Journal of Heat Transfer
Transactions of the ASME

Volume 129, Issue 1, pp. 1-98 , January 2007

GUEST EDITORIAL

TECHNICAL PAPERS

TECHNICAL BRIEFS

ERRATUM

Journal of Heat Transfer

Guest Editorial

Special Issue on Nano/Microscale Radiative Transfer

Since the mid 1980s, the heat transfer community has made significant advances in experimental and theoretical understanding of sub-continuum thermal conduction and fluidics at small length scales; these activities coincided with the increased interest in micro/nanotechnology. The special issue of the *Journal of Heat Transfer* (JHT) on Micro/Nanoscale Heat Transfer, published in April 2002, covered the breadth of most of these contributions. On the other hand, much less attention has been paid to the understanding of thermal radiation at micro/nanoscales until recently. The current special issue is devoted to Nano/Microscale Radiative Transfer and should give a glimpse of the state of the art in the field.

Thermal radiation is key to many fundamental phenomena surrounding us, from solar radiation to fire to incandescent lamp, and has played a major role in combustion and furnace design, materials processing, energy utilization, temperature measurements, remote sensing for astronomy and space exploration, food processing, and cryogenic engineering, as well as numerous agricultural, health, and military applications. At the dawn of the twentieth century, the study of thermal radiation produced two Nobel laureates (Wien in 1911 and Planck in 1918) and led to the discovery of the Bose-Einstein quantum statistics. While Planck's law has been successfully applied to a large number of practical problems for some 100 years, questions have been raised about its limitation in micro/nanosystems, especially at subwavelength distances, i.e., in the near field. The advancement of nanotechnology has enabled precise manufacturing of structures with feature sizes smaller than the characteristic wavelength of thermal radiation. Radiative properties can be largely modified by interference, diffraction, localization, and surface polaritons, resulting in engineered surface microstructures with numerous practical applications in sensors, imaging, manufacturing, and energy devices. Thus, the study of engineered microstructures for controlling surface radiative properties has become an active area of research, along with nanomaterials, nanophotonics, and nanoscale thermophysical engineering.

Discussions with a number of participants at the Fourth International Symposium on Radiative Transfer held in Istanbul, Turkey in June 2004, suggested that there was a strong desire in the community to have focused sessions on nanoscale radiative transfer in order to boost this emerging research frontier. Subsequently, we organized the first Mini-Symposium on Nano/Microscale Radiative Transfer and Properties (NanoRad), at the 2005 International Mechanical Engineering Congress and Exposition (IMECE) in Orlando, Florida in November 2005. This mini-symposium included five sessions with a total of 21 papers/posters and a panel discussion about significant issues, challenges, funding trends, and opportunities in this area. The panel consisted of Professor Alfonso Ortega, then Director of the NSF Thermal Transport and Thermal Processing Program, Professor Gang Chen from the Department of Mechanical Engineering of MIT, Professor Yongfeng Lu from the Department of Electrical Engineering of the University of Nebraska Lincoln, along with the Editors of this Special Issue. The panelists raised significant questions related to the maximum achievable thermal emissive power, applicability of Kirchhoff's law in nanostructures, and the entropy of near-field radiation, among others. The mini-symposium showed that great opportunities and needs exist in the study of near-field radiation, which we also dub as nanoscale radiative transfer. The field is rich and exciting, as it requires deeper understanding of the interplay among optical, thermal, electrical, and mechanical properties of materials and structures at nanoscales, for applications in spectral and directional control of thermal emission, photovoltaic and thermophotovoltaic devices, biological sensors, remote sensing, materials processing, and nanothermal manufacturing.

This special issue is based on selected papers presented at the mini-symposium, and additional manuscripts, which were broadly solicited. All papers went through the rigid review process of JHT, followed by extensive revisions. A total of 12 papers are included, beginning with a survey article by Ruan and Kaviany (p. 3) on the advances in laser cooling of solids. This phenomenon relies on the anti-Stokes fluorescence, where the emitted photons have a mean energy higher than that of the absorbed photons. The thermodynamics of laser cooling and the potential advantages of using nanostructured materials due to quantum size effect were discussed. In the second paper, Marquier et al. (p. 11) expanded their earlier works to study the effect of polarization and anisotropy in the emission behavior of surface relief gratings. Coherent thermal emission has been demonstrated for gratings and shows promise in thermal control and remote sensing. Lee and Zhang (p. 17) examined coherent thermal emission characteristics from a proposed multilayer structure consisting of a thin SiC layer on a one-dimensional photonic crystal. They distinguished three different mechanisms for coherent emission due to surface electromagnetic waves, optical cavity resonance, and the Brewster mode. Chandrasekharan et al. (p. 27) experimentally observed the effect of heat treatment on the optical properties of Ta_2O_5 thin films for their application as radiation shields in microcombustion systems. Attention was paid to the formation of an interfacial oxide layer on the radiative properties and the effect of wave interference. In the next paper, Jin and Xu (p. 37) focused on the near-field effect of subwavelength apertures and demonstrated nanoscale concentration of light through H-shape apertures using near-field scanning optical microscopy (NSOM). These light sources may enable lithographic fabrication and materials process at the nanometer scale. Optical microcavities have enormous applications in quantum electrodynamics (QED), enhancement and suppression of spontaneous emission, and biochemical sensors, due to their extremely high quality (Q) factor. Guo and Quan (p. 44) used the finite element method (FEM) to investigate the energy coupling (through evanescent waves) between a waveguide and a whispering-gallery-mode optical microdisk. On the other hand, Heltzel et al. (p. 53) demonstrated submicrometer manufacturing using a femtosecond laser pulse with the assistance of the patterned silica microspheres on a borosilicate glass substrate. Venkata et al. (p. 60) described a technique to characterize metallic particles and agglomerates based on surface plasmon waves. The effects of size, shape, and orientation of gold nanoparticles on their scattering patterns were explored in the visible spectrum, especially at the resonance wavelengths. Surface roughness and patterns can significantly affect the radiative properties and are important for semiconductor manufacturing. Fu and Hsu (p. 71) used a finite-difference time-domain (FDTD) to numerically solve the Maxwell equations for scattering from random rough surfaces, and showed that the results compared favorably with other methods. Chen et al. (p. 79) modeled the radiative properties of patterned silicon wafers with the smallest feature dimension down to 30 nm, considering the effects of temperature, wavelength, polarization, and angle of incidence. Rigorous coupled wave analysis (RCWA) was employed to obtain accurate solutions and to assess the applicability of the method of homogenization based on effective medium theories (EMTs). Hu et al. (p. 91) used an infrared microscope to characterize the thermal interface formed between two opposing, partially overlapped carbon nanotube (CNT) arrays. Hammonds (p. 94) employed a Green function approach based on the fluctuation-dissipation theorem to model the radiation heat transfer across an evacuated spherical cavity inside a SiC medium.

In summary, this special issue reflects a variety of contemporary research in nano/microscale radiative transfer and is expected to promote further research activities and development opportunities. We would like to thank all contributors, reviewers, and panelists, who have made the mini-symposium a success and this special issue possible. Our special thanks go to Professor Neil Wright of Michigan State University, Chair of the K-7 Thermophysical Properties Technical Committee, for his enthusiastic support and K-7's sponsorship of the mini-symposia. We appreciate the encouragement and help from the Editor of JHT, Professor Yogesh Jaluria, throughout the editorial process.

Zhuomin M. Zhang
Georgia Institute of Technology

M. Pinar Mengüç
University of Kentucky

Advances in Laser Cooling of Solids

X. L. Ruan

M. Kaviany
e-mail: kaviany@umich.edu

Department of Mechanical Engineering,
University of Michigan,
Ann Arbor, MI 48109

We review the progress on laser cooling of solids. Laser cooling of ion-doped solids and semiconductors is based on the anti-Stokes fluorescence, where the emitted photons have a mean energy higher than that of the absorbed photons. The thermodynamic analysis shows that this cooling process does not violate the second law, and that the achieved efficiency is much lower than the theoretical limit. Laser cooling has experienced rapid progress in rare-earth-ion doped solids in the last decade, with the temperature difference increasing from 0.3 to 92 K. *Further improvements can be explored from the perspectives of materials and structures. Also, theories need to be developed, to provide guidance for searching enhanced cooling performance. Theoretical predictions show that semiconductors may be cooled more than ion-doped solids, but no success in bulk cooling has been achieved yet after a few attempts (due to the fluorescence trapping and nonradiative recombination). Possible solutions are discussed, and net cooling is expected to be realized in the near future.* [DOI: 10.1115/1.2360596]

1 Introduction

Cooling is a process in which thermal energy is absorbed from a lower temperature reservoir and deposited to a higher temperature reservoir, by consuming a small amount of higher grade energy. Mechanical, electrical, and optical energies are among the high grade energies, and are expected to be used for cooling purposes. Gas compression refrigerators, which consume mechanical energy, and thermoelectric coolers, which consume electrical energy, are matured techniques which have found very broad applications. High grade optical energies like lasers are, however, well-known for their heating effects rather than cooling capabilities. In fact, gases have been cooled to the order of nano-Kelvin by lasers, and this technique resulted in the 1997 and 2001 Nobel Prizes in Physics [1–4]. Its counterpart, laser cooling of solids, has also attracted great interests recently, for the potential to develop an optical cryocooler for a variety of important applications such as the cooling of sensors.

The concept of laser cooling (optical refrigeration) of solids dates back to 1929, when Pringsheim recognized that thermal vibrational energy (phonon) can be removed by the anti-Stokes fluorescence, i.e., the photons emitted by an optical material have a mean energy higher than that of the absorbed photons [5]. Initially, it was believed that optical cooling by the anti-Stokes fluorescence contradicted the second law of thermodynamics. Predictions suggested that the cycle of excitation and fluorescence was reversible, and hence the optical cooling would be equivalent to the complete transformation of heat to work [6,7]. This issue was cleared by Landau by assigning entropy to radiation [8]. It was shown that the entropy of a radiation field increases with its frequency bandwidth and also the solid angle through which it propagates. Since the incident laser light has a very small bandwidth and propagates in a well-defined direction, it has almost zero entropy. On the other hand, the fluorescence is relatively broadband and is emitted in all directions and, therefore, it has a comparatively larger entropy. In this way, the second law of thermodynamics is satisfied.

Many attempts have been made to realize radiative refrigeration experimentally, and the associated theoretical interpretations have been discussed. The earliest experiment was performed by Kushida and Geusic on Nd: YAG [9]. Reduced heating other than net cooling was observed, which was conjectured to be a result of the impurities in the crystal. Later Djeu and Whitney laser cooled low-pressure CO_2 by 1 K from 600 K by using a CO_2 laser for pumping [10]. In 1995, Epstein et al. [11] reported the first successful experiment of laser cooling in solids. Since then, various Yb or Tm doped glasses and crystals have been cooled [12–21]. Particularly, bulk solids have been cooled from room temperature to 208 K (creating a temperature difference $\Delta T = 92$ K) [21]. Continuous progress has been made [19,22,23] towards achieving cryogenic temperatures. For semiconductors, theoretical predictions have shown their potential to be cooled to as low as 50 K starting from room temperature [24], but experimental success is yet to be achieved due to some serious challenges to be overcome.

These progresses, as well as the recent success in laser cooling of gases and the subsequent achievement of Bose-Einstein condensation [1–4], again stimulated interest in optical cooling of solids.

2 Principles of Laser Cooling of Solids

Laser cooling can be viewed as the inverse cycle of lasers, and laser materials are in principle also good candidates for laser cooling. Common laser materials in the solid state include ion-doped solids and semiconductors, which are currently being studied for laser cooling.

In Fig. 1(*a*), the fundamental energy carriers involved in the laser cooling of rare-earth-ion doped solids are shown. There is a host crystal lattice, idealized as transparent to the pumping laser. Some of its atoms are replaced by optically active, doped ions (e.g., Yb^{3+}). The ion is represented by an effective transition dipole moment, which is the matrix element of the dipole operator $e_e\boldsymbol{r}$ [25], i.e.,

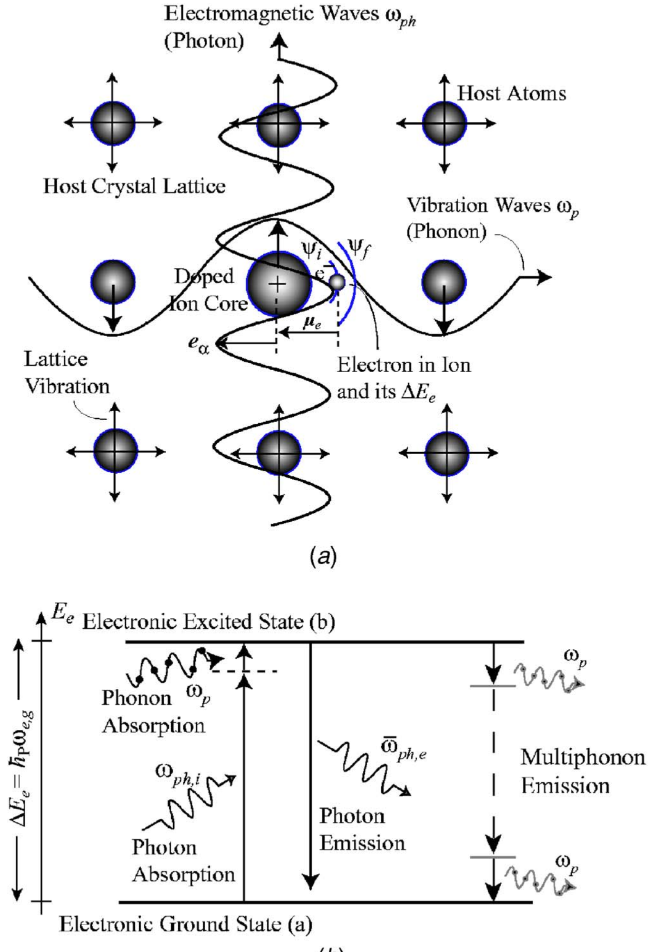
$$\boldsymbol{\mu}_e = \int \psi_f^* e_e \boldsymbol{r} \psi_i d^3 r \tag{1}$$

where, e_e is the electron charge, $\boldsymbol{r}$ is the position vector, and ψ_i and ψ_f are the initial and final state wave functions of the two level system. One can think about this matrix element as coupling states ψ_i and ψ_f, which have different parity, creating or absorbing a photon. The electromagnetic field, which has a polarization vector $\boldsymbol{e}_\alpha$, may interact with the ion if the coupling factor $\boldsymbol{e}_\alpha \cdot \boldsymbol{\mu}_e$ is nonzero (i.e., they are not orthogonal).

Shown in Fig. 1(*b*) are the principles of the photon-electron-phonon interactions which result in the cooling effect in the solid. As the pumping wavelength is tuned to the red side of the reso-

---

Contributed by the Heat Transfer Division of ASME for publication in the Journal of Heat Transfer. Manuscript received January 26, 2006; final manuscript received June 18, 2006. Review conducted by Zhuomin M. Zhang.



(*a*)

(*b*)

Fig. 1 (*a*) Three fundamental energy carriers in rare-earth ion doped solids irradiated by laser light: Photons from the pumping fields, phonons from the host crystal, and electrons of the doped ions. (*b*) Principles of laser cooling in rare-earth ion doped crystal. The electron is excited by absorbing a photon and one or more lattice phonons, and then decays by emitting a higher energy photon.

nance, the probability of a purely electronic transition between electronic sublevels, a first-order process, becomes smaller. On the other hand, the phonon-assisted transition, a second-order process, starts to contribute significantly to absorption. As a result, the absorption turns out to be a combination of the first- and second-order transitions. Since a much longer pumping wavelength than the resonance is used in laser cooling, the total transition is believed to be dominated by the second-order process. In such a process, the medium is irradiated by laser light with a frequency $\omega_{ph,i}$ that is below the resonance frequency $\omega_{e,g}$ for the energy gap (10,250 cm^{-1} for Yb^{3+} ion in Y_2O_3), the electron may still be excited by absorbing a photon from the pumping field and a phonon with a frequency ω_p from the host, such that $\omega_{ph,i}+\omega_p = \omega_{e,g}$. The electron then undergoes a radiative decay by emitting a photon with a frequency $\omega_{ph,e}$, or undergoes a nonradiative decay by emitting a few phonons (multiphonon relaxation), leading to internal heating of the system. The energy magnitudes of these carriers involved in the laser cooling process in $\text{Yb}^{3+}:\text{Y}_2\text{O}_3$ are shown in Fig. 2 [26].

