes. Patel et al., (2008) investigated the mechanisms of conduction in liquids and conduction through solid nanoparticles and the micro-convective heat transfer to the nanoparticles due to their Brownian motion in the liquid, which coupled in a more logical manner. These investigations show that nanofluids have higher heat transfer relative to conventional fluids, and also a better stability compared to fluids with suspended microparticles, making nanofluids useful. Several factors such as gravity, Brownian motion, layering at the solid/liquid interface, ballistic phonon transport between the particles and nanoparticle clusters, and the friction between the fluid and the solid particles contributes to the increase in nanofluid heat transfer.

In addition, the effects of nanolayer thickness, dispersion and volume fraction of nanoparticles enhance the heat transfer. Also, heat transfer nanofluids with carbon nanotubes are expected to possess even better heat transfer properties due to the non spherical shape of carbon nanotubes. Carbon nanotubes (CNTs) have extremely high thermal conductivity, which, at room temperature is over 3,000 times greater than that of water and over 10,000 times greater than that of engine oil (Hone et al., 1999; Berber et al., 2000; Kim et al., 2001). Therefore, fluids containing suspended carbon nanotubes are expected to exhibit significantly higher thermal conductivity relative to conventional spherical nanofluids. Sabbaghzadeh and Ebrahimi (2007) improved the semi theoretical model for effective thermal conductivity of nanofluids containing spherical nanoparticles which presented by Jang and Choi, (2004) for nanofluids containing cylindrical nanoparticles (CNTs). However four heat transfer modes exist in nanofluids. As it has been shown in figure 4 , four heat transfer modes exist in nanofluids containing cylindrical nano-particles (Fig. 4), which are by base fluid (mode 1), nanoparticle (mode 2), nanolayer (mode 3) and micro-convection (mode 4). In addition a nanofluid consisting of base fluid and complex nanoparticles is statistically homogeneous and isotropic. As shown in figure 4, the first mode is defined by collisions among the base fluid molecules, which physically represents the thermal conductivity of the base fluid.

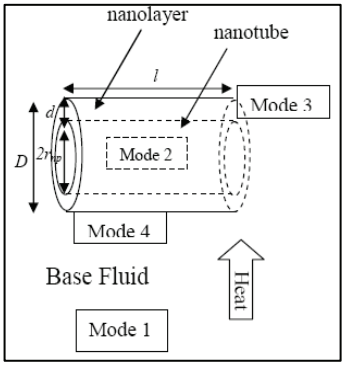


Fig. 4. Shows a cylindrical nanoparticle in base fluid and modes of energy transport in nano fluids. The mode of thermal interactions of dynamic nanoparticles with base fluid molecules (fifth mode) is not shown.

If the energy carrier travels freely over the mean-free path, l_f, (which is the collision between two fluid molecules) then the amount of transferred energy can be presented by the following equation, while, in contrary with the previous works, effect of the nanolayer thickness has not been neglected (Eastman et al., 2001).

$$J_U = -\frac{1}{3} l_f c_v \bar{u} (1 - f_{np} - f_{nl}) \frac{dT}{dz} = -K_f \frac{dT}{dz} (1 - f_{np} - f_{nl}) \tag{1}$$

where c_v, $\bar{u}$, K_f and T are the heat capacity per unit volume, mean speed, thermal conductivity of the base fluid and temperature of the base fluid molecules respectively, and f_{np} is the volume fraction of the original cylindrical nanoparticles which is given by

$$f_{np} = n\pi r_{np}^2 l_{np} \tag{2}$$

where n is the number of particles per unit volume, r_{np} is the nanoparticle radius, l_{np} is the cylinder length, and f_{nl} is the volume fraction of the nanolayer which is given by

$$f_{nl} = n\pi l_{np} \left((r_{np} + d_{nl})^2 - r_{np}^2 \right) \Rightarrow f_{nl} = f_{np} M' \tag{3}$$

Where

$$M' = ((d_{nl} / r_{np} + 1)^2 - 1) \tag{4}$$

and d_{nl} is the thickness of nanolayer. The second mode is the thermal diffusion in the nanoparticles which are covered by a nanolayer in the fluid

$$J_U = -\frac{1}{3} l_{np} c_v \bar{u} f_{np} \frac{dT}{dz} = -\kappa K_{np} \frac{dT}{dz} f_{np} \tag{5}$$

where K_{np} is the thermal conductivity of nanoparticle and κ is a constant related to the Kapitza resistance in the order of 0.01(Fand, 1965). The third mode is the nanolayer thermal diffusion in the fluid

$$J_U = -\frac{1}{3} l_f c_v \bar{u} f_{nl} \frac{dT}{dz} = -K_{nl} \frac{dT}{dz} f_{nl} \tag{6}$$

where K_{nl} is thermal conductivity of the nanolayer. The thickness, the microstructure and the physicochemical properties of these nanolayers are highly dependent on the suspended nanoparticles, the base fluid, and the interaction between them. There is no available expression for the calculation of nanolayer thermal conductivity, but as these molecules have an intermediate state between the bulk liquid and solid, therefore it is reasonable to consider this conductivity between these two states. If the thermal conductivity distribution inside the nanolayer is $k(r)$ ($r_{np} < r < r_{np} + d_{nl}$, see figure 1), then the thermal resistance of the nanolayer, R_{nl} is given as follows:

$$R_{nl} = \int_{r_{np}}^{r_{np}+d_{nl}} \frac{dr}{2\pi r l_{np} k(r)} \tag{7}$$

R_{nl} can also be expressed by the average thermal conductivity of the nanolayer, K_{nl}, as follows

$$R_{nl} = \frac{\ln(1 + d_{np}/r_{np})}{2\pi K_{nl} l_{np}} \tag{8}$$

From Eqs. (7) and (8), we have

$$K_{nl} = \frac{\ln\left(1 + d_{np}/r_{np}\right)}{\int_{r_{np}}^{r_{np}+d_{nl}} \frac{dr}{r k(r)}} \cdot \tag{9}$$

We clearly know *k(r)* is a nontrivial problem due to the complexity of physicochemical interactions between the nanoparticles and the fluid. For simplicity, we assumed a linear variation of *k(r)*, which is (Xie et al., 2003)

$$k(r) = \frac{K_f - K_{np}}{d_{nl}} r + \frac{K_{np}(r_{np} + d_{nl}) - K_f r_{np}}{d_{nl}} \tag{10}$$

by substituting Eq. (10) into (9), the following expression can be derived

$$K_{nl} = \frac{K_f r_{np}(M\varepsilon_p - 1)\ln(M)}{d_{np}(\ln(M\varepsilon_p))} \tag{11}$$

Where $M = 1 + d_{nl}/r_{np}$, and $\varepsilon_p = K_{np}/K_f$ is the ratio of the thermal conductivity of the nanoparticle to that of the base fluid. Finally the fourth mode is thermal interaction of dynamic complex nanoparticles (original nanoparticles and nanolayers) with base fluid molecules. So, the last mode can be defined by

$$J_U = h(T_{np} - T_f)(f_{nl} + f_{np}) = h\delta_T(f_{nl} + f_{np}) \times \frac{(T_{np} - T_f)}{\delta_T} \approx -h\delta_T(f_{nl} + f_{np})\frac{dT}{dz} \tag{12}$$

Where $\delta_T \sim 3d_f/\mathrm{Pr}$ is the thickness of the thermal boundary layer, in which d_f is the diameter of the base fluid molecules, and Pr is the Prandtl number defined by

$$\mathrm{Pr}_f = \nu_f / \alpha_f \tag{13}$$

where α_f is the thermal diffusivity of the base fluid and ν_f is the kinematic viscosity of the base fluid. Fand (1965) has shown that if the flow direction is perpendicular to the cylinder axes (figure 4) then the heat transfer coefficient from liquid to cylinder, is given by:

$$h = \frac{K_f}{D}\left(0.35 + 0.56\,\mathrm{Re}_f^{0.52}\right)\mathrm{Pr}_f^{0.3} \tag{14}$$

The relation is valid for $10^{-1} < \mathrm{Re}_f < 10^5$, and D is the diameter of the complex nanoparticle as shown in figure 1. The Reynolds number Re_f is defined as follows:

$$\mathrm{Re}_f = \frac{\bar{u} l_{np}}{\nu_f} \tag{15}$$

Where $\bar{u}$ is the mean velocity of the complex nanoparticles.
From the molecular heat theory, $\bar{u}$ is determined by Kim et al. (2001)

$$\bar{u} = \left(\frac{3k_B T}{m_{np}}\right)^{1/2} \tag{16}$$

where $k_B = 1.38 \times 10^{-23} JK^{-1}$ is the Boltzman constant, T is the particle temperature and m_{np} is the mass of the complex nanoparticle, and can be calculated as follows:

$$m_{np} = \pi r_{np}^2 l_{np} \rho_{np} \left[(\rho_{nl} / \rho_{np}) M^{'} + 1 \right] \tag{17}$$

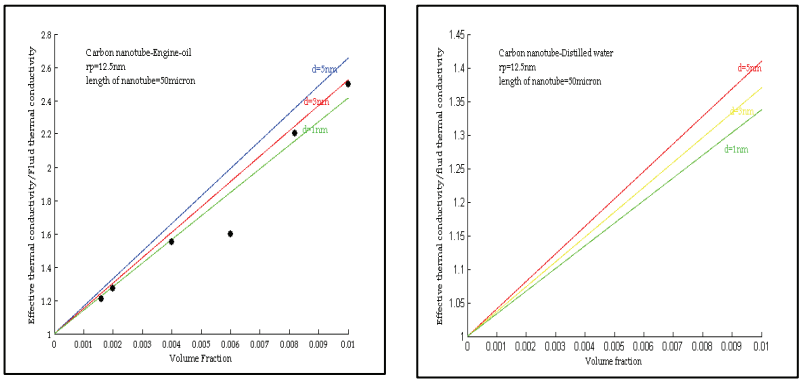
The last mode is the collision between nanoparticles due to Brownian motion. Spherical nanoparticle collision due to Brownian motion is a very slow process (Eastman et al., 2001; Ryu et al., 2002), also Brownian motion for cylindrical particles such as CNTs can be neglected. With neglecting the last mode, the following expression for the thermal conductivity of nanofluids containing cylindrical nanoparticles is derived:

$$K_{eff} = K_f \left(1 - f_{np}(1+M')\right) + f_{np}\left(\kappa K_{np} + K_{nl} M'\right) + f_{np}(1+M') \frac{df}{D} \left(0.35 + 0.56 \mathrm{Re}_f^{0.52}\right) \mathrm{Pr}_f^{0.3} K_f \tag{18}$$

The thermal conductivity of suspended nanoparticles, k_p, involving in the kapitza resistance is two order of magnitude less than the thermal conductivity of nanoparticles., Jang and Choi (2004) have constructed a theoretical model for spherical nanoparticles as follows:

$$k_{eff} = k_{BF}(1-f) + \beta k_p f + C_1 \frac{d_f}{D} k_{BF} \, \mathrm{Re}_{BF}^2 \, \mathrm{Pr}_{BF} \tag{19}$$

where $C_1 = 6 \times 10^6$, β is constant related to Kapitza resistance, and f is volume fraction of nanoparticles. In the case of cylindrical shape carbon nano tubes the results are shown in figure 5. The normalized thermal conductivity k_{eff} / k_{BF} for the nanotube in engine oil suspensions as a function of the volume fraction of naotubes is shown in figure (5-a) and in figure 5-b for distilled water. As it is clear from figure 5-a, by increasing the nanoparticle volume fraction, the thermal conductivity increases, which has a good agreement with experimental results (shown as solid circles), also by increasing the nanolayer thickness thermal conductivity will be increased. Figure 5 also shows that the effect of nanolayer thickness will become more significant by increasing the volume fraction of nanoparticles. Understanding the fundamentals of energy transport in nanofluids is important for developing extremely energy-efficient nanofluids for a range of heat transfer applications. In this section we developed a theoretical model for explaining the enhancement in the effective thermal conductivity of nanotubes (with cylindrical shape). The results show that the thermal conductivity increases if the thickness of nanolayers increases.



5-b

Fig. 5. Shows normalized thermal conductivity for the nanotube-in-engine oil suspensions at experimental (Fig. 5-a, solid circles) vs theory case (Fig. 5-b, solid lines) and nanotube in-distilled water (Fig. (5-b): as a function of the volume fraction of nanotubes with various nanolayer thicknesses.

4. Convection in nanofluids

Heat transfer contains several modes which one of them is convection that is the heat transfer from a fluid to the wall or at opposite direction. Compared to the experimental and theoretical study of the thermal conductivity of nanofluids the study on the heat transfer convection is very limited. It is very important to know that enhancement in thermal conductivity of nanofluid necessarily does not increase the heat transfer capability of it. The physical properties of nanofluids compared to pure fluid such as viscosity, heat capacity, density and stability of nanoparticles in the fluid maybe deteriorated. Convection in nanofluid is broadly divided in two types: natural convection and force convection. Natural convection is the type of convection which the flow is generated by buoyancy force during cooling or heating of the fluid. Force convection is the convection which the flow is due to external forces such as pump, fan, compressor and etc. (Raj Kemal et al, 2007). The law of convection is Fourier's conduction law (Incropera 2002):

$$Q = hA(T_W - T_f) \tag{20}$$

Where Q is the amount of heat transfer between the wall and the fluid in motion, A is the solid-liquid interface area, T_W is the wall temperature, and T_f is the fluid temperature. In nanofluids two methods can be used for the study of heat transfer (Khanafer et al., 2003): which is assumed that both the fluid phase and nanoparticles are in thermal equilibrium state and they flow at the same velocity which is named as single phase model. In the other model the fluid is considered to be a single fluid with two phases, and the coupling between them is strong which each phase has its own velocity vectors, and within a given volume fraction there is a certain volume fraction of each phase which is named as mixture model. In nanofluids due to the small particles are suspended in a ordinary fluid and their higher stability the single phase model is better applicable.

4.1. Governing Equations of Fluid Flow and Heat Transfer

Consider a two-dimensional enclosure of height H and width W filled with a fluid as shown in Fig. 6.The horizontal walls are assumed to be insulated, nonconductive and impermeable to mass transfer. The fluid in the enclosure is Newtonian, incompressible, and laminar. Moreover, it is assumed that both the fluid phase and nanoparticles are in thermal equilibrium state and they flow at the same velocity. For natural convection the right vertical wall is maintained at a high temperature T_H while the left vertical wall is kept at a low temperature T_L (Fig.6 a). In the case of no velocity on the cavity walls natural convection is dominated and in the case of velocity (V_p) on at least one of the walls the force convection is added to natural convection and convection mode is named as mix convection (Fig.6 b). In the first stage we considered a natural convection and then force convection will be applied. It is assumed that the dispersed nanoparticle of the nanofluid has a uniform cylindrical shape and size.

The variation of density in the buoyancy force is based on the Boussinesq *approximation.The initial and boundary conditions for natural convection in the present* investigation are presented as:

$$in \quad t=0 \quad u=v=T=0 \quad for \begin{cases} 0 \le x \le W \\ 0 \le y \le H \end{cases} \tag{21}$$

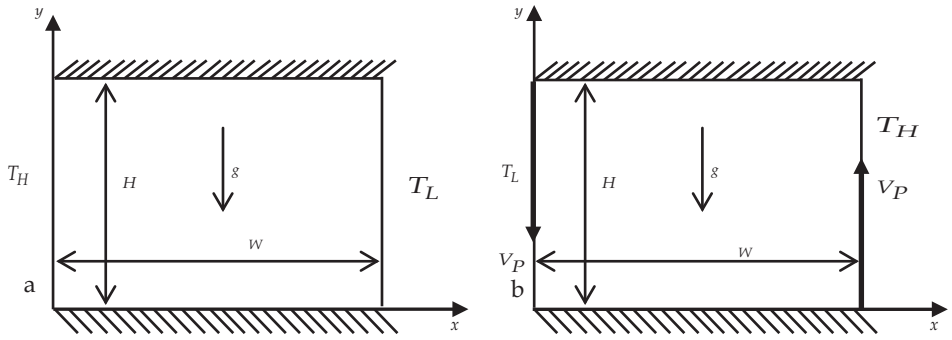


Fig. 6. Schematic of the physical model: a (natural convection) and b (force convection).

$$in \quad t\rangle 0: \begin{cases} u=v=\dfrac{\partial T}{\partial y}=0 \quad for \begin{cases} 0 \le x \le W \\ y=0,H \end{cases} \\ T=T_L, u=v=0 \quad for \begin{cases} 0 \le y \le H \\ x=0 \end{cases} \\ T=T_H, u=v=0 \quad for \begin{cases} 0 \le y \le H \\ x=W \end{cases} \end{cases} \tag{22}$$

The governing equations for the present study considering the above mentioned assumptions are written in dimensional form and the well-known stream function vorticity formulation in which the primary variables are replaced by stream function ψ and vorticity ω are used to mainly to avoid the pressure term in the momentum equation. This requires complicated algorithms in incompressible flow as follows:

Vorticity equation

$$\frac{\partial \omega}{\partial t} + u\frac{\partial \omega}{\partial x} + v\frac{\partial \omega}{\partial y} = \frac{\mu_{eff}}{\rho_{nf}}\nabla^2\omega + \frac{\beta_{nf}}{\rho_{nf}}g\frac{\partial T}{\partial x} \tag{23}$$

Energy equation

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = \frac{\partial}{\partial x}\left(\frac{K_{eff}}{(\rho c_p)_{nf}}\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{K_{eff}}{(\rho c_p)_{nf}}\frac{\partial T}{\partial y}\right) \tag{24}$$

Kinematics equation

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = -\omega \tag{25}$$

The effective density of a nanofluid at a reference temperature is given by:

$$\rho_{nf} = f_{np}\rho_{np} + (1 - f_{np})\rho_f \tag{26}$$

where ρ_f , ρ_{np} and f_{np} are the density of base fluid, density of nanoparticles, and the volume fraction of the nanoparticles, respectively. The effective viscosity of a fluid with a viscosity μ_f containing a dilute suspension of small rigid cylindrical particles is given by Brinkman equation (Brinkman 1952):

$$\mu_{eff} = \mu_f /(1 - f_{np})^{2.5} \tag{27}$$

The heat capacitance and thermal expansion coefficient of the nanofluid can be presented as

$$(\rho c_p)_{nf} = f_{np}(\rho c_p)_{np} + (1 - f_{np})(\rho c_p)_f \tag{28}$$

$$\beta_{np} = (1 - f_{np})\rho_f \beta_f + f_{np}\rho_{np}\beta_{np} \tag{29}$$

and the effective thermal conductivity of the solid-liquid mixture is considered as equation (18)

while the specific velocity is as follows:

$$Vp = \sqrt{g\beta_f \Delta TH} \tag{30}$$

The above equations for natural convection can be cast in non-dimensional form by incorporating the following dimensionless parameters

$$\begin{cases} X = \dfrac{x}{H} & Y = \dfrac{y}{H} & \tau = \dfrac{tV_p}{H} & \Omega = \dfrac{\omega H}{V_p} \\ U = \dfrac{u}{V_p} & V = \dfrac{v}{V_p} & \Psi = \dfrac{\psi}{HV_p} & \theta = \dfrac{T - T_L}{T_H - T_L} \end{cases} \tag{31}$$

$$\frac{\partial \Omega}{\partial \tau} + U\frac{\partial \Omega}{\partial X} + V\frac{\partial \Omega}{\partial Y} = \Gamma_\Omega\left(\frac{\partial^2 \Omega}{\partial X^2} + \frac{\partial^2 \Omega}{\partial Y^2}\right) + \lambda\frac{\partial \theta}{\partial X} \tag{32}$$

$$\frac{\partial \theta}{\partial \tau} + U\frac{\partial \theta}{\partial X} + V\frac{\partial \theta}{\partial Y} = \Gamma_\theta\left(\frac{\partial^2 \theta}{\partial X^2} + \frac{\partial^2 \theta}{\partial Y^2}\right) \tag{33}$$

$$\frac{\partial^2 \Psi}{\partial X^2} + \frac{\partial^2 \Psi}{\partial Y^2} = -\Omega \tag{34}$$

In the above equations Γ_Ω and Γ_θ are as:

$$\Gamma_\Omega = \frac{1}{(1 - f_{np})^{2.5}\ ((1 - f_{np}) + f_{np}\dfrac{\rho_{np}}{\rho_f})\ \sqrt{Gr}} \tag{35}$$

$$\Gamma_\theta = \frac{1}{\Pr\sqrt{Gr}}\ \frac{K_{eff}}{k_f}\ ((1 - f_{np}) + f_{np}\frac{(\rho c_p)_{np}}{(\rho c_p)_f})^{-1} \tag{36}$$

while the Grashof number is $Gr = \dfrac{g\beta_f \Delta TH^3}{\nu_f^2}$, and Prandtl number is $Pr = \dfrac{\nu_f}{\alpha_f}$. The aspect ratio is defined as $A = \dfrac{W}{H}$ and is assumed unity in this investigation. The physical dimension of the enclosure H is chosen to be 1 cm.

In equation (11) the coefficient λ that appears next to the buoyancy term is given by

$$\lambda = [(1 + \frac{(1 - f_{np})}{f_{np}} \frac{\rho_f}{\rho_{np}})^{-1} \frac{\beta_{np}}{\beta_f} + (1 + \frac{f_{np}}{(1 - f_{np})} \frac{\rho_{np}}{\rho_f})^{-1}] \tag{37}$$

The local variation of the Nusselt number of the fluid on the hot wall can be expressed as

$$Nu = -\frac{K_{eff}}{k_f} \frac{\partial \theta}{\partial X} \Big| X=0 \tag{38}$$

while the average Nusselt number is given by:

$$\overline{Nu} = -\int_0^1 (\frac{K_{eff}}{k_f} \frac{\partial \theta}{\partial X} \Big| X=0) dY \tag{39}$$

In mix convection at least one of the cavity walls is moved. In this work we supposed that the left wall is moving downward and the right wall is moving upward, also the left wall is kept at low temperature and the right wall is at high temperature. For this configuration the natural convection and force convection are in the same direction and the dimensionless form of the governing equations is presented versus wall velocity. The governing equations for mix convection can be cast in non-dimensional form as the same with natural convection except that the speed of the wall has a specific value which is named as the specific velocity. Also the dimensionless equations have the same form in the two type convection with this exception that the coefficients Γ_Ω , Γ_θ , and λ are defined as follows:

$$\Gamma_\Omega = \mu_{eff} / (\rho_{nf} \upsilon_f \, \mathrm{Re}) \tag{40}$$

$$\Gamma_\theta = \alpha_{nf} / (\alpha_f \, \mathrm{Pr} \, \mathrm{Re}) \tag{41}$$

$$\lambda = (\beta_{nf} / \beta_f \rho_{nf}) Ri \tag{42}$$

and also $\mathrm{Re} = (HV_p)/\nu_f$ is Reynolds number, and $Ri = Gr/\mathrm{Re}^2$ is Richardson number.

The numerical simulation for natural convection was developed over a wide range of Rayleigh numbers $(10^4, 10^5)$, volume fraction of nanoparticles $(0 < f_{np} < 1)$, and nanoparticle diameter $d_{np}(nm) = 10$ and also the nanolayer thickness $d_{nl}(nm) = 5$.

The average Nusselt number at the left wall for various Rayleigh numbers and different nanoparticle volume fractions $(0 < f_{np} < 1)$, and particle diameter $d_{np}(nm) = 10$ and nanolayer thickness $d_{nl}(nm) = 5$ of nanoparticles are presented in Figure 7. As it can be seen from this figure with the increase of the volume fraction of the nanoparticle, increase in average Nusselt number for high value of the Rayleigh number is more noticeable while for the low value it is weaker. The numerical simulation for mix convection for Richardson numbers $(1.0, 10.0)$, $Gr = 10^4$ and the same other parameters with the natural convection are down and the results are presented in Figure 8. Also enhancement in the average Nusselt number for mix convection in different Richardson number versus nanoparticles volume fraction are shown in Figure 10. This figure shows that with the increase in the volume fraction of nanoparticles, more increase in average Nusselt number for low Richardson number (high Reynolds number) was observed.

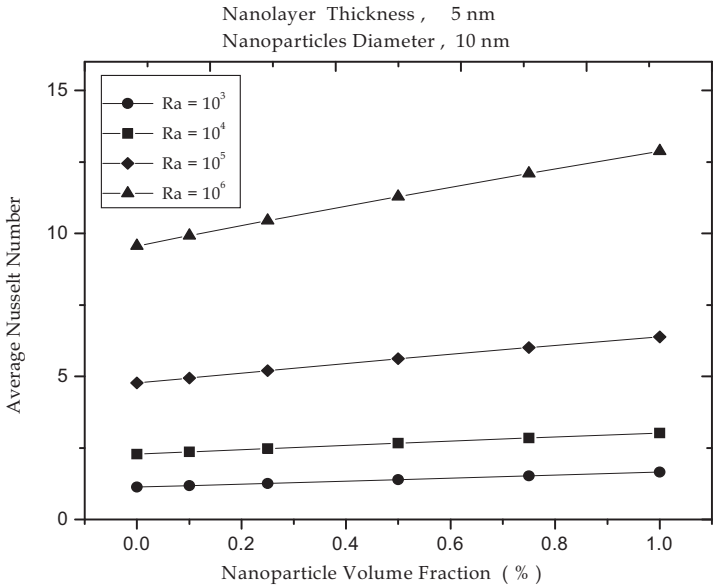


Fig. 7. Variation in average Nusselt number for natural convection in nanofluid versus the volume fraction for different Rayleigh numbers ($d_{np} = 10(nm)$, $d = 5(nm)$)

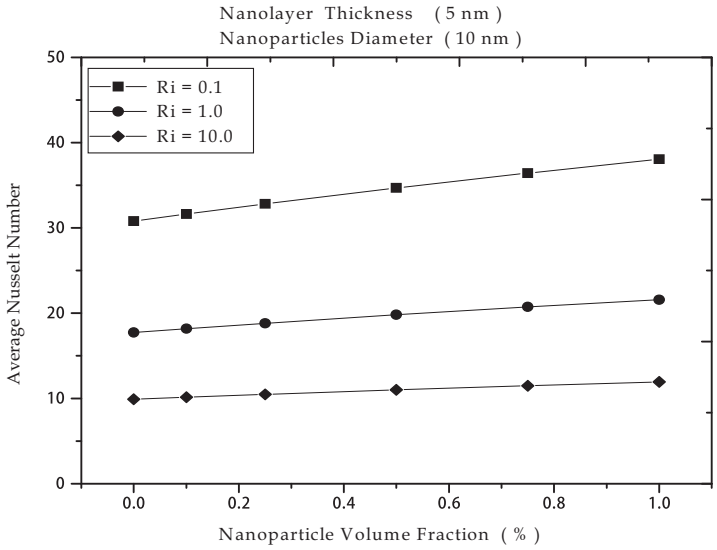


Fig. 8. Variation in average Nusselt number for mix convection in nanofluid versus nanoparticles volume fraction for different Richardson numbers($d_{np}(nm) = 10$, $d_{nl}(nm) = 5$)

5. Performance analysis of nanofluid-cooled microchannel heat sinks

With the advances in computing technology over the past few decades, electronics have become faster, smaller and more powerful. In most cases, the chips are cooled using forced air flow. However, when dealing with a component that contains billions of transistors working at high frequency, the temperature can reach a critical level where standard cooling methods are not sufficient. In addition to high-performance electronic chips, high heat flux removal is also required in devices such as laser diode arrays and high-energy mirrors. In the last two decades, many cooling technologies have been pursued to meet the high heat dissipation rate requirements and maintain a low junction temperature. Among these efforts, the microchannel heat sink (MCHS) has received much attention because of its ability to produce high heat transfer coefficient, small size and volume per heat load, and small coolant requirements. Recent progress in MCHS development was provided by Kandlikar et al. (2003).

A MCHS typically contains a large number of parallel microchannels with a hydraulic diameter ranging from 10 to 1000 μm . A coolant is forced to pass through these channels to carry the heat away from a hot surface. The MCHS cooling concept was proposed by Tuckerman and Pease (1981). Since then, MCHS performances with different substrate materials and channel dimensions have been studied extensively in the past two decades. These studies can be categorized into theoretical (Knight et al., 1992; Ambatipudi and Rahman, 2000), numerical (Fedorov and Viskanta, 2000; Lee et al., 2005; Li et al., 2004; Li and Peterson, 2006), and experimental approaches (Qu and Mudawar, 2002; Tiselj et al., 2004). In the theoretical approach, the main objective is to develop design schemes that can be used to optimize MCHS performance. Most studies in this approach employed the classical fin

theory which models the solid walls separating microchannels as thin fins. The heat transfer process is simplified as one dimensional, constant convection heat transfer coefficient and uniform fluid temperature. However, the nature of the heat transfer process in MCHS is conjugated heat conduction in the solid wall and convection to the cooling fluid. The simplifications used in the theoretical approach usually under- or over predict the MCHS performance.

To overcome the shortcomings associated with MCHS thermal performance analysis using fin theory, several investigators proposed modeling the MCHS as a porous medium. In the study by Kim and Kim (1999), laminar heat transfer in MCHS was analyzed using a modified Darcy model for fluid flow and two-equation model for heat transfer. They found that their results agreed well with those predicted using fin theory models (Knight et al., 1992) and experimental measurements by Tuckerman and Pease (1981). Zhao and Lu (2002) further extended the model developed by Kim and Kim (1999) to study the channel geometries, effective thermal conductivities and porosities on the MCHS thermal performance. Because conjugated heat transfer is involved in MCHS, and it is believed that the porous medium model is better than the classical fin theory for describing the MCHS thermal performance.

Although high thermal performance can be achieved using MCHS but further improvement is still needed to cope with the increasing demands for various device applications. From the heat transfer point of view, improving MCHS thermal performance enhances the heat transfer characteristics inside the MCHS. Extensive reviews on the techniques for heat transfer enhancement in macroscale dimensions were provided by Bergles (2002) and Webb (1993). One of the methods for enhancing heat transfer is the application of additives in the working fluids. The basic idea is to enhance the heat transfer by changing the fluid transport properties and flow features of the metal particles embedded in the liquid fluid. In earlier studies to enhance the heat exchanger performance. However, a serious clogging problem occurred due to the particle sedimentation. Recent interest based on this concept focused on heat transfer enhancement using a nanofluid in which the nanoscale metallic or non-metallic particles were suspended in the base fluids. Several experimental and analytical studies showed that nanofluids have higher thermal conductivity than pure fluids and therefore greater potential for heat transfer enhancement (Wang et al., 1999; Koo and Kleinstreuer, 2004).

Using a nanofluid as the heat transfer working fluid has gained much attention in recent years. Xuan and Roetzel (2000) proposed two theoretical models to predict the heat transfer characteristics of the nanofluid flow in the tube. Li and Xuan (2002), Xuan and Li (2003) and Pak and Cho (1998) experimentally measured the convection heat transfer and pressure drop for nanofluid tube flows. Their results indicated that the heat transfer coefficient was greatly enhanced and depended on the Reynolds number of the flow, particle Peclet number, particle size and shape, and particle volume fraction. These studies also indicated that the presence of nanoparticles did not cause an extra pressure drop in the flow. Recently, Yang et al. (2005) carried out an experimental study attempting to construct a heat transfer correlation among the parameters that affected heat transfer. For a laminar flow regime in a circular tube, they indicated that the heat transfer coefficient for the nanofluid flow had a lower increase than predicted by the conventional heat transfer correlation for the homogeneous or particle-suspended fluid. Ding et al. (2006) reported the heat transfer coefficient data for force convection in circular tubes using a carbon nanotube (CNT)

nanofluid. In most of the studies mentioned above, the nanofluid heat transfer flow characteristics were carried out in macro-scale dimensions. Only a few studies addressed the nanofluid flow and heat transfer in micro-scale dimensions. In numerical aspect, Koo and Kleinstreuer (2005) and Jang and Choi (2006) studied the MCHS performance numerically using different models for the effective thermal conductivity of the nanofluids. Sabbaghzadeh and Ebrahimi (2007) derived a general expression for the effective thermal conductivity of nanofluid containing cylindrical nanoparticles by considering the nanolayer effect. Our objective is to present the numerical method for explaining the nanolayer effect in the cooling performance of MCHS with carbon nanotube fluid suspensions.

5.1 Governing Equation

The problem of forced convective flow through a microchannel heat sink of water-based nanofluids containing CNTs with 25nm diameter and various nanolayer thicknesses is numerically considered. Figure 9 shows the physical and numerical domains. The top surface is insulated and the bottom surface is uniformly heated. A coolant passes through the microchannel heat sink and takes the heat away from a heat-dissipating component attached below.

5.2. Numerical results:

Based on the numerical results (Ebrahimi et al,2009), figure 10 shows colored temperature contours of a cross-sectional area of a microchannel heat sink containing water, CNT water-based nanofluids ($f_{np} = 0.1\%$, $d_{np} = 25$ (nm)), with various nanolayer thicknesses while the pumping power was fixed at 2.27 W and the heat flux was fixed at $300\ \mathrm{W/cm^2}$. Figure 10 also shows that with increasing the nanolayer thicknesses, the temperature gradients decrease. So it is beneficial to increase nanolayer thicknesses by CNTs surface treatment.

For nanofluids containing carbon nanotubes with 0.1% volume fraction, in fixed pumping power of 2.5W, the thermal resistance of microchannel heat sink will be decreased nine percent in respect to water. Analytical and numerical investigations show that using CNTs with good surface treatment increases heat removal in microchannel heat sinks.

A combination of microchannel heat sink (smallcharacteristic length) with nanofluids has been introduced as a new method for high cooling performance. The temperature contours of microchannel heat sink with nanofluids containing cylindrical nanoparticles have been numerically investigated. We have shown that by increasing the nanolayer thickness, enhancement in thermal conductivity and decrease in Nusselt number could remarkably be observed. The temperature gradient in microchannel heat sink will be decreased if nanolayer thickness increases. The thermal resistance of microchannel heat sinks with nanofluids containing CNTs is reduced with respect to nanofluids containing spherical nanoparticles.

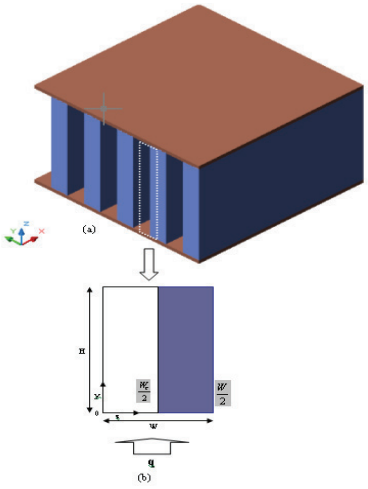


Fig. 9. Schematic of a microchannel heat sink: (a) physical domain, (b)numerical domain.

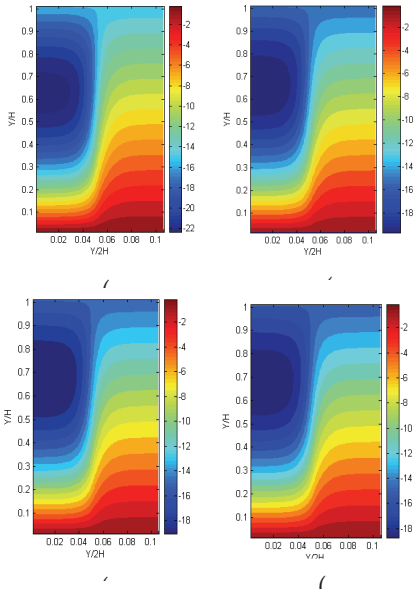


Fig. 10. Temperature counters for distilled water, 300 W/cm^2, K_f=0.6 W/mK (a), and MWCNTs dispersed in water with 0.1% volume fraction and d_{nl}=1(nm), K_{nf}=0.75W/mK (b), d_{nl}=3(nm), K_{nf}=0.8W/mK (c), and d_{nl}=5(nm), K_{nf}=0.83W/mK (d).

6. References

Ambatipudi, K.K.& Rahman, M. M. (2000). Analysis of conjugate heat transfer in microchannel heat sinks. *Numerical Heat Transfer Part A,* Vol. 37, Issue 7, 711–731, ISSN: 1040-7782

Assael, M. J.; Metaxa, I. N.; Kakosimos,K. & Constantinou, D. (2006). Thermal Conductivity of Nanofluids-Experimental and Theoretical, *International Journal of Thermophysics*, Vol.27, No.4, (July 2006)999-1017, ISSN: 0195-928X

Berber, S.; Kwon, Y. K. & Tomanek, D. (2000). Unusually high thermal conductivity of carbon nanotubes. *Phys. Rev. Lett.* Vol. 84, Issue 20, 4613-4616, ISSN: 0031-9007

Bergles, A. E. (2002). ExHFT for fourth generation heat transfer technology. *Experimental Thermal and Fluid Science,* Vol. 26, Issues 2-4, (June 2002) 335–344 ISSN: 0894-1777

Blasse, G. & Grabmeier, B. C. (1994). *Luminescent Materials,* Springer Verlag, ISBN-10: 0387580190, Berlin

Brinkman. H. C. (1952). The viscosity of concentrated suspensions and solutions. *J. Chem. Phys.,* Vol. 20, Issue 4, 571-581, ISSN: 0021-9606

Chein, R. & Hunag, G. (2005). Analysis of microchannel heat sink performance using nanofluids. *Applied Thermal Engineering*, Vol. 25, Issues 17-18, 3104–3114, ISSN: 1359-4311

Choi, S. U. S.; Zhang, Z. G.; Yu, W.; Lockwood, F. E. & Grulke, E. A. (2001). Anomalous thermal conductivity enhancement in nanotube suspensions. *Appl. Phys. Lett.,* Vol. 79, Issue 14, 2252, ISSN: 0003-6951

Chopkar, M.; Das, P.K. & Manna, I. (2006). Synthesis and characterization of nanofluid for advanced heat transfer applications. *Scripta Materialia*, Vol. 55, Issue 6, 549-552, ISSN: 1359-6462

Counio, G.; Esnouf, S.; Gacoin, T. & Boilot, J. P. (1998). Synthesis and photoluminescence of $Cd_{1-x}Mn_xS(x\leq5)$, *J. Phys. Chem B,* Vol. 102, No. 27, 5257-5260, ISSN: 0953-4075

Dalalli Isfahani, T.; Moghimi-Rad, J. & Sabbaghzadeh, J., Shaped Controlled Synthesis of Silver Nanoparticles by Surfactant Self-assembly, *Proceeding of 2nd Iran-India Joint Conference on Nanotechnology*, May 2009, University of Isfahan, Isfahan

Ding, Y., Alias, H., Wen, D.& Williams, R. A. (2006). Heat transfer of aqueous suspensions of carbon nanotubes (CNT nanofluids). *International Journal of Heat and Mass Transfer,* Vol. 49, No. 1-2, 240–250, ISSN: 0017-9310

Eastman, J. A.; Choi, S. U. S.; Li, S.; Yu, W. & Thomson, L. J. (2001). Anomalously increased effective thermal conductivities of ethylene glycol based nanofluids containing copper nanoparticles. *Appl Phys Lett.,* Vol. 78, Issue 14, 718–720, ISSN: 0003-6951

Ebrahimi, S.; Sabbaghzadeh, J.; Lajvardi, M. & Hadi, I. (2007). Nanolayer effects in thermal conductivity of nanofluids containing cylindrical nanoparticles. *Proceeding of IEEE-NANO 2007*, pp. 1195-1198, Hong Kong, Aug. 2007, IEEE Xplore, Hong Kong

Ebrahimi, S.; Sabbaghzadeh, J.; Lajvardi, M. & Hadi, I., Nanolayer effects in cooling performance of a microchannel heat sink with nanofluids containing CNTs. *Proceeding of IEEE-NEMS 2009*, pp. 845-848, China, Jan. 2009, IEEE Xplore, Shenzhen

Fedorov, A. G. & Viskanta, R. (2000). Three-dimensional conjugate heat transfer in the microchannel heat sink for electronic packaging. *International Journal of Heat and Mass Transfer,* Vol. 43, Issue 3, 399–415, ISSN: 0017-9310

Foygel. M; Morris. R. D.; Anez. D; French. S. & Sobolev, V. L. (2005). Theoretical and computational studies of carbon nanotube composites and suspensions: Electrical and thermal conductivity. *Phys. Rev. B,* Vol. 71, Issue 10, 104201-104208, ISSN: 1098-0121

Hamilton, R. L. & Crosser, O. K. (1962). Thermal conductivity of heterogeneous two-component systems. *Ind. Eng. Chem. Fundamental*., Vol. 1, Issue 3, 187–191, ISSN: 0196-4313

Hone, J.; Whitney, M. & Zettl, A. (1999). Thermal conductivity of single walled carbon nanotubes. *Synth. Met.,* Vol. 103, Issues 1-3, 2498-2499, ISSN: 0379-6779

Huignard, A.; Gacoin, T.; Chaput, F. & Boilot, J. (2001). Synthesis and luminescence properties of colloidal lanthanide. *Proceeding of Mat. Res. Soc. Symp.,* San Francisco, April 2001, MRS, San Francisco

Incropra, F. P. & DeWitt, D. P. (2002). *Introduction to Heat Transfer (Fourth Edition),* John Wiley & Sons, ISBN: 978-0-471-72226-7, New York

Jang, S. P. & Choi, S. U. S. (2004). Role of Brownian motion in the enhanced thermal conductivity of nanofluids. *Appl. Phys. Lett.*, Vol. 84, No. 21, 4316–4318, ISSN: 0003-6951

Jang, S. P. & Choi, S. U. S. (2006) Cooling performance of a microchannel heat sink with nanofluids, *Applied Thermal Engineering*, Vol. 26, Issues 17-18, 2457–2463, ISSN: 1359-4311

Kandlikar, S. G. & Grande, W. J. (2003). Evolution of microchannel flow passages – thermohydraulic performance and fabrication technology. *Heat Transfer Engineering,* Vol. 24, No. 1, 3–17, ISSN: 0145-7632

Keblinski, P.; Eastman, J.A. & Cahill, D.G. (2005). Nanofluids for thermal transport, *Materials Today*, Vol. 8, Issue 6, (June 2005), 36-44, ISSN: 1369-7021

Keblinski, P.; Phillpot, S. R.; Choi, S. U. S. & Eastman, J. A. (2002). Mechanisms of heat flow in suspensions of nano-sized particles (nanofluids). *Int. J. Heat. Mass Transfers,* Vol. 45, Issue 4, 855-863, ISSN: 0017-9310

Khanafer, K.; Vafai, K. & Lightstone, M.. (2003). Buoyancy-driven heat transfer enhancement in a two-dimensional enclosure utilizing nanofluids. *Int. J. Heat Mass Transfer,* Vol. 46, Issue 19, 3639–3653, ISSN: 0017-9310

Kim, P.; Shi, L.; Majumdar, A. & Mceuen, P. L. (2001). Thermal transport measurements of individual multiwalled nanotubes. *Phys. Rev. Lett.,* Vol. 87, Issue 21, 215502-215505, ISSN: 0031-9007

Kim, S. J. & Kim, D. (1999). Forced convection in microstructures for electronic equipment cooling. *ASME Journal of Heat Transfer,* Vol. 121, No. 3, 639–645, ISSN: 0022-1481

Knight, R. W.; Hall, D. J.; Goodling, J. S. & Jaeger, R.C. (1992). Heat sink optimization with application to microchannels. *IEEE Transactions on Components, Hybrids, and Manufacturing Technology,* Vol. 15, No. 5, 832–842, ISSN: 0148-6411

Koo, J. & Kleinstreuer, C. (2004). A new thermal conductivity model for nanofluids. *Journal of Nanoparticle Research,* Vol. 6, No. 6, 577–588, ISSN: 1388-0764

Koo, J. & Kleinstreuer, C. (2005). Laminar nanofluid flow in microheatsinks. *International Journal of Heat and Mass Transfer*, Vol. 48, No. 13, 2652–2661, ISSN: 0017-9310

Lee, P., Garimella, S. V. & Liu, D. (2005). Investigation of heat transfer in rectangular microchannels. *International Journal of Heat and Mass Transfer,* Vol. 48, Issue 9, 1688–1704, ISSN: 0017-9310

Li, J. & Peterson, G. P. (2006). Geometric optimization of a micro heat sink with liquid flow. *IEEE Transactions on Components and Packaging Technologies,* Vol. 29, No. 1, 145–154, ISSN: 1521-3331

Li, J.; Peterson, G.P.& Cheng, P. (2004). Three-dimensional analysis of heat transfer in a micro-heat sink with single phase flow. *International Journal of Heat and Mass Transfer,* Vol. 47, No. 19-20, 4215–4231, ISSN: 0017-9310

Maxwell, J. C. (1881). *A Treatise on electricity and magnetism, 2nd edn*., Vol. 1. Clarendon Press, Oxford, UK

Min, J. Y.; Jang, S. P. & Kim, S. J. (2004). Effect of tip clearance on the cooling performance of a microchannel heat sink. *International Journal of Heat and Mass Transfer,* Vol. 47, Issue 5, 1099–1103, ISSN: 0017-9310

Pak, B. C. & Cho, Y. L. (1998). Hydrodynamic and heat transfer study of dispersed fluids with submicron metallic oxide particles. Experimental Heat Transfer, Vol. 11, No. 2, 151–170, ISSN: 0891-6152

Patel, H. E.; Sundararajan, T. & Das, S. K. (2008). A cell model approach for thermal conductivity of nanofluids A micro-convection model for thermal conductivity of nanofluids. J. Nanopart Res., Vol. 10, No. 1, 87–97, ISSN: 1388-0764

Qu, W.& Mudawar, I. (2002). Experimental and numerical study of pressure drop and heat transfer in a single-phase microchannel heat sink. International Journal of Heat and Mass Transfer, Vol. 45, Issue 3-4, 2549–2565, ISSN: 0891-6152

Ren, Y.; Xie, H. & Cai, A. (2005). Effective thermal conductivity of nanofluids containing spherical nanoparticles, *J. Phys. D Appl. Phys*., Vol. 38, Issue 21, 3958–3961, ISSN: 0022-3727

Riwotzki. K & Haase. M. (1998). Wet-chemical synthesis of doped colloidal nanoparticles: YVO_4:Ln (Ln=Eu, Sm, Dy). *J. Phys. Chem B*, Vol. 102, No. 50, 10129-10135, ISSN: 1520-6106

Ryu, J. H.; Choi, D. H. & Kim, S. J. (2003). Three-dimensional numerical optimization of a manifold microchannel heat sink. *International Journal of Heat and Mass Transfer,* Vol. 46, Issue 9, 1553–1562, ISSN: 0891-6152

Ryu, J. H.; Choi, D. H. & Kim, S. J. (2002). Numerical optimization of the thermal performance of a microchannel heat sink. *International Journal of Heat and Mass Transfer*, Vol. 45, Issue 13, 2823-2827, ISSN: 0891-6152

Sabbaghzadeh, J. & Ebrahimi, S. (2007). Effective of thermal conductivity of nanofluids containing cylindrical nanoparticles. *Int. J. Nanoscience*, Vol. 6, Issue 1, 45-49, ISSN: 1793-5350

Tiselj, I.; Hetsroni, G.; Mavko, B.; Mosyak, A.; Pogrebnyak, E.& Segal, Z. (2004). Effect of axial conduction on the heat transfer in microchannels. *International Journal of Heat and Mass Transfer,* Vol. 47, Issue 12-13, 2551–2565, ISSN: 0891-6152

Tiwari R. K. & Das, M. K. (2007). Heat transfer augmentation in a two-sided lid-driven differentially heated square cavity utilizing nanofluids. *Int. J. Heat Mass Transfer*, Vol. 50, No. 9-10, 2002-2018, ISSN: 0891-6152

Trisaksri, V. & Wongwises, S. (2007). Critical review of heat transfer characteristics of nanofluids. *Renewable & Sustainable energy reviews*, Vol. 11, Issue 3, 512-523, ISSN: 1364-0321

Tuckerman, D. B. & Pease, R.F. (1981). High-performance heat sinking for VLSI. *IEEE Electronic Devices Letters EDL,* Vol. 2, Issue 5, 126–129, ISSN: 0741-3106

Wang, L. & Wei, X. (2007). Heat conduction in nanofluids. *Chaos Solitons & Fractals*, Vol. 39, Issue 5, 2211-2215, ISSN: 0960-0779

Wang, X.; Xu, X. & Choi, S. U. S. (1999). Thermal conductivity of nanoparticle–fluid mixture. *J. Thermophys. Heat Transfer*, Vol. 13, No. 4, 474–480, ISSN: 0887-8722

Webb, R.L. (1993). *Principles of Enhanced Heat Transfer*. John Wiley & Sons, ISBN-10: 0471577782, New York.

Xie, H.; Wang, J.; Xi T.; Liu, Y. & Ai, F. (2002). Thermal conductivity enhancement of suspensions containing nanosized alumina particles. *J. Appl. Phys*., Vol. 91, No. 7, 4568-4572, ISSN: 0021-8979

Xie, H.; Youn, H. W. & Choi, M. (2003). Nanofluids containing multiwalled carbon nanotubes and their enhanced thermal conductivities. *J. Appl. Phys*., Vol. 94, Issue 8, 4967-4971, ISSN: 0021-8979

Xuan, Y. & Li, Q. (2000). Heat transfer enhancement of nano-fluids. *Int. J. Heat Fluid Flow*, Vol. 21, Issue 1, 58–64, ISSN: 0142-727X

Xuan, Y.& Li, Q. (2003). Investigation on convective heat transfer and flow features of nanofluids. *ASME Journal of Heat Transfer*, Vol. 125, Issue 1, 151–155, ISSN: 0022-1481

Xuan, Y.; Li, Q. & Hu, W. (2003). Aggregation structure and thermal conductivity of nanofluids, *AIChE J* ., Vol. 49, No. 4, 1038–1043, ISSN: 0001-1541

Xuan, Y.& Roetzel, W. (2000). Conceptions for heat transfer correlation of nanofluids. *International Journal of Heat and Mass Transfer,* Vol. 43, No. 19, 3701–3707, ISSN: 0891-6152

Xue, Q. (2003). Model for effective thermal conductivity of nanofluids. *Physics Letters A*, Vol.307, No. 5-6, 313-317, ISSN: 0217-7323

Yang, Y.; Zhang, Z.; Grulke, E.A. & Anderson, W. B. (2005). Heat transfer properties of nanoparticle in fluid dispersions (nanofluids) in laminar flow. *International Journal of Heat and Mass Transfer*, Vol. 48, No. 6, 1107–1116, ISSN: 0891-6152

Zhao, C. Y. & Lu, T. J. (2002). Analysis of microchannel heat sinks for electronics cooling. *International Journal of Heat and Mass Transfer*, Vol. 45, Issue 24, 4857–4869, ISSN: 0891-6152

13

Intersubband Transitions in the Quantum Dot Layers for Quantum Confined Photodetector

Shiang-Feng Tang[1*], Tzu-Chiang Chen[2*],
Shih-Yen Lin[3*] and Hsing-Yuan Tu[4*],
[1]*Chung-Shan Institute of Science and Technology*
[2]*National Defense University*
[3]*Academia Sinica*
[4]*National Taiwan University*
Taiwan, R.O.C.

1. Introduction

In nanostructure like quantum dots (QDs) embedded in the spacing layer with high energy barrier, where electrons are three-dimensionally confined into nanometer-scale semiconductor structures, the novel physical characteristics are expected to emerge. The novel properties would greatly improve semiconductor device performance. Optoelectronic devices with quantum dot heterostructure like quantum dot infrared photodetector (QDIP) have already been proposed in the recent years. When QDs are incorporated into the layered structure of a semiconductor for optoelectronic device applications, electrical control is critical to the operation of device. It is desirable that both an electric field can be applied to change physical properties of the QDs embedded in the spacing layers i.e., the active region of the device and that the photo-induced carriers can be excited and transited to generate the photocurrent. In this chapter, The pseudopotential model of using multiple-quantum-dot (MQD) structures for detect infrared radiation can be explained by exploiting the basic principles of quantum mechanics, with the uniform and isotropic strain-induced potential to well simulate the electronic properties of InAs/In(Ga)As QDIP active region by finite element method (FEM). The vertically coupled and decoupled wave-functions of electrons on MQD with dependences of thickness of spacing layers are also calculated by means of the FEM. The method is ideally suited for numerical analysis by computer. The typical and particular QDIP structure are involved and introduced in this section. Here, the outstanding performance of the QDIP which has emerged as a potential alternative to QWIPs would be proposed. The motivation for interest in QDIPs is rooted in two characteristics of quantum dots. The first is that QDIPs are sensitive to normal-incident infrared radiation, a consequence of the 3-D confinement of electrons in the quantum dots. The other attribute is the weak thermionic coupling between the ground state and excited states. This should result in lower thermal excitation and, thus, lower dark current and higher operating temperature. The concomitant increase in the lifetimes of excited carriers should enable higher responsivities as carriers have more time to escape and contribute to the

photocurrent before relaxing to the ground state. Finally, the dependence of thickness of spacing layer affects enormously the coupling level of electron wave-function and the spectrum shifting of photon energy, where it is found to be consistent with experimentally measured intersubband transitions for the schemed QDIP.

2. Description of quantum-confined theory

An introduction into the electronic and optical properties of compound semiconductors in principle will be given. First a 3-dimensional (3D) volume is considered, followed by the lower dimensional structures and their quantum properties. The case of a quantum well (2D) will be first treated, because the principle of heterostructure properties can be explained with it. Quantum wire (1D) and quantum dot (0D) properties will then be derived in a similar way.

2.1 Wave-function in quantum structures (Alexander Weber, 1998)

2.1.1 3D bulk material

In a 3-dimensional crystal the movement of carriers (electrons, heavy holes or light holes) near to the band edge can be described as the motion of a quasi free particle, whose effective mass m* takes into account the interaction with the periodical lattice potential. In first approximation m* does not depend on the direction and a continuous energy spectrum of eigenvalues, which are isotropically distributed in the *k*-space, is obtained:

$$E^{3D}\left(\vec{k}\right)=\frac{\hbar^2}{2m^*}\left(k_x^2+k_y^2+k_z^2\right) \tag{1}$$

where k_x, k_y, k_z are the wavevectors along the *x*, *y* and *z*-axis.If the carriers are confined in lower dimensional systems such as 2-dimensional wells or 1-dimensional wires or 0D quantum dots with sizes of the order of the de Broglie wavelength of the carriers, a quantum structure in the electronic properties, for example the density of states, appears.

2.1.2 2D quantum well

For example in finite barrier, Fig. 1 shows the scheme of unstrained bandstructure of an InAs/$In_{0.52}Al_{0.48}As$ quantum well. Electrons and holes are trapped in the one-dimensional well, but can still move freely in the InAs layer. If the thickness of the active layer is of the order of the de Broglie wavelength of the carriers¸ Broglie = *hp* (*p* being the momentum of the carrier and *h* the Planck constant), quantum effects will appear, i.e., the quantization of the kinetic energy in the growth direction leads to discrete energy levels in the k_z direction in both the conduction band and the valence band.

The calculation of the eigenenergies in the confined structure using the effective mass approximation is now illustrated, with assuming that an idealized square, finite and symmetrical potential well with a thickness L_z and a potential energy

$$V_{well}\left(Z\right)=\begin{cases}0 & |z|<\frac{L_Z}{2}\\ -V_{well} & |z|>\frac{L_Z}{2}\end{cases} \tag{2}$$

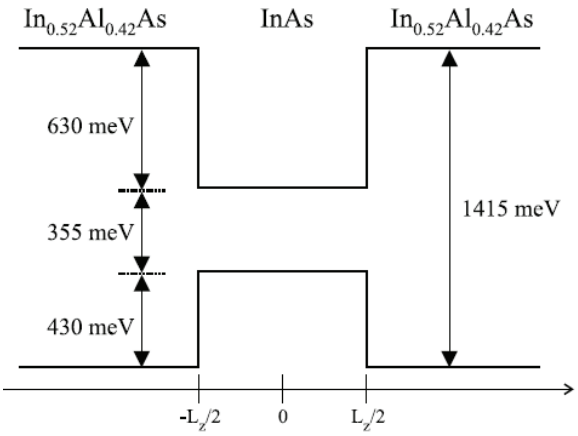


Fig. 1 Scheme of the band structure of InAs/$In_{0.52}Al_{0.48}As$.

The Schrödinger equation is:

$$\left[-\frac{\hbar^2}{2m}\Delta + v_{lattice(\vec{r})} + V_{well}(Z)\right]\Psi_{tot}(\vec{k},\vec{r}) = E(\vec{k})\Psi_{tot}(\vec{k},\vec{r}) \tag{3}$$

where V_{lattice} is the periodical, rapidly oscillating lattice potential, which describes the interaction of the carriers with the crystal lattice. Since it oscillates on a much smaller scale than the well potential V_{well}, it can be separated. This leads to the following wave-function:

$$\Psi_{tot}(\vec{k},\vec{r}) = \Psi_{env}(\vec{k},\vec{r}) \cdot \Phi_{B1}(\vec{k},\vec{r}) \tag{4}$$

The rapidly oscillating Bloch function Φ_{Bl} represents the carrier motion in the lattice potential, which can be handled with the introduction of an effective mass m^*. Ψ_{env} is the envelope function of the Bloch function and is determined by the slowly varying potential $V_{\text{well}}(z)$. Because $V_{\text{well}}(z)$ does not contain any terms in x or y, a second separation can be carried out:

$$\Psi_{env}(\vec{k},\vec{r}) = \Phi(\vec{k}_{x,y},\vec{r}_{x,y})\Theta(k_z,z) \tag{5}$$

where $\Phi(\vec{k}_{x,y},\vec{r}_{x,y})$ describes the motion in the 2 dimensions of the quantum well and leads to the energy eigenvalues:

$$E(\vec{k}_{xy}) = \frac{\hbar^2}{2m^*_{x,y}}(k_x^2,k_y^2) \tag{6}$$

with the effective mass $m^*_{x,y}$ for a motion in the layer plane.
In the z-direction a one dimensional Schrödinger equation for the motion of a free carrier with mass m^*_z in a symmetric quantum well $V_{\text{well}}(z)$ has to be solved:

$$\left[-\frac{\hbar^2}{2m_z^*}\frac{\partial^2}{\partial z^2}+V_{well}(Z)\right]\Theta(k_z,z)=E_z\Theta(k_z,z) \tag{7}$$

The following boundary conditions must be fulfilled:

1. $\Theta(k_z,z)$ is continuous everywhere;
2. by integrating equation (7) around any z_0, and for the specific potential of the square quantum well, $\frac{1}{m_z^*}\frac{d\Theta}{dz}$ is continuous everywhere;
3. $\lim_{z\to\pm\infty}|\Theta(k_z,z)| = 0.$

The second condition accounts for the discontinuity of m^*_z at the boundary layers.
The solution of the Schrödinger equation is in this case:

$$\Theta(k_z,z)=\begin{cases} A\sin\left(\sqrt{\frac{2m_z^*E_{z,nz}}{\hbar^2}}-n_z\frac{\pi}{2}\right) & |z|\le\frac{L_z}{2} \qquad n_z=1,2,3.....\\ B\exp\left(-\sqrt{\frac{2m^*_{z,barrier}(v_{well}-E_{z,nz})}{\hbar^2}}\right) & |z|\ge\frac{L_z}{2}\end{cases} \tag{8}$$

with m^*_z and $m^*_{z,\text{barrier}}$ being the effective masses in z-direction inside the quantum well and inside the barrier, respectively.
In the barrier the wavefunction is exponentially attenuated while it oscillates in the quantum well. The ground state (nz= 1) is an even function and the higher states are alternately of even or odd parity.
The boundary conditions lead to a transcendental equation for the discrete energy levels $E_{z,nz}$ in k_z-direction, which can be solved numerically:

$$\tan\left(\sqrt{\frac{2m_z^*E_{z,nz}}{\hbar^2}}\frac{L_Z}{2}-n_z\frac{\pi}{2}\right)=-\sqrt{\frac{m^*_{z,barrier}E_{z,nz}}{m_z^*(v_{well}-E_{z,nz})}} \tag{9}$$

2.1.3 1D quantum wire

If the motion of the carriers is confined in further directions of space, the additional quantization can be calculated in a way analogous to that of the quantum well. For a one dimensional system (quantum wire) with infinite barriers the energy eigenvalues are:

$$E^{1D}_{n_x,n_z}(k_y)=\frac{\hbar^2\pi^2}{2}\left(\frac{n_x^2}{m_x^*L_Z^2}+\frac{n_Z^2}{m_Z^*L_Z^2}\right)+\frac{\hbar^2k_y^2}{2m_y^*} \tag{10}$$

The energy function is again a sum of discrete and continuous eigenvalues, which leads to an uni-dimensional subband structure.

2.1.4 0D quantum dot

For a parallelepipedic quantum dot one obtains:

$$E^{0D}_{n_x,n_z,n_z} = \frac{\hbar^2\pi^2}{2}\left(\frac{n_x^2}{m_x^* L_x^2} + \frac{n_y^2}{m_y^* L_y^2} + \frac{n_Z^2}{m_Z^* L_Z^2}\right) \tag{11}$$

where $(n_x, n_y, n_z) \in (\mathrm{N3})^*$ are the quantum numbers. They are integers, but not all of them are allowed to be 0. $L_{x,y,z}$ are the sizes of the structure and $m^*_{x,y,z}$ the effective masses in the respective directions.

The carriers in a quantum dot are completely localized and only discrete energy levels exist. In the realistic case of finite potential wells, numerical calculations must be performed to find an exact solution of the Schrödinger equation (Daniel Fritsch et al., 2003). In real situations, such as self organized quantum dots, exact calculations of the discrete energy levels proved to be very difficult and have only been performed numerically for the InAs /GaAs quantum dot system. First, the exact shape of the dot is usually not known (facets of pyramids, radii in lens shape, . . .). Second, anisotropic strain largely influences the electronic properties of 1D and 0D quantum structures.

Fig. 2 shows as an example the strain distribution for an InAs/GaAs quantum dot, which has the shape of a square pyramid. Far away form the dot in the wetting layer there is biaxial strain which is entirely confined in the wetting layer: because InAs has a smaller lattice constant than GaAs, in x and y direction the InAs layer is compressed (ε_{xx} and ε_{yy} are negative) and in the z direction a tension strain can be distinguished ($\varepsilon_{zz} > 0$). The strain distribution inside the pyramid is different. Close to the lower interface, ε_{zz} is still positive but much smaller than in the wetting layer because the substrate can no longer force the interface lattice constant to be that of the substrate. With increasing height within the dot, ε_{zz} changes its sign and becomes negative at the top of the pyramid. This happens because at the very top only small forces act on the quantum dot in the xy plane, but the GaAs barrier compresses the pyramid mainly from the sides along the z direction, imposing tensile strain components in the xy plane ($\varepsilon_{xx} = \varepsilon_{yy}$ become positive). Generally, however, the strain is still compressive even at the top of the pyramid ($T_{re} < 0$). Around the pyramid the barrier also becomes significantly strained (M. Grundmann et al., 1995). Because the strain distribution and the exact shape of the dots have to be known, calculating the correct electronic structure is difficult. Most approaches have been done using a $\vec{k} \cdot \vec{p}$ model (where $\vec{k}$ is the wave vector and $\vec{p}$ is the momentum) (S. L. Chuang & C. S. Chang, 1996). It calculates the $E(\vec{k})$ relationship over a small k range around the band extreme for multiple bands. Another approach considers the dot as a structure in its own right (rather than viewing it as a perturbation of the bulk material). The model is based on a pseudopotential framework (van de Walle and Chris G., 1989) and uses no adjustable parameters outside the bulk band structure. Fig. 3 shows the electron and hole wavefunction for the ground state and excited states for a self assembled InAs/GaAs system calculated with the pseudopotential framework. As expected the ground electron state wave function lays almost completely within the dot and covers a large part of it. However, excited states represent nodes with different shapes localized at the corners of the pyramid. Hence, the small dimensions of the dot and the presence of anisotropic strain have large effects on the form of the wavefunctions.

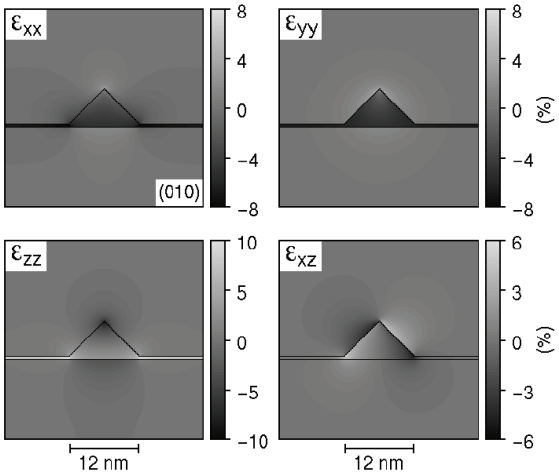


Fig. 2. Strain distribution in the (010) plane through the top of the pyramid of an InAs quantum dot grown on GaAs (O. Stier, M. et al., 1999)

Although the two approaches give a fair description of the hole and electron wavefunctions, the models strongly differ on the energetic structure of the dots especially for the excited states. They also do not account for the depletion of the wetting layer around the pyramids.

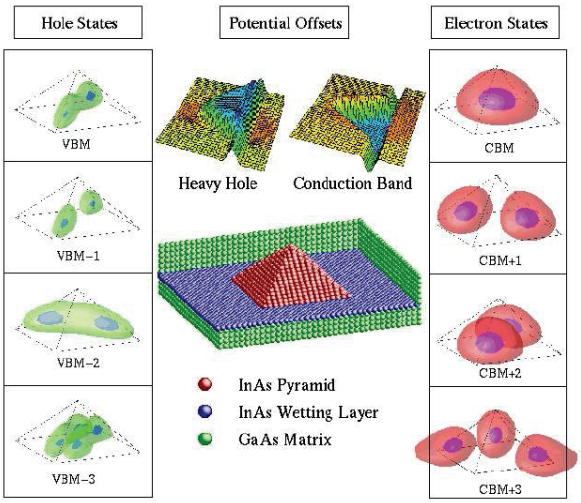


Fig. 3. The electronic structure of a strained InAs (110) pyramidal quantum dot embedded within GaAs. The strain-modified band offsets are shown above the atomic structure. They exhibit a well for both heavy holes and electrons. Isosurface plots of the four highest hole states and four lowest electron states, as obtained from pseudopotential calculations, appear

on the left and right. CBM means conduction band minimum and VBM valence band minimum.

