Transactions of the ASME®

Transactions of the ASME, Journal of Heat Transfer (ISSN 0022-1481) is published bi-monthly (Feb., Apr., June, Aug., Oct., Dec.) by The American Society of Mechanical Engineers, Three Park Avenue, New York, NY 10016. Periodicals postage paid at New York, NY and additional mailing offices. POSTMASTER: Send address changes to Transactions of the ASME, Journal of Heat Transfer, c/o THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 22 Law Drive, Box 2300, Fairfield, NJ 07007-2300. **CHANGES OF ADDRESS** must be received at Society headquarters seven weeks before they are to be effective. Please send old label and new address.
 INDEXED by Applied Mechanics Reviews and Engineering Information, Inc. Canadian Goods & Services Tax Registration #126148048.

Journal of Heat Transfer

Published Bimonthly by The American Society of Mechanical Engineers

VOLUME 124 • NUMBER 2 • APRIL 2002

(Contents continued on inside back cover)

This journal is printed on acid-free paper, which exceeds the ANSI Z39.48-1992 specification for permanence of paper and library materials. ♾™
♻ 85% recycled content, including 10% post-consumer fibers.

Journal of Heat Transfer Volume 124, Number 2 APRIL 2002

TECHNICAL NOTES

ANNOUNCEMENTS

Journal of Heat Transfer

Foreword

It is difficult to exactly define when research in micro/nanoscale heat transfer was initiated in the heat transfer community worldwide. There was no single paper that really started the field and there wasn't any single inspiring lecture such as the famous ''There is Plenty of Room at the Bottom,'' that Richard Feynman gave in 1959. The physics and chemistry community had been looking at some microscopic aspects of phase change and transport phenomena for several decades. In fact, a few publications in microscale heat transfer that appeared in this journal can be traced back to the 1960s. However, it is clear that by the end of the 1980s, there was an increasing interest within the heat transfer, as well as the engineering community at large, for studying heat and mass transfer at small length and time scales. This was partly driven by the tremendous success of microelectronics and a clear vision that devices were going to become smaller and faster, and that future developments would require an extensive knowledge base for the continuously shrinking time and length scales that would reach tens of nanometers within decades. This also coincided with the invention of a wide variety of experimental tools such as the scanning tunneling and force microscopes and femtosecond lasers, which provided immediate access to phenomena at nanoscales. The ability to make microdevices using MEMS technology became standardized at around the same time. In addition, computing power became readily available such that molecular dynamic and stochastic simulations of micro/nanoscale thermal phenomena became feasible. All these factors put together led to an intense influx and genesis of new theoretical concepts, experimental techniques, and device designs, which led to a remarkable pace of progress. It is perhaps fair to say that over the past decade, micro/nanoscale heat transfer has been the most exciting and active area of research within the heat transfer community, and has also attracted the most amount of research funding. What is most heartening is that the participation has been highly collaborative and interdisciplinary, involving researchers from mechanical, chemical and electrical engineering, as well as physics, chemistry and materials science. From the very start, research in this field had an international flavor, which led to several focused workshops and symposia involving people from around the world, but mainly from the USA, Asia, and Europe. The number of sessions in conferences as well as the number of publications in this and other scientific journals increased very rapidly. Several books were written on various micro/nanoscale heat transfer topics. New courses on this topic were started at both undergraduate and graduate levels. In fact, perhaps the most positive outcome has been the birth of a new breed of young researchers for whom the boundaries of traditional disciplines have already been torn down.

To reflect upon our rapid progress of the past decade, to define the state of the art today, and to look ahead towards a very promising and perhaps even more progressive future, we decided to put together this special issue on micro/nanoscale heat and mass transfer. The issue contains a few review papers on topics that are mature and have sufficiently well-established research content. By no means do they cover all the topics of micro/nanoscale heat and mass transfer. The purpose of the review papers is to capture the past progress in a field and present new challenges and questions for the future. We hope these papers will serve the community well, in particular, for those who want to enter the field as a new researcher and get abreast with the state of the art of the field. The issue also contains several regular publications that report new research covering various topics in micro/nanoscale heat and mass transfer.

While it is easy and reassuring for us to discuss our research achievements over the past decade, we must continue to ask some hard and sometimes unnerving questions about the field. Why is micro/nanoscale heat and mass transfer important? What impact has it made or, more importantly, can it make? Why is it attracting so much new research funding? Is it just a fad or will it last? Does it have any scientific depth? Is there a roadmap for the future? While we cannot obviously answer all these questions at once, the question of why one should miniaturize is relevant and worth some thought. Two reasons come to mind. The first is a scientific one whereas the second is more technological. The fundamental length scales of nature that are involved in heat and mass transfer happen to be at micro and nanoscales. For example, the wavelength and mean free path of electrons and phonons in solids, mean free path of molecules in gases at atmospheric pressure, and the range of intermolecular forces in liquids generally lie in the length scale range of 100 nm or below. If one can probe and understand phenomena at these length scales, it may be possible to control and engineer heat and mass transfer in unprecedented ways. The second key reason is directly related to the fact that we are able to manufacture and manipulate systems at length scales that only a few decades ago were unthinkable. It is worth noting that the commercial success of microelectronics lies not only in the fact that one can fabricate sub-micrometer devices, but more so in the low cost of manufacturing integrated systems using parallel lithographic techniques. If the same manufacturing techniques can be exploited for thermal engineering, it would be possible to design and manufacture thermal micro/nanodevices that are not only inexpensive but

also have better performance than their larger counterparts. In fact, if one could combine both these aspects—nanoscale fundamental control and low-cost microfabrication—it seems reasonable to expect commercial devices and systems with new functionality and high performance. It is the dual vision of generating scientific knowledge and enabling commercial technology that ought to stimulate further research in this field. We hope this special issue will play a positive role for achieving this dual vision.

A. Majumdar

G. P. Peterson

D. Poulikakos

Thermometry and Thermal Transport in Micro/Nanoscale Solid-State Devices and Structures

David G. Cahill
Department of Materials Science and Engineering,
and the Frederick Seitz Materials Research Laboratory,
University of Illinois,
Urbana, IL 61801

Kenneth Goodson
Department of Mechanical Engineering,
Stanford University,
Stanford, CA 94305

Arunava Majumdar
Department of Mechanical Engineering,
University of California,
Berkeley, CA 94720
e-mail: majumdar@me.berkeley.edu

We review recent advances in experimental methods for high spatial-resolution and high time-resolution thermometry, and the application of these and related methods for measurements of thermal transport in low-dimensional structures. Scanning thermal microscopy (SThM) achieves lateral resolutions of 50 nm and a measurement bandwidth of 100 kHz; SThM has been used to characterize differences in energy dissipation in single-wall and multi-wall carbon nanotubes. Picosecond thermoreflectance enables ultrahigh time-resolution in thermal diffusion experiments and characterization of heat flow across interfaces between materials; the thermal conductance G of interfaces between dissimilar materials spans a relatively small range, $20<G<200$ MW m^{-2} K^{-1} near room temperature. Scanning thermoreflectance microscopy provides nanosecond time resolution and submicron lateral resolution needed for studies of heat transfer in microelectronic, optoelectronic and micromechanical systems. A fully-micromachined solid immersion lens has been demonstrated and achieves thermal-radiation imaging with lateral resolution at far below the diffraction limit, <2 μm. Microfabricated metal bridges using electrical resistance thermometry and joule heating give precise data for thermal conductivity of single crystal films, multilayer thin films, epitaxial superlattices, polycrystalline films, and interlayer dielectrics. The room temperature thermal conductivity of single crystal films of Si is strongly reduced for layer thickness below 100 nm. The through-thickness thermal conductivity of Si-Ge and GaAs-AlAs superlattices has recently been shown to be smaller than the conductivity of the corresponding alloy. The 3ω method has been recently extended to measurements of anisotropic conduction in polyimide and superlattices. Data for carbon nanotubes measured using micromachined and suspended heaters and thermometers indicate a conductivity near room temperature greater than diamond.
[DOI: 10.1115/1.1454111]

Keywords: Conduction, Heat Transfer, Interface, Measurement Techniques, Microscale, Nanoscale, Thin Films

1 Introduction

The last 50 years have witnessed the transformation of solid-state devices from objects of mere scientific curiosity to being at the heart of information technology. Electronic and optoelectronic devices are now so ubiquitous in modern society that their use often goes unnoticed. One of the main factors that has enabled this dramatic transformation is the ability to control the flow of charge in semiconductors. While charge flows only in metals and semiconductors, heat on the other hand flows in all materials. It is, therefore, natural to ask the question: What are the fundamental and engineering limits on the control of heat flow in solids?

Heat flow in solids occurs predominantly due to two energy carriers, namely, electrons in metals, and lattice vibrations or phonons in insulators and semiconductors. If one has to control their flow, one must recognize two fundamental length scales—wavelength and mean free path—associated with both electrons and phonons. One way to gain greater control of heat flow is to design and fabricate solid structures that overlap in size with the fundamental length scales. Table 1 lists the typical range of values for the length and time scales of electrons and phonons. It is obvious that size control of solid structures must occur at nanoscales. Modern manufacturing techniques allow us to fabricate structures that are confined to 2 (films), 1 (wires), or 0 (dots) dimensions with sizes in the 1–100 nm range. Hence, it is now possible, to some extent, to manipulate heat flow in solids through the interplay of confined size with electron/phonon wavelengths and mean free paths. Although much remains to be understood about heat flow in micro/nanostructures, significant progress has been made over the last decade or so. This could not have been possible without the development of new measurement techniques that allow one to probe heat flow at length and time scales comparable to those listed in Table 1. The purpose of this paper is twofold: (i) to provide a critical review of these measurement techniques and a discussion of future areas of development; and (ii) to present our current understanding of heat flow in micro/nanostructures and identify areas that need further research.

While greater control of heat flow might indeed be possible using confined structures, the following question is rather pertinent: Why is it important and what impact could it have? As the demands of information technology keep increasing, new engineering challenges continuously emerge. For example, the rapid increase in power density and speed of integrated circuits requires the use of unconventional materials. Figure 1(*a*) shows a micrograph of a polycrystalline diamond thin film grown by chemical vapor deposition on a silicon wafer [1]. The grain size and orientation, which depend on the details of the nucleation and growth processes, influence the thermal conductivities and thereby affect heat removal rate and response speed. Passivation materials in integrated circuits are often chosen for their ability to reduce elec-

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division July 27, 2001; revision received December 7, 2001. Associate Editor: V. K. Dhir.

Table 1 Typical length and time scales for energy carriers in solids at room temperature

	Electrons in Metal	Phonons in Insulators and Semiconductors
Wavelength, λ [nm]	0.1-1	1-5
Mean Free Path, ℓ [nm]	10	10-100
Relaxation Time, τ [s]	$(10\text{-}100)\times10^{-15}$	$(1\text{-}10)\times10^{-12}$
Propagation Speed, v [m/s]	$\approx10^{6}$	$(3\text{-}10)\times10^{3}$

tromagnetic cross-talk between neighboring interconnects and devices. However, they often have very low thermal conductivities. The importance of thermal properties of passivation materials in integrated circuits is illustrated by the metal-silicon contact in Fig. 1(*b*), whose temperature rise during rapid current transients is strongly influenced by heat diffusion into the bounding silicon dioxide [2]. One of the major breakthroughs in fiber optic communication is that of dense wavelength division multiplexing (DWDM) [3] in which a single optical fiber carries encoded information from about 100 lasers operating nominally at 1550 nm wavelength but with their wavelengths separated by about 0.4–0.8 nm. Figure 2 shows the cross-sectional view [4] of vertical-cavity surface-emitting laser (VCSEL) [5,6], which is made up of multiple layers of III–V semiconducting materials nanostructured

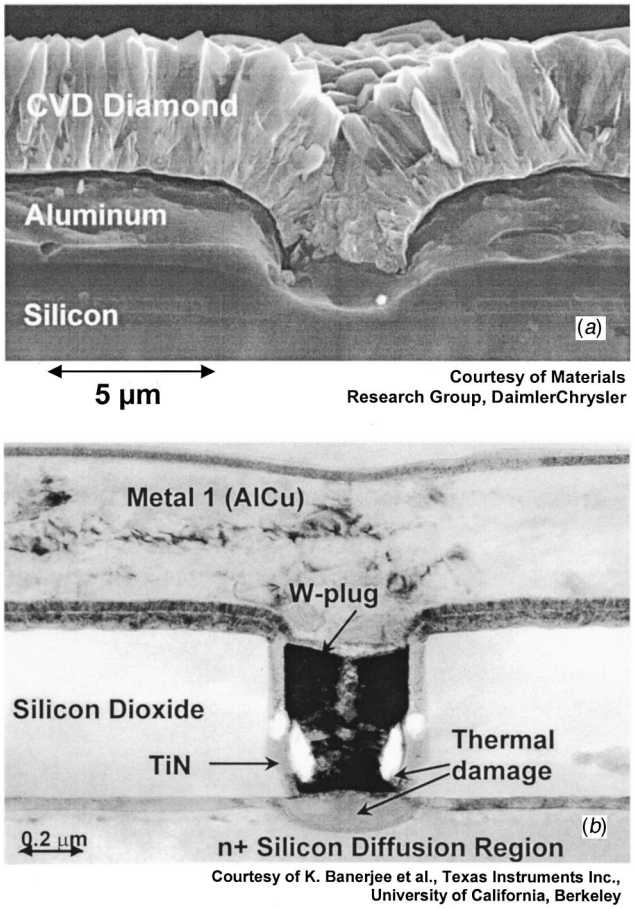


Fig. 1 Electron micrographs of novel electronic films and devices: (*a*) Diamond passivation film on aluminum. Diamond films are promising for enhanced heat removal from power integrated circuits and for fast thermal sensors. (*b*) Transmission electron micrograph of a VLSI metal-silicon contact, which has failed during a current pulse of sub-microsecond duration. The temperature rise is strongly influenced by the properties of the surrounding passive material.

in the 5–100 nm range. The maximum optical power output of these lasers depends critically on the heat removal rate. More importantly for DWDM, because the refractive index of the III–V materials is temperature dependent, the output wavelength varies with temperature at about 0.1 nm/K, which can lead to cross-talk between channels. Hence, it is necessary to actively control their temperature, which requires one to develop solid-state cooling devices. Figure 3 shows an array of thermionic microcoolers fabricated on a silicon wafer [7]. The cross-section of this device shows that it contains a Si/SiGe superlattice with a period of about 10 nm. The purpose of this superlattice is to minimize heat flow between the cold and hot ends of the refrigerator. This leads to the question as to why such solid-state refrigerators and engines cannot replace their macroscopic vapor based ones.

It is often forgotten that when semiconductors were first discovered, the initial excitement about their use was not for electronics but for refrigeration [8]. The reason we do not generally use solid-state energy conversion is because of their low device performance. This depends on the non-dimensional thermoelectric figure of merit of a material, $ZT=S^2\sigma T/k$, where S is the Seebeck coefficient, σ is the electrical conductivity, and k is the thermal conductivity. Currently, compounds such as Bi_2Te_3 offer the best ZT ($ZT\approx1$) at room temperature, which leads to coefficient of performance of refrigerators of about unity. If ZT could be increased to 3, the performance of thermoelectric refrigerators or power generators could be comparable to their gas-based counterparts. Although it is possible to control S and σ by doping a semiconductor, significant reduction of k has not been possible in the past. Recent studies have shown that by nanostructuring a solid, it would be possible to simultaneously increase $S^2\sigma$ and reduce k and thereby increase its ZT value to 3 or more [9]. This, by itself, is a compelling reason to study heat flow in micro/nanostructured solids.

The paper is divided into two main sections, one on thermometry at short length and time scales and the other on thermal transport in low-dimensional structures.

2 Thermometry

During the last decade, remarkable progress has been made in the area of microscale thermometry, which can now allow temperature measurements at length and time scales comparable to the mean free paths and relaxation times of energy carriers in solids (see Table 1). This section provides a critical review of three thermometry techniques, namely: (i) scanning thermal microscopy for high spatial resolution thermal imaging; (ii) picosecond reflectance thermometry for measuring thermal properties of interfaces and nanostructures; and (iii) scanning optical thermometry that combines high spatial and temporal resolution measurements.

2.1 Scanning Thermal Microscopy. A scanning thermal microscope (SThM) operates by bringing a sharp temperature-sensing tip in close proximity to a sample solid surface (see Fig. 4). Localized heat transfer between the tip and sample surface changes the tip temperature. By scanning the tip across the sample surface, a spatial distribution of the tip-sample heat transfer is mapped out. When the tip comes in local equilibrium with the sample, one obtains the spatial temperature distribution of sample surface, whereas if the temperature change in determined for a known heat flux, one could obtain the local thermal properties. The spatial resolution depends mainly on three factors, namely, tip sharpness, tip-sample heat transfer mechanism, and the thermal design of the probe. In the recent past, the spatial resolution has been improved to 30–50 nm, allowing one to study thermal transport phenomena at these length scales.

The idea of SThM was first proposed by Williams and Wickramasinghe [10] when they used a thermocouple at the end of a metallic wire to measure temperature. Tip-sample heat transfer was used to image the surface topography of electrically insulating materials. Since then, significant progress has

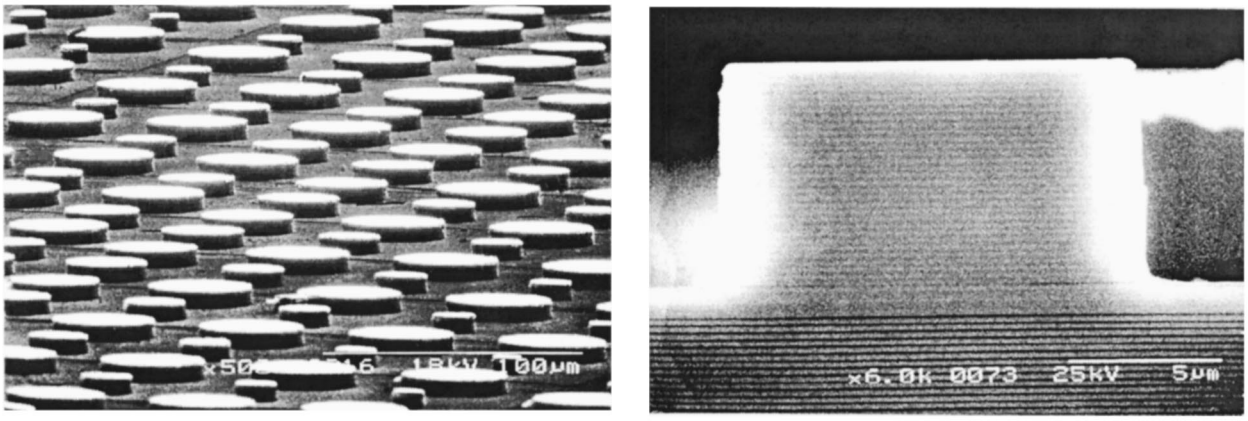


Fig. 2 Electron micrographs of (left) an array of vertical cavity surface emitting lasers (VCSEL) on a single wafer and (right) the cross of an individual VCSEL. Each laser contains multiple layers of III–V semiconductor materials which are approximately 60–100 nm in the Bragg mirrors of the laser or 1–10 nm in the active quantum well region.

been made in improving thermocouple-based measurements [11,12,13,14,15,16,17,18,19], while other techniques based on contact potential [20], electrical resistance [21,22,23,24] and thermal expansion [25,26,27] have also been developed. Although the first study utilized a wire probe [10,11], most of the subsequent development has been towards cantilever-based probes such that atomic force microscopes (AFMs) [28] could be used as a platform for SThM. Figure 4 shows a schematic diagram of an AFM, which utilizes a sharp tip mounted on a cantilever to be scanned across a sample surface. Cantilever deflections are measured by reflecting a laser beam off the cantilever and onto a position-sensitive photodiode. The deflection signal is used in a control system to maintain a constant tip-sample contact force while the tip is scanned laterally. When a temperature sensor is mounted on the very apex of the tip, such probes can be used to image both the topography and the temperature distribution of devices such as single transistors [12,13,29] and vertical cavity lasers [30]. Such studies have helped in identifying defects and failure mechanisms while providing insight about electron, photon and phonon transport in these devices. In addition, SThM has also been used to measure thermal properties of materials and perform calorimetry at nanometer scales [14,16,20,22,23,24,31]. Progress in this field till 1999 has been documented in a comprehensive review [32]. Since then, however, there have been some significant breakthroughs in both probe design and fabrication, and heat transfer studies at nanoscales [33].

The thermal design of the cantilever probes is extremely important for SThM performance. The thermal resistance network in Fig. 4 suggests that for a given ambient temperature, T_a, and sample temperature T_s, the tip temperature T_t, can be written as $T_t = T_s + (T_a - T_s)/(1+\phi)$, where $\phi = (R_c + R_t)/R_{ts}$ is the ratio of the cantilever thermal resistance, R_c, plus the tip resistance, R_t, with respect to the tip-sample thermal resistance, R_{ts}. The cantilever and tip resistance arises from both conduction through the tip and cantilever material as well as the surrounding air, while the tip-sample resistance results from liquid, air and solid-solid conduction. Changes in sample temperature can be related to changes in the tip temperature as $\Delta T_t/\Delta T_s = \phi/(1+\phi)$. This relation suggests that the accuracy and sensitivity of sample temperature measurement by the tip depend on ϕ, which must be large for better SThM performance. The spatial resolution, Δx, of SThM measurements can be expressed as $\Delta x = \Delta T_n/(dT_t/dx)$ where ΔT_n is the noise in the temperature measurement and dT_t/dx is the mea-

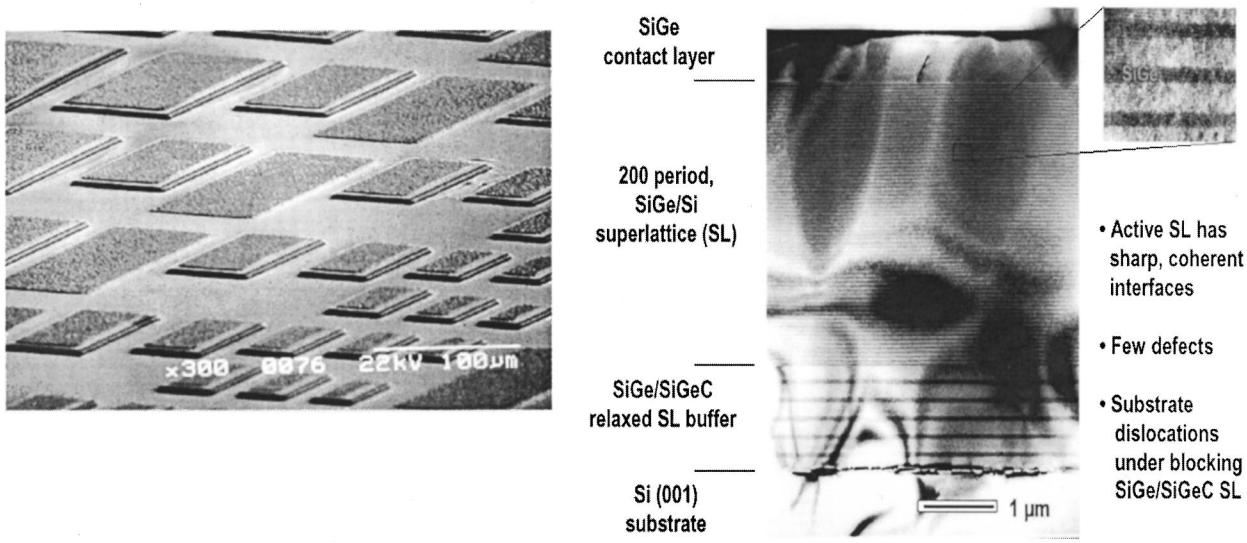


Fig. 3 Electron micrographs of (left) an array of SiGe superlattice thermionic microcoolers on a single wafer and (right) the cross section of a single device showing the SiGe/Si superlattice structure. The superlattice period is approximately 10 nm.

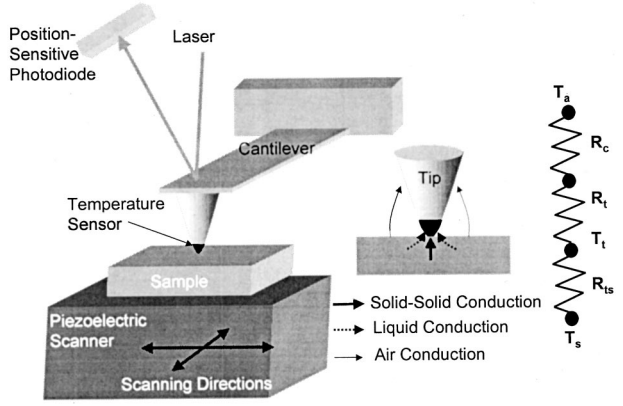


Fig. 4 Schematic diagram of a scanning thermal microscope (SThM). It consists of a sharp temperature sensing tip mounted on a cantilever probe. The sample is scanned in the lateral directions while the canitilever deflections are monitored using a laser beam-deflection technique. Topographical and thermal images can be thermally obtained. The thermal transport at the tip-sample contacts consists of air, liquid and solid-solid conduction pathways. A simple thermal resistance network model of the sample and probe combination shows that when the sample is at temperature T_s, the tip temperature T_t depends on the values of the thermal resistances of the tip-sample contact, R_{ts}, the tip, R_t, and the cantilever probe, R_c.

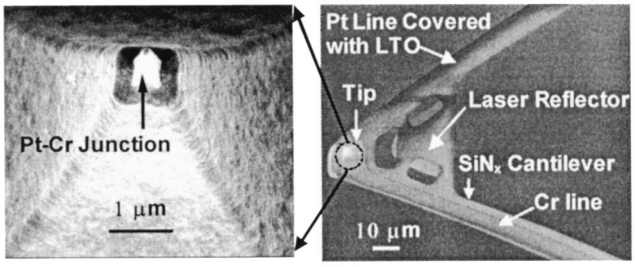


Fig. 5 Electron micrographs of a two-armed cantilever probe with a pyramid-shaped tip at the free end. The tip contains a Pt-Cr thermocouple junction at its apex which is approximately 500 nm in lateral size.

sured temperature gradient. Because the tip and sample temperatures are related through ϕ, the spatial resolution can be expressed as

$$\Delta x = \frac{\Delta T_n}{(dT_s/dx)}\left(\frac{1+\phi}{\phi}\right). \tag{1}$$

Equation (1) clearly suggests that small values of ϕ lead to poor spatial resolution of SThM. While the contributions of both air and solid conduction to R_c and R_t depends on probe geometry, that for solid conduction depends critically on the thermal conductivity of tip and cantilever material. Shi et al. [34] have performed elaborate modeling of air and solid conduction between the sample and the cantilever probe, including microscale effects near the tip-sample contact, to study the effects of material properties and probe geometry on R_c and R_t. This has provided a framework for designing SThM probes.

One of the most challenging problems in SThM is probe fabrication. Initial studies had focused on etched wire probes [10,11,12,13,14], although the high thermal conductivity of metal wires as well as irreproducibility and the large sensor size led to inaccuracies and lack of resolution. This led to development of thin film probes that could be microfabricated. However, to achieve the nanoscale sensor size at the apex of the tip, nontraditional lithography techniques had to be used, which required the probes to be made one at a time [15,17]. Subsequently, batch fabricated probes were developed, first based on e-beam lithography [19] and later using standard photolithography-based MEMS fabrication techniques [34]. Details of the design process and fabrication process can be found in Shi et al. [34]. Figure 5 shows electron micrographs of a probe [34] containing two cantilever arms a Pt-Cr thermocouple junction at the apex of the pyramid-shaped tip. The tip radius is generally about 50–100 nm while the thermocouple size can be controlled between 100–500 nm in height.

A fundamental understanding of tip-sample heat transfer mechanisms is crucial for quantitatively probing nanostructures using SThM. Although researchers initially thought heat transfer to be dominated by solid-solid and air conduction, the phenomenon turned out to be more complex. One of the surprising discoveries was that of the role of liquid film conduction. A solid surface is generally covered with a very thin layer (a few monolayers thick) of a liquid film under ambient conditions: water for hydrophilic surfaces and some hydrocarbon for hydrophobic ones. When the tip comes in contact with the sample, the liquid film forms a meniscus and bridges the tip and the sample. It was first proposed by Luo et al. [17] that this liquid film dominated tip-sample heat transfer for nanoscale point contacts, and suggested a heat transfer coefficient $h \approx 10^9$ W/m^2-K. The actual conductance depends on the radius of curvature of the tip apex. More recently, Shi et al. [35] have shown using the probes in Fig. 5, that while liquid film does play an important role, solid-solid and air conduction pathways are equally important. Figure 6 plots the cantilever deflection and the tip temperature as a function of tip-sample relative position. The measurements show that before the tip is brought into contact, air conduction dominates heat transfer. As the tip makes contact with the sample surface, liquid film conduction results in a sudden jump in tip temperature due to a marked increase in thermal conductance. Subsequently, the tip temperature rises gradually as the contact force is increased indicating the role of solid-solid conduction. Beyond a certain force, the tip temperature saturates to a constant value, which arises from the fact that the tip-sample contact area becomes constant. From this data, values of liquid, solid-solid and air conductances can be determined [35]. Although this is a typical thermal behavior of SThM probes, the exact values of forces and thermal conductances at different transition points depends on the probe design. The thermal time constant of probes shown in Fig. 5 during contact with a sample was shown to be on the order of 10 μs [34], which can allow a measurement bandwidth of about 100 kHz [36].

Figure 7 shows topographical and thermal images of the cross-section of a vertical cavity surface emitting laser (VCSEL) [30]. Because the refractive index of the III–V semiconducting materials used in VCSELs is temperature dependent, the output wavelength of VCSELs varies with temperature at about 0.1 nm/K. Hence, not only is it necessary to actively control the temperature of these devices [7], fundamental understanding of thermal phenomena inside the devices is also important. SThM images of a VCSEL showed that the maximum temperature occurred at the quantum well region where electrons and holes recombine radiatively. This suggested non-radiative recombination due to the presence of defects and crystal imperfections to be the dominant loss mechanism as opposed to other mechanisms such as Joule heating and radiative absorption. The temperature distribution also suggested a thermal lens formation due to the temperature dependence of the refractive index of these materials.

Carbon nanotubes have recently been intensively studied for their electrical, thermal, mechanical, and chemical properties [37]. They come in two kinds: multiwall carbon nanotubes (MWCNs), which are generally about 10 nm in diameter, and single wall carbon nanotubes (SWCNs) which are about 1–2 nm in diameter.

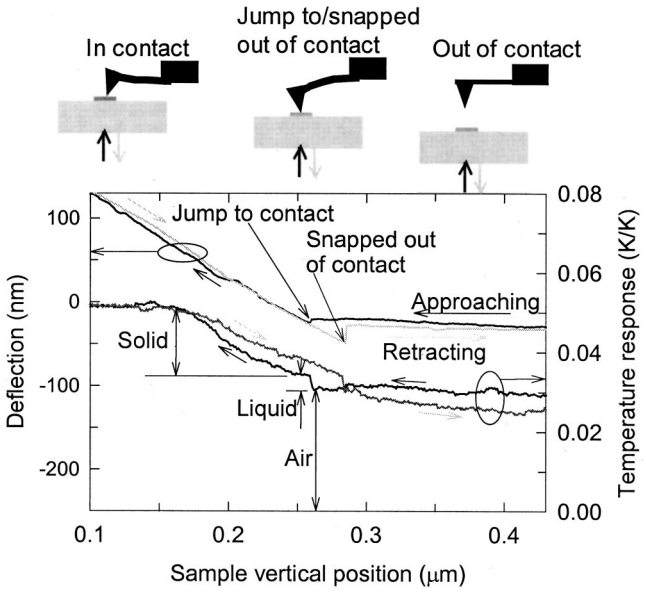


Fig. 6 Cantilever deflection and temperature response of the probe as a function of sample vertical position. The sample was a 350 nm wide electrically heated metal line. When the sample approaches the tip, the cantilever deflection remains unchanged till a jump to contact. During this time, the tip temperature rises very gradually due to air conduction. Corresponding to jump to contact the tip temperature suddenly rises due to liquid conduction. As the sample is pushed up the cantilever deflects linearly, while the tip temperature increases linearly due to increase in solid-solid contact area to a certain constant value. When the sample is retracted away, the solid-solid conductance decreases, but at a different rate, indicating some hysteresis due to plastic deformation of the tip or the sample. The tip remains in contact for a longer duration due to surface tension forces. When the tip snaps out of contact from the sample, there is a sudden drop in temperature due to loss of liquid film conduction.

MWCNs are generally metallic in nature, while SWCNs can be both metallic and semiconducting depending on the chirality in their crystal structure. Although there have been several studies on electrical transport in MWCNs and SWCNs there are several unanswered questions regarding dissipation in these nanostructures. While current flow in MWCNs is expected to be diffusive leading to uniform Joule heating, there is speculation that electron transport in SWCNs is ballistic over several micrometers in length even at room temperatures. Figure 8(*a*) shows topographical and thermal images of an electrically heated MWCN of 10 nm diameter [33,38]. The spatial resolution of SThM was about 50 nm. Detailed modeling of the measured temperature combined with an understanding of tip-sample heat transfer mechanism was used to quantitatively determine the actual temperature rise of the MWCN [33]. The temperature profile along the length of the MWCN was parabolic, indicating a uniform Joule heating and diffusive electron transport. Figure 8(*b*) shows the topographical and thermal images of a SWCN that was about 1 nm in diameter. Thermal images at low voltages showed that most of the temperature rise was near the contacts, indicating ballistic electron flow in the SWCN. At high voltages, however, bulk dissipation started to occur as indicated by the thermal images. This is presumably because of optical phonon emission by electrons when they get highly energized by transport in high electric fields [39].

There have been several important developments and breakthroughs in SThM over the last decade, which are now allowing us to routinely make thermal measurements at spatial resolutions in the sub-100 nm regime. There are three main areas where progress could be made in the future, namely: (i) fundamental understanding of tip-sample heat transfer and accuracy of temperature measurements; (ii) design and fabrication of integrated multifunctional probes; (iii) thermal probing of new devices and micro/nanostructures. Although we have made progress in our understanding of heat transfer at point contacts [17,35,40], there are still many fundamental issues unresolved. For example, can the effect of liquid conduction be controlled by tailoring surface chemistry? What is the role of near-field radiation in tip-sample heat transfer [41,42,43], especially in vacuum environment? How do phonons and electrons propagate through a point contact that is on the order of their wavelength? Can we conduct SThM experiments at low temperature and observe quantum transport effects? Because tip-sample thermal resistance depends on surface topography, topography-related artifacts in thermal images still remains a problem. There are two potential solutions: a thorough knowledge of tip-sample heat transfer in order to model the effect of topography; or a null-point measurement such that tip-sample temperature difference can be mini mized thereby eliminating any heat flow. The latter requires more sophisticated probes with multiple sensors and feedback control.

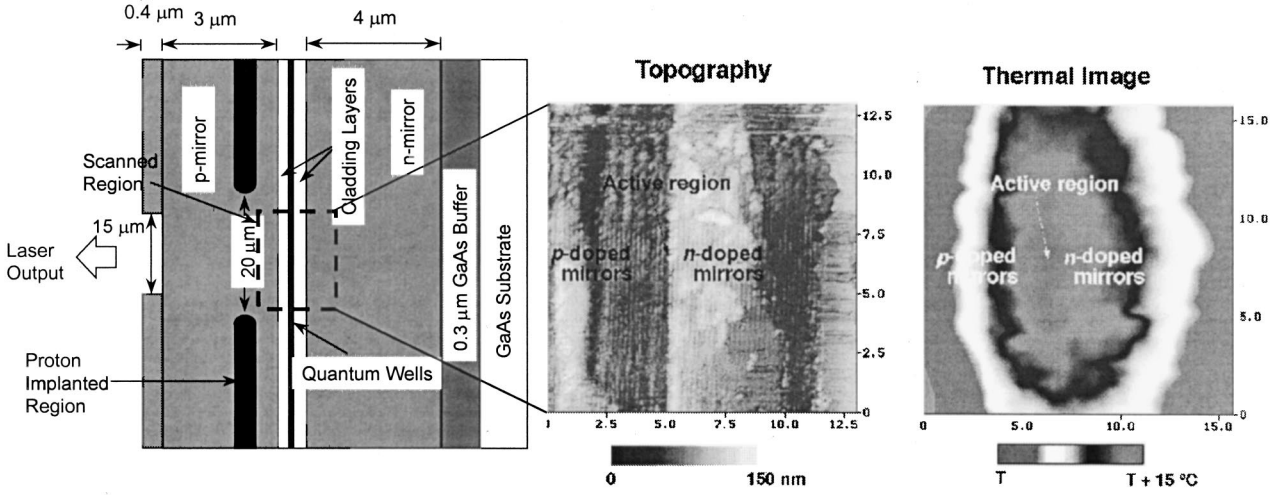


Fig. 7 (left) Schematic diagram, (middle) topographical image, and (right) thermal image of the cross-section of a vertical cavity laser. The region imaged is approximately 6 μm on either side of the active quantum well region, which includes the p-doped and n-doped Bragg mirrors consisting of alternating layers of AlAs and AlGaAs. The thermal image was obtained after a current was applied and the device was lasing.

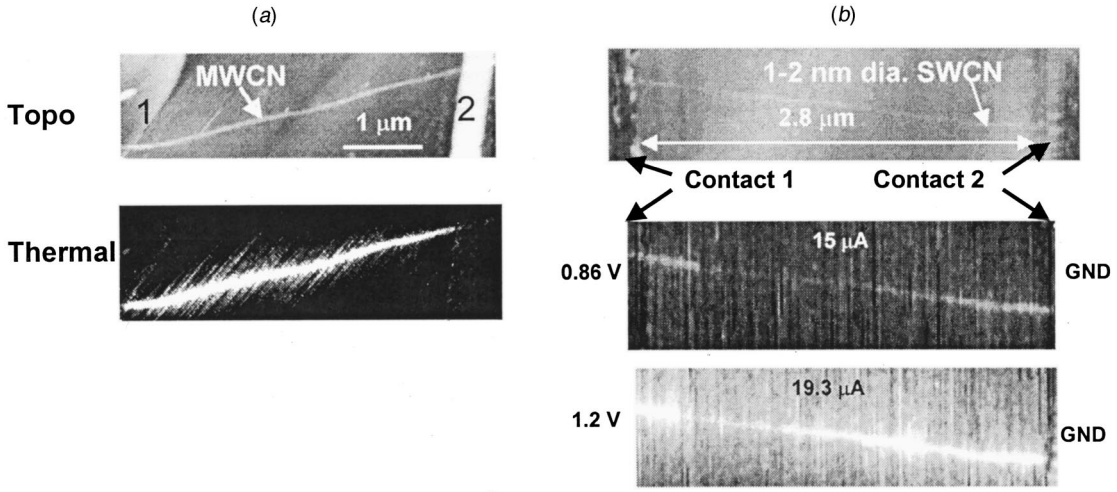


Fig. 8 Topographical and thermal images of electrically heated: (*a*) multiwall carbon nanotube (MWCN) about 10 nm in diameter with 22.4 μA flowing through it; (*b*) single wall carbon nanotube (SWCN) 1-2 nm in diameter with 15 μA (top) and 19.3 μA flowing through it. The temperature profile of the MWCN is parabolic in nature indicating a diffusive flow of electrons. At low bias, the SWCN is heated only at the contacts suggesting ballistic flow, while at higher bias the SWCN is heated in the bulk, most likely due to optical phonon emission. All measurements were made at room temperature.

The development of batch fabrication processes for making SThM probes [19,34,44,45] has made them readily available to many researchers, which itself is a significant engineering contribution. Currently, however, one can place only a single temperature sensor on SThM tips. Future progress will be driven towards multiple sensors and actuators on integrated cantilever probes such that one could simultaneously measure temperature and heat flux as well as electrical, optical and other nanoscale properties while imaging a nanostructure. Finally, because new nanoscale materials and energy conversion devices are being developed, there will also remain a need to thermal probe them at nanoscales. For example, thermoelectric properties of nanostructured semiconductors are markedly different from their bulk properties [46]. By simultaneously measuring thermal, electrical, and thermoelectric properties using a multifunctional SThM, one could characterize low-dimensional nanostructures. An area of research that is largely unexplored is SThM under liquid environment, in particular, of biological molecules. Proteins and DNA undergo structural changes and phase transitions at different temperatures. Localized calorimetry of molecules could shed light on the binding behavior of biomolecules.

2.2 Picosecond Reflectance Thermometry. Reflectance thermometry can be extended to picosecond time-scales using mode-locked lasers; mode-locked laser typically produce a series of <1 ps pules at a repetition rate of $\approx$76 MHz. Extremely high time resolution is achieved by splitting the laser output into two beam paths, a ''pump'' beam and a ''probe'' beam, and adjusting the relative optical path lengths with a mechanical delay stage. If the sample under investigation is metallic, or can be coated with a thin metal film, then a small fraction of the energy from each pulse in the pump beam produces a sudden jump in temperature near the surface of the sample. The decay of the near-surface temperature is then interrogated by the reflected energy of the series of pulses in the probe beam. Typically, the intensity of the pump beam is modulated and variations in the intensity of the reflected probe beam at the modulation frequency are measured with an rf lock-in-amplifier.

In 1986, Paddock and Eesley [47] described the first thermal transport experiments using picosecond thermoreflectance; their primary motivation—providing a direct method for measuring heat diffusion on nanometer length scales—remains valid today. Most engineering materials have thermal diffusivities D in the range $0.005<D<1\ \mathrm{cm^2\ s^{-1}}$ and at $t=200$ ps, heat diffusion lengths $\ell=\sqrt{(Dt)}$ are in the range $10<1<140$ nm. Therefore, picosecond time resolution offers nanometer-scale depth resolution and, more specifically, picosecond thermoreflectance can isolate the effects of interface conductance from the thermal conductivity of a thin layer. By contrast, thermoreflectance using Q-switched lasers [48] and the 3ω method [49] cannot distinguish between the thermal conductivity of a film and the thermal conductance of its interfaces. Greater understanding of energy transport on picosecond time-scales could also enhance capabilities for thin-film metrology. Picosecond acoustics using pump-probe optical techniques is a powerful tool for determining film thickness; in some cases, picosecond thermoreflectance could provide complementary data on film and interface microstructure [50].

Despite the promise of these methods, only a few quantitative and systematic studies of thermal properties of materials have used picosecond optical techniques [47,51,52,53,54,55]. Experiments by Maris and co-workers on interface thermal conductance [52] and superlattice conductivity [54] are the most complete. Data for the thermal decay at short times $t<50$ ps are difficult to interpret quantitatively because (i) hot-electrons can deposit energy outside of the optical absorption depth, (ii) the temperature dependence of the complex index of refraction $d\tilde{n}/dT$ is typically unknown, (iii) the diffusion equation fails on length scales comparable to the mean-free-paths of the dominant energy carriers, and (iv) the assumption of equilibrium between phonons and electrons is not always valid. At long times $t>500$ ps, measurements are often plagued by shifts in the focal plane and beam overlap created by large displacements of the delay stage. (Capinski and Maris describe a method for improving accuracy at long delay times [56].) We are optimistic, however, that improved performance and dropping costs of solid-state mode-locked lasers will expand applications of picosecond thermoreflectance. Reliable, low-noise Ti:sapphire lasers that produce 100 fs pulses are commercially available and can be operated with only a modest knowledge of the technology of ultra-fast lasers.

An example apparatus in use at the Laser Facility of the Frederick Seitz Materials Research Laboratory is shown in Fig. 9. The optical design contains some unique features: a single microscope objective focuses the pump and probe beams, collimates the reflected probe beam, and forms a dark-field microscopy image of the sample surface on a CCD camera. The integrated dark-field optical microscope greatly facilitates focusing and alignment of

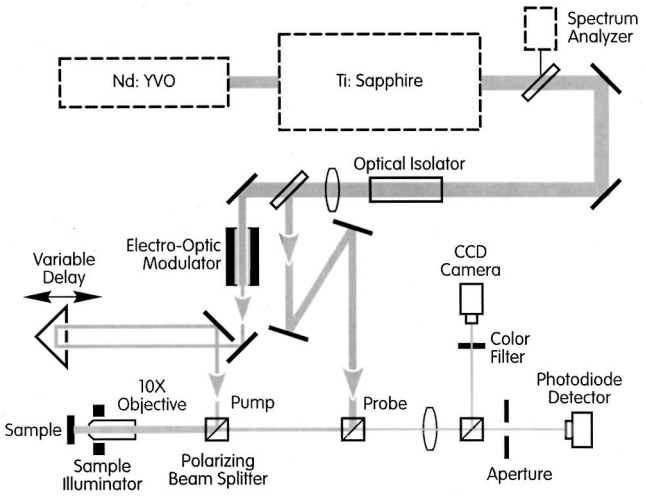


Fig. 9 Picosecond thermoreflectance and acoustics apparatus at the University of Illinois using a single objective and an integrated dark-field microscope. The pump and probe beams are parallel at the back-focal-plane of the objective lens but offset by ≈ 4 mm. The aperture in front of the detector rejects the small fraction of pump beam that leaks through the polarizing beam splitter.

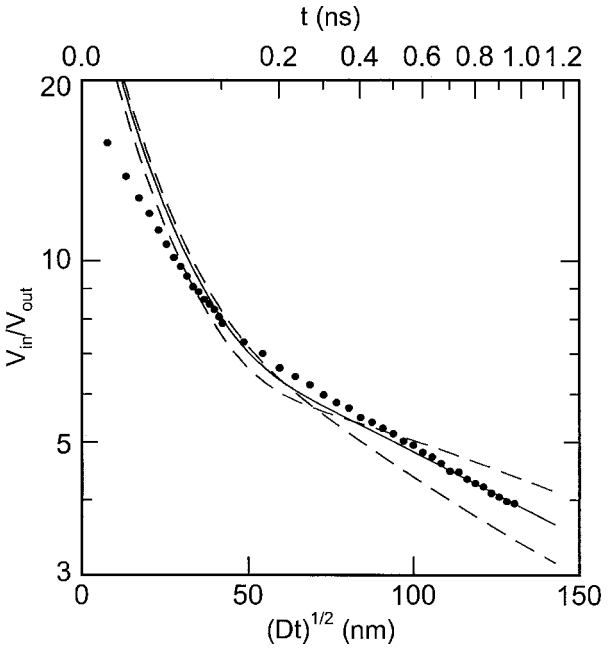


Fig. 10 Picosecond thermoreflectance data (solid circles) for thermal transport through an epitaxial TiN/MgO(001) structure at room temperature. The ratio of the in-phase to out-of-phase signals of the lock-in amplifier at $f=9.8$ MHz is plotted as function of the thermal diffusion length in TiN; the delay time t is given on the top axis. The TiN layer thickness is 100 nm. We assume that the thermal conductivity of the TiN film is given by the Wiedemann-Franz law, $\Lambda = LT/\rho = 58$ W m^{-1} K^{-1}; the thermal conductivity of MgO is 51 W m^{-1} K^{-1}. The solid and dashed lines are fits to the data using Eq. 3 with one free parameter, the interface thermal conductance G. The solid line is the best fit, $G=0.42$ GW m^{-2} K^{-1}. The upper and lower dashed lines are for $G=0.2$ and $G=1.0$ GW m^{-2} K^{-1}, respectively.

the pump and probe beams. The $1/e^2$ beam diameter at the sample is relatively large, 15 μm for both pump and probe.

In a modulated pump-probe experiment, the differences in reflected probe intensity caused by the pump pulse appear at the modulation frequency of the pump beam and are extracted with lock-in detection. But interpretation of this lock-in signal is not completely straightforward, particularly as the time delay becomes a significant fraction of the separation between pulses. Following Ref. [56], the lock-in response V_{LI} of a modulated pump-probe reflectivity measurement is

$$V_{LI}(t) = \frac{\alpha}{\tau} \sum_{q=-\infty}^{\infty} \exp(i2\pi f(q\tau + t)) \Delta R_1(q\tau + t), \qquad (2)$$

where the real part of V_{LI} is the in-phase signal of the lock-in amplifier V_{in} and the imaginary part of V_{LI} is the out-of-phase signal V_{out}; α is a constant, τ the time between pulses (typically $\approx$13 ns), f the modulation frequency, t the time delay of the probe beam, and $\Delta R_1(t)$ is the reflectivity change at a time t following an excitation of the sample by a single pump pulse. In Ref. [56], the authors state that Eq. 2 is valid only in the limit $f\tau \ll 1$ but we have found that Eq. 1 is an exact solution for all values of $f\tau$ as long as the lock-in time constant τ_{LI} satisfies $f\tau_{LI} \gg 1$. Typically, comparisons of experimental data and the prediction of Eq. 2 are used to optimize free parameters in a thermal model; for all but the shortest times, the temperature is homogeneous through the optical absorption depth of the thin metal film and $\Delta R_1(t) = (dR/dT)\Delta T_1(t)$. The surface temperature $\Delta T_1(t)$ can be accurately calculated using one-dimensional heat flow since the thermal diffusion length is usually much smaller than the radius of the focused pump beam [56].

Although Bonello and co-workers [57] discuss the in-phase signal at $t<0$, to the best of knowledge, no prior work has made use $V_{\text{out}}(t)$. But the extra information in the out-of-phase signal may provide a simple method of correcting for non-idealities in the experiment: both the in-phase and out-of-phase signals will be changed by the same amount by defocusing of the pump beam and changes in the pump-probe overlap and so the ratio of $V_{\text{in}}/V_{\text{out}}$ is often a more robust quantity than V_{in} alone. $V_{\text{out}}(t)$ varies weakly with t but decreases with modulation frequency f; we choose $f \approx 10$ MHz to minimize radial heat flow without sacrificing too much signal in V_{out}.

A frequency domain version of Eq. 2 is sometimes more convenient: instead of starting with the time domain response to a pump pulse $\Delta T_1(t)$, we calculate the frequency domain response $\Delta T(\nu)$ of a multilayer sample using matrix methods [58]. With the probe beam delayed by a time t, the ratio of the two lock-in signals is then

$$\frac{V_{in}(t)}{iV_{out}(t)} = \frac{\sum_{q=-\infty}^{\infty} (\Delta T(q/\tau + f) + \Delta T(q/\tau - f)) \exp(i2\pi tq/\tau)}{\sum_{q=-\infty}^{\infty} (\Delta T(q/\tau + f) - \Delta T(q/\tau - f)) \exp(i2\pi tq/\tau)}. \qquad (3)$$

If the pump beam is advanced instead of delaying the probe, (see Fig. 9), Eq. 3 must be corrected by a phase factor that is equivalent to shifting the phase of the lock-in reference channel by $2\pi ft$.

Figure 10 shows an example of this approach using data for an epitaxial TiN layer deposited on MgO by reactive magnetron sputtering at 800°C. Because TiN and MgO have the same crystal structure (fcc NaCl) and almost identical lattice constants, we expect that these TiN/MgO interfaces are more perfect than typical metal/ceramic interfaces. TiN is a ''metallic ceramic'', a hard, refractory material with metallic electrical properties. Our thermal model assumes that the heating produced of the pump pulse can be approximated by an instantaneous deposition of energy uniformly distributed through the optical absorption depth of TiN, $\lambda/(4\pi k) = 16$ nm. Since we assume $\Delta R(t) = (dR/dT)\Delta T(t)$, the

fit at short times is poor; i.e., we have not accounted for gradients in $\tilde{n}$ within the optical absorption depth [59]. At long times, we obtain an excellent one-parameter fit to the data but the data are relatively insensitive to the property we are interested in, the thermal conductance of the interface G: changing G by a factor of 2 produces only a $\approx$15 percent change in V_{in}/V_{out} at $t=1$ ns.

2.3 Scanning Optical Thermometry. The optical thermometry described in section 2.2 uses brief laser pulses to excite and interrogate a surface. The radiation is generally not strongly focused, which allows one-dimensional transport modeling normal to the absorbing surface. For micro and nanoscale semiconductor devices, sensors, and actuators, the goal is to measure temperature distributions resulting from *electrical* heating with simultaneously high spatial and temporal resolution. In contrast to the scanning probe methods described in section 2.1, most optical thermometry methods directly probe the microdevice without requiring heat diffusion into a solid sensor. This eliminates the associated delay and, in principle, allows measurements in the picosecond regime as discussed in section 2.2. Pump-probe microdevice thermometry with picosecond temporal resolution could be achieved using a photodiode to convert the pump pulse into an electrical current. This electro-optic excitation approach is well established for measuring electric field distributions and the voltage step response in semiconductor devices [60,61]. However, the fastest optical thermometry of electrically-heated devices used steady-state radiation and yielded temporal resolution, limited by noise and calibration complications, near 1 ns.

There has been extensive research on microdevice optical thermometry. Most measurements use visible or near-visible radiation with far-field optics, which provide diffraction-limited spatial resolution near 1 μm [62]. Examples include micro-Raman thermal imaging, for which the ratio of the Stokes and anti-Stokes lines determined the steady-state optical phonon temperature in the semiconducting regions of quantum well lasers [63] and silicon field-effect transistors [64]. Picosecond Raman thermometry of solvent-solute mixtures at large spatial dimensions [65] could be extended to microdevices. Expansion thermometry [66,67,68] used radiation interference or dilation to probe thermal expansion and temperature fields in the near-surface region. The extraction of temperature fields from surface displacement is very difficult in the absence of a dedicated surface layer with high thermal expansion coefficient, as discussed for Joule expansion microscopy using the AFM in section 2.1. The temperature dependence of photon-induced emission provides another opportunity for thermometry, although the governing electron-photon relaxation processes can limit temporal resolution. Example studies used temperature-induced shifts in the photoluminescence spectra of semiconductors [69] or, for the case of a dedicated thin-film coating, a shift in the fluorescence efficiency [70].

Thermoreflectance imaging detects temperature changes through the optical reflectance [71] and has been applied to micromachined structures [72,73,74,75]. Precise scanning of the laser focus spot over the microdevice during periodic electrical heating yields the full-field transient temperature [76]. Figure 11(*a*,*b*) show the experimental setup and temperature distribution data across the corner of an aluminum-copper alloy interconnect [77]. The temperature distribution was captured 100 ns after the initiation of an electrical current pulse and is consistent with a coupled analysis of the temperature and potential distributions in the metal. The diffraction-limited spatial resolution is comparable with the radiation wavelength of 800 nm, which was chosen to optimize the thermoreflectance coefficient, C_{TR}. Although metals have relatively low thermoreflectance coefficients, $C_{TR} \sim 10^{-5}$ K^{-1}, they are well suited for thermoreflectance imaging because their small penetration depths isolate the volume over which radiation-surface interactions are occurring and minimizes the impact on the electrical behavior of the device. The thermoreflectance coefficients of aluminum and copper achieve maxima near 800 nm and 577 nm, respectively, owing to associated high probabilities for electron interband transitions [78,79]. Because the thermoreflectance coefficient can depend on the radiation spot size, a two-step calibration procedure on a dedicated, similar micromachined structure subjected to transient electrical heating and thermometry is helpful for obtaining the relevant value of C_{TR} [76]. Reflectance thermometry can be performed through transparent passive layers, such as silicon dioxide. However, temperature and refractive index gradients in the overlayer effectively limit the temporal resolution to the thermal diffusion time normal to that layer.

Optical thermometry below the diffraction limit remains immature. From the far field, radiation with vacuum wavelength λ can be focused to a spot whose full width at half maximum is $\lambda/(2\,\mathrm{NA})$. The numerical aperture of the scanning system is $\mathrm{NA}=n\sin\theta$, the maximum angle of incidence is θ, and n is the real part of the refractive index. This is equivalent to Sparrow's criterion, in which diffraction limits the spatial resolution in a vacuum or air ($n=1$) to $\lambda/2$. Optical thermometry below the diffraction limit is based on near-field scanning optical microscopy (NSOM), which achieved resolution approaching 10 nm for optical imaging using a probe scanned a few nanometers above the surface. One NSOM method uses tapered, metal-coated optical fibers with an aperture at the tip [80]. Near-field infrared imaging used a tapered metal micropipette to collect infrared radiation emitted by a patterned metal bridge structure [81]. The fiber aperture was near 1.5 μm, which is almost an order of magnitude less than the dominant emitted wavelength at room temperature. Another technique delivered visible light through a metal-coated fiber with aperture near 50 nm [82]. The radiation reflected into the far field was used to image temperature differences between two metal bridges. For transient measurements, data interpretation was complicated by thermal expansion of the surface and probe. These exploratory scanning probe studies showed that the low transmittance through apertured probes ($\sim 10^{-6}$) can strongly limit the sensitivity for temperature measurements.

These problems may eventually be mitigated using solid immersion lens (SIL) technology, which improves spatial resolution as in oil immersion microscopy. Figure 12(*a*) indicates that a solid lens scanning in the near field can provide spatial resolution below the wavelength in air without the dramatic loss in transmission imposed by an apertured fiber. The plotted sixth-power dependence of transmission is based on the result of Bethe for apertures [83]. The SIL increases the numerical aperture of the optical path according to the real part of the refractive index of the lens material. The SIL with visible or ultraviolet radiation has achieved values of NA between 1.0 and 2.2 for optical imaging [84], data storage [85], and photolithography [86]. In principle, SILs with tailored optical properties should be appropriate for improving the spatial resolution of the various optical thermometry techniques described already in this section.

The first fully-micromachined silicon SIL [87] targeted near and mid-infrared radiation wavelengths, for which silicon is nearly transmitting with $n\sim 3.4$. Figure 12(*b*) shows the integrated silicon lens and cantilever. Lenses were fabricated by developing photoresist pillars on the silicon film on a silicon-on-insulator (SOI) substrate. The pillars were shaped into hemispheres during exposure to acetone vapor, and the shape was transferred into the silicon film using CF_4/O_2 reactive ion etching. Using transmission through a 1 μm aperture patterned in Cr/Au on a GaP substrate, the lens produced a spot size of $\lambda/5$ and $\mathrm{NA}=2.5$ with radiation wavelength $\lambda=9.3$ μm. Subsequent research imaged thermal radiation transmission through a micropatterned interconnect structure on silicon [88]. Figure 12(*c*) shows the improved resolution of the radiance images collected during a scan across two aluminum lines of width 1 μm and separated by 2.5 μm. Figure 12(*c*) is not a thermal map of the interconnect structure, but rather an image of the thermal radiation transmitted through the microstructure from an underlying heater, which shows that local measurements of temperature gradients should be possible. A

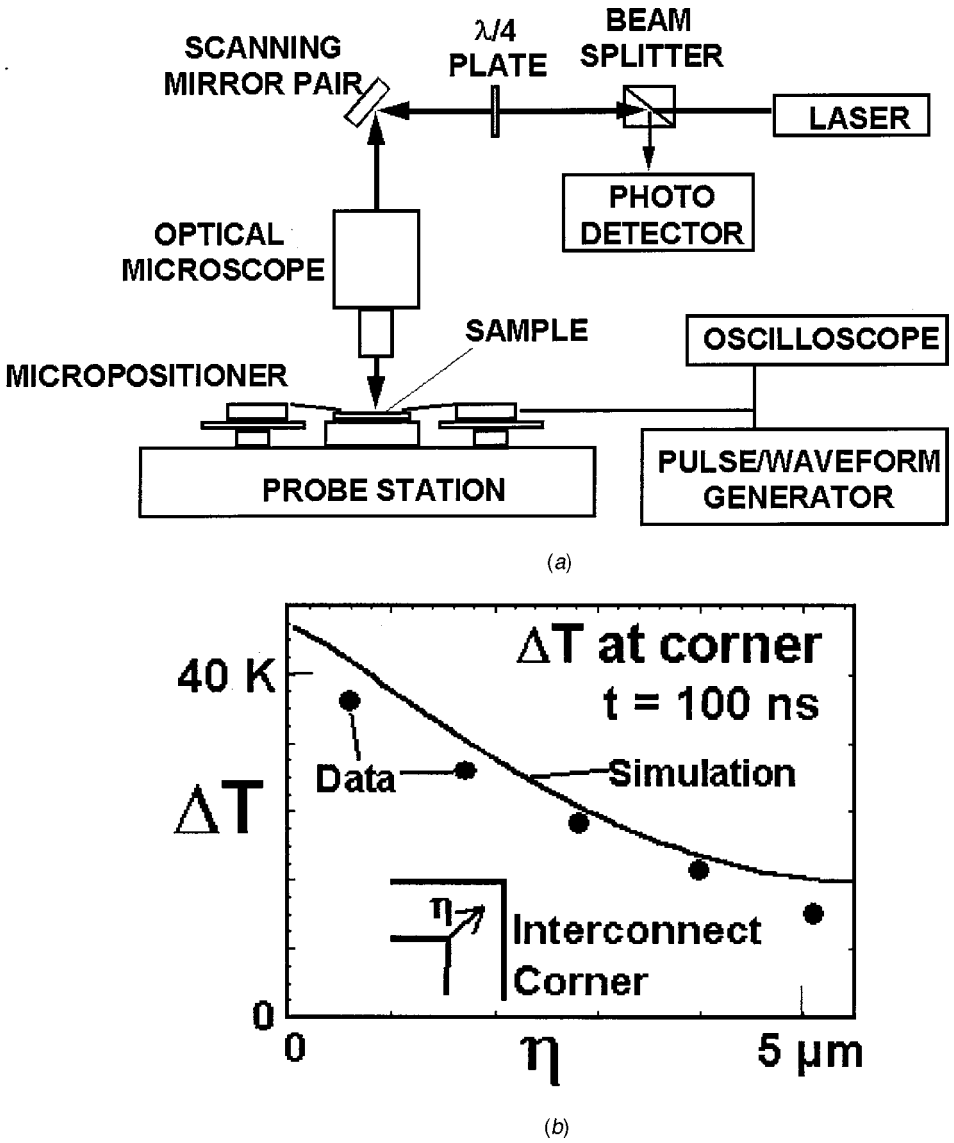


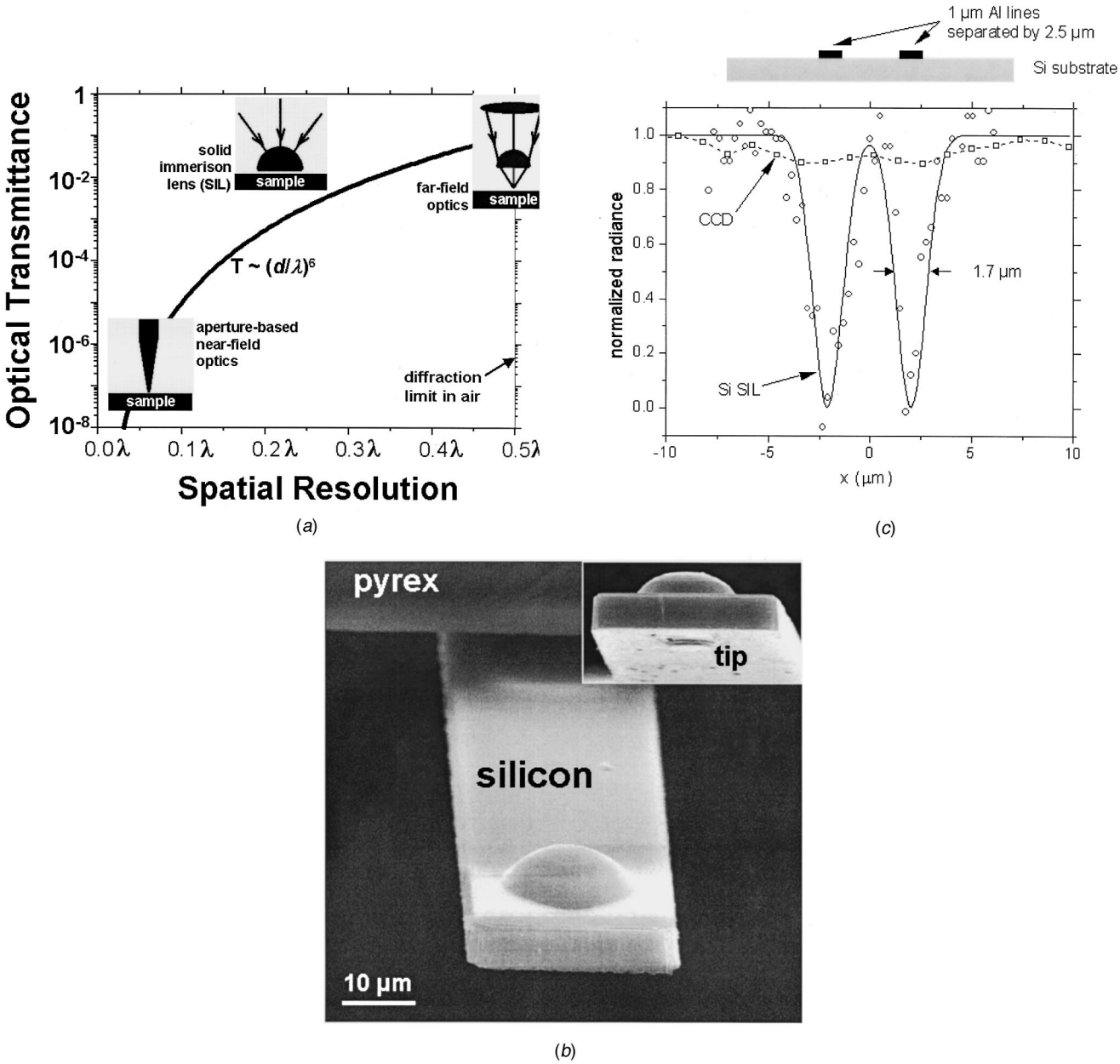
Fig. 11 (*a*) **Schematic of the setup for scanning laser-reflectance imaging of interconnects and semiconductor devices;** (*b*) **experimental data for the temperature distribution across the corner of an electrically heated metal intercorrect. For details see Refs. [62,76,77].**

noise equivalent temperature resolution below 0.2 K and a spatial resolution of 1.7 μm were predicted at 120°C using a diffraction-limited infrared microscope with InSb focal plane array detector at 20 Hz. Although the spatial resolution for optical microscopy may be improved through the use of a metal aperture patterned on the lens, the resulting reduction in optical transmittance and the noise-limited temperature sensitivity makes this approach unattractive for thermal imaging.

3 Thermal Transport in Low-Dimensional Structures

3.1 Thin Films. Thermal conduction in solid films can be strongly influenced by material imperfections and sub-continuum effects. The discrepancy between bulk and film thermal conductivities can in many cases be traced to imperfections or impurities introduced during the fabrication process. For single-crystal films of thickness comparable with the phonon mean free path, phonon-boundary scattering also reduces the effective thermal conductivity. Additional detail can be found in thin-film thermal conductivity reviews [89,90] and more focused articles on superlattices [91,92,93], interface resistance [94], diamond films [95,96], and organic films [97,98,99], and experimental methods [100].

3.1.1 Measurement Techniques. Thin film thermal conductivity measurement techniques are distinguished by the source and timescale of heating and the thermometry method. Many use electrical heating and thermometry in bridges patterned above the sample film. The 3ω method [101] sustains a current in a metal bridge at angular frequency ω, inducing fluctuations at 2ω of the heating rate and the temperature. The temperature amplitude and phase and the thermal properties of surrounding materials are extracted from the 3ω component of the voltage along the bridge. While the 3ω method first measured the out-of-plane thermal conductivity in dielectric films [102,103], recent extensions using multiple bridges of varying width measured the thermal conductivity anisotropy in superlattices [104], polyimide [105], and the in-plane conductivity of thin silicon films [106]. Measurements at heating frequencies up to 60 kHz yielded the volumetric heat capacity of silicon dioxide films [107]. Steady-state electrical heating and thermometry in parallel metal bridges is an alternative for determining film-substrate interface resistance [94] and the out-of-



(*a*) (*c*)

(*b*)

Fig. 12 (*a*) **Relationship of the transmittance and spatial resolution of optical imaging technologies, showing the promise of the solid immersion lens for thermal imaging. The curve plots the sixth power dependence of transmittance on diameter for a circular aperature.** (*b*) **Scanning electron micrograph of the silicon lens and tip and the pyrex mount.** (*c*) **Transmission data collected by an InSb CCD array through a patterned two-bridge structure with and without the SIL.**

plane thermal conductivities of disordered dielectric films [108]. Steady and transient heating in metal were used in membrane structures, generally fabricated using anisotropic chemical etching, to isolate the in-plane thermal conductivity [109,110,97].

Pulsed laser heating and thermoreflectance thermometry have been used for non-contact measurements of the out-of-plane conductivity. The instrumentation and analysis depends on the heating pulse duration. As discussed in section 2.2, measurements with picosecond-scale heating can examine transport in a variety of thin films and nanostructures. These measurements use pump-probe laser diagnostic to capture both the thermoreflectance response and the surface displacement resulting from acoustic waves. Analysis needs to consider the non-diffusive, sub-continuum nature of heat transport in the sample films and the disequilibrium between electrons, which can absorb much of the radiation, and the lattice. Nanosecond heating from a Nd:YAG laser and continuous time-domain laser-reflectance thermometry yielded the out-of-plane conductivities of silicon-dioxide [111] and polycrystalline diamond [112] films of thickness down to a few hundred nanometers. The lateral thermal conductivity of films on substrates has been measured using the transient thermal grating technique [113]. Pulsed laser radiation interferes on the sample surface, yielding a harmonic spatial variation of energy absorption. This yields spatially varying temperature and thermal strain fields, whose temporal decay is detected by the deflection of an incident probe beam. The temporal decay is governed mainly by the lateral thermal conductivity of the film and substrate and the volumetric heat capacity within a depth near the spatial period of energy deposition, which can be varied by altering the difference in incident angles between the two probes.

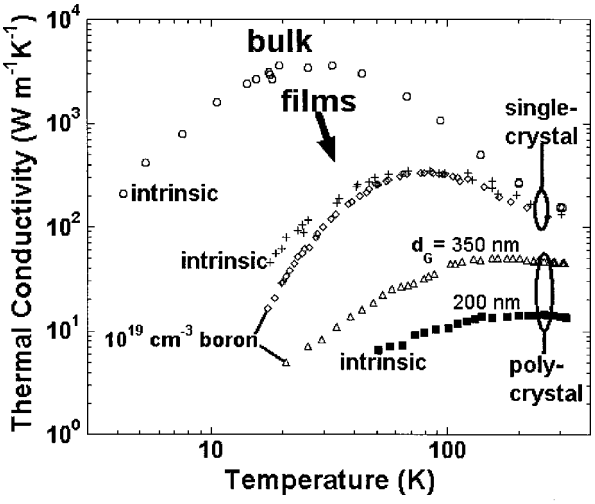
3.1.2 Single Crystal Films. The in-plane conductivities of single-crystal semiconducting monolayers are important for silicon-on-insulator (SOI) [114] and 3D circuits [115], as well as micromachined sensors and actuators involving crystalline silicon membranes [116]. Conduction is dominated by phonon transport, even for heavily-doped samples. Figure 13(*a*) shows that for nearly single-crystal silicon, phonon-interface scattering strongly reduces the effective thermal conductivity at low temperatures. The data can be predicted using the model of Holland [117] for conduction in silicon with modifications to account for scattering on film interfaces based on the phonon Boltzmann equation [118]. Crystalline monolayers are often bounded by amorphous films, which act as diffuse phonon emitter-absorbers owing to strong phonon scattering. When the mean free paths in neighboring films are comparable, which is the case in many superlattices, transport in the layers is coupled and the fraction of phonons specularly reflected must be determined from the interfacial roughness and the phonon wavelength. For single-crystal silicon layers doped at concentrations higher than 1.0×10^{17} cm^{-3} the conductivity is reduced by scattering on impurities and free electrons [119]. Figure 13(*a*) shows that this is particularly important at low temperatures. Measurements at 20 K on layers of thickness 3 μm with phosphorus and boron concentrations of 1.0×10^{18} cm^{-3} yielded reductions by approximately two and four, respectively, with the difference resulting from the disparity in mass of the two impurity types.

The anisotropy and nonlinearity of acoustic phonon dispersion relationships complicate predictions above 100 K for GaAs, Ge, and AlAs films, and above 200 K for silicon films owing to its higher Debye temperature. Figure 13(*b*) plots room-temperature in-plane thermal-conductivity data for silicon films in SOI substrates as thin as 74 nm [106]. The thermal conductivity is smaller by as much as 50 percent than the bulk value and decreases slowly with decreasing film thickness. A graybody approximation overpredicts the data because it neglects the spectral dependence of phonon scattering. Additional modeling lended support to the hypothesis that longitudinal phonons dominate transport. This hypothesis is consistent with the highly dispersive nature of high-frequency transverse acoustic phonons in silicon, which reduces their group velocities.

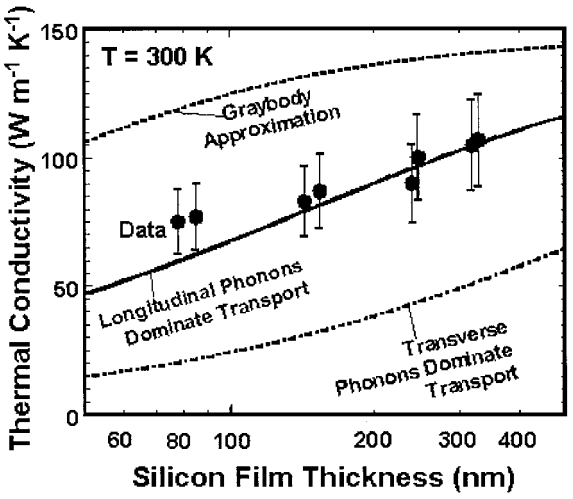
Epitaxially-grown high-temperature superconducting films, such as $YBa_2Cu_3O_7$, $EuBa_2Cu_3O_7$, and $BiSr_2Ca_1Cu_2O_8$, exhibit highly anisotropic thermal conduction properties due to their orthorhombic unit cell and the large density of oriented imperfections [120]. These films are promising for low-loss interconnects and Josephson junctions in hybrid superconductor/semiconductor circuits, as well as for thermal radiation detectors. Models have been developed for the conductivity reduction in films with the *c* axis oriented normal to the substrate [121,122]. These calculations considered the simultaneous contribution of electrons and phonons, because both carriers are significant at temperatures above a few tens of Kelvin.

The reduction is substantial for films thinner than about 100 nm and is dominated by the impact of interfaces on the phonon contribution. The electron thermal conductivity and associated thin-film size effect are strongly reduced at low temperatures due to the increasing concentration of Cooper pairs, which do not contribute to heat conduction.

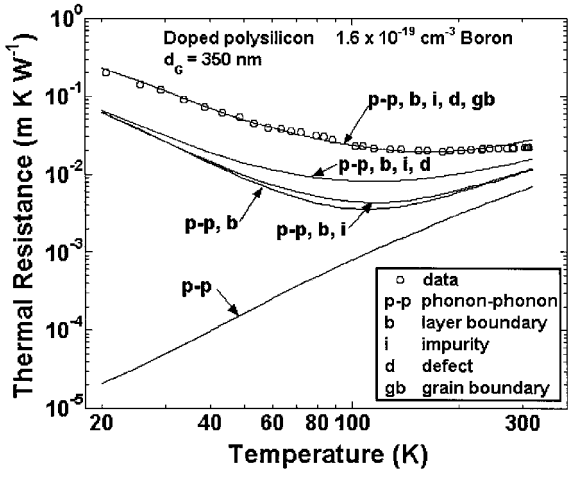
3.1.3 Interfaces. In many cases, heat transport on short length-scales and time-scales is sensitive to the properties of interfaces. Thus, an ability to control and exploit interface thermal properties could enable improved thermal management of multilayer device structures [123], simplify the fabrication of thermal sensors, and increase the efficiency of thermoelectric energy conversion [124,125]. Furthermore, because thermal phonons near room temperature are sensitive to atomic-level structure, greater scientific understanding of heat transport at interfaces may lead to convenient, nondestructive methods for evaluating the microstructure of interfaces: interface microstructure is a critical issue in materials science but the high resolution electron microscopy and diffraction needed to characterize interface structure is often difficult, time-consuming and expensive.



(*a*)



(*b*)



(*c*)

Fig. 13 (*a*) Data for bulk silicon compared with data for silicon single-crystal films and polysilicon. (*b*) Predictions and data for single-crystal silicon films of thickness down to 74 nm at room temperature as a function of film thickness. (*c*) Contributions to the thermal resistivity, 1/*k*, as a function of temperature for a doped polysilicon layer.

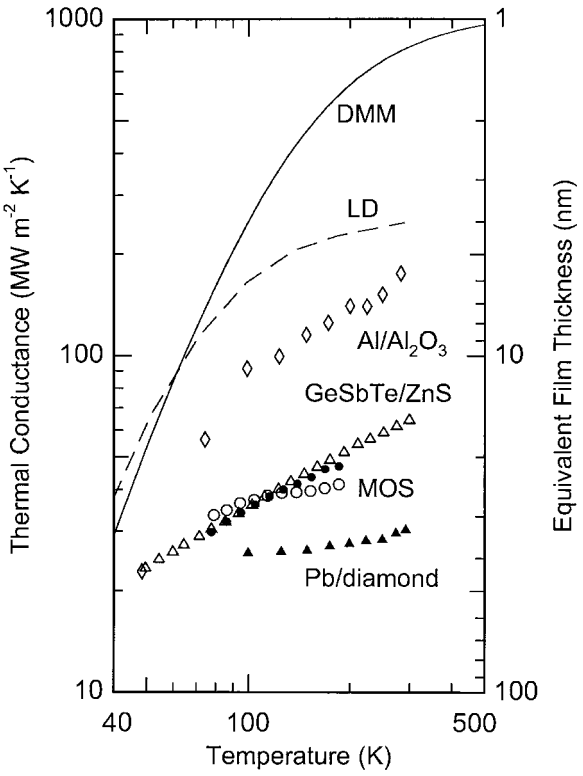


Fig. 14 Comparison of selected data for interface thermal conductance: (i) individual interfaces measured by picosecond thermoreflectance [52], Al/Al_2O_3 (open diamonds) and Pb/c-C (filled triangles); (ii) conductance of the a-$GeSbTe_{2.5}$/ZnS interface (open triangles) from a multilayer sample [133], (iii) series conductance of the top and bottom interfaces of metal-SiO_2-silicon structures (MOS, filled circles) [103]. The solid line is the calculated diffuse mismatch conductance of Al/Al_2O_3 using the Debye model; the dashed line is the theoretical prediction for Al/Al_2O_3 using a lattice-dynamical calculation of a model fcc interface [52]. The right axis gives the thickness of a film with $\Lambda = 1$ W m^{-1} K^{-1} that has the thermal conductance corresponding to the left axis.

3.1.3a Interface thermal conductance. We first consider the thermal conductance of an isolated interface, i.e., an interface that is separated from other interfaces by a distance that is large compared to the mean-free-path of the lattice vibrations that dominate heat transport in the material. In this limit, we can ignore coherent superposition of lattice waves reflected or transmitted by adjacent interfaces. The conductance of isolated interfaces has a long history in condensed matter physics beginning with the problem of the Kapitza conductance between a solid and liquid helium—an issue of great practical importance in the design of cryogenic refrigerators for low temperature physics. Swartz and Pohl [94] give an exhaustive review of thermal boundary resistance, the inverse of interface thermal conductance, through 1988.

Some recent data for interface thermal conductance are shown in Fig. 14. Surprisingly, near room temperature, the highest thermal conductance for metal/dielectric interfaces measured by picosecond thermoreflectance [52], Ti/Al_2O_3 and Al/Al_2O_3, is only a factor of ~5 larger than the lowest conductance, Pb/diamond. Other data for metal/dielectric interfaces [94,55,126,49], for the most part, fall between these two extremes. (Initial experiments on epitaxial TiN/MgO, however, indicate an interface conductance a factor of ~3 larger than Al/Al_2O_3, see Figs. 10 and 14.)

The diffuse mismatch model [94] provides a straightforward starting point for theoretical understanding of heat transport at an interface. The key assumption of this model is that phonons are randomly and elastically scattered at the interfaces with a transmission coefficient given by the relative density of vibrational states on the two sides of the interface [94,52,127]. Furthermore, if the Debye model is used for the density of states and mode dispersion, then the thermal conductance is easily calculated from the velocity of sounds and atomic densities of the two materials [94,127]. An example calculation using the diffuse mismatch model is shown in Fig. 14; this calculation for Al/Al_2O_3 overestimates the experimental data near room temperature by a factor of ≈5; if dispersion of the high energy vibrational modes is taken into account, the agreement improves [52].

The opposing limit is the assumption of no scattering at the interface; in the long-wavelength or continuum limit, phonon reflection and transmission is calculated from the mass density and anisotropic elastic constants of the materials. For shorter wavelengths, microscopic calculations of the lattice dynamics are needed; these models are typically based on idealized models of the interface structure and bonding [128,129]. The results of lattice-dynamical calculations [52] for Al/Al_2O_3, see Fig. 14, are in reasonably good agreement with the data.

While the data in Fig. 14 represent a significant advance, many more systematic studies will be needed to develop a comprehensive understanding; for example, experiments have not yet observed either the high or low extremes of conductance predicted by theory [52] and have not yet isolated contributions to heat transport at interfaces by anharmonicity, electron-phonon coupling [130], interface disorder [131] or altered interface bonding [132].

3.1.3b Multilayer thin films. Figure 14 also contains data for interface conductance extracted from studies of multilayers [133] and thin film samples [49,126] measured using the 3ω method. If the thermal conductivity of a thin layer of disordered material is independent of layer thickness d, i.e., if phonons with mean-free-paths comparable to d make a negligible contribution to heat transport, and that the microstructure of the thin layer is approximately independent of d, the measured conductance of a multilayer sample is the parallel sum of the thermal conductance of the individual layers and the thermal conductance of the interfaces [49,134,133]. Data for $GeSbTe_{2.5}$/ZnS interfaces (a component of phase-change data-storage media) nearly overlap with data for the series conductance of the metal/SiO_2 and SiO_2/Si interfaces of metal-oxide-semiconductor structures. These data for metallized SiO_2 layers on Si are in good agreement with previous room temperature measurements using nanosecond thermoreflectance [48].

While $GeSbTe_{2.5}$/ZnS multilayers show relatively small thermal conductance, our studies of multilayers of disordered oxides have failed to reveal a significant effect of interfaces: the thermal conductivity of ZrO_2:Y_2O_3/SiO_2 [135] and ZrO_2/Y_2O_3 [126] multilayers are almost independent of layer thickness with layer thickness as small as 4 nm. Apparently, the lattice vibrations of these oxides are sufficiently alike that the thermal conductance is too large to be measured by this approach. Multilayer coatings [136] do not, therefore, appear to be a promising route for lowering the conductivity of thermal barriers. Significant decreases in conductivity have been observed recently in nanocrystalline ZrO_2:Y_2O_3 [137] and attributed to the finite thermal conductance of grain boundaries.

3.1.3c Epitaxial superlattices. For a short-period superlattice with coherent interfaces, we must reevaluate our assumptions for isolated interfaces and consider superposition of lattice waves transmitted and reflected by the interfaces. Thermal transport in superlattices with perfect interfaces has been discussed in a series of papers using increasingly refined models of the lattice dynamics [138,139,140,141]. The general conclusion of this work is that a superlattice structure strongly reduces the average group velocity of phonons in the through-thickness direction and—because

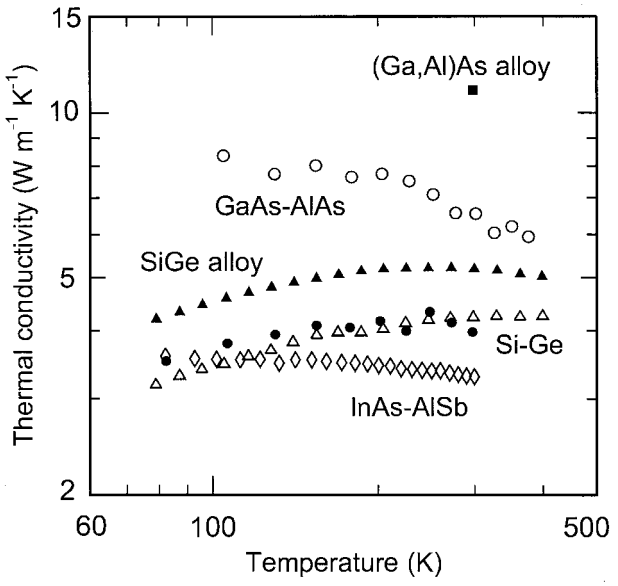


Fig. 15 Selected data for the through-thickness thermal conductivity of superlattices with bilayer periods of ≈ 5 nm; data for GaAs-AlAs (open circles, 5.67 nm period [54]) were measured by picosecond thermoreflectance; data for Si-Ge (open triangles, 5 nm period [142]; filled circles, 4.4 nm period [104]) and InAs-AlSb (open diamonds, 6.5 nm period [170]) were measured using the 3ω method. Data for $Si_{0.85}Ge_{0.15}$ [142] and $Ga_{0.4}Al_{0.6}As$ [144] alloys are included for comparison.

the thermal conductivity scales as the square of the group velocity—dramatically suppresses the thermal conductivity in comparison to the bulk components.

This suppression of thermal conductivity produced by zone-folding and phonon confinement cannot, however, fully explain the data, see Fig. 15. The observed reduction in conductivity is too large to be easily explained by the theories of perfect interfaces and both Si-Ge [142] and GaAs [54] superlattices show substantial decreases in conductivity with decreasing superlattice period; theory predicts that the conductivity should *increase* slowly with decreasing period [143].

Theory [93] based on incoherent (particle-like) scattering of phonons at interfaces predicts a superlattice thermal conductivity that decreases with superlattice period. Particle-based theory [93] can more easily explore effects of disorder but this approach requires approximations and assumptions concerning phonon dispersion, elastic vs. inelastic scattering, and specular vs. diffuse scattering that may be difficult to justify microscopically.

Interface disorder introduces diffuse scattering but disorder in a superlattice is typically modest and limited to interface roughness (a finite density of interface steps) and substitutional alloying at the interfaces caused by surface segregation during growth. Unfortunately, this physical and chemical interface roughness in an epitaxial superlattice is sensitive to the material, growth method, growth temperature, and deposition rate—and is notoriously difficult to characterize. Furthermore, asymmetry of the interface is often pronounced, i.e., the growth of component *A* on the surface of component *B* creates a different interface than the growth of *B* on *A*. Strained-layer superlattices such as Si-Ge can harbor high densities of crystalline defects when the layer thickness exceeds the critical thickness for the extension of misfit dislocations [104]; growth on relaxed buffer layers [104] reduces the density of threading dislocations but the misfit density will still be large when the critical thickness is exceeded.

Since the through-thickness conductivity of a short-period superlattice is smaller than the corresponding alloy [142,170], see Fig. 15, substitutional interface disorder acting alone is probably insufficient to explain the data. Approaches that combines both mechanisms, wave-interference and diffuse scattering [143], perhaps acting independently on different parts of the phonon spectrum, may be a more successful route to a microscopic theory. We note that emerging data on the anisotropy of superlattice thermal conductivity [145], i.e., the ratio of the in-plane to the through-thickness conductivity, may provide additional constraints on theoretical models.

3.1.4 Polycrystalline Films. In polycrystalline films, phonon scattering on grain boundaries and related defects dominates over boundary scattering. Doped polysilicon films are common in MEMS and as the gate terminal in silicon field-effect transistors. The grain structure and spatial impurity distribution depend strongly on the temperature and duration of annealing. Polysilicon conductivities are reduced compared to those of pure crystalline films [146], but the roles of impurities and grains are difficult to separate. Recent work studied films with grain sizes between 300 and 500 nm, extracted using transmission electron microscopy, and boron or phosphorus concentrations up to 4.1×10^{19} cm^{-3}, measured using secondary ion mass spectroscopy (SIMS) [147,148]. Figures 13(*a*) and 13(*c*) show that the conductivity is reduced strongly at all temperatures compared to similarly-doped single-crystal silicon layers, which illustrates the importance of grain boundary scattering. The data were predicted by reducing the mean free path using Matthiessen's rule [149] and a grain-boundary free path given by B/d_G, where d_G is the grain size and B is governed by the grain shape and grain boundary reflection coefficient. The data at 300 K are well approximated by

$$k(d_G,n)=\frac{1}{3}Cv\left(\frac{A_1}{d_G}+A_2n_i\right)^{-1},$$

where the phonon velocity and heat capacity are $v=6170$ m s^{-1} and $C=1.65\times10^6$ J m^{-3} K^{-1}, the grain size d_G has units of nm, and the impurity concentration n_i has units of cm^{-3}. The constant A_1 is 2.89×10^{10} m^2 and A_2 is 3.20×10^{-13} or -1.12×10^{-14} m^2 for boron and phosphorus, respectively.

For polycrystalline diamond films the orientation and minimum size of grains are governed by the details of the deposition process, in particular the nucleation technique, the substrate temperature, and the composition of the process gases. Verhoeven et al. [150] observed a particularly large degree of anisotropy for microwave CVD diamond films with predominantly heteroepitaxial grains, with the in-plane conductivity smaller by about one order of magnitude than the out-of-plane conductivity as shown in Fig. 16. The free path B/d_G is overly simplistic for diamond film data owing to the large concentration of other imperfections. Because imperfections populate primarily at grain boundaries, the phonon scattering rate can still be coupled to the grain size d_G. This concept was quantified using a dimensionless grain-boundary scattering strength [151], which is governed by the scattering cross sections and number density per unit grain boundary area of imperfections. For this model, the scattering rate is dictated by both the size of grains and the density of imperfections per unit grain boundary area.

3.1.5 Amorphous Oxide and Organic Films. For highly-disordered films, such as amorphous glasses and organic materials, the process-dependent material structure and stoichiometry influence the thermal conductivities and volumetric heat capacities. These films are common as passivation and thermal insulation in MEMS and integrated circuits. Silicon dioxide and silicon nitride films are often deposited at low temperatures to avoid diffusion or other failure mechanisms in micromachined structures. Lower deposition temperatures for silicon dioxide are known to reduce the mass density, which decreases the thermal conductivity [152] but increases the volumetric heat capacity owing to the large concentration of silanol [107]. Following a high-temperature anneal, the root-mean-square deviations of bond lengths, which can

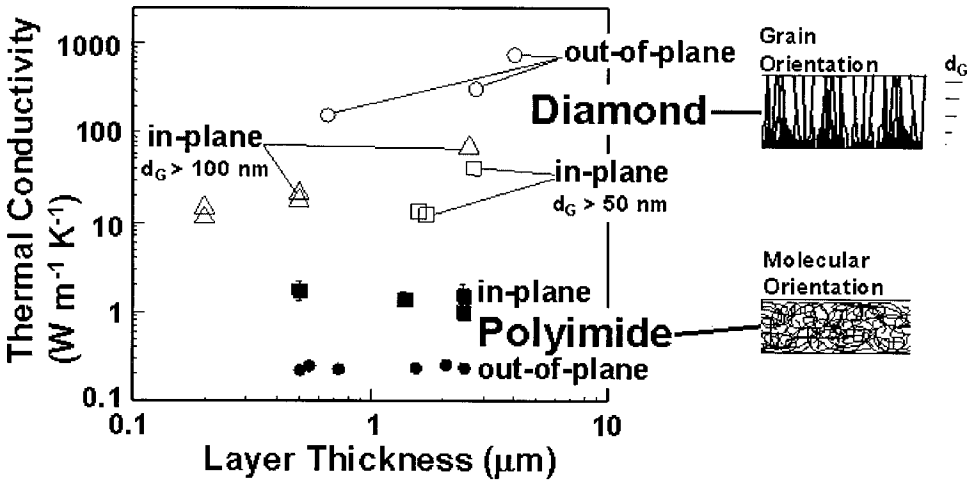


Fig. 16 The sketches indicate the impact of grain and molecular orientation on the anisotropy

serve as a measure of the degree of disorder, diminish considerably and approach those of thermally-grown silicon-dioxide films. The thermal properties of organic films are expected to be highly sensitive to their chemical composition and structural configuration. Figure 16 shows that polyimide films exhibit anisotropic conductivities due to the partial alignment of molecular strands in the film plane during spin coating. The modeling of heat conduction in disordered films is far more approximate than that of crystalline films, and is limited to estimates of the minimum conductivity [153] and simple models for the anisotropy of fully-amorphous stranded organic films [97].

3.2 Nanowires. Much of the past effort on synthesizing low-dimensional structures has focused on 0-D (quantum dots) [154] and two-dimensional (quantum wells and heterostructures) nanostructures. Except for carbon nanotubes [37], one-dimensional structures have received relatively little attention. This is mainly due to challenges in synthesis of single-crystal nanowires, which are important for determining structure-property relations. Recently, however, there have been some major advances in growth of monocrystalline semiconductor nanowires [155,156,157,158] and integrating them into devices [159,160,161]. While most of the current research is being focused on electronic and optical properties of nanowires [162], the manipulation of their thermoelectric properties through quantum confinement of electrons is extremely promising for developing solid-state energy conversion devices [46]. In this case, one must study thermal transport in these nanowires as well, which can be very different from those of bulk materials. For example, Schwab et al. [163] observed quantum thermal conductance in nanofabricated one-dimensional nanostructures, where the quantum of phonon conductance is $g_o = \pi^2 k_B^2 T/(3h)$, where k_B is the Boltzmann constant, T is the temperature, and h is the Planck constant. In this regime, a one-dimensional nanostructure behaves essentially like a phonon waveguide similar to optical ones for light. Phonon transport in nanowire can be different from that in bulk semiconductors mainly because the dispersion relation could be significantly modified due to confinement in two directions. In addition, the presence of a surface can introduce surface phonon modes. These result in many different phonon polarizations other than the two transverse and one longitudinal acoustic branches found in bulk semiconductors. Such changes in the dispersion relation can modify the group velocity and the density of states of each branch. The phonon lifetime also changes and this arises from two sources. First, the phonon-phonon interactions can change because selection rules based on energy conservation and wave-vector relations depend on the dispersion relation. Second, boundary scattering can be much stronger in nanowires (5–50 nm diameter) than in bulk semiconductors [164]. While there have been some theoretical studies on phonon conduction in nanowires [165,166], experiments have lagged behind. Recently, there have been some attempts [167] to measure the thermal conductivity of a collection or a mat of carbon nanotubes. However, due to weak coupling between carbon nanotubes, the actual thermal conductivity was found to be much lower than that predicted [168]. To accurately measure thermal properties of nanowires and compare them with theoretical studies, one must develop techniques to study the thermal behavior of isolated nanowires.

Figure 17 shows a micromachined structure that was developed by Shi [33,169]. The device is a suspended structure consisting of two adjacent silicon nitride (SiN_x) membranes or islands suspended with three 200 μm long and 2 μm wide silicon nitride beams. One 30 nm thick, 200 nm wide, and 150 μm long plati-

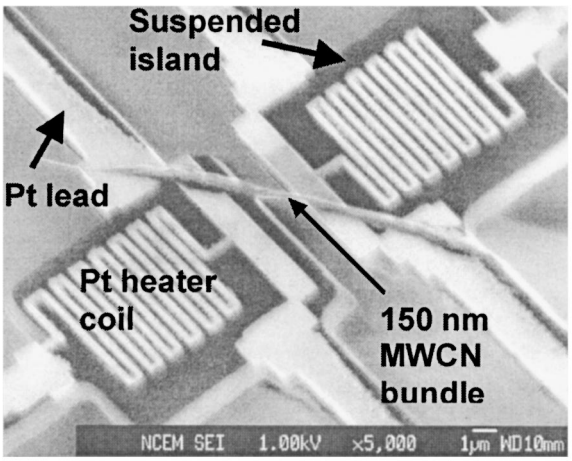


Fig. 17 Electron micrograph of the suspended heater structure used for measuring thermal conductivity of nanowires. Also laid across the heater is a multiwall carbon nanotube bundle 150 nm in diameter. The Pt heater coil on one of the suspended islands is heated and its temperature is measured by resistance thermometry. The temperature of the other island increases due to conduction through the nanowire, which is also measured by resistance thermometry. The temperature difference for a known heat flow rate from the heated island can be used to estimate the thermal conductivity of the nanowire.

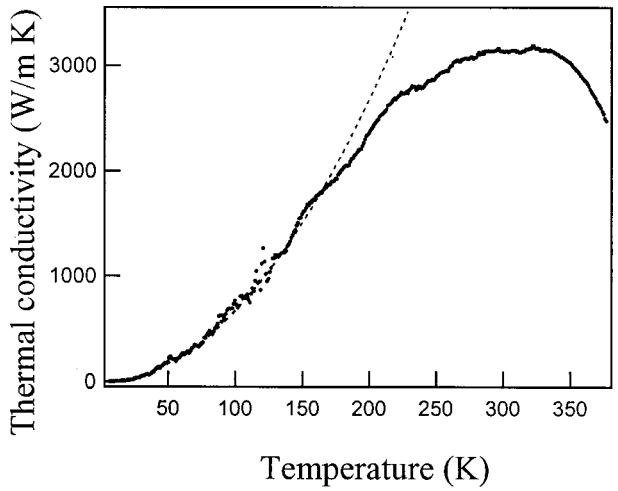


Fig. 18 Temperature dependence of thermal conductivity of a multiwall carbon nanotube. The dashed line corresponds to a T^2 temperature dependence. The thermal conductivity is about 3000 W/m-K at room temperature, which includes the contact resistances at the two islands.

num (Pt) heater/thermometer coil is built on each island. A nanowire can be placed in between and bridge the two suspended islands by several methods. A general method is to deposit a drop of aqueous solution containing nanotubes and nanowires on a wafer with densely packed suspended structures. This method has been successfully used to place SWCNs or MWCNs on the suspended structures. In the case of SWCN, nanotubes can also be grown between two islands using chemical vapor deposition (CVD). For this method, catalyst particles consisting of Fe, Al_2O_3, and Mo are first deposited on the islands from aqueous solution. Then the device is placed in a 750–900°C CVD chamber with methane gas flow. There is a good chance that one or several SWCNs can be grown and bridging the two islands.

To measure the thermal conductivity, one of the islands is heated to a temperature, T_h, which is determined by measuring the resistance of the Pt wire on the island. Conduction through the nanowire heats the sensing island to a temperature T_s, which is also measured by resistance thermometry. By noting the power dissipated by the heater, Q_h, and estimating the thermal conductance of the suspended legs connecting the islands in the absence of a nanowire, the conductance of the nanowire G_n can be calculated in a vacuum environment. Figure 18 shows the plot of thermal conductivity of a multiwall carbon nanotube as a function of temperature measured by this technique. The T^2 temperature dependence suggests that MWCN thermally behaves like a two-dimensional solid and has a thermal conductivity of about 3000 W/m-K at room temperature. This is approximately half of what is theoretically predicted [168]. Note that the measured value includes the contact resistance between the nanotube and the two islands, although it is predicted to be lower than the resistance of the tube itself [169]. Estimation of the contact resistance has remained a challenge. Nevertheless, it is envisioned that this technique can be used to measure thermal conductivities of a variety of semiconducting or metallic nanowires that are of technological importance.

4 Closing Remarks

This paper provides a critical review of two interrelated areas of research, namely, microscale thermometry and thermal transport in solid micro/nanostructures. The technology for measuring temperature and studying thermal transport is sufficiently advanced that one can now approach scales on the order of the mean free path and relaxation times of electrons and phonons at room temperature. The underlying principles, recent advances, and future challenges and research directions of scanning thermal microscopy, picosecond thermoreflectance, and scanning optical thermometry are discussed here. Experiments based on these techniques have led to fundamental understanding of electron, phonon and photon interactions and transport in both bulk and micro/nanostructured solids as well as solid-state devices. While these techniques are being developed to study thermal phenomena at microscales, it is clear that a fundamental understanding of microscale heat transfer is necessary to make further advances in these techniques. Much work remains to be done for further refining and combining these techniques. For example although one can now study short length and time scales thermal phenomena independently, it is currently very difficult to study thermal phenomena at small length and time scales simultaneously. In addition, although significant advances have occurred over the past decade, these techniques are limited to only a few laboratories. There needs to be more effort in making these techniques sufficiently robust and accessible so that the barrier for researchers to enter this field is greatly reduced.

This paper presents a critical review of thermal transport in two-dimensional and one-dimensional solid microstructures. In particular, phonon transport in single crystal, polycrystalline, and amorphous oxide and organic thin films is elucidated. The role of interfaces in multilayered polycrystalline and epitaxial thin films is also discussed. Finally, thermal transport in carbon nanotubes is presented and the opportunities for further research on nanowires are explored. For single crystal films, the effect of film thickness on thermal conductivity in materials such as Si is now fairly well understood and the thermal conductivities can be predicted using Boltzmann transport theory. For other materials, however, the non-linearity of the acoustic phonon dispersion relations makes it difficult to predict thermal conductivities at room temperature. In fact, in many materials the contribution of each phonon branch, longitudinal and transverse, to the thermal conductivity is largely unknown. For polycrystalline Si and diamond films, grain boundary scattering has been found to strongly reduce thermal conductivity. The value depends on the grain structure, which in turn depends on the growth process. Thermal properties of organic films are very sensitive to the chemical composition and structural configuration. For example, spin coating a polyimide film partially aligns the polymer chains along the film plane, which produces anisotropic thermal conductivity. Simple models have been developed to predict their thermal conductivity although detailed models based on fundamental understanding of polymer structure and chemistry remains to be developed.

Thermal transport across solid-solid interfaces remains one of the most challenging problems. While acoustic and diffuse mismatch models have been developed, their predictions rarely agree with experimental data. These models are phenomenological and do not isolate contributions to heat transport by anharmonicity, electron-phonon coupling, interfacial disorder and chemistry. Experiments have shown that interfaces play a very important role in epitaxial superlattices, where interface roughness and alloying introduces phonon scattering and thereby reduces the effective thermal conductivity. In some cases, it has been found that the thermal conductivity reduction exceeds even that offered by a collection of individual interfaces. It is conjectured that the wave interference from multiple interfaces, which produces phonon bandgaps and thereby reduction of group velocity, may play an important role. However, what is clear is that fundamental understanding of the individual and collective role of interfaces on thermal conductivity is largely missing.

While most of the paper discusses thermal transport in two-dimensional confined structures (films), it briefly discusses phonon transport in one-dimensional nanostructures, such as carbon nanotubes. It is expected that these and other nanowires are likely

to play an important role in nanostructured materials and devices, and therefore ought to be thermally studied. Confinement in an additional dimension might offer avenues of pushing the thermal conductivity to the extremes, as exemplified in carbon nanotubes. However, theoretical and experimental studies in this area have only just begun, and much remains to be understood.

Acknowledgments

DGC was supported by NSF Grant No. CTS 99-78822 and the U.S. Department of Energy, Division of Materials Sciences under Award No. DEFG02-ER9645439, through the Frederick Seitz Materials Research Laboratory at the University of Illinois at Urbana-Champaign. Picosecond thermoreflectance measurements and film thickness characterization used the Laser Facility and the Center for Microanalysis of Materials, University of Illinois. DGC thanks K. E. O'Hara (Integrated Micromachines) for invaluable assistance with the picosecond experiments and H. J. Maris, G. Chen, P. M. Norris, and G. D. Mahan for helpful discussions.

KEG appreciates the help of D. A. Fletcher and A. D. McConnell of Stanford University in preparing the sections on the SIL and polysilicon films, respectively. KEG also appreciates the many technical contributions of former students including M. Asheghi of Carnegie Mellon University, K. Kurabayashi of The University of Michigan, Y. S. Ju of IBM Corporation, M. N. Touzelbaev of AMD, and P. G. Sverdrup of Intel Corporation. Sponsorship was provided from the SRC through tasks 751 and 754.

AM greatly appreciates the support of the NSF, the DOE (Engineering Division, Basic Energy Sciences), and DARPA for their support. AM acknowledges the many contributions of several former students and post-docs, in particular, Jie Lai (Microsemi Corp.), Ke Luo (TriQuint Semiconductor), Madhavi Chandrachood (Applied Materials), Osamu Nakabeppu (Keio Univ.), John Varesi (Raytheon), and Li Shi (IBM Research Division). He also thanks Paul McEuen, Sergie Plyasunov, Adrian Bachtold and Philip Kim (Physics, UCB) for their help with nanotube related research.

Nomenclature

D = thermal diffusivity
f = modulation frequency of the probe beam
G = interface thermal conductance (per unit area)
k = imaginary part of the complex index of refraction
l = thermal diffusion length
$\tilde{n}$ = complex index of refraction
q = summation index
$\Delta R_1(t)$ = change in optical reflectivity at time t following a single pump pulse
t = time
T = temperature
$\Delta T_1(t)$ = temperature excursion at time t following a single pump pulse
$\Delta T(\nu)$ = frequency domain temperature response
V_{LI} = lock-in signal expressed as a complex number
V_{in} = in-phase lock-in signal
V_{out} = out-of-phase lock-in signal
λ = laser wavelength ($\approx$800 nm for Ti:sapphire)
ν = frequency
τ = time between mode-locked optical pulses

References

[1] Verhoeven, H., Boettger, E., Flöter, A., Reiss, H., and Zachai, R., 1997, "Thermal Resistance and Electrical Insulation of Thin Low-Temperature-Deposited Diamond Films," Diamond Relat. Mater., **6**, pp. 298–302.
[2] Banerjee, K., Amerasekera, A., Dixit, G., Cheung, N., and Hu, C., 1997, "Characterization of Contact and Via Failure under Short Duration Hight Pulsed Current Stree," *Proc. International Reliability Physics Symposium*, pp. 216–220.
[3] Cheung, N. K., Nosu, K., and Winzer, G., 1990, "Guest Editorial—Dense Wavelength Division Multiplexing Techniques for High Capacity and Multiple Access Communication Systems," IEEE J. Sel. Areas Commun., **8**, pp. 945–947.
[4] Margalit, N. M., Babic, D. I., Streubel, K., Mirin, R. P., Mars, D. E., Bowers, J. E., and Hu, E. L., 1996, "Laterally Oxidized Long Wavelength CW Vertical Cavity Lasers," Appl. Phys. Lett., **69**, pp. 471–473.
[5] Karim, A., Bjorlin, S., Piprek, J., and Bowers, J. E., 2000, "Long-Wavelength Vertical Cavity Lasers and Amplifiers," IEEE J. Sel. Top. Quantum Electron., **6**, pp. 1244–1253.
[6] Towe, E., Leheny, R. F., and Yang, A., 2000, "A Historical Perspective of the Development of the Vertical-Cavity Surface Emitting Laser," IEEE J. Sel. Top. Quantum Electron., **6**, pp. 1458–1464.
[7] Fan, X. F., Zeng, G. H., LaBounty, C., Bowers, J. E., Croke, E., Ahn, C. C., Huxtable, S., Majumdar, A., and Shakouri, A., 2001, "SiGeC/Si Superlattice Coolers," Appl. Phys. Lett., **78**, pp. 1580–1582.
[8] Mahan, G., Sales, B., and Sharp, J., 1997, "Thermoelectric Materials: New Approaches to an Old Problem," Phys. Today, **50**, pp. 42–47.
[9] Dresselhaus, M. S., Dresselhaus, G., Sun, X., Zhang, Z., Cronin, S. B., Koga, T., Ying, J. Y., and Chen, G., 1999, "The Promise of Low-Dimensional Thermoelectric Materials," Microscale Thermophys. Eng., **3**, pp. 89–100.
[10] Williams, C. C., and Wickramasinghe, H. K., 1986, "Scanning Thermal Profiler," Appl. Phys. Lett., **49**, pp. 1587–1589.
[11] Williams, C. C., and Wickramasinghe, H. K., 1988, "Photothermal Imaging With Sub-100 nm Spatial Resolution," in *Optical Sciences*, A. L. Schawlow, ed. Springer Series, pp. 364–369.
[12] Majumdar, A., Carrejo, J. P., and Lai, J., 1993, "Thermal Imaging Using the Atomic Force Microscope," Appl. Phys. Lett., **62**, pp. 2501–2503.
[13] Majumdar, A., Lai, J., Chandrachood, M., Nakabeppu, O., Wu, Y., and Shi, Z., 1995, "Thermal Imaging by Atomic Force Microscopy Using Thermocouple Cantilever Probes," Rev. Sci. Instrum., **66**, pp. 3584–3592.
[14] Stopka, M., Hadjiiski, L., Oerterschulze, E., and Kassing, R., 1995, "Surface Investigations by Scanning Thermal Microscopy," J. Vac. Sci. Technol. B, **13**, pp. 2153–2156.
[15] Luo, K., Shi, Z., Lai, J., and Majumdar, A., 1996, "Nanofabrication of Sensors on Cantilever Probe Tips for Scanning Multiprobe Microscopy," Appl. Phys. Lett., **68**, pp. 325–327.
[16] Oesterschulze, E., Stopka, M., Ackermann, L., Scholz, W., and Werner, S., 1996, "Thermal Imaging of Thin Films by Scanning Thermal Microscope," J. Vac. Sci. Technol. B, **14**, pp. 832–837.
[17] Luo, K., Shi, Z., Varesi, J., and Majumdar, A., 1997, "Sensor Nanofabrication, Performance, and Conduction Mechanisms in Scanning Thermal Microscopy," J. Vac. Sci. Technol. B, **15**, pp. 349–360.
[18] Nakabeppu, O., Igeta, M., and Hijikata, K., 1997, "Experimental Study of Point Contact Transport Phenomena Using the Atomic Force Microscope," Microscale Thermophys. Eng., **1**, pp. 201–213.
[19] Mills, G., Zhou, H., Midha, A., Donaldson, L., and Weaver, J. M. R., 1998, "Scanning Thermal Microscopy Using Batch Fabricated Thermocouple Probe," Appl. Phys. Lett., **72**, pp. 2900–2902.
[20] Nonnenmacher, M., and Wickramasinghe, H. K., 1992, "Scanning Probe Microscopy of Thermal Conductivity and Substrate Properties," Appl. Phys. Lett., **61**, pp. 168–170.
[21] Pylkki, R. J., Moyer, P. J., and West, P. E., 1994, "Scanning Near-Field Optical Microscopy and Scanning Thermal Microscopy," Jpn. J. Appl. Phys., Part 1, **33**, pp. 3785–3790.
[22] Maywald, M., Pylkki, R. J., and Balk, L. J., 1994, "Imaging of Local Thermal and Electrical Conductivity With Scanning Force Microscopy," Scanning Microsc., **8**, pp. 181–188.
[23] Hammiche, A., Hourston, D. J., Pollock, H. M., Reading, M., and Song, M., 1996, "Scanning Thermal Microscopy: Sub-Surface Imaging, Thermal Mapping of Polymer Blends, and Localized Calorimetry," J. Vac. Sci. Technol. B, **14**, pp. 1486–1491.
[24] Hammiche, A., Reading, M., Pollock, H. M., Song, M., and Hourston, D. J., 1996, "Localized Thermal Analysis Using a Miniaturized Resistive Probe," Rev. Sci. Instrum., **67**, pp. 4268–4273.
[25] Nakabeppu, O., Chandrachood, M., Wu, Y., Lai, J., and Majumdar, A., 1995, "Scanning Thermal Imaging Microscopy Using Composite Cantilever Probes," Appl. Phys. Lett., **66**, pp. 694–696.
[26] Varesi, J., and Majumdar, A., 1998, "Scanning Joule Expansion Microscopy at Nanometer Scales," Appl. Phys. Lett., **72**, pp. 37–39.
[27] Majumdar, A., and Varesi, J., 1998, "Nanoscale Temperature Distributions Measured by Scanning Joule Expansion Microscopy," ASME J. Heat Transfer, **120**, pp. 297–305.
[28] Binnig, G., Quate, C. F., and Gerber, Ch., 1986, "Atomic Force Microscope," Phys. Rev. Lett., **56**, pp. 930–933.
[29] Lai, J., Chandracood, M., Majumdar, A., and Carrejo, J. P., 1995, "Thermal Detection of Device Failure by Atomic Force Microscopy," IEEE Electron Device Lett., **16**, pp. 312–315.
[30] Luo, K., Herrick, R. W., Majumdar, A., and Petroff, P., 1997, "Scanning Thermal Microscopy of a Vertical Cavity Surface Emitting Laser," Appl. Phys. Lett., **71**, pp. 1604–1606.
[31] Ruiz, F., Sun, W. D., Pollak, F. H., and Venkatraman, C., 1998, "Determination of Thermal Conductivity of Diamond-Like Nanocomposite Films Using a Scanning Thermal Microscope," Appl. Phys. Lett., **73**, pp. 1802–1804.
[32] Majumdar, A., 1999, "Scanning Thermal Microscopy," Annu. Rev. Mater. Sci., **29**, pp. 505–585.
[33] Shi, L., 2001, "Mesoscopic Thermophysical Measurements of Microstructures and Carbon Nanotubes," Ph.D. thesis, Dept. of Mechanical Engineering, UC Berkeley.

[34] Shi, L., Kwon, O., Miner, A. C., and Majumdar, A., 2001, ''Design and Fabrication of Probes for Sub-100 nm Scanning Thermal Microscopy,'' J. of MEMS, **10**, pp. 370–378.
[35] Shi, L., and Majumdar, A., ''Thermal Transport Mechanisms at Nanoscale Point Contacts,'' ASME J. Heat Transfer (in press).
[36] Kwon, O., 2001, ''Thermal Design, Fabrication, and Imaging of MEMS and Microelectronic Structures,'' Ph.D. dissertation, Dept. of Mechanical Engineering, U.C. Berkeley.
[37] Dresselhaus, M. S., Dresselhaus, G., and Eklund, P., 1996, *Science of Fullerenes and Carbon Nanotubes*, Academic Press, New York.
[38] Shi, L., Plyasunov, S., Bachtold, A., McEuen, P., and Majumdar, A., 2000, ''Scanning Thermal Microscopy of Carbon Nanotubes Using Batch Fabricated Probes,'' Appl. Phys. Lett., **77**, pp. 4295–4297.
[39] Yao, Z., Kane, C. L., and Dekker, C., 2000, ''High-Field Electric Transport in Single-Wall Carbon Nanotubes,'' Phys. Rev. Lett., **84**, pp. 2941–2944.
[40] Phelan, P. E., Nakabeppu, O., Ito, K., Hijikata, K., Ohmori, T., and Torikoshi, K., 1993, ''Heat Transfer and Thermoelectric Voltage at Metallic Point Contacts,'' ASME J. Heat Transfer, **115**, pp. 757–762.
[41] Loomis, J. J., and Maris, H. J., 1994, ''Theory of Heat Transfer by Evanescent Electromagnetic Waves,'' Phys. Rev. B, **50**, pp. 18517–18524.
[42] Xu, J. B., Lüger, K., Möller, R., Dransfeld, K., and Wilson, I. H., 1994, ''Heat Transfer Between Two Metallic Surface at Small Distances,'' J. Appl. Phys., **76**, pp. 7209–7216.
[43] Mulet, J. P., Joulian, K., Carminati, R., and Greffet, J. J., 2001, ''Nanoscale Radiative Heat Transfer Between a Small Particle and a Plane Surface,'' Appl. Phys. Lett., **78**, pp. 2931–2933.
[44] Leinhos, T., Stopka, M., and Oesterschulze, E., 1998, ''Micromachined Fabrication of Si Cantilevers With Schottky Diodes Integrated in the Tip,'' Appl. Phys. A: Solids Surf., **66**, pp. S65–S69.
[45] Mihalcea, C., Vollkopf, A., and Oesterschulze, E., 2000, ''Reproducible Large-Area Microfabrication of Sub-100 nm Apertures on Hollow Tips,'' J. Electrochem. Soc., **147**, pp. 1970–1972.
[46] Hicks, L. D., and Dresselhaus, M. S., 1993, ''Thermoelectric Figure of Merit of a One-Dimensional conductor,'' Phys. Rev. B, **47**, pp. 16631–16634.
[47] Paddock, C. A., and Eesley, G. L., 1986, ''Transient Thermoreflectance From Thin Metal Films,'' J. Appl. Phys., **60**, pp. 285–290.
[48] Käding, O. W., Skurk, H., and Goodson, K. E., 1994, ''Thermal Conductance in Metallized Silicon-Dioxide Layers on Silicon,'' Appl. Phys. Lett., **65**, pp. 1629–1631.
[49] Lee, S.-M., and Cahill, D. G., 1997, ''Heat Transport in Thin Dielectric Films,'' J. Appl. Phys., **81**, pp. 2590–2595.
[50] Hostetler, J. L., Smith, A. N., Czajkowsky, D. M., and Norris, P. M., 1999, ''Measurement of the Electron-Phonon Coupling Factor Dependence on Film Thickness and Grain Size in Au, Cr, and Al,'' Appl. Opt., **38**, pp. 3614–3620.
[51] Clemens, B. M., Eesley, G. L., and Paddock, C. A., 1988, ''Time-Resolved Thermal Transport in Compositionally Modulated Metal Films,'' Phys. Rev. B, **37**, pp. 1085–1096.
[52] Stoner, R. J., and Maris, H. J., 1993, ''Kapitza Conductance and Heat Flow Between Solids at Temperatures From 50 to 300 K,'' Phys. Rev. B, **48**, pp. 16373–16387.
[53] Taketoshi, N., Baba, T., and Ono, A., 1999, ''Observation of Heat Diffusion Across Submicrometer Metal Thin Films Using a Picosecond Thermoreflectance Technique,'' Jpn. J. Appl. Phys., Part 2, **38**, pp. L1268–1271.
[54] Capinski, W. S., Maris, H. J., Ruf, T., Cardona, M., Ploog, K., and Katzer, D. S., 1999, ''Thermal-Conductivity Measurements of GaAs/AlAs Superlattices Using a Picosecond Optical Pump-and-Probe Technique,'' Phys. Rev. B, **59**, pp. 8105–8113.
[55] Smith, A. N., Hostetler, J. L., and Norris, P. M., 2000, ''Thermal Boundary Resistance Measurements Using a Transient Thermoreflectance Technique,'' Microscale Thermophys. Eng., **4**, pp. 51–60.
[56] Capinski, W. S., and Maris, H. J., 1996, ''Improved Apparatus for Picosecond Pump-and-Probe Optical Measurements,'' Rev. Sci. Instrum., **67**, pp. 2720–2726.
[57] Bonello, B., Perrin, B., and Rossignol, C., 1998, ''Photothermal Properties of Bulk and Layered Materials by the Picosecond Acoustics Technique,'' J. Appl. Phys., **83**, pp. 3081–3088.
[58] Carslaw, H. S., and Jaeger, J. C., 1959, *Conduction of Heat in Solids*, Oxford University Press, New York, pp. 109–112.
[59] Chen, G., and Tien, C. L., 1993, ''Internal Reflection Effects on Transient Photothermal Reflectance,'' J. Appl. Phys., **73**, pp. 3461–3466.
[60] Mertin, W., 1996, ''New Aspects in Electro-Optic Sampling,'' Microelectron. Eng., **31**, pp. 365–376.
[61] Sheridan, J. A., Bloom, D. M., and Solomon, P. M., 1995, ''System for Direct Measurement of the Step Response of Electronic Devices on the Picosecond Time-Scale,'' Opt. Lett., **20**, pp. 584–586.
[62] Ju, Y. S., and Goodson, K. E., 1999, *Microscale Heat Conduction in Integrated Circuits and Their Constituent Films*, chap. 2, Kluwer Academic Publishers, Norwell, MA.
[63] Brugger, H., and Epperlein, P. W., 1990, ''Mapping of Local Temperatures on Mirrors of GaAs/AlGaAs Laser Diodes,'' Appl. Phys. Lett., **56**, pp. 1049–1051.
[64] Ostermeier, R., Brunner, K., Abstreiter, G., and Weber, W., 1992, ''Temperature Distribution in Si-MOSFET's Studied by Micro-Ramn Spectroscopy,'' IEEE Trans. Electron Devices, **39**, pp. 858–863.
[65] Iwata, K., and Hamaguchi, H., 1997, ''Microscopic Mechanism of Solute-Solvent Energy Dissipation Probed by Picosecond Time-Resolved Raman Spectroscopy,'' J. Phys. Chem., **101**, No. 4, pp. 632–637.
[66] Martin, Y., and Wickramsinghe, H. K., 1987, ''Study of Dynamic Current Distribution in Logic Circuits by Joule Expansion Microscopy,'' Appl. Phys. Lett., **50**, pp. 167–168.
[67] Donnelly, V. M., 1993, ''Extension of Infrared-Laser Interferometric Thermometry to Silicon-Wafers Polished on Only One Side,'' Appl. Phys. Lett., **63**, No. 10, pp. 1396–1396.
[68] Glanner, G. J., Sitter, H., Faschinger, W., and Herman, M. A., 1994, ''Evaluation of Growth Temperature, Refractive-Index, and Layer Thickness of Thin znte, mnte, and cdte-films by in-situ Visible Laser Interferometry,'' Appl. Phys. Lett., **65**, No. 8, pp. 998–1000.
[69] Hall, D. C., Goldberg, L., and Mehuys, D., 1992, ''Technique for Lateral Temperature Profiling in Optoelectronic Devices Using a Photoluminescence Microscope,'' Appl. Phys. Lett., **61**, pp. 384–386.
[70] Kolodner, P., and Tyson, J. A., 1982, ''Microscopic Fluorescent Imaging of Surface Temperature Profiles with 0.01 C Resolution,'' Appl. Phys. Lett., **40**, pp. 782–784.
[71] Cardona, M., 1969, ''Modulation Spectroscopy,'' in *Solid State Physics*, Suppl. 11, F. Seitz, D. Turnbull, and H. Ehrenreich, eds., Acadmic Press, New York.
[72] Claeys, W., Dilhaire, S., Quintard, V., Dom, J. P., and Danto, Y., 1993, ''Thermoreflectance Optical Test Probe for the Measurement of Current-Induced Temperature Change in Microelectronic Components,'' Reliability Engineering International, **9**, pp. 303–308.
[73] Mansanares, A. M., Roger, J. P., Fournier, D., and Boccara, A. C., 1994, ''Temperature Field Determination of InGaAsP/InP Lasers by Photothermal Microscopy: Evidence for Weak Nonradiative Process at the Facets,'' Appl. Phys. Lett., **64**, pp. 4–6.
[74] Epperlein, P.-W., 1993, ''Micro-Temperature Measurements on Semiconductor Laser Mirrors by Reflectance Modulation: A Newly Developed Technique for Laser Characterization,'' Jpn. J. Appl. Phys., Part 1, **32**, pp. 5514–5522.
[75] Abid, R., Miserey, F., and Mezroua, F.-Z., 1996, ''Effet de la temperature sur la Reflectivite du Silicium Oxyde: Determination Experimentale de la Sensibilite Relative; Application a la Mesure sans Contact de la Temperature a la Surface d'un Thyristor GTO en Commutation,'' Journal de Physics III, **6**, pp. 279–300.
[76] Ju, Y. S., and Goodson, K. E., 1998, ''Short-Time-Scale Thermal Mapping of Microdevices using a Scanning Thermoreflectance Technique,'' ASME J. Heat Transfer, **120**, pp. 306–313.
[77] Ju, Y. S., and Goodson, K. E., 1997, ''Thermal Mapping of Interconnects Subjected to Brief Electrical Stresses,'' IEEE Electron Device Lett., **18**, pp. 512–514.
[78] Decker, D. L., and Hodgkin, V. A., 1981, ''Wavelength and Temperature Dependence of the Absolute Reflectance of Metals at Visible and Infrared Wavelengths,'' in *National Bureau of Standards Special Publication*, NBS-SP-620, Washington, D.C.
[79] Rosei, R., and Lynch, D. W., 1972, ''Thermomodulation Spectra of Al, Au, and Cu,'' Phys. Rev. B, **5**, pp. 3883–3893.
[80] Betzig, E., and Trautman, J. K., 1992, ''Near-Field Optics: Microscopy, Spectroscopy, and Surface Modification Beyond the Diffraction Limit,'' Science, **257**, pp. 189–195.
[81] Boudreau, B. D., Raja, J., Hocken, R. J., Patterson, S. R., and Patten, J., 1997, ''Thermal Imaging With Near-Field Microscopy,'' Rev. Sci. Instrum., **68**, pp. 3096–3098.
[82] Goodson, K. E., and Asheghi, M., 1997, ''Near-Field Optical Thermometry,'' Microscale Thermophys. Eng., **1**, pp. 225–235.
[83] Bethe, H. A., 1944, ''Theory of Diffraction by Small Holes,'' The Physical Review, **66**, pp. 163–182.
[84] Mansfield, S. M., and Kino, G. S., 1990, ''Solid Immersion Microscope,'' Appl. Phys. Lett., **57**, pp. 2615–2616.
[85] Terris, B. D., Mamin, H. J., Rugar, D., Studenmund, W. R., and Kino, G. S., 1994, ''Near-Field Optical-Data Storage Using a Solid Immersion Lens,'' Appl. Phys. Lett., **65**, pp. 388–390.
[86] Ghislain, L. P., Elings, V. B., Crozier, K. B., Manalis, S. R., Minne, S. C., Wilder, K., Kino, G. S., and Quate, C. F., 1999, ''Near-Field Photolithography with a Solid Immersion Lens,'' Appl. Phys. Lett., **74**, pp. 501–503.
[87] Fletcher, D. A., Crozier, K. B., Quate, C. F., Kino, G. S., Goodson, K. E., Simanovskii, D., and Palanker, D. V., 2000, ''Near-Field Infrared Imaging with a Microfabricated Solid Immersion Lens,'' Appl. Phys. Lett., **77**, pp. 2109–2111.
[88] Fletcher, D. A., 2001, ''Near-Field Microscopy with a Microfabricated Solid Immersion Lens,'' Ph.D. thesis, Department of Mechanical Engineering, Stanford University, Stanford, CA.
[89] Goodson, K. E., and Ju, Y. S., 1999, ''Heat Conduction in Novel Electronic Films,'' *Annual Review of Materials Science*, E. N. Kaufmann et al., eds., Annual Reviews, Palo Alto, CA, Vol. 29, pp. 261–293.
[90] Cahill, D. G., 1997, ''Heat Transport in Dielectric Thin-Films and at Solid-Solid Interfaces,'' Microscale Thermophys. Eng., **1**, pp. 85–109.
[91] Chen, G., and Neagu, M., 1997, ''Thermal Conductivity and Heat Transfer in Superlattices,'' Appl. Phys. Lett., **71**, pp. 2761–2763.
[92] Hyldgaard, P., and Mahan, G. D., 1997, ''Phonon Superlattice Transport,'' Phys. Rev. B, **56**, pp. 10754–10757.
[93] Chen, G., 1998, ''Thermal-Conductivity and Ballistic-Phonon Transport in the Cross-Plane Direction of Superlattices,'' Phys. Rev. B, **57**, pp. 14958–14973.
[94] Swartz, E. T., and Pohl, R. O., 1989, ''Thermal Boundary Resistance,'' Rev. Mod. Phys., **61**, pp. 605–668.
[95] Graebner, J. E., 1993, ''Thermal Conductivity of CVD Diamond: Techniques and Results,'' Diamond Films Technol., **3**, pp. 77–130.
[96] Touzelbaev, M. N., and Goodson, K. E., 1998, ''Applications of Micron-Scale

Diamond Layers for the IC and MEMS Industries,'' Diamond Relat. Mater., **7**, pp. 1–14.
[97] Kurabayashi, K., Asheghi, M., Touzelbaev, M. N., and Goodson, K. E., 1999, ''Measurement of the Thermal Conductivity Anisotropy in Polyimide Films,'' J. Microelectromech. Syst., **8**, pp. 180–191.
[98] Bauer, S., and Dereggi, A. S., 1996, ''Pulsed Electrothermal Technique for Measuring the Thermal-Diffusivity of Dielectric Films on Conducting Substrates,'' J. Appl. Phys., **80**, pp. 6124–6128.
[99] Rogers, J. A., Yang, Y., and Nelson, K. A., 1994, ''Elastic-Modulus and In-Plane Thermal Diffusivity Measurements in Thin Polyimide Films Using Symmetry-Selective Real-Time Impulsive Stimulated Thermal Scattering,'' Appl. Phys. A: Solids Surf., **58**, pp. 523–534.
[100] Goodson, K. E., and Flik, M. I., 1994, ''Solid-Layer Thermal Conductivity Measurement Techniques,'' Appl. Mech. Rev., **47**, pp. 101–112.
[101] Cahill, D. G., 1990, ''Thermal Conductivity Measurement from 30-K to 750-K: The 3-Omega Method,'' Rev. Sci. Instrum., **61**, pp. 802–808.
[102] Cahill, D. G., and Allen, T. H., 1994, ''Thermal-Conductivity of Sputtered and Evaporated SiO_2 and TiO_2 Optical Coatings,'' Appl. Phys. Lett., **65**, pp. 309–311.
[103] Lee, S. M., and Cahill, D. G., 1997, ''Heat-Transport in Thin Dielectric Films,'' J. Appl. Phys., **81**, pp. 2590–2595.
[104] BorcaTasciuc, T., Liu, W. L., Liu, J. L., Zeng, T. F., Song, D. W., Moore, C. D., Chen, G., Wang, K. L., Goorsky, M. S., Radetic, T., Gronsky, R., Koga, T., and Dresselhaus, M. S., 2000, ''Thermal Conductivity of Symmetrically Strained Si/Ge Superlattices,'' Superlattices Microstruct., **28**, pp. 199–206.
[105] Ju, Y. S., Kurabayashi, K., and Goodson, K. E., 1999, ''Thermal Characterization of Anisotropic Thin Fielectric Films using Harmonic Joule Heating,'' Thin Solid Films, **339**, pp. 160–164.
[106] Ju, Y. S., and Goodson, K. E., 1999, ''Phonon Scattering in Silicon Films of Thickness below 100 nm,'' Appl. Phys. Lett., **74**, pp. 3005–3007.
[107] Ju, Y. S., and Goodson, K. E., 1999, ''Process-Dependent Thermal Transport Properties of Silicon Dioxide Films Deposited Using Low-Pressure Chemical Vapor Deposited,'' J. Appl. Phys., **85**, pp. 7130–7134.
[108] Goodson, K. E., Flik, M. I., Su, L. T., and Antoniadis, D. A., 1994, ''Prediction and Measurement of the Thermal Conductivity of Amorphous Dielectric Layers,'' ASME J. Heat Transfer, **116**, pp. 317–324.
[109] Tai, Y. C., Mastrangelo, C. H., and Muller, R. S., 1988, ''Thermal-Conductivity of Heavily Doped Low-Pressure Chemical Vapor-Deposited Polycrystalline Silicon Films,'' J. Appl. Phys., **63**, pp. 1442–1447.
[110] Paul, O. M., Korvink, J., and Baltes, H., 1994, ''Determination of the Thermal-Conductivity of Cmos IC Polysilicon,'' Sens. Actuators A, **41**, pp. 161–164.
[111] Kaeding, O. W., Skurk, H., and Goodson, K. E., 1993, ''Thermal Conduction in Metallized Silicon-Dioxide Layers on Silicon,'' Appl. Phys. Lett., **65**, pp. 1629–1631.
[112] Goodson, K. E., Kaeding, O. W., Roesler, M., and Zachai, M., 1995, ''Experimental Investigation of Thermal Conduction normal to Diamond-Silicon Boundaries,'' J. Appl. Phys., **77**, pp. 1385–1392.
[113] Kading, O. W., Skurk, H., Maznev, A. A., and Matthias, E., 1995, ''Transient Thermal Gratings at Surfaces for Thermal Characterization of Bulk Materials and Thin-Films,'' Appl. Phys. A: Mater. Sci. Process., **61**, pp. 253–261.
[114] Special Issue of *IEEE Transactions on Electron Devices*, 1998, Vol. 45.
[115] Davis, J. A., Venkatesan, R., Kaloyeros, A., Beylansky, M., Souri, S. J., Banerjee, K., Saraswat, K. C., Rahman, A., Reif, R., and Meindl, J. D., 2001, ''Interconnect Limits on Gigascale Integration (GSI) in the 21st Century,'' Proc. IEEE, **89**, pp. 305–324.
[116] King, W. P., Kenny, T. W., Goodson, K. E., Cross, G., Despont, M., Durig, U., Rothuizen, H., Binnig, G. K., and Vettiger, P., 2001, ''Atomic Force Microscope Cantilevers for Combined Thermomechanical Data Writing and Reading,'' Appl. Phys. Lett., **78**, pp. 1300–1302.
[117] Holland, M. G., 1963, ''Analysis of Lattice Thermal Conductivity,'' Phys. Rev., **132**, pp. 2461–2471.
[118] Asheghi, M., Touzelbaev, M. N., Goodson, K. E., Leung, Y. K., and Wong, S. S., 1998, ''Temperature-Dependent Thermal Conductivity of Single-Crystal Silicon Layers in SOI Substrates,'' ASME J. Heat Transfer, **120**, pp. 30–36.
[119] Asheghi, M., Kurabayashi, K., Goodson, K. E., Kasnavi, R., and Plummer, J. D., 1999, ''Thermal Conduction in Doped Silicon Layers,'' *Proc. 33rd ASME/AIChE National Heat Transfer Conference*, Albuquerque, NM, August 8–14.
[120] Uher, C., 1990, ''Thermal Conductivity of High-T_c Superconductors,'' J. Supercond., **3**, pp. 337–389.
[121] Richardson, R. A., Peacor, S. D., Uher, C., and Nori, F., 1992, ''$YBa_2Cu_3O_{7-\delta}$ Films: Calculation of the Thermal Conductivity and Phonon Mean Free Path,'' J. Appl. Phys., **72**, pp. 4788–4791.
[122] Goodson, K. E., and Flik, M. I., 1993, ''Electron and Phonon Thermal Conduction in Epitaxial High-T_c Superconducting Films,'' ASME J. Heat Transfer, **115**, pp. 17–25.
[123] Chen, G., 1996, ''Heat Transfer in Micro and Nanoscale Photonic Devices,'' in Tien, C.-L., editor, *Annual Review of Heat Transfer*, pp. 1–57. Begell House, New York.
[124] Mahan, G. D., and Woods, L. M., 1998, ''Multilayer Thermionic Refrigeration,'' Phys. Rev. Lett., **80**, pp. 4016–4019.
[125] Mahan, G. D., 1998, ''Good Thermoelectrics,'' in H. Ehrenreich and F. Spaepen, ed., Solid State Physics, Vol. 51, Academic Press, New York, pp. 81–157.
[126] Cahill, D. G., Bullen, A., and Lee, S.-M., 2000, ''Interface Thermal Conductance and the Thermal Conductivity of Multilayer Thin Films,'' High Temperatures High Pressures, **32**, pp. 135–142.
[127] Cahill, D. G., 1998, ''Heat Transport in Dielectric Thin Films and at Solid-Solid Interfaces,'' in C.-L. Tien, A. Majumdar, and F. M. Gerner, eds., *Microscale Energy Transport*, Taylor & Francis, Washington, DC, pp. 95–117.
[128] Young, D. A., and Maris, H. J., 1989, ''Lattice-Dynamical Calculations of the Kapitza Resistance Between FCC Lattices,'' Phys. Rev. B, **40**, pp. 3685–3693.
[129] Pettersson, S., and Mahan, G. D., 1990, ''Theory of the Thermal Boundary Resistance Between Dissimilar Lattices,'' Phys. Rev. B, **42**, pp. 7386–7390.
[130] Sergeev, A. V., 1998, ''Electronic Kapitza Conductance Due to Inelastic Electron-Boundary Scattering,'' Phys. Rev. B, **58**, pp. R10199–10202.
[131] Kechrakos, D., 1991, ''The Role of Interface Disorder in the Thermal Boundary Conductivity Between Two Crystals,'' J. Phys.: Condens. Matter, **3**, pp. 1443–1452.
[132] Streib, H. M., and Mahler, G., 1987, ''Lattice Theory of Ideal Hetero Structures: Influence of Interface Models on Phonon Propagation,'' Z. Phys. B-Condensed Matter, **65**, pp. 483–490.
[133] Kim, E.-K., Kwun, S.-I., Lee, S.-M., Seo, H., and Yoon, J.-G., 2000, ''Thermal Boundary Resistance at $Ge_2Sb_2Te_5$ /ZnS:SiO_2 Interface,'' Appl. Phys. Lett., **76**, pp. 3864–3866.
[134] Cahill, D. G., and Lee, S.-M., 1997, ''Influence of Interface Conductance on the Apparent Thermal Conductivity of Thin Films,'' Microscale Thermophys. Eng., **1**, pp. 47–52.
[135] Lee, S.-M., Matamis, G., Cahill, D. G., and Allen, W. P., 1998, ''Thin Film Materials and the Minimum Thermal Conductivity,'' Microscale Thermophys. Eng., **2**, pp. 31–36.
[136] An, K., Ravichandran, K. S., Dutton, R. E., and Semiatin, S. L., 1999, ''Microstructure, Texture, and Thermal Conductivity of Single-Layer and Multilayer Thermal Barrier Coatings of Y_2O_3-stabilized ZrO_2 and Al_2O_3 Made by Physical Vapor Deposition,'' J. Am. Ceram. Soc., **82**, pp. 399–406.
[137] Soyez, G., Eastman, J. A., Thompson, L. J., Bai, G.-R., Baldo, P. M., McCormick, A. W., DiMelfi, R. J., Elmustafa, A. A., Tambwe, M. F., and Stone, D. S., 2000, ''Grain-Size-Dependent Thermal Conductivity of Nanocrystalline Yttria-Stabilized Zirconia Films Grown by Metal-Organic Chemical Vapor Deposition,'' Appl. Phys. Lett., **77**, pp. 1155–1157.
[138] Hyldegaard, P., and Mahan, G. D., 1997, ''Phonon Superlattice Transport,'' Phys. Rev. B, **56**, pp. 10754–10757.
[139] Ichiro Tamura, S., Tanaka, Y., and Maris, H. J., 1999, ''Phonon Group Velocity and Thermal Conduction in Superlattices,'' Phys. Rev. B, **60**, pp. 2627–2630.
[140] Kiselev, A. A., Kim, K. W., and Stroscio, M. A., 2000, ''Thermal Conductivity of Si/Ge Superlattices:A Realistic Model With a Diatomic Unit Cell,'' Phys. Rev. B, **62**, pp. 6896–6899.
[141] Bies, W. E., Radtke, R. J., and Ehrenreich, H., 2000, ''Phonon Dispersion Effects and the Thermal Conductivity Reduction in GaAs/AlAs Superlattices,'' J. Appl. Phys., **88**, pp. 1498–1503.
[142] Lee, S.-M., Cahill, D. G., and Venkatasubramanian, R., 1997, ''Thermal Conductivity of Si-Ge Superlattices,'' Appl. Phys. Lett., **70**, pp. 2957–2959.
[143] Simkin, M. V., and Mahan, G. D., 2000, ''Minimum Thermal Conductivity of Superlattices,'' Phys. Rev. Lett., **84**, pp. 927–930.
[144] Afromowitz, M. A., 1973, ''Thermal Conductivity of GaAlAs Alloys,'' J. Appl. Phys., **44**, pp. 1292–1294.
[145] Liu, W. L., Borca-Tasciuc, T., Chen, G., Liu, J. L., and Wang, K. L., 2001, ''Anisotropic Thermal Conductivity of Ge Quantum-Dot and Symmetrically Strained Si/Ge Superlattices,'' J. Nanosci. Nanotech., **1**, pp. 39–42.
[146] VonArx, M., Paul, O., and Baltes, H., 2000, ''Process-Dependent thin-Film Thermal Conductivities for Thermal CMOS MEMS,'' J. Microelectromech. Syst., **9**, pp. 136–145.
[147] Uma, S., McConnell, A. D., Asheghi, M., Kurabayashi, K., and Goodson, K. E., 2000, ''Temperature Dependent Thermal Conductivity of Undoped Polycrystalline Silicon Layers,'' Int. J. Thermophys., in press.
[148] McConnell, A. D., Srinivasan, U., Asheghi, M., and Goodson, K. E., 2002, ''Thermal Conductivity of Doped Polysilicon,'' J. Microelectromech. Syst., in press.
[149] Ziman, J. M., 1960, *Electrons and Phonons*, Oxford University Press, Oxford, United Kingdom.
[150] Verhoeven, H., Boettger, E., Floter, A., Reiss, H., and Zachai, R., 1997, ''Thermal-Resistance and Electrical Insulation of Thin Low-Temperature-Deposited Diamond Films,'' Diamond Relat. Mater., **6**, pp. 298–302.
[151] Goodson, K. E., 1996, ''Thermal Conduction in Nonhomogeneous CVD Diamond Layers in Electronic Microstructures,'' ASME J. Heat Transfer, **118**, pp. 279–286.
[152] Goodson, K. E., Flik, M. I., Su, L. T., and Antoniadis, D. A., 1993, ''Annealing-Temperature Dependence of the Thermal Conductivity of LPCVD Silicon-Dioxide Layers,'' IEEE Electron Device Lett., **14**, pp. 490–492.
[153] Cahill, D. G., Watson, S. K., and Pohl, R. O., 1992, ''Lower Limit to the Thermal Conductivity of Disordered Crystals,'' Phys. Rev. B, **46**, pp. 6131–6140.
[154] Alivisatos, A. P., 1996, ''Semiconductor Clusters, Nanocrystals, and Quantum Dots,'' Science, **271**, pp. 933–936.
[155] Duan, X., and Lieber, C. M., 2000, ''General Synthesis of Compound Semiconductor Nanowires,'' Adv. Mater., **12**, pp. 298–302.
[156] Morales, A. M., and Lieber, C. M., 1998, ''A Laser Ablation Method for the Synthesis of Crystalline Semiconductor Nanowires,'' Science, **279**, pp. 208–211.
[157] Yiying, W., and Yang, P., 2000, ''Germanium/Carbon Core-Sheath Nanostructures,'' Appl. Phys. Lett., **77**, pp. 43–45.
[158] Cui, Y., Lauhon, L. J., Gudiksen, M. S., Wang, J. F., and Lieber, C. M., 2001, ''Diameter-Controlled Synthesis of Single-Crystal Silicon Nanowires,'' Appl. Phys. Lett., **78**, pp. 2214–2216.
[159] Cui, Y., and Lieber, C. M., 2001, ''Functional Nanoscale Electronic Devices

Assembled Using Silicon Nanowire Building Blocks,'' Science, **291**, pp. 851–853.
[160] Lin, Y. M., Cronin, S. B., Ying, J. Y., Dresselhaus, M. S., and Heremans, J. P., 2000, ''Transport Properties of Bi Nanowire Arrays,'' Appl. Phys. Lett., **76**, pp. 3944–3946.
[161] Chung, S. W., Yu, J. W., and Heath, J. R., 2000, ''Silicon Nanowire Devices,'' Appl. Phys. Lett., **76**, pp. 2068–2070.
[162] Zhang, Z. B., Sun, X. Z., Dresselhaus, M. S., Ying, J. Y., and Heremans, J., 2000, ''Electronic Transport Properties of Single-Crystal Bismuth Nanowire Arrays,'' Phys. Rev. B, **61**, pp. 4850–4861.
[163] Schwab, K., Henriksen, E. A., Worlock, J. M., and Roukes, M. L., 2000, ''Measurement of Quantum Conductance of Thermal Conductance,'' Nature (London), **404**, pp. 974–977.
[164] Santamore, D. H., and Cross, M. C., 2001, ''Effect of Phonon Scattering by Surface Roughness on Universal Thermal Conductance,'' Phys. Rev. Lett., **87**, pp. 115502.
[165] Volz, S., and Lemonnier, D., 2000, ''Confined Phonon and Size Effects on Nanowire Thermal Conductivity. The Radiative Transfer Approach,'' Phys. Low-Dimensional Structures, **5–6**, pp. 91–107.
[166] Zou, J., and Balandin, A., 2001, ''Phonon Heat Conduction in a Semiconductor Nanowire,'' J. Appl. Phys., **89**, pp. 2932–2938.
[167] Hone, J., Whitney, M., Piskoti, C., and Zettl, A., 1999, ''Thermal Conductivity of Single-Walled Carbon Nanotubes,'' Phys. Rev. B, **59**, pp. R2514–R2516.
[168] Berber, S., Kwon, Y.-K., and Tomanek, D., 2000, ''Unusually High Thermal Conductivity of Carbon Nanotubes,'' Phys. Rev. Lett., **84**, pp. 4613–4616.
[169] Kim, P., Shi, L., Majumdar, A., and McEuen, P., 2001, ''Thermal Transport Measurements of Individual Multiwall Carbon Nanotubes,'' Phys. Rev. Lett., **87**, pp. 215502 (1–4).
[170] Borca-Tasciuc, T., Achimov, D., Liu, W. L., Chen, G., Ren, H.-W., Lin, C.-H., and Pei, S. S., 2001, ''Thermal Conductivity of InAs/AlSb superlattices,'' Microscale Thermophys. Eng., **6**, in press.