The quantum efficiency $\eta_{e\text{-}ph}$ is defined as the ratio of the radiative decay rate to the total decay rate, or, in this case, the ratio of the number of emitted photons to that of absorbed photons [26]. The net cooling power per absorbed photon P_c is given by

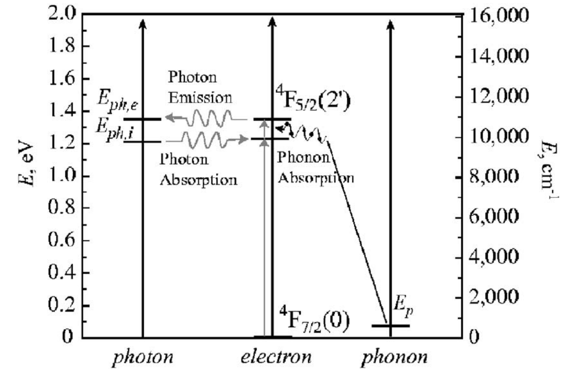


Fig. 2 Energy spectra of all three carriers in irradiated $\text{Yb}^{3+}:\text{Y}_2\text{O}_3$. Yb^{3+} has only two main electronic levels: ${}^4\text{F}_{7/2}$ and ${}^4\text{F}_{5/2}$. Carriers may interact with each other as energy and momentum conservations are met [26].

$$P_c = \eta_{e\text{-}ph}\hbar\bar{\omega}_{ph,e} - \hbar\omega_{ph,i} = \hbar\omega_{ph,i}\left(\eta_{e\text{-}ph}\frac{\bar{\omega}_{ph,e}}{\omega_{ph,i}} - 1\right) \tag{2}$$

where $\bar{\omega}_{ph,e}$ is the mean frequency of emitted photons. The cooling efficiency is defined as the ratio of the net cooling power and the absorbed power, i.e.,

$$\eta_c = \frac{P_c}{\hbar\omega_{ph,i}} = \eta_{e\text{-}ph}\frac{\bar{\omega}_{ph,e}}{\omega_{ph,i}} - 1 \tag{3}$$

Cooling is achieved as η_c is positive. Evidently, high quantum efficiency $\eta_{e\text{-}ph}$ is desirable.

The cooling process in semiconductors, as shown in Fig. 3, is similar to that of ion-doped solids. An electron originally in the valence band absorbs a laser photon and is excited to the conduction band, leaving a hole in the valence band. It then gains some energy by absorbing a phonon (intraband absorption) and climbs to a higher position in the conduction band. The electron then decays back to the valence band via either radiative recombination, or nonradiative recombination, which includes the multiphonon process and Auger process. The net cooling power is in the same form as Eq. (2).

3 Macroscopic Role of Laser Cooling

In Fig. 4 a macroscopic energy diagram is shown for a solid that is cooled by laser, where various energy flows are shown. To minimize the external thermal load to achieve the most cooling effect, in most of the existing experiments the solid was supported by very thin wires and was placed in a vacuum, to eliminate the

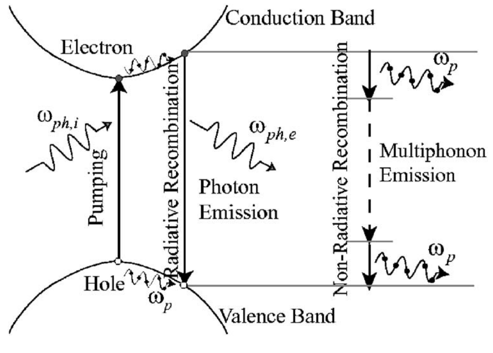


Fig. 3 Process for laser cooling of a semiconductor in which a laser photon with frequency $\omega_{ph,i}$ is absorbed followed by emission of an up-converted fluorescence photon with frequency $\omega_{ph,e}$

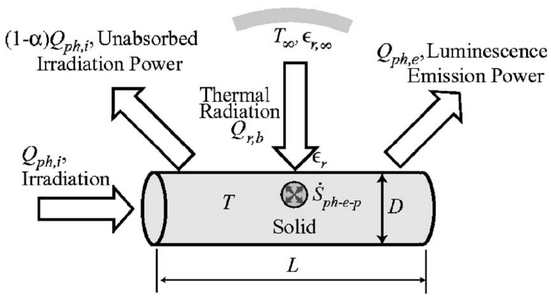


Fig. 4 The energy diagram for laser cooling of a solid, where radiation is the only external thermal load

conduction and convection. The only external load is then the thermal radiation from its surroundings, as shown in Fig. 4.

Using Eq. (3), the total cooling power $\dot{S}_{ph\text{-}e\text{-}p}$ is given by

$$\dot{S}_{ph\text{-}e\text{-}p} = P_c \frac{Q_{ph,a}}{\hbar \omega_{ph,i}} = \eta_c Q_{ph,a} = \eta_c \alpha Q_{ph,i} \tag{4}$$

where $Q_{ph,a}$ is the absorbed power, $Q_{ph,i}$ is the irradiation power, and α is the absorptivity. This equation links the macroscopic cooling behavior and the atomic level parameters. The result also indicates that the net cooling power $\dot{S}_{ph\text{-}e\text{-}p}$ is proportional to the absorbed power $Q_{ph,a}$.

The steady-state, integral-volume energy equation is

$$\dot{S}_{ph\text{-}e\text{-}p} - Q_{r,b} = 0 \tag{5}$$

Assuming that the surface area of the surrounding is much larger than that of the sample, the thermal radiation load $Q_{r,b}$ is given by [27]

$$Q_{r,b} = A_r \epsilon_r \sigma_{\text{SB}} (T_\infty^4 - T^4) \tag{6}$$

where A_r is the surface area of the solid, σ_{SB} is the Stefan-Boltzmann constant, and T_∞ is the temperature of the ambient radiation field (or the effective temperature in case the ambient field does not have a thermal spectra). Apparently the amount of cooling power $\dot{S}_{ph\text{-}e\text{-}p}$ governs how much the sample temperature can be lowered from the surrounding temperature.

4 Applications of Laser Cooling

Since the process of anti-Stokes fluorescence does not require any mechanical movement, such a solid-state cooler is likely to have a longer lifetime than other coolers. This is particularly useful in space, where reliability and lifetime are crucial considerations.

Despite the low cooling efficiency (2% to 3% at room temperature so far, and lower efficiency at lower temperatures), laser cooling is a promising candidate for cryocoolers, with the potential to cool ion-doped dielectrics and semiconductors to 50 K or even 10 K starting from room temperature [28]. Predictions have shown that at these cryogenic temperatures, thermoelectric coolers become ineffective or incapable, compared to laser coolers [29].

Another application is in the heat-balanced laser system. This process would employ fluorescence cooling to offset the heat produced in the generation of laser radiation [30–35].

5 Thermodynamics of Laser Cooling of Solids

As mentioned in the Introduction, there was a debate which stimulated the birth of radiation thermodynamics, that is, whether laser cooling is thermodynamically possible. Detailed discussions of the thermodynamics of optical cooling have been given by Mungan and Gosnell [36,37]. The following is based on their papers, with the inclusion of some more recent ideas from other researchers.

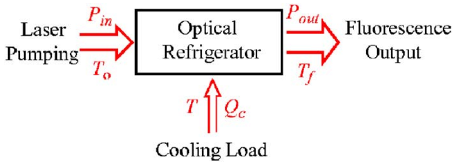


Fig. 5 A control volume showing various energies in and out in the laser cooling of a solid [37]

A simple control volume is shown in Fig. 5, in which the power flowing in and out (the pump laser, the external thermal load, and the luminescence emission) are marked. For a narrowband radiation which is independent of the angular directions θ and ϕ over a circular cone of half-angle δ, the energy flux density is given by [37]

$$I_{\dot{E}}(\omega) = \frac{1}{4\pi^2} \hbar c^{-2} \bar{f}_{\text{ph}} \omega_0^3 \Delta\omega \sin^2 \delta \tag{7}$$

where c is the speed of light, $\bar{f}_{\text{ph}}$ is the average photon distribution function over the frequency range, ω_0 is the central frequency, and $\Delta\omega$ is the bandwidth of the beam. The entropy flux density is

$$I_S = \frac{1}{4\pi^2} k_{\text{B}} c^{-2} [(1+\bar{f}_{\text{ph}}) \ln(1+\bar{f}_{\text{ph}}) - \bar{f}_{\text{ph}} \ln \bar{f}_{\text{ph}}] \omega_0^2 \Delta\omega \sin^2 \delta \tag{8}$$

If either the bandwidth $\Delta\omega$ or the divergence δ of the source collapses to zero, Eq. (7) implies that $\bar{f}_{\text{ph}} \to \infty$, to ensure that the energy flux density remains finite. In this limit, one can show that Eq. (8) yields $I_S \to 0$, i.e., the entropy carried by monochromatic or unidirectional radiation is zero, so that one can characterize an ideal laser beam as pure work or high-grade energy.

To analyze the limiting efficiency of laser cooling of solids, it is useful to define the flux temperature T_F of the radiation, which is given by

$$T_F \equiv \frac{I_{\dot{E}}}{I_S} = \frac{\hbar \omega_0}{k_{\text{B}}} \frac{\bar{f}_{\text{ph}}}{(1+\bar{f}_{\text{ph}}) \ln(1+\bar{f}_{\text{ph}}) - \bar{f}_{\text{ph}} \ln \bar{f}_{\text{ph}}} \tag{9}$$

Again, for an ideal laser, we have $T_F \to \infty$, which is consistent with the zero entropy at a finite irradiance. It can be further deduced that the flux temperature of narrowband radiation propagating in a well-defined direction is higher than that of broadband radiation propagating in all directions.

In laser cooling of solids, according to the first law of thermodynamics, we have $P_{\text{out}} = P_{\text{in}} + Q_c$. The cooling coefficient of performance is defined in the usual way for a refrigerator as $\eta = Q_c / P_{\text{in}}$. The maximum value of η is the Carnot limit, η_{C}, and is determined by the second law of thermodynamics. The entropy carried by the fluorescence cannot be less than the sum of the entropy withdrawn from the cooling sample and the entropy transported in by the pump laser, i.e.,

$$\frac{P_{\text{out}}}{T_f} \geqslant \frac{P_{\text{in}}}{T_o} + \frac{Q_c}{T} \tag{10}$$

where T is the steady-state operating temperature of the refrigerator, and T_f and T_o are the flux temperatures of the fluorescence and pump radiation fields, respectively. The reversible Carnot limit is obtained by choosing the equality sign in Eq. (10). Finally we have

$$\eta_{\text{C}} = \frac{T - \Delta T}{T_f - T} \tag{11}$$

where $\Delta T = T T_f / T_o$. In the limit of $T_o \gg T_f$, the efficiency (11) would reduce to the Carnot form.

Consider an example using actual values relevant to laser cooling of Yb^{3+}:ZBLANP [37]. The temperatures of the pump laser and the fluorescence are calculated, using Eq. (9), to be $T_o = 7$

$\times 10^{11}$ K and $T_f = 1760$ K, respectively. Thus the Carnot efficiency of this optical cooler is about 20% at room temperature, and it diminishes approximately linearly to zero as $T \rightarrow 0$. However, the actual cooling efficiency achieved to date is only around 3%, which indicates that much irreversibility has been produced in the process. One might use a longer pumping wavelength to obtain a higher cooling efficiency, but the absorption coefficient of Yb^{3+} would become too small. As a result, the trace impurity absorption will dominate over the Yb^{3+} absorption, and the cooling efficiency η decreases. To reduce these irreversibilities introduced into this process, the sample should be purified to suppress the trace absorption.

If the system is pumped hard, so that the threshold is reached, the emission will become stimulated rather than spontaneous. Although this would accelerate the cycling, the thermodynamic efficiency limit will become too small, due to the low entropy of the stimulated emission fields [38]. Anyway, one cannot expect to get cooling and laser output simultaneously by using a laser input.

6 Advance in Laser Cooling of Rare-Earth Ion-Doped Solids

Ion-doped solids were the first class of materials on which laser actions were demonstrated [39], and were attempted early for laser cooling.

6.1 Experimental Investigations. The earliest experiment was performed on Nd^{3+}:YAG by Geusic et al. at Bell Laboratories in 1968 [9], just a few years after he demonstrated the first laser action in this transition-metal doped crystal [40]. This focused on flash-lamp-pumped crystals of Nd^{3+}:YAG, with the fluorescence from one crystal being used for the cooling of another. When compared with an undoped reference sample, the neodymium-doped sample showed reduced heating, but net cooling was not observed. This was conjectured to be due to impurities in the crystals, which offset the cooling effects. A simple model yielded the rate of temperature change, and the results agreed with the experiment. Later rare-earth-ion doped solids were demonstrated for laser emission, and these materials immediately became attractive for laser cooling purpose, since the optical $4f$ levels are shielded from the surrounding by the filled $5s$ and $5p$ shells, leading to the suppression of the multiphonon relaxation. In 1995, Epstein et al. reported the first experimental success of laser cooling in solids [11]. The absorption and fluorescence spectra were measured, as shown in Fig. 6(*a*), where the mean fluorescence wavelength is marked. As the pumping wavelength was tuned longer than the mean fluorescence wave-length, a maximum local temperature decrease of 0.3 K was detected, as shown in Fig. 6(*b*). However, the cooling effect diminishes when the pumping wavelength is tuned further away, since the off-resonance absorptivity becomes too small. The cooling efficiencies achieved were up to 2%. Later, ytterbium-doped glasses have been cooled by nearly 70 K below room temperature and have reportedly cooled at temperatures as low as 77 K [16]. Edwards et al. [14] demonstrated a prototypical cryogenic refrigerator based on Yb^{3+}:ZBLANP pumped with a 1.6 W Ti:sapphire laser and measured a temperature decrease of 48 K from room temperature. Cooling from a low starting temperature in various Yb^{3+}-doped glasses has also been observed, suggesting that a cryogenic refrigerator with an extended dynamic range can be built. Mungan et al. [12] observed local cooling in a Yb^{3+}:ZBLANP sample at temperatures between 100 and 300 K, maintaining a cooling efficiency of about 0.01 through this range. Local cooling between 77 K and room temperature by photothermal deflection and spectroscopic techniques in a fluorochloride glass (Yb^{3+}:CNBZn) and a fluoride glass (Yb^{3+}:BIG) were reported by Fernandez et al. [16]. The cooling efficiency was shown to change with temperature, varying between 0.02 and 0.006 in the two materials. Gosnell cooled a Yb^{3+}-doped fiber by the amount $\Delta T = 65$ K from room temperature [15]. His experimental apparatus was very carefully designed, as shown in Fig. 7. This record was again pushed to 92 K below room temperature by Epstein et al. in 2005 [21]. Continuous progress has been made [19,22,23] towards achieving cryogenic temperatures.

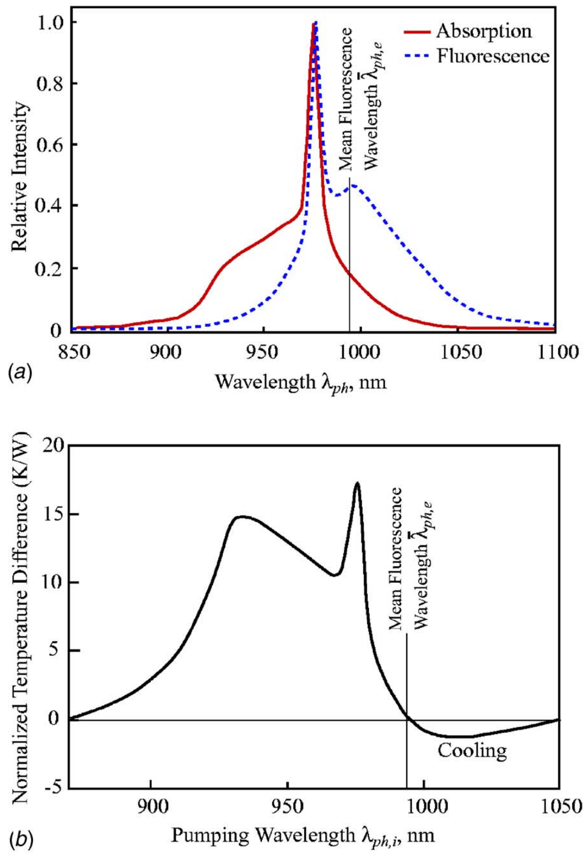


Fig. 6 (*a*) The absorption and fluorescence spectra of Yb^{3+}:ZBLANP, with the mean fluorescence wavelength marked [11]. (*b*) The normalized temperature difference with respect to the pumping wavelength. Cooling is detected as the pumping wavelength is tuned longer than the mean fluorescence wavelength [11].

In these experiments, accurate temperature measurement is of great importance. Temperature can be measured by contact or noncontact methods. A noncontact method, fluorescence thermometry, has been used in many of the existing experiments [11,15]. It is based on the fact that the fluorescence spectrum of the glass is independent of the pump laser wavelength but only dependent on temperature. A spectrum-temperature relation can be calibrated over a wide range of temperatures using a thermostat. Then the observed spectrum is compared to the reference spectra and the temperature is determined. This method is capable of measuring temperature without disturbing the original system, and is very suitable for systems as fine as laser cooling. Thermocouple is another obvious choice for its simplicity. However, it introduces an external thermal load to the cooling element and may reduce or even eliminate the cooling effect. This method is thus usually used for a rough examination of whether or not the system can be cooled, but not for an accurate temperature measurement [14].