2.2 Optical transitions

In this section, the main optical properties of low dimensional structures are discussed. The case of a single quantum well is firstly considered and lower dimensional structures are then treated in a similar way. Two kinds of optical transitions are considered (Fig. 4). Interband transitions take place between the conduction band and the valence band and involve two kinds of carriers (they are bipolar), electrons and holes. The energy of the transition is the bandgap energy plus the confinement energies of the electrons and holes minus the exciton binding energy. Intraband transitions happen inside either the conduction or the valence band and involve only one type of carrier (the transition is unipolar). In a quantum dot, intraband transitions occur between discrete energy levels. In quantum wells and quantum wires there exist subbands inside the conduction or the valence band. Intraband transitions in these structures between two subbands are called intersubband transitions. Note that another type of intraband transitions may occur which involves the transitions of a carrier from one subband to the same subband with absorption of a photon and emission of a phonon (momentum conservation). The latter is the analogue of free carrier absorption.

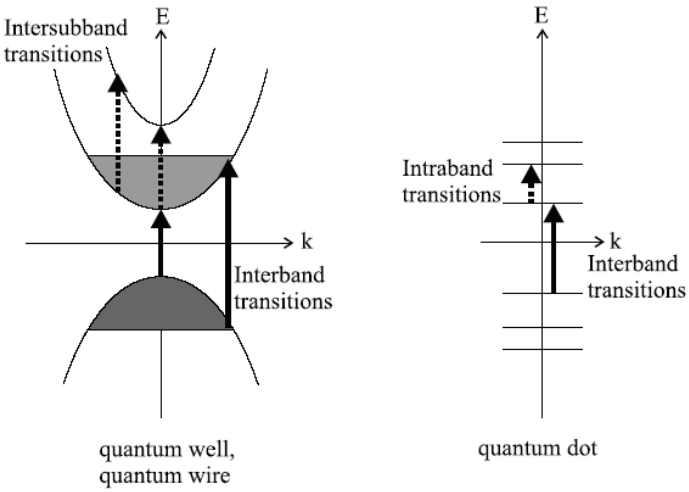


Fig. 4. Interband and intraband transitions for quantum wells, quantum wires (left) and quantum dots (right). The diagrams show a scheme of the band/level structure.

For 0D systems, there is *a priori* no forbidden direction but the actual polarization of intraband transitions between confined levels will depend on the spatial wave function symmetry of the states involved (see section 2.1.4), since the dipole is $\mu_{if} = e\left\langle \Psi_i \middle| \vec{\varepsilon} \cdot \vec{r} \middle| \Psi_f \right\rangle$. This means that the polarization of the intraband dipole can be predicted by means of

simple symmetry considerations. For example, the transition VBM→VBM-1 in Fig. 3 is polarized along the growth direction, whereas VBM→VBM-2 is polarized in the wetting layer plane. Normal incidence intraband absorption in quantum wells is usually forbidden. However, in quantum wires and quantum dots in-plane polarized transitions become possible, which is of great interest for the development of IR photodetectors irradiated at normal-incidence. Intraband spectroscopy is a validly experimental technique for investigating low dimensional semiconductors, since one can get insight on the confinement energies and on the spatial symmetries of the excited states, because the polarizations of the transitions can be measured (M. A. Cusack et al., 1997).

2.3 Oscillator strength

The oscillator strength of an intraband transition between the ground state and the first excited state is:

$$f = \frac{2m_0 E_{21}}{e^2\hbar^2}\mu_{12}^2 \tag{12}$$

where E_{21} is the energy difference between the two states. The oscillator strength of intersubband transitions does not depend on the energy of the transition, i.e., on the width of the quantum well, but only depends on the carrier effective mass which is material dependent: Because the intersubband energy $E_{21} \approx 3\hbar^2\pi^2/(2m^*L_z^2)$. Lower effective masses give larger oscillator strengths.

2.4 Intraband absorption

The intraband absorption coefficient for the 1→ 2 transition in a quantum structure can be expressed as (Shu-Shen Li, & Jian-Bai Xia, 1997):

$$\alpha(\omega) = \frac{\pi E_{21} e^2 (n_1 - n_2)}{2\varepsilon_0 c\tilde{n} m_0 \omega\Omega} \cdot f \cdot g(E_{21} - \hbar\omega) \tag{13}$$

where $n_1 \rightarrow n_2$ is the number of carriers, which can absorb in the active volume Ω of the quantum structure, ε_0 the dielectrical constant and $\breve{n}$ the refractive index. f is the oscillator strength (equation (12)) and $g(\omega)$ is a lineshape function.

The spectral lineshape $g(\omega)$ takes into account several contributions:

- The homogeneous spectral width due to the finite coherence time between the two levels. This width can be expressed by a Lorentzian lineshape:

$$L(\omega) = \frac{1}{\pi}\frac{\hbar\Gamma}{(E_{21} - \hbar\omega)^2 + \hbar^2\Gamma^2} \tag{14}$$

  where $2\hbar\Gamma$ is the full width at half maximum (FWHM) of the intraband resonance.
- The inhomogeneous broadening caused by imperfections of the structure. In the case of a quantum well, imperfections are mainly variations of the width of the layer. For quantum wires and dots, size distributions of the structures are the origin of this lineshape broadening. For the example of quantum dots, this can be explained as followed: Smaller dots have higher energy levels and also greater differences between two states in one

'band', while greater dots possess smaller energy differences between two states. As the absorption takes place in a huge number of quantum dots, the sum of all their narrow absorption lines at different energy positions will be observed. The result is that the absorption spectrum overtakes the shape of the size distribution function of the dots.

- The inhomogeneous spectral width introduced by the non-parabolicity or, more general, by the coupling with other levels. This causes an asymmetric lineshape (K. Leifer et al., 2007).

3. Introduced finite element method for confined states and preparation of quantum confined structures for quantum dot infrared photodetector

3.1 Description of study background for strained InAs QDs embedded in surfactant $In_{0.1}Ga_{0.9}As$ within thickness-modified GaAs spacer layer

The analysis of strain effects in a quantum mechanical model of semiconductor devices has only recently been attempted by C. Pryor et al. (C. Pryor et al., 1998), Williamson et al., (A. J. Williamson et al., 2000) and Grundmann et al., who calculate strain-induced potentials and wave functions in quantum dots (QDs), and Gerhard Klimeck et.al (Gerhard Klimeck et al., 2002), who reviews the topic of electronic structure in dome-shaped semiconductor QDs based on atomistic methods. Few studies have made fitting with experimental measurements; Pistol et al. model strain effects on the band gap in buried QDs, which is consistent with photoluminescence (PL) data firstly. Models of semiconductor devices by means of the finite element method (FEM) have been proposed by a number of authors for one- and two dimensionally quantum-confined problems (G. R. Liu & S .S Quek Jerry, 2002, T. Benabbas et al., 1999),. The time–independent Schrödinger equation including the strain-induced potential is solved numerically by means of the finite element method to obtain the spectrum of energies and wave functions of available states. In the technique presented here, calculations of bandstructures on the heterostructure of QDs are well suited for the FEM, which is a common tool in continuum mechanics, to the best of our knowledge. The main idea of the method is that an unknown continuous field in the domain is represented approximately in terms of its values at discrete points (nodes) within the domain; the goal is to determine optimal values for these nodal quantities. The domain is covered with areas (in two dimensions) or volumes (in three dimensions) whose boundaries are defined geometrically by the nodes; these areas or volumes are the elements. Fields are defined within each element in terms of the values of the nodal quantities on its boundary by means of a suitable interpolation scheme. In the paper, the energy band levels and wave functions of InAs QDs buried in $In_{0.1}Ga_{0.9}As$/GaAs barriers have been calculated using a three-dimensional (3D) FEM formulation incorporated with the strained structures. In earlier publication (Tzu-Huan Huang et al., 2007), it is treated in the similar quantum structure with thicker GaAs spacer i.e., 30nm and the simplified without considering the vertically electronic coupling due to the larger barrier blockade and with the coherent QD geometry and constant spacing among the QDs. It is reasonable to infer that the multiple active QD regions are merely considered as increasing the intensity of photoluminescence (PL) and photoresponse (PR). By above assuming, earlier work is made with from single QD to the array of 9 QDs in a single layer for the calculation of electron and hole confined levels contrast to the established QDIP grown structure. The simulated results are agreed with the PL and PR experimental data. And the calculations of the interband and intraband

transitions can be well predicted the morphology of QD matrix to be consistent with the cross-sectional transmission electron microscopy (XTEM) images. The QD geometry is characterized to lens-shaped with 20 nm in base diameter and 3 nm in height.
Under our calculation, we have chosen to focus on the well-established 'lens-shaped' dot geometry, because the shape of QD is considered as the experimental TEM image, which is available in most cases of QD growth. In theoretical simulations based on the schematic 3-periodic QD structures as shown in Fig. 5, it is demonstrated that the 0, 1, 2, 3, 4 and 15 nm-thick undoped GaAs spacer/5 nm-thick undoped $In_{0.1}Ga_{0.9}As$ top surfactant/3 nm-nominal thick doped InAs QD/1 nm doped $In_{0.1}Ga_{0.9}As$ bottom surfactant as the active region and considered the influences on wave-functions of the vertically electronic inter-coupling mechanism due to the modified spacer layers.

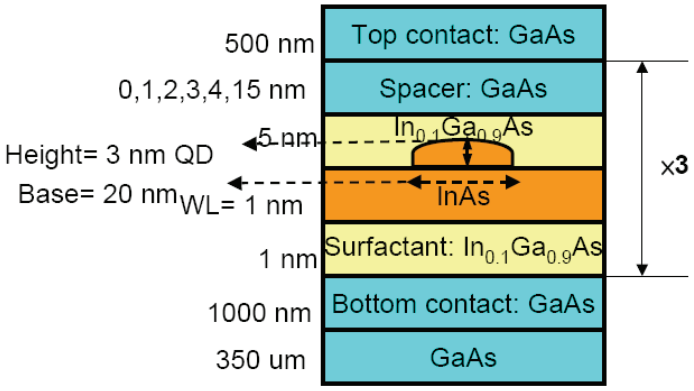


Fig. 5. The schematic 3 period of QD heterostructure with cross-sectional lens-shaped geometry.

In this section, we present a theoretical study of the electronic properties of these structures of self-assembled quantum dots (SAQD's) used the single-band effective mass approximation theory (EMA) to calculate the energy levels and wave-functions in $InAs/In_xGa_{1-x}As/GaAs$ lens-like shaped QDs. If the energy dispersion relation for a single band near k_0 ,i.e. the integration over k is mainly contributed from $k=k_0$ (an extremum or zone center), and is given by

$$E_n(k) = E_n(k_0) + \sum_{\alpha,\beta} \frac{\hbar^2}{2}\left(\frac{1}{m^*}\right)_{\alpha\beta} k_\alpha k_\beta \tag{15}$$

for the Hamiltonian H_0 with a periodic potential

$$H_0 = \frac{P^2}{2m_0} + V(r) \tag{16}$$

$$H_0\varphi_{nk}(r) = E_n(K)\varphi_{nk}(r) \tag{17}$$

then the solution for the Schrödinger equation with a perturbation $U(\mathbf{r})$ such as an impurity or well potential

$$\left[H_0 + U(r)\right]\varphi(r) = E\varphi(r) \tag{18}$$

is obtained by solving

$$\left[\sum_{\alpha,\beta}\frac{\hbar^2}{2}\left[\frac{1}{m^*}\right]_{\alpha\beta}(-i\frac{\partial}{\partial x_\alpha})(-i\frac{\partial}{\partial x_\beta}) + U(r)\right]F(r) = \left[E - E_n(K_0)\right]F(r) \tag{19}$$

for the envelope function $F(\mathbf{r})$ and the energy E. The wave function is

$$\varphi(r) = F(r)u_{nk_0}(r) \tag{20}$$

The most important result is that the periodic potential $V(\mathbf{r})$ determines the energy bands the effective masses, $\left(\frac{1}{m^*}\right)_{\alpha\beta}$, and the effective mass equation (Eq.19) contains only the extra perturbation potential $U(\mathbf{r})$, since the effective masses already take into account the periodic potential. The form of the Schrödinger equation to be solved on the finite element mesh is obtained by minimizing the total variation of the weak, or Galerkin form of the equation with respect to the wave function. The minimum in variation with respect to the wave function corresponds physically to a minimum energy. A general discussion of the variational formulation of the finite element method is given by H. T. Johnson et al., (H. T. Johnson et al. 1998). The functional corresponding to the weak form of the time-independent Schrödinger equation with a non-uniform potential is given by

$$\Pi(\Psi^{\alpha}) = \int_R \nabla\Psi^{\alpha} L^{\alpha\beta}\Psi^{\beta} dR + \int_R \Psi^{\alpha} V^{\alpha\beta}\Psi^{\beta} dR - E\int_R \Psi^{\alpha}\Psi^{\beta} dR \tag{21}$$

where Ψ^{α}, $L^{\alpha\beta}$, and $V^{\alpha\beta}$ are functions of position in the structure. The term $L^{\alpha\beta}$ is taken to be constant within each element of the mesh. The fields Ψ^{α}, $\Delta\Psi^{\alpha}$, and $V^{\alpha\beta}$ are represented by their nodal values. Values throughout each element are determined by interpolation according to the particular shape functions that are adopted. Fig.6 shows the mesh for this region with very highly refined in the neighboring QDs and wetting layer from single to the array of 9 QDs in one layer where large wave-function gradients are expected. The mesh size used in the calculations reported here results in a very accurate determination of the energies E and good spatial resolution of the wave-functions, particularly for the lower-energy states. Convergence of the method is found by comparing solutions for meshes with successively decreasing nodal separation. Element refinement is particularly important in the active region of the device, where the wave-function gradients are largest. Fig.7 shows unstrained and strained effects on band lineup and deformation potential on QD active

region, Considering without the strain effects, the confinement of electron or hole is determined by the corresponding band offset. It is worth to note that in the InAs region, the electron sees a potential well with depth 0.842 eV while the hole sees a potential well with depth 0.26 eV.

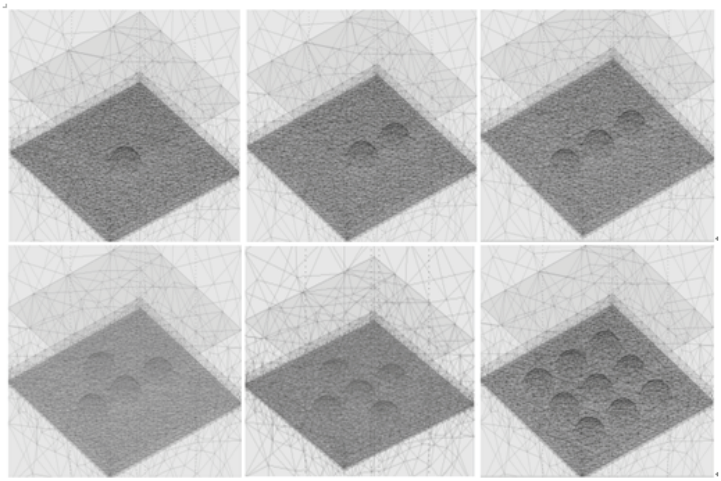


Fig. 6. The meshes in the neighboring quantum dots and wetting layer from single QD to the array of 9 QDs

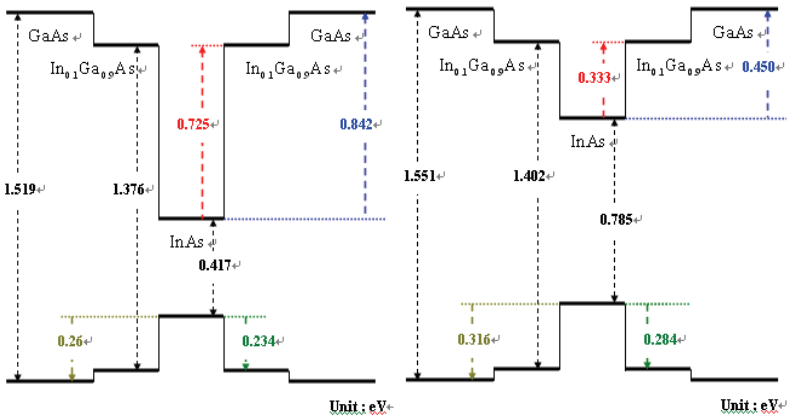


Fig. 7. Unstrained and strained effects on band lineup and deformation potential on QD active region

As well known, in the dot material, the compressive stress alters the curvature of the bulk bands causing the effective masses to differ from those of unstrained InAs. In our calculation we used the effective masses of bulk InAs under the average hydrostatic strain

present in the dot material. These values were obtained by performing semi-empirical pseudo-potential bandstructure calculations for the conduction-band valence band momentum matrix elements of InAs under strained effect. In the conduction band, these calculations yield a value for the effective mass of 0.04 m_0, a factor of two larger than in unstrained InAs of 0.023 m_0. The same trend is recovered in *ab initio* local-density calculations. However, the ab initio mass in InAs is too high and the empirical pseudo-potential result is more representative. In the absence of strain effects, the confining potential for an electron (hole) is a square well formed by the difference in the absolute energy of the conduction (valence)–band edges in InAs and GaAs. In each conventional cubic unit cell, the confining potential for each carrier type is shifted due to the strain. Since the strain varies from cell to cell, the confining potentials will also vary from cell to cell. Furthermore, degeneracies in the valence-band edge will be lifted due to deviations of the unit cells from axis symmetry. For each unit cell, the strain-induced shifts to the confining potentials are obtained by diagonalizing the 8×8 strain Hamiltonian matrix. Hence, the confining potentials including the effects of strain are piecewise continuous functions of position generated by our calculation. The compressive strain in the barrier shifts the GaAs conduction-band edge (at 1.551 eV) slightly above the unstrained level (at 1.519 eV) as shown in Fig.7.

It is assumed that the strain was taken to be a constant value between the InAs quantum dots and the surrounding InGaAs capping layer, oppositely, invalid in the surrounding GaAs spacer layer. And the main parameters including electron and heavy hole in the conduction and valence band of QD active region are listed in Table 1 and Table 2. The electron effective masses of each material system are summarized in Table1, M_GaAs, M_InGaAs_E and M_InAs_E are the electron effective masses in GaAs , InGaAs, and InAs materials, respectively. The heavy hole effective masses of each material system are summarized in Table2, M_GaAs_H, M_InGaAs_H and M_InAs_H are the heavy hole effective masses in GaAs , InGaAs, and InAs materials, respectively. The barrier height between GaAs and $In_{0.1}Ga_{0.9}As$ for electron is V1_E, oppositely, between InAs and $In_{0.1}Ga_{0.9}As$ is V2_E, and the value is 0.450 eV and 0.333 eV, respectively.

Paremeters	Values	Unit	Descriptions
C1	3.81	eV/Å	Differential coefficient $\hbar^2/2m_0$
M_GaAs_E	0.0665	m_0	Effective electron mass in GaAs, m_0 is free electron mass
M_InGaAs_E	0.064	m_0	Effective electron mass in $In_{0.1}Ga_{0.9}As$
M_InAs_E	0.04	m_0	Effective electron mass in InAs
V1_E	0.450	eV	Conduction band-edge difference between GaAs and InAs
V2_E	0.333	eV	Conduction band-edge difference between $In_{0.1}Ga_{0.9}As$ and InAs

Table 1. Simulated parameters for electron in conduction band edges (Being taken the strained effect into account)

Paremeters	Values	Unit	Descriptions
C1	3.81	eV/Å	Differential coefficient $\hbar^2/2m_0$
M_GaAs_H	0.3774	m_0	Effective hole mass in GaAs, m_0 is free electron mass
M_InGaAs_H	0.374	m_0	Effective hole mass in $In_{0.1}Ga_{0.9}As$
M_InAs_H	0.341	m_0	Effective hole mass in InAs
V1_H	0.316	eV	Valence band-edge difference between GaAs and InAs
V2_H	0.284	eV	Valence band-edge difference between $In_{0.1}Ga_{0.9}As$ and InAs

Table 2. Simulated parameters for heavy hole in valence band edges (Being taken the strained effect into account)

In the process of simulation, the conduction band-energy difference between $In_{0.1}Ga_{0.9}As$ and InAs is considered as 0.333 eV, similarly, 0.284 eV in valence band-energy difference. In the confined energy, the bound-states would be considered and extracted, once, excess the confined barrier, it is taken as non-bounded states outer QD.

3.2 Preparation for strained quantum dot infrared photodetector

The QDIP structure based on $In_{0.1}Ga_{0.9}As$/GaAs material system was grown on a (001) semi-insulting GaAs substrate by using the Riber Epineat solid-state molecular beam epitaxy (MBE) system. The samples of 30 periodic 3 ML InAs QD matrax on 1nm $In_{0.1}Ga_{0.9}As$ surfactant layer sandwiched between 5 nm $In_{0.1}Ga_{0.9}As$ and 30 nm GaAs structures with InAs QD region Si-doped to 1×10^{18} cm^{-3}. The detailed device structure and fabricated processing have been mentioned and similar to those previously proposed. Specially, it is to be noted that using the lower growth temperature of ~480 ℃ for nucleating InAs dots on GaAs substrate and As flux ≒ 80 mil (equivalent V/III ratio≦2) incorporated with 10 sec of growth interruption, the two dominant groups of InAs quantum dot matrix has been achieved and shown in Fig. 8.

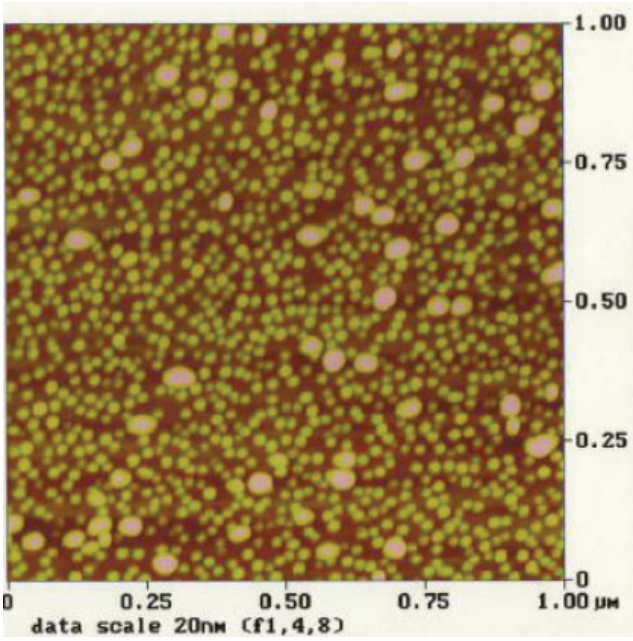


Fig. 8. The 2-diminsional 1μm×1μm AFM image with bimodal QD populations.

The 2-D morphology image is taken by DI-2000 atomic force microscopy (AFM) series. The larger and smaller-sized dot densities are 1.5×10^{10} and 2.5×10^{11} cm^{-2}, respectively. The QD lateral size and height of the two populations are 20 and 5 nm for smaller ones, 50 and 15 nm for larger ones. Due to introducing the mediate layer of 1 nm-thick $In_{0.1}Ga_{0.9}As$ under the growth temperature higher ~20℃ than that of QD and grown below the InAs QD nucleation, the surfactant layer results in high size-uniformity of QD bimodality and better epitaxial performance. Bias-dependent spectral response for the testing device measured using a Perkin-Elmer spectrum one fast Fourier-transform IR (FTIR) spectroscopy coupled with a cryostat under the normal incident scheme. The photoresponse signals were corrected by an absolute blackbody radiation source shown in Fig 9

Fig. 9. The physical FTIR and cryostat systems for temperature-dependent photoresponse measurement

4. Results of the quantum mechanical calculation with finite element method and the discussion as comparison with the experimental data

4.1 Iso-surface probability distribution for electron density of state within 3 periods of QD matrix layers calculated by finite element method

While 9 QDs are situated within 100 nmx100 nm of $In_{0.1}Ga_{0.9}As/GaAs$ unit-cell area, the equivalent dot density in each QD matrix is $9x10^{10}cm^{-2}$, as shown in Fig.6, which is comparable with that extracted from the experimental AFM image previously.

For three-dimensional FEM quantum mechanical calculations, the synthetic influences of strained interfaces of $GaAs/In_{0.1}Ga_{0.9}As/InAs$ in this work have been considered as the deformation potentials of the band offset (*i.e.*, called as band differences) to simplify the complexity of heterostructure modeling treated by the dual-cored pentium CPU. A theoretical model is first presented by Chris G. Van de Walle (Van de Walle and Chris G., 1989) to predict the band offset at pseudomorphic strained layer interfaces. The theory is based on the local-density-functional pseudopotential formalism and the model-solid approach of C. G. Van de Walle and R. M. Martin (Chris G. Van de Walle & R. M. Martin, 1986). The model can be most simply expressed in terms of an absolute energy level for each accounted for semiconductor system and deformation potentials that describe the effects of strain on the electronic bands. The model has been explored well that combinations of semiconductor materials and configurations of strains will lead to the desired electronic properties. Accordingly, using the previously mentioned and simply pseudopotential model attributed to the strained heterostructures, the differences of energy band-edges in InAs/InGaAs/GaAs material system are deformed from 0.842 ($\Delta E_{c\ GaAs\text{-}InAs}$) and 0.725 eV ($\Delta E_{c\ In0.1Ga0.9As\text{-}InAs}$) to shrink into 0.45 and 0.333 eV, respectively as seen in Fig. 7 and Table 1. However, it interesting to note that valence band offsets are enlarged in contrast with the differences of conduction band-edges. It may be explained that enlarging valence band

offsets attribute to the heavy and light hole mixing due to the strain inducement. These modified parameters for valence bands are shown in Fig. 7 and Table 2. The isotropic strains are assumed to simplify the calculation of confined states within active regions while using finite element approximation (FEA). Based on these parameters being taken account for isotropic strain effect on the interfaces between the lattice-mismatch materials, and modifying thickness from 0, 1, 2, 3, 4 to 15 nm of GaAs spacing layers, and under given 0.45 eV of the overall confined energy barrier in active region for conduction band, the eigenvalues versus quantum-confined numbers are calculated as shown in Fig. 10 and the deviation of spacing layer within 0 ~15 nm is indistinctive for extracting the eigenvalues

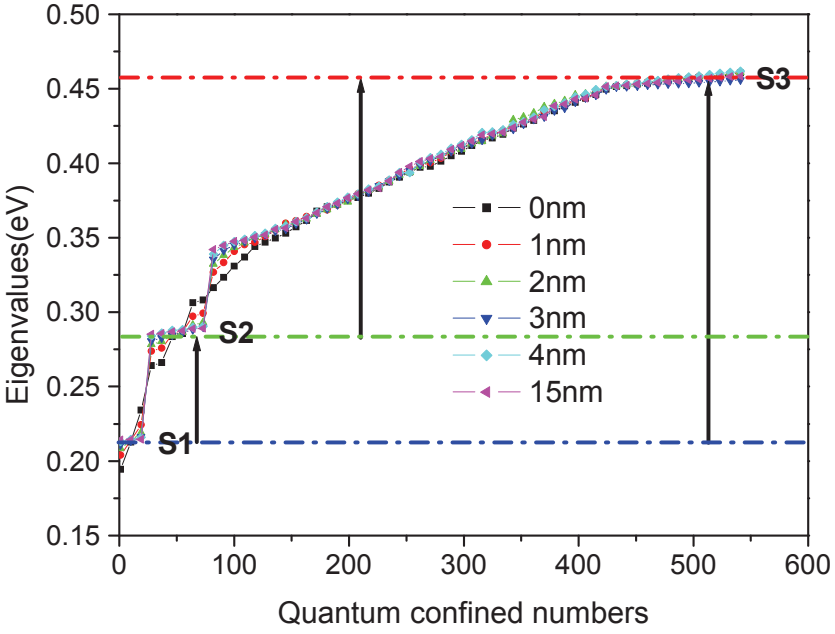


Fig. 10. The eigenvalues versus quantum-confined numbers with thickness from 0, 1, 2, 3, 4 to 15 nm of GaAs spacing layers under given 0.45 eV of the overall confined energy barrier in active region for conduction band

The electrical field of a light wave with frequency ω and wavevector $\vec{k}$ $(|k|) = \frac{n\omega}{c}$ can be expressed as

$$\vec{F}(\vec{r},t) = F\vec{\varepsilon}\cos(\omega t - \vec{k}\cdot\vec{r}) \tag{22}$$

with $\vec{\varepsilon}$ being a supposed linear polarization. In the Coulomb gauge (div $\vec{A}$ = 0), the

electrical field depends on the vector potential $\vec{A}$ as follows:

$$\vec{F} = -\frac{1}{c}\frac{\partial \vec{A}}{\partial t} \tag{23}$$

This leads to

$$\vec{A}(\vec{r},t) = -\frac{\vec{\varepsilon}\, cF}{2i\omega}\left[\exp(i(\omega t - \vec{k}\cdot\vec{r})) - \exp(-i(\omega t - \vec{k}\cdot\vec{r}))\right] \tag{24}$$

The one electron Hamiltonian of a heterostructure in presence of an electromagnetic field is in first approximation:

$$H = H_0 + \frac{e}{2m_0 c}(\vec{p}\cdot\vec{A} + \vec{A}\cdot\vec{p}) \tag{25}$$

with the electron momentum $\vec{p}$ and the electron charge *e*.
The probability of an optical stimulated transition is given by Fermi's golden rule:

$$\widetilde{P}_{if} = \frac{2\pi}{\hbar}\left|\langle f|V|i\rangle\right|^2 \cdot \delta(\varepsilon_f - \varepsilon_i - \hbar\omega) \tag{26}$$

where V is the perturbation term of the Hamiltonian $H = H_0 + V$. Under the electric dipole approximation, $\left(\exp\left(-i\vec{k}\cdot\vec{r}\right)\right) \approx 1$ which is valid for visible and infrared wavelengths, V is:

$$V = \frac{ieF}{2m_0\omega}\vec{\varepsilon}\cdot\vec{p} \tag{27}$$

If the quantum states *i* and *f* are partially occupied, the transition probability has to be weighted by the occupancy factor given by the Fermi distribution $f(\varepsilon)$:

$$P_{if} = \widetilde{P}_{if} f(\varepsilon_i)\left[1 - f(\varepsilon_f)\right] \tag{28}$$

where the Fermi distribution is the mean occupancy of the state ν:

$$f(\varepsilon_\nu) = \left[1 + \exp\left(\frac{1}{\kappa_B T}(\varepsilon_\nu - \mu)\right)\right]^{-1} \tag{29}$$

When taking into account the transitions $i \to f$ and $f \to i$, the linear absorption coefficient is

given by:

$$\alpha(\omega) = A\sum_{i,f}\frac{1}{m_0^*}\left|\vec{\varepsilon}\cdot\vec{p}_{if}\right|^2\partial\left(\varepsilon_f-\varepsilon_i-\hbar\omega\right)\left[f\left(\varepsilon_i\right)-f\left(\varepsilon_f\right)\right] \tag{30}$$

with $\vec{p}_{if}=\left\langle i\left|\vec{p}\right|f\right\rangle$, which contains the selection rules information, and $A=\frac{4\pi^2e^2}{ncm_0\omega\Omega}\cdot\Omega=SL$ is the irradiated volume of the sample.

The eigenvalues in the quantum-confined structures are derived by 3-D FEM formulations and equations (22) to (30), followed by Fermi-golden rule. It is obviously to note that more than one quantum-confined numbers in S_1, S_2 and S_3 regions associated with given similar eigenvalues, respectively. This phenomenon is known as degeneracy. The density of states of previously mentioned S regions are denser than the others. It means that higher transition probability has to be weighted by the occupancy factor given by the Fermi distribution $f(\varepsilon)$ which is followed by equations from (28) to (30). So, based on the theory of optical transition of quantum-confined structure, the transitions of $S_1\rightarrow S_2$, $S_2\rightarrow S_3$, and $S_1\rightarrow S_3$ are the most possibility for intraband transitions. The peak-wavelength positions from these 3 selection routes of optical transitions with 0 to 15 nm-thick GaAs spacing layers can be calculated and are listed in Table 3. Furthermore, according to Eq. (12), (13), and (14) associated with simplified calculations partly which are quoted from absorption spectrum of quantum well model, the absorption efficiencies versus wavelength from $S_1\rightarrow S_2$, $S_2\rightarrow S_3$, $S_1\rightarrow S_3$, and summation of above 3 routes are shown in Fig. 11. The results of absorption spectrum are well agreed with the data simulated by 3-dimensional quantum-confined FEM shown in Fig. 10 and Table 3.

Intraband Transitions \ Spacing thickness (nm)		0	1	2	3	4	15
S1-S3	E_{s1}-E_{s3} (eV)	0.2421505	0.242467	0.242726	0.239903	0.242553	0.240944
	λ_T (μm)	5.1207815	5.114108	5.108633	5.168763	5.112292	5.146414
S2-S3	E_{s2}-E_{s3} (eV)	0.1702775	0.170167	0.170794	0.168091	0.169357	0.168045
	λ_T (μm)	7.2822288	7.286958	7.2602	7.376972	7.321804	7.37897
S1-S2	E_{s1}-E_{s2} (eV)	0.071873	0.0723	0.071932	0.071812	0.073196	0.072899
	λ_T (μm)	17.252654	17.15088	17.23846	17.26731	16.94093	17.00976

Table 3. The peak-wavelength positions from these 3 selection routes of possible optical transitions with 0, 1, 2, 3, 4 and 15 nm-thick GaAs spacing layers

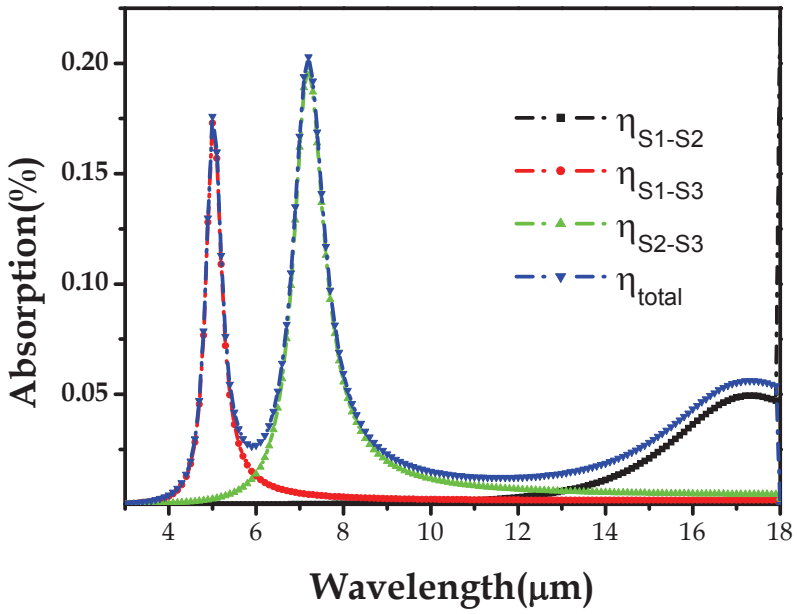


Fig. 11. Calculated absorption efficiencies versus wavelength for 3 periodic InAs QD matrix embedded in $In_{0.1}Ga_{0.9}As$/GaAs layers

The isosurface wave-functional distributions of electrons confined from ground to top-lying excited states for free GaAs spacing layer *i.e.,* electrons confined between InAs and GaAs conduction band edge minimum are shown as Fig. 12 (a) in 3-D and (b) in 2-D y-z in-plane view, respectively. Similarly, the isosurface wave-functional distributions of electrons confined from ground to top-lying excited states for 3 and 15 nm GaAs spacing layer *i.e.,* electrons confined between InAs and GaAs conduction band edge minimum are shown as Fig. 13 and 14 (a) in 3-D and (b) in 2-D y-z in-plane view, respectively. It is very interesting to observe that the wave-functions are lateral coupling easily in more top-lying excited states due to barrier lowering for higher confined states. While the thickness of GaAs spacing layer is larger than 3 nm, there are no vertical wave-couplings within the interlayers attributed to the barrier blockade of thicker GaAs spacing layer.

The similar structure based on Fig. 5 except for more than numbers of active layer are used to perform a standard semiconductor process as QDIP for spectrum measurement as mentioned section 3.2. Normal-incident photoresponsivity versus wavelength in 80K are measured and achieved from Eq. (13), (14), (28) to (30) and used the physical model from H. C. Liu are included (H. C. Liu, 1993), the simulated results can successfully interpret the measured photoresponse spectrum as shown in Fig. 15. The broadening dual-band peak wavelengths of 5.32 and 7.38 μ m for the QDIP mainly originate from the interlevel transitions due to contributions of the degeneracy of engenlevels confined in the quantum heterostructure. The other partly attributes to QD sized distributions.

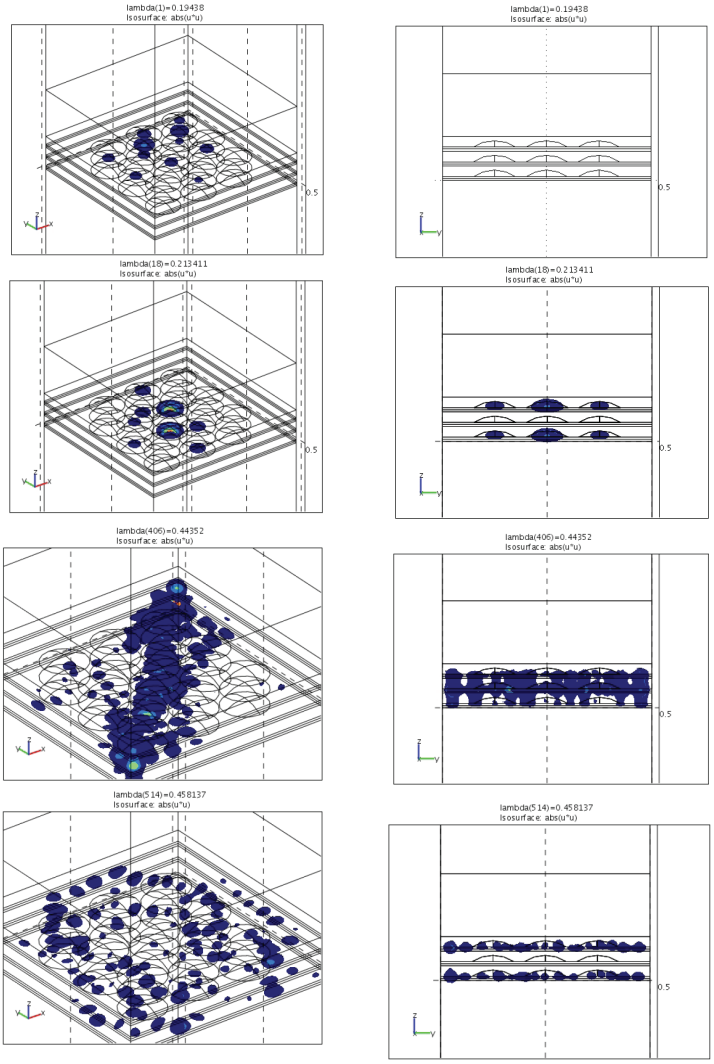


Fig. 12. Isosurface wave-functional distributions of electrons confined from ground to top-lying excited states for free GaAs spacing layer *i.e.,* electrons confined between InAs and GaAs conduction band edge minimum (a) in 3-D and (b) in 2-D y-z in-plane view

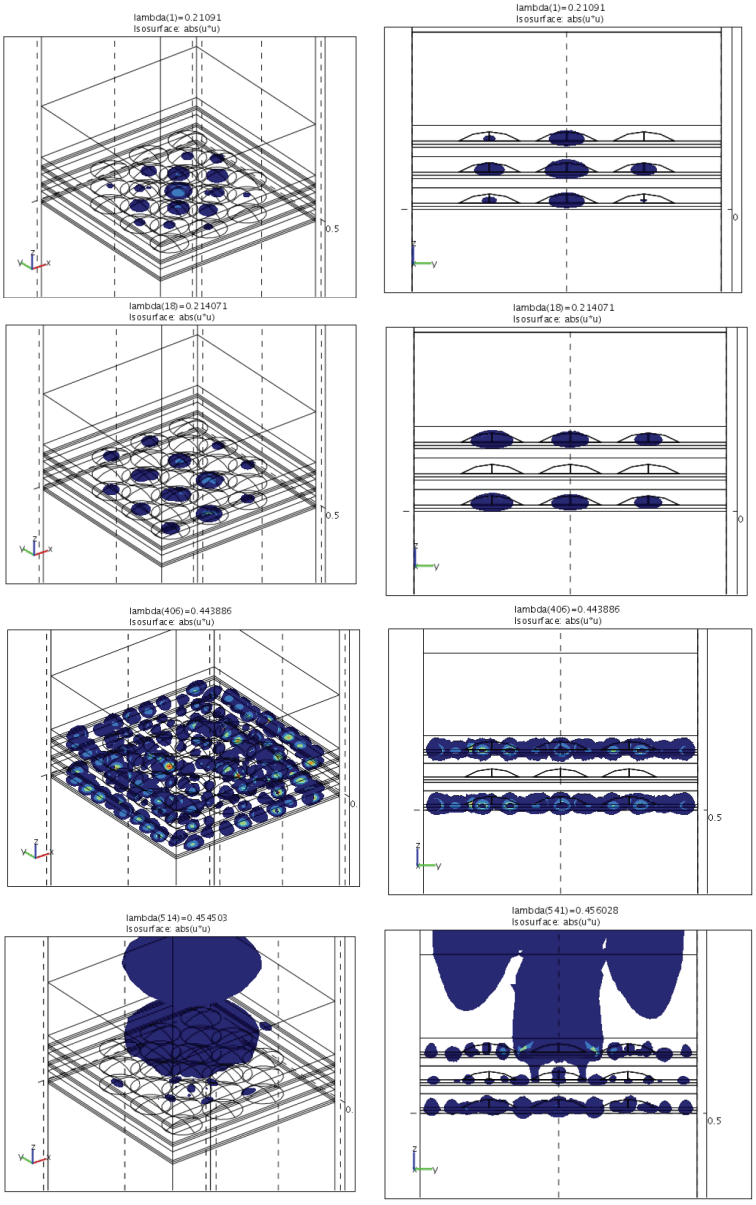


Fig. 13. Isosurface wave-functional distributions of electrons confined from ground to top-lying excited states for 3 nm-thick GaAs spacing layer *i.e.,* electrons confined between InAs and GaAs conduction band edge minimum (a) in 3-D and (b) in 2-D y-z in-plane view

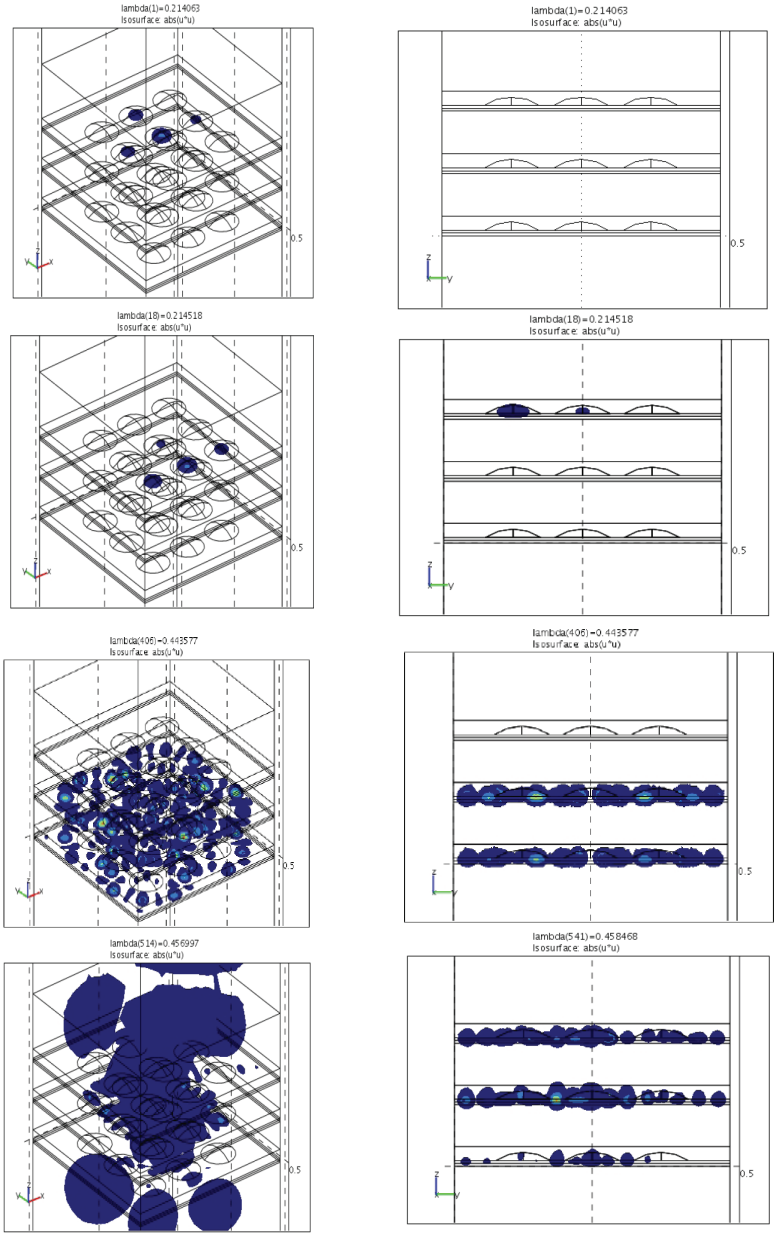


Fig. 14. Isosurface wave-functional distributions of electrons confined from ground to top-lying excited states for 15 nm-thick GaAs spacing layer *i.e.,* electrons confined between InAs and GaAs conduction band edge minimum (a) in 3-D and (b) in 2-D y-z in-plane view

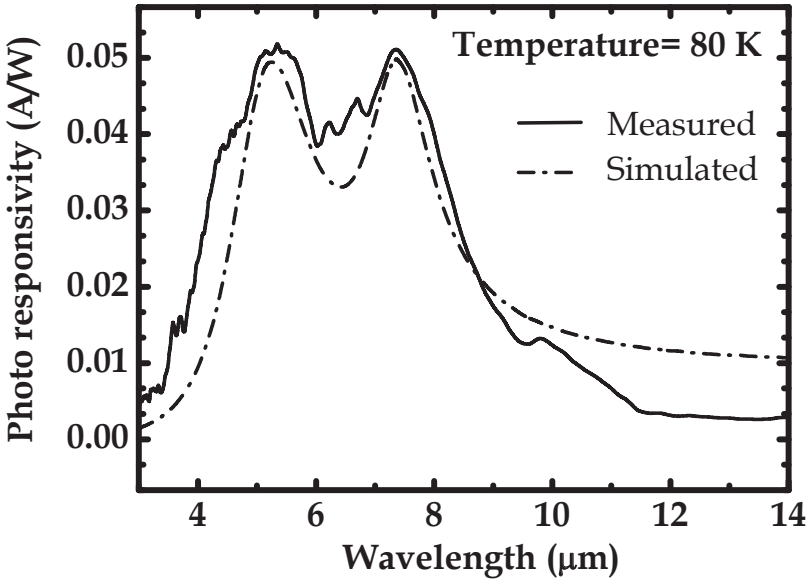


Fig. 15. The simulated results and measured photoresponse spectra

5. Conclusions

Intraband absorption coefficient with InAs/$In_{0.1}Ga_{0.9}As$ QDs covered with GaAs quantum is wells studied. The numerical values of photoinduced-intraband absorption spectra in multiple QDs layers are derived to achieve the three degeneracy regions by 3-dimensional quantum-confined FEM, which are well consistent with analytic absorption coefficient calculated from total intraband absorption coefficient $\alpha_{total}(\omega) = \sum_{i,f} \frac{\pi E_{fi} e^2 (n_i - n_f)}{2\varepsilon_0 c\tilde{n} m_0 \omega \Omega} \cdot f \cdot g(E_{fi} - \hbar\omega)$ which is inferred from the dipole $\mu_{if} = e\left\langle \Psi_i \middle| \vec{\varepsilon} \cdot \vec{r} \middle| \Psi_f \right\rangle$, and optical oscillator strength $f = \frac{2m_0 E_{21}}{e^2 \hbar^2} \mu_{12}^2$. Experimental results on photoresponse spectra combined with inter-leveling electron transitions agree satisfactorily with theory for the first time.

6. References

A. J. Williamson, L. W. Wang, and Alex Zunger (2000) "Theoretical interpretation of the experimental electronic structure of lens-shaped self-assembled InAs/GaAs quantum dots" *Physical Review B*, Vol. 62, No. 19, March 2000, pp.12963-12977

Alexander Weber (1998). "Intraband spectroscopy of semiconductor quantum dots", *Dissertation, Dr. of Phylosophy,* Julius-Maximilians-Universität, Wűrzburg

Chris G. Van de Walle and Richard M. Martin (1989) "Theoretical calculations of semiconductor heterojunction discontinuities"*J. Vac. Sci. Technol. B* .Vol. 4, No. 4, July 1986, pp. 1055-1059

C. Pryor, J. Kim, L. W. Wang, A. J. Williamson, and A. Zunger (1998) "Comparison of two methods for describing the strain profiles in quantum dots"*Journal of Applied Physics,* Vol. 83, March 1998, pp.2548-2550

Daniel Fritsch, Heidemarie Schmidt, and Marius Grundmann (2003)."Band-structure pseudopotential calculation of zinc-blende and wurtzite AlN, GaN, and InN", *Physical Review B*, Vol. 67, No. 4, June 2003, pp. 235205-1-13

Gerhard Klimeck, Marek Korkusinski, Haiying Xul, Seungwon Lee, Sebastien Goasguen, and Faisal Saied (2002). "Atomistic simulations of long-range strain and spatial asymmetry effects in multimillion-atom single and double quantum dot nanostructures", *Computer Modeling in Engineering and Science* , No. 3, February 2002, pp. 601

G R Liu and S S Quek Jerry (2002). "A finite element study of the stress and strain fields of InAs quantum dots embedded in GaAs", *Semiconductor Science and Technology,* Vol. 17, May 2002, pp. 630-643

H. C. Liu (1992). "Dependence of absorption spectrum and responsivity on the upper state position in quantum well intersubband photodetectors" *Journal of Applied Physics*, Vol. 73, No. 6, March 1992, pp. 3062-3067

H. T. Johnson, L. B. Freund, C. D. Akyüz, and A. Zaslavsky (1998). "Finite element analysis of strain effects on electronic and transport properties in quantum dots and wires", *Journal of Applied Physics*, Vol. 84, No. 7, October 1998, pp. 3714-3725

K. Leifer, E. Pelucchi, S. Watanabe, F. Michelini, B. Dwir, and E. Kapon (2007). "Narrow (≈ 4 meV) inhomogeneous broadening and its correlation with confinement potential of pyramidal quantum dot arrays", *Applied Physics Letters*, Vol. 91, No. 8, August 2007, pp. 81106-81108

M. A. Cusack, P. R. Briddon, and M. Jaros (1997). "Absorption spectra and optical transitions in InAs/GaAs self-assembled quantum dots", *Physical Review B*, Vol. 56, No. 7, August 1997, pp. 4047-4050

M. Grundmann, O. Stier, and D. Bimberg (1995)."InAs/GaAs pyramidal quantum dot: strain distribution, optical phonon, and electronic structure ", *Physical Review B*, Vol. 52, No. 16, October 1995, pp. 11969-11981

O. Stier, M. Grundmann, and D. Bimberg (1999)."Electronic and optical properties of strained quantum dots modeled by 8-band k·p theory", *Physical Review B*, Vol. 59, No. 8, February 1999, pp. 2688-2701

Shu-Shen Li and Jian-Bai Xia (1997) "Intraband optical absorption in semiconductor coupled quantum dots" *Physical Review B*, Vol. 55, No. 23, January 1997, pp.15434-15437

S. L. Chuang and C. S. Chang (1996) "k·p method for strained wurtzite semiconductors", *Physical Review B*, Vol. 54, No. 4, July 1996, pp. 2491-2504

Tzu-Huan Huang, Shiang-Feng Tang, Tzu-Chiang Chen, Fu-Fa Lu and Cheng-Der Chiang (2007) "Calculations of bandstructures on the lens and pyramid-shaped InAs quantum dot for confirming the photoluminescence and photoresponse" *Proceedings of SPIE,* Vol. 6479, February 2007, San Jose, CA, USA

T. Benabbas Y. Androussi, and A. Lefebvre (1999) "A finite-element study of strain fields in vertically aligned InAs islands in GaAs"*Journal of Applied Physics,* Vol. 86, May 1999, pp.1945-1948

van de Walle and Chris G. (1989) "Band lineups and deformation potentials in the model-solid theory" *Physical Review B*, Vol. 39, No. 3, January 1989, pp. 1871-1883

14

Intersublevel Relaxation Properties of Self-Assembled InAs/GaAs Quantum Dot Heterostructures

[1]Jiunn-Chyi Lee and [2]Ya-Fen Wu
[1]*Electrical Engineering, Technology and Science Institute of Northern Taiwan*
[2]*Electronic Engineering, Ming Chi University of Technology*
Taiwan

1. Introduction

The requirement for high performance optoelectronic devices has spurred much experimental effort directed toward understanding and exploiting the electronic and optical properties of quantum dots (QDs). The relaxation dynamics in the zero-dimensional QD systems is expected to differ qualitatively from higher-dimensional systems, since the density of states is a series of δ-functions. The limited number of states available for carriers impairs carrier relaxation toward the ground state (phonon bottleneck effect) (Benisty et al., 1991; Benisty, 1995; Hai et al., 2006). In addition, the finite degeneracy of each QD state leads already to state filling effects when few carriers populate the lowest dot states. Both effects possibly result in intersublevel relaxation rates that are comparable to interband recombination rates and have been used to explain observed photoluminescence (PL) from excited states of QDs (Bissiri et al., 2001; Smith et al., 2001).
The temperature dependence of PL emissions has been the subject of extensive studies for clarifying the mechanism of PL quenching processes in a randomly distributed dot structure (Bafna et al., 2006; Duarte et al., 2003; Polimeni et al., 1999). The PL spectra of QDs typically show peculiar temperature dependencies. A large temperature induced peak energy decrease, which is eventually sigmoidal, and a reduction of the PL full width at half maximum (FWHM) in mid-temperature range, have been reported (Dawson et al., 2005; Polimeni et al., 1999). The phenomenon is commonly attributed to effectively redistributed carriers in QDs through the channel of the wetting layer based on a model of the temperature driven carrier dynamics which takes into account the QD size distribution, random population, and carrier capture relaxation and retrapping (Nee et al., 2005; Nee et al., 2006). The physics of carrier relaxing between intersublevels in various QD systems has been extensively studied. However, the electron-phonon scattering effect on QD system is neglected and only considered in the high-temperature range to explain the increase of FWHM (Dawson et al., 2005; Nee et al., 2005; Nee et al., 2006), and the effect of dot size, density, and uniformity on this mechanism is still not fully understood (Dawson et al., 2005; Duarte et al., 2003).

In this chapter, we studied the phonon-assisted transferring of carriers in InAs QD system via an analysis of PL data in the temperature range from 15 K to 280 K. Intersublevel relaxation properties and thermally-induced activation of excitons in QD system are simulated using a rate-equation model based on carrier relaxation and thermal emission in the quantum dot system. The dot-size distribution, thermal escaping and retrapping, and electron-phonon scattering, are all considered in the model. Correlation between carrier redistribution and electron-phonon scattering effects is quantitatively discussed to explain the different temperature-dependent behaviors of the PL spectra measured from samples with different dot size distribution. Moreover, the phonon-bottleneck effect on temperature dependent PL spectra is also discussed to illustrate the significance of phonon-assisted effect on QD system. According to the simulation results, intersublevel relaxation lifetimes of QD samples are estimated under different temperatures and the carrier transferring mechanisms in the QD system are discussed in detail. The theoretical analysis confirms that the thermal redistribution of carriers and the electron phonon scattering affect the temperature dependent PL spectra simultaneously.

2. Sample Preparation

An easy way to fabricate zero-dimensional InAs QDs is to grow the InAs on GaAs in the S-K mode (Sanguinetti et al, 2002; Schmidt et al., 1996). In the S-K transformation, growth is initially two-dimensional, until the film reaches a strain dependent critical thickness. Above the critical deposition thickness of InAs on GaAs substrate, due to the 7% lattice mismatch between GaAs and InAs, the two-dimensional growth changes into a three-dimensional one. Coherent InAs islands with lateral extensions of 10-20 nm are spontaneously formed on top of the two-dimensional layer, called the wetting layer. It was traditionally believed that islands formed in S-K growth are dislocated. However, the experiment on InAs/GaAs (001) has demonstrated the formation of three-dimensional coherently strained islands.

The self-assembled InAs QD samples used in the work were created by using a metal-organic chemical vapor epitaxy system (MOCVD) system. The substrates were (100) 2°-tilted toward (111)A Si-doped GaAs. The heterostructures included a 400 nm Si-doped GaAs buffer layer, an InAs QD active region of 3 monolayers (MLs) and a 100 nm undoped GaAs capping layer. The growth rate was 0.1 MLs and the V/III ratio during the growth of InAs layer was 6.36 for samples A, B and 3 for sample C. The growth interruption (GI) introduced during dot formation for samples A, B, and C were set to 6 s, 15 s and 15 s, respectively. In order to investigate the average dot size distribution and shape, images of these samples were taken by high-resolution transmission electron microscopy (HRTEM) operating at 200 keV. PL measurements were carried out under the excitation of a continuous-wave He-Cd laser emitting at 325 nm, with the incident power intensity being 20 mW. The samples were mounted in a closed cycle He cryostat, which allowed measurements in a temperature (T) range from 15 K to 280 K. The luminescence was dispersed in a 0.5 meter monochromator, and detected with a Ge photodiode using a standard lock-in technique.

Figure 1 shows the plan-view TEM images for samples A, B and C. The quantitative data on size distribution of the InAs QD samples have been obtained from the TEM images, the average dot density of samples A, B and C are 2.4×10^{10} cm^{-2}, 1.2×10^{10} cm^{-2} and 1.4×10^{10} cm^{-2}, respectively, and the average dot diameters of the three samples are 16 nm, 19 nm and 20 nm. Generally, the application of GI time results in the formation of larger sized QDs with a

regular size distribution (Tarasov et al., 2000), as can be seen from Fig. 1(a) and Fig. 1(b). Besides, decreasing the V/III ratio during growth can increase the indium adatom surface diffusivity in the wetting layer and hence increasing the two-dimensional island size in the wetting layer. A layer composed of larger two-dimensional islands will have a more uniform strain distribution and lead to a more uniform island distribution on top of the wetting layer (Solomon et al., 1995). The highest uniformity was exhibited for sample C as can be seen in Fig. 1(c).

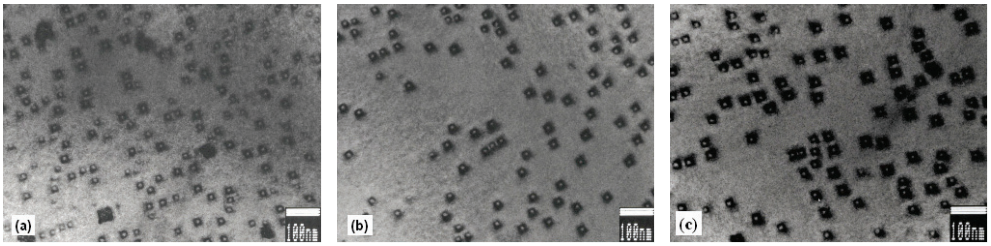


Fig. 1. Plan-view TEM images of the InAs quantum dots of (a) sample A, (b) sample B, and (c) sample C

3. Results and Discussions

3.1 Photoluminescence Characterization

The measured PL spectra at temperature T=15 K for the samples are shown in Fig. 2. All of these spectra exhibit a pronounced double-like feature and can be decomposed into two Gaussian peaks; we attribute these two main spectral features of the QDs to the ground state and excited state emissions. Sample A possesses the largest ground state and excited state transmission energies, i.e., 1.05 eV and 1.11 eV; and the values are 1.01 eV, 1.09 eV and 1.01 eV, 1.08 eV for sample B and sample C, respectively. Considering the quantum-size effect on the peak energies, we believe that the excitons localized in smaller dots will contribute to higher peak energies (Cheng et al., 1998). As a result, the highest peak energies of sample A (GI=6s) is attributed to the smallest size of the QDs in the three samples. Similarly, the peak energies for sample B and sample C are almost the same because their dot sizes are similar.

One remarkable feature in Fig. 2 is the obvious difference of the excited state peak intensities among the samples. The strongest excited state peak intensity of sample A reveals that more carriers exists in this state, and the much weaker excited state emissions in PL intensity of sample C suggests that the carriers relax rapidly into the ground state. In other words, it has shorter relaxation lifetimes than those of sample A and sample B. It indicates for sample C a restricted phonon bottleneck effect (Benisty et al., 1991; Bockelmann et al. 1990). This can be understood in terms of an improved confinement of InAs excitons and a lower defect density in sample C due to having best uniformity among the three samples.

The values of FWHM of ground state and excited state emissions are 27.1 meV and 88.3 meV for sample A, 26.8 meV and 79.6 meV for sample B, and for sample C they are 23.3 meV and 55.27 meV, respectively. The PL linewidth is mainly determined by the inhomogeneous broadening of InAs islands resulted from size fluctuation of the dot size at low temperature (Xu et al.,1996), the measured data for sample C are consistent with its better size uniformity.

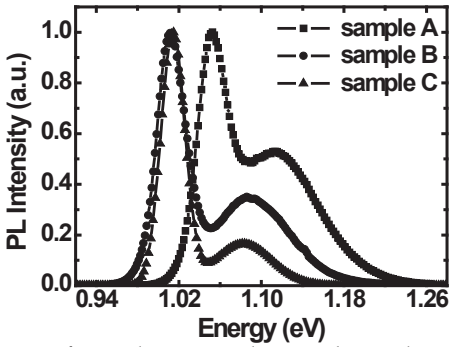


Fig. 2. Normalized PL spectra of sample A, sample B, and sample C recorded at T=15 K. The excitation energy is 20 mW

The two-dimensional contour plots in Fig. 3 display the measured temperature dependent PL intensities. The distributions of emission energy from the QD systems are clearly seen from the figures. Sample A has the widest emission band, luminescence from the excited state is apparent. The narrowest energy spreading is the contour shown for sample C. The PL intensity of excited state is too small to be observable and the PL spectra are concentrated in a narrow linewidth. Since the observation of PL from excited states transition at low

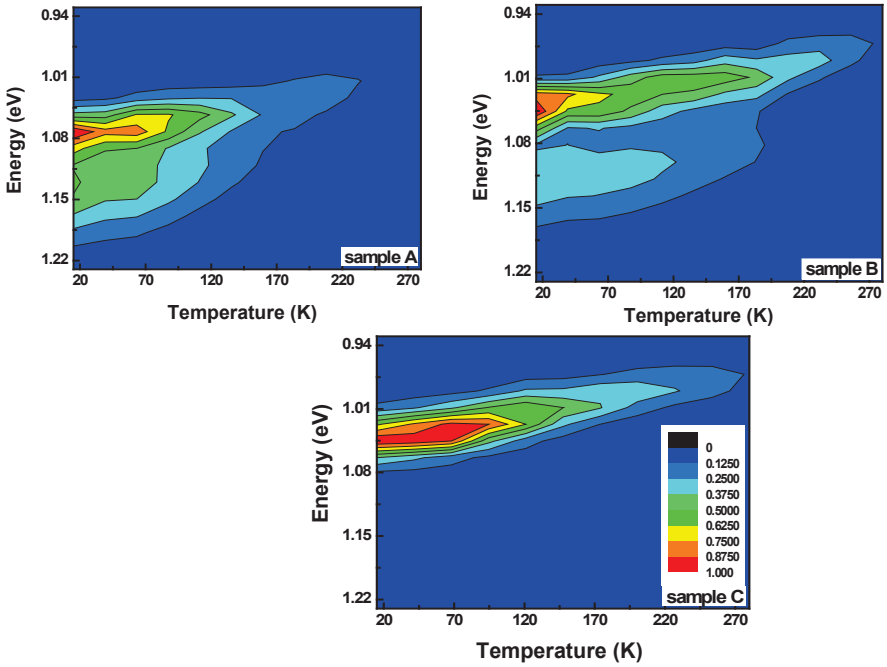


Fig. 3. Two-dimensional contour plots of the PL intensities for sample A, sample B, and sample C, measured in the temperature range from 15 to 280 K

excitation density is explained by the phonon bottleneck effect in the QD system, we attribute the inconspicuous excited state emission of sample C to the partially relaxed phonon bottleneck.

Figure 4(a) displays the temperature dependent FWHMs of PL spectra of the samples, both the ground state and the excited state are included. Observing the FWHMs of sample A and sample B, they stay constant up to 75 K and 100 K. As the temperature further increases, anomalous reduction appeared within the temperature range from 100 K to 200 K. The FWHMs decrease and the minimal FWHMs of excited state are found to be around 69 meV at 200 K for both samples. When the temperature is higher than 200 K, the PL linewidths start to increase with temperature. At low temperature, carriers are captured randomly into the QDs. With increasing temperature, carriers are thermally activated outside the dots with shallow energy minima into the wetting layer then retrapped into another dot. Carrier hopping among dots favors a drift of carriers towards the dots with lower energy emissions and leading to the decrease of FWHMs. As temperature exceeds 200 K, the FWHMs increase with temperature because the electron-phonon scattering becomes important. Figure 4(b) shows the PL excited state peak energy with increasing temperature, and the corresponding values of InAs band gap using Varshni law with the InAs parameters are also shown. As can be seen in the figure, the redshift of emission peaks for sample A and B are faster than that of the InAs bulk band gap at T=100-200 K, coincided with the carrier hopping mechanism described above.

Significantly different temperature dependent FWHMs are observed for sample C. The broadening of the PL spectra exhibits no reduction as the temperature increases, but the peak energy shifts with a slight sigmoid dependence on temperature. Thanks to the lowest PL intensity of excited state, fewer carriers exist in the state, and the thermal redistribution of carriers via wetting layer is indistinct. The slightly quick redshift of peak energy is consistent with the weak redistribution effect, whereas the increase of linewidth with temperature implies that the electron-phonon scattering is dominant in the PL spectra. Therefore, to analyze the carriers transferring mechanisms, we investigate a model for carrier dynamics in QD system under optical excitation which includes the thermal redistribution effect and the electron-phonon scattering effect.