Heat Transfer in Nanostructures for Solid-State Energy Conversion

G. Chen
Assoc. Prof. Mem. ASME
Mechanical Engineering Department,
Massachusetts Institute of Technology,
Cambridge, MA 02139

A. Shakouri
Jack Baskin School of Engineering,
University of California,
Santa Cruz, CA 95064-1077

Solid-state energy conversion technologies such as thermoelectric and thermionic refrigeration and power generation require materials with low thermal conductivity but good electrical conductivity and Seebeck coefficient, which are difficult to realize in bulk semiconductors. Nanostructures such as superlattices, quantum wires, and quantum dots provide alternative approaches to improve the solid-state energy conversion efficiency through size and interface effects on the electron and phonon transport. In this review, we discuss recent research and progress using nanostructures for solid-state energy conversion. The emphasis is placed on fundamental issues that distinguish energy transport and conversion between nanoscale and macroscale, as well as heat transfer issues related to device development and property characterization. [DOI: 10.1115/1.1448331]

Keywords: Conduction, Energy Conversion, Heat Transfer, Microscale Microstructures, Nanoscale, Thermoelectric, Thin Films

1 Introduction

Accompanying the motion of charges in conductors or semiconductors, there is also an associated energy transport. Consider a current flowing through a pair of *n*-type and *p*-type semiconductors connected in series as shown in Fig. 1(*a*). The electrons in the *n*-type material and the holes in the *p*-type material all carry heat away from the top metal-semiconductor junctions, which leads to a cooling at the junctions called the Peltier effect. Conversely, if a temperature difference is maintained between the two ends of the materials as shown in Fig. 1(*b*), higher thermal energy electrons and holes will diffuse to the cold side, creating a potential difference that can be used to power an external load. This Seebeck effect is the principle for thermocouples. For each material, the cooling effect is gauged by the Peltier coefficient Π that relates the heat carried by the charges to the electrical current through $Q=\Pi\times I$. The power generation is measured by the Seebeck coefficient S, which relates the voltage generated to the temperature difference through $\Delta V=-S\times\Delta T$. The Peltier and the Seebeck coefficients are related through the Kelvin relation $\Pi=ST$. Practical devices are made of multiple pairs of *p*-type and *n*-type semiconductors as shown in Fig. 1(*c*). Analysis shows that efficient coolers and power generators should have a large figure-of-merit [1],

$$Z=\frac{\sigma S^2}{k},$$

where σ is the electrical conductivity and k the thermal conductivity. The reason that the electrical conductivity σ enters Z is due to the Joule heating in the element. Naturally, the Joule heat should be minimized by increasing the electrical conductivity. The thermal conductivity k appears in the denominator of Z because the thermoelectric elements also act as the thermal insulation between the hot and the cold sides. A high thermal conductivity causes too much heat leakage through heat conduction. Because Z has a unit of inverse temperature, the nondimensional figure of merit ZT is often used. The best ZT materials are found in semiconductors [2]. Insulators have poor electrical conductivity. Metals have relatively low Seebeck coefficients. In addition, the thermal conductivity of a metal, which is dominated by electrons, is proportional to the electrical conductivity, as dictated by the Wiedemann-Franz law. It is thus hard to realize high ZT in metals. In semiconductors, the thermal conductivity consists of contributions from electrons (k_e) and phonons (k_p), with the majority contribution coming from phonons. The phonon thermal conductivity can be reduced without causing too much reduction in the electrical conductivity. A proven approach to reduce the phonon thermal conductivity is through alloying [3]. The mass difference scattering in an alloy reduces the lattice thermal conductivity significantly without much degradation to the electrical conductivity. The traditional cooling materials are alloys of Bi_2Te_3 with Sb_2Te_3 (such as $Bi_{0.5}Sb_{1.5}Te_3$, *p*-type) and Bi_2Te_3 with Bi_2Se_3 (such as $Bi_2Te_{2.7}Se_{0.3}$, *n*-type), with a ZT at room temperature approximately equals to one [2]. A typical power generation material is the alloy of silicon and germanium, with a $ZT\sim0.6$ at 700°C. Figure 2 plots the theoretical coefficient of performance (COP) and efficiency of thermoelectric coolers and power generators for different ZT values. Also marked in the figures for comparison are other cooling and power generation technologies. Materials with $ZT\sim1$ are not competitive against the conventional fluid-based cooling and power generation technologies. Thus, solid-state cooler and power generators have only found applications in niche areas, such as cooling of semiconductor lasers and power generation for deep space exploration, although the application areas have been steadily increasing.

While the search for high ZT materials before 1990s have been mostly limited to bulk materials, there has been extensive research in the area of artificial semiconductor structures in the last 30 years for electronic and optoelectronic applications. Various means of producing ultrathin and high quality crystalline layers (such as molecular beam epitaxy and metalorganic chemical vapor deposition) have been used to alter the ''bulk'' characteristics of the materials. Drastic changes are produced by changing the crystal periodicity (by e.g., depositing alternating layers of different crystals), or by changing the electron dimensionality (by confining the carriers in a plane (quantum well) or in a line (quantum wire, etc.)). Even though electrical and optical properties of these artificial crystalline structures have been extensively studied, much less attention has been paid to their thermal and thermoelectric properties. Thermoelectric properties of low-dimensional structures started to attract attention in the 1990s, in parallel to renewed interests in certain bulk thermoelectric materials such as skutterudites [4]. Compared to the research in bulk materials that emphasizes reducing the thermal conductivity, nanostructures offer the chance of improving both the electron and phonon transport through the use of quantum and classical size and interface effects. Several directions have been explored such as quantum size effects for electrons [5,6], thermionic emission at interfaces [7,8], and interface scattering of phonons [9,10]. Impressive ZT values have been reported in some low-dimensional structures [11,12]. Comprehensive reviews on the progress of thermoelectric

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division July 24, 2001; revision received November 20, 2001. Editor: V. K. Dhir.

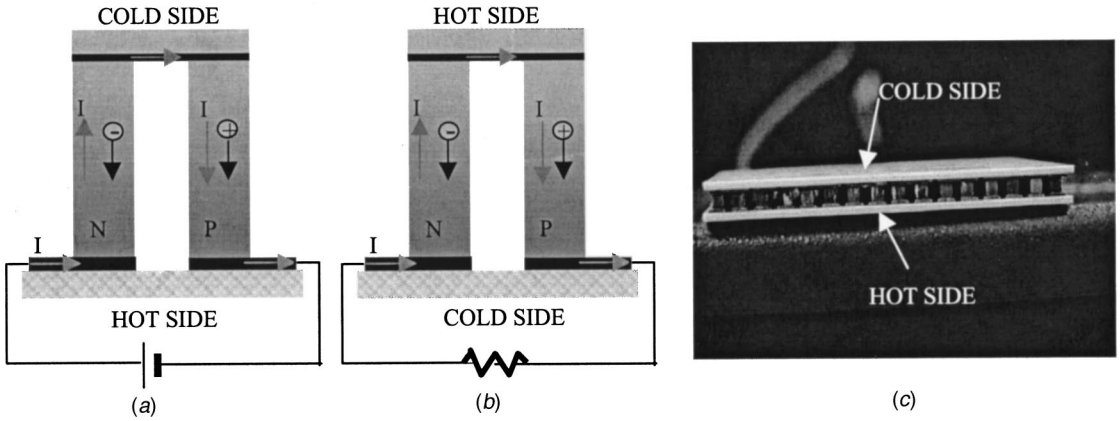


Fig. 1 Illustration of thermoelectric devices (*a*) cooler, (*b*) power generator, and (*c*) an actual device

materials research is presented in a recently published series [4] and in the proceedings of the various international conferences on thermoelectrics held in recent years.

In this article, we have in mind readers interested in nano- and microscale heat transfer and energy conversion and focus on thermoelectric energy conversion in low-dimensional structures. The emphasis is placed on fundamental issues that distinguish energy transport and conversion between nanoscale and macroscale, as well as heat transfer issues related to device development and property characterization. One of our aims is to provide the readers with an overview of recent developments. Because of the wide scope of work being carried out, the cited references are far from complete. Along with the review, we hope to stimulate the readers by pointing out unsolved, challenging questions related to the theory, characterization, and device development.

2 Formulation of Thermoelectric Effects

In solid-state coolers or power generators, heat is carried by charges from one place to another. The current density and heat flux carried by electrons can be expressed as [13]

$$\mathbf{J}(\mathbf{r}) = \frac{1}{4\pi^3}\int\int\int q\mathbf{v}(\mathbf{k})f(\mathbf{r},\mathbf{k})d^3\mathbf{k} \tag{1}$$

$$\mathbf{J}_Q(\mathbf{r}) = \frac{1}{4\pi^3}\int\int\int [E(\mathbf{k})-E_f(\mathbf{r})]\mathbf{v}(\mathbf{k})f(\mathbf{r},\mathbf{k})d^3\mathbf{k}, \tag{2}$$

where q is the unit charge of each carrier, E_f the Fermi energy, $\mathbf{v}$ the carrier velocity, and the integration is over all the possible wavevectors $\mathbf{k}$ of all the charges. The carrier probability distribution function, $f(\mathbf{r},\mathbf{k})$ is governed by the Boltzmann equation. Considering transport processes occurring much slower than the relaxation process and employing the relaxation time approximation, the Boltzmann equation can be expressed as

$$\mathbf{v}\cdot\nabla_{\mathbf{r}}f + \frac{q\boldsymbol{\varepsilon}}{\hbar}\cdot\nabla_{\mathbf{k}}f = -\frac{f(\mathbf{r},\mathbf{k})-f_{\mathrm{eq}}(\mathbf{r},\mathbf{k})}{\tau(\mathbf{k})}, \tag{3}$$

where is $\boldsymbol{\varepsilon}$ the electric field, $\tau(\mathbf{k})$ the momentum-dependent relaxation time, $\hbar$ the Planck constant divided by 2π, and f_{eq} the equilibrium distribution function. For electrons and holes,

$$f_{\mathrm{eq}}(\mathbf{r},\mathbf{k}) = \frac{1}{1+\exp\left(\dfrac{E(\mathbf{k})-E_f(\mathbf{r})}{k_BT(\mathbf{r})}\right)}, \tag{4}$$

where k_B is the Boltzmann constant, and T the local temperature. Under the further assumption that the local deviation from equilibrium is small, the Boltzmann equation can be linearized and its solution expressed as

$$f(\mathbf{r},\mathbf{k}) = f_{\mathrm{eq}}(\mathbf{r},\mathbf{k}) + \tau(\mathbf{k})\mathbf{v}\left(-\frac{\partial f_{\mathrm{eq}}}{\partial E}\right)\cdot\left[-\frac{E(k)-E_f}{T}\nabla_{\mathbf{r}}T + q\left(\boldsymbol{\varepsilon}-\frac{1}{q}\nabla_{\mathbf{r}}E_f\right)\right]. \tag{5}$$

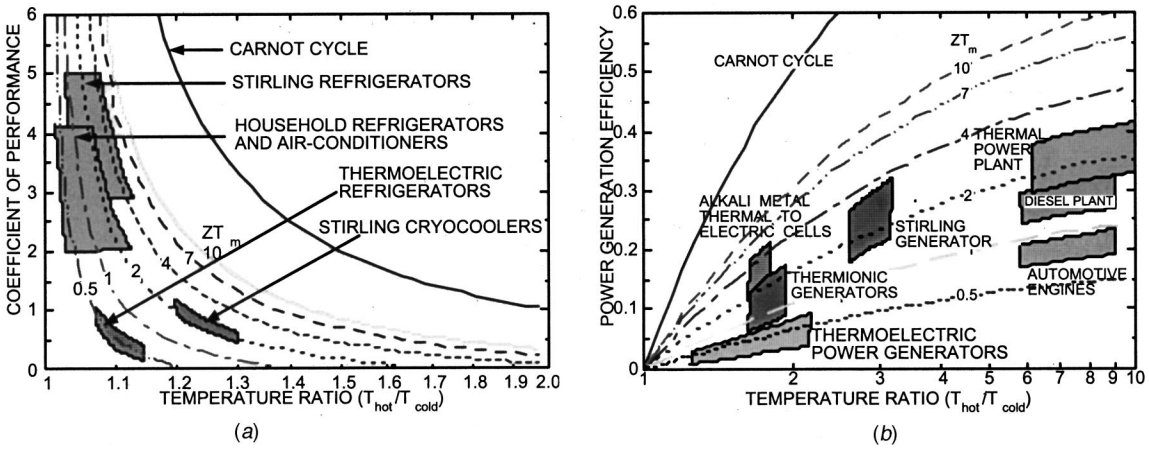


Fig. 2 Comparison of thermoelectric technology with other energy conversion methods for (*a*) cooling and (*b*) power generation

Substituting the above expression in Eqs. (1) and (2) leads to the

$$\mathbf{J}(\mathbf{r}) = q^2 L_0 \left(-\frac{1}{q}\nabla\Phi \right) + \frac{q}{T} L_1 (-\nabla T) \tag{6}$$

$$\mathbf{J}_Q(\mathbf{r}) = q L_1 \left(-\frac{1}{q}\nabla\Phi \right) + \frac{1}{T} L_2 (-\nabla_{\mathbf{r}} T), \tag{7}$$

where Φ is the electrochemical potential $(-\nabla\Phi/q = \boldsymbol{\varepsilon} - \nabla E_f/q)$. The transport coefficients L_n are defined by the following integral

$$L_n = \frac{1}{4\pi^3} \int\int\int \tau(\mathbf{k}) v(\mathbf{k}) v(\mathbf{k}) (E(\mathbf{k}) - E_F)^n \left(-\frac{\partial f_{\text{eq}}}{\partial E} \right) d^3\mathbf{k}. \tag{8}$$

From the expressions for $\mathbf{J}$ and $\mathbf{J}_Q$, various material parameters such as the electrical conductivity, thermal conductivity due to electrons, and the Seebeck coefficient can be calculated. For simplicity we assume that both the current flow and the temperature gradient are in the *x*-direction:

$$\sigma = J_x/(-\nabla\Phi/q)|_{\nabla_x T=0} = q^2 L_0 \tag{9}$$

$$S = (-\nabla\Phi/q)/\nabla_x T|_{J_x=0} = \frac{1}{qT} {L_0}^{-1} L_1 \tag{10}$$

$$k_e = J_{Q_x}/(-\nabla_x T)|_{J_x=0} = \frac{L_2 L_0 - L_1 L_1}{T L_0}. \tag{11}$$

Rewriting the expressions for electrical conductivity and the thermopower in the form of integrals over the electron energy we get

$$\sigma \equiv \int \sigma(E) \left(-\frac{\partial f_{\text{eq}}}{\partial E} \right) dE \tag{12}$$

$$S \equiv \frac{k_B}{q} \frac{\int \sigma(E) \frac{(E-E_F)}{k_B T} \left(-\frac{\partial f_{eq}}{\partial E} \right) dE}{\int \sigma(E) \left(-\frac{\partial f_{eq}}{\partial E} \right) dE} \propto \langle E - E_f \rangle, \tag{13}$$

where we introduced the ''differential'' conductivity,

$$\sigma(E) \equiv q^2 \tau(E) \int\int v_x^2(E, k_y, k_z) dk_y dk_z \cong q^2 \tau(E) \overline{v_x^2}(E) D(E), \tag{14}$$

where $D(E)$ is the density of states. $\sigma(E)$ is a measure of the contribution of electrons with energy E to the total conductivity. The Fermi ''window'' factor $(-\partial f_{eq}/\partial E)$ is a bell-shape function centered at $E = E_f$, having a width of $\sim k_B T$. At a finite temperature only electrons near the Fermi surface contribute to the conduction process. In this picture, the thermopower is the ''average'' energy transported by the charge carriers. In order to achieve the best thermoelectric properties, σ (E), within the Fermi window, should be as big as possible, and at the same time, as asymmetric as possible with respect to the Fermi energy.

The thermal conductivity of phonons is also often modeled from the Boltzmann equation under the relaxation time approximation,

$$k_p = \frac{1}{3} \sum \int C(\omega) v_p(\omega) \Lambda(\omega) d\omega, \tag{15}$$

where C is the specific heat of phonons at frequency ω, v_p the phonon group velocity, and Λ the phonon mean free path.

The above formulation for the thermoelectric properties leads to the following possibilities to increase *ZT* and thus the energy conversion efficiency of devices made of nanostructures.

1. Interfaces and boundaries of nanostructures impose constraints on the electron and phonon waves, which lead to a change in their energy states and correspondingly, their density of states and group velocity.
2. The symmetry of the differential conductivity with respect to the Fermi level can be controlled using quantum size effects and classical interface effects (as in thermionic emission).
3. The phonon thermal conductivity can be reduced through interface scattering and through the alteration of the phonon spectrum in low-dimensional structures.

3 Nanostructures for Solid-State Energy Conversion

The transport of electrons and phonons in nanostructures is affected by the presence of the interfaces and surfaces. Since electrons and phonons have both wave and particle characteristics, the transport can fall into two different regimes: totally coherent transport in which electrons or phonons must be treated as waves and totally incoherent transport in which either or both of them can be treated as particles. There is, of course, the intermediate regime where transport is partially coherent—an area that has not been studied extensively. Whether a group of carriers are coherent or incoherent depend on the strength of phase destroying scattering events (such as internal or diffuse interface scattering). In a nanostructure with no phase-destroying scattering events, a monochromatic wave can experience many coherent scatterings while preserving the phase. The coherent superposition of the incoming and scattered waves leads to the formation of new energy bands for electrons and/or phonons. For example, the quantized energy states of electrons in a quantum well are the result of the formation of standing waves inside the structure. The standing wave can be regarded as the superposition of two counter-propagating waves, each experiencing phase preserving reflections at the interfaces. On the other hand, if there is a strong internal scattering (which can be judged from the momentum relaxation time) or if the interface scattering is not phase preserving (such as due to diffuse scattering), no new energy bands form and the energy states of the carriers in such a structure are identical to these in its bulk material. Electron transport in both coherent and incoherent regimes has been considered and potential benefits of nanostructures for the power factor $(S^2\sigma)$ have been studied. Similarly, phonon heat conduction in both regimes has also be considered, although most studies are based on the particle approach. We will divide the discussion into roughly three categories: (1) improving the electronic power factor based on coherent electron states, (2) improving the electronic energy conversion based on interface filtering for incoherent electrons, (3) improving *ZT* by reducing the phonon thermal conductivity.

3.1 Electron Energy States Engineering. This approach was suggested in a pioneering work by Dresselhaus and coworkers [5,6]. The main idea is that energy states in nanostructures are very different from those in macrostructures due to the quantum size effects on electrons. Figure 3 illustrates, qualitatively, the density of states (DOS) of electrons in bulk materials,

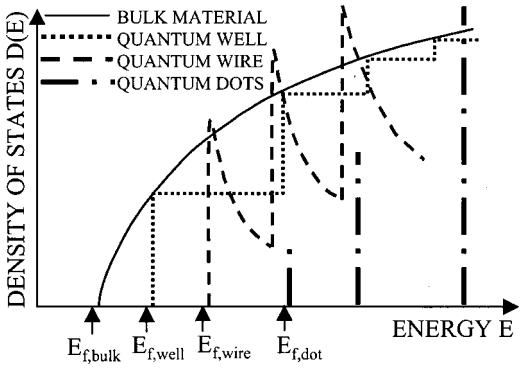


Fig. 3 Schematic illustration of the density-of-states of electrons in bulk, quantum well, quantum wire, and quantum dots materials.

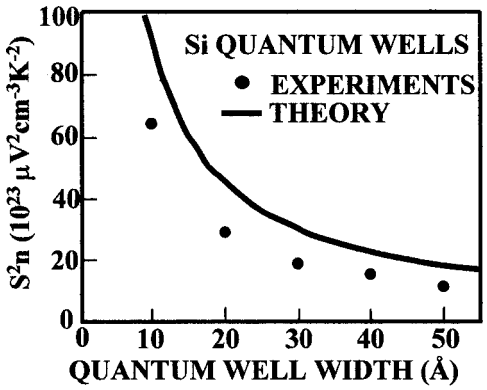


Fig. 4 Product of the Seebeck coefficient square and carrier density as a function of the silicon quantum well width [15]

quantum wells, quantum wires, and quantum dots. Examination of Eq. (13) indicates that the Seebeck coefficient is large when the average electron energy is far apart from the Fermi level. In semiconductors, a large Seebeck coefficient occurs when the Fermi level is inside the band-gap. A Fermi level deep inside the band-gap, however, leads to a low electrical conductivity. The optimized Fermi level usually is close to the band edge. Because the function $\partial f_{eq}/\partial E$ is nonzero only in an energy range $\sim k_B T$ near the Fermi level, the higher the DOS in this range, the larger power factor we can anticipate. In bulk materials, the parabolic shape of the DOS means that the electron density surrounding the Fermi level is small. In quantum structures, the steps and the spikes in the DOS suggest that $S^2\sigma$ can be increased. In a theoretical study by Mahan and Sofo [14], it was suggested that the best thermoelectric materials will have a spike like DOS. Quantum dots fit ideally into such a picture. A single quantum dot, however, is not of much interest for building into useful thermoelectric devices (but may be of interest to create localized cooling on the nanoscale). Thus the study began with quantum wells (extremely thin films) and quantum wires (extremely small wires). Experimental results for transport inside PbTe and Si/SiGe quantum well systems indicated an increase of *ZT* inside the quantum well, as shown in Fig. 4 [15].

A single quantum well, however, cannot be used to build useful devices because the film is too thin (typically less than a few hundreds angstroms). Multilayer structures were therefore used in the proof-of-concept experiments. For multilayer structures such as superlattices, three questions were raised on the effectiveness of the quantum confinement approach [16,17]. One is that electrons will tunnel through the barrier layer when the barriers are very thin. The second argument is that the barrier does not contribute to the thermoelectric transport but does contribute to the reverse heat conduction. And finally, there is also concern of interface scattering of electrons in narrow wells. A possible approach for improving performance is to utilize the quantum confinement effects inside both the quantum well and the barrier layer, and to have electrons in different carrier pockets in momentum space confined in different regions [18,19]. Additionally, the thermal conductivity of very thin superlattices can be reduced due to interface scattering—a topic we will discuss later on [20]. A natural extension of the quantum well and superlattice theory is to quantum wires. Theoretical studies predict a large enhancement of *ZT* inside quantum wires. Experimentally, different quantum wire deposition methods have been explored [6].

The experimental results that are inspired by the theoretical studies have proven to be impressive and unexpected [21]. After the proof-of-concept demonstration of *ZT* enhancement in two-dimensional quantum wells, Harman's group showed that quantum-dot superlattices have a significantly higher power factor than their corresponding bulk materials [22]. Using a thermal conductivity value estimated from bulk properties, *ZT* values as high as two have been reported. At this stage, no models exist to quantitatively explain the observed increase in *ZT*.

The above-discussed approaches are based on transport perpendicular to the confinement directions, i.e., along the film plane or wire axis. There are also considerations of the DOS change for electron transport perpendicular to the film plane of the superlattices [23,24]. These calculations, however, do not show a significant increase of the electronic power factor along these directions and suggest that the thermal conductivity reduction may be a more beneficial factor to explore along this transport direction.

Research for improving *ZT* using quantum confinement of electrons raises several interesting questions related to heat transfer [20]. First, the thermal conductivity along the film plane and wire axis should be reduced due to phonon interface scattering. We will discuss more about this point later. Second, the parasitic thermal conductivity in the barrier layer is still a concern for certain quantum structures that do not utilize the barrier region for electron transport. For example, nanowires can be deposited in an anodized alumina matrix with nanometer scale channels. Although the thermal conductivity of the anodized alumina is relatively small, it is not negligible, and further reduction of the parasitic heat conduction path through such a matrix should be considered [25]. Another very interesting yet little explored aspect is the transport processes in the synthesis of nanowires. Several different methods have been explored, including pressure injection of molten bismuth into the template [26], physical vapor deposition [27], and electrodeposition [28]. Heat and mass transport inside these nanoscale channels could be very different than that in bulk channels. So far, these techniques are developed only through trial-and-error. It is generally found that smaller diameter channels are more difficult to fill, leading to partial filling or discontinuous wires. Systematic studies of the transport processes will help with the optimization of the deposition conditions.

3.2 Heterostructure Integrated Thermionic Refrigeration. Thermionic energy conversion is based on the idea that a high work function cathode in contact with a heat source will emit electrons [29]. These electrons are absorbed by a cold, low work function anode, and they can flow back to the cathode through an external load where they do useful work. Practical vacuum thermionic generators are limited by the work function of available metals or other materials that are used for cathodes. Another important limitation is the space charge effect. The presence of charged electrons in the space between the cathode and anode will create an extra potential barrier between the cathode and anode, which reduces the thermionic current. The materials currently used for cathodes have work functions >0.7 eV, which limits the generator applications to high temperatures >500 K. Mahan [30] proposed these vacuum diodes for thermionic refrigeration. Basically, the same vacuum diodes which are used for generators will work as a cooler on the cathode side and a heat pump on the anode side under an applied bias. Mahan predicted efficiencies of over 80 percent of the Carnot value, but still these refrigerators only work at high temperatures (>500 K). In the following, we will see that heterostructures have a potential to achieve thermionic refrigeration at room temperature [7,31–33]. Vacuum thermionic refrigeration based on resonant tunneling through triangular wells and small gaps in vacuum has been proposed recently [34,35]. Theoretical calculations predict operation at room temperature and below with a cooling power density of 100 W/cm^2. Net cooling based on such vacuum thermionic coolers has yet to be confirmed experimentally.

Using various material systems one can produce different barrier heights in the anode and in the cathode (typically 0 to 0.4 eV). This is determined by the band-edge discontinuity between heterolayers. The heterostructure integrated thermionic coolers (HIT) in Fig. 5 could operate in two modes. In the nonlinear regime, electron transport is dominated by the supply of electrons in the cathode layer. Since only hot electrons (with energy greater than

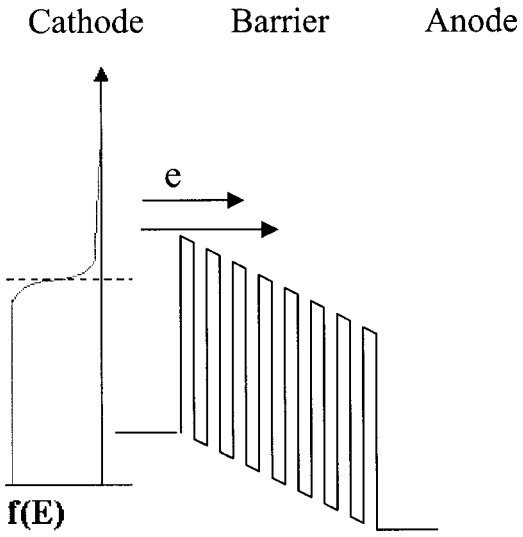


Fig. 5 Heterostructure thermionic emission for cooling at room temperatures.

E_f) are emitted above the barrier, electron-electron and electron-phonon interactions try to restore the quasi Fermi distributions in the cathode layer by absorbing heat from the lattice, thus cooling the layer. This heat is deposited on the anode side. Theoretical estimates by Shakouri and co-workers [7,32] show that there is an optimal barrier width of the order of a few electron energy relaxation lengths and an optimum barrier height of the order of k_BT, and that such heterostructure coolers can provide 20–30°C cooling with KW/cm^2 cooling density. Since the operating currents for the device is very high (10^5 A/cm^2), non-ideal effects such as the Joule heating at the metal-semiconductor contact resistance, and the reverse heat conduction have limited the experimental cooling results to <1°C. There is another regime of operation in which electron transport is dominated by the barrier structure. A superlattice is chosen so that hot electrons move easily in the materials, but the movement of cold electrons is more restricted. In this case, there will be also net cooling in the cathode layer and heating in the anode layer. Shakouri et al. [36] noted that a small barrier height on the order of k_BT does not give much improvement over bulk thermoelectric materials, and suggested tall barriers and high doping densities to achieve a large number of electrons moving in the material. To have a good HIT cooler, the barrier material should simply have an adequate electrical conductivity and a low thermal conductivity, making ternary and quaternary semiconductors good candidates.

On the experimental side, Shakouri and co-workers have fabricated thin-film thermoelectric coolers based on single heterojunction structures [37] and superlattice structures [38–42]. The SiGe/Si superlattice micro coolers can be monolithically integrated with Si-based microelectronic devices to achieve localized cooling and temperature control. Cooling by as much as 4.2 K at 25°C and 12 K at 200°C was measured on 3 μm thick, 60 $\times 60$ μm^2 devices. This corresponds to maximum cooling power densities approaching kW/cm^2. The micro cooler structure is based on cross-plane electrical transport. The main part of the cooler is a 3 μm strain-compensated SiGe/Si superlattice. It consists of 200 periods of (12 nm $Si_{0.75}Ge_{0.25}$/3 nm Si), doped with boron to about 6×10^{19} cm^{-3}. The $Si_{0.75}Ge_{0.25}$/Si superlattice has a valence band offset of about 0.16 eV, and hot holes going over this barrier can produce thermionic cooling. This superlattice was grown using molecular beam epitaxy (MBE). Its average lattice constant is that of $Si_{0.8}Ge_{0.2}$, and a buffer layer is required for it to be grown on a Si substrate. To reduce the material growth time in the MBE system, the buffer layer was grown on a p^+ (001) Si

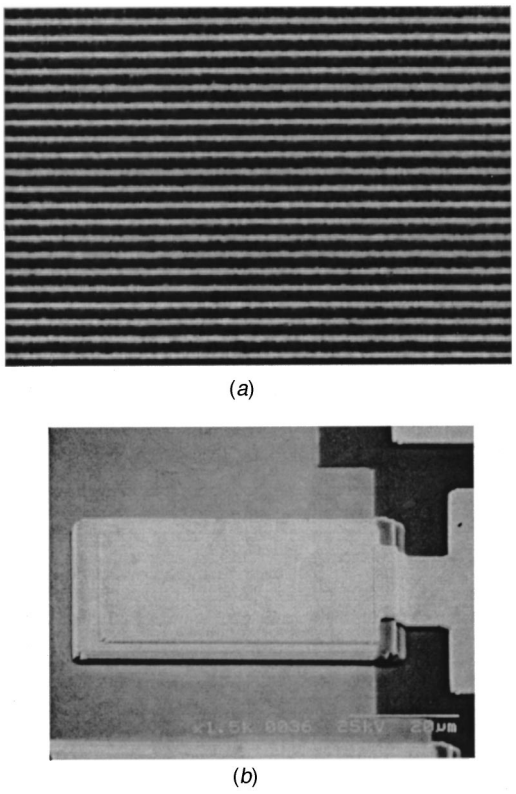


Fig. 6 (*a*) TEM image of the SiGe/Si superlattice (the dark parts are the 12 nm $Si_{0.75}Ge_{0.25}$ layers, the light parts are the 3 nm Si layers), and (*b*) a scanning electron micrograph of a fabricated micro refrigerator

substrate by chemical vapor deposition (CVD) in the form of a graded SiGe structure. The boron doping in the buffer layer is 5×10^{19} cm^{-3}. Following the superlattice growth, a 0.3 μm $Si_{0.8}Ge_{0.2}$ cap layer was grown with a boron doping level of 2×10^{20} cm^{-3} to get a good ohmic contact to the device. Figure 6 shows a cross-section transmission electron microscopy (TEM) image of the MBE-grown SiGe/Si superlattice, and also a scanning electron micrograph of a micro cooler device. Figure 7 displays the measured cooling on 60×60 μm^2 superlattice cooler and a Si cooler at the heat sink temperature of 25°C. Figure 8 shows the temperature profile at a current of ~400 mA. One notices localized and uniform cooling on top of the micro refrig-

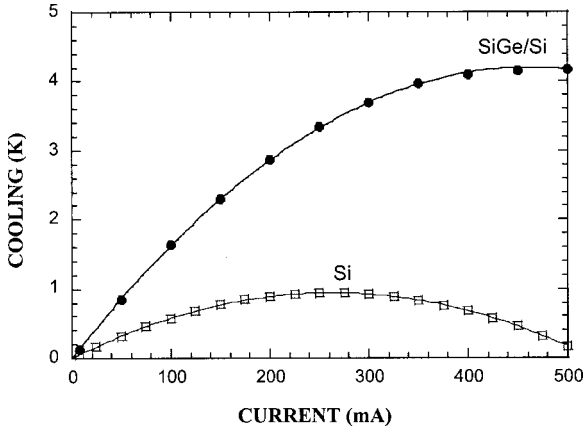


Fig. 7 Cooling measured on 60×60 μm^2 SiGe/Si superlattice coolers and on Si coolers at the heat sink temperature of 25°C

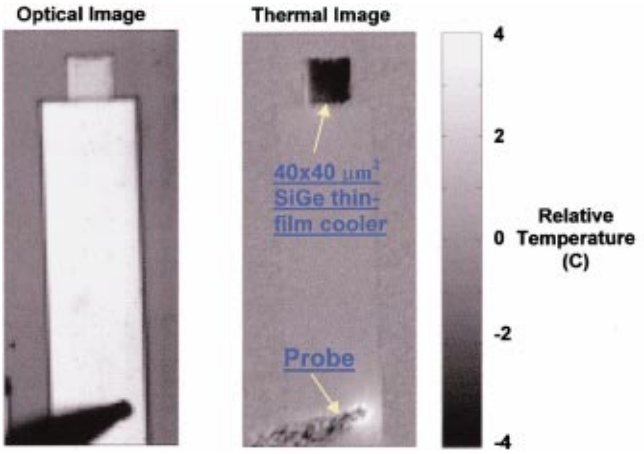


Fig. 8 Temperature distribution on top of a 40×40 micron square SiGe thin film cooler measured using thermoreflectance imaging. The applied current is ~400 mA.

erator as well as Joule heating near the probe on the side contact. With the use of a resistive heat load on top of the micro refrigerator, cooling power densities exceeding 500 W/cm^2 have been demonstrated [43].

Electron and phonon transport perpendicular to interfaces raise interesting heat transfer and energy conversion issues. One example is where heat is generated. Joule heating is often treated as a uniform volumetric heat generation. In heterostructures, the energy relaxation from electrons to phonons occurs over a distance comparable to the film thickness, and heat generation is no longer uniform. For single layer devices, this could benefit the device efficiency in principle [32,44]. Such non-uniform heat generation is a type of hot electron effect that has been studied in electronics [45], and has also be discussed quite extensively in the literature in the context of ultrafast laser-matter interactions [46]. Another example is the concurrent consideration of ballistic electron transport and ballistic phonon transport, coupled with nonequilibrium electron-phonon interaction. Zeng and Chen [47] started from the Boltzmann equations for electrons and phonons and obtained approximate solutions for the electron and phonon temperature distributions in heterostructures, as shown in Fig. 9. In this case, both electron and phonon temperatures show a discontinuity at the interface. The phonon temperature discontinuity is the familiar thermal boundary resistance phenomenon. Zeng and Chen concluded that in the nonlinear transport regime, it is the electron temperature discontinuity at the interface that determines the thermionic effect and the electron temperature gradient inside the film that determines the thermoelectric effect. Similar calculations by Vashaee and Shakouri [48] showed the importance of the electron-phonon coupling coefficient in the optimization of HIT coolers.

3.3 Phonon Thermal Conductivity Reduction Approach. Although phonons do not contribute directly to the energy conversion, the reduction of their contribution to the thermal conductivity is a central issue in thermoelectrics research. Several significant increases in the *ZT* of bulk materials were due to the introduction of thermal conductivity reduction strategies, such as the alloying [3] and phonon rattler concepts [49]. Size effects on phonon transport have long been established since the pioneering work by Casimir [50] at low-temperatures. Since the 1980s, the thermal conductivity reduction in thin films has drawn increasing attention. Naturally, the phonon thermal conductivity reduction in nanostructures has been considered as beneficial and even as a dominant approach to enhance *ZT* values.

One proposed approach is to use the thermal conductivity in the direction perpendicular to the superlattice film plane, or the cross-plane direction, while maintaining a low electronic band-edge offset, ideally no offset at all [9]. This would allow the electron transport across the interfaces without much scattering, while phonons would be scattered at the interfaces [51]. Some early experimental data [52,53] indicate that the thermal conductivity of superlattices could be significantly reduced, especially in the cross-plane direction. Tien and Chen [54] suggested the possibility of making super thermal insulators out of superlattices. Extensive experimental data on the thermal conductivity of various superlattices have been reported in recent years [51–65], mostly in the cross-plane direction. Following such a strategy, Venkatasubramanian's group has reported Bi_2Te_3/Se_2Te_3 superlattices with *ZT* values between 2-3 at room temperature. Although these results need to be confirmed due to the difficulty with the measurements, the reported data so far seem to support such claims and demonstrate that the thermal conductivity reduction is a very effective approach [9].

The mechanisms responsible for thermal conductivity reduction in low-dimensional structures thus have become a topic of considerable debate over the last few years. There have been many studies of the phonon spectrum and transport in superlattices since the original work by Narayanamurti et al. [66], but these works were focused on the phonon modes rather than on heat conduction. The first theoretical modeling predicted a small reduction of

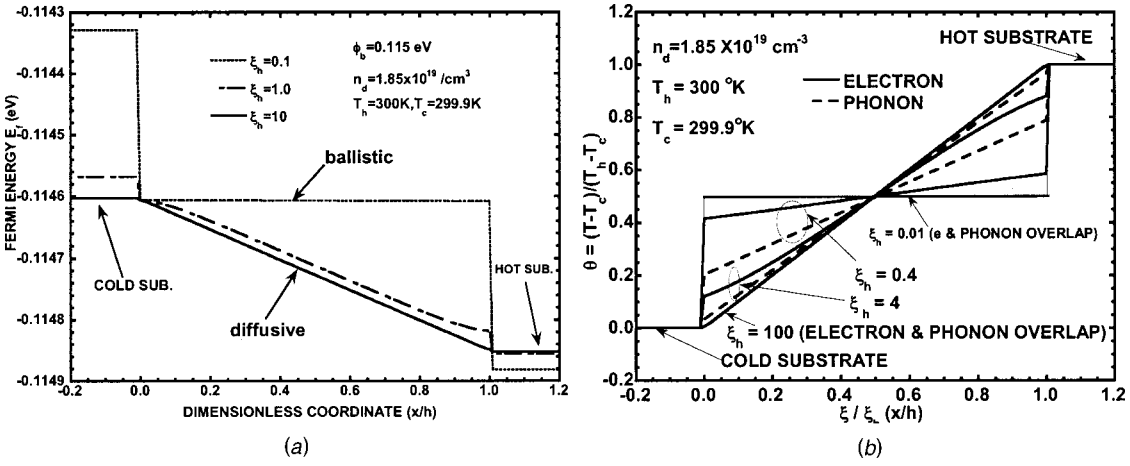


Fig. 9 Distribution of (*a*) Fermi level and (*b*) electron and phonon temperature inside double heterojunction structures. The dimensionless coordinate is normalized to the film thickness. ξ_h is the electron or phonon mean free path divided by the film thickness, n_d the carrier concentration and ϕ_b the barrier height.

the superlattice thermal conductivity [67] due to the formation of minigaps or stop bands. This predicted reduction, however, was too small compared to experiment results in recent years. Two major theoretical approaches were developed in the 1990s to explain the experimental results. One is based on solving the Boltzmann equation with the interfaces of the superlattice treated as boundary conditions [68–71]. The other is based on lattice dynamics calculation of the phonon spectrum and the corresponding change in the phonon group velocity [72–76]. More recently, there are also efforts to use molecular dynamics to simulate the thermal conductivity of superlattices directly [77,78].

Similar to the electron transport in superlattice structures, there could be several different regimes of phonon transport: the totally coherent regime, the totally incoherent regime, and the partially coherent regime. The lattice dynamics lies in the totally coherence regime. Such approaches are based on the harmonic force interaction assumption and thus do not consider anharmonic effects. A bulk relaxation time is often assumed. The main results out of the lattice dynamics models is that the phonon group velocity reduction caused by the spectrum change can lower the thermal conductivity by a factor of ~7–10 at room temperature for Si/Ge superlattices, and by a factor of 3 for GaAs/AlAs superlattices. Although it can be claimed that the predicted reduction in Si/Ge system is of the order of magnitude that is experimentally observed, the prediction clearly cannot explain the experimental results for GaAs/AlAs superlattices. The lattice dynamics model also showed that when the layers are 1–3 atomic-layers thick, there is a recovery of the thermal conductivity. The acoustic wave based model [79], which treats the superlattices as an inhomogeneous medium, shows a similar trend. It reveals that the thermal conductivity recovery is due to phonon tunneling and that the major source of the computed thermal conductivity reduction in the lattice dynamics model is the total internal reflection, which in the phonon spectrum representation, causes a group velocity reduction. For experimental results so far, the explanation of the thermal conductivity reduction based on the group velocity reduction has not been satisfactory even for the cross-plane direction. For the in-plane direction, the group velocity reduction alone leads to only a small reduction in thermal conductivity [76], and cannot explain the experimental data on GaAs/AlAs and Si/Ge superlattices [52,55,65]. There is a possibility that the change in the phonon spectrum creates a change in the relaxation time [80] but such a mechanism is unlikely to explain the experimental results for relatively thick-period superlattices since the density of states does not change in these structures [76].

Boltzmann equation-based models that treat phonons as particles transporting heat in inhomogeneous layers lie in the totally incoherent regime [68,71]. Theoretical calculations have been able to explain quantitatively the experimental data. The models are based on the solution of the Boltzmann equation using the relaxation time in the bulk materials for each layer. Phonon reflection and transmission at the interfaces are modeled based on past studies of the thermal boundary resistance. Compared to the lattice dynamics and acoustic waves models, the particle model allows the incorporation of diffuse interface scattering of phonons. In the models presented so far, the contribution of diffuse scattering has been left as a fitting parameter. In Fig. 10, we show the experimental in-plane and cross-plane thermal conductivity of a Si/Ge superlattice, together with simulation results based on the Boltzmann equation. One argument for the validity of the particle model is that thermal phonons have a short thermal wavelength, which is a measure of the coherence properties of broadband phonons inside the solid [68]. It is more likely, however, that the diffuse interface scattering, if it indeed happens as models suggested, destroys the coherence of monochromatic phonons and thus prevents the formation the superlattice phonon modes. The particle-based modeling can capture the effects of total internal reflection, which is partially responsible for the large group velocity reduction under the lattice dynamics models. Approximate

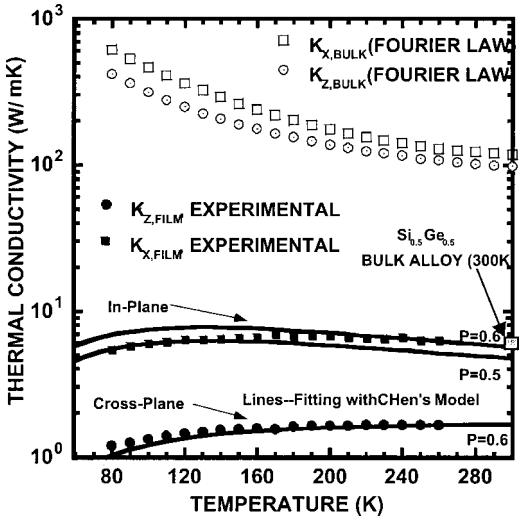


Fig. 10 Anisotropic thermal conductivity of the strained Si/Ge (20 Å /20 Å) superlattice: experimental data were fitted using Chen's models [68,71]. Also shown in the figure are comparisons of experimental data experimental data with predictions of Fourier theory based on bulk properties of each layer, and with compositionally equivalent alloy (300K) [65].

methods to incorporate phonon confinement or inelastic boundary scattering are also proposed. From the existing modeling, it can be concluded that for heat flow parallel to the interfaces, diffuse interface scattering is the key factor causing the thermal conductivity reduction. For the case of heat conduction perpendicular to the interfaces, phonon reflection, confinement, as well as diffuse scattering can greatly reduce the heat transfer and thermal conductivity. The larger the reflection coefficient, the larger is the thermal conductivity reduction in the cross-plane direction.

A key unsolved issue is what are the actual mechanisms of phonon scattering at the interfaces, particularly what causes the diffuse phonon scattering. Phonon scattering has been studied quite extensively in the past in the context of thermal boundary resistance. Superlattice structures that are grown by epitaxy techniques usually have better interface morphology than the other types of interfaces studied previously. Even for the best material system such as GaAs/AlAs, however, the interfaces are not perfect. There is interface mixing and there are also regions with monolayer thickness variation. These are naturally considered as potential sources of diffuse interface scattering for which a simplified model was developed by Ziman [82]. Another possibility is the anharmonic force between the atoms in two adjacent layers. Boltzmann equation-based modeling assumes a constant parameter p to represent the fraction of phonons specularly scattered. Ju and Goodson [81] used an approximate frequency-dependent expression for p given by Ziman [82] in the interpretation of the thermal conductivity of single layer silicon films. Chen [71] also argued that inelastic scattering occurring at interfaces can provide a path for the escape of confined phonons. A promising approach to resolve this issue is molecular dynamics simulation [77,78]. In addition to the interface scattering mechanisms, there are also several other unanswered questions. For example, experimental data of Venkatasubrmanian seems to indicate a butterfly-shaped thermal conductivity curve as a function of thickness [9,83]. Quantitative modeling of the stress and dislocation effects also needs to be further refined.

Since the lattice dynamics and the particle models present the totally coherent and totally incoherent regimes, a theoretical approach that can include both effects should be sought. Simkin and Mahan [84] proposed a new lattice dynamics model by the introduction of an imaginary wavevector that is related to the mean free path. This approach leads to the prediction of a minimum in

the thermal conductivity value as a function of the superlattice period thickness. For thicknesses larger than the minimum, the thermal conductivity increases with thickness and eventually approaches the bulk values. For thicknesses thinner than the minimum, the thermal conductivity recovers to a higher value. However, it should be pointed out that the imaginary wavevector represents an absorption process, not exactly a scattering process, as is clear in the meaning of the extinction coefficient of the optical constants. It will be interesting to see whether such an approach can explain the experimental observed trends of thermal conductivity reduction along the in-plane direction.

Aside from superlattices and thin films, other low-dimensional structures such as quantum wires and quantum dots are also being considered for thermoelectric applications. There are a few experimental and theoretical studies on the thermal conductivity of quantum dot arrays and nanostructured porous medium [85,86]. Theoretically, one can expect a larger thermal conductivity reduction in quantum wires compared to thin films [87,88]. The measurements of the thermal conductivity in quantum wires have been challenging. Recent measurements on the thermal conductivity of carbon nanotubes provide possible approach for measurements on nanowires for thermoelectric applications [89]. Nanowires for thermoelectric applications, however, usually have a low thermal conductivity, which may need different characterization techniques.

4 Characterization

The characterization of thermoelectric properties has turned out to be the most challenging issue for the development of nanostructure-based thermoelectric materials. First, the thermal conductivity measurements for even bulk materials are not easy. For thin films, these measurements become even more difficult. Even the normally easier measurements in bulk materials, such as for the electrical conductivity and the Seebeck coefficient, can be complicated due to the small thickness of the film and the contributions from the substrate.

It is generally recognized that the thermal conductivity is the most difficult parameter to measure. Fortunately, thin film thermal conductivity measurements have drawn considerable attention over the past two decades and different methods have been developed. One popular method to measure the thermal conductivity of thin films is the 3ω method [90,91]. For thermoelectric thin films such as superlattices, there are several complications. For example, thermoelectric films are semiconductors and thus an insulating film is required between the heater and the film. The superlattice thermal conductivity is highly anisotropic. The 3ω method is typically applied to measure the cross-plane thermal conductivity by ensuring that the heater width is much larger than the film thickness. Often, an additional buffer layer exists between the film and the substrate. For Si/Ge, the buffer is graded and thus has a continuously varying thermal conductivity profile. In applying the 3ω method, there is also the contrast factor that must be considered between the film and the substrate. When the film and the substrate have close properties, more complicated modeling is needed. By careful modeling and experimental design, the 3ω method can be applied to a wide range of thin films for measuring the thermal conductivity in both in-plane and cross-plane directions [92]. Other methods such as ac calorimetry, photothermal and pump-and-probe methods have also been used to measure the thermal diffusivity of superlattices. References [10], [93–96] provide detailed reviews of existing methods. By assuming that the specific heat does not change much, which is usually a valid assumption, the thermal conductivity of the structures can be calculated from the measured diffusivity.

Although the measurement of the electrical conductivity and Seebeck coefficient is considered relatively straight-forward for bulk materials, it has turned out to be much more complicated for thin films. For transport along the thin film plane, the complications arise from the fact that most thin films are deposited on semiconductor substrates and the thermoelectric effect of the substrates can overwhelm that of the films. To circumvent these difficulties, several approaches have been taken, such as removing the substrate or growing the film on insulating layers. For example, Si/Ge superlattices are grown on silicon-on-insulator structures. Even with these precautions, there are still complications such as the existence of the buffer. Thus, differential measurements are sometimes used to subtract the influence of the buffer layer. For transport in the cross-plane direction, measurements of the electrical conductivity and Seebeck effect become much more difficult because the films are usually very thin. The recently reported ZT values between 2–3 for Bi_2Te_3/Se_2Te_3 superlattices were obtained using the transient Harman method [9]. Although the method is well established for bulk materials, the application to thin film structures requires careful consideration of various heat losses and heat generation through the leads. This is a topic that has yet to be fully addressed. In addition, the transient Harman method gives ZT rather than individual thermoelectric properties such as the Seebeck coefficient. A comparative method was recently developed to measure the Seebeck coefficient in the cross-plane direction of the superlattices [97].

5 Micro Devices and Potential Applications

Although thermoelectric coolers and power generators are intrinsically smaller than conventional coolers and power generators, on the order of a few millimeters to centimeters, there are major efforts to develop microscale thermoelectric devices for several reasons. (1) Monolithic integration is desirable for the cooling and temperature stabilization of electronic and phononic devices. (2) Heat flux that can be handled by coolers increases as the device thickness decreases. Of course, smaller device sizes are attractive for their weight and volume. Several types of thermoelectric microdevices are being developed. Figures 7 and 8 show device performance based on Si/Ge superlattices with transport in the cross-plane direction. Another approach is to use electrodeposited films [98] to make microdevices. Efforts exist in using micromachining of bulk thermoelectric materials to make microdevices [99,100]. Devices based on transport along the film plane are also being pursued, for applications in detectors and power sources [101,102].

As the device length becomes smaller, several degradation factors become important and must be addressed. These include (1) the electrical contact resistance, (2) thermal contact resistance, (3) heat sink thermal resistance in both the hot and the cold sides, and (4) additional heat leaks caused by contacts. For transport perpendicular to the film plane, the most important issues are the electrical contact resistance, the thermal boundary resistance and the spreading resistance in the hot and the cold regions. For transport along the film plane, heat leakage through the support layers normally requires the removal of the substrate. Even without a substrate, the heat leakage through the supporting membrane and the buffer layers can significantly degrade the device performance, particularly when the thermoelectric film is very thin.

5.1 Applications. Thermoelectric microdevices have some immediate applications. If the reported ZT is further confirmed and enhanced, the applications will undoubtly expand into many areas. Here, we will discuss a number of potential applications: (1) temperature stabilization, (2) high cooling density spot cooling, and (3) micropower generation.

Temperature stabilization is very important for optoelectronic devices such as laser sources, switching/routing elements, and detectors. They require careful control over their operating temperature. This is especially true in current high speed and wavelength division multiplexed (WDM) optical communication networks. Long haul optical transmission systems operating around 1.55 μm wavelength typically use erbium-doped fiber amplifiers (EDFA's), and are restricted in the wavelengths they can use due to the finite bandwidth of these amplifiers. As more channels are packed into this wavelength window, the spacing between adjacent channels

becomes smaller and wavelength drift becomes very important. Temperature variations are the primary cause for the wavelength drift, and they also affect the threshold current and output power in laser sources. Most stable sources such as distributed feedback (DFB) lasers and vertical cavity surface emitting lasers (VCSEL's) can generate large heat power densities on the order of kW/cm^2 over areas as small as 100 μm^2 [103,104]. The output power for a typical DFB laser changes by approximately 0.4 dB/°C. Typical temperature-dependent wavelength shifts for these laser sources are on the order of 0.1 nm/°C [105]. Therefore, a temperature change of only a few degrees in a WDM system with a channel spacing of 0.2–0.4 nm would be enough to switch data from one channel to the adjacent one, and even less of a temperature change could dramatically increase the crosstalk between two channels. Temperature stabilization or refrigeration is commonly performed with conventional thermoelectric (TE) coolers. However since their integration with optoelectronic devices is difficult [103,106], component cost is greatly increased because of packaging. The reliability and lifetime of packaged modules are also usually limited by their TE coolers [107]. Microdevices monolithically integrated with the functioning optoelectronic devices have advantages over separate devices in terms of their response time, size, and costs.

Many electronic and optoelectronic devices dissipate high heat flux. Conventional thermoelectric devices cannot handle large heat flux. With reduced leg length, the cooling heat flux of thermoelectric devices increases, thus providing the opportunity to handle high heat flux devices. It should be remembered, however, that more heat flux must be rejected at the hot side and must be removed using conventional heat transfer technologies such as heat pipes and high thermal conductivity heat spreaders. The active cooling method is beneficial only when the device needs to be operated below ambient temperatures or for temperature stabilization. Examples are infrared detectors and quantum cascade lasers. The speed of many electronic devices increases with reduced temperature and thus it is possible to use thermoelectric coolers to gain increased speed. Instead of cooling the whole chip, thermoelectric microcoolers can potentially be applied to handle local hot spots in semiconductor chips [108]. Regions with sizes ranging from 10's to 100's of micron in diameter have a temperature 10–30°C higher than the average chip temperature. This causes clock delays and failures in digital circuits. In addition, chip reliability due to electromigration is a thermally activated process, so the mean-free time between failures decreases exponentially as the temperature rises.

Thermoelectric devices have traditionally been used as radiation detectors such as thermopiles and can be used as power sources. With the rapid developments in MEMS, microscale power supply has been in increasing demands. Thermoelectric micro-generators can be coupled with environmental heat sources to drive sensors and microdevices for automanous operation of these devices. The body temperature powered wristwatch is a recent example [109].

6 Concluding Remarks

In this paper, we discuss recent progress in nanostructure-based solid-state energy conversion. Energy transport for both electrons and phonons can differ significantly from that in bulk materials. The nanoscale size effects can be used to improve the energy conversion efficiency. Recent studies have led to quite a large increase in *ZT* values and significant new insights into thermoelectric transport in nanostructures. There is, however, much left to be done in new materials syntheses, characterization, physical understanding, and device fabrication. This is a research area that the heat transfer community can both benefit from and contribute to. Meanwhile, we would like to emphasize that thermoelectric materials research is a multidisciplinary endeavor and requires close collaboration among researchers to address issues in materials, theory, characterization, and devices. Among these issues, heat transfer plays a significant role.

Acknowledgments

G. C. would like to acknowledge supports from NSF, DoD/ONR MURI on Thermoelectrics and DARPA HERETIC project for the work reported here. He would like to thank Professors M. S. Dresselhaus, K. L. Wang, T. Sands, and R. Gronsky, and their students for their contributions to the discussed work. A. S. acknowledges support from the Packard Foundation, NSF and DARPA HERETIC project. He would like to thank Professors J. Bowers, A. Majumdar, V. Narayanamurti, their students, and Dr. Ed. Croke for their contributions to the work presented here.

Nomenclature

C = spectral volumetric specific heat, $J\,m^{-3}Hz^{-1}K^{-1}$
D = density of states per unit volume, $J^{-1}m^{-3}$
E = energy relative to band edge, J
E_f = Fermi level relative to band edge, J
f = electron distribution function
f_{eq} = equilibrium distribution function
$\mathbf{J}$ = current density, $A\,m^{-2}$
$\mathbf{J}_q$ = heat flux, $W\,m^{-2}$
k = thermal conductivity, $W\,m^{-1}K^{-1}$
$\mathbf{k}$ = wavevector
k_B = Boltzmann constant, $J\,K^{-1}$
L_n = transport coefficients defined by Eq. (8)
q = charge per carrier, C
$\mathbf{r}$ = coordinate
S = Seebeck coefficient, $V\,K^{-1}$
t = time, s
T = temperature, K
v = velocity, $m\,s^{-1}$
Z = figure of merit, K^{-1}
$\boldsymbol{\varepsilon}$ = electrical field, V/m
Φ = electrochemical potential, J
σ = electrical conductivity $\Omega^{-1}m^{-1}$, or differential conductivity, $\Omega^{-1}m^{-1}J^{-1}$
τ = relaxation time, s
ω = angular frequency, rad . Hz

References

[1] Goldsmid, H. J., 1964, *Thermoelectric Refrigeration*, Plenum Press, New York.
[2] Rowe, D. M., ed., 1995, *Handbook of Thermoelectrics*, CRC Press, Boca Raton.
[3] Ioffe, A. F., 1957, *Semiconductor Thermoelements and Thermoelectric Cooling*, Infosearch Limited, London.
[4] Tritt, T. M., ed., 2001, "Recent Trend in Thermoelectric Materials Research," in *Semiconductors and Semimetals*, Vol. 69-Vol. 71, Academic Press, San Diego.
[5] Hicks, L. D., and Dresselhaus, M. S., 1993, "Effect of Quantum-Well Structures on the Thermoelectric Figure of Merit," Phys. Rev. B, **47**, pp. 12727–12731.
[6] Dresselhaus, M. S., Lin, Y. M., Cronin, S. B., Rabin, O., Black, M. R., Dresselhaus, G., and Koga, T., 2001, "Quantum Wells and Quantum Wires for Potential Thermoelectric Applications," in Semicond. Semimetals, **71**, pp. 1–121.
[7] Shakouri, A., and Bowers, J. E., 1997, "Heterostructure Integrated Thermionic Coolers," Appl. Phys. Lett., **71**, pp. 1234–1236.
[8] Mahan, G. D., 2001, "Thermionic Refrigeration," Semicond. Semimetals, **71**, pp. 157–174.
[9] Venkatasubramanian, R., 2001, "Phonon Blocking Electron Transmitting Superlattice Structures as Advanced Thin Film Thermoelectric Materials," Semicond. Semimetals, **71**, pp. 175–201.
[10] Chen, G., 2001, "Phonon Transport in Low-Dimensional Structures," Semicond. Semimetals, **71**, pp. 203–259.
[11] Harman, T. C., Taylor, P. J., Spears, D. L., and Walsh, M. P., 2000, "Thermoelectric Quantum-Dot Superlattices with High ZT," J. Electron. Mater., **29**, pp. L1–L4.
[12] Dubios, L., 1999, "An Introduction to the DARPA Program in Advanced Thermoelectric Materials and Devices," *Proc. Int. Conf. Thermoelectrics*, ICT'99, pp. 1–4.
[13] Aschroft, N. W., and Mermin, N. D., 1976, *Solid State Physics*, Saunders College Publishing, Fort Worth.

[14] Mahan, G. D., and Sofo, J. O., 1996, ''The Best Thermoelectric,'' Proc. Natl. Acad. Sci. U.S.A., **93**, pp. 7436–7439.
[15] Sun, X., Cronin, S. B., Liu, J. L., Wang, K. L., Koga, T., Dresselhaus, M. S., and Chen, G., 1999, ''Experimental Study of the Effect of the Quantum Well Structures on the Thermoelectric Figure of Merit in Si/SixGel-x System,'' *Proceedings of Int. Conf. Thermoelectrics*, ICT'99, pp. 652–655.
[16] Broido, D. A., and Reinecke, T. L., 1995, ''Effect of Superlattice Structure on the Thermoelectric Figure of Merit,'' Phys. Rev. B, **51**, pp. 13797–13800.
[17] Sofo, J. O., and Mahan, G. D., 1994, ''Thermoelectric Figure of Merit of Superlattices,'' Appl. Phys. Lett., **65**, pp. 2690–2692.
[18] Koga, T., Sun, X., Cronin, S. B., and Dresselhaus, M. S., 1998, ''Carrier Pocket Engineering to Design Superior Thermoelectric Materials Using GaAs/AlAs Superlattices,'' Appl. Phys. Lett., **73**, pp. 2950–2952.
[19] Koga, T., Cronin, S. B., Dresselhaus, M. S., Liu, J. L., and Wang, K. L., 2000, ''Experimental Proof-of-Principle Investigation of Enhanced $Z_{3D}T$ in (001) Oriented Si/Ge Superlattices,'' Appl. Phys. Lett., **77**, pp. 1490–1492.
[20] Chen, G., 1997, ''Size and Interface Effects on Thermal Conductivity of Superlattices and Periodic Thin-Film Structures,'' ASME J. Heat Transfer , **119**, pp. 220–229 (see also Proc. 1996 Nat. Heat Transf. Conf., ASME HTD-Vol. 323, **121**, 1996).
[21] Hicks, L. D., Harman, T. C., and Dresselhaus, M. S., 1996, ''Experimental Study of the Effect of Quantum-Well Structures on the Thermoelectric Figure of Merit,'' Phys. Rev. B, **53**, pp. 10493–10496.
[22] Harman, T. C., Taylor, P. J., Spears, D. L., and Walsh, M. P., 2000, ''Thermoelectric Quantum-Dot Superlattices With High ZT,'' J. Electron. Mater., **29**, pp. L1–L4.
[23] Whitlow, L. W., and Hirano, T., 1995, ''Superlattice Application to Thermoelectricity,'' J. Appl. Phys., **78**, pp. 5460–5466.
[24] Radtke, R. J., Ehrenreich, H., and Grein, C. H., 1999, ''Multilayer Thermoelectric Refrigeration in $Hg_{1-x}Cd_xTe$ Superlattices,'' J. Appl. Phys., **86**, pp. 3195–3198.
[25] Kumar, R., Borca-Tasciuc, D., Zeng, T., and Chen, G., 2000, ''Thermal Conductivity of Nanochanneled Alumina,'' *Proc. Int. Mech. Eng. Congress and Exhibition* (IMECE2000), ASME HTD-Vol. 366-2, pp. 393–398.
[26] Zhang, Z. B., Gekhtman, D., Dresselhaus, M. S., and Ying, J. Y., 1999, ''Processing and Characterization of Single-Crystalline Ultrafine Bismuth Nanowires,'' Chem. Mater., **11**, pp. 1659–1665.
[27] Heremans, J., Thrush, C. M., Lin, Y. M., Cronin, S., and Dresselhaus, M. S., 2000, ''Bismuth Nanowire Arrays: Synthesis and Galvanomagnetic Properties,'' Phys. Rev. B, **61**, pp. 2921–2930.
[28] Behnke, J. F., Prieto, A. L., Stacy, A. M., and Sands, T., 1999, ''Electrodeposition of $CoSb_3$ Nanowires,'' ICT'99 pp. 451–453.
[29] Hatsopoulos, G. N., and Kaye, J., 1958, ''Measured Thermal Efficiencies of a Diode Configuration of a Thermo Electron Engine,'' J. Appl. Phys., **29**, pp. 1124–1125.
[30] Mahan, G. D., 1994, ''Thermionic Refrigeration,'' J. Appl. Phys., **76**, pp. 4362–4366.
[31] Mahan, G. D., and Woods, L. M., 1998, ''Multilayer Thermionic Refrigeration,'' Phys. Rev. Lett., **80**, pp. 4016–4019.
[32] Shakouri, A., Lee, E. Y., Smith, D. L., Narayanamurti, V., and Bowers, J. E., 1998, ''Thermoelectric Effects in Submicron Heterostructure Barriers,'' Microscale Thermophys. Eng., **2**, pp. 37–42.
[33] Moyzhes, B., and Nemchinsky, V., 1998, ''Thermoelectric Figure of Merit of Metal-Semiconductor Barrier Structure based on Energy Relaxation Length,'' Appl. Phys. Lett., **73**, pp. 1895–1897.
[34] Miskovsky, N. M., and Cutler, P. H., 1999, ''Microelectronic Cooling Using the Nottingham Effect and Internal Field Emission in a Diamond (Wide-Band Gap Material) Thin-Film Device,'' Appl. Phys. Lett., **75**, pp. 2147–2149.
[35] Korotkov, A. N., and Likharev, K. K., 1999, ''Possible Cooling by Resonant Fowler-Nordheim Emission,'' Appl. Phys. Lett., **75**, pp. 2491–2493.
[36] Shakouri, A., Labounty, C., Abraham, P., Piprek, J., and Bowers, J. E., 1998, ''Enhanced Thermionic Emission Cooling in High Barrier Superlattice Heterostructures,'' Mater. Res. Soc. Symp. Proc., **545**, pp. 449–458.
[37] Shakouri, A., LaBounty, C., Piprek, J., Abraham, P., and Bowers, J. E., 1999, ''Thermionic Emission Cooling in Single Barrier Heterostructures,'' Appl. Phys. Lett., **74**, pp. 88–89.
[38] Zeng, G. H., Shakouri, A., La Bounty, C., Robinson, G., Croke, E., Abraham, P., Fan, X. F., Reese, H., and Bowers, J. E., 1999, ''SiGe Micro-Cooler,'' Electron. Lett., **35**, pp. 2146–2147.
[39] Fan, X. F., Zeng, G. H., LaBounty, C., Bowers, J. E., Croke, E., Ahn, C. C., Huxtable, S., Majumdar, A., and Shakouri, A., 2001, ''SiGeC/Si Superlattice Microcoolers,'' Appl. Phys. Lett., **78**, pp. 1580–1582.
[40] Fan, X. F., Zeng, G., Croke, E., LaBounty, C., Ahn, C. C., Vashaee, D., Shakouri, A., and Bowers, J. E., 2001, ''High Cooling Power Density SiGe/Si Micro-Coolers,'' Electron. Lett., **37**, pp. 126–127.
[41] LaBounty, C., Shakouri, A., Abraham, P., and Bowers, J. E., 2000, ''Monolithic Integration of Thin-Film Coolers with Optoelectronic Devices,'' Opt. Eng., **39**, pp. 2847–2852.
[42] LaBounty, C., Shakouri, A., and Bowers, J. E., 2001, ''Design and Characterization of Thin Film Microcoolers,'' J. Appl. Phys., **89**, pp. 4059–4064.
[43] Zeng, G., Fan, X., LaBounty, C., Bowers, J. E., Croke, E., and Shakouri, A., ''Direct Measurements of Cooling Power Density for SiGe/Si Superlattice Microcoolers,'' submitted for publication.
[44] Zeng, T. F., and Chen, G., 2000, ''Energy Conversion in Heterostructures for Thermionic Cooling,'' Microscale Thermophys. Eng., **4**, pp. 39–50.
[45] Reggiani, L., ed., 1985, *Hot-Electron Transport in Semiconductors*, Springer-Verlag.
[46] Qiu, T. Q., and Tien, C. L., 1993, ''Heat Transfer Mechanisms During Short-Pulse Laser Heating of Metals,'' ASME J. Heat Transfer , **115**, pp. 835–841.
[47] Zeng, T. F., and Chen, G., 2000, ''Nonequilibrium Nonequilibrium Electron and Phonon Transport in Heterostructures for Energy Conversion,'' *Proceedings of Int. Mech. Eng. Congress and Exhibition* (IMECE2000), ASME HTD-Vol. 366-2, pp. 361–372.
[48] Vashaee, D., and Shakouri, A., 2002, unpublished.
[49] Nolas, G. S., Slack, G. A., Morelli, D. T., Tritt, T. M., and Ehrlich, A. C., 1996, ''The Effect of Rare-Earth Filling on the Lattice Thermal Conductivity of Skutterudites,'' J. Appl. Phys., **79**, pp. 4002–4008.
[50] Casimir, H. B. G., 1938, ''Note on the Conduction of Heat in Crystals,'' Physica, **5**, pp. 495–500.
[51] Venkatasubramanian, R., 1996, ''Thin-Film Superlattice and Quantum-Well Structures—A New Approach to High-Performance Thermoelectric Materials,'' Nav. Res. Rev., **58**, pp. 44–54.
[52] Yao, T., 1987, ''Thermal Properties of AlAs/GaAs Superlattices,'' Appl. Phys. Lett., **51**, pp. 1798–1800.
[53] Chen, G., Tien, C. L., Wu, X., and Smith, J. S., 1994, ''Measurement of Thermal Diffusivity of GaAs/AlGaAs Thin-Film Structures,'' ASME J. Heat Transfer , **116**, pp. 325–331.
[54] Tien, C. L., and Chen, G., 1994, ''Challenges in Microscale Conductive and Radiative Heat Transfer,'' ASME J. Heat Transfer , **116**, pp. 799–807.
[55] Yu, X. Y., Chen, G., Verma, A., and Smith, J. S., 1995, ''Temperature Dependence of Thermophysical Properties of GaAs/AlAs Periodic Structure,'' Appl. Phys. Lett. **67**, pp. 3554–3556, p. 1303.
[56] Capsinski, W. S., and Maris, H. J., 1996, ''Thermal Conductivity of GaAs/AlAs Superlattices,'' Physica B, **220**, pp. 699–701.
[57] Lee, S. M., Cahill, D. G., and Venkatasubramanian, R., 1997, ''Thermal Conductivity of Si-Ge Superlattices,'' Appl. Phys. Lett., **70**, pp. 2957–2959.
[58] Yamasaki, I., Yamanaka, R., Mikami, M., Sonobe, H., Mori, Y., and Sasaki, T., 1998, ''Thermoelectric Properties of Bi_2Te_3 /Sb_2Te_3 Superlattice Structures,'' *Proc. Int. Conf. Thermoelectrics*, ICT'98, pp. 210–213.
[59] Capinski, W. S., Maris, H. J., Ruf, T., Cardona, M., Ploog, K., and Katzer, D. S., 1999, ''Thermal-Conductivity Measurements of GaAs/AlAs Superlattices Using a Picosecond Optical Pump-and-Probe Technique,'' Phys. Rev. B, **59**, pp. 8105–8113.
[60] Borca-Tasciuc, T., Liu, W. L., Zeng, T., Song, D. W., Moore, C. D., Chen, G., Wang, K. L., Goorsky, M. S., Radetic, T., Gronsky, R., Koga, T., and Dresselhaus, M. S., 2000, ''Thermal Conductivity of Symmetrically Strained Si/Ge Superlattices,'' Superlattices Microstruct., **28**, pp. 119–206.
[61] Song, D. W., Liu, W. L., Zeng, T., Borca-Tasciuc, T., Chen, G., Caylor, C., and Sands, T. D., 2000, ''Thermal Conductivity of Skutterudite Thin Films and Superlattices,'' Appl. Phys. Lett., **77**, pp. 3854–3856.
[62] Venkatasubramanian, R., 2000, ''Lattice Thermal Conductivity Reduction and Phonon Localizationlike Behavior in Superlattice Structures,'' Phys. Rev. B, **61**, pp. 3091–3097.
[63] Huxtable, S. T., Shakouri, A., LaBounty, C., Fan, X., Abraham, P., Chiu, Y. J., Chiu, Y. J., and Majumdar, A., 2000, ''Thermal Conductivity of Indium Phosphide-Based Superlattices,'' Microscale Thermophys. Eng., **4**, pp. 197–203.
[64] Borca-Tasciuc, T., Achimov, D., Liu, W. L., Chen, G., Ren, H.-W., Lin, C.-H., and Pei, S. S., 2001, ''Thermal Conductivity of InAs/AlSb Superlattices,'' Microscale Thermophys. Eng., **5**, pp. 225–231.
[65] Liu, W. L., Borca-Tasciuc, T., Chen, G., Liu, J. L., and Wang, K. L., 2001, ''Anisotropy Thermal Conductivity of Ge-Quantum Dot and Symmetrically Strained Si/Ge Superlattice,'' J. Nanosci. Nanotech., **1**, pp. 39–42.
[66] Narayanamurti, V., Stormer, J. L., Chin, M. A., Gossard, A. C., and Wiegmann, W., 1979, ''Selective Transmission of High-Frequency Phonons by a Superlattice: the Dielectric Phonon Filter,'' Phys. Rev. Lett., **43**, pp. 2012–2015.
[67] Ren, S. Y., and Dow, J., 1982, ''Thermal Conductivity of Superlattices,'' Phys. Rev. B, **25**, pp. 3750–3755.
[68] Chen, G., 1997, ''Size and Interface Effects on Thermal Conductivity of Superlattices and Periodic Thin-Film Structures,'' ASME J. Heat Transfer, **119**, pp. 220–229.
[69] Hyldgaard, P., and Mahan, G. D., 1996, ''Phonon Knudson Flow in Superlattices,'' Therm. Conduct., **23**, p. 172–181.
[70] Chen, G., and Neagu, M., 1997, ''Thermal Conductivity and Heat Transfer in Superlattices,'' Appl. Phys. Lett., **71**, pp. 2761–2763.
[71] Chen, G., 1998, ''Thermal Conductivity and Ballistic Phonon Transport in Cross-Plane Direction of Superlattices,'' Phys. Rev. B, **57**, pp. 14958–14973.
[72] Hyldgaard, P., and Mahan, G. D., 1997, ''Phonon Superlattice Transport,'' Phys. Rev. B, **56**, pp. 10754–10757.
[73] Tamura, S., Tanaka, Y., and Maris, H. J., 1999, ''Phonon Group Velocity and Thermal Conduction in Superlattices,'' Phys. Rev. B, **60**, pp. 2627–2630.
[74] Bies, W. E., Radtke, R. J., and Ehrenreich, H., 2000, ''Phonon Dispersion Effects and the Thermal Conductivity Reduction in GaAs/AlAs Superlattices,'' J. Appl. Phys., **88**, pp. 1498–1503.
[75] Kiselev, A. A., Kim, K. W., and Stroscio, M. A., 2000, ''Thermal Conductivity of Si/Ge Superlattices: A Realistic Model with a Diatomic Unit Cell,'' Phys. Rev. B, **62**, pp. 6896–6899.
[76] Yang, B., and Chen, G., 2001, ''Anisotropy of Heat Conduction in Superlattices,'' Microscale Thermophys. Eng., **5**, pp. 107–116.
[77] Volz, S. G., Saulnier, J. B., Chen, G., and Beauchamp, P., 2000, ''Computation of Thermal Conductivity of Si/Ge Superlattices by Molecular Dynamics Techniques,'' Microelectronics J., **31**, pp. 815–819.
[78] Liang, X. G., and Shi, B., 2000, ''Two-Dimensional Molecular Dynamics

Simulation of the Thermal Conductance of Superlattices with Lennard-Jones Potential,'' Mater. Sci. Eng., **292**, pp. 198–202.
[79] Chen, G., 1999, ''Phonon Wave Heat Conduction in Thin Films and Superlattices,'' ASME J. Heat Transfer , **121**, pp. 945–953.
[80] Balandin, A., and Wang, K. L., 1998, ''Effect of Phonon Confinement on the Thermoelectric Figure of Merit of Quantum Wells,'' J. Appl. Phys., **84**, pp. 6149–6153.
[81] Ju, Y. S., and Goodson, K. E., 1999, ''Phonon Scattering in Silicon Films with Thickness of Order 100 nm,'' Appl. Phys. Lett., **74**, pp. 3005–3007.
[82] Ziman, J. M., 1960, *Electrons and Phonons*, Clarendon Press, Oxford.
[83] Venkatasubramian, R., Siivola, E., and Colpitts, T. S., 1998, ''In-Plane Thermoelectric Properties of Freestanding Si/Ge Superlattice Structures,'' *Proc. Int. Conf. Thermoelectrics*, ICT'98, pp. 191–197.
[84] Simkin, M. V., and Mahan, G. D., 2000, ''Minimum Thermal Conductivity of Superlattices,'' Phys. Rev. Lett., **84**, pp. 927–930.
[85] Arutyunyan, L. I., Bogomolov, V. N., Kartenko, N. F., Kurdyukov, D. A., Popov, V. V., Prokof'ev, A. V., Smirnov, I. A., and Sharenkova, N. V., 1997, ''Thermal Conductivity of a New Type of Regular-Structure Nanocomposites: PbSe in Opal Pores,'' Phys. Solid State, **39**, pp. 510–514.
[86] Song, D. W., Shen, W.-N., Zeng, T., Liu, W. L., Chen, G., Dunn, B., Moore, C. D., Goorsky, M. S., Radetic, R., and Gronsky, R., 1999, ''Thermal Conductivity of Nano-Porous Bismuth Thin Films for Thermoelectric Applications,'' ASME HTD-Vol. 364-1, pp. 339–344.
[87] Volz, S. G., and Chen, G., 1999, ''Molecular Dynamics Simulation of Thermal Conductivity of Si Nanowires,'' Appl. Phys. Lett., **75**, pp. 2056–2058.
[88] Walkauskas, S. G., Broido, D. A., Kempa, K., and Reinecke, T. L., 1999, ''Lattice Thermal Conductivity of Wires,'' J. Appl. Phys., **85**, pp. 2579–2582.
[89] Kim, P., Shi, L., Majumadar, A., and McEuen, P. L., 2001, ''Thermal Transport Measurements on Individual Multi-walled Nanotubes,'' Phys. Rev. Lett., **87**, p. 215502.
[90] Cahill, D. G., 1990, ''Thermal Conductivity Measurement from 30-K to 750-K—The 3-Omega Method,'' Rev. Sci. Instrum., **61**, pp. 802–808.
[91] Lee, S. M., and Cahill, D. G., 1997, ''Heat Transport in Thin Dielectric Films,'' J. Appl. Phys., **81**, pp. 2590–2595.
[92] Borca-Tasciuc, T., Kumar, R., and Chen, G., 2001, ''Data Reduction in 3ω Method for Thin Film Thermal Conductivity Measurements,'' Rev. Sci. Instrum., **72**, pp. 2139–2147.
[93] Cahill, D. G., Fischer, H. E., Klitsner, T., Swartz, E. T., and Pohl, R. O., 1989, ''Thermal Conductivity of Thin Films: Measurement and Understanding,'' J. Vac. Sci. Technol. A, **A7**, pp. 1259–1266.
[94] Hatta, I., 1990, ''Thermal Diffusivity Measurement of Thin Films and Multilayered Composites,'' Int. J. Thermophys., **11**, pp. 293–303.
[95] Volklein, F., and Starz, T., 1997, ''Thermal Conductivity of Thin Films—Experimental Methods and Theoretical Interpretation,'' *Proc. Int. Conf. Thermoelectrics*, ICT'97, pp. 711–718.
[96] Goodson, K. E., and Ju, Y. S., 1999, ''Heat Conduction in Novel Electronic Films,'' Annu. Rev. Mater. Sci., **29**, pp. 261–293.
[97] Yang, B., Liu, J. L., Wang, K. L., and Chen, G., 2002, ''Simultaneous Measurements of Seebeck Coefficient and Thermal Conductivity Across Superlattices,'' Appl. Phys. Lett., in press.
[98] Fleurial, J.-P., Snyder, G. J., Herman, J. A., Giauque, P. H., et al. 1999, ''Thick-Film Thermoelectric Microdevices,'' *Proc. Int. Conf. Thermoelectrics*, ICT'99, pp. 294–300.
[99] Kishi, M., Yoshida, Y., Okano, H., Nemoto, H., Funanami, Y., Yamamoto, M., and Kanazawa, H., 1997, ''Fabrication of a Miniature Thermoelectric Module with Elements Composed of Sintered Bi-Te Compounds,'' *Proc. Int. Conf. Thermoelectrics*, ICT'97, pp. 653–656.
[100] Rushing, L., Shakouri, A., Abraham, P., and Bowers, J. E., 1997, ''Micro Thermoelectric Coolers for Integrated Applications,'' *Proc. Int. Conf. Thermoelectrics*, ICT'97, pp. 646–649.
[101] Volklein, F., Blumers, M., and Schmitt, L., ''Thermoelectric Microsensors and Microactuators (MEMS) Fabricated by Thin Film Technology and Micromachining,'' *Proc. Int. Conf. Thermoelectrics*, ICT'99, pp. 285–293.
[102] Yao, D.-J., Kim, C.-J., and Chen, G., 2000, ''Design of Thermoelectric Thin Film Coolers,'' ASME HTD-Vol. 366-2, pp. 245–251.
[103] Dutta, N. K., Cella, T., Brown, R. L., and Huo, D. T. C., 1985, ''Monolithically Integrated Thermoelectric Controlled Laser Diode,'' Appl. Phys. Lett., **47**, pp. 222–224.
[104] Chen, G., 1996, ''Heat Transfer in Micro- and Nanoscale Photonic Devices,'' Annu. Rev. Heat Transfer, **7**, pp. 1–57.
[105] Piprek, J., Akulova, Y. A., Babic, D. I., Coldren, L. A., and Bowers, J. E., 1998, ''Minimum Temperature Sensitivity of 1.55 μm Vertical-Cavity Lasers at 30 nm Gain Offset,'' Appl. Phys. Lett., **72**, pp. 1814–1816.
[106] Berger, P. R., Dutta, N. K., Choquette, K. D., Hasnain, G., and Chand, N., 1991, ''Monolithically Peltier-Cooled Vertical-Cavity Surface-Emitting Lasers,'' Appl. Phys. Lett., **59**, pp. 117–119.
[107] Corser, T. A., 1991, ''Qualification and Reliability of Thermoelectric Coolers for Use in Laser Modules,'' 41st Electronic Components and Technology Conference, Atlanta, GA, USA, May, pp. 150–156.
[108] Cheng, Y.-K., Tsai, C.-H., Teng, C.-C., and Kang, S.-M., 2000, *Electrothermal Analysis of VLSI Systems*, Kluwer Academic Publishers.
[109] Kishi, M., Nemoto, H., Hamao, T., Yamamoto, M., Sudou, S., Mandai, M., and Yamamoto, S., 1999, ''Micro Thermoelectric Modules and Their Application to Wristwatches as an Energy Source,'' *Proc. Int. Conf. Thermoelectrics*, ICT'99, pp. 301–307.

Heat Transfer and Phase Transformations in Laser Annealing of Thin Si Films

Seung-Jae Moon

Minghong Lee

Costas P. Grigoropoulos
e-mail: cgrigoro@me.berkeley.edu

Department of Mechanical Engineering,
Laser Thermal Laboratory,
University of California,
Berkeley, CA 94720-1740

Recrystallization of thin amorphous silicon (a-Si) films can yield polysilicon (p-Si) material with functional properties suitable for fabrication of electronic devices, including high definition large area active matrix liquid crystal displays. Pulsed laser-effected melting and recrystallization is exceptionally effective since it avoids damage to the underlying insulator structure. The ensuing phase transformations and ultimately the quality of the produced p-Si material strongly depend on the temperature history. This article presents a review of research aiming to understand the complex nucleation, resolidification and crystal growth phenomena that evolve under severely non-equilibrium conditions. It is shown that elucidation of the fundamental thermodynamic processes enables conception of novel practical thin film crystal growth techniques. [DOI: 10.1115/1.1447941]

Keywords: Crystal Growth, Heat Transfer, Laser, Solidification, Thin Films

1 Introduction

Cost-effective and low-power consuming displays including hand-held devices are ubiquitous in the modern information era. Polysilicon (p-Si) thin film transistor (TFT) displays have emerged as excellent candidates for this purpose. Compared with conventional furnace annealing, pulsed laser crystallization of amorphous silicon (a-Si) thin films is an efficient technology for obtaining high-performance p-Si TFTs suitable for the advanced flat panel display applications. In addition to faster processing time, another significant advantage over furnace annealing is that the film melting and recrystallization sequence imparts minimal thermal damage to the supporting glass substrate. In order to improve both the device performance and uniformity, high-quality p-Si films with controlled grain size and location are required. Understanding the melting/solidification process induced by laser irradiation of nanosecond duration is essential. A melt-mediated transformation scenario has been proposed [1] suggesting that the recrystallized silicon morphology is determined by complex phase transformations. The transformations sensitively depend on the level of the irradiated laser energy. To verify the evolution of these melting and resolidification phenomena, it is crucial to quantify the transient temperature field by direct measurements. Optical diagnostics are appropriate for non-intrusively monitoring the melting and recrystallization phenomena. Since the optical properties depend on temperature and state of phase, the reflectivity and transmissivity are good probing indicators of the laser annealing process. For example, analysis of time-resolved reflectivity data during the laser heating of silicon and germanium has been employed to determine the onset of melting and the melting duration [2].

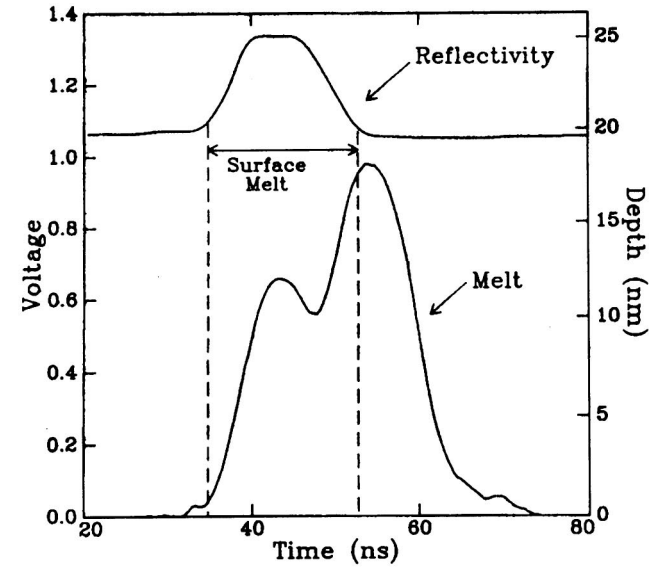
Conventional excimer laser crystallization (ELC) can produce grains of hundreds of nanometers in size depending on the a-Si film thickness. However, the processing window for the conventional technique is narrow because large grains can only be obtained in the so-called superlateral growth (SLG) regime [3], where the film is near-fully melted so that the remaining silicon solid particles on the substrate surface are thought to act as seeds for launching grain growth. In addition, the grain size produced by the conventional ELC is highly non-uniform with randomly oriented grain boundaries that result in unreliable device characteristics. Therefore, recent research efforts have been focusing on developing spatially controlled methods for laterally oriented grain growth. In broad terms, the lateral growth length is determined by the non-equilibrium conditions generated by the interfacial undercooling below the nominal melting temperature versus the competing tendency for spontaneous nucleation in the molten pool and recrystallization to fine grain material.

The thermodynamics of the complex phase transformations in thin film laser annealing is a topic of both fundamental scientific interest and practical importance. Whereas basic issues sparked the early work in the 1980s to mid 1990s, recent research is chiefly motivated by the huge technological impact in display manufacturing. In this article, the basic crystallization mechanisms, including the explosive crystallization and interface kinetics are recalled first in order to provide the necessary background. *In-situ* experiments combining time-resolved optical probes, fast pyrometry and electrical conductance measurements are presented next. Data revealing nucleation events are examined versus theoretical predictions of the relevant nucleation regimes in supercooled melts. Experimental techniques promoting lateral crystal growth are briefly reviewed next. A new double laser recrystallization experiment enhancing the lateral growth length while controlling the location of the growth is introduced. The evolution of the resolidification process that leads to lateral grain growth is visualized using high-speed laser flash photography.

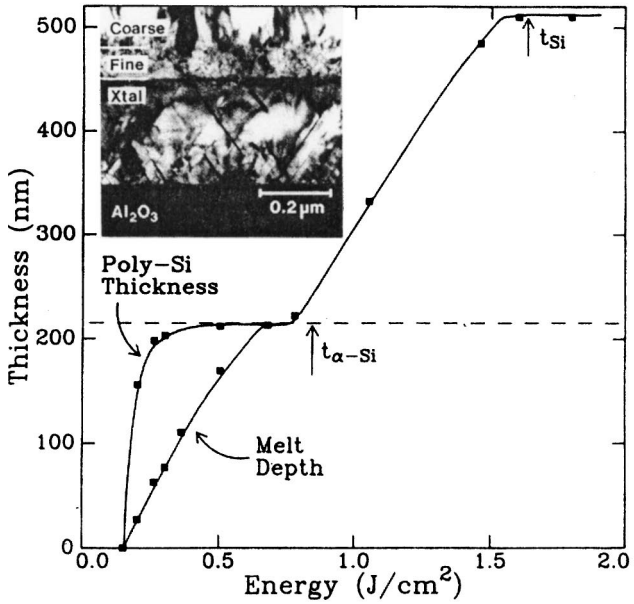
2 Crystal Growth Mechanisms

2.1 Explosive Crystallization. A variety of recrystallization phenomena occur during the pulsed laser annealing of a-Si samples. At low energy densities, fine-grain p-Si is observed. In [4], a-Si films on sapphire substrates were subjected to ruby laser pulses ($\lambda = 694$ nm) and electrical conductance transients were measured along with the reflectance response to Ar^+ laser ($\lambda = 488$ nm) probing. Figure 1(*a*) depicts transient conductance and melt-depth for a laser pulse of fluence, $F = 0.20$ J/cm^2 (the surface melt threshold is 0.15 J/cm^2) incident on a 320-nm a-Si film. Notable are (*a*) the double-peak structure of the conductance signal consisting of a lower peak followed by higher peak, indicating increased melt volume, and (*b*) the fall of the reflectivity signal at 54 ns, while the electrical conductance signal persists till almost 74 ns. Since the reflectivity is only sensitive to the first 20 nm near

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division October 10, 2001; revision received November 5, 2001. Editor: V. Dhir.



(*a*)



(*b*)

Fig. 1 (*a*) The transient conductance and surface reflectance for a 0.20 J/cm^2 pulse incident on the 320 nm amorphous film. The scale on the left axis is the observed voltage for conductance transient, with the corresponding molten thickness indicated on the right. This transient indicates a surface melt followed by an interior molten layer. (*b*) The thickness of the primary melt and the total p-Si formation as functions of the energy density on a 215 nm amorphous film. Inset: TEM cross section illustrating the formation of coarse and fine-grained p-Si at 0.5 J/cm^2. Note that the fine-grained p-Si extends to the original amorphous-crystalline interface [4]. Courtesy of Professor Michael O. Thompson.

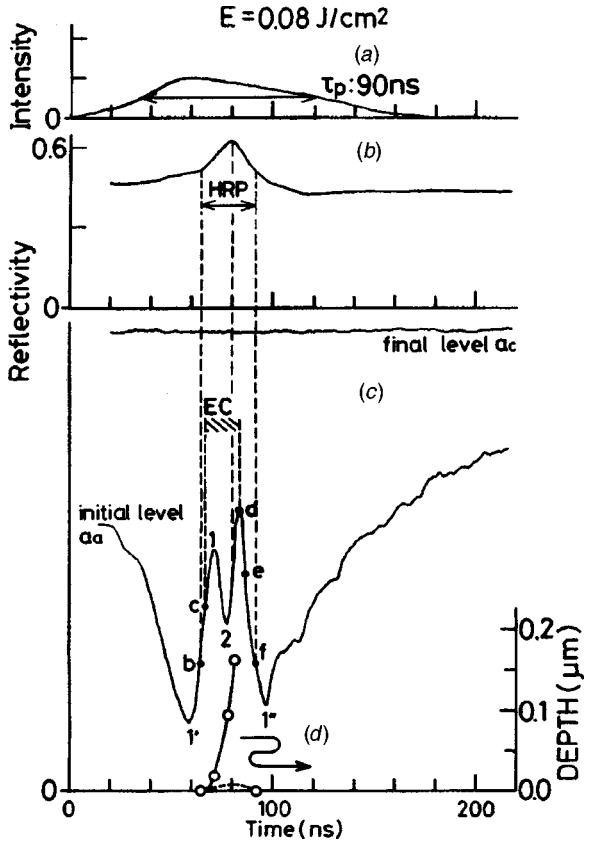


Fig. 2 (*a*) Laser pulse shape; (*b*) frontside time resolved optical reflectance (TROR) signal; (*c*) backside TROR signal; and (*d*) obtained depth of the interface between the self propagating buried ⁄-Si layer and inner a-Si. The duration of explosive crystallization (EC) is shown in (*c*) [6]. Courtesy of Professor Kuichi Murakami.

the surface, it is inferred that the second peak corresponds to melt development inside the sample. Cross-sectional transmission electron microscopy (TEM) shows that at higher energy densities, a coarse-grained p-Si layer is formed over a fine-grained layer. Figure 1(*b*) shows the primary melt depth and the total p-Si thickness observed by TEM as a function of the energy density incident on 215-nm a-Si layer. At low energy densities the p-Si thickness is significantly larger than the maximum penetration of the laser-induced melt depth. The experimental results were explained by suggesting an explosive crystallization (EC) mechanism inside the sample. At low energy densities, the laser energy melts only a thin primary liquid silicon (l-Si) layer near the surface. As this liquid solidifies to p-Si, the latent heat released from the liquid raises the temperature of the resolidified p-Si above the melting point for a-Si (presumed to be hundreds of degrees K below the equilibrium melting point for crystalline silicon) and the underlying a-Si begins to melt. This thin liquid layer is severely undercooled and can only resolidify to fine-grain material. The velocity of the explosive melt front was estimated between 10 and 20 m/s. A numerical model incorporating undercooling, interface kinetics and nucleation in heat flow calculations was constructed [5] to interpret the above outlined experimental results. The predictions captured the double-peak appearance of the transient melt volume, but the experimental results did not show the predicted oscillatory behavior.

The origin of the explosive crystallization process was examined in [6]. The samples were 600 nm crystalline silicon (c-Si) layers on sapphire, amorphized by implantation with high-energy Zn^+ ions and subjected to ruby laser pulses. Transient optical reflectance measurements were performed simultaneously from the frontside and the backside, at the 633 nm and 1152 nm HeNe laser wavelengths. Figure 2 shows the reflectance signals for the laser fluence $F = 0.08$ J/cm^2 that were interpreted using thin film optics calculations. While the front-side reflectance corresponds to a maximum surface melt depth of only 7 nm, the backside reflectance trace clearly shows interference features. Such features

would have to be generated by a l-Si/a-Si interface propagating to a maximum melt depth of 80–100 nm. Since this depth is much larger than that measured by the front reflectance of 7 nm thickness, it is reasonable to infer that the fringes are created by the interface where a buried self-propagating l-Si layer forms with the underlying a-Si. Furthermore, the onset of EC is estimated to occur at point c in Fig. 2(*c*), when the undercooled surface liquid layer is only 3 nm thick. At the time instant d when the buried l-Si layer met the interface with the underlying substrate, the front reflectance still showed evidence of surface melting. After the end of the EC process, the reflectivity drops to *e* and *f* when the thin layer solidifies. An interface velocity of about 14 m/s could be estimated for the buried interface propagation.

Interface Kinetics. The non-equilibrium solidification upon rapid laser irradiation of nanosecond and shorter duration is driven by the departure of the solid/liquid interface temperature with respect to the equilibrium melting temperature, T_m. Upon melting, the interface should in principle exhibit overheating $T_i > T_m$, while undercooling $T_i < T_m$ drives the solidification process. The fundamental nature of the phase transformation process is in essence specified by the interface response function, $v_i = v(T_i)$. Various numerical and experimental studies have been conducted to deduce the interface response function, but a definitive description has yet to be established. In the molecular dynamics study [7], the Stilliger-Weber potential was utilized to analyze the (100) solidification of c-Si. At not too high deviation from the equilibrium melting point, the slope of the response function was found to be (−9.8 K)/(m/s).

As detailed in [8], the interface response function can be analyzed via two alternative approaches. In the transition state theory (TST), the transitions between the solid and liquid phases are assumed to occur via an intermediate state, introducing a barrier to $v(T)$. The diffusion-limited theory (DLT) assumes that the interface velocity is related to the diffusivity of atoms in the liquid phase. Both descriptions result to a kinetic relation for the interface velocity of the form:

$$v(T_i) = c \exp\left(-\frac{Q}{k_B T_i}\right) \cdot \left[1 - \exp\left(-\frac{\Delta g_{ls}}{k_B T_i}\right)\right], \qquad (1)$$

where Δg_{ls} is the difference in the Gibbs free energy per atom between the liquid and the solid and Q activation energy. The kinetic prefactor is $c = f\omega_0 d$, where ω_0 is an attempt frequency, d a distance over which the interface moves for a successful jump, and f the fraction of active sites at the interface. In the collision-limited model of TST, the maximum freezing velocity is fundamentally limited by the sound velocity c_s in the solid. For c-Si, $c_s = 8433$ m/s. In DLT, the collision frequencies at the interface are presumed limited by the diffusion of atoms in the liquid. In this case, Q represents the activation energy for self-diffusion of atoms in the liquid near the interface. The attempt frequency is $\omega_0 = D_0/\lambda^2$, where λ is a characteristic distance for diffusion, and D_0 the prefactor in the equation for the diffusion constant, $D(T) = D_0 \exp(-Q/k_B T)$.

An experiment was devised to determine the interface velocities for c-Si and a-Si in solid/liquid transformations, $v_c(T)$ and $v_a(T)$, correspondingly [8]. A test sample consisting of c-Si/a-Si/c-Si layer structure was produced by ion implantation, with a 420 nm-thick amorphous layer buried underneath a 130 nm-thick c-Si layer. The specimen was subjected to irradiation by a single ruby laser ($\lambda = 694$ nm) pulse of 32 ns FWHM pulse duration. The transient optical reflectivity response was probed *in-situ* by an AlGaAs laser operating at the near-IR wavelength $\lambda = 825$ nm. Since the melting temperature of a-Si is lower than the melting temperature of c-Si, melting is initiated in the buried a-Si layer. The undercooled l-Si will crystallize at the top c-Si layer. Because of the smaller latent heat of a-Si than the latent heat of c-Si, the net transformation is exothermic, thereby promoting deeper melting of the buried a-Si via the self-sustained EC mechanism. The rapid motion of the advancing melting front creates interference of the reflected probing beam. By observing the generated fringe pattern, the EC velocity was estimated at about 15.8 m/s. Figure 3 depicts the interface response functions of c-Si and a-Si that are consistent with the constraints derived from the experimental observations. These constraints are as follows:

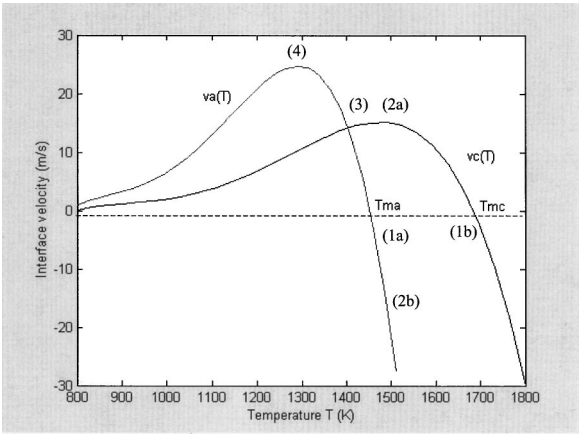


Fig. 3 Interface response function of c-Si and a-Si derived from the experimental constraints (1)–(4) defined in the text. The solid points are the experimental data used in the analysis.

1 Both v_c and v_a are equal to zero at the equilibrium melting temperature T_{mc} and T_{ma}, where T_{ma} is assumed to be about 200 K lower than T_{mc}.

2 By further assuming that the the freezing c-Si/l-Si and the melting l-Si/a-Si interfaces during expitaxial EC propagate at the same velocity, it is deduced that $v_c(T_i) = 15.8$ m/s and $v_a(T_i) = -15.8$ m/s.

3 Since the freezing of l-Si on Si(100) transforms from crystallization into amorphous growth if the interface velocity exceeds 15 m/s, the curves depicting the interface response functions must cross at 15 m/s and at temperature below T_{ma}.

4 Because the amorphization velocity in picosecond laser irradiation of c-Si saturates at 25 m/s [9], it is concluded that the a-Si interface response function attains a maximum at 25 m/s. Utilizing numerical simulations, the TST model was shown likely to be invalid and the experimental data were consistent with the DLT model [8]. The activation energy Q in the DLT model would be in the range of 0.7–1.1 eV. This activation energy is high compared to the activation energy for self-diffusion in metals, implying that the l-Si does not exhibit a purely metallic behavior. This hypothesis is supported by molecular dynamics simulations [10] suggesting that l-Si has a lower average coordination number than most liquid metals, due to persistence of covalent bonding in the liquid phase. It is also argued that the solid in the vicinity of the interface would tend to cause local ordering in the liquid, thereby increasing the barrier for self-diffusion.

Transport in Excimer Laser Crystallization of Thin Silicon Films

Figure 4 shows the experimental setup for investigating the dynamics of melting and resolidification of the a-Si thin films [11]. The transient temperature, optical properties, such as reflectivity and transmissivity, and electrical conductance signal were measured in this experiment. The same detection optics setup is placed to capture transmissivity signal on the backside of the sample. The sample consists of a 50 nm thick a-Si film deposited onto a fused quartz substrate by LPCVD. A pulsed KrF excimer laser (wavelength $\lambda = 248$ nm, FWHM = 25 ns) is used for heating the sample. Transient temperature is measured by detecting the thermal emission signal from the heated target. Emitted radiation is focused by two short focal length lenses onto a fast InGaAs

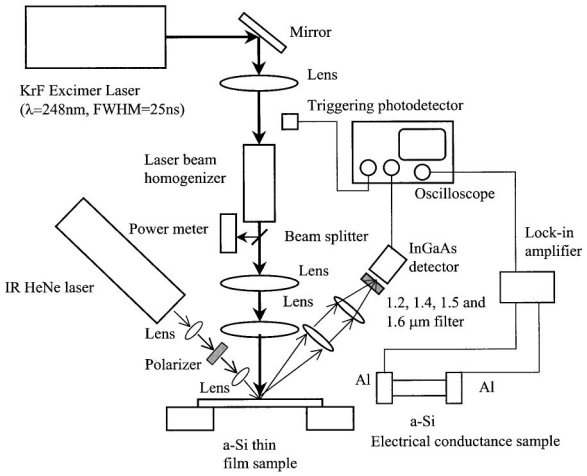


Fig. 4 The schematic diagram of excimer laser annealing and optical diagnostic setup. The IR HeNe laser is used for measuring the front reflectivity (as shown in the figure), transmissivity, emissivity and electrical conductance [11]. Reprinted by permission from AIP.

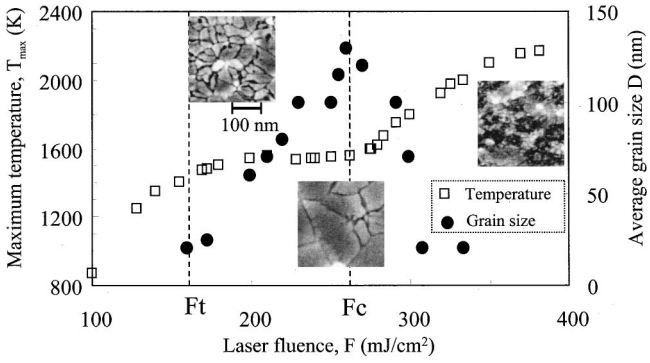


Fig. 6 Dependence of maximum temperature and average grain size on laser fluence [11]. Reprinted by permission from AIP.

photodetector with rise time of 3 ns. To enhance the accuracy of the measurement, four bandpass filters of wavelengths 1.2 μm, 1.4 μm, 1.5 μm, and 1.6 μm are employed. A focused IR HeNe laser beam ($\lambda = 1.52$ μm) is used for measuring reflectivity and transmissivity.

The temperature history of the liquid/solid phase-change process is obtained by measuring the thermal emission signals on the basis of Planck's blackbody radiation intensity distribution law. Invoking Kirchhoff's law [12], the spectral directional absorptivity is considered equal to the emissivity. The specular directional front reflectivity, ρ'_λ, and transmissivity, τ'_λ, are measured to obtain the emissivity ($\varepsilon'_\lambda = 1 - \rho'_\lambda - \tau'_\lambda$) at the $\lambda = 1.52$ μm wavelength of the IR HeNe laser. Time-resolved electrical conductance measurement [13] is applied in order to obtain the melt duration, melt depth, and the solid-liquid interface velocity. The molten Si produces an abrupt rise in the conduction electron density so that the electrical conductivity reaches values typical of liquid metals. Consequently, the total conductance of the Si is drastically increased due to the presence of a molten layer.

The maximum temperature and melt depth depend on laser fluence as shown in Fig. 5. The threshold fluence for surface melting, $F_t = 155$ mJ/cm^2 and for complete melting, $F_c = 262$ mJ/cm^2. In the partial melting regime, the melt depth increases with laser fluence. Since the absorbed laser energy in excess of the level needed for surface melting is consumed by the latent heat of

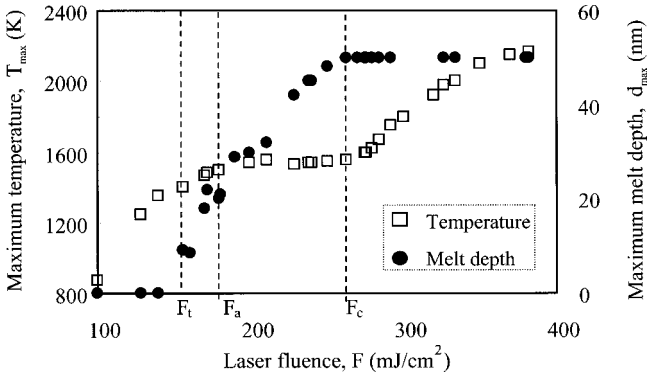


Fig. 5 Dependence of maximum temperature and melt depth on laser fluence [11]. Reprinted by permission from AIP.

phase-change from solid a-Si to l-Si, the maximum temperature remains nearly constant at about 1510 K. This is subject to the condition that the melt depth exceeds the absorption depth in l-Si, which is about 20 nm in the near-IR wavelength range. Approximately a fluence, $F_a = 179$ mJ/cm^2 is required for the maximum melt depth to reach this value. Consequently, the measured temperature rises when the fluence $F_t < F < F_a$. In the complete melting regime, $F > F_c$, the peak temperature rises with the laser energy fluence, since the excess laser energy beyond F_c is used to heat the l-Si. The average grain size, maximum temperature and SEM images belonging to each fluence regime are shown in Fig. 6. The grain size strongly depends on fluence and therefore on the temperature history. In the low fluence range corresponding to the partial melting regime, SEM shows a gradual increase in average grain size with fluence. Since the maximum temperature is still lower than the equilibrium c-Si melting temperature, crystallization may originate from unmelted silicon seeds. EC yields fine grain material underneath larger grain p-Si in the partial melting regime. In the high fluence range, which corresponds to the complete melting regime, a dramatic reversal of the micro-structural morphology is observed. This phenomenon is related to supercooling which triggers spontaneous nucleation. Substantially enlarged grain size is obtained in the 'near complete' melting regime, i.e., in the transition zone from partial to complete melting. This is explained [3] by assuming that the unmelted seeds located at the interface of the substrate may act as nucleation sites. However, the proof for this hypothesis is untenable via existing experimental tools.

As the radiant laser energy increases, the silicon layer becomes completely molten and the melting duration is prolonged. This is clearly observed in the reflectivity and transmissivity traces shown in Fig. 7. Two bumps, aligned with the melting and crystallization transitions and separated by a flat region are shown in the emissivity curve displayed in Fig. 7(*a*). The behavior of the measured optical properties can be explained by using thin film optics and invoking the effective medium theory concept [14]. The l-Si cools very rapidly, $>10^{10}$ K/s, as shown in the transient temperature signal in Fig. 7(*b*). The l-Si supercools until sufficient nucleation sites are formed. Therefore, the transient temperature exhibits a dip in the neighborhood of 60–70 ns, which exactly coincides with the end of full melting deduced by the transient conductance signal. The corresponding temperature can be considered as the nucleation temperature and is substantially lower than the melting point of c-Si by about 230 K (degree of supercooling). The temperature of the film increases due to the latent heat release during the rapid solidification triggered by the spontaneous nucleation. Following this temperature rebound, growth of the solid continues as heat is conducted into the substrate. Finally, the spontaneous nucleation results in a fine-grain structure.

P-Si samples, fabricated by XeCl excimer laser crystallization

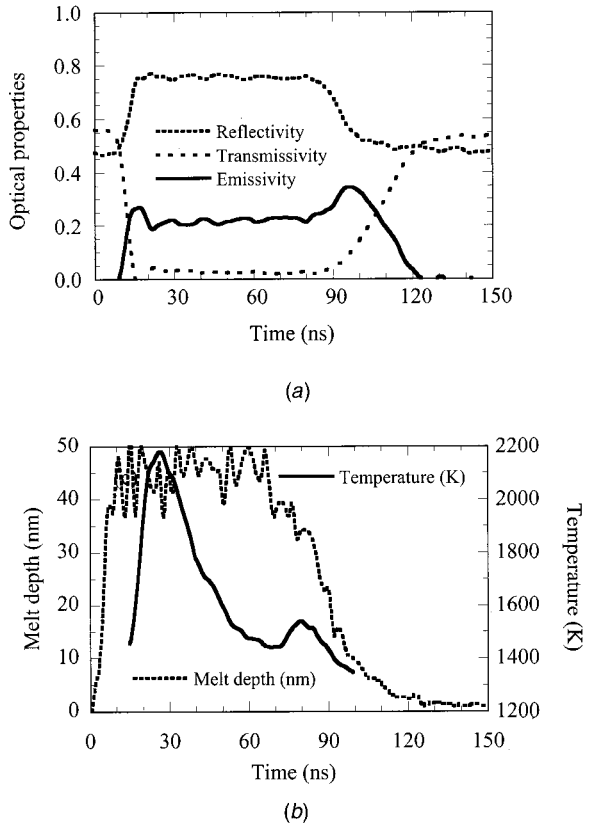


Fig. 7 (*a*) Transient front reflectivity, transmissivity, and emissivity ($\lambda = 1.52\ \mu$m) at the angle of 45 deg, (*b*) temperature and melt depth histories. The laser fluence of $F = 365$ mJ/cm^2 generates complete melting and liquid superheat [11]. Reprinted by permission from AIP.

(average grain size =120 nm), were subjected to the excimer laser irradiation. The melting behavior for initial a-Si is compared to p-Si by examining the peak temperature dependence on the irradiated laser fluence in Fig. 8. The p-Si melting temperature exhibits a plateau slightly below 1700 K. It is noted that the equilibrium melting temperature of c-Si, $T_m = 1685$ K. Hence, it is verified that a-Si melts at a temperature of about 100–150 K lower than c-Si in nanosecond laser heating. Calorimetric studies performed on thick a-Si layers prepared by high-energy implantation yielded melting temperature of 1420 K [15]. The *in-situ* study provided direct temperature measurement during the pulsed laser annealing process of a-Si and p-Si thin films.

4 Nucleation in the Supercooled Melt

The rapid quenching of the l-Si pool is inevitable due to the nanosecond time scale of a laser pulse. Because of the thinness of the a-Si film (~50 nm) and the short laser pulse, most of the heat is conducted to the substrate with minimal convection and radiation heat losses. During the timescale of a few tens of nanosecond-long laser pulse, the thermally affected zone in the poorly conducting substrate is not established. The l-Si pool tends to be supercooled as revealed in the emission temperature measurement. When the melt is quenched below its equilibrium melting temperature, the driving force for crystallization is dramatically enhanced due to the Gibbs free-energy difference, which increases with the degree of supercooling [16].

The concept of nucleation has satisfactorily explained the growth of materials in many phase transformations. Particles of a new phase are assumed to form and change in size by statistical fluctuations. Particles reaching the critical size required for continuous growth, can attain appreciable dimensions at the expense of the parent matrix and are called nuclei. Particles of sub-critical size will be called embryos in order to differentiate them from nuclei. The interface between a solid embryo and the surrounding supercooled liquid implies an activation barrier. With increasing supercooling, the Gibbs free-energy difference increases and exceeds the energy barrier so that an embryo can be transformed to a solid nucleus. The growth of embryos over the free-energy barrier is called *thermal nucleation*.

Subject to a rapid temperature transient, the distribution of embryos cannot change significantly, because of insufficient time to shrink or grow to steady-state concentration. Hence, the term *athermal nucleation* refers to the process whereby an embryo becomes a nucleus as a consequence of shrinking critical size. According to classical nucleation theory, the athermal and thermal mechanisms constitute two distinct nucleation paths through which sub-critical clusters can become supercritical [17]. It is apparent that when liquid is quenched, the temporal reduction in critical cluster size that accompanies the cooling of the liquid can lead to athermal nucleation of solid matter. The specific details of solid nucleation in supercooled liquid can influence the rates and conditions of the transformation itself, as well as determine the phase and the resulting material microstructure.

As shown in Fig. 5(*b*), the drop to a distinct temperature minimum marks the inception of nucleation in the rapidly quenched and supercooled liquid. Figure 9 shows that the nucleation temperature extracted from the transient traces is almost constant with respect to the laser fluence. It is noted that at the fluence of 224 mJ/cm^2 the melt depth is 40 nm. Since the absorption depth in the l-Si over the IR wavelength range is around 20 nm, the emission measurement of l-Si temperature is only valid when the melt depth exceeds the absorption depth. Moreover, the thermal conductivity of l-Si (~67 W/mK) is relatively high. Thus, the tem-

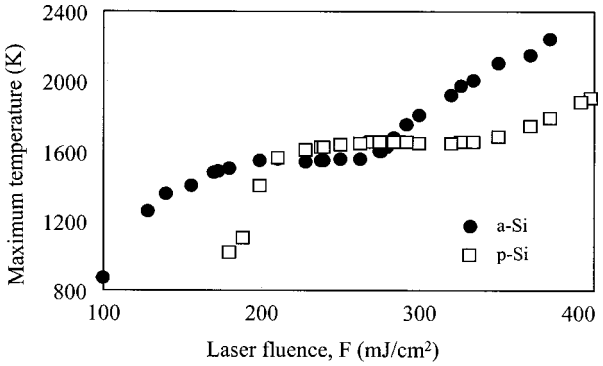


Fig. 8 Comparison of the measured maximum temperatures for a-Si and p-Si functions of the KrF excimer laser fluence [11]. Reprinted by permission from AIP.

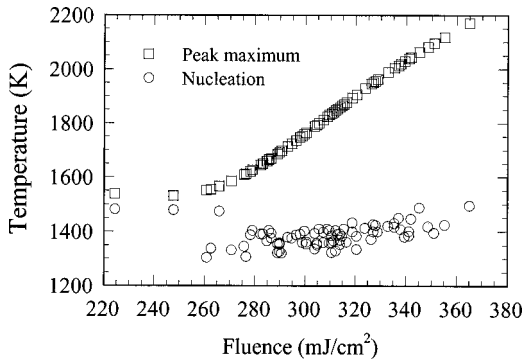


Fig. 9 Measured nucleation and peak temperatures as functions of the KrF excimer laser fluence

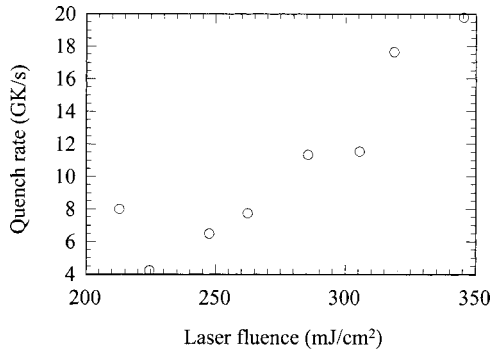


Fig. 10 Measured quench rate of the silicon melt as a function of the KrF excimer laser fluence

perature profile in the l-Si film is almost uniform over the entire liquid film. The measured data exhibit initiation of nucleation at a constant temperature level.

Figure 10 depicts the quenching rate during the cooling process. The quenching rate is herewith defined as the temporal slope of an imaginary line connecting the peak temperature with the corresponding nucleation temperature. The quenching rate increases in the full melting regime, i.e., for laser fluences greater than 262 mJ/cm^2. The total instantaneous rate describing the net rate at which supercritical clusters are produced per unit volume of supercooled liquid is expressed [18] by

$$J_{\text{total}} = -k_{n*}^{+} \left.\frac{\partial N_{n,t}}{\partial n}\right|_{n*} - N_{n*}\frac{dn*}{dt}, \tag{2}$$

where n denotes the cluster size, $N_{n,t}$ the time- and size-dependent population of solid clusters, $dn*/dt$ is the instantaneous rate of change of the critical size, k_{n*}^{+} is the forward reaction rate, N_{n*} the cluster population, and $\partial N_{n,t}/\partial n|_{n*}$ is the slope of cluster population, all evaluated at the instantaneous critical size. The first term in Eq. (2) quantifies the contribution of thermal nucleation, since it allows change in the critical cluster size. In contrast, the second term describes the instantaneous rate of formation or depletion of supercritical clusters stemming from athermal nucleation, since it does not involve growth or shrinkage of the actual critical size of the clusters. In the case of a quenched liquid, this term and hence Eq. (2) can be cast in the following form:

$$J_{\text{total}} = -k_{n*}^{+} N_{n*} \left.\frac{\partial \ln N_{n,t}}{\partial n}\right|_{n*} - N_{n*}\frac{\partial n*}{\partial T}\frac{dT}{dt}, \tag{3}$$

where dT/dt is the quenching rate. Three distinct regimes displaying the relationship between the quenching rate versus the degree of supercooling were therefore generated [18]: (i) the domain dominated by athermal nucleation, (ii) the domain dominated by the thermal mechanism, and (iii) the domain where both mechanisms overlap. The measured quenching rate in the full melting regime in Fig. 10 belongs to the athermal nucleation domain. However, in the case of partial melting, the main nucleation mechanism is thermal. In athermal nucleation, the shrinking of the nucleus critical size lowers the energy barrier. Quasi-crystalline structure tends to be formed when this happens via a rapid quenching rate. This argument could explain the formation of micro-crystalline silicon in the full melting regime. In the partial melting regime, where the initial melt depth is shallow, explosive crystallization is responsible for the formation of p-Si grains. In contrast, enhanced crystal size is produced under near-complete melting conditions via thermal nucleation.

A stochastic model for simulating nucleation of solids in supercooled liquid was constructed in [19]. Computationally, the domain was subdivided into a three-dimensional array of orthorhombic grids. Since nucleation events are independent phenomena and random birth with a well-defined mean rate, Poisson statistics were deemed to be appropriate. The probability of nucleation in a node (i,j,k) within the time interval Δt is calculated by

$$P_{ijk,t}^{\text{NUC}} = 1 - \exp\left(-\int_{t}^{t+\Delta t} \Gamma_{ijk}(\tau)\, d\tau\right), \tag{4}$$

where $\Gamma_{ijk}(\tau)$ is the mean nucleation-event frequency of the node, a function of the node temperature, composition and thermal history. At each time the probability of nucleation $P_{ijk,t}^{\text{NUC}}$, is calculated on all nodes with value ranging between 0 and 1. At the same time, a random number $R_{ijk,t}$ also in the same range from 0 to 1 is assigned to each node and compared with $P_{ijk,t}^{\text{NUC}}$. If the random number is found less than the probability of nucleation, the node is considered as containing an interface. In the simple case of steady-state homogeneous nucleation, $\Gamma_{ijk,t} \cong v_{ijk} I_o(T_{ijk,t})$ where v_{ijk} corresponds to the node volume and $I_o(T_{ijk,t})$ is the volumetric nucleation rate evaluated at the node temperature $T_{ijk,t}$. Heterogeneous nucleation is possible in nodes containing catalytic interfaces. Once nucleation is triggered within a node, its subsequent phase evolution can be computed via an interface motion algorithm. As a test case, the solidification of a 100 nm thick a-Si film on SiO_2 substrate was simulated for the laser fluence approximately 25 percent higher than the threshold for complete melting. A 3×3 μm region was divided into 100×100 nodes, and about 120 nodes were placed in the vertical direction which extended to a depth of 12 μm into the substrate. The simulation considered only homogeneous nucleation, catalyzed at the l-Si/SiO_2 interface. Even though the assumed mechanism of nucleation at the l-Si/SiO_2 interface is not precisely defined and supported by experimental evidence or physical arguments, the essence of model could be expanded to accommodate homogeneous nucleation and eventually treat complex problems of practical relevance, including lateral crystal growth.

5 Lateral Growth by Spatially Modified Irradiation

As mentioned in the introduction, recent research efforts have been focusing on developing spatially controlled crystallization methods. Several methods have been shown to produce laterally oriented grain growth. These methods include the use of beam mask [20], diffraction mask [21], anti-reflective coating [22], phase shift mask [23], and the interference effect induced by a frequency-doubled Nd:YAG laser [24]. The working principle in all these techniques relies on shaping the laser energy intensity profile that is irradiated onto an a-Si sample.

Figure 11(a) displays the experimental concept utilized for shaping the beam profile via a mask possessing a step-wise phase-shift [23]. Figure 11(b) depicts the SEM image obtained by using a phase-shift mask with $\theta = 53$ deg and mask-separation $d = 0.4$ mm. For the incoming KrF ($\lambda = 248$ nm) laser intensity $I_o = 900$ mJ/cm^2 and substrate temperature of 500°C, grains as long as 7 μm were produced in 200 nm a-Si films deposited on 850 nm thick SiO_2 on silicon.

In the laser interference experiment [24], single Nd:YAG laser pulses at $\lambda = 532$ nm were utilized to generate interference patterns on the surface of 300 nm a-Si films. The grating period of the sinusoidal patterns, P, is related to the laser wavelength λ and the angle of incidence α via the relationship $P = \lambda/2 \sin\alpha$. For appropriate laser intensities, the a-Si film melts and crystallizes only around the interference maxima. At low laser intensities the produced grain size distribution exhibits the trends typically displayed in large area annealing. At higher laser intensities, 400 mJ/cm^2 and $P = 5$ μm, lateral growth was obtained (Fig. 12). A 1 μm-wide stripe of amorphous material remains between the recrystallized stripes (region A). A narrow stripe of small grain material and width of 280 nm is distinguished in region B. The most striking feature is the growth of long grains of 1.7 μm in length extending from the boundary of region B towards the middle of the line.

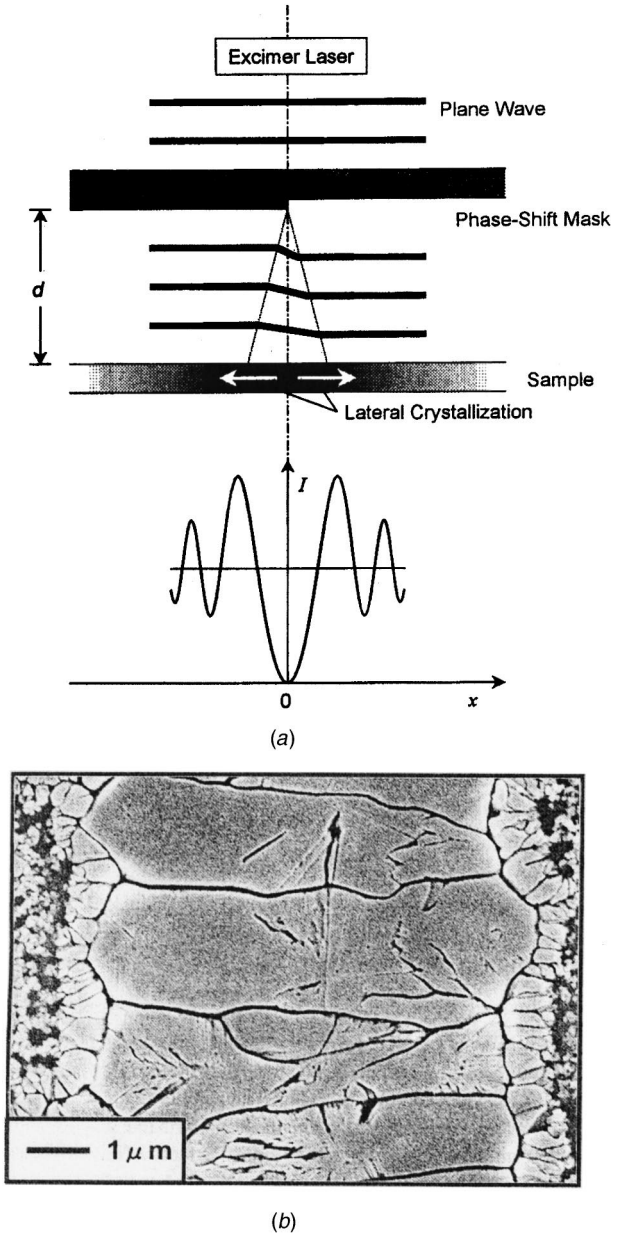


Fig. 11 (*a*) **Schematic of the proposed phase mask excimer laser crystallization (PMELC) method.** (*b*) **The top view of the crystallized Si film for the case** $d=0.4$ **mm,** $\theta=53$ **deg and** $I_o=900$ **mJ/cm^2 [23]. Courtesy of Professor Masakiyo Matsumura.**

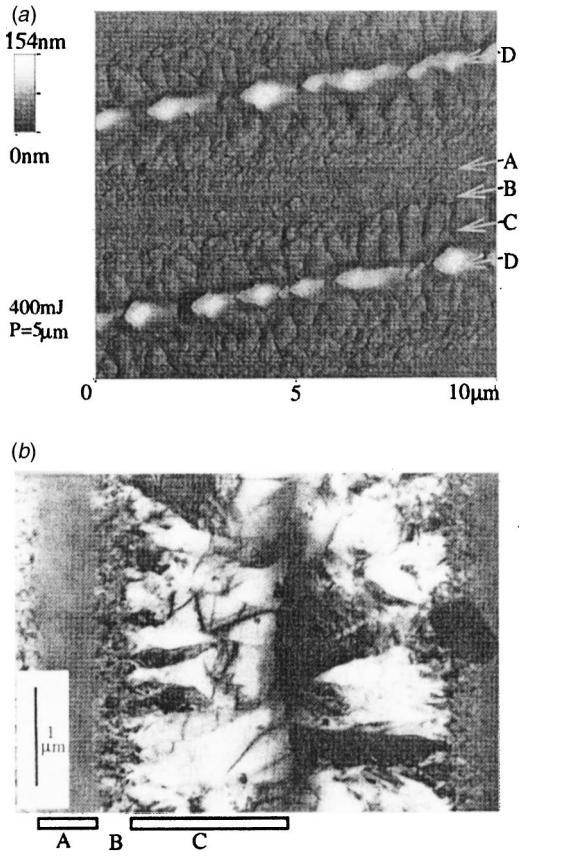


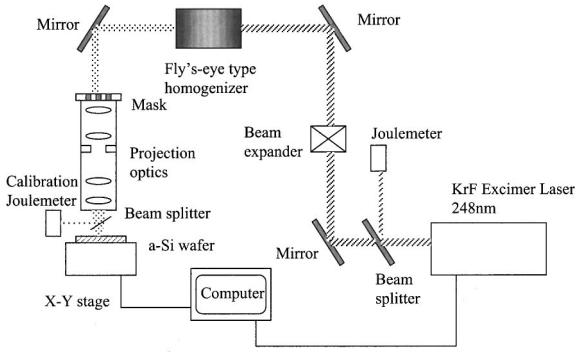
Fig. 12 (*a*) **AFM and** (*b*) **TEM micrographs of a sample fabricated by laser interference crystallization (LIC) with a pulse energy of 200 mJ/cm^2 per beam and a period of a 5** μ**m. The micrograph in** (*b*) **is an enlargement of the picture in the upper panel of Fig. 12. Fine grained material on the edges of the line (labeled B) is now flanked by long grains (labeled C), oriented towards the center of the lines and reaching lateral dimensions of almost 2** μ**m. The protrusions in the middle of the line appear flat due to the saturation of the AFM signal [24]. Courtesy of Dr. Daniel Toet.**

In order to induce lateral grain growth, a fluence gradient must be enforced such that the a-Si film is completely melted at the area exposed to higher laser fluence and partially melted at the adjacent area exposed to lower laser fluence. Under this condition, grains grow laterally towards the completely molten region. The lateral grain growth will eventually be arrested by either colliding with lateral grains grown from the other side or by spontaneous nucleation triggered in the severely supercooled molten silicon pool. Evidently, higher fluence gradients drive steeper temperature gradients. Since it takes longer time for the hotter l-Si region to cool down to the spontaneous nucleation temperature, the lateral grain growth can continue to a longer distance.

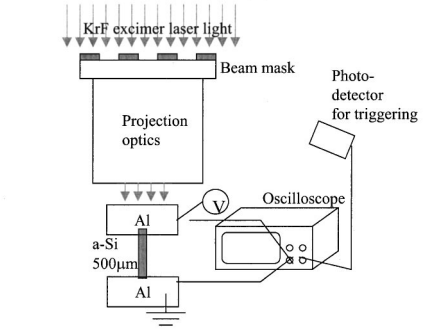
The fluence gradient and lateral growth length relationship was demonstrated in [25]. The KrF excimer laser beam (wavelength $\lambda=248$ nm) was directed into a 2:1 demagnification projection system (Fig. 13(*a*)). A beam mask consisting of 20 μm aluminum lines with 20 μm separations on the quartz substrate was placed above the mask projection assembly. In order to quantify the fluence distribution across the 20 μm line/spacing patterns, a negative photoresist was used. Time-resolved electrical conductance measurement was performed in this study in order to probe the solidification dynamics and extract the lateral solidification velocity. A schematic of the electrical conductance setup is shown in Fig. 13(*b*).

The dependence of lateral growth length on fluence gradient is shown in Fig. 14(*a*). The lateral growth length is almost constant at about 500 nm for fluence gradients below 80 mJ/cm^2μm but increases rapidly as the laser fluence gradient increases further. The directionality of the lateral grains is also improved by increasing the fluence gradient. Lateral grains of about 1.5 μm can be obtained in a 50 nm thick a-Si film by a single excimer laser pulse without any substrate heating under high fluence gradient (Fig. 14(*b*)).

Figure 15 presents a typical conductance signal obtained at high laser energy density for both masked and maskless laser irradiation. For the maskless irradiation, solidification begins at point "A" and ends in point "B," taking about 30 to 40 ns to complete. For the masked laser irradiation, a substantially longer solidification time is observed. While solidification begins at point "*a*," a change in slope on the conductance signal is noticed at point "*b*,"

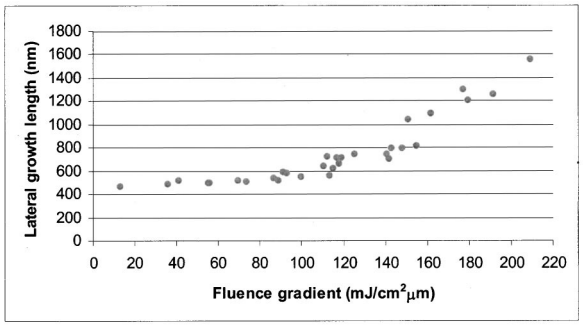


(*a*)

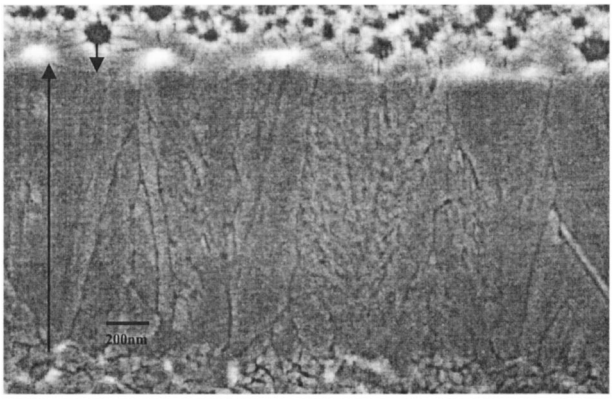


(*b*)

Fig. 13 (*a*) Projection system used for excimer laser crystallization (ELC). (*b*) Schematic of the electrical conductance setup [25]. Reprinted with permission from AIP.



(*a*)



(*b*)

Fig. 14 (*a*) Dependence of lateral growth length on fluence gradient for 50 nm a-Si film. (*b*) Lateral growth length of about 1.5 μm is obtained under high fluence gradient [25]. Reprinted with permission from AIP.

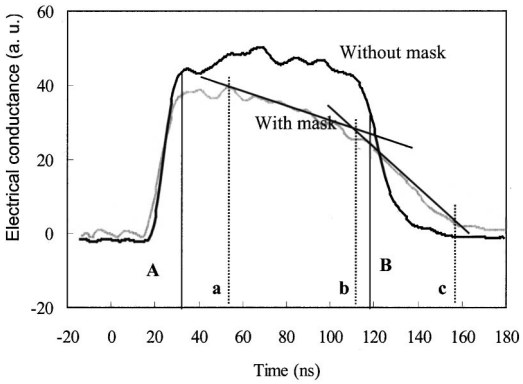


Fig. 15 Typical electrical conductance signal at high laser fluence [25]. Reprinted with permission from AIP.

and the process is concluded at point "*c*." Lateral solidification occurs between time "*a*" and "*c*." Beginning at time "*b*," spontaneous nucleation and solidification start in the bulk liquid, contributing to the increase in slope of the electrical conductance signals. Utilizing the estimated lateral solidification time, "*c-a*," and considering the lateral growth length observed from the SEM analysis, a lateral solidification velocity of about 7 m/s was obtained.

Evidently, lateral growth is limited by spontaneous nucleation in the bulk liquid. If spontaneous nucleation could be suppressed or delayed, the lateral growth would continue to a longer distance, hence producing longer lateral growth. In the case of high fluence gradient, the temperature increases rapidly from the partially molten region toward the completely molten region. Higher local temperature in the completely molten region implies correspondingly longer time to reach the deep supercooling required for spontaneous nucleation. Therefore the increase in nucleation time, i.e. the time elapsed from the beginning of the lateral growth till the inception of spontaneous nucleation, is a crucial parameter for lateral growth.

6 Ultra-large Lateral Grain Growth by Double Laser

Recrystallization of a-Si Films. Motivated by the results of the previous study, the double laser recrystallization technique was developed in order to improve grain size, uniformity and hence device performance [26]. The technique superposes a temporally modulated continuous wave (CW) Ar^+ laser ($\lambda = 514$ nm) beam with millisecond order pulse duration and a spatially homogenized KrF excimer laser pulse ($\lambda = 248$ nm, FWHM$= 25$ ns) to achieve enhanced lateral grain growth. The temporally modulated Ar^+ laser beam is released onto the sample for a few milliseconds. The absorbed Ar^+ laser energy preheats the thin film. At a specified temporal delay with respect to the Ar^+ laser pulse, a bigger-sized excimer laser beam is irradiated over the Ar^+ laser heated a-Si sample region for a few tens of nanoseconds. This process can produce ultra-large direction and location controlled p-Si grains and is insensitive to both the excimer laser fluence and the Ar^+ laser power fluctuations. Furthermore, lateral p-Si grains cannot be obtained in the non-moving sample solely with the stationary Ar^+ laser beam, even with pulse duration as long as 100 ms.

Experimental Procedures and Results. The flash laser photography experimental setup [27] is shown in Fig. 16. Since the reflectivities of a-Si (~50%), l-Si (~70%) and p-Si (~30%) are different at the illumination wavelength ($\lambda = 445$ nm) [14], the melting and resolidification sequence can be tracked. For visualizing the lateral solidification, the KrF excimer laser is replaced by a frequency-doubled Q-switched Nd:YLF solid-state laser ($\lambda = 524$ nm, FWHM $= 15$ ns), whose pulse energy is in the tens of

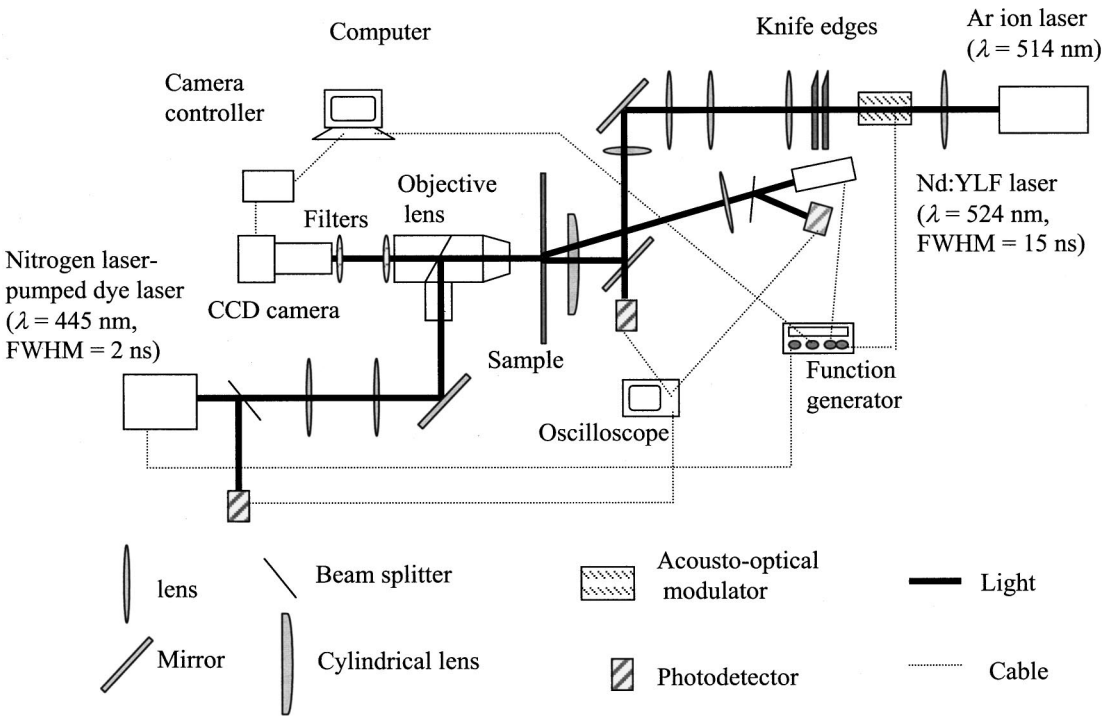


Fig. 16 A schematic of the laser flash photography experimental setup for probing the double laser recrystallization process [27]. Reprinted with permission from Elsevier Publishing.