To enhance laser cooling performance, one perspective is finding new materials, including dopants and hosts. Thulium-doped glasses are good candidates, since thulium has a transition resonance at 1.8 μm, whereas the ytterbium transition is near 1 μm. Then the thulium system is capable of obtaining the same amount of cooling power with a much smaller pumping photon energy. Hoyt et al. showed cooling of a sample of Tm^{3+}-doped ZBLANP by 1.2 K from room temperature, under vacuum, when approxi-

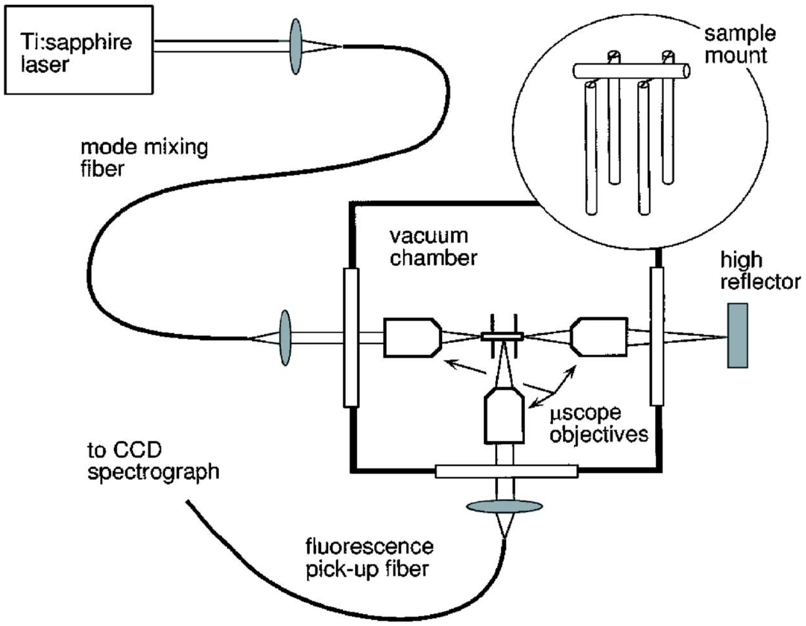


Fig. 7 Experimental apparatus used for observing laser cooling of Yb:ZBLANP [15]. Pump radiation from a cw Ti:sapphire laser undergoes mode scrambling within an external multimode silica fiber before injection into the sample fiber, which is positioned upon a sample mount (inset), imposing an extremely low conductive thermal load. Unabsorbed pump radiation is collected from the output end of the sample fiber and reinjected into the sample with the help of an external high reflector. Finally, emitted fluorescence is collected with a third internal optic and is spectrally resolved to determine the temperature.

mately 3 W of laser power at 1.9 μm was incident on the sample from a periodically poles lithium-niobate-based optical parametric oscillator in turn pumped by a 20 W cw Nd:YAG laser. Their results indicated a cooling efficiency of 3.4%, which compared favorably with the efficiencies achieved for ytterbium-doped glasses. To date, Yb^{3+} and Tm^{3+} are the only two rare-earth elements on which laser cooling has been demonstrated, although some other elements are considered to be good candidates.

A number of ytterbium-doped crystal hosts were also studied by Bowman and Mungan [22], to determine which of them can be cooled. Besides a number that did not exhibit net cooling, they found that crystals of ytterbium-doped $KGd(WO_4)_2$ can indeed be cooled. This was the first demonstration of anti-Stokes laser cooling of a crystal. Fernandez et al. achieved internal cooling of other ytterbium doped glasses (CNBZn and BIG) [16,41]. Recently, another set of results showing cooling of crystals was reported by Epstein et al. [17]. These results show cooling of 2.3% ytterbium-doped YAG crystals and 3% ytterbium-doped Y_2SiO_5. A comparison is made with ytterbium-doped ZBLAN, which shows that the cooling efficiency is slightly larger in ZBLAN, but the thermal and mechanical properties of YAG may be advantageous for some applications. Several other groups have tried to find other materials suitable for cooling, but even materials that have shown promise from an analysis of their absorption and emission spectra have had either too low a quantum efficiency or excessive absorption due to impurities for cooling to be achieved. Using the experimental data for lifetimes, Hoyt studied the possibility of laser cooling of a few Tm-doped solids, and identified those which have a potential for cooling due to their high quantum efficiencies [42].

Another perspective to enhance the cooling performance is modifying the structure. Gosnell used a long, thin optical fiber as the host, in order to increase the optical pathlength for a larger absorptivity, and to reduce the external thermal load [15]. To further enhance the absorptivity, some researchers [21,42] placed the sample between two dielectric mirrors of high reflectance at the pumping wavelength only, as shown in Fig. 8(*a*). In this way the laser pumping is reflected back and forth, while the fluorescence can escape. Recognizing that the mirrors may bring in extra loss, Heeg et al. proposed an alternative approach [20,43]. By locating the cooling medium inside a laser cavity, it was efficiently pumped by the inherent multipassing and high circulating power of the laser resonator. Recently, Ruan et al. attempted this problem from a new perspective: Nanostructure [44]. The medium in their model is a Yb^{3+} doped Y_2O_3 nanopowder, as shown in Fig. 8(*b*). They predicted that the absorption can be significantly enhanced due to the optimized dopant concentration, the size effect of the phonon density of states, and the multiple scattering of the pumping photons.

6.2 Theoretical Analysis. The basic principles of laser cooling and its thermodynamic validity provided motivation for the above mentioned experiments. Except for these, more detailed theoretical analysis achieved very limited progress, compared to the rapid improvements of laser cooling experiments. This is mainly due to very complicated physical mechanisms under the laser cooling process. Lamouche considered the temperature dependence of cooling efficiency [45]. By analyzing the temperature dependence of fluorescence and absorption spectra of Yb^{3+}:ZBLAN, they concluded that cooling would decrease with decreasing temperature, as shown in Fig. 9. Fernandez et al. used the Fermi's Golden Rule to interpret their experimental results, by assuming that the absorption is dominated by the phonon-assisted process [16]. The absorption rate $\dot{\gamma}_{e,a}$ is given by

$$\dot{\gamma}_{e,a} = \sum_f \dot{\gamma}_{e,i-f} = \frac{2\pi}{\hbar_{\rm P}} \sum_f |M_{fi}|^2 \delta(\hbar_{\rm P}\omega_{ph,i} + \hbar_{\rm P}\omega_p - \hbar_{\rm P}\omega_{e,g}) \tag{12}$$

where $\hbar_{\rm P}\omega_{e,g}$ is the energy difference of the two electronic levels, and M_{fi} is the interaction matrix element. The δ function guaran-

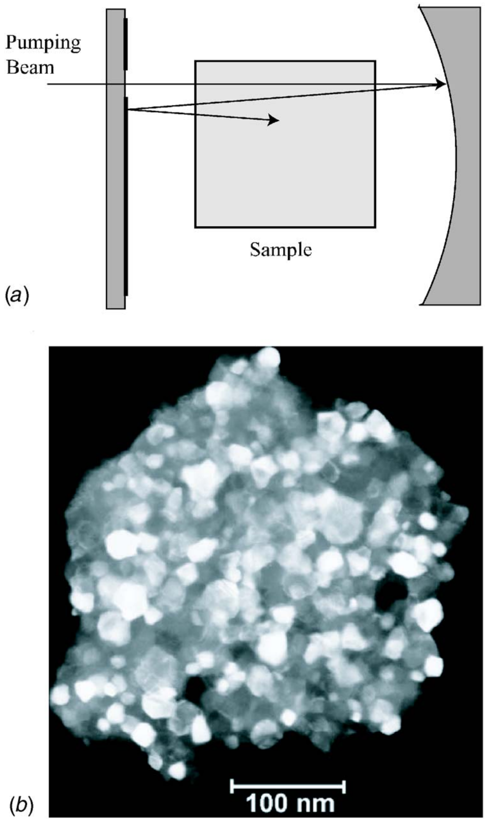


Fig. 8 Modified structures for enhanced laser cooling performance: (*a*) Cavity arrangement for multiple passes [42], (*b*) A micrograph of $Yb^{3+}:Y_2O_3$ nanopowder [51].

tees the energy conservation, that the sum of the pumping photon energy and the phonon energy should be equal to the electronic gap energy. In this theory, the decreasing cooling efficiency with decreasing temperature is due to the fact that the phonon distribution function decreases. Heeg et al. investigated the effect of another important factor, the fluorescence reabsorption, on the cooling performance [46,47]. The random process of fluorescence reabsorption and trapping were analyzed in solid-state optical materials. Using the absorption and fluorescence spectra of Yb^{3+}:ZBLAN as input data, they employed a random-walk model to test analytical approximations of the fluorescence escape efficiency and cooling efficiency, including reflections at the boundary. They concluded that moderate concentration and sample size should be used to avoid the dominance of fluorescence reabsorption.

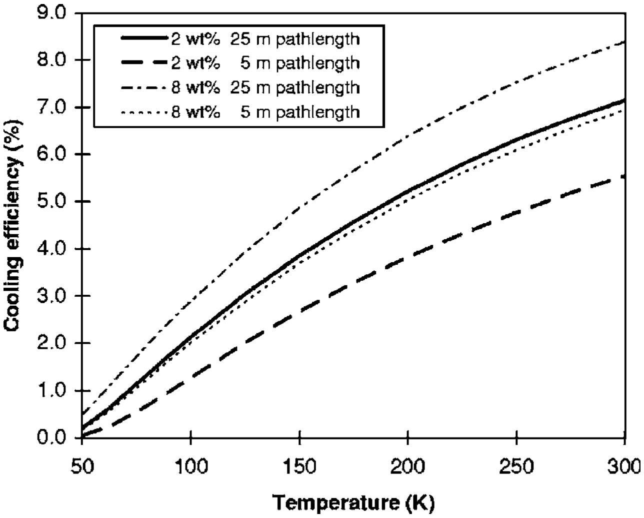


Fig. 9 Predicted cooling efficiency as a function of temperature, for the 2 and 8 wt % $Yb^{3+}:Y_2O_3$ in the linear regime for two optical pathlengths [45].

Despite these interesting theoretical studies, a more important issue, if not the most important, is to develop a theory that can provide the criterion of the material selection. Although there are many factors limiting the laser cooling performance, the materials properties, including dopants and hosts, are essential. An ideal dopant-host pair should allow for an effective ion-phonon coupling, which otherwise should not be large enough to result in multiphonon relaxation. Ruan et al. recently used the Fermi's Golden Rule to decouple various limiting factors, aiming to develop an atomic level of understanding [44]. They found that the cooling performance is limited by the population of the three carriers and their couplings. The nanostructure was proposed to be capable of enhancing the carrier populations, and the understanding of the ion-phonon coupling mechanism is to be established and is crucial for engineering the materials for laser cooling.

7 Advance of Laser Cooling in Semiconductors

Given the recent advance in fabrication of semiconductor devices and their use as lasers, it is not surprising that a number of researchers have considered these candidates for optical cooling. A GaAs/GaInP heterostructure was studied for possible cooling by Gauck et al. [48]. They observed blue-shifted luminescence but did not see net cooling because coupling inefficiency caused luminescence reabsorption. Finkeiben et al. [49] detected local cooling in the area of the pumping beam spot that was due to anti-stokes photoluminescence in a GaAs quantum-well structure, recording a temperature drop of 7 K from liquid-nitrogen temperature. Sheik-Bahae [24] et al. performed a theoretical analysis considering nonradiative decay and luminescence reabsorption, and proposed the feasibility of laser cooling in semiconductors. Nevertheless, no bulk cooling of semiconductors has yet been realized experimentally, where the luminescence trapping due to total internal reflection remains a major obstacle.

One possible solution is to use an index-matched dome lens attached to the cooling element [48], as shown in Fig. 10(*a*). Since the dome lens is much larger than the heterostructure, the fluorescence emitted at any angle would become nearly normal to the dome surface, making the extraction much more efficiently. The drawback is that the dome lens introduces more external thermal load and increases the system size. Another possible solution is to use a nanogap structure [50], as shown in Fig. 10(*b*). For the onset of the total internal reflection, there exists an evanescent wave on the other side of the surface which has an exponentially decaying amplitude and does not transfer any energy. If another surface is brought closely to the first one at a distance shorter than the wavelength, the evanescent waves will be coupled to the second surface and become propagating waves. In this way, the originally trapped fluorescence can be efficiently coupled out of the cooling element.

8 Conclusions and Outlook

Over the past decade, substantial progress has been made in the laser cooling of solids. Cooling of rare-earth-ion doped solids has been demonstrated and improved, and cooling in semiconductors are anticipated in the near future. Current work is focused on finding new materials and structures that can enhance the cooling performance, or that can be cooled (for semiconductors). Ways are also being explored in which a practical optical cooler might be engineered.

Further research into laser cooling of solids should firstly explore new materials and structures for enhanced cooling performance. Material properties determine the maximum cooling effi-

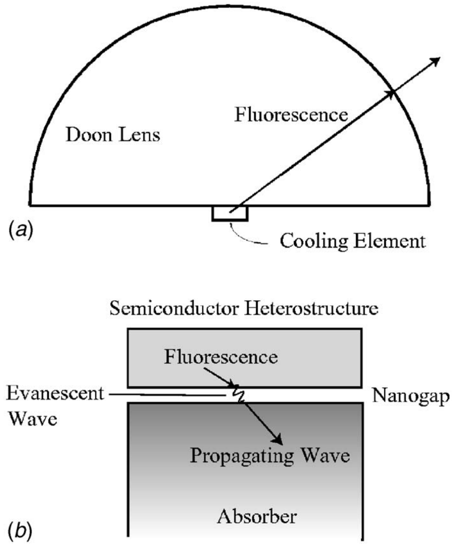


Fig. 10 Possible structures for extracting the fluorescence from the high refractive-index semiconductors: (*a*) An index-matching dome lens attached on the cooling element [48]. (*b*) A nanogap structure to couple the evanescent waves out of the cooling element [50].

ciency that can be achieved, so further advances could result from investigation of other dopants, such as thulium and others, to replace ytterbium. Consideration should also be given to other host materials, including crystals and glasses. New structures such as nanostructure are promising due to the quantum size effect, and should be considered as a major direction. Given the rate of advancement of semiconductor manufacturing techniques, they may be ideal candidates for laser cooling in the near future, since the electronic structures could be engineered to optimize the cooling efficiency for specific applications. Equally important to the experimental investigations is the development of a theory that can provide a criterion for the material selection. Although this is very challenging, considering the complexity of the electron-phonon coupling, useful results are expected to be obtained in the near future, which can then provide guidance for searching new materials and structures.

Acknowledgment

We are thankful to Professor Stephen Rand for many discussions/suggestions.

Nomenclature

A = surface area (m^2)
c_o = speed of light in vacuum (m/s)
E = complex electric field (V/m)
E = energy (J)
e_e = electron charge (C)
f_p, f_{ph} = phonon, photon distribution function
I = intensity (W/m^2)
M = interaction matrix
P = power (W)
Q = power (W)
r = interatomic distance (m)
$\dot{S}$ = energy conversion (W)
T = temperature (K)

Greek Symbols

α = absorptivity
δ = divergence
$\dot{\gamma}$ = transition rate (1/s)
$\dot{\gamma}_{e,a}$ = transition probability per unit time per unit energy density [1/s/(J/m^3)])
η_c = cooling efficiency
$\eta_{e\text{-}ph}$ = quantum efficiency
μ = magnetic permeability (H/m)
$\boldsymbol{\mu}_e$ = dipole moment vector (C-m)
ψ = wave function
ω = angular frequency (rad/s)

Subscripts

a = absorption
c = cooling
e = emission, electron
F = flux
f = fluid, final state, fluorescence
g = gap
i = incident, initial state
o = free space
p = phonon
ph = photon
r = radiative, real part of a complex number
S = entropy