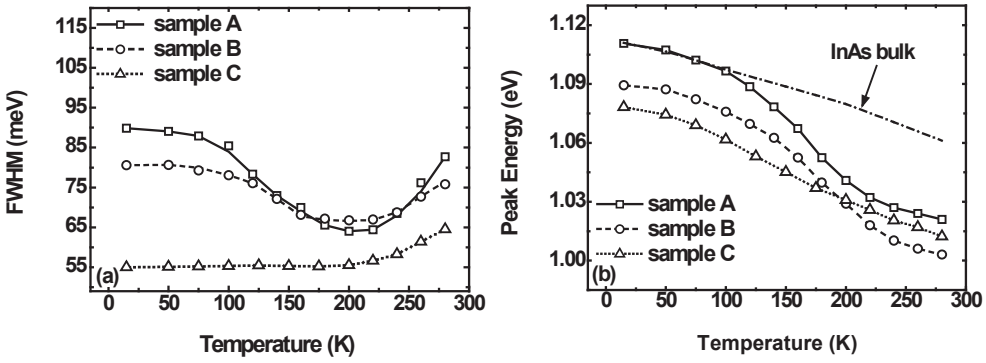


Fig. 4. Experimental values of the temperature dependent (a) FWHM and (b) peak energy of the excited state of sample A, sample B, and sample C

3.2 Theoretical Model

The possible optical transitions in a single QD consist of a series of δ function lines, whose energy positions depend on the particular three-dimensionality confined levels. In a real QD ensemble each individual dots are slightly different in size, shape, strain state, etc. The main impact of size fluctuations is a variation in the energy position of electronic levels, and subsequently to an inhomogeneous broadening of the ensemble properties. It is reasonable to assume a Gaussian distribution for it (Chang et al., 1999).

To analyze the carrier dynamics of the QD system, we develop a theoretical model that takes into account the QDs size distribution, the state filling effect, and all of the important carrier transport processes, including the carrier capture and relaxation, thermal emission and retrapping, and radiative and nonradiative recombination. Referring to the model of the QD system described schematically in Fig. 5, four discrete levels of electron (labeled as i, i=1-4) are considered in the system, namely, the ground state (E_1), the excited state (E_2), the wetting layer (E_3), and the GaAs barrier (E_4). Since the process of quantum dot population is intrinsically random, the density of states for both ground state [$n_1(E)$] and excited state [$n_2(E)$] are assumed to be proportional to their Gaussian distributions, with parameters chosen to match the peak energies and linewidths of the lowest temperature PL spectra (Lee et al., 1997; Yang et al., 1997) and taking spin into consideration, then

$$\int n_i(E)dE = \int \left[n_{if}(E) + n_{ie}(E) \right] dE \propto 2 \times i \times n_d \,, \tag{1}$$

where i=1, 2; n_{if} and n_{ie} are the filled and empty energy states of the i-th level, respectively, and n_d is the dot density of the sample.

The carrier dynamics taken into account in this model are described as follows. First, the coupling among those four carrier reservoirs is treated as a relaxation ladder process from each energy level to its lower level neighbor. Carriers are injected from the GaAs barrier into the wetting layer at rate g, from where they are captured into the excited state of QDs within a capture time τ_{32}. Further on, carriers in the excited state relax to the ground state in a time of τ_{21} or radiatively recombine. The relaxation lifetime of one electron in the i-th level ($\tau_{i,i\text{-}1}$) is proportional to the filling ratio of the (i−1)-th level (f_{i-1}), and expressed as (Mukai et al., 1996)

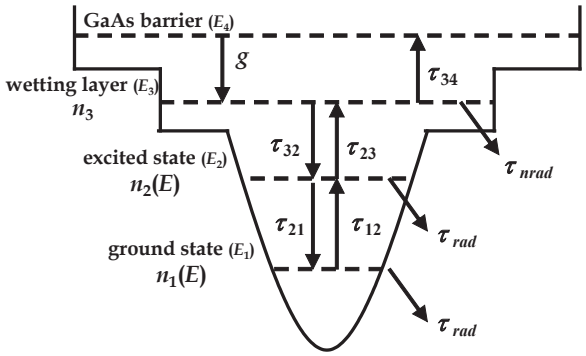


Fig. 5. Schematic representation of the processes taken into account in the rate equation model

$$\tau_{i,i-1} = \tau_{i,i-1,0} \times \left(1 - f_{i-1}\right)^{-1},\ i = 2, 3, \tag{2}$$

where $\tau_{i,i\text{-}1,0}$ is the intrinsic relaxation lifetime.
Secondly, thermal emission of the carriers toward an adjacent higher energy level arises when the temperature is sufficiently high. The coefficients corresponding to emitting from the ground state to the excited state, the excited state to the wetting layer, and the wetting layer to the GaAs barrier are given by τ_{12}, τ_{23}, and τ_{34}, respectively. In QD systems, the thermal emission and retrapping of carriers in the excited state via the wetting layer is a typical explanation for the unusual decrease of PL linewidth in the mid-temperature range (Lobo et al., 1999; Polimeni et al., 1999; Giorgi et al., 2001). We express the thermal emission time of the i-th level as $\tau_{i,i+1}$, and

$$\tau_{i,i+1} = \tau_{i,i+1,0} \times \exp\left[\left(E_{i+1} - E_i\right)/kT\right],\ i = 1, 2, 3, \tag{3}$$

where $\tau_{i,i+1,0}$ is the intrinsic thermal emission lifetime of level i. Since the peak interval between the ground state and the excited state is much larger than the value of kT, the thermal emission from ground state to excited state is neglected in our model.
The third type of carrier dynamics considered in this system is the radiative recombination. We have neglected any recombination from the second excited state of the dots, since no PL is observed at energies possible for the second excited state, and assumed that only two discrete electron levels exist inside a quantum dot, i.e., the ground state and the first excited state. The radiative recombination lifetime τ_{rad} is assumed to be the same for both of the states in all of the QDs and is constant with respect to T.
The system under steady-state conditions is then characterized by the following equations: (Lee et al., 2008; Sanguinetti et al., 1999; Wu et al., 2008)

$$\frac{dn_3}{dt} = g - \int \frac{n_3}{\tau_{32}} \times \frac{n_{2e}(E)}{n_2(E)} dE + \int \frac{n_{2f}(E)}{\tau_{23}} dE - \frac{n_3}{\tau_{34}} - \frac{n_3}{\tau_{nrad}} = 0, \tag{4}$$

$$\frac{dn_{2f}(E)}{dt} = \frac{n_3}{\tau_{32}} \times \frac{n_{2e}(E)}{n_2(E)} - \frac{n_{2f}(E)}{\tau_{21}} \times \frac{n_{1e}(E)}{n_1(E)} + \frac{n_{1f}(E)}{\tau_{12}} \times \frac{n_{2e}(E)}{n_2(E)} - \frac{n_{2f}(E)}{\tau_{23}} - \frac{n_{2f}(E)}{\tau_{rad}} = 0, \tag{5}$$

$$\frac{dn_{1f}(E)}{dt} = \frac{n_{2f}(E)}{\tau_{21}} \times \frac{n_{1e}(E)}{n_1(E)} - \frac{n_{1f}(E)}{\tau_{12}} \times \frac{n_{2e}(E)}{n_2(E)} - \frac{n_{1f}(E)}{\tau_{rad}} = 0. \tag{6}$$

The last term in (4) is the nonradiative loss of excitons in wetting layer, and τ_{nrad} is the nonradiative recombination lifetime. The state filling effect is essentially significant in the QD system because of the reduced density of states and should be taken into account. Prior to the description of the simulation process, we must discuss the parameters used in the model. To obtain τ_{rad} used in our model, we estimate the intrinsic exciton lifetime in QDs (τ_{QD}) at low temperature in terms of the exciton lifetime in a corresponding quantum well (τ_{QW}) by (Malik et al., 2001)

$$\tau_{QD} = \frac{3}{2}\tau_{QW}\left(\frac{\eta}{k_{ex}}\right)^2 . \tag{7}$$

Here $k_{ex}=2\pi n/\lambda_{PL}$ is the reciprocal wavelength of the emitted light in the quantum dot material, with the refractive index of InAs, and η is a measure of the lateral dot size. By using values of η=(1/15) nm^{-1}, n=3.6, λ_{PL}=1181 nm, and an exciton lifetime τ_{WL} of 25 ps, the radiative recombination lifetime τ_{rad} is calculated to be approximately 500 ps and assumed to be independent from temperatures.
Rewrite (4), (5) and (6) at T=15 K where the thermal emission can be neglected:

$$\frac{dn_3}{dt} = g - \int \frac{n_3}{\tau_{32}} \times \frac{n_{2e}(E)}{n_2(E)} dE - \frac{n_3}{\tau_{nrad}} = 0 , \tag{8}$$

$$\frac{dn_{2f}(E)}{dt} = \frac{n_3}{\tau_{32}} \times \frac{n_{2e}(E)}{n_2(E)} - \frac{n_{2f}(E)}{\tau_{21}} \times \frac{n_{1e}(E)}{n_1(E)} - \frac{n_{2f}(E)}{\tau_{rad}} = 0 , \tag{9}$$

$$\frac{dn_{1f}(E)}{dt} = \frac{n_{2f}(E)}{\tau_{21}} \times \frac{n_{1e}(E)}{n_1(E)} - \frac{n_{1f}(E)}{\tau_{rad}} = 0 . \tag{10}$$

The detected PL peak intensities of the ground and excited states are proportional to the values of n_1 and n_2, respectively, thus τ_{21} is determined by using (10). Combining (9) and (10) and using a value of 30 ps for the carrier capturing lifetime by QDs (Sanguinetti et al., 1999) yields the value of τ_{32}. Following a similar procedure, we get the value of g. Use these calculated parameters, the rate-equation set (4)-(6) is solved numerically by fitting the temperature dependent integrated PL intensities of the ground state and the excited state. Once the carrier distribution functions $n_{1f}(E)$ and $n_{2f}(E)$ are determined, the PL spectra of ground state (PL_1) and excited state (PL_2) can be expressed as

$$PL_1(E) = \beta \times n_{1f}(E)/\tau_{rad} , \tag{11}$$

$$PL_2(E) = \beta \times n_{2f}(E)/\tau_{rad} , \tag{12}$$

where β is a normalizing factor. From (11) and (12), the measured temperature dependent PL spectra of the ground state and the excited state are reproduced. The parameters used in our calculation are listed in Table 1.

Sample	g (s^{-1})	τ_{32} (ps)	τ_{21} (ps)	τ_{23} (ps)	τ_{34} (ps)
A	1×10^{21}	14	68	0.17	5.5×10^{-6}
B			33		
C			10		

Table 1. Parameters used in the calculations of PL peak intensities and relaxation lifetimes of the samples

3.3 Electron-Phonon Scattering Effect

The total linewidth of PL emission in QD system can be decomposed into two components: inhomogeneous and homogeneous. The nature of these two mechanisms is totally different. Inhomogeneous broadening in the QD system arises from small fluctuations in the QDs confining size, the alloy composition variations, and the shifts due to strain-field effects (Seebeck et al., 2005; Zhao et al., 2002). The major contribution to the inhomogeneous broadening comes from the size variation due to the large confining potentials and the small volumes. We can express the inhomogeneous broadening lineshape as a Gaussian function

$$G(E) = G_0 \exp\left[\frac{-(E-E_0)^2}{2\sigma^2}\right], \tag{13}$$

where G_0 and E_0 are the amplitude and peak energy position, respectively, and σ is the standard deviation of the distribution.

Homogeneous broadening is mainly due to the exciton-phonon interaction. Both acoustic and optical phonons are involved in the process (Ortner et al, 2004). The phonon contribution of the linewidth is proportional to phonon population density. In acoustic phonon case, such a density increases linearly with the temperature. On the other hand, optical phonons have a relatively fixed frequency. The number of phonons thermally excited follows Bose-Einstein statistics. The expression of the total homogeneous linewidth can be written as following (Christen & Bimberg, 1990)

$$\Gamma_{homo} = \gamma_{AC} T + \frac{\gamma_{LO}}{\left[\exp\left(\hbar\omega_{LO}/k_B T\right) - 1\right]}. \tag{14}$$

The first term represents the acoustic phonon contributions with proportionality constant γ_{AC} and the second term represents the optical phonon contributions. γ_{LO} is the longitudinal optical (LO) phonon broadening constant and $\hbar\omega_{LO}$ is the LO phonon energy. Since the phonon interactions are the results of lattice vibration, the phonon broadening(denoted as Γ_{phonon}) takes the shape of Lorentzian function (Christen & Bimberg, 1990)

$$\Gamma_{phonon}(E) = \frac{1}{(E-E_0)^2 + {\Gamma_{homo}}^2}, \tag{15}$$

where Γ_{homo} is the homogeneous linewidth given by (14).

The electron-phonon interaction in QDs and the interaction with the wetting layer continuum act as additional sources of lineshape broadening (Sanguinetti et al., 1999). In order to include carrier-phonon interaction into the model, homogeneous energy broadening has to be considered. In QD systems, all sharp excitonic transition lines at different energies are homogeneously broadened by phonons at the same time. The total transition at each energy point is the sum of the contributions of all energy points. Thus the total lineshape of a transition involving both inhomogeneous and homogeneous broadening is the convolution of the individual lineshapes. Based on the discussion, the total transition lineshape of the energy state involving both thermal redistribution and phonon scattering of

carriers is obtained by the convolution of state distribution function and the Lorentzian function $\Gamma_{phonon}(E)$.

$$n_{1f}^{ph}(E) = \int n_{1f}(E - E')\,\Gamma_{phonon}(E')dE', \tag{16}$$

$$n_{2f}^{ph}(E) = \int n_{2f}(E - E')\,\Gamma_{phonon}(E')dE'. \tag{17}$$

Calculations of the temperature dependent FWHMs for the samples, which combine thermal redistribution and electron-phonon scattering effects, are shown in Fig. 6 with adapted values of γ_{AC}=15 μeV/K, γ_{LO}=25 meV, and $\hbar\omega_{LO}$=30 meV for InAs QDs (Gammon et al., 1995; Zhao et al., 2002). The contribution from the thermal redistribution effect on FWHM is also shown. As can be seen in this figure, the experimental data for sample A are fixed to the values obtained from the contribution of redistributed carriers at T<180 K, supplying the evidence of carrier redistribution in the sample. As T>180 K, the temperature is high enough and the electron-phonon scattering starts to come into effect. However, the

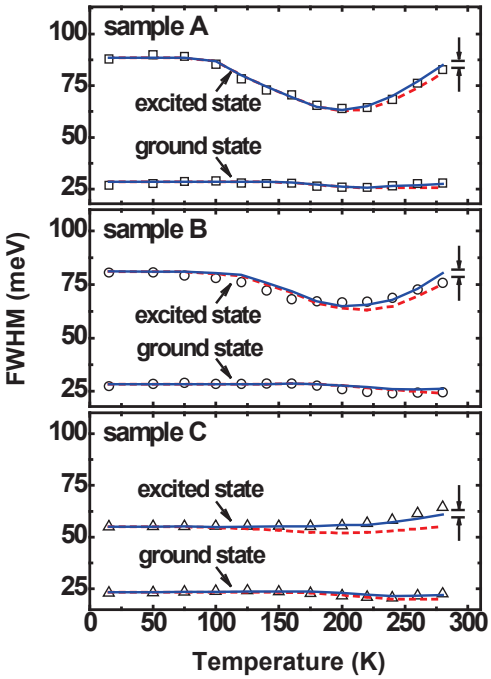


Fig. 6. Experimental and calculated FWHM of sample A, sample B, and sample C. The symbols are experimental data and the solid lines are the calculated results combining the carrier redistribution and electron-phonon scattering effects together. The dashed curves represent the contribution from thermal redistribution of carriers, compared with the linewidth resolution of 5.5 meV

effect on FWHM is unobvious and the tendency of PL linewidth with temperature is dominated by the thermal redistribution of carriers. On the other hand, referring to the simulated PL linewidth of sample C, little decrease is obtained at T=100-200 K in the curve which considers only the contribution of thermal redistribution. It indicates that repopulation of carriers among QDs is existent in this sample, while this phenomenon is too weak to be visible in the measured FWHMs. Joining in the electron-phonon scattering effect; the simulated FWHMs exhibit a monotonous broadening of the spectra as the temperature increases, coinciding with the experimental data.

3.4 Intersublevel Relaxation Process

The intersublevel relaxation lifetimes of the samples can be calculated from (2) with

$$f_1(T) = \frac{n_{1f}(E)}{n_1(E)}. \tag{18}$$

Since $f_1(T)$ is the probability of occupancy for ground state, in equilibrium condition, it is expressed as

$$f_1(T) = \frac{1}{1+\exp\left[(E_1 - E_2)/k_B T\right]}. \tag{19}$$

The relaxation lifetimes simulated by (18) for the samples, as shown in Fig. 7, are decreasing with increasing temperature. These results agree with the increase in number of phonons predicted by the Bose distribution function: $[\exp(\hbar\omega/kT)-1]^{-1}$. It is noticeable that the calculated relaxation lifetimes for sample A, sample B, and sample C at T=15 K are 347 ps, 160 ps, and 40 ps, respectively. Evidently, the lifetimes of sample A and sample B are much longer than that of sample C, resulting from their lower uniformity of QD structures. The shortest intersublevel relaxation times of sample C coincide with the hindered phonon bottleneck.

The corresponding values calculated from (19) of the samples at different temperatures are also shown in Fig. 7. Observing the calculated results from (18) and (19), the discrepancy between the curves is evident at lower temperature but the tendency of them becomes gradually similar as the temperature is raised up. At low temperature, the carrier recombination is much faster than the thermal emission, and the carrier distribution is non-equilibrium (Jiang & Singh, 1999). With the increase of temperature, the thermal emission time reduces and becomes smaller compared to the radiative recombination in the QD system. The carriers redistribute among different dots and thus approach to the thermal equilibrium distribution. Owing to the highest excited state energy and the smallest energy separation between the intersublevels of sample A, carriers start to thermally emit at a lower temperature than the other ones. As a result, sample A exhibits the lowest temperature where the relaxation lifetimes start approaching to the values that predicted under the thermal equilibrium condition.

Calculations of normalized PL peaks intensities of the samples are shown in Fig. 8, correlate well with the measured data. Observing the curves shown in the plot, peak intensities of ground state and excited state quench in the high temperature range because the carriers are

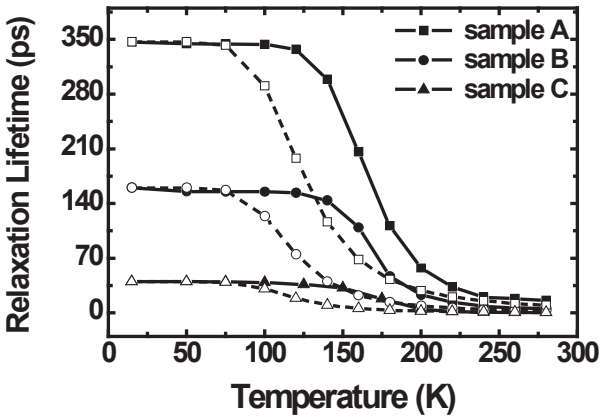


Fig. 7. Calculated intersublevel relaxation lifetimes from excited state to ground state for sample A, sample B, and sample C. The corresponding dashed lines with hollow symbols are the values that calculated under the assumption of thermal equilibrium, and normalized to the simulated relaxation lifetimes of sample A, sample B, and sample C, respectively

emitted into the GaAs barrier and irreversibly lost. We denote the temperature where the excited state starts to quench as T_Q. Sample A exhibits the lowest T_Q (160K) and the fastest quenching rate, coinciding with its highest excited state emission energy and the smallest energy separation. The temperature T_Q of sample B and sample C are 180K and 200 K, respectively. It is noticeable that the thermal quench of ground state is much slower for sample C than that for sample A and sample B. That can be explained by the different intersublevel relaxation lifetimes of the samples. The shorter relaxation lifetimes of sample C imply that the phonon bottleneck effect is partly relaxed for the sample. Through the more active phonon-assisted scatterings, more carriers relax to the ground state during the transferring process, slowing down the quenching rate of ground state.

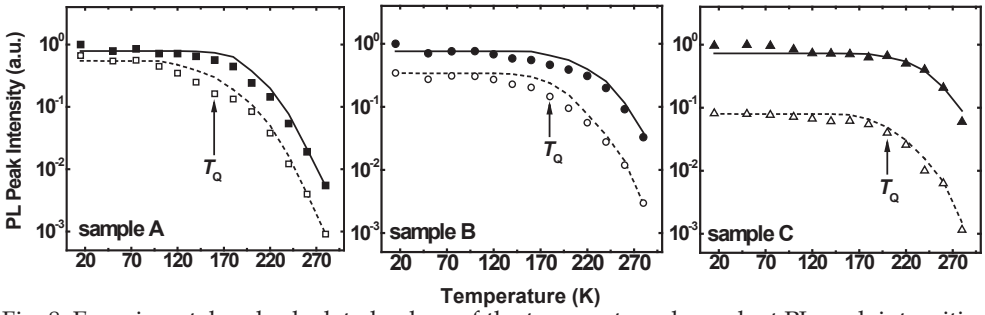


Fig. 8. Experimental and calculated values of the temperature dependent PL peak intensities of ground state and excited state of sample A, sample B, and sample C. The filled (hollow) symbols are the experimental data of ground (excited) state and the solid (dashed) lines are the calculated results of ground (excited) state. T_Q denotes the temperature where the peak intensity starts to quench

The carriers transferring mechanisms are expressed more definitely in Fig. 9 by calculating the numerical values of carriers which transferring in excited state and wetting layer for sample A and sample C. A stronger dependence on temperature is obtained in Fig. 9 for sample A. According to the curves shown in the upper panel of Fig. 9(a), thermally excited carriers from QDs to wetting layer increase rapidly within the temperature range 100-200 K. At T>200 K, the number of emitting carriers saturates and then decreases. Consulting to the plot shown in the lower panel of Fig. 9(a), carriers relaxing from wetting layer to excited state also increases at *T*=100-200 K, indicating the fact of thermal redistribution of carriers. Furthermore, emitting carriers growing up as *T*>200 K, here the temperature is high enough for carriers escaping to the GaAs barrier and irreversibly lost. The thermal loss reduces the excitons in wetting layer, which in turn suppresses the carriers transferring in the excited state.

Calculation for sample C is shown in Fig. 9(b). Comparing the simulation results to that of sample A, it is clearly seen in the upper panel that the calculated radiative recombination term possesses a less important portion of the transferring carriers. The calculated result is consistent with the shorter relaxation lifetime of sample C. Because most of the injected carriers relax to the ground state, fewer carriers exist in the excited state. The thermal redistribution of carriers in the QD system is then retarded; thermal emission occurs at a higher temperature and the amount of thermally escaping excitons is smaller than that of sample A. Consequently, the simulated results exhibit a similar but weaker response to temperature change. It is obvious that the dot size uniformity of the QD systems plays an influential role in the carrier relaxation process.

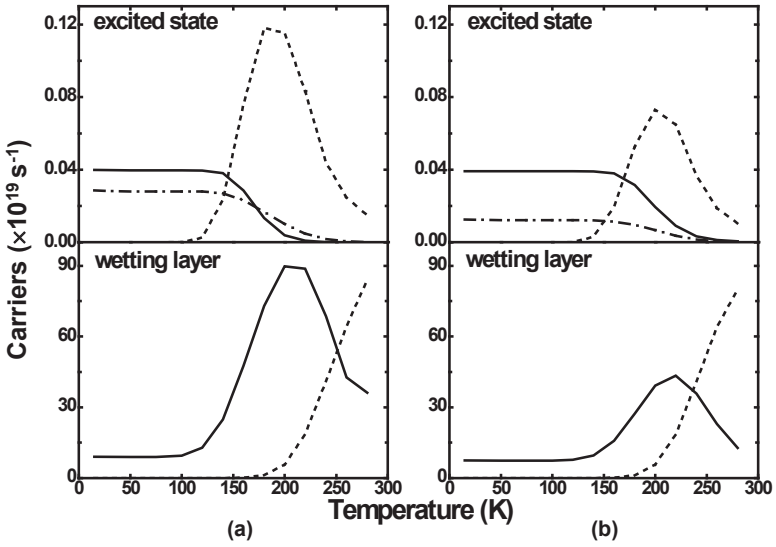


Fig. 9. Amounts of transferring carriers of each energy level in the QD system for (a) sample A, and (b) sample C. Dash-dotted lines in upper panels denote the radiative recombination terms. Solid lines and dotted lines denote the relaxation and thermal emission portions, respectively

4. Conclusion

In this chapter, we have investigated the effects of phonon-assisted transferring of carriers on QD system both experimentally and theoretically. The relaxation and thermal emission of carriers are analyzed quantitatively by a rate-equation model. The model is based on a set of rate equations which connect the ground state, the excited state, the wetting layer, and the GaAs barrier in the QD system. All of the important mechanisms for explaining the unique evolution of quantum dot PL spectra are taken into account, including the inhomogeneous broadening of QDs, the random population of density of states, thermal emission and retrapping, radiative and nonradiative recombination, and the electron-phonon scattering. The simulated results exhibit a good agreement to the experimental data measured from samples with different dot densities and size uniformities for temperatures ranging from 15 K to 280 K. Quantitative discussion of the carriers which thermally excited and relax between the excited state and the wetting layer provides an explicit proof of the thermal redistribution and lateral transition of carriers via the wetting layer.

The phonon-assisted activations of excitons with increasing temperatures are analyzed in detail as well. Homogeneous broadening is included in the rate equation model to demonstrate the correlation between thermal redistribution and electron-phonon scattering effects on the PL spectra of QD system and the intersublevel relaxation lifetimes is calculated. According to the theoretical analysis, carriers redistribute apparently with increasing temperature for sample with evident phonon-bottleneck effect and the effect of electron-phonon scattering is suppressed. On the other hand, the thermal redistribution effect is weak and compensated by the thermal-enhanced electron-phonon scattering for sample with relaxed phonon bottleneck and the electron-phonon scattering occupies an evident portion of the transferring mechanisms in the QD system. It is coinciding with the observed monotonic increase of FWHMs with temperature.

Furthermore, the numerical values of transferring carriers in discrete energy levels under different temperatures are also calculated. The shorter relaxation lifetime of the sample with better size-uniformity implies a restricted phonon bottleneck effect, and the unapparent change of excitons with temperature in each energy level reveals a better thermal stability. The simulation result confirms that the thermal redistribution of carriers and the electron phonon scattering affect the temperature dependent PL spectra simultaneously, and the size-uniformity of quantum dots is of essential importance for thermal activated mechanisms in quantum dot systems. Detailed investigations into the carrier dynamics in QD systems are of particular significance to the design of QD structures. Requirement of the relaxation lifetime is severe in the case of high-speed modulation. Therefore, our work has particular significance to the design of optoelectronic devices by QD structures which exhibiting truly three-dimensional confined state transitions.

5. Acknowledgment

This work was supported by the National Science Council of the Republic of China under Contract NSC 97-2112-M-131-001.

6. References

Bafna, M. K.; Sen, P. & Sen, P. K. (2006). Temperature dependence of the photoluminescence properties of self-assembled InGaAs/GaAs single quantum dot, *Journal of Applied Physics*, Vol. 100, pp. 103515

Benisty, H.; Sotomayor-Torres, C. M. & Weisbuch, C. (1991). Intrinsic mechanism for the poor luminescence properties of quantum-box systems, *Physical Review B*, Vol. 44, pp. 10945-10948

Benisty, H. (1995). Reduced electron-phonon relaxation rates in quantum-box systems: Theoretical analysis, *Physical Review B*, Vol. 51, pp. 13281-13293

Bissiri, M.; Hőgersthal, G. B. H.; Capizzi, M.; Frigeri, P. & Franchi, S. (2001). Quantum size and shape effects on the excited states of $In_xGa_{1-x}As$ quantum dots, *Physical Review B*, Vol. 64, pp. 245337

Bockelmann, U. & Bastard, G. (1990). Phonon scattering and energy relaxation in two-, one-, and zero-dimensional electron gases, *Physical Review B*, Vol. 42, pp. 8947-8951

Chang, W. H.; Hsu, T. M.; Tsai, K. F.; Nee, T. E.; Chyi, J. I. & N. T. Yeh, (1999). Excitation Density and Temperature Dependent Photoluminescence of InGaAs Self-Assembled Quantum Dots, *Japanese Journal of Applied Physics*, Vol. 38, pp. 554-557

Cheng, W. Q.; Xie, X. G.; Zhong, Z. Y.; Cai, L. H.; Huang, Q. & Zhou, J. M. (1998). Photoluminescence from InAs quantum dots on GaAs(100), *Thin Solid Films*, Vol. 312, pp. 287-290

Christen, J. & Bimberg, D. (1990). Line shapes of intersubband and excitonic recombination in quantum wells: Influence of final-state interaction, statistical broadening, and momentum conservation, *Physical Review B*, Vol. 42, pp. 7213-7218

Dawson, P.; Rubel, O.; Baranovskii, S. D.; Pierz, K.; Thomas, P. & Göbel, E. O. (2005). Temperature-dependent optical properties of InAs/GaAs quantum dots: Independent carrier versus exciton relaxation, *Physical Review B*, Vol. 72, pp. 235301

Duarte, C. A.; Silva, E. C. F. da; Quivy, A. A.; Silva, M. J. da; Martini, S.; Leite, J. R.; Meneses, E. A. & Lauretto, E. (2003). Influence of the temperature on the carrier capture into self-assembled InAs/GaAs quantum dots, *Journal of Applied Physics*, Vol. 93, pp. 6279-6283

Gammon, D.; Rudin, S.; Reinecke, T. L.; Katzer, D. S. & Kyono, C. S. (1995). Phonon broadening of excitons in $GaAs/Al_xGa_{1-x}As$ quantum wells, *Physical Review B*, Vol. 51, pp. 16785

Giorgi, M. De; Lingk, C.; Plessen, G. von; Feldmann, J.; Rinaldis, S. De; Passaseo, A.; Vittorio, M. De; Cingolani, R. & Lomascolo, M. (2001). Capture and thermal re-emission of carriers in long-wavelength InGaAs/GaAs quantum dots, *Applied Physics Letters*, Vol. 79, pp. 3968-3970

Grosse, S.; Sandmann, J. H. H.; Plessen, G. von & Feldmann, J. (1997). Carrier relaxation dynamics in quantum dots: Scattering mechanisms and state-filling effects, *Physical Review B*, Vol. 55, pp. 4473-4476

Hai, G. Q. & Oliveira, S. S. (2006). Electron relaxation induced by electron–longitudinal-acoustic-phonon scattering in single and coupled quantum dots in external magnetic and electric fields, *Physical Review B*, Vol. 74, pp. 193303

Jiang, H & Singh, J. (1999). Nonequilibrium distribution in quantum dots lasers and influence on laser spectral output, *Journal of Applied Physics*, Vol. 85, pp. 7438-7442

Lee, J. C.; Hu, Y. J.; Wu, Y. F.; Nee, T. E.; Wang, J. C. & Fang, J. H. (2007). Intersublevel Relaxation Properties of Self-Assembled InAs/GaAs Quantum Dot Heterostructures, *Proceedings of the 7th IEEE International Conference on Nanotechnology*, pp. 934-937, Hong Kong, August 2008

Lee, H.; Yang, W. & Sercel, P. C. (1997). Temperature and excitation dependence of photoluminescence line shape in InAs/GaAs quantum-dot structures, *Physical Review B*, Vol. 55, pp. 9757-9762

Lobo, C.; Leon, R.; Marcinkevicius, S.; Yang, W.; Sercel, P. C.; Liao, X. Z.; Zou, J. & Cockayne, D. J. (1999). Inhibited carrier transfer in ensembles of isolated quantum dots, *Physical Review B*, Vol. 60, pp. 16647-16651

Malik, S.; Ru, E. C. Le; Childs, D. & Murray, R. (2001). Time-resolved studies of annealed InAs/GaAs self-assembled quantum dots, *Physical Review B*, Vol. 63, pp. 155313

Mukai, K.; Ohtsuka, N.; Shoji, H. & Sugawara, M. (1996). Phonon bottleneck in self-formed $In_xGa_{1-x}As/GaAs$ quantum dots by electroluminescence and time-resolved photoluminescence, *Physical Review B*, Vol. 54, pp. R5243-R5246

Nee, T. E.; Wu, Y. F. & Lin, R. M. (2005). Effect of carrier hopping and relaxing on photoluminescence line shape in self-organized InAs quantum dot heterostructures, *Journal of Vacuum Science & Technology B*, Vol. 23, pp. 954-958

Nee, T. E.; Wu, Y. F.; Cheng, C. C. & Shen, H. T. (2006). Carrier dynamics study of temperature and excitation dependent photoluminescence from InAs/GaAs quantum dots, *Journal of Applied Physics*, Vol. 99, pp. 013506

Ortner, G.; Yakovlev, D. R.; Bayer, M.; Rudin, S.; Reinecke, T. L.; Fafard, S.; Wasilewski, Z. & Forchel, A. (2004). Temperature dependence of the zero-phonon linewidth in InAs/GaAs quantum dots, *Physical Review B*, Vol. 70, pp. 201301

Polimeni, A.; Patanè, A.; Henini, M.; Eaves, L. & Main, P. C. (1999). Temperature dependence of the optical properties of $InAs/Al_yGa_{1-y}As$ self-organized quantum dots, *Physical Review B*, Vol. 59, pp. 5064-5068

Raymond, S.; Fafard, S.; Poole, PJ.; Wojs, A.; Hawrylak, P.; Charbonneau, S.; Leonard, D.; Leon, R.; Petroff, PM. & Merz, JL. (1996). State filling and time-resolved photoluminescence of excited states in $In_xGa_{1-x}As/GaAs$ self-assembled quantum dots, *Physical Review B*, Vol. 54, pp. 11548-11554

Sanguinetti, S.; Henini, M.; Alessi, M. G.; Capizzi, M.; Frigeri, P. & Franchi, S. (1999). Carrier thermal escape and retrapping in self-assembled quantum dots, *Physical Review B*, Vol. 60, pp. 8276-8283

Sanguinetti, S.; Watanabe, K.; Tateno, T.; Wakaki, M.; Koguchi, N.; Kuroda, T.; Minami, F. & Gurioli, M. (2002). Role of the wetting layer in the carrier relaxation in quantum dots, *Applied Physics Letters*, Vol. 81, pp. 613-615

Schmidt, K. H.; Medeiros-Ribeiro G.; Oestreich, M. & Petroff, P. M. (1996). Carrier relaxation and electronic structure in InAs self-assembled quantum dots, *Physical Review B*, Vol. 54, pp. 11346-11353

Seebeck, J.; Nielsen, T. R.; Gartner, P. & Jahnke, F. (2005). Coherent resonant tunneling time and velocity in finite periodic systems, *Physical Review B*, Vol. 71, pp. 125317

Smith, L. M.; Leosson, K.; Erland, J.; Jensen, J. R.; Hvam, J. M. & Zwiller, V. (2001). Excited State Dynamics in $In_{0.5}Al_{0.04}Ga_{0.46}As/Al_{0.08}Ga_{0.92}As$ Self-Assembled Quantum Dots, *Physica Status Solidi (b)*, Vol. 224, pp. 447-451

Solomon, G. S.; Trezza, J. A. & Harris, J. S. J. (1995). Effects of monolayer coverage, flux ratio, and growth rate on the island density of InAs islands on GaAs, *Applied Physics Letters*, Vol. 66, pp. 3161-3163

Tarasov, G. G.; Mazur, Yu. I.; Zhuchenko, Z. Ya.; Maaßdorf, A.; Nickel, D.; Tomm, J. W.; Kissel, H.; Walther, C. & Masselink, W. T. (2000). Carrier transfer in self-assembled coupled InAs/GaAs quantum dots, *Journal of Applied Physics*, Vol. 88, pp. 7162-7170

Tomm, J. W.; Mazur, Y. I.; Tarasov, G. G.; Zhuchenko, Z. Y.; Kissel, H.; Masselink, W. T. & Elsaesser, T. (2003). Transit luminescence of dense InAs/GaAs quantum dot arrays, *Physical Review B*, Vol. 67, pp. 045326

Wu, Y. F.; Lee, J. C. & Hu, Y. J. (2008). Effect of Phonon-assisted Transferring of Excitons on Photoluminescence Spectra from InAs Quantum Dots, *Proceedings of 2008 International Electron Devices and Materials Symposia*, pp. 676-679, Taiwan, November 2008

Xu, Z. Y.; Lu, Z. D.; Yang, X. P.; Yuan, Z. L.; Zheng, B. Z.; Xu, J. Z.; Ge, W. K.; Wang, Y.; Wang, J. & Chang, L. L. (1996). Carrier relaxation and thermal activation of localized excitons in self-organized InAs multilayers grown on GaAs substrates, *Physical Review B*, Vol. 54, pp. 11528-11531

Yang, W.; Lowe-Webb, R. R.; Lee, H. & Sercel, P. C. (1997). Effect of carrier emission and retrapping on luminescence time decays in InAs/GaAs quantum dots, *Physical Review B*, Vol. 56, pp. 13314-13320

Zhang, X. Q.; Ganapathy, S.; Kumano, H.; Uesugi, K. & Suemene, I. (2002). Photoexcited carrier transfer in InGaAs quantum dot structures: Dependence on the dot density, *Journal of Applied Physics*, Vol. 92, pp. 6813-6818

Zhao, H.; Wachter, S. & Kalt, H. (2002). Effect of quantum confinement on exciton-phonon interactions, *Physical Review B*, Vol. 66, pp. 085337

15

Arithmetic Circuits Realized by Transferring Single Electrons

Wan-cheng Zhang and Nan-jian Wu
Institute of Semiconductors, Chinese Academy of Sciences
P. R. CHINA

1. Introduction

A number of challenges are facing the semiconductor industry, such as increases of power consumption and interconnects delay. The combination of current CMOS technology and novel nanotechnologies like the single-electron technology is a promising approach to solve these problems. Single-electron devices (SEDs) operate by controlling the movements of individual electrons based on the Coulomb Blockade effect (Likharev, 1999). They have potentially small device area and very low power dissipation. Single-electron circuit, in which discrete electrons are used to process information, can be viewed as the ultimate goal of electronic circuits (Ono et al., 2006). Therefore, the single-electron technology is attracting large interest in the last decade. Various single-electron circuit blocks, including memory (Yano, 1994), inverter (Ono et al., 2000), logic gates (Asahi, 1997), multiple-valued logic circuits (Inokawa et al., 2003) and sensors have been proposed and intensively studied. These circuits used the unique characteristic of SEDs and some of them have been proven to have impressive circuit performances.

Most of previous single-electron circuits have similar operation principle with their CMOS counterparts. For example, the single-electron inverter is composed of two single-electron transistors (SETs) with different I_d-V_g characteristics, just like the CMOS inverter (Ono et al., 2000). The input/output signals are represented by node voltages. The only difference is that the operations of SETs require much smaller energy than MOS transistors. We argue that, this scheme does not release the full potential of SEDs. Actually, the single-electron technology would represent a revolution not only in scaling down but also in its physical foundation of the electron charge discreteness devices. The recent development of single-electron turnstile already allows us to accurately control the transfer of single-electrons and at relative high temperature (Nishiguchi et al., 2006). In this chapter, we will show that, by using single-electron transfer devices to manipulate the transfer of single-electrons, it is able to implement smarter arithmetic circuits with very compact structures and impressive circuit performances.

Arithmetic circuits like adder and multiplier are regarded as very critical components in modern information-processing systems. A promising nanoscale adder circuit should have high integration density, low power dissipation and high speed, which is a great challenge. The worst-case propagation delay t_d of a conventional ripple-carry adder is proportional to

its operand length n, so that the operation speed of ripple-carry adder is low. Carry look-ahead adder can reduce t_d to the order of log$n$, but circuit area are largely increased. An interesting family of adders use non-binary arithmetic algorithms based on high-radix number systems or redundant number systems, such as the signed-digit (SD) adders (Avizienis, 1961) and the redundant-binary (RB) adders (Takagi, 1985). Because the carry propagations in these adders are restricted only to adjoining cells, it is possible to perform addition of two operands in constant time which is not dependent on n (Parhami, 1990; Parhami, 1993). Hereafter these non-binary adders are referred to as fast adders. Fast adders have promising characteristics, but their compact and efficient implementation still remains a big challenge. The conventional approaches use binary MOS logic gates to implement non-binary algorithms so that the adders are complex in circuit structure and are thus hard to design. Moreover, each type of fast adders requires specific consideration to optimize its performance. On the other hand, the multiple-valued current-mode logic (MVCL) approach can significantly reduce the number of devices in the circuit (Kawahito, 1988). However, MVCL suffers from relatively large power dissipation and it results in low overall area-time-power performances. Although novel approach using negative differential-resistance devices can greatly reduces the number of transistors (Gonzalez, 1998), it is only specific to one particular kind of fast adder.

In this chapter, we proposed circuit architecture and design methods to implement novel *fast adders and fast multipliers* by transferring & storing single electrons. We use the number of electrons to represent different logic values and we perform arithmetic operations by accurately manipulating electrons. We propose general fast single-electron adder architecture based on non-binary arithmetic and design methods to implement various fast adder circuits. We adopt the counter tree diagram (CTD) (Sakiyama, 2003) to represent and analyze our fast adders, and we use the MOSFET-based single-electron turnstile as the basic circuit element. We used the unique characteristic of MOSFET-based single-electron turnstile to finish complicated fast-addition arithmetic operations compactly. We propose two design styles to implement fast adder circuit blocks: the threshold approach and the periodic approach. The proposed adder circuits have several advantages: 1) The operation speeds are high; 2) The circuit structures are compact and the number of devices is small; 3) The power dissipations are much lower than conventional circuits; 4) The circuit design method based on the CTD is very simple and can be applied to a wide range of adders. In the following sections, first we introduce the background of single-electron devices and operation principle of MOSFET-based single-electron turnstile. Then we introduce fast addition algorithms and the circuit architecture to perform non-binary fast addition by transferring single-electrons. Then we introduce a family of SE transfer circuits based on MOSFET-based SE turnstile. Next we show the principles of the threshold approach and the periodic approach, and we show adder design examples for each approach. After that we study and compare the performances of the proposed adders. Finally, we summarize the results.

2. MOSFET-based Single Electron Turnstile

2.1 Working Principle

First we introduce the basic device of our work: the single-electron (SE) turnstile. MOSFET-based single-electron turnstile is a very promising SE transfer device, which could

accurately control the number of transferred electrons using the Coulomb blockade effect (Fujiwara, 2008). Room-temperature (RT) operation of a SE multilevel memory and RT operation of a digital-analog converter circuit (Nishiguchi, 2006) consisting of MOSFET-based SE turnstiles have been experimentally demonstrated. Fig. 1(a) shows the device structure of the SE turnstile. The turnstile has two MOSFETs, FET1 and FET2. The FET1 and FET2 are gate-all-around Si-nanowire MOSFETs on SOI wafer. By turning FET1 and FET2 on and off alternately, the single electrons are transferred from the source to the drain, like conventional charge-coupled devices (CCDs). Fig. 1 (b) shows a SEM picture of the SE turnstile (Fujiwara, 2004). Fig. 1(c) shows the equivalent circuit of the SE turnstile, which includes a source S, a drain D, a gate voltage terminal G and two clock voltage terminals *clk1* and *clk2*. In the following circuits, the source of the SE turnstile is connected to a supply voltage, V_{ss} or $-V_{ss}$. The drain V_D is always connected to an electron storage node (SN), and it can be regarded as virtually grounded. The single electrons can be injected into the SN or ejected from the SN through the SE turnstile (Zhang, 2007a). The number of electrons in the SN can be detected by using the single-electron transistor as an electrometer. Experimentally, a SN with small capacitance C_{SN} can be realized by a silicon nanodot on SOI wafer.

Fig. 1(d) shows the pulse sequences for V_{clk1} and V_{clk2} applied to the gates of FET1 and FET2, respectively. The SE turnstile operation requires two repulsive voltage pulses with a duty cycle less than 50%. Fig. 1(e) shows how the electrons are transferred from the source to the SN, according to steps (i)-(iv) shown in Fig. 1(d). The source of the SE turnstile is biased by $-V_{ss}$ and V_G is negative. When both FET1 and FET2 are turned off, a single-electron-box (SEB) is electrically formed. Note that, the size of the SEB is much smaller than its lithography definition, due to the barriers of the two MOSFETs. The SEB is small enough to activate the Coulomb blockade effect. Since the SEB and source are separated by FET1, the potential of the SEB is only controlled by the gate voltage V_G via electrical coupling. Therefore, the number of captured electrons N_s is determined by the difference between V_G and $-V_{ss}$. Assuming an ideal case at working temperature T=0, N_s is given by (Zhang, 2008):

$$N_s=0, \textit{ if } V_G \le -V_{ss}; \qquad (1)$$
$$N_s=[(V_G+V_{ss})/V_0+1/2],\ V_0=e/C_{ug},\ \textit{if } V_G \le -V_{ss.}$$

where V_0 is defined as constant logic value 1, C_{ug} is the capacitance between the gate and the SEB, and $[X]$ denotes the maximum integer which is smaller than X. When FET2 is turned on, the SEB is connected to the SN [step (iii)]. The capacitance of SN C_{SN} is much larger than the capacitance of the SEB, so when the electrons enter the SN, its potential change can not affect the behavior of the SE turnstile and thus can be neglected. Therefore we assume the potential of the SN is always 0V for simplicity. When V_G<0, the potential of the SEB is higher than the SN so that all single electrons flow into the SN. In this case, the number N=N_s of electrons transferred depends exclusively on V_G. On the other hand, when V_G>0, not all electrons flow out of the SEB. In this case N only depends on the potential difference between the source and the SN. Finally FET2 is turned off [step (iv)], and a transfer cycle is finished. In summary, when the SE turnstile injects electrons into the SN, N is given by:

$$\begin{aligned} N&=0, && \textit{if } V_G \le -V_{ss};\\ N&=[C_g(V_G+V_{ss})/e+1/2], && \textit{if } -V_{ss}<V_G<0;\\ N&=[C_gV_{ss}/e+1/2], && \textit{if } V_G \ge 0. \end{aligned} \qquad (2)$$

Similarly, electrons can be ejected from the SN. In this case the source of the SE turnstile is connected to V_{ss} and V_G is positive. In this case, the number of captured electrons N_d in the SEB is determined by the potential difference between the SEB and SN. In step (iii) and step

(iv), electrons flow out of the SEB to the source. When $V_G<V_{ss}$, the potential of the SEB is higher than the source, and all electrons flow out of the SEB. So $N=N_d$ also depends exclusively on V_G. On the other hand, when $V_G>V_{ss}$, the potential of the SEB is lower than the source, and not all electrons flow out of the SEB. In this case N also only depends on V_{ss}. In summary, when the SE turnstile ejects electrons from the SN, N is given by:

$$\begin{aligned} N&=0, \textit{ if } V_G\le 0;\\ N&=[C_gV_G/e+1/2], \textit{ if } 0<V_G<V_{ss};\\ N&=[C_gV_{ss}/e+1/2], \textit{ if } V_G\ge V_{ss}. \end{aligned} \tag{3}$$

The relationships between N and V_G, V_{ss} are summarized in Fig. 2(a). From equations (2) and (3) we see that, the gate voltage V_G directly controls N. Fig. 2(b) shows the relationship between N and V_G when V_{ss} is large enough. N is a periodical staircase function of V_G. This unique characteristic of SE turnstile makes the single-electron transfer highly flexible.

The parameters of the SE turnstile are determined as follows. Transfer error may occur in the transfer cycle. To reduce transfer errors, the capacitance of the SEB C_{SEB} must be small enough. Since a 0.76aF C_{SEB} was reported (Nishiguchi, 2006), we choose C_{SEB}=0.7aF and we choose T=60K to ensure low transfer error rate. We choose C_{ug}=0.53aF so that V_0=0.3V, and thus the circuit can have large noise margin.

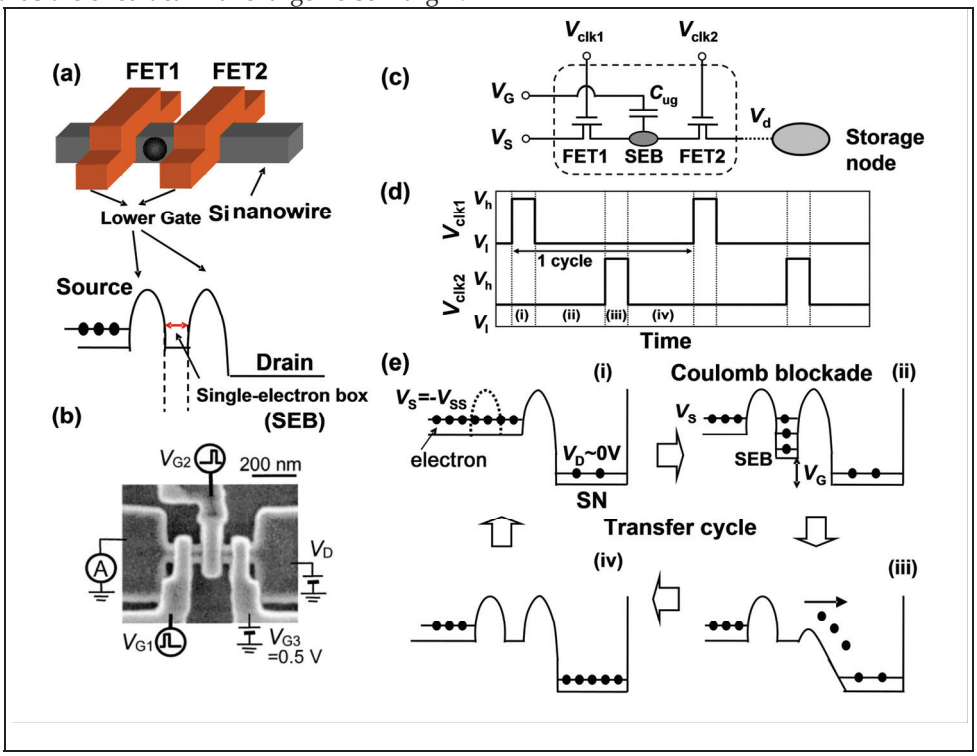


Fig. 1. (a) Device structure of the MOSFET-based SE turnstile. (b) SEM picture of the SE turnstile (Fujiwara, 2004). (c) Equivalent circuit of the SE turnstile. The drain connects to a storage node (SN). (d) Repulsive clock pulses for the SE turnstile operation. (e) Schematic diagram to accurately inject electrons into the SN. A transfer cycle has 4 steps.

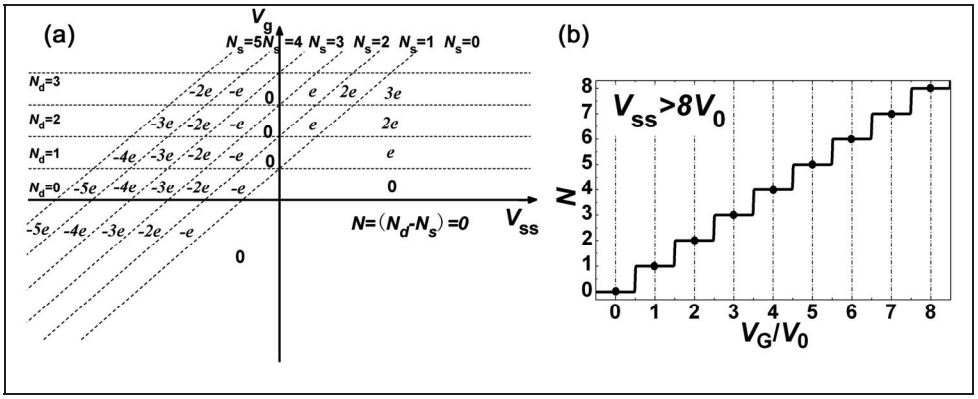


Fig. 2. (a) Relationship between the transfer electron number N with gate voltage and source voltage; (b) N as a function of gate voltage.

2.2 SPICE Simulation Model

Although the MOSFET-based turnstile is composed of MOSFETs, its behavior can not be directly simulated by SPICE because SPICE always assume that currents are continuous and thus can not handle the stochastic behavior of SE transfer. We proposed a behavioral SPICE model to simulate the MOSFET-based SE turnstile (Zhang, 2007c; Zhang 2008). The simplified model schematic is shown in Fig. 3. The most essential points of the model are: 1) to represent the discrete single-charge transfer event as a δ-type current $i_t=e\delta(t-t_0)$, where t_0 is the time of the transfer event; 2) to model the stochastic electron transfer by using random number generator. Although the transfer event is a complex random process, we assume that t_0 is the time when

$$ep = \int_0^{t_0} I_d(t)dt \tag{4}$$

where $I_d(t)$ is the current through the MOSFET and p is a random number distributed uniformly from 0 to 1, which represents the randomness of the transfer process.

In the model, the SEB is modeled as a pure capacitor with capacitance C_{SEB}. G1 is a voltage-controlled current mirror of I_1, and it is controlled by the output of module P1. With the falling clock of *clk1*, C_{SEB} is charged and its voltage V_{SEB} is feed back to module P1. The number of electrons stored in the SEB N_{SEB} is controlled by V_g. When $V_{SEB}> N_{SEB}e/C_{SEB}$, the output of P1 shut off G1, so the equivalent charge stored in the SEB is exactly $N_{SEB}e$. A noise source generates a noise voltage to module P1, so N_{SEB} is randomly changed by the noise voltage with a possibility ε, and ε corresponds to the transfer error rate of the SE turnstile. With *clk1*, N_{SN} electrons are equivalently stored in C_{SEB}. With the rising clock of *clk2*, these electrons are transferred from C_{SEB} to drain one by one. G2 is a current mirror of I_2. With the rising clock of *clk2*, G2 charges C_E. When the charge on C_E is larger than *ep* in (4), we assume that a SE transfer event will happen. Then the output of module P2 resets the charge on C_E to 0 and transiently enables G3 so that G3 outputs a delta-shape current pulse, as shown in Fig. 3. The area of the delta-shape current pulse is exactly e, and we use this current pulse to represent SE transfer. G3 is transiently opened for N times until all electron flow to the drain.

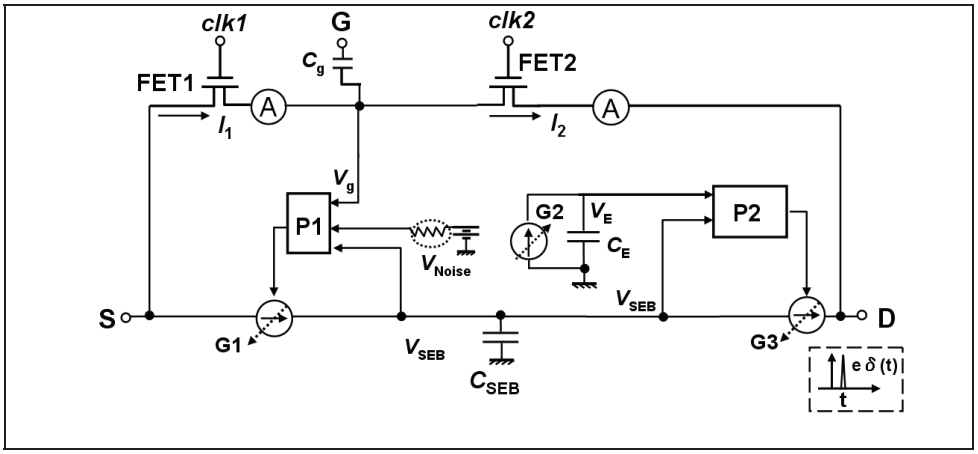


Fig. 3. SPICE circuit model of MOSFET-based single-electron turnstile

2.3 Electron Transfer Circuit Elements

Fig. 4 shows a family of SE transfer circuits using the MOSFET-based SE turnstile. These basic electron transfer circuits sever as the basic building blocks of single-electron arithmetic circuits. The circuit symbol of SE turnstile has terminals G, S and D. The linear ejector (LE) is simply a SE turnstile. It ejects N electrons from the SN per cycle and N depends on V_G according to (3). After one cycle, N_{SN} is decreased by N. In the linear injector (LI), V_G is summed with $-V_{ss}$ by a voltage adder and is then connected to the G terminal of the SE turnstile. The relationship between N and V_G is same as in the LE, according to (2). The fixed ejector (FE) has an E terminal. The *clk2* terminal of the SE turnstile connects a PMOS transistor in series, and E connects the gate of the PMOS transistor. The G terminal of the SE turnstile connects a constant bias $V_G=NV_0$. When we apply high voltage (V_{ss}) to E, the PMOS transistor cuts off, and no electrons are transferred by the SE turnstile. When we apply low voltage ($-V_{ss}$) to E, the PMOS transistor turns on and N electrons are ejected per cycle. Similarly the fixed injector (FI) also has an E terminal and it injects N electrons into the SN per cycle if we apply $-V_{ss}$ to E.

Fig. 5 shows the schematics and functions of other useful circuit elements. The voltage adder is simply composed of capacitors with equal capacitances. The voltage divider consists of two capacitors C_1 and C_2. The function of the voltage divider is $V_{out}=V_{in}/f$, where $f=(C_1+C_2)/C_1$ is the division factor. The threshold inverter is a CMOS inverter and its logic threshold value is set to $(t\text{-}1/2)V_0$, where t is a designated integer. The reset circuit is a NMOS transistor with its drain connected to the ground. When a clock pulse is applied, all electrons flow out of the SN and the logic value of the SN is reset to 0.

The charge-voltage converter is a SET/MOS hybrid circuit and can readout the number of electrons stored in the SN. It consists of a dual gate SET, a PMOS transistor as a constant current source, and a NMOS transistor as a cascode device. The operation principle and implementation details of the SET/MOS hybrid circuit were investigated (Zhang, 2007b). The SET acts as a very sensitive electrometer. The output voltage of the SET/MOS hybrid circuit depends linearly on the input voltage and it can accurately represent N_{SN}, as shown in Fig. 5.

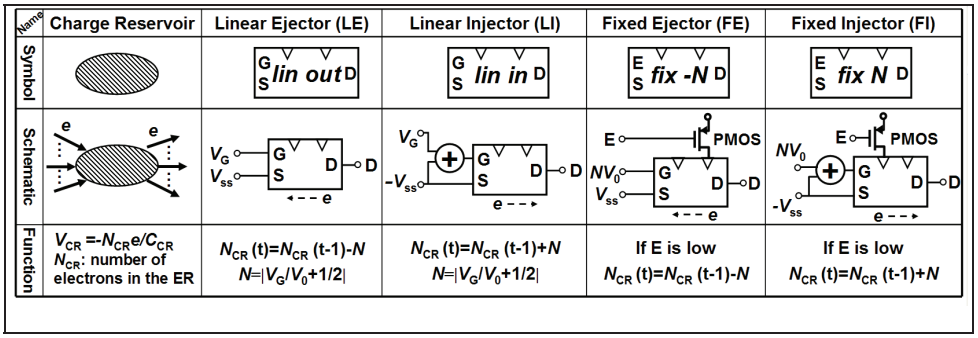


Fig. 4. A family of SE transfer circuits based on MOSFET-based SE turnstiles. The schematic includes the symbol of the SE turnstile.

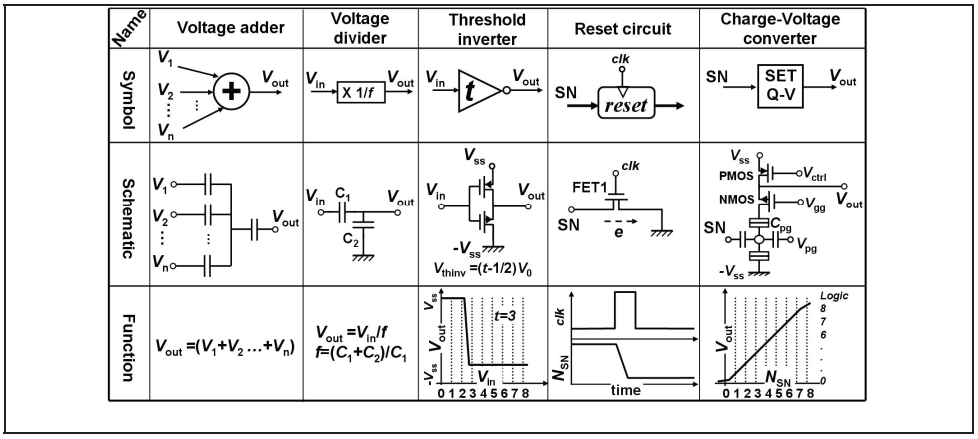


Fig. 5. Symbol, circuit schematic and function of other useful circuit elements.

3. Architecture of Single-electron Fast Adder

3.1 Fast addition algorithm and counter tree diagram (CTD)

The number system in the fast adder belongs to a generalized signed-digit (GSD) number system (Parhami, 1990). The operands belong to the digit set $\{-\alpha, \alpha+1,\ldots, \beta\}$, where both α and β are positive integers or zero. The redundancy index ρ of the GSD number system is defined as $\rho=\alpha+\beta+1-r$, where r is the number representation radix. For limited carry propagation, ρ must be larger than 0. The algorithms of the fast adders have been intensively studied. According to the number system used and its redundancy, the fast addition algorithms can be classified into many categories, such as carry-free, limited-carry, stored-carry, *etc*. Given a particular number system, there may be several valid choices for the range of carries and intermediate variables. Because these algorithms are quite complicated, it is difficult to image a circuit schematic only from definitions and equations.

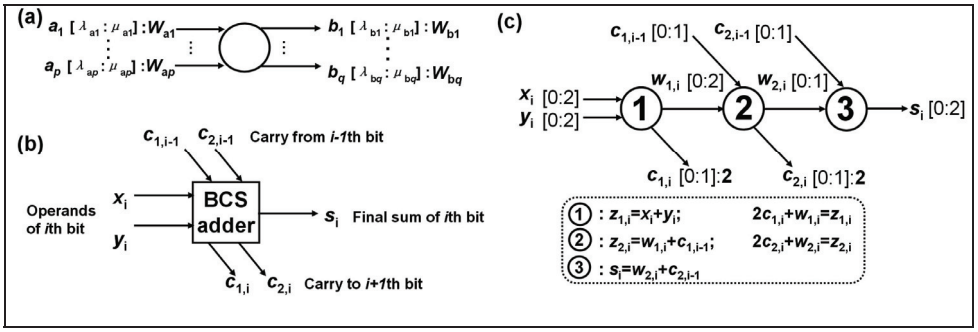


Fig. 6. (a) Symbol of a p-input & q-output counter node in the CTD. (b) Schematic of the ith bit BCS adder. (c) CTD representation of the BCS algorithm.

The counter tree diagram (CTD) is a graphic tool to represent fast addition algorithms and to design the adder circuits (Sakiyama, 2003; Homma, 2004). The CTD is a network that consists of counter nodes and directed edges. The directed edge is an abstraction of the flow of integer data between the counter nodes. Each directed edge is associated with a variable and a weighted interval $[\lambda_{a1}:\mu_{a1}]$: w_{a1}. Here a_1 is the variable name, $[\lambda_{a1}:\mu_{a1}]$ is the range of a_1 and w_{a1} is the weight. If the weight is 1, then it is not shown for simplicity. The counter node is an abstraction of a multi-operand addition function. Fig. 6(a) shows the symbol of a p-input & q-output counter node. The counter node receives p input variables a_1, a_2,…, a_p through the p directed edges and generates q output variables b_1, b_2,…, b_q through the q directed edges while keeping

$$\sum_{i=1}^{p} w_{ai} a_i = \sum_{j=1}^{q} w_{bj} b_j \tag{5}$$

Here p input variables a_1, a_2,…, a_p are associated with p ranges and weights $[\lambda_{a1}:\mu_{a1}]$: w_{a1},…, $[\lambda_{ap}:\mu_{ap}]$: w_{ap}, respectively, and q outputs variables b_1, b_2,…, b_q are associated with q ranges and weights $[\lambda_{b1}:\mu_{b1}]$: w_{b1},…, $[\lambda_{bq}:\mu_{bq}]$: w_{bq}, respectively. The weights denote the arithmetic relationships between the input variables and output variables. The variables and their ranges are clearly shown in Fig. 6(a).

The CTD provides a uniform graphic tool to depict the algorithms of different adders. We consider the addition of two n-bit operands $X=(x_{n-1}…x_i…x_0)$ and $Y=(y_{n-1}…y_i…y_0)$ using the radix-2 binary carry-save (BCS) algorithm. The addition requires $(n+1)$ identical BCS adders operating in parallel. The symbol of the ith BCS adder is shown in Fig. 6(b). The adder receives two carries $c_{1,i-1}$, $c_{2,i-1}$ from the $(i-1)$th adder and the ith input operands x_i and y_i. The adder outputs the sum s_i and two carries $c_{1,i}$, $c_{2,i}$ to the $(i+1)$th adder. Fig. 6(c) shows one type of the CTD representation of the ith BCS adder (Homma, 2004). The CTD shows that the adder consists of 3 stages and the carry propagation length is two, and it directly represents the algorithm of BCS addition. We can directly write the following steps to perform addition of X and Y:

$$\begin{aligned} 2c_{1,i}+w_{1,i} &= x_i+y_i; \\ 2c_{2,i}+w_{2,i} &= w_{1,i}+c_{1,i-1}; \\ s_i &= w_{2,i}+c_{2,i-1}; \end{aligned} \tag{6}$$

where x_i and y_i are the ith bits of the input operands, $s_i \in [0:2]$ is the final sum of the ith adder, $c_{1,i}$, $c_{1,i-1} \in [0:1]$ and $c_{2,i}$, $c_{2,i-1} \in [0:1]$ are carries (transfer digits), and $w_{1,i} \in [0:2]$ and $w_{2,i}$

$\in$[0:2] are two intermediate variables. The equations and range of variables are clearly shown in Fig. 6(c). The weight 2 associated with $c_{1,i}$ and $c_{2,i}$ represents the arithmetic relationships in (5). We can further introduce two intermediate sums $z_{1,i}$ and $z_{2,i}$ to show the calculation process more clearly. Then the algorithm of the BCS addition is

$$\begin{aligned} z_{1,i}&=x_i+y_i; \qquad 2c_{1,i}+w_{1,i}=z_{1,i};\\ z_{2,i}&=w_{1,i}+c_{1,i-1}; \qquad 2c_{2,i}+w_{2,i}=z_{2,i};\\ s_i&=w_{2,i}+c_{2,i-1}; \end{aligned} \tag{7}$$

This algorithm is a representative fast addition algorithm. Generally an n-bit adder consists of $(n+1)$ full adders, and its CTD representation includes $(n+1)m$ counter nodes in all, where m is the CTD stages. In most cases, m=2 or m=3, according to carry-free algorithms and limited-carry algorithms, respectively (Parhami, 1990). The inputs of the kth counter node in the ith adder are $w_{k-1,i}$ and $c_{k-1,i-1}$, which are the output of the $(k-1)$th counter node of the ith adder and the $(k-1)$th carry output of the $(i-1)$th adder, respectively (In some cases one counter node may receive more than one carries, as discussed later). The kth intermediate sum $z_{k,i}$ is defined as the sum of $w_{k-1,i}$ and $c_{k-1,i-1}$.The outputs of the kth counter node are $w_{k,i}$ and $c_{k,i}$, and $z_{k,i}=rc_{k,i}+ w_{k,i}$, where r is the number radix.