μJ range. The compact size and the high frequency (max 10 kHz) of this laser greatly facilitate eventual integration to the manufacturing line. The Nd:YLF laser beam is delivered from the backside of the sample to accommodate the CCD camera setup and illumination by a dye laser beam. The dye laser is used as a flashlamp in the visualization experiment at the wavelength of 445 nm and pulse duration of 2 ns (FWHM).

The grain microstructure induced by the double laser recrystallization technique is shown in Fig. 17. The applied pulse duration for the modulated Ar^+ beam is 2 ms. The results shown in Fig. 17(*a*) are produced by a 938 mW Ar^+ laser pulse and 174 mJ/cm^2 excimer laser energy fluence and in Fig. 17(*b*) are processed with 957 mW Ar^+ laser pulse and 272 mJ/cm^2 excimer laser energy

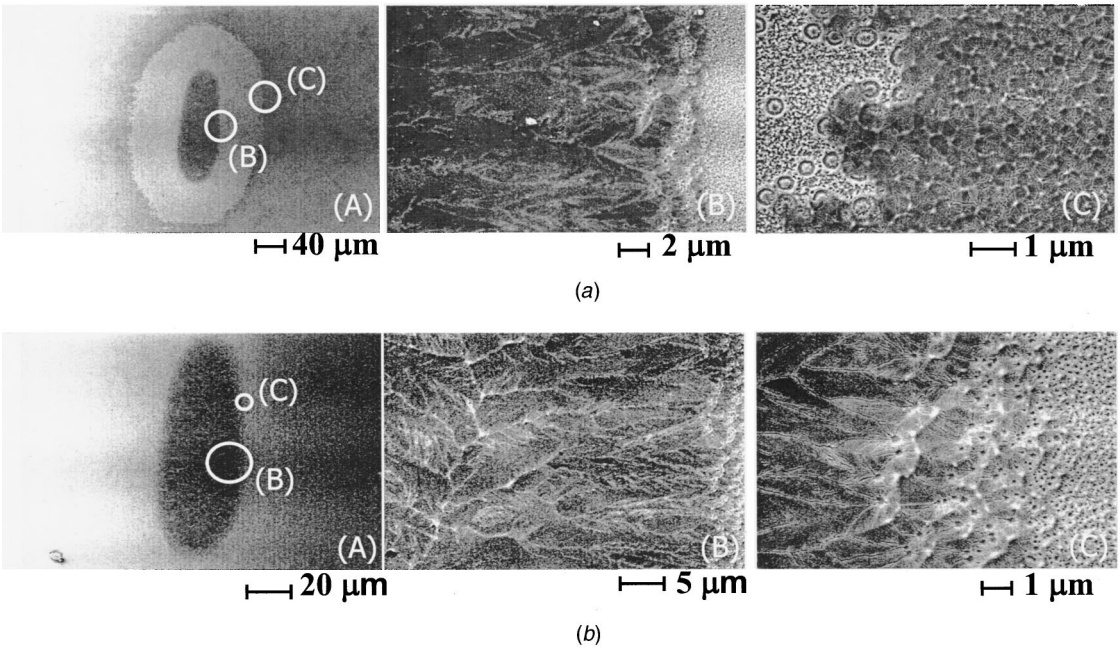


Fig. 17 Lateral grains of larger than 20 μm induced by the double laser recrystallization technique; (*a*) Ar^+ laser power, $P = 938$ mW with a 2 ms pulse, and $F = 174$ mJ/cm^2; and (*b*) Ar^+ laser power, $P = 957$ mW with a 2 ms pulse, and $F = 272$ mJ/cm^2 [26]. Reprinted with permission from Springer-Verlag.

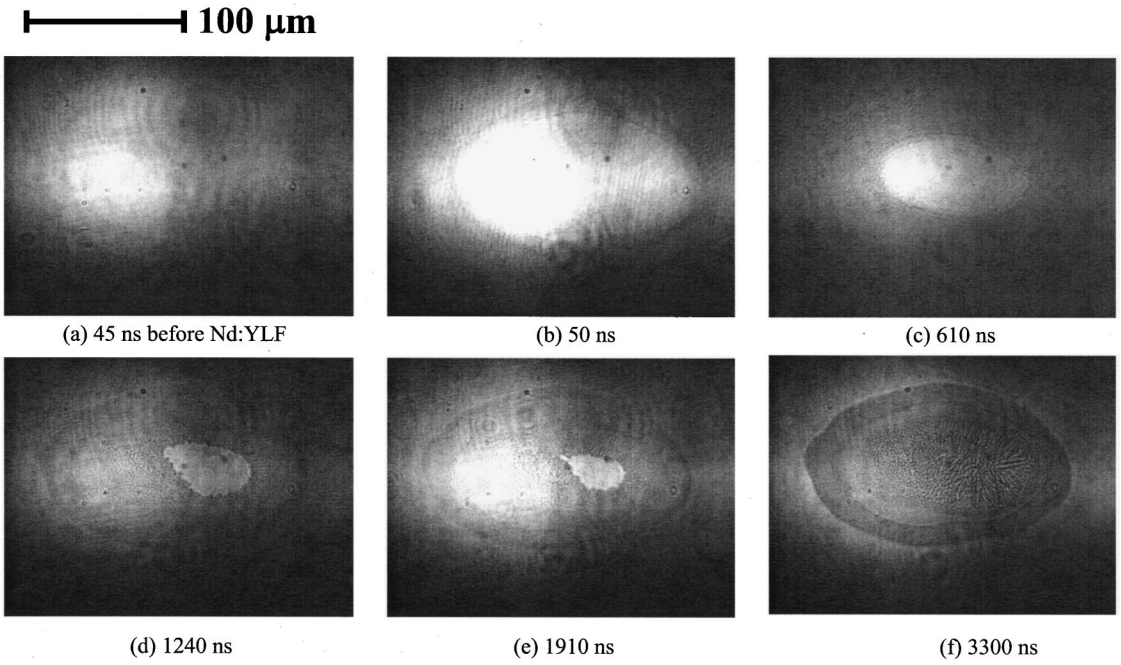


Fig. 18 Images showing the sequence of the resolidification process. The total width of an image is 200 μm [27]. Reprinted with permission from Elsevier Publishing.

fluence. The inwardly grown lateral grains are longer than 20 μm and collide across the major beam axis. Unlike in traditional ELC, the laterally grown p-Si grains are obtained regardless of the irradiated excimer laser fluence. The sequence of the resolidification process revealed by laser flash photography is shown in Fig. 18. Since the reflectivity of l-Si at the illumination wavelength ($\lambda = 445$ nm) is higher than that of a-Si, the liquid region appears bright. On the other hand, because the p-Si reflectivity is lower than that of a-Si, the resolidified p-Si appears darker. Nonetheless, at the time before the Nd:YLF laser is pulsed, the image does not show any region of enhanced brightness that is the characteristic signature of l-Si. The Ar^+ laser is irradiated for 8 ms in order to ensure the lateral solidification. A few nanoseconds following the Nd:YLF laser pulse firing, a bright l-Si region emerges, indicating that the a-Si film is then melted. The resolidification process at the initial 300 ns is carried out at a very high speed, resulting in the

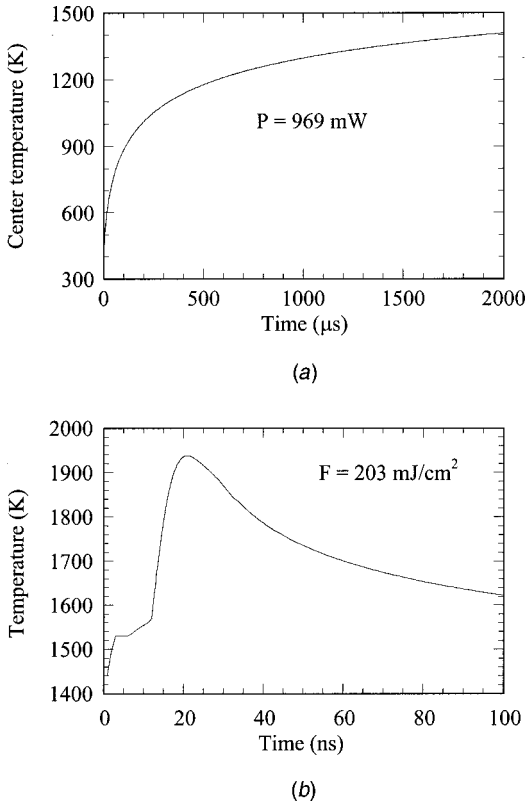


Fig. 19 The transient temperature evolution at the center point of the heated zone by (*a*) an Ar^+ laser beam. The used power is 969 mW; (*b*) an excimer laser pulse of 203 mJ/cm^2.

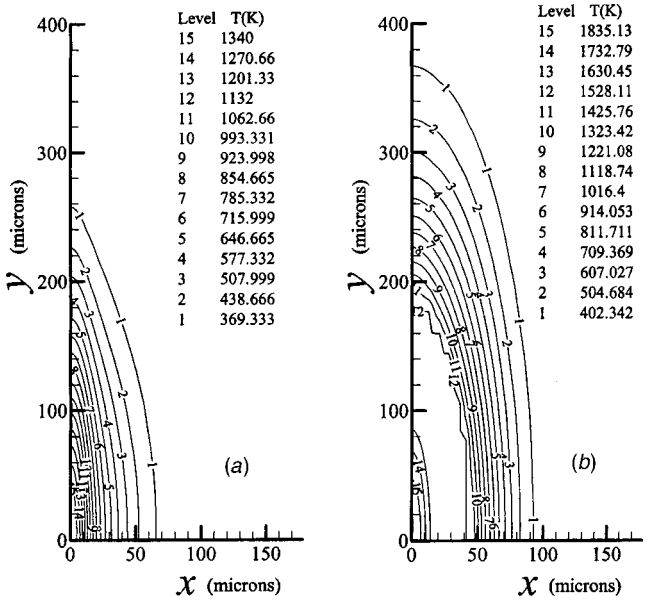


Fig. 20 The isotherms viewed from the top; (*a*) $t = 2$ ms after 2 ms Ar^+ laser pulse of 969 mW; and (*b*) $t = 21.4$ ns after the excimer laser pulse of 203 mJ/cm^2.

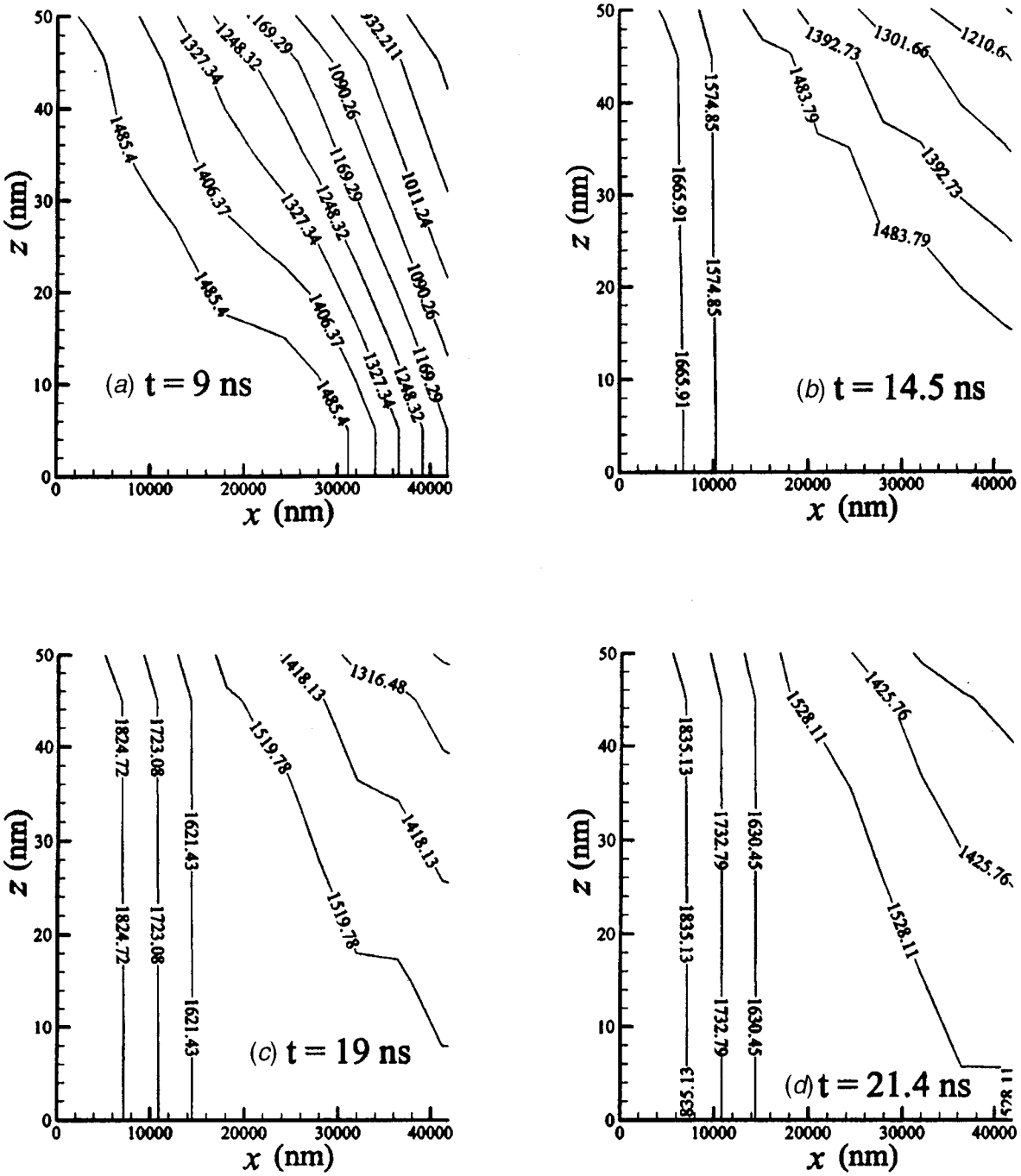


Fig. 21 Transient evolution of isotherms in the cross-section of the film; (*a*) $t=9$ ns, (*b*) $t=14.5$ ns, (*c*) $t=19$ ns, and (*d*) $t=21.4$ ns after the excimer laser pulse of 203 mJ/cm^2.

peripheral ring of microcrystalline material. After about 300 ns, the molten silicon region attains nearly the size of the final lateral growth. Beyond this time, the lateral solidification velocity is estimated to be about 10 m/s. The entire resolidification process takes about 3.3 μs to complete as compared to much shorter melt duration (<100 ns) in single beam laser recrystallization without substrate heating.

Numerical Calculation. Since the thickness of the film (50 nm) is small compared with the size of the Ar^+ laser beam (22 μm by 280 μm), and the millisecond pulse of Ar^+ laser beam drives thermal penetration to depth of tens of microns, the temperature distribution across the film thickness can be neglected. A two-dimensional conduction equation is adopted for the heat transfer in the thin film. The reflectivity, $R(T_f)$, and transmissivity, $\tau(T_f)$, are updated with thin film optics calculation. The heat diffusion to the substrate is inherently three-dimensional.

The excimer laser pulse follows immediately after the expiration of the Ar^+ laser pulse. The absorption depth, $\lambda/4\pi\kappa$, is 6 nm at the wavelength of 248 nm, while the heat diffusion length, is about 150 nm within the time scale of the excimer laser pulse, (FWHM=25 nm). Consequently, calculation in three-dimension is needed to capture the temperature variation across the film thickness: The beam size of the real excimer laser, 6 mm by 2 mm, greatly exceeds the Ar^+ laser beam dimensions of 22 μm by 280 μm. No significant error is therefore introduced by simulating the excimer laser beam with a wide Gaussian distribu-

tion for convenience. The enthalpy method [28] is used to treat the phase change of a-Si film.

Figure 19(*a*) shows that the transient temperature at the heated spot center for Ar^+ laser power of 969 mW reaches 1408 K by the end of the 2 ms-long pulse. The rapid transient upon the succeeding irradiation with the excimer laser fluence of 203 mJ/cm^2 is presented in Fig. 19(*b*). The excimer laser addition to the previously deposited Ar^+ laser energy is sufficient to completely melt the film and raise the maximum temperature to 1940 K. Figure 20(*a*) displays the top-view ($z=0$) isotherms in the thin film right at the end of Ar^+ laser pulse and Fig. 20(*b*) corresponds to $t=21.4$ ns after the firing of the excimer laser pulse. The temperature of the center of the molten zone peaks at around 21 ns, while the molten pool is of dimension comparable to the Ar^+ laser heated zone. Figure 21 shows the transient evolution of isotherms at a normal cross section through the film containing the *x*-minor axis at times elapsed with respect to the initiation of the excimer laser pulse. Isotherms virtually perpendicular to the film surfaces develop in the molten zone. Lateral isotherms would then persist in the liquid pool through the short-lived, microsecond duration solidification process. In conclusion, the size of the laterally grown p-Si is close to the extent of the region heated by the Ar^+ laser pulse. As the excimer laser-induced temperature field quickly diffuses into the substrate, nucleation is triggered in the outskirts of the heated zone when the temperature reaches the nucleation temperature. However, the Ar^+ laser-driven temperature field can maintain the liquid molten pool over the nucleation temperature, providing sufficient time for completion of the lateral growth process.

7 Conclusion

In this article, studies on the experimental probing of the melting and recrystallization of thin a-Si films with pulsed excimer laser radiation of nanosecond pulse duration have been presented. The combined optical, pyrometry and electrical conductance data allowed quantification of the temperature field and established the connection to the recrystallized p-Si morphology. Furthermore, the path of the nucleation process in the supercooled liquid was traced as a function of the quenching rate. The role of the local fluence gradient on the lateral growth was examined by applying variable beam profile distribution imposed by a mask projection system. On the basis of these findings, a new experimental technique was devised to enhance the lateral crystal growth by superposing a temporally modulated 'bias' beam and a nanosecond laser pulse. By delaying the occurrence of spontaneous nucleation, it was possible to allow sufficient time for ultra-large crystal growth that was remarkably insensitive to variations of the applied laser energy of both processing beams. Laser flash photography confirmed lateral solidification velocity of about 10 m/s.

Issues related to the role of explosive crystallization in the pulsed laser annealing of tens of nanometers-thick films clearly remain open to investigation. New experimental probes have to be developed to unveil with precision the temporal sequence of nucleation at high quenching rates and accurately quantify the microscale temperature field across the rapidly advancing solid/liquid interface. It will also be beneficial to explore nucleation and crystallization of even shorter time scales. Work on the ultra-fast laser driven phase transformations in thin films is ongoing at the Laser Thermal Laboratory. The promising, especially with the implementation of high repetition rate compact solid-state lasers, double-laser crystallization technique could be improved to achieve spatially controlled crystal growth and extended to accomplish high-throughput fabrication of devices.

References

[1] Stiffler, S. R., Thompson, M. O., and Peercy, P. S., 1988, "Supercooling and Nucleation of Silicon after Laser Melting," Phys. Rev. Lett., **60**, pp. 2519–2522.

[2] Jellison, G. E., Lowndes, D. H., Mashburn, D. N., and Wood, R. F., 1986, "Time-Resolved Reflectivity Measurements on Silicon and Germanium Using a Pulsed Excimer KrF Laser Heating Beam," Phys. Rev. B, **34**, pp. 2407–2415.

[3] Im, J. S., Kim, H. J., and Thompson, M. O., 1993, "Phase Transformation Mechanisms involved on Excimer Laser Crystallization of Amorphous Silicon Films," Appl. Phys. Lett., **63**, pp. 1969–1971.

[4] Thompson, M. O., Galvin, G. J., Mayer, J. W., Peercy, P. S., Poate, J. M., Jacobson, D. C., Cullis, A. G., and Chew, N. G., 1984, "Melting Temperature and Explosive Crystallization of Amorphous Silicon during Pulsed Laser Irradiation," Phys. Rev. Lett., **52**, pp. 2360–2363.

[5] Wood, R. F., and Geist, G. A., 1986, "Theoretical Analysis of Explosively Propagating Molten Layers in Pulsed-Laser-Irradiated A-Si," Phys. Rev. Lett., **57**, pp. 873–876.

[6] Murakami, K., Eryu, O., Takita, K., and Masuda, K., 1987, "Explosive Crystallization Starting from an Amorphous-Silicon Surface Region During Long-Pulse Laser Irradiation," Phys. Rev. Lett., **59**, pp. 2203–2206.

[7] Kluge, M. D., and Ray, J. R., 1989, "Velocity Versus Temperature Relation for Solidification and Melting of Silicon: A Molecular-Dynamics Study," Phys. Rev. B, **39**, pp. 1738–1746.

[8] Stolk, P. A., Polman, A., and Sinke, W. C., 1993, "Experimental Test of Kinetic Theories for Heterogeneous Freezing in Silicon," Phys. Rev. B, **47**, pp. 5–13.

[9] Bucksbaum, P. H., and Bokor, J., 1984, "Rapid Melting and Regrowth Velocities in Silicon Heated by Ultraviolet Picosecond Laser Pulses," Phys. Rev. Lett., **53**, pp. 182–185.

[10] Stich, I., Car, R., and Parrinnello, M., 1989, "Bonding and Disorder in Liquid Silicon," Phys. Rev. Lett., **63**, pp. 2240–2243.

[11] Hatano, M., Moon, S., Lee, M., and Grigoropoulos, C. P., 2000, "Excimer Laser-Induced Temperature Field in Melting and Resolidification of Silicon Thin Films," J. Appl. Phys., **87**, pp. 36–43.

[12] Siegel, R., and Howell, J. R., 1992, *Thermal Radiation Heat Transfer*, 3rd ed., Taylor and Francis, London.

[13] Galvin, G. J., Thompson, M. O., Mayer, J. W., Hammond, R. B., Paulter, N., and Peercy, P. S., 1982, "Measurement of the Velocity of the Crystal-Liquid Interface in Pulsed Laser Annealing of Si," Phys. Rev. Lett., **48**, pp. 33–36.

[14] Moon, S., Lee, M., Hatano, M., and Grigoropoulos, C. P., 2000, "Interpretation of Optical Diagnostics for the Analysis of Laser Crystallization of Amorphous Silicon Films," Microscale Thermophys. Eng., **4**, pp. 25–38.

[15] Donovan, E. P., Spaepen, F., Turnbull, D., Poate, J. M., and Jacobson, D. C., 1985, "Calorimetric Studies of Crystallization and Relaxation of Amorphous Si and Ge Prepared by Ion Implantation," J. Appl. Phys., **57**, pp. 1795–1804.

[16] Herlach, D. M., 1994, "Non-Equilibrium Solidification of Undercooled Metallic Melts," Mater. Sci. Eng., R., **R12**, pp. 177–272.

[17] Fisher, J. C., Hollomon, J. H., and Turnbull, D., 1948, "Nucleation," J. Appl. Phys., **19**, pp. 775–784.

[18] Im, J. S., Gupta, V. V., and Crowder, M. A., 1998, "On Determining the Relevance of Athermal Nucleation in Rapidly Quenched Liquids," Appl. Phys. Lett., **72**, pp. 662–664.

[19] Leonard, J. P., and Im, J. S., 2001, "Stochastic Modeling of Solid Nucleation in Supercooled Liquids," Appl. Phys. Lett., **78**, pp. 3454–3456.

[20] Im, J. S., Sposili, R. S., and Crowder, M. A., 1997, "Single-Crystal Si Films for Thin-Film Transistor Devices," Appl. Phys. Lett., **70**, pp. 3434–3436.

[21] Ishikawa, K., Ozawa, M., Oh, C.-H., and Matsumura, M., 1998, "Excimer-Laser-Induced Lateral Growth of Silicon Thin-Films," Jpn. J. Appl. Phys., Part 1, **37**, pp. 731–736.

[22] Ishihara, R., and Matsumura, M., 1997, "Excimer-Laser-Produced Single-Crystal Silicon Thin-Film Transistors," Jpn. J. Appl. Phys., Part 1, **36**, pp. 6167–6170.

[23] Oh, C.-H., Ozawa, M., and Matsumura, M., 1998, "A Novel Phase-Modulated Excimer-Laser Crystallization Method of Silicon Thin Films," Jpn. J. Appl. Phys., Part 1, **37**, pp. 492–495.

[24] Aichmayr, G., Toet, D., Mulato, M., Santos, P. V., Spangenberg, A., Christiansen, S., Albrecht, M., and Strunk, H. P., 1999, "Dynamics of Lateral Grain Growth during the Laser Interference Crystallization of A-Si," J. Appl. Phys., **85**, pp. 4010–4023.

[25] Lee, M., Moon, S., Hatano, M., and Grigoropoulos, C. P., 2000, "Relationship Between Fluence Gradient and Lateral Grain Growth in Spatially Controlled Excimer Laser Crystallization of Amorphous Silicon Films," J. Appl. Phys., **88**, pp. 4995.

[26] Lee, M., Moon, S., Hatano, M., and Grigoropoulos, C. P., 2001, "Ultra-Large Lateral Grain Growth by Double Laser Recrystallization of A-Si Films," Appl. Phys. A: Mater. Sci. Process., **73**, pp. 317–322.

[27] Lee, M., Moon, S., and Grigoropoulos, C. P., 2001, "In-situ Visualization of Interface Dynamics during the Double Laser Recrystallization of Amorphous Silicon Thin Film," J. Cryst. Growth, **226**, pp. 8–12.

[28] Voller, V. R., 1990, "Fast Implicit Finite-Difference Method for the Analysis of Phase Change Problems," Numer. Heat Transfer, Part B, **17**, pp. 155–169.

Molecular Dynamics Simulation of Heat Transfer and Phase Change During Laser Material Interaction

Xinwei Wang

Xianfan Xu
e-mail: xxu@ecn.purdue.edu

School of Mechanical Engineering,
Purdue University,
West Lafayette, IN 47907

In this work, heat transfer and phase change of an argon crystal irradiated by a picosecond pulsed laser are investigated using molecular dynamics simulations. The result reveals no clear interface when phase change occurs, but a transition region where the crystal structure and the liquid structure co-exist. Superheating is observed during the melting and vaporizing processes. The solid-liquid interface is found to move with a velocity of hundreds of meters per second, and the vapor is ejected from the surface with a vapor front velocity of hundreds of meters per second. [DOI: 10.1115/1.1445289]

Keywords: Ablation, Heat Transfer, Laser, Molecular Dynamics

Introduction

In recent years, ultrashort pulsed lasers have been rapidly developed and employed in materials processing. Due to the extremely short pulse duration, many difficulties exist in experimental investigation of laser material interaction, such as measuring the transient surface temperature, the velocity of the solid-liquid interface, and the material ablation rate. Ultrashort laser material interaction involves several coupled, nonlinear, and nonequilibrium processes inducing an extremely high heating rate (10^{16} K/s) and a high temperature gradient (10^{11} K/m). The continuum approach of solving the heat transfer problem becomes questionable under these extreme situations. In contrast, molecular dynamics (MD) simulation, which analyzes the movement of atoms or molecules directly, is suitable for investigating the ultrashort laser material interaction process. One aim of this work is to use MD simulation to investigate heat transfer occurring in ultrashort laser-material interaction and compares the results with those obtained with the continuum approach.

A large amount of work has been dedicated to studying laser material interaction using MD simulations. Due to the limitation of computer resources, most work was restricted to systems with a small number of atoms, thus only qualitative results such as the structural change due to heating were obtained. For instance, using quantum MD simulations, Shibahara and Kotake studied the interaction between metallic atoms and the laser beam in a system consisting of 13 atoms or less [1,2]. Their work was focused on the structural change of metallic atoms due to laser beam absorption. Häkkinen and Landman [3] studied dynamics of superheating, melting, and annealing at the Cu surface induced by laser beam irradiation using the two-step heat transfer model developed by Anisimov et al. [4]. This model describes the laser metal interaction in two steps including photon energy absorption in electrons and lattice heating through interaction with electrons. Several works in the literature investigated laser-induced ablation in various systems. Kotake and Kuroki [5] studied laser ablation of a small dielectric system consisting of 4851 atoms. Laser beam absorption was simulated by exciting the potential energy of atoms. Applying the same laser beam absorption approach, Herrmann and Campbell [6] investigated laser ablation of a silicon crystal containing approximately 23,000 atoms. Zhigileit et al. [7,8] studied laser induced ablation of organic solids using the breathing sphere model, which simulated laser irradiation by vibrational excitation of molecules. Because of the arbitrary properties chosen in the calculation, their calculation results were qualitative, and were restricted to systems with tens of thousands of atoms. Ohmura et al. [9] attempted to study laser metal interaction with the MD simulation using the Morse potential function for metals [10]. The Morse potential function simplified the potential calculation among the lattice and enabled the study of a larger system with 160,000 atoms. Heat conduction by the electron gas, which dominates heat transfer in metals, could not be predicted by the Morse potential function. Alternatively, heat conduction was simulated using the finite difference method based on the thermal conductivity of metal. Laser material interaction in a larger system was recently investigated by Etcheverry and Mesaros [11]. In their work, a crystal argon solid containing about half a million atoms was simulated. For laser induced acoustic waves, a good agreement between the MD simulation and the numerical thermoelastic calculation was obtained.

In this work, MD simulations are conducted to study laser argon interaction. The system under study has 486,000 atoms, which makes it possible to study the effect of volumetric absorption of laser energy, to define and track the solid/liquid and liquid/vapor interfaces, and to investigate whether the system is in thermal equilibrium. The number of the atoms is chosen as a compromise between the simulation accuracy and the computational time. Laser energy absorption is implemented as an increase in the kinetic energy of atoms. Laser heating of argon with different laser fluences is investigated. Laser induced heat transfer, melting, and vaporization are emphasized in this work. Phase change relevant parameters, such as the velocity of solid-liquid and liquid-vapor interfaces, ablation rate, and ablation threshold fluence are reported.

Theory of MD Simulation

In this work, argon is chosen as the material to be studied because of the physically meaningful Lennard-Jones (LJ) 12-6 potential and the less computational time required. The basic problem involves solving Newtonian equations for each atom interacting with its neighbors by means of a pairwise Lennard-Jones force:

$$m_i \frac{d^2 r_i}{dt^2} = \sum_{j \neq i} F_{ij}, \tag{1}$$

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division March 30, 2001; revision received August 15, 2001. Associate Editor: D. Poulikakos.

where m_i and r_i are the mass and position of atom i, respectively, F_{ij} is the force between atoms i and j, which is obtained from the Lennard-Jones potential as $F_{ij}=-\partial\phi_{ij}/\partial r_{ij}$. The Lennard-Jones potential ϕ_{ij} is expressed as

$$\phi_{ij}=4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12}-\left(\frac{\sigma}{r_{ij}}\right)^{6}\right], \tag{2}$$

where ε is the LJ well depth parameter, σ is the equilibrium separation parameter, and $r_{ij}=r_i-r_j$.

Many different algorithms have been developed to solve Eqs. (1) and (2), of which the Verlet algorithm is widely used due to its numerical stability, convenience, and simplicity [12]. In this calculation, the half-step leap-frog scheme is used, which is a modification to the velocity Verlet algorithm and is expressed as [12]

$$v(t+\delta t/2)=v(t-\delta t/2)+\frac{F_{ij}(t+\delta t)}{m_i}\,\delta t \tag{3a}$$

$$r_i(t+\delta t)=r_i(t)+v(t+\delta t/2)\,\delta t \tag{3b}$$

$$F_{ij}(t+\delta t)=-\frac{\partial\phi_{ij}(t+\delta t)}{\partial r_{ij}} \tag{3c}$$

$$v(t)=\frac{1}{2}(v(t+\delta t/2)+v(t-\delta t/2)). \tag{3d}$$

The criterion for choosing δt is that it should be much smaller than the typical time taken for an atom to travel its own length [12]. In this work, δt is taken as 25 fs, much smaller than the typical time (850 fs) for an argon atom to travel its own length. The distance between atoms is first compared with a potential cutoff distance r_c. Only when the distance is less than r_c is the force calculated. In this calculation, r_c is taken as 2.5σ, which is widely adopted in MD simulations using the LJ potential. At this distance, the potential is only about 1.6 percent of the well depth. The comparison of the atomic distance with r_c is organized using the cell structure and linked list method [12].

Laser energy absorption in the material is treated by scaling the velocities of all atoms in each structural cell by an appropriate factor. The amount of energy deposited in each cell is calculated assuming the laser beam is exponentially absorbed in the target. In order to prevent undesired amplification of atomic macromotion, the average velocity of atoms in each layer of the structural cells is subtracted before the velocity scaling. This laser energy absorption model de-emphasizes the details of laser material interaction, in which the quantum mechanical effect needs to be accounted for. However, the time scale for the process of laser energy absorption ($\sim$1 ps) is much smaller than the time scale considered in this work. Therefore, without knowing the details of the laser material interaction, the thermal effect can still be investigated using the current absorption model.

Results of Calculation

The target studied consists of 90 fcc unit cells in x and y directions, and 15 fcc unit cells in the z direction. Each unit cell contains 4 atoms, and the system consists of 486,000 atoms. In both x and y directions, the computational domain has a size of 48.73 nm. In the z direction, the size of the computational domain is 17.14 nm with the bottom of the target located at 4.51 nm and the top surface (the laser irradiated surface) at 12.63 nm. The schematic of the computational domain is shown in Fig. 1. The values of the parameters used in the calculation are listed in Table 1.

Thermal Equilibrium Calculation. The first step in the calculation is to initialize the system so that it is in thermal equilibrium with a minimum potential energy before laser heating. This is achieved by a thermal equilibrium calculation. Periodical boundary conditions are assumed on the surfaces in x and y directions, and free boundary conditions on the surfaces in the z direction. The target is initially constructed based on the fcc lattice

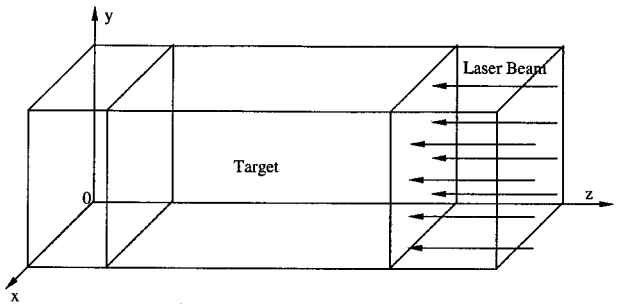


Fig. 1 Scheme of the computational domain.

structure with the (100) surface facing up. The nearest neighbor distance, r_s in the fcc lattice of argon depends on temperature T, and is initialized using the expression given by Broughton and Gilmer [13],

$$\frac{r_s}{\sigma}(T)=1.0964+0.054792\left(\frac{k_BT}{\varepsilon}\right)+0.014743\left(\frac{k_BT}{\varepsilon}\right)^2+0.083484\left(\frac{k_BT}{\varepsilon}\right)^3-0.23653\left(\frac{k_BT}{\varepsilon}\right)^4+0.25057\left(\frac{k_BT}{\varepsilon}\right)^5. \tag{4}$$

Initial velocities of atoms are specified randomly from a Gaussian distribution based on the specified temperature of 50 K using the following formula,

$$\frac{1}{2}m\sum_{i=1}^{3}v_i^2=\frac{3}{2}k_BT, \tag{5}$$

where k_B is the Boltzmann constant. During the equilibrium calculation, due to the variation of the atomic positions, the temperature of the target may change because of the exchange between the kinetic and potential energies. In order to allow the target to reach thermal equilibrium at the expected temperature, velocity scaling is necessary to adjust the temperature of the target during the early period of equilibration. The velocity scaling approach proposed by Berendsen et al. [14] is applied in this work. At each time step, velocities are scaled by a factor χ expressed as:

$$\chi=\left(1+\frac{\delta t}{t_T}\left(\frac{T}{\xi}\right)\right)^{1/2}, \tag{6}$$

Table 1 Values of the parameters used in the calculation

Parameter	Value
ε, LJ well depth parameter	1.653×10^{-21} J
σ, LJ equilibrium separation	0.3406 nm
m, Argon atomic mass	66.3×10^{-27} kg
k_B, Boltzmann constant	1.38×10^{-23} J/K
a, Lattice constant	0.5414 nm
r_c, Cut off distance	0.8515 nm
Size of the sample –x	48.726 nm
Size of the sample –y	48.726 nm
Size of the sample –z	8.121 nm
Time step	25 fs
Number of atoms	486000

where ξ is the current kinetic temperature, and t_T is a preset time constant, which is taken as 0.4 ps. This technique forces the system towards the desired temperature at a rate determined by t_T, while only slightly perturbing the forces on each atom. After scaling the velocity for 50 ps, the average temperature of the system reaches the desired value. Then the calculation is continued for another 100 ps to improve the uniformity of temperature in the target. The final equilibrium temperature of the target is 49.87 K, which is close to the desired temperature of 50 K.

At thermal equilibrium, the atomic velocity distribution should follow the Maxwellian distribution

$$P_M = 4\pi v^2 \left(\frac{m}{2\pi k_B T}\right)^{3/2} e^{-mv^2/2k_B T}, \tag{7}$$

where P_M is the probability for an atom moving with a velocity v. The velocity distribution based on the simulation results and the Maxwellian distribution are shown in Fig. 2, which indicates a good agreement between the two.

Figure 3 shows the lattice structure in the x-z plane when the system is in thermal equilibrium. For the purpose of illustration, only the atoms in the range of $0<x<12$ nm and $0<y<12.6$ nm are plotted. It is seen that atoms are located around their equilibrium positions, and the lattice structure is preserved. Another observation is that at the top and the bottom surfaces of the target, a few atoms have escaped due to the free boundary conditions.

In order to check the validity of the MD model, the thermal physical properties of argon are calculated and compared with published data. The specific heat at constant pressure (vacuum), the specific heat at constant volume, and the thermal conductivity

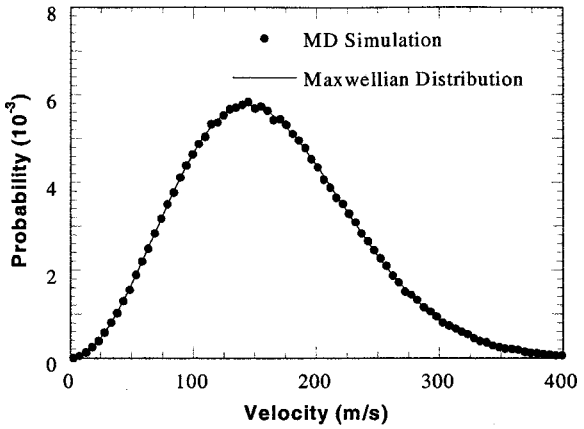


Fig. 2 Comparison of the velocity distribution by the MD simulation with the Maxwellian velocity distribution.

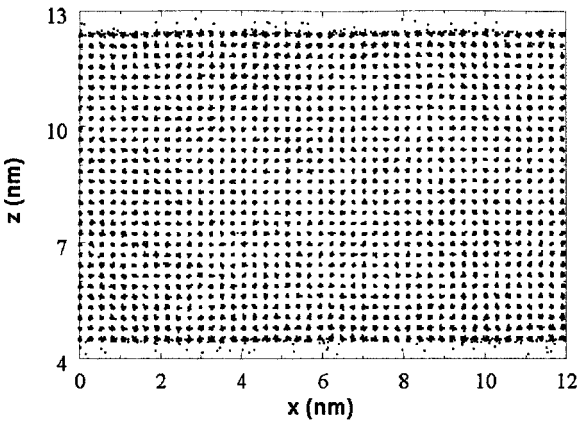


Fig. 3 Structure of the target in the x-z plane within the range of $0<x<12$ nm and $0<y<12.6$ nm.

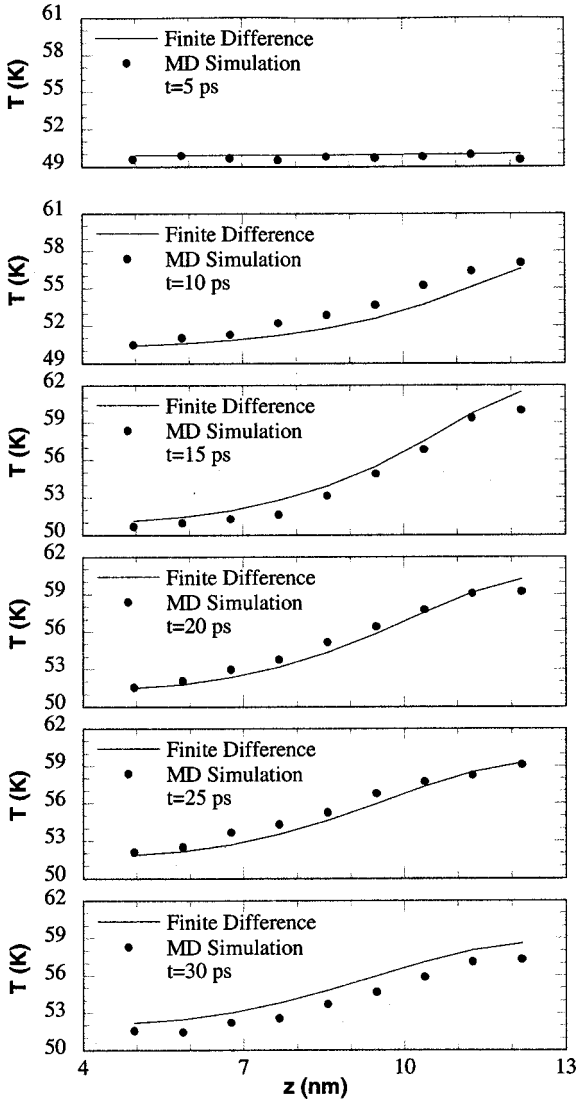


Fig. 4 Temperature distribution in the target illuminated with a laser pulse of 0.06 J/m^2.

are calculated to be 787.8 J/kg·K, 576.0 J/kg·K, and 0.304 W/m·K, while the experimental data are 637.5 J/kg·K, 543.5 J/kg·K, and 0.468 W/m·K, respectively [15]. The difference between the MD simulation results and the experimental data arises from the free boundary conditions used in MD simulations whereas the experiments were conducted under the atmospheric pressure.

Laser Material Interaction. In calculating the laser material interaction, periodical boundary conditions are used on the surfaces in x and y directions, and free boundary conditions on the surfaces in the z direction. The simulation corresponds to the problem of irradiating a block of argon in vacuum. The laser beam is uniform in space, and has a temporal Gaussian distribution with a 5 ps FWHM centered at 10 ps. The laser beam energy is absorbed exponentially in the target and is expressed as

$$\frac{dI}{dz} = -I(z)/\tau, \tag{8}$$

where I is the laser beam intensity, and τ is the characteristic absorption depth. The experimental value of τ is not available, but is expected to be intensity dependent. In the calculation, τ is chosen to be 2.5 nm.

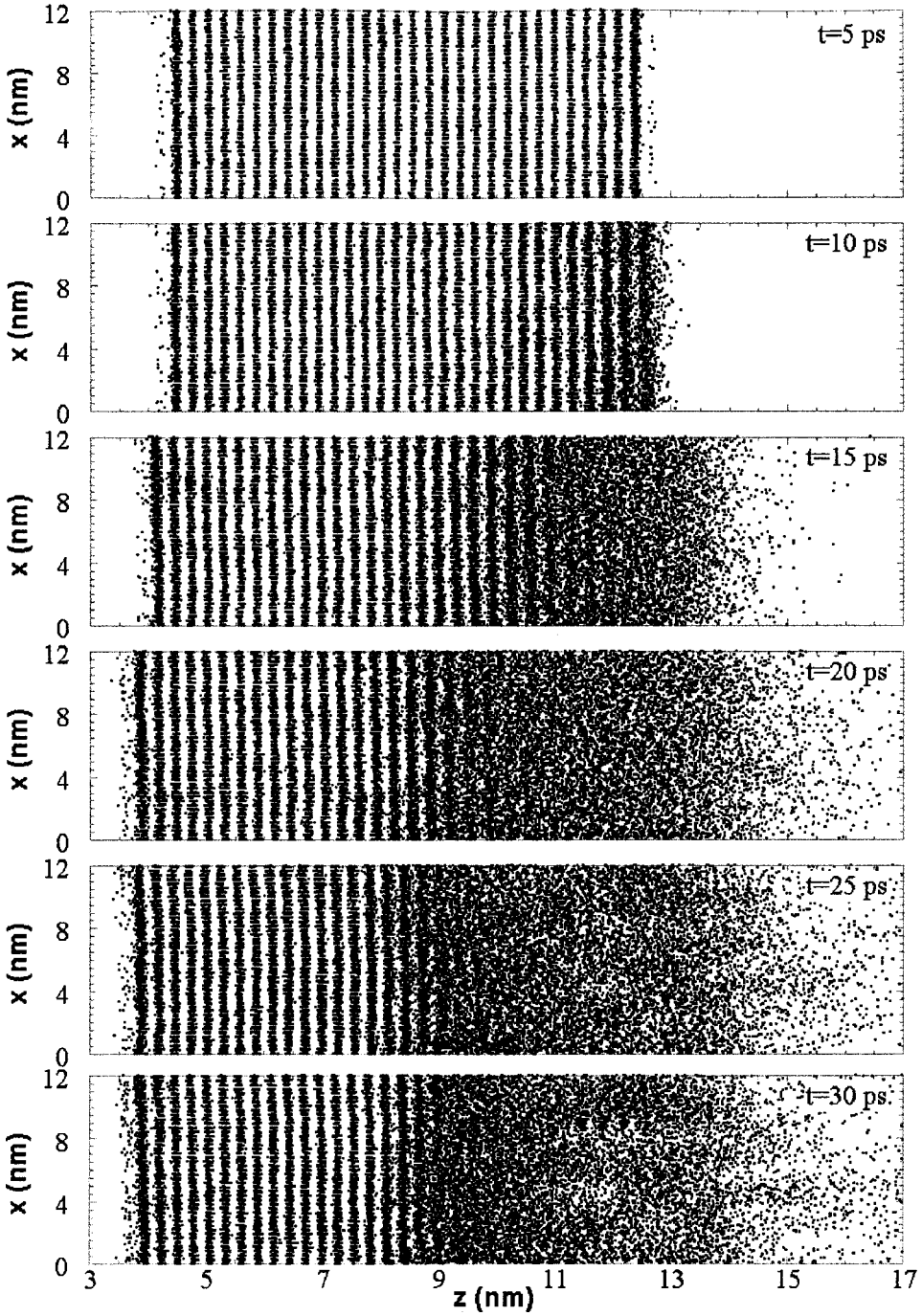


Fig. 5 Snapshots of atomic positions in argon illuminated with a laser pulse with a fluence of 0.7 J/m^2.

Laser Heating. The temperature distribution in the target illuminated with a laser pulse of 0.06 J/m^2 is first calculated and compared with the results from a finite difference calculation. No phase change occurs at this laser fluence. Figure 4 shows the temperature distribution calculated using the MD simulation and the finite difference method. In the MD simulation, the temperature at different locations is calculated as an ensemble average of a domain with a thickness of 2.5σ in the z direction. In the finite difference calculation, properties of argon obtained with the MD simulation are used in order to study the validity of the continuum approach in ps laser materials interaction, i.e., heating is very intense and localized and the temperature gradient is extremely high. It can be seen from Fig. 4 that the results obtained from the MD simulation are close to those obtained by the finite difference method. The difference between them is on the same order of the statistical uncertainty of the MD simulation. In other words, the continuum approach is still capable of predicting the heating process induced by a picosecond laser pulse. It is also noticed that no thermal wave is obtained in the MD simulation, therefore the non-Fourier effect [16] is insignificant. However, the thermal relaxation time of argon at the temperature range studied in this work is small ($\sim$1 ps). The non-Fourier effect would not be observed using the continuum approach either.

Laser Induced Phase Change. In this section, various phenomena accompanying phase change in an argon target illumi-

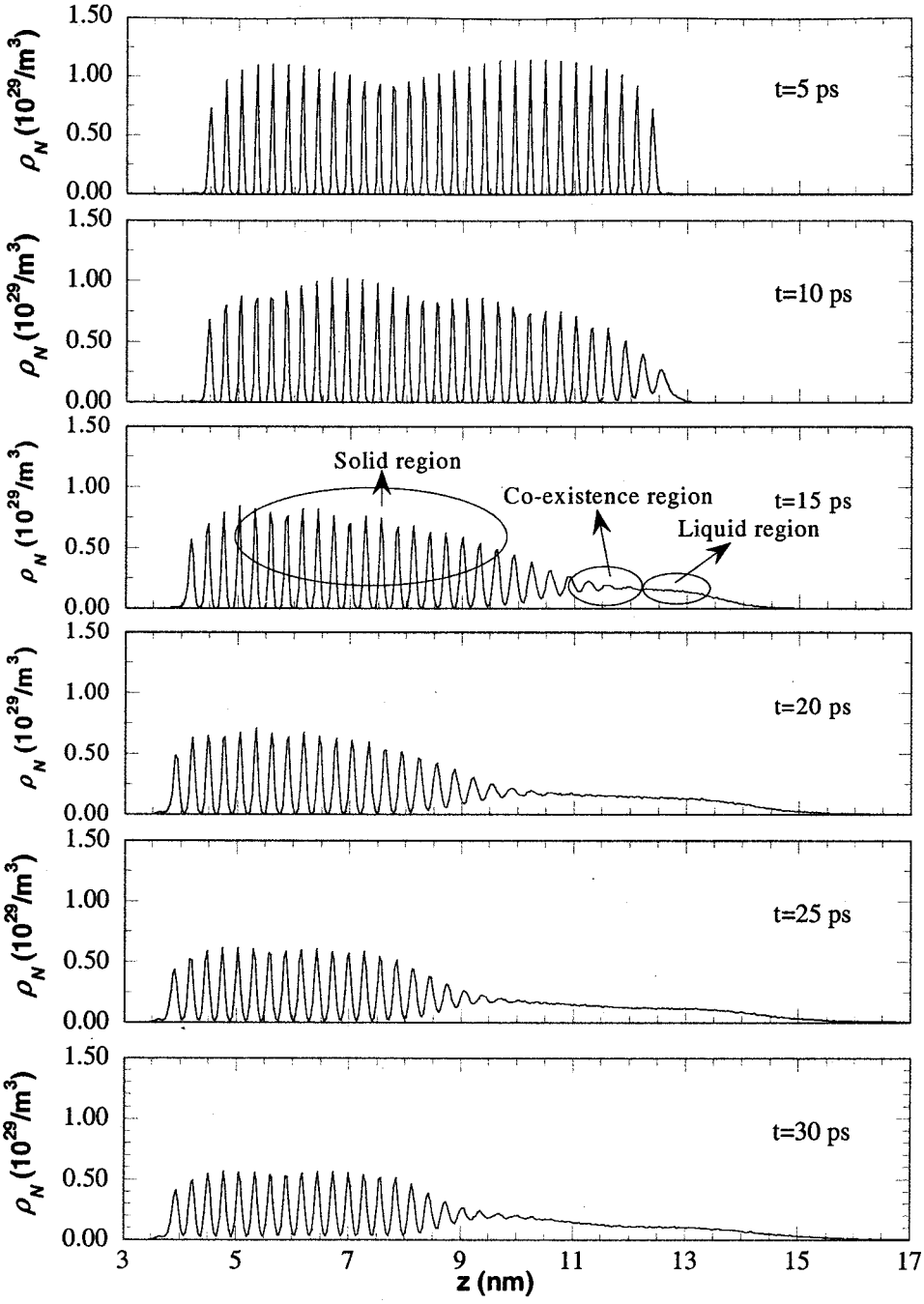


Fig. 6 Distribution of the number density of atoms at different times in argon illuminated with a laser pulse of 0.7 J/m^2.

nated with a laser pulse of 0.7 J/m^2 are investigated in detail. The threshold fluence for ablation is also studied. Towards the end of the laser pulse, the laser beam absorption is considered in both the solid and non-solid parts. The laser beam absorption depth is adjusted based on the local material density ρ_{local}, and is calculated as $\tau\rho/\rho_{\text{local}}$, with ρ denoting the density of the solid before laser heating.

A series of snapshots of atomic positions at different times is shown in Fig. 5. It is seen that until 10 ps, the lattice structure is preserved in the target. At about 10 ps, melting starts, and the lattice structure is destroyed in the melted region and is replaced by a random atomic distribution. After 20 ps, the solid liquid interface stops moving into the target, and vaporized atoms are clearly seen. Figure 6 shows the spatial distribution of the number density of atoms, ρ_N, at different times, which further illustrates the structural change during laser heating. At the early stage of laser heating, the crystal structure is preserved in the target, which is seen as the peak number density of atoms at each lattice layer. Due to the increase of atomic kinetic energy in laser heating, atoms vibrate more intensely in the solid region, resulting in a lower peak of the number density of atoms and a wider distribution. As laser heating progresses, the target is melted from its front surface, and the atomic distribution becomes random. Therefore, the number density of atoms becomes uniform over the melted region. However, no clear interface is observed between the solid and the liquid. Instead, structures of solid and liquid co-exist within a certain range, which is shown as the co-existence of the peak and the high base of the number density of atoms. Vaporiza-

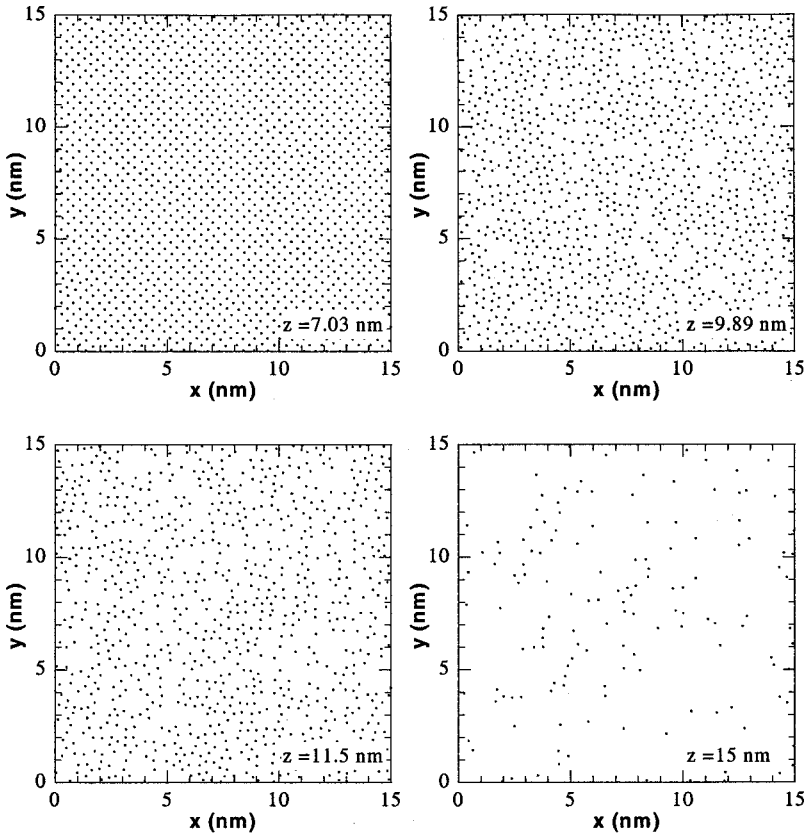


Fig. 7 In-plane structures reflected by the atomic positions at different z locations at 20 ps.

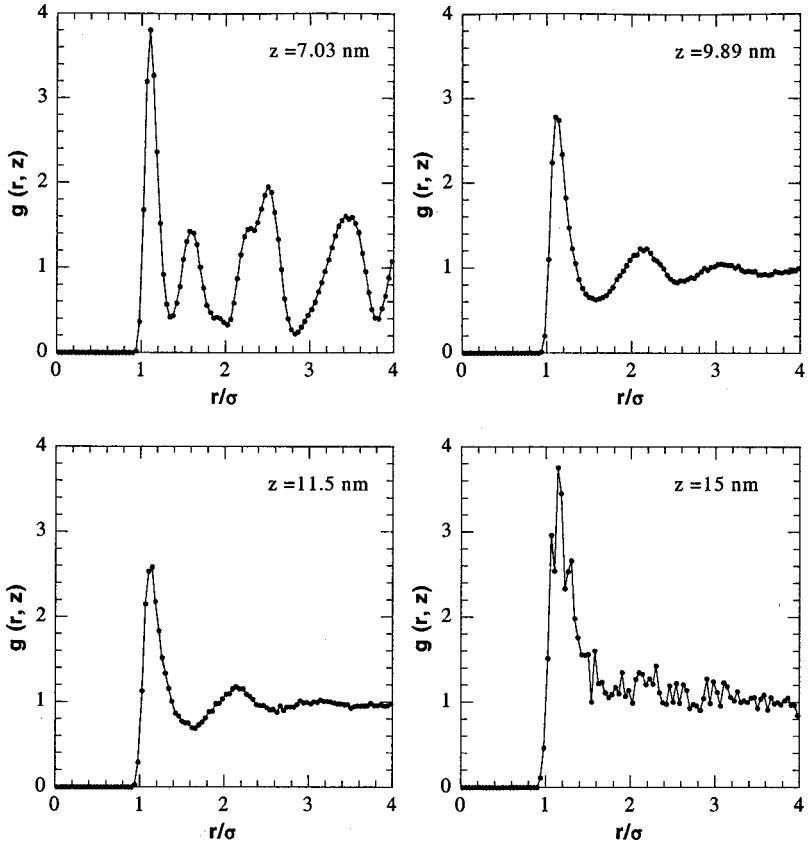


Fig. 8 In-plane radial distribution functions at different z locations at 20 ps.

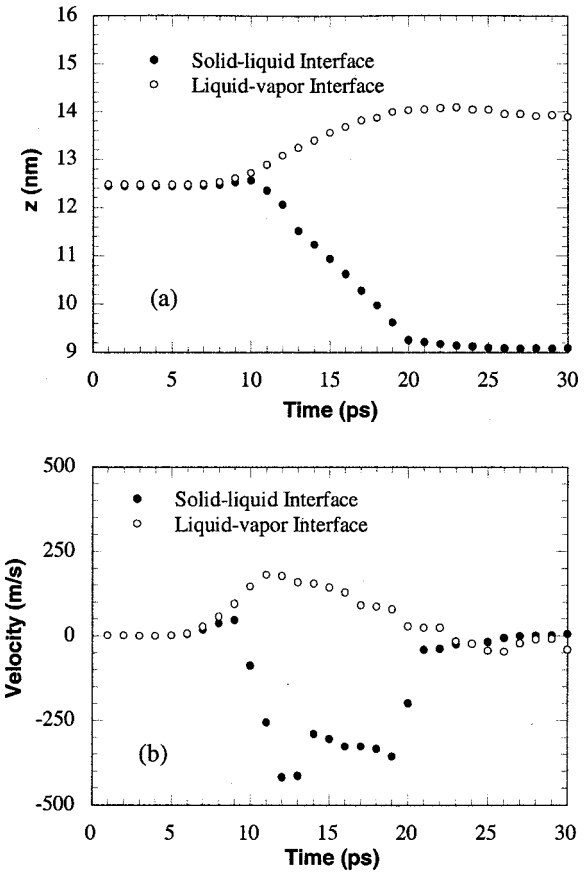


Fig. 9 (*a*) Positions, and (*b*) velocities of the solid-liquid interface and the liquid-vapor interface in argon illuminated with a laser pulse of 0.7 J/m^2.

tion happens at the surface of the target, which reduces the number density of atoms significantly at the location near the liquid surface.

In addition to the structure variation in the z direction as discussed above, the in-plane structure also reflects the phase change process. Figure 7 shows the atomic positions within a 0.27 nm thick plane centered at various z locations at 20 ps. For the purpose of illustration, only the area within $0<x<15$ nm and $0<y<15$ nm is plotted. In the solid region ($z=7.03$ nm), atoms are located regularly in space, reflecting the crystal lattice structure. In the transient region ($z=9.89$ nm), the lattice structure still can be found in small localized regions while the whole atomic distribution looks random. A similar structure is found in the liquid region ($z=11.5$ nm). However, localized regions with the short order structure are much fewer than those at $z=9.89$ nm. In the vapor region ($z=15$ nm), fewer atoms are found with no apparent structures.

One of the parameters describing the in-plane structure is the in-plane radial distribution function $g(r)$ [12,17,18]. It is a ratio of the number of atoms at a distance r from a given atom compared with the number of atoms at the same distance in an ideal gas with the same density. Figure 8 shows $g(r,z)$ in the same planes discussed in Fig. 7. From Fig. 8, structures of different phases can be quantified. At the location of 7.03 nm, both the short-range and long-range orders are preserved, showing the feature of the crystalline structure. At the location of 9.89 nm, the short-range order (the second peak) still exists while the long-range order becomes much less apparent, showing the co-existence of the solid and the liquid. At 11.5 nm, the material only has a short-range order, reflecting a liquid-like structure. The short-range order disappears at the location of 15 nm, indicating a gas-like structure.

In order to find out the rate of melting and evaporation, criteria are needed to determine the solid-liquid and liquid-vapor interfaces. Both the profile of the number of atoms [3] and the radial

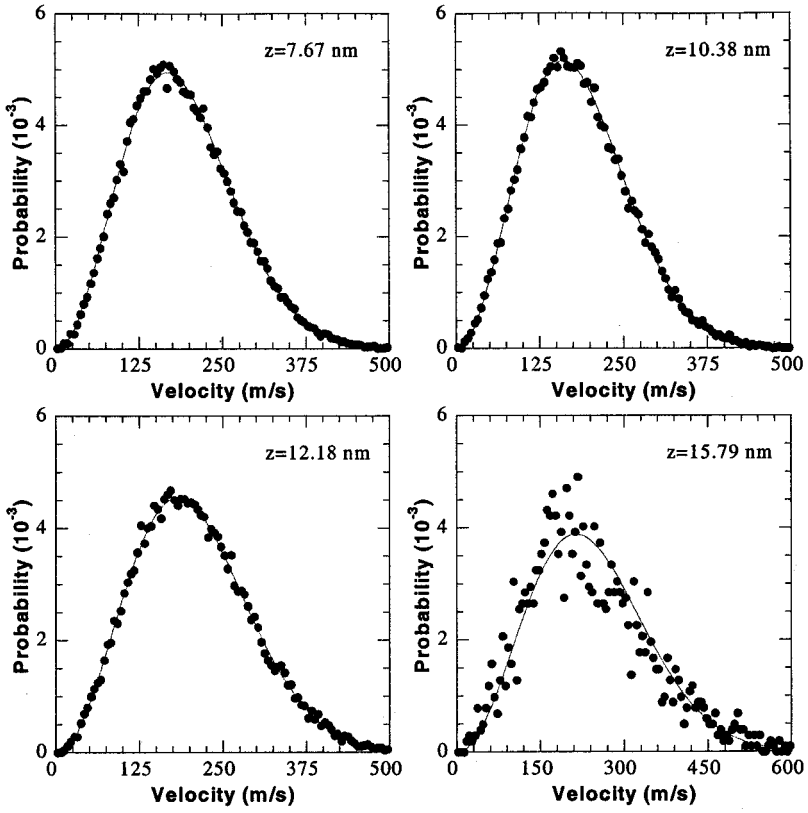


Fig. 10 Velocity distributions at different z locations at 20 ps. Solid line: Maxwellian distribution; dots: MD simulation.

distribution function [17,18] can be used to determine the locations of the interfaces. For solid argon, the average number density of atoms is 2.52×10^{28} m^{-3} with a spatial distribution shown in Fig. 6. Owing to the lattice structure, the number density of atoms is higher than the average value near the atomic layer and lower between layers. In this work, if the number density of atoms is higher than 2.52×10^{28} m^{-3}, the material is treated as solid. At the front of the melted region, the number density of atoms decreases sharply to less than 8.42×10^{27} m^{-3}, which is about one third of the number density of solid. Therefore, the material is assumed to be vapor at a number density lower than 8.42×10^{27} m^{-3}. Although the criterion used here for the liquid-vapor interface is not quite rigorous due to the large transition region from liquid to vapor, further studies on the liquid-vapor interface using the radial distribution function show that this criterion gives a good approximation of the liquid-vapor interface [16].

Applying these criteria, transient locations of the solid-liquid and liquid-vapor interfaces, as well as the velocities of the interfaces are obtained and are shown in Fig. 9. It is observed that melting and evaporation start at about 10 ps (laser heating starts at around 5 ps). The solid-liquid interface moves into the solid as melting progresses, and the liquid-vapor interface moves outward as the melted region expands because liquid argon is less dense than solid argon. At about 20 ps, both solid-liquid and liquid-vapor interfaces stop moving. The highest velocity of the liquid-vapor interface is about 200 m/s, close to the equilibrium velocity (233.5 m/s) of the argon atom at the boiling temperature. The highest velocity of the solid-liquid interface is about 400 m/s, much smaller than the local sound (longitudinal) speed ($\sim$1275 m/s) in argon [19].

Thermal equilibrium of the system at different z locations is investigated by comparing the velocity distribution with the Maxwellian distribution. Figure 10 shows the comparison at 20 ps. It can be seen that in solid ($z=7.67$ nm), the transient region ($z=10.38$ nm), liquid ($z=12.18$), and vapor ($z=15.79$ nm), the thermal equilibrium is preserved. This is also true at other times, except in vapor in the early stage of ablation when only very few atoms are evaporated.

The existence of thermal equilibrium justifies the use of temperature to discuss the energy transfer process. Temperature distributions at different times are shown in Fig. 11. At 5 ps, laser heating just starts, and the target has a spatially uniform temperature of about 50 K. Note that the initial size of the target extends from 4.5 nm to 12.6 nm. Melting starts at 10 ps as indicated in Fig. 11, and it is clear that at this moment, the temperature is higher than the melting and the boiling point in the heated region, and is even close to the critical point. At 15 ps, a flat region in the temperature distribution is observed around the location of 10 nm, which is the melting interface region. The temperature in this flat region is around 90 K, which is higher than the melting point (83.8 K), indicating superheating at the melting front. Furthermore, superheating is observed at the liquid-vapor interface ($z\sim13.5$ nm) at 15 ps, which is indicated by the much higher local temperature ($\sim$130 K) compared with the boiling temperature of argon (87.3 K).

An interesting phenomenon is observed at 20 ps, shortly after melting stops. At this moment, a local minimum of temperature is observed at 9.5 nm, which is in the solid-liquid transition region. The temperature is only 60 K, lower than the melting temperature at 1 atm. This minimum temperature disappears gradually due to heat transfer from the surrounding higher temperature regions. This low temperature is likely due to the tensile stress developed in that region. The stress can be calculated as the summation of forces exerted by atoms from both sides of a cross-section of interest, and then divided by the area. The evolution of the stress at the location of 9.5 nm is plotted in Fig. 12, which shows a peak tensile stress at 20 ps, the moment when the minimum temperature appears. A reduced pressure or a tensile stress in argon reduces the melting point [20], therefore causing a temperature drop. A detailed study of this temperature drop and the coupling between the thermal energy and the mechanical energy is going to be reported elsewhere [16]. It is worth noting that many phase change phenomena obtained in this MD simulation such as the lack of a sharp solid-liquid interface and the temperature variation in liquid could not be predicted using the continuum approach.

The average atomic velocity in the z direction at different times is shown in Fig. 13. At 10 ps, melting just starts, and the average velocity of atoms is close to zero except those on the surface, which have a higher kinetic energy due to the free boundary condition. At 15 ps, velocities are much higher. At the vapor front where the thermal equilibrium is still preserved, the velocity is about 450 m/s, while at locations near the surface, the vapor velocity is much smaller. At 30 ps, non-zero velocities are only observed at locations of 15 nm or further, showing weak evaporation on the liquid surface after laser heating stops. The negative velocity close to the bottom of the target is induced by the momentum conservation of the whole system.

The thickness of the melted and vaporized layer and the rate of phase change are shown in Fig. 14. It is seen that the melting depth is much larger than the vaporization depth. From Fig. 14(*b*) it is found that melting happens mostly between 10 and 20 ps, while evaporation goes on until 25 ps, and then reduces to a lower level corresponding to the evaporation of liquid in vacuum. The depths of ablation induced by different laser fluences are shown in Fig. 15, which shows a sudden increase of the ablation depth at the laser fluence of 0.5 J/m^2. The laser fluence when substantial ablation occurs is commonly called the threshold fluence of abla-

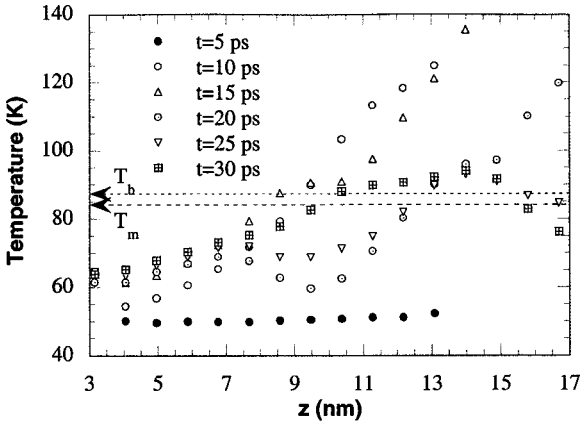


Fig. 11 Temperature distribution in argon illuminated with a laser pulse of 0.7 J/m^2.

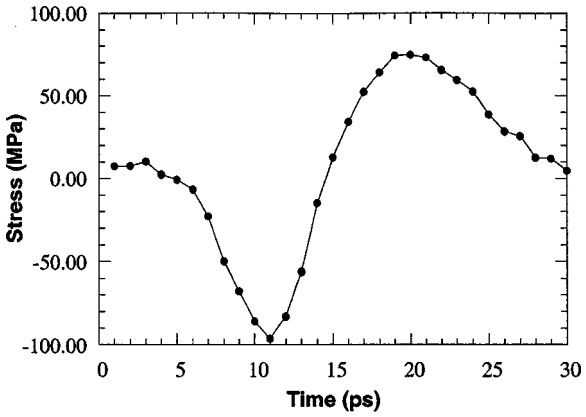


Fig. 12 The stress evolution at the location of 9.5 nm in argon illuminated with a laser pulse of 0.7 J/m^2.

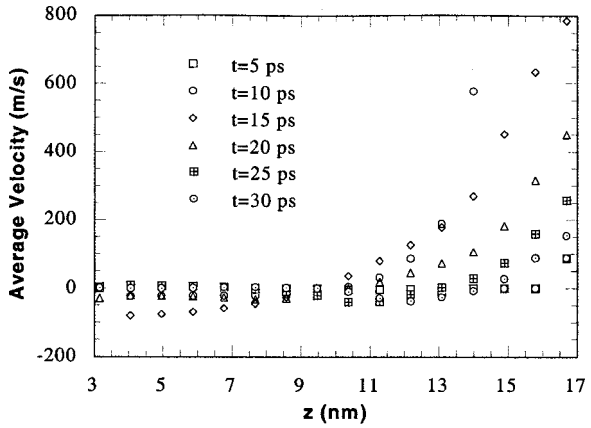


Fig. 13 Spatial distribution of the average velocity in the z direction in argon illuminated with a laser pulse of 0.7 J/m^2.

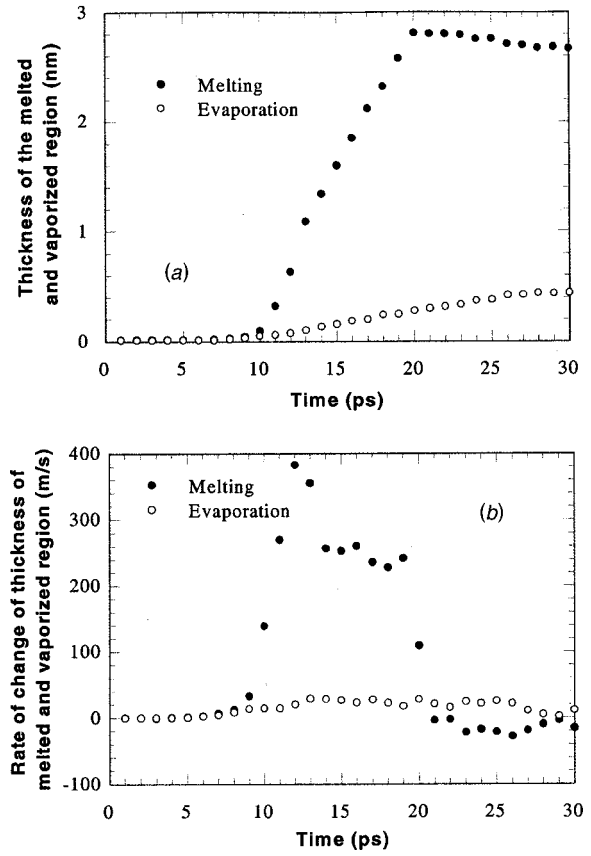


Fig. 14 (*a*) Thickness of the melted and vaporized solid, and (*b*) the rate of change in argon illuminated with a laser pulse of 0.7 J/m^2.

tion. Thus, the calculation shows that with the parameters used in this work, the ablation threshold fluence is 0.5 J/m^2.

Conclusion

In this work, laser material interaction is studied using MD simulations. It is found that during picosecond laser heating, the heat transfer process predicted using the continuum approach agrees with the result of the MD simulation. When phase change occurs, thermal equilibrium is found in solid, liquid, and vapor phases, and the solid-liquid and liquid-vapor interfaces are superheated. A transition region of about 1 nm, instead of a clear interface is found between the solid and the liquid. At a laser fluence of 0.7 J/m^2, the solid-liquid interface moves at almost a constant velocity much lower than the local sound speed, while the liquid-vapor interface moves with a velocity close to the local equilibrium velocity. The laser-ablated material burst out of the target with a vapor front velocity of about hundreds of meters per second.

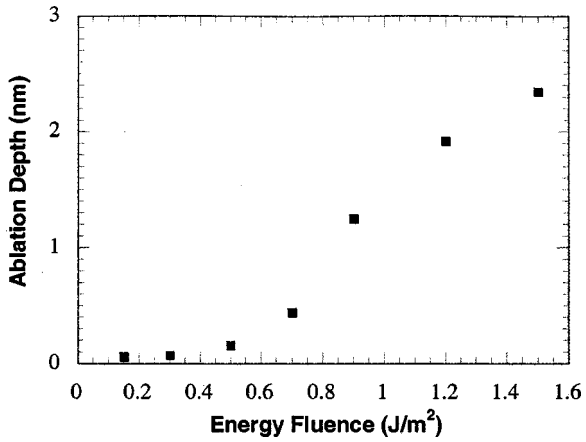


Fig. 15 The ablation depth induced by different laser fluences in argon.

Acknowledgments

Support of this work by the National Science Foundation (CTS-9624890) is gratefully acknowledged.

Nomenclature

F = force
I = laser intensity
k = thermal conductivity
k_B = Boltzmann's constant
m = atomic mass
P_M = probability for atoms moving with a velocity v
r = atomic position
r_c = cut off distance
r_s = the nearest neighbor distance
t = time
t_T = preset time constant in velocity scaling
δt = time step
T = temperature
v = velocity
x = coordinate in x-direction
y = coordinate in y-direction
z = coordinate in z-direction

Greek Symbols

χ = velocity scaling factor
ε = LJ well depth parameter
ϕ = potential
σ = equilibrium separation parameter
ξ = current kinetic temperature in velocity scaling

Subscripts

i = index of atoms

References

[1] Shibahara, M., and Kotake, S., 1997, "Quantum Molecular Dynamics Study on Light-to-Heat Absorption Mechanism: Two Metallic Atom System," Int. J. Heat Mass Transf., **40**, pp. 3209–3222.
[2] Shibahara, M., and Kotake, S., 1998, "Quantum Molecular Dynamics Study of Light-to-Heat Absorption Mechanism in Atomic Systems," Int. J. Heat Mass Transf., **41**, pp. 839–849.
[3] Häkkinen, H., and Landman, U., 1993, "Superheating, Melting, and Annealing of Copper Surfaces," Phys. Rev. Lett., **71**, pp. 1023–1026.
[4] Anisimov, S. I., Kapeliovich, B. L., and Perel'man, T. L., 1974, "Electron

Emission From Metal Surfaces Exposed to Ultra-short Laser Pulses,'' Sov. Phys. JETP, **39**, pp. 375–377.
[5] Kotake, S., and Kuroki, M., 1993, ''Molecular Dynamics Study of Solid Melting and Vaporization by Laser Irradiation,'' Int. J. Heat Mass Transf., **36**, pp. 2061–2067.
[6] Herrmann, R. F. W., and Campbell, E. E. B., 1998, ''Ultrashort Pulse Laser Ablation of Silicon: an MD Simulation Study,'' Applied Physics A, **66**, pp. 35–42.
[7] Zhigilei, L. V., Kodali, P. B. S., and Garrison, J., 1997, ''Molecular Dynamics Model for Laser Ablation and Desorption of Organic Solids,'' J. Phys. Chem. B, **101**, pp. 2028–2037.
[8] Zhigilei, L. V., Kodali, P. B. S., and Garrison, J., 1998, ''A Microscopic View of Laser Ablation,'' J. Phys. Chem. B, **102**, pp. 2845–2853.
[9] Ohmura, E., Fukumoto, I., and Miyamoto, I., 1999, ''Modified Molecular Dynamics Simulation on Ultrafast Laser Ablation of Metal,'' *Proceedings of the International Congress on Applications of Lasers and Electro-Optics*, Laser Institute of America, Orlando, pp. 219–228.
[10] Girifalco, L. A. and Weizer, V. G., 1959, ''Application of the Morse Potential Function to Cubic Metals,'' Phys. Rev., **114**, pp. 687–690.
[11] Etcheverry, J. I., and Mesaros, M., 1999, ''Molecular Dynamics Simulation of the Production of Acoustic Waves by Pulsed Laser Irradiation,'' Phys. Rev. B, **60**, pp. 9430–9434.
[12] Allen, M. P., and Tildesley, D. J., 1987, *Computer Simulation of Liquids*, Clarendon Press, Oxford.
[13] Broughton, J. Q., and Gilmer, G. H., 1983, ''Molecular Dynamics Investigation of the Crystal-Fluid Interface. I. Bulk Properties,'' J. Chem. Phys., **79**, pp. 5095–5104.
[14] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R., 1984, ''Molecular Dynamics With Coupling to an External Bath,'' J. Chem. Phys., **81**, pp. 3684–90.
[15] Peterson, O. G., Batchelder, D. N., and Simmons, R. O., 1966, ''Measurements of X-Ray Lattice Constant, Thermal Expansivity, and Isothermal Compressibility of Argon Crystals,'' Phys. Rev., **150**, pp. 703–711.
[16] Wang, X., and Xu, X., 2001, ''Molecular Dynamics Simulation of Thermal and Thermomechanical Phenomena in Picosecond Laser Material Interaction,'' submitted to the 12th International Heat Transfer Conference.
[17] Wang, X., 2001, ''Thermal and Thermomechanical Phenomena in Laser Material Interaction,'' Ph.D. thesis, Purdue University, West Lafayette, IN.
[18] Chokappa, D. K., Cook, S. J., and Clancy, P., 1989, ''Nonequilibrium Simulation Method for the Study of Directed Thermal Processing,'' Phys. Rev. B, **39**, pp. 10075–10087.
[19] Keeler, G. J., and Batchelder, D. N., 1970, ''Measurement of the Elastic Constants of Argon from 3 to 77°K,'' J. Phys. C, **3**, pp. 510–522.
[20] Ishizaki, K., Spain, I. L., and Bolsaitis, R., 1975, ''Measurements of Longitudinal and Transverse Ultrasonic Wave Velocities in Compressed Solidified Argon and Their Relationship to Melting Theory,'' J. Chem. Phys., **63**, pp. 1401–1410.

Plasma Absorption of Femtosecond Laser Pulses in Dielectrics

C. H. Fan[1]
Postdoctoral Research Associate,
Assoc. Mem. ASME;
e-mail: cfan@addall.com

J. Sun
Graduate Student Researcher,
Student Mem. ASME;
e-mail: jsun@ic.sunysb.edu

J. P. Longtin
Associate Professor, Member ASME;
e-mail: jlongtin@ms.cc.sunysb.edu

Department of Mechanical Engineering,
State University of New York at Stony Brook,
Stony Brook, NY 11794-2300

Dielectric (high bandgap) materials represent an important and diverse class of materials in micro and nanotechnology, including MEMS devices, biomedical and bioengineering systems, multilayer thin film coatings, fiber optics, etc. Micromachining dielectrics using ultrafast lasers is an exciting and promising new research area with many significant advantages, including precision material removal, negligible heating of the workpiece, micron and sub-micron-size feature fabrication, and high aspect ratio features. During ultrafast laser processing of dielectrics, the intense laser pulse ionizes the irradiated material and produces an optical breakdown region, or plasma, that is characterized by a high density of free electrons. These high-density electrons can efficiently absorb a large fraction of the laser irradiance energy, part of which will then be coupled into the bulk material, resulting in material removal through direct vaporization. The energy deposited into the material depends on the time and space-dependent breakdown region, the plasma rise time, and the plasma absorption coefficient. Higher coupling efficiency results in higher material removal rate; thus energy deposition is one of the most important issues for ultrafast laser materials processing, particularly for micron and sub-micron-scale laser materials processing. In the present work, a femtosecond breakdown model is developed to investigate energy deposition during ultrafast laser material interactions. One substantial contribution of the current work is that pulse propagation effects have been taken into account, which have been shown to become significant for pulse durations less than 10 ps. By accounting for the pulse propagation, the time and space-resolved plasma evolution can be characterized and used to determine the energy deposition through plasma absorption. With knowledge of the plasma absorption, changes in the pulse profile as it propagates in the focal region can be determined as well. Absorption of the laser pulse by plasma in water is compared with experimental data to validate the model, as water is a well characterized dielectric. The model, however, is also applicable to other transparent or moderately absorbing solid and liquid dielectric media during ultrafast laser-materials interactions. [DOI: 10.1115/1.1445135]

Keywords: Heat Transfer, Laser, Microscale, Modeling Plasma Absorption

1 Introduction

The application of ultrafast lasers for high-quality laser material processing is an exciting new field that has grown rapidly in the past few years. Due to their extremely high intensity ($>10^{12}$ W/cm^2), ultrashort pulse duration (<1 ps), and non-contact nature of processing, ultrafast lasers show great promise for precise and efficient material processing especially at the microscale. Compared with long-pulse lasers and other conventional manufacturing techniques, ultrafast lasers provide precise control of material removal, can be used with an extremely wide range of materials, produce negligible thermal damage, and provide the capability for very clean, small features, which make them a promising tool for microfabrication, thin film formation, laser cleaning, and medical and biological applications. In addition, the probability of optical breakdown induced by femtosecond laser pulses becomes deterministic, rather than statistical [1], which is the case in the longer pulse regime. Therefore, ultrafast breakdown can be more readily controlled by modulating the irradiated laser intensity, which becomes important as the feature size or material thickness approaches the nanoscale.

Laser processing of high bandgap materials, including diamond, glasses, plastics, oxide ceramics, and liquids usually requires an optical breakdown process to deposit sufficient energy into the material, as these materials generally exhibit minimal absorption at low laser intensities. During optical breakdown, a very high free electron density, i.e., a plasma, on the order of $10^{18}-10^{20}$ electrons/cm^3 is produced [1–3]. Two mechanisms are responsible for laser-induced optical breakdown: multiphoton absorption and avalanche ionization (also called cascade or impact ionization). Multiphoton absorption occurs when a molecule simultaneously absorbs enough photons to ionize the molecule, and occurs when the incoming laser irradiance exceeds a characteristic threshold intensity [4]. Avalanche ionization occurs when free electrons in the material absorb by inverse bremsstrahlung absorption, where they absorb energy from laser pulse by favorable collisions with heavier particles that results in absorption of a photon [1,3,5]. If the electrons sustain enough favorable collisions, they will eventually gain sufficient energy to impact ionize other molecules, which frees new electrons to repeat the process and results in a geometric increase in the free electron density [2,6]. For long pulses, e.g., in the nanosecond and picosecond regime, breakdown is primarily caused by avalanche ionization, while multiphoton absorption dominates breakdown in the low femtosecond regime ($\tau_p<100$ fs) [2,7]. Both multiphoton absorption and avalanche ionization depend on the laser intensity in the interaction volume, and both require a minimum—though different—threshold intensity before breakdown is initiated.

As plasma is formed during optical breakdown, a large fraction

[1]Current address: LOCI, 271 Animal Sciences, University of Wisconsin, Madison, WI 53706

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division May 2, 2001; revision received October 22, 2001. Associate Editor: A. Majumdar.

of the incident laser pulse can be absorbed by free electrons in the plasma through inverse bremsstrahlung, by which they are heated to a high energy level, e.g., ~100–1000 eV (1 eV=11,600 K) or higher; the ions, however, exhibit less absorption and have relatively lower energy, ~1–10 eV, which results in a localized non-equilibrium region in the material. As the laser pulse passes, the energetic electrons relax back to the ground state and release part of their energy to ions and the lattice, which heats the material, resulting in vaporization in the interaction volume within a picosecond to nanosecond time scale [8]. Therefore, the amount of ablated material strongly depends on the volume of the breakdown region and the energy deposited into the material.

Though an existing breakdown model by Docchio et al. [9,10] has been successfully applied to predict the optical breakdown region and plasma absorption in the long pulse regime ($\tau_p > 10$ ps), a modified model presented by Fan and Longtin [11] for femtosecond pulses that takes into account the pulse propagation is required to correctly model optical breakdown with ultrashort laser pulses. In this work, Fan and Longtin's modified breakdown model is applied to characterize the time and space-dependent breakdown region in the femtosecond regime, from which the time-dependent plasma absorption and the temporal pulse profile can be determined as well.

2 Breakdown and Plasma Absorption of Nano and Picosecond Laser Pulses

2.1 Spatial and Temporal Profiles of a Laser Pulse. For most laser systems, the pulse is approximately Gaussian in both the temporal and spatial domains. The Gaussian temporal profile of a laser pulse takes the form [9,12]:

$$P(t) = P_{\max} \exp\left[(-4 \ln 2) \left(\frac{t}{\tau_p} \right)^2 \right], \tag{1}$$

where $P_{\max}$ is the peak power of the laser pulse, and τ_p is the Full Width at Half Maximum (FWHM) pulse duration. The laser beam is also usually focused to increase the intensity, and the beam radius w as a function of axial position z is characterized by [12]:

$$w(z) = w_0 \left(1 + \frac{z^2}{z_R^2} \right)^{1/2}, \tag{2}$$

$$z_R = \frac{n \pi w_0^2}{\lambda_0}, \tag{3}$$

where w_0 is the beam waist, i.e., the minimum beam spot size, which is located at $z=0$, z_R is the Rayleigh length, n is the refractive index of the medium, and λ_0 is the laser wavelength in free space.

As a laser pulse propagates in a medium with a speed c, the spatial extent of the laser pulse is related by the pulse length $l_p = c\tau_p$. For a long laser pulse, e.g., $\tau_p \sim 1$ ns, $\lambda = 800$ nm, and $w_0 = 20$ μm, the pulse length is roughly 30 cm, which is much larger than the focal region characterized by $z_R \sim 1$ mm. In this case, the focal region is much smaller than the spatial extent of the pulse, and can be treated as a single point. Therefore the laser power is very nearly uniform within the focal region, and is time dependent only, with no spatial variation. Consequently, the laser power can be well described by Eq. (1) with $t=0$ representing the time when the pulse peak, $P_{\max}$, reaches the beam focus at $z=0$ [3,9,10,13].

2.2 Moving Breakdown Model. During laser-induced breakdown in transparent or moderately absorbing dielectrics, a minimum threshold intensity, I_{th}, is required before breakdown occurs: for $I < I_{th}$, nothing happens, while $I \geqslant I_{th}$ results in breakdown, where I is the laser intensity (W/cm^2) at the point of interest. During picosecond and longer pulse-induced breakdown, it has been experimentally observed that the breakdown region initially forms at the focal point ($z=0$), then expands up the beam path toward the laser source. This phenomenon, known as moving breakdown [9,14], was modeled by Raizer [15] and later modified by Docchio et al. [9]. In Raizer's original model, optical breakdown is assumed to take place as soon as the incoming laser irradiance exceeds the threshold intensity I_{th}. Docchio et al. [9] revised the moving breakdown model by accounting for focusing of the laser pulse. Combining the temporal pulse in Eq. (1) and the position-dependent irradiated area in Eq. (2), the pulse irradiance $I(z,t)$ can be expressed as [9]:

$$I(z,t) = \frac{P(t)}{A(z)} = \frac{P_{\max}}{\pi w^2(z)} \exp\left[(-4 \ln 2) \left(\frac{t}{\tau_p} \right)^2 \right], \tag{4}$$

where $A(z) = \pi w^2(z)$ is the irradiated area at axial location z and $t=0$ corresponds to the pulse peak intensity. It should be noted that the laser power P in Eq. (4) is uniform in the entire focal volume and depends only on time since the focal region is treated as a single point.

The time at which the laser intensity reaches the breakdown threshold at a given position, $t_B(z)$, can be obtained by letting $I(z,t) = I_{th} = P_{th}/\pi w_0^2$ in Eq. (4) and defining $\beta = P_{\max}/P_{th} = I_{\max}/I_{th}$ as the normalized pulse intensity:

$$t_B(z)_{\pm} = \pm \tau_p \left\{ \frac{1}{4 \ln 2} \ln\left[\beta \left(1 + \frac{z^2}{z_R^2} \right)^{-1} \right] \right\}^{1/2}, \tag{5}$$

with the negative and positive values referring to the starting and ending times of breakdown occurrence at a given position z, respectively. Equivalently, Eq. (5) can be solved for the time-dependent breakdown region, $z_B(t)$:

$$z_B(t)_{\pm} = \pm z_R \left\{ \beta \exp\left[-4 \ln 2 \left(\frac{t}{\tau_p} \right)^2 \right] - 1 \right\}^{1/2}. \tag{6}$$

The region between $z_B(t)_-$ and $z_B(t)_+$ obtained from Eq. (6) represents the region in which breakdown occurs at time t.

Referring to the classical model (Eq. (4)), the variation in laser intensity around the focal region with $P_{\max} = 10^{10}$ W and $w_0 = 20$ μm is plotted in Fig. 1 for different times with an arbitrarily chosen breakdown threshold ($\beta = 6$). For a fixed laser power, larger values of β represent a lower breakdown threshold. As seen in the figure, the breakdown region $z_B(t)$ reaches a maximum, $z_{\max}$, when $t=0$, which corresponds to the pulse peak $P_{\max}$ arriving at the focal region. Note that in the classical breakdown model, the plasma formation time is much shorter than the pulse duration [2], and is thus ignored [9,13], i.e., the plasma is assumed to form immediately once $I > I_{th}$.

2.3 Plasma Absorption of Long Laser Pulses. As plasma forms during optical breakdown, it will strongly absorb the energy of the pulse passing through it [10], and the time-dependent transmitted power of the laser pulse can be determined using Beer's law [13]:

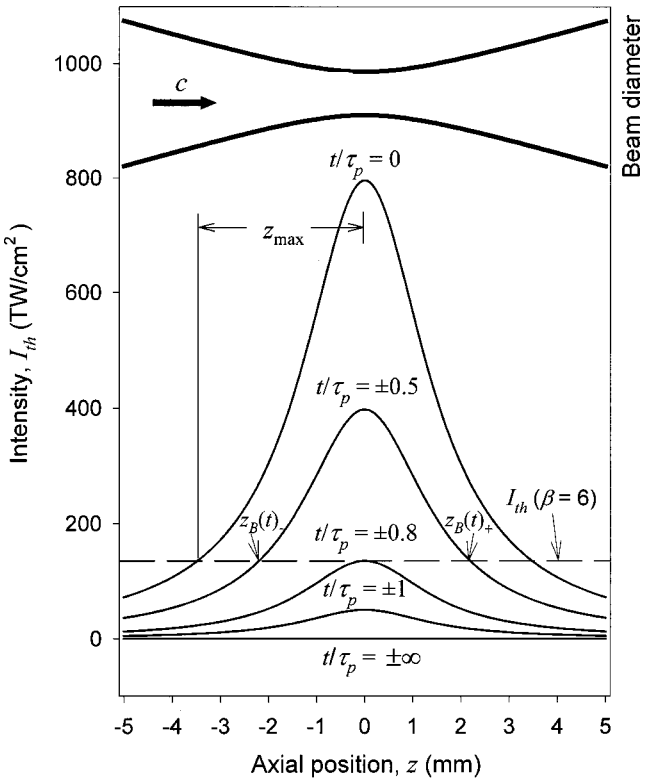


Fig. 1 Time-dependent breakdown region induced by long laser pulses

$$P_T(t) = P(t)\exp[-\alpha_p z_{pl}(t)], \tag{7}$$

where α_p is the spatially and temporally independent plasma absorption coefficient [2,10,13] and $z_{pl}(t)$ is the time-dependent plasma length.

Once formed, the plasma lifetime can approach several nanoseconds [9], hence plasma generated by the leading edge of the pulse can strongly absorb the balance of the pulse energy. Since the intensity is highest at the beam focus, plasma forms first at this point. If the plasma absorbs enough energy from the pulse, the intensity will decrease below the threshold and no new plasma will be produced. Due to strong absorption by the plasma in the focal region, the pulse intensity after the focus is not sufficient for ionization. This so-called distributed shielding effect [10] results in an asymmetric breakdown region that only forms in front of the beam focus ($z<0$). Considering the above, the time-dependent plasma length $z_{pl}(t)$ can be expressed as:

$$z_{pl}(t) = \begin{cases} 0 & t<t_0 \\ z_R\left\{\beta\exp\left[-4\ln 2\left(\dfrac{t}{\tau_p}\right)^2\right]-1\right\}^{1/2} \quad \text{(Eq.(6))} & t_0 \leqslant t<0, \\ z_{\max}=z_R\sqrt{\beta-1} & t\geqslant 0 \end{cases} \tag{8}$$

where breakdown starts at $t_0 = -\tau_p[\ln\beta/(4\ln 2)]^{1/2}$, which is the time the laser intensity first reaches the breakdown threshold at the focus $z=0$, and is a function of the normalized laser intensity β. Once the transmitted power is determined by Eqs. (7) and (8), the transmission of the plasma region can be obtained:

$$T(\beta) = \frac{\int_{-\infty}^{\infty} P_T(t)dt}{\int_{-\infty}^{\infty} P(t)dt} = \frac{1}{E_{P0}}\int_{-\infty}^{\infty} P_T(t)dt, \tag{9}$$

where E_{p0} is the initial pulse energy. Several experimental studies validate the plasma absorption model in Eqs. (7)–(9) for various pulse durations with good agreement found [10,13,16]. However, this classical absorption model can only capture the breakdown region and plasma absorption for nanosecond and picosecond laser pulses, but not for femtosecond pulses.

3 Breakdown and Plasma Absorption of Ultrashort Laser Pulses

3.1 Femtosecond Breakdown Model. As mentioned in the preceding section, the spatial extent of a laser pulse is determined by the pulse length, $l_p = c\tau_p$, while the focal region is characterized by the Rayleigh length z_R. The pulse length to focal region ratio, N_z, can be expressed as [3]:

$$N_z = \frac{l_p}{z_R} = \frac{c\tau_p}{z_R}. \tag{10}$$

Values of $N_z \gg 1$ imply the pulse length is much larger than the focal region, and thus the spatial variation of the pulse profile within the focal region can be ignored—this is the case for nanosecond pulses and is represented by the classical model. However, for femtosecond laser pulses, e.g., $\tau_p \sim 100$ fs, the pulse length is ~ 30 μm, which is much smaller than z_R (~ 1 mm), resulting in $N_z \ll 1$. In such cases, the laser pulse no longer fills the entire focal region at any given time and the propagation of the pulse must be considered.

To account for the small spatial extent and the propagation of ultrashort pulses, a femtosecond breakdown model was proposed by Fan and Longtin [11] in which the incoming laser power is both time and position-dependent with the laser intensity:

$$I(z,t) = \frac{P_{\max}}{\pi w^2(z)}\exp\left[(-4\ln 2)\left(\frac{t-z/c}{\tau_p}\right)^2\right]. \tag{11}$$

The addition of the z/c term in the exponential, compared with Eq. (4), represents the propagation of the pulse in the medium. The point $z=0$ represents the beam focus, at which the pulse peak arrives when $t=0$.

Similar to Eq. (5) from the classical model, the time and position at which the pulse intensity reaches the breakdown threshold becomes [11]:

$$t_B(z)_\pm = z/c \pm \tau_p\sqrt{\frac{1}{4\ln 2}\ln\left[\beta\left(1+\frac{z^2}{z_R^2}\right)^{-1}\right]}, \tag{12}$$

where "−" and "+" represent the times that the intensity first and last exceeds breakdown threshold. Unfortunately, no explicit expression for the time-dependent breakdown positions $z_B(t)_\pm$

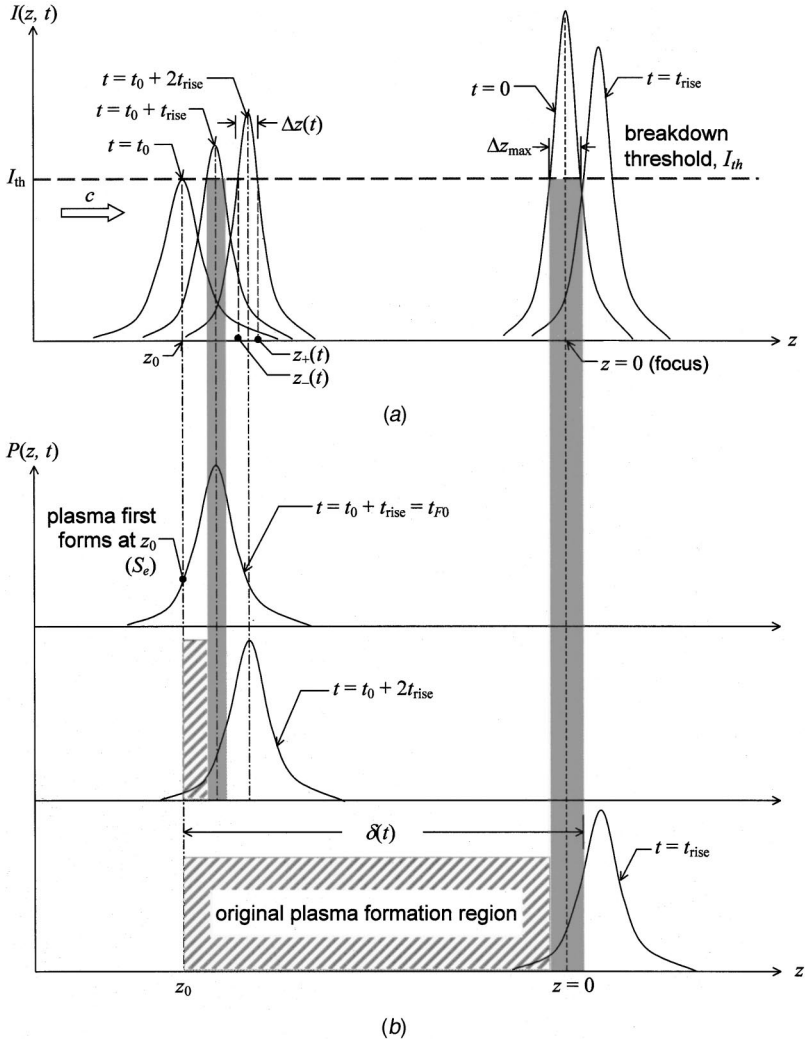


Fig. 2 Pulse propagation and corresponding plasma absorption region: (*a*) intensity increase as beam diameter decreases near focal region; and (*b*) power variation and plasma formation region

exists, though numerical procedures can readily be used. Fan and Longtin demonstrate that there is a unique point z_0 at which t_B is minimum [11], which is obtained by setting $dt_B(z)/dz=0$:

$$\frac{dt_B(z)_-}{dz}=\frac{1}{c}+\frac{\tau_p z}{z_R^2}\left\{\left(1+\frac{z^2}{z_R^2}\right)\sqrt{(4\ln 2)\ln\left[\beta\left(1+\frac{z^2}{z_R^2}\right)^{-1}\right]}\right\}^{-1}=0. \tag{13}$$

The solution to Eq. (13) yields the position z_0 of the first (earliest) breakdown point in the medium. The corresponding time t_0 can be obtained by substituting z_0 into Eq. (12). The solution (t_0, z_0) thus represents the time and position at which the laser pulse intensity first reaches the breakdown threshold as shown in Fig. 2(*a*), both of which are negative for $\beta>1$, indicating breakdown occurs before the focal point.

3.2 Time-Dependent Plasma Formation Regions. In the classical model, plasma is assumed to form once breakdown occurs; there is no time delay considered for plasma growth due to the long pulse duration. However, in the femtosecond regime, the plasma formation time can no longer be neglected: after breakdown occurs at $t=t_0$ determined by Eq. (13), a time t_{rise} is required for plasma growth to the critical electron density, after which the laser energy starts to be efficiently absorbed. Thus plasma forms at $t=t_0$ and $z=z_0$ will be completely formed at a later time

$$t_{F0}=t_0+t_{\text{rise}} \tag{14}$$

at the same location z_0 (Fig. 2(*b*)). Similarly, as the pulse propagates, the plasma induced by the pulse at a later time, for example $t_1=t_0+\Delta t$ at $z_1=z_0+c\Delta t$ will be completely formed at $t_{F1}=t_1+t_{\text{rise}}$, though with an increased spatial extent from the higher intensity due to beam focusing.

The time and position at which breakdown occurs have been determined using Eq. (12). Since the plasma always forms later than breakdown with a time delay of t_{rise}, Eq. (12) can be rewritten by adding this time delay term to establish the relationship between the time and position for plasma formation:

$$t_F(z)_\pm=t_{\text{rise}}+z/c\pm\tau_p\left\{\frac{1}{4\ln 2}\ln\left[\beta\left(1+\frac{z^2}{z_R^2}\right)^{-1}\right]\right\}^{1/2}. \tag{15}$$

Note Eq. (15) holds only for $t\geqslant t_{F0}$ and $z\geqslant z_0$, which indicates that the plasma formation region will expand toward the direction of pulse propagation.

Since the lifetime of formed plasma is ~1 ns [9], which is much longer than the pulse propagation time through the focal

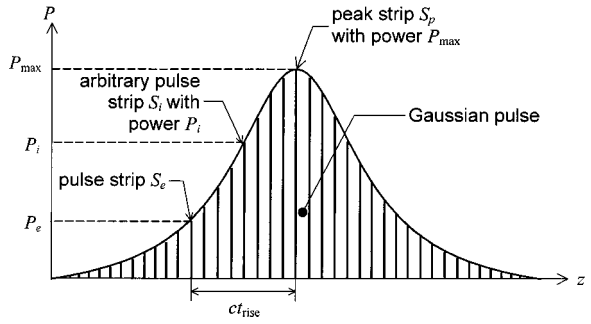


Fig. 3 Schematic of a Gaussian pulse composed of pulse strips

region, the earlier formed plasma will remain in the space and absorb the rest of the incoming laser pulse, which is shown as a shaded region in Fig. 2(*b*). Consequently, in the case shown in Fig. 2 where the pulse propagates from left to right, the left boundary of the plasma formation region at any time is always located at the first plasma formation position, i.e., z_0, while the right boundary is determined by the $z_F(t)_+$ term in Eq. (15), which is time varying and dependent on the above threshold region. Thus the time-dependent plasma formation length in Fig. 2, starting from z_0 at $t=t_{F0}$, can be expressed as:

$$\delta(t)=c(t-t_{F0})+\frac{z_F(t)_+-z_F(t)_-}{2}, \tag{16}$$

where the region between z_{F+} and z_{F-} is the above threshold region shown as a gray region in Fig. 2(*b*), and can be numerically solved by Eq. (15) for any time between t_{F0} and $-t_0+t_{\text{rise}}$.

3.3 Plasma Absorption of Ultrashort Laser Pulses. As plasma is formed, some laser pulse energy will be absorbed primarily by plasma and deposited into the material. To determine the energy absorption for femtosecond laser material interactions, it is worthwhile to consider first the energy absorption of a laser pulse with power P_0 and an infinitesimal pulse duration propagating in an absorbing medium with a constant absorption coefficient α. The position-dependent pulse power after propagating a distance of z is given by Beer's law:

$$P(z)=P_0\exp(-\alpha z). \tag{17}$$

By relating the distance z with the pulse propagation time t and speed c in the medium, the pulse power in Eq. (17) can be rewritten in a time-dependent form:

$$P(t)=P_0\exp(-\alpha ct). \tag{18}$$

Equation (18) shows that the pulse power at any position can be captured by the time that it has propagated in the absorbing medium.

Similarly, the absorption of a Gaussian laser pulse can be characterized by treating the Gaussian pulse as a collection of pulse strips *S*, each of which has infinitesimal pulse duration and a unique power determined by its position within the Gaussian envelop, as shown in Fig. 3. For example, the strip located at the center, S_p, has a maximum power of $P_{\max}$ (Fig. 3), representing the peak of the pulse. All pulse strips propagate at the same speed c in the medium, with an unchanged relative position between each other.

As plasma forms, it starts at t_{F0} and z_0, resulting in the pulse strip located at z_0 at this moment as the one that encounters plasma absorption earliest among all strips, which is denoted as S_e in Fig. 2(*b*) and Fig. 3. As shown in Fig. 2(*b*), due to the increase of the above-threshold region and pulse propagation, the plasma will first form at z_0 and then gradually expand towards the beam focus, indicating the existence of a unique absorption starting time t_F for each pulse strip. For example, the absorption starting time t_F for the strip S_e equals to t_{F0}. If the absorption starting time t_F for each pulse strip is known, then the energy absorption of every strip at any time of interest can be obtained using Eq. (18), by which the total absorption of a laser pulse will be determined.

Assume z_B is the position of an arbitrary strip when its absorption starts at t_F. Then z_B and t_F can be related to any later time t and corresponding position z through the pulse propagation as follows:

$$z=z_B+c(t-t_F). \tag{19}$$

Rewriting Eq. (19), the absorption starting time for an arbitrary strip is:

$$t_F=t-\frac{z-z_B}{c}. \tag{20}$$

As the pulse propagates from left to right shown in Fig. 2(*b*), any portion of the pulse to the *left* of S_e will eventually pass into the plasma region, which begins at z_0. Due to the long plasma lifetime of formed plasma, the absorption of each strip to the left of S_e will always start at z_0, i.e., $z_B=z_0$. Consequently, Eq. (20) can be rewritten as:

$$t_{FL}=t-\frac{z-z_0}{c}, \tag{21}$$

where the subscript L in t_{FL} means left. Note that the position z in Eq. (21) must satisfy $z\leqslant z_0+c(t-t_{F0})$ for it represents the position of any point on the pulse to the left of S_e.

For points on the pulse to the right of S_e, the situation is more complicated. When breakdown first occurs at z_0, all strips to the right of S_e have $z>z_0$, and thus have no opportunity to be absorbed by this plasma. Indeed, if the beam diameter remains constant, all strips to the right of S_e would never be absorbed, for they would always be ahead of the developing plasma region. However, recall that plasma first forms in front of the focal region, hence the pulse intensity continues to increase as it propagates towards the focus, $z=0$. As the pulse intensity increases, a larger fraction of the pulse exceeds the breakdown threshold; hence the plasma formation region will actually grow at a speed greater than the pulse propagation speed, c.[2] Eventually, some (but not all) strips to the right of S_e will be overtaken by the expanding plasma region at time t_{FR} and position z_B shown in Eq. (20).

Recall that the expanding plasma region as a function of time has been given in Eq. (15), which provides another relationship between z_B and t_{FR} besides Eq. (20). Using these two equations, the unknowns z_B and t_{FR} can be obtained:

$$t_{\text{rise}}-t+\frac{z}{c}-\tau_p\sqrt{\frac{1}{4\ln 2}\ln\left\{\beta\left(1+\frac{[z-c(t-t_{FR})]^2}{z_R^2}\right)^{-1}\right\}}=0. \tag{22}$$

Solving the above equation, an analytical solution of t_{FR}, representing the plasma absorption starting time for the strip of interest located to the right of S_e, can be found, which is a function of time t and position z:

$$t_{FR}(z,t)=\frac{\Phi(ct-z)-z_R\sqrt{\Phi(\beta-\Phi)}}{c\Phi} \tag{23}$$

with

$$\Phi=\exp\left[(4\ln 2)\left(\frac{z}{c\tau_p}-\frac{t-t_{\text{rise}}}{\tau_p}\right)^2\right]. \tag{24}$$

[2]Note that this phenomenon does not violate relativity, for it is not a particle, or disturbance, or information that is actually propagating; rather the state of the medium is changing. The situation is similar to anomalous dispersion in a dispersive linear medium [25].

Note t_{F-} in Eq. (15) is used, for this value corresponds to right leading edge of the expanding plasma region, and represents the plasma that will first encounter points to the right of S_e. A complex (non-real) solution from Eq. (23) implies that the expanding plasma front never reaches the position of interest z at the given time t, and no plasma absorption occurs for this strip.

Since the laser pulse is treated as a combination of a series of infinitesimal pulse strips, each of which has its own absorption time, the power of the pulse at a given position z and time t can be expressed as:

$$P(z,t)=\begin{cases} P_{\max}\exp\left[(-4\ln 2)\left(\dfrac{t-z/c}{\tau_p}\right)^2\right] & t<t_{F0} \\ P_{\max}\exp\left[(-4\ln 2)\left(\dfrac{t-z/c}{\tau_p}\right)^2\right]\exp\{-\alpha_p c[t-t_F(z,t)]\} & t\geqslant t_{F0} \end{cases}, \tag{25}$$

where

$$t_F(z,t)=\begin{cases} t_{FL}\ \ (\text{Eq.}(21)) & z\leqslant z_0+c(t-t_{F0}) \\ t_{FR}\ \ (\text{Eq.}(23)) & z>z_0+c(t-t_{F0}) \end{cases}. \tag{26}$$

Here t_F is the time at which the strip located at position z at time t first encountered the plasma, and began to be absorbed. As discussed above, this depends on whether the strip was to the left or right of S_e, as expressed by Eq. (26). For $t<t_{F0}$, no plasma has formed, and the pulse retains its original Gaussian shape. For $t>t_{F0}$, plasma has formed, and the attenuation of a strip of interest depends on the distance of plasma through which it has passed, $c(t-t_F)$. The absorption is taken to be linear in the plasma (Eq. (18)), and represents the additional term in Eq. (25) for $t>t_{F0}$.

The time-dependent pulse energy can be obtained by integrating over the entire pulse power profile:

$$E_p(t)=\frac{1}{c}\int_{-\infty}^{\infty}P(z,t)dz. \tag{27}$$

The time-dependent transmission through the breakdown region can also be determined using:

$$T(t)=\frac{E_p(t)}{E_{p0}} \tag{28}$$

The total transmission of the pulse is then simply obtained by letting t approach infinity.

4 Results and Discussion

4.1 Plasma Formation and Absorption of Laser Energy. The femtosecond absorption model presented in this work characterizes the interaction between a femtosecond laser pulse and dielectric materials. For optical breakdown induced by nanosecond pulses, the breakdown region expands towards the laser and the plasma absorbs the incoming pulse energy; however, the plasma induced by femtosecond pulses forms behind the pulse peak and starts absorbing energy from the tail of the pulses. According to the femtosecond absorption model, plasma absorption depends on the time and space-dependent breakdown region, plasma rise time, and plasma absorption coefficient. The propagating laser pulse that exceeds the threshold intensity produces a time and space-dependent breakdown region in the material, which has been described by the femtosecond breakdown model [11]. In this model [11], the laser beam diameter is determined exclusively by the Gaussian beam propagation in Eq. (2). Other effects, such as self-focusing and optical aberrations, are not taken into account. In addition, effects that change the Gaussian shape of the pulse, e.g., absorption and scattering, are not considered as well, as the focus of the work is on the plasma absorption of a femtosecond laser pulse itself. Incorporating absorption and scattering of the pulse by other mechanisms is the next logical step in furthering this work.

Once the laser intensity reaches the breakdown threshold, plasma will be completely formed, i.e., $\rho=10^{20}$ cm^{-3}, after the formation time, which depends on the mechanisms that induce the breakdown, including multiphoton absorption and avalanche ionization. Theoretical models to predict electron evolution in dielectrics have been reported by several researchers [2,17–19]. Noack and Vogel [2], for example, estimate plasma formation times of ~10 ns and ~200 fs for $\tau_p=6$ ns and 100 fs, respectively, for $\lambda=580$ nm in water.

Another critical parameter associated with the model is the plasma absorption coefficient, which depends on laser intensity, wavelength, electron density, and electron energy [20]. As done by others [2,10,13], an effective plasma absorption coefficient, α_p, is employed, which represents a time-averaged value and depends only on the laser wavelength and pulse duration for a given material; it does not depend on position, time, and electron density and energy.

4.2 Comparison With Experiment. To validate the present model, plasma absorption in water is compared with experimental results [16]. Hammer et al. [16] first reported plasma absorption during femtosecond breakdown in water. Laser pulses with $\lambda=580$ nm and $\tau_p=125$ and 300 fs were used to study transmission through the femtosecond breakdown region, in which several different values of β were obtained with an optical attenuation system. The estimated plasma formation times using the rate equation model [2] are ~200 and ~600 fs for $\tau_p=125$ and 300 fs, respectively, and a constant value of 200 and 180 cm^{-1} for α_p for $\tau_p=125$ and 300 fs with $\lambda=580$ nm are used [2]. Note that the linear absorption coefficient of water is only $\sim 10^{-3}$ cm^{-1} [21], which is ~10,000 times smaller than that of plasma.

The plasma rise time t_{rise} represents an averaged effective time associated with the specific absorption coefficient α_p for plasma absorption, and is chosen such that the total attenuation of a laser pulse passing through the actual, time-evolving plasma and the instant-start plasma absorption approximated in this work is the same. In practice, the effective rise time t_{rise} will be shorter than the actual rise time. Due to the lack of reported plasma rise times for femtosecond optical breakdown, no value associated with the specific α_p can be found. In this work, the plasma rise time was determined by running the model to fit the experimental results of plasma transmission measurement [16]. The plasma transmission is determined using Eq. (28), from which the absorption is then deducted. Other energy losses due to scattering, reflection, linear and nonlinear absorption are ignored.

The predicted transmissions by the new femtosecond absorption model are shown in Table 1 and compared with experimental measurement [16]. The plasma rise times used for model prediction are 120 and 310 fs for $\tau_p=125$ and 300 fs, respectively. Compared with the experimental data, deviations between experi-

Table 1 Comparison of transmission determined by present model and measurement

	β	Model prediction (τ_p = 125 fs)	Experimental measurement [16]	Model prediction (τ_p = 300 fs)	Experimental measurement [16]
No breakdown ($\beta < 1$)	0.2	1.00	0.96	1.00	0.99
	0.5	1.00	0.92	1.00	0.94
	0.9	1.00	0.85	1.00	0.89
Breakdown ($\beta > 1$)	2	0.85	0.70	0.90	0.82
	5	0.68	0.52	0.73	0.68
	10	0.54	0.37	0.61	0.54
	50	0.29	0.21	0.35	0.32
	100	0.21	0.19	0.27	0.26

ment and model prediction are found. In the experimental study, e.g., $\tau_p = 125$ fs, a transmission as low as 0.85 has been observed before breakdown occurs ($\beta < 1$), which indicates the existence of other forms of energy loss due to multiphoton absorption, avalanche ionization, scattering, and linear absorption [16,22,23]. Similar results are found for $\tau_p = 300$ fs. The model assumes no absorption or energy loss occurs until the laser intensity reaches the breakdown threshold, thus the transmission remains unity for all cases with $\beta < 1$. As β increases, the breakdown region expands accordingly; consequently, plasma absorption dominates other effects, which explains the better agreement for larger β in Table 1.

Another parameter that will affect model predictions is the plasma rise time t_{rise}. A longer plasma rise time, which increases the delay of pulse absorption, results in higher transmission. For $\tau_p = 125$ fs as an example, using $t_{\text{rise}} = 150$ fs in the absorption model will result in transmissions of 0.94, 0.74, and 0.40 for $\beta = 2$, 10, and 100 respectively [24], which differ from the results shown in Table 1 for $t_{\text{rise}} = 120$ fs. Accurate values of t_{rise} will improve the accuracy of absorption predictions using the current model.

As soon as t_{rise} is chosen, the time-dependent plasma formation length, $\delta(t)$, can be determined using Eq. (16). The resulting time-dependent plasma locations, $z_0 + \delta(t)$, are shown in Fig. 4 for $\tau_p = 300$ fs and $\beta = 2$, 10, 50, and 100, where larger β results in earlier initiation of plasma and a correspondingly larger plasma formation region.

Using $t_{\text{rise}} = 310$ fs and $\alpha_p = 180\ \text{cm}^{-1}$, the time-dependent spatial pulse profile for $\beta = 10$ and $\tau_p = 300$ fs using Eq. (25) is shown in Fig. 5. The pulse profiles at different times, $t = t_{F0}$, $t_{F0} + 5\tau_p$, $t_{F0} + 10\tau_p$, and $-t_0 + t_{\text{rise}}$ (infinity) are plotted, with corresponding transmissions of 1.00, 0.82, 0.62, and 0.61, respectively. As can be seen, the absorption starts at S_e on the pulse profile and then propagates forward with respect to the pulse as t increases. For absorption at later times or situations with larger β, the absorption region in the pulse profile can approach the pulse peak and reduce the intensity that contributes to breakdown.

Applying Eq. (28), the time-dependent transmission for $\tau_p = 300$ fs is shown in Fig. 6, where the transmission T is plotted versus the dimensionless time $(t - t_{F0})/\tau_p$ for $\beta = 2$, 10, 50, and 100. For different values of β, the deviation of T from unity, which corresponds to the beginning of plasma absorption, starts all at $t = t_{F0}$ when plasma is first formed at z_0. After this point, the transmission decreases monotonically until $t = -t_0 + t_{\text{rise}}$, indicating that the pulse energy is continuously absorbed by the plasma until plasma formation stops. Note, however, there is a turning point at $t = t_{\text{rise}}$ (circular dot in the figure) for all four curves, after which the change of transmission with time becomes negligibly small, suggesting that there is almost no absorption of the pulse energy after this point. At $t = t_{\text{rise}}$, the freshly formed plasma induced by the above-threshold laser pulse reaches a maximum. After this point, the above-threshold region decreases and no further fresh pulse strips will be overtaken by the expanding plasma region. Though fresh absorption still occurs at z_0 for the incoming tail of the pulse, its contribution to the total absorption of the pulse is very small due to the low power of the pulse tail. Therefore, though the theoretical maximum region of plasma formation ranges from z_0 to $-z_0$, the dominant absorption will stop around

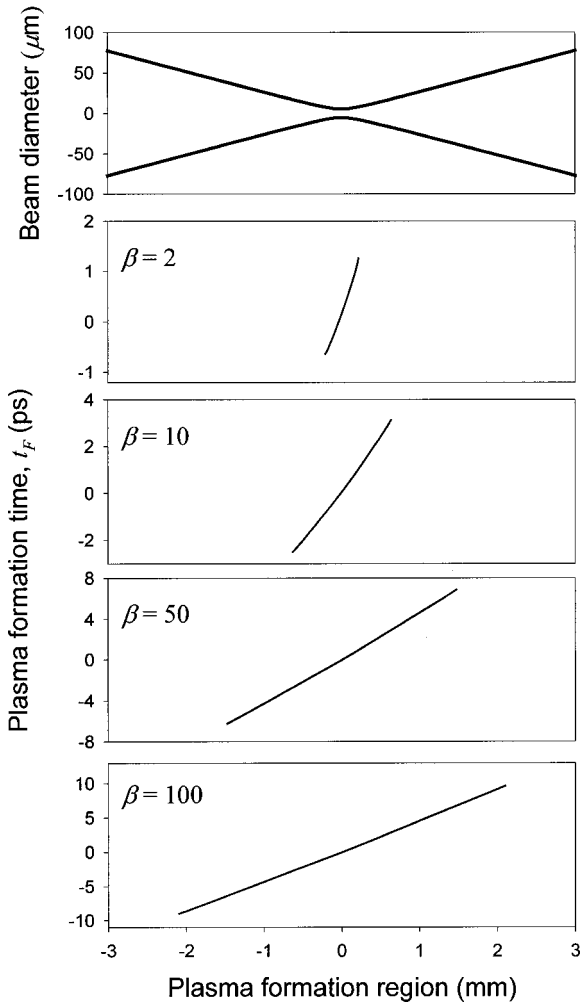


Fig. 4 Plasma formation location versus formation time

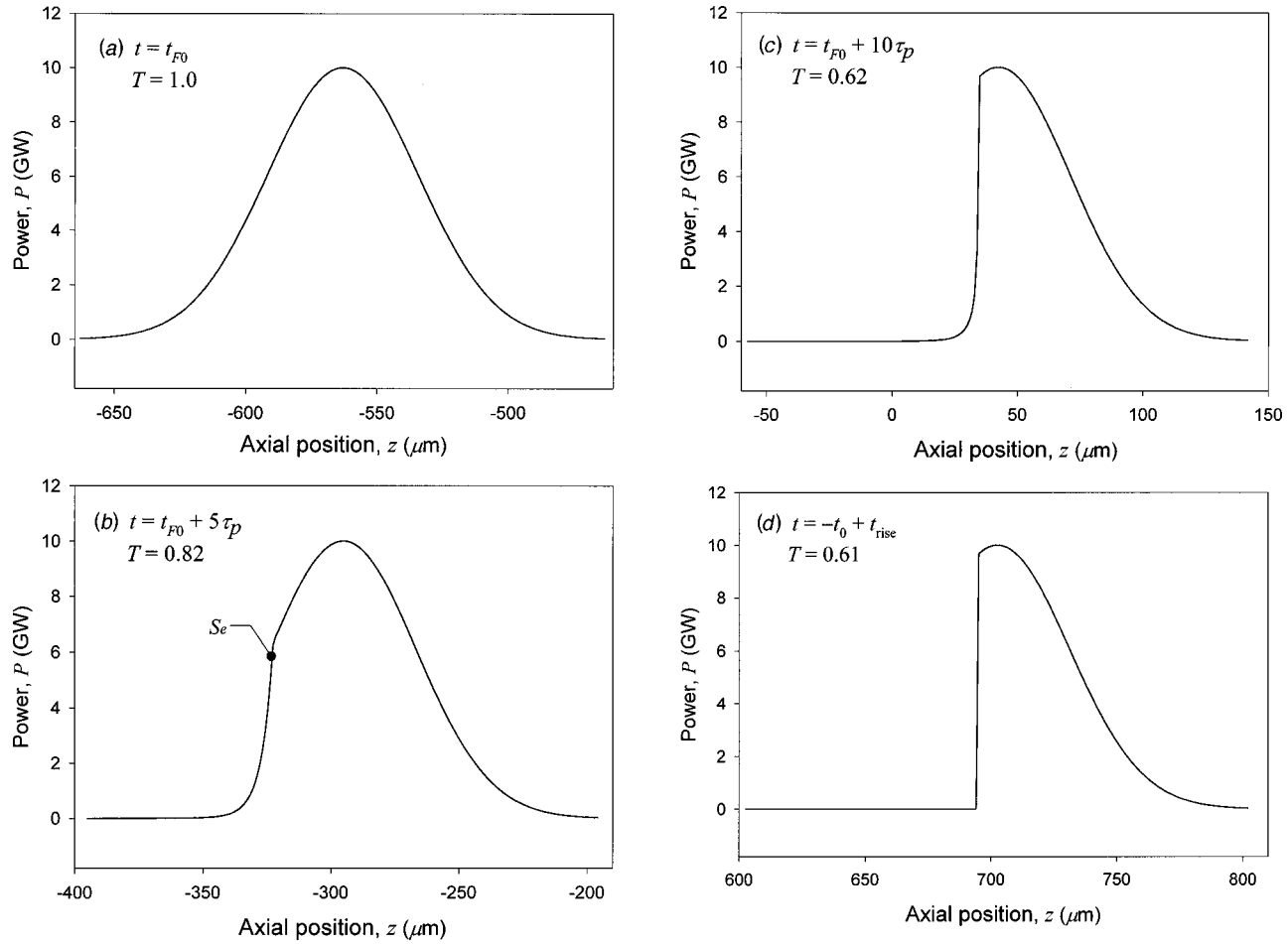


Fig. 5 Time-dependent pulse profile with plasma absorption in water (τ_p=300 fs): (*a*) $t=t_{F0}$, (*b*) $t=t_{F0}+5\tau_p$, (*c*) $t=t_{F0}+10\tau_p$, and (*d*) $t=-t_0$.

the beam focus at $z=0$ when $t=t_{\text{rise}}$, which is roughly half of the maximum breakdown region, suggesting less important for the plasma formation after the beam focus.

5 Conclusions

Precision ultrafast laser processing of dielectric materials for micro and nano-devices relies strongly on energy deposition from the laser into the materials. Theoretical models for ultrafast laser-dielectric interactions, however, are far less developed than the rapidly expanding catalog of experimental data currently being collected. This issue is addressed in this work, with a femtosecond absorption model presented to characterize plasma absorption during femtosecond laser processing of dielectric materials. Due to the short pulse length and finite plasma rise time, plasma initiated by the portion of a laser pulse that exceeds the breakdown threshold will form at a later time and absorbs energy from the tail of the pulse. This behavior differs substantially from the sequence of events for laser-material interactions in the nanosecond and picosecond regimes. Using this model, both time and space-dependent breakdown regions are determined by taking into account the pulse propagation. By comparing model predictions with experimental results, an effective plasma rise time for femtosecond pulse induced breakdown is also obtained. Based on the time and space-dependent plasma region as well as the plasma rise time, a time-dependent pulse profile can be determined, from which the energy absorption by the plasma can be quantified. This model shows promise for estimating the temperature increase and material removal rates for ultrafast laser processing.

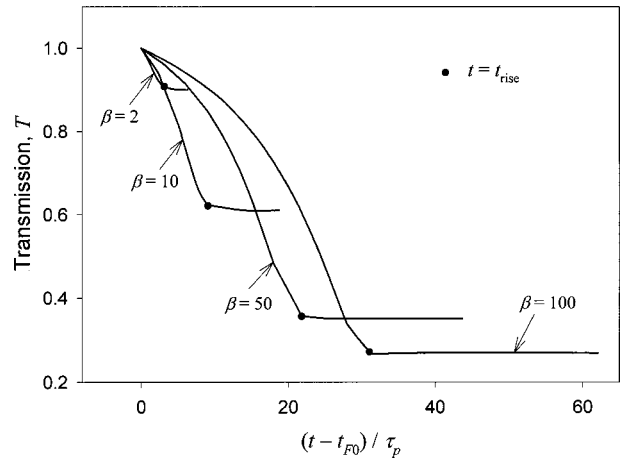


Fig. 6 Time-dependent transmission for different β (τ_p =300 fs)

Acknowledgments

The authors gratefully acknowledge support of this work by the division of Chemical and Transport Systems at the National Science Foundation under contract CTS-9702644, and would like to thank Dr. Daniel X. Hammer of Biomedical Engineering Program, University of Texas at Austin for the experimental data.

Nomenclature

A = irradiated area, m^2
c = speed of light in a medium, m/s
E_p = single pulse energy, J
I = laser intensity/irradiance, W/m^2
l_p = pulse length, m
N_z = ratio of pulse length to diffraction length
n = refractive index
P = power, W
S = pulse strip
T = transmission
t = pulse propagation time, s
t_{rise} = plasma rise time, s
w = beam radius, m
w_0 = beam waist, m
z = axial location, m
z_{pl} = plasma length, m
z_R = Rayleigh length, m

Greek Symbols

α = absorption coefficient of a medium, m^{-1}
β = normalized pulse power or intensity
δ = plasma formation length, m
λ = wavelength, m
ρ = electron density, cm^{-3}
τ_p = pulse duration (FWHM), s
Φ = dummy variable

Subscripts

0 = free space; initial
1 = dummy variable
B = breakdown
e = earliest
F = plasma formation
i = dummy variable
L = left
max = peak; maximum
p = plasma; peak
R = right
T = transmission
th = threshold

References

[1] Liu, X., Du, D., and Mourou, G., 1997, ''Laser Ablation and Micromachining With Ultrashort Laser Pulses,'' IEEE J. Quantum Electron., **33**, pp. 1706–1716.

[2] Noack, J., and Vogel, A., 1999, ''Laser-Induced Plasma Formation in Water at Nanosecond to Femtosecond Time Scales: Calculation of Thresholds, Absorption Coefficients, and Energy Density,'' IEEE J. Quantum Electron., **35**, pp. 1156–1167.

[3] Kennedy, P. K., Hammer, D. X., and Rockwell, B. A., 1997, ''Laser-Induced Breakdown in Aqueous Media,'' Prog. Quantum Electron., **21**, pp. 155–248.

[4] Nikogosyan, D. N., Oraevsky, A. A., and Rupasov, V. I., 1983, ''2 Photon Ionization and Dissociation of Liquid Water by Powerful Laser UV Radiation,'' Chem. Phys., **77**, pp. 131–143.

[5] Du, D., Liu, X., and Mourou, G., 1996, ''Reduction of Multi-Photon Ionization in Dielectrics Due to Collisions,'' Appl. Phys. B, **63**, pp. 617–621.

[6] Shen, Y. R., 1984, *The Principles of Nonlinear Optics*, John Wiley & Sons, Inc., New York.

[7] Feng, Q., Moloney, J. V., Newell, A. C., Wright, E. M., Cook, K., Kennedy, P. K., Hammer, D. X., Rockwell, B. A., and Thompson, C. R., 1997, ''Theory and Simulation on the Threshold of Water Breakdown Induced by Focused Ultrashort Laser Pulses,'' IEEE J. Quantum Electron., **33**, pp. 127–137.

[8] Korte, F., Nolte, S., Chichkov, B. N., Bauer, T., Kamlage, G., Wagner, T., Fallnich, C., and Welling, H., 1999, ''Far-Field and Near-Field Material Processing with Femtosecond Laser Pulses,'' Appl. Phys. A, **69**, pp. S7–S11.

[9] Docchio, F., Regondi, P., Capon, M. R. C., and Mellerio, J., 1988, ''Study of the Temporal and Spatial Dynamics of Plasmas Induced in Liquids by Nanosecond Nd:YAG Laser Pulses. 1. Analysis of the Plasma Starting Times,'' Appl. Opt., **27**, pp. 3661–3668.

[10] Docchio, F., Regondi, P., Capon, M. R. C., and Mellerio, J., 1988, ''Study of the Temporal and Spatial Dynamics of Plasmas Induced in Liquids by Nanosecond Nd:YAG Laser-Pulses. 2. Plasma Luminescence and Shielding,'' Appl. Opt., **27**, pp. 3669–3674.

[11] Fan, C. H., and Longtin, J. P., 2001, ''Modeling Optical Breakdown in Dielectrics during Ultrafast Laser Processing,'' Appl. Opt., **40**, pp. 3124–3131.

[12] Siegman, A. E., 1986, *Lasers*, University Science Books, Sausalito, CA.

[13] Nahen, K., and Vogel, A., 1996, ''Plasma Formation in Water by Picosecond and Nanosecond Nd:YAG Laser Pulses—Part II: Transmission, Scattering, and Reflection,'' IEEE J. Sel. Top. Quantum Electron., **2**, pp. 861–871.

[14] Vogel, A., Nahen, K., Theisen, D., and Noack, J., 1996, ''Plasma Formation in Water by Picosecond and Nanosecond Nd:YAG Laser Pulses—Part I: Optical Breakdown at Threshold and Superthreshold Irradiance,'' IEEE J. Sel. Top. Quantum Electron., **2**, pp. 847–860.

[15] Raizer, Y. P., 1966, ''Breakdown and Heating of Gases Under the Influence of a Laser Beam,'' Sov. Phys. Usp, **8**, pp. 650–673.

[16] Hammer, D. X., Jansen, E. D., Frenz, M., Noojin, G. D., Thomas, R. J., Noack, J., Vogel, A., Rockwell, B. A., and Welch, A. J., 1997, ''Shielding Properties of Laser-Induced Breakdown in Water for Pulse Durations from 5 ns to 125 fs,'' Appl. Opt., **36**, pp. 5630–5640.

[17] Perry, M. D., Stuart, B. C., Banks, P. S., Feit, M. D., Yanovsky, V., and Rubenchik, A. M., 1999, ''Ultrashort-Pulse Laser Machining of Dielectric Materials,'' J. Appl. Phys., **85**, pp. 6803–6810.

[18] Stuart, B. C., Feit, M. D., Herman, S., Rubenchik, A. M., Shore, B. W., and Perry, M. D., 1996, ''Optical Ablation by High-Power Short-Pulse Lasers,'' J. Opt. Soc. Am. B, **13**, pp. 459–468.

[19] Kennedy, P. K., 1995, ''A First-Order Model for Computation of Laser-Induced Breakdown Thresholds in Ocular and Aqueous Media. 1. Theory,'' IEEE J. Quantum Electron., **31**, pp. 2241–2249.

[20] Rae, S. C., and Burnett, K., 1992, ''Possible Production of Cold Plasmas Through Optical-Field-Induced Ionization,'' Phys. Rev. A **46**, pp. 2077–2083.

[21] Stone, J., 1972, ''Measurements of the Absorption of Light in Low-Loss Liquids,'' J. Opt. Soc. Am., **62**, pp. 327–333.

[22] Longtin, J. P., and Tien, C. L., 1997, ''Efficient Laser Heating of Transparent Liquids Using Multiphoton Absorption,'' Int. J. Heat Mass Transf., **40**, pp. 951–959.

[23] Longtin, J. P., 1999, ''Using Multiphoton Absorption With High-Intensity Lasers to Heat Transparent Liquids,'' Chem. Eng. Technol., **22**, pp. 77–80.

[24] Fan, C. H., Sun, J., and Longtin, J. P., ''Time and Space-Resolved Plasma Absorption of a Femtosecond Laser Pulse in Dielectrics,'' *35th National Heat Transfer Conference*, Anaheim, CA, NHTC01-11222.

[25] Wang, L. J., Kuzmich, A., and Dogariu, A., 2000, ''Gain-Assisted Superluminal Light Propagation,'' Nature (London), **406**, pp. 277–279.

D. Y. Tzou
James C. Dowell Professor,
Fellow ASME
Department of Mechanical and
Aerospace Engineering,
University of Missouri-Columbia,
Columbia, MO 65211

J. E. Beraun

J. K. Chen

Laser Effect Research Branch,
Air Force Research Laboratory,
Kirtland AFB, NM 87117

Ultrafast Deformation in Femtosecond Laser Heating

The hot-electron blasting model is extended in this work to describe the ultrafast deformation in thin metal films during the sub-picosecond to picosecond domain. The driving force exerting on the cold metal lattices is induced by the highly heated electrons, dictated by both the temperature and temperature gradient established in the hot electron during the picosecond transient. Since the metal lattices remain almost thermally undisturbed in this highly non-equilibrium regime, the resulting ultrafast deformation patterns cannot be described by the classical dynamical theory of thermoelasticity. The phonon-electron interaction model is used to describe the electron temperature and hence the driving force. The dominating parameters characterizing the nonlinearly coupled ultrafast heating and deformation are identified. Method of lines is used to solve the coupled field equations describing ultrafast deformation in the picosecond domain.
[DOI: 10.1115/1.1447934]

Keywords: Conduction, Heat Transfer, Microscale, Non-Equilibrium, Transient

Introduction

Femtosecond (fs, 10^{-15} s) lasers possess several unique features in thermal processing/characterization of materials, resulting from its exclusive capability in limiting the undesirable spread of the thermal process zone in the precisely heated sample. The 96 fs (about 0.1 ps, picoseconds) lasers that have been developed over a decade [1], in particular, have become mature [2] for various applications. For femtosecond laser heating on metals, for example, characterization of the Gaussian profiles describing the laser-material interactions and development of microscale heat transfer models describing the phonon-electron interactions in the picosecond domain have made it possible to further advance the laser technology in laser patterning [3], laser micro-machining [4], nondestructive evaluation of high-T_c superconducting films [5], thin-film deposition [6], local surface hardening, and tailoring of microstructures in thin-film deposition [7]. Benefiting from the noncontact nature and precisely controlled heating times in the subpicosecond domain; in addition, femtosecond lasers have been identified as an effective processing tool for assembling functional nanophase materials [8,9]. For assuring success in these applications, doubtlessly, the ways in which thermal energy is dissipated at short times in the processed sample play a central role that deeply involve fundamental studies of ultrafast heat transfer in microscale.

Research of picosecond heat transport has been mainly focused on rigid conductors, although precautions have been made in preventing possible thermal damage due to thermal exaggeration developed in the picosecond domain [10,11]. Representatives include determination of the phonon-electron coupling factors [10,14,15] in the two-step heating model [12,13] and the wave behavior due to rapid thermal relaxation in semiconductors, dielectric crystals, and insulators [16–19]. Such distinct features result from the transient time comparable to the characteristic times in electron-to-phonon (for metals, ranging from several to 10 ps) and phonon-to-phonon (for semiconductors, dielectrics, and insulators beyond 10 ps) collisions. In large, the various microscale effects could be lumped onto two characteristic times, called thermalization time and relaxation time, in thermal lagging [11,20–22]. The relaxation behavior is responsible for the temperature ripples propagating in superfluid liquid helium, while the thermalization behavior is liable for the rapid fall of electron temperature due to phonon-electron interactions in metals. When both behaviors become active at short times, most importantly, effect of rapid thermalization effectively smoothes the sharp wavefront produced by rapid thermal relaxation, rendering a much deeper thermal process zone and a much higher temperature than those predicted by the classical Fourier or thermal wave model.

The deformation pattern developed in the picosecond domain, termed ultrafast deformation in this work, has not received as much attention. Existing works include the effective Hamiltonian [23] and Lagrangian [24] formulations for the inelastic scattering of electrons by acoustic and optical phonons and effects of ultrafast thermalization and relaxation on thermally excited lattices [11,25,26]. The hot-electron blast exerting on the metal lattices [27], the basis for the present work, deserves a special attention. With the metal lattice staying almost thermally undisturbed, such a blasting force could be so strong that the ''cold'' lattices are impelled well into the domain for damage. Aiming at transient times of the order of several picoseconds, the present work employs the parabolic two-step model [2,10,14,15] to reflect the fine details in phonon-electron interactions. The driving force derived from the thermal field of hot electrons is directly fed into the equation of motion that describes the dynamic response of the metal lattices. Dominating parameters in ultrafast deformation are extracted, in forms of nondimensional groups characterizing the combined effects of laser heating, film geometry and thermal properties of phonons and electrons. Unique features include tension preceded by compression near the heated surface during the picosecond transient, and evolution from single to multiple straining in the heated film as a integral effects of film thickness and dilatational wave speed.

Kinetics

For femtosecond laser heating on metals, due to the smaller heat capacity of electrons than that of phonons by two to three orders of magnitude, electrons are excited to a very high temperature in several tenths of a picosecond. In the meantime, as illustrated later in Fig. 1(*a*), the metal lattice only experiences a minor increase of temperature, or even remains thermally undisturbed, should the laser intensity be moderate. This stage is termed nonequilibrium heating, where heat capacity and thermal conductivity of electrons both increase linearly with the electron temperature (T_e) as long as T_e is much smaller than the Fermi temperature (about 10^4 K). Energy exchange between electrons and phonons

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division April 24, 2001; revision received September 4, 2001. Associate Editor: D. Poulikakos.

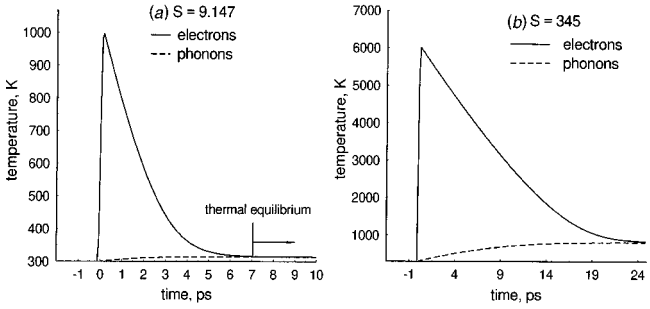


Fig. 1 Non-equilibrium electron and lattice (phonon) temperatures in the picosecond domain

is characterized by the electron-phonon coupling factor during this stage, which has been identified as a new thermal property of metals in microscale heat transfer [2,10,14,15]. For describing the deformation of metal lattices in this ultrashort domain of time, usually of the order of a few picoseconds, the classical dynamical theory of thermoelasticity has an intrinsic problem because the metal lattices remain almost thermally undisturbed. During the picosecond transient, therefore, neither the highly elevated temperature nor the large temperature gradient (of the metal lattices) is responsible for the suspected thermal damage [10,11,27].