References

[1] Chu, S., 1998, "The Manipulation of Neutral Particles," Rev. Mod. Phys., **70**, pp. 685–706.
[2] Phillips, W., 1998, "Laser Cooling and Trapping of Neutral Atoms," Rev. Mod. Phys., **70**, pp. 721–741.
[3] Cornell, E., and Wieman, C., 2002, "Nobel Lecture: Bose-Einstein Condensation in a Dilute Gas, the First 70 Years and Some Recent Experiments," Rev. Mod. Phys., **74**, pp. 875–893.
[4] Ketterle, W., 2002, "Nobel Lecture: When Atoms Behave as Waves: Bose-Einstein Condensation and the Atom Laser," Rev. Mod. Phys., **74**, pp. 1131–1151.
[5] Pringsheim, P., 1929, "Zwei Bemerkungen Uber den Unterschied von Lumineszenz und Temperaturstrahlung," Z. Phys., **57**, pp. 739–746.
[6] Vavilov, S., 1945, "Some Remarks on the Stokes Law," J. Phys. (Moscow), **9**, pp. 68–73.
[7] Vavilov, S., 1946, "Photoluminescence and Thermodynamics," J. Phys. (Moscow), **10**, pp. 499–501.
[8] Landau, L., 1946, "On the Thermodynamics of Photoluminescence," J. Phys. (Moscow), **10**, pp. 503–506.
[9] Kushida, T., and Geusic, J., 1968, "Optical Refrigeration in Nd-Doped Yttrium Aluminium Garnet," Phys. Rev. Lett., **21**, pp. 1172–1175.
[10] Djeu, N., and Whitney, W., 1981, "Laser Cooling by Spontaneous Anti-Stokes Scattering," Phys. Rev. Lett., **46**, pp. 236–239.
[11] Epstein, R., Buckwald, M., Edwards, B., Gosnell, T., and Mungan, C., 1995, "Observation of Laser-Induced Fluorescent Cooling of a Solid," Nature (London), **377**, pp. 500–503.
[12] Mungan, C., Buchwald, M., Edwards, B., Epstein, R., and Gosnell, T., 1997, "Internal Laser Cooling of Yb^{3+}-Doped Glass Measured Between 100 and 300 K," Appl. Phys. Lett., **71**, pp. 1458–1460.
[13] Luo, X., Eisaman, M., and Gosnell, T., 1998, "Laser Cooling of a Solid by 21 K Staring From Room Temperature," Opt. Lett., **23**, pp. 639–641.
[14] Edwards, B., Anderson, J., Epstein, R., Mills, G., and Mord, A., 1999, "Demonstration of a Solid-State Optical Cooler: An Approach to Cryogenic Refrigeration," J. Appl. Phys., **86**, pp. 6489–6493.
[15] Gosnell, T., 1999, "Laser Cooling of a Solid by 65 K Starting From Room Temperature," Opt. Lett., **24**, pp. 1041–1043.
[16] Fernandez, J., Mendioroz, A., Garcia, A., Balda, R., and Adam, J., 2000, "Anti-Stokes Laser-Induced Internal Cooling of Yb^{3+}-Doped Glasses," Phys. Rev. B, **62**, pp. 3213–3217.
[17] Epstein, R., Brown, J., Edwards, B., and Gibbs, A., 2001, "Measurements of Optical Refrigeration in Ytterbium-Doped Crystals," J. Appl. Phys., **90**, pp. 4815–4819.
[18] Rayner, A., Friese, M., Truscott, A., Heckenberg, N., and Rubinsztein-Dunlop, H., 2001, "Laser Cooling of a Solid From Ambient Temperature," J. Mod. Opt., **48**, pp. 103–114.
[19] Hoyt, C., Hasselbeck, M., Sheik-Bahae, M., Epstein, R., Greenfield, S., Thiede, J., Distel, J., and Valencia, J., 2003, "Advances in Laser Cooling of Thulium-Doped Glass," J. Opt. Soc. Am. B, **20**, pp. 1066–1074.
[20] Heeg, B., Stone, M., Khizhnyak, A., Rumbles, G., Mills, G., and Debarber, P., 2004, "Experimental Demonstration of Intracavity Solid-State Laser Cooling of Yb^{3+}: ZrF$_4$-BaF$_2$-LaF$_3$-AlF$_3$-NaF Glass," Phys. Rev. A, **70**, pp. 021401(R).
[21] Thiede, J., Distel, J., Greenfield, S., and Epstein, R., 2005, "Cooling to 208 K by Optical Refrigeration," Appl. Phys. Lett., **86**, p. 154107.
[22] Bowman, S., and Mungan, C., 2000, "New Materials for Optical Cooling," Appl. Phys. B, **71**, pp. 807–811.
[23] Mendioroz, A., Fernandez, J., Voda, M., Al-Saleh, M., Balda, R., and Garcia-

Adeva, A., 2002, "Anti-Stokes Laser Cooling in Yb^{3+}-Doped KPb_2Cl_5 Crystal," Opt. Lett., **27**, pp. 1525–1527.

[24] Sheik-Bahae, M., and Epstein, R., 2004, "Can Laser Light Cool Semiconductors?," Phys. Rev. Lett., **92**, p. 247403.

[25] Merzbacher, E., 1998, *Quantum Mechanics*, 3rd ed., Wiley, New York.

[26] Ruan, X., and Kaviany, M., 2005, "Enhanced Nonradiative Relaxation and Photoluminescence Quenching in Random, Doped Nanocrystalline Powders," J. Appl. Phys., **97**, p. 104331.

[27] Kaviany, M., 2002, *Principles of Heat Transfer*, Wiley, New York.

[28] Rayner, A., Heckenberg, N., and Rubinsztein-Dunlop, H., 2003, "Condensed-Phase Optical Refrigeration," J. Opt. Soc. Am. B, **20**, pp. 1037–1053.

[29] Frey, R., Micheron, F., and Pocholle, J., 2000, "Comparison of Peltier and Anti-Stokes Optical Coolings," J. Appl. Phys., **87**, pp. 4489–4498.

[30] Bowman, S., 1999, "Laser Without Internal Heat Generation," IEEE J. Quantum Electron., **35**, pp. 115–121.

[31] Lavi, R., and Jackel, S., 2000, "Thermally Boosted Pumping of Neodymium Lasers," Appl. Opt., **39**, pp. 3093–3098.

[32] Lupei, V., Lupei, A., Pavel, N., Taira, T., Shoji, I., and Ikesue, A., 2001, "Laser Emission Under Resonant Pump in the Emitting Level of Concentrated Nd:YAG Ceramics," Appl. Phys. Lett., **79**, pp. 590–592.

[33] Lupei, V., Pavel, N., and Taira, T., 2002, "Highly Efficient Continuous-Wave 946 nm Nd:YAG Laser Emission Under Direct 885 nm Pumping," Appl. Phys. Lett., **81**, pp. 2677–2679.

[34] Lupei, V., 2003, "Efficiency Enhancement and Power Scaling of Nd Lasers," Opt. Mater., **24**, pp. 353–368.

[35] Li, C., Liu, Q., Gong, M., Chen, G., and Yan, P., 2004, "Modeling of Radiation-Balanced Continuous-Wave Laser Oscillators," J. Opt. Soc. Am. B, **21**, pp. 539–542.

[36] Mungan, C., and Gosnell, T., 1999, "Laser Cooling of Solids," Appl. Opt., **40**, pp. 161–228.

[37] Mungan, C., 2005, "Radiation Thermodynamics with Applications to Lasing and Fluorescent Cooling," Am. J. Phys., **73**, pp. 1458–1460.

[38] Ruan, X., Rand, S., and Kaviany, M., 2006, "Entropy and Efficiency of Laser Cooling of Solids," Phys. Rev. B, submitted.

[39] Maiman, T., 1960, "Molecular Dynamics Study of the Structures and Bulk Moduli of Crystals in the System CaO-MgO-Al_2O_3-SiO_2," Nature (London), **187**, pp. 493–494.

[40] Geusic, J., Marcos, H., and Vanuitert, L., 1964, "Laser Oscillations in Nd-Doped Yttrium Aluminum Yttrium Gallium+Gadolinium Garnets," Appl. Phys. Lett., **4**, pp. 182–184.

[41] Fernandez, J., Mendioroz, A., Garcia, A., Balda, R., Adam, J., and Arriandiaga, M., 2001, "On the Origin of Anti-Stokes Laser-Induced Cooling of Yb^{3+}-Doped Glass," Opt. Mater., **16**, pp. 173–179.

[42] Hoyt, C., ed., 2003, "PhD Dissertation: Laser Cooling in Thulium-Doped Solids," University of New Mexico, Albuquerque, NM.

[43] Heeg, B., Rumbles, G., Khizhnyak, A., and DeBarber, A., 2002, "Comparative Intra-Versus Extra-Cavity Laser Cooling Efficiencies," J. Appl. Phys., **91**, pp. 3356–3362.

[44] Ruan, X. L., and Kaviany, M., 2006, "Enhanced Laser Cooling of Rare-Earth-Ion-Doped Nanocrystalline Powders," Phys. Rev. B, **73**, p. 155422.

[45] Lamouche, G., and Lavallard, P., 1998, "Low Temperature Laser Cooling With a Rare-Earth Doped Glass," J. Appl. Phys., **84**, pp. 509–516.

[46] Heeg, B., and Rumbles, G., 2003, "Influence of Radiative Transfer on Optical Cooling in the Condensed Phase," J. Appl. Phys., **93**, pp. 1966–1973.

[47] Heeg, B., Debarber, P., and Rumbles, G., 2005, "Influence of Fluorescence Reabsorption and Trapping on Solid-State Optical Cooling," Appl. Opt., **44**, pp. 3117–3124.

[48] Gauck, H., Gfroerer, T., Renn, M., Cornell, E., and Bertness, K., 1997, "External Radiative Quantum Efficiency of 96% From a GaAs/GaInP Heterostructure," Appl. Phys. A, **64**, pp. 143–147.

[49] Finkeißen, E., Potemski, M., Wyder, P., Vina, L., and Weimann, G., 1999, "Cooling of a Semiconductor by Luminescence Up-Conversion," Appl. Phys. Lett., **75**, pp. 1258–1260.

[50] Sheik-Bahae, M., 2005, "Laser Cooling in Semiconductors," *4th Annual Workshop on Laser Cooling of Solids*, University of New Mexico, Albuquerque, NM.

[51] Pires, A., Serra, O., Heer, S., and Gudel, H., 2005, "Low-Temperature Upconversion Spectroscopy of Nanosized Y_2O_3: Er,Yb Phosphor," J. Appl. Phys., **98**, p. 063529.

F. Marquier
M. Laroche
R. Carminati
J.-J. Greffet

Laboratoire d'Énergétique Moléculaire et
Macroscopique;
Combustion, École Centrale Paris,
Centre National de la Recherche Scientifique,
Grande Voie des Vignes,
92295 Châtenay-Malabry Cedex,
France

Anisotropic Polarized Emission of a Doped Silicon Lamellar Grating

Thermal emission of a doped silicon grating has been studied in the plane perpendicular to the grooves. We show how the excitation of surface plasmons produce a resonant emission weakly depending on the polarization and azimuthal angle. We analyze in detail the polarization and angular dependence of the emission out of the plane perpendicular to the grooves. Two kinds of thermal sources, directional and quasi-isotropic, are studied. They have been designed in a previous paper. We also compute the total hemispherical emissivity of these gratings. In addition we show that in applications such as radiative cooling, these sources are less efficient than other structures. [DOI: 10.1115/1.2360594]

1 Introduction

Emission and absorption of radiation by an air-material flat interface is controlled by the reflectivity of the interface. This property is not an intrinsic property and depends significantly on the roughness of the interface. This opens the way to the design of spatially selective or directional emission/absorption properties. This is of uppermost relevance for the design of efficient infrared radiation sources, thermophotovoltaic sources, selective absorbers for detection, photovoltaic energy conversion, or radiative cooling to name a few applications. In particular, it has been known that the interface roughness design can take advantage of the surface waves to produce remarkable properties.

Total absorption of a plane wave by a metallic grating has been known for years [1–3]. The incident wave can be coupled to a surface wave through the periodicity of the grating so that there is 100% absorption at a given wavelength and at a given angle of incidence [2,3]. According to Kirchhoff's law, the emissivity is also 100%. It has been shown that materials supporting electromagnetic surface resonances such as surface phonon-polaritons or surface plasmon-polaritons could produce thermal emission in well-defined directions by ruling a shallow grating on their surface [4–6]. Thus numerical and experimental works have been made on metals, supporting surface plasmon polaritons [7,8] as well as on polar materials, supporting surface phonon polaritons [4,5]. In this paper we show that the directional polarized emissivity can be significantly modified by ruling a grating. Extending previous works restricted to emission in the plane perpendicular to the grooves to nonzero azimuthal angles, we find that surface plasmons contribute to s-polarized emission. We consider two kinds of sources: A quasi-isotropic source at a given wavelength or a directional thermal source. In the latter case, the emission lobes are similar to those observed for antennas in the radiowave frequency range. It has been shown that this directionality is a signature of the spatial coherence of the thermal source. The spatial coherence length has been calculated numerically and measured [5].

Due to the existence of free carriers in doped silicon, this material can support surface plasmon-polaritons. Several groups have worked on this material, and they have shown that the radiative properties of doped silicon could be modified by introducing some periodicity on the surface [9–13]. In particular, Hesketh et al. have demonstrated the role of surface plasmon polaritons in the modification of the thermal emission of doped silicon structures [10,11]. The grating parameters have a great impact on the efficiency of the coupling between surface polaritons and propagating waves. The development of rigorous calculation code based on the rigorous coupled wave analysis [14] allows us to engineer the emission properties of materials such as doped silicon. We have optimized p-doped silicon lamellar gratings with two different behaviors: A directional thermal source or a quasi-isotropic source for a given wavelength [6,15]. Let us note that thermal emission has been also studied for two-dimensional photonic crystals [16–21] and interesting works have been reported for one-dimensional photonic crystals (binary layers of dielectric materials on a metallic or semiconductor bulk) [22–25].

Previous studies were restricted to a plane of incidence (or emission) perpendicular to the grating lines. In this plane, a surface wave can only be coupled to a p-polarized propagating wave but never to s-polarized waves. This is no longer true when considering a plane of emission with an azimuthal angle $\phi \neq 0$ deg (defined below in Fig. 1). A few studies have dealt with that kind of problem [26,27]. They have shown, for example, that with $\phi \neq 0$ deg one could have conversion of polarization [27] and that a surface wave could be excited by a s-polarized propagating wave [26]. In our case, it is obviously interesting to see how the azimuthal angle acts on the behavior of thermal sources.

The increase of the emissivity of p-doped silicon has led us to study the possibility of cooling silicon structures [15]. Those calculations were based on approximations related to the fact that the emissivity was calculated in the planeperpendicular to grating lines and in p-polarization only. In this paper, we study the emissivity integrated over all polarization and all emission angles. The paper is organized as follows. In Sec. 2, we present the two kinds of sources which have been previously optimized. The modifications of the radiative properties of these sources when observing in a plane which is not perpendicular to the grooves of the gratings are discussed in Sec. 3. Finally, we address the radiative cooling of p-doped silicon by these sources in Sec. 4 by performing an exact calculation of the total hemispherical emissivity of the structures.

2 Silicon Infrared Sources

In this section, we consider lamellar gratings made of heavily p-doped silicon with a carrier concentration $N = 2.5 \times 10^{20}$ cm^{-3}. We choose this large carrier concentration to have a low emissivity in the middle and far infrared. Indeed heavily doped-silicon behaves as a metal in the infrared and has a high reflectivity. Note that for n-doped silicon, the result is similar. The parameters of such a grating are the period Λ, the filling factor F and the depth h. The radiative properties of these structures in a plane of emission perpendicular to the grooves of the grating (see Fig. 1) have

---

Contributed by the Heat Transfer Division of ASME for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received January 24, 2006; final manuscript received June 21, 2006. Review conducted by Zhuomin M. Zhang. Paper presented at the Nano/microscale radiative transfer 2005.

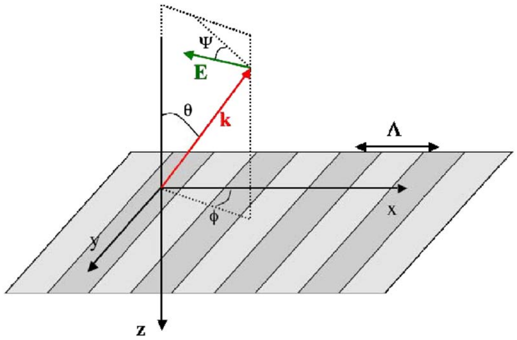


Fig. 1 Emission of a grating in the direction (θ,ϕ). The electric field E makes an angle Ψ with the plane $(\mathbf{z},\mathbf{k})$. $\Psi=0$ deg corresponds to the *p*-polarization and $\Psi=90$ deg to the *s*-polarization.

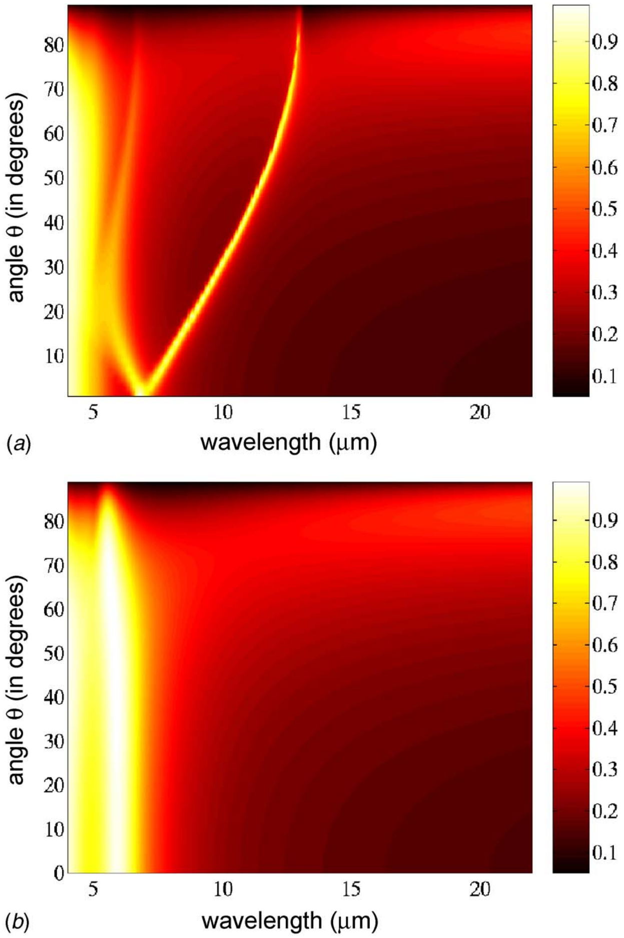


Fig. 2 Emissivity pattern in *p*-polarization for *p*-doped silicon gratings with $N=2.5\times10^{20}$ cm^{-3} (the grooves of the grating are perpendicular to the incident plane: $\phi=0$ deg). (*a*) Directional source with parameters $\Lambda=6.3$ μm, $F=0.4$ and $h=0.6$ μm. (*b*) Quasi-isotropic source with $\Lambda=2.5$ μm, $F=0.8$, and $h=0.6$ μm.

been already studied in previous papers [6,15]. These works were limited to the study of *p*-polarized waves because resonant absorption or emission due to excitation of surface plasmons cannot be observed for *s*-polarized waves in a plane perpendicular to the grooves. This is no longer the case out of this plane. In this paper, we use the same formula as in [6] for the dielectric constant of doped silicon.