3.2 General circuit architecture

We realize fast addition algorithms by manipulating SEs. Fig. 6 shows the general circuit structure of an n-bit fast adder. It consists of $(n+1)$ full adders with m-stage circuit blocks. The architecture has a similar structure to the MVCL-based adders. However, we use single electrons instead of currents to finish arithmetic operations. Given an addition algorithm, the first design step is to draw the CTD of the addition algorithm. The second step is to replace m counter nodes with m circuit blocks and replace the directed edges with the flow paths of the SEs, as shown in the bold lines in Fig. 6(a). The electron numbers are directly used to represent the variables. The arithmetic operations and the transfer of variables are finished by manipulating the SEs. The transfer of SEs in the ith adder is as follows. First, x_i and y_i electrons are transferred into the first circuit block. The first block performs the operation by following the addition algorithm, and then it outputs $w_{1,i}$ electrons and $c_{1,i}$ electrons to the 2nd block of the ith adder and to the 2nd block of the $(i+1)$th adder, respectively. Similarly, the kth block receives $w_{k-1,i}$ and $c_{k-1,i-1}$ electrons from previous circuit blocks, and it outputs $w_{k,i}$ electrons to the $(k+1)$th block of the ith adder and $c_{k,i}$ electrons to the $(k+1)$th block of the $(i+1)$th adder. Finally, the mth block outputs s_i electrons as the final sum.

Fig. 6(b) shows the circuit structure of the 1st, 2nd,..., $(m-1)$th circuit blocks in the ith adder. It consists of a electron storage node (SN), a charge-voltage converter, a reset circuit, and move electron blocks (MVEs) M_w, M_c. The SN connects to SE transfer devices, and SEs can be injected into the reservoir or ejected from the reservoir through the SE transfer devices. In the kth circuit block, first $w_{k-1,i}$ and $c_{k-1,i-1}$ electrons are transferred into the kth SN from previous blocks and are added to be $z_{k,i}=w_{k-1,i}+c_{k-1,i-1}$, as shown in Fig. 6(b). Then, $z_{k,i}$ electrons are stored there. The charge-voltage converter is a SET/MOS hybrid circuit. It readouts the number N_{SN} ($=z_{k,i}$) of the electrons stored in the SN and converts it to the output voltage signal $z_{k,i}V_0$, where V_0 is defined as the voltage of logic 1. Next, this output voltage is applied to the MVEs M_w and M_c. The proposed MVEs are mainly composed of the SE transfer devices. The number of electrons transferred through the MVEs can be accurately controlled by its input voltages. The M_w and the M_c blocks output $w_{k,i}$ and $c_{k,i}$

electrons, respectively. The combination of the SN, the charge-voltage converter, and the MVEs realizes the functions of the CTD counter nodes. The single electrons are transferred between the blocks stage by stage. After one circuit block has finished its function, the reset circuits release the electrons stored in the SN to the ground for the next operation cycle. Fig. 6(c) shows the structure of the last circuit block. It consists of a SN and a reset circuit. It receives $c_{m\text{-}1,i\text{-}1}$ and $w_{m\text{-}1,i}$ electrons so that $s_i=c_{m\text{-}1,i\text{-}1}+w_{m\text{-}1,i}$ electrons are stored in the SN as the adder's output. These electrons can be transferred to other circuits or converted to a voltage signal for further operations.

One merit of the above circuit structure is that the circuit implementation of an arbitrary addition algorithm can be directly obtained by mapping its CTD to the electrons flow paths and the MVEs. The arithmetic operations are achieved by accurately manipulating the single electrons, and the flexible SE transfer devices lead to simple and compact realization of MVEs. In the following sections, we will use the family of SE transfer circuits to build the MVEs.

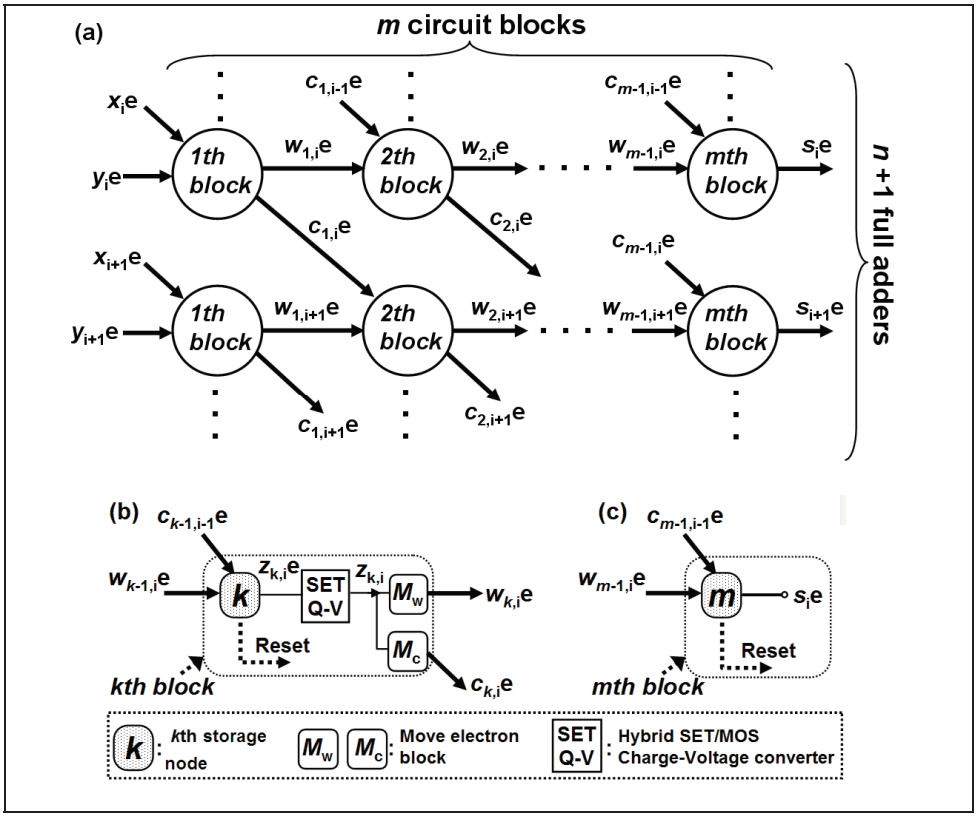


Fig. 6. (a) General circuit structure to implement fast adders by manipulating single-electrons. Each bit addition circuit has m circuit blocks. The bold arrows represent the flow paths of electrons. (b) Circuit schematic of the *1*st, 2nd,..,(*m-1*)th circuit blocks. (c) Circuit schematic of the *m*th circuit blocks.

4. Design of Single-electron Adder Circuits

In previous sections, we have shown the novel circuit architecture. The main task in circuit design is to implement the MVEs. In this section, we show two approaches to design the MVEs using the proposed family of SE transfer circuits.

4.1 Threshold Approach

The functions of the MVEs M_w and M_c are to output $w_{k,i}$ electrons and $c_{k,i}$ electrons, respectively, when a voltage signal $z_{k,i}$ is inputted. The values of $w_{k,i}$ and $c_{k,i}$ are obtained from the equation $z_{k,i}=rc_{k,i}+ w_{k,i}$. Without lost of generality, we assume $c_{k,i}$ belongs to the interval $[0:\mu]$. The idea of the threshold approach is to compare $z_{k,i}$ with μ+1 integer threshold constants $t_1,\ldots, t_{\mu+1}$, and to obtain $c_{k,i}$ and $w_{k,i}$ from the comparison results. Here $t_\mu>t_{\mu-1}>...>t_1$, $t_{\mu+1}=\infty$. If $t_j\le z_{k,i}< t_{j+1}$, then j comparison results are '1' and thus $c_{k,i}=j$. The comparison process is graphically shown in Fig. 7(a). We can obtain $c_{k,i}$ and $w_{k,i}$ from $z_{k,i}$ by performing all of the following conditional statements:

$$\begin{aligned} &\textit{Inital } c_{k,i}=0,\ w_{k,i.}= z_{k,i};\\ &\text{if } z_{k,i}\ge t_1, \text{ then } c_{k,i} \text{ -> } c_{k,i}+1,\ w_{k,i.}\text{-> } w_{k,i}-r;\\ &\text{if } z_{k,i}\ge t_2, \text{ then } c_{k,i} \text{ -> } c_{k,i}+1,\ w_{k,i.}\text{-> } w_{k,i}-r;\\ &\qquad\ldots\ldots\\ &\text{if } z_{k,i}\ge t_\mu, \text{ then } c_{k,i} \text{ -> } \mu,\ w_{k,i.}\text{-> } w_{k,i}-r. \end{aligned} \tag{8}$$

From this algorithm, the circuit implementation of the M_c block has μ conditional FIs and μ inverters, as shown in Fig. 7(b). We use the output results of the inverters to control the operations of the SE turnstiles. The logic threshold value of the jth threshold inverter is t_j, and its output terminal is connected to the E terminal of the jth FI. The jth threshold inverter and the jth FI correspond to the $(j+1)$th statement in (8). If $t_j\le z_{k,i}<t_{j+1}$, then j FIs are enabled and each FI ejects one electron into the SN. Therfore the overall M_c block injects $c_{k,i}=j$ electrons into the SN. Similarly, the M_w block has one LI, μ threshold inverters and μ FEs, as shown in Fig. 7(c). The LI's G terminal connects the input voltage, so LI injects $z_{k,i}$ electrons into the kth SN. If $t_j\le z_{k,i}<t_{j+1}$, j FEs are enabled, and each FE ejects r electrons from the SN. Thus, the overall M_w block injects $w_{k,i.}= z_{k,i}-jr$ electrons into the SN.

In summary, the threshold approach uses threshold inverters as comparators. The intermediate variables are obtained from the comparison results. To optimize the adder circuits, μ should be as small as possible.

Example1: Design of a BCS adder

The algorithm for the BCS adder was shown in (1). Fig. 8(a) shows one type of CTD of the BCS adder. Both $c_{1,i}$ and $c_{2,i}$ belong to [0:1]. Fig. 8(b) shows the characteristics of $w_{k,i}$ and $c_{k,i}$ as functions of $z_{k,i}$. We see that $c_{k,i}$ =1 is a simple threshold function of $z_{k,I}$, and $w_{k,i}$ and $c_{k,i}$ are:

$$\begin{aligned} &w_{1,i.}= z_{k,i},\ c_{k,i}=0 \quad \text{if } z_{k,i} < 2,\\ &w_{1,i.}= z_{k,i}-2,\ c_{k,i}=1 \quad \text{if } z_{k,i} \ge 2. \ (k=1 \text{ or } 2). \end{aligned} \tag{9}$$

Fig. 8(c) shows the schematic of the BCS adder using the threshold approach. The circuit consists of 3 blocks, since the CTD has 3 stages. The M_w block consists of a LI, an inverter, and a FE, since μ=1. The M_c block simply consists of a FI and it shares the inverter with the M_w block. The circuit has 29 transistors.

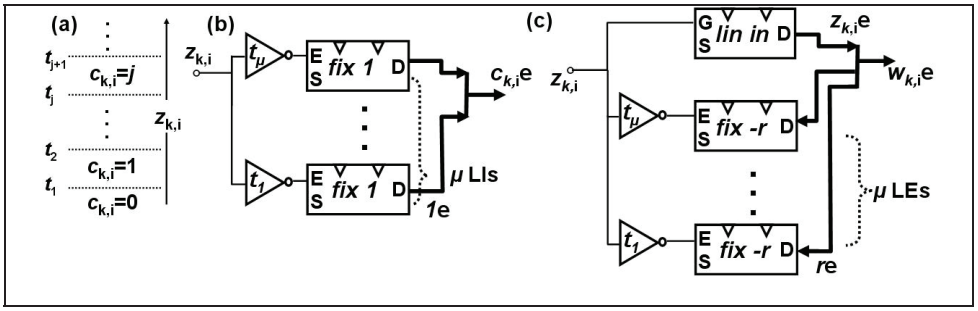


Fig. 7. (a) The threshold approach compares $z_{k,i}$ with comparison constants and obtains $c_{k,i}$ from comparison results. (b) Schematic of the M_c block and (c) Schematic of the M_w block.

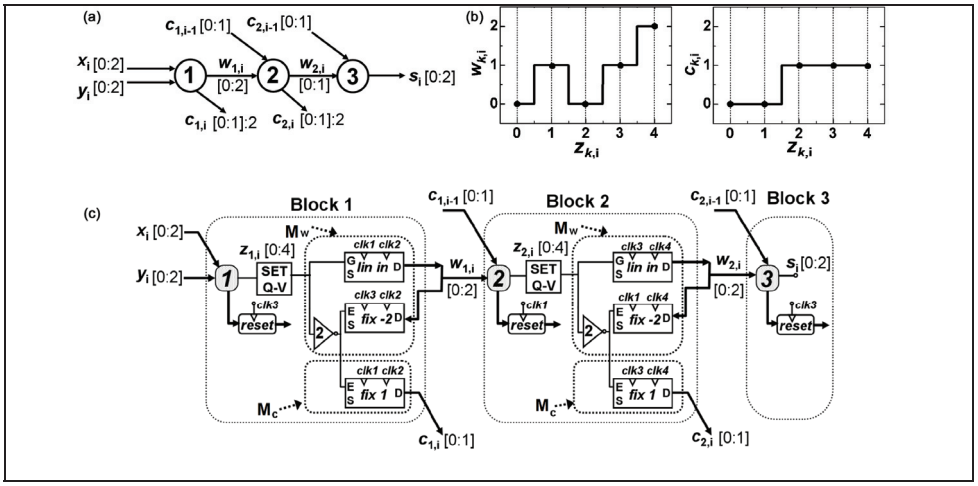


Fig. 8. (a) CTD representation of the BCS adder. The carries $c_{1,i}$ and $c_{2,i}$ belongs to interval [0:1]. (b) Characteristics of $w_{k,i}$ and $c_{k,i}$ as functions of $z_{k,i}$. (c) Schematic of the BCS adder circuit using threshold approach. The circuit consists of three circuit blocks.

The circuit requires 4 repulsive clock signals, *clk1, clk2, clk3* and *clk4*, as shown in Fig. 9(a). Each SE turnstile is associated with two clock signals. Fig. 9(b) shows the change in the number of electrons in the 2nd SN and 3rd SN with time. With the rising edge of *clk2*, $z_{1,i}$ electrons enter into the 2nd SN from the LI, as shown in Fig. 9(b). Also $c_{1,i-1}$ electrons are injected to 2nd SN by the (*i-1*)th adder. With the falling edge of *clk2*, electrons are ejected from the 2nd SN by the FE of the first block. Then $w_{1,i}$=2 electrons remain in the 2nd SN. After *clk2* the first block finishes its function. Similarly, with the rising edge of *clk4*, $w_{1,i}$ electrons enter into the 3rd SN from the LI and $c_{1,i-1}$ electrons enter into the 3rd SN from the FI. With the falling edge of *clk4*, electrons are ejected from the 3rd SN by the FE. After the falling edge of *clk4*, exactly s_i=1 electron remains in the 3rd SN, and the second block finishes its function. Electrons retain in the 2nd SN until the second block finished its function, and the electron number is reset to 0 by *clk1*. The circuit performs the addition operation by repeating the clock sequences.

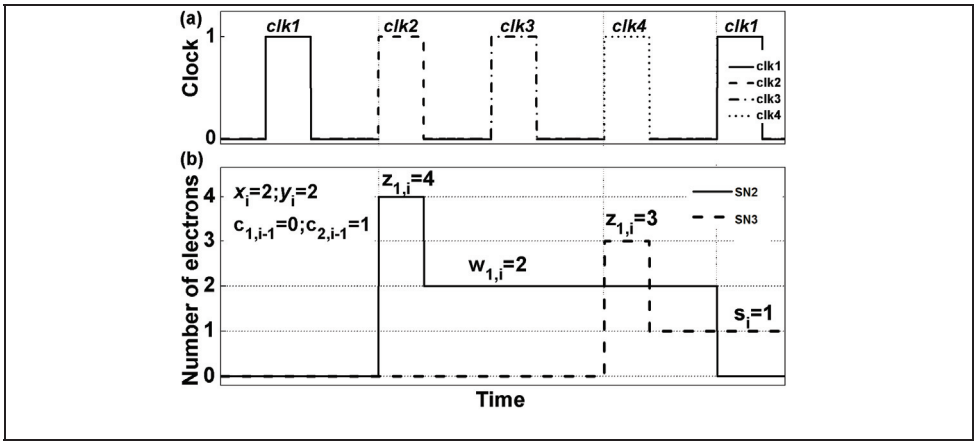


Fig. 9. (a) Clock sequences for the circuit. (b) Number of electrons in SN2 and SN3 as a function of time. After *clk4,* electrons number of SN3 corresponds to the final sum s_i.

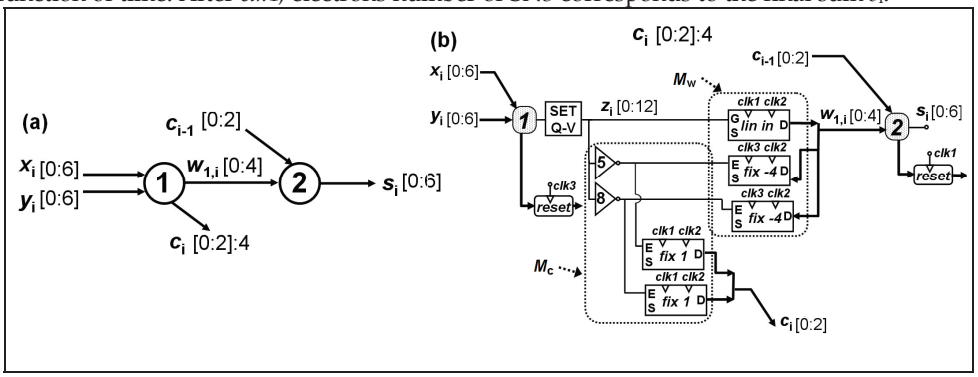


Fig. 10. (a) CTD of the SD4,3 adder. (b) Schematic of the SD4,3 adder circuit using threshold approach.

Example2: Design of a SD4,3 adder

The number system of the SD4,3 adder is the radix-4 signed-digit number system with a digit set [-3:3]. The SD4,3 adder has a redundancy index ρ=3 and it belongs to carry-free adders (Kawahito, 1988). However, in our circuit blocks the transfer of electrons is not bi-directional. To prevent complexity in circuit design, we use the digit set [0:6] to replace the former digit set [-3:3] of the SD4,3 algorithm. Fig. 10(a) shows the CTD of the SD4,3 algorithm:

$$\begin{gathered} z_{i.}=x_i+y_i; \\ 4c_{1,i}+w_{1,i}=z_i; \\ s_i=w_i+c_{1,i-1}. \\ w_{1,i}=z_i, \quad c_{1,i}=0, \quad \text{if } z_{1,i} \le 4; \\ w_{1,i}=z_i\text{-}4, \; c_{1,i}=1, \quad \text{if } 4 < z_{1,i} < 8; \\ w_{1,i}=z_i\text{-}8, \quad c_{1,i}=2, \quad \text{if } z_{1,i} \ge 8. \end{gathered} \tag{10}$$

Therefore $c_{1,i} \in [0:2]$ and μ=2. Fig. 10(b) shows the adder schematic. The M_w block consists of a LI, two inverters, and two FEs. The M_c block has two FIs. The logic threshold values of the two inverters are 4.5 and 7.5, respectively, according to (10). Since the circuit has only two stages, it requires 3 repulsive clock signals. The circuit performs radix-4 addition using 23 transistors.

4.2 Periodic Approach

The periodic approach realizes the M_w block by circuit elements that have *periodic* transfer functions. In other words, the number of electrons outputted by the M_w block is a periodic symmetric function (PSF) of its input. Assume $z_{k,i}$ belongs to the interval $[0:Z]$ and $c_{k,i}$ belongs to the interval $[0:\mu]$. The periodic approach requires that: (1) $u=[Z/r]$; (2) $c_{k,i}=j$ if $t_j \le z_{k,i} < t_{j+1}$, where $t_j=jr$, $0<j\le\mu$. In this case, $c_{k,i}$ and $w_{k,i}$ are given by:

$$c_{k,i}= [(z_{k,i}+1/2)/r],$$
$$w_{k,i}= z_{k,i}-rc_{k,i}. \quad (11)$$

We propose to use the unique transfer characteristics of the SE turnstile to realize the PSF circuit and thus to finish the fast addition algorithm compactly. Fig. 11(a) shows the implementation of the M_c block. The circuit consists of SE turnstiles and voltage dividers. When f=1, the number of electrons transferred N_f equals $z_{k,i}$. Fig. 11(b) shows this case. The division factor f of the voltage divider is designed to be r (r>1). The input voltage is added with $V_1=-(f-1)V_0/2$, so according to (2) the number N_f of transferred electrons is $|(z_{k,i}V_0+V_1)/fV_0+1/2|$. This value corresponds to $c_{k,i}$ according to (7). Fig. 11(c) and Fig. 11(d) show the relationships between N_f and $z_{k,i}$ when r=2 and r=4, respectively.

Fig. 11(e) shows the implementation of the M_w block. Generally, the circuit consists of two voltage adders, two voltage dividers with division ratio $f1$ and $f2$, respectively, one LI, and $f2/f1$ LEs. In the special case of $f1$=1 and $f2$=r, the circuit includes one voltage divider, one voltage adder, one LI, and r LEs. The LI injects N_{f1} electrons and each LE ejects N_{f2} electrons, and the overall M_w block injects $N=N_{f1}-f2N_{f2}$ electrons per cycle. Since N_{f2} corresponds to $c_{k,i}$, N directly corresponds to $w_{k,i}$ according to (8). Fig. 11(f) and Fig. 11(g) show the relationships between N and $z_{k,i}$ when $f2$=2 and $f2$=4, respectively. Both N_{f1} and $f2N_{f2}$ increase monotonically with $z_{k,i}$ and thus their difference N is a PSF of $z_{k,i}$. Fig. 11(h) shows another case when $f1$=2 and $f2$=4. In this case, N also has a periodic relationship with $z_{k,i}$. Actually a 3-bit A/D converter can be realized by combing the circuits represented by Fig. 11(d), Fig. 11(h) and Fig. 11(f). It can convert an input voltage to 3-bit binary signal $D_0D_1D_2$.

In summary, the periodic approach specifies the range of intermediate variables so that $c_{k,i}$ is a staircase function of $z_{k,i}$, and $w_{k,i}$ is a PSF of $z_{k,i}$. Since N_f is always a staircase function of $z_{k,i}$, the PSF circuits can be realized using the difference of the number of electrons transferred by the LI and LEs. Compared with conventional PSF circuits, the SE turnstile-based PSF circuit has much fewer transistors.

Example1: Design of a BCS adder

In the BCS adder using periodic approach, $z_{1,i}$ and $c_{1,i}$ belong to [0:4] and [0:2], respectively. Fig. 12(a) shows the BCS adder CTD representation that is suitable for the circuit structure in the periodic approach. Note that the only difference between Fig. 12(a) and Fig. 8(a) is the range of $c_{1,i}$ and $w_{1,i}$. In Fig. 12(a), $c_{1,i}$ belongs to [0:2] while in Fig. 8(a), $c_{1,i}$ belongs to [0:1]. Fig. 12(b) shows the relationship between $w_{k,i}$ and $z_{k,i}$. We see that $w_{k,i}$ is a periodic function of $z_{k,i}$. The circuit of the BCS adder using the periodic approach has the same structure as

Fig. 8(c), except for the realization of MVEs. Fig. 12(c) shows the circuit schematic of the first block. Since r=2, M_w consist of a voltage divider, a voltage adder, a LI and two LEs, while M_c consists of a LI. The circuit has 25 transistors.

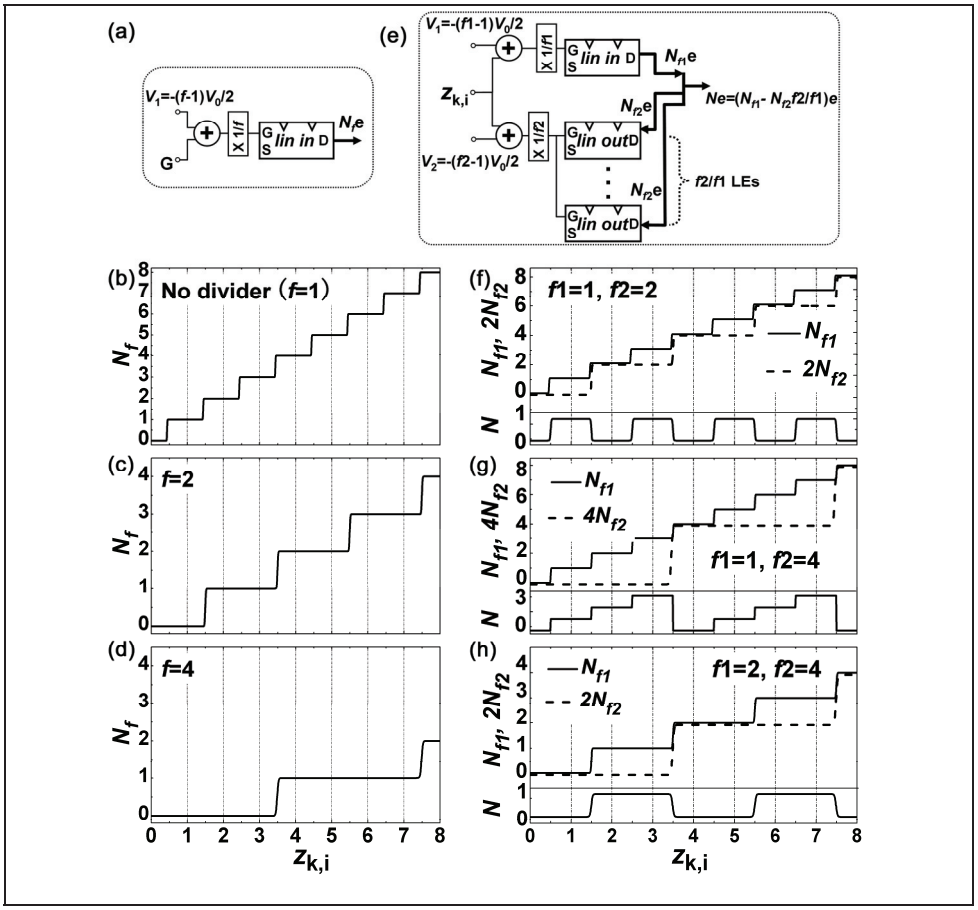


Fig. 11. (a) Schematic of the M_c block using the periodic approach. The circuit has a LI and a voltage divider. (b) Relationship between N_1 and $z_{k,i}$ when f=1 (no voltage dividers). (c) Relationship between N_2 and $z_{k,i}$ when f=2. (d) Relationship between N_4 and $z_{k,i}$ when f=4. (e) Schematic of the M_w block using the periodic approach. The circuit has a LI and $f2/f1$ linear ejectors. (f) Relationship between N and $z_{k,i}$ when $f1$=1, $f2$=2. (g) Relationship between N and $z_{k,i}$ when $f1$=1, $f2$=4. (d) Relationship between N and $z_{k,i}$ when $f1$=2, $f2$=4.

Fig. 8(a) and Fig. 12(a) represent the only two valid 3-stage BCS algorithms. The algorithm represented in Fig. 8(a) is suitable for the threshold approach, while the algorithm shown in Fig. 12(a) is suitable for the periodic approach. Traditional implementations of the BCS adder used the algorithm shown in Fig. 8(a). The algorithm shown in Fig. 12(a) leads to a more compact circuit by using the periodic approach.

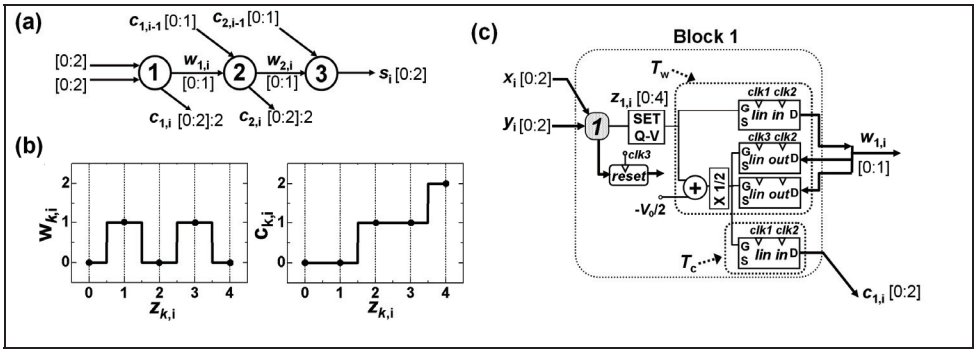


Fig. 12. (a) Another CTD representation of the BCS adder. The carries $c_{1,i}$ and $c_{2,i}$ belongs to interval [0:2]. (b) Characteristics of $w_{k,i}$ and $c_{k,i}$ as functions of $z_{k,i}$. (b) Schematic of the BCS adder circuit using periodic approach.

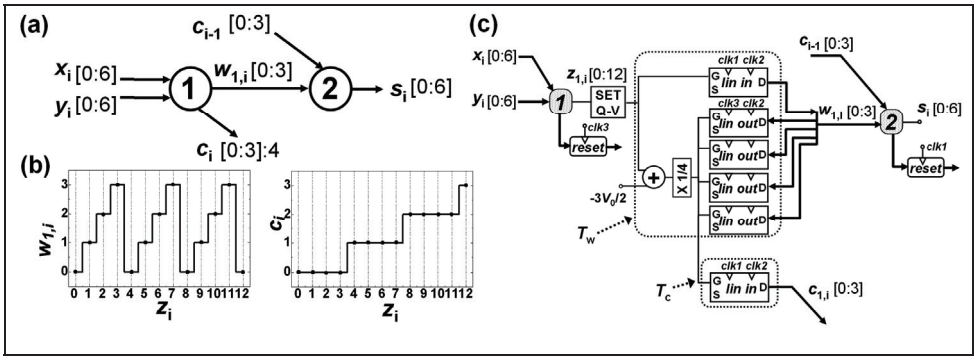


Fig. 13. (a) CTD representation of the SD4,3 adder using the periodic approach. (b) Characteristics of $w_{k,i}$ and $c_{k,i}$ as functions of $z_{k,i}$. (c) Schematic of the SD4,3 adder circuit using periodic approach.

Example2: Design of a SD4,3 adder

In the SD4,3 adder using the periodic approach, $z_{1,i}$ and $c_{1,i}$ belong to [0:12] and [0:3], respectively. Fig. 13(a) shows the other CTD of the SD4,3 adder. The only difference between Fig. 13(a) and Fig. 10(a) is the range of $c_{1,i}$ and $w_{1,i}$. In Fig. 13(a), $c_{1,i}$ belongs to [0:3] while in Fig. 10(a), $c_{1,i}$ belongs to [0:2]. Fig. 13(b) shows $w_{1,i}$ and $c_{1,i}$ as functions of $z_{1,i}$. In fact, the characteristics of $w_{1,i}$ are same as in Fig. 11(g), and the characteristics of $c_{1,i}$ are same as in Fig. 11(d). Fig. 13(c) shows the circuit schematic. Since r=4, the M_w block has one LI and 4 LEs, and the M_c block consists of a LI. Multiple-valued periodic $w_{1,i}$-$z_{1,i}$ characteristics are compactly achieved using the SE turnstiles. The circuit has only 17 transistors in all. Actually, Fig. 10(a) and Fig. 13(a) represent the only two valid SD4,3 algorithms. Conventional SD4,3 adders use the algorithm shown in Fig. 10(a). As far as we know, the algorithm shown in Fig. 13(a) has not been implemented by conventional approaches yet.

Example3: Design of a PD2,3 adder

The number system of the PD2,3 adder is the radix-2 positive-digit number system with a digit set [0:3]. The PD2,3 adder is a special adder in which more than 1 carries are transferred between two stages (Parhami, 1989). The CTD of the PD2,3 algorithm is shown in Fig. 14(a). The algorithm for the PD2,3 adder is:

$$\begin{aligned} z_i &= x_i + y_i; \\ 4c_{1,i} + 2c_{2,i} + w_{1,i} &= z_i; \\ s_i &= w_i + c_{1,i-2} + c_{2,i-1}. \end{aligned} \tag{12}$$

Since one circuit block outputs two carries $c_{1,i}$ and $c_{2,i}$, now three blocks M_{c1}, M_{c2}, and M_w are required. Fig. 14(b) shows the characteristics of $w_{1,i}$, $c_{1,i}$, and $c_{2,i}$. If z_i is converted to a 3-bit binary signal $D_0D_1D_2$, then $w_{1,i}$, $c_{2,i}$ and $c_{1,i}$ correspond to D_0, D_1 and D_2, respectively. So we can directly use the circuits represented by Fig. 11(h), Fig. 11(d) and Fig. 11(f) to implement the M_w, M_{c1} and M_{c2} blocks, respectively. Fig. 14(c) shows the circuit schematic. The M_w block has one LI and two LEs. The M_{c2} block has one LI and two LEs. The M_{c1} block has one LI. The circuit has 19 transistors. This algorithm is particularly suitable for the periodic approach since all intermediate variables are PSFs of z_i.

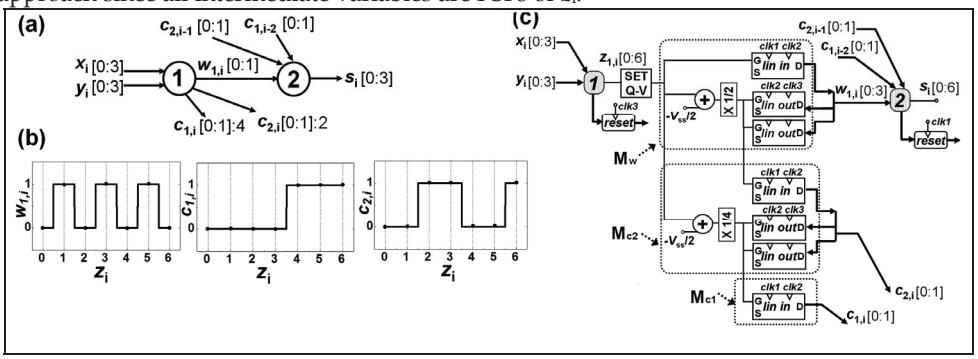


Fig. 14. (a) CTD representation of the PD2,3 adder. (b) Characteristics of $w_{1,i}$, $c_{1,i}$ and $c_{2,i}$ as functions of $z_{k,i}$. (c) Schematic of the PD2,3 adder using periodic approach.

4.3 Single-electron fast multiplier

We also used the above PD2,3 carry-free adders to construct a tree-structure single-electron fast multiplier. Fig. 15 shows the block diagram of an $X \times Y$ [0:11] multiplier. The circuit structure is same as the CMVL PD2,3 multiplier circuit (Kawahito et al., 1988). The multiplication algorithm has 4 steps. In the first step, 12×12 partial-product generator (PPG) generates 12×12 partial-products $p_{i,j}$ of multiplicand Y and multiplier X, $p_{i,j} \in \{0:1\}$. In our circuit the PPG is simply a SE turnstile. Only when both x_i and y_i is 1, a SE is injected to the SN. In step2, the output electrons of PPGs are grouped and transferred into SNs and are then are converted to voltage signals by the SET/MOS hybrid circuit. The SNs naturally finish the sum operation of partial-products. After this step, 4 operands with PD2,3 representation $\boldsymbol{P}_j' = (p'_{12,j} \ldots p'_{i,j} \ldots p'_{1,j} p'_{0,j})$ (j=0,1,2,3) are generated, where $p'_{i,j} = p_{i,3j} + p_{i-1,3j+1} + p_{i-2,3j+2}$ corresponds to the number of electrons stored in the SN. In step3, these operands are added by two levels of parallel single-electron PD2,3 adders. Finally in step 4, the outputs of the second level PD2,3 adders are converted by conventional PD-binary converters.

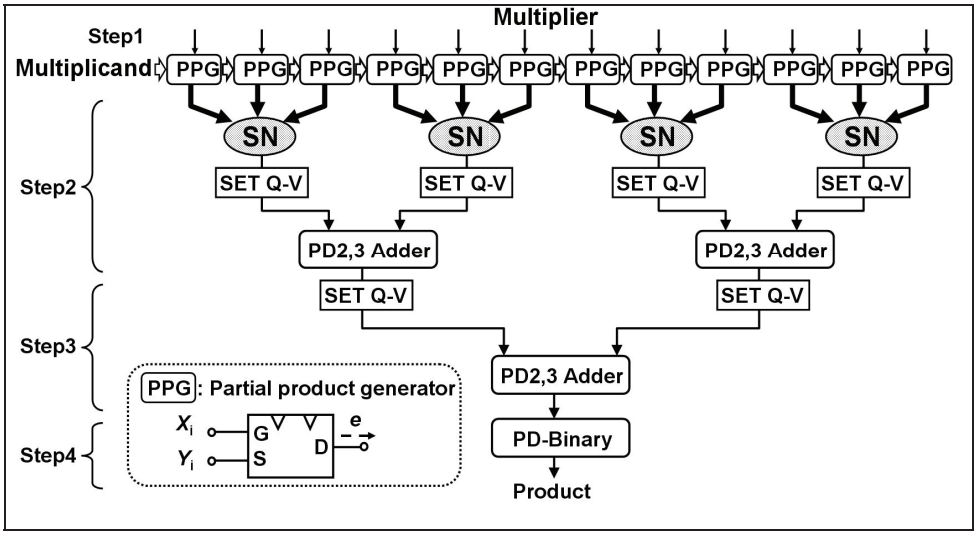


Fig. 15. Block diagram of the 12×12bit single-electron fast multiplier using two levels of single-electron PD2,3 fast adders. The bold lines show the flow paths of electrons.

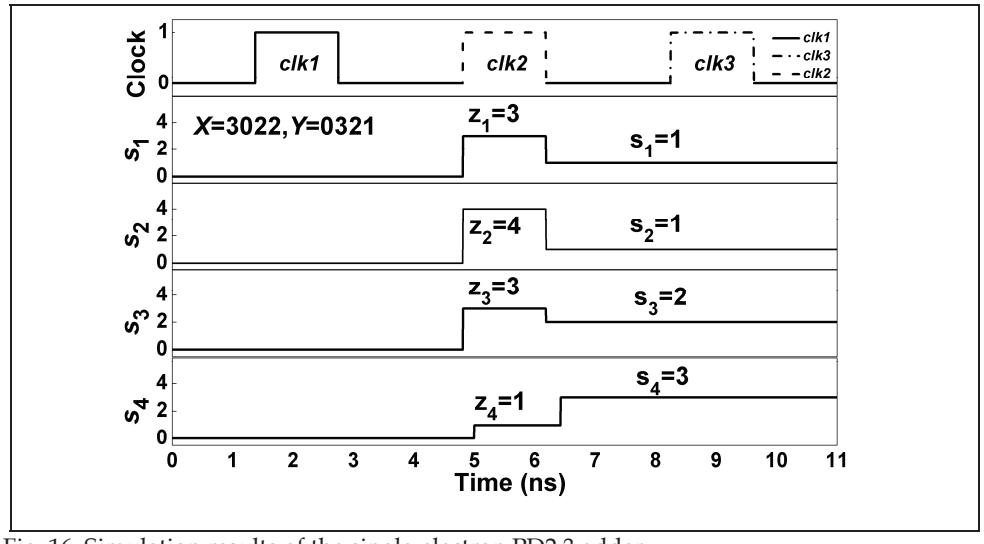


Fig. 16. Simulation results of the single-electron PD2,3 adder.

5. Circuit Performances

We use the HSPICE simulator and the proposed SPICE model of SE turnstile to investigate the performances of the adders. The SETs are described by the compact SPICE model of SET (Inokawa & Takahashi, 2003). The MOS transistors are described by 65nm technology node Predictive Technology SPICE models (Zhao & Cao, 2006).

5.1 Simulation Examples

Fig. 16 shows the simulated waveforms of the PD2,3 adder. The operation frequency f is 100MHz. The circuit finishes the addition function successfully. Correct operations of the other proposed adders were also verified.

5.2 Area

The adder circuits have compact structures and a small number of devices. This small device number is due to: 1) the use of electron counting paradigm to finish the sum operation and 2) the use of SE turnstiles to realize PSF circuits efficiently. Another advantage of the adders is that the transistors number does not depend on the radix of the number system. This can greatly reduce the number of transistors in large-size adders. For example, a 64-bit adder which is composed of 32 SD4,3 adders shown in Fig. 12(c) has only 608 transistors. On the other hand, one of the most area-efficient CMOS 64-bit adders has more than 1800 transistors (Ono, 2002), and the 64-bit SET adder has more than 1900 SETs (Lageweg et al., 2004) or 1136 SETs (Sulieman & Beiu, 2004).

5.3 Speed

The speed of the adders is limited by two factors. The first factor is that the inherent arithmetic operation frequency is limited by the transfer error rate of the SE turnstile. The electron transfer speed of the SE turnstile is very high. Unlike SETs, the MOSFET-based SE turnstiles have a CCD-like structure and they do not have static tunneling barriers when transferring single electrons. Therefore, the equivalent RC delay of the SE turnstile is very small compared to SETs. However, transfer error is inevitable in SE circuits. Transfer error in SE turnstile has many different origins, such as thermal error and dynamic error (Zimmermana et al., 2004). The thermal error rate is around 5e-10 with C_{SEB}=0.7aF at T=60K. The dynamic error rate dominates when the falling times of the clock signals are too small. We calculate the dynamic error rate. Considering a total error rate ε=1e-8 at T=60K, the falling time t_{fall} should be larger than 0.1ns (Zhang, 2008). Assuming the rising times of the clock signals are the same as t_{fall}, the transfer error rate limits the maximum frequency to $f_{max}=1/8t_{fall}$, which is higher than 1GHz. At lower temperature T=10K, f_{max} can be increased up to 5GHz.

The second factor is the delay of the charge-voltage converter. The delay of the SET/MOS hybrid circuit t_d is inverse proportion to the bias current I_0. Simulation results show that $t_d \approx 0.5$ns when $I_0 \approx 200$nA. In adders that require 4 repulsive clock signals, $f_{max}=1/4t_d$, which is around 500MHz.

The inherent speed of the fast adder is regardless of its length n. In practical applications, the fast adders may require extra conversion circuits to convert binary operands/results to non-binary operands/results. This will introduce an additional delay in the order of O($logn$). In applications consisting of multiple stages of adders like fast multiplier, several stages of adders work in non-binary mode in series. Only one conversion circuits is required to convert the final outputs to binary. In this case, the speed merit of fast adders can be fully exploited.

5.4 Power dissipation

The power dissipation can be grouped into two parts. The first part is the power dissipated by performing the *arithmetic operations*. Since all arithmetic operations are achieved by moving single electrons, the power dissipation of the arithmetic part is $W_{arth}=N_{arch}eV_{ss}f$, where N_{arch} is the number of electrons transferred in one operation cycle. At f=100MHz and V_{ss}=2V, W_{arth} of the periodic approach BCS adder is 0.67nW, and W_{arth} of the periodic approach SD4,3 adder is 0.85nW. The second part is the power dissipated by the conversion circuits W_{conv}, which is usually much larger than the W_{arth}. W_{conv} of the periodic approach BCS adder is 0.48uW, while W_{conv} of the threshold approach BCS adder is around 2uW due to the existence of threshold inverters.

5.5 Comparison with the state-of-the-art

Comparison of area

Table 1 shows comparison of the transistors numbers of several 1-bit *fast* adders implemented by the threshold approach, the periodic approach, the CMOS approach, and the CMVL approach. The SE fast adders have a small number of transistors. Table 2 shows comparison of circuit elements with other *binary* adders based on single-electron technology. One MOS transistor of the SE turnstile is treated as one tunneling junction of the SET. All coupling capacitors and load capacitors are also counted, since these capacitors may have larger area than the tunneling junctions. The results indicate that the proposed SE non-binary adder has similar number of circuit elements with other binary SE adders. Moreover, one SD4,3 adder can replace two binary adders due to its high radix.

Comparison of speed and power dissipation

Table 3 shows the comparison of several 32-bit PD2,3 fast adders using different approaches. Compared with CMOS, the proposed adder has a more than 98% reduction in power dissipation and more than 90% reduction in power-delay-product (PDP). The speed of the proposed adder is lower than its CMOS counterpart, due to the low speed of the charge-voltage converter. With the scaling down of the MOS transistors, the adder speed can be further increased.

Name	Threshold	Periodic	CMOS	CMVL
BCS	29	25	58	50
SD4,3	23	17	>80	52
PD2,3	29	19	>80	28

Table 1. Number of Transistors Comparison of Fast Adders

Adder Name	Junctions	Capacitors	Sum
TLG-SET (Sulieman and Beiu, 2005)	8	20	28
MAJ-SET (Sulieman and Beiu, 2005)	14	29	43
SE adder (Cotofana, 2005)	9	18	27
SD4,3 periodic	18	12	30
PD2,3 periodic	20	17	37

Table 2. Comparison of Number of Elements with Other Single-Electron Adders

Name	Delay	Power	PDP
CMOS	~0.4ns	~0.8mW	~0.3pJ
CMVL	~0.15ns	~3mW	~0.5pJ
This chapter	2ns	15uW	30fJ

Table 3. Number of Transistors Comparison of Fast Adders

Name	Delay	W_{arch} at f=100MHz	PDP		
	Typical value	Scale with	Typical value	Scale with	
TLG adder (Sulieman, 2004) V_{ss}=6.5mV, T=4K	~3ns	>O(logn)	~20pW	$nC_L V_{ss}^2$	~0.4eV
SE adder (Cotofana, 2005) V_{ss}=16mV	~20ns	O(n/r)*	~500pW	$n2^{r+1}eV_{ss}/r$	~65eV
PD2,3 this chapter V_{ss}=2V, T=60K	2ns	O(1)**	20nW	$nN_{arch}eV_{ss}$	250eV

* n is the adder length and r is a technology dependent constant.
** Additional O(logn) delay when output is converted to binary

Table 4. Comparison of Delay & Power with Other 32-bit Single-Electron Adders

Table 4 shows the comparison of delay and power dissipation with other 32-bit SE binary adders. Although the proposed adder performs high-radix non-binary addition algorithm, it still has higher speed than other SE adders.

In Table 4 we did not consider the power dissipation W_{conv} of the converter circuits. This is reasonable because practically all SE circuits require converters or amplifiers to obtain large voltage gain and to drive output loads. Therefore we only compare W_{arch} of the adders. The power of the TLG-SET adder (Sulieman, 2004) is much lower than others. However, this value is obtained with a V_{ss}=6.5mV and T=4K. To work at higher temperature, the junction capacitances of the SETs should be decreased, and thus V_{ss} has to be increased. Moreover, power of the CMOS-alike SET adder is proportional to the square of V_{ss}. It will be larger than the power of our PD2,3 adder when V_{ss} is larger than 160meV. On the other hand, W_{arch} of the SE adder by (Cotofana, 2005) and W_{arch} of our PD2,3 adder scale linearly with V_{ss} and they are proportional to the maximum total number of electrons N_{tot} involved in the arithmetic operation. The 32-bit SE adder by (Cotofana, 2005) has $N_{tot} \approx 2^{(r+1)}n/r \approx 2048$ when r=8. The 32-bit PD2,3 adder has $N_{tot}=nN_{arch}=672$. Therefore, the two adders have similar W_{arch} values with same V_{ss}. In this chapter, we use a 2V V_{ss} to obtain a large noise margin (V_0=0.3V) and we use experimental parameters of the SE turnstile in simulation. With lower V_{ss}=0.1V, lower temperature T=5K, and larger C_{seb}=15aF, the adder can also work well with a PDP=12eV. Therefore, with same temperature and V_{ss}, the PDPs of the proposed SE fast adders are not worse than previous SE binary adders.

In summary, the proposed SE fast adders have high overall time-area-power performances. The PDP of the fast adder is much lower than advanced CMOS adders (Zlatanovici, 2006). This point is rather critical in system-on-chip applications. Consider the 54x54 bit fast multiplier circuit based on radix-2 signed-digit arithmetic (Mochizuki, 2005). This CMVL circuit consumes more than 70mW with 500MHz operation frequency and it would be difficult to integrate hundreds of these multipliers in a single chip. However, using the SE fast adders to construct multipliers, the power dissipation will be largely reduced and thus thousands of multipliers can be integrated.

5.5 Comparison of two approaches

Each of the proposed approach has its merits and disadvantages. In some sense, the two approaches are complementary. Here we briefly compare the two approaches.

1) The threshold approach is more general. The periodic approach is better in adders in which more than 1 carry is transferred between two stages.

2) Number of transistors. The threshold approach has more transistors with small r, and the periodic approach has more transistors with large r.

3) Power dissipation. The threshold approach has much larger power dissipation due to existence of inverters.

6. Conclusion

In this chapter, we proposed novel fast adders based on single-electron devices. We proposed a family of SE transfer circuits based on the MOSFET-based SE turnstile. The adders can be designed by directly mapping the CTD representation of the addition algorithm to circuits using the threshold approach or the periodic approach. The two design approaches have their own advantages. The threshold approach can be viewed as a special kind of CMVL circuit with ultimate small current. The periodic approach uses the flexible staircase N-V_g characteristics of the SE turnstile to compactly realize PSFs. We used SPICE to simulate the adders. The high time-area-power performances of the adders were demonstrated. Parameter dispersions should be considered for practical implementation of single-electron circuits. For the adder circuits, methods to reduce the calculation error rate and fabrication-related analysis are important, which will be our future work. In conclusion, the single-electron fast adder circuits have simple structure, fast operation speed and low power dissipation, and are promising in future nanoscale information-processors.

7. References

Asahi, N.; Akazawa, M. & Amemiya, Y. (1997). Single-electron logic device based on the binary decision diagram. *IEEE Trans. Electrons Devices*, Vol. 44, No. 7, pp. 1109.

Avizienis, A. (1961). Signed-Digit Number Representations for Fast Parallel Arithmetic. *IRE Trans. Electronic Computers*, Vol. 10, Sept. 1961, pp. 389-400

Cotofana, S.; Lageweg, C. & Vassiliadis, S. (2005). Addition Related Arithmetic Operations via Controlled Transport of Charge. *IEEE Trans. Comput.*, Vol. 54, No. 3, pp. 243-256.

Fujiwara, A.; Zimmerman, N. M.; Ono, Y. & Takahashi, Y. (2004). Current quantization due to single-electron transfer in Si-wire charge-coupled devices. *Appl. Phys. Lett.* Vol. 84, pp. 1323.

Fujiwara, A.; Nishiguchi, K.; & Ono, Y. (2008). Nanoampere charge pump by single-electron ratchet using silicon nanowire metal-oxide-semiconductor field-effect transistor. *Appl. Phys. Lett.* Vol. 92, pp. 042102.

Gonzalez, F. A. & Mazumder, P. (1998). Multiple-valued Signed-digit Adder Using Negative Differential Resistance Devices. *IEEE Trans. Comput.*, Vol. 47, No. 9, pp. 947 – 959.

Homma, N.; Sakiyama, J.; Wakamatsu, T.; Aoki, T. & Higuchi, T. (2004). Systematic Approach for Analyzing Fast Addition Algorithms Using Counter Tree Diagrams. *Proceedings of the 2004 International Symposium on Circuits and Systems*, vol. 5, pp. 197-200, May. 2004., IEEE, Vancouver

Inokawa, H.; Fujiwara, A. & Takahashi, Y. (2003). A Multiple-valued Logic and Memory with Combined Single-electron and Metal-oxide-semiconductor Devices. *IEEE Trans. Electron Devices*, Vol. 50, No. 2, pp. 462-470.

Inokawa, H. & Takahashi, Y. (2003). A Compact Analytical Model for Asymmetric Single-Electron Tunneling Transistors. *IEEE Trans. Electron Devices*, Vol. 50, pp. 455-461.

Kawahito, S.; Kameyama, M.; Higuchi, T. & Yamada, H. (1988). A 32 x 32-bit Multiplier Using Multiple-valued MOS Current-mode Circuits. *IEEE J. Solid-State Circuits*, Vol. 23, No. 1, pp. 124 – 132.

Kawahito, S.; Ishida, M.; Nakamura, T.; Kameyama, M.; & Higuchi, T. (1994) High-speed Area-efficient Multiplier Design Using Multiple-valued Current-mode Circuits. *IEEE Trans. Comput*, Vol. 43, pp. 34-42.

Lageweg, C.; Cotofana, S. & Vassiliadis, S. (2004). Binary Addition Based on Single Electron Tunneling Devices. *Proceedings of 4th IEEE Conf. Nanotechnology*, pp. 327–330, Aug. 2004, IEEE, Munich.

Likharev, K. K. (1999). Single-electron devices and their applications. *Proceeding of the IEEE.*, Vol. 87, No. 4, pp. 606-632.

Mochizuki, A. & Hanyu, T. (2005) A 1.88ns 54×54-bit Multiplier in 0.18μm CMOS Based on Multiple-valued Differential-pair Circuitry. *Digest of Technical Papers, 2005 IEEE Symposium on VLSI Circuits*, vol. 17-3, pp. 264-267, June 2005.

Nishiguchi, K.; Fujiwara, A.; Ono, Y.; Inokawa, H. & Takahashi, Y. (2006). Room-temperature-operating Data Processing Circuit Based on Single-electron Transfer and Detection with Metal-oxide-semiconductor Field-effect transistor Technology. *Appl. Phys. Lett.* Vol. 88, pp. 183101.

Ono, Y.; Takahashi, Y; Yamazaki, K.; Nagase, M.; Namatsu, H.; Kurihara, K. & Murase, K. (2000). Si complementary single-electron inverter with voltage gain. *IEEE Trans. Electron Devices*, Vol. 47, pp. 147.

Ono, Y.; Inokawa, H. & Takahashi, Y. (2002). Binary Adders of Multi-gate Single-electron Transistor: Specific Design Using Pass-transistor Logic. *IEEE Trans. Nanotechnol.*, Vol. 1, No. 3, pp. 93–99.

Ono, Y.; Fujiwara, A.; Nishiguchi, K.; Inokawa, H. & Takahashi, Y. (2006). Manipulation and detection of single electrons for future information processing. *J. Appl. Phys.*, Vol. 97, pp. 031101.

Parhami, B. (1989) A New Method for Designing Highly Parallel Binary Multipliers," *Proceeding of the Third Ann. Parallel Processing Symp.*, pp. 176-185, Mar. 1989.

Parhami, B. (1990). Generalized Signed-digit Number Systems: A Unifying Framework for Redundant Number Representations. *IEEE Trans. Comput.*, Vol. 39, No. 1, pp. 89 – 98.

Parhami, B. (1993). On the Implementation of Arithmetic Support Functions for Generalized Signed-Digit Number Systems. *IEEE Trans. Comput.*, Vol. 42, No. 3, pp. 379 – 384.

Sakiyama, J.; Aoki, T. & Higuchi, T. (2003). Counter Tree Diagrams for Design and Analysis of Fast Addition Algorithms. *Proceedings of the 33rd IEEE International Symposium on Multiple-Valued Logic*, pp. 91 - 98, May. 2003., IEEE.

Sulieman, M.H. & Beiu, V. (2004). Characterization of a 16-bit Threshold Logic Single Electron Technology Adder. *Proceedings of the 2004 International Symposium on Circuits and Systems*, pp. III- 681-4, May. 2004, IEEE, Vancouver

Sulieman, M.H. & Beiu, V. (2005) On Single-electron Technology Full Adders. *IEEE Trans. Nanotechnol.*, Vol. 4, No. 6, pp. 669-680.

Takagi, N.; Yasuura, H.; & Yajima, S. (1985). High-speed VLSI Multiplication Algorithm with a Redundant Binary Addition Tree. *IEEE Trans. Comput.*, Vol. 34, No. 9, pp. 789 - 796.

Yano, K.; Ishii, T.; Hashimoto, T.; Kobayashi, T.; Murai, F. & Seki, K. (1994). Room-temperature single-electron memory. *IEEE Trans. Electron Devices*., Vol. 41, pp. 1628–1638.

Zhang, W-C.; Nishiguchi, K.; Ono, Y.; Fujiwara, A.; Hiroshi, Y.; Inokawa, H.; Takahashi, Y. & Wu, N-J. (2007 a). Transfer and Detection of Single Electrons using Metal-Oxide-Semiconductor Field-Effect-Transistors. *IEICE Trans. Electron*, Vol. E90-C, pp. 943-948.

Zhang, W-C. & Wu, N-J. (2007 b). Novel Hybrid Voltage Controlled Ring Oscillators Using Single Electron and MOS Transistors. *IEEE Trans. Nanotechnol*., Vol. 6, pp. 146-157.

Zhang, W-C. & Wu, N-J. (2007 c). A Novel Hybrid Phase-locked-loop Frequency Synthesizer Using Single Electron Devices and CMOS Transistors. *IEEE Trans. Circuits Syst. I,* Vol. 54, No. 11, pp. 2516-2527.

Zhang, W-C.; Wu, N-J. & Yang, F-H. (2007 d). Compact Universal Logic Gates realized using Quantization of Current in Nanodevices. *Nanotechnology*. Vol. 18, pp. 495201.

Zhang, W-C. & Wu, N-J. (2008). Smart Universal Multiple-Valued Logic Gates by Transferring Single Electrons. *IEEE Trans. Nanotechnol.* Vol. 7, No. 4, pp. 440-450.

Zhao, W. & Cao, Y. (2006). New generation of Predictive Technology Model for sub-45nm design exploration. *IEEE International Symposium on Quality Electronic Design,* pp. 585-590, IEEE, San Jose.

Zimmermana, N.M.; Hourdakis, E.; Ono, Y.; Fujiwara, A. & Takahashi, Y. (2004) Error Mechanisms and Rates in Tunable-barrier Single-electron Turnstiles and Charge-coupled devices. *J. Appl. Phys.*, Vol. 96, pp. 5254.

Zlatanovici Kao, S. & Nikolic, B.R. (2006) A 240ps 64b Carry-lookahead Adder in 90nm CMOS. 2006 *Digest of Technical Papers, IEEE International Solid-State Circuits Conference*, pp. 1735 – 1744, Feb. 2006.

16

Single-Electron Circuits for Sigma-Delta Domain Signal Processing

Hisato Fujisaka
Hiroshima City University
Japan

1. Introduction

Contemporary micro-electronics fabrication technology downsizes electronic circuits to the nanometer scale. In nanometer-scale electronic circuits, physical quantities are quantized and device behaviour is discontinuous. The following three physical phenomena are examples of the quantization and the discontinuity:

Single-electron tunnelling (Grabert & Devoret, 1992): Figure 1(a) shows a physical system consisting of two electrodes and an island. They are capacitively coupled. Suppose that the left gap of the island is so narrow that electrons can tunnel. When voltage Vg is applied to the right electrode and the left electrode is grounded, electrons tunnelled from the left electrode accumulate on the island. The equivalent circuit of the physical system is shown in Fig. 1(b). The two coupling capacitances are denoted by Ct and Cg. Suppose further that capacitances Ct and Cg are small enough to satisfy

$$\frac{e^2}{C_t + C_g} >> k_B T \,, \tag{1}$$

where e, k_B, and T denote elementary charge, Boltzmann constant, and temperature. The above inequality means that a small number of excess electrons can remain on the island withstanding collision with phonons. Then the negative charge on the island cannot be regarded as continuous any more. As voltage Vg increases, the number n of electrons on the island increases. However, small increase of Vg do not necessarily increase n because of repulsive force between electrons, which is called Coulomb blockade. Figure 1(c) shows the relation between Vg and n. The figure implies that only a single-electron tunnels at a moment.

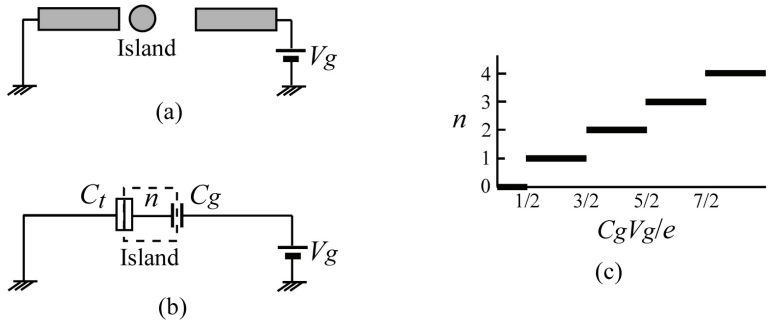


Fig. 1. Coulomb blockade and single-electron tunneling

Single-flux quantum (Tinkham, 1975): Figure 2(a) shows a superconductive ring in magnetic field H. The ring is cooled under its critical temperature. Magnetic flux also exists in the air surrounded by the ring because of diamagnetic current in the ring. Even after magnetic field H is switched off, the ring holds magnetic flux ϕ as shown in Fig 2(b). The wave function of electrons distributes so that it satisfies a periodic boundary condition of the ring. The condition quantizes the flux . Figure 2(c) shows the relation between magnetic field H and magnetic flux . In the figure, ϕ 0 $=h/e$, h: Planck constant. This quantization can be seen even if the ring is macroscopic in scale.

Point contact (van Wees et al., 1988): A heterojunction between GaAs-AlAs and GaAs layers makes a two-dimensional electron system as shown in Fig. 3(a). Two electrodes with a y-directional gap narrower than mean free path of electrons are made on the surface. When negative voltage Vg is applied to the electrodes, no electron exists under the electrodes and the two-dimensional system is divided into two parts. The gap between the electrodes makes a narrow path of electrons between the two parts. The path is called a point contact. In the contact, wave function of electrons distributes nonuniformly as shown in Fig. 3(b). Then the y-directional kinetic energy of electrons is quantized. This quantization also leads to the quantization of the x-directional conductance G of the electron system. The conductance varies depending on electrode voltage Vg but it takes quantized values as shown in Fig. 3(c).

Like these phenomena, both passive and active nanometer scale devices behave discontinuously and signals are no more continuous in nanometer-scale circuits. In near-future nano-electronics era, although analog signal processing (ASP) based on quantum wave interference (Sols et al., 1989) will find a niche for itself, present type of ASP is considered to be destined to decay.

Nanoelectronic devices are seriously vulnerable to electronic magnetic interferences, atmospheric neutrons, and varying environmental conditions such as temperature. In Nyquist-rate multi-bit digital signal processing (DSP) circuits built of nanoelectronic devices, large surge noise is caused in internal signals and in outputs when the device errors lead to the reversal of the sign-bits or significant bits of the signals. Irregularly structured nanoelectronic circuits are difficult to construct because special technology for nanoscale fabrication such as self-organization of quantum dots cannot be used. Long routing lines among nanoelectronic devices badly affect device behavior because of parasitic components of the lines. Regularly arranged and locally connected cellular array structure is required to

the architecture of nanoelectronic signal processing circuits. Conventional circuit techniques for Nyquist-rate multi-bit DSP do not always satisfy the requirements.

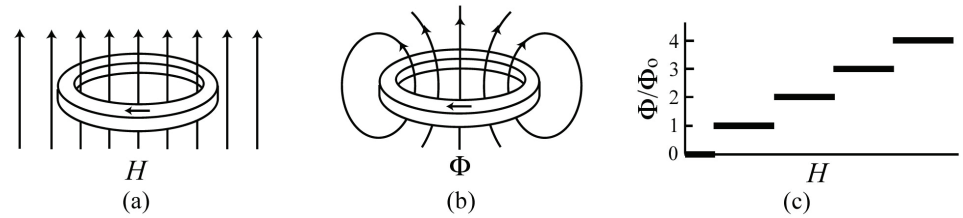


Fig. 2. Quantization of magnetic flux

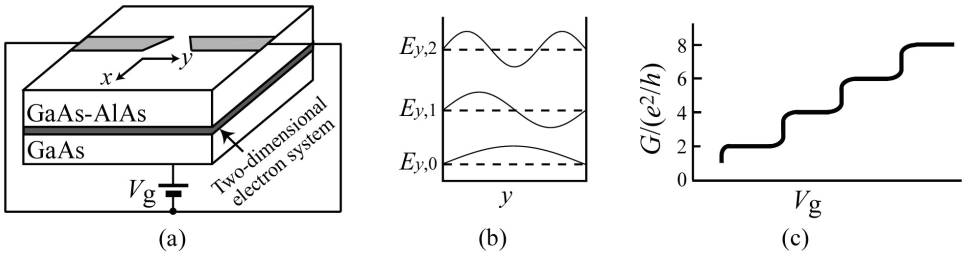


Fig. 3. Quantization of conductance

Pulsed signal processing (PSP) has the potential to grow up to be the new nanoelectronic signal processing. Its signal form is a stream of positive and negative pulses whose average represents signal level, as shown in Fig. 4. As later sections of this chapter will show, PSP will be developed to be an ASP-like signal processing and inherit the properties of ASP. It also will be seen in the later sections that the PSP scheme will solve the above-mentioned problems of Nyquist-rate multi-bit DSP.

Several types of PSP schemes have been proposed in various areas such as communications and controls. Stochastic computing (Gaines, 1969), (Hori et al., 2006) is one of the PSP schemes. Bernoulli sequences used in stochastic computing have uniformly distributed frequency spectra of quantization errors. Pulsed signals with large quantization spacing contain large power of quantization errors. Therefore, the quantization error spectra of Bernoulli sequences are high in signal band. Thus the sequences are not high quality pulse streams. Sigma-delta domain signal processing (SDSP) (Fujisaka et al., 2002), (Fujisaka et al., 2003) is another PSP scheme. Its signal form is sigma-delta (SD) modulated signals (Candy & Temes, 1992) whose quantization error components are small in signal band. Then, compared with stochastic computing, SDSP is a high quality signal processing.