In the sub-pico to picosecond domain, hot electrons may influence cold lattices in two ways [23,24,27]. First, renormalization of elastic moduli (E_{ijlm}, for $i,j,l,m=1, 2, 3$ in a three-dimensional Cartesian space) may take place as a result of the electron loop in the phonon self-energy function. For transient time (t) longer than the electron relaxation time ($\sim$10 fs), the dominant contribution for the change of elastic moduli (ΔE_{ijlm}) is from the local equilibrium Fermi-Dirac partition function:

$$\Delta E_{ijlm}=E_{ijlm}-\langle d_{ij}(\mathbf{p})d_{lm}(\mathbf{p})\rangle \cong -\frac{\pi T_e^2}{6}\frac{\partial}{\partial \varepsilon_F^2}\langle d_{ij}(\mathbf{p})d_{lm}(\mathbf{p})\rangle \quad \text{with} \quad \langle d_{ij}d_{lm}\rangle=\int_F \frac{d_{ij}d_{lm}}{4\nu\pi^3}dS_F, \tag{1}$$

where $\mathbf{p}$ represents the energy state variables defining the local electron spectrum (ε), d_{ij} is the deformation potential tensor of the lattice surface whose unit normal is in the direction of x_i and is driven to deform in the direction of x_j, F is the Fermi surface, the brackets denote integration over the Fermi surface, and υ denotes the mean velocity of electrons. The change of elastic moduli due to the renormalization process, as shown in Eq. (1), is of the order of $(T_e/T_F)^2$, since $\varepsilon_F=k_BT_F$ with k_B being the Boltzmann constant. Unless the electron temperature approaches the order of 10^4 K (T_F), the elastic moduli can be assumed constant without a significant difference.

Second, the driving force produced by the hot-electron blast is a much more important factor in ultrafast lattice deformation. Mathematically, the equation of dynamical theory of thermoelasticity can be recapitulated as [27]

$$\rho\frac{\partial^2 u_i}{\partial t^2}-E_{ijlm}\frac{\partial^2 u_l}{\partial x_j\partial x_m}=\Lambda_{ij}\left(T_e\frac{\partial T_e}{\partial x_j}\right) \quad \text{where} \quad \Lambda_{ij}=\frac{1}{16\pi}\frac{\partial}{\partial \varepsilon_F}\int\frac{d_{ij}(\mathbf{p})dS_F}{\nu}, \tag{2}$$

where ρ stands for mass density, u_i is the displacement vector of acoustic phonons, x_j and t are space and time coordinates, respectively, and a repeated index (j or m) implies summation from 1 to 3 in the Cartesian coordinates. Equation (2) describes the first-order effects of dynamic responses in (T_e/T_F). Under moderate laser heating where $T_e\sim10^2$ to 10^3 K, effect of hot electron blast dominates over that of the lattice renormalization described in Eq. (1). Note that the driving force shown by the right-hand side of Eq. (2) is proportional to the *gradient* of the electron temperature squared, i.e., $2T_e(\partial T_e/\partial x_j)=\partial(T_e^2)/\partial x_j$. It depends on both the electron temperature and its gradient. The proportional constants, Λ_{ij} of the unit in J/m^{3}K^2, results from the integral of the deformation potential (d_{ij}) over all the energy variables ($\mathbf{p}$) along the Fermi surface (S_F). For most metals, the value of Λ_{ij} is of the same order of the electron density states at the Fermi surface [27]. For noble metals such as gold, the number of free electrons is taken as one per atom and the number density of free electrons is taken as a fraction of the valence electrons in the *s*-band. A precise estimate for the value of Λ_{ij} is difficult to make because the number of free electrons is not sufficiently clear [14]. For most metals with an electron density states on the order of 10^{-2}–10^{-1} levels/(eV-atom), the value of Λ_{ij} should be in the range from 10 to 10^3 J/m^{3}K^2.

Note that Eq. (2) only applies to acoustic phonons that are active during the picosecond transient. Optical phonons describing the deformation of sublattices are active during the much earlier femtosecond transient, possessing deformation potential vectors, in contrast to the deformation potential tensor in Eq. (1). Consequently, the equations of motion as well as the driving force are different from those shown by Eq. (2).

Two-Step Model

Temperature distribution in the electron gas plays a dominating role in determining the driving force, and hence the ultrafast lattice deformation during the picosecond transient, as shown by Eq. (2). Derived on the same basis of the Boltzmann equation [15], the two-step model describing the phonon-electron interactions in metals is readily applicable for calculating the electron temperature [2,10,11,14,15]. For thin metal films ($\sim$0.1 μm) heated by an ultrashort laser, in parallel to the three-dimensional expression of the temperature gradient in Eq. (2), the energy equations are

$$C_e(T_e)\frac{\partial T_e}{\partial t}=\frac{\partial}{\partial x_j}\left[k_e(T_e,T_l)\frac{\partial T_e}{\partial x_j}\right]-G(T_e-T_l)+0.94J\left(\frac{1-R}{\delta t_p}\right)\exp\left[-\frac{x}{\delta}-4\ln(2)\left(\frac{t}{t_p}\right)^2\right];$$
$$C_l\frac{\partial T_l}{\partial t}=G(T_e-T_l), \quad \text{where} \quad C_e(T_e)=C_{e0}\left(\frac{T_e}{T_0}\right) \quad \text{and} \quad k_e(T_e,T_l)=k_0\left(\frac{T_e}{T_l}\right) \tag{3}$$

with subscripts e and l referring to electrons and lattices, respectively, C the volumetric heat capacity, k the thermal conductivity, and G the phonon-electron coupling factor (in the unit of W/m^{3}K). In the energy absorption rate described by the last term in the first expression of Eq. (3),[1] J represents the laser fluence (W/m^2), R the surface reflectivity, t_p the laser pulse width, δ the radiation penetration depth and the constant $4\ln(2)\cong2.77$ describes the Gaussian profile of the time-varying laser intensity. Also in Eq. (3), C_{e0} is the volumetric heat capacity of electrons at room temperature (T_0), whereas k_0 is the thermal conductivity of electrons when coming to thermal equilibrium with metal lattices ($T_e=T_l$). Since $T_e\gg T_l$ during the first few picoseconds in non-equilibrium heating, the value of k_e may be several orders of magnitude larger than the equilibrium value (k_0). Heat capacity (C_e) and conductivity (k_e) of electrons increase linearly with the electron temperature (T_e), while heat capacity of the metal lattice (C_l) is assumed

[1]The Gaussian profile describing the energy absorption rate in Eq. (3), and later in Eq. (5), is only valid for shallow (near-surface) heating with constant optical properties. Deep heating may involve optical properties that sensitively vary with the light intensity, pulse duration, and temperature in the heated domain, which are not modeled in the present formulation.

constant due to the minor change of lattice temperature (T_l) during the picosecond transient (shown later in Fig. 1). Effect of heat conduction in phonons is neglected in Eq. (3) since heat flux is mainly carried by free electrons in metals. At last, the phonon-electron coupling factor (G) is insensitive to electron temperature, whose values are typically from 10^{14} to 10^{16} W/m^{3}K.

Thin-Film Heating

To illustrate the fundamental behavior of ultrafast deformation induced by the hot-electron blast, we consider a simply supported, thin metal film of thickness l heated by an ultrashort laser. The local response focuses on the laser-heated area near the center, which is sufficiently away from the supported ends. The film thickness ($l \sim 0.1\,\mu$m) and the radiation penetration depth ($\delta \sim 0.01\,\mu$m) are both small comparing to the laser beam diameter, supporting a one-dimensional analysis for the thermomechanical response at short times. The coordinate x is measured downward from the front (heated) surface of the film, and the displacement in the x-direction is denoted by u. The one-dimensional form of Eq. (2) becomes

$$\rho \frac{\partial^2 u}{\partial t^2} - E_0 \frac{\partial^2 u}{\partial x^2} = \Lambda \left(T_e \frac{\partial T_e}{\partial x} \right), \tag{4}$$

where $E_0 \equiv E_{1111}$ and $\Lambda \equiv \Lambda_{11}$ are the only components of E_{ijlm} and Λ_{ij} that appear in the one-dimensional equation of motion. Two types of one-dimensional approximations must be noted. For a slender (wire-like) rod that has zero constraint in the lateral direction, E_0 is the regular Young's modulus (E). For a thin film in which lateral constraints are imposed on the deformation developed in the x-direction, $E_0 = E(1-\nu)/(1+\nu)(1-2\nu)$ with ν being the Poisson ratio. Effect of compressibility is thus included in E_0, with $\nu = 1/2$ representing the case of an incompressible medium.

The one-dimensional form of Eq. (3) is

$$\left(\frac{C_{e0}}{T_0} \right) T_e \frac{\partial T_e}{\partial t} = k_0 \frac{\partial}{\partial x} \left(\frac{T_e}{T_l} \frac{\partial T_e}{\partial x} \right) - G(T_e - T_l) + 0.94J \left(\frac{1-R}{\delta t_p} \right) \times \exp\left[-\frac{x}{\delta} - 4\ln(2) \left(\frac{t}{t_p} \right)^2 \right];$$
$$C_l \frac{\partial T_l}{\partial t} = G(T_e - T_l). \tag{5}$$

The thin metal film is assumed heated from a stationary state (zero initial displacement and velocity), whereas electrons and phonons are in thermal equilibrium at room temperature (T_0):

$$T_e = T_0, \quad T_l = T_0, \quad u = 0, \quad \text{and} \quad \frac{\partial u}{\partial t} = 0 \quad \text{as} \quad t = t_0 = -2t_p \tag{6}$$

with t_0 particularly chosen to describe the temporal shape of the laser pulse by a Gaussian distribution peaking at $t = 0$, with a full width at half maximum (FWHM) duration t_p [12–14,19–22]. During the picosecond transient, heat losses from the front and rear surfaces of the film can be neglected:

$$\frac{\partial T_e}{\partial x} = 0, \quad \frac{\partial T_l}{\partial x} = 0 \quad \text{at} \quad x = 0 \quad \text{and} \quad x = l. \tag{7}$$

Mechanically, both front and rear surfaces of the film are assumed free of deformation,

$$\frac{\partial u}{\partial x} = 0 \quad \text{at} \quad x = 0 \quad \text{and} \quad l, \tag{8}$$

which corresponds to the stress-free boundary conditions in the elastic response already assumed in Eq. (4).

Numerous thermomechanical properties are involved in Eqs. (4) to (8). To identify the dominating groups that characterize the ultrafast deformation, the following nondimensional parameters are introduced:

$$\xi = \frac{x}{\delta}, \quad \beta = \frac{t}{t_p}, \quad \theta_{e(l)} = \frac{T_{e(l)}}{T_0}, \quad U = \frac{u}{(\Lambda T_0^2 t_p^2 / \rho\delta)},$$
$$Y = \frac{E_0/\rho}{(\delta/t_p)^2}, \quad \beta_0 = \frac{t_0}{t_p},$$
$$S = 0.94J \left(\frac{1-R}{\delta C_{e0} T_0} \right), \quad H = \frac{G t_p}{C_{e0}}, \quad K = \frac{\alpha_0 t_p}{\delta^2}, \quad C = \frac{C_{e0}}{C_l},$$
$$L = \frac{l}{\delta}, \quad \alpha_0 = \frac{k_0}{C_{e0}}. \tag{9}$$

Equations (4) to (8) become

$$\frac{\partial^2 U}{\partial \beta^2} - Y \frac{\partial^2 U}{\partial \xi^2} = 2\theta_e \frac{\partial \theta_e}{\partial \xi}; \tag{10}$$

$$\theta_e \frac{\partial \theta_e}{\partial \beta} = K \frac{\partial}{\partial \xi} \left(\frac{\theta_e}{\theta_l} \frac{\partial \theta_e}{\partial \xi} \right) - H(\theta_e - \theta_l) + S \exp[-\xi - 4\ln(2)\beta^2];$$
$$\frac{\partial \theta_l}{\partial \beta} = HC(\theta_e - \theta_l); \tag{11}$$

$$\theta_e = 1, \quad \theta_l = 1, \quad U = 0, \quad \text{and} \quad \frac{\partial U}{\partial \beta} = 0 \quad \text{as} \quad \beta = -2 \tag{12}$$

$$\frac{\partial \theta_e}{\partial \xi} = 0, \quad \frac{\partial \theta_l}{\partial \xi} = 0, \quad \frac{\partial U}{\partial \xi} = 0 \quad \text{at} \quad \xi = 0 \quad \text{and} \quad \xi = L. \tag{13}$$

Clearly, ultrafast heating is characterized by three parameters. They are (1) parameter K, the squared ratio of the diffusion length of electrons at room temperature to the radiation penetration depth, (2) parameter H, the amount of thermal energy exchange between electrons and phonons per volume in one pulse t_p relative to the volumetric heat capacity of electrons at room temperature, and (3) parameter C, the ratio of volumetric specific heat between electrons and phonons at room temperature. Ultrafast deformation is further characterized by an additional parameter Y, which is the squared velocity ratio between the dilatational wave speed $\sqrt{E_0/\rho}$ and (δ/t_p). The nondimensional intensity (S) and film thickness (L), in addition, are two external parameters characterizing the heating condition and geometry of the metal film. Mathematically, Eqs. (10) and (11) display a situation of ''one-way'' coupling, in the sense that electron and phonon temperatures (θ_e and θ_l) can be obtained independently from Eq. (11), without interfering with the displacement of acoustic phonons (U) described by Eq. (10). The electron temperature (θ_e) thus obtained is then fed into Eq. (10) for determining the acoustic displacement.

Equations (10) and (11) are solved by modifying the subroutine MOLCH in the IMSL numerical package. It employs the method of lines, a collocation method in nature, which reduces a system of coupled partial differential equations (PDEs) to a system of coupled ordinary differential equations (ODEs). All PDEs must be expressed in a standard form of a diffusion equation containing the first-order derivative in time (β). Since Eq. (10) is already in the standard form, Eq. (11) is decomposed into two coupled diffusion equations to satisfy this requirement:

$$\begin{cases} \dfrac{\partial U}{\partial \beta} = V \\ \dfrac{\partial V}{\partial \beta} = Y\theta_e^2 \dfrac{\partial^2 U}{\partial \xi^2} + 2\theta_e \dfrac{\partial \theta_e}{\partial \xi} \end{cases} \tag{14}$$

Solutions for θ_e, θ_l, U, and V are expanded in terms of the cubic Hermite polynomials (of ξ) with time-dependent coefficients in

the space domain (ξ) that is divided into subintervals separated by a number of breakpoints. Continuities of the standard basis functions for the cubic Hermite polynomials and their first-order derivative in ξ are imposed at all breakpoints. The time-dependent coefficients are determined by ensuring the approximate solutions satisfy all differential equations at the two Gaussian points in each subinterval, which results in a system of coupled ODEs governing the time-dependent coefficients in the approximate solutions. Another subroutine, IVPAG in IMSL, is then used to solve such a system of coupled ODEs for the time-dependent coefficients. The fifth-order Gear's backward differentiation formulas is selected to suppress the stiffness in such a set of ODEs. An adaptive algorithm is employed to continuously increase the number of subintervals dividing the space domain, until the maximum error in the changes of θ_e, θ_l, U, and V is less than 1.0×10^{-5}. The time step used in BDF is determined adaptively, which turns out to be less than 0.05 for all cases in this work.

Numerical Results

The electron temperature (θ_e) is first compared to the numerical result obtained by Qiu and Tien [2] for a 0.02 μm thin gold film heated by a 100 fs (0.1 ps), with a wavelength typically from 0.2 to 0.7 μm. The properties of gold and the heating conditions used in the analysis are $k_0=315$ W/mK, $C_{e0}=2.1\times10^4$ J/m^{3}K, $C_l=2.5\times10^6$ J/m^{3}K, $G=2.6\times10^{16}$ J/m^{3}K, $l=0.02$ μm, $t_p=0.1$ ps, $\delta=15.3$ nm, $J=13.4$ J/m^2, $R=0.93$, and $T_0=300$ K, which, in terms of the nondimensional groups, result in $K\cong6.41$, $L\cong1.307$, $H\cong0.124$, $C=8.4\times10^{-3}$, and $S=9.147$ in Eq. (11). These values are used in all numerical calculations below unless stated otherwise. The normalized temperature change at the front (heated) surface of the film was examined by the method of lines. The result agrees very well with that obtained by the semi-implicit Crank-Nicholson algorithm [10].

Emphasizing non-equilibrium heating during the picosecond transient, Fig. 1 displays the electron and phonon temperatures in degrees Kelvin. For $S=9.147$ ($J=13.4$ J/m^2), Fig. 1(a), electrons shoot to a temperature of about 1000 K at the peak of the laser intensity, while the lattices gradually rise from 300 K (the assumed initial temperature) to 313.6 K at 7.1 ps. The metal lattice remains cold in this time domain. Electron temperature may drastically increase to 10^3 to 10^4 K as the laser intensity increases, exemplified by Fig. 1(b) for $S=345$ ($J=505.4$ J/m^2). The peak temperature of electrons reaches 6000 K, approximately, while thermal equilibrium between electrons and phonons is prolonged until 24.5 ps at 825 K. The dilatational rate effect on the lattice temperature, which is not included in Fig. 1(b) for the time being, needs to be accommodated in this case due to the excessively high-rate heating experienced by the metal lattice [11]. Femtosecond lasers with such a high intensity is mainly for the purpose of material removal rather than thermal processing, necessitating a simultaneous inclusion of the temperature-gradient and strain-rate effects in the energy and momentum equations. We restrict ourselves in this work to the value of $J=13.4$ J/m^2 ($S=9.147$) unless stated otherwise.

Electron temperatures are more sensitive to parameters H (nondimensional phonon-electron coupling factor) and L (nondimensional film thickness). As shown in Fig. 2(a), where $\beta=1$ is equivalent to $t=0.1$ ps and $\theta_e=1$ is equivalent to $T_e=300$ K, the electron temperature significantly decreases as the value of H ($\sim G$) increases. A larger value of H implies a higher rate of energy exchange between electrons and phonons. As a result, the electron temperature is lower and thermal equilibrium (between phonons and electrons) is achieved earlier at a larger value of H, at approximately $\beta=8$ ($t=0.8$ ps) for $H=1.24$. A smaller value of K implies a shorter diffusion length of electrons over t_p comparing to the depth of laser penetration (δ), as defined in Eq. (9). The laser energy is consequently confined within a domain closer the heated surface, which results in a higher peak temperature of electrons as shown in Fig. 2(b). As the value of K enters the order of one and beyond, however, no significant change of θ_e is observed, even when the K value is increased by two orders of magnitude.

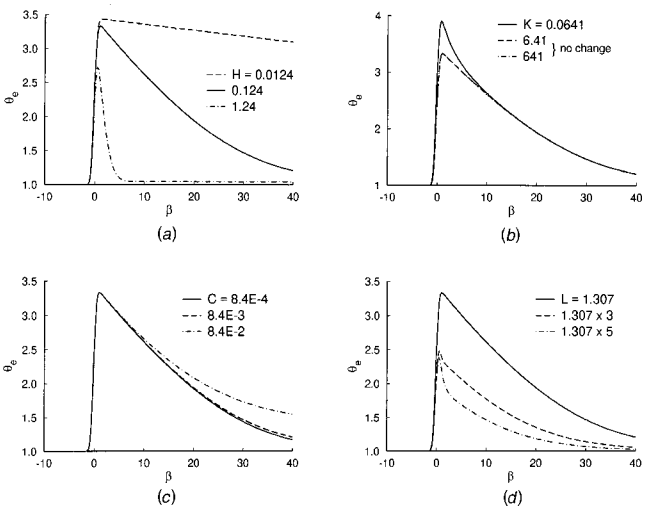


Fig. 2 Effect of (*a*) *H*, (*b*) *K*, (*c*) *C*, and (*d*) *L* on electron temperatures at the front (heated) surface of the thin metal film ($\xi=0$)

A smaller value of C, Fig. 2(c), implies a larger heat capacity of phonons relative to that of electrons. The metal lattices in this case absorb the thermal energy transported from hot electrons more effectively, resulting in a lower temperature of electrons, as expected. A thicker film (a larger value of L) provides a wider physical domain for phonon-electron interactions to take place. Consequently, the surface temperature of electrons becomes significantly lower in thicker films, as shown in Fig. 2(d). While having a similar pattern of thermalization beyond 1 ps ($\beta>10$), however, a thicker film results in a faster thermalization *rate* as the electron temperature departs from the peak at short times ($\beta<3$, approximately).

Note that more pronounced effects of H, K, C, and L shown in Fig. 2 only appear during the post-peak thermalization between electrons and phonons. During the heating stage of electrons, for $\beta<1$ or $t<0.1$ ps approximately, electron temperatures are basically the same at all the values of H, K, C, and L.

Under the same conditions in Fig. 1(a), the nondimensional driving force represented by $2\theta_e(\partial\theta_e/\partial\xi)$ is shown in Fig. 3. Significant amplitude induced by the femtosecond laser ($t_p=0.1$ ps) is only present for $\beta<1$. The insulated conditions im-

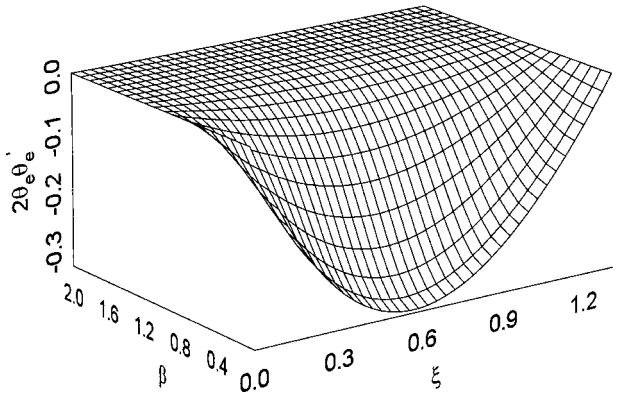


Fig. 3 Compressive driving force, $2\theta_e(\partial\theta_e/\partial\xi)$ in Eq. (10), in gold films heated by a 100 fs laser. Same conditions as those in Fig. 1(*a*).

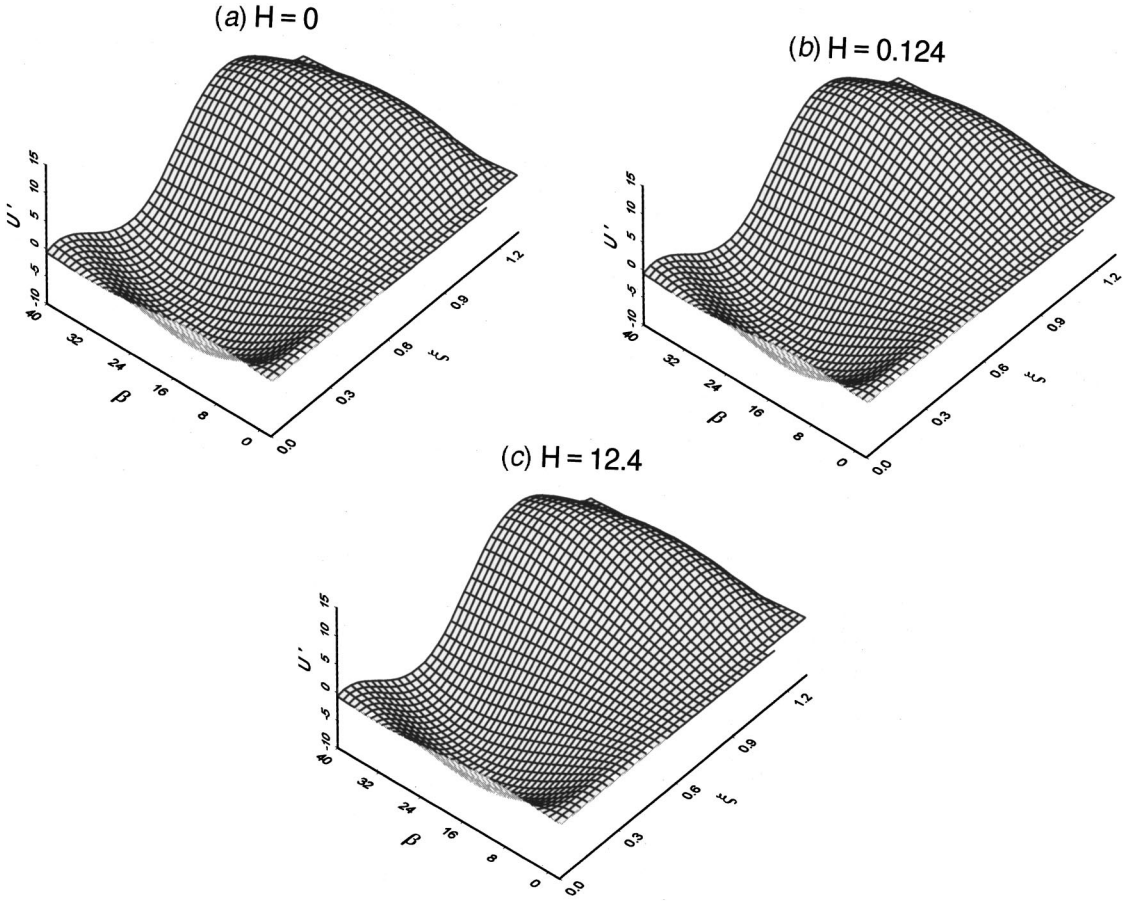


Fig. 4 Straining patterns, $U'(\xi)$, developed in the thin gold film with different values of H. Same conditions as those in Fig. 1(a).

posed at the front and rear surfaces of the film, $\partial\theta_e/\partial\xi=0$ at $\xi=0$ and L (1.307) in Eq. (13), result in zero driving force at both surfaces of the film. The driving force is negative (compressive) throughout the film, peaking at $\xi\cong 0.58$ closer to the heated (front) surface. During the heating stage of $\beta<1$, as indicated in Fig. 2, both θ_e and $(\partial\theta_e/\partial\xi)$ are insensitive to the values of H, K, C, and L. The distribution of the driving force shown in Fig. 3, therefore, applies to all values of H, K, C, and L considered in this study.

The compressive driving force results in compressive strains near the heated surface, as shown in Figs. 4(a) to 4(c) in the domain of β from 0 to 40 ($0<t<4$ ps). For gold with Λ being of the same order of magnitude as the density of state of electrons, parameter $Y\cong 1.71\times 10^{-4}$, which is used in Fig. 4 throughout. The value of H is varied from 0 (no exchange of thermal energy between electrons and phonons), 0.124 (a typical value for gold), to 12.4 (the threshold values for Ti, Nb, and V whose G values are greater than the value of gold by two orders of magnitude) to show the weak dependence of ultrafast straining (induced by the hot electron blast during the picosecond transient) on the thermal properties. The nondimensional strain pattern $(U'(\xi,\beta))$ stays the same as the value of H increases by several orders of magnitude. The front ($\xi=0$) and rear ($\xi=L=1.307$) surfaces of the film are maintained strain-free, $U'=0$ as required by the boundary conditions in Eq. (13).

Although laser heating results in compression during the initial contact, both compressive and tensile regions can be observed in Fig. 4. The early-time compression $(U'<0)$ drastically switches into tension $(U'>0)$ across the mechanical wavefront propagating along the characteristic line $x=\sqrt{E_0/\rho}t$, or $\xi=\sqrt{Y}\beta$ in terms of ξ and β. Such a wave behavior in mechanical straining is dictated by Eq. (4), while alterations from compression to tension is commanded by the thermally induced wavefront, as further illustrated in Fig. 5 that views Fig. 4(b) from the $\xi-U'$ plane. For $Y\cong 1.71\times 10^{-4}$, the nondimensional mechanical wave speed is approximately 0.0131 and the mechanical wavefront is located at $\xi\cong 0.0131\beta$, laying on the plane defined by $(\xi,\beta)=(0,0)$ and (0.523,40). This plane is represented by the two dash lines marked in Fig. 5 for clearer visualization. The physical domain behind the wavefront, $\xi<0.0131\beta$ to the right of this plane, is entirely in compression due to the initial effect of heating. Such compressive strain reaches a maximum value at $\xi\cong 0.0131\beta$, the mechanical wavefront. Crossing the wavefront and approaching the rear end of the film, in the domain of $\xi>0.0131\beta$ to the left of the dashed

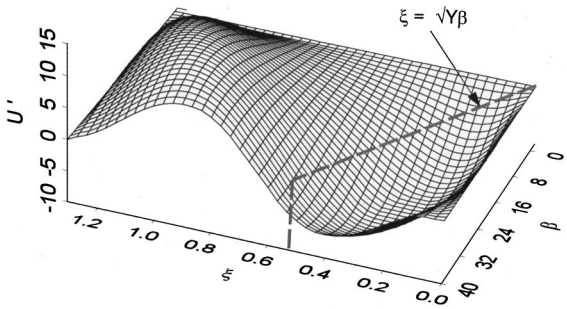


Fig. 5 Maximum compressive strain at the mechanical wavefront at $\xi=\sqrt{Y}\beta$. Fig. 6(b) viewing from the $\xi-U'$ plane.

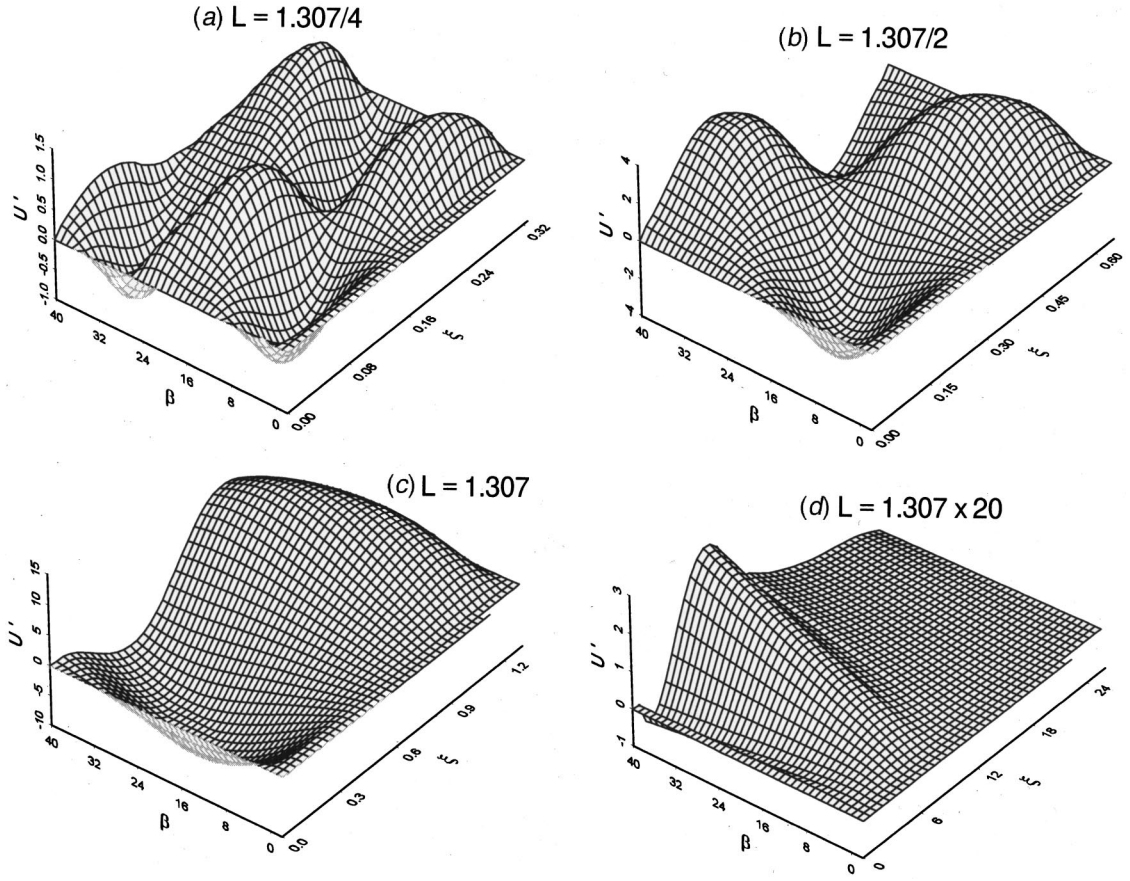


Fig. 6 Straining patterns, $U'(\xi)$, developed in the thin gold film with different thickness (L). Same conditions as those in Fig. 1(a).

plane, the magnitude of compression rapidly decreases and the compression switches into tension in several tenths of δ (note that $\xi=1$ is equivalent to $x=15.3$ nm). As a general trend, metal lattices are weaker in resisting tension than compression. The maximum value in tension, $U'\cong 13.11$ at $(\xi,\beta)\cong(0.96,25.5)$ shown in Fig. 5, therefore, is most noteworthy since it poses a critical stage for *internal* tensile damage should the laser intensity become exceedingly high.

Equation (11) describes ultrafast thermalization between phonons and electrons in the picosecond domain, which is diffusive in nature. Inheriting the general nature of diffusion, the heat affected zone (of electrons) spreads over the entire film and the driving force exerts on all metal lattices throughout the film, including those beyond the mechanical wavefront. The highly elevated tensile strain in front of the mechanical wave, as shown in Figs. 4 and 5, is comparable in magnitude to the compressive strain behind the wavefront, which is a distinct feature that cannot be described by the classical thermoelastic waves employing Fourier's law [11].

Effect of film thickness (L) on the ultrafast deformation is displayed in Fig. 6. The film thickness is reduced by four times in Fig. 6(a), and increased by 20 times in Fig. 6(d), to study the thickness effects in general. Since the dilatational wave speed remains constant by neglecting the renormalization effect in the metal lattices, the mechanical waves (induced by ultrafast heating) bounce back and forth more frequently in a thinner film. Repetitive interactions between the thermally induced and the reflected waves from the two surfaces of the film produce a capricious deformation pattern, as shown in Fig. 6(a) for $L\cong 0.327$. In the short time domain with $t<4$ ps ($\beta<40$), three pairs of compressive and tensile waves are present, resulting from the interactions of incipient waves ($\xi=0.0131\beta$) and reflected waves ($\xi=-0.0131\beta$) whose characteristic lines are perpendicular to each other. The number of deformation ripples decreases as the film thickness increases, Figs. 6(b) and 6(c), due to reduction of the number of traveling waves in the film. The tension-compression waves reduce from two pairs to a couple as the film thickness (L) increases from 0.654 (Fig. 6(b)) to 1.307 (Fig. 6(c), same as Fig. 4(b)). Meanwhile, the amplitude of the interacting strain waves increases as the number of the interacting pairs decreases. Fig. 6(d) further stretches the film thickness to twenty times of the reference value, $L\cong 26.14$, to eliminate the effects of reflected waves from the rear surface. As clearly shown, tension preceded by compression near the heated surface is a preserved feature regardless of the reflected waves. The amplitude of deformation, either in compression or in tension, however, is smaller due to the absence of excitations by the reflected wave.

Parameter Y, the nondimensional dilatational wave speed, significantly changes the ultrafast deformation pattern since it dictates the frequency at which the thermally induced mechanical waves travel between the front (heated) and rear surfaces of the film. Under a constant film thickness, a larger value of Y implies a faster speed of the dilatational wave, resulting in more deformation ripples traveling in the film. This is shown from Fig. 7(a), where the value of Y is reduced by one order of magnitude from the reference value (1.71×10^{-4}), to Fig. 7(d), where the value of Y is increased by two orders of magnitude. The deformation patterns are similar to those observed in Fig. 6, with the effect of a larger value of Y (under a constant film thickness L) in correspondence with the effect of a smaller film thickness (under a constant dilatational speed Y). The number of interacting waves traveling in the film, clearly, is a dominating factor in the evolution of the

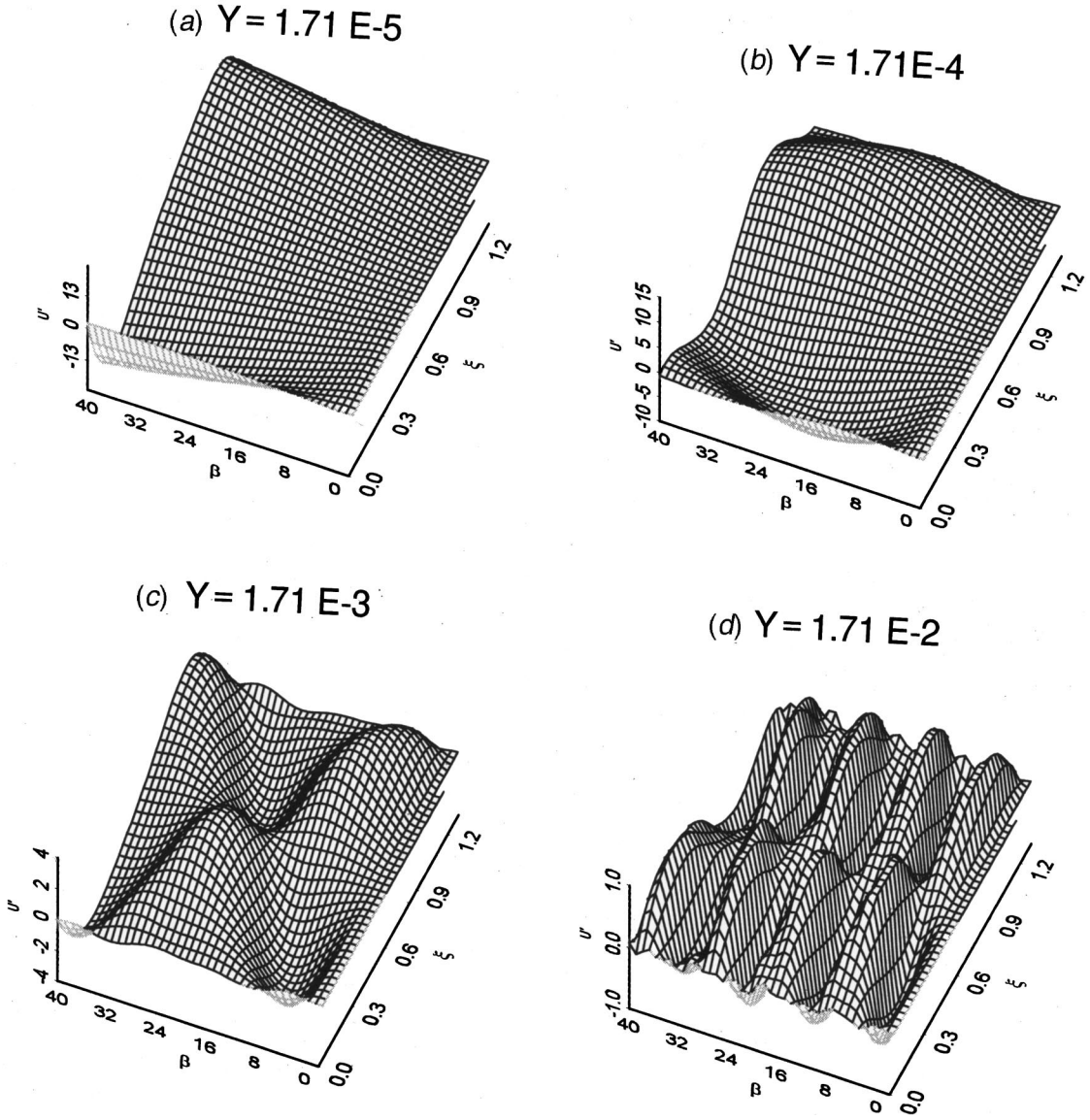


Fig. 7 Straining patterns, $U'(\xi)$, developed in the thin gold film with different values of Y. Same conditions as those in Fig. 1(a).

ultrafast deformation pattern. As such a number increases, either due to reduced film thickness (a smaller value of L) or increased film rigidity (a larger value of Y), multiple damage could be resulted near the tensile peaks.

Ultrafast deformation strongly depends on the constraints imposed on the film surfaces. Fig. 8 shows the stain patterns developed in the thin film heated at the front surface ($\xi=0$) but fixed at the rear surface ($\xi=L$), which simulates the local response in thermal processing of thin films coated on a hard (rigid) substrate. The freely deformed condition at the front surface remains the same, $U'=0$ at $\xi=0$, but the fixed-edge condition, $U=0$ at $\xi=L$, replaces the last condition in Eq. (13). The solution procedure remains the same regardless of the change in this boundary condition. In a shorter domain of time, Fig. 8(a) for $0<\beta<40$, compression prevails near the heated surface. The tensile waves developed in front of the mechanical waves, which was shown earlier in Fig. 5, however, highly elevate the strain level when approaching the fixed (rear) surface of the film. The magnitude of tensile strains thereby is much larger than the compressive strains near the front surface, peaking at $U'\cong 24.6$ as $\beta\cong 40$. As transient time lengthens, $0<\beta<100$ in Fig. 8(b), another characteristic line is present at $\xi=-0.0131\beta$ due to the reflected waves from the rear surface. Alteration from compression to tension across the mechanical wavefront is similar to that depicted in Fig. 5, except for a change of the characteristic line from $\xi=0.0131\beta$ to $\xi=-0.0131\beta$. Exaggerated strains, in tension, exist to the left of the characteristic line, at $(\xi,\beta)\cong(0.8,100)$, as well as at the rear surface of film at $(\xi,\beta)\cong(1.307,54.5)$. More strain ripples are generated in longer time, as shown by Fig. 8(c) where β is further extended to 250. Frequent interactions among the reflected waves result in repetitive tension-compression cycles, which produce high frequency, fatigue-like straining near the rear surface.

Conclusion

Ultrafast deformation studied in this work is a highly nonequilibrium phenomenon in metals that cannot be depicted by classical thermoelasticity. It is induced by the hot-electron blast in the sub-picosecond domain, where electrons shoot to a very high temperature while metal lattices remain almost thermally undisturbed, in contrast to the large temperature gradient across the metal lattices as described in the classical thermoelasticity. The

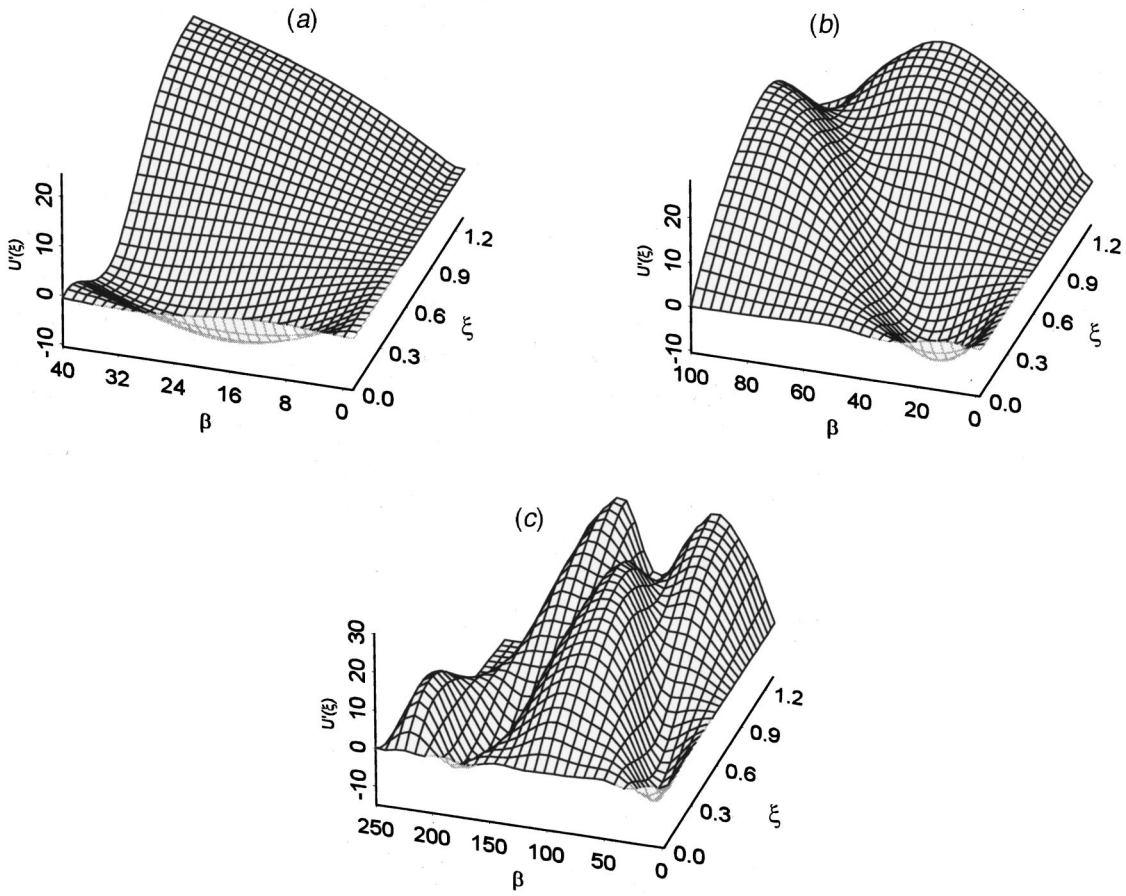


Fig. 8 Straining patterns, $U'(\xi)$, developed in the thin gold film with a fixed surface at $\xi = L$ in (*a*) $0<\beta<40$, (*b*) $0<\beta<100$, and (*c*) $0<\beta<250$. Same conditions as those in Fig. 1(*a*).

applicable domain for the hot electron blast model ranges from sub-picosecond to a few picoseconds, where electron-heating results in a driving force that sets the cold lattices in motion. Such a driving force is proportional to both the temperature and temperature gradient in the hot electron gas. It produces compression when heated by the femtosecond laser, switching into tension with a comparable magnitude when traversing across the mechanical wavefront.

Ultrafast heating and deformation in thin metal films are characterized by four parameters: the nondimensional length of diffusion of electrons at room temperature (K), the nondimensional phonon-electron coupling factor (H), the heat-capacity ratio of electrons (at room temperature) to phonons (C), and the nondimensional dilatational wave speed (Y). The ultrafast straining patterns result from the combined effects of these parameters, along with the heating intensity (S) and film thickness (L) as external parameters. Their individual effects may not be separable in general.

The failure patterns are dictated by the number of strain ripples traveling in the metal film during the picosecond transient. Single-site failure is likely to occur in a thicker film with a slower speed of dilatational waves. Multiple-site failure, on the other hand, tends to develop in a thinner film with a faster speed of dilatational waves. In both cases, of courses, the laser energy must be sufficiently high to produce an abundant driving force that pushes the strain level to the damage threshold. Ultrafast deformation strongly depends on the conditions imposed on the surfaces of the film. For metal films with freely deforming surfaces, damage always occurs in the interior. For metal films with a fixed surface, on the other hand, damage tends to develop in the vicinity of the fixed surface. In addition to the heating intensity of lasers, the amplitude of mechanical strain, in both tension and compression, depends on the values of (Λ/ρ) and $(T_0 t_p/\delta)^2$. Along with the magnitude of the driving force (Λ), therefore, the initial temperature (T_0), mass density of the metal film (ρ), pulse duration of the femtosecond laser (t_p), and the depth of laser penetration (δ) need to be considered altogether to characterize the damage developed in the picosecond domain.

Transition of the deformation pattern from the stage of the hot electron blast (the present work) to the stage of lattice heating (classical thermoelasticity) is an ongoing effort. Lattice heating, including both the high strain-rate effect in transporting heat and the lattice expansion effect in transmitting load, occurs later, which may produce a stronger effect than the hot electron blast. A metal target that can sustain the hot electron blast in sub-picoseconds, therefore, is not a guarantee that the target can survive through the more severe thermomechanical coupling developed in tens/hundreds of picoseconds. We shall report our findings on this subject soon.

References

[1] Brorson, S. D., Fujimoto, J. G., and Ippen, E. P., 1987, "Femtosecond Electron Heat-Transport Dynamics in Thin Gold Film," Phys. Rev. Lett., **59**, pp. 1962–1965.

[2] Qiu, T. Q., Juhasz, T., Suarez, C., Bron, W. E., and Tien, C. L., 1994, "Femtosecond Laser Heating of Multi-Layered Metals—II. Experiments," Int. J. Heat Mass Transf., **37**, pp. 2799–2808.

[3] Elliott, D. J., and Piwczyk, B. P., 1989, "Single and Multiple Pulse Ablation of Polymeric and High Density Materials With Excimer Laser Radiation at 193 nm and 248 nm," Mater. Res. Soc. Symp. Proc., **129**, pp. 627–636.

[4] Knapp, J. A., Børgesen, P., and Zuhr, R. A., eds., 1990, "*Beam-Solid Interactions: Physical Phenomena*," Mater. Res. Soc. Symp. Proc., **157**, Pittsburgh, PA.

[5] Mandelis, A., and Peralta, S. B., 1992, ''Thermal-Wave Based Materials Characterization and Nondestructive Evaluation of High-Temperature Superconductors: A Critical Review,'' *Physics and Materials Science of High Temperature Superconductors II*, R. Kossowsky, ed, Kluwer, Boston, MA, pp. 413–440.
[6] Narayan, J., Godbole, V. P., and White, G. W., 1991, ''Laser Method for Synthesis and Processing of Continuous Diamond Films on Non-Diamond Substrates,'' Science, **52**, pp. 416–418.
[7] Grigoropoulos, C. P., 1994, ''Heat Transfer in Laser Processing of Thin Films,'' *Annual Review of Heat Transfer V*, C. L. Tien, ed., Hemisphere, New York.
[8] Department of Energy (DOE) Basic Energy Sciences (BES) Workshop, 1999, ''Complex Systems—Science for the 21st Century,'' www.sc.doe.gov/production/bes/complexsystems.htm.
[9] Department of Energy (DOE) Basic Energy Sciences (BES) Workshop, 1999, ''Nanoscale Science, Engineering, and Technology—Research Direction,'' www.sc.doe.gov/production/bes/nanoscale.html.
[10] Qiu, T. Q., and Tien, C. L., 1994, ''Femtosecond Laser Heating of Multi-Layered Metals—I. Analysis,'' Int. J. Heat Mass Transf., **37**, pp. 2789–2797.
[11] Tzou, D. Y., 1997, *Macro to Microscale Heat Transfer: The Lagging Behavior*, Taylor & Francis, Washington, DC.
[12] Kaganov, M. I., Lifshitz, I. M., and Tanatarov, M. V., 1957, ''Relaxation Between Electrons and Crystalline Lattices,'' Sov. Phys. JETP, **4**, pp. 173–178.
[13] Anisimov, S. I., Kapeliovich, B. L., and Perel'man, T. L., 1974, ''Electron Emission From Metal Surfaces Exposed to Ultra-Short Laser Pulses,'' Sov. Phys. JETP, **39**, pp. 375–377.
[14] Qiu, T. Q., and Tien, C. L., 1992, ''Short-Pulse Laser Heating on Metals,'' Int. J. Heat Mass Transf., **35**, pp. 719–726.
[15] Qiu, T. Q., and Tien, C. L., 1993, ''Heat Transfer Mechanisms During Short-Pulse Laser Heating of Metals,'' ASME J. Heat Transfer, **115**, pp. 835–841.
[16] Joseph, D. D., and Preziosi, L., 1989, ''Heat Waves,'' Rev. Mod. Phys., **61**, pp. 41–73.
[17] Joseph, D. D., and Preziosi, L., 1990, ''Addendum to the Paper on Heat Waves,'' Rev. Mod. Phys., **62**, pp. 375–391.
[18] Tzou, D. Y., 1992, ''Thermal Shock Phenomena under High-Rate Response in Solids,'' *Annual Review of Heat Transfer IV*, chap. 3, C. L. Tien, ed., Hemisphere Washington, DC, pp. 111–185.
[19] Özisik, M. N., and Tzou, D. Y., 1994, ''On the Wave Theory in Heat Conduction,'' ASME J. Heat Transfer, **116**, pp. 526–535.
[20] Tzou, D. Y., 1995, ''A Unified Field Approach for Heat Conduction from Micro to Macro-Scales,'' ASME J. Heat Transfer, **117**, pp. 8–16.
[21] Tzou, D. Y., 1995, ''The Generalized Lagging Response in Small-Scale and High-Rate Heating,'' Int. J. Heat Mass Transf., **38**, pp. 3231–3240.
[22] Tzou, D. Y., 1995, ''Experimental Support for the Lagging Response in Heat Propagation,'' AIAA J Thermophysics Heat Transfer, **9**, pp. 686–693.
[23] Falkovsky, L. A., and Mishchenko, E. G., 1995, ''Surface Excitations in Metals: Brillouin and Raman Light Scattering,'' Phys. Rev. B, **51**, pp. 7239–7249.
[24] Mishchenko, E. Zh., and Falkovsky, L. A., 1995, ''Long-Wavelength Optical Phonons: Damping, Surface Oscillations, and Raman Scattering,'' J. Exp. Theor. Phys., **80**, pp. 531–538.
[25] Hetnarski, R. B., and Ignaczak, J., 1999, ''Generalized Thermoelasticity,'' J. Therm. Stresses, **22**, pp. 451–476.
[26] Hetnarski, R. B., and Ignaczak, J., 2000, ''Nonclassical Dynamical Thermoelasticity,'' Int. J. Solids Struct., **37**, pp. 215–224.
[27] Falkovsky, L. A., and Mishchenko, E. G., 1999, ''Electron-Lattice Kinetics of Metals Heated by Ultrashort Laser Pulses,'' J. Exp. Theor. Phys., **88**, pp. 84–88.
[28] Myers, H. P., 1997, *Introductory Solid State Physics*, Taylor & Francis, Bristol, PA.

Non-Equilibrium Phase Change in Metal Induced by Nanosecond Pulsed Laser Irradiation

Xianfan Xu[1]
e-mail: xxu@ecn.purdue.edu

David A. Willis[2]

School of Mechanical Engineering,
Purdue University,
West Lafayette, IN 47907

Materials processing using high power pulsed lasers involves complex phenomena including rapid heating, superheating of the laser-melted material, rapid nucleation, and phase explosion. With a heating rate on the order of 10^9 K/s or higher, the surface layer melted by laser irradiation can reach a temperature higher than the normal boiling point. On the other hand, the vapor pressure does not build up as fast and thus falls below the saturation pressure at the surface temperature, resulting in a superheated, metastable state. As the temperature of the melt approaches the thermodynamic critical point, the liquid undergoes a phase explosion that turns the melt into a mixture of liquid and vapor. This article describes heat transfer and phase change phenomena during nanosecond pulsed laser ablation of a metal, with an emphasis on phase explosion and non-equilibrium phase change. The time required for nucleation in a superheated liquid, which determines the time needed for phase explosion to occur, is also investigated from both theoretical and experimental viewpoints. [DOI: 10.1115/1.1445792]

Keywords: Ablation, Experimental, Heat Transfer, Laser, Phase Change

1 Introduction

High power lasers are being used in a variety of advanced engineering applications, including micromachining, pulsed laser deposition (PLD) of thin films, and fabrication of nanometer size particles and carbon nanotubes [1–4]. Most of these processes involve complex thermal phenomena, including rapid heating, non-equilibrium phase change, superheating, and rapid nucleation in a superheated liquid. Under intense radiation of a laser pulse, the surface of a target is heated with a heating rate of 10^9 K/s or higher. At laser fluences (energy per unit area) of about 1 J/cm^2 or higher, melting and ablation (rapid removal of material) will occur.

There are several mechanisms of laser thermal ablation, namely normal evaporation at the surface, heterogeneous boiling, and homogeneous boiling. During high power pulsed laser ablation, homogeneous boiling, or phase explosion could be an important ablation mechanism [5–8]. The phase explosion phenomenon was first investigated in the earlier work of rapid heating of metal wires using a high current electric pulse [9,10]. It occurs when a liquid is rapidly heated and approaches the thermal dynamic critical temperature, and the instability in the liquid causes an explosive type of liquid-vapor phase change. Miotello and Kelly [5] pointed out that phase explosion was a likely mechanism in nanosecond pulsed laser ablation. Song and Xu [6] were the first to provide experimental evidence of phase explosion induced by a nanosecond pulsed laser. They also showed that surface temperature-pressure relation could depart from the equilibrium Clausius-Clapeyron relation [11,12]. It has also been suggested that phase explosion occurred during sub-picosecond laser ablation [13]. Using molecular dynamics simulations, Zhigelei et al. showed phase explosion occurred when the laser fluence was above a threshold value, while surface desorption occurred at lower laser fluences [14].

This paper is concerned with energy transport and phase change in metal induced by a high power nanosecond pulsed laser, with an emphasis on phase change mechanisms and nonequilibrium phase change kinetics at the evaporating surface. Phase explosion induced by rapid heating will be described first. A brief review of the experimental evidence of phase explosion will then be given. The experiments were performed in the laser fluence range from 2.5 J/cm^2 to 9 J/cm^2, which is commonly used for many applications including PLD and micromachining. The nucleation process in liquid leading to phase explosion is discussed in detail.

2 Thermal Mechanisms of Laser Ablation and Phase Explosion

The phase change process induced by pulsed laser heating can be best illustrated using the pressure-temperature diagram as shown in Fig. 1 [7]. The ''normal heating'' line indicates heating of a liquid metal when the temperature is below the boiling temperature. At the boiling temperature, the liquid and vapor phases are in equilibrium, which is shown in Fig. 1 as the intersection between the normal heating line and the binode line. The binode line represents the equilibrium relation between the surface temperature and the vapor saturation pressure, which is calculated from the Clausius-Clapeyron equation. Evaporation occurs at the liquid surface, which is a type of heterogeneous evaporation, or normal surface evaporation.

The surface evaporation process can be computed. The rate of atomic flux (atoms/m^{2}s) leaving the surface during normal evaporation is given as [15]:

$$\dot{m} = \frac{p_s}{(2\pi m k_B T)^{1/2}}, \tag{1}$$

where $\dot{m}$ is the mass of the evaporating molecule or atom, k_B is the Boltzmann constant, and p_s is the saturation pressure at the liquid surface temperature T, which are related by the Clausius-Clapeyron equation:

$$p_s = p_o \exp\left\{\frac{H_{lv}(T-T_b)}{RTT_b}\right\}. \tag{2}$$

In Eq. (2), p_o is the ambient pressure, H_{lv} is the enthalpy of vaporization, and T_b is the equilibrium boiling temperature at the ambient pressure (the normal boiling temperature).

[1]Corresponding author. Phone: (765) 494-5639, Fax: (765) 494-0539.
[2]Current address: Department of Mechanical Engineering, Southern Methodist University, Dallas, TX 75275.

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division April 30, 2001; revision received November 16, 2001. Associate Editor: V. P. Carey.

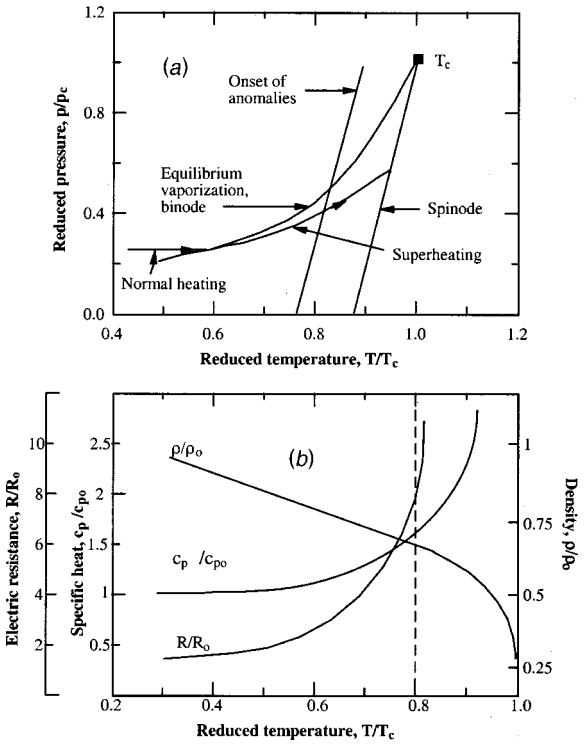


Fig. 1 (*a*) *p-T* Diagram and (*b*) typical variations of physical properties of liquid metal near the critical point. The substrate "o" denotes properties at the normal boiling temperature.

In a slow heating process, the surface temperature-pressure relation follows Eq. (2). On the other hand, when the heating rate is high enough such as what occurs during high power pulsed laser heating, it is possible to superheat a liquid metal to temperatures above the boiling point while the surface vapor pressure is not built up as rapidly. The liquid is then superheated, i.e., its temperature is higher than the vaporization temperature corresponding to its surface pressure. In this case, the heating process deviates from the binode, but follows a superheating line shown in Fig. 1, and the liquid is in a *metastable* state. The exact details of the superheating are not known, but should depend upon the heating rate. There is an upper limit for superheating of a liquid, the spinode [16–18], which is the boundary of thermodynamic phase stability and is determined by the second derivatives of the Gibbs' thermodynamic potential [19]:

$$\left(\frac{\partial p}{\partial v}\right)_T = 0, \tag{3}$$

where v is the specific volume. Using Eq. (3), the spinode equation can be derived from empirical equations of state such as the van der Waals equation or the Berthelot Eq. [20]. As the temperature approaches the spinode, fluctuations of local density of a liquid metal increase rapidly, and $(\partial p/\partial v)_T \rightarrow 0$, resulting in a loss of thermodynamic stability. These fluctuations begin when the temperature approaches 0.8 T_c, which drastically affect other physical properties. Figure 1(*b*) shows properties of a liquid metal near the critical point. Rapid changes of properties can been seen when the liquid temperature is above 0.8 T_c. These drastic property changes are called anomalies, which are also indicated in Fig. 1(*a*). Usually, the onset of anomalies concurrently marks the onset of significant reduction or even disappearance of electrical conductivity of a liquid metal due to many isolated regions with few free electrons [21,22]. Thus, at the onset of anomalies, a liquid metal is transferred from a conductor to a dielectric. Its transmission to optical radiation increases and surface reflectivity decreases.

A competing process that prevents superheating of a liquid is spontaneous nucleation. If the rate of spontaneous nucleation is high enough, homogeneous liquid-vapor phase change would occur. Therefore, the existence of the superheated state requires a low rate of spontaneous nucleation. The rate of spontaneous nucleation can be determined from the Döring and Volmer's theory [23,24]. According to this theory, the frequency of spontaneous nucleation is calculated as

$$J = \eta \exp\left(\frac{-W_{cr}}{k_B T}\right), \tag{4}$$

where W_{cr} is the energy needed for vapor embryos to grow to nuclei at temperature T, or the work of formation of nuclei. (Embryos smaller than a critical size will collapse, while those larger than the critical radius, called nuclei, will favor growing in order to reduce free energy.) η is on the order of magnitude of the number of liquid molecules per unit volume, calculated as [23]:

$$\eta = N\left(\frac{3\sigma}{\pi m}\right)^{1/2}, \tag{5}$$

where N is the number of liquid molecules per unit volume, and σ is the surface tension. Note that the above results are derived based on the assumption that the heating rate is slow enough that an equilibrium distribution of embryos exists in the liquid.

According to Eqs. (4) and (5), the spontaneous nucleation rate increases exponentially with temperature. It has been calculated that the frequency of spontaneous nucleation is only about 0.1 $s^{-1}cm^{-3}$ at the temperature near 0.89 T_c, but increases to 10^{21} $s^{-1}cm^{-3}$ at 0.91 T_c [5]. For a slowly heated liquid, the number of nuclei generated by spontaneous nucleation will be high enough to cause homogeneous phase change at the normal boiling temperature. Therefore, a superheated state cannot be sustained. On the other hand, during high power pulsed laser heating considered in this work for which the time duration is on the order of tens of nanoseconds, the amount of nuclei generated by spontaneous nucleation is negligible at temperatures lower than 0.9 T_c. Therefore, the liquid could possess considerable stability with respect to spontaneous nucleation. At a temperature of about 0.9 T_c, a significant number of nuclei can be formed within a short period of time. Hence, homogeneous nucleation, or explosive phase change occurs, which turns the liquid into a mixture of liquid and vapor, leaving the surface like an explosion.

To analyze phase change induced by pulsed laser heating, it is also necessary to consider the time required for a vapor embryo to grow to a critical nucleus, which is called the time lag for nucleation. For most engineering applications, the time to form critical nuclei is too short to be considered. However, during pulsed laser heating when the heating time is on the order of nanoseconds or shorter, this time lag could be on the same order of the time period under consideration. Equation (4) can be modified to account for this time lag, τ, which can be expressed as [24]:

$$J = \eta \exp\left(\frac{-W_{cr}}{k_B T}\right)\exp\left(\frac{-\tau}{t}\right), \tag{6}$$

where t is the time duration for which the liquid is superheated. The time lag τ has been estimated to be [24]:

$$\tau = \left(\frac{2\pi M}{RT}\right)^{1/2} \frac{4\pi\sigma p_s}{(p_s - p_t)^2}, \tag{7}$$

where M is the molar weight of the substance. Skripov performed calculations based on Eq. (7) for metals and found the time lag to be approximately 1–10 ns [24].

3 Experimental Investigation of Phase Explosion and Its Time Lag

This section describes studies on phase change mechanisms during laser ablation through a number of experimental investigations on laser-induced vapor. Although it is more desirable to measure the surface temperature and pressure for the study of phase change kinetics, direct measurement of the surface temperature is hampered by the strong radiation from the laser-induced vapor. On the other hand, properties of the vapor are strongly linked to the surface thermodynamic parameters. Therefore, knowing the properties of the laser-induced vapor could help to understand the phase change phenomenon occurring at the surface.

3.1 Summary of Experimental Study on Phase Explosion. The laser used for the experimental study is a KrF excimer laser with a wavelength of 248 nm and a pulse width of 25 ns (FWHM). The center, uniform portion of the excimer laser beam is passed through a rectangular aperture (10 mm by 5 mm) to produce a laser beam with a uniform intensity profile. A single 150 mm focal length CaF_2 lens is used to focus the laser beam on the target. Polished nickel (50 nm RMS roughness) is used as the target.

The following parameters are measured: transient transmissivity of laser beam through the laser-induced vapor plume, scattering of laser beam from the laser-induced vapor plume, transient location and velocity of the laser-induced vapor front, and ablation depth per laser pulse. Details of the experiments have been given elsewhere [7]. The experimental results are provided here for further discussions.

Figure 2(*a*) shows the transient location of the vapor front as a function of laser fluence. The measurement is based on an optical deflection technique, which is highly accurate (better than ± 3 percent) and repeatable. Figure 2(*b*) shows the averaged velocity of the laser-evaporated vapor. These are time-averaged vapor velocities from the vapor onset to 50 ns calculated from Fig. 2(*a*). It is seen that the vapor velocity increases with the laser fluence increase from about 2,000 m/s at the lowest fluence to about 7000 m/s at the highest fluence. However, the increase of velocity is not monotonous. A sudden jump of the velocity is seen at the laser fluence of 4.2 J/cm^2. In the laser fluence range between 5.2 and 9 J/cm^2, the velocity is almost a constant.

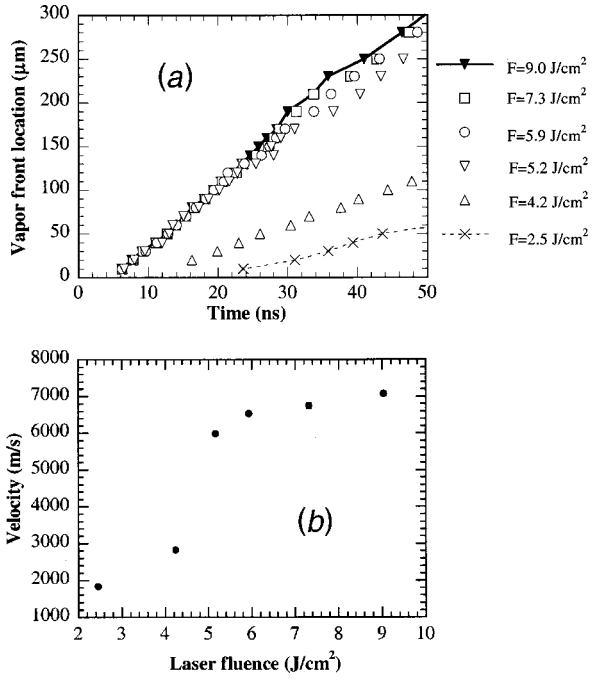


Fig. 2 (*a*) Transient locations of the vapor front as a function of laser fluence; and (*b*) vapor velocity as a function of laser fluence.

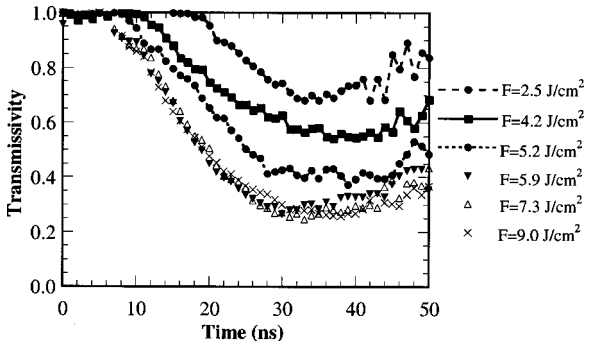


Fig. 3 Transient transmissivity of vapor as a function of laser fluence

The different relations between the vapor velocity and the laser fluence indicate different laser ablation mechanisms. The velocity of the vapor plume is determined by the pressure and the temperature at the target surface. The constant velocity at high laser fluences indicates that the surface temperature is not affected by the increase of the laser fluence. Such a constant surface temperature can be explained as a result of phase explosion. As discussed earlier, the surface temperature during phase explosion is about 0.9 T_c, the spinodal temperature. Once the laser fluence is high enough to raise the surface temperature to the spinode, increasing the laser fluence would not raise the surface temperature further. On the other hand, when the laser fluence is below 4.2 J/cm^2, the velocity increases over 50 percent. Therefore, the surface temperature increases with the laser fluence increase; normal vaporization occurs at the surface.

Figure 3 shows the transient transmissivity of the vapor at different laser fluences. The uncertainty of the measurement is less than a few percent in the time duration from a few nanoseconds to about 45 ns. Near the end of the laser pulse, the uncertainty of the measurement is larger ($\sim$10 percent), since the pulse intensity is weak. The data show that the transient transmissivity is almost identical for laser fluences higher than 5.2 J/cm^2, which is exactly the same fluence region in which the velocity of the vapor changes little. This is also explained as a result of explosive phase change occurring at laser fluences above 5.2 J/cm^2. Extinction of the laser beam is determined by the cross section of the energized atoms, which in turn is determined by the temperature of the vapor plume. The temperatures of vapor are about the same when the laser fluence is higher than 5.2 J/cm^2, since the temperatures at the surface are all about 0.9 T_c. Thus, transmission through vapor stays at a constant value.

Figure 4 shows the percentage of laser energy scattered from the vapor plume as a function of laser fluence. Scattering of laser energy is due to large size (on the order of sub-micron or larger) droplets in the vapor plume instead of (atomic) vapor. It is seen from Fig. 4 that there is almost no scattering (less than 0.5 percent, the measurement resolution) in the low laser fluence region. Therefore, there are no droplets in the vapor plume. When the laser fluence is higher than 5.2 J/cm^2, the percentage of laser energy scattered by the plume is about 4 to 5 percent, indicating the existence of liquid droplets. This phenomenon again can be explained by the different ablation mechanisms. When explosive phase change occurs, the melted layer is turned into a liquid-vapor mixture. Therefore, the increase of scattering at the laser fluence of 5.2 J/cm^2 also indicates the transition from surface evaporation to phase explosion.

Figure 5 shows the ablation depth per laser pulse at different laser fluences. The ablation depth increases almost linearly from 14 to 20 nm with laser fluence when the laser fluence is less than

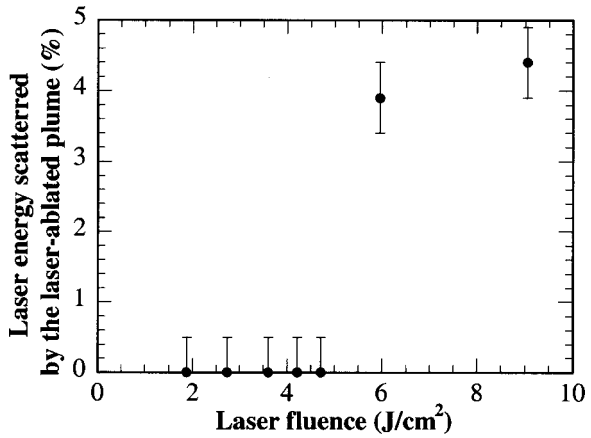


Fig. 4 Percent of laser energy scattered to the ambient as a function of laser fluence

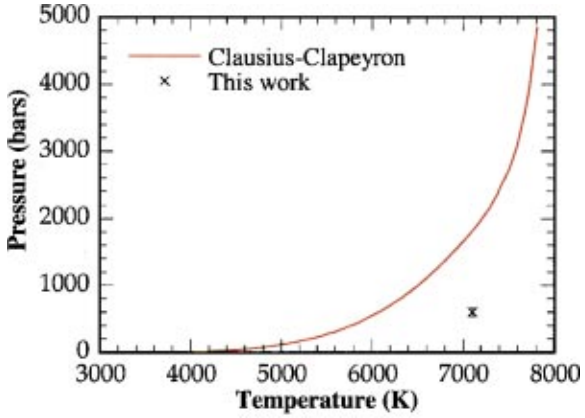


Fig. 6 Comparison between the Clausius-Clapeyron relation and the measured pressure at 0.9 T_c

4.0 J/cm^2. When the laser fluence increases from 4.2 to 5.2 J/cm^2, a jump increase in the ablation depth is observed, and stays relatively a constant at higher laser fluences. This again is explained as surface normal vaporization versus volumetric phase explosion. When phase explosion occurs, the liquid layer is ablated, therefore, the ablation depth is much greater than that of surface evaporation.

Since the surface temperature is maintained at a relatively constant value when phase explosion occurs, one question arises as to how the additional laser energy dissipates when the laser fluence is further increased. A possible explanation is that once the temperature reaches above 0.8 T_c, the material is heated up less quickly, since it becomes less absorbing as seen in Fig. 1(*b*) allowing the laser energy to penetrate deeper into the material. Another reason is that when the temperature approaches the spinodal temperature, most of the additional incoming laser energy is consumed by nucleation instead of raising the temperature, and the nucleation rate increases exponentially around the spinode.

In a brief summary, these four independent experiments all show that surface evaporation occurs at laser fluences below 4.2 J/cm^2, while an explosive phase change occurs when the laser fluence is higher than 5.2 J/cm^2.

3.2 Kinetics at the Evaporating Surface. To understand the kinetics at the evaporating surface, the transient pressure of the evaporating surface is measured with the use of a PVDF transducer attached to the back of a thin nickel target. Details of the experiment have been given elsewhere [25]. The transient surface pressure is obtained at various laser fluences. Of particular interest is the pressure when phase explosion occurs, which is determined to be about 600 bars ($\pm$10 percent) at 5.2 J/cm^2. Figure 6 shows the Clausius-Clapeyron equation for Ni, together with the experimental data point at 5.2 J/cm^2. It can be seen that the pressure obtained from the experiment is well below the equilibrium pressure, showing that the liquid is indeed superheated under pulsed laser irradiation.

Another way to estimate the validity of the equilibrium evaporation kinetics is to compute the evaporation depth from the measured pressure using the Clausius-Clapeyron equation and compare the calculated results with the measured data. To do so, the transient surface temperature T is first calculated from the measured transient surface pressure p and the Clausius-Clapeyron equation, Eq. (2). Knowing the surface temperature and pressure, the evaporation velocity, V_{lv}, can be calculated from the atomic flux $\dot{m}$ using Eq. 1, modified by a factor of m/ρ_l. The ablation depth per laser pulse is obtained by integrating the evaporation velocity over time. Note that this calculation can only be carried out for surface evaporation.

The calculated ablation depths at different laser fluences are shown in Fig. 7. It can be seen that the calculated ablation depths are greater than the measured values, by as much as a factor of seven to eight. This large discrepancy again indicates that the equilibrium interface kinetics and the Clausius-Clapeyron equa-

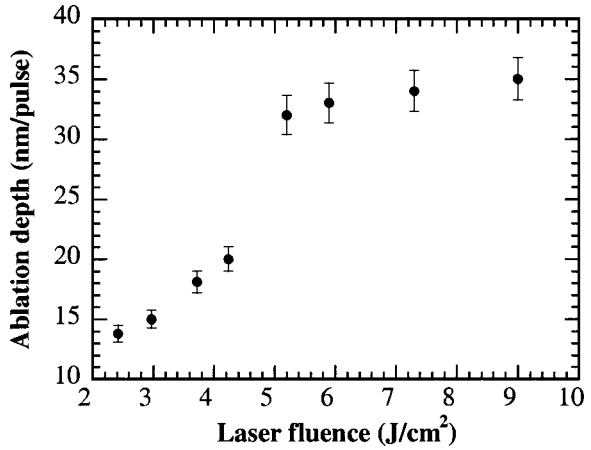


Fig. 5 Ablation depth as a function of laser fluence

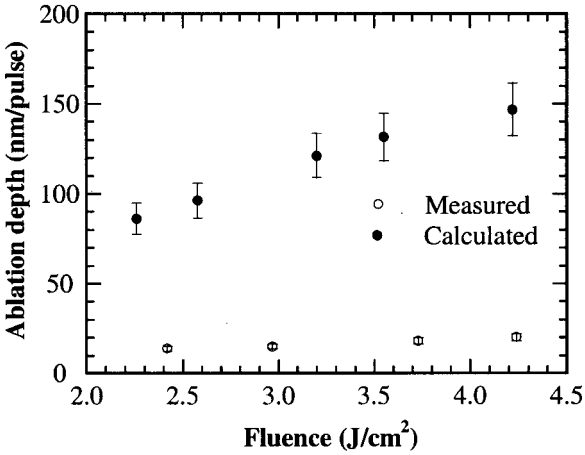


Fig. 7 Comparison between the measured ablation depth and the values calculated using transient pressure data and the equilibrium kinetic relation

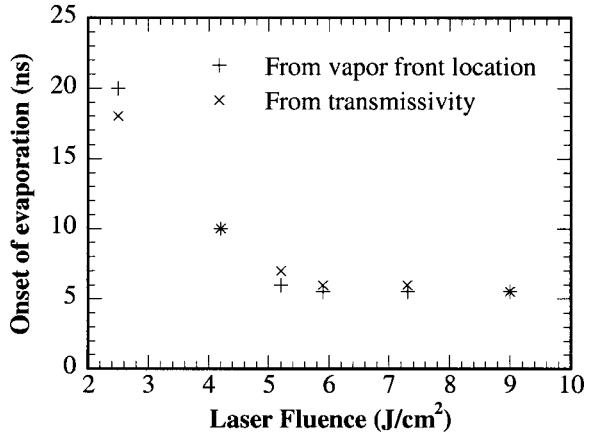


Fig. 8 Onset of evaporation as a function of laser fluence determined from the transient location of the vapor front and from the transient transmissivity through the vapor

tion do not correctly represent the actual surface temperature-pressure relation during pulsed laser evaporation.

3.3 Time Lag in Phase Explosion. Further examinations of Fig. 2(*a*) and Fig. 3 reveal another phenomenon: the onset of ablation, indicated as the time when the vapor front leaves the surface (Fig. 2(*a*)) and the time that transmission starts to decrease (Fig. 3), is about the same when the laser fluence is higher than 5.2 J/cm^2. The onset of ablation is re-plotted in Fig. 8. It can be seen that, when the laser fluence is higher than 5.2 J/cm^2, the onset of ablation does not change with the laser fluence, but remains at around 5.5 ns after the beginning of the laser pulse. The accuracy of this measurement depends on the time resolution of the measurement instrument, which is about 0.5 ns. The two independent measurements provide almost identical results.

The constant value of the onset of ablation can be explained as the time needed for phase explosion to occur, or the time lag for phase explosion. The experimental results discussed previously indicate that the phase explosion occurs when the laser fluence is higher than 5.2 J/cm^2. At these laser fluences, the measured results of the vapor front location and the optical transmission are dictated by the mass removal due to phase explosion. Thus, the constant onset of ablation at laser fluences higher than 5.2 J/cm^2 indicates that the time lag prevents phase explosion to occur at an earlier time, and this time lag is about a few nanoseconds.

The experiments described in this paper are performed with the use of a 25 ns pulsed excimer laser on a nickel target. It is believed that the phase change phenomena discussed here should occur for other metals as well. On the other hand, if the laser fluence is *much* higher than the threshold fluence for phase explosion, it is possible that the surface temperature can be raised higher. In our experiments with a laser fluence above 10 J/cm^2, it was indeed found that the velocity of vapor increases, and transmission and onset of evaporation reduces from the values of the constant region. Experiments with very high laser fluences should be conducted to investigate the possibility of heating the material above the limit of thermodynamic stability. A last note is on the phase change mechanisms induced by sub-nanosecond laser ablation. The threshold nature of ablation has been observed in many pico-and femtosecond laser ablation experiments [e.g., [8,26]]. Phase explosion is explained as the ablation mechanism induced by a femtosecond laser irradiation [13]. However, since the heating time is much less than the time lag of nucleation, much work is needed to gain a thorough understanding of ablation induced by a pico or femtosecond laser.

4 Conclusions

Heat transfer and non-equilibrium phase change during nanosecond pulsed excimer laser ablation of nickel were investigated. Results of experiments showed surface evaporation occurred when the laser fluence was below 4 J/cm^2. When the laser fluence was higher than 5 J/cm^2, the liquid reached a metastable state during laser heating and its temperature approached the critical point, causing an explosive type of phase change. The kinetic relation between the surface temperature and pressure was found to deviate from the equilibrium Clausius-Clapeyron equation. With the given experimental conditions, the time lag of phase explosion was found to be around a few nanoseconds.

Acknowledgments

Support of this work by the National Science Foundation and the Office of Naval Research is gratefully acknowledged.

Nomenclature

H_{lv} = latent heat of evaporation [J/kmole]
J = frequency of spontaneous nucleation [$m^{-3}\,s^{-1}$]
k_B = Boltzmann's constant, 1.380×10^{-23} J/K
m = atomic mass [kg]
$\dot{m}$ = atomic flux, [$s^{-1}\,m^{-2}$]
M = molar weight, [kg/kmol]
N = number density of atoms [m^{-3}]
p = pressure [N/m^2]
p_l = pressure in liquid [N/m^2]
p_s = saturation pressure [N/m^2]
R = universal gas constant, 8.314 kJ/kmol·K
t = time [s]
T = temperature [K]
T_b = normal boiling temperature [K]
T_c = critical temperature [K]
v = specific volume [m^3/kg]
W_{cr} = energy required to form critical nuclei [J]

Greek Symbols

η = factor in Eqs. (4) and (5) [$m^{-3}\,s^{-1}$]
ρ_l = density of liquid [kg/m^3]
σ = surface tension [N/m]
τ = time lag of nucleation [s]

References

[1] Crafer, R. C., and Oakley, P. J., eds., 1993, *Laser Processing in Manufacturing*, Chapman & Hall, New York.
[2] Chrisey, D. B., and Hubler, G. K., eds., 1994, *Pulsed Laser Deposition of Thin Films*, John Wiley & Sons, New York.
[3] Wood, R. F., Leboeuf, J. N., Chen, K. R., Geohegan, D. B., and Puretzky, A. A., 1998, "Dynamics of Plume Propagation, Splitting, and Nanoparticle Formation during Pulsed-Laser Ablation," Appl. Surf. Sci., **127–129**, pp. 151–158.
[4] Puretzky, A. A., Geohegan, D. B., Fan, X., and Pennycook, S. J., 2000, "Dynamics of Single-wall Carbon Nanotube Synthesis by Laser Vaporization," Appl. Phys. A, **70**, pp. 153–160.
[5] Miotello, A., and Kelly, R., 1995, "Critical Assessment of Thermal Models for Laser Sputtering at High Fluences," Appl. Phys. Lett., **67**, pp. 3535–3537.
[6] Song, K. H., and Xu, X., 1998, "Explosive Phase Transformation in Pulsed Laser Ablation," Appl. Surf. Sci., **127–129**, pp. 111–116.
[7] Xu, X., 2001, "Heat Transfer and Phase Change during High Power Laser Interaction With Metal," Annu. Rev. Heat Transfer, C.-L. Tien, V. Prasad, and F. P. Incropera, eds., Bell House, New York, **12**, pp. 79–115.
[8] Willis, D. A., and Xu, X., 2001, "Heat Transfer and Phase Change during Picosecond Laser Ablation of Nickel," Int. J. Heat Mass Transf., submitted.
[9] Martynyuk, M. M., 1974, "Vaporization and Boiling of Liquid Metal in an Exploding Wire," Sov. Phys. Tech. Phys., **19**, pp. 793–797.
[10] Martynyuk, M. M., 1983, "Critical Constants of Metals," Russ. J. Phys. Chem., **57**, pp. 494–501.
[11] Xu, X., and Song, K. H., 1999, "Interface Kinetics during Pulsed Laser Ablation," Appl. Phys. A., **69**, pp. S869–S873.
[12] Xu, X., and Song, K. H., 2000, "Phase Change Phenomena during High Power Laser-Materials Interaction," Mater. Sci. Eng., A, **292**, pp. 162–168.
[13] Sokolowski-Tinken, K., Bialkowski, J., Boing, M., Cavalleri, A., and von der Linde, D., 1999, "Bulk Phase Explosion and Surface Boiling during Short Pulsed Laser Ablation of Semiconductor," in *Quantum Electronics and Laser*

Science Conference, OSA Technical Digest, Optical Society of America, Washington, DC, pp. 231–232.
[14] Zhigilei, L. V., and Garrison, B. J., 1999, ''Molecular Dynamics Simulation Study of the Fluence Dependence of Particle Yield and Plume Composition in Laser Desorption and Ablation of Organic Solids,'' Appl. Phys. Lett., **74**, pp. 1341–1343.
[15] Von Allmen, M., 1987, *Laser-Beam Interactions with Materials, Springer Series in Materials Science 2*, Springer-Verlag, Berlin.
[16] Fucke, W., and Seydel, U., 1980, ''Improved Experimental Determination of Critical-Point Data for Tungsten,'' High Temp.-High Press., **12**, pp. 419–432.
[17] Lienhard, L. H., and Karimi, A., 1981, ''Homogeneous Nucleation and the Spinodal Line,'' ASME J. Heat Transfer **103**, pp. 61–64.
[18] Avedisian, C. T., 1985, ''The Homogeneous Nucleation Limits of Liquids,'' J. Phys. Chem. Ref. Data, **14**, pp. 695–729.
[19] Martynyuk, M. M., 1977, ''Phase Explosion of a Metastable Fluid,'' Fiz. Goreniya Vzryva, **13**, pp. 213–229.
[20] Domb, C., 1996, *The Critical Point*, Taylor & Francis, New York.
[21] Martynyuk, M. M., 1975, ''Liquid-Vapor and Metal-Dielectric Transitions in Mercury,'' Russ. J. Phys. Chem., **49**, pp. 1545–1547.
[22] Batanov, V. A., Bunkin, F. V., Prokhorov, A. M., and Fedorov, V. B., 1973, ''Evaporation of Metallic Targets Caused by Intense Optical Radiation,'' Sov. Phys. JETP, **36**, pp. 311–322.
[23] Carey, V. P., 1992, *Liquid-Vapor Phase-Change Phenomena: An Introduction to the Thermophysics of Vaporization and Condensation Processes in Heat Transfer Equipment*, Hemisphere Publishing Corp., Washington.
[24] Skripov, V. P., 1974, *Metastable Liquids*, John Wiley & Sons, New York.
[25] Song, K. H., and Xu, X., 1998, ''Measurement of Laser-induced Shock Wave in Thin Nickel Foil,'' *Proc. of the 1998 ASME International Mechanical Engineering Congress and Exposition*, HTD-Vol. 361-4, pp. 79–86.
[26] Cavalleri, A., Sokolowski-Tinten, K., Bialkowski, J., and von der Linde, D., 1998, ''Femtosecond Laser Ablation of Gallium Arsenide Investigated with Time-of-Flight Spectroscopy,'' Appl. Phys. Lett., **72**, pp. 2385–2387.

Transient Heat Conduction in a Heat Generating Layer Between Two Semi-Infinite Media

Leendert van der Tempel
Philips Research Laboratories,
5656 AA Eindhoven,
The Netherlands,
e-mail: Leendert.van.der.Tempel@Philips.com

The problem of transient heat conduction in a heat generating layer between two semi-infinite media has been solved. The one-dimensional thermal model is Laplace transformed. Three analytical temperature solutions are derived: two approximation solutions and an exact series solution. They are compared with respect to accuracy, convergence and computational efficiency. The approximations are computationally more efficient, and the series converge to the exact solution. The presented accurate solutions enable quick thermal analysis in terms of just 2 parameter groups, but overestimate the temperature during initialization of rewritable optical disks due to lateral heat conduction.
[DOI: 10.1115/1.1447930]

Keywords: Analytical, Conduction, Heat Transfer, Temperature, Transient

Introduction

The problem of transient heat conduction in a heat generating layer between two semi-infinite media occurs in rewritable (*RW*) optical disks during initialization and writing by a laser. Initialization of rewritable disks is done in production to crystallize the amorphous sputtered phase change alloy layer. A laser spot describes a spiral track over the disk. The absorption heats the disk locally to about 700 °C.

Holtslag proposed three-dimensional analytical temperature approximations for optical recording [1] with fair accuracy.

This paper concentrates on one-dimensional models applicable to initialization of rewritable disks. Tangential heat diffusion is neglected for simplicity, overestimating the temperature rise. General solutions to the case of *composite slabs* or *multi-layers* have been presented [2–4]. These general solutions tend to be complex and ask often for the evaluation of integrals or infinite series with an unclear truncation error.

A well known analytical solution technique is [5]:

- Laplace transform of the energy equations
- Solution of the transformed *ordinary* differential equations
- Inverse Laplace transform.

However, all these steps constitute problems in our case:

- Direct Laplace transform of a Gaussian heat distribution is impossible
- Symbolic solution of the equation set for disks with a multilayer optical stack is unfeasible
- I did not find inverse Laplace transforms for the temperature in the literature.

To get around, this paper presents:

- An infinite time shift
- An expansion of the three-layer model to multi-layer stacks
- Both *exact series* and *rational approximations* with inverse Laplace transforms, like in [6]

The temperature, solutions fulfill our aim to provide a computationally efficient analytical temperature solution for initialization, giving insight and allowing sensitivity analysis.

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division April 27, 2001; revision received October 30, 2001. Associate Editor: G. S. Dulidravich.

Thermal Model

Lateral heat conduction is neglected to simplify the solution.

The one-dimensional axial geometry is shown in Fig. 1.

An (optical stack) layer *o* with thickness *l* is placed between two semi-infinite media *c* and *s* with uniform initial temperature T_0.

We assume that

- each medium has constant uniform thermal properties: conductivity *k*, density ρ, and specific heat c_p
- the thermal contact at all interfaces is perfect
- conduction can be modeled by Fourier's law
- a radiative heat flux q'' is absorbed in the absorbing (phase change) layer
- the phase change energy is negligible

The energy equations for the media read then

$$\frac{\partial T_c}{\partial t} = \alpha_c \frac{\partial^2 T_c}{\partial x^2}$$

$$\frac{\partial T_o}{\partial t} = \alpha_o \frac{\partial^2 T_o}{\partial x^2} + \nu G(t - t_0) \tag{1}$$

$$\frac{\partial T_s}{\partial t} = \alpha_s \frac{\partial^2 T_s}{\partial x^2},$$

where t_0 is the time when the (center of a Gaussian) radiative heating distribution passes, $\nu = q''/(l\rho c_p)_{abs}$ is the heating rate and the usual Gaussian heating rate distribution with effective heating time τ reads

$$G(t) = \sqrt{\frac{2}{\pi}} e^{-2(t/\tau)^2}. \tag{2}$$

The initial conditions are

$$T_c(x,0) = T_0 \quad T_o(x,0) = T_0 \quad T_s(x,0) = T_0. \tag{3}$$

The boundary conditions for perfect thermal contact are

$$T_c(0,t) = T_o(0,t) \quad T_o(l,t) = T_s(l,t)$$

$$\lim_{x \to -\infty} T_c = \text{finite} \qquad \lim_{x \to \infty} T_s = \text{finite} \tag{4}$$

$$k_c \frac{\partial T_c}{\partial x}(0,t) = k_o \frac{\partial T_o}{\partial x}(0,t) \quad k_o \frac{\partial T_o}{\partial x}(l,t) = k_s \frac{\partial T_s}{\partial x}(l,t)$$

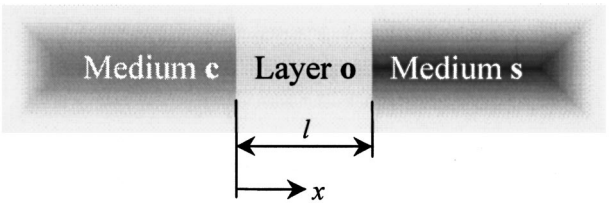


Fig. 1 One-dimensional geometry

Table 1 Test conditions

	Medium c	Layer o	Medium s
Material	lacquer	optical stack	polycarbonate
$H=\sqrt{k\rho c_p}$ [W√s/m²K]	750	3000	750
Diffusivity α [m²/s]	$1.2\cdot10^{-7}$	$1.6\cdot10^{-6}$	$1.2\cdot10^{-7}$
Thickness l [m]		$0.26\cdot10^{-6}$	

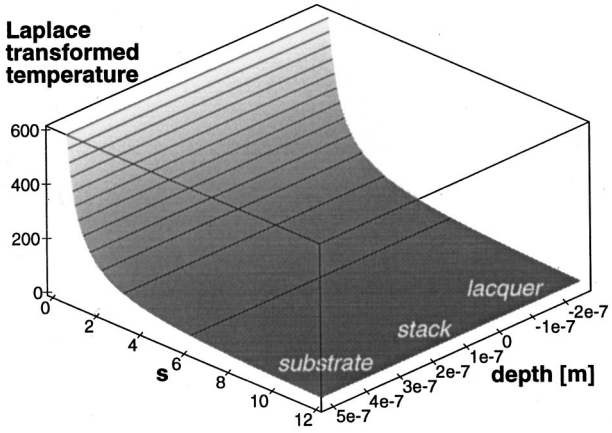


Fig. 2 Laplace transformed temperature distribution

Test Case

The test case that I use in all calculation result plots is the initialization of a 0.26 μm thick optical layer stack between semi-infinite media of polycarbonate and polymethyl methacrylate coating lacquer (CD-RW). The initial temperature $T_0=300$ K. The tangential heat flux distribution of the linear laser beam is Gaussian with heat flux $q''=3.9\cdot10^9$ W/m^2, spot width $b=1$ μm and effective heating time $\tau=0.5$ μs (Table 1).

The neglect of lateral heat conduction by the metal mirror in the optical stack is the most important flaw in the thermal model, despite the stack thickness/spot width ratio b/l of only 0.26.

Minor deviations from the model assumptions are:

The temperature dependence of the thermal properties of polycarbonate is up to 15 percent [7].

The estimated thermal resistance between stack layers is 10^{-8} Km^2/W [8].

The phase change energy from amorphous to crystalline is only about 0.2 MW/m^3 [9].

The reflectivity rises from 7 percent to about 27 percent during initialization (Fig. 13), reducing merely the absorption in the tail of the Gaussian distribution. Apart from the phase change metal, also the metal mirror layer absorbs some radiation.

Analytical Solution of the Temperature

Laplace Transform. Laplace transformation of the model simplifies the *partial* differential equations to *ordinary* differential equations.

The Laplace transformed energy equations in the media are

$$\frac{d^2\Theta_c}{dx^2}-\frac{s\Theta_c}{\alpha_c}=-\frac{T_0}{\alpha_c}$$
$$\frac{d^2\Theta_o}{dx^2}-\frac{s\Theta_o}{\alpha_o}=-\frac{T_0}{\alpha_o}-\frac{\nu g(s)}{\alpha_o} \tag{5}$$
$$\frac{d^2\Theta_s}{dx^2}-\frac{s\Theta_s}{\alpha_s}=-\frac{T_0}{\alpha_s}$$

The shifting property is used to shift the center t_0 of a Gaussian heat flux distribution to infinity to overcome the Laplace transform definition for $t\geqslant0$ only.

$$g(s)=\frac{1}{2}\tau e^{s^2\tau^2/8-st_0}\,\mathrm{erfc}\left(\sqrt{2}\left(s\tau/4-\frac{t_0}{\tau}\right)\right) \tag{6}$$

The Laplace transformed boundary conditions are

$$\Theta_c(0,s)=\Theta_o(0,s) \qquad \Theta_o(l,s)=\Theta_s(l,s)$$
$$\lim_{x\to-\infty}\Theta_c=\text{finite} \qquad \lim_{x\to\infty}\Theta_s=\text{finite} \tag{7}$$
$$k_c\frac{d\Theta_c}{dx}(0,s)=k_o\frac{d\Theta_o}{dx}(0,s) \qquad k_o\frac{d\Theta_o}{dx}(l,s)=k_s\frac{d\Theta_s}{dx}(l,s)$$

Solution of the Laplace Transformed Temperature. The complementary solution of the Laplace transformed media equations reads with $q_i=\sqrt{s/\alpha_i}$ with $i=c$, o, or s

$$\Theta_c(x,s)=A(s)e^{xq_c}+\frac{T_0}{s}$$
$$\Theta_o(x,s)=B(s)e^{-xq_o}+C(s)e^{xq_o}+\frac{T_0}{s}+\frac{\nu g(s)}{s} \tag{8}$$
$$\Theta_s(x,s)=D(s)e^{-xq_s}+\frac{T_0}{s}.$$

The Laplace transformed boundary conditions become

$$A(s)=B(s)+C(s)+\frac{\nu g(s)}{s}$$
$$B(s)e^{-lq_o}+C(s)e^{lq_o}+\frac{\nu g(s)}{s}=D(s)e^{-lq_s}$$
$$k_cq_cA(s)=k_oq_o(-B(s)+C(s)) \tag{9}$$
$$k_oq_o(-B(s)e^{-lq_o}+C(s)e^{lq_o})=-k_sq_sD(s)e^{-lq_s}.$$

The integration constants $A(s)-D(s)$ can be solved to obtain the particular solution of the Laplace transformed temperature $\Theta(x,s)$ (Fig. 2)

$$\Theta_c=\frac{T_0}{s}+\frac{\nu g}{sD}((H_s+H_o)e^{lq_o}+(H_s-H_o)e^{-lq_o}-2H_s)H_oe^{-|x|q_c}$$
$$\Theta_o=\frac{T_0}{s}+\frac{\nu g}{sD}(D+(H_s(H_c-H_o)-H_c(H_s+H_o)e^{lq_o})e^{-xq_o}$$
$$+(H_c(H_s-H_o)e^{-lq_o}-H_s(H_c+H_o))e^{xq_o}) \tag{10}$$
$$\Theta_s=\frac{T_0}{s}+\frac{\nu g}{sD}((H_c+H_o)e^{lq_o}+(H_c-H_o)e^{-lq_o}-2H_c)H_oe^{(l-s)q_s}$$

with denominator

$$D=(H_s+H_o)(H_o+H_c)e^{lq_o}+(H_s-H_o)(H_o-H_c)e^{-lq_o},$$

where $H=\sqrt{k\rho c_p}$ is the coefficient of heat penetration in the pertinent medium.

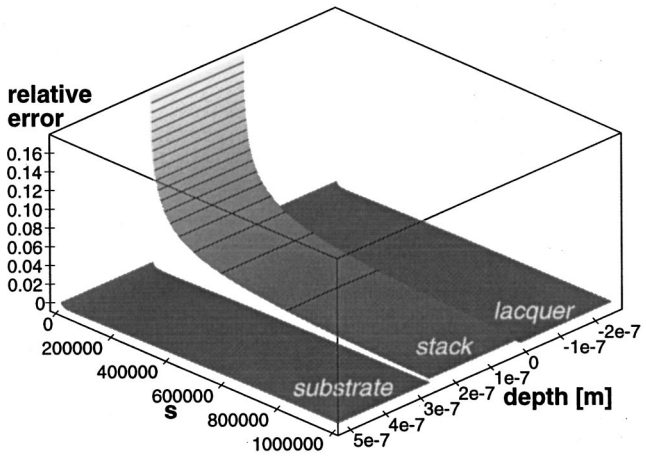


Fig. 3 Relative truncation error after four terms of the series

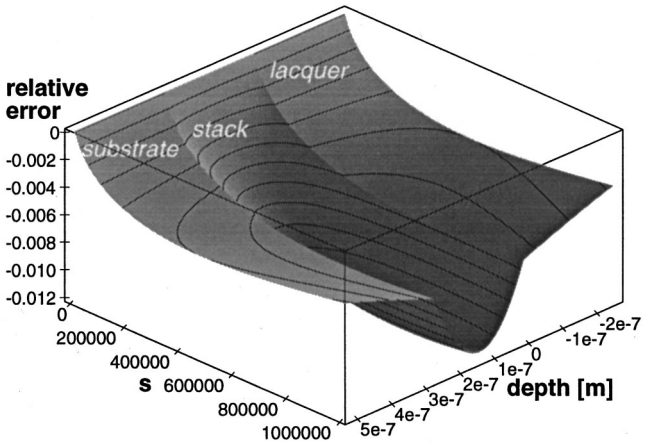


Fig. 4 Relative error of the rational approximation

I did not find inverse Laplace transforms for the temperature expressions (10) in the literature. To get around this problem, we propose both a geometric series expansion and rational approximations with available inverse Laplace transforms.

Series Expansion. Put

$$h=\frac{(H_s-H_o)(H_o-H_c)}{(H_s+H_o)(H_o+H_c)} \tag{11}$$

and notice that $|h|<1$ ($h\approx-0.36$ in the test case) and $0<e^{-2lq}\leqslant 1$ so $|he^{-2lq}|<1$.

This condition allows to use the *geometric series* expansion for $|f|<1$

$$\frac{1}{1-f}=1+f+f^2+f^3\ldots=\sum_{i=0}^{\infty} f^i=\sum_{i=0}^{m} f^i+\frac{f^{m+1}}{1-f}. \tag{12}$$

So the required number of terms for *truncation error* ε_m is

$$m=\frac{\log|\varepsilon_m|+\log(1-f)}{\log|f|}-1. \tag{13}$$

Substitution of $f=-he^{-2lq_o}$ in Eq. (12) yields a series expansion for the Laplace transformed temperature

$$\Theta_c=\frac{T_0}{s}+\frac{\nu g}{s}\frac{H_o(1-2r_se^{-lq_o}+r_{so}e^{-2lq_o})}{H_o+H_c}\sum_{i=0}^{\infty}(-h)^ie^{-X_{2i}}$$

$$\Theta_o=\frac{T_0}{s}+\frac{\nu g}{s}\Bigg(1+((r_sr_{co}-r_ce^{lq_o})e^{-\Omega_0}+(r_cr_{so}e^{-lq_o}-r_s)e^{\Omega_0})\sum_{i=0}^{\infty}(-h)^ie^{-(1+2i)lq_o}\Bigg) \tag{14}$$

$$\Theta_s=\frac{T_0}{s}+\frac{\nu g}{s}\frac{H_o(1-2r_ce^{-lq_o}+r_{co}e^{-2lq_o})}{H_o+H_s}\sum_{i=0}^{\infty}(-h)^ie^{-\Sigma_{2i}}.$$

The nomenclature defines the parameter groups.

Convergence is poor when $h\approx-1$ for layers with much higher or much lower conductivity than the semi-infinite media.

Figure 3 shows the distribution of the relative truncation error $(\Theta_{\text{series}}-\Theta)/\Theta$ after the fourth term for the test case. The relative truncation error has a maximum in the heat-generating layer for short times.

Rational Approximations. A linear approximation for $0<lq_o<1$ to the denominators D in the Laplace transformed temperature $\Theta(x,s)$ is

$$D\approx 2(H_o(H_c+H_s)+(H_o^2+H_cH_s)lq_o)=2\Delta. \tag{15}$$

Linearizing the denominator in this way yields a *rational approximation* to the Laplace transformed temperature about $lq_o=0$

$$\Theta_c\approx\frac{T_0}{s}+\frac{\nu g}{s\Delta}H_o^2lq_oe^{-|x|q_c}$$

$$\Theta_o\approx\frac{T_0}{s}+\frac{\nu g}{s\Delta}H_o^2lq_o \tag{16}$$

$$\Theta_s\approx\frac{T_0}{s}+\frac{\nu g}{s\Delta}H_o^2lq_oe^{(l-x)q_s}$$

Figure 4 shows the distribution of the relative error of this rational approximation. The relative approximation error $(\Theta_{\text{appr}}-\Theta)/\Theta$ has a maximum for $x=l/2$ and $s\approx 8\cdot 10^5$ for the test case.

The error can be reduced by *corrections* yielding a Padé-like rational approximation. The corrections should

- improve the accuracy
- have an inverse Laplace transform
- not introduce singularities
- yield positive erfc arguments

The third requirement limits the order of Padé approximations [10] to one half: only s^0 and $s^{1/2}$ terms are allowed in the numerator and denominator. The second and fourth requirement prohibits us to multiply or divide both the numerator and denominator by e^{lq_o}. However $lq_oe^{-lq_o}$ is a low cost correction to the numerator in medium s. The general corrected approximation looks like

$$\Theta\approx\frac{T_0}{s}+(c_0+c_1lq_o+(c_2+c_3lq_o)e^{-c_elq_o})\frac{e^{-lq}}{s\Delta}. \tag{17}$$

The constants c_i are solved to yield correct limits for $s\to\infty$ and correct behavior for $s\downarrow 0$ applying Viète's theorem on roots [11].

The resulting *Padé-like approximation* to the Laplace transformed stack temperature is

$$\Theta_c\approx\frac{T_0}{s}+\frac{H_o\nu g}{(H_o+H_c)\Delta}(-c_c+(H_o^2+H_cH_s)lq_o+(c_c-H_o(H_c+H_s)lq_o)e^{-lq_o})e^{-|x|q_c}$$

$$\Theta_o\approx\frac{T_0}{s}+\frac{\nu g}{\Delta}(-c_o+(H_o^2+H_cH_s)lq_o+(c_o+(c_o-H_cH_s)lq_o)e^{-lq_o}) \tag{18}$$

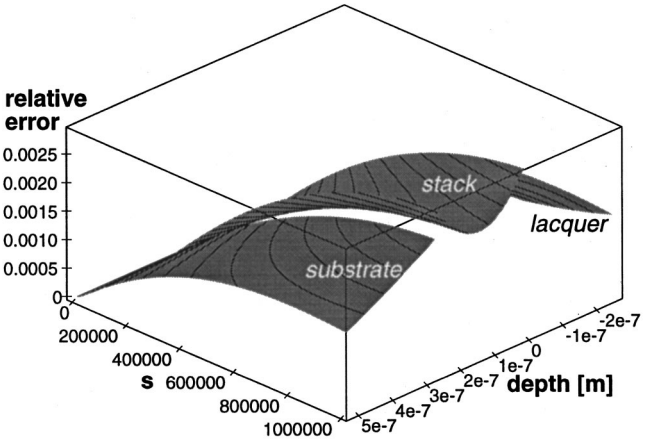


Fig. 5 Relative error of the Padé approximation

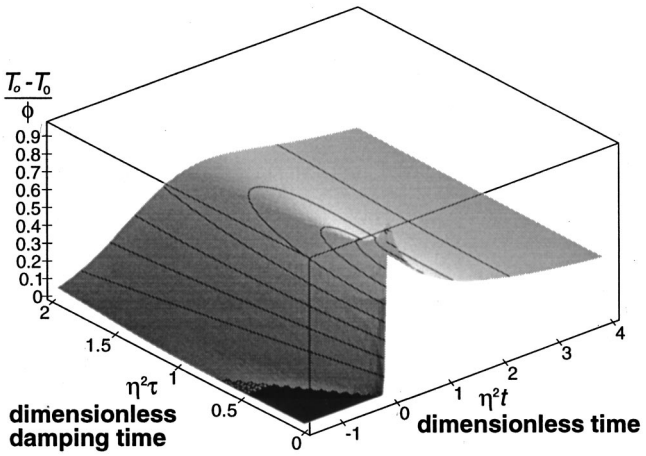


Fig. 6 Temperature rise in layer *o* as a function of dimensionless damping time $\eta^2 T$ and dimensionless time $\eta^2 t$

$$\Theta_s \approx \frac{T_0}{s} + \frac{H_o \nu g}{(H_o+H_s)\Delta}(-c_s + (H_o^2 + H_c H_s) l q_o + (c_s - H_o(H_c+H_s) l q_o) e^{-l q_o}) e^{(l-x) q_s}$$

with

$$c_c = -2H_c H_o + H_s(H_c - H_o)$$

$$c_o = H_s(2H_c - H_o) + H_o \frac{x}{l}\left(-2H_c + (H_c+H_s)\frac{x}{l}\right).$$

$$c_s = -2H_s H_o + H_c(H_s - H_o)$$

Figure 5 shows the distribution of the relative error of the Padé-like approximation. The relative approximation error has a maximum for $x=0$ and $x=l$ at $s \approx 12 \cdot 10^5$ for the test case.

Hybrid Combination of Series and Rational Approximation. The rational approximations can also evaluate the truncation error of the series expansion (12) yielding a hybrid combination of both approaches. This will not be elaborated.

Inverse Laplace Transform. The Appendix and the convolution theorem

$$\mathcal{L}^{-1}\{f(s)g(s)\} = \int_0^t F(u)G(t-u-t_0)du \tag{19}$$

have been used to derive the inverse Laplace transform $T(x,t)$ of the Laplace transformed temperature $\Theta(x,s)$.

Shifting the center of a Gaussian heat flux distribution back to the origin ($t_0 = 0$)

$$\lim_{t_0 \to \infty} \int_0^t F(u)G(t-u-t_0)du = \int_0^\infty F(u)G(t-u)du. \tag{20}$$

The integral can be evaluated numerically to sufficient accuracy by

- (~10 point) Gaussian integration for $t < 2\tau$

$$T = T_0 + 1/2 t_\infty \sum_{i=1}^{10} w_i F(t_\infty/2(1+x_i)) G(t - 1/2 t_\infty (1+x_i)) \tag{21}$$

with integration time $t_\infty = \max(\tau, t+2\tau)$

- (~5 point) Hermite integration for $t > 2\tau$

$$T = T_0 + 1/2\sqrt{2}\tau \sum_{i=1}^{5} w_i F(x_i), \tag{22}$$

where w_i are weight factors and x_i abscissas [12].

Series Expansion. The inverse Laplace transform of the series expansion is

$$T_c = T_0 + \frac{\nu H_o}{H_c + H_o}\int_0^\infty \sum_{i=0}^{\infty} (-h)^i S_{ci} G(t-u)du$$

$$T_o = T_0 + \nu \int_0^\infty \left(1 + \sum_{i=0}^{\infty} (-h)^i S_{oi}\right) G(t-u)du \tag{23}$$

$$T_s = T_0 + \frac{\nu H_o}{H_s + H_o}\int_0^\infty \sum_{i=0}^{\infty} (-h)^i S_{si} G(t-u)du$$

with

$$S_{ci} = \operatorname{erfc} \chi_{2i} - 2r_s \operatorname{erfc} \chi_{2i+1} + r_{so} \operatorname{erfc} \chi_{2i+2}$$

$$S_{oi} = r_s r_{co} \operatorname{erfc} \omega_{1,2i+1} - r_c \operatorname{erfc} \omega_{1,2i} + r_c r_{so} \operatorname{erfc} \omega_{-1,2i+2} - r_s \operatorname{erfc} \omega_{-1,2i+1}$$

$$S_{si} = \operatorname{erfc} \sigma_{2i} - 2r_c \operatorname{erfc} \sigma_{2i+1} + r_{co} \operatorname{erfc} \sigma_{2i+2}$$

requiring $15(m+1)$ to $50(m+1)$ function evaluations, where m is the number of terms.

Rational Approximation. The inverse Laplace transform of the rational approximation is

$$T_c \approx T_0 + \varphi \int_0^\infty e^{(2\chi_0 + \eta\sqrt{u})\eta\sqrt{u}} \operatorname{erfc}(\chi_0 + \eta\sqrt{u}) G(t-u)du$$

$$T_o \approx T_0 + \varphi \int_0^\infty e^{\eta^2 u} \operatorname{erfc}(\eta\sqrt{u}) G(t-u)du \tag{24}$$

$$T_s \approx T_0 + \varphi \int_0^\infty e^{(2\sigma_0 + \eta\sqrt{u})\eta\sqrt{u}} \operatorname{erfc}(\sigma_0 + \eta\sqrt{u}) G(t-u)du$$

requiring only 10 to 20 function evaluations (one spreadsheet cell).

Just two parameter groups determine the stack temperature rise $T_o - T_0$: the heating rate φ and the dimensionless damping time $\eta^2\tau$. The heating rate φ determines the temperature amplitude, whereas the dimensionless damping time $\eta^2\tau$ smoothes the temperature transient (Fig. 6).

In the test case the heating rate $\varphi = 4.8 \cdot 10^9$ K/s and the dimensionless damping time $\eta^2\tau = 1.9$.

Expansion of the three-layer model to four and more layers yields the same basic solution and η and φ definitions with a sum over the stack layers i: $\eta=(H_s+H_c)/\Sigma_i[(H_{oi}^2+H_cH_s)l_i/k_{oi}]$ and $\varphi=q''/\Sigma_i[(H_{oi}^2+H_cH_s)l_i/k_{oi}]$.

The inverse Laplace transform of the *Padé-like temperature approximation* requires 25 to 70 function evaluations (six spreadsheet cells) and reads

$$T_c \approx T_0+\nu\int_0^\infty \begin{pmatrix} c_{c1}(\operatorname{erfc}\chi_1-\operatorname{erfc}\chi_0) \\ +c_{c2}e^{(2\chi_0+\eta\sqrt{u})\eta\sqrt{u}}\operatorname{erfc}(\chi_0+\eta\sqrt{u}) \\ -c_{c3}e^{(2\chi_1+\eta\sqrt{u})\eta\sqrt{u}}\operatorname{erfc}(\chi_1+\eta\sqrt{u} \end{pmatrix} \times G(t-u)du$$

$$T_o \approx T_0+\nu\int_0^\infty \begin{pmatrix} c_{o1}(\operatorname{erfc}\omega_{0,1}-1)+c_{o2}e^{\eta^2 u}\operatorname{erfc}(\eta\sqrt{u}) \\ -c_{o3}e^{(2\omega_{0,1}+\eta\sqrt{u})\eta\sqrt{u}}\operatorname{erfc}(\omega_{0,1}+\eta\sqrt{u}) \end{pmatrix} \times G(t-u)du \quad (25)$$

$$T_s \approx T_0+\nu\int_0^\infty \begin{pmatrix} c_{s1}(\operatorname{erfc}\sigma_1-\operatorname{erfc}\sigma_0) \\ +c_{s2}e^{(2\sigma_0+\eta\sqrt{u})\eta\sqrt{u}}\operatorname{erfc}(\sigma_0+\eta\sqrt{u}) \\ -c_{s3}e^{(2\sigma_1+\eta\sqrt{u})\eta\sqrt{u}}\operatorname{erfc}(\sigma_1+\eta\sqrt{u}) \end{pmatrix} \times G(t-u)du$$

with

$$c_{c1}=\frac{c_c}{(H_c+H_o)(H_c+H_s)}$$

$$c_{c2}=c_{c1}+\frac{H_o}{H_c+H_o}$$

$$c_{c3}=c_{c1}+\frac{H_o^2(H_c+H_s)}{(H_c+H_o)(H_o^2+H_cH_s)}$$

$$c_{o1}=\frac{c_o}{H_o(H_c+H_s)}$$

$$c_{o2}=c_{o1}+1$$

$$c_{o3}=c_{o1}+\frac{H_cH_s-c_o}{H_o^2+H_cH_s}$$

$$c_{s1}=\frac{c_s}{(H_s+H_o)(H_c+H_s)}$$

$$c_{s2}=c_{s1}+\frac{H_o}{(H_s+H_o)}$$

$$c_{s3}=c_{s1}+\frac{H_o^2(H_c+H_s)}{(H_s+H_o)(H_o^2+H_cH_s)}.$$

Results

Figure 7 shows the temperature evolution for the initialization test case. The heating of the stack and the heat diffusion into the substrate and coating are clearly visible.

Accuracy and Convergence

The absolute maximum temperature solution error is determined by comparing to the ten term series solution serving as a reference for the test case. Figure 8 shows a comparison of the solutions with respect to the absolute maximum temperature solution error and the computational work for the test case. The convergence of the series and hybrid solution depends strongly on h, so on the layer conductivity.

In the initialization test case, physical model simplifications dominate these solution errors. The three-layer approximation to six-layer disks introduces only a few percent model error, but neglecting the tangential heat conduction overestimates the temperature rise by 63 percent compared to a fine grid numerical 3DT model [13] simulation provided by courtesy of Erwin Meinders. Figure 9 compares the initialization temperatures calculated by the

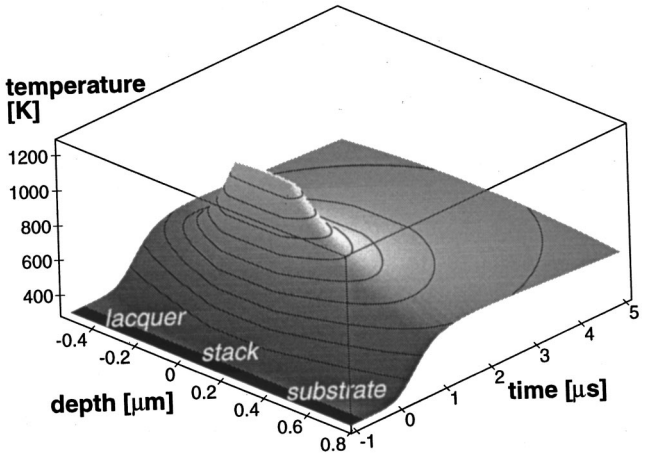


Fig. 7 Temperature distribution during initialization of CD-RWs

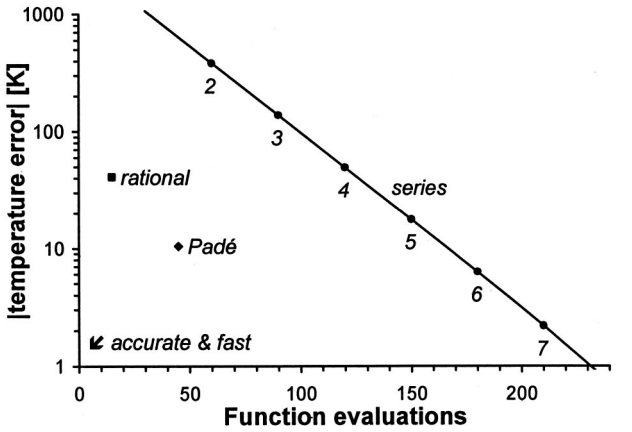


Fig. 8 Relation between absolute maximum temperature solution error and computational work

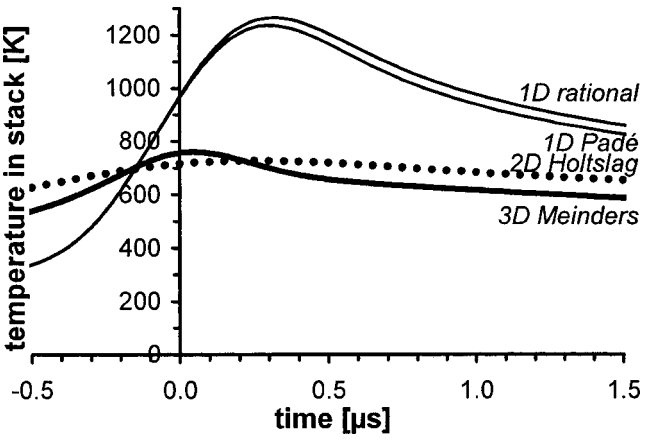


Fig. 9 Comparison of stack temperature calculations by 1DT, 2DT, and 3DT models

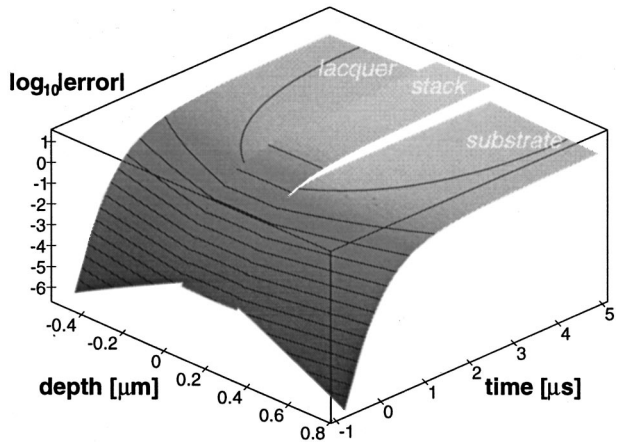


Fig. 10 Common logarithm of the absolute truncation error after 4 terms of the temperature series

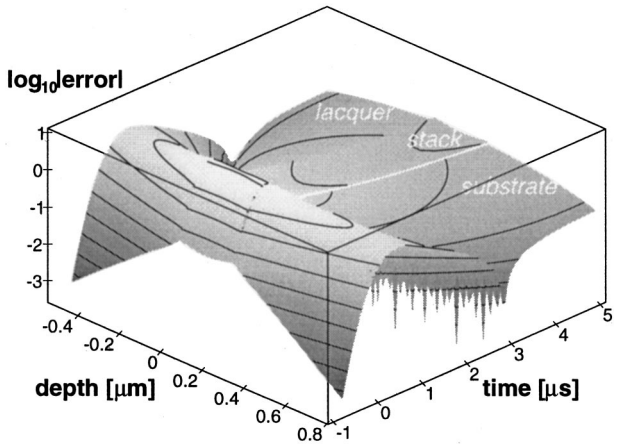


Fig. 12 Common logarithm of the absolute error of the Padé-like temperature approximation

numerical 3DT model of Erwin Meinders [13], the 2DT approximate model of Toon Holtslag [1] and the proposed 1DT rational and Padé approximations (22) and (23).

Accuracy and Convergence of the Series Expansion. Figure 10 shows the distribution of the truncation error of the temperature series expansion after the fourth term (Fig. 3 shows the Laplace transformed counterpart). The negative scale numbers on the vertical axis indicate the number of correct decimal places. The truncation error has a maximum in the stack for long times. The maximum truncation error after four terms is 50 K for the test case with $h \approx -0.36$.

The series solution converges exponentially (Fig. 8). However the convergence rate depends strongly on h according to Eq. (13).

Accuracy of the Rational Approximations. Figure 11 shows the distribution of the accuracy of the temperature approximation (Fig. 4 shows the Laplace transformed counterpart). The negative scale numbers on the vertical axis indicate the number of correct decimal places. The solution error reaches a maximum of 41 K for $x = \frac{1}{2}l$ and $t = 0.8\tau$ for the test case.

Figure 12 shows the distribution of the accuracy of the Padé-like temperature approximation (Fig. 5 shows the Laplace transformed counterpart). The negative scale numbers on the vertical axis indicate the number of correct decimal places. The solution error reaches a maximum of 10 K for $x \approx 0$ or $x \approx l$ and $t \approx 0$ for the test case.

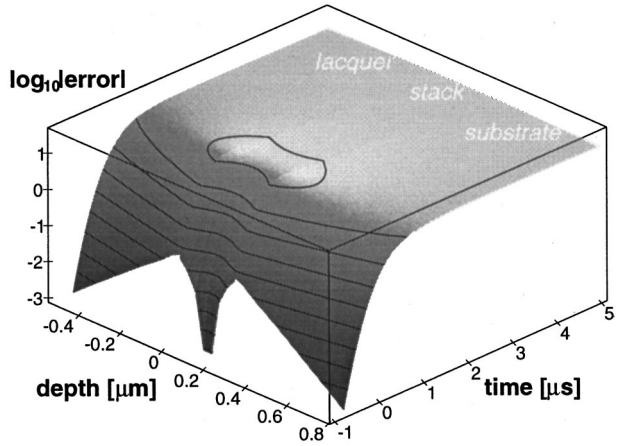


Fig. 11 Common logarithm of the absolute error of the rational temperature approximation

Qualitative Experimental Validation

Direct temperature validation measurements are complicated by the small length and time scale.

An alternative qualitative experimental validation was performed by initializing 175 Philips CD-RWs varying four parameters. Next the *reflectivity* and the *radial tilt increase* of the disks have been measured at 55 mm radius using a Dr. Schenk Prometeus disk tester.

The *reflectivity* is a measure for the *temperature* dependent crystallization of the phase change alloy [9]. So the goodness of a fit to these reflectivity measurements provides a *qualitative* validation of the temperature solutions, without any further assumption.

The *radial tilt increase* is a measure for the *warpage* of the disk by initialization. The radial deviation (of the laser beam) is 2× the radial tilt angle of the disk. Its temperature dependence is investigated below. The disks were left for a week in a climate room to equilibrate humidity before each radial tilt measurement.

Measurement Program. Four parameters or factors have been varied (Table 2).

All 84 combinations have been measured (full factorial design) twofold. A few additional measurements have been carried out with 386 J/m^2 initialization energy. The number of samples was 175. The 52 °C initial temperature was monitored by an infrared camera and realized by preheating disks in a furnace at 85 °C. Preheating warped the disks by 0.17–0.24 deg due to cross-linking of the lacquer and aging of the substrate. This has been excluded from the radial tilt increase.

Measurement Results. The measured average *reflectivity* and *radial tilt increase* at 55 mm radius are plotted in Figs. 13 and 14 against the calculated maximum (stack) temperature during initialization according to Eq. (24).

The (production and measurement) *repeatability* of the reflectivity is 0.0040 and of the radial tilt increase 0.010 deg.

Physical models are not available for these measurements, leaving empirical models. Empirical 4 and 5-parameter Padé approxi-

Table 2 Measurement program

Parameter (factor)	Values (levels)				
Lacquer thickness	4.26	9.29			[μm]
Initial disk temperature T_0	22	52			[°C]
Initialization velocity b/τ	2	2.5	3		[m/s]
Initialization heat flux q''	0	643	965		
	1286	1608	1929	2251	[J/m²]

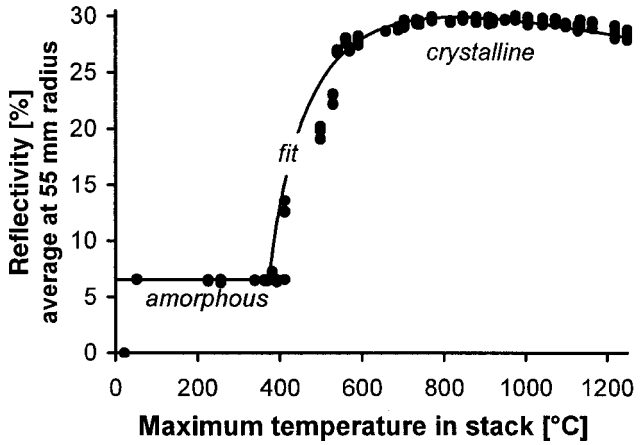


Fig. 13 Measured reflectivity against the calculated maximum stack temperature

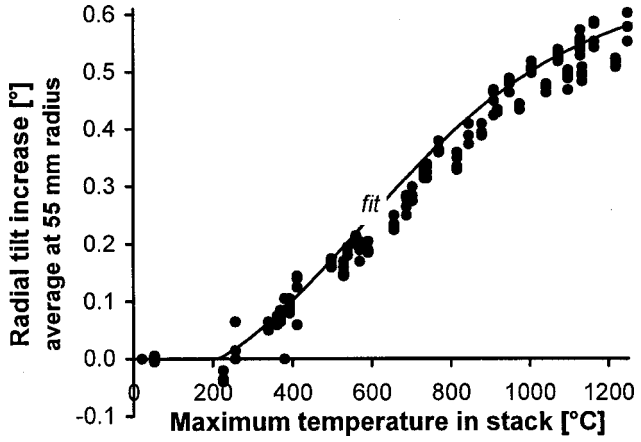


Fig. 14 Measured radial tilt increase against the calculated maximum stack temperature

Table 3 Goodness of fit R^2

Temperature approximation	Reflectivity	Radial tilt
1DT rational	0.9922	0.9870
1DT Padé	0.9928	0.9873
2DT Holtslag	0.9951	0.9877

mations (with a threshold) fit the measurement data (Fig. 13 and Fig. 14). The *goodness of fit* R^2 of these approximations is tabulated in Table 3.

All approximations, especially the 2DT Holtslag approximation, explain the measurements very well.

The maximum stack temperature explains also the observed radial tilt well. Initialization shrinks the stack, introducing a bending moment into the disk. Notice that the onset of the warpage (Fig. 14) occurs at a lower temperature than the onset of crystallinity (Fig. 13). So crystallization shrinkage of the phase change layer *cannot* cause the warpage.

Conclusions

A temperature *series solution* and two *approximations* have been derived for the problem of transient heat conduction in a heat-generating layer between two semi-infinite media at uniform initial temperature.

The *series solution* converges within ten terms to 0.1 K truncation error in the test case.

The *rational approximation* has about 3 percent solution accuracy in the test case and is most efficient, making it an attractive solution. It describes (initialization) heating by only *two parameter groups*.

The *Padé-like approximation* has about 1 percent solution accuracy in the test case and is quite efficient. The Padé-like correction improves the accuracy only in the heat-generating layer a little.

Both temperature approximations explain the observed temperature dependent reflectivity and warpage of initialized CD-RWs very well in terms of just two parameter groups, although they overestimate the initialization temperature neglecting tangential conduction.

Nomenclature

b = laser spot width [m]
c_c = $-2H_cH_o+H_s(H_c-H_o)$
c_o = $H_s(2H_c-H_o)+H_ox/l(-2H_c+(H_c+H_s)x/l)$
c_p = specific heat, [J/kg K]
c_s = $-2H_sH_o+H_c(H_s-H_o)$
f = $-he^{-2lq_o}$
g = Laplace transformed heating distribution
G = heating distribution
h = $(H_s-H_o)(H_o-H_c)/[(H_s+H_o)(H_o+H_c)]$
H = coefficient of heat penetration $\sqrt{(k\rho c_p)}$, [W$\sqrt{}$s/m^{2}K]
k = thermal conductivity, [W/mK]
l = layer thickness, [m]
m = required number of terms of series
q_i = $\sqrt{(s/a_i)}$
q'' = heat flux in the absorbing layer, [W/m^2]
r_c = $H_c/(H_c+H_o)$
r_{co} = $(H_c-H_o)/(H_c+H_o)$
r_s = $H_s/(H_s+H_o)$
r_{so} = $(H_s-H_o)/(H_s+H_o)$
R^2 = goodness of fit (explained variation/total variation)
s = Laplace transform parameter
t = time, [s]
t_0 = time when the (center of a Gaussian) distribution passes, [s]
T = temperature, [K]
u = integrated time, [s]
w_i = weight factor (numerical integration)
x = position, [m]
x_i = abscissa (numerical integration)
α = thermal diffusivity $k/(\rho c_p)$, [m^2/s]
Δ = linearized denominator $H_o(H_c+H_s)+(H_o^2+H_cH_s)l/q_c$
η = $(H_s+H_c)/\Sigma_i[(H_{oi}^2+H_cH_s)l_i/k_{oi}]$, [l/$\sqrt{}$s]
Θ = Laplace transformed temperature
ν = heating rate $q''/(l\rho c_p)_{\text{abs}}$, [K/s]
ρ = density, [kg/m^3]
σ_i = $(x-l)/\sqrt{(4\alpha_s t)}+il/\sqrt{(4\alpha_o t)}$
Σ_i = $(x-l)q_s+ilq_o$
τ = heating time, [s]
φ = heating rate $q''/\Sigma_i[(H_{oi}^2+H_cH_s)l_i/k_{oi}]$, [K/s]
χ_i = $|x|/\sqrt{(4\alpha_c t)}+il/\sqrt{(4\alpha_o t)}$
X_i = $|x|q_c+ilq_o$
$\omega_{j,i}$ = $(jx+il)/\sqrt{(4\alpha_o t)}$
Ω_i = $(x+il)q_o$

Subscripts

abs = absorbing (phase change alloy) layer in stack
c = layer c (coating)
o = layer o (optical stack)
s = layer s (substrate)
0 = initial

Appendix

Inverse Laplace Transforms. The inverse Laplace transforms (Prudnikov [14])

$$\mathcal{L}^{-1}\left\{\frac{e^{-a\sqrt{s}}}{s}\right\}=\operatorname{erfc}\frac{a}{2\sqrt{t}} \tag{26}$$

$$\mathcal{L}^{-1}\left\{\frac{e^{-a\sqrt{s}}}{\sqrt{s}(b+\sqrt{s})}\right\}=e^{b(a+bt)}\operatorname{erfc}\left(\frac{a}{2\sqrt{t}}+b\sqrt{t}\right) \tag{27}$$

$$\mathcal{L}^{-1}\left\{\frac{e^{-a\sqrt{s}}}{s(b+\sqrt{s})}\right\}=\frac{1}{b}\left(\operatorname{erfc}\frac{a}{2\sqrt{t}}-e^{b(a+bt)}\operatorname{erfc}\left(\frac{a}{2\sqrt{t}}+b\sqrt{t}\right)\right) \tag{28}$$

$$\mathcal{L}^{-1}\left\{\frac{e^{-a\sqrt{s}}}{s\sqrt{s}(b+\sqrt{s})}\right\}=\frac{2}{b}\sqrt{\frac{t}{\pi}}e^{-a^2/4t}-\frac{ab+1}{b^2}\operatorname{erfc}\frac{a}{2\sqrt{t}}+\frac{1}{b^2}e^{b(a+bt)}\operatorname{erfc}\left(\frac{a}{2\sqrt{t}}+b\sqrt{t}\right) \tag{29}$$

are used to derive the inverse Laplace transform of the temperature approximation.

The exp·erfc product has to be implemented properly, without overflow or underflow for large arguments, typical for 4 byte REAL implementations.

References

[1] Holtslag, A. H. M., 1989, "Calculations on Temperature Profiles in Optical Recording," J. Appl. Phys., **66**, pp. 1530–1543.

[2] Mulholland, G. P., and Cobble, M. H., 1972, "Diffusion Through Composite Media," Int. J. Heat Mass Transf., **15**, pp. 147–160.

[3] Mikhailov, M. D., and Özişik, M. N., 1984, *Unified Analysis and Solutions of Heat and Mass Diffusion*, John Wiley & Sons, New York.

[4] Mikhailov, M. D., Özişik, M. N., and Vulchanov, N. L., 1983, "Diffusion in Composite Layers With Automatic Solution of the Eigenvalue Problem," Int. J. Heat Mass Transf., **26**, pp. 1131–1141.

[5] Myers, G. E., 1998, *Analytical Methods in Conduction Heat Transfer*, 2nd ed., AMCHT Publications, Madison, WI.

[6] van der Tempel, L., Potze, W., and Lammers, J. H., 1999, "Transient Heat Conduction in a Layer Between Two Semi-Infinite Media," *Proceedings of 33rd National Heat Transfer Conference*, Albuquerque, NM, HTD99–100.

[7] Zhang, X., Hendro, W., Fujii, M., Tomimura, T., and Imaishi, N., 2000, "Measurements of the Thermal Conductivity and Thermal Diffusivity of Polymer Melts with the Short-Hot-Wire Method," *14th Symposium on Thermophysical Properties 2000*, Boulder, CO.

[8] Kim, E. K., Kwun, S. I., Lee, S. M., Seo, H., and Yoon, J. G., 2000, "Thermal Boundary Resistance at $Ge_2Sb_2Te_5$/ZnS:SiO_2 Interface," Appl. Phys. Lett., **76**, pp. 3864–3866.

[9] Peng, C., Cheng, L., and Mansuripur, M., 1997, "Experimental and Theoretical Investigations of Laser-Induced Crystallization and Amorphization in Phase-Change Optical Recording Media," J. Appl. Phys., **82**, pp. 4183–4191.

[10] Padé, H., 1892, "Sur la réprésentation approchée d'une fonction par des fractions rationelles," *Annales Sci. École Normale Supérieure*, **9** Suppl., pp. 1–93.

[11] Viète, F., 1646, *Opera Mathematica*, Leiden, the Netherlands.

[12] Abramowitz, M., and Stegun, I. A., *Handbook of Mathematical Functions*, Dover Publications, New York.

[13] Meinders, E., Borg, H., Lankhorst, M., Hellmig, J., Mijiritskii, A., 2001, "Advances in Thermal Modeling of Dual-Layer DVR-Blue Fast-Growth Media," *17th ODS Topical Meeting 2001*, Santa Fe, NM.

[14] Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., 1992, *Integrals and Series, Volume 5: Inverse Laplace Transforms*, Gordon & Breach Science Publishers, Philadelphia, PA, pp. 55–56.

Certain Anomalies in the Analysis of Hyperbolic Heat Conduction

A. Haji-Sheikh
Department of Mechanical and Aerospace Engineering, The University of Texas at Arlington, Arlington, TX 76019-0023

W. J. Minkowycz
Department of Mechanical and Industrial Engineering, University of Illinois at Chicago, Chicago, IL 60607-7022

E. M. Sparrow
Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN 55455-0111

The hyperbolic diffusion equation is often used to analyze laser heating of dielectric materials and in thermal processing of nonhomogeneous materials. In this paper, anomalies in existing solutions of the hyperbolic heat equation are identified. In particular, the singularities associated with the interaction of a wave front and a boundary may cause a violation of the imposed boundary condition. This violation may give rise to physically unacceptable results such as a temperature drop due to heating or a temperature rise due to cooling. The development of appropriate remedies for these happenings is a major focus of this paper. In addition, the unique mathematical features of the hyperbolic heat equation are studied and set forth. Green's function solutions for semi-infinite and infinite bodies are presented. For finite bodies, it is demonstrated that the relevant series solutions need special attention to accelerate their convergence and to deal with certain anomalies. [DOI: 10.1115/1.1447935]

Keywords: Conduction, Heat Transfer, Microstructures, Nanoscale, Thin Fims, Transient

Introduction

The hyperbolic heat equation is often used to study the temperature field and related heat transfer quantities in dielectric materials with nanoscale thickness and in tightly packed microelectronic devices. The solution to this equation may exhibit some anomalies that must be precluded in an admissible solution. The anomalies may arise when a wave front encounters a boundary or another temperature discontinuity.

The focus of this paper is to identify the anomalies and to suggest appropriate remedies. In particular, the conditions for which highly accurate temperature-field solutions can be obtained will be established. Such solutions are of special importance for the verification of numerical simulations. It is demonstrated herein that an acceptable solution must satisfy the governing equation and the boundary conditions at all times between 0 and t.

When the dimension of the thin film is of the order of the carrier's mean free path, the classical solution methods of thermal conduction are no longer valid. The energy transport has wave characteristics and it is possible to study the thermal behavior of the system through a thermal wave model [1]. Various studies [1–7] report and/or studied the appearance of a wave behavior in heterogeneous materials, in dielectric materials with submicron thickness, and other applications. Minkowycz et al. [8] studied a similar wave-like phenomenon in porous media. The generalization diffusion equation for conductors, non-conductors, and semiconductors with submicron structures, assuming constant thermal properties, is

$$\nabla\cdot(k_e\nabla T)+\tau_t\frac{\partial[\nabla\cdot(k_e\nabla T)]}{\partial t}+\left(g+\tau_q\frac{\partial g}{\partial t}\right)=C\frac{\partial T}{\partial t}+C(\tau_e+\tau_q)\frac{\partial T}{\partial t}, \tag{1a}$$

where $T=T(\vec{r},t)$ is the lattice temperature [9–11] and $\vec{r}$ is the position vector. The relation between the electron and lattice temperatures, selected values of relaxation times τ_t, τ_q, and τ_e are in [9–11]. The electron temperature T_e depends on the lattice temperature T by the relation

$$T_e(\vec{r},t)=T(\vec{r},t)+\tau_t\frac{\partial T(\vec{r},t)}{\partial t}. \tag{1b}$$

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division May 7, 2001; revision received August 15, 2001. Associate Editor: D. Poulikakos.