We have implemented the rigorous coupled wave analysis method (RCWA) [14,28,29] to numerically calculate the directional emissivity at a given wavelength λ for any polarization and for any plane of emission which is not perpendicular to the lines of the grating. The aim of the RCWA method is the calculation of the transmitted and reflected efficiencies for each diffraction order of the grating for a given incident plane wave. For an infinite periodic structure, the electromagneticfield is pseudoperiodic (Bloch modes). The method is based on Fourier's decomposition of electric and magnetic fields and also of the geometry of the grating. Maxwell's equations yields a linear system of equations for Fourier's order of the fields. The number of equations depends on the truncature order of Fourier's series. The greater this number is, the better the numerical convergence of the calculation. Solving this system we calculate the reflected and transmitted efficiencies. Furthermore the algorithm described by Hugonin and Chateau in Ref. [14] is a S-matrix algorithm, which is more efficient than the T-matrix algorithm. The calculation of the efficiencies with 1% of precision for a given wavelength and angle of incidence, takes 0.1 s using a 4 Gbytes computer with a 1.5 Ghz CPU frequency.

Figure 1 summarizes the different notations to define the direction of emission. The structure is periodic along the $\mathbf{x}$-axis. A direction of observation is given by the wave vector $\mathbf{k}$ and defined by two angles θ and ϕ (see Fig. 1). The plane $(\mathbf{z},\mathbf{k})$ is the plane of emission or sagittal plane. The angle Ψ gives the polarization: $\Psi=0$ deg corresponds to the *p*-polarization and $\Psi=90$ deg to the *s*-polarization.

Depending on the grating parameters, the emissivity can be either highly directional or quasi isotropic as shown in Refs. [6,15]. In the present study we work on the same sources. For convenience, we give the parameters of such gratings. A directional source is produced with the following parameters: Period $\Lambda=6.3$ μm, filling factor $F=0.4$, and depth $h=0.6$ μm. A quasi-isotropic source has been optimized for different parameters: $\Lambda=2.5$ μm, filling factor $F=0.8$, and depth $h=0.6$ μm. In Fig. 2, we represent the directional emissivity in a plane (θ,λ) for the directional source (Fig. 2(*a*)) and for the quasi-isotropic one (Fig. 2(*b*)). For both sources, we represent the emissivity in *p*-polarization and for an azimuthal angle $\phi=0$ deg.

For both figures, the emissivity is high for wavelengths below $\lambda=4.2$ μm which corresponds to the plasma frequency of *p*-doped silicon for a carrier concentration $N=2.5\times10^{20}$ cm^{-3}. In this wavelength range the doped silicon is a dielectric medium. Beyond the plasma wavelength, the heavily doped-silicon behaves as a metal and has a low emissivity. This is the typical emissivity of a plane interface of *p*-doped silicon. In the region of low emissivity, one can see a line corresponding to a high emissivity. This line is due to the resonant thermal excitation of surface waves (surface plasmon-polaritons) which are electromagnetic resonances of the system. The fact that the line is narrow shows that the emission pattern has a lobe with a small angular width. The mechanism of this emission can be described as follows. A surface wave can be excited by the random currents thermally excited. Since the surface wave has a decay length when propagating along the surface that is much larger than the period of the grating, it is scattered by many grooves. The interference of the field scattered by all the grooves produces a lobe in a well-defined direction. A detailed study of the physical mechanism is made in Refs. [4–6,30].

For the first source (Fig. 2(*a*)), the wavelength of the emission varies with the angle of observation θ. Therefore, at a given wavelength, the emissivity is high at an angle θ well defined and the source is directional. For the second source (Fig. 2(*b*)), the line is broader and the wavelength of emission is fixed at $\lambda=5.9$ μm for any angles of observation. Therefore, at this wavelength, the emissivity reaches unity for a broad range of angles and the source can be considered as quasi-isotropic at this wavelength. This is a direct consequence of the structure of the dispersion relation. At the particular wavelength of 5.9 μm, the dispersion relation is almost

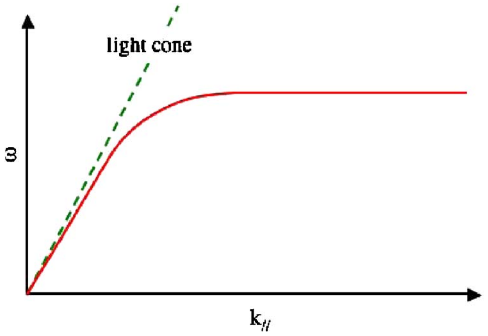


Fig. 3 Solid line: Schematic dispersion relation of a surface plasmon polariton. Dashed line: Light cone. The solid curve lies out of the light cone, so that the wave is evanescent. $k_{\|}$ is the projection of the wave vector in the plane (x,y), and $\omega=2\pi c/\lambda$ is the pulsation.

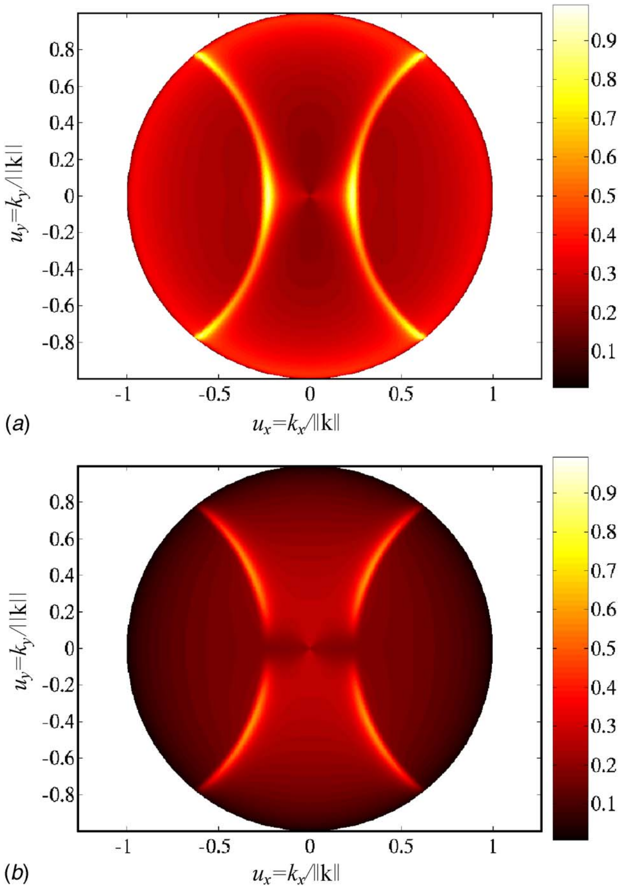


Fig. 4 Polar representation of the emissivity at $\lambda=8.0$ μm for the directional source in both *p*-polarization (*a*) and *s*-polarization (*b*)

flat as seen in Fig. 3. This entails that the condition of emission by a surface plasmon given by the grating relation $k_x^{\text{inc}}=k_{sp}+p2\pi/\Lambda$ is satisfied for any value of the emission angle since k_{sp} takes any value larger than ω/c. The surface wave is a *p*-polarized phenomenon, and the high emissivity cannot appear when observing the source in *s*-polarization. However, when the azimuthal angle ϕ is varying, the surface waves can be coupled to a *s*-polarized propagating one. Therefore, the emission pattern is not the same in other planes of emission.

3 Variation of the Emissivity with the Azimuthal Angle ϕ

In this section, we focus on the emissivity out of the plane perpendicular to the grooves, i.e., for $\phi \neq 0$. Only a few studies have dealt with this issue [26,27]. In what follows, we shall consider the same gratings as in Sec. 2. We introduce the factors $u_x=\sin\theta\cos\phi$ and $u_y=\sin\theta\sin\phi$ which are the projections on, respectively, the *x*-axis and *y*-axis of the unit vector $\mathbf{u}=\mathbf{k}/\|\mathbf{k}\|$ where $\mathbf{k}$ is the wave vector of the emitted wave.

We start by considering the emission at a wavelength of $\lambda=8$ μm of the directional source. The results are shown in Fig. 4. We observe an emissivity close to the value of a plane surface except along two lines where the emissivity approaches 1. In order to analyze the emissivity as a function of angle and wavelength, it is easier to take advantage of Kirchhoff's law and analyze the absorptivity. The two lines observed in Fig. 4 can be interpreted as a resonant absorption due to the excitation of surface plasmons. Indeed, at this wavelength, the dispersion relation of a surface plasmon yields a well-defined value of the wave vector given by $k_{\text{SP}}=\omega/c\sqrt{\epsilon/(\epsilon+1)}$ [31]. A resonant absorption due to the excitation of a surface plasmon is thus expected if the condition $\mathbf{k}_{//}^{\text{inc}}+(p2\pi/\Lambda)\mathbf{x}=\mathbf{k}_{\text{SP}}$ is fulfilled. We have schematically represented the dispersion relation in Fig. 5. We have also represented the location of the points given by $\mathbf{k}_{\text{SP}}-(p2\pi/\Lambda)\mathbf{x}$ with $p=\pm 1$. When the extremity of the incident wave vector matches the dispersion relation of the surface plasmon polariton, a resonance is expected. It is seen in Fig. 4 that the emissivity maxima are along these lines.

Another feature is observed in Fig. 4, the variation of the value of the maximum emissivity with the angle ϕ. It is seen that for values of ϕ close to zero, the emissivity is maximum for *p*-polarization whereas it is very small for *s*-polarization. This behavior can be explained by taking into account the polarization of the electric field associated with surface waves. A surface wave propagating along the interface with a wave vector $\mathbf{k}_{\text{SP}}$ has two components of the electric field: one is parallel to $\mathbf{k}_{\text{SP}}$ and the other is parallel to the normal of the surface. It follows that the surface current density associated with the surface plasmon in the material has a component parallel to $\mathbf{k}_{\text{SP}}$. An external electric field can excite a surface plasmon provided that it has a component parallel to $\mathbf{k}_{\text{SP}}$. This condition explains the ϕ dependence and the polarization dependence of the emissivity. Let us consider, for

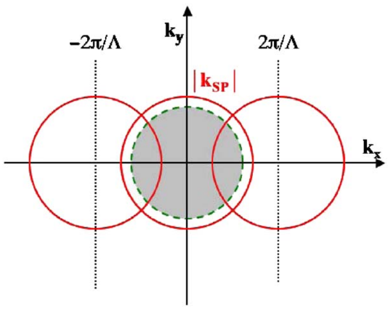


Fig. 5 Dispersion relation of surface plasmon-polaritons in a (k_x,k_y) plane for a fixed frequency on a grating. The radius of the dashed circle is the wave vector modulus in free space. Inside this circle (in gray), the emitted waves have a real *z*-component of k, so that they are propagating waves. Outside the dashed circle, the waves are evanescent waves. The solid circle at the center has a radius equal to the modulus of the wave vector of the surface plasmon-polariton k_{SP}. With a grating of period Λ, this solid circle is reproduced with a period $2\pi/\Lambda$: It represents the surface waves diffraction at the orders ±1. One can see that a part of the solid line is lying now in the light cone, so that the corresponding surface plasmon-polaritons are coupled to propagating waves.

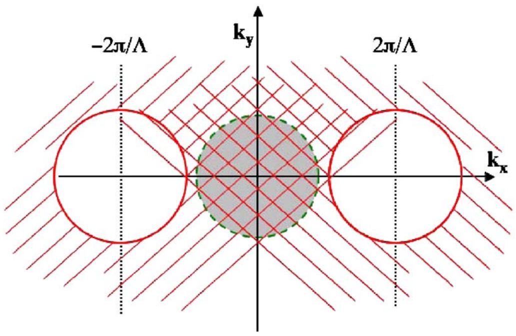


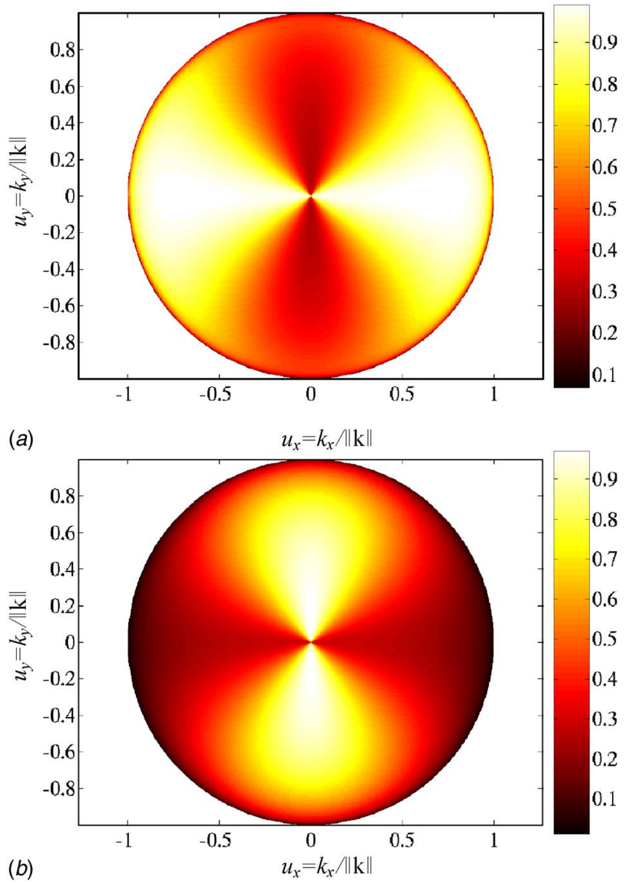
Fig. 6 Wave vector diagram of a lamellar grating periodic along the *x*-axis. The material can support surface waves.

example, the absorption of a plane wave with an angle $\phi=0$ so that the incident wavector is along the x-axis $(k_x^{\mathrm{inc}},0)$. A plasmon can in principle be excited with a wavector $(k_x=k_{\mathrm{SP}},\ k_y=0)$ provided that $k_x^{\mathrm{inc}}=k_{\mathrm{SP}}-2\pi/\Lambda$. If the incident plane wave is p-polarized, the incident field can excite a current density parallel to the x-axis and, therefore, excite the surface wave. Conversely, if the wave is s-polarized, the incident electric field is along the y-axis and, therefore, cannot excite a current density along the x-axis.

Let us note that similar patterns have been also observed in studies on diffuse light scattered by diffraction gratings [32–35]. Those studies consider periodic structures on which a microroughness exists. The grating or the microroughness can couple the incident plane wave into any surface mode. These modes are then scattered by the grating. The light re-emitted in this way forms what is called the diffuse bands. In the case of thermal emission by p-doped silicon gratings, the surface plasmon-polaritons are excited by the random currents due to thermal motion of electrons. The subsequent coupling of surface plasmons into propagating waves through the grating is similar in both cases explaining the similarity between both patterns.

We now consider the absorption at the frequency corresponding to the asymptote of the dispersion relation (see Fig. 3). At this particular frequency, the dispersion relation is flat as discussed previously. It follows that any incident wave will excite a surface plasmon provided that $|\mathbf{k}_{//}^{\mathrm{inc}}+(p2\pi/\Lambda)\mathbf{x}|>\omega/c$. For a flat interface, the domain of existence of surface plasmon is the whole plane (k_x,k_y) except for the area of a disk as depicted in Fig. 6. The condition of excitation of a surface plasmon is given by $\mathbf{k}_{//}^{\mathrm{inc}}=\mathbf{k}_{\mathrm{SP}}-(p2\pi/\Lambda)\mathbf{x}$. From Fig. 6, we see that any incident direction excites a surface plasmon when the period Λ is larger than the wavelength. As previously mentioned, we must pay attention to the polarization to see how efficient is the excitation of the surface wave. To have an efficient coupling between the incident wave and the surface plasmon, one knows that the incident electric field must have a component parallel to $\mathbf{k}_{\mathrm{SP}}$. In p-polarization the maximal effect should be obtained for a zero azimuthal angle. Regarding s-polarization, the maximal effect should be obtained for $\phi=90$ deg. The numericalresults are shown in Fig. 7. The angular and polarization dependence is clearly observed.

Finally, we consider the unpolarized emissivity by summing the two polarized contributions (see Fig. 8). It is remarkable to note that despite the strong anisotropy of the surface, the total emissivity does not depend on the angle i and is slightly above 0.6. These results strongly suggest that by texturing the interface along both x and y directions, one should be able to produce a surface that would emit unpolarized light in all directions with an emissivity close to 1. A possible application is the design of an emitter or absorber at a given frequency.