This chapter is devoted to the principles and the design of SDSP circuits with single-electron tunneling (SET) devices (Grabert & Devoret, 1992). The rest of this chapter is organized as follows: In section 2, a basic SD modulation circuit and the characteristics of SD modulated signals are presented. Section 3 describes the concept, architecture and distinctive features of SDSP. In section 4, logic circuits built of SET devices are presented. In section 5, arithmetic and piecewise linear circuit modules for SDSP are built of the logic gates introduced in

section 4. In section 6, linear and nonlinear filters are constructed by using the circuit modules as examples of nanoelectronic SDSP circuits.

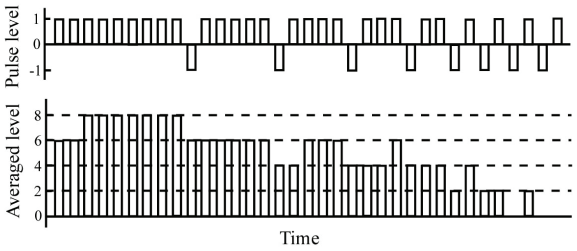


Fig. 4. A pulsed signal and its local time average

2. Sigma-Delta Modulation

Sigma-delta modulation (Candy & Temes, 1992) is applied to analog-to-digital and digital-to analog conversions. Figure 5(a) shows first-order SD modulators. The two modulators are equivalent. Although the lower diagram shows the popular structure of the SD modulator, we will use the upper diagram to illustrate the modulation. The block Q in the SD modulator performs one or a few-bit quantization. In this article SD modulated signals are limited to bit-streams of one-bit pulses with height $\pm\Delta/2$, that is $y(n) \in \{+\Delta/2, -\Delta/2\}$, Δ: quantization spacing.

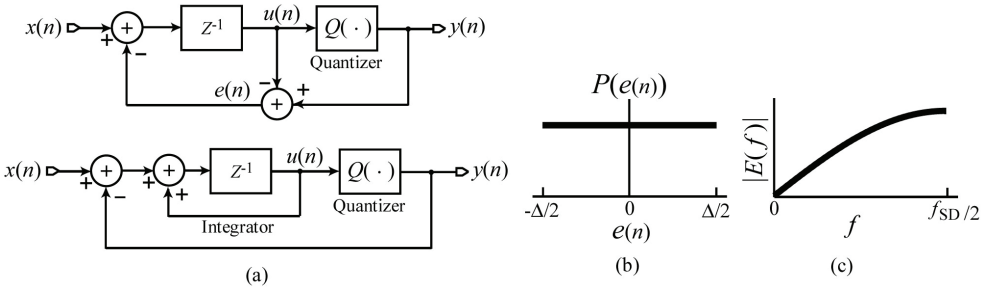


Fig. 5. Sigma-delta modulator and its noise characteristics

The bit-rate f_{SD} of the streams is higher than Nyquist-rate f_N. The ratio f_{SD}/f_N is referred as oversampling ratio (OSR). Let quantization error be defined with the variables in Fig. 5 as

$$e(n) = y(n) - u(n). \tag{2}$$

The probability distribution of the error is assumed to be uniform as Fig. 5(b) shows. Then the variance of the error, that is its power, is given by

$$e_{rms}{}^2 = < e(n)^2 > = \frac{\Delta^2}{12}. \tag{3}$$

The first-order SD modulated signal $y(n)$ contains the error in the following difference form:

$$y(n) = x(n-1) + e(n) - e(n-1). \tag{4}$$

The difference of the error $e(n)$-$e(n$-1) is called first-order SD modulation noise. Assume further that the quantization error $e(n)$ is white noise. Then the spectral density $E(f)$ of the SD modulation noise is given by

$$E(f) = e_{rms} \frac{\left|1 - \exp(-j2\pi f / f_{SD})\right|}{\sqrt{f_{SD}/2}} = \Delta \sqrt{\frac{2}{3 f_{SD}}} \sin(\pi f / f_{SD}), \quad j^2 = -1. \tag{5}$$

Figure 5(c) shows the spectral density of the SD modulation noise. The spectrum is low in low frequency band in which signal components locate. Thus SD modulation signals with high OSR are high quality. Conversion of original white noise $e(n)$ to its difference $e(n)$-$e(n$-1) is called noise shaping.

Sigma-delta modulators built of nanoelectronic devices have been developed (Yokoyama et al., 2001), (Chibashi et al., 2004). A circuit shown in Fig. 6 is a first-order SD modulator proposed in (Chibashi et al., 2004). It is a continuous-time correspondence to the lower SD modulator in Fig. 5(a). It employs resonant-tunneling diodes (RTD). By exploiting the negative resistance of RTDs, a monostable-bistable transition logic element (MOBILE) (Maezawa, 1994) is composed with a pair of RTDs. A D-type flip flop is built of a MOBILE and an FET transistor. The components of the modulator circuit work in the following way: Transistors Tr1 and Tr2 convert the voltage difference between input *Vin* and *Vout* to a current. Capacitance *C* integrates the current. A D-type flip flop consisting of a MOBILE and transistor Tr3 operates as a quantizer with a unit-delay element. Then, we find that the circuit operates as a SD modulator.

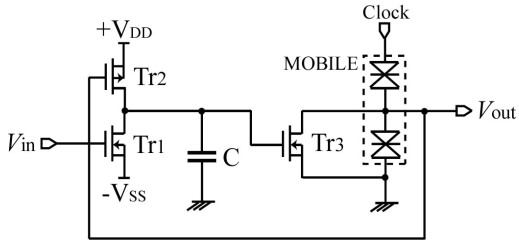
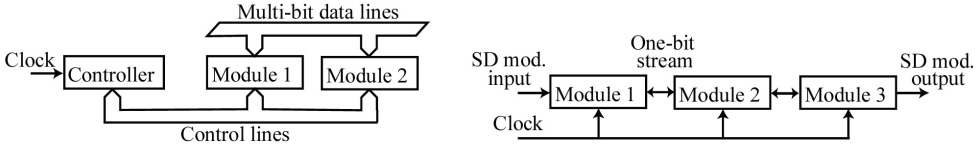


Fig. 6. A first-order sigma-delta modulator using resonant-tunneling diodes

3. Sigma-Delta Domain Signal Processing

As mentioned in section 1, SDSP (Fujisaka et al., 2002), (Fujisaka et al., 2003), (Katao et al., 2007), (Hayashi et al., 2007) is a PSP scheme employing SD modulated signal forms. Figure 7 shows the architecture of Nyquist-rate multi-bit DSP and SDSP systems. In Nyquist-rate multi-bit DSP circuits built of nanoelectronic devices, large surge noise is caused in internal signals and in outputs when transient device errors lead to the reversal of the sign-bits or significant bits of the signals. On the other hand, a few bits of errors in a SD modulated bit-stream slightly decrease its signal quality but do not cause large surge because the signal level is represented by the average of the bit-stream in a time window. Even one transient device error in the sequence controller of Nyquist-rate multi-bit DSP systems causes chain-

reaction errors and leads the systems to failure. In SDSP systems, all the circuit modules are driven by a clock synchronized to the SD modulated bit streams. The modules complete their tasks within each one period of the clock. Thus, unlike conventional processors, SDSP processors need no operation sequence controller. Therefore, SDSP systems never go out of control by transient device errors. The pulsed signal form and the controller-less architecture equip SDSP systems with fault tolerance of transient device errors.



(a) Conventional DSP architecture (b) SDSP architecture

Fig. 7. The architecture of signal processors

Stochastic computing (Gaines, 1969), (Hori et al., 2006) mentioned briefly in section 1 uses binary Bernoulli sequences $x(n) \in \{+\Delta/2, -\Delta/2\}$ as signals. The advantage of stochastic computing is the simplicity of processing circuits. The signal level is represented by

$$\overline{x(n)} = \left\{ \mathrm{Prob}(x(n) = +\frac{\Delta}{2}) - \mathrm{Prob}(x(n) = -\frac{\Delta}{2}) \right\} \frac{\Delta}{2}. \tag{6}$$

Figure 8(a) shows the probability distribution of the quantization error $e_{SC}(n)$ defined by

$$e_{SC}(n) = x(n) - \overline{x(n)}. \tag{7}$$

Then the power of error $e_{SC}(n)$ is given by

$$e_{SC,rms}{}^2 = < e_{SC}(n)^2 > = \frac{\Delta^2}{6}. \tag{8}$$

Because signal $x(n)$ is a Bernoulli sequence, error $e_{SC}(n)$ is considered as white noise. Then, the spectral density $E_{SC}(f)$ of $e_{SC}(n)$ is given by

$$E_{SC}(f) = e_{SC,rms} \frac{1}{\sqrt{f_{SC}/2}} = \Delta \sqrt{\frac{1}{3 f_{SC}}}, \tag{9}$$

where f_{SC} is the bit rate of the signal $x(n)$. Figure 8 compares spectral density between error $e_{SC}(n)$ in stochastic computing and SD modulation noise $e_{SD}(n)$ in SDSP. As signal components locate in low frequency band, SD modulated signals have higher quality than Bernoulli sequences. Figure 9 shows the power of the components of $e_{SC}(n)$ and $e_{SD}(n)$ located in signal band $[0, f_N]$. The power decreases by -3dB and -9dB when OSR is doubled in stochastic computing and SDSP respectively. The figure implies that SDSP can perform higher quality signal processing at lower speed than stochastic computing.

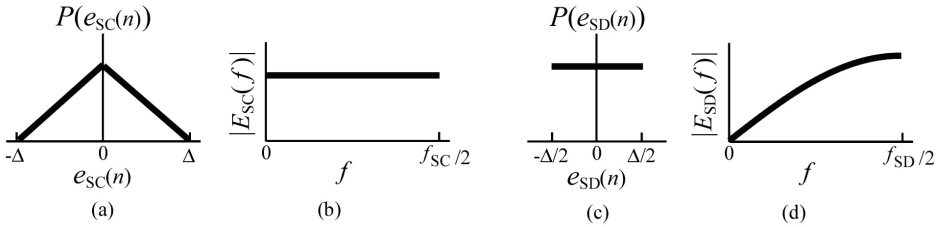


Fig. 8. Error characteristics of signals in stochastic computing and SDSP

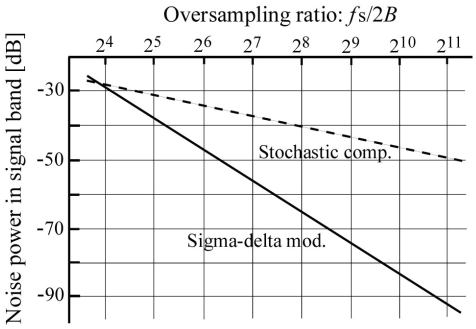


Fig. 9. The quality of signals in stochastic computing and SDSP

4. Logic Circuits based on Single-Electron Tunnelling

Coulomb blockade and single-electron tunneling phenomena mentioned in section 1 make it possible to confine excess n (: integer) electrons in a nanoscale island. Logic gates exploiting the phenomena are introduced in this section. In most of the logic gates, the number of excess electrons in every island of the logic gates is $n \in \{0, 1\}$ or $\{0, -1\}$.

4.1 Linear Threshold Gates

Consider a linear threshold gate (LTG) (Lewis & Coates, 1967) expressed by

$$y = \mathrm{sgn}(f(X)) = \begin{cases} 1, & \text{if } f(X) \geq 0, \\ -1, & \text{if } f(X) < 0, \end{cases}$$

$$f(X) = \sum_{i=1}^{N} \varpi_i x_i - \theta, \quad X = (x_1, \ x_2, \ \cdots, \ x_N). \tag{10}$$

The LTG is constructed of SET devices as shown in Fig. 10(a) (Lageweg et al., 2001). The circuit possessing two islands P- and P+ takes the following two states.

State 1: Islands P- and P+ respectively hold one excess hole and one excess electron.

State 0: Both islands P- and P+ hold no excess charge.

When the circuit is in state 1/0, the output level is high/low. In state 0, voltage V_j across the tunneling junction between P- and P+ is given by

$$V_j = \sum_{i=1}^{N+} w_{i,+} v_{in,i,+} - \sum_{i=1}^{N-} w_{i,-} v_{in,i,-} + w_b V_b,$$
$$w_{i,+} = C_{\Sigma-} C_{i,+} / C_r^2, \quad w_{i,-} = C_{\Sigma+} C_{i,-} / C_r^2, \quad w_b = C_{\Sigma-} C_b / C_r^2,$$
$$C_{\Sigma+} = C_b + \sum_{i=1}^{N+} C_{i,+}, \quad C_{\Sigma-} = C_o + \sum_{i=1}^{N-} C_{i,-}, \quad C_r^2 = C_{\Sigma+} C_j + C_{\Sigma+} C_{\Sigma-} + C_j C_{\Sigma-}. \tag{11}$$

The circuit gets into state 1 when Vj is greater than critical voltage Vc given by

$$Vc = (C_{\Sigma+} + C_{\Sigma-}) e / 2C_r^2. \tag{12}$$

Then it is found that the circuit behaves as LTG described by Eq. (10).

The LTG circuits work as Boolean logic gates. A circuit shown in Fig. 10(b) operates as AND or OR gate when its threshold level is set at a high or a low level. A circuit shown in Fig. 10(c) operates as NOR or NAND gate depending on its threshold level. Circuit parameters to make the LTGs operate as various Boolean logic gates are shown in Tab. 1.

4.2 Logic Circuits based on Binary Decision Diagram

A binary decision diagram (BDD) is a directed graph representing a Boolean logic function (Miller et al., 2006). Examples of BDD expression are shown in Fig. 11. All the paths from the top of the graph to terminal-1/0 assign values of variables x_1, x_2, …, x_N that make Boolean function true/false.

Logic circuits based on BDD can also be constructed by using SET devices (Asai et al., 1997). In BDD-based logic circuits, paths are represented by series of islands located like stepping-stones. In the logic circuits, a single-electron entered at the top is transferred from one island to another through SET junctions. A basic circuit transferring a single-electron is shown in Fig. 12(a). The circuit is named single-electron pump (Pothier, 1992). Let terminal voltages be fixed at $V_1 = 0$ and $V_2 = 0$. When the two islands hold n_1 and n_2 excess electrons, the electrostatic energy $E(n_1, n_2)$ of the circuit is given by

$$E(n_1, n_2) = \frac{1}{6C_j}\left\{(n_1 e - C_g V_{g1})^2 + (n_2 e - C_g V_{g2})^2 + \left((n_1 + n_2)e - C_g(V_{g1} - V_{g2})\right)^2\right\}, \tag{13}$$

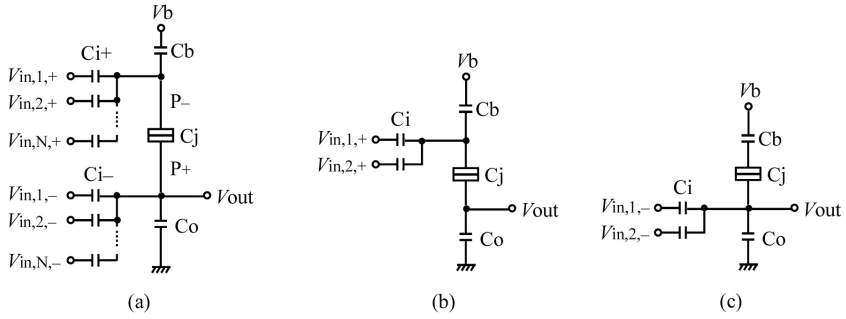


Fig. 10. Linear threshold gates built of SET devices

Parameters \ Gates	OR	AND	NOR	NAND	4-input AND
*V*in High, *V*in Low	0V, 16mV	0V, 16mV	0V, 16mV	0V, 16mV	0V, 16mV
*V*b	16mV	16mV	16mV	16mV	16mV
Cb	10.1aF	10aF	9.9aF	10.1aF	9.7aF
Ci	0.1aF	0.1aF	0.1aF	0.1aF	0.1aF
Co	10aF	10aF	10aF	10aF	10aF
Cj	0.1aF	0.1aF	0.1aF	0.1aF	0.1aF

Table 1. Circuit parameters of LTG-based Boolean logic gates

where C_j is the capacitance of the tunneling junctions, C_g is the gate capacitance, and V_{g1} and V_{g2} are the gate voltages. Figure 12(b) shows the region in which energy $E(n_1, n_2)$ becomes low and the two islands can stably hold (n_1, n_2) electrons in (V_{g1}, V_{g2})-plane. By changing (V_{g1}, V_{g2}) along the circle in Fig. 12(b), a single-electron moves from the left to the right terminal. The node cells in BDD-based logic circuits are built by applying the single-electron pumps, as shown in Fig. 13. A clocked logic circuit which functions as $f(x_1, x_2, x_3, x_4) = x_1 x_2 + x_3 x_4$ is constructed by connecting the node cells as shown in Fig. 14.

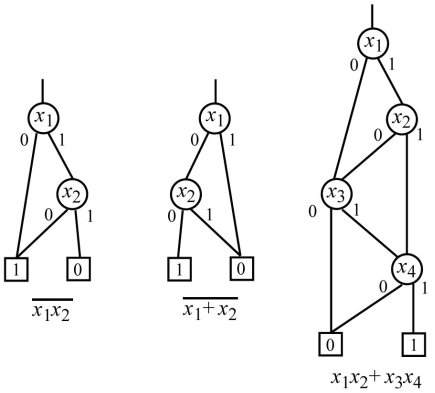


Fig. 11. BDD expression of Boolean logic functions

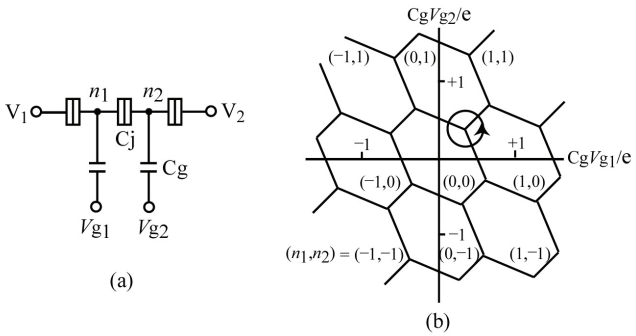


Fig. 12. Single-electron pump

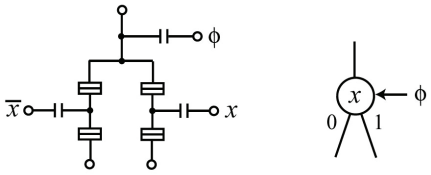


Fig. 13. Single-electron node cell of BDD-based logic circuits (Asai et al., 1997)

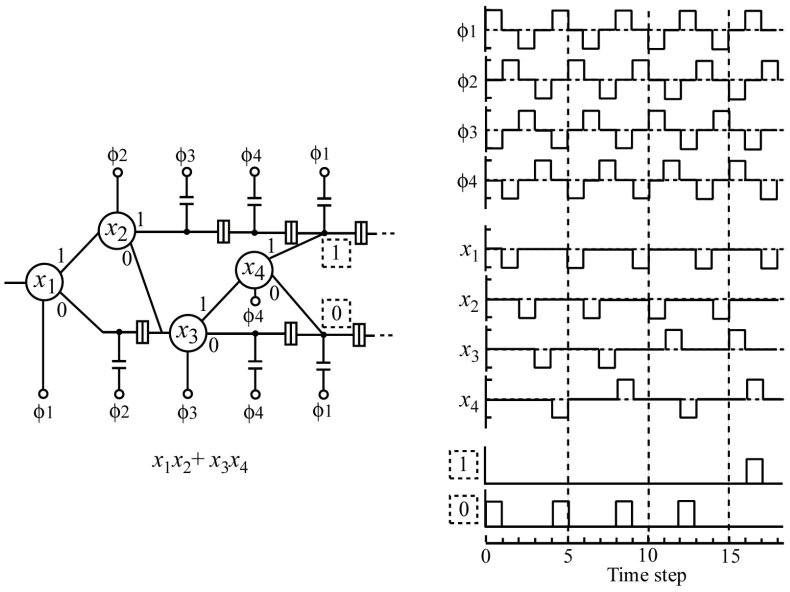


Fig. 14. A BDD-based clocked logic circuit (Asai et al., 1997)

4.3 Delay Elements

Delay elements are built by applying the single-electron pumps. Figure 15 shows a unit delay driven by a four-phase clock set. The delay element consists of seven tunneling junctions (Keller et al., 1996). Connecting additional 4(K-1) junctions to the unit delay changes its delay time to K(: integer).

4.4 Pseudo-CMOS Inverter Buffer

Figure 16(a) shows what is called a single-electron transistor and its gate voltage V_g versus drain-to-source current I_{DS} characteristic. The current peaks locate between the regions of gate voltage in which the island of the transistor holds integer number of excess electrons stably. The periodic characteristic is called Coulomb oscillation. Figure 16(b) shows a multi-gate single-electron transistor. When $C_{g1} = C_{g2} = C_g$ and $C_{g2}V_{g2} = e/2$, its V_{g1} versus I_{DS} characteristic shifts by $e/2C_g$ horizontally. Figure 16(c) shows an inverter (Tucker, 1992) consisting of the two two-gate single-electron transistors. Gate bias voltages V_p and V_n are determined so that current I_{DS} passes through the upper and the lower transistors at zero and high levels of input voltage respectively. The inverter is used as a buffer since its voltage gain $|dVout/dVin|$ can be greater than 1.0.

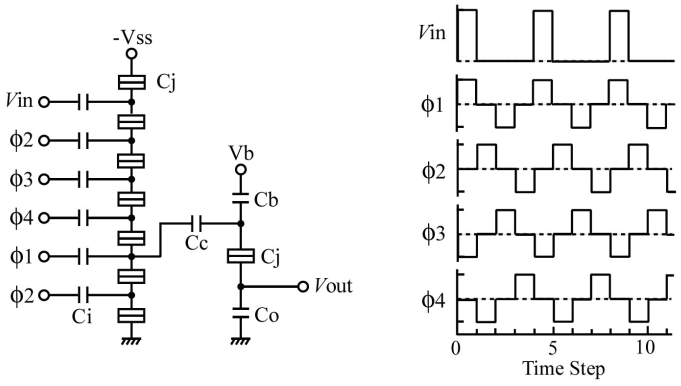


Fig. 15. A unit delay driven by a four-phase clock set

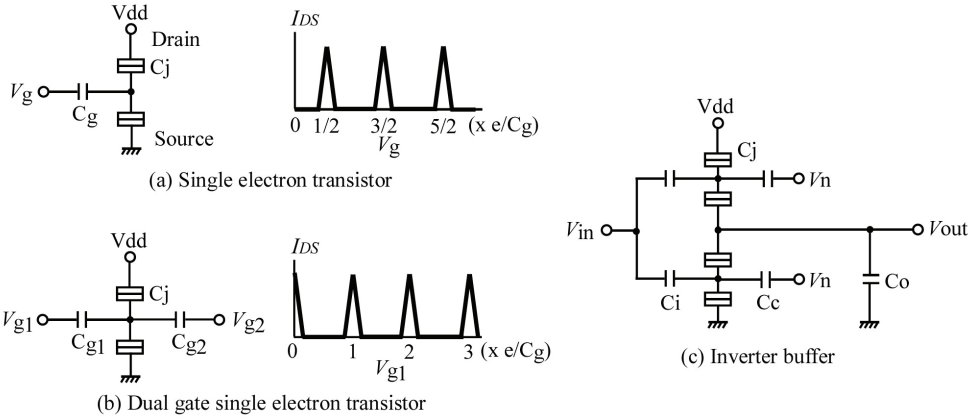


Fig. 16. A puseudo-CMOS inverter buffer

5. SET Circuit Modules for SDSP

In this section we will build arithmetic and piecewise linear (PWL) circuit modules (Fujisaka et al., 2002), (Katao et al., 2007), (Fujisaka et al., 2003), (Hayashi et al., 2007) for nanoelectronic SDSP using SET devices. The circuit modules operate on SD modulated signals and shape quantization error into the form of SD modulation noise.

5.1 Arithmetic Circuits

5.1.1 Adder

Figure 17 shows digital SD modulators converting 3-level signals represented by equally weighted two bits $x(n)$, and $y(n)$ into a SD modulated one-bit stream $z_{1/2}(n)$. The two digital SD modulators are respectively described by

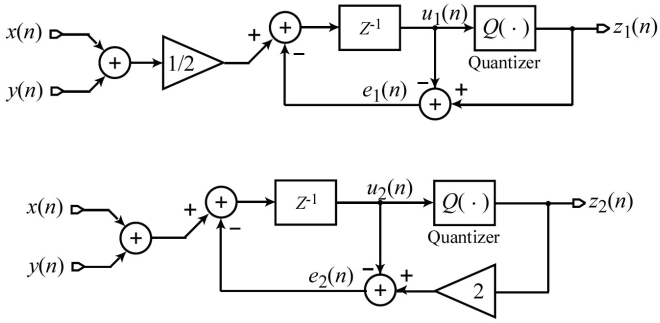


Fig. 17. Two SD modulators with equally weighted two-bit input

$$\begin{aligned} u_1(n) &= \frac{1}{2}\{x(n)+y(n)\}-e_1(n-1), \\ z_1(n) &= Q(u_1(n)), \\ e_1(n) &= z_1(n)-u_1(n), \end{aligned} \tag{14}$$

and

$$\begin{aligned} u_2(n) &= x(n)+y(n)-e_2(n-1), \\ z_2(n) &= Q(u_2(n)), \\ e_2(n) &= z_2(n)-u_2(n), \end{aligned} \tag{15}$$

where $Q(\bullet)$ denotes two-level quantization. The above two equation systems have the following relations between their variables and outputs:

$$u_2(n) = 2u_1(n), \quad e_2(n) = 2e_1(n), \quad z_2(n) = z_1(n). \tag{16}$$

Thus, the signal conversions by the two digital SD modulators are equivalent. When two SD modulated one-bit streams are applied to the inputs of the digital SD modulators, they operate as two-input adders. We will construct a SET circuit described by Eq. (15). Since $x(n)$, $y(n) \in \{+1, -1\}$, the state of the digital modulator and the quantization noise are $u_2(n) \in \{-3, -1, +1, +3\}$ and $e_2(n) \in \{-1, +1\}$. Thus, $u_2(n)$ can be represented by the sum of three binary signals as

$$u_2(n) = \sum_{i=1}^{3} u_{2,i}(n), \quad u_{2,i}(n) \in \{+1, \quad -1\}. \tag{17}$$

The three binary signals $u_{2,i}(n)$, i=1,2,3, are given values as shown in Tab. 2. Then, the output can be

$$z_2(n) = Q(u_2(n)) = u_{2,2}(n). \tag{18}$$

Fig. 18 shows an adder built of the SET gates introduced in section 4 (Katao et al., 2008).

$u_2(n)$	$u_{2,1}(n)$	$u_{2,2}(n)$	$u_{2,3}(n)$
–3	–1	–1	–1
–1	+1	–1	–1
+1	+1	+1	–1
+3	+1	+1	+1

Table 2. Three-bit representation of state $u_2(n)$

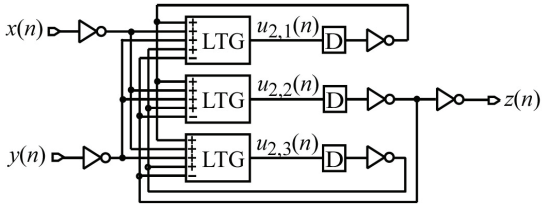


Fig. 18. Block diagram of the adder built of SET devices

5.1.2 Multiplier

Although SD modulation noise does not contain large low frequency components, the direct multiplication of the SD modulated signals generates large low frequency noise components. Pre-filters are necessary to remove SD modulation noise from a multiplier signal and a multiplicand signal. Let a multiplier and a multiplicand be denoted by $x(n)$ and $y(n)$ and let their filtered signals be given by

$$\overline{x(n)} = \frac{1}{N}\sum_{i=0}^{N-1} x(n-i), \qquad \overline{y(n)} = \frac{1}{N}\sum_{i=0}^{N-1} y(n-i). \tag{19}$$

The filtered signals are no more SD modulated single-bit signals and the multiplication cannot be performed in SD domain. The product of the filtered signals is expanded as the right hand side of the following equation:

$$\overline{x(n)}\,\overline{y(n)} = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} x(n-i)y(n-j). \tag{20}$$

The expanded product can be computed with exclusive-OR gates and the above-mentioned SD domain adders. Then the product is computed in SD domain. Figure 19 shows an example of SD domain multipliers. When N=2 in Eq. (20), the multiplier is built of SET gates as shown in Fig. 20 (Katao et al., 2008).

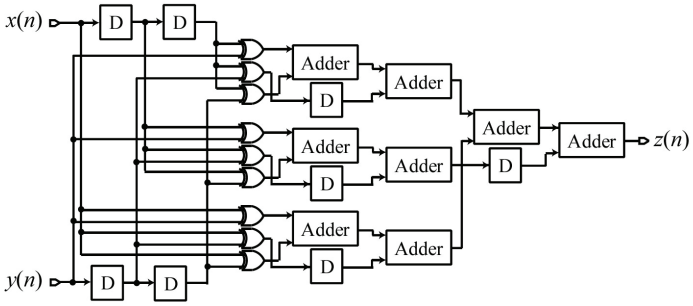


Fig. 19. Structure of multipliers for SDSP

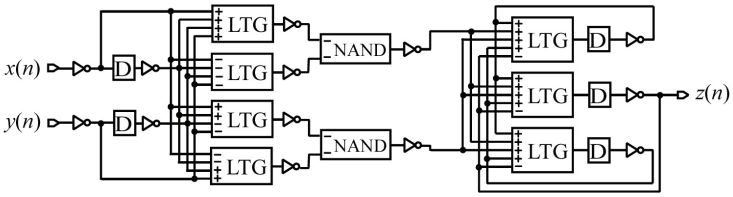


Fig. 20. Block diagram of a multiplier built of SET devices

5.1.3 Circuit Simulation of the Arithmetic Circuits

The SD modulated sum and product obtained by the SET arithmetic circuits in Figs. 18 and 20 are shown in the time and frequency domain in Figs. 21 and 22. The results are obtained by circuit simulation with SIMON (Wasshuber et al., 1997). Figure 22 validates that the SET arithmetic circuits have the function of quantization noise shaping. Figure 21 shows the averaged outputs of the SET circuits. As the transient error rate of SET junctions becomes

higher, the circuits increase noise components and decrease signal components in their outputs. However, the circuits never go out of control. The signal-to-noise ratio (SNR) versus SET junction error rate curves of the adder and the multiplier are shown in Fig. 23. It is found from the figure and Eq. (9) in section 3 that the outputs of the arithmetic circuits with the SET junction error rate of 10^{-2} have the same signal quality with that of errorless stochastic computing.

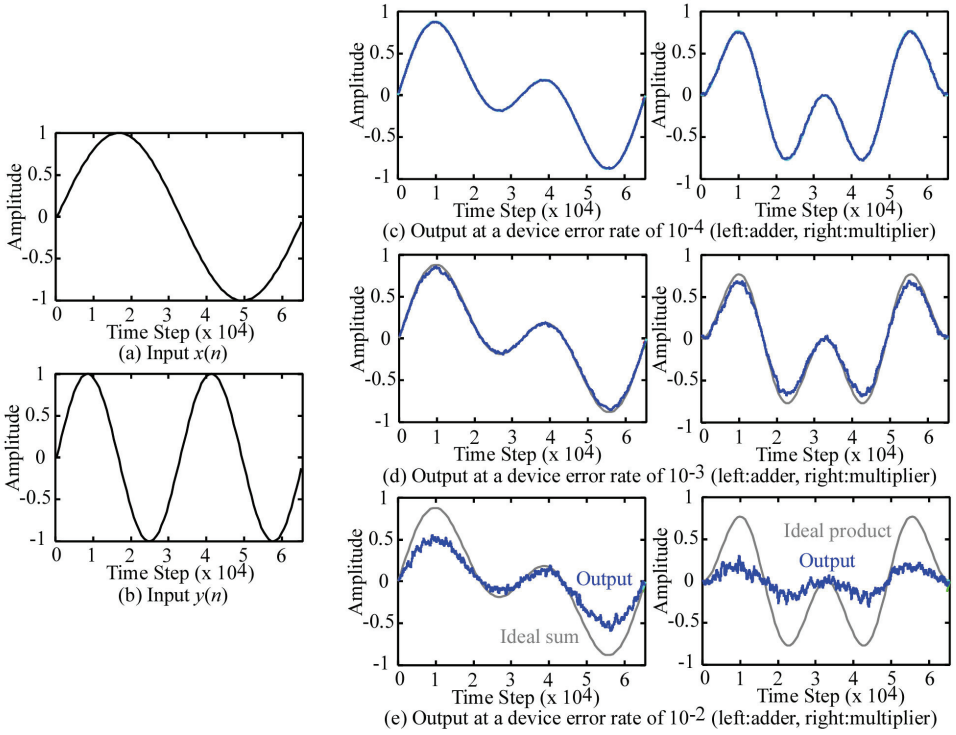


Fig. 21. Averaged inputs and outputs of the arithmetic circuits

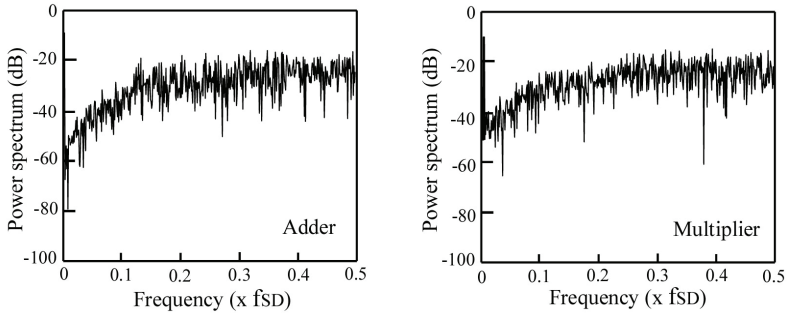


Fig. 22. Frequency spectra of the outputs of the arithmetic circuits

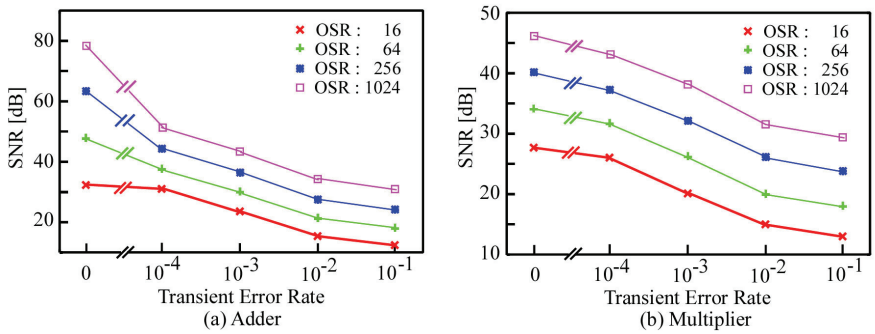


Fig. 23. Fault tolerance characteristics of the arithmetic circuits: Output SNR versus transient error rate of SET junctions

5.2 Piecewise Linear Circuits

In ASP, PWL modules such as limiters and rectifiers are easily build by using diodes and frequently used. In SDSP as a successor of ASP, PWL modules are required.

5.2.1 Absolute Circuit

There are several methods of computing absolute (ABS) values in SDSP. We will show a method proposed in (Hayashi et al., 2007). The first-order binary SD modulated signals possess the following characteristic:

[Theorem] Let a first-order binary SD modulator be described by

$$\begin{aligned} &y(n)=Q(u(n)) \in \{+1, \quad -1\}, \\ &e(n)=y(n)-u(n), \\ &u(n)=x(n-1)-e(n-1), \quad x(n) \in [-1, \quad +1]. \end{aligned} \tag{21}$$

The output bit-stream $y(n)$ is a sequence such that

$$(y(n), \quad y(n-1)) \neq \begin{cases} (-1, \quad -1) & \text{if } x(n) \geq 0, \\ (+1, \quad +1) & \text{if } x(n) \leq 0. \end{cases} \tag{22}$$

[Proof] Consider the SD modulation in the case that $0 \leq x(n) \leq 1$ and $-1 \leq u(m) \leq 2$, $n \geq m$. When state $u(m)$ is in

$$0 \leq u(m) \leq 2, \tag{23}$$

the quantization error is

$$-1 \leq e(m)=Q(u(m))-u(m) \leq 1. \tag{24}$$

Then, from the last equation in Eq. (21), we have

$$-1 \leq u(m+1) \leq 2. \tag{25}$$

When state $u(m)$ is in

$$-1 \leq u(m) \leq 0, \quad y(m)=Q(u(m))=-1, \tag{26}$$

the quantization error is

$$-1 \leq e(m)=Q(u(m))-u(m) \leq 0. \tag{27}$$

Then,

$$0 \leq u(m+1) \leq 2, \quad y(m+1)=Q(u(m+1))=+1. \tag{28}$$

From Eqs. (23) to (28), the range of state $u(n)$ is

$$-1 \le u(n) \le 2 \quad \text{for} \quad n \ge m. \tag{29}$$

From Eqs. (26), (28) and (29), the upper part of the right hand side of Eq. (22) is obtained. The lower part is similarly proved. An ABS circuit for SDSP is shown in Fig. 24. When the sub-circuit consisting of two AND gates and a unit delay detects ($y(n)$, $y(n\text{-}1)$) = (-1, -1), the SR flip flop is cleared and the exclusive-OR gate inverts input $x(n)$. A SET implementation of the ABS circuit is built by using LTG-based Boolean logic gates mainly, as shown in Fig. 25.

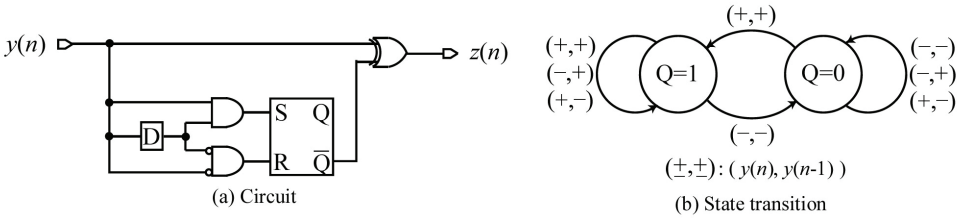


Fig. 24. ABS circuit and the state transition of its SR flip-flop

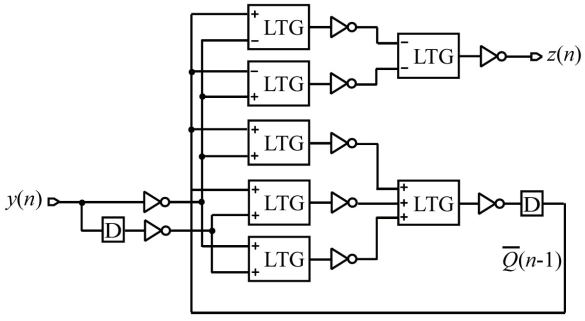


Fig. 25. Block diagram of the ABS circuit built of SET devices

5.2.2 Min/Max and Other PWL Circuits

Operation of a Min/Max selector is expressed by ABS operation and addition/subtraction (Li et al., 1999) as

$$\begin{aligned} \mathrm{Max}(x,y) &= c + |d|, \\ \mathrm{Min}(x,y) &= c - |d|, \\ c &= (x+y)/2, \\ d &= (x-y)/2. \end{aligned} \tag{30}$$

Since ABS and adder modules for SDSP have been developed, the operations given by Eq. (30) can be carried out in SDSP. Figure 26 shows a Min/Max circuit built according to Eq. (30).

A Limiter, a dead zone circuit, and a negative resistance whose functions are described in Fig. 27 are built by using other PWL circuits and adders as shown in Fig. 28. SET circuits for these three PWL operations in SDSP can be built by using the adder, ABS, and Min/Max modules.

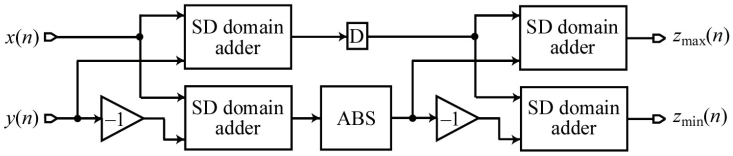
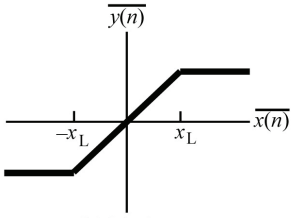
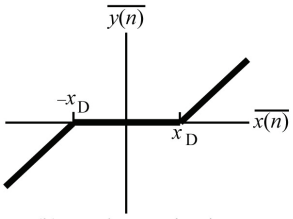


Fig. 26. Block diagram of the Min/Max circuit





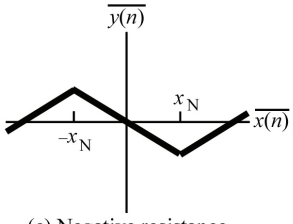
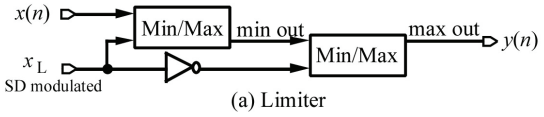
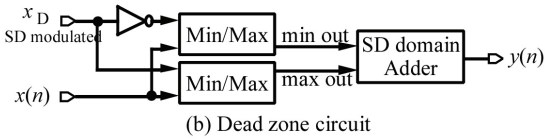


Fig. 27. PWL functions





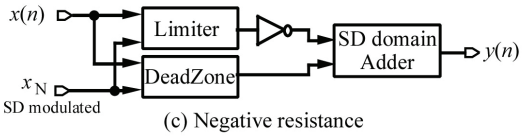


Fig. 28. Block diagrams of the PWL circuits

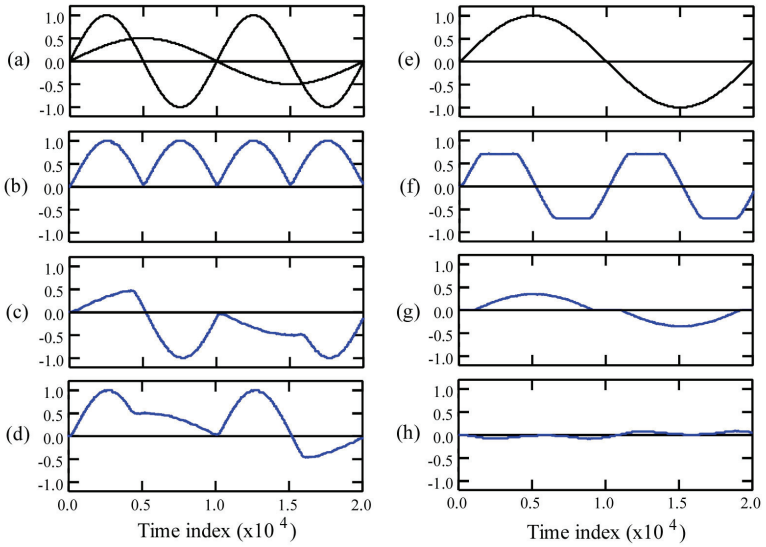


Fig. 29. Averaged inputs and outputs of the PWL circuits; (a) inputs to the ABS and Min/Max circuits, (b) output of the ABS circuit, (c, d) the Min and Max outputs, (e) input to the limiter, dead zone circuit, and negative resistance, (f) output of the limiter, (g) output of the dead zone circuit, (h) output of the negative resistance

5.2.3 Circuit Simulation of the PWL Circuits

Results of simulating the ABS, Min/Max, limiter, dead-zone circuit, and negative resistance are shown in Fig. 29.

6. Applications

Filters have been the most important circuits in both analog and digital signal processing, and they will be in SDSP too. In this section, using the SET modules presented in the previous section, we will construct linear and nonlinear filters.

6.1 Linear Filter

To integrate signals in SD domain is not simple because of the bit-stream signal form. Linear filters for SDSP should be constructed not based on lumped analog filters with integrators but based on analog distributed-parameter filters (ADPF) (Mattaei, et al., 1980). ADPFs are built of transmission lines connected directly, or coupled capacitively and/or inductively, as shown in Fig. 30. In some kinds of ADPF, waves propagating on their transmission lines reflect at the junctions between the transmission lines with different characteristic impedances. Interference between incidental and reflected waves effects filtering. In other kind of ADPF, standing waves are generated on the transmission lines. The resonance causes bandpass or bandstop effect.

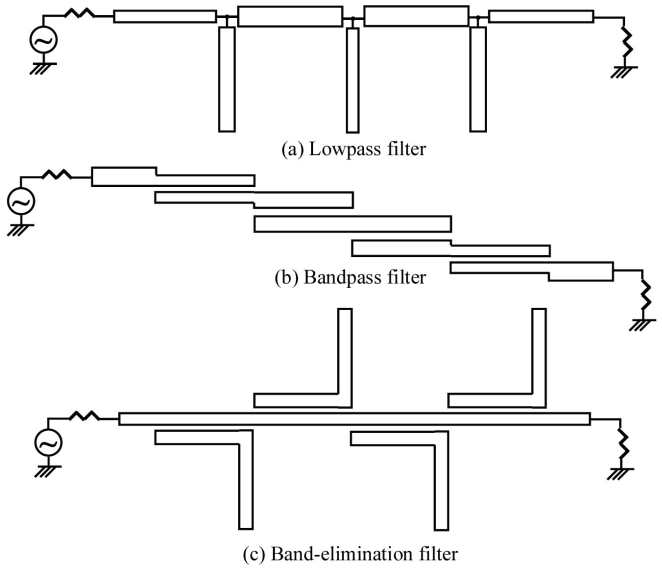


Fig. 30. Analog distributed parameter filters

Digital wave filters based on ADPFs have been proposed in (Suzuki et al., 2007a & 2007b). Wave propagation on lossless transmission lines are described by

$$\frac{\partial^2 w(x,t)}{\partial t^2} = c^2 \frac{\partial^2 w(x,t)}{\partial x^2}, \tag{31}$$

where c denotes propagation speed. By the spatial and temporal discretization $x = i\Delta x$, $t = n\Delta t$ with $\Delta x/2\Delta t = c$ and the binary quantization of variable w, Eq. (31) is transformed to the following equations.

$$\begin{aligned} &w(i,n) = w_L(i,n) + w_R(i,n), \\ &w_L(i,n+1) = w_L(i+1,n-1), \quad n : \text{even}, \\ &w_R(i,n+1) = w_R(i-1,n-1), \quad n : \text{odd}, \\ &w_L(i,n), w_R(i,n) \in \{+1, \quad -1\}. \end{aligned} \tag{32}$$

The equations mean that the discrete wave w is the sum of binary-quantized waves w_L and w_R traveling left and rightwards. Thus, as shown in Fig. 31, a transmission line is represented by a pair of left and right shift integer delays in digital wave filters. The arithmetic modules presented in section 5 are placed between integer delays to cause wave phenomena corresponding to wave penetration and reflection at a junction of transmission lines and to coupling between transmission lines of ADPFs.

Consider a bandpass digital wave filter corresponding to the ADPF shown in Fig. 30(b). Building blocks for the ADPF are coupled parallel transmission line pairs with open ends. Let the lines of a pair be referred as line 1 and 2.

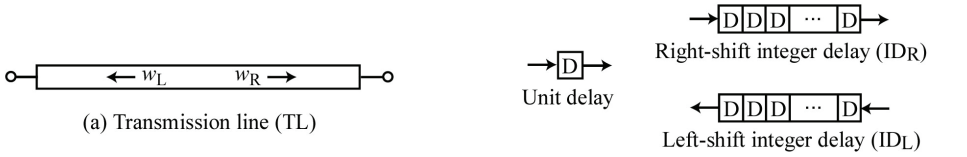


Fig. 31. Analog transmission line and its binary-quantized discrete model

Waves traveling left and rightward on lines 1 and 2 are denoted by $w_{1,L}$, $w_{1,R}$, $w_{2,L}$, and $w_{2,R}$. Even and odd mode waves propagating left and rightward on the line pair are defined by

$$\begin{aligned} w_{e,L} &= (w_{1,L} + w_{2,L})/2, \\ w_{o,L} &= (w_{1,L} - w_{2,L})/2, \\ w_{e,R} &= (w_{1,R} + w_{2,R})/2, \\ w_{o,R} &= (w_{1,R} - w_{2,R})/2. \end{aligned} \tag{33}$$

Because the even and odd mode waves are different in propagation speed, the line pair has two resonance frequencies. The center frequency and the bandwidth of a bandpass ADPF consisting of the coupled line pairs are determined by the resonance frequencies. A digital model of the coupled line pair is shown in Fig. 32(a). Integer delays ID$_{e/o,L/R}$ representing transmission lines are built as mentioned in section 4. Type-1 adders are the same circuits as the one presented in Fig. 18. Type-2 adders are different from type-1 adders in that the variable $u_{2,2}$ is not fedback. Both types of adders are used to make the coupling phenomena between the lines in the following way: The type-1 adders synthesize even and odd mode waves $w_{e,L/R}$ and $w_{o,L/R}$. The integer delay pairs propagate the two modes of waves. The length of the integer delays in a pair are set differently instead of setting the propagation speed of the even and odd mode waves differently. The type-2 adders reconstruct waves $w_{1,L/R}$, and $w_{2,L/R}$. Reflection at the open ends of the lines of the pair is represented by inputting waves outputted from the type-2 adders to the type-1 adders. A bandpass digital wave filter is constructed by connecting the building blocks like the ADPF, as shown in Fig. 32(b).

Figure 33 shows the frequency response of the bandpass filter with 4 blocks. In each block, the integer delays ID$_{o,L/R}$ on which odd mode waves propagate are three times as long as the integer delays ID$_{e,L/R}$ on which even mode waves propagate. One building block with ID$_{e,L/R}$ of length $0.5\times l_{ID}$ and ID$_{o,L/R}$ of length $1.5\times l_{ID}$, l_{ID} = 64, is simulated with SIMON. Simulation results are shown in Fig. 34. In the figure, $f_0 = 1/4l_{ID} = 1/256$. Inputs are passed and attenuated in the passband and the stopband as theoretically estimated. Figure 35 shows the signal-to-noise ratio (SNR) versus SET junction error rate curves. In the figure, OSR = $f_{SD}/2f_0 = 2\times l_{ID}$. The figure shows that the signal quality does not affected by low rate junction error.

6.2 Nonlinear Filter

Nonlinear digital filters are categorized into two main classes, ε filters and order statistic filters (Pitas & Venetsanopoulos, 1990). Order statistic filters permute input series sampled in a time window in ascending or descending order and weight the samples depending on the order. Median filters, mid-range filters, α-trimmed mean filters belong to the class of

order statistic filters. They can remove wide band noise contained in wide band signals without changing the waveform of the signals. For example, median filters effectively remove pulse-shaped noise.

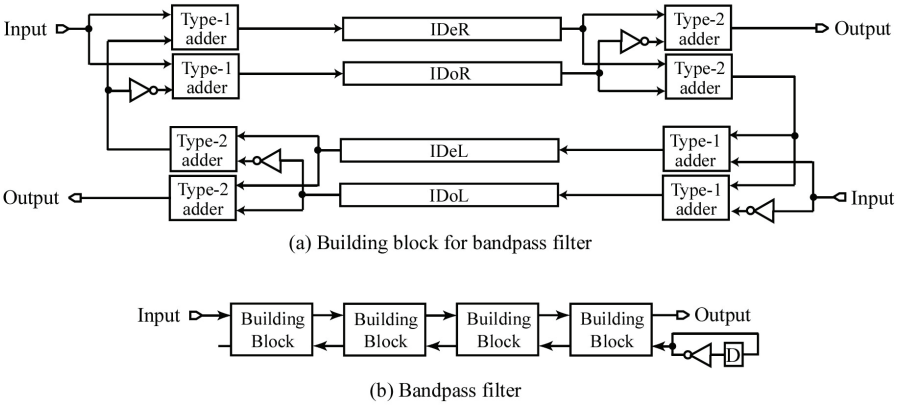


Fig. 32. Digital bandpass filter

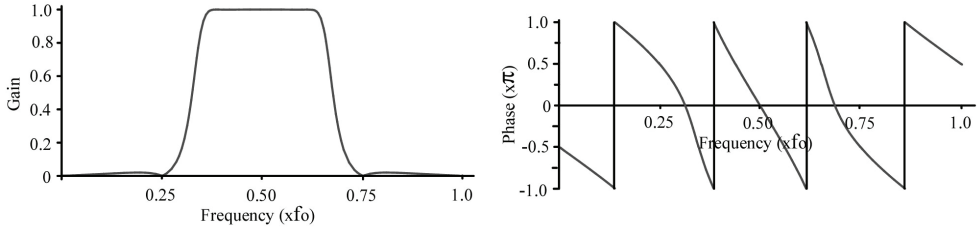


Fig. 33. Theoretical frequency response of the bandpass filter

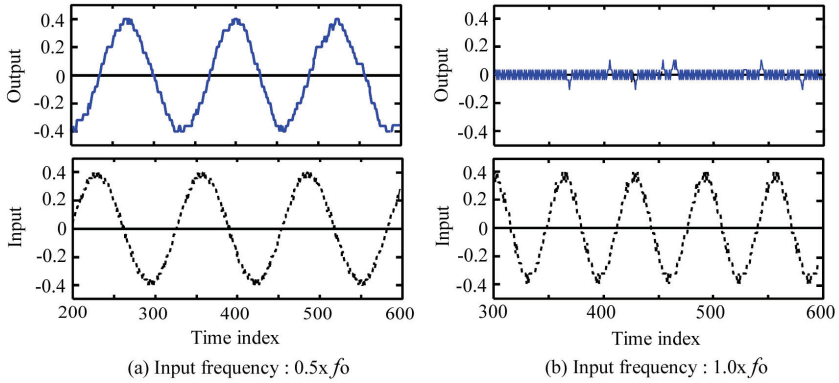


Fig. 34. Examples of simulation results of the bandpass filter block with SIMON

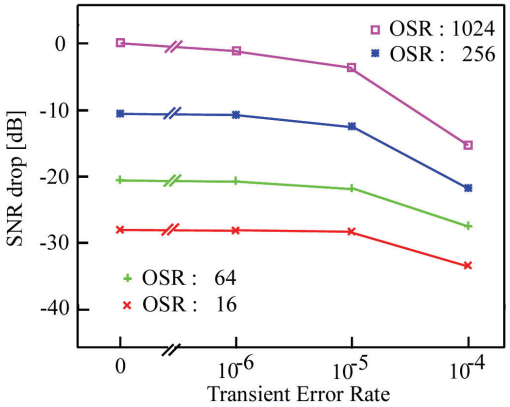
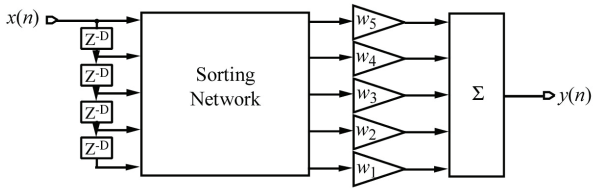
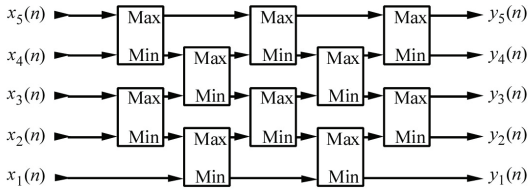


Fig. 35. Fault tolerance characteristic of the bandpass filter: Output SNR drop versus transient error rate of SET junctions

Figure 36(a) shows the schematic diagram of order statistic filters. They consist of front delay elements, a sorting network permuting input samples, and a weighted sum computation block. In Nyquist rate operation, the delay time of the front delay elements is D=1. Several kind of efficient sorting networks have been proposed. A sorter shown in Fig. 36(b) is what is called an odd-even transposition sorter. A median filter, a mid-range filters, and an α-trimmed mean filter have weight sets ($w_3 = 1$, $w_1 = w_2 = w_4 = w_5 = 0$), ($w_1 = w_5 = 1/2$, $w_2 = w_3 = w_4 = 0$), and ($w_2 = w_3 = w_4 = 1/3$, $w_1 = w_5 = 0$) respectively.



(a) Order Statistic Filter



(b) Odd-even ransposition sorting network

Fig. 36. Order statistic filters

A median filter operating on SD modulated signals was first presented in (Fujisaka et al., 2003). An example of impulse noise suppression by the median filter is shown in Fig. 37. The result is obtained by logic level simulation carried out under the following conditions: The

delay time of the front delay elements is D = 20. The signal component is a sinusoidal wave with amplitude 0.5 and period 10^4. The positive impulse noise is 20D in interval, 0.2D in width and +0.4 in height. The negative impulse noise is 35D in interval, 0.9D in width and -0.5 in height. The SET circuit of the median filter can be constructed by employing integer delays presented in section 4 as the front delay elements and Min/Max selectors shown in section 5 in the odd-even transposition sorter.

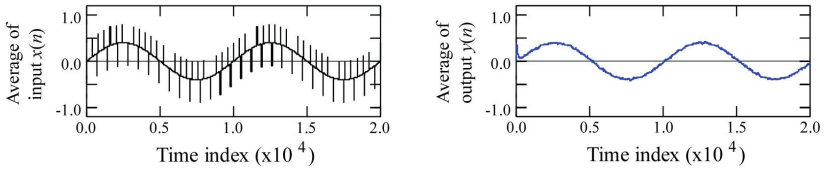


Fig. 37 Average input and output of the median filter

7. Conclusions

This chapter has presented single-electron pulsed signal processing circuits. The pulsed signal processing scheme is like analog signal processing which is considered to decay in the near-future nanoelectronics age. The pulsed signal processing employs high quality and fault-tolerant sigma-delta modulated signal form, which derived the name of signal processing, sigma-delta domain signal processing (SDSP).
The research on nanoelectronic SDSP has just begun. It contains a lot of works to be challenged. First, circuit architecture of local connection style needs to be developed since parasitic capacitors, inductors, and resistors of the wires between nano-electronic devices cause errors in the device operation. The filters presented in section 6 partly take the style.

Secondly, the circuits operating on high-order SD modulated signals need to be developed since high-order SD modulated signal form is useful for high quality and low clock-rate signal processing. Thirdly, massively parallel computing for high performance processing and spatially redundant computing for defect and fault tolerance (Shukla & Bahar, 2004) will be challenged by taking advantage of the small size of SDSP modules. Lastly, new functions and new applications of SDSP will be explored since SDSP systems which are essentially digital systems can be functionally higher than ASP circuits.

8. Acknowledgment

The author appreciates the help and the advice to this research from the staff members, Dr. Takeshi Kamio, Dr. Chang-Jun Ahn, Prof. Kazuhisa Haeiwa, Dr. Masahiro Sakamoto, and the graduate students, Daisuke Hamano, Yuji Hidaka, Tsubasa Katao, Keita Hayashi, Yoshinao Suzuki, Takasi Yasuno who are(were) associated with the Communication and Signal Processing Laboratory at the Faculty of Information Sciences, Hiroshima City University.

9. References

Asai, N.; Akazawa, M. & Amemiya, Y. (1997). Single-Electron Logic Device Based on the Binary Decision Diagram, *IEEE Transactions on Electron Devices*, Vol. 44, No. 7, pp. 1109-1116, ISSN 0018-9383.

Chibashi, M.; Eguchi, K. & Waho, T. (2004). A Novel Delta-Sigma Modulator Using Resonant-Tunneling Quantizers, *Proceedings of IEEE International Symposium on Circuits and Systems*, pp. I-533-I-536, ISBN 0-7803-8252-8, Vancouver, May, 2004, IEEE.

Candy, J. C. & Temes, G. C. (1992). Oversampling Methods for A/D and D/A Conversion, In: *Oversampling Delta-Sigma Data Converters*, Candy, J. C. & Temes, G. C. (Ed.), pp. 1-29, IEEE Press, ISBN 978-0-87942-285-1, New York.

Fujisaka, H.; Kurata, R.; Sakamoto, M. & Morisue, M. (2002). Bit-Stream Signal Processing and its Application to Communication Systems, *IEE Proceedings on Circuits, Devices and Systems*, Vol. 149, No. 3, pp. 159-166, ISSN 1350-2409.

Fujisaka, H.; Hidaka, Y.; Kajita, S. & Morisue, M. (2003). Piecewise Linear Operators on Sigma-Delta Modulated Signals and their Application, *IEICE Transactions on Fundamentals*, Vol. E86-A, No. 5, pp. 1249-1255, ISSN 0916-8505.

Gaines, B. R. (1969). Stochasitic Computing Systems, In: *Advances in Information Systems Sciences*, Tou, J. F. (Ed.), Vol. 2, pp. 37-172, Plenum Press, ISBN 978-0306416446, New York.

Grabert, H. & Devoret, M. H. (Ed.) (1992). *Single Charge Tunneling - Coulomb Blockade Phenomena in Nanostructures*, Plenum Press, ISBN 0-306-44229-9, New York.

Graham, P. & Gokhale, M. (2004). Nanocomputing in the Presence of Defects and Faults: A Survey, In: *Nano, Quantum and Molecular Computing*, Shukla, S. K. & Bahar, R. I. (Ed.), pp. 39-72, Kluwer Academic Publishers, ISBN 1-4020-8067-0, Dordrecht.

Hayashi, K.; Katao, T.; Fujisaka, H.; Kamio, T. & Haeiwa, K. (2007). Piecewise Linear Circuits Operating on First-order Multi-level and Second-order Binary Sigma-Delta Modulated Signals, *Proceedings of European Conference on Circuit Theory and Design*, pp. 683-686, ISBN 1-4244-1342-7, Seville, August, 2007, IEEE.

Hori, M.; Ueda, M. & Iwata, A. (2006). Stochastic Computing Chip for Measurement of Manhattan Distance, *Journal of Applied Physics*, Vol. 45, No. 4B, pp. 3301-3306, ISSN 0021-8979.

Katao, T.; Hayashi, K.; Fujisaka, H.; Kamio, T. & Haeiwa, K. (2007). Sorter-based Sigma-Delta Domain Arithmetic Circuits, *Proceedings of European Conference on Circuit Theory and Design*, pp. 679-682, ISBN 1-4244-1342-7, Seville, August, 2007, IEEE.

Katao, T.; Suzuki, Y.; Fujisaka, H.; Kamio, T.; Ahn, C. & Haeiwa, K. (2008). Single-Electron Arithmetic Circuits for Sigma-Delta Domain Signal Processing, *Proceedings of IEEE International Conference on Nanotechnology*, pp.729-732, ISBN 1-4244-2104-6, Texas, August, 2008, IEEE.

Keller, M. W.; Martinis, J. M.; Zimmerman, N. M. & Steinbach, A.H. (1996). Accuracy of Electron Counting Using a 7-junction Electron Pump, *Applied Physics Letters*, Vol. 69, pp. 1804-1806, ISSN 0003-6951.

Lageweg, C.; Cotofana, S. & Vassiliadis, S. (2001). A Linear Threshold Gate Implementation in Single-Electron Technology, *Proceedings of IEEE Computer Society VLSI Workshop*, pp. 93-98, ISBN 978-0769510569, Florida, May, 2001, IEEE.

Lewis, P. M. & Coates, C. L. (1967). *Threshold Logic*, John Wiley and Sons, ISBN 978-0471532507, New York.

Li, W.; Lin, J. & Unbehauen, R. (1999). Unification of Order-Statistics Based Filters to Piecewise-Linear Filters, *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Vol. 46, No. 11, pp. 1397-1403, ISSN 1057-7130.

Maezawa, K.; Akeyoshi, T. & Mizutani, T. (1994). Functions and Applications of Monostable-Bistable Transition Logic Elements (MOBILE's) Having Multiple-input Terminals, *IEEE Transactions on Electron Devices*, Vol. 41, No. 2, pp. 148-154, ISSN 0018-9383.

Mattaei, G.; Young, L.; & Jones, E.M.T. (1980). *Microwave Filters, Impedance Matching Networks, and Coupling Structures*, Artech House, ISBN 978-0890060995, Norwood, MA.

Miller, D. M.; Shmerko, V. P.; Stankovic, R. S. & Yanushkevich, S. N. (2006). *Decision Diagram Techniques for Micro- and Nanoelectronic Design*, CRC Press, ISBN 0-8493-3424-1, North Wales.

Pitas, I. & Venetsanopoulos, A. N. (1990). *Nonlinear Digital Filters - Principles and Applications*, Kluwer Academic Publishers, ISBN 0-7923-9049-0, Massachusetts.

Pothier, H.; Lafarge, P.; Urbina, C.; Esteve, D. & Devoret, M. H. (1992). Single-Electron Pump Based on Charging Effects, *Europhysics Letters*, Vol. 17, pp. 249-254, ISSN 0295-5075.

Sols, F.; Macucci, M.; Ravaioli, U. & Hess, K. (1989). On the Possibility of Transistor Action Based on Quantum Interference Phenomena, *Applied Physics Letters*, Vol. 54, pp. 350-352, ISSN 0003-6951.

Suzuki, Y.; Fujisaka, H.; Kamio, T. & Haeiwa, K. (2007a). Sorter-based Digital Wave Filters: An Approach to Nanoelectronic Signal processing, *Proceedings of IEEE International Workshop on Nonlinear Dynamics of Electronic Systems*, pp. 18-21, Tokushima, July, 2007.

Suzuki, Y.; Fujisaka, H.; Kamio, T. & Haeiwa, K. (2007b). Digital Filters Built of Locally Connected Nanoelectronic Devices, *Proceedings of IEEE International Conference on Nanotechnology*, pp. 873-878, ISBN 1-4244-0608-0, Hong Kong, August, 2007, IEEE.

Tinkham, M. (1975). *Introduction to Superconductivity*, McGraw-Hill, ISBN 978-0070648777, New York.

Tucker, J. R. (1992). Complementary Digital Logic Based on the Coulomb Blockade, *Journal of Applied Physics*, Vol. 72, No. 9, pp. 4399-4413, ISSN 0021-8979.

Wasshuber, C.; Kosina, H. & Selberherr, S. (1997). A Simulator for Single-Electron Tunnel Devices and Circuits, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 6, Issue 9, pp. 937-944, ISSN 0278-0070.

van Wees, B. J.; van Houten, H.; Beenakker, C. W.; Williamson, J. G.; Kouwenhoven, L. P.; van der Marel, D. & Foxon, C. T. (1988). Quantized Conductance of Point Contacts in a Two-Dimensional Electron Gas, *Physical Review Letters*, Vol. 60, Issue 9, pp.848–850, ISSN 0031-9007.