The other variables are: $\tau_e=\tau_t C_e/C$, $C=C_l+C_e$, C_e and C_l are capacitances of electron and lattice, $g=g(\vec{r},t)$ is volumetric heat source, k_e is the equilibrium thermal conductivity, and $\tau_e=\tau_t C_s/C$. When $\tau_e=\tau_t=0$, Eq. (1a) reduces to the thermal wave equation, the subject of this study, then the hyperbolic form of the diffusion equation is

$$\nabla\cdot[k_e\nabla T(\vec{r},t)]+\left[g(\vec{r},t)+\tau_q\frac{\partial g(\vec{r},t)}{\partial t}\right]=C\frac{\partial T(\vec{r},t)}{\partial t}+C\tau_q\frac{\partial^2 T(\vec{r},t)}{\partial t^2}. \tag{2}$$

It is possible to obtain Eq. (2) directly by combining the energy equation

$$-\nabla\cdot\vec{q}+g(\vec{r},t)=C\frac{\partial T(\vec{r},t)}{\partial t} \tag{3a}$$

with the non-Fourier heat conduction relation

$$\vec{q}+\tau_q\frac{\partial\vec{q}}{\partial t}=-k_e\nabla T(\vec{r},t), \tag{3b}$$

and k replaces k_e in the subsequent analyses. Apart from the study of the wave effect, Eq. (2) also can serve as a limiting solution when τ_e and τ_t are small.

Mathematical Analyses

Numerical and analytical solutions of Eq. (2) are widely available in the literature [12–17]. Some variances observed in different numerical results motivated this presentation. The presentation begins by providing a one-dimensional solution to the hyperbolic diffusion equation in a finite body. This solution is selected to show certain anomalies that are often ignored by researchers. Later, the solutions in infinite and semi-infinite space are discussed. The solution is cast in the form of the Green's function solution that can accommodate the volumetric heat source, different initial conditions, and boundary conditions.

Thermal Wave in Finite Bodies. For convenience of algebra, it is possible to transform the differential equation for the thermal-wave form of the diffusion equation into a modified form of the classical wave equation. To accomplish this, and to provide a reference equation for comparison with a classical wave equation, the parameter k/σ^2 replaces $C\tau_q$ [1] to obtain

$$k\nabla^2 T + g(\vec{r},t) + \frac{\alpha}{\sigma^2}\frac{\partial g(\vec{r},t)}{\partial t} = C\frac{\partial T}{\partial t} + \frac{k}{\sigma^2}\frac{\partial^2 T}{\partial t^2} \tag{4}$$

assuming constant thermophysical properties. When the wave speed σ is finite, this equation reduces to a modified wave equation using a transformation $T=\exp(-\sigma^2 t/2\alpha)\Psi$ and, substituting for T in Eq. (4), the result is

$$\nabla^2\Psi(\vec{r},t) + \frac{\sigma^2}{4\alpha^2}\Psi(\vec{r},t) + S^*(\vec{r},t) = \frac{1}{\sigma^2}\frac{\partial^2\Psi(\vec{r},t)}{\partial t^2}, \tag{5a}$$

where

$$S^*(\vec{r},t) = \left[g(\vec{r},t) + \frac{\alpha}{\sigma^2}\frac{\partial g(\vec{r},t)}{\partial t}\right]\exp\left[\frac{\sigma^2 t}{2\alpha}\right]. \tag{5b}$$

Clearly, Eq. (5*a*) is the wave equation with two source terms, $S^*(\vec{r},t)$ and $\sigma^2\Psi/(4\alpha^2)$. In a finite body, this equation can be solved conveniently in a classical manner, using separation of variables, as shown below.

Here, the so-called fundamental solution is presented for a body having homogeneous boundary conditions of the first, second, or third kind on the surface and having prescribed initial conditions. Furthermore, in this derivation, there is no volumetric heat source in the plate, $g(\vec{r},t)=0$. Following these assumptions, one can obtain the solution of Eq. (5*a*) using the method of separation of variables, by assuming $\Psi(\vec{r},t)=R(\vec{r})\Gamma(t)$ and then replacing Ψ in Eq. (5*a*) to obtain

$$\frac{\nabla^2 R(\vec{r})}{R(\vec{r})} + \frac{\sigma^2}{4\alpha^2} = \frac{1}{\sigma^2}\frac{\Gamma''(t)}{\Gamma(t)}. \tag{6}$$

As usual, the right side of Eq. (6) depends on time t, while the left side depends on the position vector $\vec{r}$, therefore, each must be a constant. For convenience of algebra, set $\nabla^2 R(\vec{r})/R(\vec{r}) = v^2/\alpha$ and $\Gamma''(t)/\Gamma(t) = -\omega^2$ to get

$$\omega^2 = (\sigma^2/\alpha)\left[v^2 - \frac{\sigma^2}{4\alpha}\right]. \tag{7}$$

Since v is a deterministic quantity, the general form of the temperature solution when $v \to v_n$, is

$$T = \sum_{n=1}^{\infty} R_n(\vec{r})[C_n\cos(\omega_n t) + D_n\sin(\omega_n t)]\exp[-\sigma^2 t/(2\alpha)]. \tag{8}$$

Two initial conditions are needed in order to determine the coefficients C_n and D_n. Let $T_i(\vec{r}) = T(\vec{r},0)$ and $T_i'(\vec{r}) = \partial T(\vec{r},t)/\partial t|_{t=0}$, then

$$C_n = \frac{1}{N_n}\int_V T_i(\vec{r})R_n(\vec{r})dV, \tag{9a}$$

where

$$N_n = \int_V [R_n(\vec{r})]^2 dV, \tag{9b}$$

and

$$D_n = \frac{1}{\omega_n N_n}\left[\frac{\sigma^2}{2\alpha}\int_{x=0} T_i(\vec{r})R_n(\vec{r})dV + \int_V T_i'(\vec{r})R_n(\vec{r})dV\right]. \tag{9c}$$

The function $R(\vec{r})$ for regular geometries is known [18,19]. As an illustration, using a plate with thickness L insulated on both sides, $\vec{r}$ replaces x in the relations listed above, and the eigenfunction becomes $R_n(x)=\cos[(n-1)\pi x/L]$ yielding $N_n = L$ for $n=1$ and $N_n = L/2$ for $n>1$. Therefore, in the absence of a volumetric heat source and in the presence of homogeneous boundary conditions, the solution of the thermal wave equation in a multi-dimensional space is

$$T(\vec{r},t) = \sum_{n=1}^{\infty}\frac{R_n(\vec{r})}{N_n}\cos(\omega_n t)\exp[-\sigma^2 t/(2\alpha)]$$
$$\int_V T_i(\vec{r}')R_n(\vec{r}')dV' + \sum_{n=1}^{\infty}\frac{R_n(\vec{r})}{\omega_n N_n}\sin(\omega_n t)\exp[-\sigma^2 t/(2\alpha)] \tag{10}$$
$$\left[\frac{\sigma^2}{2\alpha}\int_V T_i(\vec{r}')R_n(\vec{r}')dV' + \int_V T_i'(\vec{r}')R_n(\vec{r}')dV'\right].$$

To describe the contribution of a volumetric heat source $S^*(\vec{r},t)$ in Eq. (5*a*), consider a finite body with volume V subject to homogeneous boundary conditions of the first, second, or third kind. The temperature and its derivative with respect to time are zero at $t=0$, and there is a volumetric heat source within this body. The objective is to obtain an equation that describes temperature as a function of coordinates $\vec{r}$ and time t. As before, separating the variables by using $\Psi(\vec{r},t)=R(\vec{r})\Gamma(t)$ yields the working second order differential equation,

$$\Gamma''(t)R(\vec{r}) + \omega^2\Gamma(t)R(\vec{r}) = \frac{\sigma^2}{k}\left[g(\vec{r},t) + \frac{\alpha}{\sigma^2}\frac{\partial g(\vec{r},t)}{\partial t}\right]\exp\left(\frac{\sigma^2}{2\alpha}\right). \tag{11}$$

For a solution in a finite body it is appropriate to replace v in Eq. (7) by v_n, and hence ω becomes $\omega_n^2 = (\sigma^2/\alpha)[v_n^2 - \sigma^2/(4\alpha)]$. Using the orthogonality condition

$$\int_V R_n(\vec{r})R_m(\vec{r})dV' = \begin{cases} 0 & \text{when } n\neq m \\ N_n & \text{when } n=m \end{cases}, \tag{12}$$

where $R(\vec{r})$ is replaced by $R_n(\vec{r})$, Eq. (11), for any given n reduces to

$$\Gamma_n''(t) + \omega_n^2\Gamma_n(t) = \frac{\sigma^2}{kN_n}e^{\sigma^2 t/2\alpha}\int_V\left[g(\vec{r},t) + \frac{\alpha}{\sigma^2}\frac{\partial g(\vec{r},t)}{\partial t}\right]R_n(\vec{r})dV. \tag{13}$$

The function $\Gamma_n(t)$ has two particular solutions $\sin(\omega_n t)$ and $\cos(\omega_n t)$. Since the prescribed initial conditions suggest that $\Gamma(0)=\Gamma'(0)=0$, the particular solution will not contribute to the final solution (see Eq. (10)). However, the contribution of the source term is

$$T(\vec{r},t) = e^{-\sigma^2 t/2\alpha}\Psi(\vec{r},t)$$
$$= \sum_{n=1}^{\infty}\frac{\sigma^2}{k\omega_n}\int_{\tau=0}^{t} e^{-\sigma^2(t-\tau)/2\alpha}\sin[\omega_n(t-\tau)]$$
$$\times\int_V\frac{R_n(\vec{r})R_n(\vec{r}')}{N_n}\left[g(\vec{r}',\tau) + \frac{\alpha}{\sigma^2}\frac{\partial g(\vec{r}',\tau)}{\partial\tau}\right]dV'd\tau. \tag{14}$$

Based on Eq. (14), the contribution of the volumetric heat source can be mathematically determined for any regular-shaped body.

To complete this presentation, it remains to find the contribution of the nonhomogeneous boundary conditions. This can be accomplished by using a suitable transformation that reduces the problem with nonhomogeneous boundary conditions to one with homogeneous boundary conditions. Alternatively, one can use Eq. (14) to find the Green's function, and then, for the contribution of a nonhomogeneous boundary condition, use the Green's function solution, see Appendix A, Eq. (*A*.10). A Green's function can be obtained indirectly by comparing the volumetric heat source contribution in Eq. (14) with Eq. (*A*.10), or directly by substituting for $g(\vec{r}',\tau) = C\delta(r-r')\delta(t-\tau)$ in Eq. (14). A comparison between Eq. (*A*.10) and Eq. (14) yields the function G_1,

$$G_1(r,t|\vec{r}',\tau)=e^{-\sigma^2(t-\tau)/2\alpha}\times\sum_{n=1}^{\infty}\frac{\sigma^2/\alpha}{\omega_n}\frac{R_n(\vec{r})R_n(\vec{r}')}{N_n}\sin[\omega_n(t-\tau)]. \quad (15a)$$

Once G_1 is known, the function G_2, the second component of $G=G_1+G_2$, Eq. (*A*.4), is readily available using the relation $G_2=(\alpha/\sigma^2)\partial G_1/\partial t$ as

$$G_2(r,t|\vec{r}',\tau)=-\frac{1}{2}G_1(r,t|\vec{r}',\tau)+e^{-\sigma^2(t-\tau)/2\alpha}\times\sum_{n=1}^{\infty}\frac{R_n(\vec{r})R_n(\vec{r}')}{N_n}\cos[\omega_n(t-\tau)]. \quad (15b)$$

The function $R_n(\vec{r})$ in Eqs. (15a–b) now becomes the solution of differential equation $X''(x)/X(x)=-v^2/\alpha$. The functional form of $X(x)$ depends on the type of boundary condition at $y=0$, i.e., $\sin(vx/\sqrt{\alpha})$ for the first kind, $\cos(vx/\sqrt{\alpha})$ for the second kind, and $(h\sqrt{\alpha}/kv)\sin(vx/\sqrt{\alpha})+\cos(vx/\sqrt{\alpha})$ for the third kind. The boundary condition at $y=L$ determines the value of v. Two initial conditions are needed for inclusion in Eq. (10), $T_i(x)=T(x,0)$ and $T_i'(x)=\partial T(x,t)/\partial t|_{t=0}$. In the absence of a volumetric heat source and in the presence of homogeneous boundary conditions at $x=0$ and $x=L$, the contribution of initial conditions from Eq. (*A*.10), designated as $T_I(x,y)$, is the temperature solution,

$$T_I(x,t)=\sum_{n=0}^{\infty}\frac{X_n(x)}{N_n}\cos(\omega_n t)\exp[-\sigma^2 t/(2\alpha)]\times\int_{x=0}^{L}T_i(x')X_n(x')dx'+\sum_{n=0}^{\infty}\frac{X_n(x)}{\omega_n N_n}\sin(\omega_n t)\exp[-\sigma^2 t/(2\alpha)]\times\left[\frac{\sigma^2}{2\alpha}\int_{x=0}^{L}T_i(x')X_n(x')dx'+\int_{x=0}^{L}T_i'(x')X_n(x')dx'\right]. \quad (16)$$

A casual examination of Eq. (16) shows that the first term on the right side is a series whose convergence is, at best, poor. This and other anomalies of this solution are demonstrated through a numerical example. The following example is chosen to accentuate the problems that are inherent in this solution method.

Example 1. For the purpose of discussing this solution technique, consider a slab in Fig. 1 having a thickness L and boundary conditions of the first kind over $x=0$ surface and the second kind over $x=L$ surface. The volumetric heat source within the plate is $g(x,t)=0$ while the initial conditions are also prescribed.

Solution. In this numerical illustration, it is assumed that the initial temperature T_i=constant when $0\leqslant x\leqslant L$, making $T_i'=0$. Moreover, the temperature at $x=0$ is $T=T_o$=constant and the surface at $x=L$ is insulated, therefore, the plate is subject to boundary conditions of the first kind at $y=0$ and the second kind at $y=L$. The numerically obtained solution to this problem should compare well, when $t\leqslant L/\sigma$, with a solution that considers the plate to be a semi-infinite body; the two solutions should be in near perfect agreement.

For the boundary conditions specified in this example, the eigenfunctions are $\sin(v_n x/\sqrt{\alpha})$, for which the eigenvalues are $v_n=(n-1/2)\pi\sqrt{\alpha}/L$ with $n=1,2,3,\ldots$, and the parameter N_n in

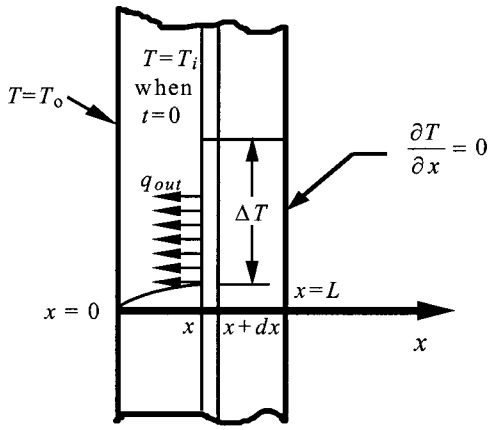


Fig. 1 Schematic of a plate with given boundary conditions

Eq. (55) is equal to $L/2$. Without loss of generality, one can use a dimensionless temperature $[T(x,t)-T_o]/(T_i-T_o)$. Once the eigenvalue v_n is known, Eq. (7) yields the parameter

$$\omega_n=(\sigma v_n/\sqrt{\alpha})[1-\sigma^2/(4\alpha v_n^2)]=\varepsilon_n\sigma(n-1/2)\pi/L,$$

where

$$\varepsilon_n=\sqrt{1-(\alpha L/\alpha)^2/[(2n-1)\pi]^2}.$$

As an important property, $\varepsilon_n\to 1$ as $n\to\infty$, and this helps to maintain the wave nature of this solution. Following appropriate substitutions, Eq. (16) provides the solution for dimensionless temperature with homogeneous boundary conditions,

$$\frac{T(x,t)-T_o}{T_i-T_o}=\left\{\frac{1}{\pi^2}\left(\frac{\sigma L}{\alpha}\right)\times\sum_{n=1}^{\infty}\frac{\sin[(n-1/2)\pi x/L]\sin[(n-1/2)\pi\varepsilon_n\sigma t/L]}{\varepsilon_n(n-1/2)^2}+\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\sin[(n-1/2)\pi x/L]\cos[(n-1/2)\pi\varepsilon_n\sigma t/L]}{n-1/2}\right\}\times\exp[-\sigma^2 t/(2\alpha)]. \quad (17)$$

The convergence of the second term inside the curly bracket is poor since the terms inside the summation reduce to $1/n$ as n increases. However, using the Heaviside function

$$H(\sigma x/\alpha-\sigma^2 t/\alpha)=0 \text{ when } x<\sigma t \text{ or } t>x/\sigma$$
$$=1 \text{ when } x\geqslant\sigma t \text{ or } t\leqslant x/\sigma$$

enhances its convergence of this series. The Fourier series expansion of $H(\sigma x/\alpha-\sigma^2 t/\alpha)$, when $0<\sigma t/L\leqslant 1$, is

$$\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\sin[(n-1/2)\pi x/L]\cos[(n-1/2)\pi\sigma t/L]}{n-1/2}=+H(\sigma x/\alpha-\sigma^2 t/\alpha). \quad (18)$$

However, when $1<\sigma t/L\leqslant 3$, the "+" sign on the right side of Eq. (18) becomes a "−" sign and this sign alternates every time the wave front arrives at the $x=L$ surface. Because $\varepsilon_n\to 1$, as n becomes large, multiplying both sides of Eq. (18) by $\exp[-\sigma^2 t(2\alpha)]$ and then subtracting the resulting equation from Eq. (17) produces an alternative temperature solution

$$\frac{T(x,t)-T_o}{T_i-T_o}=H(\sigma x/\alpha-\sigma^2 t/\alpha)\exp[-\sigma^2 t/(2\alpha)]$$
$$+\left\{\frac{1}{\pi^2}\left(\frac{\sigma L}{\alpha}\right)\sum_{n=1}^{\infty}\frac{\sin[(n-1/2)\pi x/L]\sin[(n-1/2)\pi\varepsilon_n\sigma t/L]}{\varepsilon_n(n-1/2)^2}+\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\sin[(n-1/2)\pi x/L]\Phi_n(x,t)}{n-1/2}\right\}$$
$$\times\exp[-\sigma^2 t/(2\alpha)]\quad\text{when}\quad t<x/\sigma, \tag{19}$$

where

$$\Phi_n(x,t)=\cos[(n-1/2)\pi\varepsilon_n\sigma t/L]-\cos[(n-1/2)\pi\sigma t/L],$$

and $\Phi_n(x,t)$ terms, in the second summation, more rapidly approach zero as n increases. For instance, set $\sigma L/\alpha=1$, then at a temperature jump, when $\sigma x/\alpha=x/\sqrt{\alpha\tau_q}=0.5$ and $\mathrm{Fo}=\alpha t/x^2=2$, numerical evaluation of Eq. (17) converges at near the midpoint of the step at $(T-T_o)/(T_i-T_o)=0.6098$; however, Eq. (19) yields the value of 0.2210 using 200 terms while the accurate value is 0.2212. As time increases and the temperature step passes by, Eq. (17) produces more accurate results. For example, when $\sigma x/\alpha=x/\sqrt{\alpha\tau_q}=0.5$ and $\mathrm{Fo}=\alpha t/x^2=2.5$, Eq. (17), using 200, 500, and 4000 terms yields $(T-T_o)/(T_i-T_o)=0.212899$, 0.214811, 0.215408, respectively. In comparison, Eq. (19) provides a dimensionless temperature of 0.2152888 using 200 terms while the reference value is 0.2152887. The accurate values, mentioned above, are obtained using a solution for semi-infinite bodies as addressed later in this paper. A graphical demonstration of the temperature variation when $\sigma t/L\leqslant 1$ is shown in Fig. 2. An arrow in the figure shows the direction of the wave front. The solid lines are obtained using data from Eq. (17) and the dash lines are from Eq. (19). Except near the jump, both solutions are satisfactory. However, the poor convergence of Eq. (17) in the neighborhood of a step is an expected property of the Fourier series.

Figure 3 shows the solutions using Eq. (17) and Eq. (19). They clearly show an unacceptable negative temperature; an anomaly often ignored by researchers. Moreover, the selection of the Dirichlet boundary condition at $x=0$ has accentuated this anomaly. These solutions, in Fig. 3, are not admissible when $1\leqslant\sigma t/L\leqslant 2$, because it can be demonstrated that they do not satisfy the specified boundary condition of the second kind at $x=L$ when $\sigma^2 t/\alpha=\sigma L/\alpha$ or $\sigma t/L=1$. An admissible solution of a partial differential equation, at time t, must satisfy the partial differential equation, initial conditions, and all boundary conditions when t goes from 0 to t. To demonstrate this, the energy balance over a volume element between x and $x+dx$, as depicted in Fig. 1, is examined. The time required for the front located at x to move from x to $x+dx$ is $dt=dx/\sigma$. During this period, the heat flux entering at $x+dx$ surface is zero, therefore, the thermal energy leaving this volume element at surface x is

$$1\cdot q_{\mathrm{out}}dt=\rho c_p\Delta T(1\cdot dx).$$

Since the temperature step ΔT is the first term on the right side of Eq. (19), then

$$\Delta q(x,t)=-\rho c_p\sigma(T_i-T_o)\exp[-\sigma^2 t/(2\alpha)], \tag{20}$$

where σ has replaced dx/dt and $\Delta q(x,t)=-q_{\mathrm{out}}$. According to Eq. (3*b*), the function $\Delta q(x,t)=\tau_q\partial q(x,t)/\partial t$ since $q(x,t)=0$ at time t just before the arrival of the wave front. As the front arrives to the $x=L$ surface, Eq. (20) indicates the existence of a surface heat flux pulse when $\sigma t/L=1-|\varepsilon|$ as $\varepsilon\to 0$. An examination of the data plotted in Fig. 3 shows that the line for $t/\tau_q=\sigma^2 t/\alpha=1$ does not satisfy the specified homogeneous boundary condition of the second kind at $x=L$; that is, $\partial T/\partial x=0$, unless the solution accounts for this energy pulse. Indeed, without this energy pulse, $\partial T/\partial x\to-\infty$ when $\sigma t/L=1+|\varepsilon|$, as $\varepsilon\to 0$. To correct solutions given by Eqs. (17) and (19), one could use superposition and add an auxiliary solution to Eqs. (17) and (19) in order to satisfy the condition $\partial T/\partial x=0$ at $x=L$ when $\sigma t/L\to 1+|\varepsilon|$ as $\varepsilon\to 0$. This auxiliary solution will have zero conditions everywhere except satisfying the boundary condition

$$k\partial T/\partial x|_{x=L}=\rho c_p\sigma(T_i-T_o)$$
$$\times\exp[-\sigma^2 t/(2\alpha)]\,\delta(x/L-\sigma t/L)|_{x=L}. \tag{21a}$$

The Dirac delta function describes the pulse effect and it is dimensionless. An approximate auxiliary function that satisfies the needed requirement is readily available, see Eq. (19), it is

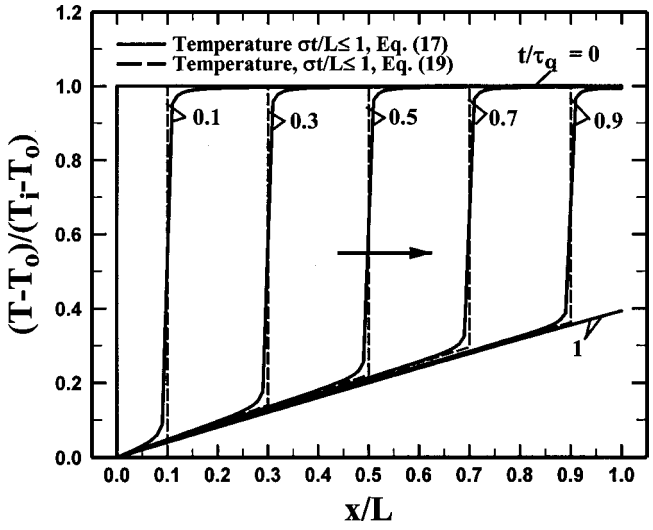


Fig. 2 Dimensionless temperature as a function of x/L for different values of $\sigma t/L$ when $\sigma t/L\leqslant 1$

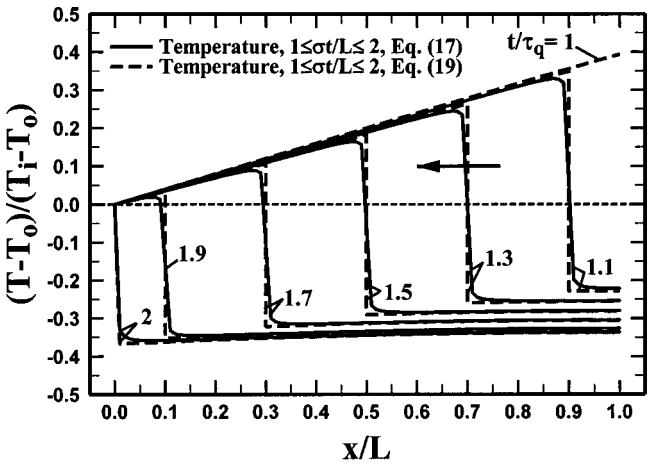


Fig. 3 Dimensionless temperature from Eqs. (17) and (19) as a function of x/L for different values of $\sigma t/L$ when $1\leqslant\sigma t/L\leqslant 2$

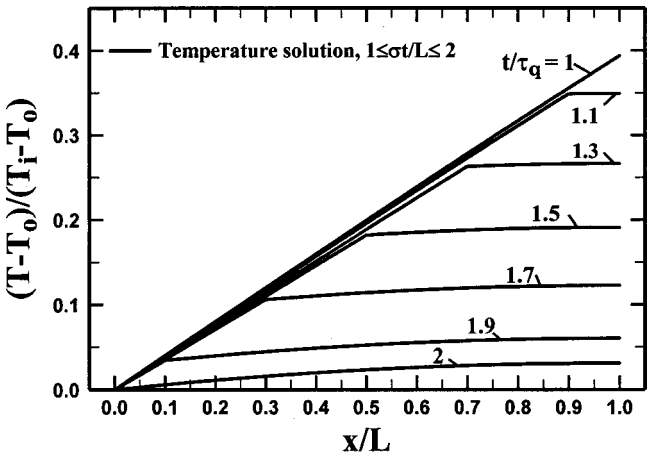


Fig. 4 Corrected dimensionless temperature as a function of x/L for different values of $\sigma t/L$ when $1 \leq \sigma t/L \leq 2$

$$(T_a - T_o)(T_i - T_o) = H[\sigma(L-x)/\alpha - \sigma(\sigma t - L)/\alpha] \times \exp[-\sigma^2 t/(2\alpha)]. \quad (21b)$$

Formally, to satisfy the energy equation, a similar relation is obtainable by using Eq. (*A*.10) retaining only the effect of the boundary condition of the second kind as

$$T_a(x,t) - T_o = \frac{1}{C}\int_{\tau=0}^{t} k\left(G_1 \frac{\partial T}{\partial x}\right)_{x=L} d\tau. \quad (21c)$$

Using G_1 from Eq. (15*a*) and after appropriate substitution, the auxiliary solution when $\sigma t/L > 1$ is

$$\frac{T_a(x,t) - T_o}{T_i - T_o} = \frac{2}{\pi}\sum_{n=1}^{\infty} (-1)^{n+1} \exp[-\sigma^2 t/(2\alpha)] \times \frac{\sin[(n-1/2)\pi x/L]\sin[(n-1/2)\pi\varepsilon_n(\sigma t/L - 1)]}{\varepsilon_n \sigma (n-1/2)}. \quad (21d)$$

Similar to Eq. (21*b*), this auxiliary solution despite the effect of $\sin[(n-1/2)\pi x/L]$ converges to a step similar to Eq. (21*b*) when $\sigma t/L > 1$ and $\varepsilon_n = 1$.

Equation (21*d*) is added to Eq. (17), and the dimensionless temperature is calculated and plotted in Fig. 4. The data in Fig. 4 are physically reasonable and they show the effect of surface heat pulse in Eq. (21*a*) as the front travels back toward $x=0$ plane. The solution whose numerical data are plotted in Fig. 4 is only valid when $1 \leq \sigma t/L < 2$. The auxiliary solution, Eq. (21*d*) does not satisfy the homogeneous boundary condition of the first kind as the wave front arrives to $x=0$ surface when $\sigma t/L = 2$. Briefly, one can replace the temperature step by an equivalent source term eliminating the temperature jump at $x=0$ when $t = 2L/\sigma$. The necessary steps are lengthy and are not presented here.

As an interesting feature of the hyperbolic heat diffusion, many problems in finite bodies can be handled using the solutions for infinite and semi-infinite bodies. Because of the finite propagation speed, this type of analysis is very accurate since it does not use infinite series. Remarks concerning the accuracy of data using Eqs. (17) and (19) are in a later section.

Temperature Solution in Infinite and Semi-Infinite Bodies

Solution in an Infinite Body. The initial goal is to calculate $T(x,t)$ in an infinite solid shown in Fig. 5. The governing equation is Eq. (2) in a one-dimensional space in the absence of a volumetric heat source. The initial temperature of the infinite solid

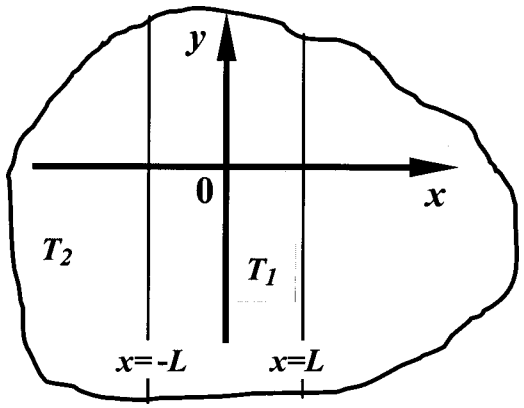


Fig. 5 Schematic of an infinite two-region body

is a constant, T_o, over a finite region between $x=-L$ and $x=+L$ planes, and it is zero elsewhere. Furthermore, all time derivatives of the initial temperature distribution are equal to zero. For simplicity of this presentation, two different temperature solutions are considered: the first region between $x=-L$ and $x=+L$ planes, where

$$\alpha \frac{\partial^2 T_1}{\partial x^2} = \frac{\partial T_1}{\partial t} + \tau_q \frac{\partial^2 T_1}{\partial t^2} \quad (22)$$

subject to the initial conditions $T_1(x,0) = T_o$ for $-L \leq x \leq +L$, and the second region outside of the first region, where

$$\alpha \frac{\partial^2 T_2}{\partial x^2} = \frac{\partial T_2}{\partial t} + \tau_q \frac{\partial^2 T_2}{\partial t^2} \quad (23)$$

with $T_2(x,0) = 0$ for $x < -L$ and $x > +L$. For symmetry about the $x=0$ plane, it is required to have

$$\left(\frac{\partial T_1}{\partial x}\right)_{x=0} = 0. \quad (24)$$

This condition implies that $T_1(-x,t) = T_1(x,t)$ and $T_2(-x,t) = T_2(x,t)$. In addition, the temperature is finite throughout this region; therefore, $|T_2(x,t)|$ is finite. The Laplace transforms of Eqs. (22) and (23) are

$$\alpha \frac{d^2 \overline{T}_1(x,s)}{dx^2} = (s + \tau_q s^2)\overline{T}_1(x,s) - (1 + \tau_q s) T_o \quad (25)$$

and

$$\alpha \frac{d^2 \overline{T}_2(x,s)}{dx^2} = (s + \tau_q s^2)\overline{T}_2(x,s), \quad (26)$$

where $\overline{T}_1 = \mathscr{L}\{T_1\}$ and $\overline{T}_2 = \mathscr{L}\{T_2\}$ are the Laplace transforms of T_1 and T_2. The solution of Eq. (25) using the condition of symmetry at $x=0$ is

$$\overline{T}_1(x,s) = C_1 \cosh[\sqrt{s(\tau_q s + 1)/\alpha}\, x] + \frac{T_o}{s}; \quad \text{for } 0 \leq x \leq L, \quad (27)$$

where C_1 is a constant to be determined. Since $\overline{T}_2(x,s)$ must be finite as $x \to \infty$, the solution of Eq. (26) becomes

$$\overline{T}_2(x,s) = C_2 \exp(-\sqrt{s(\tau_q s + 1)/\alpha}\, x); \quad \text{for } x > L. \quad (28)$$

Next, the conditions at $x = \pm L$ determines the values of C_1 and C_2. Since the temperature must be continuous at $x=L$,

$$\overline{T}_1(L-0,s) = \overline{T}_2(L+0,s). \quad (29a)$$

Also, the heat flux and its time derivative must be continuous because the region is homogeneous and, therefore, the following condition must be satisfied:

$$\frac{d\overline{T}_1(L-0,s)}{dx}=\frac{d\overline{T}_2(L+0,s)}{dx}. \tag{29b}$$

In view of the conditions described by Eqs. (29*a*) and (29*b*), the coefficients C_1 and C_2 can be evaluated. The solution of interest is $\overline{T}_2(x,s)$ as given by Eq. (28). This is because the objective of this example is to allow L to approach to zero. Using the calculated value of C_2, Eq. (28) becomes

$$\overline{T}_2(x,s)=\frac{T_o}{s}\sinh(\sqrt{s(\tau_q s+1)/\alpha}L)\times\exp(-\sqrt{s(\tau_q s+1)/\alpha}x);\quad \text{for } x>L. \tag{30}$$

The solution given by Eq. (30) holds for any value of T_o. The temperature solution when the total excess energy between $x=-L$ and $x=L$ planes is $e_o=\rho c_p(2L\times 1)T_o$ becomes $T_o=e_o/(2L\rho c_p)$, and Eq. (30) reduces to

$$\overline{T}_2(x,s)=\frac{e_o/(\rho c_p)}{2s}\left[\frac{\sinh(\sqrt{s(\tau_q s+1)/\alpha}L)}{L}\right]\times\exp(-\sqrt{s(\tau_q s+1)/\alpha}x);\quad \text{when } x>L. \tag{31}$$

To obtain the temperature solution when an energy pulse q_o is applied to the unit area of the $x=0$ plane, it is necessary to find the limit as $L\to 0$. When L goes to zero, the value of T_o approaches infinity, however, the term in the square brackets, Eq. (31), becomes $\sqrt{s(\tau_q s+1)/\alpha}$, and Eq. (31) reduces to

$$\overline{T}_2(x,s)=\left(\frac{e_o}{\rho c_p}\right)\frac{\sqrt{s(\tau_q s+1)/\alpha}}{2s}\exp(-\sqrt{s(\tau_q s+1)/\alpha}x), \tag{32}$$

which can be written as

$$\overline{T}_2(x,s)=\left(\frac{e_o}{\rho c_p}\right)\frac{(1+\tau_q s)}{2\sqrt{\alpha\tau_q}}\times\frac{\exp[-\sqrt{s(s+1/\tau_q)}(x\sqrt{\tau_q/\alpha})]}{\sqrt{s(s+1/\tau_q)}};\quad \text{for } x>0. \tag{33}$$

Simultaneously, to obtain the Green's function, it is appropriate to consider $(2L\times 1)T_o=1$ and set $T_o=1/(2L)$ in Eq. (3). Then, the Laplace transform of the Green's function is

$$\overline{G}(x,s|0,0)=\frac{(1+\tau_q s)}{2\sqrt{\alpha\tau_q}}\times\frac{\exp[-\sqrt{s(s+1/\tau_q)}(x\sqrt{\tau_q/\alpha})]}{\sqrt{s(s+1/\tau_q)}};\quad \text{for } x>0. \tag{34}$$

The inverse Laplace transform of function $\overline{T}_2(x,s)$ or $\overline{G}(x,s|0,0)$ has two components, here $\overline{G}(x,t|0,0)=\overline{G}_1(x,t|0,0)+\overline{G}_2(x,t|0,0)$. The first component is

$$\overline{G}_1(x,t|0,0)=\frac{1}{2\sqrt{\alpha\tau_q}}\left\{\frac{e^{-\sqrt{s(s+1/\tau_q)}x\sqrt{\tau_q/\alpha}}}{\sqrt{s(s+1/\tau_q)}}\right\}, \tag{35}$$

and the second component is

$$\overline{G}_2(x,t|0,0)=\frac{\tau_q}{2\sqrt{\alpha\tau_q}}\left\{\frac{se^{-\sqrt{s(s+1/\tau_q)}x\sqrt{\tau_q/\alpha}}}{\sqrt{s(s+1/\tau_q)}}\right\}. \tag{36}$$

Using the table of Laplace transforms in Churchill [20] (Appendix 3, Eq. 88), the inverse transformation of Eq. (35) is

$$G_1(x,t|0,0)=\frac{1}{2\sqrt{\alpha\tau_q}}\exp\left[-\frac{1}{2}(t/\tau_q)\right]\times I_0\left[\frac{1}{2\tau_q}\sqrt{t^2-\tau_q x^2/\alpha}\right]H(t-x/\sigma), \tag{37}$$

where H is the Heaviside function, I_0 is the modified Bessel function of 0 order, and $\sigma=\sqrt{\tau_q/\alpha}$ is the speed of the front. These algebraic steps can be repeated for $\overline{T}_2(x,s)$. The Green's function notation $G_1(x,t|x',\tau)$ implies that one unit of energy is released at a time τ over a plane located at $x=x'$, and the effect is observed at a point x at time t. Accordingly, when the Heaviside function $H=1$, Eq. (37) can be written as

$$G_1(x,t|x',\tau)=\frac{1}{2\sqrt{\alpha\tau_q}}\exp\left[-\frac{1}{2}(t-\tau)/\tau_q\right]\times I_0\left[\frac{1}{2}\sqrt{[(t-\tau)/\tau_q]^2-(x-x')^2/(\alpha\tau_q)}\right]. \tag{38}$$

The second component of the Green's function is readily obtainable as

$$G_2(x,t|x',\tau)=\tau_q\partial G_1(x,t|x',\tau)/\partial t. \tag{39}$$

Now, the components $G_1(x,t|x',\tau)$ and $G_2(x,t|x',\tau)$ of the Green's function $G(x,t|x',\tau)$ can be used to obtain the temperature distribution in the presence of a variable heat source and/or initial temperature in an infinite body. This solution is extended to include semi-infinite solids with boundary conditions of the first or second kind at $x=0$.

Solution in Semi-Infinite Bodies. In this case, for the boundary condition of the second kind at $x=0$, one obtains the Green's function using the methodology for Fourier conduction in Beck et al. [18] as

$$G_1(x,t|x',\tau)=\frac{1}{2\sqrt{\alpha\tau_q}}\exp\left[-\frac{1}{2}(t-\tau)/\tau_q\right]\times\left\{I_0\left[\frac{1}{2}\sqrt{[(t-\tau)/\tau_q]^2-(x-x')^2/(\alpha\tau_q)}\right]+I_0\left[\frac{1}{2}\sqrt{[(t-\tau)/\tau_q]^2-(x+x')^2/(\alpha\tau_q)}\right]\right\} \tag{40a}$$

and for boundary condition of the first kind at $x=0$ as

$$G_1(x,t|x',\tau)=\frac{1}{2\sqrt{\alpha\tau_q}}\exp\left[-\frac{1}{2}(t-\tau)/\tau_q\right]\times\left\{I_0\left[\frac{1}{2}\sqrt{[(t-\tau)/\tau_q]^2-(x-x')^2/(\alpha\tau_q)}\right]-I_0\left[\frac{1}{2}\sqrt{[(t-\tau)/\tau_q]^2-(x+x')^2/(\alpha\tau_q)}\right]\right\}. \tag{40b}$$

Example 2. In this example, consideration is given to a semi-infinite solid, Fig. 6, initially at temperature T_i, whose surface at $x=0$ receives a fixed amount of energy e_o, in J/m^2, at time $t=0$.

Solution. A parametric study shows that the dimensionless temperature, using Eq. (40*a*), is related to $\xi=x/\sqrt{\alpha\tau_q}$ and $\mathrm{Fo}=\alpha t/x^2$ since $\tau/\tau_q=\xi^2\mathrm{Fo}$. The inverse Laplace transform of Eq. (33), after replacing e_o by $2e_o$, T_2 by $T-T_i$, $x/\sqrt{\alpha\tau_q}$ by ξ, and τ/τ_q by $\xi^2\mathrm{Fo}$, is

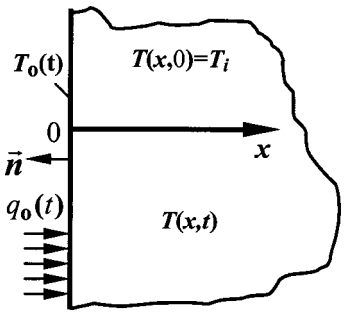


Fig. 6 Schematic of a semi-infinite solid with initial and boundary conditions

$$\frac{x\rho c_p(T-T_i)}{e_o}=\frac{\xi}{2}\exp(-\xi^2\mathrm{Fo}/2)\left\{I_0\left[\frac{1}{2}\sqrt{(\xi^2\mathrm{Fo})^2-\xi^2}\right]+\frac{\xi^2\mathrm{Fo}}{\sqrt{(\xi^2\mathrm{Fo})^2-\xi^2}}I_1\left[\frac{1}{2}\sqrt{(\xi^2\mathrm{Fo})^2-\xi^2}\right]\right\}$$

$$\text{when } \mathrm{Fo}\geq 1/\xi, \tag{41}$$

otherwise $x\rho c_p(T-T_i)/e_o=0$. The dimensionless temperature $x\rho c_p(T-T_i)/e_o$ for different values of $\xi=x/\sqrt{\alpha\tau_q}$ is plotted as a function of the Fourier number $\mathrm{Fo}=\alpha t/x^2$ in Fig. 7. The solid lines in the figure show the discrete nature of the temperature profile. The dashed line is the solution for the same problem using the Fourier conduction [18]. The data show that, when $x/\sqrt{\alpha\tau_q}>2$, the solution of the hyperbolic equation rapidly approaches the Fourier conduction solution. The right side of Eq. (41), when $\mathrm{Fo}=1/\xi$ yields $(\xi/2+\xi^2/8)\exp(-\xi/2)$, the locus of dimensionless temperature at the wave front.

For infinite and semi-infinite regions, the Green's functions are given above. Equation (*A*.10) in Appendix A shows a superposition of three basic solutions for three important contributions from: (*a*) the boundary conditions, (*b*) the volumetric heat source, and (*c*) the initial conditions. Examples describing each of these three contributions are given below.

Example 3. In this example, consideration is given to a semi-infinite solid with a constant initial temperature T_i. It is assumed that the surface temperature is $T(0,t)=T_o$ when $t\geq 0$, and T_o is also a constant.

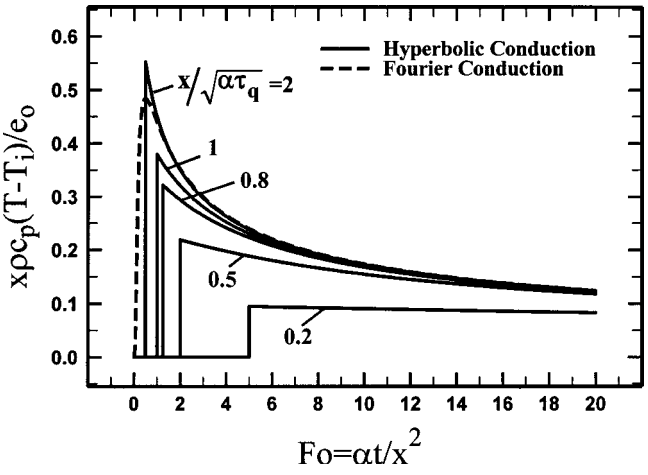


Fig. 7 Temperature solution in a semi-infinite body irradiated with instantaneous energy e_o in J/m^2 at $x=0$ surface

Solution. The temperature solution after casting it in a reduced temperature form $\theta(x,t)=T(x,t)-T_i$, with the surface temperature $\theta_o=T_o(t)-T_i$, becomes

$$\theta=\frac{1}{C}\int_{\tau=0}^{t}d\tau\int_S k\left(-\theta_o(\tau)\frac{\partial G_1}{\partial n}\right)_{S'}dS'$$

$$=\frac{k}{C}\int_{\tau=0}^{t}\theta_o(\tau)\left(\frac{\partial G_1}{\partial x'}\right)\Bigg|_{x'=0}d\tau. \tag{42}$$

After differentiating $G_1=G_1(x,t|x',\tau)$, using Eq. (40*b*), and setting $x'=0$, the result is

$$\left(\frac{\partial G_1}{\partial x'}\right)\Bigg|_{x'=0}=\exp\left[-\frac{\eta}{2}\right]\left\{\frac{\delta[\eta-\xi]}{\alpha\tau_q}+\frac{\xi}{2}\frac{I_1\left[\frac{1}{2}\sqrt{\eta^2-\xi^2}\right]}{\alpha\tau_q\sqrt{\eta^2-\xi^2}}\right\}, \tag{43}$$

where $\eta=(t-\tau)/\tau_q$, $\xi=x/\sqrt{\alpha\tau_q}$. It is important to note an anomaly that the derivative of G_1 in Eq. (43) must have the Dirac delta function $\delta[\eta-\xi]$, see the definition of H in Eq. (37). Following appropriate substitution for $\partial G_1/\partial x'$ at $x'=0$ from Eq. (43) into Eq. (42), the solution for θ is

$$\theta(x,t)=\int_{\eta=\xi}^{t/\tau_q}\theta_o(t/\tau_q-\eta)\exp\left[-\frac{\eta}{2}\right]\times\left\{\delta[\eta-\xi]+\frac{\xi}{2}\frac{I_1\left[\frac{1}{2}\sqrt{\eta^2-\xi^2}\right]}{\sqrt{\eta^2-\xi^2}}\right\}d\eta. \tag{44a}$$

When $\theta_o=$constant, this solution reduces to

$$\theta(x,t)/\theta_o=e^{-\xi/2}H(t/\tau_q-\xi)+\frac{1}{2}\xi\int_{\eta=\xi}^{t/\tau_q}\frac{e^{-\eta/2}I_1\left[\frac{1}{2}\sqrt{\eta^2-\xi^2}\right]}{\sqrt{\eta^2-\xi^2}}d\eta. \tag{44b}$$

The solutions for $\theta(x,t)/\theta_o=[T(x,t)-T_i]/(T_o-T_i)$ as a function of the Fourier number $\mathrm{Fo}=\alpha t/x^2$ and $t/\tau_q=(\mathrm{Fo})\xi^2$ are given in Table 1. The data in Table 1 are accurate to all significant figures appearing in this table. The last column in Table 1 is for $\xi=\infty$; the data are obtained from the relation $\theta(x,t)/\theta_o=\mathrm{erfc}[x/\sqrt{4\alpha t}]$ for Fourier conduction. The right side of Eq. (44*b*), when $\mathrm{Fo}=1/\xi$ or $t/\tau_q=\xi$, yields $\exp(-\xi/2)$, the locus dimensionless temperature at the wave front.

The data in Table 1 are calculated using Mathematica [21]. A function "semx10" in Appendix B.1 has two arguments: "fo" stands for $\mathrm{Fo}=\alpha t/x^2$ and "xi" for $\xi=x/\sqrt{\alpha\tau_q}$. This function was used to export numerical data to a computer data file reported in Table 1.

Alternatively, for a semi-infinite solid, one can use the contribution of the initial conditions $T_i=T(x,0)$ in Eq. (*A*.10) for constant surface temperature condition, T_o, to obtain

$$T(x,t)-T_o=\int_{x'=0}^{\infty}G_1|_{\tau=0}[T_i(x)-T_o]dx'+\tau_q\int_{x'=0}^{\infty}G_1|_{\tau=0}T_i'(x)dx'+\int_{x'=0}^{\infty}G_2|_{\tau=0}[T_i(x)-T_o]dx'. \tag{45}$$

In this example, G_1 is given by Eq. (40*b*) and G_2 by Eq. (39). When $T_i=T(x,0)$ in Eq. (45) is also a constant, following some algebraic steps, the temperature solution is

Table 1 Dimensionless temperature $\theta(x,t)/\theta_i$ for different values of the Fourier number Fo$=\alpha t/x^2$ and $\xi=x/\sqrt{\alpha\tau_q}$

Fo No.	$\theta(x,t)/\theta_i = [T(x,t)-T_i]/(T_o-T_i)$					
	ξ=0.2	ξ=0.5	ξ=0.8	ξ=1	ξ=2	Fourier
0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.50	0.000000	0.000000	0.000000	0.000000	0.367879	0.317311
1.00	0.000000	0.000000	0.000000	0.606531	0.501182	0.479500
1.25	0.000000	0.000000	0.670320	0.624496	0.543561	0.527089
1.50	0.000000	0.000000	0.680671	0.640661	0.576882	0.563703
2.00	0.000000	0.778801	0.699322	0.668492	0.626321	0.617075
2.50	0.000000	0.784711	0.715623	0.691515	0.661699	0.654721
3.00	0.000000	0.790291	0.729954	0.710820	0.688605	0.683091
3.50	0.000000	0.795563	0.742629	0.727215	0.709957	0.705457
4.00	0.000000	0.800549	0.753903	0.741301	0.727437	0.723674
4.50	0.000000	0.805270	0.763984	0.753534	0.742090	0.738883
5.00	0.904837	0.809744	0.773047	0.764261	0.754606	0.751830
5.50	0.905288	0.813987	0.781233	0.773752	0.765458	0.763025
6.00	0.905734	0.818015	0.788663	0.782216	0.774986	0.772830
6.50	0.906175	0.821843	0.795437	0.789818	0.783439	0.781511
7.00	0.906612	0.825484	0.801638	0.796692	0.791005	0.789268
7.50	0.907045	0.828950	0.807336	0.802943	0.797829	0.796253
8.00	0.907474	0.832253	0.812593	0.808659	0.804025	0.802587
8.50	0.907899	0.835402	0.817459	0.813909	0.809684	0.808365
9.00	0.908320	0.838408	0.821978	0.818753	0.814879	0.813664
9.50	0.908737	0.841280	0.826187	0.823241	0.819671	0.818546
10.0	0.909150	0.844025	0.830119	0.827412	0.824108	0.823063

$$\frac{T(x,t)-T_o}{T_i-T_o}=\frac{1}{2}e^{-\eta/2}\int_{\xi'=0}^{\xi}\left\{I_0\left[\frac{1}{2}\sqrt{\eta^2-\xi'^2}\right]+\eta I_1\left[\frac{1}{2}\sqrt{\eta^2-\xi'^2}\right]\Big/\sqrt{\eta^2-\xi'^2}\right\}d\xi'. \quad (46)$$

At the wave front, when Fo$=1/\xi$ or $\eta=t/\tau_q=\xi$, the integral in Eq. (46) reduces to

$$\frac{\pi\xi}{2}\{I_{1/2}(\xi/4)I_{-1/2}(\xi/4)+[I_{1/2}(\xi/4)]^2\}$$
$$=2\sinh(\xi/4)\cosh(\xi/4)+2\sinh^2(\xi/4)$$

and the right side of Eq. (46) yields the locus of dimensionless temperature at the wave front, $1-\exp(-\xi/2)$.

According to Eq. (37), the integrand in Eq. (46) vanishes when $\xi>\eta$, therefore, the upper limit in Eq. (46) should not exceed the value of η. As before, Mathematica [21] is used to generate numerical data. The results, using function "semi1" in Appendix B.2, show that these two Mathematica programs provide identical results. The output of the "semx10" function is $\theta(x,t)/\theta_o$ while the output of "semi1" function is $[T(x,t)-T_o]/(T_i-T_o)=1-\theta(x,t)/\theta_o$. Note that using "WorkingPrecision" statement enables one to achieve a high degree of accuracy. In this illustration, the data from "semx10" and "semi1" programs in Appendix B are identical to all 40 digits in the "WorkingPrecision" statements. Certainly, even accuracy provided by standard precision is sufficient for verification of the accuracy of different numerical schemes.

The numerical output from a Mathematica program, using function "semx10" (or "semi1") with standard precision, were sent to a graphical package that produced Fig. 8. Figure 8 enables one to compare the results from hyperbolic and parabolic conduction models. When $\xi=x/\sqrt{\alpha\tau_q}=2$, data in Fig. 8 show a trend similar to those in Fig. 7, indicating that the solution for the hyperbolic heat conduction model approaches that of the Fourier heat conduction model when $x/\sqrt{\alpha\tau_q}>2$.

Example 4. The reason for this example is to study the contribution of the volumetric heat source in Eq. (*A*.10). Consideration is given to a semi-infinite body initially at temperature T_i and whose surface, at $x=0$, is being irradiated by the amount of $g(x,t)=\dot{q}_o(t)\,\delta(x-0)$, beginning at time $t=0$.

Solution. By viewing $g(x,t)$ as a volumetric heat source one can consider the $x=0$ surface to be insulated, therefore Eq. (*A*.10) reduces to

$$T(x,t)-T_i=\frac{1}{C}\int_{\tau=0}^{t}d\tau\int_{x'=0}^{\infty}G_1\left[g(x',\tau)+\tau_q\frac{\partial g(x',\tau)}{\partial\tau}\right]dx'. \quad (47)$$

Substituting for $g(x',\tau)=\dot{q}_o(\tau)\,\delta(x'-0)$ in Eq. (47), integrating over x' and utilizing the property of the Dirac delta function produces equation

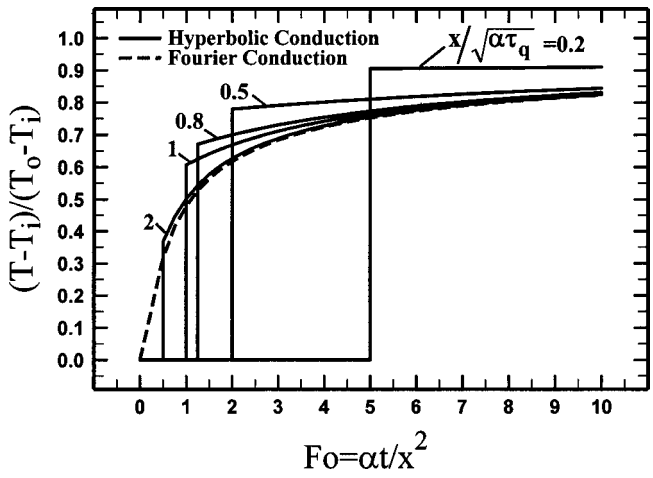


Fig. 8 Temperature solution in a semi-infinite body with a change in the surface temperature

$$T(x,t)-T_i=\frac{1}{C}\int_{\tau=0}^{t} G_1(x,t|\tau,0)\left[\dot{q}_o(\tau)+\tau_q\frac{\partial\dot{q}_o(\tau)}{\partial\tau}\right]d\tau. \quad (48)$$

Alternatively, one can obtain the same equation by considering $\dot{q}_o(t)$ to be the surface heat flux at $x=0$. In the absence of a volumetric heat source, Eq. (*A*.10) yields

$$T(\vec{r},t)-T_i=\frac{-k}{C}\int_{\tau=0}^{t} G_1(x,t|\tau,0)\left(\frac{\partial T}{\partial x'}\right)_{x'=0} d\tau. \quad (49a)$$

According to the definition of the heat flux relation for non-Fourier heat conduction, one can substitute for $-k\partial T/\partial x=q_o(t)+\tau_q\partial[q_o(t)]/\partial t$ in Eq. (49*a*) and produce Eq. (48). In Eq. (48), the surface heat flux can be time dependent and Eq. (39*a*) provides the function $G_1(x,t|\tau,0)$; the result is

$$T(x,t)-T_i=\frac{\alpha}{\sqrt{\alpha\tau_q}}\int_{\tau=0}^{t}\exp\left[-\frac{1}{2}(t-\tau)/\tau_q\right]\times I_0\left[\frac{1}{2}\sqrt{[(t-\tau)/\tau_q]^2-x^2/(\alpha\tau_q)}\right]\times\left[q(\tau)+\tau_q\frac{\partial q(\tau)}{\partial\tau}\right]d\tau. \quad (49b)$$

However, when $q_o(t)$ has a jump or assumes a constant value, the problem requires special attention. Whenever there is a jump in this function, one must account for its derivative during the jump. In this case of a jump at $t=t_o$, the derivative at the jump is $\partial[q_o(t)]/\partial t=[q_o(t_o+)-q_o(t_o-)]\delta(t-0)$. Therefore, when q_o is a constant and heating begins at $t=0$, Eq. (49*b*) reduces to

$$T(x,t)-T_i=\frac{q_o}{C\sqrt{\alpha\tau_q}}\int_{\tau=0}^{t}\exp\left[-\frac{1}{2}(t-\tau)/\tau_q\right]\times I_0\left[\frac{1}{2}\sqrt{[(t-\tau)/\tau_q]^2-x^2/(\alpha\tau_q)}\right]\times[1+\tau_q\delta(\tau-0)]d\tau \quad (50a)$$

and after replacing $(t-\tau)/\tau_q$ by η to get

$$\frac{k[T(x,t)-T_i]}{xq_o}=\frac{1}{\xi}\int_{\eta=\xi}^{t/\tau_q} e^{-\eta/2}I_0\left[\frac{1}{2}\sqrt{\eta^2-\xi^2}\right]d\eta+\frac{1}{\xi}\exp(-t/2\tau_q)I_0\left[\frac{1}{2}\sqrt{(t/\tau_q)^2-\xi^2}\right] \quad (50b)$$

and $k[T(x,t)-T_i]/xq_o=0$ when $t/\tau_q<\xi=x/\sqrt{\alpha\tau_q}$. The right side of Eq. (50*b*), when $\mathrm{Fo}=1/\xi$ or $t/\tau_q=\xi$, yields $\exp(-\xi/2)/\xi$, the locus of dimensionless temperature at the wave front. Equation (49*a*) can be obtained directly from the hyperbolic diffusion equation using the Laplace transform solution technique and the inversion formula used for Eq. (36). Equation (50*b*) is evaluated numerically using Mathematica, and the computed data are listed in Table 2. The data in Table 2 have accuracies similar to those in Table 1. The Mathematica program ''waveg0'', in Appendix B.3, is used to produce the data and the data were exported to graphical software to produce Fig. 9. Figure 9 also shows a trend observed in Figs. 7 and 8, that is, the data show a tendency to approach the Fourier conduction solution

$$\frac{k[T(x,t)-T_i]}{xq_o}=\frac{2\sqrt{\alpha t}}{x}\exp[-x^2/(4\alpha t)]-\mathrm{erfc}[x/\sqrt{4\alpha t}] \quad (51)$$

when $x/\sqrt{\alpha\tau_q}>2$.

Remarks

The exact solutions in infinite and semi-infinite bodies are capable of determining accurate temperature distribution. A Green's solution presented here can provide the contribution of the initial conditions, boundary conditions, and the volumetric heat source for complex problems. Solutions using Mathematica [21] were tested, and they provided solutions with accuracies exceeding 50 significant figures. The infinite series solution suffers from errors caused by the truncation of terms, but it has special characteristics that were addressed earlier and are further elaborated on below.

The solution of the hyperbolic heat conduction equation in finite bodies has received significant attention [13,16,22]. The so-

Table 2 Dimensionless temperature $kT(x,t)/xq_o$ for different values of the Fourier number $\mathrm{Fo}=\alpha t/x^2$ and $\xi=x/\sqrt{\alpha\tau_q}$

Fo No.	$kT(x,t)/xq_o$					
	ξ=0.2	ξ=0.5	ξ=0.8	ξ=1	ξ=2	Fourier
0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.50	0.000000	0.000000	0.000000	0.000000	0.183940	0.166631
1.00	0.000000	0.000000	0.000000	0.606531	0.431908	0.399282
1.25	0.000000	0.000000	0.837900	0.699158	0.540185	0.505794
1.50	0.000000	0.000000	0.917082	0.787789	0.641038	0.606115
2.00	0.000000	1.557602	1.068410	0.954580	0.825625	0.791186
2.50	0.000000	1.665652	1.211363	1.109479	0.992864	0.959621
3.00	0.000000	1.770875	1.346972	1.254461	1.146959	1.115053
3.50	0.000000	1.873429	1.476103	1.391061	1.290627	1.260020
4.00	0.000000	1.973459	1.599485	1.520490	1.425752	1.396355
4.50	0.000000	2.071100	1.717737	1.643715	1.553706	1.525417
5.00	4.524187	2.166480	1.831388	1.761515	1.675528	1.648248
5.50	4.571579	2.259714	1.940889	1.874528	1.792025	1.765666
6.00	4.618747	2.350913	2.046630	1.983281	1.903843	1.878325
6.50	4.665693	2.440178	2.148949	2.088215	2.011504	1.986757
7.00	4.712421	2.527605	2.248140	2.189701	2.115441	2.091402
7.50	4.758932	2.613283	2.344459	2.288055	2.216017	2.192632
8.00	4.805227	2.697295	2.438131	2.383549	2.313538	2.290758
8.50	4.851310	2.779718	2.529357	2.476417	2.408267	2.386049
9.00	4.897182	2.860626	2.618312	2.566864	2.500432	2.478736
9.50	4.942844	2.940087	2.705154	2.655071	2.590229	2.569022
10.0	4.988300	3.018165	2.790024	2.741194	2.677833	2.657085

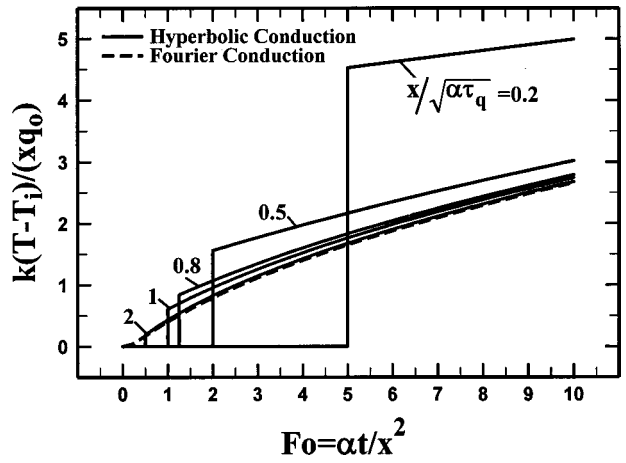


Fig. 9 Temperature solution in a semi-infinite body with a step change in the surface heat flux

lution presented here can be extended to describe radial variation of temperature in spherical coordinates, as described in Guo and Xu [12] for highly packed integrated circuit devices. To provide a broader understanding of solutions in finite bodies, the data from Example 1 are compared with the dimensionless temperature data for a semi-infinite solid in Table 1. The value of $\theta/\theta_i=1-[T(x,t)-T_o]/(T_i-T_o)$ for a plate is computed for $\sigma x/\alpha = x/\sqrt{\alpha\tau_q}=0.2$, 0.5, 1, and for different $\mathrm{Fo}=\alpha t/x^2$. The computed temperature values from Eq. (19) and Eq. (21*d*) (with included Heaviside function) are in columns 2, 4, and 6 of Table 3. Also, temperature data when $t>2L/\sigma$ is included in the table. Similar data for a semi-infinite solid are included in columns 3, 5, and 7. The temperature data for the plate agree well with the temperature data for a semi-infinite body when $\alpha t/x^2 \leqslant [\alpha/(\sigma x)](2L/x-1)$. This dimensionless time also corresponds to the time $(2L-x)/\sigma$ for the effect of surface temperature change at $t=0$ to travel to the insulated boundary at $x=L$ and then return to location x. As stated in [23], these solutions unconditionally satisfy Parsaval's theorem and, therefore, they are unique. This was numerically demonstrated in these comparisons. However, the convergence of the solution using Eq. (17) remains poor in the neighborhood of the wave front.

As discussed earlier, the solution in a finite body exhibits certain anomalies that can result in significant errors when a thermal wave arrives at a surface. It can be shown that this behavior is independent of the solution method. One can obtain an alternative solution to those from Eq. (17) or (19) by using the Green's function in an infinite body or a semi-infinite body. To accomplish this task, an infinite body is selected that has a periodic initial temperature: $[T(x,t)-T_o]/(T_i-T_o)=+1$ when $0\leqslant x/L\leqslant 2$, -1 when $2\leqslant x/L\leqslant 4$, $+1$ when $4\leqslant x/L\leqslant 6, \ldots$, in the positive direction and -1 when $-2\leqslant x/L\leqslant 0$, $+1$ when $-4\leqslant x/L\leqslant -2$, -1 when $-6\leqslant x/L\leqslant -4, \ldots$, in the negative direction. The solution following standard procedure described earlier is

$$\frac{T(x,t)-T_o}{T_i-T_o}=e^{-\bar{\eta}/2}\times\sum_{m=-\infty}^{+\infty}(-1)^m[U(\bar{\eta},\bar{\xi}-\bar{\xi}_m) - U(\bar{\eta},\bar{\xi}-\bar{\xi}_{m-1})], \quad (52a)$$

where, for this illustration, $\bar{\eta}=\sigma^2 t/\alpha$, $\bar{\xi}=x/L$, $\bar{\xi}_m=2m$, $\bar{\xi}_{m-1}=2(m-1)$, and the function U is

$$U(\bar{\eta},\bar{\xi}-\bar{\xi}_m)=\int_0^{\bar{\xi}-\bar{\xi}_m} I_0\left(\frac{1}{2}\sqrt{\bar{\eta}^2-z^2}\right)dz+H(|\bar{\xi}-\bar{\xi}_m|-\bar{\eta}) \quad (52b)$$

and $H(|\bar{\xi}-\bar{\xi}_m|-\bar{\eta})$ is the Heaviside function defined earlier. This solution is simple to implement because of the finiteness of σ; only a very few regions will contribute to this solution. Equation (52a-b) was programmed in Mathematica computer language [21]. It is possible to get results with a specified number of accurate digits. A comparison with Eq. (19) shows that less than 100 terms can provide accurate results. This confirms the earlier statement that a large number of terms are needed when using Eq. (17). This solution also reveals the same anomalies as depicted in Fig. 3 indicating that they could appear in all numerical computations and should be dealt with accordingly.

Table 3 Dimensionless temperature $\theta(x,t)/\theta_i$ for different values of the Fourier number $\mathrm{Fo}=\alpha t/x^2$ and $\xi=\sigma x/\alpha$ when $\sigma L/\alpha=1$

	$\xi=0.2$		$\xi=0.5$		$\xi=1$	
Fo No.	Plate	Semi-Inf.	Plate	Semi-Inf.	Plate	Semi-Inf.
0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.50	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1.00	0.000000	0.000000	0.000000	0.000000	0.606531	0.606531
1.50	0.000000	0.000000	0.000000	0.000000	0.801545	0.640661
2.00	0.000000	0.000000	0.779023	0.778801	0.945750	0.668492
2.50	0.000000	0.000000	0.784711	0.784711	0.973421	0.691515
3.00	0.000000	0.000000	0.790291	0.790291	0.994394	0.710820
3.50	0.000000	0.000000	0.795563	0.795563		
4.00	0.000000	0.000000	0.800549	0.800549		
4.50	0.000003	0.000000	0.805270	0.805270		
5.00	0.905078	0.904837	0.809744	0.809744		
5.50	0.905290	0.905288	0.813987	0.813987		
6.00	0.905734	0.905734	0.818015	0.818015		
6.50	0.906175	0.906175	0.857350	0.821843		
7.00	0.906613	0.906612	0.893874	0.825484		
7.50	0.907046	0.907045	0.927837	0.828950		
8.00	0.907475	0.907474	0.959462	0.832253		
8.50	0.907899	0.907899	0.966664	0.835402		
9.00	0.908320	0.908320	0.973265	0.838408		
9.50	0.908737	0.908737	0.979349	0.841280		
10.0	0.909150	0.909150	0.984989	0.844025		

There are other implications to be mentioned, without presenting mathematical details. If the plate in Fig. 1 is replaced by another plate $2L$ thick, also with a Dirichlet condition at $x=L$, the solution by Eq. (52*a*) should not change. In this case, two waves arrive from two directions and, as they meet, there will be an energy pulse that should be treated as a volumetric heat source and viewed as an anomaly. When these two interacting waves have different intensities, this leads to another interesting problem. The anomalies concerning reflection and interaction of thermal waves will become more demanding for multidimensional bodies. The anomalies are peculiar to specific problems and they should be treated accordingly.

Conclusion

The solution of the thermal wave equation is important in the emerging fields of microscale and nanoscale heat transfer. An exact and accurate solution of the thermal wave equation is valuable for verification and validation of numerical simulations. Verification refers to the accurate solution of the known relations describing the effect of lattice vibration and analogous processes. Usually these partial differential equations describe physical processes when the diffusion equation is inadequate and quantum analyses are too cumbersome. The emphasis is upon determining if the prescribed equations do describe the physical phenomena that are being modeled. A marked improvement of temperature data makes the thermal wave equation a viable model for analyzing thermal characteristics of materials having dimensions of a few nanometers. Although it is difficult to measure temperatures within small devices of nanometer size, a microscale heat conduction equation remains the only viable method for future study of thermally induced stresses and delaminations in small structure thinner than 10 nm; e.g., diffusion barriers, isolation barriers, etc.

Exact analytical solutions, presented in this paper, are useful for verification of large numerical codes of complex processes. The anomalies addressed in this paper appear in exact analysis, Fourier series solution, and numerical simulations. The choice of a solution method depends on the specific application and the desired accuracy. For finite bodies, a Fourier solution is easy to obtain but may require a large number of terms for sufficient accuracy. The exact Green's function in one-dimensional infinite bodies has a limited scope and is useful for finding temperature in some one-dimensional finite bodies and semi-infinite bodies. An analytical or numerical solution should address anomalies that emerge when (1) a wave arrives at a boundary, (2) when two waves interact within a domain, and (3) the effects associated with other discontinuities. In general, one can deal with an anomaly by replacing the temperature discontinuity with an equivalent volumetric heat source for inclusion in the temperature solution. Depending on the system geometry and its boundary conditions, different strategies may be required and, for each case, one should be cognizant of the temperature and heat flux discontinuities.

Nomenclature

A = area, m^2
C = C_e+C_l, $J/m^3\cdot K$
C_e = electron heat capacitance, $J/m^3\cdot K$
C_l = lattice heat capacitance, $J/m^3\cdot K$
e_o = internal energy per unit area, J/m^2
Fo = Fourier number
$G(\cdot)$ = Green's function
$G_i(\cdot)$ = ith Green's function component
g = volumetric heat source, W/m^3
H = Heaviside function
i,j = indices
k = thermal conductivity, $W/m\cdot K$
k_e = effective thermal conductivity, $W/m\cdot K$
L = plate thickness, m
m,n = indices
q = heat flux, W/m^2
R = eigenfunction
$\vec{r}$ = position vector, m
S^* = see equation (5*b*)
s = Laplace transform variable
T = temperature, K
T_e = electron temperature, K
t = time, s
t_p = pulse period, s
U = function, see Eq. (52*b*)
V = volume, m^3
$X(x)$ = eigenfunction
x = coordinate, m
z = dummy variable

Greek

α = thermal diffusivity, m^2/s
γ_n = eigenvalue
δ = Dirac delta function
η = t/τ_q
θ = $(T-T_i)/(T_o-T_i)$
ν = eigenvalue, s^{-1}
ξ = $x/\sqrt{\alpha\tau_q}=\sigma x/\alpha$
σ = thermal wave speed, m/s
τ = dummy variable and Green's function parameter, s
τ_e = lag time, energy equation, s
τ_q = lag time, heat flux, s
τ_t = lag time, temperature, s
Φ = function in Eq. (19)
Ψ = $T\exp(\sigma^2 t/2\alpha)$, K
ω = see Eq. (7), s^{-1}

Subscripts

i = initial condition
o = at $x=0$ surface
m,n = indices

Appendix A

Formulation of the Green's Function Solution. The Green's functions for infinite and semi-infinite bodies are derived for specific boundary conditions. Using the Green's function solution, one can obtain solutions for different boundary and initial conditions. First, it is convenient to establish a definition for the Green's function. Later, a Green's function solution is derived and analyzed through numerical examples.

Definitions. As shorthand notations, let $\vec{r}$ and $\vec{r}'$ be the position vectors and let G stand for the Green's function $G(\vec{r},t|\vec{r}',\tau)$. Now, it is appropriate to find a response at location $\vec{r}$ when there is a source at $\vec{r}'$. Equation (2), describes the thermal wave in the hyperbolic diffusion equation. Replacing $g(\vec{r},t)$ with $C\delta(\vec{r}-\vec{r}')\delta(t-\tau)$ in Eq. (2) results in a partial differential equation that provides the Green's function

$$\nabla\cdot(k_e\nabla G)+C\delta(\vec{r}-\vec{r}')\delta(t-\tau)+C\tau_q\times\frac{\partial}{\partial t}[\delta(\vec{r}-\vec{r}')\delta(t-\tau)]$$

$$=C\frac{\partial G}{\partial t}+C\tau_q\frac{\partial^2 G}{\partial t^2} \qquad (A.1)$$

For convenience of algebra, this equation can be reduced to a simpler form. To accomplish this task, one can define a Green's function $G=G_1+G_2$ so that

$$\nabla\cdot(k_e\nabla G_1)+C\delta(\vec{r}-\vec{r}')\delta(t-\tau)=C\frac{\partial G_1}{\partial t}+C\tau_q\frac{\partial^2 G_1}{\partial t^2} \qquad (A.2a)$$

$$\nabla\cdot(k_e\nabla G_2)+C\tau_q\frac{\partial}{\partial t}[\delta(\vec{r}-\vec{r}')\delta(t-\tau)]=C\frac{\partial G_2}{\partial t}+C\tau_q\frac{\partial^2 G_2}{\partial t^2} \qquad (A.2b)$$

Examining the above two equations, one can hypothesize that $G_2=\tau_q\partial G_1/\partial t$; this is acceptable since both G_1 and G_2 have homogeneous boundary conditions and their initial conditions, including all time derivatives, are zero. To show this relation between G_1 and G_2, simply substitute for G_2 in Eq. (*A*.2*b*) and get

$$\tau_q\frac{\partial}{\partial t}[\nabla\cdot(k_e\nabla G_1)]+C\tau_q\frac{\partial}{\partial t}[\delta(\vec{r}-\vec{r}')\delta(t-\tau)]$$
$$=C\tau_q\frac{\partial^2 G_1}{\partial t^2}+C(\tau_q)^2\frac{\partial^3 G_1}{\partial t^3} \quad (A.3a)$$

that reduces to the equation,

$$\frac{\partial}{\partial t}[\nabla\cdot(k_e\nabla G_1)+C\delta(\vec{r}-\vec{r}')\delta(t-\tau)]=\frac{\partial}{\partial t}\left[C\frac{\partial G_1}{\partial t}+C\tau_q\frac{\partial^2 G_1}{\partial t^2}\right] \quad (A.3b)$$

Notice that any function G_1 that satisfies Eq. (*A*.2*a*) also satisfies Eq. (*A*.3*b*). Therefore, instead of solving for *G* from Eq. (22), it is sufficient to solve a simpler relation, Eq. (*A*.2*a*), and then utilize the relation

$$G=G_1+G_2=G_1+\tau_q\frac{\partial G_1}{\partial t} \quad (A.4)$$

This is consistent with the derivation of Green's function that appeared in an earlier section. The next task is finding a Green's function solution to be used in numerical examples within the main body of this paper.

Green's Function Solution. It is possible to find a generalized Green's function solution to the hyperbolic form of the diffusion equation. For convenience of algebra, Eq. (*A*.2*a*) is used instead of Eq. (*A*.1). Changing the spatial variables in Eq. (*A*.2*a*) to "prime" space and time from t to τ yields

$$\nabla_o\cdot(k_e\nabla_o G_1)+C\delta(\vec{r}'-\vec{r})\delta(t-\tau)=-C\frac{\partial G_1}{\partial\tau}+C\tau_q\frac{\partial^2 G_1}{\partial\tau^2} \quad (A.5)$$

where ∇_o is the "grad" operator in the $\vec{r}'$ space. Moreover, the hyperbolic form of the diffusion equation in $(\vec{r}',\tau)$ space is

$$\nabla_o\cdot(k_e\nabla_o T)+\left[g(\vec{r}',\tau)+\tau_q\frac{\partial g(\vec{r}',\tau)}{\partial\tau}\right]=C\frac{\partial T}{\partial\tau}+C\tau_q\frac{\partial^2 T}{\partial\tau^2} \quad (A.6)$$

Multiplying Eq. (*A*.5) by T and Eq. (*A*.6) by G_1, then subtracting the results to produce equation

$$G_1\nabla_o\cdot(k_e\nabla_o T)-T\nabla_o\cdot(k_e\nabla_o G_1)+G_1\left[g(\vec{r}',\tau)+\tau_q\frac{\partial g(\vec{r}',\tau)}{\partial\tau}\right]$$
$$-C\delta(\vec{r}'-\vec{r})\delta(t-\tau)T$$
$$=C\frac{\partial(TG_1)}{\partial\tau}+C\tau_q\left(\frac{\partial^2 T}{\partial\tau^2}G_1-\frac{\partial^2 G_1}{\partial\tau^2}T\right)$$
$$=C\frac{\partial(TG_1)}{\partial\tau}+C\tau_q\left[\frac{\partial}{\partial\tau}\left(G_1\frac{\partial T}{\partial\tau}\right)-\frac{\partial}{\partial\tau}\left(T\frac{\partial G_1}{\partial\tau}\right)\right] \quad (A.7)$$

First, using standard procedures [17,18], both sides of Eq. (*A*.7) are integrated, $\vec{r}'$ over volume V and τ from 0 to $t+\varepsilon$, where ε is a small positive number. Then, following the application of the Green's theorem and after letting ε to go to zero, one gets

$$T(\vec{r},t)=\frac{1}{C}\int_{\tau=0}^{t}d\tau\int_S k\left(G_1\frac{\partial T}{\partial n}-T\frac{\partial G_1}{\partial n}\right)_{S'}dS'$$
$$+\frac{1}{C}\int_{\tau=0}^{t}d\tau\int_V G_1\left[g(\vec{r}',\tau)+\tau_q\frac{\partial g(\vec{r}',\tau)}{\partial\tau}\right]dV'$$
$$+\int_V G_1|_{\tau=0}T(\vec{r}',0)dV'+\tau_q\int_V G_1|_{\tau=0}T'(\vec{r}',0)dV'$$
$$-\tau_q\int_V\left.\frac{\partial G_1}{\partial\tau}\right|_{\tau=0}T(\vec{r}',0)dV' \quad (A.8)$$

Since $\partial G_1/\partial\tau=-\partial G_1/\partial t$ then

$$-\tau_q\int_V\left.\frac{\partial G_1}{\partial\tau}\right|_{\tau=0}T(\vec{r}',0)dV'=\tau_q\int_V\left.\frac{\partial G_1}{\partial t}\right|_{\tau=0}T(\vec{r}',0)dV'$$
$$=\int_V G_2|_{\tau=0}T(\vec{r}',0)dV' \quad (A.9)$$

The final form of the Green's function solution is

$$T(\vec{r},t)=\frac{1}{C}\int_{\tau=0}^{t}d\tau\int_S k\left(G_1\frac{\partial T}{\partial n}-T\frac{\partial G_1}{\partial n}\right)_{S'}dS'$$
$$+\frac{1}{C}\int_{\tau=0}^{t}d\tau\int_V G_1\left[g(\vec{r}',\tau)+\tau_q\frac{\partial g(\vec{r}',\tau)}{\partial\tau}\right]dV'$$
$$+\int_V G_1|_{\tau=0}T(\vec{r}',0)dV'+\tau_q\int_V G_1|_{\tau=0}T'(\vec{r}',0)dV'$$
$$+\int_V G_2|_{\tau=0}T(\vec{r}',0)dV' \quad (A.10)$$

On the right side of Eq. (*A*.10) there are three contributions: boundary conditions is the first term, volumetric heat source is the second term, and the last three terms describe the effects of the initial conditions.

Appendix B

1 The following Mathematica program uses a function "semx10" to calculate temperature in a semi-infinite body with boundary condition of the first kind at $x=0$.

```
In[1]:=
semx10[fo_,xi_]:=(
tb=fo*xi^2;
f1=Exp[-eta/2]/2;
f2=BesselI[1,Sqrt[eta^2-xi^2]/2];
f=f1*f2/Sqrt[eta^2-xi^2];
rt=Exp[-xi/2]+xi*(NIntegrate[f,{eta,xi,tb}, WorkingPrecision→40]);
Return[rt])
N[semx10[2,8/10],40]
N[semx10[4,1/2],40]
Out[2]=
0.699322278263895295711011962194785941123
Out[3]=
0.800549421193494458700016148728201640524
```

2 The following Mathematica program uses a function "semi1" to calculate temperature in a semi-infinite body with prescribed initial conditions.

```
In[1]:=
semi1[fo_,xi_]:=(
tb=fo*xi^2;
eta=tb;
f1=Exp[-eta/2]/2;
f2=Besse1I[0,Sqrt[eta^2-x^2]/2];
f3=BesselI[1,Sqrt[eta^2-x^2]/2]/Sqrt[eta^2-x^2];
rt=f1*NIntegrate[f2+eta*f3,{x,0,xi}, WorkingPrecision→40];
Return[rt])
N[1-semi1[2,8/10],40]
N[1-semi1[4,1/2],40]
Out[2]=
0.699322278263895295711011962194785941123
Out[3]=
0.800549421193494458700016148728201640524
```

3 The following Mathematica program uses a function "semg0" to calculate temperature in a semi-infinite body with

known heat source.

```
In[1]:=
semg0[fo_,xi_]:=(
tb=fo*xi^2;
f1=Exp[-eta/2];
f2=BesselI[0,Sqrt[eta^2-xi^2]/2];
f=f1*f2;
f3=f/.eta->tb;
rt=(NIntegrate[f,{eta,xi,tb}]+f3)/xi;
Return[rt])
N[semg0[2,8/10],16]
N[semg0[4,1/2],16]
Out[2]=
1.068409638680114
Out[3]=
1.973459030378494
```

References

[1] Tzou, D. Y., 1997, *Macro-to-Microscale Heat Transfer*, Taylor and Francis, New York.

[2] Chester, M., 1963, ''Second Sound in Solids,'' Phys. Rev., **131**, pp. 2013–2015.

[3] Maurer, M. J., and Thompson, H. A., 1973, ''Non-Fourier Effects at High Heat Flux,'' ASME J. Heat Transfer, **95**, pp. 284–286.

[4] Chandrasekharaiah, D. S., 1986, ''Thermoelasticity With Second Sound: A Review,'' Appl. Mech. Rev., **39**, pp. 355–376.

[5] Bertman, B., and Sandiford, D. J., 1970, ''Second Sound in Solid Helium,'' Sci. Am., **222**, pp. 92–101.

[6] Xu, Yun-Sheng, and Guo, Zeng-Yuan, 1995, ''Heat Wave Phenomena in IC Chips,'' Int. J. Heat Mass Transf., **38**, pp. 2919–2922.

[7] Kaminski, W., 1990, ''Hyperbolic Heat Conduction Equation for Materials With a Nonhomogeneous Inner Structure,'' ASME J. Heat Transfer, **112**, pp. 555–560.

[8] Minkowycz, W. J., Haji-Sheikh, A., and Vafai, K., 1999, ''On Departure From Local Thermal Equilibrium in Porous Media Due to a Rapidly Changing Heat Source. The Sparrow Number,'' Int. J. Heat Mass Transf., **42**, No. 18, pp. 3373–3385.

[9] Hays-Stang, Kathy J., and Haji-Sheikh, A., 1999, ''A Unified Solution For Heat Conduction In Thin Films,'' Int. J. Heat Mass Transf., **42**, pp. 455–465.

[10] Qui, T. Q., and Tien, C. L., 1992, ''Short-Pulse Laser Heating on Metals,'' Int. J. Heat Mass Transf., **35**, pp. 719–726.

[11] Qui, T. Q., and Tien, C. L., 1993, ''Heat Transfer Mechanisms During Short-Pulse Laser Heating on Metals,'' ASME J. Heat Transfer, **115**, pp. 835–841.

[12] Guo, Z.-Y., and Xu, Y.-S., 1995, ''Non-Fourier Heat Conduction in IC Chip,'' ASME J. Electron. Packag., **117**, pp. 174–177.

[13] Ozisik, M. N., and Vick, B., 1984, ''Propagation and Reflection of Thermal Waves in a Finite Medium,'' Int. J. Heat Mass Transf., **27**, pp. 1845–1854.

[14] Ozisik, M. N., and Tzou, D. Y., 1994, ''On the Wave Theory in Heat Conduction,'' ASME J. Heat Transfer, **116**, pp. 526–535.

[15] Chen, H.-T., and Lin, J.-Y., 1994, ''Analysis of Two-Dimensional Hyperbolic Heat Conduction Problems,'' Int. J. Heat Mass Transf., **37**, pp. 153–164.

[16] Haji-Sheikh, A., and Beck, J. V., 1994, ''Green's Function Solution for Thermal Wave Equation in Finite Bodies,'' Int. J. Heat Mass Transf., **37**, No. 17, pp. 2615–2626.

[17] Tzou, D. Y., 1995, ''A Unified Field Approach for Heat Conduction from Macro to Micro-Scales,'' ASME J. Heat Transfer, **117**, p. 8.

[18] Beck, J. V., Cole, K., Haji-Sheikh, A., and Litkouhi, B., 1992, *Heat Conduction Using Green's Functions*, Hemisphere Publ. Corp., Washington D.C.

[19] Ozisik, M. N., 1980, *Heat Conduction*, Wiley, New York.

[20] Churchill, R. V., 1958, *Operational Mathematics*, McGraw-Hill, New York.

[21] Wolfram, S., 1996, *The Mathematica*, Third Edition, Cambridge University Press, Cambridge, UK.

[22] Kar, A., Chan, C. L., and Mazumder, J., 1992, ''Comparative Studies on Nonlinear Hyperbolic and Parabolic Heat Conduction for Various Boundary Conditions: Analytic and Numerical Solutions,'' ASME J. Heat Transfer, **114**, pp. 14–40.

[23] Haji-Sheikh, A., Payne, F. R., and Hays-Stang, and Kathy, J., 2000, ''On Convergence and Uniqueness of Microscale Heat Transfer Equation,'' in *Integral Methods in Science and Engineering*, B. Bertram, C. Constanda, and A. Struthers, eds., Champman & Hall/CRC, London, UK, pp. 167–172.

Ballistic-Diffusive Equations for Transient Heat Conduction From Nano to Macroscales

Gang Chen
Member ASME
Mechanical Engineering Department,
Massachusetts Institute of Technology,
Cambridge, MA 02139-4307

In this paper, we present the derivation of a new type of heat conduction equations, named as the ballistic-diffusive equations, that are suitable for describing transient heat conduction from nano to macroscale. The equations are derived from the Boltzmann equation under the relaxation time approximation. The distribution function is divided into two parts. One represents the ballistic transport originating from the boundaries and the other is the transport of the scattered and excited carriers. The latter is further approximated as a diffusive process. The obtained ballistic-diffusive equations are applied to the problem of transient heat conduction across a thin film and the results are compared to the solutions for the same problem based on the Boltzmann equation, the Fourier law, and the Cattaneo equation. The comparison suggests that the ballistic-diffusive equations can be a useful tool in dealing with transient heat conduction problems from nano to macroscales. Boundary conditions for the derived equations are also discussed. Special emphasis is placed on the consistency of temperature used in the boundary conditions and in the equations. [DOI: 10.1115/1.1447938]

Keywords: Conduction, Heat Tansfer, Microscale, Microstructures, Nanoscale

Introduction

It is well-recognized that heat conduction in micro- and nanoscale and in ultrafast processes may deviate significantly from the predictions of the Fourier law, due to the boundary and interface scattering and the finite relaxation time of heat carriers. In the regime where the particle description of electrons and phonons is valid, the Boltzmann equation (BE) is often used in conjunction with the relaxation time approximation. Solving the Boltzmann equation is generally difficult, although past work done in the radiative heat transfer and neutron transport can often be consulted for the possible solution methods [1]. So far, the solution of the Boltzmann equation for heat conduction has been limited to few simple geometrical configurations such as thin films and superlattices [2,3]. Most engineering problems encountered in heat conduction are much more complex. The inherent difficulties associated with the solution of the Boltzmann equation have significantly limited the consideration of the size and transient effects in these problems. This forms a large contrast with the simplicity of the Fourier heat conduction law.

In this work, the ballistic-diffusive approximation is introduced to derive the ballistic-diffusive heat conduction equations. A previous paper has briefly described the strategy [4]. Here we will give details of the derivations and computational results. This approximation divides the heat flux into two components. One originates from the boundaries and represents the ballistic nature of the transport. The other is contributed by the scattered and excited carriers that are approximated as a diffusive transport process. This approach has its origin in thermal radiation [5,6]. The major difference is that we consider the time dependence of heat conduction. The ballistic-diffusive equations we derive will facilitate the solution of complex engineering problems in nanoscales and in ultrafast processes. In the next section, we will derive these equations. The equations will be applied to transient heat conduction across a thin film.

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division August 31, 2000; revision received August 6, 2001. Associate Editor: A. Majumdar.

Ballistic-Diffusive Equations

We start from the BE under the relaxation time approximation,

$$\frac{\partial f}{\partial t}+\mathbf{v}\cdot\nabla_{\mathbf{r}} f=-\frac{f-f_o}{\tau_\omega}, \tag{1}$$

where f is the carrier distribution function, and f_o is the equilibrium distribution, and $\mathbf{v}$ is the carrier group velocity. The force term that is important for electron transport under an electric field is neglected. This paper emphasizes phonon transport, although we believe that the same strategies can be applied for electron transport. An isotropic relaxation time approximation is used in Eq. (1). It has been shown in radiation heat transfer literature that the ballistic-diffusive approximation to be introduced is also reasonable for anisotropic scattering [5]. Although this work will be limited to isotropic scattering only, the extension to deal with anisotropic scattering, if deemed necessary, can be done following similar derivation for photon transport in thermal radiation.

Rather than working with the distribution function f, we use the intensity notation and transform the Boltzmann equation into [1]

$$\frac{\partial I_\omega}{\partial t}+\mathbf{v}\cdot\nabla_{\mathbf{r}} I_\omega=-\frac{I_\omega-I_{o\omega}}{\tau_\omega}, \tag{2}$$

where $I_\omega(t,\mathbf{r})$ is the phonon intensity defined as

$$I_\omega(t,\mathbf{r},\mathbf{v})|\mathbf{v}|\hbar\omega f(t,\mathbf{r},\mathbf{v})D(\omega)/4\pi, \tag{3}$$

and $D(\omega)$ the phonon density of states per unit volume. The equilibrium distribution $I_{o\omega}$ is determined by the local equilibrium temperature through the Bose-Einstein statistics.

The essence of the ballistic-diffusive approximation is to divide the intensity at any point into two parts,

$$I_\omega(t,\mathbf{r})=I_{b\omega}(t,\mathbf{r})+I_{m\omega}(t,\mathbf{r}), \tag{4}$$

where $I_{b\omega}(t,\mathbf{r})$ represents carriers originating from the boundaries and experiencing out-scattering only,

$$\frac{1}{|\mathbf{v}|}\frac{\partial I_{b\omega}}{\partial t}+\hat{\Omega}\cdot\nabla I_{b\omega}=-\frac{I_{b\omega}}{|\mathbf{v}|\tau_\omega}, \tag{5}$$

and $\hat{\Omega}$ is the unit vector in the direction of carrier propagation. This is the ballistic part since it originates from the boundaries. A general solution for this ballistic part is [7]

$$I_{b\omega}(t,\mathbf{r},\hat{\Omega}) = I_{w\omega}[t-(s-s_o)/|\mathbf{v}|, \mathbf{r}-(s-s_o)\hat{\Omega}] \times \exp\left(-\int_{s_o}^{s} \frac{ds}{|\mathbf{v}|\tau_\omega}\right), \quad (6)$$

where $I_{w\omega}$ is the value of intensity at the boundary point $\mathbf{r}$-$(s-s_o)\hat{\Omega}$. It includes those reflected and also emitted carriers at the boundary. The time retardation is due to the finite speed of the heat carriers.

The other part in Eq. (4) represents those originating from inside the medium due to the excitation and the boundary contributions converted into scattered or emitted phonons after absorption.

$$\frac{\partial I_{m\omega}}{\partial t} + \mathbf{v}\cdot\nabla I_{m\omega} = -\frac{I_{m\omega}-I_{o\omega}}{\tau_\omega}. \quad (7)$$

For this part, we can use the diffusion approximation that is familiar in thermal radiation. This can be derived from the spherical harmonic expansion of the intensity [6],

$$I_{m\omega}(t,\mathbf{r},\hat{\Omega}) = J_{o\omega}(t,\mathbf{r}) + \mathbf{J}_{1\omega}(t,\mathbf{r})\cdot\hat{\Omega}, \quad (8)$$

where $\mathbf{J}_{1\omega}$ is a vector. Substituting Eq. (8) into Eq. (7) yields

$$\frac{1}{|\mathbf{v}|}\frac{\partial J_{o\omega}}{\partial t} + \frac{1}{|\mathbf{v}|}\frac{\partial \mathbf{J}_{1\omega}}{\partial t}\cdot\hat{\Omega} + \hat{\Omega}\cdot\nabla J_{o\omega} + \hat{\Omega}\cdot[\nabla \mathbf{J}_{1\omega}\cdot\hat{\Omega}] = \frac{I_{o\omega}-J_{o\omega}}{\Lambda_\omega} - \frac{\mathbf{J}_{1\omega}\cdot\hat{\Omega}}{\Lambda_\omega}, \quad (9)$$

where $\Lambda_\omega (=|\mathbf{v}|\tau_\omega)$ is the heat carrier mean free path at frequency ω. Integrating the above equation over the solid angle of whole space leads to

$$\frac{4\pi}{|\mathbf{v}|}\frac{\partial J_{o\omega}}{\partial t} + \frac{4\pi}{3}\nabla\cdot\mathbf{J}_{1\omega} = -\frac{4\pi}{\Lambda_\omega}(J_{o\omega}-I_{o\omega}). \quad (10)$$

Multiplying Eq. (8) by unit vector $\hat{\Omega}$ and integrating the obtained equation over the solid angle of whole space gives [6]

$$\frac{1}{|\mathbf{v}|}\frac{\partial \mathbf{J}_{1\omega}}{\partial t} + \nabla J_{o\omega} = -\frac{\mathbf{J}_{1\omega}}{\Lambda_\omega}. \quad (11)$$

Based on the decomposition of intensity, Eq. (4), the heat flux can be expressed as

$$\mathbf{q}(t,\mathbf{r}) = \int\left[\int I_\omega \cos\theta d\Omega\right] d\omega = \mathbf{q}_b + \mathbf{q}_m, \quad (12)$$

where the ballistic and diffusive heat flux are defined as

$$\mathbf{q}_b(t,\mathbf{r}) = \int\left[\int I_{b\omega}(t,\mathbf{r})\cos\theta d\Omega\right] d\omega, \quad (13)$$

$$\mathbf{q}_m(t,\mathbf{r}) = \frac{4\pi}{3}\int \mathbf{J}_{1\omega} d\omega. \quad (14)$$

The energy conservation equation gives

$$-\nabla\cdot\mathbf{q} + \dot{q}_e = \frac{\partial u}{\partial t}, \quad (15)$$

where $\dot{q}_e$ is heat generation per unit volume due to external heat sources and the internal heat generation region is assumed to be much larger than the mean free path of heat carriers. When the later is not true, the internal heat source should also be treated as ballistic term [8]. The internal energy of the heat carriers per unit volume, $u = \int \hbar\omega f D(\omega) d\omega d\Omega/4\pi$, can be written as

$$u = \int\left[\int\left[\frac{I_{b\omega}}{|\mathbf{v}|}d\Omega\right]\right] d\omega + \int\frac{4\pi J_{o\omega}}{|\mathbf{v}|}d\omega = u_b + u_m, \quad (16)$$

where u_b and u_m are the internal energy of the ballistic and the diffusive components, respectively. The total internal energy is related to the equivalent temperature through

$$\frac{du}{\partial t} = C\frac{\partial T}{\partial t} = \frac{\partial u_m}{\partial t} + \frac{\partial u_b}{\partial t}. \quad (17)$$

It should be reminded that in the ballistic regime, the statistical distribution of heat carriers deviates far from equilibrium. The local temperature is best considered as a measure of the local internal energy.

Integrating Eq. (11) over frequency and solid angle, and utilizing Eq. (14), we have the following constitutive relation for the diffusive component,

$$\tau\frac{\partial \mathbf{q}_m}{\partial t} + \mathbf{q}_m = -\frac{k}{C}\nabla u_m, \quad (18)$$

where k is the thermal conductivity,

$$k = \int C_\omega |\mathbf{v}| \Lambda d\omega/3, \quad (19)$$

and τ is an average of the relaxation time that is spectrally weighed by the heat flux at each carrier frequency. Inclusion of the frequency-dependence of the relaxation time can be done through such an averaging procedure but is not the emphasis of this work. If we treat u_m/C as an effective temperature representing the local energy of the diffusive carriers, Eq. (18) is the familiar Cattaneo equation [9]. In the current model, however, q_m and u_m are only part of the heat flux and internal energy. From Eqs. (15) and (17) we have

$$-\nabla\cdot\mathbf{q}_m = \frac{\partial u_m}{\partial t} + \frac{\partial u_b}{\partial t} + \nabla\cdot\mathbf{q}_b - \dot{\mathbf{q}}_e. \quad (20)$$

Taking the divergence of Eq. (18) and utilizing Eq. (20) to eliminate $\mathbf{q}_m$ leads to

$$\tau\frac{\partial^2 u_m}{\partial t^2} + \frac{\partial u_m}{\partial t} = \nabla\left(\frac{k}{C}\nabla u_m\right) - \nabla\cdot\mathbf{q}_b + \left(\dot{q}_e + \tau\frac{\partial \dot{q}_e}{\partial t}\right). \quad (21)$$

In deriving Eq. (21), we have used the following equation:

$$\tau\frac{\partial u_b}{\partial t} + \nabla\cdot\mathbf{q}_b = -u_b \quad (22)$$

that can be derived from Eq. (5). Equation (21) differs from the hyperbolic heat conduction, derived based on the Cattaneo constitutive relation, mainly in the additional ballistic term $\nabla\cdot\mathbf{q}_b$. This additional term, together with the appropriate boundary conditions, however, makes a large difference in the final results, as will be seen later. From Eq. (6) and (13), we can express the ballistic heat flux as

$$\mathbf{q}_b(t,\mathbf{r}) = \int\left[\int I_{w\omega}[t-(s-s_o)/|\mathbf{v}|, \mathbf{r}-(s-s_o)\hat{\Omega}] \times \exp\left(-\int_{s_0}^{s}\frac{ds}{|\mathbf{v}|\tau_\omega}\right)\cos\theta d\Omega\right] d\omega. \quad (23)$$

We will call Eqs. (6) and (21) as the ballistic-diffusive heat conduction equations, or the ballistic-diffusive equations. In Ref. [4], we introduced temperatures $T_m = u_m/C$ and $T_b = u_b/C$ and thus wrote Eq. (21) in terms of temperature T_m rather than u_m. Upon the suggestion of a reviewer, we discard the symbol T_m to avoid the confusion that temperatures may seem additive while in reality, T_m and T_b represent only the local internal energy constitutions. Among the two equations, only the solution of Eq. (21) is mathematically involved. The ballistic term can be expressed explicitly in terms of the values at the boundaries, as in Eq. (23). Solving Eq. (21) is much simpler than solving the Boltzmann equation, which involves seven coordinates in the phase space

(three spatial, three momentum, and one time), because Eq. (21) is averaged over the momentum space and thus involves only three spatial coordinates plus time, as in the Fourier equation. This equation can be solved using standard finite difference or finite volume methods.

Boundary Conditions. The formulation of the ballistic-diffusive equations implies that heat carriers originating from the boundaries be treated as the ballistic components. The boundary heat carriers may be emitted/transmitted from another medium or reflected from the same medium. Because the boundary does not contribute to the diffusive component, thus the diffusive heat flux at the boundary is made of the incident diffusive carriers only,

$$\mathbf{q}_m.\mathbf{n} = -\int_{\hat{\Omega}\cdot\mathbf{n}<0} I_{m\omega}\hat{\Omega}\cdot\mathbf{n}d\Omega, \tag{24}$$

where **n** is the unit vector perpendicular to the boundary. Substituting Eqs. (8) and (14) into the above equation yields,

$$\mathbf{q}_m.\mathbf{n} = -v u_m/2. \tag{25}$$

This is the Marshak boundary condition [6] in the limit of black walls for thermal radiation or neutron diffusion. In the studies by Olfe [4], the boundaries are limited to the blackbody radiation. Several authors [10,11] modified this boundary condition for nonblack surfaces by including the reflected components into the diffusive heat flux. We argue here, without substantial simulation results, that Eq. (25) can be applied to nonblack surfaces by grouping the reflected component into the ballistic part. This means that the boundary values in the ballistic components may be coupled to the diffusion components at the same location. Such a treatment will not increase the mathematical complexity significantly since only algebraic equations are involved for nonblack surfaces.

We can eliminate q_m in Eq. (25) based on the constitutive relation, Eq. (18), to obtain the following boundary conditions for the diffusion components,

$$\tau\frac{\partial u_m}{\partial t} + u_m = \frac{2\Lambda}{3}\nabla u_m\cdot\mathbf{n}, \tag{26}$$

where Λ is the average mean free path of heat carriers.

Application to Transient Heat Conduction Across Thin Films

The validity of the ballistic-diffusive approximation has been widely tested in thermal radiation for steady-state radiative heat transfer in both one-dimensional and two-dimensional geometry [5,6] and we thus anticipate that it should have the same accuracy for steady-state heat conduction problems in similar geometrical configurations. In this section, we will apply the ballistic-diffusive heat conduction equation to transient phonon heat conduction across thin films. Results obtained will be compared to the solutions of the Boltzmann equation, the Cattaneo equation, and the Fourier heat conduction equation. Joshi and Majumdar [12] have previously made a comparison among the Boltzmann equation, the Fourier law, and the Cattaneo equation for this case. Through this example, we will also illustrate deficiencies of the ballistic-diffusive approximation and emphasize the importance of the consistent use of the definition of temperature.

We consider a thin film initially at ambient temperature T_o with a constant specific heat. At time $t=0$, one surface emits phonons at temperature T_l. Both of the surfaces are black phonon emitters. There exists no heat generation inside the film. The Eq. (21) becomes

$$\tau\frac{\partial^2 u_m}{\partial t^2} + \frac{\partial u_m}{\partial t} = \frac{k}{C}\frac{d^2 u_m}{dx^2} - \frac{dq_b}{dx}. \tag{27}$$

The boundary and initial conditions are

$$t=0, \quad T(x,0)=T_o, \quad \left.\frac{\partial T(x,t)}{\partial t}\right|_{t=0} = 0 \tag{28}$$

or in terms of the phonon energy density,

$$t=0 \quad u(x,0)=u_m(x,0)+u_b(x,0)=CT_o, \quad \left.\frac{\partial u(x,t)}{\partial t}\right|_{t=0} = 0 \tag{29}$$

$$x=0, \quad \left(\tau\frac{\partial u_m}{\partial t}+u_m\right)_{x=0} = \frac{2\Lambda}{3}\left(\frac{du_m}{dx}\right)_{x=0} \tag{30}$$

$$\text{and} \quad u_{b1}=CT_1 \tag{31}$$

$$x=L, \quad \left(\tau\frac{\partial u_m}{\partial t}+u_m\right)_{x=L} = -\frac{2\Lambda}{3}\left(\frac{du_m}{dx}\right)_{x=L} \tag{32}$$

$$\text{and} \quad u_{b2}=CT_o. \tag{33}$$

To solve Eq. (27), we need to first determine the ballistic heat flux. From Eq. (6) the intensity of the ballistic components can be expressed as

$$I_b(x,t,\mu) = I_{b1}\left(t-\frac{x}{\mu v}\right)e^{-(x/\Lambda\mu)} \quad (0\leqslant\mu\leqslant 1) \tag{34}$$

$$I_b(x,t,\mu) = I_{b2}\left(t+\frac{x}{\mu v}\right)e^{(L-x)/\mu\Lambda} \quad (-1\leqslant\mu\leqslant 0), \tag{35}$$

where I_{b1} and I_{b2} are intensity of phonons emitted from the wall at the specified time. Based on Eqs. (34) and (35) and boundaries conditions (31) and (33), we can express the boundary heat flux, its derivative, and the internal energy of the ballistic components as

$$q_b(x,t) = \begin{cases} \dfrac{Cv\Delta T}{2}\displaystyle\int_{\mu_t}^{1}\mu e^{-(x/\mu\Lambda)}d\mu + q_{bo}(x,0) & [0\leqslant\mu_t=x/(vt)\leqslant 1] \\ q_{bo}(x,0) & \text{other } \mu_t \end{cases} \tag{36}$$

$$\frac{\partial q_b(x,t)}{\partial x} = \begin{cases} -\dfrac{Cv\Delta T}{2\Lambda}\left(\displaystyle\int_{\mu_t}^{1} e^{-(x/\mu\Lambda)}d\mu + \frac{x\Lambda}{v^2t^2}e^{-vt/\Lambda}\right) + \dfrac{\partial q_{bo}(x,0)}{\partial x} & (0\leqslant\mu_t\leqslant 1) \\ \dfrac{\partial q_{bo}(x,0)}{\partial x} & \text{other } \mu_t \end{cases} \tag{37}$$

$$u_b(x,t) = \begin{cases} \dfrac{C\Delta T}{2}\displaystyle\int_{\mu_t}^{1} e^{-(x/\mu\Lambda)}d\mu + u_{bo}(x,0) & (0\leqslant\mu_t\leqslant 1) \\ u_{bo}(x,0) & \text{other } \mu_t \end{cases}, \tag{38}$$

where $\Delta T = T_1 - T_o$, q_{bo} and u_{bo} are the ballistic terms at the equilibrium conditions,

$$q_{bo}(x,0) = \frac{Cv}{2}\left\{E_3\left(\frac{x}{\Lambda}\right) - E_3\left(\frac{L-x}{\Lambda}\right)\right\} \tag{39}$$

$$u_{bo}(x,0) = \frac{u_o}{2}\left\{E_2\left(\frac{x}{\Lambda}\right) + E_2\left(\frac{L-x}{\Lambda}\right)\right\}, \tag{40}$$

and $u_o = CT_o$. Here E_n is the exponential integral function of order n [6].

Normalization Procedure. In Eqs. (36–38), we deliberately separated the ballistic terms at the equilibrium conditions, q_{bo} and u_{bo} from those involved in the transport. This separation not only facilitates the mathematical operation, but is also necessary to minimize the error caused by the inconsistency inherent in the ballistic-diffusive approximation. One can examine this inconsistency by considering the equilibrium limit, i.e., when the thin film is at thermal equilibrium with an equilibrium temperature T_o. In this case, the left-handed side of Eq. (27) disappears and the resultant equation is

$$\frac{k}{C}\frac{d^2 u_{mo}}{dx^2} - \frac{dq_{bo}}{dx} = 0. \tag{41}$$

Solving this equation with the corresponding boundary conditions and q_{bo} as given in Eq. (39) leads to an expression for the diffusion component, u_{mo}, at equilibrium. This u_{mo}, together with u_{bo} given by Eq. (40), does not give the anticipated uniform internal energy ($u_o = u_{bo} + u_{mo} = CT_o$) and thus temperature T_o. The reason for this inconsistency is that Eq. (8) is only an approximation of the real intensity distribution for I_m. To minimize this inconsistency, we introduce a normalization procedure and define the following non-dimensional parameters,

$$\theta_m = \frac{u_m - u_{mo}}{C\Delta T} \quad \theta_b = \frac{u_b - u_{bo}}{C\Delta T} \quad \theta = \frac{u - u_o}{C\Delta T} = \frac{T - T_o}{\Delta T}$$

$$q_m^* = \frac{q_m - q_{mo}}{Cv\Delta T} \quad q_b^* = \frac{q_b - q_{bo}}{Cv\Delta T} \quad q^* = \frac{q - q_o}{Cv\Delta T}$$

$$t^* = t/\tau \quad \eta = x/L \quad Kn = \Lambda/L, \tag{42}$$

where Kn is the phonon Knudsen number. With these normalized parameters, the inconsistency of the initial distributions for u_{bo} and u_{mo} can be eliminated. This, however, does not mean that the above-discussed inconsistency is all avoided since it is inherent in the diffusion approximation used. We will see later that the temperature distributions calculated from the ballistic-diffusive equations are not as good as the heat flux when compared to the solutions of the Boltzmann equation for the same problem. Equation (27) and its associated boundary conditions can be rewritten as

$$\frac{\partial^2 \theta_m}{\partial t^{*2}} + \frac{\partial \theta_m}{\partial t^*} = \frac{\mathrm{Kn}^2}{3}\frac{d^2 \theta_m}{d\eta^2} - \mathrm{Kn}\frac{dq_b^*}{d\eta} \tag{43}$$

$$t^* = 0, \quad \theta(\eta,0) = 0 \quad \left.\frac{\partial \theta(\eta,t^*)}{\partial t^*}\right|_{t^*=0} = 0 \tag{44}$$

$$\eta = 0, \quad \left(\frac{\partial \theta_m}{\partial t^*} + \theta_m\right)_{\eta=0} = \frac{2\mathrm{Kn}}{3}\left(\frac{d\theta_m}{d\eta}\right)_{\eta=0} \tag{45}$$

$$\text{and} \quad \theta_{b1} = 1 \tag{46}$$

$$\eta = 1, \quad \left(\frac{\partial \theta_m}{\partial t^*} + \theta_m\right)_{\eta=1} = -\frac{2\mathrm{Kn}}{3}\left(\frac{d\theta_m}{d\eta}\right)_{\eta=1} \tag{47}$$

$$\text{and} \quad \theta_{b2} = 0 \tag{48}$$

$$q_b^*(\eta,t^*) = \begin{cases} \frac{1}{2}\int_{\mu_t}^1 \mu e^{-(\eta/\mu\mathrm{Kn})} d\mu & (0 \leqslant \mu_t \leqslant 1) \\ 0 & \text{other } \mu_t \end{cases} \tag{49}$$

$$\frac{\partial q_b^*(\eta,t^*)}{\partial \eta} = \begin{cases} -\frac{1}{2\mathrm{Kn}}\left(\int_{\mu_t}^1 e^{-(\eta/\mu\mathrm{Kn})} d\mu + \frac{\eta}{\mathrm{Kn}t^{*2}} e^{-t^*}\right) & (0 \leqslant \mu_t \leqslant 1) \\ 0 & \text{other } \mu_t \end{cases} \tag{50}$$

$$\theta_b(\eta,t^*) = \begin{cases} \frac{1}{2}\int_{\mu_t}^1 e^{-(\eta/\mu\mathrm{Kn})} d\mu & (0 \leqslant \mu_t \leqslant 1) \\ 0 & \text{other } \mu_t \end{cases}. \tag{51}$$

Equation (43) and its associated boundary conditions are solved numerically using an implicit finite difference scheme. In addition, the Boltzmann equation, i.e., Eq. (2), the hyperbolic heat conduction equation based on the Cattaneo equation, and the diffusion heat conduction equation based on the Fourier law are also solved for the identical geometry and boundary conditions. The Boltzmann equation is solved using Gauss-Legendre quadrature in the angular direction, the forward difference for $0 < \mu < 1$ and the backward difference for $-1 < \mu < 0$, and an implicit scheme in time [1]. The accuracy and convergence are ensured by increasing the number of points in both angular and coordinate directions. The corresponding hyperbolic equation is

$$C\left(\tau\frac{\partial^2 T}{\partial t^2} + \frac{\partial T}{\partial t}\right) = k\frac{\partial^2 T}{\partial x^2} \tag{52}$$

with the following boundary and initial conditions:

$$t = 0 \quad T(x,t) = 0 \quad \frac{\partial T}{\partial t} = 0 \tag{53}$$

$$x = 0 \quad T(0,t) = T_1 \tag{54}$$

$$x = L \quad T(0,t) = T_o. \tag{55}$$

Analytical solutions for Eq. (52) with above boundary conditions have been given before [12], but we found that some discrepancies exist in the final expressions. So we give the temperature and heat flux distributions here,

$$\theta(\eta,t^*) = 1 - \eta - \frac{e^{-t^*/2}}{2\pi}\sum_{n=1}^{\infty}\frac{1}{n}\left[\frac{1+2\delta}{\delta}e^{\delta t^*} - \frac{1-2\delta}{\delta}e^{-\delta t^*}\right]\sin(n\pi\eta) \tag{56}$$

$$q^*(\eta,t^*) = \frac{Kn}{3}\left\{1 - e^{-t^*} + e^{-t^*/2} \times \sum_{n=1}^{\infty}\frac{1}{n\delta}[e^{\delta t^*} - e^{-\delta t^*}]\cos(n\pi\pi)\right\}, \tag{57}$$

where $\delta = \sqrt{1 - n^2\pi^2 Kn^2/3}$. We have checked the analytical solution with direct numerical solution of the Eq. (52), and they compare well with each other.

In addition to the major difference in the governing equations between the ballistic-diffusive and the hyperbolic heat conduction equations, another major difference lies in the boundary condition. In the hyperbolic heat conduction equation, the boundary temperature is assumed to be the temperature of the medium. In the boundary conditions for the ballistic-diffusive heat conduction, the boundary temperature is not specified. Rather, they are deter-

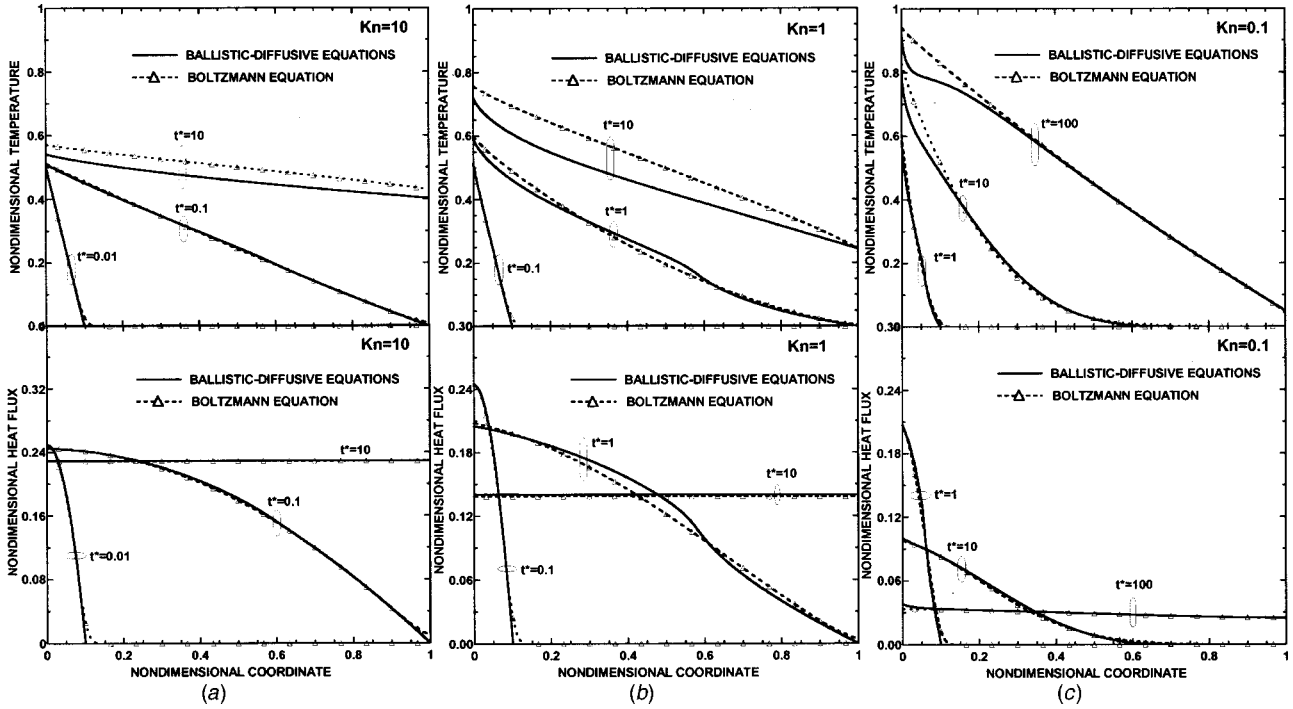


Fig. 1 Nondimensional temperature and heat flux distributions calculated from the Boltzmann equation and the ballistic-diffusive equations for different carrier Knudsen number (Kn) and differential nondimensional time (t^*)

mined based on the calculated u_m and u_b values at the boundaries, for example, $CT(0)=u_m(0)+u_b(0)$. The temperatures specified T_{b1} and T_{b2} at the boundaries are the equivalent temperature of the heat carriers entering the medium from the boundaries. We will discuss this point in more detail in conjunction with the results.

Results and Discussion. Figure 1 shows the nondimensional temperature (or internal energy) and heat flux distributions obtained from the Boltzmann equation and the ballistic-diffusive equations for different values of the Knudsen number at different times. The heat flux calculated from the two approaches generally agrees better with each other than the temperature. Some differences exist between the temperature distributions calculated from the two approaches, particularly as the transport reaches the steady state. This is not surprising since we have discussed above that the diffusion approximation causes some inconsistency in the temperature definition. At short time, the dominant heat carriers are ballistic ones experiencing little scattering. Since the ballistic component originating from the boundaries is treated rigorously, the agreements between the two equations are good. In Fig. 2, we show the ballistic and diffusive contributions to the total local internal energy and heat flux. The back scattered heat carriers flow out of the boundary, thus leading to a negative diffusive heat flux at $\eta=0$. The diffusive component of the internal energy shows an artificial but weak wave front, which cannot be seen from the Boltzmann equation simulation results. This is because even the scattered carriers still travel a distance of the order of one mean free path. This nonlocal transport is replaced by the local gradient in the diffusion approximation.

Figure 3 compares the temperature and heat flux distributions obtained from the ballistic-diffusive equations, the Boltzmann equation, the Cattaneo equation, and the Fourier law. It clearly illustrates that the Cattaneo equation leads to wrong results in the heat flux and temperature distribution inside the medium. Figure 4 gives the surface heat flux history based on the four different equations. The Fourier law leads to unrealistic infinite heat flux as time approaches zero. The Cattaneo equation creates artificial surface heat flux oscillation. The agreement between the ballistic-diffusive equation and the Boltzmann equation is excellent over a wide range of time and length scales.

Further Discussion on Temperature. Both the Boltzmann equation and the ballistic-diffusive equations produce a temperature jump at the interfaces. This temperature jump phenomenon has been discussed in several studies [1,13–16]. However, it still needs further clarification because a consistent picture is crucial when dealing with transport in both sides of the interfaces. One could ask what this temperature jump means and where it occurs. In this article, we have assumed that no interface reflection of phonons occurs, i.e., the interfaces are black phonon emitters and absorbers. The simplest way to create such a situation is to place an imaginary interface in an otherwise continuous thermal conductor. In this case, there should be no temperature jump across this imaginary interface! This dilemma was pointed out by Little [17] in his original work on the thermal boundary resistance but was not explained satisfactory. We mentioned previously that the temperatures used in the boundary conditions for the Boltzmann equation and the ballistic-diffusive equation are the temperatures of the emitted phonons, while those in the Cattaneo and Fourier equations are those of the equilibrium phonons. We should emphasize that the temperature distribution obtained from the Boltzmann or the ballistic-diffusive equations are also the equilibrium temperature, as determined by f_o in Eq. (1). In the case where transport is highly nonequilibrium inside the film, the temperature can only be understood as a measure of the local total energy. Chen [16] called this equivalent equilibrium temperature by imagining that these phonons thermallize adiabatically to equilibrium. It is the inconsistency between the temperatures used in the boundary condition and in the Boltzmann and the ballistic-diffusive equations that created this artificial temperature jump! This statement should be understood on the premise that the interfaces are nonreflecting. In the case that the interfaces are reflecting, the thermal boundary resistance [13] exists and, consequently, a true temperature discontinuity exists. In addition, this

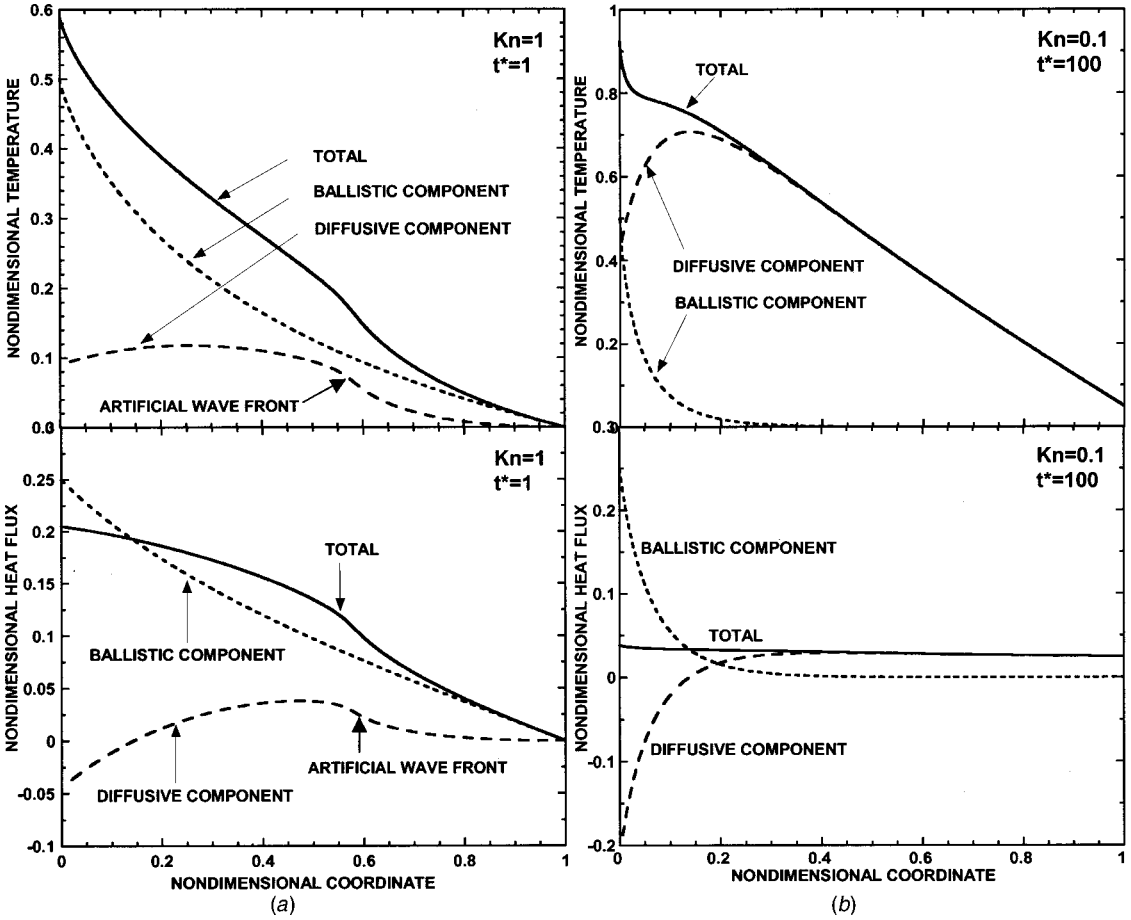


Fig. 2 Contribution of the ballistic and the diffusive components to the nondimensional total internal energy (temperature) and heat flux. The weak wave front in the diffusive component is artificially caused by the diffusion approximation.

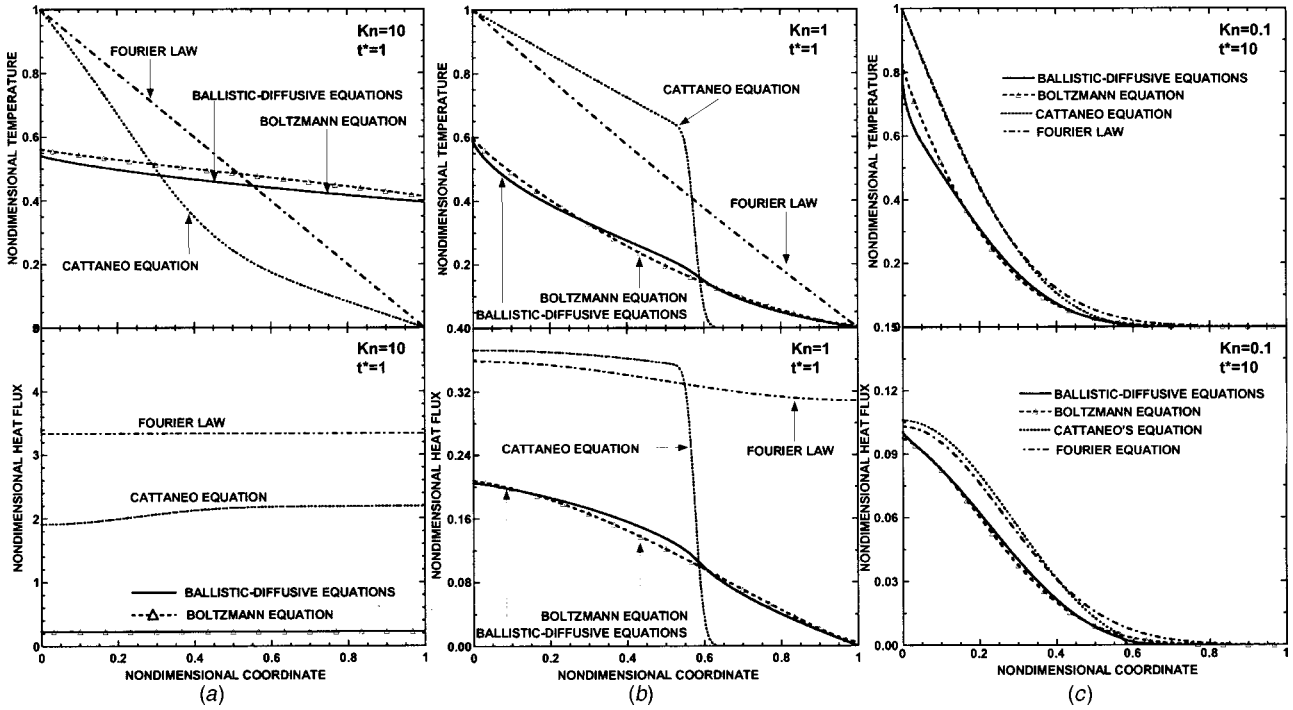


Fig. 3 Comparison of temperature and heat flux distributions obtained from the Boltzmann equation, the ballistic-diffusive equations, the Cattaneo equation, and the Fourier law for different time and Knudsen number

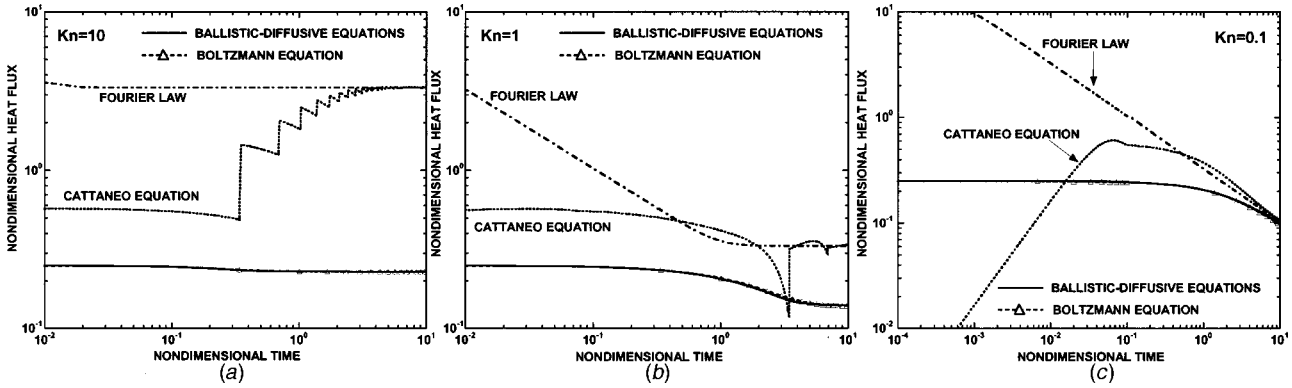


Fig. 4 Comparison of surface heat flux obtained from the Boltzmann equation, the ballistic-diffusive equations, the Cattaneo equation, and the Fourier law as a function of time for different Knudsen numbers

statement does not render the presented results useless, because certain experimental configurations may be close to what is calculated, as explained below.

We will use Fig. 5 to facilitate the discussion. Consider a thin film situated on a substrate. A heater is deposited on the film and it also serves as a temperature sensor. In the modeling, one needs to deal with the interface between the film and the substrate. If the Boltzmann equation or the ballistic-diffusive equations are used to solve the heat conduction problem in the whole structure without distinguishing the difference between the emitted temperature and the equivalent equilibrium temperature, an error will occur. This is particularly true when the thermal boundary resistance is small, as we just discussed for the limiting case when the film and the substrate are identical. Chen and Zeng [18] employed a consistent definition for temperature, i.e., all based on the equivalent equilibrium temperature, for the case of simple diffusion approximation with discontinuous boundary conditions. They were able to show that in the limiting case that when the transmissivity is one, there is no temperature jump or thermal boundary resistance. This approach has been taken in the past in several studies on the issue of thermal boundary resistance [19,20] and is particularly important when one deals with nanostructures with multiple interfaces [16]. Having emphasized the interface between the film and the substrate, we now examine the heater and the temperature sensor. It can be made of a metal or a semiconductor. In a typical temperature measurement, the average temperature of the layer is measured from its electrical resistance. If phonons transmitted from the film into the heater is thermalized within a small distance from the interface and if this distance is much smaller than the heater thickness, the measured temperature is not the temperature of the equivalent equilibrium phonons at the interface, but the average temperature over the whole heater thickness. Most of the time in reality, one can only measure the average temperature of the whole heater, not the equilibrium temperature at the interface. This measured temperature can be considered as the temperature emitted into the film. In this case, there is a temperature drop inside the heater close to the interface, not inside the film. This temperature jump is what is drawn in Figs. 1–3. As long as the temperature drop is not so large that the temperature dependence of specific heat becomes important, however, the results on heat flux and temperature distribution inside the film should not be affected.

The next question we should answer is what are the differences among the four approaches if the temperature and heat flux are all defined based on the equivalent equilibrium temperature. We answer this question by rescaling the temperature and heat flux with the calculated medium temperature at $x=0$ and $x=L$ based on the Boltzmann equation,

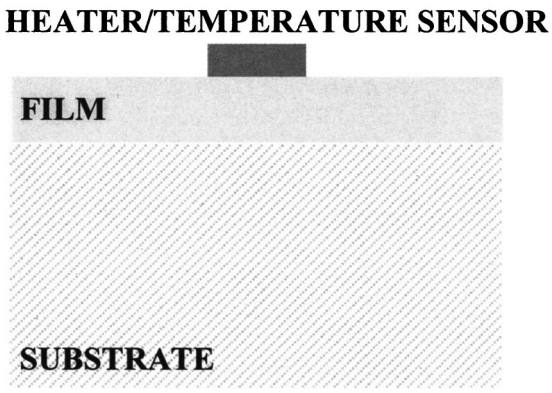


Fig. 5 Schematic drawing for discussing the consistency of temperature used in the equations and the boundary conditions

$$\theta(\eta,t^*)=\frac{T(\eta,t^*)-T_{\mathrm{BE}}(\eta=1,t^*)}{T(\eta=0,t^*)-T_{\mathrm{BE}}(\eta=1,t^*)} \tag{58}$$

and

$$q^*(\eta,t^*)=\frac{q(\eta,t^*)-q_o}{T(\eta=0,t^*)-T_{\mathrm{BE}}(\eta=1,t^*)} \tag{59}$$

Figures 6(*a*) and (*b*) replot the results of Fig. 3(*b*) based on the above definition of heat flux and temperature. They clearly show that the ballistic-diffusive equation is an appropriate tool dealing with transient problems from nanoscale to macroscale. More consistent comparisons between the results of the ballistic-diffusive equations and those of the Fourier law based on the equivalent equilibrium temperature specified at the boundaries are given in Ref. [8].

Finally, we would like to remind that the results is less accurate in the steady state and when the diffusive component is dominant, because the diffusive component is only an approximation to the scattered heat carriers, the transport of which is still nonlocal in nature. Figure 2(*b*) shows that the largest error occurs at the boundaries because of the ballistic transport of the scattered carriers. This may be improved through establishing boundary conditions that compensate for the ballistic carriers close to the boundary. It should also be pointed out that this work is limited to the transport of single type of carrier and that the equation has not been extensively investigated for complicated geometrical configurations. Through further study, we hope that the ballistic-

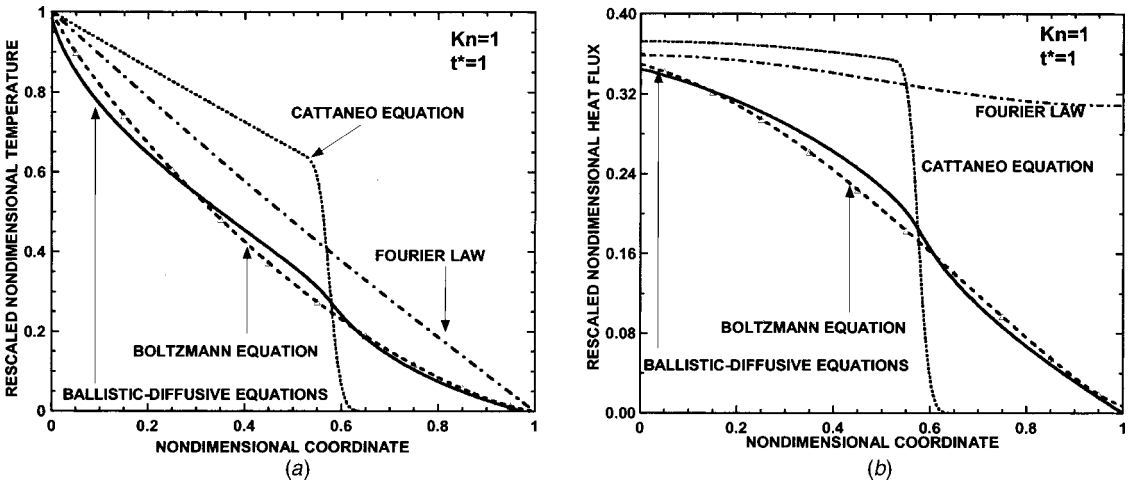


Fig. 6 Normalized temperature and heat flux distributions rescaled to the difference of the instantaneous medium temperatures at the two boundaries

diffusive equations could be a useful alternative to the Boltzmann equation for complicated geometrical configurations. We believe that the same approach may be extended for rarefied gas flow in microstructures and for electron transport in submicron devices.

Conclusion

This work establishes the transient ballistic-diffusive heat conduction equations that are applicable to transient heat conduction in nanostructures. The equations are derived from the Boltzmann equation under the relaxation time approximation. Heat carriers originating from boundaries travel ballistically, experiencing outscattering only. Heat flux and equivalent temperature expressions for this ballistic component can be obtained analytically. The transport of the scattered and emitted heat carriers is treated as diffusive component with the diffusion approximation. The resultant set of equations, named as the ballistic-diffusive equations, is much simpler to solve than the Boltzmann equations. Computational results of these equations applied to the transient heat conduction problem across a thin film suggest that they are a much better alternative to the Fourier and the Cattaneo equations at scales when the mean free path is comparable to the system size and when the time is comparable to the carrier relaxation time. Particular emphasis is given to the consistency of the definition of temperature, which will be a critical point in applying the equations to complicated structures.

Acknowledgments

This work is supported by the department of energy (DE-FG03-00ER45854) and the DARPA HERETIC project (through CFDRC).

Nomenclature

C = volumetric specific heat, $JK^{-1}m^{-3}$
D = density of states per unit volume, m^{-3}
E_n = exponential integral function of order n
f = carrier distribution function
$\hbar$ = Planck constant divided by 2π, Js^{-1}
I = intensity, $Wm^{-2}sr^{-1}$
J_o = first term in the spherical harmonics expansion, $Wm^{-2}sr^{-1}$
$\mathbf{J}_1$ = second term in the spherical harmonics expansion, Wm^{-2}
k = thermal conductivity, $Wm^{-1}K^{-1}$
Kn = carrier Knuden number ($=\Lambda/L$)
L = film thickness, m
$\mathbf{n}$ = direction normal to boundary
q = heat flux, Wm^{-2}
$\dot{q}_e$ = volumetric heat generation rate, Wm^{-3}
s = distance along direction of propagation, m
t = time, s
T = temperature, K
u = internal energy, Jm^{-3}
v = group velocity, ms^{-1}
x = coordinate
δ = nondimensional quantity
η = nondimensional coordinate
θ = polar angle, rad
Λ = mean free path, m
μ = directional cosine
μ_t = $x/(vt)$
τ = relaxation time, s
ω = angular frequency, Hz
Ω = solid angle, sr
$\hat{\Omega}$ = unit vector along solid angle direction

Subscripts

1 = first interface, first order term
b = ballistic
BE = Boltzmann equation based quantities
m = diffusive
o = equilibrium quantities, boundary points
r = coordinate
ω = frequency dependent quantities

Superscript

* = nondimensional

References

[1] Majumdar, A., 1993, "Microscale Heat Conduction in Dielectric Thin Films," Journal of Heat Transfer, , **115**, pp. 7–16.
[2] Goodson, K. E., and Ju, Y. S., 1999, "Heat Conduction in Novel Electronic Films," Annu. Rev. Mater. Sci., **29**, pp. 261–293.
[3] Chen, G., 2001, "Phonon Heat Conduction in Superlattices and Nanostructures," Semiconductors and Semimetals, **71**, pp. 203–259.
[4] Chen, G., 2001, "Ballistic-Diffusive Heat Conduction Equations," Phys. Rev. Lett., **86**, pp. 2297–2230.
[5] Olfe, D. B., 1967, "A Modification of the Differential Approximation for Radiative Transfer," AIAA J., **5**, pp. 638–643.
[6] Modest, M. F., 1993, *Radiative Heat Transfer*, McGraw-Hill, New York.
[7] Pomraning, G. C., 1973, *The Equation of Radiation Hydrodynamics*, Pergamon, New York.
[8] Yang, R. G., Chen, G., and Taur, Y., 2002, "Ballistic-Diffusive Equations for Multidimensional Nanoscale Heat Conduction," to be presented at 12 Int. Heat Transfer Conf., Grenoble, France, Aug. 18–23, 2002.
[9] Joseph, D. D., and Preziosi, L., 1990, "Heat Waves," Rev. Mod. Phys., **62**, pp. 375–391.

[10] Wu, C. Y., 1987, ''Successive Improvement of the Modified Differential Approximation in Radiative Heat Transfer,'' J. Thermophys. Heat Transfer, **1**, pp. 296–300.
[11] Modest, M. F., 1989, ''Modified Differential Approximation for Radiative Transfer in General Three-Dimensional Media,'' J. Thermophys. Heat Transfer, **3**, pp. 283–288.
[12] Joshi, A. A., and Majumdar, A., 1993, ''Transient Ballistic and Diffusive Phonon Heat Transport in Thin Films,'' J. Appl. Phys., **74**, pp. 31–39.
[13] Swartz, E. T., and Pohl, R. O., 1989, ''Thermal Boundary Resistance,'' Rev. Mod. Phys., **61**, pp. 605–668.
[14] Klitsner, T., VanCleve, J. E., Fischer, H. E., and Pohl, R. O., 1988, ''Phonon Radiative Heat Transfer and Surface Scattering,'' Phys. Rev. B, **38**, pp. 7576–7594.
[15] Chen, G., 1996, ''Nonlocal and Nonequilibrium Heat Conduction in the Vicinity of Nanoparticles,'' J. Heat Transf., , **118**, pp. 539–545.
[16] Chen, G., 1998, ''Thermal Conductivity and Ballistic Phonon Transport in Cross-Plane Direction of Superlattices,'' Phys. Rev. B, **57**, pp. 14958–14973.
[17] Little, W. A., 1959, ''The Transport of Heat Between Dissimilar Solids at Low Temperatures,'' Can. J. Phys., **37**, pp. 334–349.
[18] Chen, G., and Zeng, T., 2001, ''Nonequilibrium Phonon and Electron Transport in Heterostructures and Superlattices,'' Microscale Thermophys. Eng., **5**, pp. 71–88.
[19] Simons, S., 1974, ''On the Thermal Boundary Resistance between Insulators,'' J. Phys. C, **7**, pp. 4048–4052.
[20] Katerberg, J. A., Reynolds, C. L., Jr., and Anderson, A. C., 1977, ''Calculations of Thermal Boundary Resistance,'' Phys. Rev. B, **16**, pp. 673–679.