(*a*)

(*b*)

Fig. 7 Polar representation of the emissivity at λ=5.9 μm for the quasi-isotropic source in both *p*-polarization (*a*) and *s*-polarization (*b*)

4 Total Hemispherical Emissivity

Let us now consider the total hemispherical emissivity of such sources. To obtain this value, we have to spectrally and angularly integrate the previously calculated emissivity. The total hemispherical emissivity of the doped silicon gratings has been calculated in a previous paper [15]. However, the calculations of the directional emissivity were performed only in p-polarization and for a plane of incidence perpendicular to the grooves of the grat-

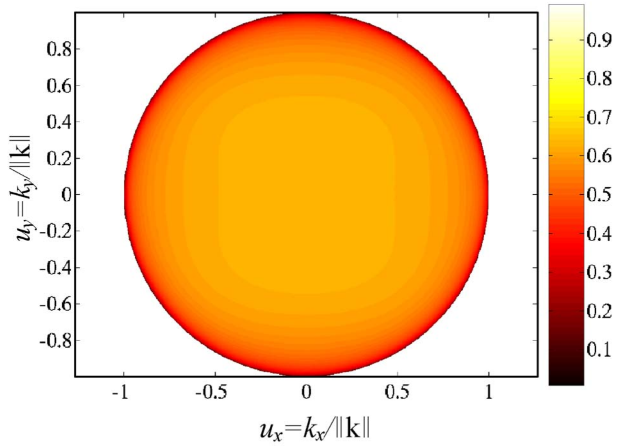


Fig. 8 Polar representation of the average emissivity in both *p*- and *s*-polarization at λ=5.9 μm for the quasi-isotropic source

Table 1 Total hemispherical emissivity at $T=100\,^\circ$C

	Directional source	Isotropic source	Plane interface
1D approximation (*p*-polarization)	0.29	0.376	0.24
exact result (*p*-polarization)	0.28	0.31	0.24
exact result (both polarizations)	**0.218**	**0.249**	**0.179**

ings. The emissivity integrated over all angles was estimated by assuming that the emissivity was the same in 2π steradians. With the calculation of the directional emissivity in a plane which is not perpendicular to the grooves of the grating, we can now perform an exact calculation of the total hemispherical emissivity. The total hemispherical emissivity at the temperature T is defined by the ratio of the hemispherical emitted flux with the emitted flux of the blackbody at the same temperature T

$$\epsilon(T)=\frac{1}{\sigma T^4}\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi/2}\int_{0}^{\infty}\epsilon(\theta,\phi,\lambda)I_\lambda^0(T)\cos\theta\sin\theta d\theta d\phi d\lambda \tag{1}$$

where $\epsilon(\theta,\phi,\lambda)$ is the directional emissivity of the surface for a given wavelength λ, $I_\lambda^0(T)$ is the Planck's function and σ the Stefan's constant. We perform the calculation in both polarizations by a 2 deg step in θ, a 5 deg step in ϕ, and about a 100 nm step in wavelength. It corresponds to about 350,000 points, i.e., about 10 h computational time. The convergence of the integration has been checked by multiplying the number of points by 4: We obtained the same results. The average of the two results yields a value of the total hemispherical emissivity at temperature T. The results for the directional and quasi-isotropic sources previously defined are given in Table 1 for a temperature $T=100\,^\circ$C. In order to have a reference, we have also calculated the total hemispherical emissivity for a plane interface of p-doped silicon with the same carrier concentration. In p-polarization, the approximation that we made in Ref. [15] was in fair agreement. In particular, the values of the total hemispherical emissivity for the directional source are closer than expected. Let us consider now the average values. One can see that the quasi-isotropic source has a higher emitted flux than the directional source. The quasi-isotropic source emits about 40% more radiation than the plane interface. However, when decreasing the doping of silicon, the emissivity of the plane interface increases. In fact there is an optimum for the doping about $N=10^{18}$ cm^{-3}. Indeed, below this doping level the p-doped silicon becomes less absorbing so that the emissivity decreases. Therefore, the gratings are not interesting in applications such as radiative cooling. Their interest is to allow the design of efficient sources with spectral or directional selectivity. As indicated in the previous section, the overall emissivity could be significantly improved by using a two-dimensional structure of the surface.

5 Conclusions

We have reported the design of p-doped silicon gratings to emit thermal light in a directional or quasi-isotropic manner. As demonstrated in previous works, such thermal sources can be engineered with metals, polar materials and doped silicon [4–6]. We are interested here in p-doped silicon. However, it is obviously possible to optimize such gratings with n-doped silicon. We have shown that s-polarized waves out of the plane perpendicular to the grooves can be coupled with surface modes. We have highlighted the fact that the source remains directional or quasi-isotropic even when considering a plane not perpendicular to the grating lines. In that case, both polarizations must be taken into account. Finally, performing an exact calculation of the total hemispherical emissivity, we have shown that these sources radiate more power than a plane interface with the same doping. They are interesting for their high spectral and directional selectivity.

Nomenclature

c = light speed
F = filling factor of a grating
h = depth of a grating
$I_\lambda^0(T)$ = specific intensity at temperature T
$\mathbf{k}$ = wave vector
n = diffraction order (integer)
N = free carrier concentration in doped silicon (holes concentration in this paper)
T = temperature
$\mathbf{u}$ = unit vector supporting the wave vector
$\mathbf{x}$ = unit vector of the in-plane axis perpendicular to the lines of a grating
$\mathbf{y}$ = unit vector of the in-plane axis parallel to the lines of a grating
$\mathbf{z}$ = unit vector of the axis perpendicular to the structure
$\epsilon(T)$ = total hemispherical emissivity at temperature T
$\epsilon(\theta,\phi,\lambda)$ = directional spectral emissivity of a structure
θ = polar angle of emission
λ = wavelength
Λ = period of a grating
ω = pulsation ($\omega=2\pi/\lambda$)
σ = Stefan's constant
ϕ = azimuthal angle of emission
Ψ = polarization angle

Subscripts

inc = incident
SP = surface plasmon
x = coordinate along the x-axis
y = coordinate along the y-axis
// = projection in the plane (x,y)

References

[1] Wood, R. W., 1902, "On a Remarkable Case of Uneven Distribution of Light in a Diffraction Grating Spectrum," Philos. Mag., **4**, pp. 396–402.
[2] Maystre, D., and Hutley, M. C., 1976, "The Total Absorption of Light by a Diffraction Grating," Opt. Commun., **19**, pp. 431–436.
[3] Petit, R., 1980, *Electromagnetic theory of gratings*, Springer Verlag, Berlin.
[4] Greffet, J. J., Carminati, R., Joulain, K., Mulet, J. P., Mainguy, S., and Chen, Y., 2002, "Coherent Thermal Emission of Light by Thermal Sources," Nature (London), **416**, pp. 61–64.
[5] Marquier, F., Joulain, K., Mulet, J. P., Carminati, R., Greffet, J. J., and Chen, Y., 2004, Phys. Rev. B **69**, p. 155412.
[6] Marquier, F., Joulain, K., Mulet, J. P., Carminati, R., and Greffet, J. J., 2004, "Engineering Infrared Emission Properties of Silicon in the Near Field and the Far Field," Opt. Commun., **237**, pp. 379–388.
[7] Kreiter, M., Oster, J., Sambles, R., Herminghaus, S., Mittler-Neher, S., and Knoll, W., 1999, "Thermally Induced Emission of Light From a Metallic Diffraction Grating, Mediated by Surface Plasmons," Opt. Commun., **200**, pp. 1–4.
[8] Laroche, M., Arnold, C., Marquier, F., Carminati, R., Greffet, J. J., Collin, S., Bardou, N., and Pelouard, J. L., 2005, "Highly Directional Radiation Generated by Tungsten Thermal Source," Opt. Lett., **30**, pp. 2623–2625.
[9] Hesketh, P. J., Zemel, J. N., and Gebhart, B., 1986, "Organ Pipe Radiant Modes of Periodic Micromachined Silicon Surfaces," Nature (London), **324**, pp. 549–551.
[10] Hesketh, P. J., Zemel, J. N., and Gebhart, B., 1988, "Polarized Spectral Emittance From Periodic Micromachined Surfaces. I. Doped Silicon: The Normal Direction," Phys. Rev. B, **37**, pp. 10795–10802.
[11] Hesketh, P. J., Zemel, J. N., and Gebhart, B., 1988, "Polarized Spectral Emittance From Periodic Micromachined Surfaces. II. Doped Silicon: Angular Variation," Phys. Rev. B, **37**, pp. 10803–10813.
[12] Hava, S., Auslender, M., Lacquet, B. M., Coetzer, P. J., and Swart, P. L., 1995, "IR Transmission and Reflection Study of Lamellar Silicon Grating-Wafer Structures," Infrared Phys. Technol., **36**, pp. 639–647.
[13] Auslender, M., and Hava, S., 1995, "Zero Infrared Reflectance Anomaly in Doped Silicon Lamellar Gratings. I. From Antireflection to Total Absorption," Infrared Phys. Technol., **36**, pp. 1077–1088.
[14] Chateau, N., and Hugonin, J. P., 1994, "Algorithm for the Rigorous Coupled-

Wave Analysis of Grating Diffraction," J. Opt. Soc. Am. A, **11**, pp. 1321–1331.

[15] Laroche, M., Marquier, F., Carminati, R., and Greffet, J. J., 2005, "Tailoring Silicon Radiative Properties," Opt. Commun., **250**, pp. 316–320.

[16] Sai, H., Yugami, H., Akiyama, Y., Kanamori, Y., and Hane, K., 2001, "Spectral Control of Thermal Emission by Periodic Microstructured Surfaces in the Near-Infrared Region," J. Opt. Soc. Am. A, **18**, pp. 1471–1476.

[17] Maruyama, S., Kashiwa, T., Yugami, H., and Esachi, M., 2001, "Thermal Radiation From Two-Dimensionally Confined Modes in Microcavities," Appl. Phys. Lett., **79**, pp. 1393–1395.

[18] Pralle, M. U., Moelders, N., McNeal, M. P., Puscasu, I., Greenwald, A. C., Daly, J. T., Johnson, E. A., George, T., Choi, D. S., El-Kady, I., and Biswas, R., 2002, "Photonic Crystal Enhanced Narrow-Band Infrared Emitters," Appl. Phys. Lett., **81**, pp. 4685–4687.

[19] Kusunoki, F., Kohama, T., Hiroshima, T., Fukumoto, S., Takahara, J., and Kobayashi, T., 2004, "Narrow-Band Thermal Radiation With Low Directivity by Resonant Modes Inside Tungsten Microcavities," Jpn. J. Appl. Phys., Part 1, **43**, pp. 5253–5258.

[20] Enoch, S., Simon, J. J., Ascoubas, L., Elamy, Z., Lemarquis, F., Torchio, P., and Albrand, G., 2005, "Simple Layer-By-Layer Photonic Crystal for the Control of Thermal Emission," Appl. Phys. Lett., **86**, p. 261101.

[21] Laroche, M., Carminati, R., and Greffet, J. J., 2006, "Coherent Thermal Antenna Using a Photonic Crystal Slab," Phys. Rev. Lett., **96**, p. 123903.

[22] Hu, L., Schmidt, A., Narayanaswamy, A., and Chen, G., 2004, "Effects of Periodic Structures on the Coherence Properties of Blackbody Radiation," ASME J. Heat Transfer, **126**, pp. 786–792.

[23] Celanovic, I., Perreault, D., and Kassakian, J., 2005, "Resonant-Cavity Enhanced Thermal Emission," Phys. Rev. B, **72**, p. 075127.

[24] Lee, B. J., Fu, C. J., and Zhang, Z. M., 2005, "Coherent Thermal Emission From One-Dimensional Photonic Crystals," Appl. Phys. Lett., **87**, p. 071904.

[25] Lee, B. J., and Zhang, Z. M., 2006, "Coherent Thermal Emission From Modified Periodic Multilayer Structures," ASME J. Heat Transfer (to be published).

[26] Kretschmann, M., Leskova, T. A., and Maradudin, A. A., 2003, "Conical Propagation of a Surface Polariton Across a Grating," Opt. Commun., **215**, pp. 205–223.

[27] Elston, S. J., Bryan-Brown, G. P., and Sambles, J. R., 1991, "Polarization Conversion From Diffraction Gratings," Phys. Rev. B, **44**, pp. 6393–6400.

[28] Moharam, M. G., Grann, E. B., Pommet, D. P., and Gaylord, T. K., 1995, "Formulation for Stable and Efficient Implementation of the Rigorous Coupled-Wave Analysis of Binary Gratings," J. Opt. Soc. Am. A, **12**, pp. 1068–1076.

[29] Li, L., 1997, "New Formulation of the Fourier Modal Method for Crossed Surface-Relief Gratings," J. Opt. Soc. Am. A, **14**, pp. 2758–2767.

[30] Joulain, K., Mulet, J. P., Marquier, F., Carminati, R., and Greffet, J. J., 2005, "Surface Electromagnetic Waves Thermally Excited: Radiative Heat Transfer, Coherence Properties and Casimir Forces Revisited in the Near-Field," Surf. Sci. Rep., **57**, pp. 59–112.

[31] Raether, H., 1988, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer Verlag, Berlin.

[32] Simon, J. M., and Gonzãlez Pagliere, J. M., 1986, "Diffuse Light in Diffraction Gratings and Surface Plasma Oscillations," Opt. Acta, **33**, pp. 1035–1049.

[33] Simon, J. M., and Gonzãlez Pagliere, J. M., 1988, "Diffuse Light Bands From Diffraction Gratings: Polarization Dependence and Phase Behaviour," J. Mod. Opt., **35**, pp. 1549–1555.

[34] Depine, R. A., and Brudny, V. L., 1991, "Speckle Patterns Generated by Rough Surfaces With a Periodic Component," J. Mod. Opt., **38**, pp. 2281–2293.

[35] Brudny, V. L., Iemmi, C. C., and Ledesma, S. A., 1994, "Diffuse Light Bands for *p* and *s* Polarization in Microrough Diffraction Gratings," Optik (Jena), **97**, pp. 23–30.

Coherent Thermal Emission From Modified Periodic Multilayer Structures

B. J. Lee

Z. M. Zhang[1]
Fellow ASME
e-mail: zzhang@me.gatech.edu

George W. Woodruff School of Mechanical Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332

Enhancement of thermal emission and control of its direction are important for applications in optoelectronics and energy conversion. A number of structures have been proposed as coherent emission sources, which exhibit a large emissivity peak within a narrow wavelength band and at a well-defined direction. A commonly used structure is the grating, in which the excited surface polaritons or surface waves are coupled with propagating waves in air, resulting in coherent emission for p polarization only. One-dimensional photonic crystals can also support surface waves and may be modified to construct coherent emission sources. The present study investigates coherent emission from a multilayer structure consisting of a SiC *film coated atop a dielectric photonic crystal (PC). By exciting surface waves at the interface between* SiC *and the PC, coherent emission is predicted for both p and s polarizations. In addition to the excitation of surface waves, the emission from the proposed multilayer structure can be greatly enhanced by the cavity resonance mode and the Brewster mode.* [DOI: 10.1115/1.2401194]

Keywords: electromagnetic, emitting, microstructures, radiation, thin films

1 Introduction

Surface microstructures can strongly affect the directional behavior of absorption and emission due to multiple reflections [1] and diffraction [2,3], allowing the radiative properties to be controlled. A number of studies have dealt with the effects of microstructures on radiative properties, especially thermal emission from structured surfaces [4–6]. In fact, enhancement of thermal emission and control of its direction are crucial for the applications in energy conversion systems, such as solar cells [7], thermophotovoltaic devices [8], and radiation filters [9]. Because of the important applications to radiative energy transfer and energy conversion devices, the study of engineered micro/nanostructures with desired thermal radiative characteristics has become a new frontier of radiation heat transfer research.

Coherent emission is associated with sharp spectral peaks (temporal coherence) and/or narrow angular lobes in well-defined directions (spatial coherence). Recently, coherent thermal emission has been demonstrated using gratings made of a metal [10] or a polar material [11] by exciting surface waves, which propagate along the interface and decay into both media. However, surface waves from grating structures are limited to the *p* polarization (when the magnetic field is parallel to the grooves, and the emission direction is perpendicular to the grooves). Metamaterials with a refractive index much less than unity [12] or single negative materials (with alternative negative permittivity and negative permeability layers) [13] have also been proposed as coherent emission sources; however, these materials have not been fabricated for the near-infrared and midinfrared spectral regions, where thermal radiation is the most important.

Recently, periodic microstructures such as photonic crystals (PCs) were extensively investigated to engineer radiative properties for specific applications [14–18]. A PC is a periodic array of unit cells (i.e., photonic lattices in analog to those in real crystals), which replicate infinitely in one, two, or three dimensions. Analogously to electron movement in crystals, electromagnetic wave propagation in PCs should also satisfy the Bloch condition [19]. Similarly, due to the periodicity, a PC exhibits band structures consisting of pass bands and stop bands when the frequency is plotted against the wave vector. In the pass band, for instance, electromagnetic waves can propagate through a PC, whereas in the stop band, no energy-carrier waves can exist inside a PC, and only an oscillating but evanescently decaying field may exist. The existence of stop bands enables a PC to be used in many optoelectronic devices such as band-pass filters and waveguides. One-dimensional (1D) PCs are usually made of alternating layers with two lossless dielectrics. Since no energy can transmit into the PC in the stop band, the 1D PC composed of several unit cells can be used as an excellent reflector in the resonance cavity [20]. Furthermore, Yeh et al. [21] showed that a 1D PC can support a surface mode or surface wave for both *p* and *s* polarizations in the stop band. Since surface waves are nonradiative, an attenuated total reflection (ATR) configuration is normally employed to excite the surface waves at the interface between a PC and air [22]. It has been demonstrated that, if a metallic layer is coated on a 1D PC, surface waves can be directly excited by propagating waves in air, resulting in a sharp reduction in the reflectance [23].