Yokoyama, Y.; Ohno, Y.; Kishimoto, S.; Maezawa, K. & Mizutani, T. (2001). A Delta-Sigma Analog-to-Digital Converter Using Resonant Tunneling Diodes, *Japanese Journal of Applied Physics*, Vol. 40, pp. L1005-L1007, ISSN 0021-4922.

17

Carbon Nanotube Capacitors

Mark M. Budnik and Eric W. Johnson
Valparaiso University
United States of America

1. Introduction

Capacitors are important components in many integrated circuits. They serve numerous roles in analog and mixed signal circuits, including switched capacitor filters and sample-and-hold circuits. In addition, capacitors provide a vital role in the decoupling of microprocessors, digital signal processors, and microcontrollers from power supply variations.

Traditional integrated circuit capacitors use a parallel plate structure. The capacitance of these parallel plate structures is fundamentally limited by the die area they consume and the thickness of the dielectric material between the parallel plates. However, increasing the capacitance of traditional parallel plate capacitors by increasing the die area (resulting in a more expensive component) or decreasing the dielectric thickness (resulting in a higher leakage current) contradicts two of the fundamental tenements of integrated circuit design (ITRS, 2008).

Carbon nanotubes (CNTs) are nanotechnology materials that have been in prominence for the last several years. They are cylinders of graphene that can exhibit radii on the order of nanometers. CNTs exhibit a number of properties that make them attractive as potential horizontal and vertical interconnects in future integrated circuits (Naeemi et al., 2004, Naeemi et al., 2005, Raychowdhury & Roy, 2004). The same properties allow CNTs to be used to create a new capacitor structure suitable for use in future integrated circuits (Budnik et al., 2006). In this chapter we introduce these CNT capacitor structures and compare their capacitance per unit area with existing technologies.

The chapter is organized as follows. Section 2 reviews the structure and limitations of traditional parallel plate capacitors in integrated circuits. Section 3 briefly introduces CNTs, their varieties, and how they can be manufactured. Section 4 reviews the electrical models that have been developed for CNTs for use in the development of a CNT capacitor model. Section 5 formally introduces CNT capacitor devices as an extrapolation of the currently proposed CNT interconnects. Three different CNT capacitor structures and their electrical models are presented in this section. Finally, the conclusions are presented in Section 6.

2. Capacitors

Traditional semiconductor capacitors use a parallel plate structure. The various semiconductor capacitors, however, use different electrode materials. The metal-oxide-

semiconductor capacitor (MOSC) uses a metal-oxide-semiconductor field effect transistor (MOSFET) to create the capacitor. The MOSFET's gate serves as one electrode, and its source and drain regions serve as the second electrode. The gate dielectric serves as the capacitor dielectric. Therefore, the inter-plate spacing is the MOSFET gate oxide thickness. Another integrated circuit parallel plate semiconductor capacitor is the metal-insulator-metal capacitor (MIM). The electrodes of the MIM are metal layers located in the higher metal levels of an integrated circuit. A thin dielectric material separates the plates. The capacitance of both the MOSC and MIM devices is approximated by the equation:

$$C = \frac{A \in}{d} \tag{1}$$

where A is the area of each parallel plate, d is the thickness of the dielectric, and $\in$ is the permittivity of the dielectric material separating the parallel plates.

However, due to their parallel plate structure, MOSC and MIM capacitors face numerous challenges in future integrated circuit technologies (ITRS, 2008). First, as component dimensions are scaled, decreased parallel plate area (scaled in width and length, an S^2 decrease) will result in a reduction in capacitance. Decreasing the dielectric thickness can offset the decrease in plate area, but the increase in capacitance is only linear (an S increase) with respect to reduced inter-plate spacing. In addition, decreasing the dielectric thickness can lead to higher leakage currents. Therefore, new capacitor structures are being investigated for future integrated circuits. For example, interdigitated Metal-Oxide-Metal (MOM) capacitors are similar to MIM capacitors and do not require additional processing steps (ITRS, 2008). However, even MOM devices are expected to have capacitance densities of less than 30fF/μm^2 through 2022 (see Figure 1).

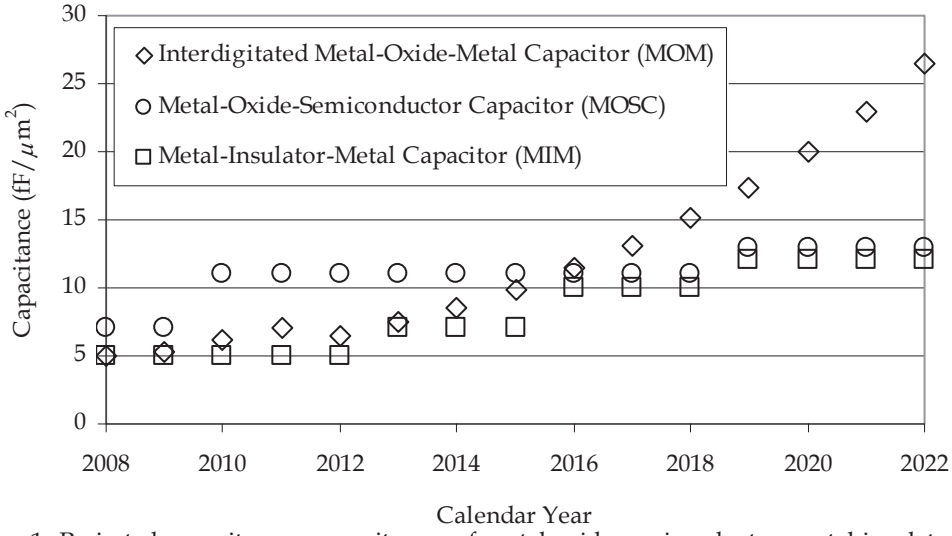


Fig. 1. Projected capacitance per unit area of metal-oxide-semiconductor, metal-insulator-metal, and interdigitated metal-oxide-metal capacitors (ITRS, 2008).

3. Carbon Nanotubes

CNTs are nanoscale devices that have unique strength and electrical properties. They are formed when a graphene sheet of carbon atoms is rolled into a cylindrical structure. These structures have a wall thickness of one atom, a diameter on the order of ten atoms and are typically over several micrometers in length (Dresselhaus et al., 2000).

A single cylinder nanotube is called a single wall carbon nanotube (SWCNT). Multi-wall carbon nanotubes (MWCNTs) contain multiple coaxial cylinders which are concentric SWCNTs. Both SWCNTs and MWCNTs can have varying diameter dimensions. SWCNTs have diameters from one to a few nanometers, while MWCNT diameters are from the tens to a hundred nanometers (Naeemi & Meindl, 2007). CNTs can be analyzed either in isolation (singularly) or in bundles (arrays of CNTs along a common axis).

One of the reasons CNTs have attracted so much interest is their electrical characteristics. CNTs can act as either metallic or semiconducting materials, depending on their structure. The circumference of a CNT can be represented by a chiral vector, which describes how carbon atoms within the tube are connected together. Depending on this vector, different types of nanotube structures can be constructed when the graphene sheet is rolled into a cylinder. If the chiral vector angle is 0°, then a *zigzag* structure is formed, and the CNT acts like a semiconductor. If the angle is 30°, an *armchair* structure occurs, and the CNT acts like a metal. All other nanotubes are labeled chiral nanotubes and will also act as semiconductors. Scientists have had limited success controlling the chiral vector; therefore, only about one third of carbon nanotubes that are manufactured are metallic (Dresselhaus et al., 2000).

CNTs are primarily manufactured using three different methods: arc-discharge, laser ablation, or chemical vapor deposition (Dresselhaus et al., 2000). Arc-discharge and laser ablation are similar in that both involve evaporating solid carbon sources to produce a condensation of carbon atoms. In contrast, chemical vapor deposition grows carbon nanotubes on top of a metal catalyst.

In the arc-discharge method, a carbon anode and cathode are set up in a chamber filled with helium gas. High currents are passed through the circuit, causing the evaporation of carbon atoms (Dresselhaus et al., 2000). The resulting carbon soot contains both MWCNTs and graphitic particles that are removed in a high-temperature oxygen environment. The growth of the nanotubes can be controlled by changing either the pressure of the gas in the chamber or the amount of arcing current used. To create SWCNTs, a metal catalyst, such as cobalt, must be added to the system.

Laser ablation is a similar method for manufacturing CNTs (Dresselhaus et al., 2000). A laser is used to ablate a carbon target within a tube furnace heated to 1200°C. The MWCNTs created from the process are carried down the chamber on an inert gas and then collected. SWCNTs are created if the carbon target contains small amounts of nickel and cobalt. Varying the temperature changes the diameters of the carbon nanotubes.

The final method used for manufacturing carbon nanotubes is chemical vapor deposition (CVD). In CVD, a metal catalyst such as nickel or cobalt is placed on a substrate in a tube furnace. The catalyst is heated at a high temperature (500-1000°C) while a hydrocarbon gas flows through the furnace (Dresselhaus et al., 2000). During the process, CNTs grow on the catalyst. The properties of the CNTs (single vs. multiwall, diameter) are determined by the catalyst used and the temperature of the furnace. This method has the most potential for commercial applications because the nanotubes can be grown on an existing substrate rather than being attached to new material after a collection process.

4. Electrical Models of Carbon Nanotube Interconnects

This section summarizes the primary electrical models that have been developed for SWCNTs. An understanding of these models is essential for the development of our CNT capacitor models. The (Naeemi & Meindl, 2007) SWCNT electrical model will be used to determine the electrical models for our CNT capacitors. However, for completeness, we also present two other models. The models in this section were based on the work that developed transmission line and *RLC* models for ideal SWCNTs (Burke, 2002; Burke, 2003).
An equivalent *RLC* circuit model for SWCNTs originally developed for interconnect applications is shown in Figure 2 (Naeemi & Meindl, 2007). This model is valid for isolated metallic SWCNTs of all lengths used in interconnect applications.

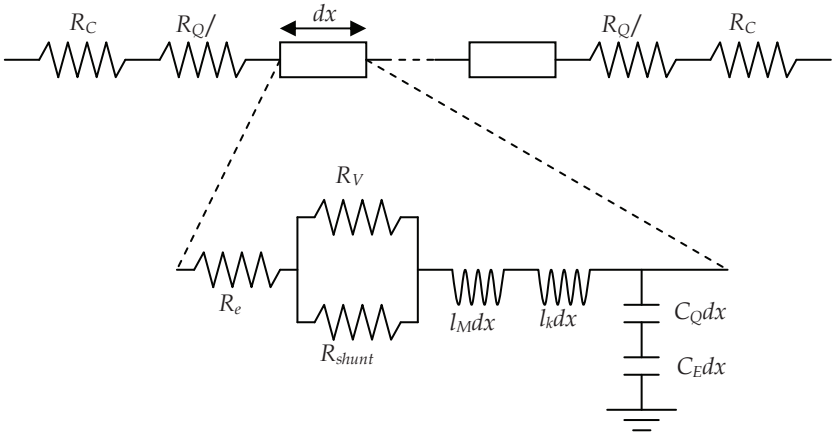


Fig. 2. Equivalent circuit for metallic SWCNTs developed in (Naeemi & Meindl, 2007).

The resistance components in the model include a lumped resistance at the two ends of the SWCNT and a distributed resistance throughout the SWCNT. The lumped resistance consists of the contact resistances (R_C) and the quantum resistance per channel (R_Q). The contact resistance can vary depending on the quality of the connection. If the connection is poor, the electron scattering at the contacts can cause a significant increase in the resistance. The quantum resistance is the minimum resistance associated with a quantum wire, which for a SWCNT is 6.45kΩ. It can be combined with the distributed resistances (R_e, R_V, and R_{shunt}) and simplified to form the overall resistance, R, which can be approximated by the equation:

$$R = R_Q\left(1+\frac{\ell}{\ell_{mpf,low}}\right)+\frac{V_{DROP}}{I_0} \tag{2}$$

where ℓ is the length of the SWCNT, $\ell_{mfp,low}$ is the SWCNT low-bias mean free path, V_{DROP} is the voltage drop across the SWCNT, and I_0 is the low-bias current through the SWCNT.
The capacitive components include a quantum capacitance (C_Q) in series with an electrostatic capacitance (C_E). The quantum capacitance is approximately 100aF/μm of CNT length. The electrostatic capacitance is typically much smaller than the quantum

capacitance. Since the two capacitances are in series, the quantum capacitance has little impact on the overall capacitance.

The kinetic (l_k) and magnetic (l_M) inductances are the inductive components of the SWCNT. The kinetic inductance, which is the kinetic energy of the electrons in the nanotube, is approximately 4nH/μm. This value is orders of magnitude larger than the magnetic inductance, so l_M can be disregarded when calculating the overall inductance in the SWCNT model.

To further improve the conductive performance of SWCNTs, they can be arranged in bundles. Fortunately, many of the electrical properties associated with isolated SWCNTs can be extended to SWCNTs bundles. For example, (Naeemi & Meindl, 2007) computes the conductivity of a SWCNT bundle using

$$\sigma_{SWCNT} = (n / A_{bundle}) / \left(\frac{R_C}{\ell} + R_Q \left(\frac{1}{\ell} + \frac{1}{\ell_{eff}} \right) \right) \tag{3}$$

where ℓ_{eff} is the mean free path length of electrons in the SWCNTs, n is the number of SWCNTs in the bundle, A_{bundle} is the cross sectional area of the bundle, and R_C is the contact resistance which quantifies the quality of the connections at each end. The quantum capacitance of SWCNT bundles is calculated by summing the quantum capacitances of all metallic SWCNTs in the bundle. Since all the nanotubes in the bundle have the same potential, the electrostatic capacitances between nanotubes can be ignored and, similar to metal interconnects, only the electrostatic capacitances to ground and neighboring bundles must be considered. The electrostatic capacitance values are again significantly less than the total quantum capacitance. Since the two capacitances are still in series, the quantum capacitance again has only a slight impact on the overall SWCNT bundle capacitance, however, it will be included in our development of the CNT capacitor electrical model.

For comparison, the circuit model by (Srivanstava & Banerjee, 2005) is given in Figure 3. It assumes that the length of the SWCNT is less than the mean free path of electrons and that there are ideal contacts at each end.

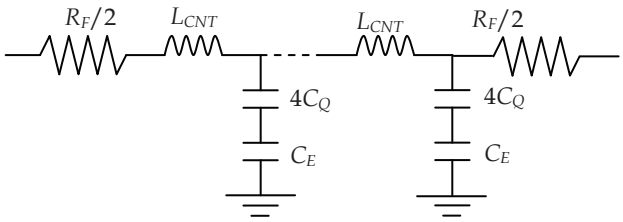


Fig. 3. Equivalent circuit for isolated SWCNTs developed by (Srivanstava & Banerjee, 2005).

The above model is similar to Figure 2 for isolated SWCNTs in that both have a lumped and distributed portion of the model. The inductance, L_{CNT}, contains both the mutual and kinetic inductance, which is also similar to the (Naeemi & Meindl, 2007) model. The main difference between the two SWCNT models is their quantum capacitance. In the (Srivanstava & Banerjee, 2005) model, C_Q is multiplied by a factor of four to account for the four conducting channels found in SWCNTs (with $4C_Q$ = 388aF/μm of SWCNT length).

There are also differences between the (Srivanstava & Banerjee, 2005) and (Naeemi & Meindl, 2007) models of SWCNT bundles for the calculation of the bundle resistance. The (Srivanstava & Banerjee, 2005) model assume that all SWCNTs in the bundle are metallic (and conducting), which makes the overall bundle resistance equal to the resistance of an isolated SWCNT divided by the number of tubes in the bundle.
The final model for comparison was developed by (Nieuwoudt et al., 2007) and is shown in Figure 4. They have created an *RLC* model that can be scaled to include bundles of SWCNTs.

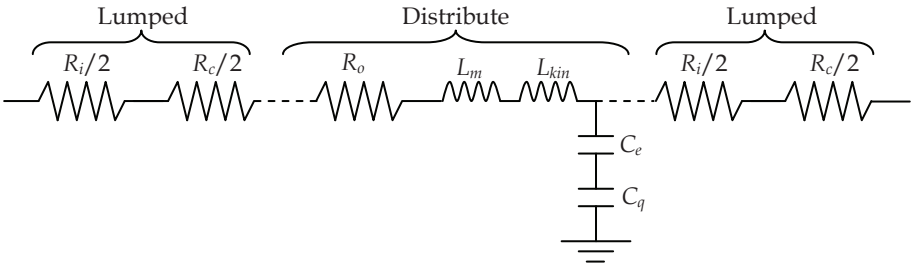


Fig. 4. Equivalent *RLC* circuit for isolated SWCNTs developed by (Nieuwoudt & Massoud, 2006a, 2006b).

The lumped portion of the model contains both an intrinsic and a contact resistance (R_i and R_c), while the distributed portion has an ohmic resistance (R_o), the kinetic and magnetic inductance (L_{kin} and L_m) and the electrostatic and quantum capacitance (C_e and C_q) of the SWCNT. This model of an isolated SWCNT is similar to the previous two models except that it also includes an ohmic resistance of the SWCNT. In addition, the model uses the diameters of the SWCNT as a variable affecting the ohmic and contact resistances (Nieuwouldt & Massoud, 2006a). Also, the model also considers the entire current loop when modelling the magnetic inductance. They argue this value can be as large as the kinetic inductance in some interconnect applications, and that it must also be considered when computing the overall inductance (Nieuwouldt & Massoud, 2006b). Finally, the capacitance of a bundle of SWCNTs is represented as a single conductor with a given width and height, which significantly decreases the computation needed to calculate the overall capacitance.

5. Carbon Nanotube Capacitor Structures

One potential application for SWCNT bundles in future integrated circuits is as replacements for vertical interconnect (via) structures. In such structures, the SWCNT bundle will connect two consecutive horizontal metal layers (see Figure 5). (Nihei et al., 2004) has shown how bundles of CNTs can be implemented as vias. They fabricated a bundle of about one thousand CNTs between two metal layers and observed no drop in current density through the via over a one hundred hour period.

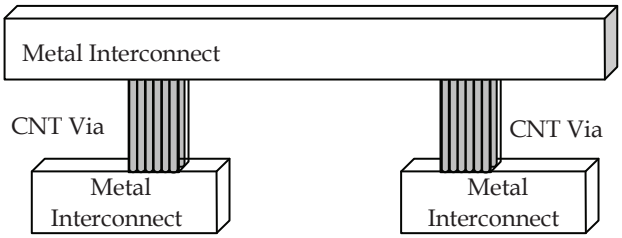


Fig. 5. Example of how a bundle of single wall carbon nanotubes can be used as vertical interconnect (via) structures in future integrated circuits.

If we refer back to Figure 5, we can readily see that by cutting the upper metal layer, we can create a vertical CNT capacitor (Wood & Budnik, 2007) (see Figure 6).

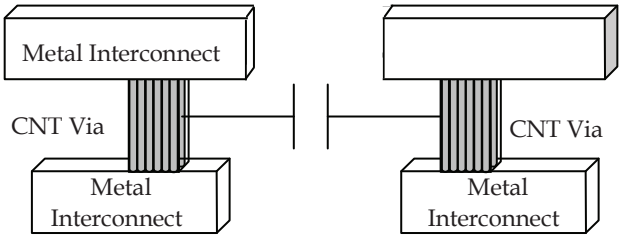


Fig. 6. Example of how a vertical carbon nanotube capacitor can be constructed from two carbon nanotube vias. (Wood & Budnik, 2007)

In the following subsections, we will examine three different structures for CNT capacitors:
- CNT vertical parallel plate capacitor
- CNT vertical bundle capacitor
- CNT vertical interleaved sheet capacitor

In each subsection, we show how the capacitors can be constructed. In addition, we develop a three element electrical model (Budnik et al., 2006) for each capacitor (see Figure 7). R_{CNCAP} is the CNT capacitor's equivalent series resistance, L_{CNCAP} is the CNT capacitor's equivalent series inductance, and C_{CNCAP} is the device's capacitance.

5.1 Carbon nanotube vertical parallel plate capacitor

The first CNT capacitor we introduce is the vertical parallel plate capacitor (Wood & Budnik, 2007). For this device, two bundles of CNTs are grown vertically in close proximity

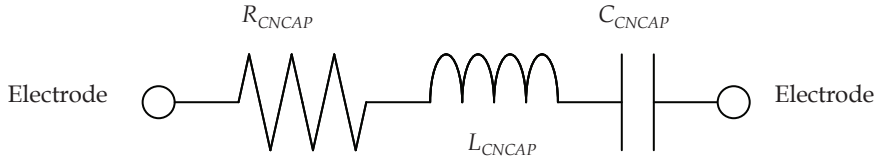


Fig. 7. Three element electrical model of a carbon nanotube capacitor (Budnik et al., 2006) to each other (see Figure 8). The bundles have a separation distance, s, width, w, and height, h. Each CNT has a radius, r.

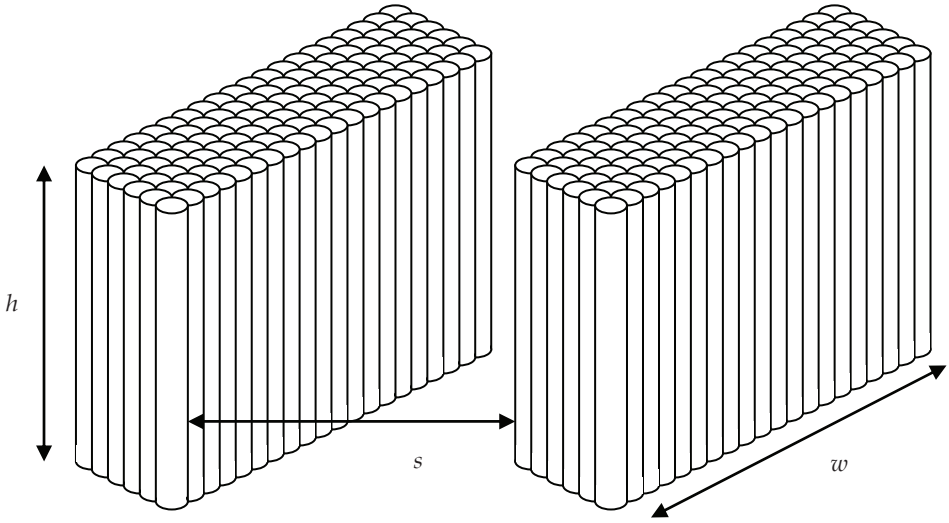


Fig. 8. Carbon nanotube vertical parallel plate capacitor.

The only difference between the ideally packed bundles in Figure 8 and perfectly smooth plates is the surface roughness of the bundles. When r is less than an order of magnitude less than s, however, the impact of the surface roughness on the electrostatic coupling capacitance is less than 3% (Naeemi & Meindl, 2007), based upon simulations with the field solver, RAPHAEL (RAPHAEL, 1999). This is consistent with the models developed by (Bruce et al., 1999). Therefore, we take the electrostatic coupling capacitance of the structure from (1) as:

$$C_{Plates} = \frac{wh\epsilon}{s} \tag{4}$$

For the two bundles, the number of SWCNTs on each vertical parallel plate is given by:

$$N_{Plate} = \frac{w}{2r} \tag{5}$$

where r is the radius of the individual SWCNTs. For SWCNTs, r is typically 0.5nm (Dresselhaus, 2000).

Since C_Q is in units of F/unit length, the total quantum capacitance of each bundle is:

$$C_{Q,Bundle} = N_{Plate}\left(hC_Q\right) \tag{6}$$

With C_{Plates} in series with the quantum capacitance of each bundle, the capacitance of a vertical SWCNT parallel plate capacitor is:

$$C_{CNCAP} = \frac{1}{\frac{1}{C_{Q,Bundle}} + \frac{1}{C_{Plates}} + \frac{1}{C_{Q,Bundle}}} \tag{7}$$

Substituting (4-6) into (7) yields:

$$C_{CNCAP} = \frac{1}{\frac{1}{N_{Plate}\left(hC_Q\right)} + \frac{1}{\frac{wh\in}{s}} + \frac{1}{N_{Plate}\left(hC_Q\right)}} \tag{8}$$

Simplifying (8) results in:

$$C_{CNCAP} = \frac{h}{\frac{2}{N_{Plate}\left(C_Q\right)} + \frac{s}{w\in}} \tag{9}$$

As an example, if we let $h = 1\mu$m, $\in = 3.9(8.854\text{x}10^{-12}\text{F/m})$, and $C_Q = 388\text{aF}/\mu\text{m}$, we can plot C_{CNCAP} vs. s and w as shown in Figure 9. At an example point in this graph where $s = 10$nm and $w = 100\mu$m, $C_{CNCAP} = 18.3$fF.

Next, we turn our attention to the equivalent series resistance and inductance of the vertical SWCNT parallel plate capacitor. If the current in each SWCNT continues to flow in only a vertical direction inside each bundle, then the total number of SWCNTs conducting in each bundle is again N_{Plate}. Using a lumped model for the device, R_{CNCAP} and L_{CNCAP} will be given by:

$$R_{CNCAP} = \frac{1}{N_{Plate}}\left[2R_C + R_Q\left(1 + \frac{h}{\ell_{mpf,low}}\right) + \frac{V_{DROP}}{I_0}\right] \tag{10}$$

$$L_{CNCAP} = \frac{hL_k}{N_{Plate}} \tag{11}$$

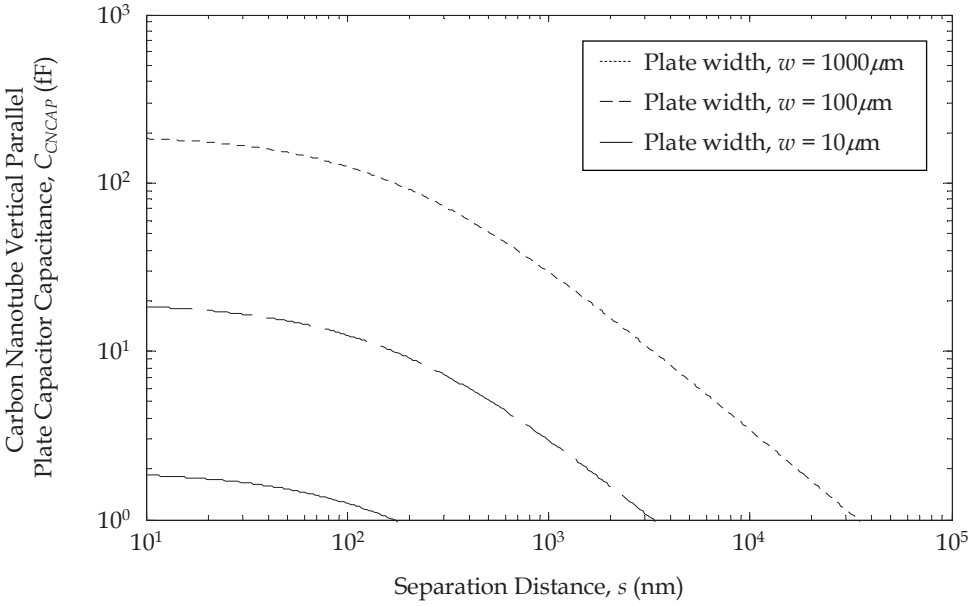


Fig. 9. Carbon nanotube vertical parallel plate capacitor capacitance, C_{CNCAP}, vs. plate separation distance, s, and plate width, w.

5.2 Carbon nanotube vertical bundle capacitor

From the CNT capacitor structure in the above sub-section, we can readily imagine an array of CNT bundles alternately connected to the capacitor's anode (A) and cathode (C). The capacitance between nearest neighbor bundles is C_{Bundle}. The cross section of the device is shown in Figure 10 (Wood & Budnik, 2007, Budnik & Johnson, 2009). The bundles have four sides of width w and are of height h. The inter-bundle spacing is s.

Extrapolating from the previous section, the electrostatic coupling capacitance C_{Bundle} is:

$$C_{Bundle} = \frac{wh\epsilon}{s} \tag{12}$$

For the bundles, the number of CNTs on each bundle side is given by:

$$N_{Bundle\ Side} = \frac{w}{2r} \tag{13}$$

Since C_Q is in units of F/unit length, the total quantum capacitance of each bundle face is again:

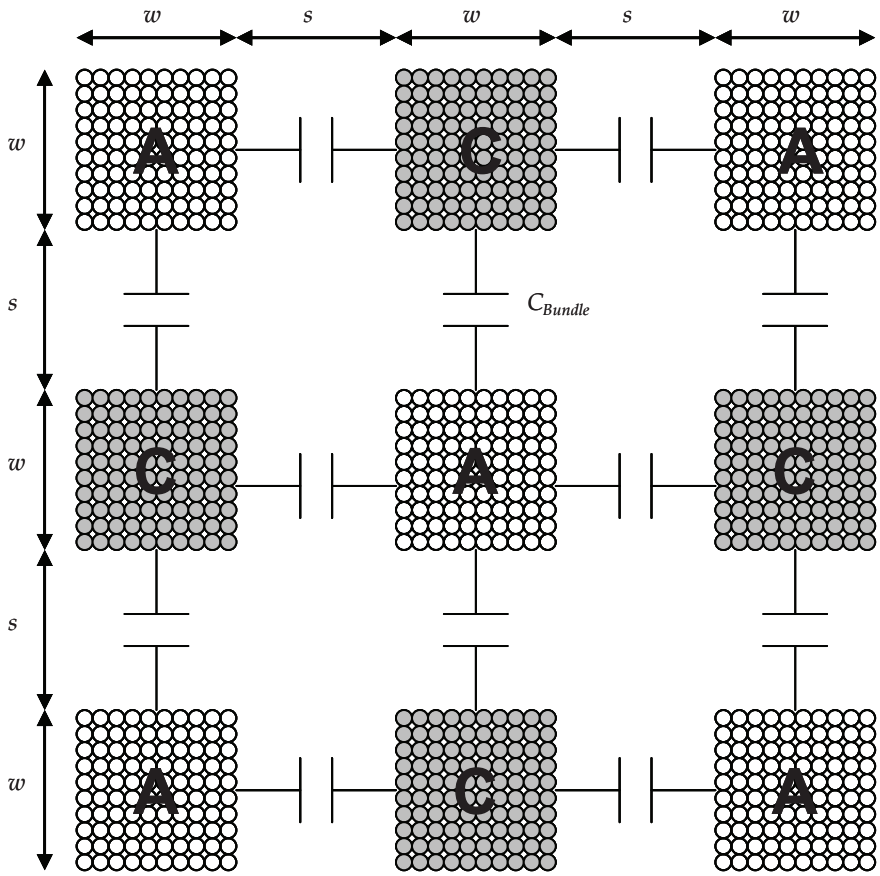


Fig. 10. Cross section of a carbon nanotube vertical bundle capacitor. Bundles are alternately connected to the device anode (A) or cathode (C).

$$C_{Q,Bundle} = N_{Bundle\ Side}\left(hC_Q\right) \tag{14}$$

With C_{Bundle} in series with the quantum capacitance of each bundle face, from (12-14) the capacitance between two nearest neighbor SWCNT bundle faces is:

$$C_{Bundle\ Faces,Total} = \frac{1}{\frac{1}{C_{Q,Bundle}} + \frac{1}{C_{Bundle}} + \frac{1}{C_{Q,Bundle}}} \tag{15}$$

$$C_{Bundle\ Faces,Total} = \frac{1}{\frac{1}{N_{Bundle\ Side}\left(hC_Q\right)} + \frac{1}{\frac{wh\in}{s}} + \frac{1}{N_{Bundle\ Side}\left(hC_Q\right)}} \tag{16}$$

Simplifying (16) results in:

$$C_{Bundle\ Faces,Total} = \frac{h}{\frac{2}{N_{Plate}\left(C_Q\right)} + \frac{s}{w\epsilon}} \tag{17}$$

To determine C_{CNCAP}, we need to determine the number of bundles per unit area. The unit cell of the vertical SWCNT bundle capacitor shown above in Figure 10 is:

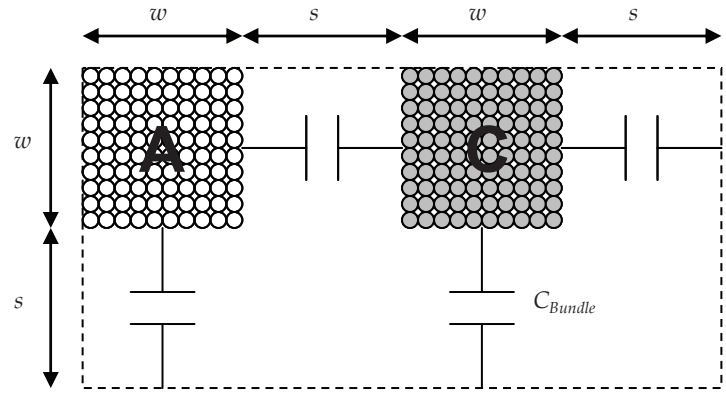


Fig. 11. Unit cell of vertical bundle capacitor in Figure 10.

The area of the unit cell is:

$$A_{Unit\ Cell} = \left(2w + 2s\right)\left(w + s\right) \tag{18}$$

When $w = s$, (18) simplifies to:

$$A_{Unit\ Cell} = \left(4w\right)\left(2w\right) = 8w^2 \tag{19}$$

There are four capacitors in the vertical SWCNT bundle capacitor unit cell. Neglecting fringe effects, from (17) and (19) the total capacitance per unit area of the device is:

$$C_{CNCAP} = \frac{4C_{Bundle\ Faces,Total}}{8w^2} \tag{20}$$

$$C_{CNCAP} = \frac{1}{2w^2}\left(\frac{h}{\frac{2}{N_{Plate}\left(C_Q\right)} + \frac{1}{\epsilon}}\right) \tag{21}$$

As an example, we again let r = 0.5nm, ϵ = 3.9(8.854x10^{-12}F/m), and C_Q = 388aF/μm. We plot C_{CNCAP} vs. s and h in Figure 12. At an example point in this graph where $s = w$ = 10nm and h = 1μm, C_{CNCAP} = 169fF/μm^2. This is more than one order of magnitude higher than the 2022 projections for MOSC and MIM devices (ITRS, 2008). If taller bundles of the SWCNTs are used in the CNT vertical bundle capacitor, the capacitance per unit area will increase linearly.

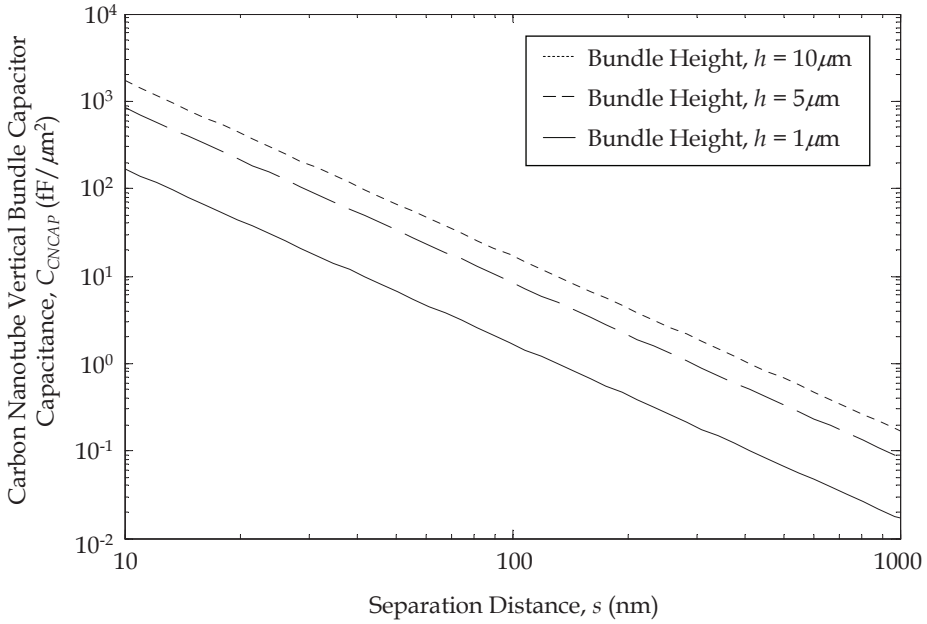


Fig. 12. Carbon nanotube vertical bundle capacitor capacitance per unit area (C_{CNCAP}) vs. bundle separation distance, s, and bundle height, h.

Next, we calculate the equivalent series resistance and inductance of the vertical SWCNT bundle parallel plate capacitor. Again, if the current in each SWCNT continues to flow in only a vertical direction inside each bundle, then the total number of SWCNTs conducting in each bundle face is again:

$$N_{Bundle\ Face} = \frac{w}{2r} \tag{22}$$

where r is the radius of the SWCNTs. However, because there are four capacitors in each unit cell, the equivalent series resistance-unit area of the vertical SWCNT bundle capacitor is:

$$R_{CNCAP} = \frac{A_{Unit\ Cell}}{4N_{Bundle\ Face}}\left[2R_C + R_Q\left(1 + \frac{h}{\ell_{mpf,low}}\right) + \frac{V_{DROP}}{I_0}\right] \tag{23}$$

Likewise, the equivalent series inductance-unit area is:

$$L_{CNCAP} = \frac{A_{Unit\ Cell}}{4N_{Bundle\ Face}}\left(h\,L_k\right) \tag{24}$$

5.3 Carbon nanotube vertical interleaved sheet capacitor

The cross section of a third CNT capacitor structure is shown in Fig. 13. It consists of densely packed, thin vertical plates of CNTs connected to alternating electrode polarities

(anode and cathode) (Budnik et al, 2008). The plates are formed by a single row of CNTs with radius *r* and height *h*. The plate width is *w* and the spacing between adjacent plates is *s*.

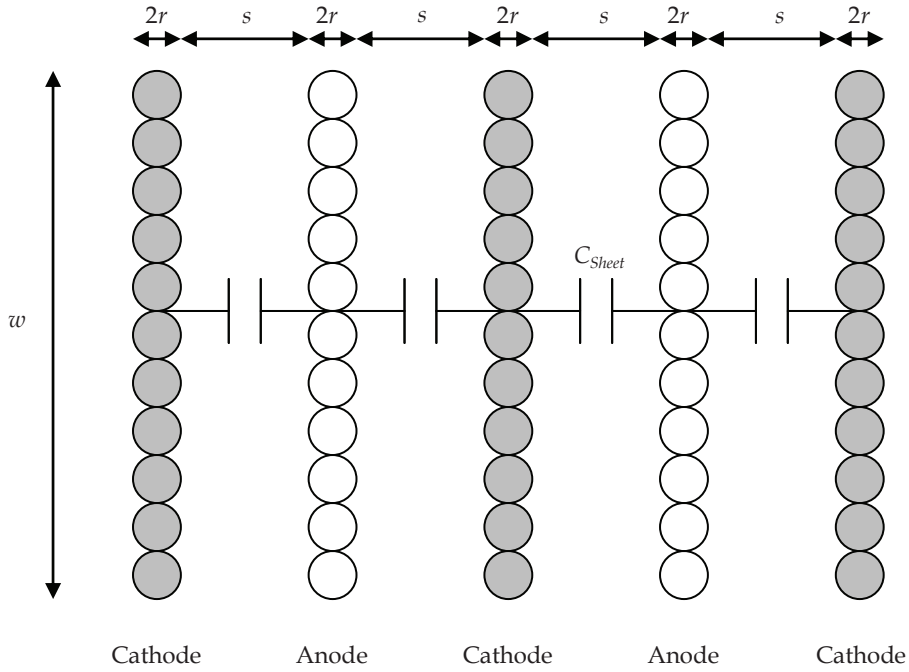


Fig. 13. Cross section of a carbon nanotube vertical interleaved sheet capacitor. Sheets are alternately connected to the device anode or cathode.

C_{Sheet} is given by:

$$C_{Sheet} = \frac{wh\in}{s} \tag{25}$$

The number of SWCNTs in each vertical sheet is given by:

$$N_{Sheet} = \frac{w}{2r} \tag{26}$$

where *r* is the radius of the individual SWCNTs.
For SWCNTs, since C_Q is in units of F/unit length, the total quantum capacitance of each sheet is:

$$C_{Q,Sheet} = N_{Sheet}\left(hC_Q\right) \tag{27}$$

With C_{Sheet} in series with the quantum capacitance of each sheet, the total capacitance between two vertical sheets is:

$$C_{2\,Sheets} = \frac{h}{\frac{2}{N_{Sheet}C_Q} + \frac{s}{w\epsilon}} \tag{28}$$

To determine C_{CNCAP}, we need to determine the number of sheets per unit area. The area of the unit cell of the vertical SWCNT interleaved sheet capacitor in Figure 13 is:

$$A_{Unit\ Cell} = \left[2(2r) + 2s\right](w) \tag{29}$$

There are two capacitors in the vertical SWCNT interleaved sheet capacitor unit cell. Neglecting fringe effects, from (28) and (29) the total capacitance per unit area of the device is:

$$C_{CNCAP} = \frac{2C_{2\,Sheets}}{\left[2(2r) + 2s\right](w)} \tag{30}$$

$$C_{CNCAP} = \frac{1}{2r + s}\left(\frac{h}{\frac{2w}{N_{Sheet}C_Q} + \frac{s}{\epsilon}}\right) \tag{31}$$

As an example, we again let h = 1μm, r = 0.5nm, ϵ = 3.9(8.854x10^{-12}F/m), and C_Q = 388aF/μm. We plot C_{CNCAP} vs. s in Fig. 14. At an example point in this graph where s = 10nm and w = h = 1μm, C_{CNCAP} = 282fF/μm^2. This is again more than one order of magnitude higher than the 2022 projections for MOSC and MIM devices (ITRS, 2008).

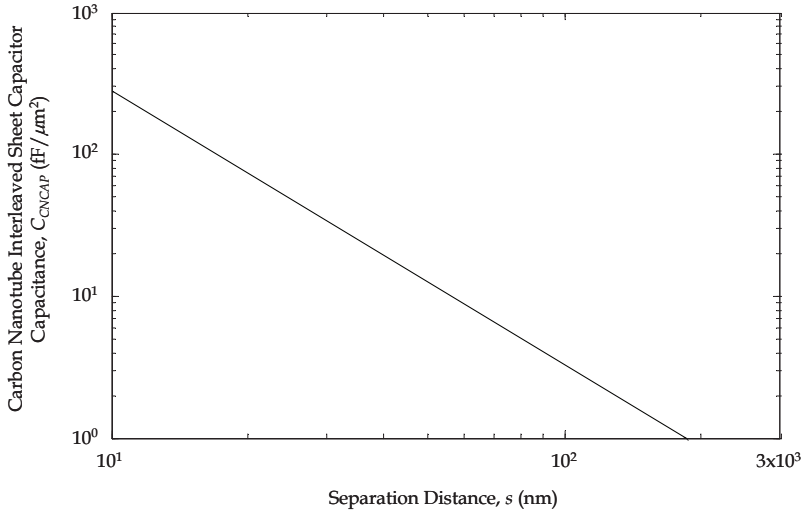


Fig. 14. Carbon nanotube vertical interleaved sheet capacitor capacitance per unit area (C_{CNCAP}) vs. bundle separation distance, s.

Finally, we turn our attention to the equivalent series resistance and inductance of the vertical SWCNT interleaved sheet capacitor. Using a lumped model for the device, R_{CNCAP} and L_{CNCAP} will be given by:

$$R_{CNCAP} = \frac{1}{N_{Sheet}}\left[2R_C + R_Q\left(1 + \frac{h}{\ell_{mpf,low}}\right) + \frac{V_{DROP}}{I_0}\right] \tag{32}$$

$$L_{CNCAP} = \frac{h\,L_k}{N_{Sheet}} \tag{33}$$

6. Conclusion

We began by examining the properties and shortcomings of various parallel plate capacitors used in traditional integrated circuits. Based on these shortcomings, it is projected that MOSC and MIM devices will exhibit capacitances per unit area of less than 15fF/μm^2 by 2022. Therefore, because of their attractive properties, we investigated the feasibility of using a vertical capacitor structure with CNT electrodes in future integrated circuit technologies.

Three separate CNT capacitor structures and their electrical models were presented. The first device was the CNT vertical parallel plate capacitor. While it had a limited capacitance per unit area, it was a simple transition from the CNT vertical parallel plate capacitor to our second device, the CNT vertical bundle capacitor. Through our analysis, we demonstrated how the CNT vertical bundle capacitor can exhibit capacitances per unit area of 169fF/μm^2 (based upon a 1μm tall structure). Finally, we presented a CNT interleaved sheet capacitor which can demonstrate an improved capacitance per unit area of 282fF/μm^2 (again based upon a 1μm tall structure).

CNT fabrication in integrated circuit technologies is still in its infancy. However, vertical CNT electrode capacitors have the potential of significantly improving the options available for future integrated circuit designers.

7. References

Bruce, N. C.; Garciá-Valenzuela, A.; & Kouznetsov, D. (1999). Rough-surface capacitor: Approximations of the capacitance with elementary functions. *J. Phys. D, Appl. Phys.*, vol. 32, no. 20, (Oct. 1999), pp. 2692–2702, ISSN 1361-6463.

Budnik, M. M.; Raychowdhury, A. ; Bansal, A.; & Roy, A. (2006). A High Density, Carbon Nanotube Capacitor for Decoupling Applications, *Proceedings of the ACM Design Automation Conference*, pp. 935-938, ISBN: 1-59593-381-6, San Francisco, California, USA, July, 2006.

Budnik, M. M.; Johnson, E. W.; & Wood, J. D. (2008). Electrical Models for Vertical Carbon Nanotube Capacitors, *Proceedings of the ACM Great Lakes Symposium on VLSI*, pp. 367-370, ISBN:1-58113-677-3, Orlando, Florida, USA, May, 2008.

Budnik, M. M.; Johnson, E. W.; & Wood, J. D. (2008). A Thin, Vertical, Parallel-Plate Capacitor with Multi-Wall Carbon Nanotube Electrodes, *Proceedings of the IEEE Conference on Nanotechnology*, pp. 274-276, ISBN: 978-1-4244-2103-9, Arlington, Texas, USA, August, 2008.

Budnik, M. M.; Johnson, E. W. (2009). A Carbon Nanotube Capacitor, To appear in *Proceedings of the IEEE Nanotechnology Materials and Devices Conference*, Traverse City, Michigan, USA, June, 2009.

Burke, P. J. (2002), Luttinger liquid theory as a model of the gigahertz electrical properties of carbon nanotubes, *IEEE Transactions on Nanotechnology*, vol. 1, no. 3, (Sept. 2002), pp. 129–144, ISSN: 1536-125X.

Burke, P. J. (2003), An RF Circuit Model for Carbon Nanotubes, *IEEE Transactions on Nanotechnology*, vol. 2, no. 1, (March 2003), pp. 55-58, ISSN: 1536-125X.

Dresselhaus, M. S.; Dresselhaus, G.; & Avouris, P.; Eds. (2000). *Topics in Applied Physics, Carbon Nanotubes: Synthesis, Structure, Properties and Applications*, Springer-Verlag, ISBN 978-3540728641, New York.

International Technology Roadmap for Semiconductors (ITRS) (2008). International Technology Roadmap for Semiconductors, 2008 Update. Online: www.itrs.net/ Links/2008ITRS/Home2008.htm.

Naeemi, A.; Sarvari, R.; & Meindl, J. D. (2004). Performance comparison between carbon nanotube and copper interconnects for GSI, *IEDM Technical Digest*, pp. 699–702, San Francisco, California, USA, (December, 2004), ISBN: 0-7803-8684-1.

Naeemi, A.; Sarvari, R.; & Meindl, J. D. (2005). Performance comparison between carbon nanotube and copper interconnects for gigascale integration (GSI). *IEEE Electron Device Letters*, vol. 26, no. 2, (February 2005), pp. 84–86, ISSN: 0741-3106.

Naeemi, A.; Meindl, J. D. (2007). Design and Performance Modeling for Single-Walled Carbon Nanotubes as Local, Semiglobal, and Global Interconnects in Gigascale Integrated Systems, *IEEE Transactions on Electron Devices*, vol. 54, no. 1, (Jan. 2007), pp. 26 37, ISSN: 0018-9383.

Naeemi, A; Meindl, J. D. (2008). Performance Modeling for Single- and Multiwall Carbon Nanotubes as Signal and Power Interconnects in Gigascale Systems, *IEEE Electron Devices*, vol. 55, no. 10, (October, 2008), pp. 2574-2582, ISSN: 0741-3106.

Nieuwoudt, A.; Mondal, M. & Massoud, Y. (2007) Predicting the Performance and Reliability of Carbon Nanotube Bundles for On-Chip Interconnect, in *Proceedings of the IEEE Asia and South Pacific Design Automation Conference*, pp. 708–713, ISBN: 1-4244-0630-7, Yokohama, Japan, Jan. 2007.

Nieuwoudt, A.; Massoud, Y. (2006) Accurate Resistance Modeling or Carbon Nanotube Bundles in VLSI Interconnect, *Proceedings of the IEEE Conference on Nanotechnology*, pp. 288-291, ISBN: 1-4244-0077-5, Cincinnati, OH, USA, June 2006.

Nieuwoudt, A.; & Massoud, Y. (2006) Scalable Modeling of Magnetic Inductance in Carbon Nanotube Bundles for BLSI Interconnect, *Proceedings of the IEEE Conference on Nanotechnology*, pp. 254-257, ISBN: 1-4244-0077-5, Cincinnati, OH, USA, June 2006.

Nihei, M.; Horibe, M.; Kawabata, A.; Awano, Y. (2004) Carbon Nanotube Vias for Future LSI Interconnects, *Proceedings of the 2004 International Interconnect Technology Conference*, pp. 251 – 253, ISBN: 0-7803-8308-7, Burlingame, CA, USA, June 2004.,.

RAPHAEL, (1996). "Interconnect Analysis Program," TMA Inc.

Raychowdhury, A.; & Roy, K.; (2004). A circuit model for carbon nanotube interconnects: Comparative study with Cu interconnects for scaled technologies, *Proceedings of the IEEE International Conference on Computer Aided Design*, pp. 237–240, ISBN: 0-7803-8702-3, San Jose, CA, USA, November, 2004..

Srivanstava, N.; & Banerjee, K. (2005) Performance Analysis of Carbon Nanotubes Interconnects for VLSI Applications, *Proceedings of the IEEE International Conference on Computer Aided Design*, pp. 383–390, ISBN: 0-7803-9254-X, San Jose, CA, USA, Nov. 2005..

Wood, J. D.; Budnik, M. M. (2007). A Carbon Nanotube Capacitor Structure, *Proceedings of the IEEE International Semiconductor Device Research Symposium*, ISBN: 978-1-4244-1892-3, pp. 1-2, College Park, Maryland, USA, December, 2007.

Zhang, C; Cott, D; Chiodarelli, N.; Vereecken P.; Robertson J. & Whelan, C. M. (2008). Growth of Carbon Nanotubes as Horizontal Interconnects, *Physica Status Solidi (b)*, Vol. 245, No. 10, (October 2008), pp. 2308-2310, ISSN: 0370-1972.

18

Electrochemical Co-deposition of Metal-Nanoparticle Composites for Microsystem Applications

Kyle Jiang
School of Mechanical Engineering, University of Birmingham, UK, B15 2TT

1. Introduction

Electrochemical deposition has been regarded as one of the key disciplines to microtechnology [1], as it provides a way to produce high precision microcomponents in various pure metals, alloys and composites in cost effective processes. Among the first microdevices demonstrated, were the electrochemical deposition of lithographically patterned features introduced in the early 1960's by IBM to increase memory density and reduce the physical size of memory devices [2]. The electrochemical deposition onto microfabricated template is also known as electrodeposition and microelectroforming in literature.

Electrochemical deposition is widely used for producing metallic components from sub-millimeters to a few millimeters in overall dimensions and up to a few hundred microns in thickness, i.e. in a range other metal deposition methods cannot challenge. As electrodeposition is particularly suitable for producing thick metallic microstructures, early efforts in this area were concentrated in producing relatively deep micromoulds. A renown process in this field is LIGA, pioneered in Germany in the early 1980s [3]. LIGA is the abbreviation of three German words representing three process steps, i.e. lithography (Lithographie), electroforming (Galvanoformung) and moulding (Abformung). It has been applied to fabrication of metallic microstructures and plastic replicates.

LIGA fabricated Ni microstructures and microcomponents have been applied in micro sensors and actuators. However, applications in harsh service conditions such as high temperature, high pressure, constant corrosion and friction put forward a requirement of better material properties on LIGA Ni, as numerous studies in characterizing LIGA fabricated samples have showed the insufficient mechanical and tribological properties of pure Ni. For example, Ni has been found to have high friction coefficient during the wear testing of LIGA fabricated microrotors [4]. It implies that Ni lifetime was well below the required level. The high friction coefficient of Ni also leads to problems in obtaining dimensionally accurate microcomponents with high aspect ratio during the plastic moulding process [5]. The stress-life experiments have shown that LIGA Ni structures have endurance limits that scale with tensile strength. [6].

Electroformed alloys and metal matrix composites (MMC) are promising alternatives to pure metals, which is proposed in literature for improving properties of LIGA Ni or replace Ni for microsystem applications. The former introduces alloy elements into Ni during electroforming process. For example, nanocrystalline nickel–tungsten (Ni-W) alloys are being considered as an attractive alternative to electroformed Ni for applications such as mould inserts and micro motors [7]. So far, however, only a few number of binary alloys, i.e. NiCu, NiFe, NiCo and NiP, have been attempted in LIGA fabrication of microstructures [8]. The latter introduces non-metallic particles usually to Ni matrix for improving the properties of the components. It is well documented in the literature that the solid and dispersed particulate composites possess a greatly enhanced mechanical properties compared with their metallic counterparts in strength, hardness, tribological properties and high temperature properties [9]. Recent research developments on nanomaterials provide scientists and engineers many new opportunities to synthesize nanocomposites with enhanced properties and design novel devices for microsystems. This chapter gives a brief introduction to this research with the latest results reported in literature.

2. Fabrication of Micromoulds in Photoresist

Central to the LIGA process is the fabrication of high quality LIGA moulds, and the X-ray sensitivity of thick resists is very critical to the LIGA process **Error! Reference source not found.**]. Polymethylmethacrylate (PMMA) was widely used in deep X-ray lithography initially. It is not highly sensitive to X-rays and a thick piece of PMMA sheet (>1mm) needs a long time to get exposed and results in excessive cost or fabrication difficulties **Error! Reference source not found.**]. In addition, the high costs for access to X-ray from a synchrotron limit the technique from wide applications. The commercial appearance of SU-8 in 90s started to change the situation. SU-8 is imageable with X-ray, UV and electron beam. Compared with PMMA, it requires much less exposure energy. In fabrication of 600µm standing microsctructures with X-ray exposure, PMMA requires a typical bottom dose of 4000 J/cm^3, while SU-8 requires 30 – 52 J/cm^3, only about a hundredth exposure energy of what PMMA requires [12]. Such high sensitivity to exposure proves sufficient even under UV exposure. Therefore, SU-8 photoresist gradually becomes the most popular resist for the LIGA process and UV-LIGA technique with wavelength of 356-405nm has been rapidly developed.

SU-8 is a negative tone, near-UV photoresist developed and patented by IBM in 1989 [13]. The photoresist is based on the EPON SU-8 epoxy resin from Shell Chemical. Depending on the percentage of the organic solvent and so the viscosity of the resist, the thickness of a single-spin SU-8 film can reach more than 500µm. SU-8 resist is well known for its excellent mechanical properties and low optical absorption in the near-UV range which results in uniform exposure conditions across the thickness [14] [15]. It has been reported that 1000µm thick microstructures with 40:1 aspect ratio was produced in a single lithography cycle [16] and a 1200µm thick microstructure with 18:1 aspect ratio in a double spin lithography process [17], both of which are of good vertical sidewalls and geometry. However, SU-8 is difficult to be stripped, and frequently, oxygen plasma process is required to remove SU-8 after electroforming. This property limits its applications to permanent components or micromoulds of only convex shapes most time. Meanwhile the needs for thick and strippable micromoulds lead to the development of some new resists.

Niedermann et al. [] presented DiaPlate 133 as a negative tone strippable photoresist for electroforming. A 60μm thick nickel foil with 32μm diameter holes and a 300μm thick nickel microgear were successfully electroformed using this resist. However, distortion seemed severe when the feature size was less than 50μm and a complicated long time soft bake is required when the thickness of the resist was greater than 200μm. Srinivasa Rao et al [] used JSR THB 151N resist, which requires a short time in soft baking, to develop 130μm thick and 2.6:1 aspect ratio electroforming moulds with near vertical sidewalls. Lee and Jiang [] studied KMPR, a photoresist newly developed by Microchem™, and microcomponents with an aspect ratio of 18:1 were fabricated.
BPR100 is another negative tone thick photoresist with a single-spin film thickness ranging from 40 to 130μm. It has been applied to a variety of electroplating and etching processes used in wafer level packaging. However, its dark blue colour absorbs irradiated beam energy, causing uneven exposure from top to bottom of a thick layer. This make it difficult to be used for thick moulds. The fabrication processes and the results of SU-8, KMPR and Shipley BPR 100 are presented as follows.

2.1 SU-8 Fabrication Process

SU-8 photoresist series have many grades for applications of different thickness. Take SU-8 2075 for example, it contains 73.45% solid and is intended for a thickness range from 75 to 225μm. In its process, the attention should be focused on softbake time and exposure dose for the optimum dimensional accuracy and sidewall profile of the microstructures. The processing parameters can be found in [21]. An optimized process for producing 500 μm thick SU-8 2075 moulds is used here as an example to introduce SU-8 process.
First, a 6ml SU-8 2075 photoresist is directly cast on a 4 inch silicon wafer coated with a thin layer of gold deposited by using thermal evaporation. After the photoresist spreads evenly, a soft bake at 65°C for 2 hrs and then 95°C for 8hrs is carried out on a levelled hotplate. The softbake step is performed to evaporate all the solvent in the resist and solidify the coated resist layer. UV exposure density is 12.6mW/cm^2. In the UV exposures, hard contacts are used between the SU-8 layer and the mask. A chrome mask is used in the exposure step. The total exposure dose is 3019mJ/cm^2. A postbake at 65°C for 2min follows and then 95°C for 20min. The polymerisation process occurs during the exposure and postbake steps. More specifically, the photoacid is generated in the resist during the exposure step and the crosslinking of epoxy groups takes place in the post bake step. The crosslinked SU-8 structures are then retained on the wafer after developing in the EC solvent for 10min.
Figure 1 shows an SU-8 mould for electroforming a microgear. Other SU-8 grades such as SU-8 2002, SU-8 2005 and SU-8 2050 are for different thicknesses. Their fabrication process are similar to SU-8 2075, but the processing parameters such as spin speed, baking time and exposure do need to be determined from experiments and optimised for each application case.

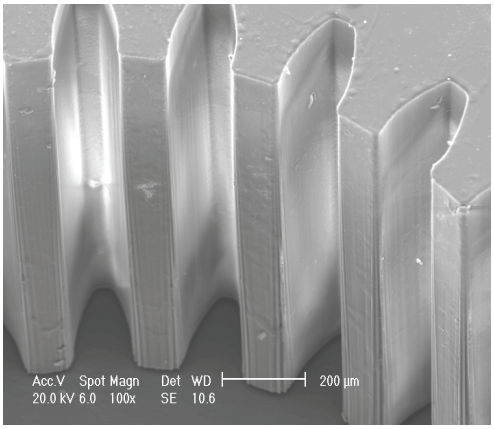


Fig. 1. SU-8 mould for producing microgears using electroforming.

2.2 Fabrication of KMPR micromoulds

KMPR photoresist is a chemically amplified negative tone epoxy photoresist. It was developed with the intention to resolve the stripping problem inherent with SU-8. Similar to SU-8, a KMPR process includes spin coat, prebake, exposure, post-exposure bake and development, and an additional resist stripping step is possible after electroforming. The particular KMPR series commercially available from MicroChem include 1005, 1010, 1025 and 1050 to cover thicknesses from 5 to 115 µm. More process details can be found from the MCC KMPR datasheet [22].

KMPR 1025 is taken as an example to demonstrate its process here. In a single layer lithography process, KMPR 1025 photoresist is first spun coated up to 60 µm thick on a metalized wafer and soft baked. If the required micromoulds are thicker than 60 µm, a multiple spin deposition can be applied. A multiple spin deposition process consists of spin deposition, soft bake, spin deposition and soft bake again, until the thickness reaches the requirement before the UV exposure, PEB and development steps.

In a single spin deposition process for making 60 µm KMPR photoresist layers, the following steps are used. (1) The spin deposition is conducted at 1000 rpm for 30 seconds. (2) The wafer is soft baked on a hotplate at 100oC for 20 mins. (3)Then the coated wafer is exposed to UV light at a density of 1135 mJ/cm2 using a chromium coated glass mask. The UV light source can be provided by a mask aligner with a wavelength range of 365 to 436 nm. (4) Post exposure bake (PEB) is carried out on a hotplate immediately after the exposure for 3 mins at 100°C. (5) The wafer is developed in 2.38% TMAH MF26A solution (Chestech, UK) for 6 mins in room temperature and the micromoulds should appear on the substrate. (6) Finally, the wafer is blown dry using N_2 gas. In a multiple layer deposition process, Steps (1) and (2) above will be repeated until the thickness reaches the requirement before moving to step (3). It is noted that the soft bake time for a multilayer deposition is different from that for a single layer. For the first layer, 100°C for 20 mins baking condition is adequate. However, for the second layer and onwards, 100°C for 30 mins baking time is more appropriate. The increase in baking time is understood as that the previous layer tends to soak some solution from the newly deposited resist and it takes longer time to evaporate the solution from the both layers.

Figure 2 illustrates an SEM image of a 180 μm thick micropiston mould made of KMPR. The resist was deposited in two spin and soft bake cycles, in which the spin deposition was carried out at 700 rpm for 30 seconds and the soft bake temperatures were at 100oC for 20 mins for the first layer and 30 mins for the second layer. The SEM images show that dimensions can be controlled at a high resolution and sidewall about in 90±1°.

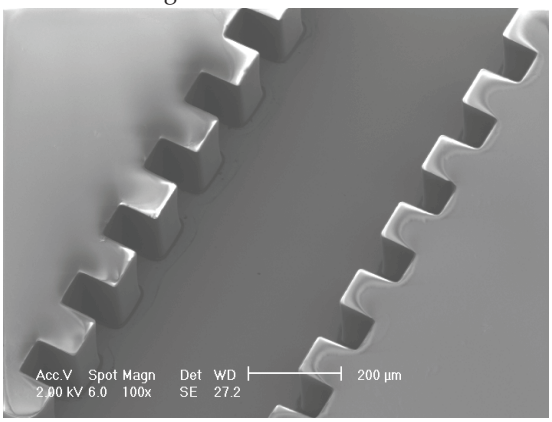


Fig. 2. An SEM image of a 180 μm thick KMPR photoresist moulds deposited using a multiple spin process.

2.3 BPR100 Fabrication Process

The fabrication process of BPR100 mould is slightly different from that of SU-8 and KMPR in that a post bake is not required before development. An optimized process of producing 100μm BPR100 micromoulds is introduced as an example.

A metalized substrate is first cleaned using acetone and isopropyl alcohol (IPA) separately. Then the substrate is pre-baked on a levelled hotplate at 120°C for 90s. After being naturally cooled down to room temperature, the substrate is spun coated with one layer of BPR100 photoresist. The coating consists of spreading spin and level spin. In the spread spin, the spin speed is ramped to 500 rpm at 100 rpm/second acceleration and held for 10 seconds. In the level spin, the speed is ramped to 800 rpm at 100 rpm/second acceleration and held for 30 seconds. This will result in a 100 μm layer. Further details for achieving different thickness can be found in [23]. A thicker layer can be achieved with a multilayer deposition process, involving repeated spin coating and baking cycle. After the photoresist had been coated, the substrate is soft baked to evaporate the solvent. This is done in two steps. First, the baking temperature is ramped from 20°C to 65°C at a rate of 5°C/min, held for 3 mins at 65°C, and then from 65°C to 95°C at the same rate, held for 7 mins at 95°C. After being cooled down, the baked substrate is then exposed using the same mask aligner for 80 seconds. The exposed substrate is subsequently developed in a solution of Eagle 2005 and deionized water in the ratio of 1:24 at 39°C for 5~10 mins depending on agitation of the developing solution. A hard bake is followed at 105°C for 5min. The resultant BPR100 moulds are examined using an SEM. Figure 3 shows two micromoulds made in Shipley BPR 100. The moulds are 200 μm deep and have straight sidewalls, providing excellent conditions for forming quality thick metallic components.

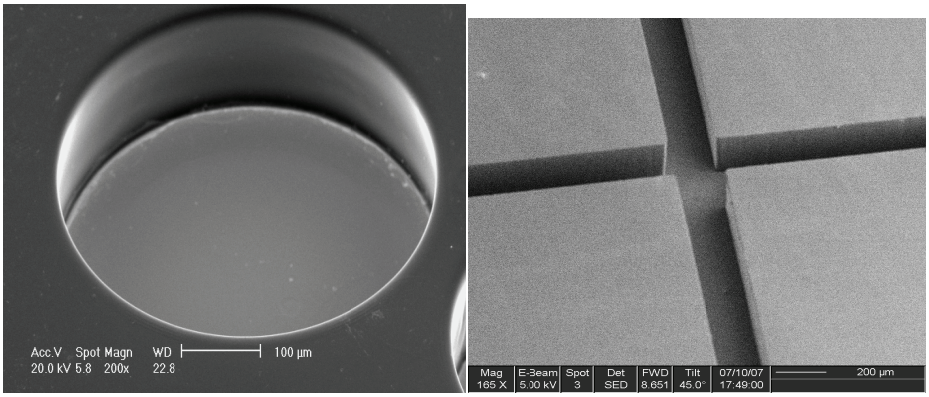


Fig. 3. Micromoulds made in Shipley BPR 100 resist.

3. Electrochemical co-deposition of metal/nanoparticles

3.1 Electrochemical Deposition

In a typical electrodeposition setup, shown in Figure 4, two electrodes are submerged in a plating solution. One of the electrodes is the anode, linked to a block of the source metal, while the other is the cathode, linked to the micromoulds with areas to be deposited with the metal. The two electrodes are powered by an external power supply. The plating solution is a salt of the metal for deposition. For Nickel deposition, nickel sulfamate solution is used, which consists of nickel (85~95g/L), nickel chloride (8~12g/L) and boric acid (25~35g/L).