Thermal Transport Mechanisms at Nanoscale Point Contacts

Li Shi

Arunava Majumdar
e-mail: majumdar@me.berkeley.edu

Department of Mechanical Engineering,
University of California,
Berkeley, CA 94720

We have experimentally investigated the heat transfer mechanisms at a 90±10 nm diameter point contact between a sample and a probe tip of a scanning thermal microscope (SThM). For large heated regions on the sample, air conduction is the dominant tip-sample heat transfer mechanism. For micro/nano devices with a submicron localized heated region, the air conduction contribution decreases, whereas conduction through the solid-solid contact and a liquid meniscus bridging the tip-sample junction become important, resulting in the sub-100 nm spatial resolution found in the SThM images. Using a one dimensional heat transfer model, we extracted from experimental data a liquid film thermal conductance of 6.7±1.5 nW/K. Solid-solid conduction increased linearly as contact force increased, with a contact conductance of 0.76±0.38 W/m^2-K-Pa, and saturated for contact forces larger than 38±11 nN. This is most likely due to the elastic-plastic contact between the sample and an asperity at the tip end.
[DOI: 10.1115/1.1447939]

Keywords: Contact Resistance, Heat Transfer, Microscale, Nanoscale, Probes

Introduction

The continuous scaling of semiconductor devices has produced devices with submicron feature sizes. Localized Joule heating in these devices is becoming a serious issue affecting device reliability. For example, current crowding and localized heating in deep submicron vias are known to strongly impact reliability of interconnects in very large-scale integrated (VLSI) circuits [1]. Experimental techniques for investigating thermal transport in semiconductor devices are needed for improving device design and for understanding device physics. Such techniques are also desirable for studying new thermophysical phenomena in nanoscale devices made of novel nanostructures, such as carbon electronics [2].

Scanning thermal microscopy (SThM) [3] has been developed to meet the need for thermally imaging devices and nanostructures. While the spatial resolution of other thermometry techniques based on far-field optics [4,5] are diffraction limited to the order of several microns, spatial resolution of 50 nm has been demonstrated for SThM [6].

A SThM operates by raster scanning a sharp temperature-sensing tip on a solid surface. The temperature-sensing tip is usually mounted on a micro cantilever of an atomic force microscope (AFM) probe so that tip-sample constant contact force is maintained by the force feedback loop of the AFM. While the tip scans on a sample, tip-sample heat transfer changes the tip temperature, which is measured and used to calculate the temperature or thermal properties of the sample at the tip-sample contact.

SThM has been used to locate ''hot spots'' in electronic devices and to image contrast in thermal properties of composite thin film materials [3]. In both cases, qualitative rather than quantitative results have been obtained. To accurately interpret a temperature map or a thermal property image obtained by SThM, a thorough understanding of heat transfer mechanisms at the tip-sample contact is required. The knowledge of heat transfer at micro or nanoscale contacts is, however, still limited [3,7,8]. Figure 1 shows the schematic diagram of a SThM tip in contact with a sample surface. Also shown are the various heat transfer mechanisms. The temperature sensor is a thin-film metal thermocouple junction fabricated at the end of the tip. The various tip-sample heat transfer mechanisms include solid-solid conduction at the contact, conduction through the air gap between the probe and the sample, radiation, and heat conduction through a liquid meniscus formed at the tip-sample junction. The liquid meniscus is formed from water molecules and/or contaminations adsorbed on the sample and tip surfaces. Luo et al. [7] speculated that conduction through the liquid meniscus could be the dominant heat transfer mechanism. This was used to explain the sub-100 nm spatial resolution obtained in the thermal image. Using a heated Pt wire probe to image thermal property contrast on a sample, Gomes et al. [8] recognized the importance of liquid-film conduction while suggesting that the predominant heat transfer mechanism depends on the sample thermal conductivity.

Recently, we have thermally designed and batch-fabricated SThM probes. Using these probes, we have thermally investigated localized heating in VLSI interconnects [9] and Joule heating in current-carrying carbon nanotubes [6] with a spatial resolution as small as 50 nm. Despite achieving sub-100 nm spatial resolution, there still remain questions regarding what is the dominant heat transfer mechanism in this nano-thermometry and what is the thermal contact resistance between the tip and the micro/nano structures. It is important to address these questions for quantitatively interpreting the thermal imaging results and for further improving the technique. In the following sections, we will (i) present several experiments for characterizing the heat transfer mechanisms at the nanoscale tip-sample point contact; (ii) describe a model considering various microscale heat transfer mechanisms for extracting point contact thermal resistance from the experimental data; and (iii) discuss the relative contribution of various conduction mechanisms under different conditions.

Experiments

Figure 2 shows the cross section and scanning electron micrographs of one of the thermal probes that we have thermally-designed and batch-fabricated [10]. The silicon nitride (SiN_x) cantilever was 128±1 μm long, 18.4±0.5 μm wide, and 0.89 ±0.01 μm thick. The tip was made of silicon dioxide (SiO_2), and was 8±0.5 μm high and the cone angle of the tip was 36 ±4 deg. Platinum (Pt) and chromium (Cr) films were sputtered and patterned on the cantilever and the tip. The thickness of each metal was 75±5 nm on the cantilever, as measured by AFM. Because of the steep slope of the tip, the metals deposited on the tip were much thinner, in the range of 20–25 nm for each metal as imaged by scanning electron microscopy (SEM). The two metals

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division February 8, 2001; revision received July 27, 2001. Associate Editor: D. Poulikakos.

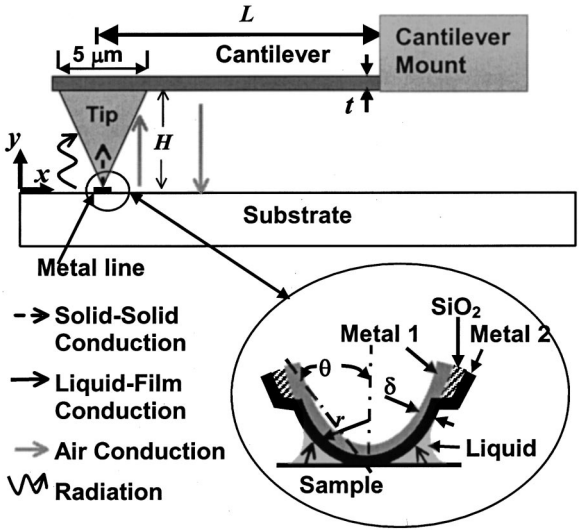


Fig. 1 Schematic diagram of a SThM probe in contact with a Joule heated metal line. Also shown are various tip-sample heat transfer mechanisms.

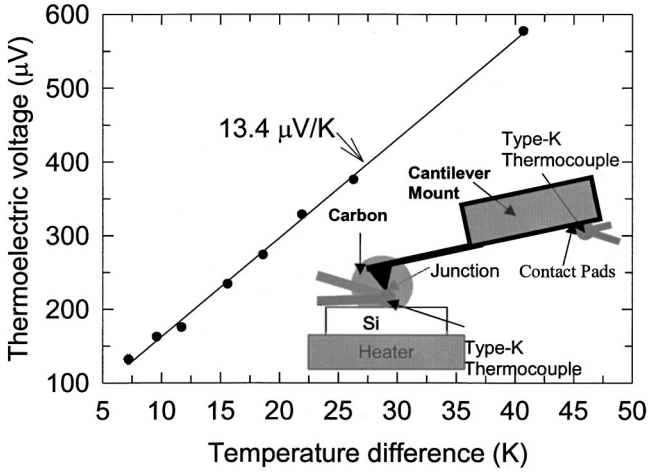


Fig. 3 Thermoelectric voltage as a function of temperature difference between the Pt-Cr junction and the contact pads. The inset shows the measurement setup

were separated by a 270 ± 10 nm thick SiO_2 film except at the very end of the tip. The junction at the tip end was 900 ± 60 nm high and 600 ± 30 nm wide at the base. The tip diameter was 90 ± 10 nm as measured by SEM. It should be noted that the cantilever design was not the optimized thermal design. However, because this probe has simpler geometry than other designs with cantilevers of *V* or other shapes, it is easier to model the heat transfer processes involved. Thus, this probe has been chosen for the following experiments.

We first measured the thermopower of the Pt-Cr junction using the experimental setup shown in the inset of Fig. 3. Under an optical microscope, a tiny drop of carbon paint was used to fix a probe tip to a fine Omega® type-K thermocouple mounted on a silicon substrate by silver paste. The tip was assumed to be at thermal equilibrium with the type-K thermocouple. Another type-K thermocouple was attached in close proximity to the Pt and Cr contact pads located at the end of the Pt and Cr lines. While the silicon substrate was heated to different temperatures, a thermoelectric voltage was created between the Pt and Cr contact pads due to the temperature difference between the Pt-Cr junction

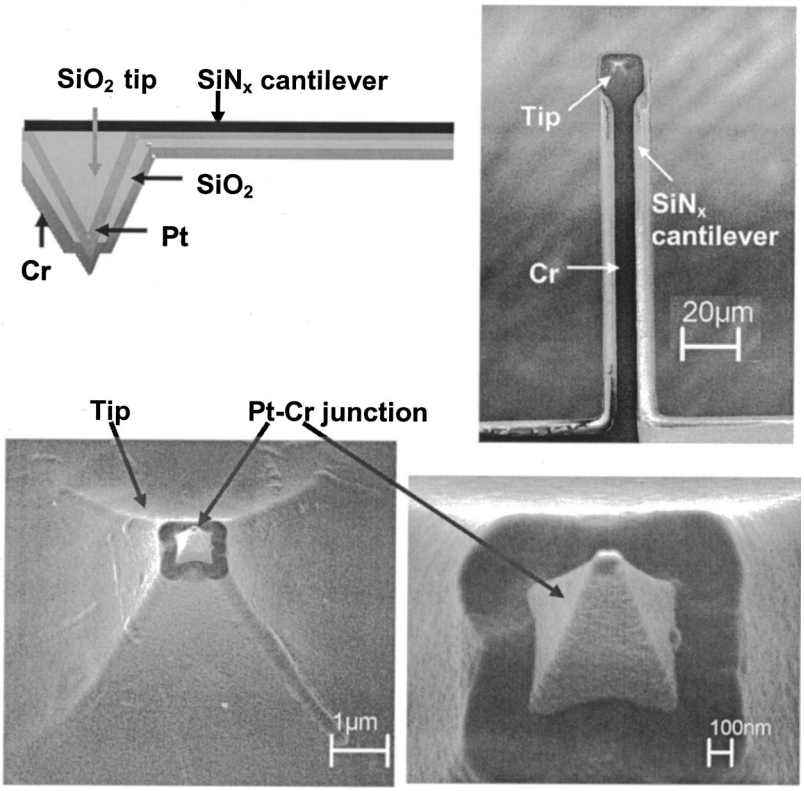


Fig. 2 Cross section (upper left) and scanning electron micrographs of a SThM probe (upper right), the probe tip (lower left), and the Pt-Cr junction (lower right) at the end of the tip

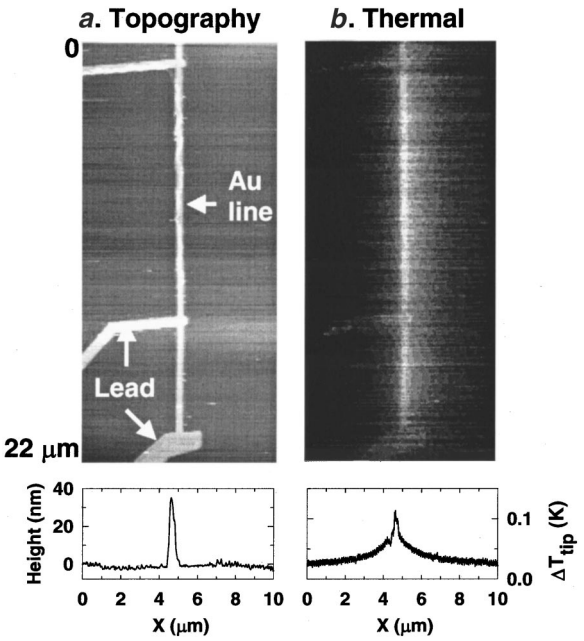


Fig. 4 Topographic (*a*) and thermal (*b*) images of a 350 nm wide Au line. One of the four leads to the Au line is located on top of the line and is not shown in the figure.

and the contact pads. This voltage was recorded as a function of the temperature difference between the two type-K thermocouples. From the same wafer containing the probe shown in Fig. 2, we chose a few probes for this destructive measurement. One of the measurement results is shown in Fig. 3. The thermopower of the Pt-Cr junction was found to be 13.4 ± 0.06 μV/K, which is about half of that of a pure bulk Pt-Cr junction. Due to the thin-film confinement effect and differences in structure and purity, the thin metal films are expected to have different Seebeck coefficients as well as electrical and thermal conductivities from those of corresponding pure bulk metals.

The samples for studying tip-sample heat transfer were thin film metal lines with different line widths and lengths fabricated on silicon wafers containing a 1 μm thick SiO_2 film. The topographic image of a Au line obtained by the SThM probe is shown on Fig. 4(*a*). This particular sample was patterned using electron beam lithography and metal lift-off technique. The line was 350 ± 50 nm wide and 20 μm long. When a current passed between the two leads at the two ends of the line, the voltage drop at the middle segment was measured using the two middle leads that was 13 μm apart and was used to calculate the four-probe resistance. With a current of 1 μA, the resistance was measured while the sample was heated to different temperatures using a hot plate. From the resistance versus temperature curve shown in Fig. 5, we calculated the temperature coefficient of resistance (TCR) of the Au line to be $(152\pm0.6)\times10^{-6}$ K^{-1}. When the substrate was at room temperature and the line was Joule heated by a current in the range of 100–1500 μA, the temperature rise ΔT of the metal line can be determined as $\Delta T=(R-R_0)/(\mathrm{TCR}^*R_0)$, where R and R_0 are the four-probe electrical resistances when the line was Joule heated and at room temperature, respectively. The measured temperature rise, ΔT, is plotted as a function of current and is shown in Fig. 6.

When a current of 781.8 ± 0.1 μA passed the Au line, a thermal image of the Joule heated line was obtained by the thermal probe, as shown in Fig. 4(*b*). During thermal imaging and all of the following experiments, the cantilever was parallel to the substrate and oriented perpendicular to the metal line. The thermal image shows that the temperature was uniform along the line. With the probe in contact with the heated metal line, we measured the

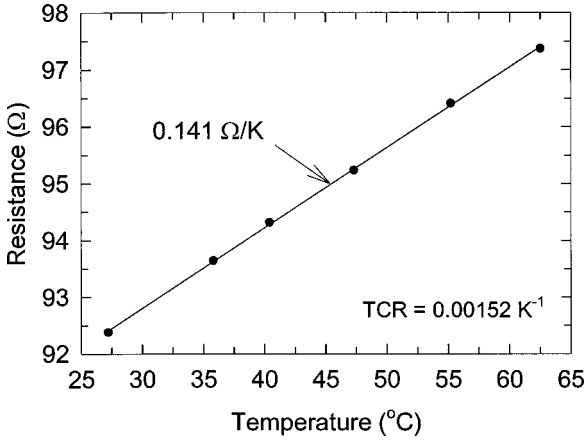


Fig. 5 Four-probe resistance of the Au line as a function of temperature

temperature response of the probe, which is defined as the ratio of the temperature rise at the Pt-Cr junction to that in the sample and has a unit of K/K. For the 350 nm wide line, the measured temperature response was only 0.05 ± 0.01 K/K, as shown in Fig. 7. At first sight, this result surprised us because the temperature response of the probes were usually about 0.5 K/K when Al and Au metal lines with larger (3–50 μm) line widths were used to calibrate other probe designs [10], as shown in Table 1. To confirm that the low temperature response in Fig. 7 was not due to the specific design of the probe, we used the 350 nm wide line to

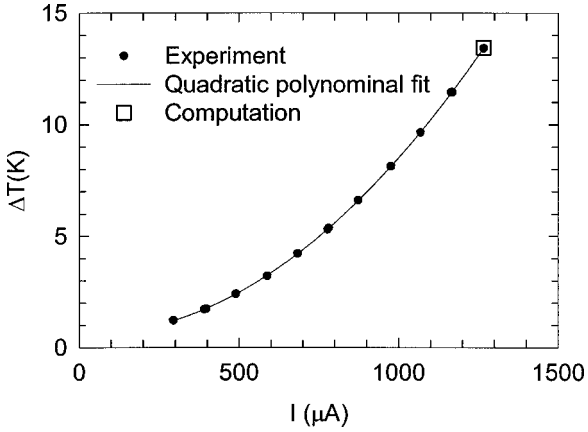


Fig. 6 Temperature of the Au line as a function of current

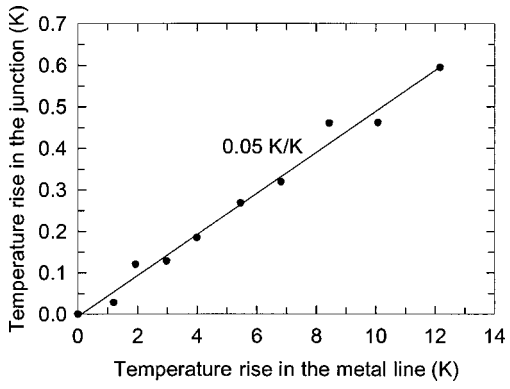


Fig. 7 Temperature rise in the Pt-Cr junction as a function of temperature rise in the 350 nm wide line in contact with the probe tip

Table 1 Temperature response of thermal probes with different dimensions on different samples. The cantilever is a V-shape. *L* is the cantilever length. *W* is the cantilever width and is expressed as two times the width of each of the two SiN_x arms. Cantilever thickness $t=0.89\pm0.01$ μm, tip height $H=8\pm0.5$ μm, Pt-Cr junction, junction height $h=900\pm60$ nm, thickness of each metal $=75\pm5$ nm, metal line width $w=5$ μm. The calibration samples were 2000 μm long Joule-heated thin film Al lines with different line widths.

Sample Line Width / Probe Dimension	3 μm	5 μm	50 μm
L = 100 μm, W = 2 x 8 μm	-	0.50±0.02	0.56±0.02
L = 100 μm, W = 2 x18 μm	-	0.50±0.02	-
L = 200 μm, W = 2 x 8 μm	0.46±0.02	-	0.53±0.02

measure the temperature responses of those probes used in Table 1 and obtained similar low temperature responses. The probe used for Fig. 7 was also calibrated with a 5.8 ± 0.2 μm wide, 45 ±5 nm thick, and 1500 μm long Joule heated Au line patterned on 1 μm thick oxide of a silicon wafer. As shown in Fig. 8, the measured temperature response was 0.64 ± 0.02 K/K, which was more than an order of magnitude larger than that for the 350 nm wide line.

Table 1 shows that the temperature response was larger for the 50 μm wide line than for the 3 μm wide line. This fact suggested that the probes was heated more by air conduction between the tip and the larger hot area for the 50 μm wide line than for the 3 μm one. Here, the much lower temperature response for the 350 nm wide Au line led us to suspect that air conduction might be responsible for the large temperature response obtained for the 5.8 μm wide lines.

To determine the relative contribution of various tip-sample heat transfer mechanisms, we Joule heated the 350 nm wide line to 5.3 ± 0.1 K above room temperature and recorded the cantilever deflection and probe temperature response simultaneously when the sample was raised toward and then retracted from the thermal probe. When the sample approached the tip, the cantilever deflection signal remained nearly zero before the sample contacted the tip, as shown in the deflection curve in Fig. 9. In this region, the temperature response was mostly due to air conduction between the probe and the sample, because radiation contribution is negligible when both the sample and the tip are close to room temperature [3]. As the tip-sample distance reduced, the temperature response due to air conduction increased slowly. Before the sample made solid-solid contact to the tip, the adsorbed liquid layers on the tip and the sample bridged each other. Initially, this liquid bridge pulled the tip down by van der Waals force, as being seen in the dip labeled as "jump to contact" in the deflection curve. Coincidentally, there was a small jump in the temperature response curve due to conduction through the liquid bridge. As the sample was raised further, both the solid-solid contact force and the temperature response increased gradually, until the cantilever deflection reached 100 nm higher than its previous position when the initial solid contact was just made. After this point, the temperature response remained almost constant as contact force increased.

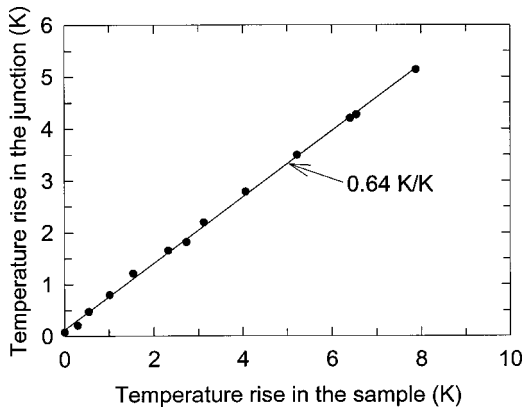


Fig. 8 Temperature rise in the Pt-Cr junction as a function of temperature rise in the 5.8 μm wide line in contact with the probe tip

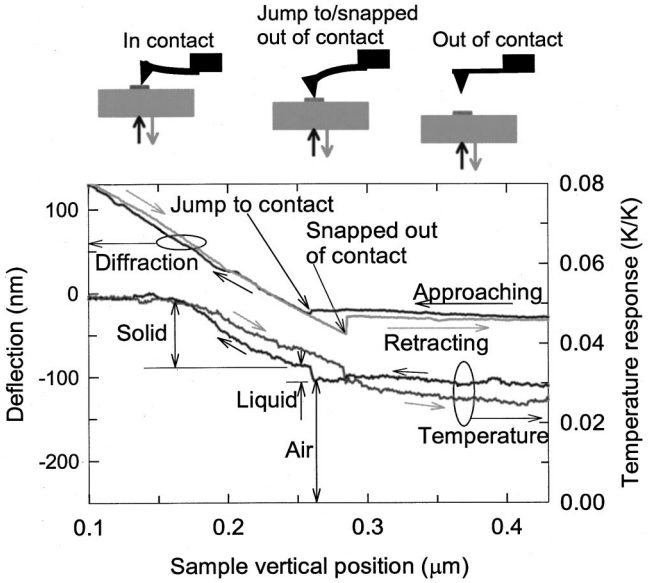


Fig. 9 Cantilever deflection and temperature response of the probe as a function of sample vertical position when the 350 nm wide line was raised toward and then retracted from the tip

As the sample was retracted from the tip, the temperature response again remained almost constant until at a cantilever deflection of 100 nm, the temperature response reduced roughly linearly but at a smaller slope than that found in the approaching cycle. As the sample was lowered further, the tip was pulled down together with the sample by surface tension of the liquid bridge until after a certain point, the restoring spring force of the cantilever exceeded the surface tension and the tip "snapped out of contact" with the sample. Associated with the breaking of the liquid bridge, there was a small drop in temperature response.

The above experiment shows several mechanisms. First, before tip-sample contact, air conduction contributed to a temperature response up to 0.03 K/K, which was about 60 percent of the maximum temperature response of 0.05 K/K at large contact force. Second, conduction through a liquid meniscus was responsible to the sudden jump and drop in temperature response when the tip "jumped to contact" to and "snapped out of contact" from the sample, respectively. Third, solid-solid conduction resulted in the almost linear increase or decrease of temperature response with contact force, which is a well understood feature for macroscopic solid-solid contacts [11]. Since the temperature response decreased at a slower slope during unloading (decreasing contact force), there must have been plastic deformation during loading (increasing contact force). For plastic deformation, contact area increases with load [12], resulting in the linear increase of solid-solid contact conductance with contact force. However, since the conductance was still a function of load during unloading, elastic recovery (spring back) must have been significant.

Besides the above-discussed hysteresis of temperature response

Table 2 Width, thickness, and Young's modulus of the four materials constituting the composite cantilever

Material	Width (μm)	Thickness (μm)	Young's modulus (GPa)
SiN_x	18.4 ± 0.5	0.89 ± 0.01	95 ± 10 (ref 14)
Pt	10 ± 0.5	0.075 ± 0.005	170 (ref 15)
SiO_2	18.4 ± 0.5	0.28 ± 0.2	57 ± 11 (ref 16)
Cr	7.7 ± 0.5	0.075 ± 0.005	140 (ref 15)

occurred during tip-sample contact, there was another hysteresis in the temperature response curve after the tip was released from the sample. As shown in Fig. 9, after the tip was snapped out of contact and when the substrate was retracted away from the tip, the temperature response was slightly different from those when the tip approached the sample at the same tip-sample distance. This indicates change in cantilever-sample heat transfer between approaching and retraction when tip and sample were out of contact. One possible scenario is that resonance vibration in the cantilever was excited after the tip was snapped out of contact with the sample. The vibration gave rise to convection and also change in air conduction between the cantilever and the sample, resulting in the hysteresis during approaching and retracting when tip was out of contact with the sample.

One question still remains as to why the temperature response saturated for cantilever deflection larger than 100 nm. To clarify this question, we first calculated the contact force corresponding to 100 nm deflection. The spring constant of the composite cantilever beam was calculated following Roark's formulas [13]. For a composite beam consisting of n different materials, the equivalent moment of inertia

$$I_{eq}=\sum_{i=1}^{n}\left(\frac{w_i^* t_i^3}{12}+w_i^* t_i^*(y_i-\bar{y})^2\right);\quad w_i^*=\frac{E_i}{E_1}w_i,$$

$$y_i=\frac{t_i}{2}+\sum_{j=i+1}^{n} t_k, \tag{1}$$

where w, t, and E are the width, thickness, and Young's Modulus. The centroidal axis

$$\bar{y}=\frac{\sum_{i=1}^{n} w_i^* t_i y_i}{\sum_{i=1}^{n} w_i^* t_i}. \tag{2}$$

The spring constant

$$K=\frac{E_1 I_{eq}}{L^3}, \tag{3}$$

where L is the cantilever length. Using the parameters listed in Table 2, we calculated $K=0.38\pm0.11$ N/m. Therefore, the 100 nm deflection corresponds to a contact force $F=38\pm11$ nN.

Assuming plastic deformation, this contact force resulted in a contact spot with a diameter

$$d_c=\sqrt{\frac{4F}{\pi H}}, \tag{4}$$

where H is the hardness of the tip or sample, whichever is softer. Among the tip and sample materials, the hardness of Au is the lowest [17]. For thin gold films, two measurement results for H were reported to be 500 MPa [18] and 1.5 GPa [19], respectively. For H on the order of 1 GPa, $d_c\approx8$ nm. As shown in Fig. 2, the tip diameter was about 90 nm. However, the details on the tip end

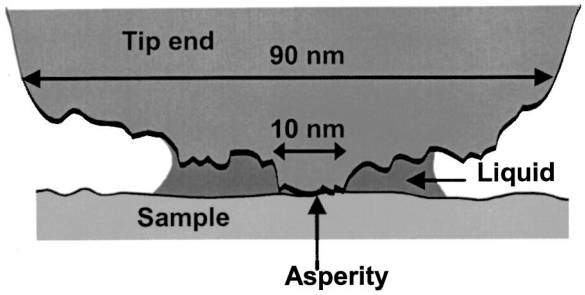


Fig. 10 A schematic diagram showing the contact between an asperity on the tip end and the sample

are not clearly shown in the SEM image due to the lack of resolution. We suspect that there were asperities of about 10 nm diameter on the tip end, as shown in Fig. 10. One of the asperities protruded out and first made contact with the sample. The linear increase in contact area is probably due to the roughness on the surfaces of the asperity and the sample, since it is well known that in the junction of random rough surfaces, the contact area increases linearly with contact force [20]. As the contact force increased to about 38 nN, the contact size approached the diameter of the asperity. At this point, the contact area could not increased further with contact force, until the asperity could be completely pressed into the soft sample by a contact force much larger than those used in the experiment. As a result, the temperature response in Fig. 9 remained almost constant for deflection larger than 100 nm. The feature, i.e., saturation of contact conductance at large contact forces, was not unique only to this particular thermal probe, but was also found for several other probes. Some of these probes had been used extensively for imaging before this force-calibration experiment, and it was possible that some of the asperities at the tip end have been worn out. For this case, another asperity could protrude out at the tip end and was responsible for the conductance saturation.

Now we will investigate why the temperature responses for the 5.8 μm wide heated line and the 350 nm one were so different. We repeated the point contact experiment for the 5.8 μm wide line and the result is plotted in Fig. 11. We can see that while the increase of temperature response due to solid and liquid conduction was similar in magnitude to that in Fig. 9, the temperature response due to air conduction, i.e., about 0.6 K/K, is one order of magnitude higher than the corresponding one (0.03 K/K) in Fig. 9.

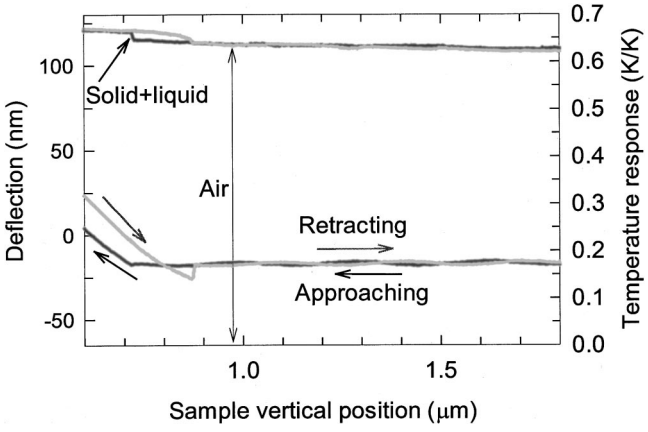


Fig. 11 Cantilever deflection and temperature response of the probe as a function of sample vertical position when the 5.8 μm wide line approached and then retracted from the tip

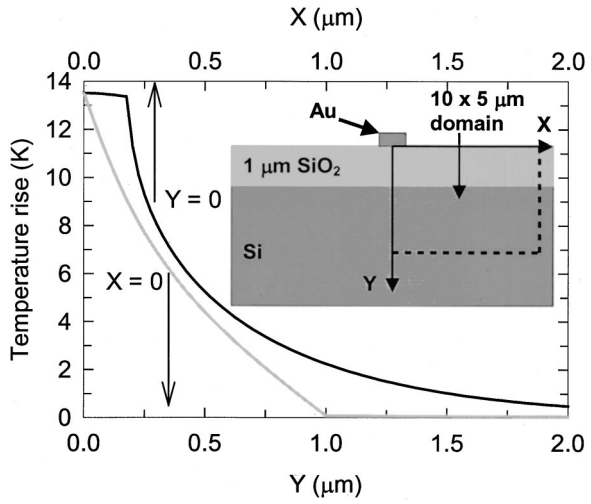


Fig. 12 Modeled temperature profile along the *X* and *Y* axis on the substrate of the 350 nm wide line. The inset shows the computation domain.

Therefore, it is clear that for the 5.8 μm wide line, air conduction dominated tip-sample conduction and is responsible for the large temperature response.

The above experiment reveals that the contribution of air conduction in tip-sample heat transfer depends on the size of the heated region on the sample. For larger heated regions, air conduction may dominate tip-sample heat transfer. As the characteristic size of the heat source reduces, the contribution of air conduction decreases and solid and liquid conduction become important. For micro/nano devices with localized submicron heated features, such as carbon nanotube circuits, air conduction contribution may reduce to a level smaller than that from solid-solid and liquid film conduction [6]. In the thermal images of these devices [6,9], the sub-100 nm resolution was a result of the increased contribution of local solid and liquid conduction; whereas, air contribution was insensitive to the distance between the tip and the heat source, giving rise to a slowly varying background signal.

Modeling

The purpose of the model is (i) to evaluate the thermal conductance of the liquid meniscus, and that of the solid-solid contact; and (ii) to examine the possible deviation of SThM-measured temperature profiles from the true ones due to the influence of air conduction. To do this, it is necessary to calculate the temperature response of the probe for different tip-sample distance. In the calculation, it was assumed that the temperature was constant at each horizontal cross sections of the tip and only varied along the height or y direction, as shown in Fig. 1. This assumption can be justified because the external thermal resistance through the air is much larger than the internal thermal resistance in the tip across each horizontal cross section. Although Fig. 1 shows that the tip is in contact with the sample, we consider a general case that the tip end is at a distance d above the sample. Tip-sample contact corresponds to $d \leq 0$. In addition, y is measured from the end of the tip instead of from the sample surface.

The sample was the 350 nm wide Au line Joule heated to a temperature T_s. The resulting temperature profile on the SiO_2 substrate surface was calculated using a finite difference method. It was assumed that the temperature does not change along the length of the metal line, yielding a two-dimensional problem. This assumption can be justified by the SThM image in Fig. 4(b) and is expected because the length of the line (20 μm) was much larger than the thickness (1 μm) of the SiO_2 film. We considered a computation domain of $0 \leq X \leq 10$ μm and $0 \leq Y \leq 5$ μm, as shown in the inset of Fig. 12. The temperature rise sufficiently far away at $X=10$ μm and $Y=5$ μm is assumed to be zero. We ignored heat dissipation through the air and assumed an adiabatic boundary condition on the sample surface. The symmetric geometry results in an adiabatic boundary condition at $X=0$. For a grid size of 25 nm by 25 nm, we solved the temperature distribution in the computation domain for an electrical current of 1267 μA flowing in the line. We confirmed that the modeled temperature profile did not change when the grid size was increased by a factor of 2. The temperature profile on the top surface, i.e., $Y=0$, and that of $X=0$ are plotted in Fig. 12. The average temperature rise of the Au line was calculated to be 13.45 K, which agrees well with the measurement result of 13.42 K obtained from Fig. 6. The temperature decays to almost room temperature at the SiO_2 and Si interface, which is expected because of the high thermal conductivity of Si compared to that of SiO_2 (see Table 3). The temperature also approaches room temperature for $X>2$ μm. The modeling results, hence, justify the assumption that temperature rise is zero at $X=10$ μm or $Y=5$ μm.

Table 3 Thermophysical properties of candidate probe materials

	Thermal Conductivity at 300 K (W/m-K) (ref. 21)
Cr	93.7
Pt	71.6
Si	148
SiO_2	1.4
SiN_x	5.5 (ref. 22)

With the known temperature profile on the substrate, the one-dimensional heat conduction equation in the tip can be written as

$$\frac{d}{dy}\left[(A_t(y)k_t(y)+A_m k_m)\frac{dT(y)}{dy}\right]-p(y)h_a(y)\tan\theta(T(y)-T_{\text{sub}}(y))=0. \quad (5)$$

Here, θ and r are the half angle and the radius of the conical tip, respectively, as shown in Fig. 1, A_t and k_t are the cross section area and thermal conductivity of the SiO_2 tip, respectively, k_m and A_m are the cross section area and thermal conductivity of the metal coating on the tip, respectively, and p is the perimeter of the square cross section of the tip. For each point on the perimeter, we assumed that heat was conducted by air between this point and a point right below on the substrate, and treated these two points as two parallel plates. This simplified picture of tip-sample air conduction is represented by the second term in Eq. (5), where T_{sub} was the temperature of the point on the substrate. The distance between the two points on the tip and substrate, respectively, is $y+d$. The air conduction coefficient h_a needs to be written in different forms for different values of $(y+d)/\lambda$, where λ is the mean free path of air molecules and is about 60 nm under ambient condition and at sea level [23]. For $(y+d)/\lambda > 100$, we assumed a constant temperature gradient at the air gap and used $h_a = \alpha k_a/(y+d)$, where k_a is the thermal conductivity of bulk air and α is a geometry factor to accommodate the fact that the tip and the substrate is not exactly two parallel plates. We will obtain α by fitting the modeling results with measurement data. For $1 < (y+d)/\lambda < 100$, significant temperature discontinuity may develop at the air-solid boundaries because intermolecular collisions

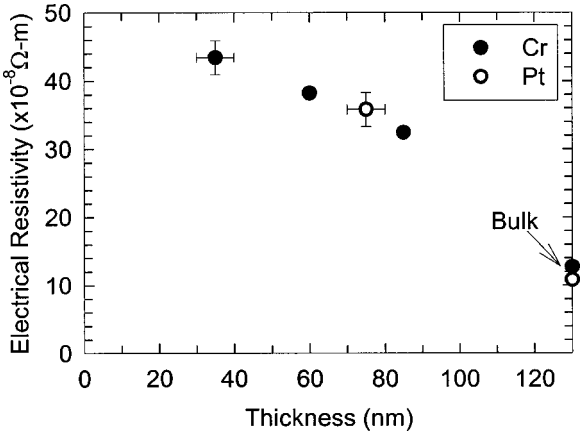


Fig. 13 Electrical resistivity as a function of the thickness of the Pt and Cr films

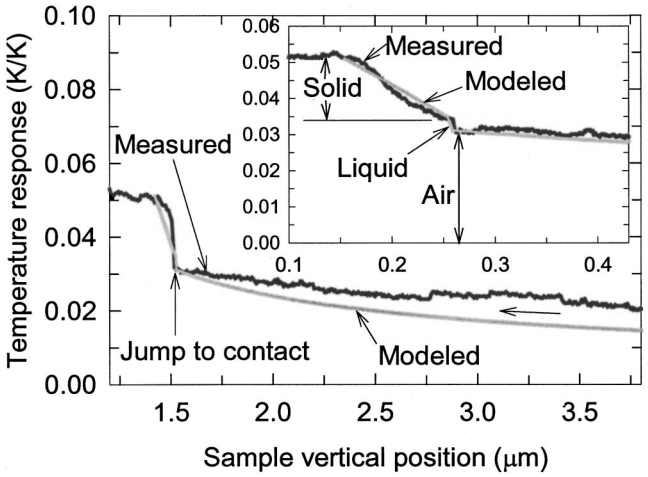


Fig. 14 Modeled and measured temperature responses of the probe as a function of sample vertical position when the 350 nm wide line approached the tip

become less frequent and molecules arriving at the solid surfaces are unable to come into equilibrium with the surface [23]. In this so-called slip regime,

$$h_a=\frac{\alpha k_a/(y+d)}{1+2\,f\lambda/(y+d)};\quad f=\frac{2(2-A)\gamma}{A(\gamma+1)\mathrm{Pr}},\tag{6}$$

where A is a thermal accommodation coefficient and is about 0.9 for air, γ is the ratio of air heat capacity, and Pr is the Prandtl number. For $(y+d)/\lambda<1$,

$$h_a=\frac{\alpha k_a'}{(y+d)(1+2\,f)};\quad k_a'=CV(y+d)/3,\tag{7}$$

where k_a' is the thermal conductivity of air in the free molecule flow regime, and C and V is the heat capacity and velocity of air molecules, respectively.

The boundary conditions are

$$(A_t k_t+A_m k_m)\frac{dT}{dy}=\frac{T-T_s}{R_{ts}},\quad \text{at}\quad y=0\tag{8a}$$

$$(A_t k_t+A_m k_m)\frac{dT}{dy}=\frac{T_0-T}{R_c},\quad \text{at}\quad y=H,\tag{8b}$$

where R_{ts} is the tip-sample thermal resistance. When the tip is not in contact with the sample, $R_{ts}=[h_a(A_t+A_m)]^{-1}|_{y=0}$, as a result of air conduction between the tip end and the sample. When the tip was in contact with the sample, or $d\leqslant 0$, R_{ts} is due to solid-solid and liquid film conduction and will be obtained by fitting the modeling result with measurement data. T_0 is the ambient temperature.

When the cantilever was oriented perpendicular to the 350 nm wide metal line with the tip above the center of the line, the substrate right below the cantilever arm was more than 2.5 μm away from the metal line, because the base of the tip was 5 μm wide, as shown in Fig. 1. For the 350 nm wide line, the substrate temperature approaches room temperature for $X>2\,\mu$m, as shown in Fig. 12. Accounting for air conduction between the cantilever and the room temperature substrate below, the thermal resistance of the cantilever, R_c is solved using fin theory [24]

$$R_c=\frac{\tanh(mL)}{mk_c wt};\quad m=\sqrt{\frac{h_a}{k_c t}},\tag{9}$$

where L, w, t, and k_c are the length, width, thickness, and thermal conductivity of the cantilever, respectively, h_a is the heat conduction coefficient of the air gap between the cantilever and the substrate and was calculated following the discussion in the preceding paragraph with $\alpha=1$ because the cantilever and the substrate can be treated as two parallel plates.

The thermophysical properties of candidate probe materials are tabulated in Table 3, where the thermal conductivity data of the metals are for pure bulk metals. To determine the thermal conductivities of the thin metal films, we measured their electrical conductivities as a function of thickness, as shown in Fig. 13. Using Wiedemann-Franz law and assuming that the reduced thermal conductivities of the thin metal films had the same proportionality with the reduced electrical conductivities [25], we estimated the thermal conductivities of the thin metal films on the cantilever. From the electrical conductivity measurements, the correction factors with respect to bulk values were 0.3 for Pt and 0.39 for Cr. For the even thinner metals on the tip, the correction factor were 0.22 for Pt and 0.28 for Cr. It is unnecessary to correct the thermal conductivity of SiO_2 at the SiO_2 tip end, because the phonon mean free path in amorphous SiO_2 is expected to be shorter than the size of the end of the SiO_2 tip [26], which was about 20 nm.

Equation (5) was solved using a finite difference method to obtain the temperature distribution in the tip, i.e., $T(y)$, for different value of tip-sample distance d. A non-dimensional temperature at a distance y away from the sample was defined as

$$\phi(y)=\frac{T(y)-T_0}{T_s-T_0}.\tag{10}$$

Due to the particular geometry of the junction at the tip end, the measured thermal signal corresponds to the temperature difference between the room temperature and that at about 900 nm away from the tip end for the probe shown in Fig. 2. We plotted the modeled temperature response $\phi(y=900\,\mathrm{nm})$ as a function of tip-sample distance in Fig. 14. Two measurement results are also shown in Fig. 14. The sample travelling range was 0.5 μm for the measurement result in the inset and 3.8 μm for the other one. For the one with 3.8 μm travelling range, the distinction between the sudden jump due to liquid conduction and the gradual increase due to solid-solid conduction is unclear due to the lack of resolution in sample position. To fit the modeling results with the experimental ones while the tip was not in contact with the sample, we used $\alpha=0.8$ to correct tip-sample air conduction for the deviation from that between two parallel plates. Considering the 0.01 K/K uncertainty in the measured temperature response, we estimated the uncertainty in α to be ± 0.1. After obtaining α, we used $R_{ts}=(1.5\pm 0.3)\times 10^8$ K/W to fit the magnitude of the temperature jump due to liquid film conduction. This resistance value corresponds to a liquid film thermal conductance $G_{lf}=6.7\pm 1.5$ nW/K. To fit the modeling result with the measured saturated value of temperature response at the critical contact force

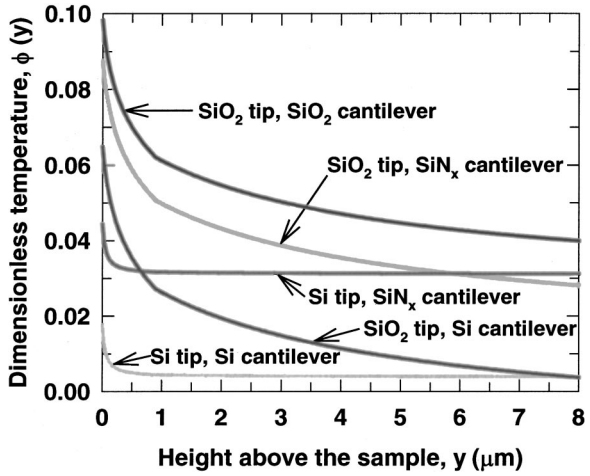


Fig. 15 Temperature in the tip as a function of *y* for different tip and cantilever materials

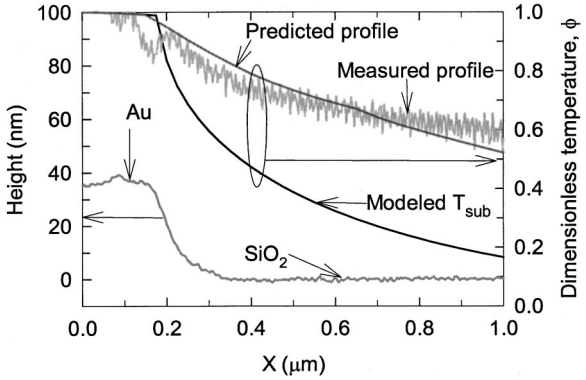


Fig. 16 Modeled temperature profile, predicted SThM-measured temperature and actual SThM measured temperature and height profiles across the 350 nm wide line

$F = 38 \pm 11$ nN, we used a solid-solid contact conductance of $G_{ss} = (29 \pm 6)$ nW/K, assuming that liquid film conductance did not change for contact forces smaller than 38 ± 11 nN. Therefore, the proportionality between solid-solid contact thermal conductance and contact pressure is $G_{ss}/F = 0.76 \pm 0.38$ W/m^2-K-Pa.

The estimated liquid and solid-solid conductances are different from what were suggested by previous work [7]. In that work, the solid-solid conductance was estimated to be on the order of 10–100 nW/K, which was close to what was found in this work. However, the liquid film conductance was estimated to be on the order of 1 μW/K, two orders of magnitude larger than the current result. The estimation of liquid film conductance did not include the contact conductance at the solid-liquid interfaces, which may be much lower than the conductance through the liquid film of monolayer thickness. Therefore, the liquid film conductance could be quite low, as found in this study.

The low solid-solid and liquid-solid contact conductance results in low temperature responses for samples with a localized submicron heated region, such as the 350 nm wide line as well as defective submicron VLSI vias or current carrying-carbon nanotubes. To improve the temperature response, there is still some room to improve regarding the thermal design of the probes. In fact, the current probes made of low-thermal conductivity materials was expected to have larger temperature response than those constructed from high-thermal conductivity ones. To confirm this, we calculated the temperature response of the probe for different probe materials. For a liquid film conductance $G_{lf} = 6.7$ nW/K and a maximum solid-solid contact conductance $G_{ss} = 29$ nW/K, the temperature distribution in the tip is calculated for different combinations of tip and cantilever materials and is plotted in Fig. 15. Clearly, the temperature response of the probe is improved for using low thermal conductivity materials such as SiN_x and SiO_2. The temperature response can be further improved for smaller (100–300 nm high) junctions and narrower cantilever width, etc, as discussed in another paper [10].

Last, due to the influence of air conduction, the measured temperature profile may deviate from the true one. This may occur even when air conduction contribution decreases for the 350 nm wide line and other micro/nano devices with submicron localized heated features. Using the above model with $G_{lf} = 6.7$ nW/K, and a solid-solid conductance $G_{ss} = 14.5$ nW/K corresponding to 50 nm cantilever deflection, and $\alpha = 0.8$, we predicted that the SThM-obtained temperature profile across the 350 nm wide line decays slower than the true substrate temperature shown in Fig. 12. The predicted SThM measurement result and the substrate temperature profile from Fig. 12 are plotted in Fig. 16. For the ease of comparison, we plotted a dimensionless temperature ϕ defined as $\phi = \Delta T(X)/\Delta T(0)$, where $\Delta T(X)$ is the temperature rise of the sample surface at a distance X away from the center of the Au line. To confirm the prediction, the SThM-measured temperature and height profiles are also shown in Fig. 16. Note that the cantilever deflection was maintained at 50 nm during tip scanning. Figure 16 shows that the measurement result agrees well with the prediction. This confirms that the modeling results of G_{lf}, G_{ss}, and α are in the correct ranges. Furthermore, the results in Fig. 16 suggest that it is desirable to perform SThM at vacuum, as has been demonstrated by Nakabeppu et al., [27], in order to eliminate the unwanted influence of air conduction.

Conclusion

We have experimentally investigated the heat transfer mechanisms at the SThM tip-sample contact with a contact diameter of 90 ± 10 nm. The contribution of tip-sample air conduction depends on the size of the heated region on the sample and is not sensitive to tip-heat source distance. For large heated regions, air conduction dominates tip-sample heat transfer, resulting in large temperature responses. For micro/nano devices with submicron localized heated region, the contribution of air conduction decreases; whereas, conduction through the solid-solid contact and a liquid meniscus become important, resulting in sub-100 nm spatial resolutions achieved in SThM images. Despite the superior spatial resolution, SThM-measured temperature profiles deviate from the true one due to the influence of air conduction. Therefore, it is desirable to perform SThM in vacuum to eliminate the unwanted influence of air conduction.

To estimate the magnitude of solid-solid and liquid film conductance, we developed a one-dimensional heat conduction model considering various microscale heat transfer mechanisms. Using the model, we inferred from the experimental data that the thermal conductance of the liquid meniscus was $G_{lf} = 6.7 \pm 1.5$ nW/K. Solid-solid conduction increased linearly with the contact force with a contact conductance of 0.76 ± 0.38 W/m^2-K-Pa. For contact forces larger than 38 ± 11 nN, solid-solid conductance saturated. This is most likely because at large contact forces, the contact size between the sample and an asperity on the tip end approached the asperity diameter of about 10 nm, and could not increased further with contact force.

Acknowledgment

The authors thank DOE (Engineering Division, Basic Engineering Sciences) and NSF (Chemical and Transport Systems) for providing financial support for this work.

Nomenclature

A = thermal accommodation factor for air conduction when the length scale is close to the mean free path of air molecules
C = specific heat of air (W/kg-K)
d = contact diameter (m) or tip-sample distance (m)
G = thermal conductance (W/K)
h = air conduction coefficient (W/m^2-K)
H = height of the tip (m) or hardness (Pa)
I = current (A)
k = thermal conductivity (W/m-K)
K = spring constant (N/m)
L = length of the cantilever (m)
n = number of materials of the composite cantilever
p = perimeter of a horizontal cross section of the tip (m)
r = tip radius (m)
R = thermal or electrical resistance (K/W or Ω)
t = cantilever thickness (m)
T = temperature (K)
V = velocity of air molecules
W = cantilever width (m)
x, X, y, Y = coordinate (m)

Greek Letters

α = correction factor for air conduction between the tip and sample
δ = thickness of the metal film (m)
ΔT = temperature rise (K)
ϕ = dimensionless temperature
λ = mean free path of air molecules (m)
θ = half angle of the conical tip (°)

Subscripts

a = air
c = cantilever or contact
lf = liquid film
m = metal film
t = tip
ts = tip to sample
s = sample
ss = solid-solid
0 = ambient condition

References

[1] Kwok, T., Nguyen, T., Ho, P., and Yip, S., 1987, ''Current Density and Temperature Distributions in Multilevel Interconnections With Studs and Vias,'' IEEE Proc. Int. Reliab. Phys. Symp., **25**, pp. 130–135.
[2] Dekker, C., 1999, ''Carbon Nanotubes as Molecular Quantum Wires,'' Phys. Today, **52**, No. 5, pp. 22–8.
[3] Majumdar, A., 1999, ''Scanning Thermal Microscopy,'' Annu. Rev. Mater. Sci., **29**, pp. 505–585.
[4] Khurana, N., and Chiang, C.-L., 1986, ''Analysis of Product Hot Electron Problems by Gated Emission Microscopy,'' IEEE Proc. Int. Reliab. Physics Symp., **25**, pp. 189–194.
[5] Ju, Y. S., and Goodson, K. E., 1998, ''Short-Time-Scale Thermal Mapping of Micro Devices Using a Scanning Thermoreflectance Technique,'' ASME J. Heat Transfer, **120**, No. 2, pp. 306–313.
[6] Shi, L., Plyasunov, S., Bachtold, A., McEuen, P. L., and Majumdar, A., 2000, ''Scanning Thermal Microscopy of Carbon Nanotubes using Batch Fabricated Probes,'' Appl. Phys. Lett., **77**, No. 26, pp. 4295–4297.
[7] Luo, K., Shi, Z., Varesi, J., and Majumdar, A., 1997, ''Sensor Nanofabrication, Performance, and Conduction Mechanisms in Scanning Thermal Microscopy,'' J. Vac. Sci. Technol. B, **15**, No. 2, pp. 349–360.
[8] Gomes, S., Trannoy, N., and Grossel, P., 1999, ''DC Thermal Microscopy: Study of the Thermal Exchange Between a Probe and a Sample,'' Meas. Sci. Technol., **10**, pp. 805–811.
[9] Shi, L., Kwon, O., Wu, G., and Majumdar, A., 2000, ''Quantitative Thermal Probing of Devices at Sub-100 nm Resolution,'' IEEE Proc. Int. Reliab. Phys. Symp., **38**, pp. 394–398.
[10] Shi, L., Kwon, O., Miner, A., and Majumdar, A., 2001, ''Design and Batch Fabrication of Probes for Sub-100 nm Scanning Thermal Microscopy,'' J. Microelectromech. Syst., **10**, pp. 370–378.
[11] Williamson, M., and Majumdar, A., 1992, ''Effect of Surface Deformations on Contact Conductance,'' ASME J. Heat Transfer, **114**, pp. 802–810.
[12] Johnson, K. L., 1985, *Contact Mechanics*, Cambridge University Press, NY.
[13] Roark, R. J., and Young, W. C., 1989, *Roark's Formulas for Stress and Strain*, McGraw-Hill, NY.
[14] Kiesewetter, L., Zhang, J.-M., Houdeau, D., and Steckenborn, A., 1992, ''Determination of Young's Moduli of Micromechanical Thin Films Using the Resonance Method,'' Sens. Actuators A, **35**, pp. 153–159.
[15] Helg, B., 1990, ''On a Nonvolatile Memory Cell Based on Micro-Electro-Mechanics,'' IEEE Proc. Micro Electro Mechanical Systems, Napa Valley, CA., **90**, pp. 172–176.
[16] Peterson, K. E., 1978, ''Dynamic Micromechanics on Silicon: Techniques and Devices,'' IEEE Trans. Electron Devices, **ED25**, pp. 1241–1250.
[17] Samsonov, G. V., ed., 1968, *Handbook of the Physicochemical Properties of the Elements*, IFI-Plenum, New York.
[18] Insepov, Z., Manory, R., Matso, J., Yamada, I., 1999, ''Ionized Cluster Beam as a Hardness Measurement Tool,'' Nucl. Instrum. Methods Phys. Res. B, **148**, pp. 47–52.
[19] Jiang, D. S., and Kim, D. E., 1996, ''Tribological Behavior of Ultra-Thin Soft Metallic Deposits on Hard Substrate,'' Wear, **196**, pp. 171–179.
[20] Cooper, M. G., Mikic, B. B., and Yovanovich, M. M., 1969, ''Thermal Contact Conductance,'' Int. J. Heat Mass Transf., **12**, pp. 279–300.
[21] Touloukian, Y. S., and Ho, C. Y., 1972, *Thermophysical Properties of Matter*, Vols. 1 and 2, Plenum Press, New York.
[22] The thermal conductivity of low stress SiN_x was measured by S. Huxtable in our group using the 3-ω method.
[23] Rohsenow, W., and Choi, H., 1961, *Heat, Mass, and Momentum Transfer*, Prentice-Hall, Englewood Cliffs, NJ.
[24] Bejan, A., 1993, *Heat Transfer*, John Wiley & Sons, New York.
[25] Tien, C.-L., Armaly, B. F., and Jagannathan, P. S., 1969, ''Thermal Conductivity of Thin Metallic Films and Wires at Cryogenic Temperatures,'' *Thermal Conductivity*, Plenum Press, New York, pp. 13–19.
[26] Tien, C.-L., Majumdar, A., and Gerner, F. M., 1998, *Microscale Energy Transport*, Taylor & Francis, Washington, DC.
[27] Nakabeppu, O., Igeta, M., and Inoue, T., 1999, ''Microscale Real Temperature Measurement by the AFM Using Thermal Feedback Method,'' Therm. Sci. Eng., **7**, No. 6, pp. 1–7.

Computations of the Flow and Heat Transfer in Microdevices Using DSMC With Implicit Boundary Conditions

Yichuan Fang
e-mail: Yichuan.Fang@wmich.edu

William W. Liou
Associate Professor
e-mail: William.Liou@wmich.edu

Department of Mechanical and Aeronautical Engineering,
Western Michigan University,
Kalamazoo, MI 49008

The heat transfer and the fluid dynamics characteristics of subsonic gas flows through microchannels are examined using the direct simulation Monte Carlo (DSMC) method. A simple implicit treatment for the low-speed inflow and outflow boundaries for the DSMC of the flows in microelectromechanical systems (MEMS) is used. Micro-Couette flows and micro-Poiseuille flows are simulated with the value of the Knudsen numbers ranging between 0.06 and 0.72. Where appropriate, the calculated velocity slip and temperature distribution are compared with analytical solutions derived from the Navier-Stokes equations with slip-boundary conditions. A patterned microstructure with nonuniform surface temperature is also simulated. The computational results show that the Knudsen number and the geometric complexity have significant effects on the heat transfer as well as the fluid dynamics properties of the microfluid flows studied. [DOI: 10.1115/1.1447933]

Keywords: Channel Flow, Computational, Heat Transfer, Microscale, Monte Carlo

1 Introduction

MEMS is an emerging technology with a significant potential for future growth. MEMS devices are manufactured using processes similar to those used in the fabrication of microprocessors, such as optical lithography and etching. MEMS are capable of sensing and controlling physical processes with length scales in the order of one micron [1]. Current applications for such devices include, for example, thermomechanical data storage, high temperature pressure sensors, and laminar flow control. Many of these proposed designs involve internal fluid flows, such as microchannel flows. A thorough understanding of the heat transfer properties of such flows is important to the design, fabrication, and operation of MEMS.

Despite the growing number of realized applications of MEMS in scientific and engineering devices, there is only a minimum level of understanding of the fluid dynamics and heat transfer processes in fluidic MEMS. Performance of MEMS often defies predictions made using scaling laws developed for large systems. In addition, heat can easily build up in a densely packed MEMS protective housing. The heat can cause undesirable deformation and becomes destructive. This could be of particular concern to MEMS structures that operate in a high temperature environment. As the functions of the future MEMS expand, there is a pressing need for reliable computational capabilities for the thermal and fluid dynamic processes in MEMS.

In most MEMS flows, the mean free path of the fluid, λ, is of the same order as the characteristic system size, h. A ratio of the two length scales, λ/h, is commonly referred to as the Knudsen number, Kn. For many MEMS flows, the value of Kn is high and the fluid can no longer be regarded as a continuum. The fluid motions are better described from the molecular point of view, as opposed to the continuum point of view. The use of continuum-based techniques in MEMS flow analyses may therefore lead to large errors in the predictions.

The DSMC method of Bird [2] is a well-developed technology. This technique models a gas flow as thousands or millions of simulated "molecules," with each of them representing a large number of real gas molecules. As the simulated molecules move in the computational domain, they may collide with other molecules or with physical boundaries. The macroscopic flow quantities, such as the velocity and temperature, are determined by sampling the calculated molecular properties. The method has been widely used for problems in, for example, rarefied gas dynamics [3], contaminant pollution over space platform [4], and aerodynamics of reentry vehicles [5]. Computations of the high-Kn flow in microchannels have recently been demonstrated [6–9]. Comprehensive reviews can be found in Muntz [10] and Oran et al. [11] DSMC has also been applied successfully in the prediction of the heat transfer of high-speed flows in MEMS [12].

The numerical treatment of the flow boundaries is important to a successful DSMC simulation. For the flow in MEMS, the flow speed can be much lower than the most probable thermal speed. As a result, the numerical treatments that have been applied successfully [2–11] to high-speed flows become physically unrealistic for low-speed MEMS applications, as they do not properly model the mass flux due to thermal fluctuations. Liou and Fang [13] proposed an implicit treatment for the low-speed inflow and outflow boundaries. In the implicit treatment, gas molecules are allowed to enter the flow domain from both the upstream inlet and the downstream exit boundaries. The local mean velocities, temperature, and number density at the boundaries at every sample average step determine the number of the entering molecules, their velocities, and internal energies.

In this study, the heat transfer and flow characteristics of microchannel flows are examined using DSMC and the implicit treatment of the boundary conditions. The Knudsen number of the microflows considered ranges between 0.06 and 0.72. The effects of the wall temperature and the Knudsen number on the temperature and the velocity variations are examined. In the following, the DSMC method and the implicit boundary conditions used in this study are briefly described.

2 Direct Simulation Monte Carlo Method

The DSMC method was introduced by Bird in the early 1960s to simulate hypersonic rarefied gas flows. It is a numerical solution procedure to solve the dynamic equations for real gases. A

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division May 2, 2001; revision received October 1, 2001. Associate Editor: H. Bau.

real gas is modeled through thousands or millions of simulated particles. As the simulated molecules move, they may collide with the other simulated molecules or interact with the physical boundaries, such as a wall. The DSMC simulation solves for the time-dependant variation of the locations and velocity of the simulated molecules. The locations and the velocity components of the simulated molecules are determined and stored in sequences of time. The solution for a steady state case is obtained as an asymptotic limit of an associated unsteady one.

Bird's DSMC procedure consists of the following steps: (1) index molecules into cells, (2) track the movement of molecules, and (3) select collision pairs, and calculate post-collision properties. Starting from the initial data, the flowfield and surface quantities are sampled repeatedly. The physical space is divided into computational cells and each cell is also divided into subcells. The topology of the computational cells is used to facilitate the choice of molecules for collisions and for the sampling of the macroscopic flow properties such as pressure, density, and temperature. The time step is set such that a typical molecule moves about one third of the cell dimension at each time step. Also, this time step is much less than the mean collision time, defined as the mean period of time between the successive molecular collisions.

3 Implicit Boundary Treatment

Microflows often operate with a given pressure (gradient) at the inlet and the outlet boundaries. In this section, an implicit treatment for low-speed inlet and exit boundaries for the DSMC of microflows in such operating conditions is briefly described for completeness. A detailed derivation can be found in Liou and Fang [13].

From the microscopic point of view, gas molecules translate, in addition to a mean molecular velocity, by the thermal or random velocity. For flows at high speeds, such as hypersonic flows, the thermal velocity can be smaller in magnitude compared with the mean velocity. For a DSMC simulation of high-speed flow, a conventional approach is to impose a ''vacuum'' condition at the exit boundary, where no molecules enter the computational domain from the region external to the flow domain. For the low-speed flows in fluidic MEMS, the thermal motion can be of the same order of magnitude as the mean molecular motion. It then becomes inappropriate to neglect the mass influxes at a flow boundary.

A boundary treatment has been proposed for such MEMS flow simulations in Liou and Fang [13]. In this method, the number of molecules entering the computational domain and their corresponding internal energy and velocity components are determined in an implicit manner by the local mean flow velocity, temperature, and number density.

The number flux of the molecules entering the computational domain can be described by using the Maxwellian distribution function,

$$F_j=\frac{n_j}{2\sqrt{\pi}\beta_j}[\exp(-s_j^2\cos^2\vartheta)+\sqrt{\pi}s_j\cos\vartheta\{1+\mathrm{erf}(s_j\cos\vartheta)\}], \tag{1}$$

where

$$s_j=U_j\beta_j \tag{2}$$

$$\beta_j=1/\sqrt{2RT_j}. \tag{3}$$

F_j represents the number flux through a cell face of the boundary cell j. ''erf'' denotes the error function, R the gas constant, and n_j the number density of molecules in cell j. T_j and U_j denote the local temperature and the streamwise mean velocity component, respectively. The value of ϑ is zero for the upstream boundary and π for the downstream exit boundary.

The velocity components of the entering molecule can be determined by using the acceptance-rejection method [2] and the Maxwellian distribution function. At the upstream inlet boundary, the streamwise velocity u and cross-stream velocities, v and w, of the molecules entering the computational domain through the cell face of a boundary cell j can be written as

$$u=(U_j+3C_{mp})R_f \tag{4}$$

$$v=A\cos\varphi \tag{5}$$

$$w=A\sin\varphi, \tag{6}$$

where

$$A=\sqrt{-\ln(R_f)}C_{mp},\quad \varphi=2\pi R_f,\quad C_{mp}=\sqrt{2RT_j},$$

and R_f represents a random fraction number, and C_{mp} the local most probable thermal speed of molecules. At the downstream, the velocity components for the molecule entering the computational domain through the exit flow boundary are

$$u=(U_j-3C_{mp})R_f \tag{7}$$

$$v=A\cos\varphi+V_j \tag{8}$$

$$w=A\sin\varphi, \tag{9}$$

where V_j denotes the local transverse mean velocity.

With the vibrational energy neglected, the internal energy of the entering equilibrium gases of diatomic molecule consists of translational energy, e_{tr}, and rotational energy e_{rot},

$$e_{tr}=mc^2/2 \tag{10}$$

$$e_{\mathrm{rot}}=-\ln(R_f)KT_j, \tag{11}$$

where c is the speed of an entering molecule, m the mass of the simulated gas, and K the Boltzmann constant.

In order to implement equations (1)–(11), the number density, temperature, and mean velocity at the flow boundaries are needed. At the upstream boundary, the pressure, p_{in}, and density, ρ_{in}, are the given parameters of the flow. The number density, n_{in}, and temperature, T_{in}, can then be obtained according to the conservation of mass and the equation of state. That is

$$n_{in}=\rho_{in}/m \tag{12}$$

$$T_{in}=p_{in}/(\rho_{in}R). \tag{13}$$

The transverse mean velocity, $(V_{in})_j$, is set zero. A first-order extrapolation is used to determine the streamwise mean velocity, $(U_{in})_j$, from that of the computed for cell j. That is

$$(U_{in})_j=U_j. \tag{14}$$

At the downstream boundary, the only given flow parameter is the exit pressure, p_e. The other mean properties of the flow are to be determined implicitly as the calculation proceeds. In the present method, the flow variables are first computed by the following characteristics-theory-based equations

$$(n_e)_j^k=n_j^k+\frac{p_e-p_j^k}{m(a_j^k)^2} \tag{15}$$

$$(u_e)_j^k=u_j^k+\frac{p_j^k-p_e}{mn_j^ka_j^k} \tag{16}$$

$$(v_e)_j^k=v_j^k \tag{17}$$

$$(T_e)_j^k=p_e/[(n_e)_j^kmR]. \tag{18}$$

The subscript e denotes the exit boundary, superscript k the computed quantities at the kth step, and a_j^k the local exit speed of sound. The exit mean flow velocities can then be obtained through a sampling of the following form:

$$(U_e)_j=\frac{1}{N_j}\sum_{i=1}^{N_j}u_{e,i} \tag{19}$$

$$(V_e)_j = \frac{1}{N_j} \sum_{i=1}^{N_j} v_{e,i}. \tag{20}$$

Nance et al. [14] first proposed an application of the characteristics theory that has been used extensively in continuum computational fluid dynamics (CFD) [15] in DSMC simulations. The current implementation of the characteristic equations involves an updating procedure for the mean flow properties at the exit boundary as the DSMC calculation proceeds. These information, in turn, are used in equations (1), (7), (8), and (9) to calculate the properties of the molecules entering the computational domain.

It is shown in the following that the implicit boundary treatment successfully drives a microchannel flow to a stationary state with matching exit pressure and overall mass balance, which are the fundamental criteria for validating internal flow computation. Because of the statistical scatter of the DSMC method, this technique becomes inappropriate for flows of extremely low speeds. The information preservation (IP) method [16,17] has been shown to work well in such conditions. In the following, the results for three types of microchannel flow calculations are presented. The effects of the Knudsen number on the heat transfer characteristics are discussed.

4 Results and Discussions

Micro-Couette flows and micro-Poiseuille flows with various values for the ratio of wall temperature to the inlet flow temperature are simulated. Sketches of the simulated geometry are shown in Fig. 1. The variable hard sphere (VHS) model [2] and diffusive wall boundary condition are used for the collision processes. Some of the computational results are presented and, where appropriate, compared with analytical results based on the Navier-Stokes equations with temperature-jump and velocity-slip boundary conditions.

Micro-Couette Flow. For the simulated micro-Couette flow of nitrogen, the upper wall moves with a speed U of 100 m/s. The pressure at the inlet and the outlet boundaries are both set at 0.83 atm. The inlet flow temperature, T_{in}, and the wall temperature, T_w, are set equal to 300 K. The channel height h is 0.8 μm, and length L 4.0 μm. The Knudsen number, based on the channel height and the inlet conditions, is 0.08. The physical parameters

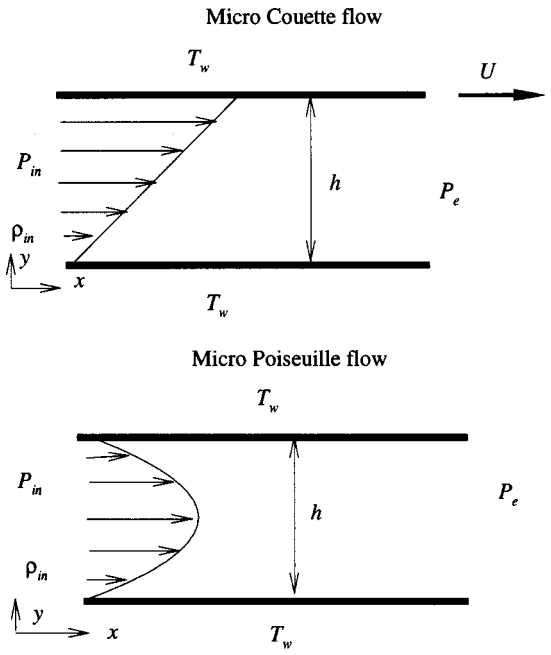


Fig. 1 Simulated micro-geometries

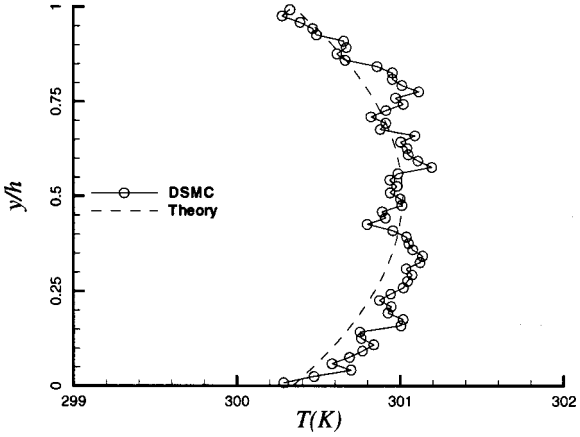


Fig. 2 Comparison of the temperature profile for micro-Couette flow

are set such that a low-speed microflow is simulated. The computational grid consists of 100×60 uniform rectangular cells. The number of the simulated molecules is about 320,000. Runs of 200,000 time steps and sampling after 2000 time steps of development on SGI Octane typically take roughly 48 hours of CPU time with a single processor.

The distribution of temperature in the wall-normal direction, at the station $x/L=2.5$, is shown in Fig. 2 for the micro-Couette flow. Also included is an analytical approximation, which is derived from a simplified form of the Navier-Stokes equations using the first-order wall velocity-slip and the first-order wall temperature-jump condition of Maxwell. The analytical approximation can be written as

$$T = T_w + \frac{1}{2}\frac{\mu U^2}{k}\left(\frac{1}{2\dfrac{2-\sigma}{\sigma}\mathrm{Kn}+1}\right)^2 \times\left[-\left(\frac{y}{h}\right)^2+\frac{y}{h}+\frac{2-\sigma}{\sigma}\frac{2\gamma}{\gamma+1}\frac{\mathrm{Kn}}{\mathrm{Pr}}\right], \tag{21}$$

where μ represents the viscosity, σ the coefficient of heat accommodation, k the conductivity and Pr the Prandtl number. The values used for μ, σ, k, and Pr for nitrogen are [18] 1.656×10^{-5} kg/(ms^2), 1.0, 0.023 J/(msK), and 0.72, respectively. The analytical form gives a distribution that, in general, agrees well with the present DSMC results. The wall temperature-jump, about 0.3 K, is predicted by both methods. The relative scattering error is less than 0.1 percent. Results at other locations are similar because of the one-dimensional nature of the micro-Couette flow. For the flow velocity, a distribution that shows a slight deviation from a linear variation is observed. A detailed comparison of the velocity field has been described in [13].

Micro-Poiseuille Flow. Two micro-Poiseuille flows are calculated. The height of the microchannel, h, is 0.4 μm, and the length L is 2.0 μm for both cases. The inlet flow temperature is 300 K and the wall temperature is 323 K. The pressure ratios are 2.5 for Case 1 and 4.5 for Case 2. The inlet pressure is 2.5 atm and 0.73 atm for Case 1 and Case 2, respectively. These operating conditions result in the local Kn in the microchannels varies between 0.055 and 0.123 for Case 1 and between 0.19 and 0.72 for Case 2. The aspect ratio of the microchannels simulated is 5, which is significantly lower than those found in experiments [19,20]. The computational grid consists of 100×60 uniform rectangular cells. The number of the simulated molecules is about 180,000. Runs of 200,000 time steps, with sampling after 2000 time steps of development, on SGI Octane typically take nearly 36 hours of CPU time with a single processor. Significant speedup

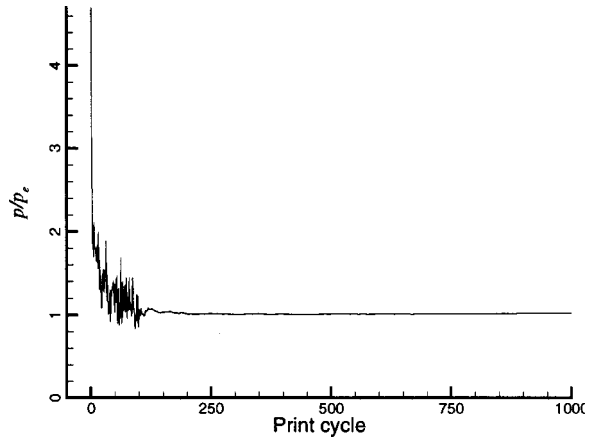


Fig. 3 Evolution of the downstream pressure, Case 2

can be achieved with a parallel version of the DSMC code [21]. A multi-fold speedup is normally obtained. The results shown in the following for Case 2 are obtained using a dual-processor SGI Octane. A constant velocity of 100 m/s is used uniformly in the computational region, including the flow boundaries, to initiate the calculation.

Figure 3 shows the evolution of the calculated pressure at the downstream boundary, nondimensionalized by the imposed exit pressure, p_e, as the solution progresses for Case 2. In each print cycle, there are ten time steps of sampling. The calculated pressure converges to the imposed value ($p/p_e \cong 1$) after a transient variation from the uniform initial conditions has subsided.

Figure 4 shows the approach to a steady state solution using the same print cycles as in Fig. 3 for Case 2. As was described earlier, an initial velocity field (U_{initial}) is required to start the calculations and $U_{\text{initial}} = 100$ m/s is used. For a pressure-driven internal flow, this computational initial condition set at the interior flow domain and the flow boundaries where the boundary treatment is used should have no bearing on the final numerical solution. For validation, Fig. 4 also includes results for a second U_{initial} ($=15$ m/s). Figure 4(*a*) shows the variation of the mass fluxes at the upstream and downstream boundaries. The mass flux is defined as

$$\dot{m} = \frac{1}{h}\int_0^h \rho u dy. \tag{22}$$

The corresponding initial values are denoted by the solid symbols in Figs. 4(*a*) and 4(*b*). The initial mass flux is about 12 kg/ms for the case with $U_{\text{initial}} = 15$ m/s and 80 kg/ms for $U_{\text{initial}} = 100$ m/s. For both cases Fig. 4(*a*) shows that, after a transient from the uniform initial conditions, the mass fluxes at the upstream and the downstream converge to the same constant value and an overall mass balance in the microchannel is established. Figure 4(*a*) also shows that the converged mass fluxes for the two different initial fields agree well, which is about 39.4 kg/ms with a 0.25 percent variation. Figure 4(*b*) shows the variation of the average streamwise velocity for an inlet cell and an outlet cell located at $(0,h/2)$ and $(L,h/2)$, respectively. For both initial conditions, the streamwise velocity for the inlet cell converges to 58.78 (± 0.54 percent) m/s. For the outlet cell, the velocity converges to 232.05 (± 0.32 percent) m/s. These results agree well with Bird's calculation [22] using extended buffer domains and vacuum boundary conditions. The results shown in Figs. 3 and 4 indicate that the present implicit treatment of the low-speed boundaries supports a stable and efficient solution process for the DSMC of the internal microflows studied.

Figure 5 shows the contours of the streamwise mean velocity magnitude for Case 1 and Case 2. The flow velocity increases as it develops downstream. Partly due to the high-pressure gradient,

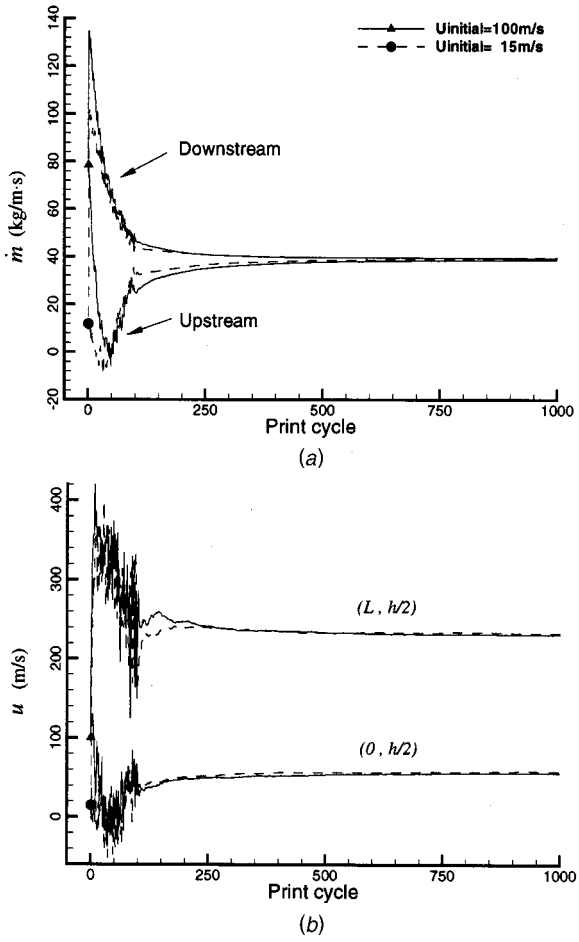


Fig. 4 (*a*) Evolution of mass fluxes, Case 2; and (*b*) evolution of the streamwise velocity, Case 2.

the flow speed at the exit of Case 2 is higher than that of Case 1. There is only small statistical scattering in the computed velocity for both cases.

Figure 6 shows the variations of the gas slip-velocity along the wall for both cases. The slip-velocity is defined as

$$u_s = u_g - u_w, \tag{23}$$

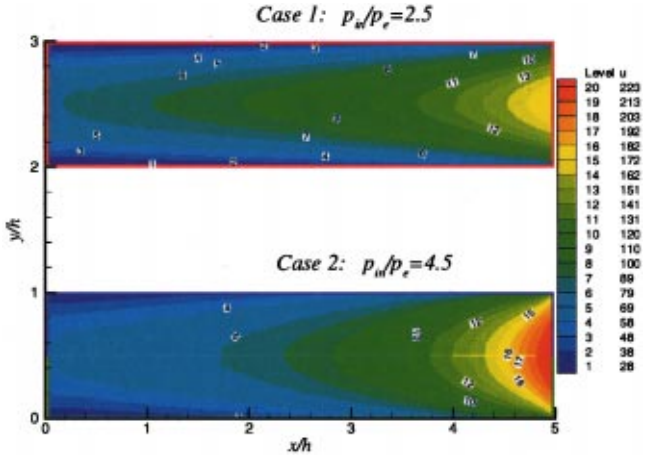


Fig. 5 Streamwise velocity magnitude contours

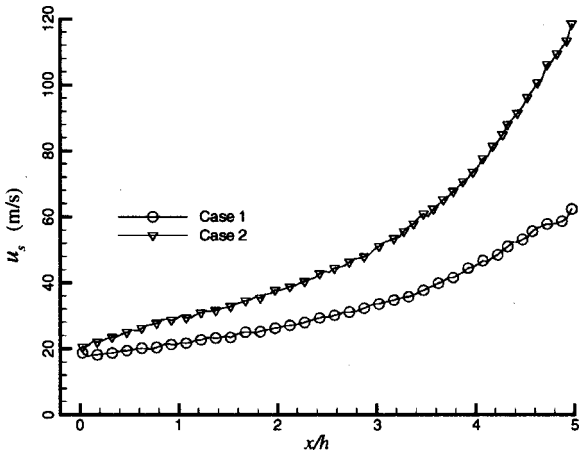


Fig. 6 Slip velocity on the wall

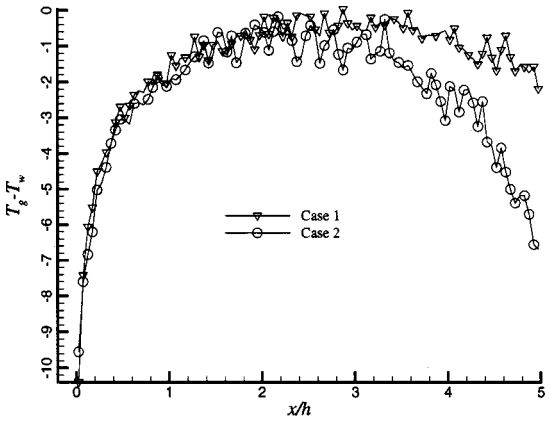


Fig. 8 Wall temperature-jump

where u_g represents the gas velocity on a wall, and u_w the wall velocity. It can be seen that the slip-velocity increases as the flow develops downstream. For the high Kn case, Case 2, the increase is more significant. At the exit of the microchannel, the slip-velocity for Case 2 is nearly twice as high as that of Case 1.

Analytical forms [13] of the velocity distribution can be obtained using the Navier-Stokes equations with slip-wall conditions. The resulting wall slip-velocity can be written

$$\frac{u_s}{u_c} = \mathrm{Kn} \Big/ \left(\frac{1}{4} + \mathrm{Kn}\right) \tag{24}$$

$$\frac{u_s}{u_c} = \frac{\mathrm{Kn}}{1+\mathrm{Kn}} \Big/ \left(\frac{1}{4} + \frac{\mathrm{Kn}}{1+\mathrm{Kn}}\right), \tag{25}$$

where Kn is the local Knudsen number on the wall, which varies with the flow developing downstream. Equation (24) is obtained by using the first-order slip-wall condition proposed by Maxwell and equation (25) by using a high-order condition [23]. In Fig. 7, the wall slip-velocity given by Eqs. (24) and (25) are compared with the present DSMC predictions. For Case 1 with the lower Kn, the continuum based analytical values agree well with the DSMC results in the entire microchannel. As the Kn is low, it is not surprising to observe such a good agreement between the continuum-based solutions and the current DSMC results. This agreement also further validates the DSMC solver used in the present study. The difference between the first-order and the high-order approximations becomes more apparent for Case 2, where the Kn is large (>0.19). In fact, Fig. 7 shows that the first-order approximation is no longer valid for Case 2, producing a velocity slip significantly higher than both the DSMC result and the high-order analysis.

The difference between the wall temperature and the gas temperature near the wall, or the temperature-jump, for these two cases are given in Fig. 8. For both cases, there is a gradual decrease of temperature-jump in the first half of the channel, from about 10 K to nearly the same, suggesting a corresponding gradual decrease of the heat transfer from the wall to the microflow. In the second half of the channel, the wall temperature-jump for Case 2 shows a significant increase as the flow develops toward the exit than that for Case 1.

The calculated distributions of temperature are shown in Fig. 9. The flows are seen to have an increase of temperature as the flow develops in the first half of the channel. This is likely to be caused by the temperature difference between the wall and the flow. However, the contours show a subsequent decrease of temperature in the downstream half of the microchannel for both cases, with a more pronounced reduction in Case 2.

A different view of the temperature distribution is given in Fig. 10. The temperature distribution across the channel for these two cases develops in a similar manner before $x/h = 3.0$. Further downstream, the temperature decreases, with a more significant reduction for Case 2 with high Kn. Near the exit, the temperature in the centerline region of Case 2 is about 15 K lower than that predicted for Case 1. Figure 11 shows the wall heat transfer distribution. The net heat flux on a wall element with length ox can be evaluated as

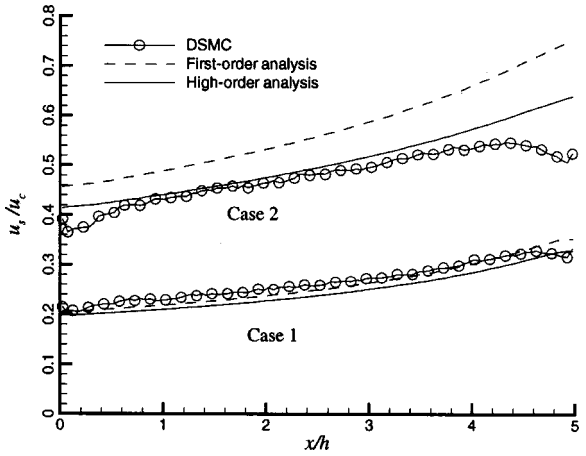


Fig. 7 Comparison of normalized slip velocity distributions

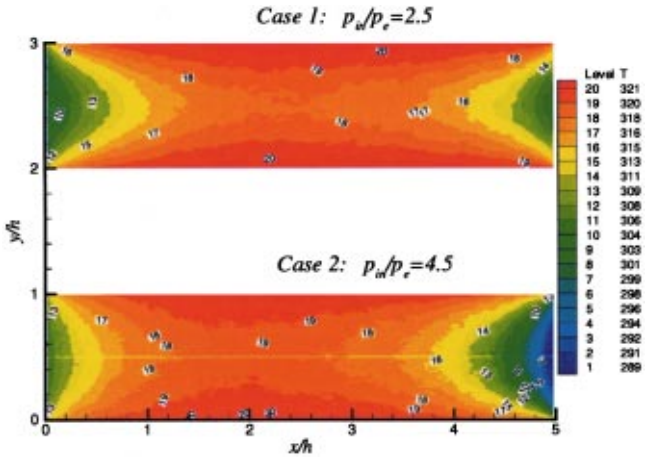


Fig. 9 Temperature contours

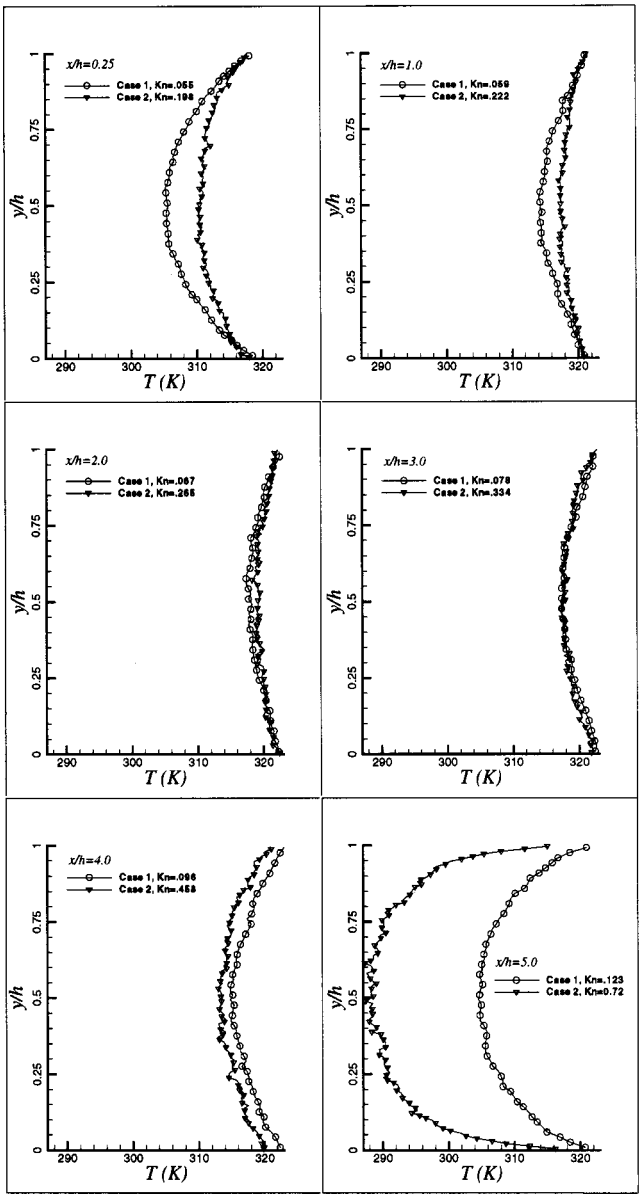


Fig. 10 Comparison of temperature profiles

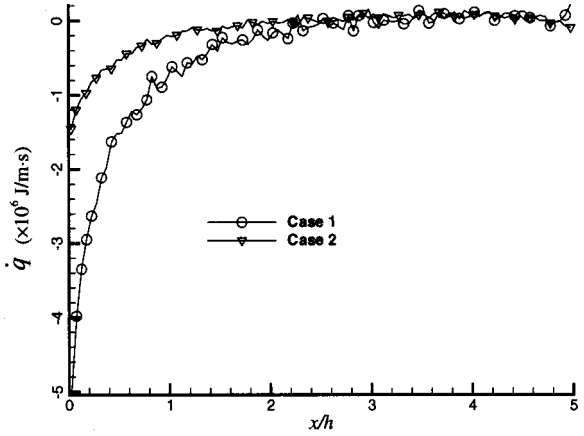


Fig. 11 Wall heat transfer

the wall and the flow is not accompanied by any significant amount of wall heat transfer. Note that the heat transfer characteristics of supersonic flows with three different values of Kn in a microchannel have been studied in Liou and Fang [12]. For the case with the highest Kn (0.186), the results showed a significant increase of wall heat transfer with large temperature difference between the wall and the flow. On the other hand, the temperature along the streamwise direction develops in a manner similar to the current low-speed cases. That is, the temperature for the flow with large Kn is higher than that with low Kn at the upstream and, as the flow develops downstream, a more substantial drop of the flow temperature is found for cases with high Kn. These computational

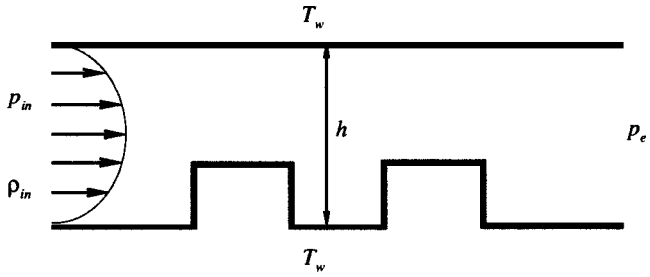


Fig. 12 Sketch for patterned microchannel

$$\dot{q} = \frac{[(\Sigma_{i=1}^{n}(e_{tr,i}+e_{\mathrm{rot},i})_{inc} - \Sigma_{i=1}^{n}(e_{tr,i}+e_{\mathrm{rot},i})_{\mathrm{ref}}]\cdot N_0}{t_s(1\cdot\Delta x)}, \quad (26)$$

where n is the total number of simulated molecules that strike the wall element during sampling; N_0 is the number of gaseous molecules associated with a simulated molecule; and t_s is the time period of the sampling. The subscripts "inc" and "ref" denote the values before and after the molecule impacts the wall, respectively. Near the entrance region, there is significant transfer of heat from the wall for both cases due to the high wall temperature. The level is higher for Case 1 than for Case 2 with higher Kn. As the flow develops downstream the wall heat transfer diminishes and the difference between Case 1 and Case 2 falls within statistical error. The results suggest that the flow expansion observed in Fig. 5 in the second half of the channel is nearly adiabatic for both cases, despite the finite temperature-jump at the wall. This is particularly true for Case 2. Recall that the gas temperature near the exit for Case 2 is lower than that for Case 1 by about 15 K and than the wall temperature by about 40 K. It suggests that, for the high Kn case (Case 2), the large temperature difference between

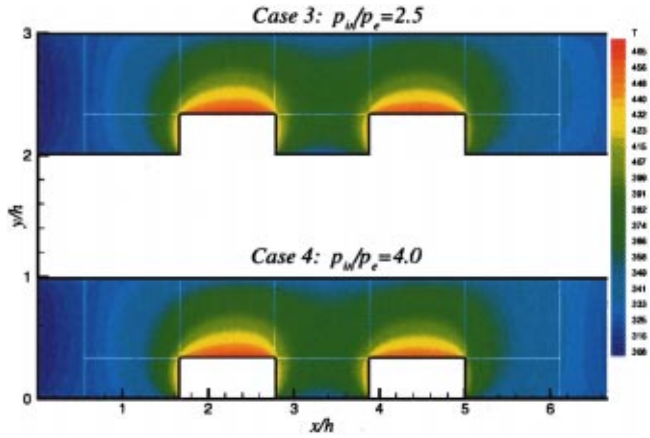


Fig. 13 Temperature contours

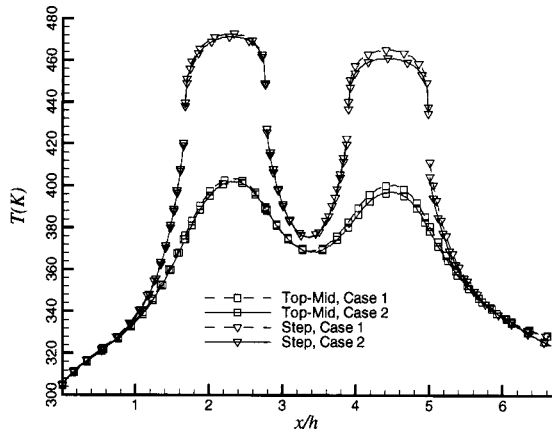


Fig. 14 Temperature distributions

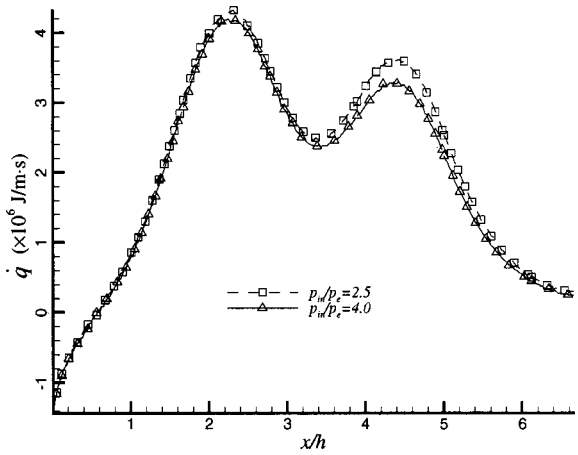


Fig. 16 Comparison of heat transfer on the upper wall

results appear to indicate that the heat transfer characteristics of high Kn microfluid flows at low speeds can be different from that at high speeds. A detailed, systematic study is apparently needed to gain a better understanding of the possible differences in the heat transfer characteristics between low-speed and high-speed microchannel flows.

Patterned Microchannel Flow. Depending upon the specific applications, the internal flow passage in MEMS may contain various forms of blocks and cavities. These patterns can be micromachined with surface etching that selectively removes materials using, for example, imaged photoresist as a masking template. Figure 12 shows a typical cross-sectional view of such patterns. Compared with the micro-Couette and micro-Poiseuille flows described in the sections above, the flow inside the geometry shown in Fig. 12 is highly two-dimensional. To examine the thermal and microflow phenomena in a more realistic geometry, the present DSMC procedure is applied to the microchannel shown in Fig. 12. The inlet pressure is 0.73 atm and temperature is 300 K. For comparison, two cases with different exit pressure have been computed. The corresponding pressure ratios are 2.5 (Case 3) and 4.0 (Case 4), which are comparable to those of Case 1 and Case 2, respectively. The channel height is 0.9 μm with an aspect ratio of 6.7 and the block height is 0.3 μm. The wall temperature is 323 K, except for the top of the blocks with a temperature of 523 K. The Knudsen number, based on the inlet conditions, is about 0.08. The simulations have been performed in the parallel mode by domain decomposition [21]. The number of the simulated molecules is about 1.62 million in each case, with 16,000 uniform rectangular cells. Runs of 200,000 time steps, with sampling after 2000 time steps of development, on ten processors of Pentium III-800 take roughly 50 hours.

Figure 13 shows the temperature contours for both cases. Mushroom-like regions of high temperature can be observed surrounding the blocks, due to the high temperature at the top of the blocks. These high temperature regions for the two cases are geometrically similar. Figure 14 shows the streamwise variation of temperature at two height levels for both cases; one at the top of the blocks and the second at midway between the top of the blocks and the top wall. The temperature in the regions above the blocks is significantly higher than the regions upstream and downstream of the blocks. The highest temperature change is roughly 60 K between these two levels, which are 0.3 μm apart. Except for the region above the second block, there is a minimal difference in the streamwise temperature distributions between the two cases.

Figure 15 shows the temperature-jump distributions along the upper and the horizontal part of the lower walls. The temperature-jumps along both the upper and the lower walls are high in the regions around the blocks. The absolute value of the temperature-jump is higher on the top surface of the block than on the upper wall. For instance, on the top surface of the first block the temperature-jump is nearly 80 K, while that for the upper surface is about 20 K. Overall, the temperature-jump distributions for Case 3 are similar to those of Case 4. A comparison of the heat transfer at the top channel wall for the two cases is given in Fig. 16. The peak values roughly correspond to the locations of the blocks. The heat transfer for the low pressure-ratio case (low Kn) is slightly higher than that for the higher value. This agrees well with the results for the micro-Poiseuille flows presented above where the flow with high Kn shows less heat transfer compared to that with lower Kn.

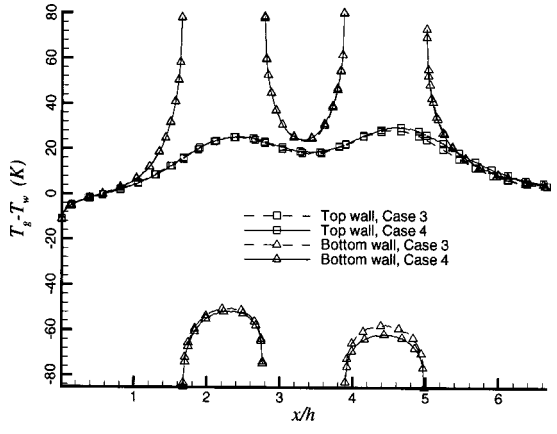


Fig. 15 Temperature-jump on the upper and lower wall

Concluding Remarks

Three types of subsonic microchannel flows have been computed using the DSMC method and an implicit treatment of the flow boundaries developed specifically for the simulation of low-speed microfluid flows. The thermal and the fluid dynamics of the computed flows are examined. Where applicable, the present calculations agree well with the analytical results. The wall heat transfer in the calculated subsonic microfluid flows decreases with an increase of Knudsen number. The present results show that the Knudsen number and the geometric complexity of the channel have significant effects on the fluid dynamic and the thermodynamic behavior of the microfluid flows studied.

Acknowledgment

The authors thank Professor G.A. Bird for his comments on the implicit boundary conditions. The helpful suggestions from the reviewers are also acknowledged.

References

[1] Ho, C. M., and Tai, Y. C., 1994, *MEMS: Science and Technology, Application of Microfabrication to Fluid Mechanics*, ASME, pp. 39–49.
[2] Bird, G. A., 1994, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*, Oxford Science Publications.
[3] Moss, J. N., Mitcheltree, R. A., Dogra, V. K., and Wilmoth, R. G., 1994, ''Direct Simulation Monte Carlo and Navier-Stokes Simulations of Blunt Body Wake Flow,'' AIAA J., **32**, pp. 1399–1406.
[4] Rault, A. G., and Woronowicz, M. S., 1995, ''Application of Direct Simulation Monte Carlo Satelite Contamination Studies,'' J. Spacecr. Rockets, **32**, pp. 392–397.
[5] Ivanov, M., Markelov, G. N., Gimelshein, S. F., Mishna, L. V., Krylov, A. N., and Grechko, N. V, 1997, ''Capsule Aerodynamics With Real Gas Effects From Free Molecular to Near-Continuum Regions,'' AIAA paper 97-0476.
[6] Oh, C. K., Oran, E. S., and Cybyk, B. Z., 1995, ''Microchannel Flow Computed with DSMC-MLG,'' AIAA Paper 95-2090.
[7] Piekos, E. S., and Breuer, K. S., 1995, ''DSMC Modeling of Micromechanical Devices,'' AIAA Paper 95-2089.
[8] Xue, H., Fan, Q., and Shu, C., 2000, ''Prediction of Micro-channel Flows Using Direct Simulation Monte Carlo,'' Probab. Eng. Mech., **15**, pp. 213–219.
[9] Sun, H., and Faghri, M., 2000, ''Effects of Rarefaction and Compressibility of Gaseous Flow in Microchannel Using DSMC,'' Numer. Heat Transfer, Part A, **38**, pp. 153–168.
[10] Muntz, E. P., 1989, ''Rarefied Gas Dynamics,'' Annu. Rev. Fluid Mech., **21**, pp. 387–417.
[11] Oran, E. S., Oh, C. K., and Cybyk, B. Z., 1998, ''Direct Simulation Monte Carlo: Recent Advances and Applications,'' Annu. Rev. Fluid Mech., **30**, pp. 403–441.
[12] Liou, W. W., and Fang, Y. C., 2001, ''Heat Transfer in Microchannel Devices Using DSMC,'' J. Microelectromech. Syst., **10**, pp. 274–279.
[13] Liou, W. W., and Fang, Y. C., 2000, ''Implicit Boundary Conditions for Direct Simulation Monte Carlo Method in MEMS Flow Predictions,'' Computer Modeling in Engineering & Science, **1**, No. 4, pp. 119–128.
[14] Nance, R. P., Hash, D. B., and Hassan, H. A., 1998, ''Role of Boundary Conditions in Monte Carlo Simulation of MEMS Devices,'' J. Thermophys. Heat Transfer, **12**, No. 3, pp. 447–449.
[15] Whitfield, D. L., and Janus, J. M., 1984, ''Three-Dimensional Unsteady Euler Equations Solution Using Flux Vector Splitting,'' AIAA Paper 84-1552.
[16] Cai, C., Boyd, I. D., and Fan, J., 2000, ''Direct Simulation Methods for Low-Speed Microchannel Flows,'' J. Thermophys. Heat Transfer, **14**, No. 3, pp. 368–378.
[17] Fan, J., and Shen, C., 2001, ''Statistical Simulation of Low-Speed Rarefied Gas Flows,'' Journal of Computational Physics, **167**, 393–412.
[18] Weast, R. C., Astle, M. J., and Beyer, W. H., 1984, *CRC Handbook of Chemistry and Physics*, 64th Edition, CRC Press.
[19] Harley, J., Huang, Y., Bau, H., and Zemel, J., 1995, ''Gas Flow in Microchannels,'' J. Fluid Mech., **284**, pp. 257–274.
[20] Arkilic, E., Breuer, K., and Schmidt, M., 1994, ''Gaseous Flow in Microchannels,'' *Application of Microfabrication to Fluid Mechanics*, FED-Vol. 197, ASME 1994.
[21] Fang, Y. C., and Liou, W. W., 2002, ''Microfluid Flow Computations Using a Parallel DSMC Code,'' AIAA Paper 2002–1057.
[22] Bird, G. A., 2000, private communication.
[23] Beskok, A., Karniadakis, G. E., and Trimmer, W., 1996, ''Rarefaction and Compressibility Effects in Gas Microflows,'' ASME J. Fluids Eng., **118**, pp. 448–456.

Slip Flow Convection in Isoflux Rectangular Microchannels

Shiping Yu

Timothy A. Ameel
e-mail: ameel@mech.utah.edu

Department of Mechanical Engineering,
University of Utah,
Salt Lake City, UT 84112

Laminar forced convection in thermally developing slip flow through isoflux rectangular microchannels is analytically investigated. Local and fully developed Nusselt numbers, fluid temperatures, and wall temperatures are obtained by solving the continuum energy equation for hydrodynamically fully developed slip flow with the velocity slip and temperature jump condition at the walls. It is found that heat transfer may increase, decrease, or remain unchanged, compared to nonslip flow conditions, depending on aspect ratios and two-dimensionless variables that include effects of the microchannel size or rarefaction and the fluid/wall interaction. The transition points that separate heat transfer enhancement from reduction are acquired for different aspect ratios.
[DOI: 10.1115/1.1447932]

Keywords: Analytical, Forced Convection, Gaseous, Internal, Microscale

1 Introduction

The developing microscale thermal and fluidic systems typically have characteristic length on the order of 1 to 100 μm and are often operated in gaseous environments at standard conditions, where the molecular mean free path is on the order of 100 nm. Under these constraints, noncontinuum dynamics will take place. Knudsen number (Kn) is the parameter that is usually used to describe the departure from the continuum regime and is defined as the ratio of the molecular mean free path to the characteristic length of the system concerned. The slip flow regime is in the region near continuum, and is classified as $10^{-3}<\mathrm{Kn}<0.1$. Therefore, some microscale thermal and fluidie systems must take into account slip flow effects.

There is strong evidence to support the use of continuum momentum and energy equations to model the slip flow heat transfer problem, while the boundary conditions are appropriately modified by imposing velocity slip and temperature jump at the walls [1]. Liu et al. [2] and Arkilic et al. [3] found that the Navier-Stokes equations, if combined with slip flow conditions, would fit the experimental data in some microchannel flows. Otherwise, significant departures were observed between the numerical and experimental results without applying the slip conditions. Ameel et al. [4] found that the temperature jump effect is also important and should be included in the modeling for the slip flow heat transfer problem.

Slip flow heat transfer in circular tubes has been extensively studied and is treated as an extended Graetz problem. Shih et al. [5], Barron et al. [6,7], and X. M. Wang [8] investigated this problem subjected to isothermal boundary conditions. Their results showed that heat transfer appeared to be strengthened by the slip effects; however, the temperature jump condition was actually not included in their models as a primary boundary condition. Ameel et al. [4] and M. L. Wang et al. [9] studied the related problem involving constant heat flux, and found that the slip effects tend to reduce heat transfer. In all of these studies, both the thermal accommodation coefficient and tangential momentum accommodation coefficient were assumed to be unity, which limits the applicability of their results.

More recently, Larrode et al. [10] conducted a more generalized investigation into the extended Graetz problem for the isothermal wall condition by introducing two dimensionless variables. One-dimensionless variable, $\beta_v\mathrm{Kn}$, is the product of Knudsen number Kn and a parameter β_v that includes the tangential momentum accommodation coefficient, which thus represents rarefaction effects, or in a micro thermal fluid system, microscale effects. Another dimensionless variable, β, includes both the thermal and tangential momentum accommodation coefficients, which thus characterizes the fluid/wall interaction. It was found that heat transfer may increase, decrease, or remain unchanged, depending on these two dimensionless variables, $\beta_v\mathrm{Kn}$ and β. In their investigation, Larrode et al. [10] considered a wide range for β and $\beta_v\mathrm{Kn}$, which increases the applicability of their results.

The fact that conduits of rectangular or trapezoidal cross section are easier to manufacture than circular tubes in the actual development of microscale thermal fluid systems has provided strong motivation for the study of slip flow heat transfer in rectangular microchannels. Yu and Ameel [11] first analytically studied a limiting case for the rectangular geometry, i.e., the laminar slip flow thermal entry problem in a micro gap formed by two parallel surfaces with constant wall temperature. This work is of great importance for many proposed microscale thermal fluid systems that have very large microchannel aspect ratios. In addition to slip flow effects, the axial conduction effects were included, which is appropriate since many microscale thermal fluid systems operate with extremely low Reynolds numbers. It was found that the axial conduction always increases heat transfer. Yu and Ameel [12] furthered their analytical studies by considering the laminar slip flow developing convection problem in the more complex three-dimensional rectangular microchannels with constant wall temperature. The above two investigations showed that the two dimensionless variables $\beta_v\mathrm{Kn}$ and β affect the heat transfer in ways similar to the circular case [10]; that is, heat transfer may increase, decrease, or remain unchanged subjected to $\beta_v\mathrm{Kn}$ and β. As for rectangular microchannels, aspect ratio is another parameter that affects the heat transfer. The analytical solutions for the above two problems (parallel plates [11] and rectangular ducts [12]) were obtained by employing a modified integral transform technique [13,14]. Typical solution methods, such as separation of variables and the conventional integral transform technique, were unable to accommodate the nonseparable nature of the corresponding eigenvalue problems. In Yu and Ameel's other work [15], an interesting behavior for the entrance Nusselt number in internal slip flow conditions was observed. They found that the entrance Nusselt number is finite due to slip flow effects and this finite Nusselt number can be evaluated from a simple equation in terms of the two slip flow parameters $\beta_v\mathrm{Kn}$ and β. Furthermore, the expression for the entrance Nusselt number is universal and

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division February 20, 2001; revision received June 28, 2001. Associate Editor: H. Bau.

applicable for any cross-sectional geometry (circular, rectangular, and others) and for both isothermal and isoflux boundary conditions.

In addition to the isothermal wall condition, the other interesting boundary condition, for the practical application of microscale thermal fluid systems, is the well known H_2 condition, defined as constant wall heat flux on all surfaces [16]. Following the previous work [11,12,15], the laminar slip flow thermal entry problem in rectangular microchannels with the H_2 boundary condition is investigated in this paper. The flow is assumed to be hydrodynamically fully developed at the channel entrance and conduction in the channel walls is neglected. An analytical solution is developed using the modified integral transform technique [13,14]. Both velocity slip and temperature jump conditions will be included in the model. Both accommodation coefficient will be explicitly included for increasing applicability. The results will be analyzed by introducing the two dimensionless variables $\beta_v \mathrm{Kn}$ and β.

2 Analysis

2.1 Solutions to Energy Equation. Slip flow heat transfer can be appropriately modeled by utilizing continuum momentum and energy equations with the modified slip flow boundary conditions [1]. Using the coordinate system shown in Fig. 1, the modified boundary conditions including velocity slip and temperature jump at $y=b$ in the y-direction are given as [1,17,18]

$$u(x,b)=\left[-\beta_v \lambda \frac{\partial u}{\partial y}+3\left(\frac{\Re T}{8\pi}\right)^{1/2}\frac{\lambda}{T}\frac{\partial T}{\partial z}\right]_{y=b} \tag{1}$$

$$T(x,b,z)-T_w=\left[-\beta_t \lambda \frac{\partial T}{\partial y}+\frac{1}{4\Re}u^2\right]_{y=b}. \tag{2}$$

For typical conditions in which Eckert number is small and Kn is moderate, only the first terms in each boundary condition need be retained [10,11,12]. These conditions have been assumed here; therefore, the second terms on the right side of both Eqs. (1) and (2) are not included in the analysis. The two terms on the right hand side of Eq. (1) represent velocity slip due to the y-direction transverse velocity gradient and thermal creep due to the axial temperature gradient, respectively. The two terms on the right hand side of Eq. (2) represent temperature jump due to the y-direction transverse temperature gradient and slip velocity viscous heating, respectively. The slip flow boundary conditions on the other three surfaces have forms similar to Eqs. (1) and (2). In Eqs. (1) and (2), β_v and β_t are two dimensionless variables defined as

$$\beta_v=\frac{2-F_v}{F_v} \tag{3}$$

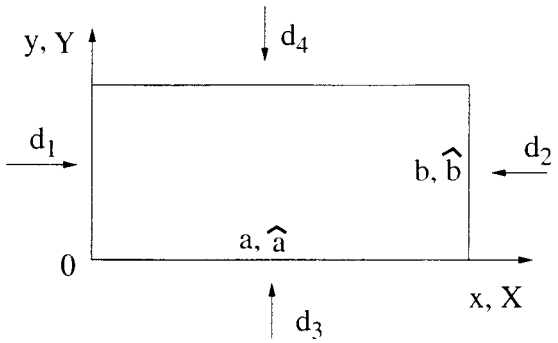


Fig. 1 Coordinate system (flow direction is *Z*)

$$\beta_t=\frac{2-F_t}{F_t}\frac{2\gamma}{\gamma+1}\frac{1}{\mathrm{Pr}}, \tag{4}$$

where F_v and F_t are the tangential momentum accommodation coefficient and thermal accommodation coefficient, respectively. F_v and F_t are close to unity for typical engineering applications [1,17], but F_t could vary in a wide range from 10^{-2} to 1.0 [1]. The ratio of β_t and β_v is defined as

$$\beta=\frac{\beta_t}{\beta_v} \tag{5}$$

and includes all of the parameters associated with the fluid/wall interaction. This parameter is generated in the nondimensionalization of the temperature jump boundary condition given by Eq. (2).

Ebert and Sparrow [19] have obtained the fully developed slip flow velocity profile in a rectangular duct by solving the z-direction momentum equation with slip boundary conditions similar to Eq. (1). This velocity solution has been reproduced by Yu and Ameel [12] using the dimensionless groups consistent with that defined in this paper. Since the origin of the coordinate system used by Yu and Ameel [12] is at the center of the rectangular duct, the dimensionless velocity solution when considering the current coordinate system shown in Fig. 1 and making the required coordinate shift will be

$$U(X,Y)=A\sum_{k=1}^{\infty} B_k F_k(Y) G_k(X), \tag{6}$$

where

$$A=\frac{\hat{a}\hat{b}}{4\Sigma_{i=1}^{\infty}\phi_i^{-5}\left[\dfrac{\sin^2(\phi_i\hat{b}/2)}{\hat{b}/2+(\beta_v \mathrm{Kn}\sin^2(\phi_i\hat{b}/2)}\right]\left[\phi_i\hat{a}/2-\dfrac{\tan h(\phi_i\hat{a}/2)}{1+(\beta_v \mathrm{Kn})(\phi_i)\tan h(\phi_i\hat{a}/2)}\right]} \tag{7}$$

$$B_k=\frac{\sin(\phi_k\hat{b}/2)}{\phi_k^3[\hat{b}/2+(\beta_v \mathrm{Kn})\sin^2(\phi_k\hat{b}/2)]} \tag{8}$$

$$F_k(Y)=\cos[\phi_k(Y-\hat{b}/2)] \tag{9}$$

$$G_k(X)=1-\frac{\cos h[\phi_k(X-\hat{a}/2)]}{\cos h(\phi_k\hat{a}/2)+(\beta_v \mathrm{Kn})(\phi_k)\sin h(\phi_k\hat{a}/2)} \tag{10}$$

$$\cot(\phi_i\hat{b}/2)=(\beta_v \mathrm{Kn})\phi_i \quad (i=1,2,3,\ldots). \tag{11}$$

Consistent with Ebert and Sparrow [19], both u and u_m have been normalized with $D^2/\mu\, dp/dz$ so that the pressure term is not explicitly shown in U.

Under conditions of high Peclet number, negligible energy dissipation, and constant properties, the dimensionless energy equation, the H_2 heat flux boundary conditions, and the inlet condition may be expressed as

$$\frac{\partial^2\Theta}{\partial X^2}+\frac{\partial^2\Theta}{\partial Y^2}=U(X,Y)\frac{\partial\Theta}{\partial Z} \tag{12}$$

$$\left[\frac{\partial\Theta}{\partial X}\right]_{X=0}=d_1 \tag{13}$$

$$\left[\frac{\partial\Theta}{\partial X}\right]_{X=\hat{a}}=d_2 \tag{14}$$

$$\left[\frac{\partial\Theta}{\partial Y}\right]_{Y=0}=d_3 \tag{15}$$

$$\left[\frac{\partial\Theta}{\partial Y}\right]_{Y=\hat{b}}=d_4. \tag{16}$$

$$\Theta|_{Z=0}=0. \tag{17}$$

The boundary conditions given by Eqs. (13)–(16) have been developed by neglecting the contribution from sliding friction. This idealization is appropriate for small slip velocities. Thus, Fourier's law may be applied to the fluid as a thermal boundary condition since the imposed heat flux is directly related to the temperature gradient. Upon the determination of the fluid temperature distribution, the temperature jump condition may be applied to obtain the wall temperature and the Nusselt number. The fact that the fluid temperature $\Theta(X,Y,Z)$ is not a function of the temperature jump effect differs from the isothermal wall case for rectangular microchannels studied by Yu and Ameel [12].

It can be seen from Eqs. (6)–(11) that $U(X,Y)$ is a function of R (implied by $\hat{a}$ and $\hat{b}$) and β_vKn, where β_vKn represents the effects of rarefaction and/or the microscale. Since $\Theta(X,Y,Z)$ includes $U(X,Y)$, the non-dimensional temperature will be a function of R and β_vKn as well. In addition, both the fluid velocity and temperature profiles are functions of β_vKn; thus, $\beta_v\mathrm{Kn}=0$ represents a particular case where both the velocity slip and temperature jump are absent, recovering the classical nonslip case. β_vKn has a range from 10^{-3} to 10^{-1} in the slip flow regime.

In Eqs. (13)–(16), d_1 through d_4 represent the dimensionless heat fluxes at the walls. The convention used here is that heat flux is always imposed inwardly to the fluid (heating mode). For this convention and considering the coordinate system shown in Fig. 1, the value of d_i ($i=1$, 2, 3, 4) may be $+1$ or -1, while $d_i=0$ represents an abiabatic wall. The eight possible combinations of heat fluxes may be expressed in terms of the associated values of d_i. For the case considered here in which all four channel walls are equally heated, $d_1=d_3=-1$ and $d_2=d_4=1$.

The difficulty associated with solving the energy equation, given by Eq. (12), lies in the nonseparable nature of the related eigenvalue problem which is due to the inclusion of the velocity term. This difficulty may be overcome by means of a modified integral transform technique, similar to that employed by Aparecido and Cotta [13] and Cotta [14] who solved a nonslip convection problem in a rectangular duct. This solution technique has been found to be efficient and relatively computationally cheap [13]. The solution procedure for the current problem starts by constructing the following two auxiliary problems:

$$\frac{d^2\psi(\eta,X)}{dX^2}+\eta^2\psi(\eta,X)=0 \tag{18}$$

$$\left[\frac{d\psi(\eta,X)}{dX}\right]_{X=\hat{a}/2}=0 \tag{19}$$

$$\left[\frac{d\psi(\eta,X)}{dX}\right]_{X=\hat{a}}=0 \tag{20}$$

and

$$\frac{d^2\omega(\xi,Y)}{dY^2}+\xi^2\omega(\xi,Y)=0 \tag{21}$$

$$\left[\frac{d\omega(\xi,Y)}{dY}\right]_{Y=\hat{b}/2}=0 \tag{22}$$

$$\left[\frac{d\omega(\xi,Y)}{dY}\right]_{Y=\hat{b}}=0. \tag{23}$$

Eigenfunctions are found from the above two groups of equations to be $\psi_i=\cos[\eta_i(X-\hat{a}/2)]$, and $\omega_m=\cos[\xi_m(Y-\hat{b}/2)]$, respectively, while the eigenvalues are determined to be

$$\eta_i=2i\pi/\hat{a},\quad \xi_m=2m\pi/\hat{b}\quad (i,m=0,1,2,3,\ldots). \tag{24}$$

Then, the two auxiliary problems in Eqs. (18)–(23) allow the creation of the following integral transform pair, utilizing the above eigenvalues and eigenfunctions:

Transform.

$$\overline{\Theta}^*_{im}(Z)=\int_0^{\hat{a}}\int_0^{\hat{b}}\cos[\eta_i(X-\hat{a}/2)]\times\cos[\xi_m(Y-\hat{b}/2)]\Theta(X,Y,Z)dXdY \tag{25}$$

Inversion.

$$\Theta(X,Y,Z)=\sum_{i=0}^{\infty}\sum_{m=0}^{\infty}(M_iM_m)^{-1}\cos[\eta_i(X-\hat{a}/2)]\times\cos[\xi_m(Y-\hat{b}/2)]\overline{\Theta}^*_{im}(Z), \tag{26}$$

where

$$M_i=\frac{\hat{a}}{2}+\frac{\sin(\eta_i\hat{a})}{2\eta_i}=\begin{cases}\hat{a}/2 & \text{when } i\neq 0\\ \hat{a} & \text{when } i=0\end{cases},\qquad M_m=\frac{\hat{b}}{2}+\frac{\sin(\xi_m\hat{b})}{2\xi_m}=\begin{cases}\hat{b}/2 & \text{when } m\neq 0\\ \hat{b} & \text{when } m=0\end{cases}. \tag{27}$$

The energy Eq. (12) is now operated on by

$$\int_0^{\hat{a}}\int_0^{\hat{b}}\cos[\eta_i(X-\hat{a}/2)]\cos[\xi_m(Y-\hat{b}/2)]dXdY. \tag{28}$$

After employing the inversion formula (26) and using the nature of orthogonality of the eigenfunctions, this procedure will yield

$$\sum_{j=0}^{\infty}\sum_{n=0}^{\infty}D_{ijmn}\frac{d\overline{\Theta}^*_{jn}(Z)}{dZ}+\begin{cases}0=\hat{b}(-d_1+d_2)+a(-d_3+d_4) & \text{when } i=m=0\\ \eta_i^2\overline{\Theta}^*_{i0}(Z)=(-1)^i\hat{b}(-d_1+d_2) & \text{when } i\neq 0,\quad m=0\\ \xi_m^2\overline{\Theta}^*_{0m}(Z)=(-1)^m\hat{a}(-d_3+d_4) & \text{when } i=0,\quad m\neq 0\\ (\eta_i^2+\xi_m^2)\overline{\Theta}^*_{im}(Z)=0 & \text{when } i\neq 0,\quad m\neq 0\end{cases} \tag{29}$$

where

$$D_{ijmn}=\frac{1}{M_jM_n}\int_0^{\hat{a}}\int_0^{\hat{b}}\cos[\eta_i(X-a/2)]\cos[\eta_j(X-\hat{a}/2)]\times\cos[\xi_m(Y-\hat{b}/2)]\cos[\xi_n(Y-\hat{b}/2)]U(X,Y)dXdY. \tag{30}$$

The transformed inlet condition for $\overline{\Theta}^*_{im}$ becomes

$$\overline{\Theta}^*_{im}(0)=0. \tag{31}$$

The integral in Eq. (30) is evaluated to give

$$D_{ijmn}=\frac{A}{M_jM_n}\sum_{k=1}^{\infty}B_k\Phi_{ijk}\Phi^*_{mnk}, \tag{32}$$

where

$$\Phi_{ijk}=\Phi_c-\frac{(-1)^{i+j}}{[(1+(\beta_v\mathrm{Kn})\phi_k\tan h(\phi_k\hat{a}))]}\times\left[\frac{1}{(\eta_i+\eta_j)^2+\phi_k^2}+\frac{1}{(\eta_i-\eta_j)^2+\phi_k^2}\right] \tag{33}$$

$$\Phi_c=\begin{cases}0 & \text{when } j\neq i\\ \hat{a} & \text{when } j=i=0\\ \hat{a}/2 & \text{when } j=i\neq 0\end{cases} \tag{34}$$

$$\Phi^*_{mnk}=\frac{(-1)^{m+n}}{2}\sin\left(\frac{\phi_k\hat{b}}{2}\right)\left[\frac{1}{\xi_m+\xi_n+\phi_k}-\frac{1}{\xi_m+\xi_n-\phi_k}+\frac{1}{\xi_m-\xi_n+\phi_k}-\frac{1}{\xi_m-\xi_n-\phi_k}\right]. \tag{35}$$

After $\overline{\Theta}^*_{im}(Z)$ is determined, other properties of interest are readily derived. The dimensionless fluid mixed mean temperature $\Theta_m(Z)$ is first evaluated from its definition

$$\Theta_m(Z)=\frac{1}{\hat{a}\hat{b}}\int_0^{\hat{a}}\int_0^{\hat{b}}U(X,Y)\Theta(X,Y,Z)dXdY=\frac{2A}{\hat{a}\hat{b}}\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}\sum_{m=0}^{\infty}(M_iM_m)^{-1}B_kQ_{km}P_{ki}\overline{\Theta}^*_{im}(Z), \tag{36}$$

where

$$Q_{km}=2(-1)^m\sin\left(\frac{\phi_k\hat{b}}{2}\right)\frac{\phi_k}{\phi_k^2-\xi_m^2} \tag{37}$$

$$P_{ki}=P_c-\frac{(-1)^i\phi_k\tan h(\phi_k a/2)}{(\eta_i^2+\phi_k^2)[1+(\beta_v\mathrm{Kn})(\phi_k)\tan h(\phi_k\hat{a}/2)]} \tag{38}$$

$$P_c=\begin{cases}0 & \text{when } i\neq 0\\ \hat{a}/2 & \text{when } i=0\end{cases}. \tag{39}$$

Alternatively, $\Theta_m(Z)$ can also be evaluated directly from an energy balance of the fluid

$$\Theta_m(Z)=\frac{\hat{b}(-d_1+d_2)+\hat{a}(-d_3+d_4)}{\hat{a}\hat{b}}Z. \tag{40}$$

Equation (40) gives the same result as that from Eq. (36). Note that $\Theta_m(Z)$ varies linearly with the flow direction and the slope of the variation only depends on the aspect ratio (R) and thermal boundary conditions.

The local heat transfer coefficient is defined as

$$h(z)=\frac{-k(\partial T/\partial n|_{\mathrm{wall}})}{T_{w,m}-T_m}=\frac{q_0}{T_{w,m}-T_m}, \tag{41}$$

where T_m is the fluid mixed mean temperature, and $T_{w,m}$ is the peripheral average wall temperature defined as

$$T_{w,m}(z)=\frac{1}{l_p}\int_0^{l_p}T_w(l_p,z)dl_p, \tag{42}$$

where l_p is the heated perimeter length.

$T_{w,m}$ in Eq. (42) is now related to the temperature jump condition given in Eq. (2) with only the first term on the right hand side retained. This modified boundary condition (2), after the temperature gradient term ($\partial T/\partial y$) is replaced with constant heat flux (q_0) and after applied to the ith side of the walls, becomes

$$T_{w,i}(z)=T_{fw,i}(z)+\beta(\beta_v\mathrm{Kn})\frac{q_{0,i}D}{k}, \tag{43}$$

where $T_{fw,i}$ is the fluid temperature at the ith side of the walls.

Considering Eq. (42), Eq. (43) is nondimensionalized to give

$$\Theta_{w,m}(Z)=\Theta_{fw,m}(Z)+\beta(\beta_v\mathrm{Kn}), \tag{44}$$

where $\Theta_{fw,m}$ is the dimensionless peripheral average fluid temperature, and is determined to be

$$\Theta_{fw,m}(Z)=\frac{1}{L_p}\int_0^{L_p}\Theta(L_p,Z)dL_p=\frac{1}{\hat{a}\hat{b}}\left[\overline{\Theta}^*_{00}(Z)+a_1\sum_{m=1}^{\infty}(-1)^m\overline{\Theta}^*_{0m}(Z)+a_2\sum_{i=1}^{\infty}(-1)^i\overline{\Theta}^*_{i0}(Z)\right] \tag{45}$$

$$a_1=\frac{2(-d_3+d_4)}{R(-d_1+d_2)+(-d_3+d_4)},\qquad a_2=\frac{2R(-d_1+d_2)}{R(-d_1+d_2)+(-d_3+d_4)}, \tag{46}$$

where L_p refers to the dimensionless heated perimeter length and $L_p=\hat{b}(-d_1+d_2)+\hat{a}(-d_3+d_4)$.

The local Nusselt number, after nondimensionalizing Eq. (41), is given as

$$\mathrm{Nu}(Z)=\frac{h(z)D}{k}=\frac{1}{\Theta_{w,m}(Z)-\Theta_m(Z)}. \tag{47}$$

Then applying Eq. (44) to this equation yields

$$\mathrm{Nu}(Z)=\frac{1}{\Theta_{fw,m}(Z)-\Theta_m(Z)+\beta(\beta_v\mathrm{Kn})}. \tag{48}$$

Since the fluid temperature field $\Theta(X,Y,Z)$ is a function of the velocity slip effect indicated by $\beta_v\mathrm{Kn}$, the dimensionless peripheral average fluid temperature $\Theta_{fw,m}$ given in Eq. (45) is a function of $\beta_v\mathrm{Kn}$ as well. However, it is recognized that the temperature jump condition represented by β is not involved in the solution process for the energy equation, which means that $\Theta_{fw,m}$ is not a function of β. Also note that the dimensionless fluid mixed mean temperature $\Theta_m(Z)$ given in Eq. (40) is a function of neither $\beta_v\mathrm{Kn}$ nor β although it has a complex alternative expression given in Eq. (36). Therefore, the β that explicitly appears in the product $\beta(\beta_v\mathrm{Kn})$ in Eq. (48) is the single way in which the β effect is imparted to $\mathrm{Nu}(Z)$.

Referring to the values of F_t and F_v in Refs. [1], [17], β defined in Eq. (5) may be in a wide range from close to unity to more than 100, and noticeably will never be zero for actual physical surfaces. Hence, setting $\beta=0$ will artificially disable the temperature jump effect as seen from Eq. (43), permitting observation of the temperature jump effect. β together with $\beta_v\mathrm{Kn}$ was first introduced by Larrode et al. [10] in their studies on slip flow heat transfer for the extended Graetz problem in circular tubes. It should be mentioned that $\beta\approx 1.667$ is a typical value for most engineering applications, corresponding to $F_v=1$, $F_t=1$, $\gamma=1.4$, and $\mathrm{Pr}=0.7$. In the present study, β will be investigated in a wide range from 0 to 10.

The slip flow solutions obtained above are exact and are in the form of infinite series. In order to acquire numerical results from these exact solutions, the infinite series must be truncated to finite series. When the truncation number (N) is chosen to be 20 for all infinite summations, the solutions show good convergence and desired accuracy for $Z>10^{-4}$. Values less than 20 resulted in poor convergence or accuracy while values greater than 20 increased the computational time significantly without an improvement in accuracy. The initial value problem of the first order nonhomogenous ordinary differential equation (ODE) system for the transformed potential $\overline{\Theta}^*$ given in Eqs. (29) and (31) can be readily solved by using the built-in matrix functions in MATLAB or other equation solvers.

2.2 Entrance Nusselt Number and Nusselt Number Approximation. From Eq. (48), two interesting results for $\mathrm{Nu}(Z)$ can be observed. First, at the entrance of the duct, $\mathrm{Nu}(0)$ is found not to be infinite due to the term $\beta(\beta_v\mathrm{Kn})$ although the tempera-

ture difference $(\Theta_{fw,m}(0)-\Theta_m(0))$ is always equal to zero. This behavior has been previously repeated by Yu and Ameel [15]. Furthermore, an identical equation for Nu(0) to that from Ref. [15] can also be derived from Eq. (48):

$$\mathrm{Nu}(0)=\frac{h(0)D}{k}=\frac{1}{\beta(\beta_v\mathrm{Kn})}. \tag{49}$$

It has been shown in Ref. [15] that the above equation holds universally regardless of boundary conditions (isoflux or isothermal) and the cross-sectional geometry (rectangular ducts, circular tubes and others). Note that for the nonslip case, Kn=0, producing the characteristic infinite value for Nu(0).

Secondly, $\beta(\beta_v\mathrm{Kn})$ may be the dominant term in the determination of Eq. (48) (much larger than $\Theta_{fw,m}(Z)-\Theta_m(Z)$) in the downstream region $(Z>0)$. In this case, the slip flow Nusselt number can be approximated as

$$\mathrm{Nu}(Z>0)\approx\frac{1}{\beta(\beta_v\mathrm{Kn})} \quad \text{if } \beta(\beta_v\mathrm{Kn}) \text{ is sufficiently large.} \tag{50}$$

A large value of $\beta(\beta_v\mathrm{Kn})$ may result from a large magnitude of β alone in the slip flow regime, where $10^{-3}<\beta_v\mathrm{Kn}<0.1$. Note that the above relation (50) is the same as Eq. (49) except that the equality sign in Eq. (49) strictly holds. Combination of Eq. (49) and (50) indicates that Nu(Z) will approximately be constant if $\beta(\beta_v\mathrm{Kn})$ is sufficiently large.

The above results (Eqs. (49) and (50)) may also be observed in Yu and Ameel [12] for slip flow heat transfer in isothermal rectangular microchannels and the work by Larrode et al. [10] on slip flow heat transfer in isothermal circular tubes.

2.3 The Fully Developed Nusselt Number. The fully developed Nusselt number Nu_∞ can also be determined using the following approach which can lead to much faster convergence than that using the local Nu derived above by letting $Z\to\infty$. The computational time is approximately reduced by a factor of $(N+1)^2/2$ with this method. Equation (29) for the transformed potentials $\bar{\Theta}^*_{im}$ can be expressed as a matrix equation in the form

$$C\left[\frac{d\bar{\Theta}^*_{im}(Z)}{dZ}\right]+E[\bar{\Theta}^*_{im}(Z)]=J$$

$$(i=0,1,2,\ldots,\overbrace{N}^{N},\overbrace{0,0,\ldots,0}^{N},\overbrace{1,1,\ldots,1}^{N},\overbrace{2,2,\ldots,2}^{N},\ldots,\overbrace{N,N,\ldots,N}^{N}), \tag{51}$$

$$(m=\overbrace{0,0,\ldots,0}^{N},\overbrace{1,2,\ldots,N,1,2,\ldots,N,1,2,\ldots,N,\ldots,1,2,\ldots,N}^{N+1}),$$

where C and E are $(N+1)^2\times(N+1)^2$ square coefficient matrices, respectively, J is a $(N+1)^2\times1$ vector and is formed by the right-hand terms in Eqs. (29), and both $[d\bar{\Theta}^*_{im}(Z)/dZ]$ and $[\bar{\Theta}^*_{im}(Z)]$ are $(N+1)^2\times1$ vectors. Note that the variables N and $N+1$ above the brackets in the sequences for i and m represent the number of repetitions for the number in each of the subsequences.

Note that the coefficient matrix E is singular since its first row elements are all zeros which stem from the first condition in Eq. (29). This singularity can be eliminated by the following two steps. First, find $\bar{\Theta}^*_{00}(Z)$ from the first order ODE equation formed by the first row in Eq. (51). After integrating once and applying the inlet condition in Eq. (31), this equation gives

$$\bar{\Theta}^*_{00}(Z)=\frac{1}{C(1,1)}\{[\hat{b}(-d_1+d_2)+\hat{a}(-d_3+d_4)]Z-C_1[\bar{\Theta}^*_{im}(Z)]\}$$

$$(i=1,2,\ldots,\overbrace{N}^{N},\overbrace{0,0,\ldots,0}^{N},\overbrace{1,1,\ldots,1}^{N},\overbrace{2,2,\ldots,2}^{N},\ldots,\overbrace{N,N,\ldots,N}^{N}), \tag{52}$$

$$(m=\overbrace{0,0,\ldots,0}^{N},\overbrace{1,2,\ldots,N,1,2,\ldots,N,1,2,\ldots,N,\ldots,1,2,\ldots,N}^{N+1}),$$

where C_1 represents a $1\times[(N+1)^2-1]$ vector that consists of all the elements except the first one of the first row in C, and $C(1,1)$ is the first element in C and is found from Eq. (30) to be

$$C(1,1)=\frac{1}{\hat{a}\hat{b}}\int_0^{\hat{a}}\int_0^{\hat{b}}U(X,Y)dXdY=1. \tag{53}$$

After all the $\bar{\Theta}^*_{00}(Z)$ and $[d\bar{\Theta}^*_{00}(Z)/dZ]$ in the other rows in C are replaced with Eq. (52), the singularity is eliminated and then Eq. (51) becomes

$$C^*\left[\frac{d\bar{\Theta}^*_{im}(Z)}{dZ}\right]+E^*[\bar{\Theta}^*_{im}(Z)]$$
$$=J^*(i,m \text{ subject to conditions in Eq. (52)}), \tag{54}$$

where C^* and E^* are $[(N+1)^2-1]\times[(N+1)^2-1]$ square coefficient matrices modified from C and E, respectively, and J^* is a $[(N+1)^2-1]\times1$ vector modified from J.

Table 1 Comparison of nonslip Nusselt number $Nu_{\infty,ns}$

R	1	2	3	4	5	6	7	8	10
Present Results	3.135	3.065	3.00	2.975	2.961	2.954	2.950	2.947	2.943
Iqbal et al. [16]	3.09	3.02	2.97	2.94	2.93	2.93	2.94	2.94	2.95
Chandrupatla and Sastri [16]	3.095	—	—	—	—	—	—	—	—
Lyczkowski et al. [16]	3.230	—	—	—	—	—	—	—	—

Since $d\overline{\Theta}^*_{im}(\infty)/dZ=0$, Eq. (54) has an asymptotic solution in the form of

$$[\overline{\Theta}^*_{im}(\infty)]=E^{*-1}D^*\equiv G \quad \times(i,m \text{ subject to conditions in Eq. (52)}), \tag{55}$$

where E^{*-1} is the inverse of E^*, and G is a $[(N+1)^2-1]\times 1$ constant vector. Then, $\overline{\Theta}^*_{00}(\infty)$ is determined by inserting Eq. (55) into Eq. (52)

$$\Theta^*_{00}(Z)=[\hat{b}(-d_1+d_2)+\hat{a}(-d_3+d_1)]Z-C_1G \quad (Z\to\infty). \tag{56}$$

In the fully developed region, then, the dimensionless peripheral average fluid temperature $\Theta_{fw,m}(Z)$ is found by inserting Eqs. (55) and (56) into Eq. (45):

$$\Theta_{fw,m}(Z)=\frac{\hat{b}(-d_1+d_2)+\hat{a}(-d_3+d_4)}{\hat{a}\hat{b}}Z+C_d \quad (Z\to\infty), \tag{57}$$

where C_d is a constant and is given as

$$C_d=\frac{1}{\hat{a}\hat{b}}\left[a_1\sum_{i=N+1}^{2N}(-1)^{i-N}G(i)+a_2\sum_{i=1}^{N}(-1)^iG(i)-\sum_{i=1}^{(N+1)^2-1}C_1(i)G(i)\right]. \tag{58}$$

Given $\Theta_{fw,m}(\infty)$ in Eq. (57), the dimensionless peripheral average wall temperature $\Theta_{w,m}(Z)$ when $Z\to\infty$ is determined immediately from Eq. (44)

$$\Theta_{w,m}(Z)=\frac{\hat{b}(-d_1+d_2)+\hat{a}(-d_3+d_4)}{\hat{a}\hat{b}}Z+C_d+\beta(\beta_v\mathrm{Kn}) \quad (Z\to\infty). \tag{59}$$

Note that both $\Theta_{fw,m}(\infty)$ in Eq. (57) and $\Theta_{w,m}(\infty)$ in Eq. (59) have the same linear coefficient as that of the dimensionless fluid mixed mean temperature $\Theta_m(Z)$, given in Eq. (40), as expected. Therefore, the fully developed Nu is determined from Eq. (48) by considering Eqs. (57) and (40)

$$\mathrm{Nu}_\infty=\frac{1}{C_d+\beta(\beta_v\mathrm{Kn})}. \tag{60}$$

3 Results and Discussion

The solutions are first validated by comparing the nonslip fully developed Nusselt numbers $Nu_{\infty,ns}$ derived here with those in the literature, and the results of this comparison are listed in Table 1. Values of $Nu_{\infty,ns}$ may be calculated either from Eq. (48) for the local Nu or from Eq. (60) for Nu_∞. When using Eq. (48), the axial dimensionless distance $Z=30$ and the truncation number $N=20$. Both methods produce completely identical results. It is seen from the table that the present solutions have a quite satisfactory accuracy.

The fully developed Nusselt number Nu_∞, local $Nu(Z)$, fluid temperatures $\Theta_m(Z)$ and $\Theta_{fw,m}(Z)$, and wall temperature $\Theta_{w,m}(Z)$ are of interest in the current investigations. Some of these parameters are functions of aspect ratio R, the two dimensionless variables β and β_vKn, the heat flux boundary conditions, and the axial dimensionless distance Z. Of the eight types of heat flux boundary conditions, the version for four equally heated sides (version 4) is the most referenced in the literature and therefore only this version will be considered in this paper.