The present paper describes a comprehensive investigation of the multilayer structure proposed by Lee et al. [24] for coherent emission. The structure is made of a polar material and a semi-infinite 1D PC, as shown in Fig. 1. Here, the PC is terminated or modified by placing a polar material on top of it to couple surface waves with propagating waves in air. SiC is an ideal polar material for coherent emission in the midinfrared spectral region. The unit cell of a 1D PC is composed of the alternating dielectrics with different refractive indices n_a and n_b, respectively, and defined by the following thicknesses: a_1, d_b, and a_2, where $a_1+a_2=d_a$ with a period (or lattice constant) $\Lambda=d_a+d_b$. In practice, only a finite number of unit cells are required and the thin films can be deposited onto a suitable substrate, such as silicon. The thickness a_1, which may be varied from 0 to d_a, plays an important role in tuning the coherent emission wavelength. The spectral–directional emissivity of the SiC–PC structure is calculated using the 1D transfer matrix formulation [25–27]. By examining conditions that

---

[1] Corresponding author.

Contributed by the Heat Transfer Division of ASME for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received November 6, 2005; final manuscript received April 5, 2006. Review conducted by Suresh V. Garimella. Paper presented at the 2005 ASME International Mechanical Engineering Congress (IMECE2005), Orlando, FL, November 5–11, 2005.

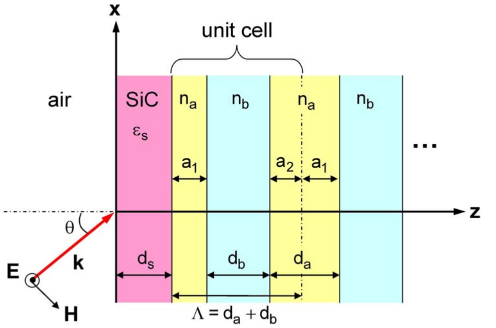


Fig. 1 Schematic of the multilayer structure made of a SiC layer coated on a semi-infinite 1D PC for an *s*-polarized wave incident from air. The unit cell of the PC consists of a dielectric (type *a*) on both sides of a dielectric (type *b*) with a total thickness (lattice constant) $\Lambda = d_a + d_b$, where $d_a = a_1 + a_2$. The surface termination is determined by the thickness of the dielectric (a_1) next to the SiC film with a dielectric function of ε_s.

cause a large emission in a narrow spectral range and an angular lobe in a well-defined direction, a regime map is developed to identify the emissivity enhancement due to three different mechanisms: the excitation of surface waves, cavity resonance mode, and the Brewster mode. The effects of geometric parameters and arrangement of the high- and low-index dielectrics in the PC on the emission characteristics are investigated.

2 Theory

Similar to grating structures [11,12] and single negative materials [14], the excitation of surface waves plays an important role to enable coherent thermal emission from the proposed SiC–PC structure. However, surface waves involving PCs possess some unique features different from those involving gratings or single negative materials. For clear understanding of coherent thermal emission from the SiC–PC structure, brief discussions about PCs as well as the surface waves from PCs are presented below with key equations.

2.1 Fundamentals of Photonic Crystals. While three-dimensional photonic crystals with complicated structures can be fabricated and used in a number of applications, the fundamental physics can be illustrated using 1D PCs and can be easily generalized into two- or three-dimensional structures [19]. As illustrated in Fig. 2, a 1D PC is a periodic multilayer structure, whose unit cell is composed of alternating dielectrics with different refractive indices. It is assumed that the PC is infinitely extended to the right, and type-*a* dielectric starts at $z=0$. Wave propagation in periodic media is analogous to the motion of electrons in crystalline materials [28]. From such analogy, the electric field vector in the 1D PC for a monochromatic electromagnetic wave of angular frequency ω should satisfy the Bloch condition given by [19]

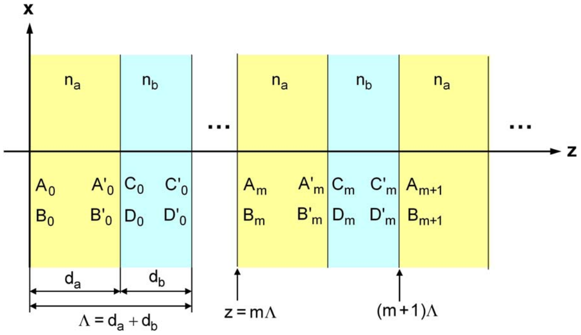


Fig. 2 Schematic of the amplitudes of forward and backward waves in a semi-infinite 1D PC, in the right half space. The unit cell of the 1D PC is made of two dielectric layers, type *a* and type *b*, and has a period $\Lambda = d_a + d_b$.

$$\boldsymbol{E}(x,y,z,t) = \boldsymbol{u}(z)e^{iKz}e^{i(k_x x + k_y y - \omega t)} \tag{1}$$

where $\boldsymbol{u}(z+\Lambda)=\boldsymbol{u}(z)$ is a periodic function of z, k_x, and k_y are the parallel components of the wave vector and are the same in all layers as required by the phase-matching condition, and K is the magnitude of the Bloch wave vector, which is parallel to the z axis for 1D PC. Here, K is a characteristic parameter of the PC and is the same for all layers. The wave vector components in the z direction are $k_{z,a}$ (in medium a) and $k_{z,b}$ (in medium b), which are determined by $k_x^2+k_y^2+k_{z,a}^2=n_a^2\omega^2/c^2$ and $k_x^2+k_y^2+k_{z,b}^2=n_b^2\omega^2/c^2$, respectively. Here, c is the speed of light in air, whose optical properties are assumed to be the same as those of vacuum. From the Bloch condition, the electric field in the 1D PC satisfies

$$\boldsymbol{E}(x,y,z+\Lambda,t) = \boldsymbol{E}(x,y,z,t)e^{iK\Lambda} \tag{2}$$

Because of axial symmetry, the coordinate can always be rotated around the z axis to make $k_y=0$. It is a common practice to analyze the wave propagation for two separate polarizations. In the case of s polarization, the electric field is perpendicular to the x-z plane. Therefore, the electric field component in the y direction can be expressed as [21]

$$E_y(x,z) = \begin{cases} [A_m e^{ik_{z,a}(z-m\Lambda)} + B_m e^{-ik_{z,a}(z-m\Lambda)}]e^{ik_x x}, & m\Lambda \le z \le (m\Lambda + d_a) \\ [C_m e^{ik_{z,b}(z-m\Lambda-d_a)} + D_m e^{-ik_{z,b}(z-m\Lambda-d_a)}]e^{ik_x x}, & (m\Lambda + d_a) \le z \le (m+1)\Lambda \end{cases} \tag{3}$$

where the time dependence term $\exp(-i\omega t)$ is omitted for simplicity; m is a nonnegative integer; A_m and C_m are the amplitudes of forward waves; and B_m and D_m are those of backward waves at the interfaces, as shown in Fig. 2. Due to phase shifts from wave propagation through the medium, the amplitudes at the other interface of the corresponding layer are $A'_m=e^{ik_{z,a}d_a}A_m$, $B'_m=e^{-ik_{z,a}d_a}B_m$, $C'_m=e^{ik_{z,b}d_b}C_m$, and $D'_m=e^{-ik_{z,b}d_b}D_m$. Boundary conditions require that the tangential components of the electric and magnetic fields (E_y and H_x) be continuous at each interface. The coefficients A_m and B_m at $z=m\Lambda$ are related to those at $z=(m+1)\Lambda$ via the propagation matrix (**P**) and dynamical matrix (**D**)

$$\begin{pmatrix} A_m \\ B_m \end{pmatrix} = (\mathbf{P}_a\mathbf{D}_a^{-1}\mathbf{D}_b)(\mathbf{P}_b\mathbf{D}_b^{-1}\mathbf{D}_a)\begin{pmatrix} A_{m+1} \\ B_{m+1} \end{pmatrix} \tag{4}$$

Detailed discussions about the matrix formulation and expressions of the matrix elements can be found elsewhere [25–27]. From Eq. (2), the ratio of the electric fields at two points separated by a period Λ along the z direction is equal to $\exp(iK\Lambda)$, thus

$$\begin{pmatrix} A_{m+1} \\ B_{m+1} \end{pmatrix} = e^{iK\Lambda}\begin{pmatrix} A_m \\ B_m \end{pmatrix} \tag{5}$$

The Bloch wave vector parameter K can be obtained by solving the eigenvalue equation

$$\mathbf{M}\begin{pmatrix} A_{m+1} \\ B_{m+1} \end{pmatrix} = e^{-iK\Lambda}\begin{pmatrix} A_{m+1} \\ B_{m+1} \end{pmatrix} \tag{6}$$

where $\mathbf{M}=(\mathbf{P}_a\mathbf{D}_a^{-1}\mathbf{D}_b)(\mathbf{P}_b\mathbf{D}_b^{-1}\mathbf{D}_a)$. In general, K depends on the frequency ω and the parallel wave vector component k_x for given geometry and refractive indices of constituent dielectrics. Once K is determined, the electric field in the PC can be expressed in the Bloch wave form

$$E_y(x,z)=u(z)\exp(iKz)\exp(ik_x x) \qquad (7)$$

In the above equation

$$u(z)=[A_0 e^{ik_{z,a}(z-m\Lambda)}+B_0 e^{-ik_{z,a}(z-m\Lambda)}]e^{-iK(z-m\Lambda)} \qquad (8a)$$

for $m\Lambda \le z \le (m\Lambda + d_a)$, where A_0 and B_0 are amplitudes of the first layer at $z=0$ and

$$u(z)=[C_0 e^{ik_{z,b}(z-m\Lambda-d_a)}+D_0 e^{-ik_{z,b}(z-m\Lambda-d_a)}]e^{-iK(z-m\Lambda)} \qquad (8b)$$

for $(m\Lambda + d_a) \le z \le (m+1)\Lambda$, where

$$\begin{pmatrix} C_0 \\ D_0 \end{pmatrix} = (\mathbf{P}_a \mathbf{D}_a^{-1} \mathbf{D}_b)^{-1} \begin{pmatrix} A_0 \\ B_0 \end{pmatrix}$$

Obviously, $u(z)$ is a periodic function of z.

The expressions for the magnetic field can be obtained from the electric field using Maxwell's equations. For p polarization, the magnetic field is parallel to the y axis. The same procedure can be used to determine the magnetic field first and then the electric field [19,25,26]. The amplitudes A_0 and B_0 depend on the boundary condition at $z=0$, that is, the interaction of the PC with the medium in the left half space.

For a given PC, the Bloch wave vector can be solved from the eigenvalue problem described by Eq. (6) for any real positive values of ω and k_x. In general, K is complex. When K is purely real, i.e., $\mathrm{Im}(K)=0$, the electric field oscillates in the z direction, and the Bloch wave propagates into the positive z direction, which is called an *extended mode*. When $\mathrm{Im}(K) \neq 0$, on the other hand, the amplitude of the Bloch wave decays exponentially into the positive z direction, and the wave is confined to the first few unit cells of the photonic crystal; this is called a *localized mode* [19,21]. Localized mode can also occur inside a PC in the stop band if defects are introduced. In analogue to real crystals, a defect is a local change of material properties or geometric parameters within the PC. Further studies about the field localization effects in the disordered periodic multilayers due to the random thickness variation can be found in Refs. [29,30]. Notice that $K = K(k_x, \omega)$, and the regions with $\mathrm{Im}(K)=0$ in the $\omega - k_x$ plane are called pass bands, and those with $\mathrm{Im}(K) \neq 0$ are called stop bands. Suppose light is incident from air (in the left half space) on the PC at $z=0$; in the stop band, the PC will act like a perfect mirror, which is also called a Bragg reflector, especially when the thickness of each dielectric is a quarter of the wavelength in that medium. A diagram in the $\omega - k_x$ domain showing the different regions allows one to study the band structure of a PC.

Figure 3 shows the calculated band structure of the 1D PC for both polarizations. Here, it is assumed that $d_a = d_b$, so that Λ is the only geometric parameter that affects the band structure. For convenience, the refractive indices for the two types of dielectrics are assumed to be constant, $n_a = 2.4$ and $n_b = 1.5$, in the wavelength region of interest. These values approximate those of ZnSe ($n \approx 2.4$) and KBr ($n \approx 1.52$) near $\lambda = 11$ μm [31]. The band structure is expressed by the reduced frequency and wave vector, so that Fig. 3 is applicable to arbitrary values of Λ. The shaded regions represent stop bands, and the unshaded regions are for pass bands. The light line in air is plotted as a dash-dot line. On the upper-left side of this line, propagating waves exist in air. Here, the parallel component of the wave vector is given by $k_x = (\omega/c)\sin\theta$, where θ is the angle of incidence. On the lower-right side of this line, evanescent waves exist in air. Alternatively, one can think that the light line corresponds to an incidence angle (or emission angle) of 90 deg. Note that the width of the first two stop bands shrinks to zero only for p polarization. The crossover point corresponds to the Brewster angle between constituent dielectric types a and b. At the Brewster angle, the reflectivity at the interface between the two dielectrics is zero; thus, waves can always propagate into the PC.

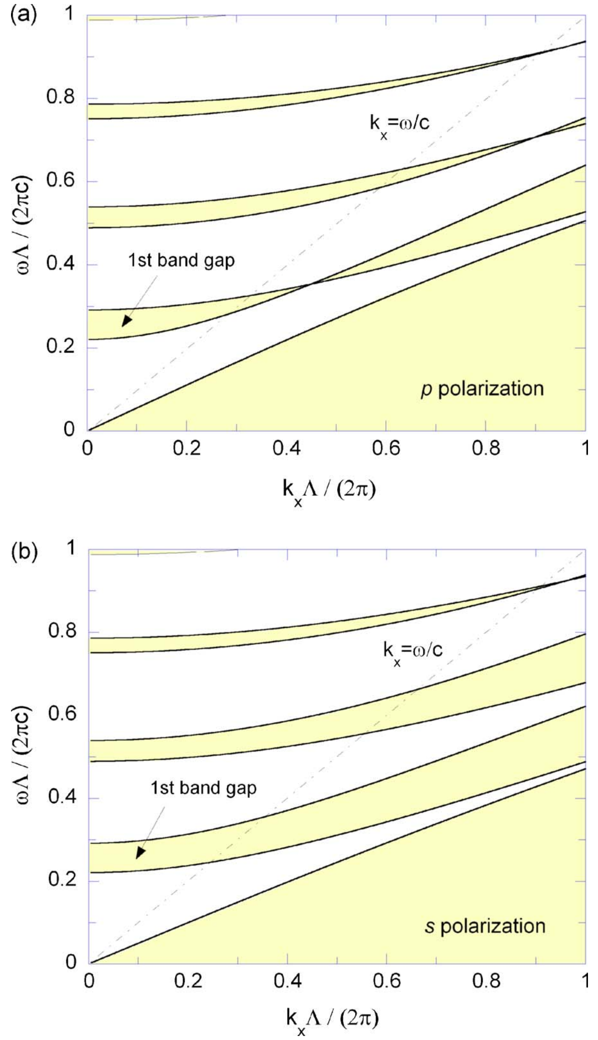


Fig. 3 Band structure of the 1D PC shown in Fig. 2, with $d_a = d_b$ and n_a=2.4 and n_b=1.5 in the reduced $\omega - k_x$ domain: (*a*) *p* polarization; and (*b*) *s* polarization. The dash-dot line denotes the light line in air. The crossovers for *p* polarization are due to the Brewster angle between media *a* and *b*.

2.2 Dispersion Relation of Surface Waves. A surface wave is an electromagnetic wave, which propagates along the interface and decays exponentially into both media. In general, the dispersion relation of surface waves at the interface between two semi-infinite homogeneous media is given by [13,32,33]

$$\frac{k_{z,1}}{\varepsilon_1} + \frac{k_{z,2}}{\varepsilon_2} = 0, \quad \text{for } p \text{ polarization} \qquad (9a)$$

$$\frac{k_{z,1}}{\mu_1} + \frac{k_{z,2}}{\mu_2} = 0, \quad \text{for } s \text{ polarization} \qquad (9b)$$

where ε_j and μ_j are the relative permittivity and permeability of medium j ($j=1$ or 2), respectively, and $k_{z,j} = i\sqrt{k_x^2 - \varepsilon_j \mu_j \omega^2/c^2}$ is purely imaginary when loss is not considered, suggesting evanescent waves exist in both media. Equation (9) can be obtained by matching the boundary conditions for the electric and magnetic fields. For two semi-infinite media, the Fresnel reflection coefficients are $r_p = (k_{z,1}/\varepsilon_1 - k_{z,2}/\varepsilon_2)/(k_{z,1}/\varepsilon_1 + k_{z,2}/\varepsilon_2)$ for p polarization and $r_s=(k_{z,1}/\mu_1-k_{z,2}/\mu_2)/(k_{z,1}/\mu_1+k_{z,2}/\mu_2)$ for s polarization [26]. Therefore, the conditions given in Eq. (9) correspond to the case when the denominator of the Fresnel coefficients becomes zero [32]. The resonance characteristics of the surface wave excitation can be understood from the point of view that the reflection coefficients go to infinity.