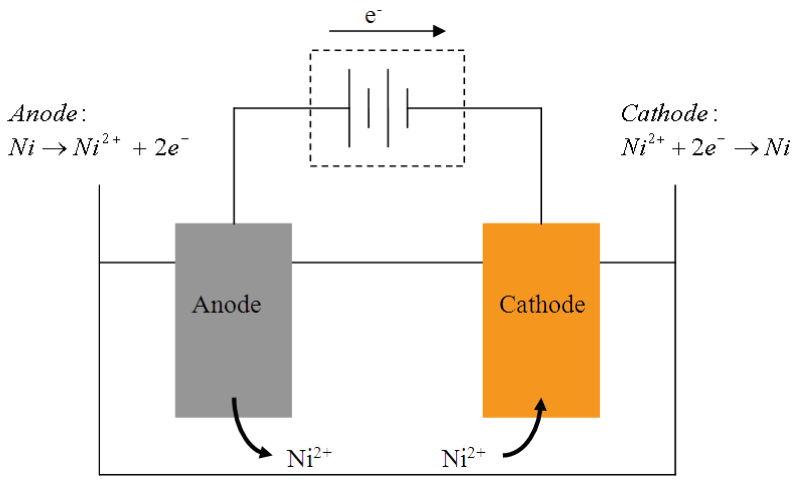


Fig. 4. Schematic representation of an electrodeposition cell

The mechanism of how a cation merges with an electron from the cathode to form a metal atom on the substrate is arguable in two main theories [24]. According to one theory, the cation is discharged and neutralised preferentially in to an active site on the substrate. If

there is no active site within a distance for the cation to discharged, a new nucleus is formed. This theory postulates a lower potential is needed for the cation to grow on an existing crystal than to create a new crystal [24]. The other theory suggests that the cation is neutralised and discharged at random, and form into a metal atom on the surface of the substrate. The metal atoms then orient themselves into the lattice and grain size that is characteristic to the metal. This theory assumes that the discharged metal atoms have the appropriate mobility [24].

In electroforming practice, the usual problems are air bubbles, uneven growing of deposited metal and grain size control. Continuous research has been conducted in this arer for improving the quality of deposition. Apart from improving agitation conditions, including using ultrasonic agitator, pH and temperature control, and small additives, the configurations of the electrodeposition, such as Hull cell, and variations on voltage patterns have been explored, and good results have been achieved for particular applications. Further reading can be found in references [25, 26, 27, 28].

3.2 Nanoparticles and their preparation for electrochemical co-deposition

Composite materials provide a useful way to change the properties of the materials. Electro co-deposition of nanoparticles with metals is one of the recent developments in electroforming field. The micro composite components often demonstrate superior properties. Ceramic nanoparticles are widely adopted as the fillers and they are commercially available. For example, fumed Al_2O_3 nanoparticles of high purity can be purchased from CABOT Cooperation^{TM}.

It is well known that the nanoparticles tend to agglomerate into clusters due to their high surface energy. The phenomena can be observed in SEM analysis. A small amount of Al_2O_3 nanoparticles is first diluted in DI water and dispersed using a mechanical stirrer for 30 mins. An ultrasonic agitation of the Al_2O_3 suspension is followed for 15 mins. A little of diluted Al_2O_3 suspension is dropped on a metal stub and dried up slowly. After gold sputtering, the prepared sample can be examined under the SEM microscope. It can be seen from Figure 5 that the nanoparticles are still agglomerated to some extent and the particle size is estimated to be around 30 nm. The HRTEM micrograph in Figure 6 indicates that the nanoparticle size is in the range from 25nm to 50nm.

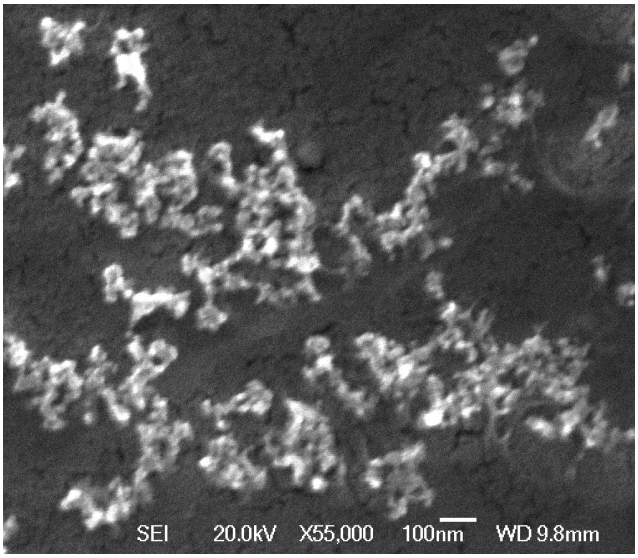


Fig. 5. High magnification SEM micrograph of Al_2O_3 nanoparticles

Nanoparticles should be well dispersed in the electrolyte in order to achieve uniform distribution in the metal matrix. Therefore, it is necessary to avoid agglomeration of nanoparticles in the solution as much as possible. For this reason, the dispersion property of the nanoparticles in the nickel electrolyte was investigated by performing zeta potential measurements. The tests were carried out in a Zeta Potential Analyzer (Zetamaster, Malvern instruments, U.K.). The suspension was diluted to 0.2g/L by DI water and the pH value was adjusted with NaOH and HCl. The obtained results will be present and discussed in the next section.

In the composite electroforming processes, the Al_2O_3 nanoparticles is firstly dispersed in the DI water using a mechanical stirrer for 30mins and then the suspension is ultrasonically agitated for 15mins. After the nickel sulfamate electrolyte is heated to the desired temperate, the Al_2O_3 suspension was slowly mixed into the electroforming bath, which is agitated continuously using a magnetic stirrer. The prepared composite bath is maintained at the constant temperature and agitation for about 30mins prior to electrodeposition. The loading of Al_2O_3 nanoparticles in the bath, operating temperature and the current density have been studied to find the effects of electroforming conditions on the properties of electroformed microcomponents.

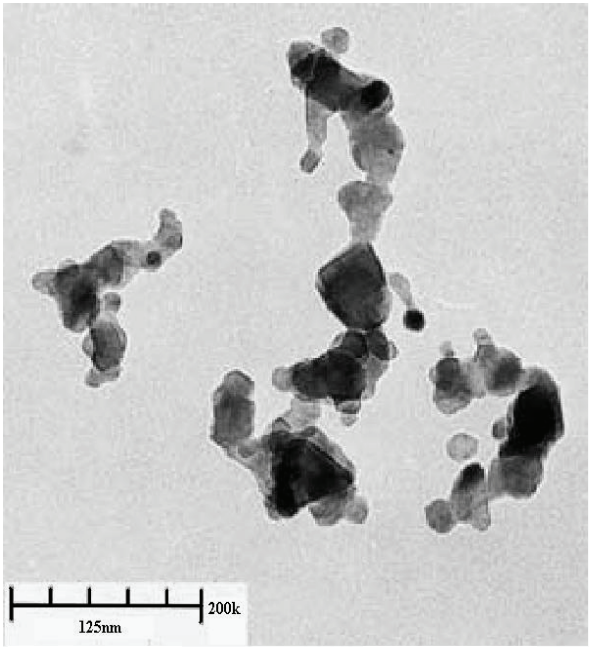


Fig. 6. HRTEM image of Al_2O_3 nanoparticles used in this study (By courtesy of CABOT Cooperation^{TM})

Carbon nanotubes have been co-deposited into Ni matrix, together with Al_2O_3 nanoparticles, as reported by Wei and Jiang [29]. A commercial nickel sulfamate solution (PMD Chemicals Ltd, UK) was used as the host electrolyte for preparing the nickel and nanoparticles colloidal bath. Al_2O_3 nanoparticles (SpectrAl^{TM} 81) of a high purity and raw Multi-Walled Carbon Nanotubes (MWCNTs) were used as the second phase materials for the composite deposition. First, Al_2O_3 nanoparticles and MWCNTs were added into two beakers containing nickel sulfamate electrolyte. However, the nanoparticles and CNTs tend to agglomerate due to the Van Der Walls force. Besides, the hydrophobic nature of CNTs makes them difficult to dissolve in most solutions. Sodium dodecyl sulphate (SDS) as a surfactant was used to improve the dispersion of nanoparticles and CNTs in the solution. Then, the two beakers of colloidal solution were placed in an ultrasonic bath agitation for 30 minutes before mixing the solutions in one beaker. In order to improve the contact between the Al_2O_3 nanoparticles and MWCNTs, the mixture was alternatively agitated using the ultrasonic and mechanical stirring methods for a total of 30 mins before conducting any experiment. Table.1 shows the composition of the prepared electrodeposition bath.

Symbol	Name	Value
$Ni(NH_2SO_3)_2$	Nickel Sulfamate	380~425g/L
$NiCl_2$	Nickel Chloride	8~12g/L
H_3BO_3	Boric Acid	25~35g/L
$NaC_{12}H_{25}SO_4$	Sodium Dodecyl Sulphate	~1g/L
Al_2O_3	Alumina Nanoparticles	3g/L
MWCNTs	Multi-Walled Carbon Nanotubes	0.05~0.3g/L

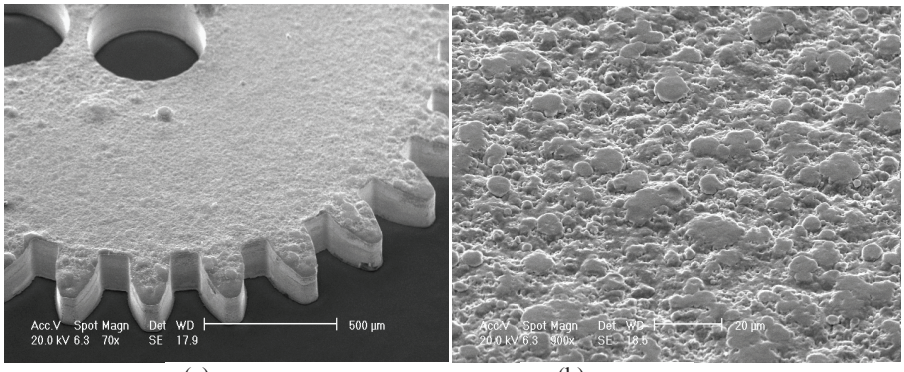
Table 1. Composition of the bath for Ni-Al_2O_3/CNTs electrodeposition.

A study on their dispersion behaviour in the nickel sulfamate bath was conducted. It was found that the addition of SDS makes the dispersion of Al_2O_3 nanoparticles and MWCNTs easier with the aid of sonication. Zeta potential measurements of the diluted solution taking from the prepared nickel composite bath show that the surfaces of Al_2O_3/MWCNTs are positively charged in the pH value of a wide range from 2 to 14. Therefore, Al_2O_3 nanoparticles and MWCNTs will migrate to the cathode in the electrical field.
The experiments were performed in a 1L glass beaker using a potentiostat/galvanostat power supply (TENMA^{TM} 72-6153). A magnetic stirrer was used to agitate the bath to keep the solution with a uniform concentration and reduce the agglomeration of nanoparticles and CNTs during the experiments. The current density was varied in the range of 10~50mA/cm^2. The temperature and pH were maintained at 50±2^{o}C and 4.3±0.1 respectively during the experiments.

3.3 Analysis of microstructures, surface morphology and composition

After electrodeposition, the resultant microcomponents were ultrasonically cleaned in deionized water for 30s to get rid of the loosely attached Al_2O_3 nanoparticles and MWCNTs. The surface morphology was characterized using a scanning electron microscope, and the content of nanoparticles and MWCNT incorporated in the Ni matrix was evaluated from the aluminium and carbon signals using Energy Dispersive X-ray Spectroscopy (EDX) coupled with the SEM. Microhardness tests were conducted in a hardness machine (MVK-H1, Mitutoyo) using a load of 100g force and a dwelling time of 15s.
Fig. a shows an SEM image of a micro gear electrodeposited from a bath containing 0.05g/L MWCNTs and 3g/L Al_2O_3 nanoparticles. Clearly, the gear is uniformly grown through the SU8 micromould and its sidewalls are very smooth, while its top surface morphology is distinct from that of electroformed pure Ni, as shown in Figure 7b. The previous study has shown that the peaks of nickel pyramids are round and even flat in the co-deposition of Ni-Al_2O_3 nanocomposite. However, unlike Al_2O_3 nanoparticles, CNTs are electrically conductive material. Hence, the embedment of MWCNTs on the cathode surface enlarges the electrodeposition area, and accordingly leads to a spherical growth of deposit. Even a low MWCNT loading in the bath still shows its effects as observed in Figure 7b. Increasing the MWCNT loading in the bath enhances the trend of spherical growth of the deposit.
Figure 8 shows a surface of a micro gear electroformed from a bath containing 0.3g/L MWCNTs and 3g/L Al_2O_3 nanoparticles. It can be seen that the surface is very rough and

there are lots of micro gaps. It was also found that the two ends of MWCNTs are embedded in the nickel matrix and other parts of tube are above the nickel surface attached by Al_2O_3 nanoparticles as shown in the micrograph of a high magnification, Fig9. This could be explained as that CNTs have a high electrical conductivity along their axial direction, and therefore, the tube ends of CNTs are easily deposited in the nickel matrix during the deposition process. The high curvature of nanotubes leads to high surface energy of carbon atoms, which makes MWCNTs surface attract Al_2O_3 nanoparticles. The density of applied current has a dramatic effect on the morphology of the deposit, as it is found that a high current density will lead to breakdown of the uniform growth of the deposit, forming a very rough surface. A low current density was therefore used in the experiments to fabricate microcomponents.



(a) (b)

Fig. 7. (a) An SEM image of a microgear deposited in a nickel sulfamate solution containing 0.05g/L MWCNTs and 3g/L Al_2O_3 nanoparticles. (b) A zoomed view of its surface morphology.

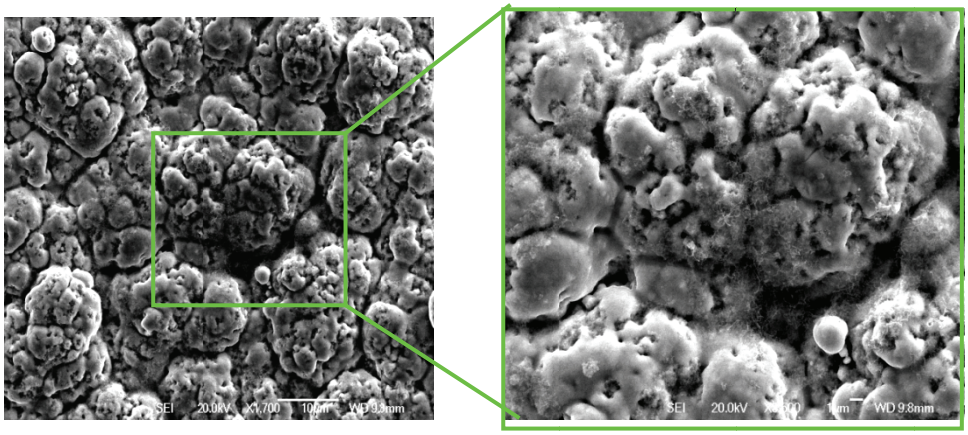


Fig. 8. SEM micrographs of Ni-Al_2O_3/MWCNTs Nanocomposite surface morphology

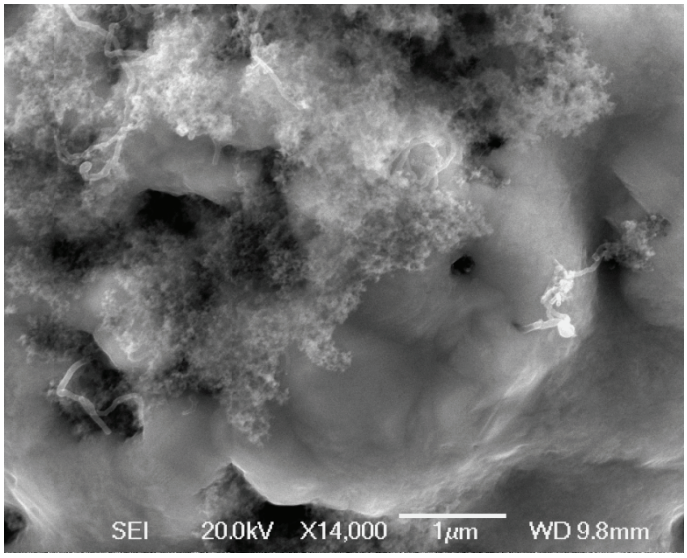


Fig. 9. An SEM Micrograph of Ni-Al_2O_3/MWCNTs Nanocomposite in high magnification

The content of Al_2O_3 nanoparticles in the nanocomposite is affected by the bath composition and electroforming conditions. Figure 10 shows a spectrum of EDX examination on the surface of Ni-Al_2O_3/MWCNTs nanocomposite. It was found that the average volume percentage of Al_2O_3 embedded in the nanocomposite electroformed from a bath containing 3g/L nanoparticles is around 9±2.5%. The average volume percentage of embedded Al2O3 electrodeposited from a bath containing 3g/L Al2O3 nanoparticles and 0.05g/L MWCNTs is around 6.65%. The results of EDX examination and x-ray mapping indicate that the nanoparticles are uniformly distributed in the Ni matrix. As expected, the higher the concentration of Al_2O_3 nanoparticles in the bath, the more nanoparticles are incorporated into the composite electroform with no observed tendency to precipitation.

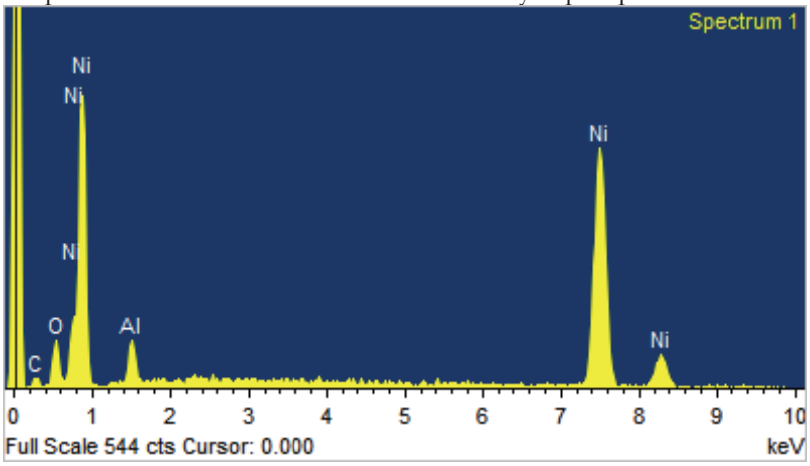


Fig. 10. Energy dispersive spectrum of Ni-Al_2O_3/MWCNTs nanocomposite.

4. Mechanical properties of nanocomposites

4.1 Microhardness

As many factors influence the hardness of electroformed Ni-Al_2O_3 microcomponents, a systematic study was carried out to investigate the effects of the major factors, such as temperature, current density and loading of Al_2O_3 nanoparticles. In this study, the microhardness of the Ni-Al_2O_3 nanocomposite electrodeposited at 45°C and 55°C respectively with the similar pH and current density was compared. It was found that the lower temperature and higher current density result in a relatively higher hardness as shown in Fig. 11. The hardness of deposits usually falls with the rise in temperature of the plating conditions. It can be explained that increasing temperature will lead to high cathodic overpotential, which favours the formation of fine grains and therefore results in high hardness.

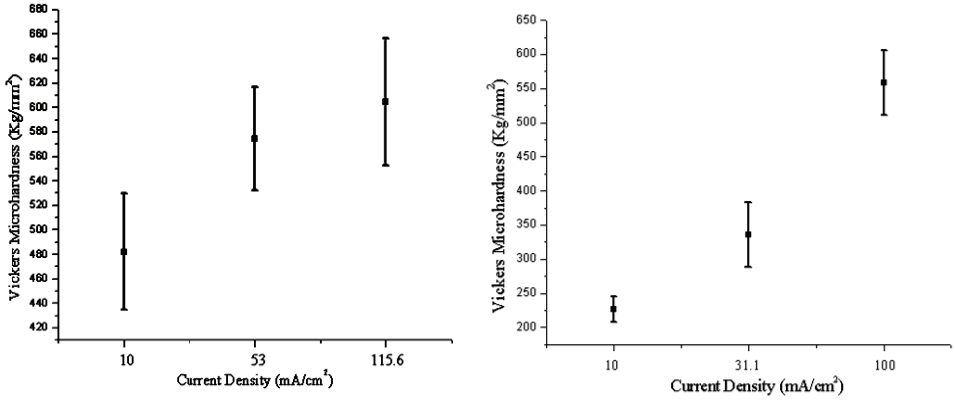


Fig. 11. Microhardness of nanocomposites made from (a) 45°C and (b) 55°C respectively at the similar pH value and current density.

For the high temperature applications, the effect of post heat treatment on mechanical strength of Ni-Al_2O_3 composite was characterized by using a microhardness tester. Ni-Al_2O_3 microcomponents were electroformed using a current density of 25mA/cm^2 from the bath containing 20g/L Al_2O_3 nanoparticles. After stripping off the photoresist moulds, microcomponents were treated in a vacuum heating furnace with a vacuum pressure of 10^{-5}Pa. The applied temperature was gradually ramped up to 600°C at a ramping rate of 5°C/min and then it was cooled down to room temperature slowly after dwelling at 600°C for 1hr. Microhardness tests were performed on the post treated samples and the hardness was found $HV_{0.2}$ 300, which is still higher than that of electroformed pure nickel. It was found that a higher annealing temperature softened the Ni-Al_2O_3 nanocomposite while a relatively lower annealing temperature less than 200°C almost had no obvious effects on the microhardness of the sample. In comparison between the electroformed pure Ni and Ni nanocomposites, it can be concluded that the annealing effect is minimized due to the existence of Al_2O_3 nanoparticles, which hinders the nickel grains from growing bigger during the annealing process [31].

4.2 Elastic property

Since the appearance of LIGA technology, Ni as a MEMS material has been widely used, but its mechanical properties were not paid much attention only until the recent 10 years. Measurements of mechanical properties of electroplated or LIGA-deposited Ni now have been extensively conducted by many researchers using different techniques. The reported Young's moduli range from 148.04GPa ~ 159.90GPa by Kim [32], 175±7GPa by Ballandras *et al* [3], and 190.5GPa by Zhou *et al* [4]. Apart from the difference of methods they used, the Young's modulus of electrodeposited Ni is also strongly influenced by the fabrication conditions, including the temperature and current density. Luo *et al* Fig.[5] and Fritz [6] investigated the effects of temperature and current density in electrodeposition. They both found that Young's modulus decreased with respect to increasing the mean current density. In Figure 12, the Young's moduli of Ni reported in references [35] and [36] are compared with that of Ni–Al_2O_3 nanocomposite obtained from bending and nanoindentation tests in this study.

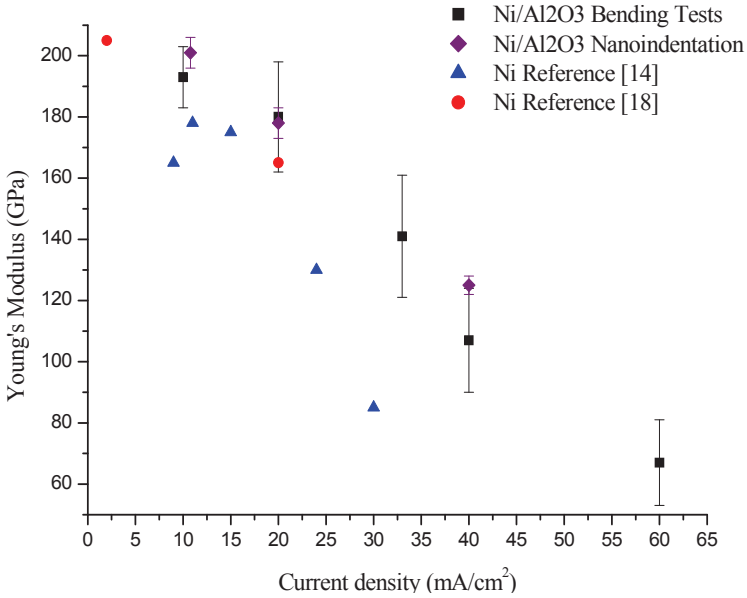


Fig. 12. Comparison of Young's Modulus of electrodeposited pure Ni and Ni–Al_2O_3 nanocomposite at different current density

It is found that the Young's modulus of Ni–Al_2O_3 nanocomposites is dramatically affected by the current density. Nevertheless, Ni–Al_2O_3 nanocomposite still has an increased elastic property than the pure Ni thin films as indicated in the comparison. According to the well-known rule of mixtures [7], such an enhancement is likely due to the addition of Al_2O_3 nanoparticles, which have a higher Young's modulus of roughly 300GPa. In addition, the Al_2O_3 nanoparticles also influence the growth of microstructures of the nickel matrix in the course of electroforming process, which will accordingly affect the intrinsic property, such as the Young's Modulus.

5. Examples of potential applications

Nickel has been used as a dominant metal in LIGA technology for over twenty years, and accordingly research on strengthening nickel properties by incorporating particles into nickel matrix is expected to improve the serving performances of microdevices and microcomponents made of pure nickel.

Teh and co-workers [8] proposed that the incorporation of diamond nanoparticles into nickel matrix formed in electroless electroforming to enhance greatly the overall stiffness of the nickel film attributed to the extreme hardness, stiffness, and temperature resistance of the diamond particle. As more diamond nanoparticles are incorporated into the matrix, the residual stress in the composite film is reduced, making it more suitable for fabricating suspended microstructures. It was found that an electrothermal microactuator made of Ni-nanodiamond composite film could reduce the power requirement by 73% from 0.924 to 0.248W, compared with a pure nickel device for the same 3μm displacement. In addition, the nanocomposite microactuator can exhibit reversible displacement over 3μm, which is larger than pure nickel one for only with 1.8μm displacement [9]. The authors then further investigated the effects on the performance of electrothermal microactuators deposited by the incorporation of diamond nanoparticles in nickel electrolytic bath, in terms of the mechanical strength, coefficient of thermal expansion (CTE) of the nanocomposite material, and the improvements of power consumption and operational reliability in the devices [40]. Similar effects were also found in the Ni-Cordierite nanocomposite thin film by these researchers [41]. The as-deposited nickel-cordierite films exhibit better thermal compatibility with silicon than nickel. The addition of cordierite particles significantly reduces the residual stress as shown in Figure 13. The inset shows a magnified view of the released double folded beam.

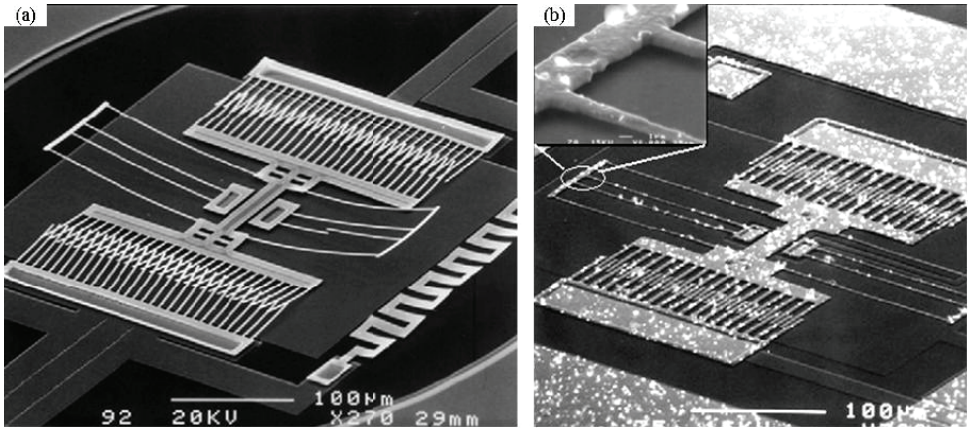


Fig. 13. (a) A partially released, internally stressed nickel microresonator; (b) A fully released, unannealed, residual stress free, nickel-cordierite microresonator

Wang and co-workers [2] fabricated Ni–Al_2O_3 micro pillars using PMMA moulds and the conventional nickel sulfamate bath containing Al_2O_3 particles. In the process, they monitored the changes of current density for pure Ni and Ni–Al_2O_3 deposition. It was found that the

addition of the particles in the plating bath causes the deposition process to change from mass transport controlled to partially kinetic controlled. This is reasonable because Al_2O_3 particles are non-conductive and their distribution in the vicinity of the cathode and their embedment in the deposit affect the reduction process of nickel ions. They also found that the deposition rate decreased as compared to the theoretical prediction. Fig. 14.14 shows an array of Ni–Al2O3 micro pillars (500μm tall and 200μm diameter) electrodeposited from a bath containing 7g/l Al2O3 particles. Results of mechanical tests showed that microhardness was substantially improved with the incorporation of particles. Following the same route, they also studied other MMC microcomponents, for example, Ni–SiC , NiCo–SiC and NiCu–Al_2O_3.

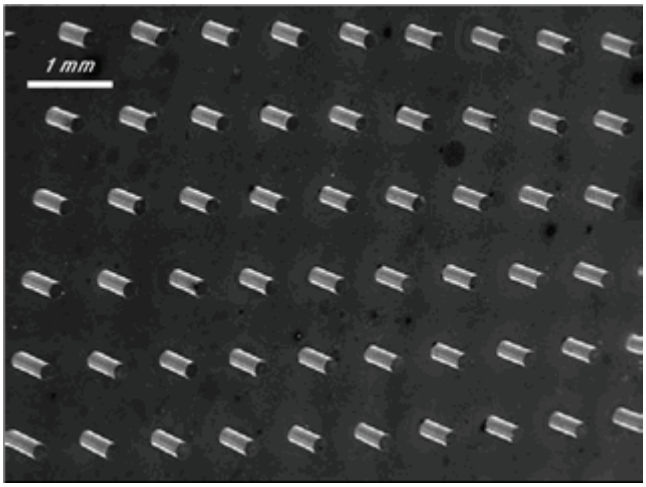


Fig. 14. Ni–Al_2O_3 posts grown from a bath containing 7g/l Al_2O_3 particles

It was found that the average volume percentage of Al2O3 embedded in the nanocomposite electrodeposited from a bath containing 3g/L Al2O3 nanoparticles and 0.05g/L MWCNTs is around 6.65%. The results of EDX examination and x-ray mapping indicate that the nanoparticles are uniformly distributed in the Ni matrix. The microhardness of the Ni-Al2O3-MWCNT composite microcomponents is found in the range of $HV_{0.1}$ 350~400. This value is greater than that of a pure Ni microcomponent produced from a similar bath without nanoparticles ($HV_{0.1}$ 280 ~ 330).

Wei and Jiang [29] fabricated Ni-Al_2O_3 microcomponents for a micro engine. Figure 7a is a composite microgear of Ni, Al_2O_3 and MWCNT for the device. The composition and mechanical properties have been discussed above. Figure 15 shows a Ni-Al_2O_3 composite micropiston for the microengine. It was electrodeposited from a bath containing 7g/l Al_2O_3 particles. The improved hardness of the nanocomposite Ni components will reduce the wear and extend the life, which is important in the design of microengines.

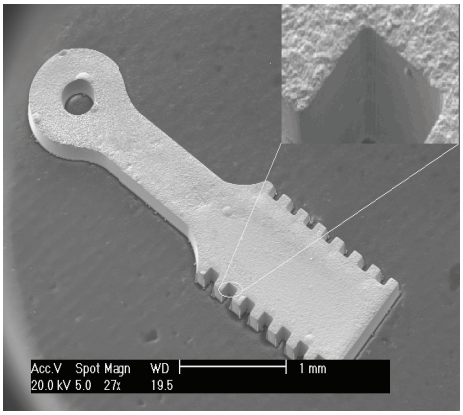


Fig. 15. A micro Ni–Al_2O_3 piston electroformed from a bath containing 7g/l Al_2O_3 particles.

Compared with nickel which has been widely used as a structure material for building up microcomponents and devices, copper is mainly used as a material for electronic interconnections and electrical microdevices. Gan and co-workers [43] studied the feasibility of electrodepositing copper matrix nanocomposite with Al_2O_3 nanoparticles onto silicon wafers. They successfully fabricated a microchannel array of Cu–Al_2O_3 nanocomposite as shown in Figure 16. Microstructural examination and nanoindentation experiments were performed on the nanocomposite. It was found that the presence of the Al_2O_3 nanoparticles in the bath interfere with the grain growth process during the electrodeposition and accordingly lead to smaller grain size compared with pure copper. The mechanical tests showed that the microhardness increased from 1.29GPa to 1.62GPa with the addition of the nanoparticles, which is a 25% increase.

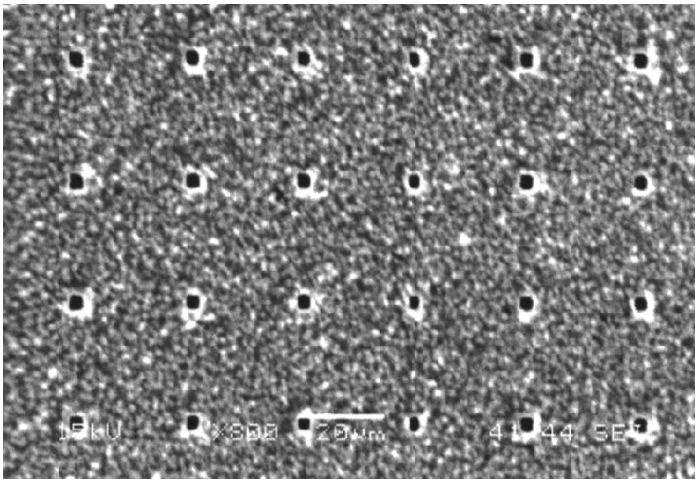


Fig. 16. vertically aligned microchannels in the Cu–Al_2O_3 nanocomposite film

6. Summary and conclusions

Electrodeposition of microcomponents is suitable for mass fabrication of high precision metallic microcomponents. Recent research has combined electrodeposition with nanomaterials and produced nanocomposite components. The research results currently available have demonstrated obvious advantages of the new development. This chapter introduces the latest nanocomposite electrodeposition techniques, characterization results and some applications.

After a general introduction to electrodeposition, the chapter presents the processes of producing high quality and thick micromoulds using three photoresists in SU-8, KMPR and Shipley BPR 100. SU-8 has been used for making moulds of over 1 mm in depth using UV exposure, but it may be difficult to strip. Oxygen plasma is widely used to remove the SU-8 moulds after metal deposition. KMPR was designed with the intention to overcome the stripping difficulty in electrodeposition applications. So far, the reported thickness of the moulds exposed using UV is limited to a few hundred microns. Shipley BPR 100 is another thick photoresist. Its dark blue colour absorbs irradiated light, making an even exposure through a thick layer very difficult. So far the reported thickness of Shipley BPR 100 layers is limited to about 200 μm. However, Shipley BPR 100 moulds can be completely dissolved, which makes it attractive for making relatively thin structures.

Setting and material preparation for nanocomposite co-deposition are introduced. Al_2O_3 nanoparticles and MWCNT have been used as the fillers. The proposed treatment of the nanomaterials against agglomeration in electrodeposition is proven effective. The co-deposited components have been examined. Because of the composite materials, the top of the components is a bit bumpy, but the side walls are smooth. Microstructure analysis shows that carbon nanotubes have been deposited into the structure in separate distribution. EDX analysis shows that the average volume percentage of Al_2O_3 embedded in the nanocomposite electroformed from a bath containing 3g/L nanoparticles is around 9±2.5%. The composite components have been analysed in term of hardness and elastic property. Overall, nickel composite is found harder than pure nickel. In addition, the deposition conditions are related to the hardness. It is found that lower temperature and higher current density result in a relatively higher hardness. Young's modulus of the composite components are generally higher than pure electrodeposited nickel. However, low deposition current resulted in higher Young's modulus.

The chapter provides some practical applications of the co-deposited nickel composite components, ranging from microresonators to microengines. In these applications, the microcomposite components show some excellent properties in thermal expansion, hardness, and wear resistance. These properties vary with the percentage of embedded naoparticles, which makes the composite components tuneable to suit their applications. Currently, the technology for electrochemical co-deposition of Ni-nanoparticles is just emerging, but it has demonstrated its obvious advantages for various applications. This technology has potential to be much developed to suit wide engineering applications.

7. References

[1] Wolfgang Ehrfeld, Electrochemistry and microsystem, Electrochimica Acta, 48 (2003) 2857-2868.

[2] L.T. Romankiw, A path: from electroplating through lithographic masks in electronics to LIGA in MEMS, Electrochimica Acta, 42 (1997) 2985-3005.
[3] E.W. Becker, W. Ehrfeld and et al., Production of separation-nozzle systems for uranium enrichment by a combination of X-ray lithography and galvanoplastics, Naturwissenschaften, 69 (1982) 520.
[4] T. Bieger and U. Wallrabe, Tribological investigations of LIGA microstructures, *Microsyst. Technol.,* 2 (1996) 63–70.
[5] Z. Peng, L. Gang, T. Yangchao and T. Xuehong, The properties of demoulding of Ni and Ni-PTFE moulding inserts, Sens. Actuators A: Phys. 118 (2005) 338–341.
[6] S. M. Allameh, J. Lou, F. Kavishe, T. Buchheit and W. O. Soboyejo, An investigation of fatigue in LIGA Ni MEMS thin films, Materials Science and Engineering A, 371 (2004) 256-266.
[7] A.S.M.A. Haseeb, U. Albers and K. Bade, Friction and wear characteristics of electrodeposited nanocrystalline nickel-tungsten alloy films, Wear, 264 (2008) 106-112.
[8] J. Hormes, J. Gottert, and et al, Materials for LIGA and LIGA based microsystems, Nuclear Instruments and Methods in Physics research B, 199 (2003) 332-341.
[9] R. K. Everett and R. J. Arsenault, Metal Matrix Composites: Mechanisms and Properties, Academic Press Limitted, 1991.
[10] E. W. Becker, W. Ehrfeld and et al, Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming, and plastic molding (LIGA process), Microelectronic Engineering, 4 (1986) 35.
[11] W. Ehrfeld, V. Hessel, H. Lowe, Ch. Schuiz and L. Weber., Materials of LIGA technology, Microsystem Technologies, 5 (1999) 105-112.
[12] L Singleton; A L Bogdanov., S S Peredkov, O Wilhelmi, A Schneoder, C Cremers, S Megtert, A Schmidt, Deep X-ray lithography with the SU-8 resist, Proceedings of Emerging Lithographic Technologies V, vol. 4343, pp. 182-192, 2001.
[13] H. Ito, CG. Willson, in: Proceedings of the SPE Regional Technical Conference on Photopolymers, Society of Plastic Engineers, 1982, p. 331.
[14] Lorenz H, Despont M, Fahrni N, LaBianca N, Renaud P and Vettiger P, SU-8: a low-cost negative resist for MEMS, J.Micromech. Microeng, 7 (1997) 121– 124.
[15] Lorenz H, Laudon M and Renaud P, Mechanical Characterization of a New High-Aspect-Ratio Near UV-Photoresist, *Microelectronic Engineering*, 41 (1998) 371 - 374.
[16] K. Jiang, M. J. Lancaster, I. Llamas-Garro and P. Jin, SU-8 Ka-Band Filter and Microfabrication, *J. Micromech. Microeng.,* 15 (2006) 1522-1526
[17] Lorenz H, Despont M and et al, High-aspect-ratio, Ultrathick, negative-tone near-UV photoresist and its applications for MEMS, *Sensors and Actuators* ,64 (1998) 33-9
[18] Niedermann Ph, Berthou H and et al. A novel thick photoresist for microsystem technology, *Microelectronic Eng*, 67-68 (2003) 259-265
[19]Rao VS, Kripesh V and et al, A thick photoresist process for advanced wafer level packaging applications using JSR THB-151N negative tone UV photoresist *J. Micromech. Microeng*. 16 (2006)1841-6
[20] Lee, CH and Jiang K, Fabrication of thick electroforming micro mould using KMPR negative tone photoresist, *J. Micromech. Microeng*., 18 (2008) 055032.

[21] K. Jiang, M. J. Lancaster, I. Llamas-Garro and P. Jin, SU-8 Ka-Band Filter and Microfabrication, J. Micromechanics and Microengineering, 15, 1522-1526, 2005
[22] Microchem and Nipon Kayaku, KMPR 1000 Guidelines,
[23] Rohm and Haas Electronic Materials, BPR 100 ™ Photoresist, http://www.sts-israel.com/files/pdf/BPR-100_PF04N022R1.pdf, August 2004
[24] Gray, A.G., Modern Electroplating. 1953, New York: London : John Wiley & Sons ; Chapman & Hall. 563p.
[25] R. Yu. Bek, T. P. Aleksandrova, and L. I. Shuraeva, Dependence of the Rate of Electrodeposition of Metals at Mono- and Multielectrodes on Voltage in the Case of an Exaltation Mechanism of Mass Transport, Russian Journal of Electrochemistry, Vol. 41, No. 10, pp. 1027–1031, 2005.
[26] B Marquardt1, L Eude1, M Gowtham1, G Cho1, H J Jeong1,MChâtelet1, C S Cojocaru1, B S Kim2 and D Pribat1, Density control of electrodeposited Ni nanoparticles/nanowires inside porous anodic alumina templates by an exponential anodization voltage decrease, Nanotechnology, 19 (2008) 405607
[27] R.O. Hull, Journal of American Electroplating Society, 27 (1939) 52-60.
[28] T.D. MCCOLM and J.W. EVANS, A modied Hull cell and its application to the electrodeposition of zinc, Journal of Applied Electrochemistry 31: 411-419, 2001.
[29] Xueyong Wei1 and Kyle Jiang, "Synthesis and characterization of nanoparticulate strengthened nickel microcomponents", Advances in Science and Technology Vol. 54, pp 299-304, 2008
[30] X. Wei, Z. G. Zhu, P.D. Prewett and K. Jiang, Microelectronic Engineering, Vol.84, (2007), p.1256-1259.
[31] C. S. Lin, C. Y. Lee, C. F. Chang and C. H. Chang, Annealing behavior of electrodeposited Ni-TiO2 composite coatings, Surface and Technology, 200 (12-13) 3690-3697, 2006.
[32] S. H. Kim, Determination of mechanical properties of electroplated Ni thin film using the resonance method, Mater. Lett., 61 (2007) 3589-3592.
[33] S. Ballandaras, S. Basrour, and et al. Microgrippers fabricated by the LIGA technique, Sensors and Actuators, A, 58(1997): 265.
[34] N. M. Zhou, Y. Zhou and et al., The evaluation of Young's modulus and residual stress of nickel films by microbridge testings, Measurement Science and Technology, 15 (2004): 2389.
[35] T. Fritz, M. Griepentrog and et al, Determination of Young's modulus of electroplated nickel, Electrochimica Acta, 48 (2003) 3029 - 3035
[36] T. W. Clyne and P. J. Withers, An introduction to metal matrix composites, London: Cambridge University Press, 1993
[37] T. W. Clyne and P. J. Withers, An introduction to metal matrix composites, London: Cambridge University Press, 1993
[38] K. S. Teh, Y. T. Cheng and L. W. Lin, Nickel nanocomposite film for MEMS applications, in Proc. 12th International Conference on Solid State Sensors, Actuators and Microsystems, Boston, MA, 2003, pp. 1534–1537.
[39] L. N. Tsai, G. R. Shen, Y. T. Cheng, and W. S. Hsu, The 54th Electronic Components and Technology Conference, June 2004, pp. 472–476.

[40] L. N. Tsai, G. R. Shen, Y. T. Cheng, Performance improvement of an electrothermal micro actuator fabricated using Ni-Diamond nanocomposite, Journal of MEMS, 15 (1) 149-158, 2006.

[41] K.S. Teh, Y. T, Chen and L.W. Lin, MEMS fabrication based on nickel-nanocomposite: film deposition and characterization, J. Micromech. Microeng. 15 (2005) 2205-2215.

[42] T. Wang and K. W Kelly, Particulate strengthened Ni–Al2O3 microcomposite HARMs for harsh-environmental micromechanical applications. J. Micromech. Microeng, 15 (2005) 81–90.

[43] Y. Gan, D.Y. Lee and et al., Structure and properties of electrodeposited Cu-Al2O3 Nanocomposite thin films, Journal of Engineering Materials and Technology, 127 (2005) 451-456.

19

Single Cell Analysis inside Environmental Scanning Electron Microscope (ESEM)-Nanomanipulator System

Mohd Ridzuan Ahmad[1], Masahiro Nakajima[1], Seiji Kojima[2], Michio Homma[2] and Toshio Fukuda[1]
[1]*Department of Micro-Nano Systems Engineering*
[2]*Division of Biological Science*
Nagoya University
JAPAN

1. Introduction

The conventional approach to characterize cellular biology is called biochemistry. This developed science is used for studying physiological aspects, mainly genetics, by characterizing protein and other biomaterials. Since single cells are difficult to study, a collection of cells are used for characterizing cellular physiology and inturn used to describe behavior of single cell (Brehm-Stecher & Johson, 2004). However, in addition to this advance understanding of cellular genetics, information about mechanical properties of cells is also needed. The molecular structure of the cell-wall is only partially understood, and its mechanical properties are an area of "near-total darkness" (Harold, 2005). Moreover, the approximation of single cell behavior from a group used in conventional approach also requires further justification whether it can be applied to all cell types (Shapiro, 2000). The knowledge of the cell mechanics could be valuable in the future for biomedical applications, for example, variations in cell mechanics of healthy and unhealthy cells can be linked to a specific disease.

Available experimental techniques to probe single cells include micropipette aspiration, optical tweezers, magnetic tweezers (Bausch et al., 1999), atomic/molecular force probes (Gueta et al., 2006), nanoindenters, microplate manipulators, optical stretchers (Thoumine et al., 1999) and nanoneedle (Obataya et al., 2005).. The functionality of nanoneedles is not limited only to the stiffness measurements but it can also be used for single cell surgery (Leary et al., 2006) which can be further applied to a novel single cell drug delivery system (Bianco et al., 2005) or as a delivery tool for nanoparticles (Brigger et al., 2002). Conventional drug therapy suffers from the problems of inefficacy or nonspecific effects; hence, nanosystems are being developed for targeted drug delivery (Stylios et al., 2005). In order to successfully deliver materials; e.g. bioactive peptide, proteins, nucleic acids or drugs inside the cell, carriers must be able to penetrate the cell wall or cell membrane without causing death or create any mechanotransduction to the cell (Goodman et al., 2004), i.e. the process

of converting physical forces to biochemical signals and integrating them into cellular responses (Huang et al., 2004). Therefore, the knowledge of biomechanics of the cell is crucial in providing *prior-estimation* of required insertion force to deliver drug material inside the cell. Without having this information, the insertion process may be unsuccessful due to inadequate applied force or the cell may be seriously damaged due to the excessive applied force.

This chapter focuses on the following two needs, i.e. the needs for the understanding of the mechanical properties of single cells and the needs for the novel nanotools for the single cells probing. The first need was fed by highlighting our findings on the effects of three factors, i.e. cell sizes, environmental conditions and growth phases, on the strength of the single *W303* yeast cells. The second need was served by showing our findings on the development of nanoneedles which can be used for single cell local stiffness characterizations and single cell surgery.

2. ESEM-Nanomanipulation System

We have developed an integrated environmental scanning electron microscope (ESEM)-nanomanipulator system. Unlike the conventional SEM, ESEM enables soft, moist, and/or electrically insulating materials to be imaged without prior specimen preparation. Fig. 1(a) shows the image of the ESEM instrument. A low pressure (up to around 1333 Pa) gas can be accommodated around the sample. When this gas is water, hydrated samples can be maintained in their native state. Sample temperature can be controlled (0−40 °C) by using the cooling stage assembled inside the ESEM chamber. By controlling the chamber pressure (10−2600 Pa) and sample temperature, single cell mechanical characterization and analysis can be conducted by using a nanomanipulator and image analysis.

There are two main advantages of the ESEM system compared to the AFM system. The first advantage is that AFM system is difficult to provide a real-time sample observation and the image is mainly constructed after the scanning of the AFM tip on the sample surface is finished. Therefore, it is difficult, if not impossible, to directly observe the response of the object during the sample manipulation. On the other hand, ESEM system provides real-time sample observation which can be obtained before, during and after the manipulation. This real-time observation capability of the ESEM system has many advantages such as in analyzing the dynamic response of the sample and sample selection/sorting for certain properties like cell size, become much easier. The mechanical characterization and analysis capabilities which can be done on the sample can also be enhanced.

The second advantage is related to the manipulation aspect. The AFM system can only provide two-dimensional (2D) manipulation, due to the difficulty of the real-time observation. Although 2D manipulation can measure the stiffness property of the sample, it lacks in the flexibility of manipulation thus limiting the area of sample which can be analyzed. Unlike AFM system, ESEM-nanomanipulation system can provide both 2D and 3D manipulations on the sample, thus increasing the flexibility of the measurement such as the stiffness characterization in different area of the sample like bottom surface can be realized.

The characterization of mechanical properties of the cell requires no interference from other external pressures or forces on the cell surface. These external forces such as the osmotic pressure may prevent investigation of the true mechanical properties of cell experimentally.

ESEM system can provide environmental conditions for the cells in their native state without any effect of the osmotic pressure. Cells can be sustained by releasing low pressure of H_2O molecule gas inside the ESEM chamber to provide high relative humidity (up to 100%). Therefore, the stiffness data obtained from this experimental setting shows better representation of the mechanical properties of the cell.

Nanomanipulation is an effective strategy for the characterization of basic properties of individual nano-scale objects and to construct nano-scale devices quickly and effectively. Previously, we have constructed a hybrid nanorobotic manipulation system integrated with a transmission electron microscope (TEM) - nanorobotic manipulator (TEM manipulator) and a scanning electron microscope (SEM) - nanorobotic manipulator (SEM manipulator) (Nakajima et al., 2006). This system allows effective sample preparation inside SEM with wide working area and many degrees of freedom (DOFs) of manipulation. It has high resolution measurement and evaluation of samples inside a TEM capability. The sample chambers of these electron microscopes are set under the high vacuum (HV) condition to reduce the disturbance of electron beam for observation. However, to observe the water-containing samples, e.g. bio-cells, drying treatment processes are additionally needed. Hence, direct observations of water-containing samples are normally quite difficult in these electron microscopes. This limitation was overcome by using ESEM.

In the present study, we used the nanorobotic manipulators inside ESEM. It has been constructed with 3 units and 7 degrees of freedom (DOFs) in total (Fig. 1(b)). The ESEM-nanomanipulator system can be easily incorporated with various kinds of nanoprobes for single cell analysis as shown in Fig. 1 (c).

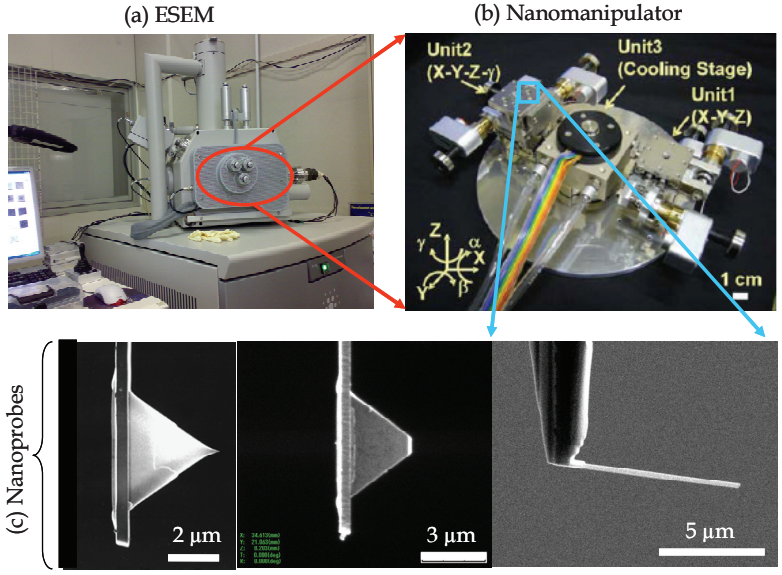


Fig. 1. Overview of the (a-b) ESEM-nanomanipulator system incorporated with (c) various kind of nanoprobes.

3. ESEM Observation of Single Cells

Wild type yeast cells (W303) were used for observation and measurements by ESEM nanomanipulation system. The W303 cells were cultured on YPD plates (1% yeast extract, 2% peptone, 2% glucose, 2% agar) in a 37 °C incubator for 24 hours and dispersed in pure water. Several micro liters of dispersed cells were placed on aluminum block on the cooling stage. The HV mode is operated at 16.7 °C and 3.03 x 10-3 Pa pressure. All yeast cells appeared concave and broken under HV mode. The ESEM mode is operated at 0.0 °C and ~600 Pa pressure and acceleration voltage is set at 15 kV. Under the ESEM mode, decreasing the pressure from ~700 Pa, water gradually evaporates and the samples can be seen underneath. Their images are shown with HV and ESEM modes as shown in Figs. 2(a)-(b). The remaining water can be seen at the intercellular spaces of yeast cells as black contrast (Fig. 2(b)). Almost all yeast cells appear sphere shaped by water-contained condition under the ESEM operation.

In order to reveal the influence of an electron beam observation under HV and ESEM modes, yeast cells were cultured once again after observation. Yeast cells were observed under the ESEM mode, and after the observation, we collected the cells from the observation stage (aluminum block) by a sterilized toothpick and inoculated onto the fresh YPD plate, and grew them overnight. To compare the cell viability with untreated cell, water-dispersed yeast cells were inoculated on the same plate for the control. The cultured plate is shown in Fig. 3. The plate was divided into three regions; cultured after water dispersion, SEM observation and ESEM observation. The numbers of yeast cells colony on the ESEM mode were higher than on the HV mode. From this experiment, the living cell rate on the ESEM mode is almost same order with initial condition of the water dispersion method.

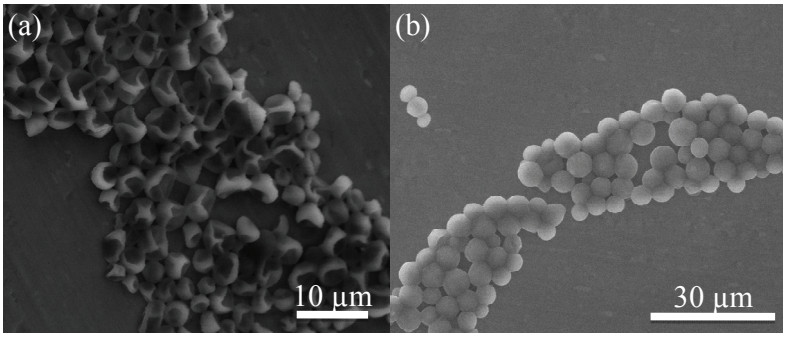


Fig. 2. Images of *W303* yeast cells under (a) HV mode and (b) ESEM mode.

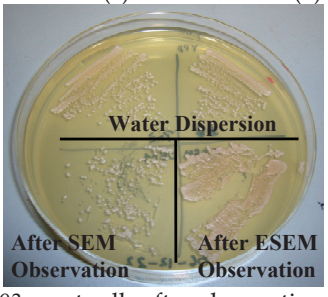


Fig. 3. Cell culture plate of *W303* yeast cells after observations under SEM condition (bottom

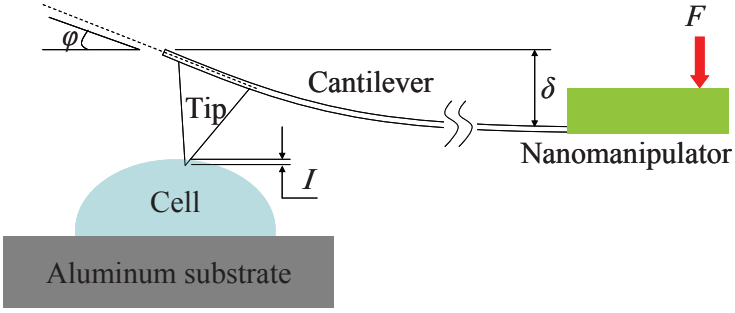
left of the plate area) and ESEM condition (bottom right of the plate area) as compared to the control cell culture by a standard water dispersion method (upper area of the plate).

4. The Effects of the Cells Sizes, Environmental Conditions and Growth Phases on the Strength of Individual *W303* Yeast Cells

4.1 Determination of the Penetration Force and the Cell Stiffness

The indentation experiments (Fig. 4(a)) were conducted by using the integrated ESEM-nanomanipulation system. The cells were fixed on the cooling stage in perpendicular direction with an AFM cantilever (0.02 or 0.7 N/m spring constant). The AFM cantilever is controlled by the nanomanipulator and used to apply the compression force on each cell. The penetration of a single cell by the AFM cantilever in this experiment was based on a direct observation of the cell when it was bursted by the AFM cantilever. Due to the shape and size of the indenters, i.e. pyramidal or tetrahedral tips (sharp or flat), the cell bursted phenomenon can not be avoided as soon as the cell wall is penetrated (Obataya et al., 2005a).

The ESEM can provide real-time observation, i.e. a cell being approached, touched, indented, and finally penetrated/bursted by the AFM cantilever can be seen clearly. Therefore, as far as this paper is concerned, the penetration of a single cell by the AFM cantilever was based on the occurrence of the cell bursting via real-time observation.



(a) Cell indentation experiment

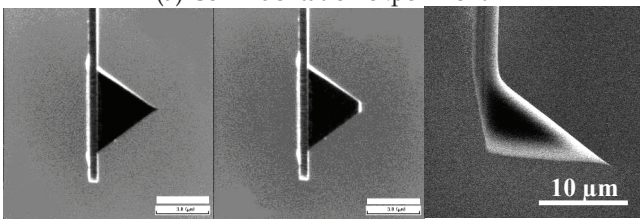


Fig. 4. Schematic diagram of (a) an indentation experiment by using (b) sharp and (c) flat AFM pyramidal cantilever tips under the ESEM mode. For the HV mode, (d) a tetrahedral cantilever tip was used.

Force (F)-cantilever indentation (I) curves were determined from the experiment. A typical F-I curve using a sharp cantilever tip shows a response of a cell before and after the penetration as shown in Fig. 5. The value of F was calculated by using (1), where k, φ and L are the spring constant, the displacement angle in radians and the length of the cantilever, respectively. Values of φ, L and I were determined from direct measurements of ESEM images using an image analysis software (ImageJ, developed at National Institute of Health), while k was obtained from the manufacturer (Olympus Corp.). Equation (1) was derived from the Hooke's law, i.e. $F=k\delta$, where δ is the displacement of the cantilever which was obtained by using $\delta = \varphi(2/3)L$.

$$F = k\varphi(2/3)L \tag{1}$$

The Hertz-Sneddon models which based on the shape of the tips, i.e. conical, spherical, and cylindrical, were used to estimate the Young's modulus of the cells as expressed in (2)-(4). Parameters of E, v, a, R and a are the Young's modulus, the Poisson's ratio (v = 0.5 for soft biological materials (Lanero et al., 2006)) of the elastic half space (cell's surface), the half opening angle of a conical tip, the radius of curvature of a spherical tip and the radius of a cylindrical tip, respectively. Values for a, R and a were obtained from ESEM images. Values for these parameters are summarized in Table 1.

$$F_{cone} = \frac{2}{\pi}\tan\alpha\frac{E}{\left(1-v^2\right)}I^2 \tag{2}$$

$$F_{spherical} = \frac{4}{3}\frac{E}{\left(1-v^2\right)}R^{1/2}I^{3/2} \tag{3}$$

$$F_{cylindrical} = 2\frac{E}{\left(1-v^2\right)}aI \tag{4}$$

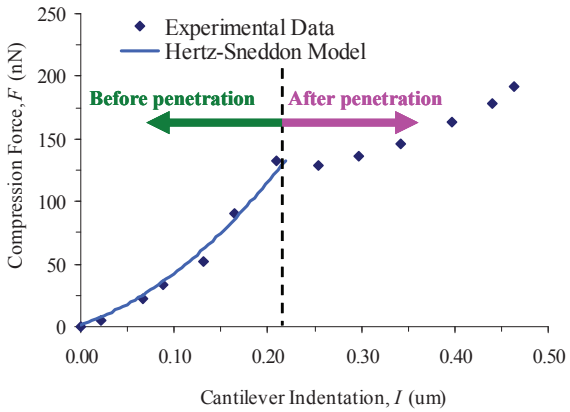


Fig. 5. Typical force – indentation (F-I) curve under ESEM mode using a sharp tip. The curve shows points before penetration (green arrow) and after penetration (pink arrow) of the cell.

The blue line shows the best fit of the Hertz-Sneddon model of the experimental data until the penetration point (dotted line).

Parameters	Values
v	0.5
α	10º
R	20 nm
a	316 nm (ESEM mode) 206 nm (HV mode)

Table 1. Values for parameters used in the estimation of the Young's modulus of the single cells.

The models predict that the load depends on the indentation according to a power law related to the tip geometry (Lanero et al., 2006). In order to choose the correct tip geometry an equation of the form $F=mI^b$ was fitted to force versus indentation curves using commercial fitting software (DataFit), where the exponent b depends on the tip shape.
For the sharp pyramidal tip under ESEM mode, we obtained a value of b close to 3/2, characteristic of the spherical tip. The curves were then fitted by using the spherical model. Interestingly, under HV mode, the value of b close to 2 was obtained. Therefore, the canonical model was applied under HV mode. Finally, for the flat pyramidal tip, the b value close to 1 was obtained which is for the cylindrical model. The fitting model for the sharp tip using the spherical tip was reported by Lanero et al. (Lanero et al., 2006) while Touhami et al. (Touhami et al., 2003) used the conical tip as their fitting equation. Obataya et al. (Obataya et al., 2005a) used the cylindrical equation to model their flat nanoneedles.
The Hertz-Sneddon models were derived from the classic Hertz's mechanics model for linear elastic material (Sneddon, 1965). For flat cylindrical punch relating to an applied force, F, with an indentation depth, I, yields the following expression for the initial unloading (load is released after indentation) response, $dF/dI^(1/2)(1-v^2)/E$, where c is coefficient (c=2/(π)1/2) and A is the contact area. The equation was based on the flat-ended punch, but holds true for any punch that can be described by a smooth solid revolution (spherical, conical, elliptical, etc.) (Pharr et al., 2002). Berkovich and Vickers indenters (3 and 4 sided pyramidal cones, respectively) commonly applied in instrumented indentation techniques cannot be described as bodies of revolution. However, it has been found experimentally and by means of finite element simulations that the deviation from the above equation of pyramidal and other geometrical shapes is negligible (Pharr et al., 2002). The constant c=1.142 for the pyramidal indenter and c=1.167 for tetrahedral indenter differ little from $c = 2/(\pi)^(1/2) = 1.1284$ from the above equation. In other words, the derived equation for the sharp conical tip can be used without large error, even the sharp indenter is not a true body of revolution.
The final equations of the Young's modulus based on the Hertz-Sneddon models, i.e. (2)-(3), and the parameter's values (Table 1) for three different tip shapes, i.e. conical (HV mode), spherical (ESEM mode) and cylindrical (ESEM mode and HV mode), are expressed in (5)-(7).

$$E_{cell} = (6.681)\frac{F_{cone}}{I^2} \tag{5}$$

$$E_{cell} = \left(3.977 \times 10^{3}\right) \frac{F_{spherical}}{I^{3/2}} \tag{6}$$

$$\begin{cases} E_{cell} = (1.187 \times 10^{6}) \dfrac{F_{cylindrical}}{I} & \text{(ESEM mode)} \\ E_{cell} = (1.825 \times 10^{6}) \dfrac{F_{cylindrical}}{I} & \text{(HV mode)} \end{cases} \tag{7}$$

Fig. 6 shows typical *F-I* curves using the spherical and cylindrical models under ESEM mode while the inset shows a curve using the conical model under HV mode as obtained in our experiments. The curves are shown until the bursting point of the cells which is the maximum force which the cell can sustain before the cell is bursted. After the bursting point, the force will drop for a while due to the reversed force and will increase again when the tip traverses inside the cell or touch the bottom of the cell wall (Fig. 5). Starting from Fig. 6, data after the bursting points are not shown in the following figures. The reason for this exclusion is that the data were no longer representing the cell stiffness property since the cells were seriously damaged.

The curve leading up to the failure is expected to give a combined elastic property of indenting the whole yeast cell and the localized deformation of the surface ruled by the properties of the cell-wall surface. Therefore, the reported *E* values in this paper indicate the maximum strength of the cell before bursting from a local indentation point. The *E* values must be higher than the local elastic cell property as the strength of the cell is dominated by the whole cell stiffness (Smith et al., 2000). Nevertheless, examining the cell strength based on the Young's modulus parameter per se is not adequate since its value may remains constant at different situations, e.g. cell growth phases (Smith et al., 2000). Therefore, the penetration force was used to determine the cell strength under several factors, i.e. cell sizes, environmental conditions and cellular growth phases.

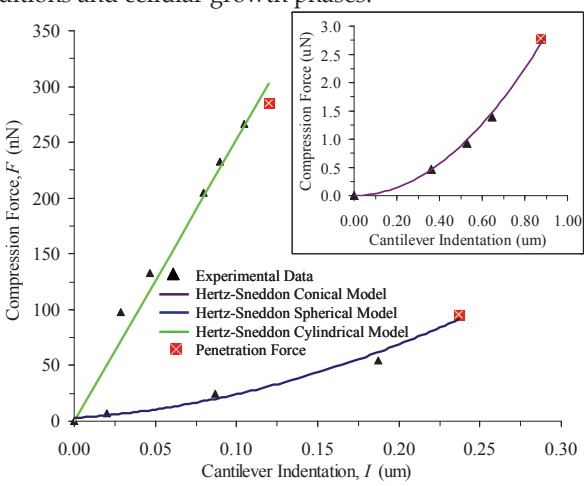


Fig. 6. Typical force – indentation (*F-I*) curves under ESEM mode using spherical and cylindrical models for sharp and flat tips, respectively. Inset shows the *F-I* curve under HV mode using conical model.

4.2 The Effect of the Cell Sizes on the Strength of the Single Cells

The penetrations of single yeast cells using sharp and flat cantilever tips (Figs. 4(b)-(c)) are shown in Figs. 7(a)-(b). The required force to penetrate a single yeast cell under the ESEM mode using a sharp tip for different cell sizes are shown in Fig. 8. The comparison between sharp and flat tips to penetrate each cell is shown in Fig. 9. The environmental settings for ESEM mode are set to 600 Pa and 0.0°C. These parameters are chosen to vaporize water so that the majority of the cell surface can be seen for mechanical manipulation.