The fully developed Nusselt number Nu_∞ is of importance in designing microscale thermal fluid systems from long microchannels. Figure 2 shows the Nu_∞ dependence on β_vKn, β, and R. Nu_∞ has been normalized with respect to its nonslip value $Nu_{\infty,ns}$ for the same R. Nu_∞ is calculated directly from Eq. (60) with truncation number $N=20$. It is seen that heat transfer may in-

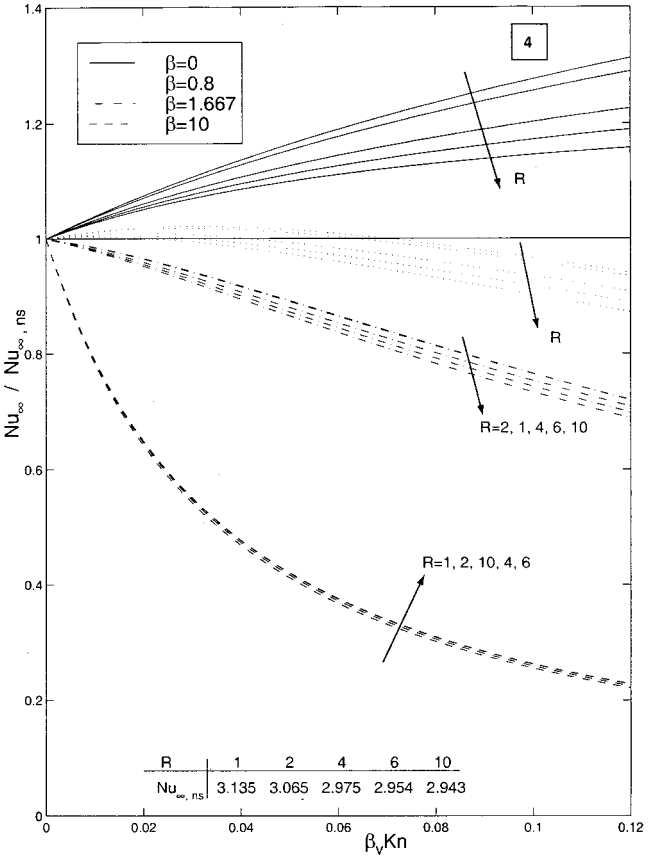


Fig. 2 Normalized Nu_∞ as a function of β, β_vKn, and R

Table 2 β transition values (β_c) for version 4

R	1	2	3	4	5	6	8	10
β_c	0.81	0.80	0.75	0.68	0.67	0.60	0.57	0.53

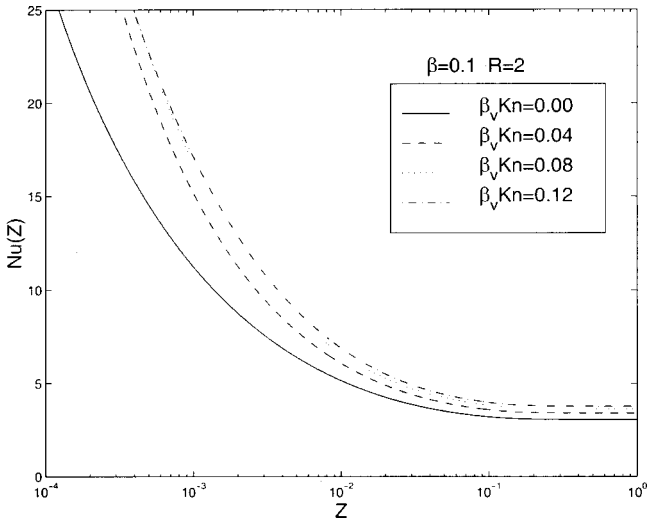


Fig. 3 Local Nu as a function of β_vKn for small β

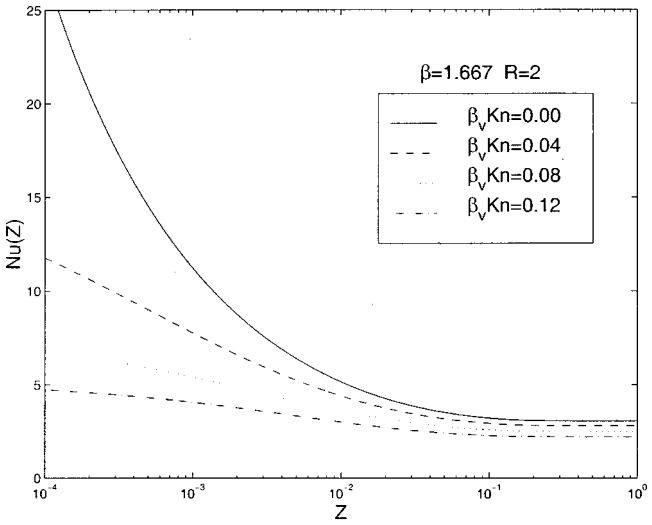


Fig. 5 Local Nu as a function of β_vKn for large β

crease, decrease, or remain unchanged in comparison with the nonslip case, subject to β_vKn, β, and R. A set of dotted curves that denotes $\beta=0.8$ forms a band that approximately separates the heat transfer increase from decrease. Above this band where values of β are small, heat transfer always increases with increasing β_vKn; while below the band, heat transfer always decreases with increasing β_vKn. Physically, β is an indication of the degree of temperature jump at the walls; the larger the value of β, the larger the temperature jump considering Eq. (2) where the temperature gradient is a constant corresponding to the constant heat flux q_0. This large temperature jump results in reduction of the heat transfer coefficient when observing Eq. (41) where the denominator is increased. Note that the mechanism for decreasing heat transfer due to increasing β is different from that for the isothermal boundary condition [12] in which increasing β leads to a decrease in the temperature gradient in the numerator of Eq. (41). On the other hand, β_vKn is a measure of the degree of the momentum exchange between the fluid and the walls; large values of β_vKn correspond to large velocity slip at the walls, thus producing large momentum exchange at the wall. For small values of β, the momentum exchange dominates over the increase in the temperature jump, and, therefore, heat transfer is enhanced by the increasing momentum exchange due to the increasing β_vKn. However, heat transfer is decreased for large β when the increase in temperature jump prevails over the increasing momentum exchange at the wall with increasing β_vKn. Values of β that define the transition point separating the heat transfer enhancement from reduction can be identified for various R. This determination is based on the value of β for which $\mathrm{Nu}_\infty/\mathrm{Nu}_{\infty,ns}=1$ at $\beta_v\mathrm{Kn}\approx 0.07$ (mid range in the slip regime). Accordingly, $\beta_c=0.8$ for $R=2$ from Fig. 2. Values of the transition β_c for different R have been determined and are presented in Table 2.

In Fig. 2, the effects of R on the normalized Nu_∞ are also depicted. As shown, increasing R does not always decrease the normalized Nu_∞. At $\beta=0$ and 0.8, the normalized Nu_∞ drops monotonically with increasing R. However, at $\beta=1.667$, the normalized Nu_∞ for $R=2$ is slightly higher than that for $R=1$ such that the normalized Nu_∞ trend is no longer a monotonic decay with increasing R. At $\beta=10$, no readily identificable trend is apparent. The reason behind this is unclear. Thus, it is noted that when values of β are large, the effects of R on the normalized Nu_∞ become less obvious.

It is also seen from Fig. 2 that the reduction in heat transfer is more significant than the enhancement. At $\beta=10$, the reduction in

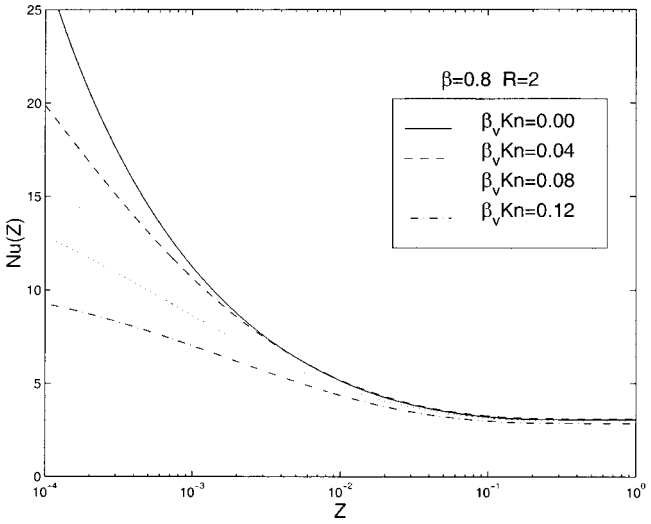


Fig. 4 Local Nu as a function of β_vKn for transition β

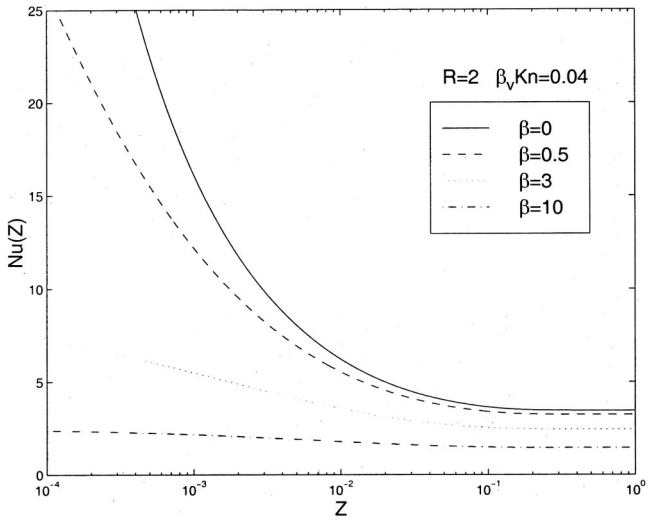


Fig. 6 Local Nu as a function of β at a modest slip condition

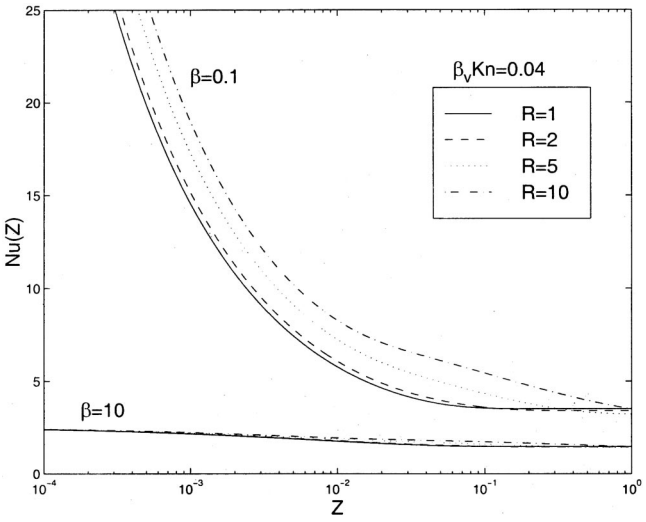


Fig. 7 Local Nu as a function of *R* at a modest slip condition for small and large β

heat transfer is as large as 70 percent, while at $\beta=0$, the promotion in heat transfer is approximately 20 percent. For most engineering applications where $\beta=1.667$, heat transfer is always reduced and the degree of the reduction is on the order of 20 percent for the upper region of the slip regime.

The variations of the local Nu(Z) with β_vKn for small β, transition β_c, and large β are shown in Figs. 3, 4, and 5, respectively. In these three figures, aspect ratio $R=2$ for which β_c has a value of 0.8. For $\beta=0.1$, it is seen from Fig. 3 that the local Nu always increases with increasing β_vKn, compared to the nonslip case (β_vKn$=0$). On the other hand, the case for $\beta=1.667$. shown in Fig. 5, has a reversed trend, i.e., Nu(Z) decreases as β_vKn increases. This is again consistent with the explanations in terms of the molecular dynamics occurring between the fluid and the walls for fully developed conditions. The increase in the momentum exchange with increasing β_vKn is dominant over the slight increase in the temperature jump for small β (Fig. 3); however, when β is large (Fig. 5), the significant increase in the temperature jump will override the increase in the momentum exchange such

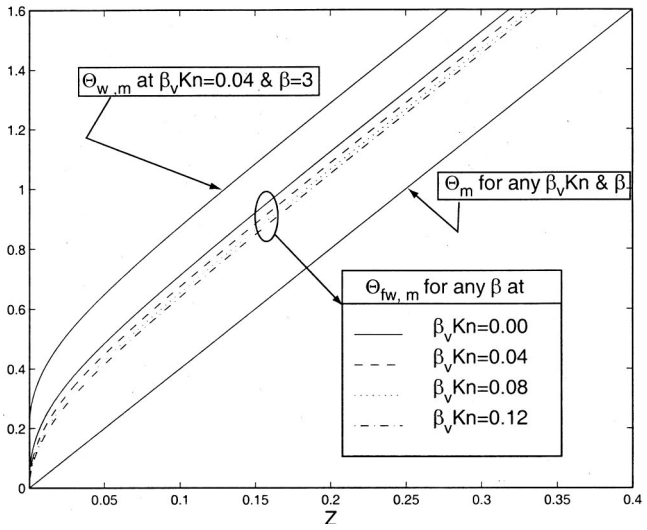


Fig. 8 Effects of β and β_vKn on dimensionless fluid mixed mean temperature Θ_m, dimensionless peripheral average fluid temperature $\Theta_{fw,m}$, and dimensionless peripheral average wall temperature $\Theta_{w,m}$ for any *R*

that the trend for Nu(Z) is reversed. In Fig. 4, β has a mid-range value ($\beta=0.8$) in which the increase in the momentum exchange almost offsets the increase in the temperature jump so that Nu(Z) for $Z>0.01$ either increases slightly for β_vKn$=0.04$, or decreases slightly for β_vKn$=0.08$ and 0.12.

Figure 6 reveals the effects of β on the local Nu for $R=2$ when β_vKn has a mid-range value which represents a modest slip. As expected, heat transfer always decreases with increasing β, which accompanies the increase in the temperature jump at the walls. It is noted that as β increases, the Nu(Z) profile becomes flatter, which means that the entrance length is shortened for increasing β. It is also seen that Nu values at the entrance are finite when $\beta(\beta_v\text{Kn})\neq 0$, which is different from the nonslip situation. Moreover, values of the entrance Nu can be predicted from the simple expression given by Eq. (49) which has been shown to hold for any cross-sectional geometry and boundary condition [15]. For example, for $\beta=10$, Nu$(0)=1/(10\times0.04)=2.5$. which is the approximate value shown in Fig. 6 for the minimum $Z(Z_{\min}=10^{-4}\neq 0)$. In addition, note that in the downstream region ($Z>0$), when $\beta(\beta_v\text{Kn})$ is sufficiently large, Nu(Z) can be approximated by Eq. (50). Again, for $\beta=10$, Nu$(Z>0)\approx 1/(10\times0.04)=2.5$ from Eq. (50). This approximation for Nu_∞ is quite good for the nearly constant profile as shown in Fig. 6.

Figure 7 shows the aspect ratio R influence on the local Nu for a modest slip condition (β_vKn$=0.04$) when β has either a small value (0.1) or a large value (10), respectively. For small β, Nu(Z) increases with increasing R; however, in the region far from the entrance, Nu nearly converges to the same value. On the other hand, for large β, Nu(Z) is quite insensitive to R. Once again, it is noted that for large β the value of Nu(0) agrees well with the prediction from Eq. (49), and the almost constant profile of Nu(Z) can also be well approximated by Eq. (50).

Typical fluid temperature and wall temperature distributions are shown in Fig. 8. As discussed above, the dimensionless fluid mixed mean temperature Θ_m is not a function of either β_vKn or β, and only depends on R and the heat flux boundary conditions. For the four side heating case (version 4) studied here, the slope of $\Theta_m(Z)$ is determined from Eq. (40) to always equal 4 regardless of the value of R. The dimensionless peripheral average fluid temperature $\Theta_{fw,m}$ shows dependence on β_vKn, but is not a function of β, also as mentioned earlier in section 2.1. It is seen that $\Theta_{fw,m}$ decreases with increasing β_vKn in comparison with the nonslip case (β_vKn$=0$). However, the dimensionless peripheral average wall temperature $\Theta_{w,m}$ is a function of both β_vKn and β, as shown in Eq. (44). From Eq. (44), it is recognized that $\Theta_{w,m}$ is $\beta(\beta_v\text{Kn})$ greater than $\Theta_{w,m}$. It is also noted that the value of $\Theta_{w,m}$ is not zero at the entrance and is equal to β_vKn, while on the other hand, $\Theta_{fw,m}$ is always equal to zero at the entrance regardless of the value of β and β_vKn. This nonzero value for $\Theta_{w,m}$ is the reason for the finite Nu at the entrance. In addition, in the fully developed region, it is seen that both $\Theta_{w,m}$ and $\Theta_{fw,m}$ have the same linear slope as that of Θ_m as discussed early from Eqs. (59), (57), and (40).

4 Conclusions

The laminar slip flow thermal entry problem in rectangular microchannels with constant wall heat flux has been analytically studied by a modified integral transform technique. It is found that heat transfer may increase, decrease, or remain unchanged, compared to the nonslip conditions, subjected to two dimensionless variables β_vKn and β, and aspect ratio R. β_vKn indicates the rarefaction effects or microscale effects and is a measure of the momentum exchange between the fluid and the walls, while β represents the properties of fluid/wall interaction and is a measure of temperature jump effects at the walls. For small values of β, heat transfer increases with increasing β_vKn, and this increase can be as great as 20 percent. While β is large, heat transfer decreases with increasing β_vKn, and this decrease can be as high

as 70 percent. A transition β value that separates the heat transfer enhancement from reduction for different R can be identified. A large value of β produces a flatter Nu(Z) profile and tends to reduce the thermal entrance length. This nearly constant local Nu value (as well as Nu_∞) can be approximated well by a simple expression given by Eq. (50). A large value of β also leads to the insensitivity of the local Nu to variations of R. The values of the entrance Nusselt number are again shown, for this isoflux rectangular duct case, to be finite due to slip effects, and can be determined from the simple expression given by Eq. (49). The dimensionless fluid mixed mean temperature is not a function of either β_vKn or β because of the direct energy balance established between the fluid and the imposed heat fluxes at the walls. The dimensionless peripheral average fluid temperature is a function of β_vKn but not β. The dimensionless peripheral average wall temperature depends on both β_vKn and β. These three kinds of dimensionless temperatures have the same linearly varying slope in the region far from the entrance.

Nomenclature

a = channel width
a_1, a_2 = constants, Eq. (46)
$\hat{a}$ = a/D
A = Eq. (7)
b = channel height
$\hat{b}$ = b/D
B_k = Eq. (8)
C = matrix, Eq. (51)
C_1 = vector, Eq. (52)
C_d = constant, Eq. (58)
$d_1 \sim d_4$ = dimensionless heat flux
D = hydraulic diameter, $2ab/(a+b)$
D_{ijmn} = Eq. (30)
E = matrix, Eq. (51)
E^* = matrix, Eq. (54)
F_k = Eq. (9)
F_t = thermal accommodation coefficient
F_v = tangential momentum accommodation coefficient
G = vector, defined in Eq. (55)
G_k = Eq. (10)
h = heat transfer coefficient, defined in Eq. (41)
i = index
j = index
J = vector, Eq. (51)
J^* = vector, Eq. (54)
k = index/thermal conductivity
Kn = Knudsen number, λ/D
l_p = heated perimeter length
L_p = dimensionless heated perimeter length
m = index
M = Eq. (27)
n = index/normal direction to the wall
N = summation truncation number
Nu(Z) = local Nusselt number
Nu_∞ = fully developed Nusselt number for slip flow
$Nu_{\infty,ns}$ = fully developed Nusselt number for nonslip flow
P_c = Eq. (39)
Pe = Peclet number RePr
P_{ki} = Eq. (38)
Pr = Prandtl number
q_0 = constant wall heat flux
Q_{km} = Eq. (37)
R = aspect ratio, b/a
Re = Reynolds number, $u_m D/\nu$
$\mathfrak{R}$ = universal gas constant
T = fluid temperature
T_o = inlet fluid temperature
T_m = fluid mixed mean temperature
T_w = wall temperature
$T_{w,m}$ = peripheral average wall temperature, defined in Eq. (42)
$T_{f,w}$ = fluid temperature at the walls
$T_{fw,m}$ = peripheral average fluid temperature, $1/l_p \int_0^{l_p} T_{fw}(l_p, z)\, dl_p$
u = streamwise velocity
u_m = mean velocity
U = dimensionless velocity, u/u_m
x = transverse coordinate
X = dimensionless transverse coordinate, x/D
y = transverse coordinate
Y = dimensionless transverse coordinate, y/D
z = axial coordinate
Z = dimensionless axial coordinate, $z/(\mathrm{Pe}D)$

Greek Symbols

β = dimensionless variable, β_t/β_v
β_c = transition value of β
β_t = dimensionless variable, defined in Eq. (4)
β_v = dimensionless variable, defined in Eq. (3)
γ = ratio of specific heats
ν = kinematic viscosity
η = eigenvalue, Eq. (24)
Θ = dimensionless fluid temperature, $(T-T_0)/(q_0 D/k)$
Θ_w = dimensionless wall temperature, $(T_w-T_0)/(q_0 D/k)$
$\Theta_{w,m}$ = dimensionless peripheral average wall temperature, $1/L_p \int_0^{L_p} \Theta_w(L_p, Z)\, dL_p$
$\Theta_{fw,m}$ = dimensionless peripheral average fluid temperature, defined in Eq. (45)
Θ_m = dimensionless fluid mixed mean temperature, defined in Eq. (36)
$\bar{\Theta}^*_{im}$ = transformed potentials
λ = molecular mean free path
ψ = eigenfunction
ξ = eigenvalue, Eq. (24)
ϕ = eigenvalue, Eq. (11)
Φ_c = Eq. (34)
Φ_{ijk} = Eq. (33)
Θ^*_{mnk} = Eq. (35)
ω = eigenfunction

References

[1] Eckert, E. G. R., and Drake, R. M., Jr., 1972, *Analysis of Heat and Mass Transfer*, McGraw-Hill, pp. 467–486.
[2] Liu, J. Q., Tai, Y. C., and Ho, C. M., 1995, "MEMS for Pressure Distribution Studies of Gaseous Flows in Microchannels," *Proceedings, IEEE Micro Electro Mechanical Systems*, pp. 209–215.
[3] Arkilic, E. B., Breuer, K. S., and Schmidt, M. A., 1994, "Gaseous Flow in Microchannels, Application of Micro-Fabrication to Fluid Mechanics," ASME FED-Vol. 197, pp. 57–66.
[4] Ameel, T. A., Barron, R. F., Wang, X. M., and Warrington, R. O., 1997, "Laminar Forced Convection in a Circular Tube with Constant Heat Flux and Slip Flow," Microscale Thermophysical Engineering, **1**, No. 4, pp. 303–320.
[5] Shih, Y. P., Huang, C. C., and Tsay, S. Y., 1995, "Extended Leveque Solution for Laminar Heat Transfer to Power-law Fluids in Pipes with Wall Slip," Int. J. Heat Mass Transf., **38**, No. 3, pp. 403–408.
[6] Barron, R. F., Wang, X. M., Warrington, R. O., and Ameel, T. A., 1996, "Evaluation of the Eigenvalues for the Graetz Problem in Slip-Flow," Int. J. Heat Mass Transf., **23**, No. 4, pp. 563–574.
[7] Barron, R. F., Wang, X. M., Ameel, T. A., and Warrington, R. O., 1997, "The Graetz Problem Extended to Slip Flow," Int. J. Heat Mass Transf., **40**, No. 8, pp. 1817–1823.
[8] Wang, X. M., 1996, "Evaluation of the Eigenvalues of the Graetz Problem in Slip-Flow," M.S. thesis. Louisiana Tech University, Ruston, LA.
[9] Wang, M. L., Ameel, T. A., Frazier, A. B., and Warrington, R. O., 1998, "Microtube Convection Heat Transfer For a Power-Law Fluid in Laminar Slip Flow with an Isoflux Boundary Condition," International Mechanical Engineering Congress and Exposition, CA., ASME HTD-Vol.361-3/PID-Vol.3, pp. 157–164.
[10] Larrode, F. E., Housiadas, C., and Drossinos, Y., 2000, "Slip-Flow Heat Transfer in Circular Tubes," Int. J. Heat Mass Transf., **43**, pp. 2669–2680.

[11] Yu, S. P., and Ameel, T. A., 2000, ''Slip-Flow Low Peclet Number Thermal Entry Problem Within a Flat Microchannel Subject to Constant Wall Temperature,'' *Proceedings, Heat Transfer and Transport Phenomena in Microsystems*, Banff, Alberta, Canada, October.
[12] Yu, S. P., and Ameel, T. A., 2001, ''Slip Flow Heat Transfer in Rectangular Microchannels,'' Int. J. Heat Mass Transf., **44**, No. 22, pp. 4225–4234.
[13] Aparecido, J. B., and Cotta, R. M., 1990, ''Thermally Developing Laminar Flow Inside Rectangular Ducts,'' Int. J. Heat Mass Transf., **33**, No. 2, pp. 341–347.
[14] Cotta, R. M., 1993, *Integral Transforms in Computational Heat and Fluid Flow*, CRC Press, FL, pp. 180–188.
[15] Yu, S. P., and Ameel, T. A., 2001, ''A Universal Entrance Nusselt Number for Internal Slip Flow,'' Int. Commun. Heat Mass Transfer, **28**, No. 7, pp. 905–910.
[16] Shah, R. K., and London, A. L., 1978, *Laminar Flow Forced Convection in Ducts, Advances in Heat Transfer, Supplement 1*, Academic Press, New York, pp. 196–222.
[17] Rohsenow, W. M., and Hartnett, J. P., 1973, *Handbook of Heat Transfer*, McGraw-Hill, pp. 9.1–9.8.
[18] Goniak, R., and Duffa, G., 1995, ''Corrective Term in Wall Slip Equations for Knudsen Layer,'' J. Thermophys. Heat Transfer, **9**, No. 2, pp. 383–384.
[19] Ebert, W. A., and Sparrow, E. M., 1965, ''Slip Flow in Rectangular and Annular Ducts,'' Journal of Basic Engineering, Transactions of the ASME, pp. 1018–1024.

Constant-Wall-Temperature Nusselt Number in Micro and Nano-Channels[1]

Nicolas G. Hadjiconstantinou

Olga Simek

Mechanical Engineering Department,
Massachusetts Institute of Technology,
Cambridge, MA 02139

We investigate the constant-wall-temperature convective heat-transfer characteristics of a model gaseous flow in two-dimensional micro and nano-channels under hydrodynamically and thermally fully developed conditions. Our investigation covers both the slip-flow regime $0 \leqslant Kn \leqslant 0.1$, and most of the transition regime $0.1 < Kn \leqslant 10$, where Kn, the Knudsen number, is defined as the ratio between the molecular mean free path and the channel height. We use slip-flow theory in the presence of axial heat conduction to calculate the Nusselt number in the range $0 \leqslant Kn \leqslant 0.2$, and a stochastic molecular simulation technique known as the direct simulation Monte Carlo (DSMC) to calculate the Nusselt number in the range $0.02 < Kn < 2$. Inclusion of the effects of axial heat conduction in the continuum model is necessary since small-scale internal flows are typically characterized by finite Peclet numbers. Our results show that the slip-flow prediction is in good agreement with the DSMC results for $Kn \leqslant 0.1$, but also remains a good approximation beyond its expected range of applicability. We also show that the Nusselt number decreases monotonically with increasing Knudsen number in the fully accommodating case, both in the slip-flow and transition regimes. In the slip-flow regime, axial heat conduction is found to increase the Nusselt number; this effect is largest at $Kn = 0$ and is of the order of 10 percent. Qualitatively similar results are obtained for slip-flow heat transfer in circular tubes. [DOI: 10.1115/1.1447931]

Keywords: Heat Transfer, Microscale, Molecular Dynamics, Monte Carlo, Nanoscale

1 Introduction

Over the last decade there has been an enormous interest in micro and nano-technology. As systems approach microscopic scales, increasing deviations from the well-established continuum laws are reported [17]. In dilute gaseous flows the failure of the continuum description is quantified by the Knudsen number, defined here as the ratio of the molecular mean free path λ to the channel height H. The regime $0 \leqslant Kn \leqslant 0.1$ is referred to as slip-flow; no-slip is captured by $Kn = 0$. For $Kn \gtrsim 0.1$ the continuum description is expected to fail [7], and the regime $0.1 < Kn \leqslant 10$ is referred to as the transition regime because the molecular motion undergoes a transition from diffusive (continuum) for $Kn \lesssim 0.1$, to ballistic for $Kn \gtrsim 10$ (free molecular flow).

In this paper we consider the constant-wall-temperature heat-transfer characteristics of two-dimensional channels that are sufficiently long for flow to be fully developed. Our objective is to understand convective heat transfer in the transition regime, and in particular, to bridge the continuum slip-flow physics to the transition-regime physics. This is a difficult task; the flow physics includes the effects of compressibility, viscous heating, expansion cooling, and thermal creep, which are typically neglected in continuum analyses. For the above reasons, we have performed our molecular simulations on a model problem that requires minimal modeling on the simulation side such that critical comparison between continuum and molecular solutions can be directly translated into information about the validity of the continuum model that is ultimately the most convenient modeling tool. Our model problem focuses on monoatomic gases that do not require modeling of the rotational energy exchange; the latter will be undertaken in a future study. For reasons of computational convenience we have used the simplest monoatomic gas model, the hard sphere gas; similarity guarantees that our non-dimensional results are valid for all dilute monoatomic gases. Our continuum solutions are, of course, not specific to monoatomic gases; this will be further discussed below. Our molecular simulations were performed using a stochastic simulation technique known as the direct simulation Monte Carlo (DSMC).

The constant-wall-temperature Nusselt number in the slip-flow regime can be determined using continuum theory subject to slip boundary conditions (both velocity and temperature); this was first undertaken by Inman [18]. The work of Inman shows that for fully accommodating walls the Nusselt number decreases in the presence of slip, as expected. This was verified by Kavehpour et al. [19], who performed a series of continuum calculations in the slip-flow regime investigating the relative effects of rarefaction and compressibility. However, the work of Inman and subsequent investigators did not include the effects of axial heat conduction that become important for low Peclet numbers that are typical in small scale geometries.

The effects of axial heat conduction have only been investigated in the no-slip limit. The original investigation by Pahor and Strnad [29] shows that in the presence of axial heat conduction the Nusselt number monotonically increases, approaching the limiting value $Nu_T = 8.1174$ as the Peclet number goes to zero. This represents an increase of approximately 10 percent compared to the typically studied infinite Peclet number limit ($Nu_T = 7.5407$).

The constant-wall-temperature convective heat-transfer characteristics of gaseous internal flows in the transition regime are not known. Hadjiconstantinou [13] performed direct Monte Carlo simulations to show that in the constant-wall-heat-flux case the Nusselt number decreases monotonically with increasing rarefaction (Kn) in the fully accommodating case.

The use of the direct simulation Monte Carlo method for modeling flows in microdevices was introduced by Piekos and Breuer [30] and Alexander et al. [2] for the simulation of flows in channels and disk-drive mechanisms respectively. Subsequently, Na-

[1]A significant part of this work has been presented at the 2001 International Mechanical Engineering Congress and Exposition, Heat Transfer Division, session "Fundamentals of Single-Phase Convection"

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division April 6, 2001; revision received August 21, 2001. Associate Editor: J. G. Georgiadis.

gayama et al. [26] investigated heat transfer in channels for Kn >0.001 at a fairly high Mach number (Ma$=0.2$) using DSMC. Recently, Sun and Faghri [35] investigated the effects of rarefaction and compressibility in flows in two-dimensional channels using DSMC.

In the next section we describe the model problem investigated here. In section 3, we calculate the fully-developed slip-flow Nusselt number in two-dimensional channels in the presence of axial heat conduction. The Nusselt number is found to increase in the presence of axial heat conduction throughout the slip-flow regime; however, the effects of axial heat conduction become less important as the Knudsen number increases. In section 4, we describe the computational model used to extend our calculations to the transition regime; the results obtained using this model are presented and discussed in section 5. In section 5 we also present slip-flow heat transfer results for cylindrical tubes. Those are found to be qualitatively similar to the two-dimensional channel results. We finish with our conclusions.

2 Problem Description

We consider the flow of a gas through a two-dimensional channel of length L with walls that are a distance H apart (see Fig. 1). We denote $T=T(x,y)$ and $P=P(x,y)$ the temperature and pressure fields, respectively. The temperature and pressure at the channel entrance and exit are assumed uniform and denoted (T_i,P_i) and (T_o,P_o), respectively. The gas velocity field is denoted $\vec{\mathbf{u}}=\vec{\mathbf{u}}(x,y)=(u_x(x,y),u_y(x,y),u_z(x,y))$.

The wall temperature is given by

$$T_w(x)=T_i, \quad x<\bar{L}$$

$$T_w(x)=T_o, \quad x\geq\bar{L}.$$

The length $\bar{L}$ is sufficiently large for the flow to develop hydrodynamically before heat transfer between the gas and the wall commences. Similarly, L is sufficiently large for the flow to be thermally fully developed and conditions at the outlet to be uniform. In the next section we derive an expression for the thermal development length. The relative temperature jump was taken to be small, i.e., $|T_o-T_i|/T_i\ll 1$, so that transport coefficients can be assumed constant and acceleration effects small.

As is customary, we define the Reynolds number

$$\mathrm{Re}=\frac{\rho u_b 2H}{\mu}, \tag{1}$$

where μ is the gas viscosity and u_b is the bulk velocity evaluated at the location of interest, the bulk velocity at any station being defined by

$$u_b=\frac{\int_A u_x dA}{\int_A dA}, \tag{2}$$

where A is the cross-sectional area of the channel. The Peclet number is defined as Pe$=$Re Pr, where Pr is the gas Prandtl number.

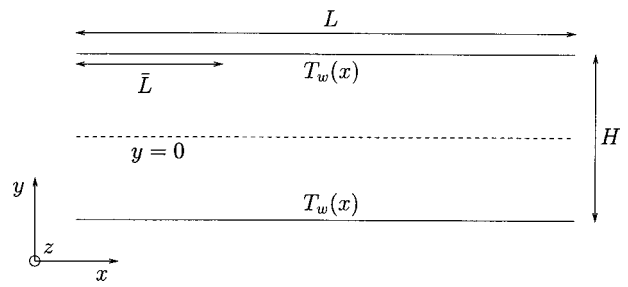


Fig. 1 Channel geometry

A more appropriate measure of the relative importance of convective forces to viscous forces resulting from the fluid acceleration due to compressibility is the modified Reynolds number

$$\mathrm{Re}_L=\frac{\rho u_b 2H}{\mu}\frac{H}{L}. \tag{3}$$

Arkilic et al. [4] have shown that the non-linear pressure distribution measured experimentally can be captured by a "locally fully developed flow" model. The effect of fluid acceleration along the channel was considered by Harley et al. [16] who show that for the pressure ratios investigated here, neglecting the fluid acceleration is justified. An alternative approach that is valid throughout the transition regime is discussed in section 5; DSMC simulations show that for $\mathrm{Re}_L<0.1$, the skin friction coefficient equals the fully developed value. Since fluid acceleration directly influences the skin friction coefficient, this is a good indication that "locally fully developed" models provide good approximations of the flow physics.

The Nusselt number (non-dimensional heat transfer coefficient h) is defined as

$$\mathrm{Nu}=\frac{h2H}{\kappa}=\frac{q2H}{\kappa(T_w-T_b)}, \tag{4}$$

where κ is the gas thermal conductivity, and T_b is the bulk temperature defined by

$$T_b=\frac{\int_A \rho u_x T dA}{\int_A \rho u_x dA}. \tag{5}$$

Note that in the transition regime linear transport theory is not valid. Although the gas viscosity and thermal conductivity can still be defined, these quantities cease to represent the well-defined transport coefficients associated with linear transport in the continuum regime. They are used here to provide continuity with the already existing continuum frameworks; they also arguably still serve as measures of diffusive transport suitable for use in constructing non-dimensional parameters, especially for Knudsen numbers close to slip flow. Alternatively they can be viewed as the transport coefficients of a reference (unbounded) gas at the temperature of the simulation.

Viscous heat generation can be important in small-scale flows. The (continuum) Brinkman number that quantifies the effect of viscous dissipation is typically defined as

$$\mathrm{Br}^*=\frac{\mu u_b^2}{\kappa(T_w-T_b)}. \tag{6}$$

However, because of the presence of slip, both in temperature and velocity, and because for Kn>0.1 the temperature rise due to dissipation is not correctly captured by the continuum description used to construct the above measure, we constructed a second measure of dissipation; we compare T_w-T_b to the temperature rise due to dissipation in "isothermal" $(T_w(x)=T_i)$ simulations at flow velocities equal to the ones encountered in the flows with heat transfer. The modified Brinkman number is defined as

$$\mathrm{Br}=\frac{|T_w-T_b|_{is}}{|T_w-T_b|}, \tag{7}$$

where *is* denotes the "isothermal" simulation for the same bulk velocity. In the latter simulations, the flow was gravity driven and periodic, thus avoiding the effects of expansion cooling. As will be discussed below, viscous heat generation will be neglected in the slip-flow analysis and will be very small in our molecular simulations.

The effect of expansion cooling on the temperature field has been investigated by van den Berg et al. [6] for circular capillaries. This work showed that in the continuum regime, the contribution of expansion cooling can also be scaled with the Brinkman number and is of the same order of magnitude as viscous heat

generation. In fact, due to their competing nature, the combined effect of expansion cooling and viscous heat dissipation is smaller than either of the two. DSMC simulations showed that in our molecular results the effects of expansion cooling are expected to be less than 5 percent, and $\mathrm{Br}<0.03$.

3 Slip-Flow Heat Transfer With Axial Heat Conduction

We now present the calculation of the constant-wall-temperature Nusselt number in the slip-flow regime with axial heat conduction. According to slip-fow theory, the gas velocity at the wall differs from the velocity of the wall by an amount that is proportional to the local velocity gradient. The temperature of the gas at the wall is similarly different from the wall temperature by an amount that is proportional to the local temperature gradient. More specifically, for a stationary wall, the velocity slip is given by

$$u_{\mathrm{gas}}|_{\mathrm{wall}}=\alpha\frac{2-\sigma_v}{\sigma_v}\lambda\left.\frac{du}{d\tilde{\eta}}\right|_{\mathrm{wall}}, \tag{8}$$

where σ_v is the momentum accommodation coefficient that is equal to zero for specular reflections and equal to 1 for diffuse reflections [7], and $\tilde{\eta}$ is the coordinate normal to the wall. The thermal slip at the wall is given by a similar expression

$$T_{\mathrm{gas}}|_{\mathrm{wall}}-T_w=\beta\frac{2\gamma}{\gamma+1}\frac{2-\sigma_T}{\sigma_T}\frac{\lambda}{\mathrm{Pr}}\left.\frac{dT}{d\tilde{\eta}}\right|_{\mathrm{wall}}, \tag{9}$$

where σ_T is the energy accommodation coefficient, and γ is the ratio of specific heats.

The coefficients α and β introduce corrections to the original results of Maxwell ($\alpha=\beta=1$) that were obtained through an approximate method [10]. For air, α and β are usually taken to be equal to unity [7]. Hard sphere molecular dynamics simulations, direct Monte Carlo simulations [25,37] and linearized solutions of the Boltzmann equation [23] show that for hard spheres $\alpha\approx\beta\approx1.1$.

The Nusselt number for constant-wall-temperature is obtained by solution of the differential equation governing energy conservation for the system of interest. In the absence of viscous heat generation expansion cooling and transient effects, this equation can be expressed as [24]

$$\rho c_p u_x\frac{\partial T}{\partial x}=\kappa\left(\frac{\partial^2T}{\partial y^2}+\frac{\partial^2T}{\partial x^2}\right), \tag{10}$$

where c_p is the constant pressure heat capacity of the gas. We seek solution of this equation in the region $x\geq\bar{L}$. According to the problem statement in the previous section, the fluid enters the region $x\geq\bar{L}$ at a temperature

$$T=T_i \quad \text{at } x=\bar{L}. \tag{11}$$

The remaining boundary conditions are

$$\frac{\partial T}{\partial y}=0, \quad \text{at } y=0 \tag{12}$$

$$T-T_w=-\beta\frac{2\gamma}{\gamma+1}\frac{2-\sigma_T}{\sigma_T}\frac{\lambda}{\mathrm{Pr}}\frac{\partial T}{\partial y} \quad \text{at } y=\frac{H}{2}, \tag{13}$$

where in the region of interest $x\geq\bar{L}$, $T_w=T_o$.

The velocity distribution in the channel is assumed to be known based on our assumption of hydrodynamically fully developed flow for $x\geq\bar{L}$. It can be written in terms of the bulk velocity u_b as

$$u_x=\frac{u_b}{\mathrm{Kn}+\frac{1}{6}}\left[\left(\mathrm{Kn}+\frac{1}{4}\right)-\frac{y^2}{H^2}\right], \tag{14}$$

or

$$u_\zeta=\frac{3}{2}u_b\frac{1+4\mathrm{Kn}-\eta^2}{1+6\mathrm{Kn}}, \tag{15}$$

where we have used the definitions $\zeta=2(x-\bar{L})/H$ and $\eta=2y/H$ for the axial and transverse non-dimensional coordinates respectively. In the above equation and in what follows, we will absorb α, β, and the accommodation coefficients in the Knudsen number. The origin of the various terms in the final expressions is clear so that these coefficients can be added to the various expressions if desired.

Using the above definitions for the non-dimensional coordinates, the governing differential equation can be written as

$$\mathrm{Pe}\frac{u_\zeta}{4u_b}\frac{\partial\theta}{\partial\zeta}=\frac{\partial^2\theta}{\partial\eta^2}+\frac{\partial^2\theta}{\partial\zeta^2} \tag{16}$$

where $\theta=T-T_w$. Under the assumption of a separable solution of the form

$$\theta=Y(\eta)\exp\left(-\frac{\omega^2}{\mathrm{Pe}}\zeta\right) \tag{17}$$

the original problem is transformed into the following eigenvalue problem

$$\frac{d^2Y}{d\eta^2}+\left(\frac{u_\zeta}{4u_b}+\frac{\omega^2}{\mathrm{Pe}^2}\right)\omega^2Y=0 \tag{18}$$

$$\frac{dY}{d\eta}=0 \quad \text{at } \eta=0 \tag{19}$$

$$Y=-\frac{4\gamma}{\gamma+1}\frac{\mathrm{Kn}}{\mathrm{Pr}}\frac{dY}{d\eta} \quad \text{at } \eta=1 \tag{20}$$

with ω being the corresponding eigenvalue. Solution of this eigevalue problem was obtained through a standard series-solution method [5,34]. More details can be found in [14].

Fully developed conditions correspond to ζ sufficiently large such that only the first eigenvalue ω_0 is important. An approximate criterion for the realization of fully developed conditions can be formulated by requiring that the contribution of the exponential term based on the second eigenvalue (ω_1) is less than 1 percent of the corresponding term based on the first eigenvalue. This can be expressed as

$$\zeta>\zeta_t=\frac{4.6\mathrm{Pe}}{\omega_1^2-\omega_0^2}. \tag{21}$$

This criterion was used to extract the fully developed Nusselt number in our molecular simulations. Note that we will use the superscript s to denote the fully developed Nusselt number calculated using slip-flow theory. We will reserve the symbol Nu_T to denote the Nusselt number for all Kn including the Nusselt number calculated using the molecular simulation technique described in the next section.

Our slip-flow solution neglects the effects of thermal creep that are important in the presence of walls along which a temperature gradient exists. Inman [18] estimates the region in which the effects of thermal creep may be important to be of the order of $0.01\mathrm{Pe}H/2$; based on our results for ω_0 and ω_1, this is much smaller than $\zeta_tH/2$ and hence we expect the effects of thermal creep to be small. This is further discussed in section 5.

4 Transition-Regime Model

Due to the failure of the continuum description in the transition regime, in this section we use a molecular simulation technique to extend our investigation of the constant-wall-temperature heat transfer characteristics beyond slip-flow theory. Molecular modeling will not only provide valuable information about the Nusselt number in the transition regime that very little is known about, but

will also indicate the limits of applicability of slip-flow theory and the rate at which this theory deteriorates in the transition regime.

The molecular simulation technique used here is known as the direct simulation Monte Carlo and in the form used here it models a dilute hard-sphere gas ($\gamma=5/3$, Pr$=2/3$). The properties of a dilute hard-sphere gas are well known, making the connection to the continuum theory of the previous sections fairly straightforward. The viscosity is given by

$$\mu=\frac{5}{16\sigma^2}\sqrt{\frac{mkT}{\pi}},$$

and the thermal conductivity is given by

$$\kappa=\frac{75k}{64\sigma^2}\sqrt{\frac{kT}{\pi m}},$$

where σ is the hard sphere diameter, k is Boltzmann's constant and m is the molecular mass (the gas constant is given by $R=k/m$).

Note that in the constant-wall-temperature problem and its extensions to non-zero Knudsen numbers and non-negligible axial heat conduction, the Nusselt number is independent of the transport coefficients, in the sense that those can be scaled out through proper non-dimensionalization. We thus did not use any collision models that reproduce the temperature variation of transport coefficients of a real monoatomic gas (variable hard sphere, variable soft sphere [9]), since similarity guarantees that our results should be valid for any dilute monoatomic gas in the limit of small temperature changes ($|T_o-T_i|/T_i\ll 1$). To facilitate comparison with our continuum results in which transport coefficients were assumed constant, our molecular results are non-dimensionalized using transport coefficients evaluated at the local gas temperature.

The simulations will exactly follow the problem description given before, and will involve the same assumptions made in the continuum analysis, namely that viscous heat generation and expansion cooling are negligible, transient effects are absent, the flow is fully developed thermally, and that the effect of fluid acceleration is negligible in the dynamics but enters through a ''locally fully developed'' model. The various criteria developed in this paper have been used to ensure that the molecular simulations satisfy these assumptions. In particular, in our simulations Br, $|\mathrm{Br}^*|$, $\mathrm{Re}_L<0.1$; the Nusselt number given in the next section is determined for $\zeta>\zeta_t$. For $\zeta>\zeta_t$ the fluid bulk temperature has typically decayed to within 2 percent of T_w, i.e., $|T_b-T_w|/T_w<0.02$, thus indicating that acceleration effects due to temperature differences should also be small, as assumed in the continuum analysis.

4.1 The Direct Simulation Monte Carlo. The DSMC method [9] is a particle-based stochastic numerical scheme for solving the nonlinear Boltzmann equation [36]. The motion of a representative set of particles is simulated in time in a series of timesteps, each of which involves a ballistic advection of each molecule and stochastic collisions between pairs of molecules. This ''coarse-grained'' molecular description contains the essential physics to fully capture both the hydrodynamic and kinetic regimes [9]. DSMC offers significant modeling advantages compared to continuum techniques in situations where molecular information is required to achieve closure of the governing hydrodynamic equations, or when the continuum hydrodynamic equations are not valid. As with all molecular simulation techniques, macroscopic properties are defined as averages over molecular data. The standard deviation associated with the distribution of the molecular quantities (population standard deviation) determines the uncertainty in our estimation that manifests itself as a noisy signal; this uncertainty is inversely proportional to the square root of the number of samples taken.

For the sake of brevity we will not present a description of the DSMC algorithm. Excellent introductory [1] and detailed [9] descriptions can be found in the literature, as well as comparisons of DSMC simulation results with solutions of the linearized Boltzmann equation [8] for flows in microchannels. Comparisons of DSMC results with experiments for diverse non-equilibrium phenomena spanning the whole Knudsen range can be found in [28,9].

4.2 DSMC Simulations. We simulated gaseous argon ($\sigma=3.66\times10^{-10}$ m, $m=6.63\times10^{-26}$ Kg) at atmospheric density ($\rho=1.78$ Kg/m^3). The hard sphere diameter for argon is well known to reproduce equilibrium and non-equilibrium properties accurately. Gaseous argon has been historically used in a large number of DSMC studies because it provides instant availability to a substantial literature of simulation and experimental results for code validation. This choice should have no effect on our non-dimensionalized results that should apply to any dilute monoatomic gas.

Although continuum calculations can easily be performed for different values of the accommodation coefficients, DSMC simulations are much more computationally expensive and a complete investigation of the full parameter space is beyond our computational capabilities. Additionally, gas-surface interactions for partially accommodating walls are, at present, not well-characterized [9]. We also expect the assumption of full accommodation to be adequate for low speed flows over rough surfaces [9] that are typical of the engineering systems investigated here. This assumption will be relaxed in subsequent studies.

A pressure ratio of $\Pi=P_i/P_o\sim1.4$ is applied at the ends of the channel that causes the gas flow. This pressure ratio balances the conflicting requirements of small velocities for low Mach number and low viscous dissipation, and large velocities for good signal-to-noise ratio. This is particularly important for the calculation of the heat flux at the wall that is the third moment of the molecular velocity distribution function and is very sensitive to noise. The ends of the channels are subject to the inflow and outflow temperatures T_i and T_o respectively, with $\bar{L}=0.2L$. We have found that this satisfies both the hydrodynamic and thermal development criteria.

The heat flux at the wall is defined as the average (over all collisions in a period of time t) energy exchange between the wall and colliding molecules per unit time and area. For hard spheres, only kinetic energy contributes to the energy exchange, and, thus,

$$q=\frac{1}{St}\sum_t\left(\frac{1}{2}m|\vec{V}'|^2-\frac{1}{2}m|\vec{V}|^2\right), \tag{22}$$

where S is the area over which q is defined, t is the period over which averages are taken, $\vec{V}$ is the molecule velocity before collision, and $\vec{V}'$ is the molecule velocity after collision with the wall. Note that the above definition includes the shear work done by the slipping gas at the wall; this contribution should be small in our simulations since it scales with the Brinkman number.

The total number of molecules used (400,000) was such that the average number of molecules per cell was more than 35; that is substantially more than the number (20) empirically determined to be required for accurate solutions. The number of cells was chosen so that the cell linear dimension is at most 0.7 of a mean free path but typically much smaller than that. Alexander et al. [3] have shown that the transport coefficients deviate from the dilute gas Enskog values as the square of the cell size Δx with the proportionality constant such that for cell sizes of the order of one mean free path an error of the order of 10 percent occurs. For $\Delta x\leqslant0.7\lambda$ the difference between the viscosity of the gas and the viscosity of a dilute hard-sphere gas is less than 4.6 percent. We used the results of Alexander et al. [3] to correct all our results for the transport coefficients (viscosity and thermal conductivity).

The timestep of the simulation Δt was taken to be significantly smaller than the mean free time λ/c_o, where $c_o=\sqrt{2RT}$ is the most probable velocity. It has been shown [11,15] that the error in the transport coefficients is proportional to the square of the timestep, with a proportionality constant such that for timesteps of

the order of one mean free time the error is of the order of 5 percent. In our simulations, $\Delta t < \lambda/(7c_o)$, thus making the error negligible.

Due to the impulsive change in wall temperature in the constant-wall-temperature problem, a relatively large signal $(q, T_b - T_w)$ exists; this results in a relatively small uncertainty in the determination of the fully developed Nusselt number, which has additional benefits in terms of driving forces used in the simulation. This is in contrast to the constant-wall-heat-flux case that required large pressure and temperature ratios that resulted in some cases in non-negligible (but small) Brinkman numbers and thermal creep velocities [13].

5 Results and Discussion

5.1 Fully Developed Flow and the Skin Friction Coefficient. In this section we present an expression for the skin friction coefficient

$$C_f = \frac{\tau_w}{(1/2)\rho u_b^2} \tag{23}$$

in fully developed flow that is valid for all Knudsen numbers. This expression is used here to verify the assumption of locally fully developed flow used in our heat transfer calculations.

An expression for the skin friction coefficient in fully developed flow for arbitrary Knudsen numbers can be obtained through the following scaling relation

$$\dot{Q} = u_b H = -\frac{1}{P}\frac{dP}{dx} H^2 \sqrt{\frac{RT}{2}}\,\overline{Q}, \tag{24}$$

which describes the flowrate, $\dot{Q}$, in pressure-driven flow in two-dimensional channels for all Knudsen numbers [10,27]. Here $\overline{Q} = \overline{Q}(\mathrm{Kn})$ has been determined for hard sphere gases by solution of the linearized Boltzmann equation [27,8]. In the absence of fluid acceleration, the pressure gradient must balance the wall shear stress leading to

$$\frac{2\tau_w}{H} = -\frac{dP}{dx}. \tag{25}$$

Combining the above two equations (24,25) leads to

$$C_f = \frac{32}{5\sqrt{\pi}\,\mathrm{Re}\,\mathrm{Kn}\overline{Q}}. \tag{26}$$

Use of a more general viscosity-based mean free path [25] results in an expression that differs by 2 percent.

Figure 2 shows a comparison between Eq. (26) and DSMC simulations in channels with a length to height ratio $L/H = 20$ and fully accommodating walls. The shear stress at the wall is defined and evaluated as the average (over all collisions in a period of time t) momentum exchange between the wall and colliding molecules per unit time and area

$$\tau_w = \frac{m}{St}\sum_t (\vec{V}' \cdot \hat{t} - \vec{V} \cdot \hat{t}), \tag{27}$$

where S is the area over which τ_w is defined, and $\hat{t}$ is the tangent to the wall (positive in the streamwise direction).

The solid line in Figure 2 denotes Eq. (26), the dashed line denotes the slip flow result

$$C_f = \frac{24}{\mathrm{Re}(1 + 6\alpha\mathrm{Kn})} \tag{28}$$

and the stars denote simulation results with error estimates. The bulk velocities in the above simulations where equal to or greater than the corresponding bulk velocities in the calculations with heat transfer. The largest Reynolds number encountered $\mathrm{Re}_L \approx 0.06$ corresponded to the largest channel height H ($\mathrm{Re} \approx 1.2$, $\mathrm{Kn} \approx 0.06$); smaller channel heights were characterized by smaller Reynolds numbers. The agreement between Eq. (26) and simulations is very good and suggests that the flow is well approximated by the locally fully developed assumption.

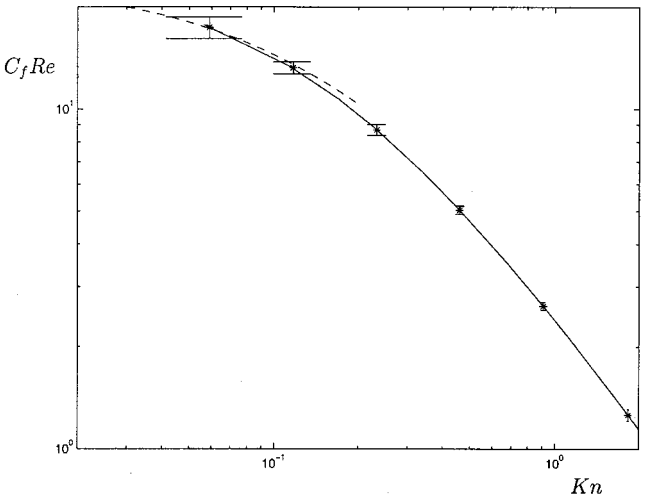


Fig. 2 Comparison between the theoretical prediction of Eq. (26) denoted by a solid line, the slip-flow result (28) with α=1.1 denoted by a dashed line, and the simulation results denoted by stars.

Due to the fairly wide minimum in $\overline{Q}$ centered around $\mathrm{Kn} \approx 1$, equation (26) can be approximated in the region $0.3 < \mathrm{Kn} < 3$ with less than 12 percent error by

$$C_f \approx \frac{2.4}{\mathrm{Kn}\,\mathrm{Re}}. \tag{29}$$

By using the minimum value of $\overline{Q}$ we ensure that Eq. (29) gives an upper bound to the skin friction coefficient in the range $0.3 < \mathrm{Kn} < 3$.

5.2 Slip-Flow Heat Transfer. Figure 3 shows the results of the slip-flow calculation in the range $0 \leq \mathrm{Kn} \leq 0.2$ for air ($\mathrm{Pr} = 0.7$, $\gamma = 1.4$, $\alpha = \beta = 1$) for $\sigma_v = \sigma_T = 1$; we have extended our calcula-

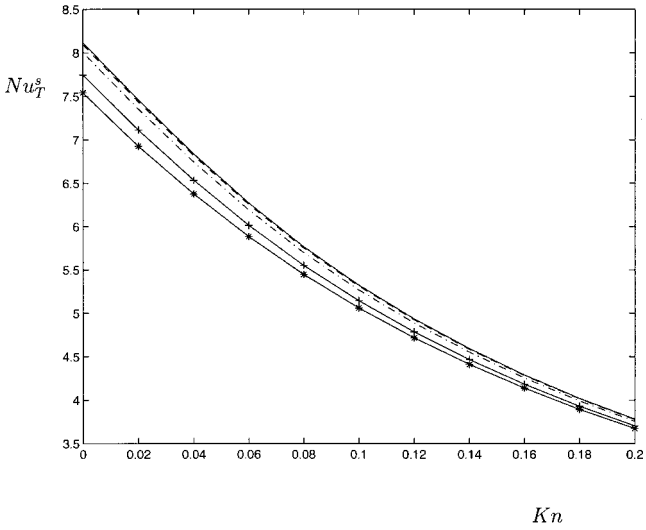


Fig. 3 Variation of Nusselt number Nu_T^s with Knudsen number Kn for air in the fully accommodating case. The solid line denotes Pe→0, the dashed line Pe=0.2, the dash-dotted line Pe=1, the line with crosses Pe=5, and the line with stars Pe→∞.

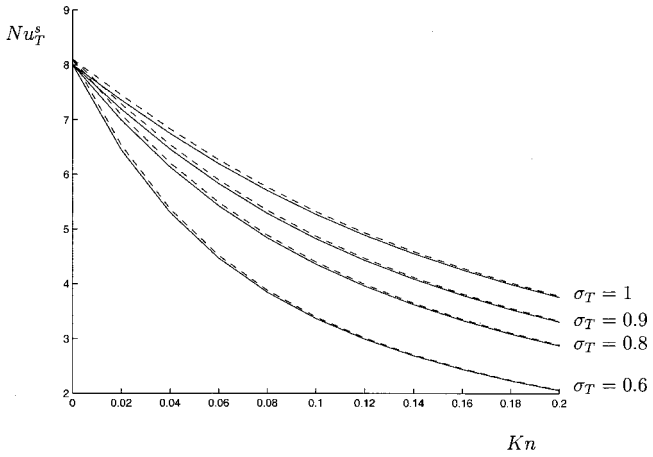


Fig. 4 Variation of Nusselt number Nu_T^s with Knudsen number Kn for air for various values of the thermal accommodation coefficient σ_T with σ_v=1. The solid lines denote Pe=1, and the dashed lines Pe=0.1.

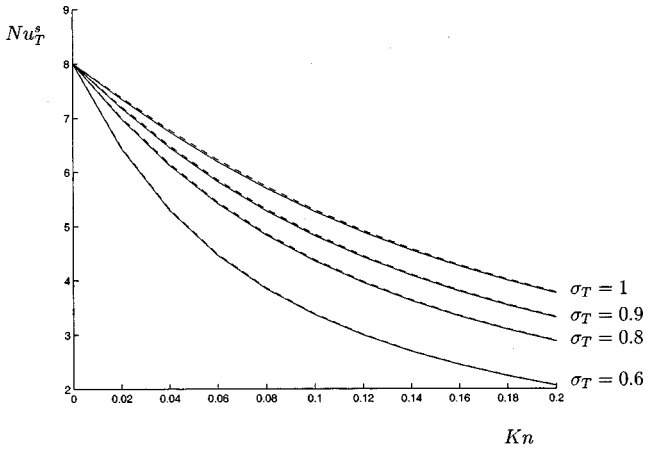


Fig. 5 Variation of Nusselt number Nu_T with Knudsen number Kn for air for various values of the accommodation coefficients for Pe=1. The solid lines denote σ_v=1, and the dashed lines σ_v=0.9.

tions to Kn=0.2 because, as we show later, the slip-flow results are found to be a good approximation to the DSMC results for Kn≤0.2. This figure shows that the Nusselt number decreases monotonically with increasing Knudsen number at constant Pe, for all values of Pe, as expected. The effect of axial heat conduction is to increase the Nusselt number throughout the slip-flow regime. The maximum change occurs at Kn=0 and is of the order of 10 percent, whereas the effects of axial heat conduction are less pronounced as the Knudsen number increases. The dependence of the Nusselt number on the Peclet number in the Pe≪1, Kn=0 limit, is captured by the following asymptotic expression

$$\mathrm{Nu}_T^s(\mathrm{Pe},\mathrm{Kn}=0)=8.11742(1-0.0154295\mathrm{Pe}+0.0017359\mathrm{Pe}^2-\ldots) \tag{30}$$

developed by Grosjean et al. [12], and given here in terms of the Peclet number used in our work.

Although we have presented solutions in the range 0<Pe<∞, the following caveat is in order. Due to the small characteristic dimensions of the channels investigated here, Pe≫1 corresponds to large mean flow velocities that are not consistent with the assumptions of incompressible flow and small Brinkman number made in this study. Hence the solutions for Pe≫1, presented here in the full range 0<Kn<0.2 to facilitate comparison with previous work and illustrate the effects of axial heat conduction, are only physically realistic as Kn→0 and should be interpreted in this way. As we show later, comparison with our DSMC simulations is limited to Pe≤1, since in the latter, Pe≲1.

Figures 4 and 5 show slip-flow results for selected values of the accommodation coefficients. These results show that the dependence of the Nusselt number on the accommodation coefficients is qualitatively similar to the infinite-Peclet number limit studied before [18]: the Nusselt number decreases when the thermal accommodation coefficient decreases; the Nusselt number increases slightly when the momentum accommodation coefficient decreases while the thermal accommodation coefficient remains constant, because a lower momentum accommodation coefficient leads to higher velocities close to the wall and hence enhanced heat transfer. The effect of axial heat conduction is found to be of diminishing importance as the Knudsen number increases, similarly to the fully accommodating case.

The predictions of slip-flow theory for a hard sphere gas (Pr =2/3, γ=5/3, α=β=1.1) are qualitatively the same. A comparison between the two gases for σ_v=σ_T=1 can be found in Fig. 6.

5.3 Transition-Regime Heat Transfer. Figure 7 shows our DSMC results in the range 0.02<Kn<2 for σ_v=σ_T=1. Simulations with both T_o>T_i and T_o<T_i have been performed. The Peclet number in our simulations was less than 1 (except for Kn =0.029 for which Pe=1.35) that is typical of the small scale flows studied here. This feature also makes the comparison between slip-flow theory and DSMC results easier: due to the diminished importance of axial heat conduction as Kn increases, the dependence of Nusselt number on the Peclet number is very weak. Our slip-flow results show that, in the fully accommodating case, the maximum difference between Nu_T^s(Pe=0.1, 0≤Kn≤0.2) and Nu_T^s(Pe=1, 0≤Kn≤0.2) occurs at Kn=0, and is equal to 1.2 percent. The maximum difference between Nu_T^s(Pe=0.01,0≤Kn ≤0.2) and Nu_T^s(Pe=0.1, 0≤Kn≤0.2) also occurs at Kn=0 as expected, and is less than 0.2 percent. Figure 7 also demonstrates this by comparing the hard-sphere slip-flow results for Pe=0.01, 0.1, and 1.0, which are found to be indistinguishable for Kn>0.1.

Our DSMC results show that in the transition regime the Nusselt number continues to decrease monotonically with increasing Knudsen number when the momentum and thermal accommodation coefficients are equal to unity. The results for Kn<0.1 are in very good agreement with the slip-flow calculation for a hard

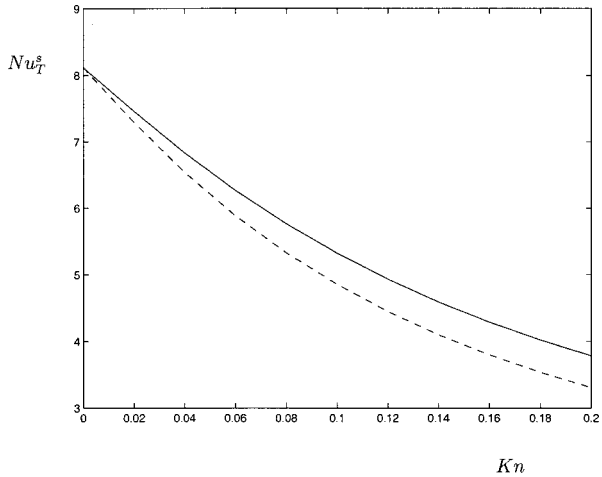


Fig. 6 Variation of Nusselt number Nu_T^s with Knudsen number Kn for Pe→0, σ_v=σ_T=1, for air (solid line) and hard sphere gas (dashed line).

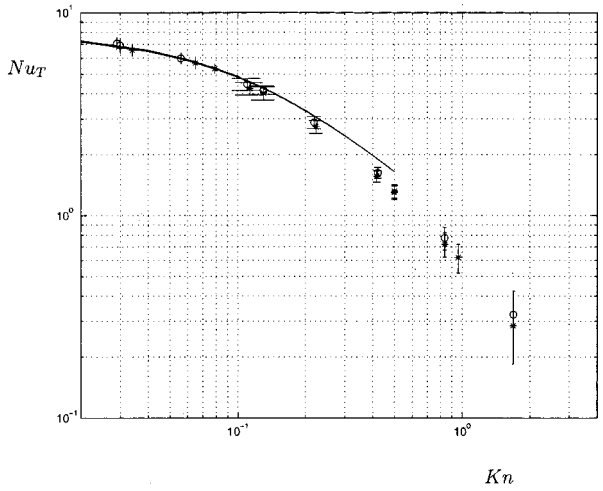


Fig. 7 Variation of Nusselt number Nu_T with Knudsen number Kn. The stars denote DSMC simulation data with $T_o > T_i$ and circles denote DSMC simulation data with $T_i > T_o$. The solid lines denote hard sphere slip-flow results for Pe=0.01, 0.1, and 1.0 which are virtually indistinguishable.

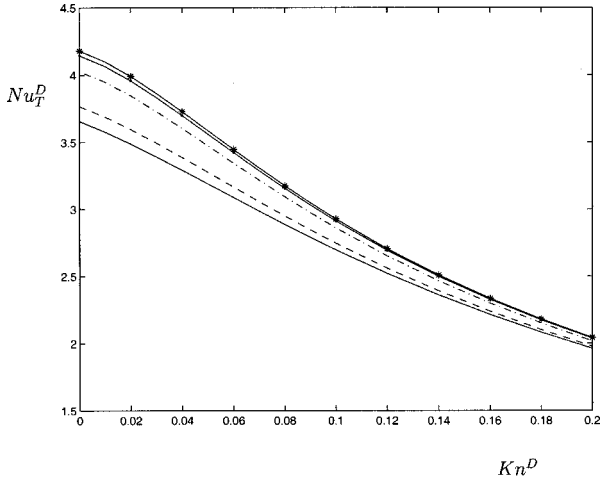


Fig. 8 Variation of Nusselt number Nu_T^D with Knudsen number Kn^D for air in the fully accommodating case. The solid line denotes $Pe^D \to \infty$, the dashed line $Pe^D = 5$, the dash-dotted line $Pe^D = 1$, the line with dots $Pe^D = 0.2$ and the line with stars $Pe^D \to 0$.

sphere gas; as the Knudsen number increases, slip-flow theory overpredicts the Nusselt number. The error (or difference between DSMC and slip-flow theory) is less than 5 percent at Kn≈0.12, and less than 10 percent at Kn≈0.2, indicating that slip-flow theory provides a reasonable approximation beyond Kn=0.1. This is particularly remarkable given that Kn=0.1 corresponds to $Kn' = 0.2$, where $Kn' = \lambda/(H/2)$ is an alternative definition that is frequently used in the literature.

For Kn>0.1, our DSMC results predict that for the same value of the Knudsen number, $Nu_T(T_o > T_i) < Nu_T(T_o < T_i)$. It is possible that this is due to thermal creep flow that provides a ''symmetry breaking'' mechanism since it reverses direction when the temperature gradient along the channel wall reverses direction. Viscous heat generation (including shear work at the boundary) and expansion cooling are also possible sources; they also provide ''symmetry breaking'' mechanisms by introducing a heat flux that is always directed from the fluid to the walls (viscous heat generation) or from the walls to the fluid (expansion cooling), in contrast to the total heat flux that depends on the sign of $T_o - T_i$. As discussed before, the combined effect of viscous heat generation and expansion cooling should be small (Br<0.03), but it is still possible that the *difference* between the two cases (leading to an effect of double magnitude) may be responsible for a discrepancy of the order of 10 percent. The effect is in fact of the order of 5 percent at Kn≈0.4 and increases to approximately 15 percent at Kn≈2. Solution of the continuum equations in the presence of thermal creep is difficult because thermal creep depends on the local axial temperature gradient. However, in the constant-wall-heat-flux case, where the temperature gradient is constant, it is estimated [34] that positive thermal creep ($T_o > T_i$) tends to increase the Nusselt number and negative thermal creep ($T_i > T_o$) decreases the Nusselt number. Here we observe the opposite; however, the above estimate is based on an idealized, plug-flow, thermal creep contribution superimposed on a parabolic (Poiseuille) profile. However, for Kn≫0.1 the velocity profile becomes fairly flat [8]. Additionally, the thermal creep flow profile is steep close to the channel walls in most of the Knudsen regime [22]. This may explain why the opposite trend is observed.

Another possible explanation for the difference between $Nu_T(T_o > T_i)$ and $Nu_T(T_o < T_i)$ is related to the thermal development length that may be longer in the transition regime than the continuum prediction (21). Our DSMC results for Kn>0.4 indicate that the Nusselt number reaches an asymptotic value (within the statistical uncertainty) for $\zeta > 2\zeta_t$ and not $\zeta > \zeta_t$. It is possible that the increased development period is related to thermal creep [33] that decays slowly away from the temperature discontinuity, or due to acceleration effects due to variable fluid temperature despite the fact that at $\zeta = \zeta_t$, $|T_b - T_w|/T_w < 0.02$. Unfortunately, allowing for a larger development length does not resolve this uncertainty because at larger sampling distances than the ones used here, $T_b - T_w$ decays to very small values and the Brinkman number becomes appreciable.

Further work is required to understand the effects of thermal creep and development length and the possible connection between the two. Despite this, our results are still capable of bounding $Nu_T(|T_o - T_i| \to 0)$ to within 15 percent (at most), which is comparable to the statistical uncertainty of these calculations. They also suggest that thermal creep effects in the slip-flow regime are small, thus justifying the neglect of these phenomena in slip-flow analyses.

5.4 Slip-Flow Heat Transfer in Circular Tubes. In this section we present results for slip-flow heat transfer in cylindrical tubes of length L and diameter D. The results presented here include the effects of axial heat conduction that have not been included in previous investigations [5,18,21,34]. The problem description and solution approach is otherwise identical to those in sections 2 and 3 and in the interest of brevity, they will not be included here. More details can be found in [32]. The Reynolds number and Nusselt number for a circular geometry are defined as

$$Re^D = \frac{\rho u_b D}{\mu}, \tag{31}$$

and

$$Nu_T^D = \frac{hD}{\kappa}, \tag{32}$$

respectively. The Knudsen number is defined as $Kn^D = \lambda/D$.

Figure 8 summarizes our results for $Nu_T^D (0 \leqslant Kn^D \leqslant 0.2, 0 < Pe^D < \infty, \sigma_v = \sigma_T = 1)$. In agreement with previous work, for $Kn^D = 0$ the Nusselt number increases as axial heat conduction becomes important (Peclet number decreases). The increase of Nusselt number in the presence of axial heat conduction is less pronounced for non-zero Knudsen numbers: the maximum increase in the Nusselt number occurs at $Kn^D = 0$ and is of the order of 15 percent. In fact, in the regime $Pe^D < 1$ that is of most practical interest, the dependence of Nusselt number on Peclet number

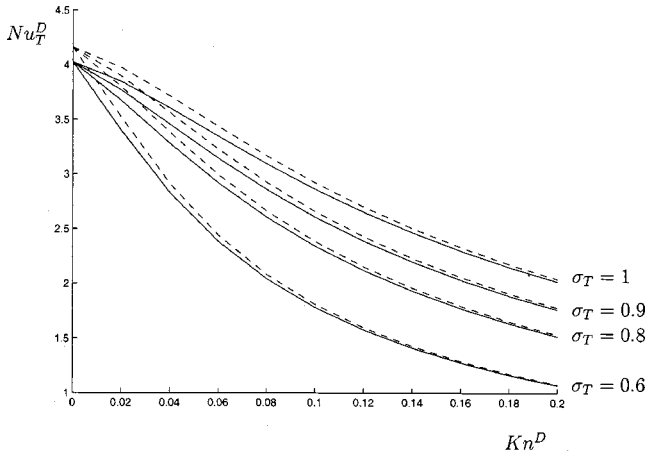


Fig. 9 Variation of Nusselt number Nu_T^D with Knudsen number Kn^D for air for various values of the thermal accommodation coefficient σ_T with σ_v=1. The solid lines denote Pe^D=1, and the dashed lines Pe^D=0.1.

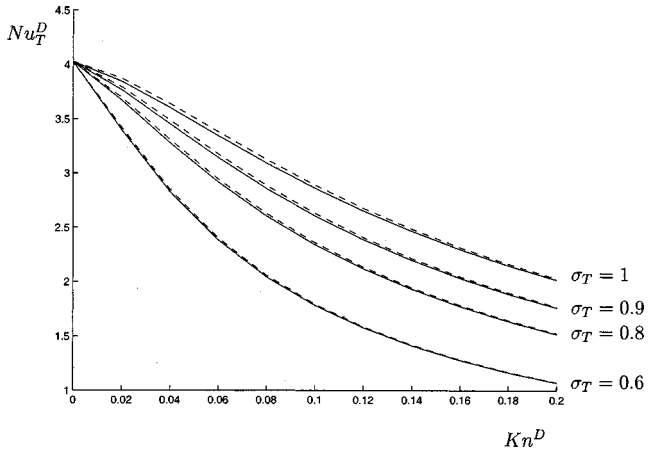


Fig. 10 Variation of Nusselt number Nu_T^D with Knudsen number Kn^D for air for various values of the accommodation coefficients for Pe^D=1. The solid lines denote σ_v=1, and the dashed lines σ_v=0.9.

is weak. The maximum difference between $Nu_T^D(Pe^D=0.2, 0 \leqslant Kn^D \leqslant 0.2)$ and $Nu_T^D(Pe^D=1, 0 \leqslant Kn^D \leqslant 0.2)$ occurs at $Kn^D=0$, and is less than 3 percent. The maximum difference between $Nu_T^D(Pe^D=0.05, 0 \leqslant Kn^D \leqslant 0.2)$ and $Nu_T^D(Pe^D=0.2, 0 \leqslant Kn^D \leqslant 0.2)$ also occurs at $Kn^D=0$ as expected, and is less than 0.7 percent.

Figures 9 and 10 show the variation of the fully developed Nusselt number as a function of Knudsen number for various values of the accommodation coefficients. The qualitative behavior is very similar to the two-dimensional channel case.

6 Conclusions

From the good agreement between the continuum and molecular results in the slip-flow regime and beyond, we conclude that convective heat transfer in low speed flows in the slip-flow regime can be captured by slip-corrected continuum flows that neglect viscous heat dissipation, expansion cooling and thermal creep. The slip coefficients α and β, that are strictly applicable in the presence of a constant gradient in an infinite system, provide good accuracy in channel flows.

The effect of axial heat conduction is to increase the Nusselt number throughout the slip-flow regime. The effect of axial heat conduction decreases as the Knudsen number increases. These findings may explain the weak dependence of Nusselt number on the Reynolds number reported in early experimental work [19].

In the transition regime, the Nusselt number decreases monotonically with increasing Knudsen number for unity momentum and thermal accommodation coefficients. Our transition-regime results, in the non-dimensional form presented here, should be applicable to any dilute monoatomic gas. They are also in qualitative agreement with the previous study that investigated the Nusselt number in the transition regime under constant-wall-heat-flux conditions [13]. The similarity between the Nusselt number and skin friction coefficient dependence on the Knudsen number (see Figs. 2 and 7) suggests that a Reynolds-type analogy may exist. At this time no simple expression for this analogy has been found. The ability of the hard sphere gas model to accurately predict average flow rates of real gases [8] suggests that the skin friction relation (Eq. (26)) should also capture real-gas behavior accurately.

Future work will focus on molecular simulations with accommodation coefficients different from unity. Although our current molecular results suggest that deductions from the slip-flow regime can be qualitatively extended to the transition regime, it would be useful to have exact theoretical predictions. Extensions of this work to polyatomic gases will follow.

Acknowledgments

The authors would like to thank Professors John Lienhard V and Alejandro Garcia for helpful suggestions and discussions, and Professor Triantafyllos Akylas for critically commenting on the manuscript. This work was made possible by the computer resources made available to the authors by the Center for Applied Scientific Computing (CASC) at Lawrence Livermore National Laboratory through the efforts of Dr. Xabier Garaizar. This work was supported in part by Dr. Kyran Mish, Director, Center for Computational Engineering, Lawrence Livermore National Laboratory, U.S. Department of Energy under contract W-7405-ENG-48.

Nomenclature

A = channel cross sectional area (m^2)
Br = modified Brinkman number
Br* = continuum Brinkman number
c_o = most probable (particle) velocity (m/s)
c_p = specific heat at constant pressure (J/Kg K)
c_v = specific heat at constant volume (J/Kg K)
D = tube diameter (m)
H = channel height (m)
h = heat transfer coefficient (W/m^2 K)
Kn = Knudsen number=λ/H
Kn^D = Knudsen number for circular tubes=λ/D
k = Boltzmann's constant (J/K)
L = channel length (m)
$\underline{L}$ = isothermal entry length (m)
m = molecular mass (Kg)
Nu = Nusselt number
Nu_T = constant-wall-temperature Nusselt number
Nu_T^D = constant-wall-temperature Nusselt number for circular tubes
Nu_T^s = slip-flow constant-wall-temperature Nusselt number
P = pressure (Pa)
Pe = Peclet number=RePr
Pe^D = Peclet number for circular tubes=Re^D Pr
P_i = inlet pressure (Pa)
P_o = exit pressure (Pa)
Pr = Prandtl number
$\dot{Q}$ = volume flow rate per unit depth (m^2/s)

$\overline{Q}$ = proportionality constant in volume-flow-rate scaling relation (24)
q = wall heat flux (W/m^2)
R = gas constant=k/m (J/Kg K)
Re = Reynolds number=$\rho u_b 2H/\mu$
Re_L = Re H/L
Re^D = Reynolds number for circular tubes=$\rho u_b D/\mu$
S = flux sampling area (m^2)
T = temperature (K)
T_b = bulk temperature (K)
T_i = inlet temperature (K)
T_o = exit temperature (K)
T_w = wall temperature (K)
t = flux sampling time (s)
$\hat{t}$ = channel wall tangent
$\vec{u}$ = flow velocity (m/s)
u_b = bulk velocity (axial direction) (m/s)
$\vec{V}$ = particle velocity (m/s)
$\vec{V}'$ = post-wall-collision particle velocity (m/s)
x = axial coordinate
y = transverse coordinate
z = out of plane coordinate

Greek Symbols

α = velocity slip coefficient
β = temperature jump coefficient
γ = ratio of specific heats=c_p/c_v
Δx = DSMC cell size (m)
Δt = DSMC timestep (s)
ζ = non-dimensional axial coordinate=$2(x-\overline{L})/H$
ζ_t = non-dimensional thermal development length
η = non-dimensional transverse coordinate=$2y/H$
$\tilde{\eta}$ = coordinate normal to the wall
θ = $T-T_w$ (K)
κ = thermal conductivity (W/m K)
λ = molecular mean free path (m)
μ = coefficient of viscosity (Kg/m s)
Π = pressure ratio=P_i/P_o
ρ = mass density (Kg/m^3)
σ = molecular diameter (m)
σ_T = thermal accommodation coefficient
σ_v = momentum accommodation coefficient
τ_w = wall shear stress ($Kg/m\,s^2$)
ω_i = i-th eigenvalue of equation (18)

References

[1] Alexander, F. J., and Garcia, A. L., 1997, "The Direct Simulation Monte Carlo Method," Comput. Phys., **11**, pp. 588–593.
[2] Alexander, F. J., Garcia, A. L., and Alder, B. J., 1994, "Direct Simulation Monte Carlo for Thin-Film Bearings," Phys. Fluids, **6**, pp. 3854–3860.
[3] Alexander, F. J., Garcia, A. L., and Alder, B. J., 1998, "Cell Size Dependence of Transport Coefficients in Stochastic Particle Algorithms," Phys. Fluids, **10**, pp. 1540–1542.
[4] Arkilic, E., Breuer, K. S., and Schmidt, M. A., 1994, "Gaseous Flow FED-197, in Microchannels," *Application of Microfabrication to Fluid Mechanics*, ASME, New York, pp. 57–66.
[5] Barron, R. F., Wang, X., Ameel, T. A., and Washington, R. O., 1997, "The Graetz Problem Extended to Slip-Flow," Int. J. Heat Mass Transf., **40**, pp. 1817–1823.
[6] van den Berg, H. R., Seldam, C. A., and van der Gulik, P. S., 1993, "Thermal Effects in Compressible Viscous Flow in a Capillary," Int. J. Thermophys., **14**, pp. 865–892.
[7] Beskok, A., and Karniadakis, G. E., 1994, "Simulation of Heat and Momentum Transfer in Complex Microgeometries," J. Thermophys. Heat Transfer, **8**, pp. 647–655.
[8] Beskok, A., and Karniadakis, G. E., 1999, "A Model for Flows in Channels and Ducts at Micro and Nano Scales," Microscale Thermophys. Eng., **3**, pp. 43–77.
[9] Bird, G. A., 1994, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*, Clarendon Press, Oxford.
[10] Cercignani, C., 1988, *The Boltzmann Equation and Its Applications*, Springer-Verlag, New York.
[11] Garcia, A., and Wagner, W., 2000, "Time Step Truncation Error in Direct Simulation Monte Carlo," Phys. Fluids, **12**, pp. 2621–2633.
[12] Grosjean, C. C., Pahor, S., and Strnad, J., 1963, "Heat Transfer in Laminar Flow Through a Gap," Appl. Sci. Res., **11**, pp. 292–294.
[13] Hadjiconstantinou, N. G., 2000, "Convective Heat Transfer in Micro and Nano Channels: Nusselt Number Beyond Slip Flow," *Proceedings of the 2000 IMECE*, HTD-Vol. 366-2, pp. 13–22.
[14] Hadjiconstantinou, N., and Simek, O., 2001, "Nusselt Number in Micro and Nano Channels under Conditions of Constant Wall Temperature," *Proceedings of the 2001 IMECE*.
[15] Hadjiconstantinou, N. G., 2000, "Analysis of Discretization in the Direct Simulation Monte Carlo," Phys. Fluids, **12**, pp. 2634–2638.
[16] Harley, J. C., Huang, Y., Bau, H. H., and Zemel, J. N., 1995, "Gas Flow in Microchannels," J. Fluid Mech., **284**, pp. 257–274.
[17] Ho, C. M., and Tai, Y. C., 1998, "Micro-Electro-Mechanical Systems (MEMS) and Fluid Flows," Annu. Rev. Fluid Mech., **30**, pp. 579–612.
[18] Inman, R. M., 1964, "Heat Transfer for Laminar Slip Flow of a Rarefied Gas in a Parallel Plate Channel or a Circular Tube With Uniform Wall Temperature," NASA TN D-2213.
[19] Kavehpour, H. P., Faghri, M., and Asako, Y., 1997, "Effects of Compressibility and Rarefaction on Gaseous Flows in Microchannels," Numer. Heat Transfer, Part A, **32**, pp. 677–696.
[20] Knudsen, M., 1909, "Die Gesetze der molecular Stromung und dieinneren Reibungstromung der Gase durch Rohren," Annals of Physics, **28**, pp. 75–130.
[21] Larrode, F. E., Housiadas, C., and Drossinos, Y., 2000, "Slip-Flow Heat Transfer in Circular Tubes," Int. J. Heat Mass Transf., **43**, pp. 2669–2690.
[22] Loyalka, S. K., 1975, "Kinetic Theory of Thermal Transpiration and Mechanocaloric Effect. II.," J. Chem. Phys., **63**, pp. 4054–4060.
[23] Loyalka, S. K., 1989, "Temperature Jump and Thermal Creep Slip: Rigid Sphere Gas," Phys. Fluids, **1**, pp. 403–408.
[24] Mills, A. F., 1992, *Heat Transfer*, Irwin.
[25] Morris, D. L., Hannon, L., and Garcia, A. L., 1992, "Slip Length in a Dilute Gas," Phys. Rev. A, **46**, pp. 5279–5281.
[26] Nagayama, K., Farouk, B., and Oh, C. K., 1998, "Heat Transfer in Low Pressure (High Knudsen Number) Developing Flows Through Parallel Plates," *Heat Transfer 1998: Proceedings of the Eleventh International Heat Transfer Conference*, Vol. 3, pp. 127–132.
[27] Ohwada, T., Sone, Y., and Aoki, K., 1989, "Numerical Analysis of the Poiseuille and Thermal Transpiration Flows Between Parallel Plates on the Basis of the Boltzmann Equation for Hard-Sphere Molecules," Phys. Fluids, **1**, pp. 2042–2049.
[28] Oran, E. S., Oh, C. K., and Cybyk, B. Z., 1999, "Direct Simulation Monte Carlo: Recent Advances and Applications," Annu. Rev. Fluid Mech., **30**, pp. 403–431.
[29] Pahor, S., and Strnad, J., 1961, "A Note on Heat Transfer in Laminar Flow Through a Gap," Appl. Sci. Res., **10**, pp. 81–84.
[30] Piekos, E. S., and Breuer, K. S., 1996, "Numerical Modeling of Micromechanical Devices Using the Direct Simulation Monte Carlo Method," ASME J. Fluids Eng., **118**, pp. 464–469.
[31] Shah, R. K., and London, A. L., 1978, *Laminar Flow Forced Convection in Ducts*, Academic Press.
[32] Simek, O., and Hadjiconstantinou, N. G., 2001, "Slip-Flow Constant-Wall-Temperature Nusselt Number in Circular Tubes in the Presence of Axial Heat Conduction," *Proceedings of the 2001 IMECE*.
[33] Sone, Y., 2000, "Flows Induced by Temperature Fields in a Rarefied Gas and Their Ghost Effect on the Behavior of a Gas in Continuum Limit," Annu. Rev. Fluid Mech., **32**, pp. 779–811.
[34] Sparrow, E. M., and Lin, S. H., 1962, "Laminar Heat Transfer in Tubes Under Slip Flow Conditions," ASME J. Heat Transfer, **84**, pp. 363–369.
[35] Sun, H., and Faghri, M., 2000, "Effects of Rarefaction and Compressibility of Gaseous Flow in Microchannel Using DSMC," Numer. Heat Transfer, Part A, **38**, pp. 153–168.
[36] Wagner, W., 1992, "A Convergence Proof for Bird's Direct Simulation Monte Carlo Method for the Boltzmann Equation," J. Stat. Phys., **66**, pp. 1011–1044.
[37] Wijesinghe, S., and Hadjiconstantinou, N. G., 2001, "Velocity Slip and Temperature Jump in Dilute Hard Sphere Gases at Finite Knudsen Numbers," *Proceedings of the First MIT Conference on Computational Fluid and Solid Mechanics*, Vol. 2, pp. 1019–1021.

Effect of Microscale Mass Transport and Phase Change on Numerical Prediction of Freezing in Biological Tissues

Ramachandra V. Devireddy[1]
Materials Research Science and Engineering Center, Department of Chemical Engineering, University of Minnesota, Minneapolis, MN 55455

David J. Smith

John C. Bischof

Bioheat and Mass Transfer Laboratory, Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN 55455

A numerical model incorporating the microscale heat and mass transport in biological tissue during freezing is developed. The heat transfer problem is formulated in a general one-dimensional coordinate system (cartesian, cylindrical or spherical), and a finite control volume discretization is used. The latent heat release for each control volume in the domain is determined by the cellular water transport and intracellular ice formation processes occurring there (a coupled thermal/biophysical approach). The coupled model is applied to two cryobiological freezing problems, with different geometry and boundary conditions. The temperature dependent thermal properties of water and the biophysical properties of two biological tissues, normal rat liver and Dunning AT-1 rat prostate tumor tissue are used to simulate both the micro and macroscale freezing processes. A major advantage of the coupled thermal/biophysical model is its unique ability to predict both the macroscale thermal response and the microscale biophysical response at various locations within the tissue domain during a freezing process, simultaneously. Thermal histories predicted by the coupled model are compared to predictions of a standard enthalpy-method model in which the temperature dependence of the latent heat release, $\Lambda(T)$ is an explicit function adapted from the water-NaCl phase diagram, and phase change is not rate-limited by microscale biophysical processes (i.e., an uncoupled approach). The results for both models are very similar; this suggests that the microscale biophysical processes which occur in the chosen biological tissues during freezing do little to limit the rate at which phase change occurs. Additional simulations suggest that the predicted macroscale thermal history results are not significantly affected (<2 percent variation) even with significantly altered biophysical parameters (i.e., a factor of 100 times lower or higher), as long as the magnitude of the latent heat is constant.
[DOI: 10.1115/1.1445134]

Keywords: Bioengineering, Heat Transfer, Low Temperature, Microscale, Phase Change

Introduction

The process of freezing biological tissue continues to be actively studied because of its central role in two important biomedical applications, cryosurgery, and cryopreservation. Cryosurgery has been in clinical use for years as a method for tumor ablation, see recent review by Gage and Baust [1], and tissue cryopreservation has the potential to become a long-term storage method for donated tissues and organs and tissue-engineered equivalents [2]. Widespread clinical use of these techniques will require further understanding of the relationship between the freezing process and tissue survival. The cellular-level biophysical events which occur in tissue as it freezes (i.e., cell dehydration and intracellular ice formation) have long been recognized as a mechanistic way to explain cell injury as a result of the freezing process [3]. The dynamics of both the thermal conditions inside frozen tissue and the cellular-level biophysical events must be studied to understand the tissue freezing problem. The two important biophysical responses experienced by embedded tissue cells in the presence of vascular/extracellular ice during freezing are water transport (or mass transfer) out of the tissue cells (i.e., cellular dehydration) and intracellular ice formation (IIF). These biophysical responses are directly coupled to tissue injury as described by Mazur's two factor hypothesis: a) at ''low'' cooling rates freezing injury occurs due to high solute concentration effects caused by cellular dehydration and decreasing unfrozen fraction of the vascular/extracellular space and b) at ''high'' cooling rates freezing injury occurs due to IIF [3]. The exact magnitude of the ''low'' and ''high'' cooling rates are specific to a cell type and is dependent on the biophysical (water transport and IIF) parameters.

Numerical modeling has proven to be extremely useful in studying tissue freezing problems, as evidenced by the volume of available literature [4–7]. Most models of tissue thermal response assume the tissue to be a single-compartment, homogeneous medium, e.g., pure water or isotonic saline. In this case, phase change is described by the phase diagram for the chosen medium. Mathematically, latent heat content is assumed to be a function of temperature $\Lambda(T)$, and the energy equation is solved in the enthalpy form [8–10] or with a modified specific heat which incorporates this function [11,12]. In native biological tissues, a significant portion of the water in tissue is intracellular, separated from the surrounding tissue by a cell membrane. Therefore, phase change in tissue is governed not only by the phase diagram for tissue fluids but also by the dynamics of the biophysical processes of cellular dehydration and intracellular ice formation (IIF). Rubinsky and Pegg [13] proposed a mathematical model for tissue freezing in which the thermal model used a cell dehydration model and the phase diagram for isotonic saline to govern the rate of phase change. Bischof and Rubinsky [6] added a model for IIF,

[1]Current affiliation: Department of Mechanical Engineering, Louisiana State University, Baton Rouge, LA 70803

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division May 4, 2001; revision received October 22, 2001. Associate Editor: A. Majumdar.

and solved the coupled thermal and biophysical model numerically for a one-dimensional cartesian coordinate system. The model developed by Bischof and Rubinsky [6] was somewhat limited by the requirement that the spatial grid spacing be equal to a tissue cell diameter, and that extension to cylindrical, or spherical coordinates was not possible. Additionally, the front tracking scheme employed by Bischof and Rubinsky [6] is not as accurate as source term or enthalpy formulations used in the current study. And finally, an improved and mechanistic model of IIF developed by Toner et al. [14] is incorporated in the current study.

In the present work, the coupling of the biophysical phenomena into the calculation of phase change is accomplished by creating a latent heat function similar to those used in enthalpy or modified specific heat formulations, but now this is a function of both temperature and time, $\Lambda(T,t)$. This is in direct contrast to previous studies where the latent heat content is assumed to be a function of temperature $\Lambda(T)$, only. The coupled heat transfer and biophysics problem is then set up for solution in a generalized one-dimensional coordinate system (Cartesian, cylindrical or spherical coordinates) using an iterative source term calculation technique [11]. The freezing of biological tissue is solved for two tissue freezing problems of cryobiological relevance; one is slow cooling (5 °C/min in Cartesian coordinates); and the other is fast cooling (50 °C/min in cylindrical coordinates). Cooling was applied to the internal boundary with a fixed temperature boundary condition on the opposing or outer boundary. In order to distinguish these cases in the text, the slow cooling case will be termed a ''cryopreservation'' case while the fast cooling case will be termed a ''cryosurgical'' case. However, it should be noted that the solutions presented here are not necessarily typical of either cryopreservation or cryosurgery. The terminology is used here simply to compare the results from the coupled microscale freezing model with the solution of the enthalpy model under the same boundary conditions.

Governing Equations

Heat transfer during phase change in a general one-dimensional Cartesian, cylindrical or spherical domain can be described by the following governing equation:

$$\frac{\partial}{\partial t}(\rho c T+\rho L\Lambda)=\frac{1}{r^n}\frac{\partial}{\partial r}\left(r^n k\frac{\partial T}{\partial r}\right), \tag{1}$$

where T is temperature, t is time, r is a spatial coordinate, ρ is density, c is specific heat, L is latent heat of freezing per unit mass of pure water, and k is thermal conductivity. Λ is the mass fraction of the medium which has not yet released latent heat during freezing; it is a function of distance from the cooling boundary, temperature and time. The coordinate system is determined by the geometric parameter n: $n=0$ for cartesian, 1 for cylindrical and 2 for spherical coordinates. Similar formulations have been used by previous researchers to solve phase change problems with biomedical or organic applications [11,12,15]. Note that Eq. (1) represents a lumped energy balance over both the cellular and extracellular constituents within a differential volume, with the thermophysical properties determined by the volume fraction of ice and water within the differential volume [6]. In biological tissue, the function Λ is composed of contributions from the extracellular water (Λ_{ec}) and the intracellular water (Λ_{ic}):

$$\Lambda=f_{ec}\Lambda_{ec}+f_{ic}\Lambda_{ic}, \tag{2}$$

where f_{ec} and f_{ic} are the fractional masses of extracellular and intracellular water in the medium, respectively. The fractional mass is the mass of water divided by the total mass of medium. Assuming that biological tissue has a density equal to water, the fractional masses f_{ec} and f_{ic} can be approximated by the volume fractions of the extra and intracellular water respectively; the quantities f_{ec} and f_{ic} will therefore be referred to as fractional volumes in the remainder of the paper, since this is how they are calculated. In general, because of the presence of non-water components in biological tissue, the sum $f_{ec}+f_{ic}$ will be <1. The extracellular latent heat fraction Λ_{ec} is defined using the binary phase diagram for a 150 mM NaCl aqueous solution, as given in by Pitt [16]:

$$\Lambda_{ec}(T)=\begin{cases}1 & T\geqslant 272.62K\\ \dfrac{0.53}{273.15-T} & T<272.62K\end{cases}. \tag{3}$$

The intracellular latent heat fraction Λ_{ic} is an integral function of cell volume (V) and probability of intracellular ice formation (PIF):

$$\Lambda_{ic}(V(T,t),\mathrm{PIF}(T,t))=\int_{V_o}^{V}(1-\mathrm{PIF})\frac{dV}{V_o-V_b}-\int_0^{\mathrm{PIF}}\left(\frac{V-V_b}{V_o-V_b}\right)d\,\mathrm{PIF}, \tag{4}$$

where V_o is the initial cell volume (before freezing begins), and V_b is the volume of osmotically-inactive (i.e., non-water) cell components. The first integral represents the contribution of intracellular water which freezes after being transported through the cell membrane to the extracellular space, and the second integral represents the contribution of water which freezes intracellularly. In order to evaluate these two integrals, both V and PIF must be determined during the freezing process. The model for cell volume V is presented by Mazur [17] as:

$$\frac{dV(T,t)}{dt}=-\frac{L_p\mathrm{ART}}{\nu_w}\left(\ln\left[\frac{(V-V_b)}{(V-V_b)+\phi\nu_w n_s}\right]-\frac{L}{R}\left[\frac{1}{T_r}-\frac{1}{T}\right]\right), \tag{5}$$

where $T<272.62$ K (or -0.53°C), L_p is the cell membrane hydraulic permeability, A is the surface area of the cell available for water transport, R is the universal gas constant (8.314 J/mol-K), T is temperature, v_w is the molar volume of water ($18.0(10^{-6})$ m^3/mol), ϕ is the salt dissociation constant (assumed equal to 2), n_s is the moles of salt in the cell, and T_r is the reference temperature, 273.15 K. The permeability L_p is temperature dependent, according to an Arrhenius relationship [18]:

$$L_p=L_{pg}\exp\left[\frac{-E_{Lp}}{R}\left(\frac{1}{T}-\frac{1}{T_r}\right)\right], \tag{6}$$

where L_{pg} is the hydraulic permeability at the reference temperature T_r, and E_{Lp} is the activation energy for the water transport process. The probability of intracellular ice formation is determined using the model proposed by Toner [19]

$$\mathrm{PIF}(T,t)=1-\exp\left[-\int_{T_{\mathrm{seed}}}^{T}A\Omega_o\left(\frac{T}{T_{f0}}\right)^{1/2}\frac{\eta_0}{\eta}\frac{A}{A_o}\times\exp\left[\frac{-\kappa_o(T_f/T_{fo})^{1/4}}{(T-T_f)^2T^3}\right]dt\right], \tag{7}$$

where $T<272.62$ K (or -0.53°C), T_{seed} is the temperature at which ice is seeded in the extracellular space (i.e., the phase change temperature), T_f is the equilibrium phase change temperature of the intracellular water, η is the intracellular solution viscosity, and κ and Ω are cell type-dependent constants contributing to the thermodynamic and kinetic components of crystal nucleation on the cell membrane surface, respectively. The subscript ''o'' on any parameter denotes the value of that parameter under isotonic conditions. The viscosity η is approximated using the following expression [20]:

$$\eta=0.139(10^{-3})\left[\frac{T}{225}-1\right]^{-1.64}\frac{kg}{m\cdot s}. \tag{8}$$

Equations (1) to (8), if all the constants are given appropriate values, fully define the governing equations for heat transfer and

phase change during freezing in a biological tissue. These equations can be used to solve for the thermal and biophysical response during freezing in any one-dimensional spatial tissue domain with properly defined initial and boundary conditions.

Numerical Solution

The discretization and general algorithm for numerical solution of the heat transfer problem (Eq. 1) was performed as described in a previous study of freezing in bulk saline solutions [21] which used explicit definitions of Λ (e.g., Eq. 3 for Λ instead of Λ_{ec}). Eq. (1) is discretized using the finite control volume method [22], with a spatial grid spacing of $\Delta r = 0.1$ mm in the cryopreservation case study and 0.02 mm in the cryosurgery case study (the two case studies are described in more detail in the Simulations section). First-order implicit discretization in time is used to promote numerical stability, with a temporal step size of $\Delta t = 0.1$ sec. Phase change is handled using an iterative procedure described by Hayes and Diller [11]. The phase change term $\rho L(\partial\Lambda/\partial t)$ is treated as a volumetric heat source term in the governing equation; the proper value of the source term at each time step is determined iteratively. Briefly, a guess is made of the value of $\partial\Lambda/\partial t$ for each control volume in which phase change is occurring (the "mushy zone"), and the heat transfer equation is solved for the temperatures in the entire domain. The "mushy zone" is defined here as the set of control volumes in which T is less than the phase change temperature of isotonic saline (272.62 K). Based on the change in temperature, a value of $\partial\Lambda/\partial t$ can be calculated using the change in the known explicit function of Λ over one time step. The guessed value of $\partial\Lambda/\partial t$ for each control volume is adjusted and the process repeated until the guessed and calculated values of $\partial\Lambda/\partial t$ match within a small tolerance (± 0.1 percent). This procedure is shown schematically in the flow chart shown in Fig. 1. Smith et al. [21] benchmarked a numerical model using this algorithm against analytical and numerical solutions to phase change problems in the literature.

One advantage of the iterative source term calculation procedure is that more complicated definitions of Λ such as the integral form described above for biological tissue, Eqs. (2) to (8), can be used, as long as a value of $\partial\Lambda/\partial t$ can be calculated for each control volume in the mushy zone. Consider any control volume in the mushy zone at time t, for which the temperature $T(t)$, cell volume $V(t)$ and probability of intracellular ice formation PIF(t) are known. A guess is made for the value of $\partial\Lambda/\partial t$, and the heat transfer equation is solved for the entire domain; now $T(t+\Delta t)$ is known. $V(t+\Delta t)$ is calculated by integration of Eqs. (5) to (6) using a fourth-order Runge Kutta algorithm, and PIF$(t+\Delta t)$ is calculated by Simpson's rule integration of Eqs. (7) to (8) over Δt. $\partial\Lambda/\partial t$ is approximated over Δt by $\Delta\Lambda/\Delta t$, where $\Delta\Lambda$ is represented as $\Lambda(T(t+\Delta t),t+\Delta t)-\Lambda(T(t),t)$ and is calculated from Eq. (2):

$$\Delta\Lambda = f_{ec}\Delta\Lambda_{ec} + f_{ic}\Delta\Lambda_{ic}. \tag{9}$$

$\Delta\Lambda_{ec}$ is calculated from Eq. (3):

$$\Delta\Lambda_{ec} = \frac{0.53}{273.15 - T(t+\Delta t)} - \frac{0.53}{273.15 - T(t)}. \tag{10}$$

$\Delta\Lambda_{ic}$ is calculated from Eq. (4):

$$\Delta\Lambda_{ic} = \int_{V_o}^{V(t+\Delta t)} (1-\mathrm{PIF})\frac{dV}{V_o - V_b} - \int_{V_o}^{V(t)} (1-\mathrm{PIF})\frac{dV}{V_o - V_b} - \int_0^{\mathrm{PIF}(t+\Delta t)} \left(\frac{V - V_b}{V_o - V_b}\right) d\mathrm{PIF} + \int_0^{\mathrm{PIF}(t)} \left(\frac{V - V_b}{V_o - V_b}\right) d\mathrm{PIF} \tag{11}$$

or,

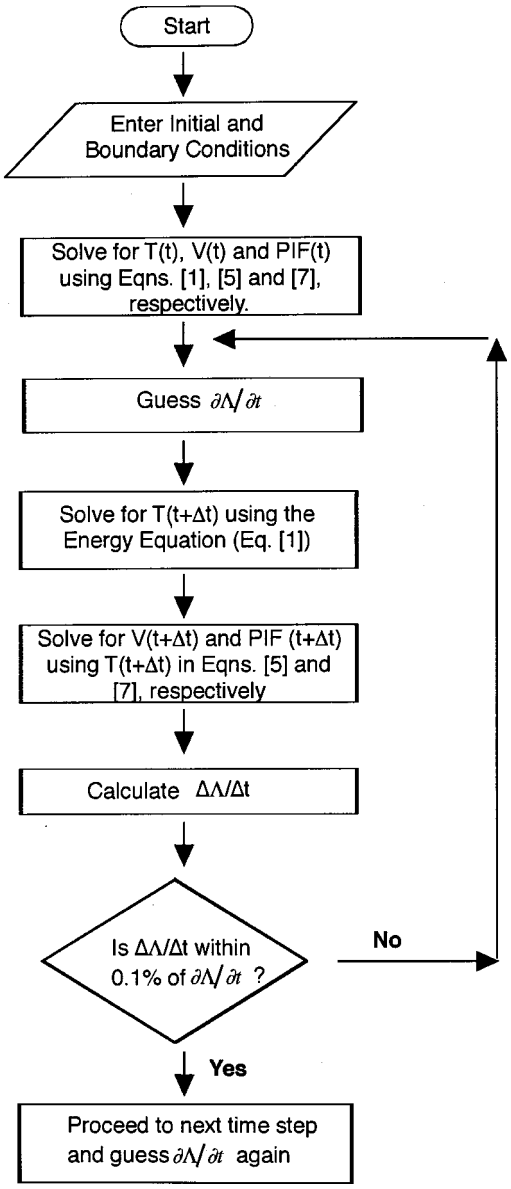


Fig. 1 A schematic depiction (flow chart) of the source term matching algorithm used in the coupled thermal/biophysical model

$$\Delta\Lambda_{ic} = \int_{V(t)}^{V(t+\Delta t)} (1-\mathrm{PIF})\frac{dV}{V_o - V_b} - \int_{\mathrm{PIF}(t)}^{\mathrm{PIF}(t+\Delta t)} \left(\frac{V - V_b}{V_o - V_b}\right) d\mathrm{PIF} \tag{12}$$

To calculate these integrals, it is assumed that the water transport process occurs first between t and $t+\Delta t$, then PIF changes from PIF(t) to PIF$(t+\Delta t)$ instantaneously at $t+\Delta t$. This assumption results in the following expression for $\Delta\Lambda_{ic}$:

$$\Delta\Lambda_{ic} = \int_{V(t)}^{V(t+\Delta t)} (1-\mathrm{PIF}(t))\frac{dV}{V_o - V_b} - \int_{\mathrm{PIF}(t)}^{\mathrm{PIF}(t+\Delta t)} \left(\frac{V(t) - V_b}{V_o - V_b}\right) d\mathrm{PIF} \tag{13}$$

or

$$\Delta\Lambda_{ic} = (1 - \mathrm{PIF}(t)) \frac{V(t+\Delta t) - V(t)}{V_o - V_b} - \left(\frac{V(t) - V_b}{V_o - V_b} \right) (\mathrm{PIF}(t+\Delta t) - \mathrm{PIF}(t)). \quad (14)$$

The value of $\Delta\Lambda/\Delta t$ calculated using Eqs. (9) to (14) is compared to the guessed value of the source term used in the energy equation, and a new value of $\partial\Lambda/\partial t$ is guessed for use in the heat equation solution in the next iteration. The new value guessed for $\partial\Lambda/\partial t$ is determined by adding a fraction of the difference between calculated and guessed values of $\partial\Lambda/\partial t$ to the previous guess. This process continues until the calculated value of $\Delta\Lambda/\Delta t$ matches the guessed value of $\partial\Lambda/\partial t$ for each control volume within 0.1 percent.

Model Parameters

Several thermophysical, biophysical and geometric constants and parameters must be specified to use Eqs. (1) to (14). All of the necessary constants and parameters are summarized in Tables 1–4. The thermophysical properties (k, ρ, and c) of the medium are assumed to be those of water (Table 1). Although, this assumption is not accurate for most biological tissues, the use of common thermophysical properties in both the coupled and uncoupled models facilitates an easy comparison of the results between the models, which is the main focus of the present study. The remaining constants shown in Table 1, include the reference temperature for the Arrhenius expression of L_p (T_r), the universal gas constant (R), the molar volume of water (v_w), the number of moles of salt in a single cell (n_s, calculated from the isotonic concentration and the cell volume) and the salt dissociation constant (ϕ, assumed equal to 2).

The structural and biophysical characteristics of Sprague-Dawley rat liver and Dunning AT-1 prostate tumor tissue were used for the tissue medium because they have been quantitatively determined and published in the literature by Pazhayannur and Bischof [23] and Devireddy et al. [24], respectively. The Krogh cylinder is used to describe the cellular-level structure of the tissue (Fig. 2). The Krogh model is defined by the sinusoid (vascular) radius (r_v), the distance between adjacent sinusoid centers (ΔX) and the axial length of the cylinder (1). The cellular space with volume V is modeled as the box surrounding the cylinder ($V = l\Delta X^2 - \pi r_v{}^2 l$) while the extracellular/vascular area is denoted by the inner cylinder. The effective membrane surface area available for water transport during the freezing process is assumed to be a constant, $A_c = 2\pi r_{vo} L$, where r_{vo} is the initial sinusoid radius. The values for each cylinder dimension ΔX, r_{vo}, and l were obtained by stereological analysis, as reported in literature (Table 2). Figure 2 also depicts schematically the different biophysical responses that occur in the tissue (light micrographs) and in the freezing domain (series of Krogh cylinders). A more detailed description and analysis of the light micrographs for rat liver tissue and AT-1 tumor tissue is provided by Pazhayannur and Bischof [23] and Devireddy et al. [24], respectively. In addition, the osmotically inactive cell volumes have also been reported (Table 2). The parameter bf_{ec} is the fraction of the extracellular medium which is either dissolved solids (salts, proteins, other biomolecules) or water which is "bound" to those solids and therefore unable to change phase and was chosen to be 0.6, based on latent heat measurements obtained in our laboratory [25].

With the tissue geometry defined, the fractional volumes f_{ec} and f_{ic} can be calculated. The formulas for each are shown in Table 2. The value of f_{ec} is much smaller than that of f_{ic} because of the small extracellular space. The sum of f_{ec} and f_{ic} is the

Table 1 Thermophysical and other constants in Eqs. (1) to (8)

Thermophysical Constant	Value and Units
k[1]	0.6 W/m-K (unfrozen) 2.24+0.005975(273-T)$^{1.156}$ W/m-K (frozen)
c[1]	4200 J/kg-K (unfrozen) 7.16T+138 J/kg-K (frozen)
ρ	999 kg/m^3 (unfrozen) 921 kg/m^3 (frozen)
L	335 J/g
T_r	273.15 K
R	8.314 J/mol-K
v_w	18.1(10^{-6}) m^3/mol
n_s	(0.15 mol/m^3) • V_o
ϕ	2

[1]Taken from Alexiades and Solomon [9].

Table 2 Krogh cylinder dimensions and non-water fractions for rat liver and AT-2 tumor tissue

Dimension or Parameter	Value and Units	
	Rat Liver[1]	AT-1 Tumor[2]
ΔX	22 μm	17.9 μm
r_{vo}	3.8 μm	6.5 μm
l	11.4 μm	18.1 μm
V_b/V_o	0.35	0.25
bf_{ec}	0.6	0.6
$V_o = (\Delta X^2 - \pi r_{vo}^2) \cdot l$	5000 μm^3	3400 μm^3
$f_{ec} = (1 - bf_{ec}) \frac{\pi r_{vo}^2}{\Delta X^2}$	0.0375	0.166
$f_{ic} = \frac{(V_o - V_b)}{\Delta X^2 l}$	0.589	0.44

[1]Pazhayannur and Bischof [23]; [2]Devireddy et al. [27];

Table 3 Biophysical parameters for rat liver and AT-1 tumor tissue

Biophysical Parameters	Value and Units	
	Rat Liver Tissue	AT-1 Tumor Tissue
L_{pg}[1,2]	3.0 (10^{-13}) m^3/N-s	0.5 (10^{-13}) m^3/N-s
E_{Lp}[1,2]	265 kJ/mol	96.3 kJ/mol
Ω_o[3,4]	85 (10^8) 1/m^{2}s	857 (10^8) 1/m^{2}s
κ_o[3,4]	2.3 (10^9) K^5	4.19 (10^9) K^5

[1]Pazhayannur and Bischof [23]; [2]Devireddy et al. [27];
[3]Schulte [28]; [4]Smith et al. [29];

Table 4 Problem definitions for cryopreservation and cryosurgery test cases

	Cryopreservation Case	Cryosurgery Case
Coordinate System	1D Cartesian	1D Cylindrical
Domain Size	8 cm	8 cm
Cooling Boundary Location	r = 0 cm	r = 0.15 cm
Spatial Grid Resolution	$\Delta r = 0.1$ mm	$\Delta r = 0.02$ mm
Time Step Size	$\Delta t = 0.1$ sec	$\Delta t = 0.1$ sec
Cooling Boundary Condition	$T(r=0,t)=\begin{cases}10°C-(5°C/\min)\cdot t & t\le 18\min\\ -80°C & t>18\min\end{cases}$	$T(r=0,t)=\begin{cases}37°C-(50°C/\min)\cdot t & t\le 3.9\min\\ -160°C & t>3.9\min\end{cases}$
Condition at r --> ∞	$T(r=8cm,t)=10°C$	$T(r=8cm,t)=37°C$
Initial Condition	$T(r,t=0)=10°C$	$T(r,t=0)=37°C$

fraction of the tissue volume which is water capable of phase change. With the assumption that biological tissue has the same density as water, the actual latent heat per gram of tissue is equal to the latent heat per gram of pure water multiplied by the fraction of tissue which is water, i.e.,

$$\text{Latent heat (rat liver)} = (f_{ec}+f_{ic})*335 \text{ J/g, or } 210 \text{ J/g}$$
$$\text{and Latent heat(AT}-1\text{ tumor)} = (f_{ec}+f_{ic})*335 \text{ J/g, or } 203 \text{ J/g} \quad (15)$$

These values lies within the wide range of latent heat values for mammalian tissues reported in the food science literature, e.g., 120 J/g for pork, 210 J/g for beef and 250 J/g for poultry [26]. The remaining parameters and constants listed in Table 3 are required in the biophysical models of water transport and IIF, Eqs. (5) to (7). The water permeability parameters L_{pg} and E_{Lp} were measured in Sprague-Dawley rat liver and AT-1 tumor using differential scanning calorimetry by Devireddy and Bischof [24] and Devireddy et al. [27], respectively. The IIF parameters Ω_o and κ_o have not been measured in either rat liver tissue or AT-1 tumor tissue, but have been measured in isolated rat (Sprague-Dawley) hepatocytes [28]; (note that IIF parameters on other rat hepatocytes have been reported in the literature, 20) and in isolated AT-1 cells [29]; these parameters were used in Eq. (7). With the constants and parameters specified in Tables 1–3, Eqs. (1) to (8) can be solved to determine the thermal history in a biological tissue domain with defined initial and boundary conditions.

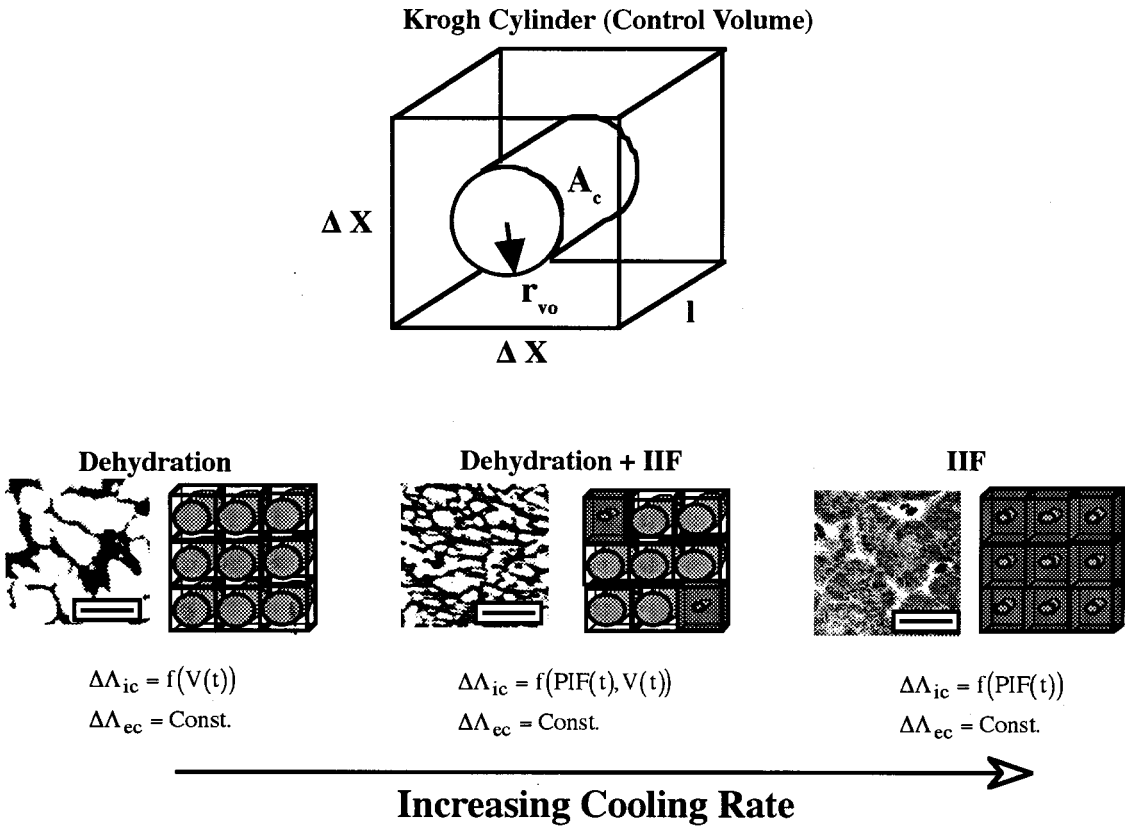


Fig. 2 The Krogh cylinder used to describe cellular-level geometry in the tissue medium. The values of the cylinder dimensions ΔX, r_{vo}, and l used in the solution of Eqs. (1) to (14) are shown in Table 2 for both the rat liver and AT-1 tumor tissue system. The lower part of the figure schematically depicts the biophysical processes (dehydration and intracellular ice formation, IIF) that occur during freezing inside the tissue as a function of cooling rate. In the light micrographs shown the optically dense or "dark" spaces represent the "cell space" while the optically transparent or "white" spaces represent "ice" in the tissue. The scale bar in the micrographs represents 20 μm.

Simulations

Two cryobiologically relevant case studies are considered to determine if incorporation of the phase-change-limiting effects of cell biophysical response during freezing will have significant impact on the predicted thermal history in biological tissue. The two cases are designed to be analogous to cryopreservation and cryosurgery applications. In both cases, a one-dimensional domain is considered with a constant cooling rate imposed at one boundary and a constant temperature assumed for regions far from the cooling surface. The initial and boundary conditions for each case are specified in Table 4. The thermal history is predicted for both cases using two solution methods: (1) numerical solution of Eqs. (1) to (8) in either rat liver tissue or AT-1 tumor tissue using the algorithm described above in the Numerical Solution section; and (2) numerical solution of Eq. (1) using the enthalpy method algorithm [10], a standard technique for solving phase change problems in a homogeneous medium. For method 2, Λ is defined using Eq. (3) with Λ substituted for Λ_{ec}, and L is assumed to be 210 J/g or 203 J/g, the equivalent latent heat per gram of rat liver tissue or AT-1 tumor tissue, respectively (see Eq. 15). Fortran codes using each solution method were developed, and simulations were performed on Sun or SGI workstations in the UNIX operating system. The sensitivity of the coupled model solution was tested by decreasing the time step to 0.05 sec from 0.1 sec and the cryosurgical and cryopreservation simulations were repeated. This test had no effect on the results, thus confirming the convergence of our solution.

Benchmarking the coupled model results for both thermal and biophysical response of a biological tissue during freezing is not possible due to lack of quantitative data for such a case in the literature, but simulations were performed to insure that the coupled model-predicted response for tissue frozen at constant cooling rate was consistent with explicit models for this case, Eqs. (5) to (8). The benchmark case was defined using the cryopreservation case characteristics (Table 2) with a smaller computational domain (1 cm) and a smaller spatial grid resolution (0.02 mm). The coupled model was run with several constant cooling rates at the boundary ($B=1$, 2, 5, 10, 20, 50 and 100°C/min), such that the control volume immediately adjacent to the boundary, essentially followed the same thermal response, i.e., constant rate cooling. The biophysical response ($V(t)$ and PIF(t)) calculated for the control volume immediately adjacent to the boundary was com-

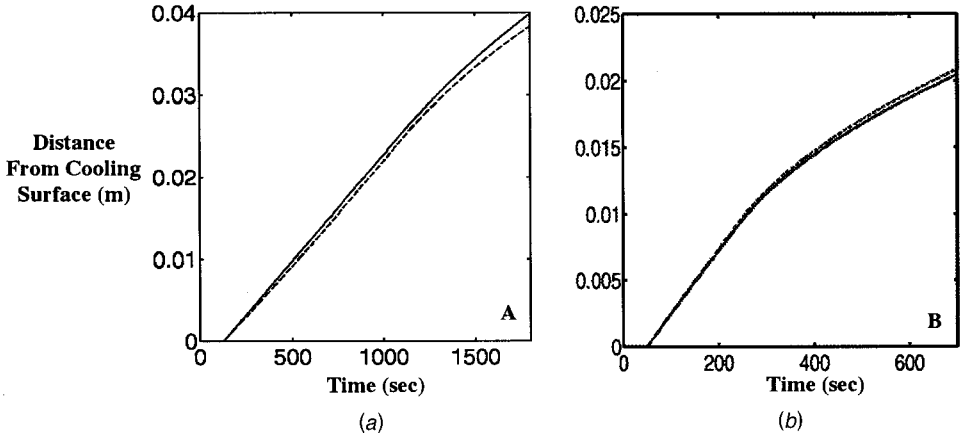


Fig. 3 Predicted interface propagation in rat liver tissue under (*a*) cryopreservation and (*b*) cryosurgery conditions. The cooling conditions are described in Table 4. Microscale freezing model predictions are given by the dotted line (- - - -), and enthalpy method model predictions are given by the solid line (—). Time (secs) is shown on the *x*-axis while the distance from the cooling surface (*m*) is shown on the *y*-axis.

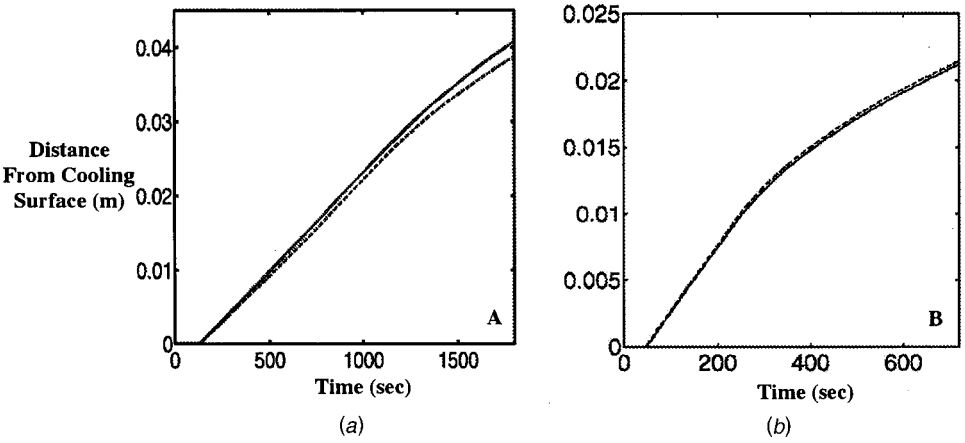


Fig. 4 Predicted interface propagation in AT-1 tumor tissue under (*a*) cryopreservation and (*b*) cryosurgery conditions. The cooling conditions are described in Table 4. Microscale freezing model predictions are given by the dotted line (- - - -), and enthalpy method model predictions are given by the solid line (——). Time (secs) is shown on the *x*-axis while the distance from the cooling surface (*m*) is shown on the *y*-axis.

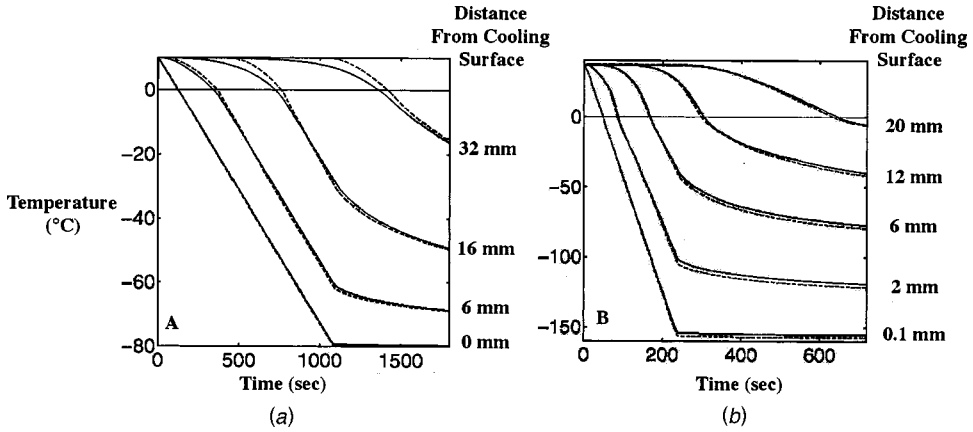


Fig. 5 Predicted thermal history at various locations (as indicated) under (*a*) cryopreservation and (*b*) cryosurgery conditions (the results are essentially identical for both rat liver and AT-1 tumor tissue). The cooling conditions are described in Table 4. Microscale freezing model predictions are given by the dotted line (- - - -), and enthalpy method model predictions are given by the solid line (——). Time (secs) is shown on the *x*-axis while the temperature (°C) is shown on the *y*-axis.

pared to explicit calculations of V and PIF obtained by integrating Eqs. (5) to (8) directly for constant cooling rate, i.e., $T = 272.62\ (\mathrm{K}) - B \cdot t$. The results for both V and PIF agreed with 1 percent (data not shown).

Results

Figures 3 and 4 show the model results for interface propagation in the cryopreservation (Figs. 3(*a*) and 4(*a*)) and cryosurgical (Figs. 3(*b*) and 4(*b*)) cooling conditions for rat liver and AT-1 tumor tissue, respectively. Microscale freezing model predictions are given by the dotted line (- - - -), and enthalpy method model predictions are given by the solid line (——). The interface propagation predictions of the two model simulations are very similar in both the tissue systems. The propagation of both the leading (at early times) and trailing (at late times) edges of the mushy zone are shown in Figs. 3 and 4. The trailing edge of the mushy zone is assumed to be where $\Lambda = 0.005$ for graphing convenience. The leading and trailing edge propagations predicted by both simulations are fairly close (<3 percent difference) to one another (see Figs. 3 and 4). The difference in trailing edge (or the interface at late times) propagation between the two models looks significant, but this is an artifact of the plateau in the saline phase diagram (i.e., 90 percent of the water is frozen at −5°C). The unfrozen fraction in the enthalpy method solution is about 0.03 (or 97 percent of the solution is frozen) along the 0.005 contour for the tissue freezing solution, or essentially frozen from a practical standpoint.

A comparison of the maximum interface propagated for the cryopreservation case (slow cooling conditions; as shown in Table 4) shows that it is ~1.5 percent greater in AT-1 tumor tissue (Fig. 4(*a*)) than in the rat liver tissue (Fig. 3(*a*)). A similar result is also obtained in the cryosurgical case as shown in Figs. 3(*b*) and 4(*b*) for rat liver and AT-1 tumor tissue, respectively (~1.2 percent larger radius or ~3.3 percent larger volume in AT-1 tumor tissue). This effect is clearly due to the lower latent heat value of AT-1 tumor tissue in comparison with rat liver (see Eq. 15). However, this decrease in the latent heat value was found to have "no effect" (<0.1 percent difference) on the thermal history predicted in the rat liver and AT-1 tumor tissue. Figures 5(*a*) and 5(*b*) show the model results for thermal history in the cryopreservation (5°C/

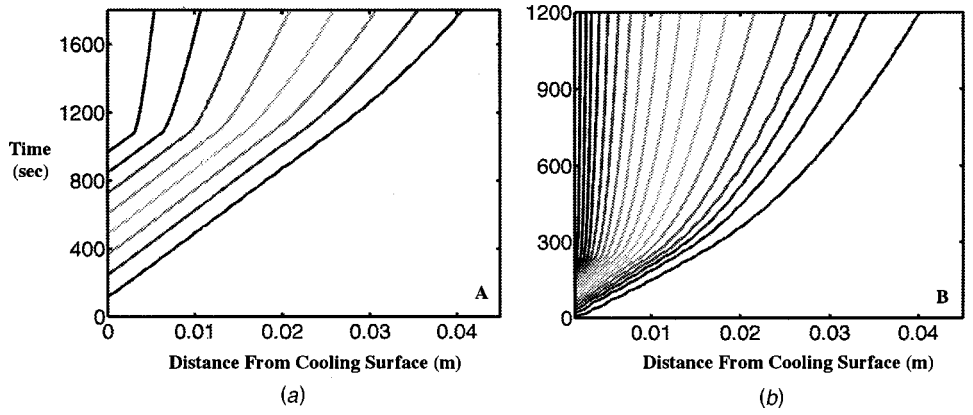


Fig. 6 Predicted isotherms under (*a*) cryopreservation and (*b*) cryosurgery conditions (the isotherms are essentially identical for both rat liver and AT-1 tumor tissue). The cooling conditions are described in Table 4. In Fig. (*a*) the contours from left to right represent isotherms of 203 to 273 K, in increments of 10 K, respectively. In Fig. (*b*) the lines from left to right represent isotherms of 133 to 303 K, in increments of 10 K. The distance from the cooling surface (*m*) is shown on the *x*-axis while time (secs) is shown on the *y*-axis.

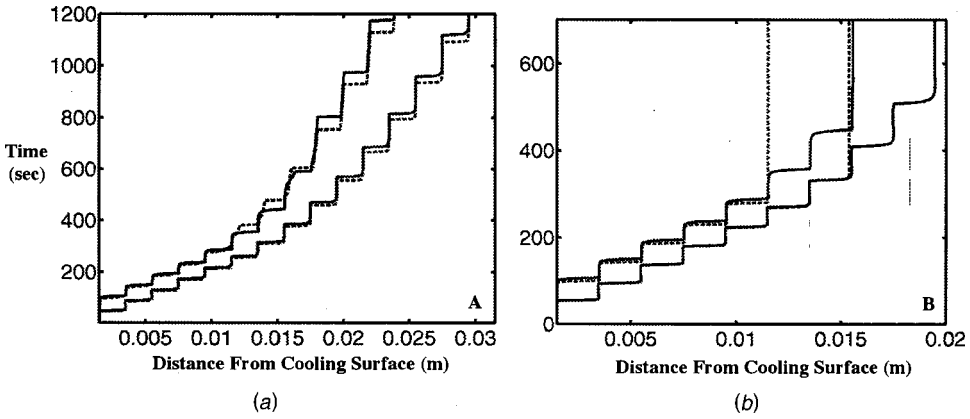


Fig. 7 (*a*) Predicted unfrozen fractions and the probability of intracellular ice formation, and (*b*) PIF under cryosurgery conditions in the rat liver (solid lines,——) and AT-1 tumor tissue (dotted lines,- - - -). The cooling conditions are described in Table 4. Note that the "stepped" behavior of the contours is not intrinsic to the tissue freezing process and is caused by the computational limitation of generating and plotting data at discrete intervals. In Fig. 7(a) the region to the left of the inner contours is essentially frozen (i.e., 99 percent frozen) and the region to the right of the right contours is essentially unfrozen (i.e., 99 percent unfrozen) while the region between the contours represents the "mushy" or partially frozen region. In Fig. 7(*b*) as in Fig. 7(*a*) the region to the left of the inner contour represents the freezing domain with 99 percent PIF and the region to the right of the outer contour represents the freezing domain with 1 percent PIF. The distance from the cooling surface (*m*) is shown on the *x*-axis while time (secs) is shown on the *y*-axis.

min; as shown in Table 4) and cryosurgical (50°C/min; as shown in Table 4) cooling conditions at selected locations within the tissue (these thermal history simulations are valid for both rat liver and AT-1 tumor tissue). As before, the microscale freezing model predictions are given by the dotted line (- - - -), and enthalpy method model predictions are given by the solid line (——). The similarity between the predictions of the two model simulations (coupled and uncoupled) continues in the thermal history results, as well.

Figure 6 shows the model (coupled thermal/biophysical model) simulated isotherms (lines of constant temperature) in the computational domain for both the cryopreservation (Fig. 6(*a*)) and cryosurgical cooling conditions (Fig. 6(*b*)). In Fig. 6(*a*), the lines from left to right represent isotherms from 203 to 273 K, in increments of 10 K, respectively. While in Fig. 6(*b*), these lines represent isotherms from 133 to 303 K, in increments of 10 K. As was the case in the thermal history results (Fig. 5), these isotherms are identical in both the tissue systems studied and are also in good agreement with those simulated by the uncoupled or the enthalpy model (data not shown).

Figure 7 shows a comparison of the amount of unfrozen fraction (Fig. 7(*a*)) and the probability of intracellular ice formation (Fig. 7(*b*)) in the rat liver (solid line, ——) and AT-1 tumor tissue (dotted lines, · · · ·) at various locations within the computational domain (*x*-axis) as a function of temperature (*y*-axis) for the cryosurgical (as shown in Table 4) cooling conditions. In Fig. 7(*a*), the region to the left of the inner most solid (or dashed) line represents the frozen region (or <1 percent is unfrozen) while the region to the right of the outer most solid (or dashed) line represents the unfrozen region (or >99 percent unfrozen fraction). The domain between the two solid (or dashed) lines represent the "mushy" or partially frozen region. Similarly, the contour lines, from left to right, in Fig. 7(*b*) represent >99 percent and <1 percent probability of intracellular ice formation in the tissue. An

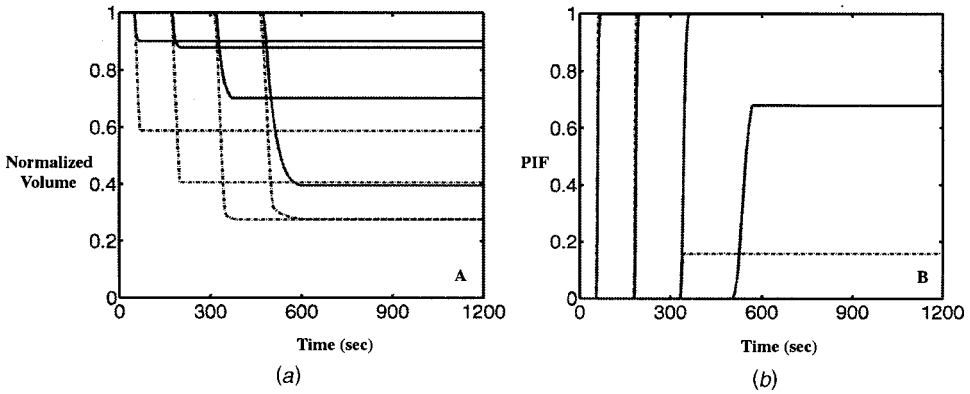


Fig. 8 (*a*) Predicted volumetric shrinkage response and the probability of intracellular ice formation, and (*b*) PIF under cryosurgery conditions in the rat liver (solid lines,——) and AT-1 tumor tissue (dashed-dotted lines,—·—·—·—), at 4 different locations within the tissue. The cooling conditions are described in Table 4. In Fig. 8, the lines from left to right are at 0.1 mm, 6 mm, 12 mm, and 16 mm from the cooling surface, respectively. The distance from the cooling surface (*m*) is shown on the *x*-axis while time (secs) is shown on the *y*-axis.

examination of Fig. 7 shows that there are significant and clear differences in both the amount of unfrozen fraction and the probability of intracellular formation between rat liver and AT-1 tumor tissue. These differences are due to the differences in the characteristic dimensions, shown in Table 2, and in the membrane permeability and intracellular ice formation parameters (shown in Table 3) between rat liver and AT-1 tumor tissue. However, identical thermal history (Fig. 5) and isotherm contours (Fig. 6) results were obtained in both the tissue systems studied. Thus, suggesting that the different microscale biophysical response (cellular dehydration and intracellular ice formation) exhibited during freezing (shown in Fig. 7) does not affect the macroscale freezing problem.

And finally, Fig. 8 shows the unique capability of the coupled thermal/biophysical model to generate both the volumetric shrinkage of cells due to cellular dehydration (or water transport out of the cells) and the probability of intracellular ice formation, PIF at various locations in the tissue for the cryosurgical cooling conditions (the exact conditions are shown in Table 4). Figures 8(*a*) and 8(*b*) show the normalized cellular volume and PIF as a function of time at four different locations in both the rat liver (solid lines,) and the AT-1 tumor tissue (dashed-dotted lines, - - - -), respectively. In Fig. 8(*a*), the volumetric shrinkage curves (from left to right) are at 0.1 mm, 6 mm, 12 mm, and 16 mm from the cooling surface. Similarly, the PIF curves (from left to right) are also at 0.1 mm, 6 mm, 12 mm, and 16 mm from the cooling surface. As seen in Fig. 8, the cells closest to the cooling surface experience the least amount of cellular dehydration (and all cells undergo intracellular ice formation or the maximum PIF is 1) due to the relatively high cooling rate (50°C/min) imposed on the boundary. As the cooling rate experienced by the cells inside the tissue is lower than the imposed cooling rate at the boundary, the cells within the tissue undergo progressively larger volumetric shrinkage (as shown in Fig. 8(*a*)) and exhibit smaller probability of intracellular ice formation (as shown in Fig. 8(*b*)). Note that the maximum PIF in AT-1 tumor tissue 16 mm from the cooling surface is 0 while it is greater than 0.6 in rat liver tissue (as shown in Fig. 8(*b*)), illustrating the starkly different microscale biophysical response exhibited by these tissues during the freezing process.

Discussion

This study presents a unique combination of the microscale phenomena (water transport and IIF) with the macroscale phase change process in biological tissues. The microscale processes are included in the bio-heat transfer equation using a source term methodology. However, it should be noted that the model is limited by the lack of tissue intracellular ice formation (IIF) parameters. Whether the IIF parameters in tissue are significantly different from the cell level parameters used in this study, can only be answered in future studies.

The current model approximates the phase transition in the biological tissues based on the microscale phenomena. Although the microscale freezing, response is clearly different between both the tissues as shown by previous studies [23–25] and also in the present study (Figs. 7 and 8), the macroscale freezing response is not (as shown in Figs. 5 and 6). Thus, suggesting that the uncoupled approach (where the latent heat is a function of temperature, alone) is adequate to predict the macroscale freezing response in the tissue systems, investigated. A thorough parametric investigation varying the Krogh cylinder dimensions, non-water fractions, fractional volumes of extra- and intracellular water, the biophysical parameters and cooling conditions should further elucidate the effect of including the microscale phenomena in the larger macroscale freezing problem. However, preliminary simulations show that the predicted thermal history results are not significantly affected (<2 percent variation) even when the biophysical parameters are drastically altered (i.e., a factor of 100 lower or higher). Thus, indicating that the computationally less intensive enthalpy model (or the uncoupled approach) is adequate to describe the freezing process in most biological tissue systems during typical cryosurgical and cryopreservation cooling conditions (shown in Table 4).

The main result obtained in the current study, i.e., the similarity in the macroscale freezing response predicted using the coupled thermal/biophysical model and the uncoupled enthalpy model has also been noted previously by Smith et al. [21]. The main difference between the current study and Smith et al. [21] is the methodology used to calculate the temperature and time dependence of the latent heat release. Smith et al. [21] curve fitted to the experimentally measured temperature and time dependence of the measured latent heat release during freezing in solute laden aqueous solutions. In the current study, the temperature and time dependence of the latent heat release is based on mechanistic microscale (heat transfer or IIF and mass transfer or water transport) phenomena for a particular biological tissue. Additionally, Smith et al. [21] showed that the choice of latent heat magnitude and the inclusion of temperature dependent thermophysical properties has a much larger effect on the predicted cooling rates, the end temperatures and the total volume of the frozen region than the inclusion of a temperature (and time) dependent release of latent heat (i.e., inclusion of microscale freezing phenomena). Therefore, accurate characterization of the temperature dependent thermophysical properties (of the relevant biological tissue) are necessary to improve the accuracy and applicability of the model predicted macroscale freezing response.

The present model neglects the effect of metabolic heat generation and the effect of blood flow on the freezing process, see Eq. (1). This formulation is clearly less valid for simulating an in-vivo cryosurgical scenario than an in-vitro cryopreservation case. In addition, the coupled model presented in this study assumes that the latent heat released by the extracellular solution during the freezing process follows the phase diagram, as stated in Eq. (3). Recent studies in our lab [21,25] suggest that the latent heat released by the extracellular solution has a cooling rate dependence (i.e., it is both temperature and time dependent). Future improvements to the models will include these effects appropriately.

And finally, the coupled thermal/biophysical model presented here, has several advantages and disadvantages over the more commonly adopted uncoupled enthalpy models. The main advantages of the coupled thermal/biophysical model are as follows: (1) an unique capability to predict the continuous volume/PIF at various locations within the tissue as a function of time; (2) an easily adaptable model that can be used to predict both the thermal and biophysical response in several different tissue types as well as in engineered tissue equivalents; (3) one complete model that describes both the thermal and biophysical behavior of tissue freezing under a variety of cooling conditions; (4) an easily extendable model that can predict the thermal and biophysical response during freezing of tissues in the presence of cryoprotective agents or CPAs (CPAs are used to limit the freezing damage experienced by tissues); (5) vascular distension throughout the tissue could also be easily predicted using the model predicted cell volume changes through the Krogh cylinder model and used to develop vascular injury theories; and (6) the model with a few modifications can predict partial or complete vitrification in cells and extracellular/vascular space. The main disadvantages of the coupled thermal/biophysical model are as follows: (1) computationally more intensive than the uncoupled enthalpy models, i.e., on an SGI workstation the uncoupled model requires approximately 4 to 6 hours of computational time while the corresponding coupled thermal/biophysical code requires $\sim$10 days of processing time (or an increase in the required computational time by 30 to 60 times); and (2) requires a priori knowledge of several tissue level biophysical parameters, several of which are still unknown (for example, the intracellular ice formation parameters in whole tissue slices).

Conclusion

A coupled thermal/biophysical numerical model for understanding the effect of microscale heat and mass transport during freezing in biological tissues was developed. The numerical results from the coupled model were compared with the results from the more commonly used uncoupled model (enthalpy model) in two test cases (rat liver tissue and AT-1 tumor tissue) under two different cooling conditions (cryopreservation and cryosurgical). The results for both the coupled and uncoupled models are very similar; this suggests that the microscale biophysical processes which occur in the biological tissue during freezing do little to limit the rate at which phase change can occur, and that the uncoupled approach to numerical solution would be adequate to accurately determine thermal history for the tissue types studied. Further application of the coupled thermal/biophysical model to different tissue types with various biophysical properties could demonstrate what values, if any, of the biophysical parameters (L_{pg}, E_{Lp}, Ω_o, and κ_o) significantly affect predicted thermal history, and thus which (if any) tissue types would require use of the coupled approach to numerical solution of the thermal and biophysical processes in tissue freezing. However, preliminary simulations suggest that the predicted thermal history results are not significantly affected (<2 percent variation) even when the biophysical parameters are drastically altered (i.e., a factor of 100 lower or higher).

Acknowledgments

This work was supported by the National Science Foundation (NSF-BES # 9703326) and a grant from the Materials Research Science and Engineering Center (MRSEC) at the University of Minnesota.

References

[1] Gage, A. A., and Baust, J. 1998, ''Mechanism of Tissue Injury in Cryosurgery,'' Cryobiology, **37**, pp. 171–186.
[2] Nerem, R. M. 2000, ''Tissue Engineering: Confronting the Transplantation Crisis,'' Proc. Instn. Mech. Engrs., **214**, (Part H), pp. 95–99.
[3] Mazur, P. 1984, ''Freezing of Living Cells: Mechanisms and Implications,'' (review), Am. J. Physiol., **143**, pp. C125–C142.
[4] Budman, H., Shitzer, A., and Del Giudice, S. 1986, ''Investigation of Temperature Fields Around Embedded Cryoprobes,'' ASME J. Biomech. Eng., **108**, pp. 42–48.
[5] Keanini, R. G., and Rubinsky, B. 1992, ''Optimization of Multiprobe Cryosurgery,'' ASME J. Heat Transfer , **114**, pp. 796–801.
[6] Bischof, J. C., and Rubinsky, B. 1993, ''Microscale Heat and Mass Transfer of Vascular and Intracellular Freezing in the Liver,'' ASME J. Heat Transfer **115**, pp. 1029–1035.
[7] Rabin, Y., and Shitzer, A. 1998, ''Numerical Solution of the Multidimensional Freezing Problem During Cryosurgery,'' ASME J. Biomech. Eng., **120**, pp. 32–37.
[8] Lunardini, V., 1981, ''Finite Difference Methods for Freezing and Thawing,'' in *Heat Transfer in Cold Climates*, Van Nostrand Reinhold, Co., New York.
[9] Alexiades, V., and Solomon, A. D., 1993, *Mathematical Modeling of Melting and Freezing Processes*, Hemisphere Publishing Corp., Washington.
[10] Ozisik, M. N., 1994, *Finite Difference Methods in Heat Transfer*, CRC Press, Boca Raton, FL.
[11] Hayes, L. J., and Diller, K. R. 1983, ''Implementation of Phase Change in Numerical Models of Heat Transfer,'' Journal Energy Research Technology, **105**, pp. 431–435.
[12] Hayes, L. J., Diller, K. R., Chang, H.-J., and Lee, H. S. 1988, ''Prediction of Local Cooling Rates and Cell Survival During the Freezing of a Cylindrical Specimen,'' Cryobiology, **25**, pp. 67–82.
[13] Rubinsky, B., and Pegg, D. E., 1988, ''A mathematical model for the freezing process in biological tissue,'' Proc. Phys. Soc. London, Sect. B, **234**, pp. 343–358.
[14] Toner, M., Cravalho, E. G., and Karel, M. 1990, ''Thermodynamics and Kinetics of Intracellular Ice Formation During Freezing of Biological Cells,'' J. Appl. Phys., **67**, pp. 1582–1593.
[15] Hayes, L. J., Diller, K. R., and Chang, H. J., 1986, ''A Robust Numerical Method for Latent Heat Release During Phase Change,'' *Advances in Heat and Mass Transfer in Biotechnology*, HTD-Vol. **62,** pp. 63–69.
[16] Pitt, R. E. 1990, ''Cryobiological Implications of Different Methods of Calculating the Chemical Potential of Water in Partially Frozen Suspending Media,'' Cryo-Letters, **11**, pp. 227–240.
[17] Mazur, P. 1963, ''Kinetics of Water Loss From Cells at Subzero Temperatures and the Likelihood of Intracellular Freezing,'' J. Gen. Physiol., **47**, pp. 347–369.
[18] Levin, R. L., Cravalho, E. G., and Huggins, C. E. 1976, ''A Membrane Model Describing the Effect of Temperature on the Water Conductivity of Erythrocyte Membranes at Subzero Temperatures,'' Cryobiology, **13**, pp. 415–429.
[19] Toner, M., 1993, ''Nucleation of Ice Crystals Inside Biological Cells,'' in *Advances In Low-Temperature Biology*, P. Steponkus ed., JAI Press, London, pp. 1–52.
[20] Toner, M., Tompkins, R. G., Cravalho, E. G., and Yarmush, M. L. 1992, ''Transport Phenomena During Freezing of Isolated Hepatocytes,'' AIChE J., **38**, pp. 1512–1522.
[21] Smith, D. J., Devireddy, R. V., and Bischof, J. C., 1999, ''Prediction of Thermal History and Interface Propagation During Freezing in Biological Systems—Latent Heat and Temperature-Dependent Property Effects,'' *Proc. 5th ASME/JSME Thermal Eng. Joint Conf.*, San Diego CA. CD-ROM Publication, (http://www.me.umn.edu/divisions/tht/bhmt/publications/pdf/ASMEJSME-freezing.pdf).
[22] Patankar, S. V., 1980, *Numerical Heat Transfer and Fluid Flow*, Hemisphere Publishing Co., New York.
[23] Pazhayannur, P. V., and Bischof, J. C. 1997, ''Measurement and Simulation of Water Transport During Freezing in Mammalian Liver Tissue,'' ASME J. Biomech. Eng., **119**, pp. 269–277.
[24] Devireddy, R. V., and Bischof, J. C. 1998, ''Measurement of Water Transport During Freezing in Mammalian Liver Tissue: Part II—The Use of Differential Scanning Calorimetry,'' ASME J. Biomech. Eng., **120**, pp. 559–569.
[25] Devireddy, R. V., Bischof, J. C., Leo, P. H., and Lowengrub, J. S., 2000, ''Measurement and Modeling of Latent Heat Release During Freezing of Aqueous Solutions in a Small Container,'' *Advances in Heat and Mass Transfer in Biotechnology*, HTD-Vol. 368/BED-Vol. 47, pp. 23–31.
[26] Fennema, O. R., Powrie, W. D., and Marth, E. H., 1973, *Low-Temperature Preservation of Foods and Living Matter*, Marcel Dekker, Inc., New York.
[27] Devireddy, R. V., Smith, D. J., and Bischof, J. C. 1999, ''Mass Transfer During Freezing in Rat Prostate Tumor Tissue,'' AIChE J., **45**, No. 3, pp. 639–654.
[28] Schulte, M. G., 1996, ''The Effects of Varying Concentrations of DMSO on the Biophysical Parameters of Isolated Mammalian Cells During Freezing in the Presence of Extracellular Ice,'' M.S. thesis, University of Minnesota, Minneapolis, MN.
[29] Smith, D. J., Josephson, S. S., and Bischof, J. C., 1997, ''A Model of Cryosurgical Destruction in AT-1 Prostate Tumor Based on Cellular Damage Mechanisms,'' *Advances in Heat and Mass Transfer in Biotechnology*, HTD-Vol. 355, pp. 149–150.

Transient Thermal Bubble Formation on Polysilicon Micro-Resisters

Jr-Hung Tsai
Mechanical Engineering Department,
University of Michigan,
Ann Arbor, MI
e-mail: jhtsai@argon.eecs.berkeley.edu

Liwei Lin
Mechanical Engineering Department,
University of California at Berkeley,
Mechanical Engineering,
1113 Etcheverry Hall,
University of California,
Berkeley, CA 94720-1740

Transient bubble formation experiments are investigated on polysilicon micro-resisters having dimensions of 95 μm in length, 10 μm or 5 μm in width, and 0.5 μm in thickness. Micro resisters act as both resistive heating sources and temperature transducers simultaneously to measure the transient temperature responses beneath the thermal bubbles. The micro bubble nucleation processes can be classified into three groups depending on the levels of the input current. When the input current level is low, no bubble is nucleated. In the middle range of the input current, a single spherical bubble is nucleated with a waiting period up to 2 sec while the wall temperature can drop up to 8°C depending on the magnitude of the input current. After the formation of a thermal bubble, the resister temperature rises and reaches a steady state eventually. The bubble growth rate is found proportional to the square root of time that is similar to the heat diffusion controlled model as proposed in the macro scale boiling experiments. In the group of high input current, a single bubble is nucleated immediately after the current is applied. A first-order model is proposed to characterize the transient bubble nucleation behavior in the micro-scale and compared with experimental measurements. [DOI: 10.1115/1.1445136]

Keywords: Bubble Growth, Heat Transfer, Microscale, Phase Change, Transient

1 Introduction

Device miniaturization has been the technology driver in both semiconductor and microelectromechanical system (MEMS) research. As a result, heat dissipation has become a major problem in integrated circuit (IC) industry due to the fact that more microelectronic devices are densely placed in an extremely tiny area. Because boiling heat transfer is the most effective way for heat removal, several research have been proposed to use the phase change phenomenon as the heat dissipation mechanism for semiconductor chips [1,2]. On the other hand, boiling phenomenon has also been applied as the driving mechanism in various MEMS devices, such as thermal bubbles for ink-jet printers [3], bubble powered microactuators [4], and bubble powered pumping effects [5]. The possible applications for micro-bubble powered devices are abundant and the potential of using the phase change process to solve the heat dissipation problem in IC industry is enormous. Therefore, it is important to investigate the fundamental mechanisms of bubble nucleation in the micro-scale.

Among the demonstrated devices using thermal bubbles, they can be classified into two groups according the devices structures. In the case of open environment, there is ink-jet printer head [3]. In the case of close environment, there are micro-thermal bubble pumps [6,7] and micro-thermal bubble valves [8]. In order to make better bubble-powered micro-devices, it is important to study micro-bubble formation in both open environment and inside microchannels, including the transient bubble formation behavior. Previously, Lin et al. investigated the bubble formation mechanisms on polysilicon micro resistor in open environment [5] and inside microchannels [9]. Yang [10] reported an overview of boiling on microstructures. Although these reports contributed various aspects of bubble formation by using micro resisters, none of them discussed the transient behavior of micro-bubble formation that plays an important role in the design and the operation of micro-bubble-based devices. In macro-scale, the bubble nucleation process has been studied from many aspects, such as ultimate boiling temperature limit [11], nucleation from trapped gas [12], bubble growth [13], and the local temperature fluctuation at the heating surface [14]. This paper concentrates on the transient temperature fluctuation during bubble formation and bubble growth on a polysilicon micro-resistor. Both mathematical models and experimental results are established for the characterization of transient bubble formation in the micro-scale.

2 Experiments

A Petri dish containing the testing chip and experimental liquid is placed on the observation chuck of a probe station as shown in Fig. 1. The bulk liquid temperature is measured before and after each experiment by a thermal couple, and is found to be within 1°C of variation. The microscope is connected to a CCD camera to capture images during the transient bubble formation experiments. HP4145 semiconductor parameter analyzer is used as the power source and signal recording system. Polysilicon micro-resisters are fabricated by using the MCNC MUMPs service [15] and they work as resistive heaters. The cross sectional view of the device is shown in Fig. 2(*a*) with silicon nitride as the insulation material and gold as the metal contact. The typical dimension of the polysilicon resistive heater is 95 μm in length, 5 μm, or 10 μm in width and 0.5 μm in thickness. There are two pairs of electrical contact pads for measuring current and voltage simultaneously to monitor the resistance changes of the polysilicon micro resister as shown in Fig. 2(*b*). When a current is applied to the resistive heater, the voltage change at the central portion, 2.5 μm away from both ends of the resister, is recorded. The temperature coefficient of resistance of polysilicon is characterized as $1.2*10^{-3}$/K by measuring the resistance changes with respect to environmental temperature changes. The transient temperature response of the resister is calculated based on this coefficient. It is noted that this coefficient is experimentally characterized and valid when temperature is less than 300°C that is within the temperature range of this work. Second-order effects may need to be considered when the operation temperature is high. A fabricated device is shown in Fig. 2(*c*) where gold pads have the shining bright color.

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division April 26, 2001; revision received October 18, 2001. Associate Editor: G. P. Peterson.

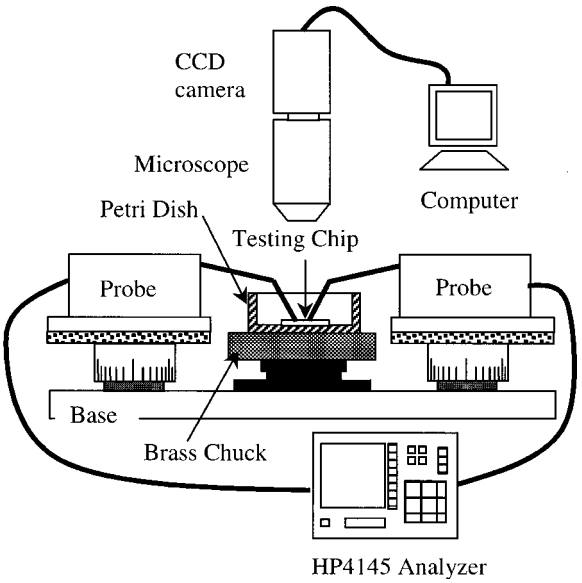


Fig. 1 Schematic drawing of the micro boiling experiment

3 A Lumped Heat Transfer Model

A lumped heat transfer model is developed to investigate the thermal responses of the micro resister theoretically. This model is graphically shown in Fig. 3 with the following assumptions. First, the polysilicon micro resister has a uniform temperature along the length of the resister. This assumption is based on a previous work [17] that line shape polysilicon micro-resister has relatively uniform temperature distribution away from the boundaries. Since the model is only built on the central portion of the resister in between the two voltage detecting points, the temperature distribution is assumed to be uniform. Second, the heat loss through the two ends of the micro-resister are neglected because the cross-section areas of the two ends of the micro-resister are comparatively much smaller than the areas along the micro resister. Third, the temperature distribution across the silicon nitride layer underneath the polysilicon resister is linear. This assumption is based on the fact that the width of the micro-heater is much larger than the thickness of nitride layer. Therefore, the temperature change is approximated as linear within the portion of nitride layer that is underneath the polysilicon resister. Fourth, the substrate warming

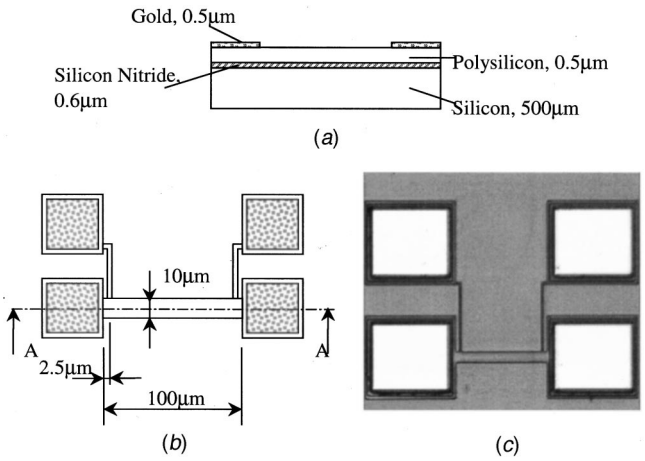


Fig. 2 Schematic diagrams of the polysilicon micro resister: (*a*) cross-sectional view; (*b*) top view; and (*c*) photograph of a fabricated polysilicon micro resister.

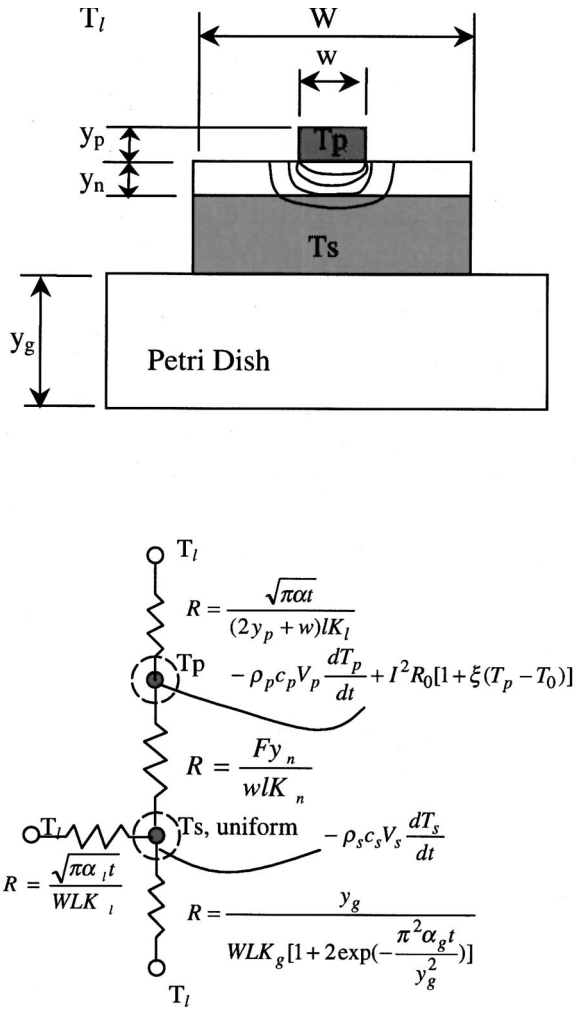


Fig. 3 Schematic drawing of the lumped heat transfer model. The electric analogy is drawn on the right.

effect is an important parameter to be characterized, and the silicon substrate is assumed to have a uniform temperature of *T*s to simplify the analysis. However, non-uniform temperature distribution is expected at regions close to the polysilicon micro-resister. An experimentally determined factor, *F*, is used to adjust possible errors coming from this assumption. This factor is defined as the extension factor to the thickness of the nitride layer to accommodate the region of non-uniform temperature. Intuitively, this factor will be determined by the thermal conductivity of nitride, silicon, and their geometrical relationships. In this work, this number is experimentally characterized by comparing the analytical prediction with the measured wall temperature. Fifth, heat conduction in liquid is modeled by semi-infinite thermal conduction and both heat convection and radiation are neglected [17]. Sixth, it is assumed that the bottom surface of the Petri dish is at room temperature, since it is in contact with a big brass chuck that is a good heat sink. A lumped, one-dimensional heat transfer model is established graphically as shown in Fig. 3. According to the transient thermal conduction equation, the transient temperature of a slab with temperature fixed at two sides is modeled by [18]

$$T_y = T_1 + (T_2 - T_1)\frac{y}{l} + \frac{2}{\pi}\sum_{1}^{\infty}\frac{T_2\cos n\pi - T_1}{n}\sin\frac{n\pi y}{l}e^{-\alpha n^2\pi^2 t/l^2}, \quad (1)$$

Table 1

Parameter	Value	Parameter	Value	Parameter	Value
ρ_p	2320 Kg/m^3	c_p	700 J/KgK	K_n	3.2 W/mK
K_s	141 W/mK	K_l	0.136 W/mK	y_n	0.6 μm
W	10 μm	l	95 μm	y_p	0.5 μm
R_0	300 Ω	ξ	1.2E-3 /K	α_l	5.9E-8 m^2/s
T_1	294 K	T_0	300 K	ρ_s	2320 Kg/m^3
c_s	700 J/KgK	K_g	1.2 W/mK	y_s	500 μm
W	1 cm	L	1 cm	y_g	1.8 mm
α_g	5.1E-7 m^2/s	E	6.4E5 J/Kg		

where $T_{1,2}$ are temperatures at two sides of a slab, y is the coordinate along the slab thickness, l is the slab thickness, and α is the thermal diffusivity of the slab. Taking derivative with respect to y and multiplying with material conductivity K, one can obtain the heat flux as

$$K\frac{\partial T_y}{\partial y}=K\left[\frac{(T_2-T_1)}{l}+\frac{2}{l}\sum_{l}^{\infty}(T_2\cos n\pi-T_1)\times\cos\frac{n\pi y}{l}e^{-\alpha n^2\pi^2 t/l^2}\right]. \quad (2)$$

For first-order approximation, only the first term in the summation is considered in the above equation. Based on the above assumptions and energy balance, the heat transfer equations of polysilicon and silicon can be derived as

$$\rho_p c_p\frac{dT_p}{dt}=-\frac{K_n}{y_p}\frac{(T_p-T_s)}{Fy_n}-\left(\frac{2}{w}+\frac{1}{y_p}\right)K_l\frac{(T_p-T_l)}{\sqrt{\pi\alpha_l t}}+\frac{I^2R_0[1+\xi(T_p-T_0)]}{wly_p} \quad (3)$$

$$\rho_s c_s\frac{dT_s}{dt}=\frac{wl}{WL}\frac{K_n}{y_s}\frac{(T_p-T_s)}{Fy_n}-\frac{K_l}{y_s}\frac{(T_s-T_l)}{\sqrt{\pi\alpha_l t}}-\frac{K_g}{y_s}\frac{T_s-T_l}{y_g}\times\left[1+2*\exp\left(-\frac{\pi^2\alpha_g t}{y_g^2}\right)\right]. \quad (4)$$

The values used in the equations are listed in Table 1. These two equations can be solved simultaneously to obtain the values of T_p (polysilicon micro resister temperature) and T_s (silicon substrate temperature) with respect to time. The liquid used in the experiments is IPA (Isopropyl Alcohol) [19].

4 Results and Discussion

4.1 General Behavior of Micro-Boiling Phenomenon. The general behavior of micro-boiling phenomenon is recorded based on the micro-resister with size of $100\times10\times0.5\,\mu m^3$ under various constant input currents from 15 to 40 mA. The measured wall temperatures for the first 10 sec are shown in Fig. 4 and the close look of relative temperature changes from Fig. 4 are plotted in Fig. 5. As expected, the measured wall temperature increases as the input current increases. It is found that bubble formation characteristics with respect to input current levels can be classified into three groups. In group I, where the input current is less than 22 mA, no bubble is nucleated and the temperature of the polysilicon micro-resister remains nearly the same. Group II represents the middle range of input currents between 25 to 30 mA and is the most interesting group in terms of transient bubble formation behavior. The wall temperature increases at the beginning and drops ahead of the bubble formation when a blurry cloud covers the polysilicon micro-resister as observed through the microscope. The slope of this temperature drop seems to become steeper as the magnitude of the input current increases and disappears under high input power. For example, this phenomenon is observed when the input current is less than 40 mA for the 10 μm wide resisters and 15 mA for the 5 μm wide resisters. Once a bubble is nucleated, the wall temperature increases and goes above the initial temperature before reaching a steady state. In the final group where the input current is equal or higher than 40 mA, a bubble is formed as soon as the electrical current is applied. There is no detectable temperature drop when the thermal bubble is generated and the resister temperature continues to increase and the bubble grows before a steady state is reached. In all three groups, it is measured that the temperature of the silicon substrate has also risen and gradually reaches a steady state temperature about 1 to 10°C higher than the room temperature depending on

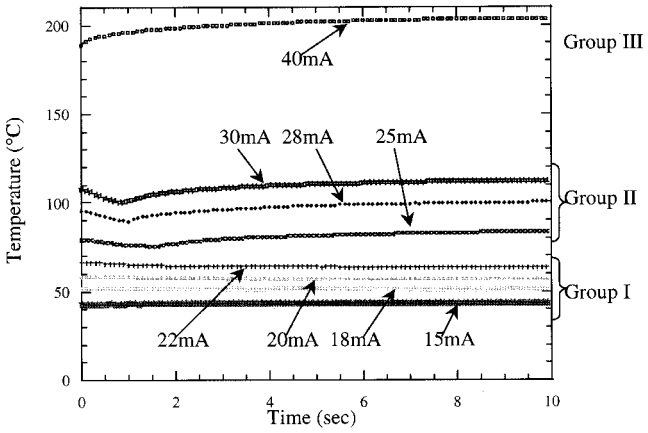


Fig. 4 Measured wall temperature with respect to time on the polysilicon micro resister of 10 μm wide

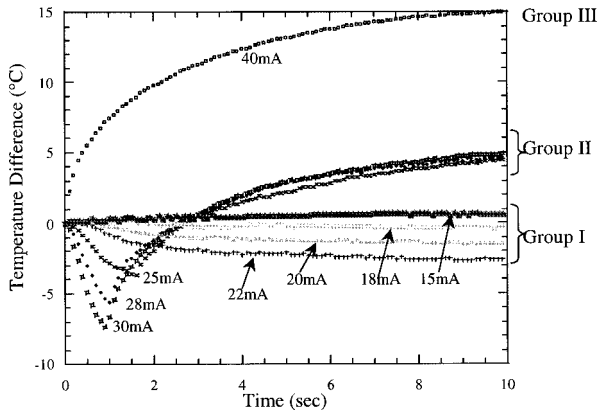


Fig. 5 Relative temperature changes for the transient micro bubble formation on the 10 μm wide polysilicon micro resister

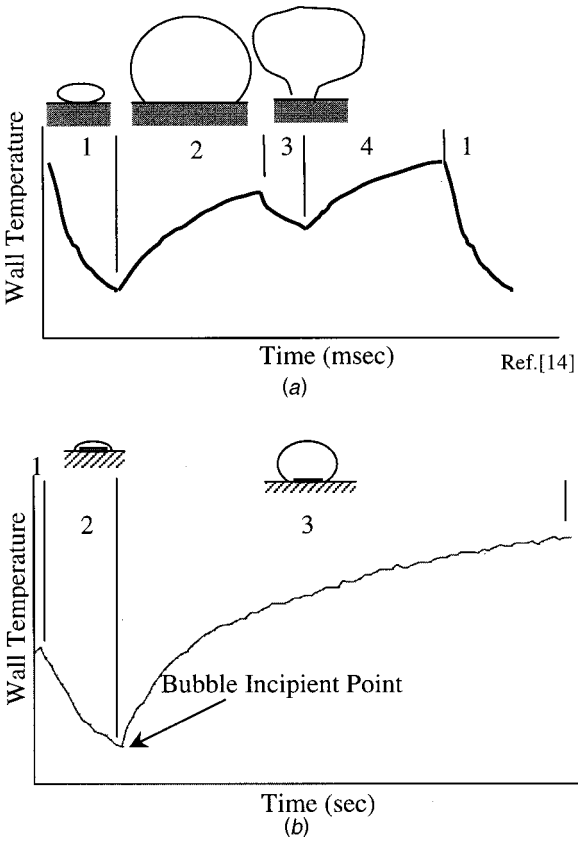


Fig. 6 Comparison of bubble nucleation in macro and micro scales: (*a*) typical wall temperature history in macro scale boiling experiments; and (*b*) typical wall temperature history in micro scale boiling on a polysilicon micro resister.

the level of the input power. It is noted that the dissolved air in bulk liquid may affect the experimental results but is not studied in this work.

4.2 Comparison of Macro and Micro-Boiling Experiments. Figure 6 illustrates the typical wall temperature of bubble nucleation processes for both macro [20] and micro-scale experiments, respectively. The macro-scale bubble nucleation experiments are generally conducted at pool boiling condition where the liquid is at the saturation state. The micro-scale bubble nucleation test, on the other hand, refers to bubble formation on micro-scale resistive heaters in a pool of cool liquid. For a typical macro-scale boiling process under a constant heating power as shown in Fig. 6(*a*) [20], formation of bubbles is observed in stage 1. The temperature drop in stage 1 has been explained as the evaporation of a microlayer liquid between the bubble and the heating wall. When the microlayer evaporates and dries out, the boiling process moves to stage 2, where the bubble continues to grow. The wall temperature rises in this stage because of the blockage of the heat flux from the wall to the liquid due to the existence of the bubble. When the bubble grows to a certain size and starts to lift off the wall as shown in stage 3, cool liquid is sucked to the wall to cause a quenching effect. The cool liquid is then heated up gradually in stage 4 to the original temperature, where a new bubble is nucleated and the process repeats itself continuously.

Group II in Fig. 4 and 5 of the bubble nucleation experiments in the micro scale, is discussed and compared with macro scaled boiling experiments due to similarities. One may identify three stages of wall temperature responses on transient bubble formation. In stage 1, the wall temperature jumps right after an electric current is applied and moves up gradually as the substrate temperature rises. The period of this stage is shorter than 0.1 sec for the 10 μm wide resister and becomes shorter if the input current level is increased. It is believed that liquid is superheated locally in this stage until naturally generated disturbance induces the evaporation process and moves the micro-boiling process to stage 2. According to the homogeneous bubble nucleation theory [20], there is a nucleus radius under thermodynamic equilibrium. Once the initial bubble nucleus radius is larger than the critical value, the bubble will tend to grow. Otherwise, it will shrink. High current inputs generate high temperature that implies smaller equilibrium nucleus radius. Therefore, the probability that the bubble embryos are larger than the critical size is increased and a large single spherical bubble is induced. This can be the reason causing the temperature profile transits from stage 2 into stage 3. Moreover, high current inputs seem to speed up the evaporation process that causes the wall temperature to drop faster as shown in Fig. 5. As soon as a bubble emerges, the wall temperature increases as shown in stage 3, in which a spherical bubble grows and covers the whole polysilicon micro resister. The wall temperature increases in this stage because the vapor bubble is blocking the heat dissipation process. The nucleated micro-bubble continues to grow slowly and sticks to the polysilicon micro-resister due to Marangoni effect [17] and the wall temperature gradually reaches equilibrium.

4.3 Thermal Conduction. According to Eqs. (3) and (4), one can estimate the time constants of the polysilicon micro-resister and silicon substrate by the following expressions:

$$1/\tau_p \sim \frac{1}{\rho_p c_p y_p}\left(\frac{K_n}{F y_n} - \frac{I^2 R_0 \xi}{wl}\right) \tag{5}$$

$$1/\tau_s \sim \frac{1}{\rho_s c_s y_s}\left(\frac{wlK_n}{WLFy_n} + \frac{K_g}{y_g}\right) \tag{6}$$

where τ_p and τ_s are time constants of polysilicon micro-resister and silicon substrate, respectively. Substituting values in Table 1 into the above equations, one can find that τ_p is in the order of microseconds, and τ_s is in the order of milliseconds. It implies that in the early stage of the heating process, the temperature of polysilicon micro resister is not affected by the silicon substrate warming effect and can be estimated by Eq. (3). When the operation period is longer than milliseconds, the substrate warming effect has to be considered.

A modification factor, F, is introduced as a measure of the effective thickness of nitride layer, and this factor accounts modeling errors coming from the simplifications of the model and its magnitude is determined by experiments. Figure 7 shows the measured temperature of the polysilicon micro resister as well as the predictions from various values of F. It can be observed that when F is equal to 1.47, a good match is achieved. This value is then used in further simulation. Figure 8 shows the simulated temperatures of the 10 μm wide polysilicon micro-resister and the silicon substrate under an input current of 28 mA. As the result shown in Fig. 8, the polysilicon micro-resister temperature rises quickly and reaches 95°C in less than 1 millisecond. This result explains the fast temperature responses in Fig. 4 after an input current is applied. Comparatively, it takes a longer time to create a small temperature change on the silicon substrate. In the same figure, the temperature rise for the silicon substrate is not observable until 0.5 second later when the temperature rise of 1°C is predicted. This substrate warming effect explains qualitatively the slow temperature rise on the polysilicon micro resister before a steady state is reached as shown in Fig. 4 and 5. However, the heat conduction model only predicts the increase but not the drop of temperature with respect to time. Other heat dissipation mechanisms must be considered to characterize the wall temperature drops as observed.

4.4 Thermal Convection. The critical Rayleigh number for the onset of natural convection is calculated to investigate the role of thermal convection in this time dependent micro-boiling experiment

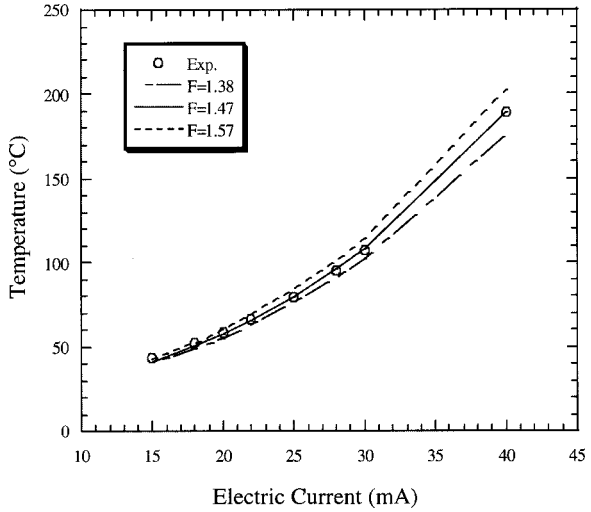


Fig. 7 Simulated initial temperatures of the 10 μm wide micro resister by using various F values as compared with experimental measurements

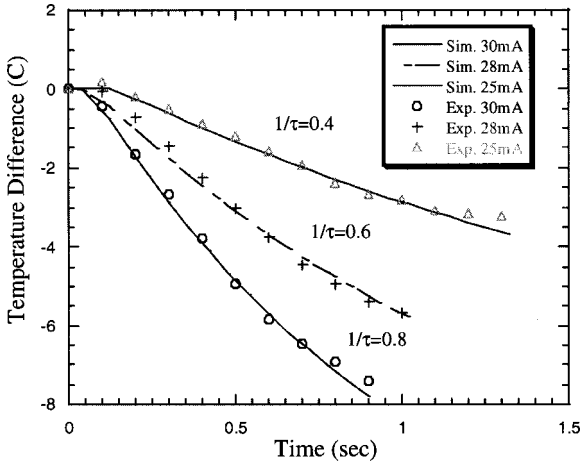


Fig. 9 Comparison of experimental and simulated wall temperatures before the nucleation of a micro-bubble. The equivalent heat transfer coefficient is found to be 10^5 W/m^2°C

$$\mathrm{Ra}_t = \frac{2\pi g \beta q \alpha t^2}{\nu K}, \tag{7}$$

where q is the heat flux to liquid, g is gravity, β is the volume expansion coefficient, ν is dynamic viscosity, and K is thermal conductivity. The model developed in the previous section is used to estimate the heat flux to liquid and other variables. The calculated Rayleigh number is found at least one order of magnitude smaller than the reported critical Rayleigh number [21] when the time period is less than two seconds. Therefore, natural convection is not considered in the mathematical model of Eq. (3) and (4) in the initial stage of bubble formation.

4.5 Evaporation. According to the simulation result presented in Fig. 8, the heating speed in this work is fast in the order of 10^7 K/s. Moreover, a blurry cloud is observed around the micro resister before the bubble formation. It indicates the film boiling of liquid after the liquid is heated into the metastable state [20] and condensed where the vapor meets with cool liquid. This evaporation of liquid may be considered as the primary reason of temperature drop before a micro bubble is nucleated in the micro bubble nucleation process. The heat flux of evaporation is modeled as an equivalent heat transfer coefficient and the temperature difference between the polysilicon micro resister and bulk liquid.

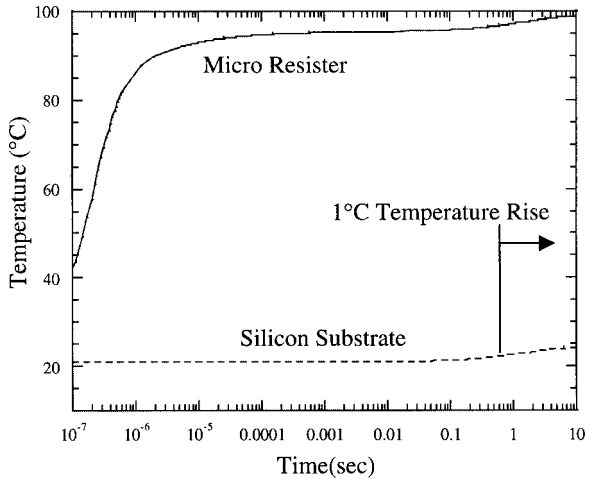


Fig. 8 Simulated results of temperature on polysilicon micro resister of 10 μm wide and silicon substrate versus time under an input current of 28 mA

$$h_{\mathrm{evp}}(T_p - T_l) = \frac{\rho_l E}{(w+2y_p)l}\frac{dV}{dt}, \tag{8}$$

where h_{evp} is the equivalent heat transfer coefficient during the evaporation stage, T_p is the polysilicon micro resister temperature, T_l is the bulk liquid temperature, E is the latent heat of liquid, and $(w+2y_p)l$ is the heating area. In order to determine the value of h_{evp}, the model is compared with experimental data from Fig. 5. It is found that the wall temperature drops at a higher rate at the beginning and at a lower rate when approaching the bubble incipient point. Therefore, the equivalent heat transfer coefficient is proposed as

$$h_{\mathrm{evp}} = h_{\max}[1 - e^{(-t-t_{\mathrm{init}}/\tau)}], \tag{9}$$

where t is time, t_{init} is the time when resister temperature starts to drop, and τ is an evaporation time constant depending on the level of input power. Therefore, the governing equation of micro resister of Eq. (3) can be modified in the second stage of bubble formation as

$$\rho_p c_p \frac{dT_p}{dt} = -\frac{K_n}{y_p}\frac{(T_p - T_s)}{Fy_n} - \left(\frac{2}{w} + \frac{1}{y_p}\right) h_{\mathrm{evp}}(T_p - T_l) + \frac{I^2 R_0[1+\xi(T_p - T_0)]}{wly_p}. \tag{10}$$

The two parameters in Eq. (9) are found experimentally as shown in Fig. 9 by means of curve fitting. It is found that $h_{\max}$ is 7×10^5 W/m^2°C, and $1/\tau$ is 0.4, 0.6, and 0.8 for input current levels at 25, 28, and 30 mA, respectively. This equivalent heat transfer coefficient is at least two orders of magnitude higher than the natural convection heat transfer coefficient from the macro-scale heat transfer models. Therefore, this result further supports the previous assumption that heat convection only plays a minor role in this stage. The calculated evaporation time constants suggest that higher input current corresponds to smaller time constant and higher evaporation rate. Because the heat flux calculated by the heat evaporation term is much larger than the thermal conduction term, the heat dissipation process in this stage is mainly attributed to liquid evaporation. By substituting these numbers into Eq. (8), where ρ_l is 770 kg/m^3 and E is 640 kJ/kg, the rate of evaporation volume is calculated in the order of 10 nl/sec. In most cases, the

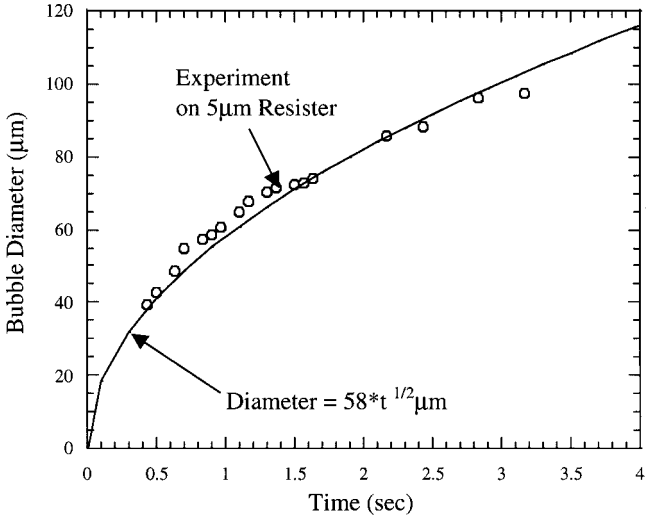


Fig. 10 Bubble size growth rate on a 5 μm micro resister compared with the heat diffusion controlled bubble growth mode

temperature drop lasts longer than one second and more than 10 nanoliter of liquid has gone through the evaporation-condensation process to cause the observed temperature drop.

4.6 Bubble Growth Rate. It is observed that the bubble formation in Group II of Fig. 4 is randomly scattered along the temperature drop curve in stage 2 of the micro bubble formation process. The possible reason of the transition into a spherical bubble may be the instability of vapor-liquid interface that is agitated by local fluctuation in combination with the readiness of enough metastable state of liquid that can overcome the cool liquid to form a spherical bubble. When a bubble is formed, the wall temperature rises in a manner similar to the second stage in the macro-scaled boiling process. It is believed that the heat dissipation path from the polysilicon micro resister surface to liquid is partially blocked by the vapor bubble on top. Therefore, temperature rises with the increase of bubble size to reach a new equilibrium temperature.

As described in a previous report [20], thermal bubble growth can mainly classified into two modes in macro-scale bubble nucleation experiments. The first mode occurs at the initial stage of bubble growth that is hydrodynamically controlled and dominated by liquid inertia and the bubble size increases proportionally to time, $D \propto t$. The second mode occurs at the later stage of bubble growth that is heat diffusion controlled. The bubble grows at a slower rate than in the hydrodynamically controlled mode and is proportional to the square root of time, $D \propto t^{1/2}$. The experimental data on a 5 μm wide micro-resister is extracted by moving the start point of time to the time when a bubble is nucleated. The bubble diameter is measured from the captured images taking on top of the micro-resister by the CCD camera, and graphed along with time by circular markers as shown in Fig. 10. It is found that the bubble diameter, D, can be represented as $D=2Ct^{1/2}$, where C is a constant and t is time. The close match between the markers and the curve indicates that the bubble growth rate on the micro resister at this stage is dominated by heat diffusion as $D \propto t^{1/2}$.

This relation between bubble size and time is further applied to extend the mathematical model to the bubble growth stage. Some assumptions are made and graphically presented in Fig. 11. First, the bubble is spherical with a uniform temperature at the saturation point. Although the temperature at the bottom portion of the bubble is expected to be higher than the top portion, this assumption is made to simplify the analysis. Second, it is assumed that the bottom of the bubble is in contact with the substrate as shown in Fig. 11 due to a strong local Marangoni effect and the contact

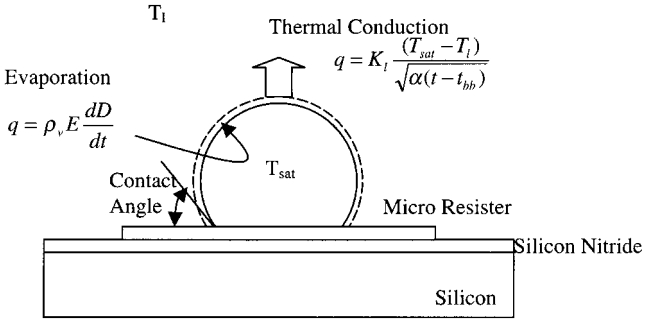


Fig. 11 Schematic drawing of the heat transfer mechanism in bubble growth model

angle is adopted from a previous report [22] as 30 deg. Third, it is further assumed that the bottom of the bubble is in thermal equilibrium with the substrate (no heat transfer) to simplify the problem. In reality, a thin liquid film and thermal gradient with magnitudes may exist depending on the temperature profile of the substrate surrounding the micro heater and a three-dimensional analysis is necessary to model this effect. Fourth, the heat conduction between the bubble/liquid interfaces is assumed as a constant heat source to an infinite field because the bubble temperature is assumed staying at the saturation point. By the above assumptions and energy balance of micro-resister with heat generation of Joule heating, heat dissipation of the substrate, liquid, and bubble growth, Eq. (1) can be modified as

$$\rho_p c_p \frac{dT_p}{dt} = -\frac{K_n}{y_p}\frac{(T_p - T_s)}{Fy_n} - \frac{A_{\text{evp}} h_{\text{evp}}(T_p - T_l) + A_{bb} h_{bb}(T_p - T_{\text{sat}})}{wly_p} + \frac{I^2 R_0[1 + \xi(T_p - T_0)]}{wly_p} \tag{11}$$

$$A_{\text{evp}} = l(w + 2y_p) - A_{bb}; \quad A_{bb} = D(w + 2y_p)\sin\theta$$

$$h_{\text{evp}} = h_{\max}[1 - e^{(-t - t_{\text{init}}/\tau)}],$$

where A_{evp} and A_{bb} are areas of the micro resister that are not covered by the bubble, h_{evp} is the equivalent heat transfer coefficient of evaporation experimentally determined previously, h_{bb} is the equivalent heat transfer coefficient of the bubble, T_{sat} is the saturation temperature of the liquid, D is the diameter of the bubble, and θ is the contact angle of the bubble with the heater that is approximated as 30 deg from the picture in a previous work [22]. The variation of this angle can affect the magnitude the bubble growth rate and this number is used for a first-order approximation. The equivalent heat transfer coefficient of the bubble is derived from the energy balance of the heat transferred into the bubble from the micro resister with the heat dissipation out of the bubble by evaporation and heat conduction to the liquid as illustrated in Fig. 11:

$$A_{bb} h_{bb}(T_p - T_{\text{sat}}) = \left[\rho_v E \frac{dD}{dt} + K_l \frac{(T_{\text{sat}} - T_l)}{\sqrt{\alpha(t - t_{bb})}}\right] A_{\text{sph}} \tag{12}$$

$$A_{\text{sph}} = \frac{\pi D^2 (1 + \cos\theta)}{2},$$

where ρ_v is the vapor density, E is the latent heat of the liquid, t_{bb} is the time when the bubble is formed, and A_{sph} is the interface area of the bubble and the liquid. In this stage, the evaporation and condensation processes are competing with each other. If the diameter of the bubble increases with time, extra energy is supplied to the bubble for the extra evaporation. Otherwise, the bubble

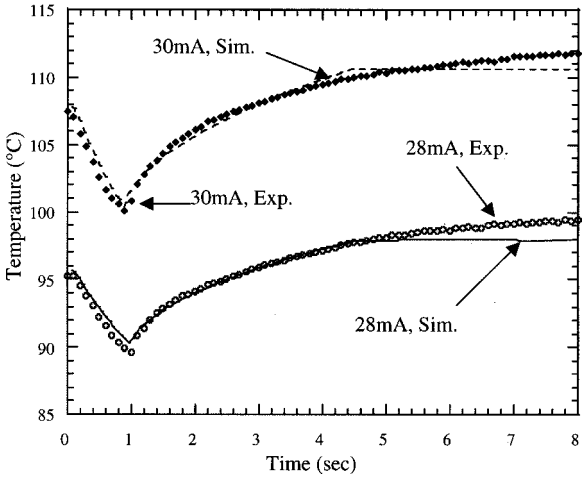


Fig. 12 Comparison between experimental data and simulation on a 10 μm wide of micro-resister under constant electrical current input. The constant of the bubble size growth rate is 50 μm/sec$^{1/2}$ in each current level.

shrinks due to extra condensation. By substituting the areas A_{bb} and A_{sph} into Eq. (12), the equivalent heat transfer coefficient is derived as

$$h_{bb}=\left[\rho_v E\frac{dD}{dt}+K_l\frac{(T_{\text{sat}}-T_l)}{\sqrt{\alpha(t-t_{bb})}}\right]\frac{\pi D(1+\cos\theta)}{2(w+2y_p)\sin\theta(T_p-T_{\text{sat}})}. \tag{13}$$

According to the result obtained from Fig. 10, the bubble size growth rate is modeled by the thermal diffusion controlled bubble growth mode as

$$\frac{dD}{dt}=\frac{C}{\sqrt{t-t_{bb}}}, \tag{14}$$

where C is the bubble growth rate constant and is determined by curve fitting with the experimental results as 29 and 50 μm/sec$^{1/2}$ for 5 μm and 10 μm wide resisters, respectively. Combining Eqs. (2), (11), (13), and (14), a first-order approximation model is developed for this bubble growth stage.

Figure 12 shows the simulation results based on the models of Eq. (10) and Eq. (11) and experimental data. It is found that the models match the experimental data well until the time reaches 5 sec when the models underestimate the experimental temperature. The deviation of the simulated temperature from the experimental data may be resulted from the decrease of the bubble growth rate, the simplifications of the model in various aspects such as the contact angle, and the warming effect of the petri dish. For example, the bubble growth rate decreases more than the mode prediction as shown in the later stage in Fig. 10 in reality. It causes the over prediction of the cooling effect in the model such that the simulated temperature is lower than the experimental data.

5 Conclusion

This paper investigates transient bubble formation on micro strip resisters under various levels of constant electrical current input. The transient temperature variations are classified into three groups by the level of input currents. When the constant input current is lower than 25 mA, the wall temperature stays nearly constant and no bubble is nucleated. For the constant input current between 25 to 30 mA, the wall temperature increases initially, quickly drops afterwards until a bubble is formed, and then increases until reaching steady state. For the third group of high input current, the wall temperature rises sharply as soon as the current is applied and a bubble is formed. Among the three groups, Group II has very interesting transition phenomena that are classified into three stages with respect to time and are investigated extensively in this paper. Mathematical models are established to estimate the resister temperature during the micro boiling process. It is found that this heat conduction based model is capable of predicting the resister temperature in the first stage of the boiling heat transfer. The time constant for the resister temperature is calculated in the microsecond range and in the millisecond range for the silicon substrate. The evaporation of liquid is identified as the key mechanism that causes the resister temperature to drop before the bubble nucleation in the second stage. Based on the experimental measurements and a simplified, first-order model, an equivalent heat transfer coefficient for the second stage is found to be in order of 7×10^5 W/m^2°C, and the evaporation time constants are found as, 2.5, 1.67, and 1.25 sec for the input current levels of 25, 28, and 30 mA, respectively on a 10 μm wide micro-resister. The incipient point of micro-bubble formation is found randomly distributed during this wall temperature drop period with a trend of earlier occurrence in higher input current. This is probably due to the instability of vapor-liquid interface in the stage of strong evaporation process. Higher electrical current will generate bubbles in shorter time at a higher temperature. After a spherical bubble is found, the bubble growth rate is found proportional to the square root of time on both 5 μm and 10 μm wide of micro resisters. The bubble diameter growth rate constant C is found as 29 μm/sec1/2 in the case of 5 μm wide micro-resister, and 50 μm/sec$^{1/2}$ in the case of 10 μm wide micro-resister.

Acknowledgment

The authors would like to thank Dr. Chuan-Hsien Ku for discussion and assistance on experimental equipments. The work is supported in part by an NSF CAREER award (ECS-0096098) and a DARPA/MTO/MEMS grant (F30602-98-2-0227).

Nomenclature

C = bubble diameter growth rate constant
c = specific heat
D = diameter of spherical bubble
E = latent heat
F = modification factor
g = gravity
h = heat transfer coefficient
K = thermal conductivity
l = length q heat flux
R = electrical resistance
T = temperature
t = time
w = width
y = thickness

Greek Symbols

ρ = density
ξ = thermal coefficient of resistance
α = thermal diffusivity
β = thermal expansion coefficient
ν = dynamic viscosity
τ = time constant

Subscripts

0 = reference temperature at 300 K
bb = bubble
evp = evaporation
g = Petri dish
i = interface
l = liquid
n = silicon nitride
p = polysilicon
s = silicon
sat = saturation

References

[1] Cotter, T. P., 1984, ''Principles and Prospects of Micro heat Pipes,'' *Proc. 5th Int'l. Heat Pipe Conf.*, Tsukuba, Japan, pp. 328–335.

[2] Bar-Cohen, A., 1983, ''Thermal Design of Immersion Cooling Modules for Electronic Components,'' Heat Transfer Eng., **4**, No. 3-4, pp. 35–50.

[3] Niellsen, N. J., 1985, ''History of Thinkjet Printerhead Development,'' HP Journal, **36**, No. 5, pp. 4–10.

[4] Lin, L. and Pisano, A. P., 1994, ''Thermal Bubble Powered Microactuators,'' Microsystem Technol., **1**, pp. 51–58.

[5] Lin, L. and Pisano, A. P., 1998, ''Thermal Bubble Formation on Polysilicon Micro Resisters,'' ASME Journal of Heat Transfer, **120**, No. 3, pp. 735–742.

[6] Yuan, H., Oguz, H. N., Prosperetti, A., 1999, ''Growth and Collapse of a Vapor Bubble in a Small Tube,'' Int. J. Heat Mass Transf., **42**, pp. 3643–3657.

[7] Tsai, J. H., and Lin, L., 2001, ''A Thermal Bubble Actuated Micro Nozzle-Diffuser Pump,'' IEEE 2001 Micro Electro Mechanical System Workshop, Interlaken, Switzerland, pp. 409–412.

[8] Evans, J. D., and Liepmann, D., 1999, ''The Bubble Spring and Channel (BSAC) Valve: An Actuated, Bi-stable Mechanical Valve for In-Plane Fluid Control,'' *Transducer'99*, Sendai, Japan, pp. 1122–1125.

[9] Lin, L., Udell, K. S., and Pisano, A. P., 1994, ''Liquid-Vapor Phase Transition and Bubble Formation in Micro Structures,'' Therm. Sci. Eng., **2 pp**, 52–59.

[10] Yang, W. J., and Tsutsui, K., 2000, ''Overview of Boiling on Microstructures-Macro Bubbles From Micro Heaters,'' Microscale Thermophys. Eng., **4**, No. 1, pp. 7–24.

[11] Cole, R. 1974, ''Boiling Nucleation,'' in *Advances in Heat Transfer*, Vol. 10, Academic, New York.

[12] Hsu, Y. Y., 1962, ''On the size Range of Active Nucleation Cavities on a Heating Surface,'' ASME Journal of Heat Transfer, **84C**, pp. 207–216.

[13] Sernas, V. and Hooper, F. C., 1969, ''The Initial Vapor Bubble Growth on a Heated Wall during Nucleation Boiling,'' Int. J. Heat Mass Transf., **12**, pp. 1627–1639.

[14] Moore, F. D. and Mesler, R. B., 1961, ''The Measurement of Rapid Surface Temperature Fluctuations During Nucleate Boiling of Water,'' AIChE J., **7**, No. 4, pp. 620–624.

[15] Koester, D., Mhedevan, R., and Marcus, K., 1999, *Multi-User MEMS Processes (MUMPS) Design Handbook*, rev. 4, May 1999, Cronos Integrated Microsystems, 3021 Cornwallis Road, Research Triangle Park, NC, 27709.

[16] Mastrangelo, C. H., 1991, Ph.D. thesis, University of California at Berkeley, Berkeley, CA.

[17] Lin, L., 1993, Ph.D. thesis, University of California at Berkeley, Berkeley, CA.

[18] Carslaw, H. S., and Jaeger, J. C., 1959, *Conduction of Heat in Solids*, 2nd ed., Oxford University Press, London, pp. 99–100.

[19] Vargaftik, N. B., 1975, *Tables on the Thermophysical Properties of Liquids and Gases*, Hemisphere Pub. Corp., Bristol, PA, pp. 418–420.

[20] Stralen, S. V., and Cole, R., 1979, *Boiling Phenomena*, Vol. 1, Hemisphere Pub. Crop., Bristol, PA, pp. 454–456.

[21] Goldstein, R. J., Volino, R. J., 1995, ''Onset and Development of Natural Convection Above a Suddenly Heated Horizontal Surface,'' ASME Journal of Heat Transfer, , **117**, No. 4, pp. 808–821.

[22] Shimokubo, T., Takagi, S., and Inoue, T., 1998, ''Boiling Bubble Behavior form the Micro Heater,'' *Proc. 35th Heat Transfer Symp. of Japan*, pp. 173–174.

Enhanced Boiling of FC-72 on Silicon Chips With Micro-Pin-Fins and Submicron-Scale Roughness

H. Honda

H. Takamastu

Institute of Advanced Material Study,
Kyushu University,
Kasuga, Fukuoka 816-8580, Japan

J. J. Wei

Interdisciplinary Graduate School of Engineering Sciences,
Kyushu University,
Kasuga, Fukuoka 816-8580, Japan

Experiments were conducted to study the effects of micro-pin-fins and submicron-scale roughness on the boiling heat transfer from a silicon chip immersed in a pool of degassed and gas-dissolved FC-72. Square pin-fins with fin dimensions of $50\times50\times60\ \mu m^3$ (width $\times$thickness$\times$height) and submicron-scale roughness (RMS roughness of 25 to 32 nm) were fabricated on the surface of square silicon chip ($10\times10\times0.5\ mm^3$) by use of microelectronic fabrication techniques. Experiments were conducted at the liquid subcoolings of 0, 3, 25, and 45 K. Both the micro-pin-finned chip and the chip with submicron-scale roughness showed a considerable heat transfer enhancement as compared to a smooth chip in the nucleate boiling region. The chip with submicron-scale roughness showed a higher heat transfer performance than the micro-pin-finned chip in the low-heat-flux region. The micro-pin-finned chip showed a steep increase in the heat flux with increasing wall superheat. This chip showed a higher heat transfer performance than the chip with submicron-scale roughness in the high-heat-flux region. The micro-pin-finned chip with submicron-scale roughness on it showed the highest heat transfer performance in the high-heat-flux region. While the wall superheat at boiling incipience was strongly dependent on the dissolved gas content, it was little affected by the liquid subcooling. [DOI: 10.1115/1.1447937]

Keywords: Boiling, Devices, Enhancement, Finned Surfaces, Heat Transfer, Roughness

Introduction

Direct liquid cooling by use of dielectric liquids has been considered as one of the promising cooling schemes for high-powered electronic components. Since 1970s, active studies have been conducted in this area. A comprehensive review of relevant literature has been given by Bar-Cohen [1]. One of the difficulties associated with the use of dielectric liquids with a high wettability is a high wall superheat required for boiling incipience. You et al. [2] studied the boiling incipience of R-113 on smooth surfaces. The boiling incipience superheat was highly non-repeatable and the data was best presented in statistical form. A number of studies have dealt with the enhancement of boiling heat transfer from electronic devices by use of treated surfaces. The enhancement techniques studied include a ''dendritic heat sink'' (a brush-like structure on a silicon chip) [3], laser drilled holes on a silicon chip [4], mechanically manufactured heat sinks with drilled holes, micro-fins, micro-channels etc. [5–7], hexagonal dimples fabricated on a silicon chip by photo-etching [8], alumina particle spraying on a simulated chip [9], painting of silver flakes or diamond particles on a simulated chip [10,11] and micro-reentrant cavities fabricated on a silicon chip by microelectronic fabrication techniques [12]. In the painting technique, the highest enhancement level was achieved by the painting of 3–10 μm silver flakes with cavity feature sizes of about 1 μm [10] or 8–12 μm diamond particles with cavity feature sizes of 1–10 μm [11].

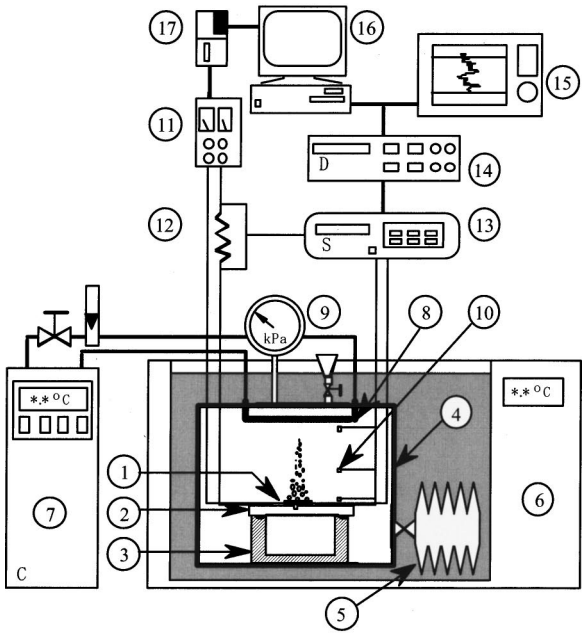
While most of the foregoing experiments with treated surfaces have been conducted for saturated conditions with negligible dissolved gas content in the liquid, Kubo et al. [12] and O'Connor et al. [13] studied the effect of dissolved gas for smooth and enhanced surfaces. In their experiments FC-72 was gas-saturated. Thus the dissolved gas content of the bulk liquid y_g was dependent on the total pressure and the degree of subcooling. For the smooth surface, the boiling incipience and nucleate boiling wall superheats decreased with increasing y_g. However, the effect was negligible at low y_g ($<$0.004 mole air/mole liquid). For the diamond painted surface, on the other hand, the wall superheats for a given total pressure were less sensitive to the dissolved gas content due to localized degassing of the liquid.

The present paper reports on the boiling heat transfer of FC-72 from silicon chips with micro-pin-fins and submicron-scale roughness. The advantage of these enhanced surfaces is that they can be fabricated on the back surface of actual LSI chips by use of microelectronic fabrication techniques. The micro-pin-fins were fabricated on a silicon chip by use of a dry etching technique. The submicron-scale roughness was fabricated by depositing a thin layer of SiO_2 on a silicon chip by sputtering and then wet etching of the SiO_2 surface. Experiments are conducted under the atmospheric condition with degassed and gas-dissolved FC-72 (T_{sat} = 56 deg) at the liquid subcoolings of 0, 3, 25, and 45 K as test liquids. The results are compared with those for a smooth chip and previously developed enhanced surfaces.

Experimental Apparatus and Procedure

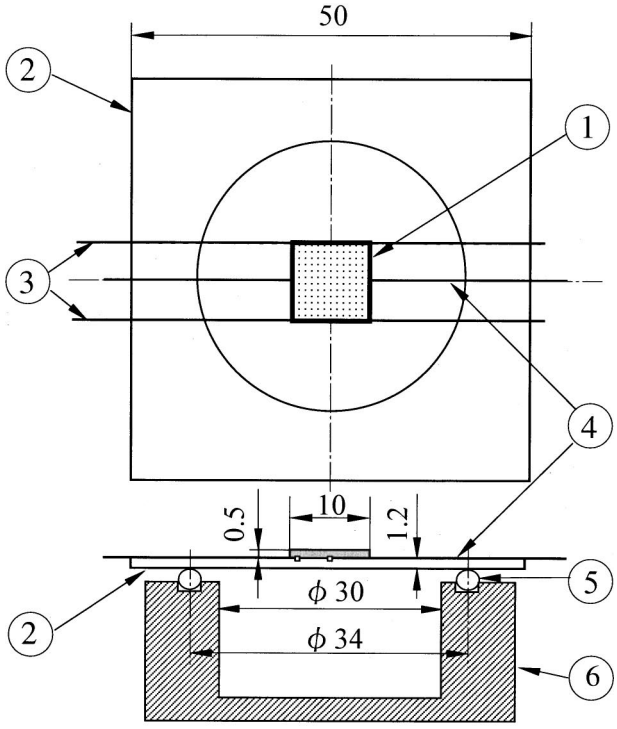
The experimental apparatus is shown schematically in Fig. 1. A rectangular test vessel filled with FC-72 was fixed inside a water bath. A test section consisting of a test chip bonded on a glass plate and a vacuum chuck made of brass was immersed in the test chamber. The bulk temperature of FC-72 in the test vessel was maintained at a prescribed value by controlling the water temperature inside the water bath. In order to maintain the pressure in the test chamber at the atmospheric pressure, a rubber bag was attached to the test chamber. The test chamber was fitted with glass windows for the visual observation of boiling phenomena. The local temperatures of the test liquid at the chip level, and 40 mm and 80 mm above the chip level were measured by *T*-type thermocouples that were fixed to a vertical wire located 25 mm apart from the edge of the test chip. Most of the experiments were conducted at a horizontal upward orientation. A number of experi-

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division April 30, 2001; revision received August 21, 2001. Associate Editor: G. P. Peterson.



1. Test chip
2. Glass plate
3. Vacuum chuck
4. Test vessel
5. Rubber bag
6. Water bath
7. Cooling unit
8. Condenser
9. Pressure gauge
10. Thermocouples
11. DC power supply
12. Standard resistor
13. Scanner
14. Digital multimeter
15. Pen recorder
16. Computer
17. Power supply controller

Fig. 1 Schematic diagram of experimental apparatus



1. Silicon chip
2. Pyrex glass plate
3. Copper lead wire
4. Thermocouple
5. O ring
6. Vacuum chuck

Fig. 2 Details of test section

ments were also conducted at a vertical orientation to take the pictures of boiling phenomena using a still camera and a high-speed video camera.

Details of the test section are shown in Fig. 2. The glass plate on which the test chip was bonded was pressed firmly to the o-ring on the vacuum chuck when the inside of the chuck was evacuated by using a vacuum pump. This facilitated an easy exchange of the test chip and minimization of conduction heat loss. The test chip was a *P* doped N-type silicon chip with dimensions of $10\times10\times0.5\ \text{mm}^3$. The resistance of the smooth chip and the chip with submicron-scale roughness was about 38 Ω at 60°C. The resistance was somewhat larger for chips with micro-pin-fins. The chip was Joule heated by using a regulated d.c. power supply. Two 0.25 mm dia. copper wires for power supply and voltage drop measurement were soldered to the side surfaces at the opposing ends. A special solder with the melting point of 300°C was first used to secure the Ohmic contact between the test chip and the copper wire. Then a low temperature solder with the melting point of 180°C was applied to increase the mechanical strength of the joint. The local wall temperatures at the center and at about 1.5 mm from the edge on the centerline of the chip were measured by 0.12 mm dia. T-type thermocouples bonded under the test chip. The thermocouples for the measurements of liquid and wall temperatures were calibrated by using temperature controlled water and oil baths. The bath temperature was measured by a platinum resistance thermometer with an accuracy of 0.03 K. The e.m.f.s of the thermocouples were read and recorded consecutively eight times by a data logger to 1 μV and the average values of the eight measurements were adopted as the experimental data. The average chip temperature was also measured by the resistance thermometry. In order to obtain a calibration curve between the temperature and electric resistance of the test chip, the test chip was immersed in a bath of FC-40 ($T_{sat}=150$°C at atmospheric pressure) that was immersed in a temperature-controlled oil bath. The test chip and a standard resistor of 1 Ω were connected in series to a d.c. power supply and a constant current of about 1 mA was passed through the circuit. The bath temperature was measured by the platinum resistance thermometer and the voltage drops of the test chip (of the order of 40 mV) and the standard resistor (about 1 mV) were measured by the data logger in the same manner as the case of the thermocouples. The sensitivity of the resistance thermometry was about 0.24 mV/K, which was much higher than that of the T-type thermocouple (about 0.043 mV/K). The uncertainties in the chip temperature measurements by the thermocouple and the resistance thermometry are estimated to be less than 0.1 K.

Micro-pin-fins with square cross-sections were fabricated on the surface of silicon chip by use of the dry etching technique. The fin dimensions were $50\times50\times60\ \mu\text{m}^3$ (width×thickness ×height), and the fin pitch was twice the fin width. Thus the surface area enhancement ratio as compared to a smooth chip is 2.2. This chip is denoted as chip PF. The scanning-electron-micrograph (SEM) image of chip PF is shown in Fig. 3. The surface with submicron-scale roughness was fabricated by depositing a thin layer (about 3 μm) of SiO_2 on the smooth chip by sputtering and then dipping it into the mixture of HF, NH_4F and water for wet etching. This chip is denoted as chip E. The micro-pin-finned chip with submicron-scale roughness was fabricated in the same manner as the case of chip E. This chip is denoted as chip EPF. The atomic-force-microscope (AFM) images of the smooth chip (chip S) and the fin tip of chip EPF are shown in Figs. 4(*a*) and 4(*b*), respectively. Chip S has a very smooth surface

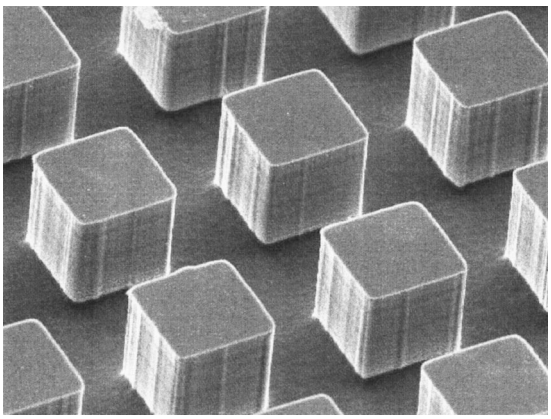
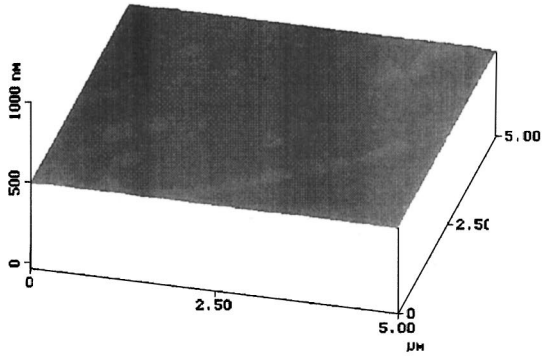


Fig. 3 SEM image of chip PF

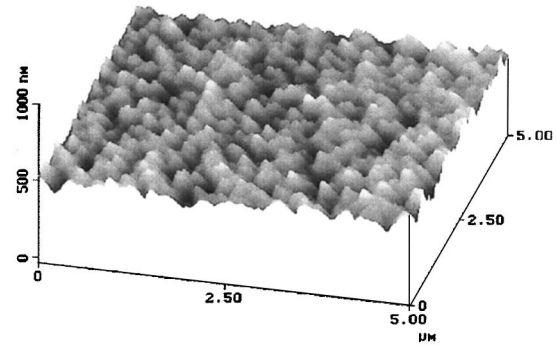
with the root-mean-square (RMS) roughness of 2.3 nm, whereas chip EPF has a very fine roughness with the RMS roughness of 32.4 nm. It is relevant to note here that chip E had the RMS roughness of about 25 nm. Considering the fact that the vertical scale of Fig. 4 is 2.1 times as large as that of the horizontal scale, the surface area enhancement ratios of chips E and EPF are close to 1.0 and 2.2, respectively.



(*a*) Chip S

RMS Roughness: 2.3 nm

Max height: 32.5 nm



(*b*) Chip EPF

RMS Roughness: 32.4 nm

Max height: 271.3 nm

Fig. 4 AFM images of chips S and EPF

Experiments were conducted at the liquid subcoolings ΔT_{sub} of 0, 3, 25, and 45 K. Prior to the experiments, the test fluid was degassed by maintaining the water temperature inside the water bath at about 63°C (i.e., about 7 K above the normal boiling point of FC-72) for 4 hours while operating the downward facing condenser located at the ceiling of the test vessel. For the cases of $\Delta T_{sub}=25$ and 45 K, experiments were also conducted with a gas-dissolved FC-72. In these cases a pressurized air was bubbled into the test liquid for about 12 hours and then the test liquid was exposed to ambient air for 12 hours. The mole fraction of dissolved air y_g was measured by a gas chromatograph. The measured value of y_g was about $2.1\times10^{-4}-3.4\times10^{-4}$ and $2.6\times10^{-3}-3.3\times10^{-3}$ for the degassed and gas-dissolved FC-72, respectively. The measured air contents of the test liquid before and after a series of experiments were almost unchanged.

Since the boiling incipience was affected by the initial condition of the chip surface, experiments were started one day after the test liquid was degassed or gas-dissolved. Power input to the test chip was increased in small steps up to the critical heat flux point (CHF point). In order to prevent the occurrence of real burn out, an overheating protection system was incorporated in the power circuit. The temperatures of the test chip and the bulk liquid were monitored by a pen recorder. After a steady state was reached, e.m.f.s of the thermocouples for the wall temperature and liquid temperature measurements and the voltage drops of the test chip and the standard resistor were read and recorded eight times using the data logger to 1 μV and the average values of these measurements were adopted as the experimental data. The highest value of the wall temperature just before the boiling incipience was read from the chart of the pen recorder. The wall temperature T_w was defined by the measured value at the center on the back surface of test chip. The bulk temperature of test liquid T_b was defined by the measured value at the chip level. The uncertainty in the measured T_b is estimated to be less than 0.1 K. The heat flux q (defined on the basis of base area) was obtained from the voltage drop of the test chip and the electric current. The uncertainty in the calculated value of q is estimated to be less than 5 percent. The major source of error in the measurement of q is the conduction heat loss to the bulk liquid through the copper lead wires and the glass substrate. Experimental results for degassed FC-72 (see Figs. 9, 10, and 12) revealed that the measured heat flux in the free convection regime was about 80 percent higher than the prediction of heat transfer correlation for upward-facing horizontal flat plates [14]. Typically, the increase in the measured heat flux above the free convection value was 0.7, 1.4, and 2.4 W/m^2 at $\Delta T_{sat}=20$ K for $\Delta T_{sub}=0$, 25 and 45 K, respectively.

The experiment was repeated three times for all chips. The time interval between the subsequent runs was about 0.5 hours. The boiling curves showed a good repeatability for all cases except for the boiling incipience point. Thus, only the results for the third runs are presented in the next section. Most of the experiments were conducted at the horizontal upward orientation of the chip. Additional experiments with the vertically mounted chip EPF was conducted to take video pictures of the boiling phenomena.

Experimental Results and Discussion

Figure 5 shows the variation of measured liquid and wall temperatures with q for the case of horizontally mounted chip EPF immersed in degassed FC-72 at $\Delta T_{sub}=25$ K. In Fig. 5, T_{wc} and T_{we} denote the local wall temperatures at the center and edge, T_{wr} the average wall temperature measured by the resistance thermometry, and T_{b1}, T_{b2}, and T_{b3} the local liquid temperatures at the chip level, 40 mm above the chip level and 80 mm above the chip level, respectively. The values of T_{wc}, T_{we}, and T_{wr} increase rapidly as q increases up to about 76°C. Then they show a large dip of about 14 K and then increase first rapidly, then slowly, up to the CHF point. As expected, the value of T_{wc} is greater than that

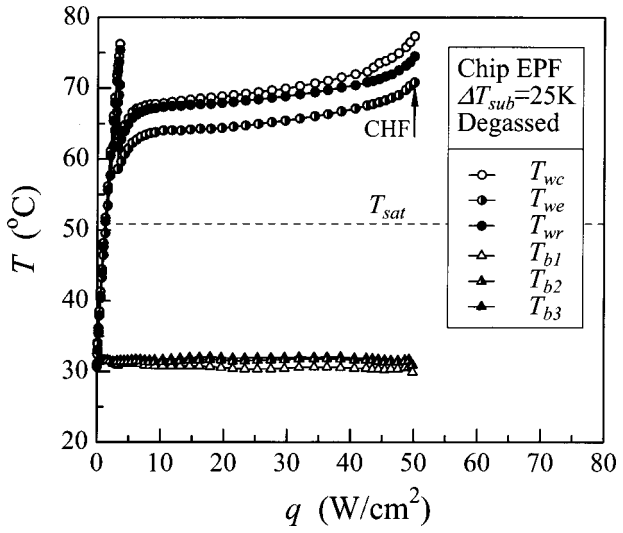


Fig. 5 Variation of measured wall and liquid temperatures with heat flux; ΔT_{sub}=25 K, y_g=2.1×10^{-4}~3.4×10^{-4}

of T_{we}. The difference between them increases as q increases. The maximum difference is about 6.5 K at the CHF point. The value of T_{wr} lies in between T_{wc} and T_{we}. The value of T_{b1} is maintained closely at 31°C. The values of T_{b2} and T_{b3} are slightly higher than T_{b1}.

Figures 6(*a*)–6(*c*) show the photographs of boiling phenomena of gas dissolved FC-72 at ΔT_{sub}= 25 K on the vertically mounted chip EPF. The positions on the boiling curve corresponding to these figures are shown in Fig. 8. In Fig. 6(*a*) for the low-heat-flux region, the number of vapor bubbles increases as the height from the lower edge of chip increases. It is also seen that the distribution of bubble nucleation site is not uniform on the surface. In Fig. 6(*b*) for the medium-heat-flux region, the distribution of the bubbles is almost uniform except for a small portion near the lower left end. The bubbles have a range of diameters and larger bubbles are observed at the upper part of the chip. This indicates that the bubbles increase their size as they move upward along the surface and merge with the other bubbles. In Fig. 6(*c*) for the high-heat-flux region, large secondary bubbles that cover most of the chip surface are observed.

Figure 7 shows a sequence of boiling phenomena that corresponds to point (*d*) on the boiling curve in Fig. 8. The arrow in the figure indicates one of the points where bubble growth occurs. At

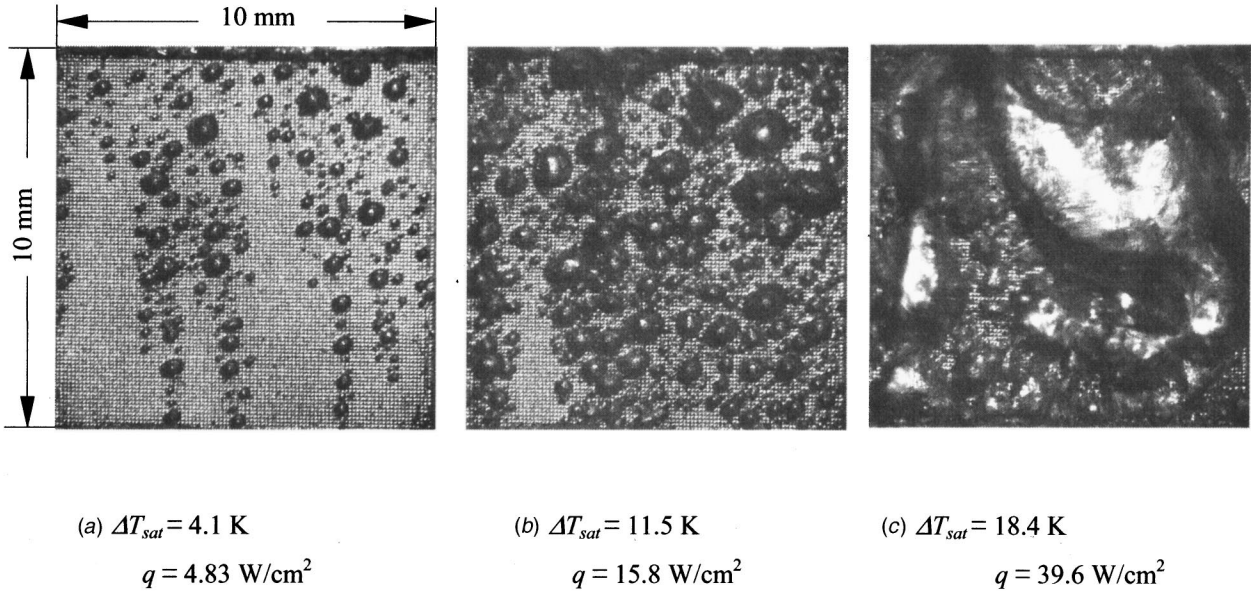


Fig. 6 Photographs of boiling phenomena on vertically mounted chip EPF; ΔT_{sub}=25 K, gas dissolved (y_g=2.6×10^{-3} ~3.3×10^{-3})

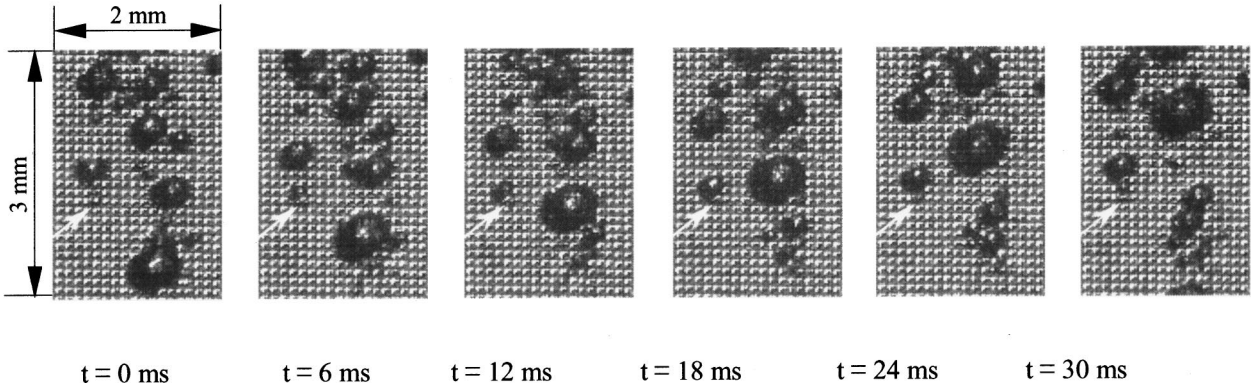


Fig. 7 Sequence of boiling phenomena on vertically mounted chip EPF; ΔT_{sub}=25 K, gas dissolved (y_g=2.6×10^{-3}~3.3 ×10^{-3}); ΔT_{sat}=9.1 K; q=8.74 W/cm^2

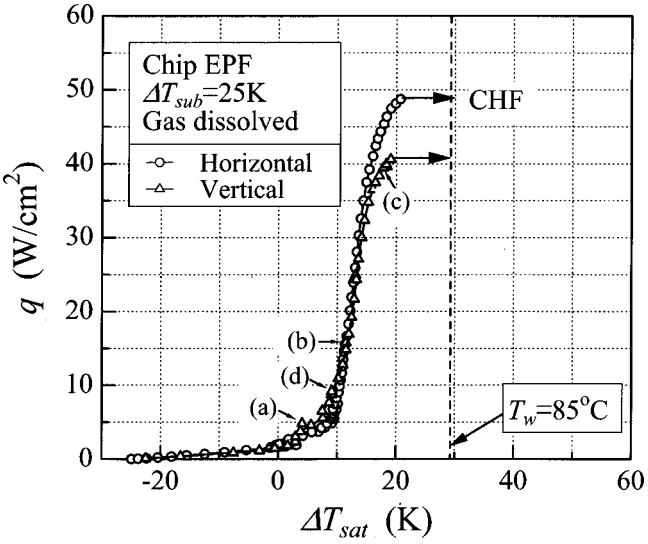


Fig. 8 Comparison of boiling curves for vertically mounted and horizontally mounted chip EPF; ΔT_{sub}=25 K, y_g=2.6 ×10^{-3}~3.3×10^{-3}

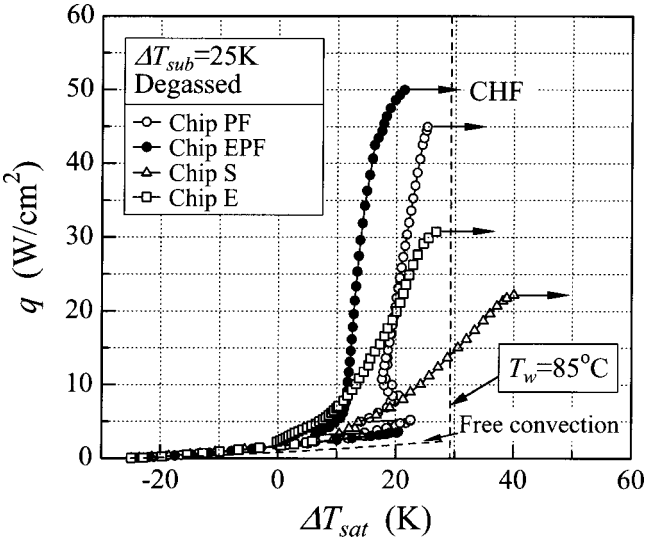


Fig. 10 Comparison of boiling curves; ΔT_{sub}=25 K, y_g=2.1 ×10^{-4}~3.4×10^{-4}

t=0 ms, a very small bubble is left within the gap between adjacent pin fins as a larger bubble moves upward. Then the small bubble grows as time passes. The bubble slides upward along the surface due to the buoyancy force, thereby leaving a small amount of vapor within the gap. At t=30 ms, the next cycle of bubble growth is started. The above results suggest that a sharp increase in the heat flux with increasing wall superheat observed for chips PF and EPF (see Figs. 8–13) is associated with the evaporation that occurred within the small gap.

In Fig. 8, the boiling curve for horizontally mounted chip EPF is also shown for comparison. The other experimental conditions are the same as the case of vertically mounted one. It is seen from Fig. 8 that the vertically mounted chip shows a slightly higher heat flux in the low-wall-superheat region (ΔT_{sat}<12 K). In the high-wall-superheat region (ΔT_{sat}>12 K), on the other hand, the horizontally mounted chip shows a higher heat flux. The values of the critical heat flux q_{CHF} are 41 and 49 W/cm^2 for the vertical and horizontal orientations, respectively. This difference may be attributed to the difference in the behavior of large secondary bubbles formed on the chip surface in the high heat flux region. It was revealed by the observation of boiling phenomena using the high speed video that the average residence time of large secondary bubbles on the chip surface was 43 and 39 ms for the vertical and horizontal orientations, respectively. The critical heat flux point is reached when a thin liquid film formed under a large secondary bubble evaporates completely and dry out of the surface begins. Thus, q_{CHF} is supposed to be inversely proportional to the residence time of the secondary bubble if the same amount of liquid film is left on the surface when the secondary bubble grows. Another reason may be ascribed to the flow direction of large secondary bubbles. In the case of vertically mounted chip, the bubble moves upward along the surface. Thus the inflow of liquid toward the chip after the bubble detaches is mainly from the lower end. In the case of horizontally mounted chip, the bubble moves upward and detaches the chip surface. This causes inflow of liquid from all sides of the chip. Thus the amount of liquid supplied to the portion on the chip where the dry out first occurs at the CHF point (i.e., center of the upper end for the vertical orien-

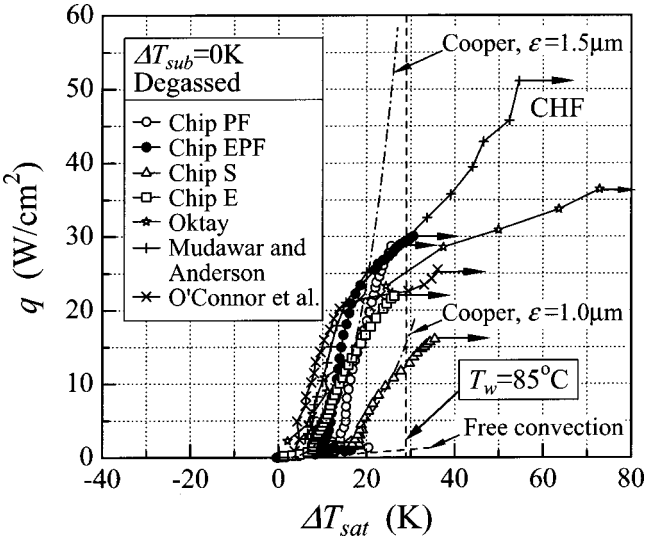


Fig. 9 Comparison of boiling curves; ΔT_{sub}=0 K, y_g=2.1 ×10^{-4}~3.4×10^{-4}

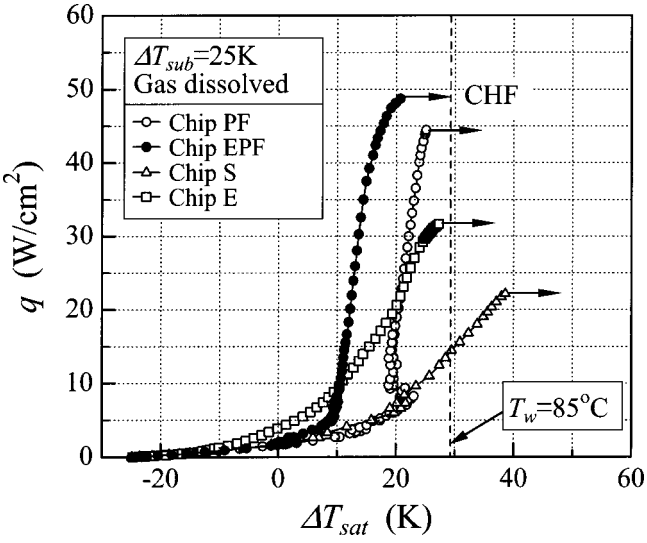


Fig. 11 Comparison of boiling curves; ΔT_{sub}=25 K, y_g=2.6 ×10^{-3}~3.3×10^{-3}

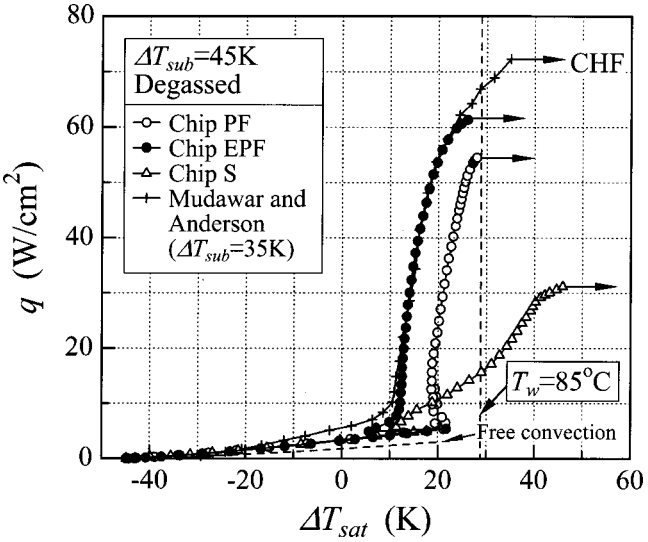


Fig. 12 Comparison of boiling curves; ΔT_{sub}=45 K, y_g=2.1 ×10^{-4}~3.4×10^{-4}

tation and center of the chip for the horizontal orientation) is supposed to be larger for the horizontally mounted chip. It is relevant to note here that the ratio of q_{CHF} between the two cases agrees well with the prediction of the empirical equation proposed by Chang and You [15] that correlated the heater orientation effect on the critical heat flux for smooth and enhanced surfaces.

Figure 9 compares the boiling curves for all chips at ΔT_{sub} = 0 K (degassed FC-72). As expected, all chips show almost the same heat transfer performance in the non-boiling region. The wall superheat at the boiling incipience $\Delta T_{sat,i}$ is about 18, 11.8, 20.4, and 16.2 K for chips S, E, PF, and EPF, respectively. However, it should be mentioned here that for all chips the boiling incipience occurred at the edge of the chip where the lead wire was soldered. Thus it is probable that a much higher $\Delta T_{sat,i}$ is realized when the end effects are completely removed. In this connection, it is relevant to note here that the boiling curve for chip S agrees closely with that for a polished copper surface (5 ×18 mm^2) reported by O'Connor et al. [13] except that the value of $\Delta T_{sat,i}$ is a little higher for the latter (20 K). The boiling curves for chips E, PF and EPF move toward a smaller ΔT_{sat} region as

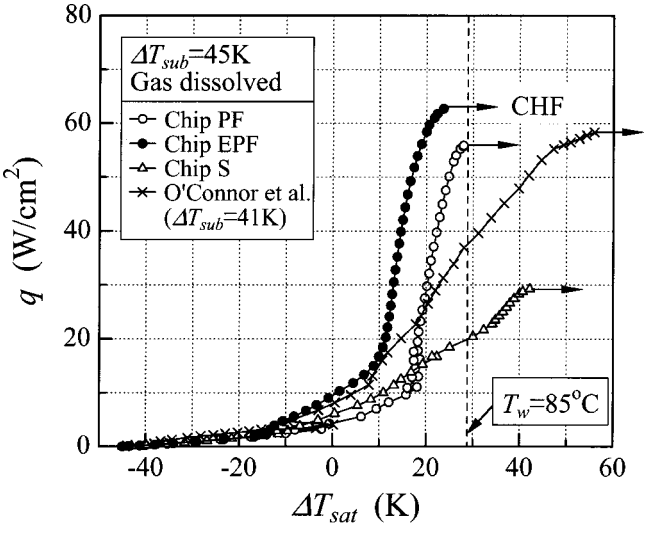


Fig. 13 Comparison of boiling curves; ΔT_{sub}=45 K, y_g=2.6 ×10^{-3}~3.3×10^{-3}

compared to chip S. Thus both the submicron-scale roughness and micro-pin-fins are effective in enhancing boiling heat transfer. It is also seen from Fig. 9 that chips PF and EPF show a sharp increase in q in the nucleate boiling region. For chip EPF with a thin layer of SiO_2 on the surface, the slope decreases as the CHF point is approached. The critical heat flux increases in the order of chips S, E, PF, and EPF. In Fig. 9, the prediction of empirical equation for free convection on upward-facing horizontal surfaces proposed by Fishenden and Saunders [14], that for nucleate boiling proposed by Cooper [16], and previous experimental results for enhanced surfaces reported by O'Connor et al. [13], Oktay [17] and Mudawar and Anderson [18], are also shown for comparison. The heat flux in the non-boiling region is about 80 percent higher than the predicted value. This is probably due to the heat loss to the bulk liquid through the copper lead wires and glass substrate. Since the Cooper correlation [16] includes the surface roughness as a parameter, it is fitted to the experimental data for chips S and E by using appropriate values of the surface roughness ε (i.e., 1.0 and 1.5 μm, respectively). These values are about 430 and 60 times as large as the measured RMS roughnesses, respectively. The test surface of O'Connor et al. [13] had a porous layer consisting of 8-to-12 μm diamond particles produced by the painting technique [11]. That of Oktay [17] had about 1-mm high dendritic structure. The test surface of Mudawar and Anderson [18] had square micro-pin-fins with the dimensions of 0.305×0.305 ×0.508 mm^3 (width×thickness×height). These fin dimensions are one order-of-magnitude greater than those of chips PF and EPF in the present experiments. The fin pitch was twice the fin width. Generally, the boiling curves for these surfaces show a trend close to that for chip EPF. The ΔT_{sat} values in the low-heat-flux region are smaller than that for chip EPF, whereas the wall superheats at the CHF point $\Delta T_{sat,CHF}$ are greater than that for the latter.

Figure 10 shows the boiling curves for all chips immersed in the degassed FC-72 with ΔT_{sub}= 25 K. All chips follow almost the same q versus ΔT_{sat} relation in the non-boiling region. Thus the micro-pin-fin and submicron-scale roughness are not effective in enhancing free convection. This indicates that the heights of these surface structures are below the thickness of thermal boundary layer formed on the chip surface. In Fig. 10, the prediction of free convection correlation [14] is also shown for comparison. As was the case for ΔT_{sub}= 0 K shown in Fig. 9, the experimental data in the non-boiling region is about 80 percent higher than the correlation. The temperature overshoot at the boiling incipience is clearly observed for chips S, PF, and EPF, and the values of $\Delta T_{sat,i}$ are fairly close to those for ΔT_{sub}= 0 K. It is interesting to note that Chip PF shows a decrease in ΔT_{sat} as q increases at around q= 8 W/cm^2. Visual observation revealed that this was due to a rapid increase in the number of nucleation sites. In the nucleate boiling region, the trend of experimental data is basically the same as the case of ΔT_{sub}= 0 K except for the fact that the critical heat flux increases considerably.

Figure 11 shows the boiling curves for all chips immersed in the gas-dissolved FC-72 with ΔT_{sub}= 25 K. In this case boiling begins at a negative value of ΔT_{sat} and the temperature overshoot at the boiling incipience is not observed. Comparison of Figs. 10 and 11 reveals that while the value of q in the low-heat-flux region is higher for the gas-dissolved FC-72, the boiling curve in the high-heat-flux region agrees closely between the degassed and gas-dissolved FC-72. This may be ascribed to the decrease in y_g near the chip surface with increasing q that occurs for the gas-dissolved FC-72. As described in [13], the dissolved gas enhances bubble nucleation on the chip surface in the low-heat-flux region. The dissolved gas is removed from the surrounding liquid as the vapor bubbles grow. Thus, the dissolved gas content in the liquid near the chip surface will decrease as q increases. Consequently, in the high-heat-flux region, the value of y_g near the chip surface will be virtually the same between the cases of degassed and gas-dissolved FC-72.

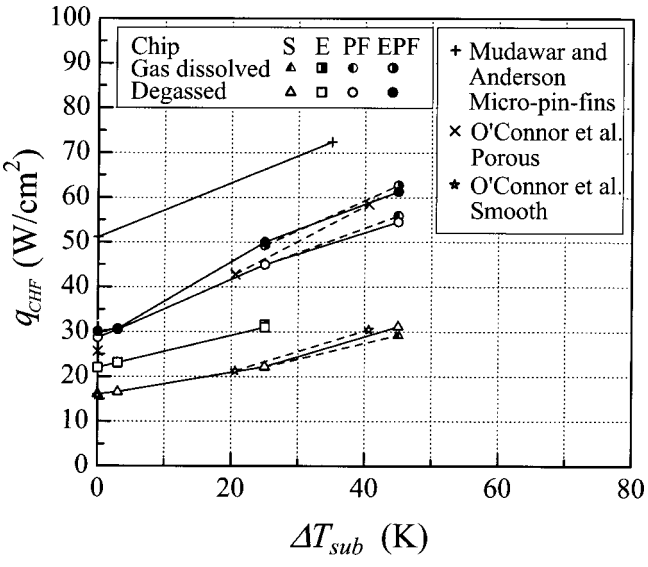


Fig. 14 Variation of q_{CHF} with ΔT_{sub}

Figures 12 and 13, respectively, show the boiling curves for chips S, PF, and EPF immersed in the degassed and gas-dissolved FC-72 with $\Delta T_{sub}=45$ K. As expected, the value of q at a prescribed value of ΔT_{sat} is higher than the corresponding case for $\Delta T_{sub}=25$ K shown in Figs. 10 and 11. The results for enhanced surfaces reported by Mudawar and Anderson [18] and O'Connor et al. [13] are respectively shown in Figs. 12 and 13 for comparison. It should be mentioned here that the values of ΔT_{sub} for these data are smaller than the present case. The boiling curve of Mudawar and Anderson [18] for the surface with micro-pin-fins is close to that of chip EPF and shows a higher q_{CHF} than the latter. The boiling curve of O'Connor et al. [13] for the surface fabricated by the painting technique shows a smaller slope than chips PF and EPF.

In Figs. 8–13, the upper limit of wall temperature for the reliable operation of LSI chips ($T_w=85$°C) is also shown. Comparison with the experimental results reveals that the wall temperature at the CHF point $T_{w,CHF}$ is higher than 85°C for chip S and the previous results reported by O'Connor et al. [13], Oktay [17], and Mudawar and Anderson [18]. On the other hand, chips E, PF, and EPF show $T_{w,CHF}$ smaller than 85°C (excepting the case of chip EPF at $\Delta T_{sub}=0$ K where $T_{w,CHF}=87$°C).

Figure 14 shows q_{CHF} for all chips plotted as a function of ΔT_{sub}. The q_{CHF} increases almost linearly as ΔT_{sub} increases. The values of q_{CHF} for the degassed and gas-dissolved FC-72 are fairly close to each other. As stated previously, q_{CHF} increases in the order of chips S, E, PF, and EPF. Chip E shows the q_{CHF} value about 1.4 times as large as that for chip S. Chips PF and EPF, respectively, show the q_{CHF} values about 1.8 to 2.0 times and 1.8 to 2.3 times as large as that for chip S. These values are a little smaller than or comparable to the surface area enhancement ratio (=2.2). The enhancement in q_{CHF} is more significant for subcooled boiling than for saturated boiling. In Fig. 14, the experimental data by O'Connor et al. [13] and Mudawar and Anderson [18] are also shown for comparison. The data for a smooth copper surface reported by O'Connor et al. [13] agree well with that for chip S. The porous surface of O'Connor et al. [13] shows the q_{CHF} value close to that for chip EPF at high ΔT_{sub}. However, the former shows a smaller q_{CHF} than the latter as ΔT_{sub} decreases (i.e., as the size of the secondary bubble increases). The surface of Mudawar and Anderson [18] shows the highest q_{CHF}. However, as seen from Fig. 13, $T_{w,CHF}$ is higher than 85°C and q at $T_w=85$°C is close to q_{CHF} for chip EPF.

Conclusions

Three types of micro-scale surface structures (submicron-scale roughness, micro-pin-fins and micro-pin-fins with submicron-scale roughness on it) were tested to enhance boiling of FC-72 on a silicon chip. Most of the experiments were conducted at the horizontal upward orientation of the chip. Photographic observation of the boiling phenomena was made at the vertical orientation. The results may be summarized as follows:

1 Both the micro-pin-fins ($50\times50\times60\ \mu m^3$ in width $\times$thickness$\times$height) and the submicron-scale roughness (RMS roughness of 25 to 32 nm) were effective in enhancing heat transfer in the nucleate boiling region and increasing the critical heat flux. While the chip with submicron-scale roughness showed a higher heat transfer performance than the micro-pin-finned chip in the low-heat-flux region, the former showed a lower heat transfer performance than the latter in the high-heat-flux region.

2 The boiling curve of the micro-pin-finned chip was characterized by a sharp increase in the heat flux with increasing wall superheat. The wall temperature at the CHF point was less than the upper limit for the reliable operation of LSI chips ($T_w=85$°C).

3 Photographic observation of boiling phenomena on the micro-pin-finned chip revealed that a small amount of vapor is left within the gap between pin fins when a growing bubble leaves the surface. This suggested that the evaporation that occurred within the small gap was responsible for a sharp increase in the heat flux with increasing wall superheat.

4 The micro-pin-finned chip with submicron-scale roughness showed the highest heat transfer performance in the high-heat-flux region. The critical heat flux was 1.8 to 2.3 times as large as that for a smooth chip. These values were a little smaller than or comparable to the surface area enhancement (=2.2). The enhancement was more significant for subcooled boiling than for saturated boiling.

5 For all chips, the critical heat flux increased almost linearly with liquid subcooling. This trend was in accord with the previous results for enhanced surfaces reported by O'Connor et al. [13] and Mudawar and Anderson [18].

6 The gas-dissolved FC-72 showed a marked decrease in the boiling incipience temperature. As a result, the heat transfer performance in the low-heat-flux region was higher than the case of degassed FC-72. However, the heat transfer performance in the high-heat-flux region was close to each other.

Acknowledgment

The authors acknowledge the support of the Ministry of Education, Science, Sports and Culture of Japan under Grant-in-Aid for Scientific Research No. 12450089.

Nomenclature

q = heat flux based on projected surface area (W/cm^2)
q_{CHF} = critical heat flux (W/cm^2)
T_b = temperature of bulk liquid (°C)
T_{sat} = saturation temperature (°C)
T_w = wall temperature (°C)
$T_{w,CHF}$ = wall temperature at CHF point (°C)
$T_{w,i}$ = wall temperature at boiling incipience (°C)
ΔT_{sat} = wall superheat $=T_w-T_{sat}$ (K)
$\Delta T_{sat,i}$ = wall superheat at boiling incipience $=T_{w,i}-T_{sat}$ (K)
ΔT_{sub} = liquid subcooling $=T_{sat}-T_b$ (K)
y_g = mole fraction of dissolved air (−)
ε = surface roughness (μm)

References

[1] Bar-Cohen, A., 1993, "Thermal Management of Electronic Components With Dielectric Liquids," JSME Int. J., Ser. B **36**, pp. 1–25.
[2] You, S. M., Simon, T. W., Bar-Cohen, A., and Tong, W., 1990, "Experimental

Investigation of Nucleate Boiling Incipience With a Highly Wetting Dielectric Fluid (R-113),'' Int. J. Heat Mass Transf., **33**, pp. 105–117.
[3] Oktay, S., and Schmeckenbecher, A., 1972, ''Method for Forming Heat Sinks on Semiconductor Device Chips,'' U.S. Patent No. 3,706,127.
[4] Chu, R. C., and Moran, K. P., 1977, ''Method for Customizing Nucleate Boiling Heat Transfer from Electronic Units Immersed in Dielectric Coolant,'' U.S. Patent No. 4,650,507.
[5] Nakayama, W., Nakajima, T., and Hirasawa, S., 1984, ''Heat Sink Studs Having Enhanced Boiling Surfaces for Cooling Microelectronic Components,'' ASME Paper No. 84-WA/HT-89.
[6] Anderson, T. M., and Mudawar, I., 1989, ''Microelectronic Cooling by Enhanced Pool Boiling of a Dielectric Fluorocarbon Liquid,'' ASME J. Heat Transfer, **111**, pp. 752–759.
[7] Carvalho, R. D., and Bergles, A. E., 1990, ''The Influence of Subcooling on the Pool Nucleate Boiling and Critical Heat Flux of Simulated Electronic Chips,'' *Proc. 9th Int. Heat Transfer Conf.*, Vol. **2**, pp. 289–294.
[8] Miller, W. J., Gebhart, B., and Wright, N. T., 1990, ''Effects of Boiling History on a Micro-configured Surface in a Dielectric Liquid,'' Int. Commun. Heat Mass Transfer, **17**, pp. 389–398.
[9] You, S. M., Simon, T. W., and Bar-Cohen, A., 1992, ''A Technique for Enhancing Boiling Heat Transfer with Application to Cooling of Electronic Equipment,'' IEEE Trans. Compon., Hybrids, Manuf. Technol., **15**, pp. 90–96.
[10] O'Connor, J. P., and You, S. M., 1995, ''A Painting Technique to Enhance Pool Boiling Heat Transfer in Saturated FC-72,'' ASME J. Heat Transfer, **117**, pp. 387–393.
[11] O'Connor, J. P., You, S. M., and Price, D. C., 1995, ''A Dielectric Surface Coating Technique to Enhance Boiling Heat Transfer from High Power Microelectronics,'' IEEE Trans. Compon., Packag. Manuf. Technol., Part A, **18**, pp. 656–663.
[12] Kubo, H., Takamatsu, H., and Honda, H., 1999, ''Effects of Size and Number Density of Micro-Reentrant Cavities on Boiling Heat Transfer from a Silicon Chip Immersed in Degassed and Gas-dissolved FC-72,'' J. Enhanced Heat Transf., **6**, pp. 151–160.
[13] O'Connor, J. P., You, S. M., and Chang, J. Y., 1996, ''Gas-Saturated Pool Boiling Heat Transfer From Smooth and Microporous Surfaces in FC-72,'' ASME J. Heat Transfer, **118**, pp. 662–667.
[14] Fishenden, M., and Saunders, O. A., 1950, *An Introduction to Heat Transfer*, Oxford University Press, New York.
[15] Chang, J. Y., and You, S. M., 1996, ''Heater Orientation Effects on Pool Boiling of Micro-Porous-Enhanced Surfaces in Saturated FC-72,'' ASME J. Heat Transfer, **118**, pp. 937–943.
[16] Cooper, M. G., 1984, ''Heat Transfer Rates in Saturated Pool Boiling—A Wide-Ranging Examination Using Reduced Properties,'' *Advances in Heat Transfer*, **16**, Academic Press, Orlando, pp. 157–239.
[17] Oktay, S., 1982, ''Departure from Natural Convection (DNC) in Low-Temperature Boiling Heat Transfer Encountered in Cooling Micro-Electronic LSI Devices,'' *Proc. 7th Int. Heat Transf. Conf., Vol.* **4**, pp. 113–118.
[18] Mudawar, I., and Anderson, T. M., 1989, ''High Flux Electronic Cooling by Means of Pool Boiling—Part II: Optimization of Enhanced Surface Geometry,'' ASME HTD, **111**, pp. 35–49.

Journal of Heat Transfer

Technical Notes

Thermal Analysis of Multiphoton Absorption at 193 nm in Volume-Absorbing Glass

Z. M. Zhang and D. H. Chen
Department of Mechanical Engineering,
University of Florida, Gainesville, FL 32611

Two-photon absorption in volume-absorbing materials may be significant for high-power, short-pulse excimer laser irradiation. In the present work, the effects of two-photon absorption on the laser intensity propagation and the penetration depth are investigated. The temperature history of a glass plate during pulsed ultraviolet (UV) laser heating is predicted. The results show that the two-photon absorption can compress the volume-absorbing effect to surface absorption and increase the peak temperature on the glass surface. This will reduce the damage threshold, increase the energy loss, and reduce the accuracy of pulsed-laser calorimeters. [DOI: 10.1115/1.1447940]

Keywords: Heating, Laser, Microscale, Radiation, Thermophysical

1 Introduction

The microelectronics industry has made organized, concentrated efforts to reduce the printing feature of photolithography in the past decade. ArF excimer laser at the wavelength of 193 nm is an immediate candidate for printing features below 0.18 μm. Accurate measurements of the pulse energy are required for the development of new optical lithographic processes. Therefore, 193 nm pulsed-laser calorimeters have been designed and built at the National Institute of Standards and Technology (NIST) [1]. In the 193 nm laser calorimeter, volume-absorbing borosilicate glass is used to absorb the ultraviolet (UV) laser energy instead of the surface-absorbing material, which could be damaged by the high-power excimer laser pulses.

Chen and Zhang [2] studied the thermal response of volume-absorbing glass under single- and multiple-pulse heating. The temperature history and distribution were predicted and compared with those under continuous-wave laser heating. However, the nonlinear behavior due to multiphoton absorption was not considered, which may be significant at high laser intensities. Multiphoton absorption has been studied by many researchers in recent years along with the development of high-power, short-pulse lasers. Taylor et al. [3] measured the two-photon absorption coefficient at 248 nm wavelength for a number of window materials including fused silica and CaF_2. Longtin and Tien [4] investigated the mechanisms of multiphoton absorption in liquids and developed a thermal model from microscopic considerations. The two-photon absorption coefficients of window materials and fused silica fibers have been measured at excimer laser wavelengths [5,6]. The two-photon absorption at 193 nm has been shown to cause color center formation (additional absorption) and densification (change in optical path length) in synthetic fused silica [7,8]. Our present study aims at predicting laser energy absorption and temperature distributions in the UV-absorbing glass used for the 193 nm laser calorimeter.

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division March 1, 2001; revision received August 23, 2001. Associate Editor: D. Poulikakos.

2 Laser Intensity and Penetration Depth

It is assumed that the single-photon absorption is due to the trap-site ionization and the two-photon absorption excites electrons from the valance band to the conduction band [3,9]. The probability of multiphoton absorption increases with the photon energy, which is determined by the laser wavelength. The band gap of fused silica is on the order of 10 eV [9], while the photon energy is 6.4 eV at the wavelength of 193 nm. The thermal model of the volume-absorbing glass and copper disks involved in this study is shown in Fig. 1. The radius of the glass and copper is 1.6 cm, the thickness of the glass is 0.1 cm, and the thickness of the copper is 0.02 cm. The laser beam is incident on the glass surface normally, and most of the laser energy is transmitted into the glass and absorbed inside the glass. In the volume-absorbing glass, the laser intensity attenuation in the direction of propagation (z-direction) due to one and two-photon absorption is governed by [9]

$$\frac{dI}{dz} = -\alpha I - \beta I^2, \tag{1}$$

where I is the laser intensity, α and β are the single-photon and two-photon absorption coefficients, respectively. By integrating Eq. (1), the laser intensity distribution in the z-direction is obtained as

$$I(z) = \frac{\alpha I_0 e^{-\alpha z}}{\alpha + \beta I_0 (1 - e^{-\alpha z})}, \tag{2}$$

where I_0 is the transmitted intensity at the incident surface. For the glass material with a refractive index of approximately 1.5, the reflectivity of glass is calculated to be 0.04 [2], indicating that the transmitted intensity is 96 percent of the incident intensity at the glass surface.

The penetration depth δ is defined as the distance z at which $I(z)/I_0$ is equal to $1/e$. For single-photon absorption, the penetration depth is equal to $1/\alpha$ and is independent of

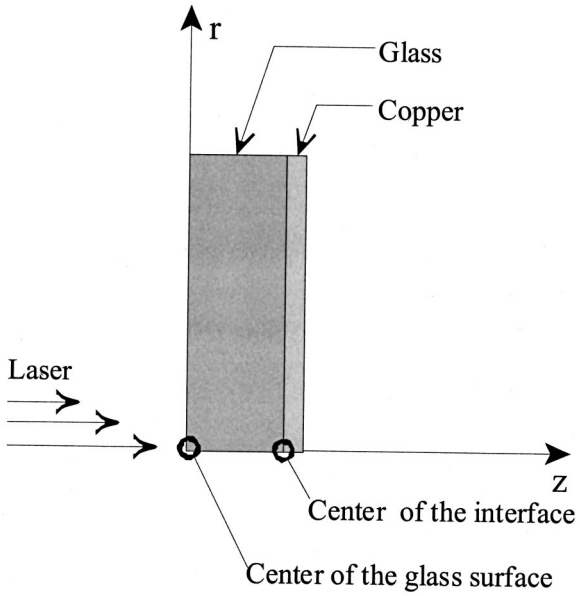


Fig. 1 Schematic of the thermal model (not to scale)

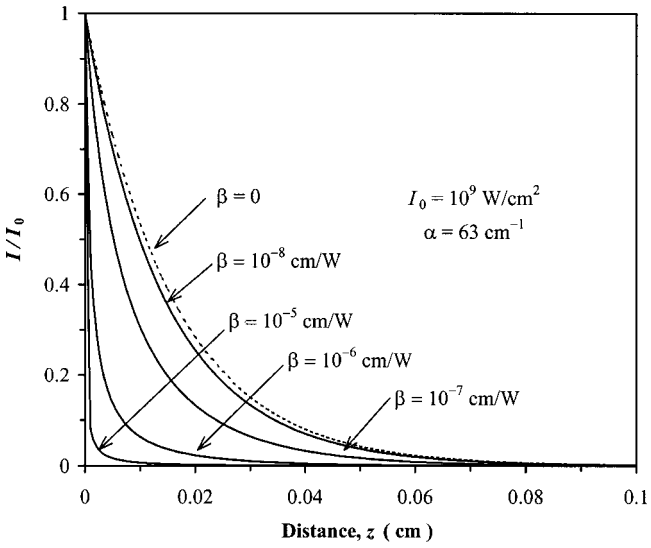


Fig. 2 Intensity distribution in the *z* direction

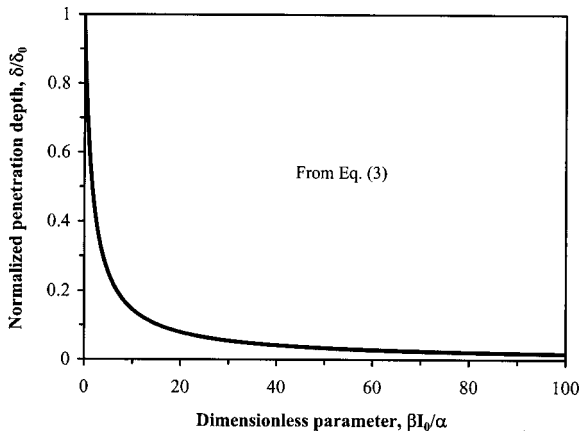


Fig. 3 Normalized penetration depth as a function of $\beta I_0/\alpha$

the laser intensity. When the laser beam is subject to one and two-photon absorption, the penetration depth is derived from Eq. (2) as

$$\delta = \frac{1}{\alpha}\ln\left(\frac{e+\beta I_0/\alpha}{1+\beta I_0/\alpha}\right). \tag{3}$$

In the present study, the value of $\alpha = 63\ \text{cm}^{-1}$ is used for the one-photon absorption coefficient of the glass material [2]. Without the two-photon absorption, the internal transmittance is less than 0.2 percent; hence the optically thick assumption is appropriate. Brimacombe et al. [6] measured two-photon absorption coefficients of glass fibers and obtained $\beta = 1.8\times10^{-9}$ cm/W for $\alpha = 5.4\times10^{-3}\ \text{cm}^{-1}$. The one-photon absorption coefficient of the glass used in the present study is 4 orders of magnitude greater than that of the glass fibers studied in [6]. Because data on the two-photon absorption coefficient for the UV-absorbing glass is not available, in the present study, the effect of the two-photon absorption is investigated by varying β from 0 to 10^{-5} cm/W. Figure 2 shows the intensity distribution calculated from Eq. (2) in the 0.1 cm thick glass with $I_0 = 10^9\ \text{W/cm}^2$. If there is only single-photon absorption ($\beta = 0$), the laser intensity decays exponentially. It can be seen from Eqs. (1) and (2) that the two-photon absorption will not be so important when $\beta I_0/\alpha \ll 1$. For $\beta = 10^{-8}$ cm/W ($\beta I_0/\alpha = 0.16$), the intensity distribution deviates slightly from the exponential decay. Further increase of the two-photon absorption coefficient causes the intensity to decay much faster and the distribution to deviate significantly from the single exponential function. When $\beta = 10^{-5}$ cm/W, $\beta I_0/\alpha = 160$ and hence the absorption process is dominated by two-photon absorption. For $\beta = 10^{-5}$ cm/W about 90 percent of the laser energy is absorbed within a 10 μm thickness; whereas for $\beta = 0$ only 10 percent of the laser energy is absorbed within the same thickness.

The penetration depth is reduced with two-photon absorption. Figure 3 shows the ratio of the penetration depth considering two-photon absorption calculated from Eq. (3) to $\delta_0 = 1/\alpha$ (without two-phonon absorption), as a function of the dimensionless parameter $\beta I_0/\alpha$. The penetration depth decreases rapidly as $\beta I_0/\alpha$ increases. The normalized penetration depth δ/δ_0 becomes 0.1 for $\beta I_0/\alpha = 15$ and 0.02 for $\beta I_0/\alpha = 84$. For high two-photon absorption coefficients, the penetration depth decreases rapidly as the laser intensity increases. The volume-absorbing effect can be compressed to surface absorption with two-photon absorption at sufficiently high laser intensities.

3 The Axisymmetric Pulse Heating Model

A finite element model has been developed to study the temperature distribution in the laser calorimeter (see Fig. 2). In this model, it is assumed that the laser intensity follows Gaussian distributions in both time and space. Although the spatial distribution for a typical 193 nm excimer laser is Gaussian in one direction and uniform in the other, the axisymmetric Gaussian distribution is used here to simplify the modeling. The transmitted intensity I_0 at the incident glass surface ($z = 0$) can be expressed as

$$I_0(r,t) = \frac{(1-R)Q}{\pi^{3/2}b^2\tau}\, e^{-r^2/b^2} e^{-(t-t_m)^2/\tau^2}, \tag{4}$$

where R is the reflectance at the glass surface ($R = 0.04$), Q is the energy per pulse, t_m is the time at which the power is maximum, b is related to the full width at half maximum (FWHM) beam diameter D by $D = 2b\sqrt{\ln 2}$, and τ is related to the FWHM pulse width by $\tau_p = 2\tau\sqrt{\ln 2}$.

When the laser beam propagates in the volume-absorbing glass, its energy is absorbed due to the one and two-photon absorption. The absorbed laser energy is treated as an internal heat generation source in the thermal modeling, and, hence,

$$\dot{q}(r,z,t)=\frac{\alpha^2 I_0(\alpha+\beta I_0)e^{-\alpha z}}{[\alpha+\beta I_0(1-e^{-\alpha z})]^2}, \tag{5}$$

where $I_0=I_0(r,t)$ given in Eq. (4). Although the Gaussian distribution extends from $-\infty$ to $+\infty$, it decreases quickly away from the peak. The heat generation rate is computed only for $r\leqslant 2D$ and from $t_m-2\tau_p$ to $t_m+2\tau_p$.

The thermal conductivity, density, and specific heat of the glass are taken from [2] as $k=0.73$ W/m·K, $\rho=2300$ kg/m^3, and $c_p=700$ J/kg·K, respectively. The copper plate attached to the volume-absorbing glass is also involved in the two-dimensional modeling without considering the thermal contact resistance. For copper, $k=401$ W/m·K, $\rho=8933$ kg/m^3, and $c_p=385$ J/kg·K [2]. In the modeling, thermal properties are treated as constants for the sake of simplicity. The Fourier heat conduction equations are solved using the heat generation rate given by Eq. (5). Convection and radiation conditions are modeled using a convection coefficient of 8 W/cm^2·K and emissivity of 0.8 for the glass surface and 0.01 for the gold-coated copper surface. The initial temperature of the glass and copper is equal to the ambient temperature T_0, which is set at 300 K. The time at peak power for the mth pulse is $t_m=2\tau_p+(m-1)/f$, where f is the pulse repetition rate. The temperature distribution and history are modeled at different two-photon absorption coefficients ($\beta=0$, 10^{-7}, and 10^{-6} cm/W) with the one-photon absorption coefficient α fixed at 63 cm^{-1}.

The element sizes and time steps are carefully selected and tested for numerical convergence. Variable time steps are used for the heating and cooling process. An adiabatic analysis is further used to test the amount of energy deposited in the heating model. Under adiabatic conditions, the relative errors between the temperature rise calculated theoretically and that predicted by the finite element model are less than 0.5 percent.

The peak temperature at the center of the glass surface increases with two-photon absorption due to the shift of the laser intensity distribution to the surface. The peak temperature for various two-photon absorption coefficients at pulse energies from 0.1 to 5 J are shown in Fig. 4, for $\tau_p=15$ ns and $D=0.5$ cm. The nonlinear dependence of the peak temperature on pulse energy is clearly seen with two-photon absorption, especially at low pulse energies. For large β, the peak temperature increases rapidly with the pulse energy and quickly reaches the threshold temperature (800 K) estimated from the phase transformation temperature, which is somewhere between 800 and 1000 K for fused silica. The strong two-photon absorption can significantly reduce the damage threshold of the volume-absorbing glass and hence decrease the dynamic range of laser calorimeters. The critical intensity calculated based on the temperature threshold is 3×10^8 W/cm^2 for $\beta=10^{-6}$ cm/W. Note that the predicted high temperatures are merely indicative because of the use of constant properties without considering phase transformation. It should be noted that the forgoing analysis assumes that ionization or laser-induced breakdown did not occur. Optical breakdown can happen with extremely high laser intensity, resulting in plasma formation and ablation of materials [10].

The temperature histories for multiple pulse heating are shown in Fig. 5, for $Q=1$ J, $f=100$ Hz, $\tau_p=15$ ns, and $D=0.5$ cm. The peak temperature for $\beta=10^{-6}$ cm/W is much higher than those for $\beta=0$ and 10^{-7} cm/W and increases as the number of pulses increases. At the beginning of each cooling process, the temperature drops much more for $\beta=10^{-6}$ cm/W than for the other cases due to heat conduction, as well as convection and radiation losses at the glass surface. The temperature at the center of the glass-copper interface increases slightly during the first 10 pulses. The temperature at the interface is lower for $\beta=10^{-6}$ cm/W than for the other cases. The excess temperature rise at the glass surface due to the two-photon absorption may cause extra heat loss (by

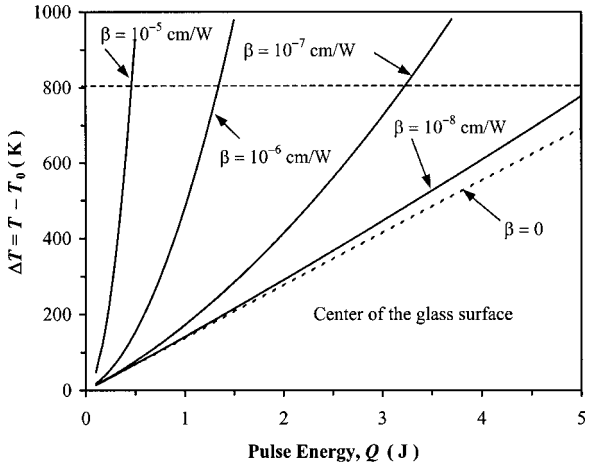


Fig. 4 Peak temperature rise versus the pulse energy for different β values

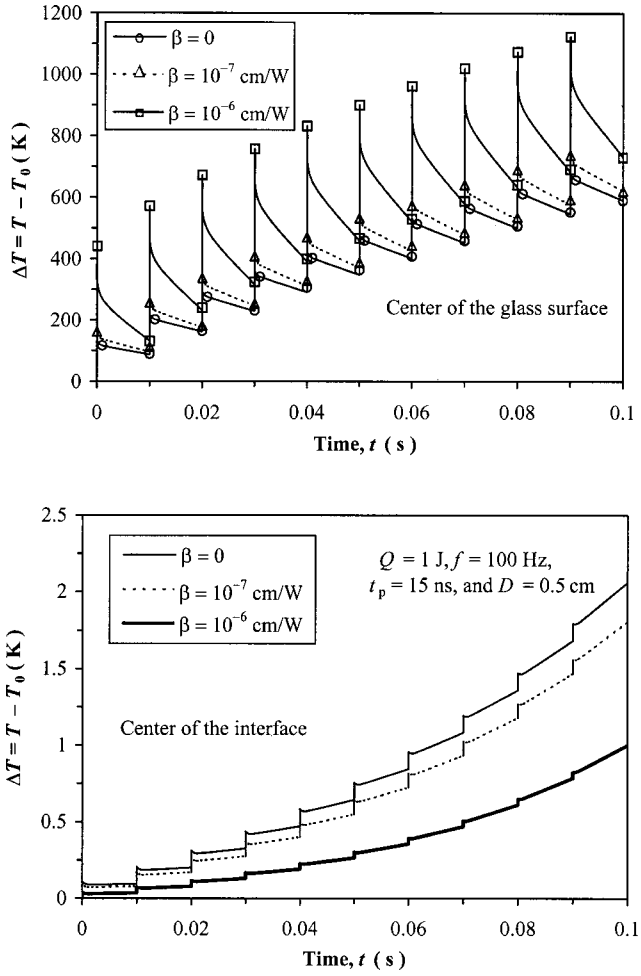


Fig. 5 Temperature histories for multiple-pulse heating at the center of the glass surface (upper) and the center of the interface (lower)

convection and radiation at the surface) as compared with the volume-absorbing process. The extra heat loss can introduce nonequivalence between the laser heating and the electrical heating; this will subsequently increase the uncertainty of the laser calorimeter.

4 Conclusions

Two-photon absorption in the UV-absorbing glass causes much faster attenuation of the laser intensity than the exponential decay. The penetration depth decreases rapidly with increasing two-photon absorption coefficient and laser intensity, and the volume-absorbing effect can eventually be compressed to surface absorption. For laser heating, two-photon absorption causes a high peak temperature on the glass surface and a large temperature gradient in the glass. Hence, it will reduce the damage threshold and accuracy of pulsed-laser calorimeters. This work provides quantitative information of the two-photon absorption effect in volume-absorbing materials and is important for the future design of laser calorimeters. Future research is needed to measure the two-photon absorption coefficient for UV absorbing glasses and to experimentally validate the thermal model presented here.

Acknowledgments

The authors wish to thank Drs. Marla Dowell, David Hahn, and Joseph Simmons for valuable discussions. This work was partially supported by NSF (Grant No. CTS-9875441) and the NIST Optoelectronics Division.

References

[1] Dowell, M. L., Cromer, C. L., Leonhardt, R. W., and Scott, T. R., 1998, ''Deep Ultraviolet Laser Metrology for Semiconductor Photolithography,'' in *Characterization and Metrology for ULSI Technology*, D. G. Seiler et al., eds. (AIP, College Park, MD), CP449, pp. 539–541.

[2] Chen, D. H., and Zhang, Z. M., 2000, ''Thermal Analysis of the Volume Absorber in Pulsed Excimer Laser Calorimeters,'' Int. J. Heat Mass Transf., **43**, pp. 3061–3072.

[3] Taylor, A. J., Gibson, R. B., and Roberts, J. P., 1988, ''Two-Photon Absorption at 248 nm in Ultraviolet Window Materials,'' Opt. Lett., **13**, pp. 814–816.

[4] Longtin, J. P., and Tien, C. L., 1997, ''Efficient Laser Heating of Transparent Liquids Using Multiphoton Absorption,'' Int. J. Heat Mass Transf., **40**, pp. 951–959.

[5] Kittelmann, O., and Ringling, J., 1994, ''Intensity-Dependent Transmission Properties of Window Materials at 193-nm Irradiation,'' Opt. Lett., **19**, pp. 2053–2055.

[6] Brimacombe, R. K., Taylor, R. S., and Leopold, K. E., 1989, ''Dependence of the Nonlinear Transmission Properties of Fused Silica Fibers on Excimer Laser Wavelength,'' J. Appl. Phys., **66**, pp. 4035–4040.

[7] Seward, III, T. P., Smith, C., Borrelli, N. F., and Allan, D. C., 1997, ''Densification of Synthetic Fused Silica Under Ultraviolet Irradiation,'' J. Non-Cryst. Solids, **222**, pp. 407–414.

[8] Mann, K., Apel, O., and Eva, E., 1998, ''Characterization of Absorption and Scatter Losses on Optical Components for ArF Excimer Lasers,'' *SPIE Proc.*, **3578**, pp. 614–624.

[9] Weitzman, P. S., and Osterberg, U., 1996, ''Two-Photon Absorption and Photoconductivity in Photosensitive Glasses,'' J. Appl. Phys., **79**, pp. 8648–8655.

[10] Stuart, B. C., Feit, M. D., Herman, S., Rubenchik, A. M., Shore, B. W., and Perry, M. D., 1996, ''Nanosecond-to-Femtosecond Laser-Induced Breakdown in Dielectrics,'' Phys. Rev. B, **53**, pp. 1749–1761.

Nanosecond Laser Repair of Adhered MEMS Structures

James W. Rogers

Leslie M. Phinney
e-mail: phinney@uiuc.edu

Department of Mechanical and Industrial Engineering,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801

Due to the small component sizes in modern microdevices, surface forces can create undesirable adhesion between microstructures, which is referred to as stiction. The current study uses a 100 ns, 1064 nm, Nd:YAG laser to repair polycrystalline silicon microcantilevers stuck to the underlying substrate. The results show that a Nd:YAG, 1064 nm laser is capable of repairing failed microstructures with yields exceeding those reported in earlier studies. Yields of 100 percent for cantilevers up to 1 mm in length were demonstrated for several laser operating conditions. The yields increase strongly with increased laser fluence and increase slightly with longer exposure times. [DOI: 10.1115/1.1447936]

Keywords: Experimental, Laser, Microscale, Microstructures, Surface Adhesion

Introduction

Surface adhesion, also termed stiction, is one of the major barriers to full-scale MEMS production and marketing since stiction failures limit device yields and operating lifetimes leading to a decrease in profitability [1]. Previous studies [2–6] investigate the use of ultrashort-pulse lasers to reduce microstructure adhesion in MEMS devices. Experiments using a Ti:Sapphire, 800 nm laser successfully repaired polycrystalline silicon microcantilevers which had failed due to stiction [3]; however, these experiments were only able to repair microcantilevers of lengths up to 140 μm and needed an expensive and difficult to maintain laser system. In this paper, experiments using a Nd:YAG, 1064 nm laser to repair stiction-failed microcantilevers are described. The primary motivation for using a Nd:YAG laser is the increased potential for industrial application as compared to a Ti:Sapphire femtosecond laser. The experiments described in this paper determine the dependence on fluence and exposure time for the microcantilever repair process.

Experimental Setup and Procedure

Arrays of doped polycrystalline silicon (polysilicon) cantilevers with lengths varying from 60 μm to 1000 μm were fabricated using the Multi-User MEMS Processes, MUMPs™ [7]. The cantilevers were 10 μm wide, 2 μm thick, and suspended 2 μm above the substrate. After being released through an HF etch and subsequent rinse processes, each array of microcantilevers was inspected to determine which beams were freestanding and which were stiction-failed. This inspection was accomplished using a Michelson interferometric attachment on an optical microscope similar to that described by de Boer et al. [8].

The laser used for this investigation was a Q-switched, Nd:YAG, 1064 nm laser manufactured by Molectron. The repetition rate of the laser was 20 Hz, and the pulse duration was 100 ns at full-width half-maximum (FWHM). The center of the beam was directed onto the device through an aperture that is approximately 3 mm in diameter. A low-power He-Ne laser, which was co-linear to the Nd:YAG beam, was used to align the Nd:YAG beam onto a given array on a die. The device was mounted on an X-Y stage in order to align the specified microcantilever array without readjusting either of the lasers. The Nd:YAG beam diameter was measured to be $3.0\ \text{mm} \pm 0.1\ \text{mm}$ at the $1/e^2$ points.

Experiments studying the dependence of laser fluence and exposure time were performed. To determine the effects of the laser fluence, the exposure time was kept constant at 5 minutes while the laser fluence was varied from $14 \pm 1\ \text{mJ/cm}^2$ to $85 \pm 4\ \text{mJ/cm}^2$ in steps of approximately $7\ \text{mJ/cm}^2$. Three different dice were irradiated to determine the variability in the results. The exposure time dependence was investigated using a constant laser fluence of $70 \pm 3\ \text{mJ/cm}^2$ and varying the time during which the

Contributed by the Heat Transfer Division for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received by the Heat Transfer Division December 18, 2000; revision received August 6, 2001. Associate Editor: A. Majumdar.

sample was exposed to the laser irradiation from a single pulse to 5 minutes (~6000 pulses). In order to obtain exposure times of less than 5 seconds (~100 pulses), the single-shot mode of the Nd:YAG laser was used. The laser beam was controlled with the manual shutter, which was timed using a digital timer, for times 5 seconds and longer. Two different dice were irradiated for the exposure time experiments in order to investigate sample-to-sample variation.

After the dice were irradiated, they were immediately inspected under the interferometric microscope. All of the dice were irradiated within 12 hours of being released in order to decrease the uncertainty involved in the magnitude of the surface adhesion forces increasing over time as shown by Rogers and Phinney [9].

Results and Discussion

The percentage yields reported in this investigation are defined as the ratio of microcantilevers repaired after laser irradiation to the number of microcantilevers stiction-failed after release. This includes every length of beam on the array. Yields of less than 100 percent usually had the shorter beams freestanding while the longer beams remained stiction-failed.

The results for the laser fluence experiments are summarized in Fig. 1. As seen in the figure, a laser fluence dependence on yield exists with laser fluences above approximately 70 ± 3 mJ/cm^2 resulting in 100 percent yield, which means every microcantilever on the array was freestanding after laser irradiation. Sample-to-sample variation is also apparent in Fig. 1 with a single laser fluence varying in yields by 40 percent in some cases. A likely reason for this is the variation in conditions after release. After irradiation, all of the beams, which after release were deflected near the base of the cantilever were deflected at the tip after irradiation suggesting that the beams ''peel'' away from the substrate.

The results for the exposure time experiments, performed at a laser fluence of 70 ± 3 mJ/cm^2, are summarized in Fig. 2. As seen in Fig. 2, a single pulse resulted in yields greater than 55%, and the yields increase with longer exposure times. Furthermore, times longer than approximately 30 seconds (~600 pulses) resulted in yields greater than 90 percent. Exposure times longer than 30 seconds usually repaired beams up to 700 μm with any failed beams longer than 700 μm only stuck at the tip. Tien et al.'s [3] experiments also repaired stiction-failed microcantilevers with a single pulse but not with the yields seen in this investigation.

The mechanism responsible for microstructure repair has been investigated by Rogers et al. [10] where it was found that a temperature difference between the microcantilever and substrate causes an increase of strain energy in the beam due to thermal expansion. The increased strain energy causes the beam to peel from the substrate in a manner similar to the propagation of a crack, and the beam eventually releases from the substrate resulting in a successful repair. Since this process requires a temperature difference between the microcantilever and substrate, it is important to consider the possible magnitudes of this temperature difference for the experimental laser conditions. The absorption depth of polysilicon will vary from sample to sample depending on doping and grain-size; however, an absorption depth of approximately 5 μm at 1064 nm is consistent with results in the literature [11,12]. Therefore, for a laser fluence of 70 mJ/cm^2 and neglecting interface effects, a temperature difference between the microcantilever and substrate of 50°C is reasonable and sufficient to initiate repair. Furthermore, a thermal expansion mechanism is consistent with the strong fluence dependence of the repair process as higher fluences create larger temperature differences between the microstructure and substrate resulting in longer beams being repaired. The yields achieved using the Nd:YAG laser are higher than those obtained with the Ti:Sapphire laser [3] because the range of laser fluences, and subsequent lattice temperatures, were higher for the Nd:YAG process. Also, the dependence on exposure time suggests thermal energy is accumulated in the microcantilever from one pulse to the next.

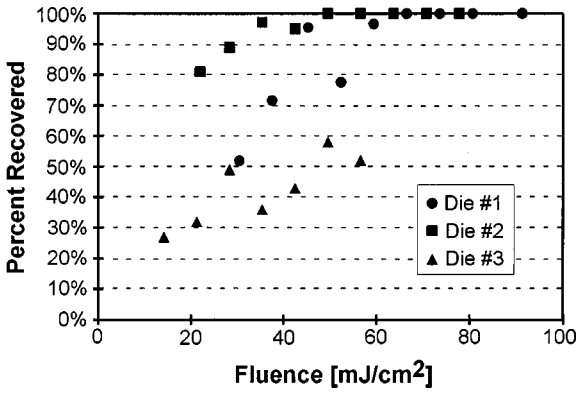


Fig. 1 Fluence variation for repair of stiction-failed microcantilevers

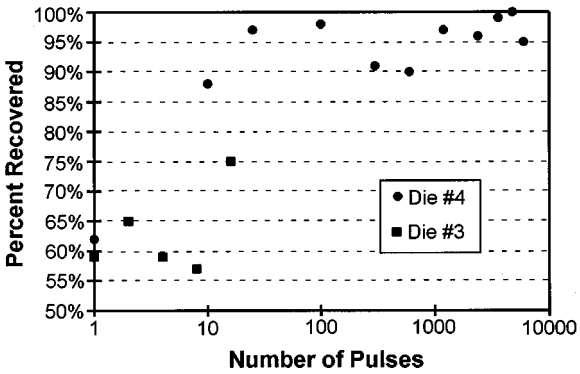


Fig. 2 Exposure time variation for repair of stiction-failed microcantilevers

It should be noted that no visible damage was observed after laser irradiation of the microcantilevers. Due to the desire to have an operating device after laser irradiation, it is very important that the laser not change the operating characteristics of the device. At this point, it cannot be concluded that the devices were not altered, but no visible damage, such as a darkening of the material or melting, occurred.

Conclusions

The objective of this investigation was to determine the feasibility of using a Nd:YAG, 1064 nm laser to repair stiction-failed silicon microstructures. The Nd:YAG process detailed in this work is very promising for industrial applications. The yields produced using the Nd:YAG laser are much greater than were produced using a Ti:Sapphire laser by Tien et al. [3].

The Nd:YAG laser experiments displayed that the yields are strongly dependent on laser fluence and slightly dependent on exposure time. From the results reported in this work, repair yields near 100 percent for beams as long as 1000 μm are achievable using a Nd:YAG laser with a fluence of 70 mJ/cm^2 and an exposure time of 30 seconds to one minute. Thermal expansion of a microstructure heated to a temperature greater than the substrate is the mechanism responsible for the repair process.

Acknowledgments

The authors would like to acknowledge financial support from the National Science Foundation under grant CTS-9984979 and from the University of Illinois at Urbana-Champaign.

References

[1] Maboudian, R., and Howe, R. T., 1997, ''Critical Review: Adhesion in Surface Micromechanical Structures,'' J. Vac. Sci. Technol. B, **15**, pp. 1–20.

[2] Fushinobu, K., Phinney, L. M., and Tien, N. C., 1996, ''Ultrashort-Pulse Laser Heating of Silicon to Reduce Microstructure Adhesion,'' Int. J. Heat Mass Transf., **39**, pp. 3181–3186.

[3] Tien, N. C., Jeong, S., Phinney, L. M., Fushinobu, K., and Bokor, J., 1996, ''Surface Adhesion Reduction in Silicon Microstructures Using Femtosecond Laser Pulses,'' Appl. Phys. Lett., **68**, pp. 197–199.

[4] Phinney, L. M., and Tien, C. L., 1998, ''Recovery Mechanisms for Stiction-Failed Microcantilevers Using Short-Pulse Lasers,'' *Heat Transfer 1998: Proceedings of the 11th International Heat Transfer Conference*, Vol. 5, pp. 145–150.

[5] Fushinobu, K., Phinney, L. M., Kurosaki, Y., and Tien, C. L., 1999, ''Optimization of Laser Parameters for Ultrashort-Pulse Laser Recovery of Stiction-Failed Microstructures,'' Numer. Heat Transfer, Part A, **A36**, pp. 345–357.

[6] Phinney, L. M., Fushinobu, K., and Tien, C. L., 2000, ''Subpicosecond Laser Processing of Polycrystalline Silicon Microstructures,'' Microscale Thermophys. Eng., **4**, pp. 61–75.

[7] Koester, D. A., Mahadevan, R., Hardy, B., and Markus, K. W., 2001, ''MUMPs Design Handbook,'' Cronos Integrated Microsystems, Research Triangle Park, NC, pp. 1–41.

[8] de Boer, M. P., Clews, P. J., Smith, B. K., and Michalske, T. A., 1998, ''Adhesion of Polysilicon Microbeams in Controlled Humidity Ambients,'' *Proceeding of the 1998 Spring Materials Research Society Symposium*, **518**, pp. 131–6.

[9] Rogers, J. W., and Phinney, L. M., 2001, ''Process Yields for Laser Repair of Aged, Stiction-Failed, MEMS Devices,'' J. Microelectromech. Syst., **10**, pp. 280–285.

[10] Rogers, J. W., Mackin, T. J., and Phinney, L. M., 2001, ''A Thermomechanical Model for Adhesion Reduction of MEMS Microcantilevers,'' *Proceedings of the 2001 ASME International Mechanical Engineering Congress and Exposition*, **2**, MEMS-23823, ASME, New York.

[11] Jones, Jr., R. E., and Wesolowski, S. P., 1984, ''Electrical, Thermoelectric, and Optical Properties of Strongly Degenerate Polycrystalline Silicon Films,'' J. Appl. Phys., **56**, pp. 1701–1706.

[12] Globus, T., Fonash, S. J., and Gildenblat, G., 1996, ''Optical Characterization of Hydrogenated Silicon Films in the Extended Energy Range,'' in *Diagnostic Techniques for Semiconductor Materials Processing II*, Pang et al., eds., pp. 313–318.