Results clearly show a strong relationship between cell size and strength, the latter increases with an increase in cell size (Fig. 8). Yeast cell walls consists predominantly of glucan with (1,3)-β and (1,6)-β linkages, and mannan covalently linked to protein (mannoprotein) (Klis et al., 2006; Lesage & Bussey, 2006). Mechanical properties are a function of β-glucans, whereas mannoproteins control porosity (Lesage & Bussey, 2006). Thus it can be inferred that the β-glucans composition increases when the yeast cell size increases. This prediction is consistent with other studies stating that small and large yeast cells exhibit similar glucan and mannan contents, but further fractionation of the glucan showed five-times less alkali-insoluble glucan in the smaller sized cells (Srinorakutara, 1998). This may be another reason why the large yeast cells appear to be stronger than the smaller cells (Srinorakutara, 1998). The alkali-insoluble glucan plays a role in maintaining cell rigidity (Fleet & Manners, 1976).

Nevertheless, the elastic properties of the cells at different sizes remain unchanged (Table 2). The small and large yeast cells contain different composition levels of alkali-insoluble glucan (Srinorakutara, 1998). The increase in this glucan in the large cells does not change the properties of the molecule itself but rather increase the number of molecules, without changing its elasticity properties. Consequently, the elasticity of the whole cell also remains unchanged and independent of cell size. On the other hand, since there are more glucan molecules in the larger cells, the thickness of the cell wall also increases (Smith et al., 2000). Hence, the amount of force needed to penetrate the cell also increases due to more molecule-barriers during the indentation which leads to penetration.

The average penetration forces for 3, 4, 5 and 6 μm cell diameter ranges are 96 ± 2, 124 ± 10, 163 ± 1 and 234 ± 14 nN, respectively (Fig. 8). This was expected since the yeast's cell wall is remarkably thick (100 - 200 nm) (Walker, 1998), whereas its plasma membrane is only about 7 nm thick (Walker, 1998).

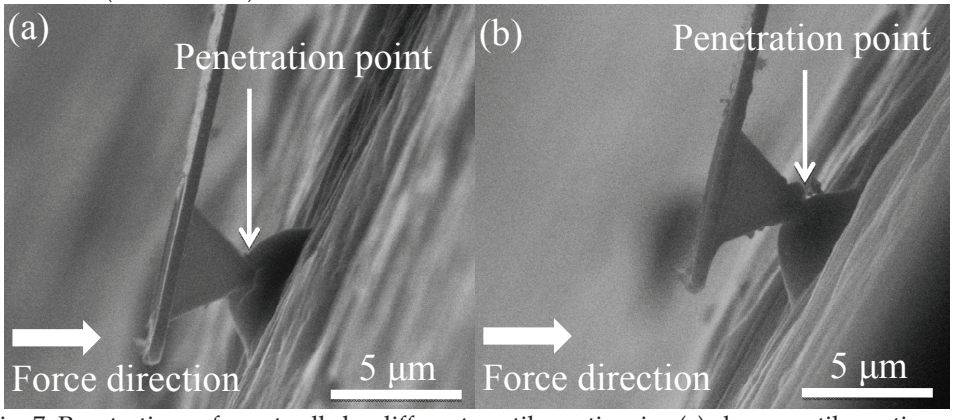


Fig. 7. Penetrations of yeast cells by different cantilever tips, i.e. (a) sharp cantilever tip and (b) flat cantilever tip.

These results are comparable to reported penetration force of 1 nN for a human epidermal melanocytes that is the only cell membrane without cell wall structure (Obataya, 2005). Since the mechanical strength of the cell is provided by the cell wall and not the cell membrane, therefore the melanocytes are easily penetrated compared to yeast cells. Additionally, the effect of local (small contact area between tip and cell surface and global (large contact area between tip and cell surface) penetrations are also an important factor in the amount of applied force. This can be seen from our results for the local penetration (96 to 234 nN) compared to reported average of 149 ± 56 μN for global bursting of the yeast cell (Smith et al., 2000). The average Young's modulus obtained from (6), i.e. 3.24 ± 0.09 MPa is reasonable compared to reported local cell stiffness values of 0.73 MPa (Pelling et al., 2004) and 1.12 MPa (Lanero, 2006), since the *E* values obtained in this paper representing not only the local elastic property of the cell but also the whole cell stiffness.

In addition to the cell size factor, we showed that the shape of the indenter also plays an important role in predicting the penetration force of a single cell. This can be confirmed from Fig. 9 where the penetration force can be reduced 40-50% by using a sharp cantilever tip compared to a flat tip. This provides a good indicator for future cell surgery where a minimal penetration force is required for preventing serious cell damage or undesirable mechanotransductions due to a large cell deformation (Huang et al., 2004). The Young's modulus (E) of the cells estimated using the spherical model (sharp tip) and the cylindrical model (flat tip) obtained from (6) and (7), respectively, showed that the E values of the cell remained constant at about 3 MPa.

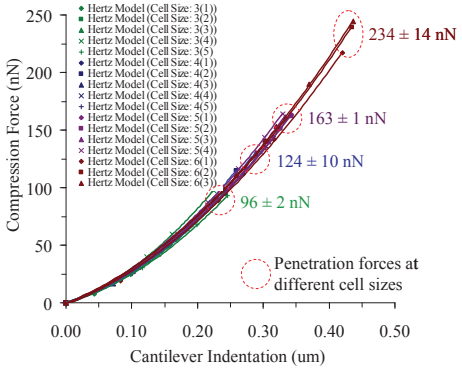


Fig. 8. Penetration force using a sharp tip at different cell sizes. Experimental data are fitted using the spherical model. The curves are shown until the penetration points. Inner bracket () indicates the cell number for each cell size. Cell size is in μm.

Cell Size (μm)	Young Modulus (MPa)
~ 3	3.27 ± 0.23
~ 4	3.28 ± 0.12
~ 5	3.31 ± 0.11
~ 6	3.29 ± 0.10

Table 2. Young's modulus of the single cells at different cell sizes.

The link between the cell strength with the increase cell size as reported in this paper is somehow against the Laplace law. Laplace law governs the tension of an ideal thin-wall sphere, i.e. *Wall tension=Transmural pressure*Radius of the curvature*½* (Morris & Homann, 2000). From this equation, the more curved a membrane (small cell), the less tension experienced by the cell wall for a given transmural pressure (the pressure difference between inside and outside of the cell). The less tension the cell wall, the less possibility the cell wall to failure. This mean a smaller cell is less likely to be ruptured than a larger cell.
This notion could be true if the thickness of the cell remains constant independent of cell sizes. Since the cell wall plays a main role in cell stiffness than the membrane cell, any different in the cell wall thickness would give different stiffness property of the cell. Smith et al. (Smith et al., 2000) showed that the thickness of the cell wall increases as the cell growth. Kliss et al. (Klis et al., 2006) also revealed that cells increase their size during their lifetime before undergone a cell division. During this growing process, the composition of the β-glucan is increasing which inturn increases the thickness of the cell wall (Smith et al., 2000). The heterogeneity of the cell wall composition and cell wall thickness will give different stiffness property to the cell and therefore the Laplace law which governs the surface tension of the cell may play a minor role in predicting the strength of the cell.

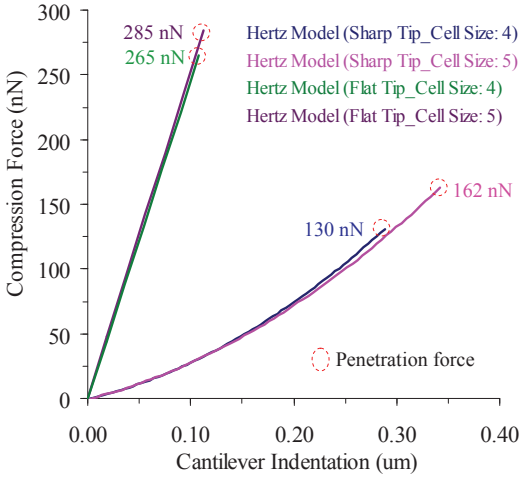


Fig. 9. Comparison of penetration force using sharp (using the spherical model) and flat tips (using the cylindrical model) for 4 and 5 μm cell sizes range. Young's modulus of the cells remains constant at about 3 MPa. The curves are shown until the penetration points.

4.3 The Effect of the Environmental Conditions on the Strength of the Single Cells

The required force to penetrate a single yeast cell under HV mode using sharp and flat tips for different cell sizes are shown in Figs. 10 and 11, respectively. Insets are images of penetration of a single yeast cell using sharp (Fig. 10) and flat (Fig. 11) cantilever tips. Higher spring constant of the cantilever, i.e. 0.7 N/m was used for the compression experiment as lower spring constant values failed to penetrate the cell. Fig. 4(d) shows the image of a tetrahedral cantilever tip used in this experiment. The penetration force to burst a

cell wall increases about twenty times from ESEM mode to high vacuum (HV) condition (cells which 5 µm diameter were used for comparison and analysis) (Figs. 8 and 10). Since this data under high vacuum condition is reported for the first time, it is difficult to compare with other reports. Nevertheless, the increase in penetration force from ESEM to HV modes which was obtained in our experiment can be logically understood.

Cell wall construction is tightly controlled since polysaccharide composition, structure and thickness of the cell wall vary considerably, depending on the environmental conditions (Klis et al., 2006). Therefore, we can infer that under extreme condition of high vacuum, yeast cell tries to survive by increasing its wall thickness as suggested by the increased compression force needed to penetrate its cell wall (Figs. 10 and 11). Other possibility is that under high vacuum condition, cells become too dry which eventually increases their cell hardness. These findings reveal that besides the shape of a cantilever tip, other external factors such as environmental conditions also influence the penetration force of a single cell. Like the ESEM mode, cell size remains the internal factor that influences the penetration force under high vacuum condition. By using a sharp tip, an average of 3 µN penetration force was required for a 5 µm diameter cell. As expected, this force increased (up to 14 µN) when a flat cantilever tip was applied, for cells with an average diameter of 4 to 6 µm.

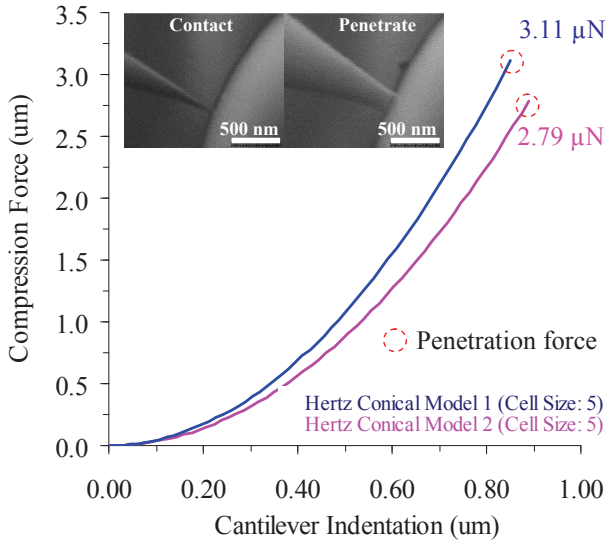


Fig. 10. Force-indentation curves until the penetration points using a sharp tip (conical model) for about 5 µm cell size under HV mode.

For a quantitative evaluation of the elasticity of the cell, the Hertz-Sneddon mechanics model for conical and cylindrical tips as expressed in (2) and (4) are used for sharp and flat tips, respectively. Average elastic properties of yeast cells under the influence of two environmental conditions are shown in Table 3, where the cell stiffness increases about ten times under HV mode. The cell surface becomes rigid when it is placed in the HV mode. It is believed that in the absence of water, cell surface becomes harder than in the presence of

water. This leads to a significant increase in the Young's modulus (26.02 ± 3.66 MPa) obtained from (5) under HV mode as compared to the ESEM mode (3.31 ± 011 MPa) as expected.

4.4 The Effect of the Growth Phases on the Strength of the Single Cells

All cells have a unique growth phase curve during their life cycle which is normally divided into four phases; lag, log, saturation and death phases. In the present study we did not investigate the death phase. These phases can be easily identified from the optical density (OD) absorbance-time curve (Fig. 12). Lag phase located at the initial lower horizontal line where the cells adapt themselves to growth conditions and the cells mature and divide slowly. Log phase, can be seen from an exponential line where the cells have started to divide and grow exponentially. The actual growth rate depends upon the growth conditions. Saturation phase resembles a final upper horizontal line where the growth rate slows and ceases mainly due to lack of nutrients in the medium.

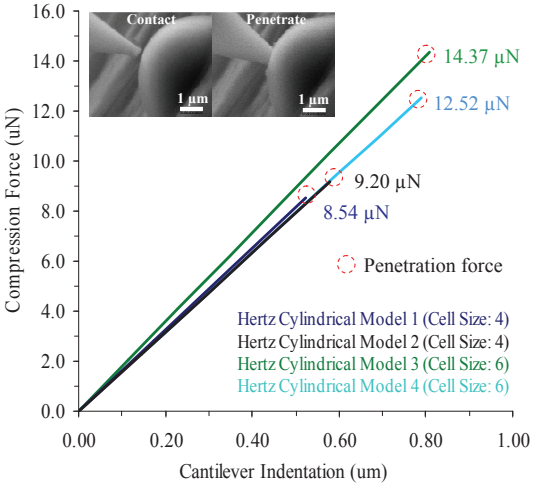


Fig. 11. Force-indentation curves until the penetration points using a flat tip (cylindrical model) for 4 to 6 µm cell sizes under HV mode.

Environmental Conditions (Pa)	Young Modulus (MPa)
ESEM mode (~ 600)	3.31 ± 0.11 (n=4)
HV mode (~ 0.00314)	26.02 ± 3.66 (n=6)

Table 3. Young's modulus of the single cells at different environmental conditions.

During the death phase, all of nutrients are exhausted and the cells die (Tortora et al., 2003). It is believed that the mechanics of the normal cells have different properties at each of the phases. In this section, the mechanical properties of *W303* yeast cells at four different growth phases (early, mid, late log and saturation) are discussed.

Since each cell has a unique growth phase, we determined the growth phase of *W303* yeast cells by measuring the absorbance at 660 nm using a Spectrophotometer (Hitachi: U-2000) at one hour intervals (Fig. 12). The growth rate (absorbance per hour) was about 0.23 as can be seen from the slope of the curve. Samples were stored at -80°C (Sanyo: MDF-291AT) to hold its growth before the compression experiment.

The compression experiments were done at 6 hour intervals starting at early log phase and ending at saturation phase. Penetration force was analyzed for each phase as shown in Figs. 13(a)-(d). As expected, the force needed to penetrate a single cell increased from early log to saturation phase (161 ± 25, 216 ± 15, 255 ± 21 and 408 ± 41 nN), whereas, the elastic properties of the cells seem constant for all the phases obtained from (6), i.e. 3.28 ± 0.17, 3.34 ± 0.14, 3.24 ± 0.11 and 3.38 ± 0.11 MPa (Table 4). These mechanical properties of the *W303* yeast cells at different growth phases are in an agreement with reported increment of average surface modulus of *S .cerevisiae* cell wall to be 11.1 ± 0.6 N/m (log phase) and 12.9 ± 0.7 N/m (saturation phase) with no significant increase in the elastic modulus of the cell, i.e. 112 ± 6 MPa (log phase) and 107 ± 6 MPa (saturation phase) (Smith et al., 2000). Their values for elastic modulus are quite high is also reasonable since they measured the whole elastic properties of the cell by compressing a single cell between two big flat indenters as compared to a local cell indentation in our case.

From this analysis, it can be inferred that the thickness of the *W303* cell wall is increased during the growth phase since the penetration force increases at each phase, and at the same time, maintaining their elastic properties during the growth phase as verified by the constant elastic modulus values shown in Table 4. The increase in the cell thickness does not affect the elastic properties of the cell but the penetration force will be increased due to more molecules need to be penetrated. We believe that the increase in the thickness of cell wall is mainly due to an increment of β-glucans molecules (Lesage & Bussey, 2006).

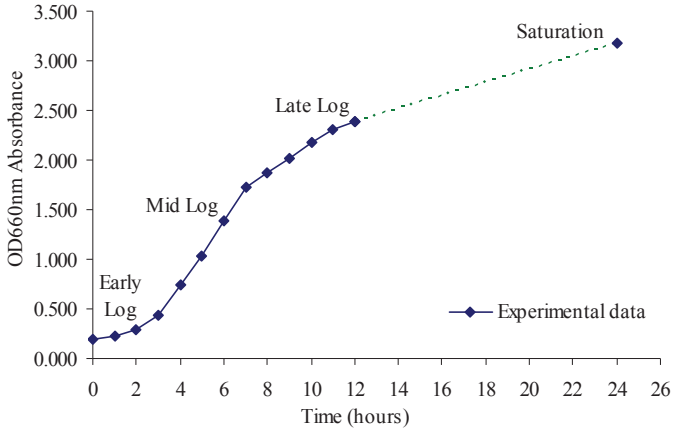


Fig. 12. The growth phase cycle for *W303* yeast cells based on the OD_{660nm} absorbance measurement.

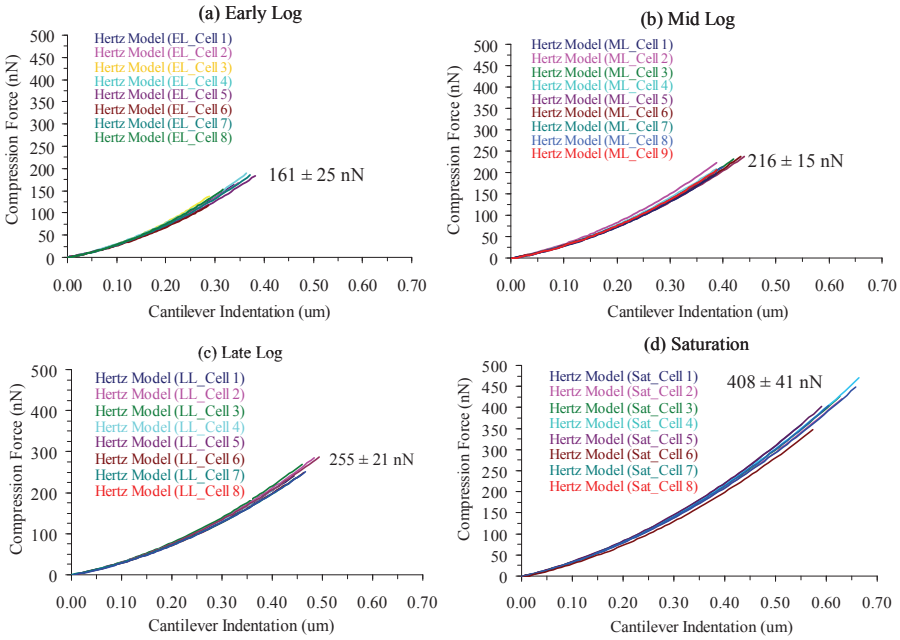


Fig. 13. The force-indentation curves (fitted using spherical model) for single cells at each growth phases (a) early log, (b) mid log, (c) late log and (d) saturation. The curves are shown until the penetration points.

Cell Growth Phase	Young Modulus (MPa)
Early Log	3.28 ± 0.17 (n=8)
Mid Log	3.34 ± 0.14 (n=9)
Late Log	3.24 ± 0.11 (n=8)
Saturation	3.38 ± 0.11 (n=8)

Table 4. Young's modulus of the single cells at different cell growth phases.

5. Nanoneedles for Single Cells Mechanical Characterizations

5.1 Determination of Cell Stiffness and Cell Surgery Based on Nanomanipulation

We propose several approaches for achieving two biomechanic tasks, i.e. determining the stiffness of a single cell and performing single cell surgery as shown in Fig. 14.

In the cell stiffness measurement, two approaches are designed. The first approach, to the best of our knowledge, is a novel technique in determining the stiffness of the cell. In this technique, the nanoneedle and the cell can be modeled as two springs in series. In order to model the nanoneedle as a spring, firstly, the nanoneedle should be able to buckle linearly and secondly, it should have lower or the same spring constant as the cell. This technique prevents the damage to the cell since the indentation is minimized from the buckling effect

of the nanoneedle which avoids excessive indentation being applied to the cell, the so-called "soft nanoneedle". The schematic of this approach is shown in Fig. 14(a).
The second approach for the measurement of cell stiffness is based on a "hard nanoneedle" which relies totally on the deformation of the cell in order to measure its stiffness. By knowing the applied compression force and the deformation of the cell, the stiffness of the cell can be measured (Fig. 14b). In order to fabricate the hard nanoneedle, strong material which is hard to bend is needed. We used Tungsten material to construct the hard nanoneedle. To prevent excessive indentation force on the cell by the hard nanoneedle, this process has to be performed by avoiding any sudden pressure which can interrupt chemical activities of the cell, e.g. mechanotransduction effect. The other preventive step is to use a lower cantilever spring constant. We used 0.09 N/m cantilever spring constants.
The hard nanoneedle can also be used for single cell surgery. In theory, all hard nanoneedles must be able to penetrate the cell, however, in practice, for the yeast cells, only some nanoneedles (Si-Ti and W_2 nanoneedles) can penetrate the cells while others (e.g. $W_{0.09}$ nanoneedle) are only limited to cell stiffness measurement only. This is because to penetrate the cell, more compression force is needed. This excessive force may exceed the force limits of the nanoneedle causing failure. This so-called "cell surgery" operation is shown schematically in Fig. 14(c).

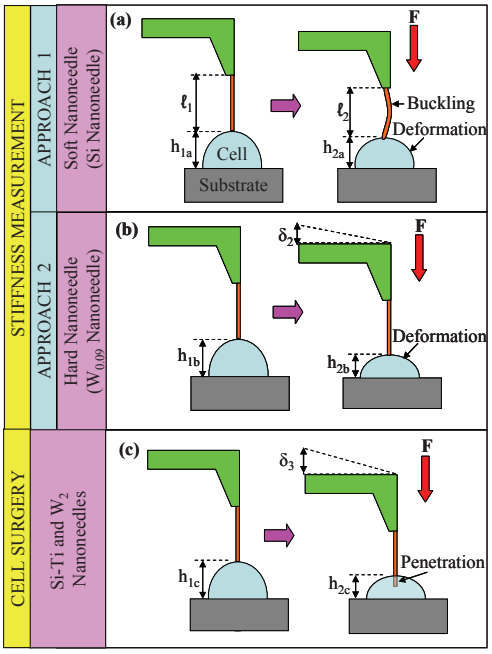


Fig. 14. The schematics diagrams for the nanoneedles indentation experiments indicate local single cell stiffness measurement using (a) soft Silicon (Si) nanoneedle and (b) hard nanoneedle which is based on Tungsten deposition on the cantilever (spring constant = 0.09 N/m) ($W_{0.09}$). Single cell surgery is shown in (c) by using hard Si-Ti and W_2 nanoneedles.

5.2 Fabrication of Soft and Hard Nanoneedles

Four kinds of nanoneedles were fabricated. Fig. 15 shows the general fabrication processes for each of the nanoneedles. The standard Platinum coated tetrahedral cantilever tips having a spring constant of 2 N/m (Olympus Corp.) were used in the fabrication of Si, Si-Ti and W_2 nanoneedles. Whereas for the $W_{0.09}$ nanoneedle, the standard sharp pyramidal cantilever tip which has a 0.09 N/m spring constant (Olympus Corp.) was used.

The soft Silicon (Si) nanoneedle was fabricated by using only Focused Ion Beam (FIB) etching as shown in Fig. 15(a). The parameters for the etching process are 30 kV of acceleration voltage and 8.561×10^{18} of ion dose. The time required to complete the process is between 20 to 45 min. The diameter and the length of the Si nanoneedle are about 150-170 nm and 7 μm, respectively.

The quantitative analysis of the detailed material of the soft nanoneedle was performed using an energy dispersion spectrometry (EDS) method (Ametex Inc.). In general, this type of spectroscope identifies and counts the impinging X-rays based upon their characteristic energy levels. The X-ray spectrum, used in quantitative analysis, is generated by the nanoneedle in response to the spot from an electron beam on the center area of the nanoneedle. The intensity of the X-rays is measured by EDS as shown in Fig. 16. From the EDS analysis, the main element of the soft nanoneedle is Silicon (83.03 %) while the rest is Oxygen.

The first type of hard nanoneedle, i.e. Silicon-Titanium (Si-Ti) nanoneedle, was fabricated by coating the former Si nanoneedle with Ti material by sputtering as shown in Fig. 15(b). The time for the sputtering operation is about 3 min to obtain a coating of approximately 20 nm of Ti material. The coated area of the Ti material is only on the one side of the nanoneedle facing the Ti source during the sputtering operation. Nevertheless, this coating increases the strength of the basic Si nanoneedle from easily buckling to unbuckling. The typical diameter and length of the Si-Ti nanoneedles are about 185-200 nm and 4 μm respectively.

The second type of hard nanoneedle ($W_{0.09}$) was fabricated by first flattening the apex of the sharp pyramidal cantilever tip by using FIB etching. Then W material was deposited on the flat surface by using a small area of FIB deposition, i.e. 100 nm^2 as shown in Fig. 15(c). The ion dose parameter for deposition is 2.672×10^{20}. The time for the deposition is about 13 min. The typical diameter and the length of the $W_{0.09}$ nanoneedles are about 250 nm and 3 μm respectively.

Finally, the third type of hard nanoneedle (W_2), was fabricated by first flattening the apex of the sharp tetrahedral cantilever tip by using FIB etching, then followed by W deposition of large area of FIB deposition, i.e. 160 000 nm^2, and finally trimmed to produce nanoneedle structure by using FIB etching as shown in Fig. 15(d). The diameter and length of the W_2 nanoneedle are around 150-170 nm and 3 μm. The actual images of each nanoneedle are shown in the Fig 17. All of the fabricated nanoneedles have flat ended tips.

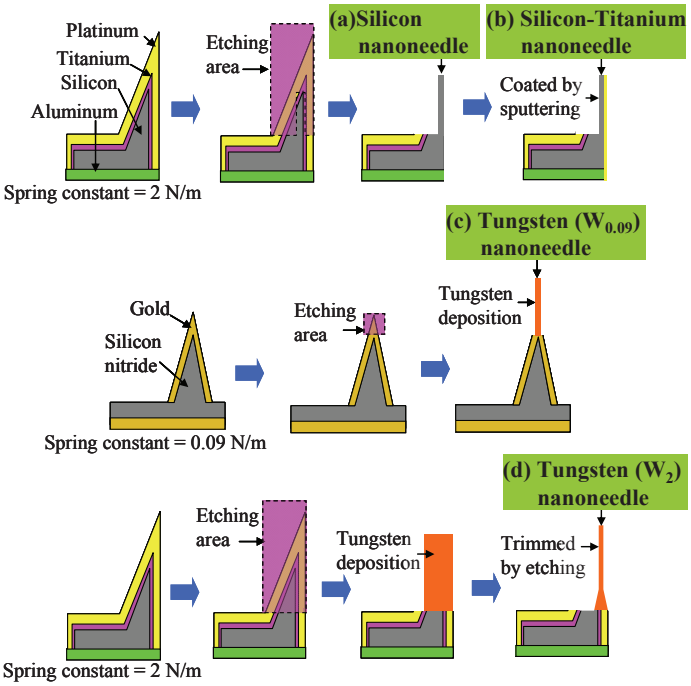


Fig. 15. Fabrication procedures of nanoneedle: (a) Silicon nanoneedle, (b) Silicon-Titanium nanoneedle, (c) Tungsten ($W_{0.09}$) nanoneedle and (d) Tungsten (W_2) nanoneedle, respectively.

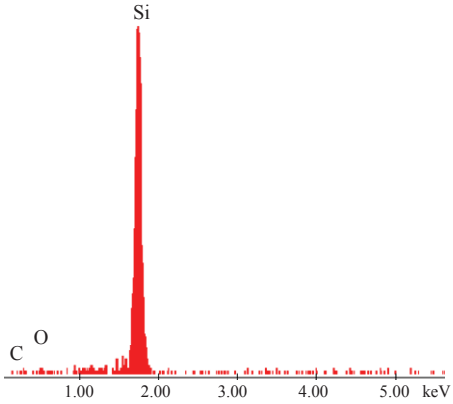


Fig. 16. The determination of the precise material of the soft nanoneedle was done by using energy dispersion spectrometry (EDS) method. From this data, the main component of the soft nanoneedle is Silicon as expected.

5.3 Application of Soft Nanoneedle: Single Cell Whole Stiffness Characterization

The whole stiffness response of a single cell, i.e. the deformation of the entire cell upon the applied load from a single point indentation, to the best of our knowledge, has not been previously reported. The difficulty to achieve such data is due to the stiff indenter which may penetrate or burst the cell. Our soft nanoneedle can be used for preventing the cell penetration by its ability to buckle during indentation. The mechanism of the cell deformation is stated as the following; upon indentation of the soft nanoneedle on the cell, local deformation occurred below the tip of the nanoneedle. Further indentation has induced the whole cell to deform in addition to the local cell deformation. The ability to produce a large local point indentation could give more information regarding the mechanical property of the organelles inside the cell.

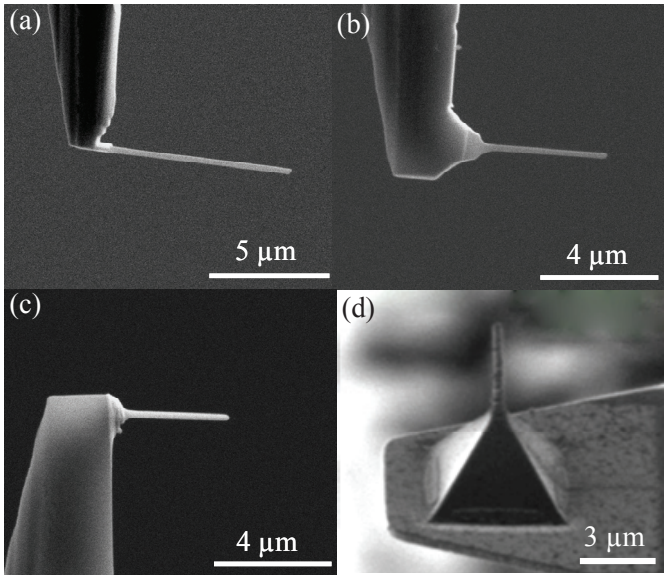


Fig. 17. Actual images of (a) Si, (b) Si-Ti, (c) W_2 and (d) $W_{0.09}$ nanoneedles.

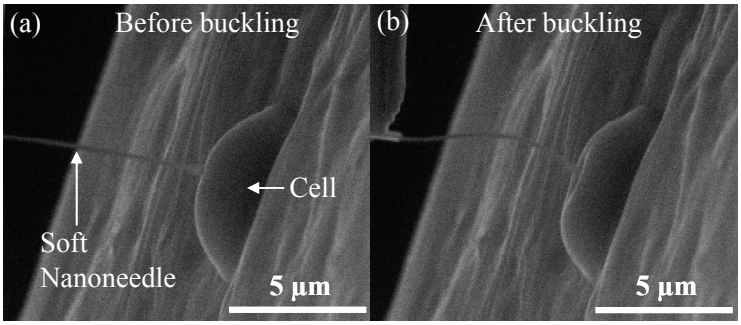


Fig. 18. Single cell global stiffness measurement from a single point using the Si nanoneedle. Two images show the Si nanoneedle at (a) straight condition and (b) buckling condition.

The global stiffness measurement of single cells from single point indentation was performed using a Si nanoneedle as shown in Fig. 18. The measurement was performed using a standard indentation procedure. The Si nanoneedle started to buckle after the cell deformed for about 0.5 μm. From this point, the buckling rate of the Si nanoneedle increased with the increase of the indentation depth. Fig. 19 shows the force-cell deformation response of a single cell using the Si nanoneedles. By using (5) (based on Hertz-Sneddon conical tip model) as described earlier, the spring constant and Young modulus of single *W303* yeast cells were estimated as presented in Table 5. The values of k_{cell} for about the same physical parameters of two yeast cells, i.e. 0.92 ± 0.12 N/m and 0.95 ± 0.36 N/m show strong mechanical property similarity. The values which represent the whole cell spring constants are reasonable as compared to the reported local spring constant of the *Saccharomyces cerevisiae* yeast cell (0.06 ± 0.025 N/m) (Pelling et al., 2004). The values of whole E_{cell}, i.e. 3.64 MPa and 3.92 MPa are also rational since they represent the whole cell stiffness property as compared to the reported local Young modulus of the yeast cell, i.e. 0.72 ± 0.06 MPa (Pelling et al., 2004).

Our method shows improved data sensitivity as compared to the other global stiffness measurement method that relies totally in a compression of a single cell between one big indenter and a substrate. Data from this method may not represent the actual global cell stiffness as extra dissipation force may be included in the measurement as reported by (Smith et al., 2000) for the global cell spring constant and Young modulus of yeast cell as 11.1 N/m and 112 MPa.

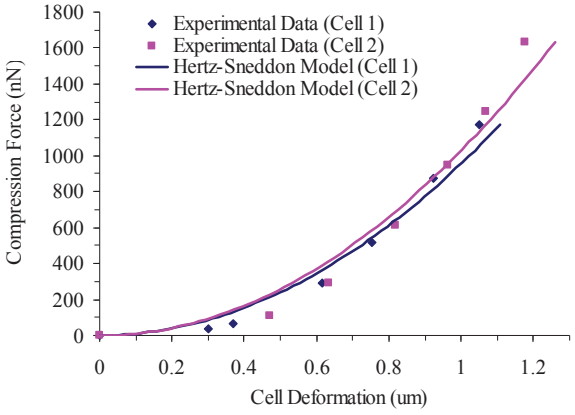


Fig. 19. Force-deformation curves of single cells using the Si nanoneedle. Experimental data were fitted with the Hertz-Sneddon conical tip model.

	Cell Physical Parameters		Cell Stiffness Characteristics	
	Height (μm)	Diameter (μm)	Spring Constant (N/m)	Young Modulus (MPa)
Cell 1	2.824	6.524	0.92	3.64
Cell 2	3.062	6.417	0.95	3.92

Table 5. The global characteristics of two single cells from local single point indentation using buckling Si nanoneedle.

5.4 Application of Hard Nanoneedle: Single Cell Local Stiffness Characterization

The local stiffness property, i.e. the deformation of the surface of the cell upon the applied load, on several points on mother cell and daughter cell, to the best of our knowledge, has never been reported. In order to achieve multi-point indentations on a single cell, we have used the standard indentation technique together with the image analysis to measure the cell deformation and compression force. The compression force was calculated using (1) and the value of φ was obtained from the bending of the cantilever. These values (cell deformation and compression force) were then fitted in (4) (Hertz-Sneddon cylindrical tip model) to obtain the local stiffness of the cell. $W_{0.09}$ nanoneedle which has small value of cantilever's spring constant (0.09 N/m) was used in this experiment as shown in Fig. 20.

Three random indentation points on each mother and daughter cell were performed as shown in Figs. 20(a) and (b) respectively. The cells can be recognized as mother and daughter cells from a direct image observation or from the cell volume comparison. The shape of the cells was assumed to be a prolate spheroid, and the volumes were calculated by using the formula $V=(\pi/6)lw^2$, where V is the volume, l is the height, and w is the width at maximum circumference (diameter) (Johnston et al., 1977). Thomas et al. (Thomas et al., 1980) reported about 38% difference in cell volume between fully separated mother and daughter cells of *Candida utilis* type of yeast cell. In our experiment, the volume difference of the mother and daughter cells is about 45% and this is reasonable since the daughter was not fully grown as it was still intact with its mother at the time of measurement.

The results from the indentations are shown in Fig. 21 and data were summarized in Table 6. From the data, all the three points on the mother and daughter cells show very good agreement to each other in local stiffness profile, i.e. for mother cell (1.47 MPa, 1.43 MPa and 1.47 MPa) and for daughter cell (1.11 MPa, 1.09 MPa and 1.09 MPa). On the other hand, the data revealed the heterogeneous local stiffness property between mother and daughter cell. This can be comprehended since the daughter cell was still in its early growth phase and the insertion of new cell wall macromolecules, as required for its strength, into its existing polymer network (Klis et al., 2006) was still carried out at the time of measurement and therefore its Young's modulus exhibited slightly lower value than the mother cell.

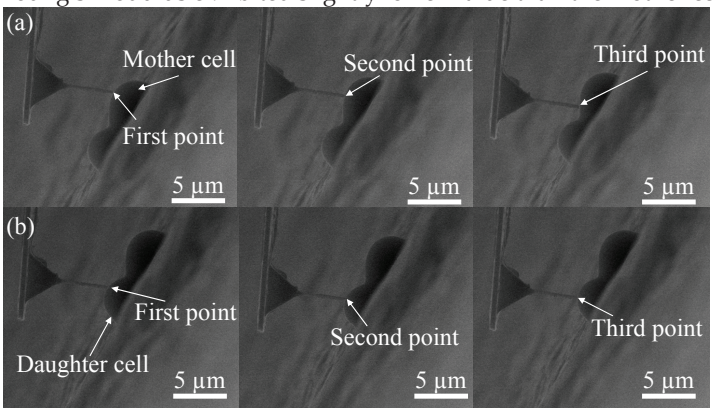


Fig. 20. Local stiffness measurement using the $W_{0.09}$ nanoneedle on three points of the (a) mother cell and (b) daughter cell.

The local stiffness characteristic of the mother cell as reported here falls in the same stiffness range as the data obtained from the literature. Peeling et al. (Pelling et al., 2004) reported 0.72 ± 0.06 MPa of the local stiffness property of the yeast cell. Lanero et al. (Lanero et al., 2006) reported 1.79 ± 0.08 MPa, 1.12 ± 0.02 and 2.0 ± 0.2 MPa for coated cell, uncoated cell and coated cell-bud scar of the *Saccharomyces cerevisiae* respectively. As compared to the *Saccharomyces cerevisiae* strain, baker-type yeast cells which has been widely used for the cell stiffness analysis, we have shown, for the first time, the mechanical property of the *W303* strain, wild-type yeast cell.

5.5 Application of Hard Nanoneedle: Single Cell Surgery

The large stiffness and the small diameter of the Si-Ti and W_2 nanoneedles have bring additional functionality to those nanoneedles. Fig. 14(c) shows the schematic of single cell surgery. In this experiment, we are not interested in the measurement of the elastic property of the cells but rather the ability of the nanoneedles to penetrate the cell without creating any cell bursting. As compared to our previous work (Ahmad et al., 2007), the penetration activities were followed by the cell bursting which killed the cell. This is unwanted scenario if we want to perform the future drug delivery and the single cell surgery. From the experimental real-time observation under ESEM, we can see clearly that a single cell was punctured by a nanoneedle without any cell bursting as can be seen from Fig. 22. Fig. 23 shows the force-cell deformation response of a single cell using Si-Ti nanoneedle and W_2 nanoneedle at three stages of indentation, i.e. early indentation (blue line), after penetration (pink line), and late indentation, i.e. the nanoneedle was touching the substrate (red line).

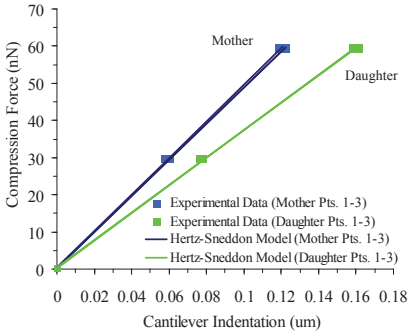


Fig. 21. Force-indentation curves of mother and daughter cells using the $W_{0.09}$ nanoneedle. Experimental data were fitted with the Hertz-Sneddon cylindrical tip model.

		Cell Height (μm)	Cell Diameter (μm)	Young Modulus (MPa)
Mother cell	Point 1	2.981	5.565	1.47
	Point 2			1.43
	Point 3			1.47
Daughter cell	Point 1	2.488	4.467	1.11
	Point 2			1.09
	Point 3			1.09

Table 6. The local stiffness characteristics of mother cell and daughter cell using $w_{0.09}$ nanoneedle.

At the first stage, the nanoneedle was indenting the cell without penetration of the cell as can be seen from the first inset of Fig. 23(a). The second stage begins when the nanoneedle started to penetrate the cell. The values of force and cell deformation at the point of penetration by using Si-Ti and W_2 nanoneedles were about 550 nN (0.42 μm) and 778 nN (0.56 μm). The increase in the penetration force by using the W_2 nanoneedle was due to the small bending of the nanoneedle prior penetration. This stage can be recognized from the drop in force value as can be seen from the second inset of Fig. 23(b). Since the penetration force before the penetration was quite high, the drop force was large due to the high momentum of the reversed force. Further indentation in the second stage had brought the nanoneedle in contact with the other cell wall that was supported by the substrate as can be seen by the sudden rise of force value as shown in the third inset of Fig. 23(c). In order to confirm the success of the injection of the nanoneedle without cell bursting, we intentionally hit the nanoneedle on the other side of the cell wall to break it as shown in Fig. 24. The injection without cell bursting is very important in single cell surgery and drug delivery systems. Nevertheless, the work on single cell surgery is still under investigation and more researchers are needed in this area to develop better devices for single cell surgery. One of the works in single cell surgery was done by (Obataya et al., 2005b). They measured force between 1 to 2 nN to penetrate a human epidermal melanocyte cell. The value of the force is small compared to the penetration force obtained in this paper because unlike the yeast cell, the human epidermal melanocyte cell does not have the cell wall. The cell wall which exists in the yeast cell is the main component which governs the stiffness properties of the cell (Klis et al., 2006).

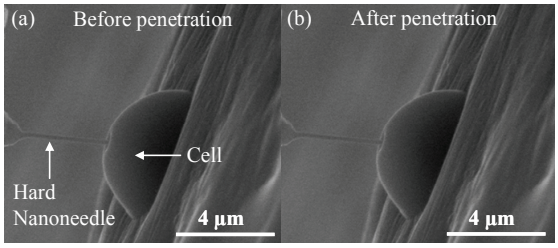


Fig. 22. Single cell surgery without cell bursting using Si-Ti nanoneedle.

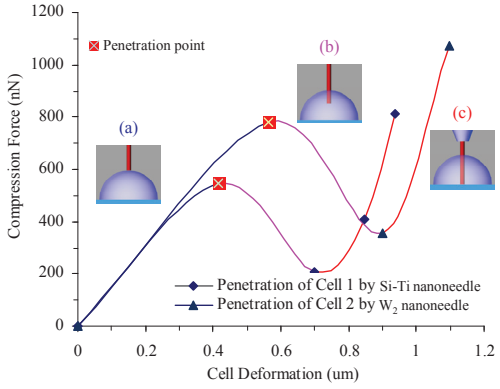


Fig. 23. Force-cell deformation curve using Ti-Si and W_2 nanoneedles at three different stages, i.e. (a) before penetration, (b) after penetration and (c) touching the substrate.

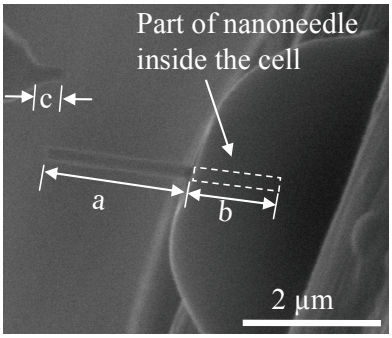


Fig. 24. Single cell surgery without cell bursting using Si-Ti nanoneedle. The dotted line represents the inserted nanoneedle inside the cell. The values for a, b and c are 2.16 μm, 1.19 μm and 0.30 μm respectively.

6. Conclusion

In this chapter, two important needs for the single cells analysis, i.e. the understanding of the mechanical properties of single cells and the implementation of the nanodevices in single cells mechanical property characterizations, have been addressed. The advantages of the integrated ESEM-nanomanipulation system rely on its capability to perform *in-situ* local direct observation and manipulation of a biological sample and the ability to control the environmental conditions. The ESEM-nanomanipulation system has a great capability for conducting single cells analysis. To the best of our knowledge, we, for the first time, have demonstrated the effect of the internal influences (cell size and growth phases) and the external influence (environmental conditions) on the cell strength. The penetration force is proportional to the increase in cell size, during the cell growth and under high vacuum condition. The elastic modulus of the cell increases dramatically under high vacuum condition but remain unaffected during the cell growth. Furthermore, we have highlighted the mechanical properties characterization of individual yeast cells from *W303* strain using four types of nanoneedles, i.e. Si, Si-Ti, $W_{0.09}$ and W_2 nanoneedles. We have demonstrated via experimental verification that all four types of nanoneedles described are very effective for yeast cell local stiffness characterization. This capability has numerous future applications especially in human disease detection. In addition to the mechanical characterization, Si-Ti and W_2 nanoneedles can also be applied in single cell surgery due to their strength. Penetrations of single cells have been successfully performed using either of these nanoneedles. This single cell surgery can be further applied in the future single cell drug delivery applications.

7. References

Brehm-Stecher, B.F., & Johson, E.A. (2004). Single-cell microbiology: Tools, technologies, and applications. *Microbiology and Molecular Biology Reviews 68*, 538-559.

Harold, F.M. (2005). Molecules into Cells: Specifying Spatial Architecture. *Microbiology and Molecular Biology 69*, 544-564.

Shapiro, H.M. (2000). Microbial analysis at the single-cell level: tasks and techniques. *Journal of Microbiological Methods 42*, 3-16.

Bausch, A.R., Moller, W., & Sachmann, E. (1999). Measurement of local viscoelasticity and forces in living cells by magnetic tweezers. *Biophysical Journal 76*, 573-579.

Gueta, R., Barlam, D., Shneck, R.Z., & Rousso, I. (2006). Measurement of the mechanical properties of isolated tectorial membrane using atomic force microscopy. *Proceedings of the National Academy of Sciences 103*, 14790–14795.

Thoumine, O., Ott, A., Cardoso, O., & Meister, J.J. (1999). Microplates: a new tool for manipulation and mechanical pertubation of individual cells. *J. Biochem Biophys Meth 39*, 47-62.

Obataya, I., Nakamura, C., Han, S.W., Nakamura, N., & Miyake, J. (2005a). Nanoscale Operation of a Living Cell Using an Atomic Force Microscope with a Nanoneedle. *Nano Letters 5*, 27-30.

Leary, S.P., Liu, C.Y., & Apuzzo, M.L.I. (2006). Toward the emergence of nanoneurosurgery: Part III - Nanomedicine: Targeted nanotherapy, nanosurgery, and progress toward the realization of nanoneurosurgery. *Neurosurgery 58*, 1009-1025.

Bianco, A., Kostarelos, K., & Prato, M. (2005). Applications of carbon nanotubes in drug delivery. *Current Opinion in Chemical Biology 9*, 674-679.

Brigger, I., Dubernet, C., & Couvreur, P. (2002). Nanoparticles in cancer theraphy and diagnosis. *Advance Drug Delivery Reviews 54*, 631-651.

Stylios, G.K., Giannoudis, P.V., & Wan, T. (2005). Applications of nanotechnologies in medical practice. *Injury, Int. J. Care Injured 36*, 6-13.

Goodman, M.B., Lumpkin, E.A., Ricci, A., Tracey, W.D., Kernan, M., & Nicolson, T. (2004). Molecules and Mechanisms of Mechanotransduction. *The Journal of Neuroscience 24*, 9220-9222.

Huang, H., Kamm, R.D., & Lee, R.T. (2004). Cell mechanics and mechanotransduction: pathway, probes, and physiology. *Am J Physiol Cell Physiol 287*, C1-C11.

Nakajima, M., Arai, F., & Fukuda, T. (2006). In situ Measurement of Young's Modulus of Carbon Nanotube inside TEM through Hybrid Nanorobotic Manipulation System. *IEEE Transactions on Nanotechnology 5*, 243-248.

Lanero, T.S., Cavalleri, O., Krol, S., Rolandi, R., & Gliozzi, A. (2006). Mechanical properties of single living cells encapsulated in polyelectrolyte matrixes. *Journal of Biotechnology 124*, 723-731.

Touhami, A., Nysten, B., & Dufrene, Y.F. (2003). Nanoscale Mapping of the Elasticity of Microbial Cells by Atomic Force Microscopy. *Langmuir 19*, 4539-4543.

Sneddon, I.N. (1965). The Relation Between Load and Penetration in the Axisymmetric Boussinesq Problem for a Punch of Arbitrary Profile. *Int. J. Engng Sci. 3*, 47-57.

Pharr, G.M., Oliver, W.C., & Brotzen, F.R. (2002). On the generality of the relationship among contact stiffness, contact area, and elastic modulus during indentation. *J. Mat. Res. 7*, 619-627.

Smith, A.E., Zhang, Z., Thomas, C.R., Moxham, K.E., & Middelberg, A.P.J. (2000). The mechanical properties of *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences 97*, 9871-9874.

Klis, F.M., Boorsma, A., & Groot, P.W.J.D. (2006). Cell wall construction in *Saccharomyces cerevisiae*. *Yeast 23*, 185-202.

Lesage, G., & Bussey, H. (2006). Cell wall assembly in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews 70*, 317-343.

Srinorakutara, T. (1998). Study on the effect of diluent osmotic pressure on yeast cell strength and cell size using a novel micromanipulation technique. *World Journal of Microbiology & Biotechnology 14*, 719-725.

Fleet, G.H., & Manners, D.J. (1976). Isolation and composition of an alkali-soluble glucan from the cell walls of *Saccharomyces cerevisiae*. *Journal of General Microbiology 94*, 180-192.

Walker, G.M. (1998). *Yeast Physiology and Biotechnology*. Chichester: John Wiley & Sons.

Obataya, I. (2005). Nanoscale Operation of a Living Cell Using an Atomic Force Microscope with a Nanoneedle. *Nano Letters 5*, 27-30.

Pelling, A.E., Sehati, S., Gralla, E.B., Valentine, J.S., & Gimzewski, J.K. (2004). Local Nanomechanical Motion of the Cell Wall of Saccharomyces cerevisiae. *Science 305*, 1147-1150.

Lanero, T.S. (2006). Mechanical properties of single living cells encapsulated in polyelectrolyte matrixes. *Journal of Biotechnology 124*, 723-731.

Morris, C.E., & Homann, U. (2000). Cell Surface Area Regulation and Membrane Tension. *Journal of Membrane Biolology 179*, 79-102.

Tortora, G.J., Funke, B.R., & Case, C.L. (2003). *Microbiology: An Introduction*: Benjamin Cummings.

Johnston, G.C., Pringle, J.R., & Hartwell, L.H. (1977). Coordination of growth with the cell division in the yeast *Saccharomyces cerevisiae*. *Experimental Cell Research 105*, 79-98.

Thomas, K.C., dawson, P.S.S., & Gamborg, B.L. (1980). Differential Growth Rates of *Candida utilis* Mother and Daughter Cells Under Phase Cultivation. *Journal of Bacteriology 141*, 1-9.

Ahmad, M.R., Nakajima, M., Kojima, S., Homma, M., & Fukuda, T. (2007). In-situ Single Cell Mechanical Characterization of W303 Yeast Cells inside Environmental-SEM. *Proc. of the 7th IEEE International Conference on Nanotechnology, vol. 1*. Hong Kong, pp. 1022-1027.

Obataya, F., Nakamura, C., & Han, S.W. (2005b). Mechanical sensing of the penetration of various nanoneedles into a living cell using atomic force microscopy. *Journal of Biosensors & Bioelectronics 20*, 1652-1655.

20

Anti silver nanoparticle bacteria

Chi-Yea Yang
Department of Biotechnology, Vanung University
TAIWAN

1. Introduction

Nanoparticles (NPs) are in the 1- to 100-nm size range and can be composed of many different base materials. Particles in the nanometer size range do occur both in nature and as a result of industrial processes. NPs have been widely used today, especially in the latest developed science materials. Silver nanoparticles (Ag-NPs), one of the most popular antimicrobial materials had been generally utilized in the textile industry and medical engineering (7, 13). Different Ag-NPs fabricating methods will make different physical or chemical property. Besides from silver nanofibers and plasma nanoparticles, most textile processing methods involve the dipping method that produce a lot of Ag-NPs containing waste water. The entire industrial Ag-NPs containing waste water still has not any proper efficient method to grip them out from the waste before it was drained into the environment (1, 6). A novel bacteria mutant was isolated from the ordinary laboratory analysis, and was revealed and identified. This mutant strain could degrade 5 different trade mark Ag-NPs, directly. Under the textile NPs operation concentration, which exhibit a complete antimicrobial activity, all these five different products of commercial Ag-NPs will be degraded into none antimicrobial activity by this mutant within 4 hours. Atomic absorbance analysis observed that the Ag-NPs were absorbed by the cell directly.

Nanotechnology was expected to have a revolutionary impact on life science. A variety of material production processes occur at nanometer length scales (4). Such as, textile Ag-NPs antimicrobial activity, gold nanoparticles super catalase activity and most NPs offer some unique advantages as sensing and drug delivery (2). Due to different fabrication methods, such many varieties of NPs were available: polymeric nanoparticles, dendrimers, metal nanoparticles, liposomes and other types of nano-assemblies (11). They are multipurpose agents with a diversity of biomedical applications including high sensitive diagnostic assays and radiotherapy enhancement (2), as well as drug and gene delivery analysis (5). Since Ag-NPs excellent antimicrobial activity had been applied onto textile industry, cosmetic products and food preservation followed, but the nanoparticle safety problems were confirmed by less positive reports when some others status on the contrary side (3, 4).

Most NPs were never been presented in human life. When more and more NPs have been developed and applied into the environment, how could we estimate the environmental impact (9)? Anyway, today NPs have to be used incessantly and enormously; here we isolate a novel *Klebsiella* mutant strain, an opportunistic pathogen found in the environment

and on mammalian mucosal surfaces, which could degrade many different kinds of Ag-NPs, directly, within 2~4 hours. After this *Klebsiella* mutant had precipitate down the NPs, the solution turn to none antimicrobial activity, at all, and the waste could be drained out into the aquatic environment. It means that, this *Klebsiella* mutant could be served as a powerful bioremediation tool for industrial Ag-NPs containing waste management.

2. Experiments

2.1. Bacteria strains

Bacteria strains used in this study were *Staphylococcus aureus*（ATCC6538P, *Escherichia coli*（BCRC 16081） and *Salmonella enterotidis*（BCRC10744）. These strains were obtained from American Type Culture Collection (ATCC, Rockville, MD, U.S.A.); Culture Collection and Research Center (CCRC, Hsin-Chu, Taiwan).

2.2. PCR primers design

The primer sequences were chosen from the conserved regions of the bacteria 16S rDNA. The sequences of the PCR primers designed were as follows: forward primer, 16S-350F (5′-CCTACGGGAGGCAGCAGT-3′), corresponding to *E. coli* 16S rDNA (accession number J01859), and reverse primer, 16S-820R (5′-CGTTTACGGCGTGGACTAC-3′). Primers were synthesized with a model 394 DNA-RNA synthesizer (Applied Biosystems, Foster City, Calif.).

2.3. DNA isolation and PCR amplification

Genomic DNA from *Klebsiella* sp. was extracted and purified using Genomic Tips (Qiagen, Valencia, Calif.) with Genomic DNA Buffer Set (Qiagen). The concentration of DNA was determined by UV spectrophotometer (Sigma, St. Louis, Mo.). DNA fragment was amplified by using the primer set 16S-350F and 16S-820R that were targeted to universally conserved regions. The amplification reactions were performed with a GeneAmp 2400 PCR thermocycler (Perkin-Elmer, Norwalk, Conn.). The reaction mixtures contained 15 pmol of primers, 200 μM each deoxyribonucleoside triphosphate, 10 mM Tris-HCl (pH 8.0), 1.5 mM $MgCl_2$, 1 U of DynaZyme DNA polymerase (Finnzymes Oy, Espoo, Finland) and 3 μl of the cell-free DNA, in a final volume of 50 μl. The PCR mixtures were then subjected to 95℃ for 5 min followed by 35 cycles of 95℃ for 30 s, 56℃for 30 s, and 72℃ for 30 s, and after amplification an extended step of 5 min at 72℃. The PCR products were confirmed by 2% agarose gel electrophoresis, and proceeded for sequence analysis by the Sequencer (Perkin-Elmer, Norwalk, Conn.). All these sequences were alignment and blasted in the NCBI databank.

2.4. Silver nanoparticle antimicrobial activity analysis

The antimicrobial activity of silver nanoparticle was tested quantitatively by a viable cell count method, depending on JIS-L-1902-2000 standard method.

S. aureus was grown in BHI broth (Difco) and incubated aerobically for 16h at 37 ℃. Another two Gram-negative bacteria, *S. typhimurium* and *E. coli* , were separately cultivated in TS broth (Difco) at 37 ℃ for 16 h. Each 30 mL tube of bacterial cell culture was then

centrifuged for 5 min at 4 °C and 7000 x*g*, decanted, washed with 0.1% peptone (Difco), centrifuged for 5 min, and decanted. The cell pellet was placed into 100 mL of BHI or TS broth and diluted to 10% of the original broth concentration with 900 mL of sterile distilled water to obtain an inoculum of (1.0~2.5) $\times 10^6$ colony-forming units (CFU)/ mL. Then, 0.5 mL of the inoculum was aseptically added to each of the microtube containing 0.5 mL Ag-NPs solution, and keep agitation at 50 rpm in 37 °C for 18 hours. For each analyzed bacteria, an inoculum of cell suspension in a microtube without Ag-NPs solution was used as a control. Aliquots of 0.1 mL of cell suspension were periodically taken from the microtube, and plated in duplicate on BHI agar for *S. aureus* cells or on TS agar for *S. typhimurium* and *E. coli* cells. The plates were incubated in an aerobic chamber for 16 hours at 37 °C. The number of colonies on each plate was counted and calculated to the inhibit activity.

2.5. Antagonism analysis of the mutant with silver nanoparticle

Klebsiella sp. mutant was cultured at 37 °C overnight from a single colony. Aspirate 0.5 ml cultured broth (1.0~2.0X10^9 cfu/ml) mixed with different Ag-NPs to the final concentration of 0.5, 1.0, 1.5, 2.0 and 2.5 %. After incubate at 37 °C for different period of time (1~4 hours), the mixtures were boiled for 10 minutes to inactive the bacteria, and centrifuged in 10,000 xg for 10 minutes, followed by aspirate 0.5ml, cell free, clear supernatant for antimicrobial activity analysis.

3. Conclusion

Today, nanotechnology is a major innovative scientific and economic growth area. However nanomaterial residues may have a detrimental effect on human health and the environment. To date there is a lack of quantitative ecotoxicity data, and recently there has been great scientific concern about the possible adverse effects that may be associated with manufactured nanomaterials (6, 8, 10). Nanomaterial properties were different from the parent compounds, more than 50% of the atoms in NPs are on the surface, resulting in greater activity than bulk materials this make the NPs have a complete dissimilar chemical, biological activity than parent compounds (12).

Ag-NPs have been known for it's inhibit and bactericidal activity and more than 85% major pathogens were inhibited by this new material (7). Unfortunately, in our laboratory routine analysis, we isolate a mutant strain which could resistant to Ag-NPs bactericidal activity. The occurrence was seems like to the bacterial multidrug resistance revolution.

This *Klebsiella* mutant is a gram negative rod bacilli, and all the biochemical identification analysis of API 40H exhibit that this novel mutant was *Klebsiella* sp.. By the bacteria conserved 16S rRNA sequence analysis and alignment in the GeneBank, we could make sure that this mutant was *Klebsiella pneumoniae*.

All these analyzed Ag-NPs (NTX-205, 305, LG-05, 06, VY707) were purchased from the cooperative companies products (Table 1), and renumbered. The antimicrobial activity analysis was preceded by the modified standard method of JIS-L-1902-2000, and all these five Ag-NPs exhibit a strong antimicrobial activity with less than 1.0 % concentration against gram positive or gram negative bacteria (Table 1). After this *Klebsiella* mutant had inactive Ag-NPs containing solutions, the cell free supernatant exhibits no antimicrobial

activity anymore, then the analyzed bacteria could grow up on the plate again rather than the control group without any colony on the plate. It means that, this *Klebsiella* mutant could degrade these five different kinds of Ag-NPs, properly, and make the Ag-NPs containing solution turned to harmless and safety normal waste water.

The industrial scale production and wide variety of applications of manufactured NPs and their possible release in considerable amounts into the natural aquatic environment have produced an increasing concern among the nanotechnology and environmental science community. Assessing the risks of these NPs in the environment requires an understanding of their mobility, reactivity, ecotoxicity and persistency (6, 12).

In addition, release of engineered NPs containing waste into the aquatic environment is largely an unknown. NPs provide surfaces that could bind and transport toxic chemical pollutants, which maybe a mutagen or a nutrition molecular. Industrial NPs was the newest molecular of microorganism who had never been touched within the past few billion years. Unfortunately, microorganism must could faced up the industrial NPs, but the quantity have to be concerned, because that must will make the ecotoxicological problems turned to more complex, and the environment may will be no more predictable.

How much environmental damage could be estimated? Never know. Oberdorster *et al.* investigated the effect of water-soluble fullerene aggregates, nano-C_{60}, on HDF, HepG2, and NHA cells in culture. Nano-C_{60} demonstrated significant toxicity in cell culture studies, while a highly hydroxylated, water-soluble fullerene, $C_{60}(OH)_{24}$ did not (10). It means that each new engineered nanoparticle should be analyzed relative ecotoxicological analysis.

In contrast to environmental harmful effects, NPs could interact with some toxic chemical compounds and alleviate the chemical toxicity. But we almost could make sure that no engineered NPs was produced to treatment the toxic chemical compounds containing waste, today.

Different nanoparticles fabricated from different carrier, cortex, protocol and modification. In these five different Ag-NPs analysis results, different kinds of NPs will not influence this *Klebsiella* mutant degradation activity, even the Ag-NPs concentration was increased to 2.0 %, the mutant still could degrade Ag-NPs containing solution down to none bactericidal activity within 4 hours.

At the cellular level, bacteria may be largely protected against the uptake of many types of NPs since they do not have mechanisms for transport of colloidal particles across the cell wall (7). But quite a lot of results were available for Ag-NPs due to their use as bactericides. The bacterial cell membrane proteins were damaged in the presence of Ag-NPs, finally resulting in death of the organisms (13). Another study reported that C_{60} adsorption onto the gram-negative *E. coli* was 10 times higher than on gram-positive *Bacillus subtilis* (11). The interaction of NPs with the cell is size-dependent and seems to depend on the shape of the NPs (11).

Most of the countries, NPs containing waste had been extensively, directly drained into environment through the industrial manipulation of such many Ag-NPs containing items. Due to the environmental impact of Ag-NPs was under evaluated, most country governments were still uncertain, unknown, how to manage the relative products. People could not withdraw the science development, but we have to evaluate the risk assessment of NPs, carefully, before the bioterrorism was happened on us. We proposed that this *Klebsiella*

mutant could be served as a powerful tool for NPs containing waste management and slow down the environmental impact.

Instead, ecotoxicology in this emerging area must provide a basis for predicting systematically how Ag-NPs biological behavior relates to its structure, composition and morphology. In vitro cell culture experiments are well suited for developing mechanistic models to inform material development. All these results could set as reference data for future efforts to characterize the environmental and health impacts of the other engineered NPs. Ultimately, such positive environmental and toxicological studies will be imperative to ensure the NPs design process yields both effective and safety engineered products.

4. References

Farre M., Gajda-Schrantz K., Kantiani L. and Barcelo D. 2009. Ecotoxicity and analysis of nanomaterials in the aquatic environment. Anal. Bioanal. Chem. 393:81-95.

Hainfeld J. F., Slatkin D. N. and Smilowitz H. M. 2004. The use of gold nanoparticles to enhance radiotherapy in mice. Phys. Med. Biol. 49:309–315.

Handy R. D., Owen R. and Valsami-Jones E. 2008. The ecotoxicology of nanoparticles and nanomaterials: current status, knowledge gaps, challenges, and future needs. Ecotoxicology. 17:315-325.

Ju-Nam Y. and Lead J. R. 2008. Manufactured nanoparticles: an overview of their chemistry, interactions and potential environmental implications. Sci. Total. Environ. 400:396-414.

Lord M. S., Cousins B. G., Doherty P. J., Whitelock J. M., Simmons A., Williams R. L. and Milthorpe B. K. 2006. The effect of silica nanoparticulate coatings on serum protein adsorption and cellular response. Biomaterials. 27:4856-4862.

Moore M. N. 2006 Do nanoparticles present ecotoxicologiocal risks for the health of the aquatic environment. Environ. Int. 32:967–976.

Morones J. R., Elechiguerra J. L., Camacho A., Holt K., Kouri J. B., Ramirez J. T. and Yacaman M. J. 2005. The bactericidal effect of silver nanoparticles, Nanotechnol. 16:2346–2353.

Nel A., Xia T., Madler L. and Li N. 2006. Toxic potential of materials at the nano level. Science 311:622–627.

Nowack B. and Bucheli T. D. 2007. Occurrence, behavior and effects of nanoparticles in the environment. Environ. Pollut. 150:5-22.

Oberdorster E., Zhu S., Blickley T. M., McClellan-Green P. and Haasch M. L. 2006. Ecotoxicology of carbon-based engineered nanoparticles: effects of fullerene (C_{60}) on aquatic organisms. Carbon 44:1112–1120.

Pal S., Tak Y. K. and Song J. M. 2007. Does the antibacterial activity of silver nanoparticles depend on the shape of the nanoparticle? A study of the gram-negative bacterium *Escherichia coli*. Appl. Environ. Microbiol. 73:1712–1720.

Phenrat T., Saleh N., Sirk K., Tilton R. D. and Lowry G. V. 2007. Aggregation and sedimentation of aqueous nanoscale zerovalent iron dispersions. Environ. Sci. Technol. 41:284–290.

Sondi, I. and Salopek-Sondi, B. 2004. Silver nanoparticles as antimicrobial agent: a case study on *E. coli* as a model for Gram-negative bacteria. J. of Colloid and Interface Science. 275:177-182.

Nanoparticles	**NTX-205**	**NTX-305**	**LG-05**	**LG-06**	**VY-707**
Original concentration	6.0%	6.0%	8.0%	8.0%	12.0%
Concentration[1] of 100% inhibit activity[2]	0.5%	0.45%	0.6%	0.65%	0.4%
Production country	Taiwan	Taiwan	Korea	Korea	China

1. the concentration was refer to the original solution by volumn/volumn (v/v).
antimicrobial inhibit activity=[1－sample cfu/control cfu] x 100%

Table 1. The antimicrobial activity of the five different silver nanoparticles.