Evanescent waves can exist in a nonabsorbing medium when $k_x>\sqrt{\varepsilon_j\mu_j}(\omega/c)$, which can occur for light incident from a medium of larger refractive index to another with a smaller refractive index at an angle of incidence greater than the critical angle; this phenomenon is called total internal reflection. It is important to notice that, for surface wave to occur, the signs of ε_1 and ε_2 must be opposite in the case of p polarization to satisfy Eq. (9*a*). For most metals and some polar materials, the real part of ϵ becomes negative in certain frequency regions; however, materials with a negative μ rarely exist at optical frequencies to satisfy the dispersion relation for s polarization given in Eq. (9*b*). This is the reason why grating structures made of a metal or a polar material support surface waves only for p polarization. Metamaterials with simultaneously negative ε and μ have recently been demonstrated in the microwave and far-infrared regions, and hold promise for use to excite surface waves for both polarizations [13,33].

In order to excite a surface wave at the interface between SiC and 1D PC, evanescent waves are required in both SiC and the PC. In SiC, an evanescent wave can exist regardless of the angle of incidence when the dielectric function (ε_s) becomes negative because the normal component of the wave vector $k_{z,s}=(\omega^2\varepsilon_s/c^2-k_x^2)^{1/2}$ would be purely imaginary. In reality, the dielectric function of SiC is a complex quantity due to loss or damping in SiC; in which case, an evanescent wave can exist when the real part of ε_s is negative and the imaginary part is much smaller than unity. This is the case in the strong phonon absorption band of SiC for wavelengths between 10.32 μm and 12.61 μm. In the infrared region, the optical constants (i.e., refractive index and extinction coefficient) of SiC are calculated from the functional expression of ε_s given by [34]

$$\varepsilon_s=(n_s+i\kappa_s)^2=\varepsilon_\infty\left(1+\frac{\nu_{\mathrm{LO}}^2-\nu_{\mathrm{TO}}^2}{\nu_{\mathrm{TO}}^2-\nu^2-i\Gamma\nu}\right) \qquad (10)$$

where $\nu=1/\lambda$ is the wave number in cm^{-1}; $\varepsilon_\infty=6.7$ is a high-frequency constant; $\nu_{\mathrm{LO}}=969$ cm^{-1} is the longitudinal optical (LO) phonon frequency; $\nu_{\mathrm{TO}}=793$ cm^{-1} is the transverse optical (TO) phonon frequency; and $\Gamma=4.76$ cm^{-1} is the damping coefficient (or scattering rate). In the present study, λ is defined as the wavelength in air. From Eq. (10), it can be easily seen that the real part of ε_s is negative for $\nu_{\mathrm{TO}}<\nu<\nu_{\mathrm{LO}}$. In Fig. 4, the optical constants of SiC are plotted against the wavelength in the vicinity of the phonon absorption band, and the inset of Fig. 4 will be used in the later discussion. It is clear that κ_s much larger than n_s within the phonon absorption band.

A photonic crystal is a heterogeneous structure, and for the PC discussed previously, $\varepsilon_a=n_a^2$, $\varepsilon_b=n_b^2$, and $\mu_a=\mu_b=1$ (nonmagnetic). Hence, it is inappropriate to define equivalent ϵ and μ of the PC separately by regarding it as a homogeneous medium. However, surface waves can be excited at the stop band of the PC because there exists in the PC an *effective* evanescent wave, which is an oscillating field whose amplitude gradually decays to zero as z approaches infinity [24]. It can be shown that the effective evanescent wave does not carry energy into a semi-infinite PC. Since the wavelength range corresponding to stop bands of the PC can be scaled by changing the thickness of the unit cell, Λ is chosen to be 3 μm in order to approximately match the wavelengths corresponding to the first band gap of the 1D PC (shown in Fig. 3) with the phonon absorption band of SiC. Surface waves can be excited at the SiC–PC interface within the SiC phonon absorption band for both p and s polarizations. This is the key to constructing coherent thermal emission sources with directional and spectral selectivity for both polarizations without using magnetic materials. By using the equivalent layer method [23] or the supercell method [35], it is possible to obtain the dispersion relation of surface waves between a PC and another medium similar to Eqs. (9*a*) and (9*b*).

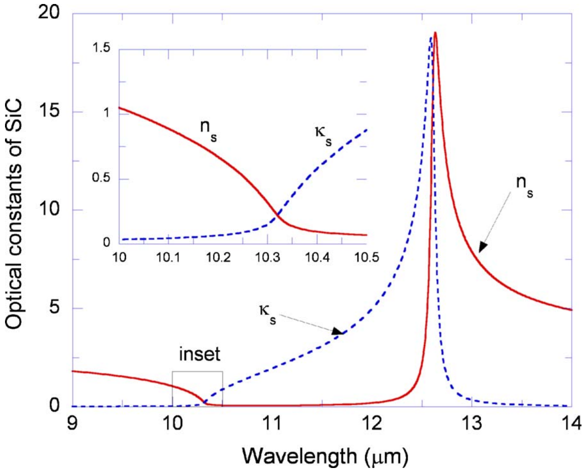


Fig. 4 The optical constants of SiC predicted by the phonon oscillator model given in Eq. (10) at wavelengths from 9 to 14 μm. The inset is for wavelength from 10 to 10.5 μm.

3 Results and Discussion

Figure 5(*a*) shows the directional–hemispherical reflectance R of the 1D PC coated with a SiC film (solid line) at normal incidence. The parameters for the 1D PC are the same as those in Fig. 3, except that the surface termination occurs in the type-a dielectric at $a_1=d_a/2$. The effect of thickness a_1 will be discussed later. Calculations show that 30 periods of unit cells are sufficient to be approximated as semi-infinite in the stop band. The thickness of SiC is set to be 1.45 μm. The reflectance of 30-period PC without SiC is plotted as a dashed line for comparison. The reflectance of the PC without SiC is unity in the shaded region that corresponds to the first stop band of the 1D PC with $\Lambda=3$ μm. Several dips appear in the reflectance spectrum with SiC within the stop band. The first dip around $\lambda=11.5$ μm is located in the phonon absorption band of SiC and is caused by the excitation of surface waves. Two reflectance dips also appear in the stop band near $\lambda=13$ μm, where propagating waves exist in the SiC film. These reflectance dips are caused by cavity resonance, when a standing wave exists inside the SiC layer. Note that interference effects cause some oscillations at wavelengths shorter than 10.2 μm and longer than 13.6 μm. The period of oscillation varies with the number of unit cells and the reduction in the reflectance corresponds to a large transmittance through the photonic crystal if it is not infinitely extended. Since the present study is focused on the coherent emission from the SiC layer, only features within the stop band will be further discussed.

The reflectance also depends on the angle of incidence. A regime map in the $\lambda-\theta$ plane can be constructed and is shown in Fig. 5(*b*) to identify regions where surface wave or cavity resonance mode can be excited. Region I represents the domain where the PC's stop band overlaps the SiC phonon absorption band. In this region, surface waves can be excited at the interface between SiC and 1D PC. Cavity resonance can occur in Region II, which is outside the SiC absorption band, for the SiC–PC structure. In Region III, another phenomenon, the Brewster mode may introduce coherent emission but only for p polarization at sufficiently large incidence angles.

The spectral–directional emissivity is illustrated by the contour plot in Fig. 6 for both polarizations as a function of the wave-

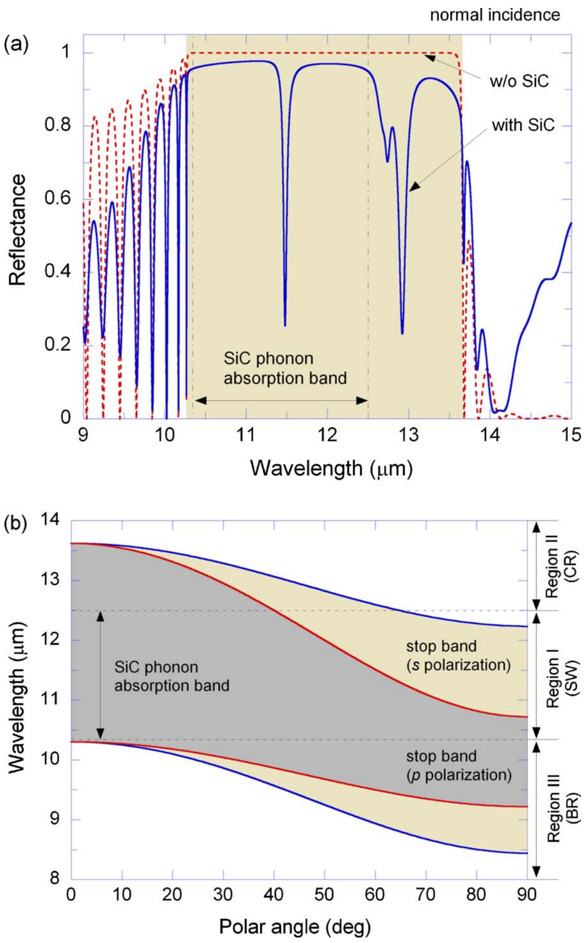


Fig. 5 Identification of the important regimes where the radiative properties are dominated by different mechanisms: (*a*) the reflectance of the 1D PC with a 1.45-μm-thick SiC layer, where the dotted line is for the 1D PC without SiC; and (*b*) the regime map in $\lambda-\theta$ space. CR stands for the cavity resonance mode, SW for surface wave, and BR for the Brewster mode. The shaded areas are corresponding to the first band gap of the 1D PC for *s* polarization (lighter) and *p* polarization (darker). The phonon absorption band of SiC is between the two dashed horizontal lines.

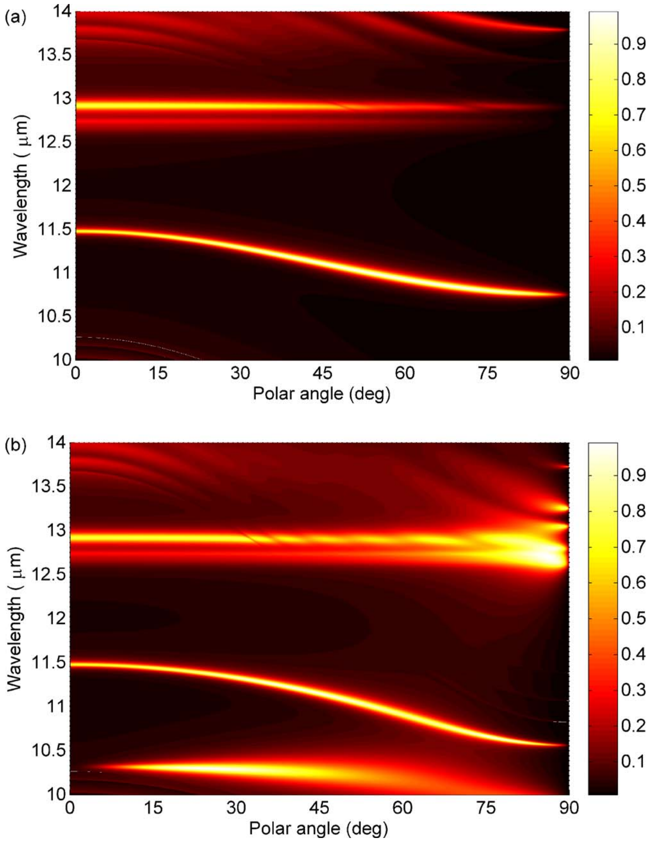


Fig. 6 Contour plot of the spectral–directional emissivity of the SiC–PC structure: (*a*) *s* polarization; and (*b*) *p* polarization. The thickness of SiC is set to be 1.45 μm.

length and emission angle. From Kirchhoff's law [36], the spectral–directional emissivity $\varepsilon_{\lambda,\theta}$ can be obtained by $\varepsilon_{\lambda,\theta}=1-R-T$, where T is the directional–hemispherical transmittance. The large emissivity values can be seen in a certain range of wavelengths and emission angles. As discussed previously, there exist three different mechanisms for the SiC–PC structure. The surface wave excitation and cavity resonance [15,18,20] can happen for both *s* and *p* polarizations even at normal incidence or emission, while the Brewster mode [37] can occur only for *p* polarization and at oblique incidence. For the bands that are relatively flat (e.g., upper branches near 13 μm for both figures), the emissivity is almost independent of the angle. Therefore, relatively diffuse emission can be achieved from the flat bands. In Figs. 6(*a*) and 6(*b*), the top band around 13 μm is due to cavity resonance. For the lower band in Fig. 6(*a*) and middle band in Fig. 6(*b*), on the other hand, coherent emission is due to the excitation of surface waves and is confined in a narrow angular range at a given resonance wavelength. Finally, the bottom band in Fig. 6(*b*) is due to the Brewster mode. For the SiC–PC structure, the Brewster mode also results in relatively flat band like the cavity resonance mode. In the following, each mode contributing to coherent thermal emission from the SiC–PC structure is separately considered.

3.1 Region I: Excitation of Surface Waves. Figure 7(*a*) shows the spectral–directional emissivity spectra in the wavelengths between 10.5 μm and 12 μm at $\theta=0$ deg, 30 deg, and 60 deg for *s* polarization. Three cases are considered with different thicknesses of the dielectric layer adjacent to SiC, such that a_1 is equal to $0.6d_a$, $0.5d_a$, and $0.4d_a$. When $a_1=0.5d_a$, for instance, the geometric parameters of the 1D PC are identical to those in Fig. 5. Here, the thickness of SiC is set to be $d_s=1.45$ μm, which results in a near-unity emissivity at $\theta=60$ deg if $a_1=0.5d_a$. Notice that since the emission peak values depend on the thickness of SiC, d_s can be tuned to maximize the emissivity for any given emission angle and polarization. The temporal coherence of thermal emission is evident from the sharp spectral peak in the emissivity. It is shown in the previous study [24] that the quality factor $Q=\lambda_c/\Delta\lambda$, where $\Delta\lambda$ is the full width at half maximum, from the SiC–PC structure is comparable to that for gratings [12] and single negative materials [14].

In general, the resonance wavelength of the surface wave at a given incidence angle can change for different values of a_1 [35,38], i.e., the dispersion relation of the surface wave largely depends on a_1. It can be seen from Fig. 7(*a*) that in addition to the resonance wavelengths, the emissivity peak values vary with a_1 as well. For example, the emissivity peak value due to surface waves decreases as a_1 changes from $0.5d_a$ to $0.6d_a$. On the other hand, when $a_1=0.4d_a$ the emissivity values for $\theta=0$ deg and 30 deg increase, but emissivity for 60 deg slightly decreases. Similarly, the peak emissivity value changes with the thickness a_1 for *p* polarization. Consequently, the emissivity from the proposed

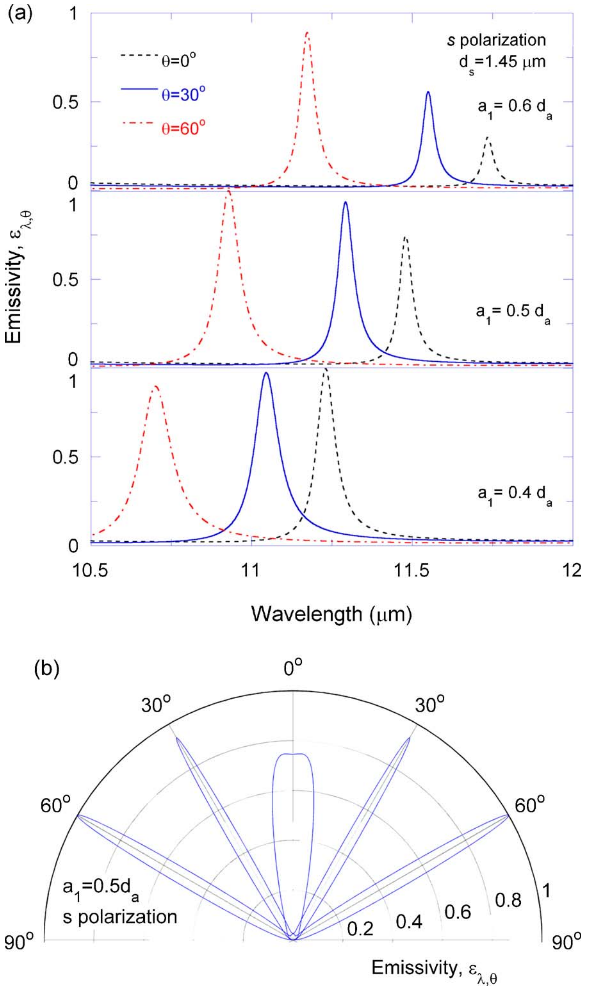


Fig. 7 The spectral–directional emissivity of the SiC–PC structure when surface wave is excited. (*a*) Spectral dependence at θ=0 deg, 30 deg, and 60 deg for *s* polarization with different locations of the surface termination. The upper panel illustrates the emissivity when surface termination occurs at $a_1=0.6d_a$. The middle and lower panels show the cases when $a_1=0.5d_a$ and $a_1=0.4d_a$, respectively. (*b*) Angular distributions at λ_c=11.479 μm, 11.293 μm, and 10.929 μm for *s* polarization when $a_1=0.5d_a$.

structure can be tuned by changing the value of a_1. This is an advantage of the SiC–PC structure over the conventional ATR configuration (consisting of a prism, a metal layer, and air) in which surfaces waves can also be excited. In the ATR configuration, the only parameter that can affect the emissivity peak values at the resonance condition is the thickness of the metallic layer next to the prism. However, in the SiC–PC structure, the thickness of the SiC layer as well as the surface termination determined by the thickness a_1 can affect the emissivity values from the structure, which allows more freedom to tune the emissivity.

The spatial coherence of the proposed emission source can be seen by the angular distributions of the emissivity, shown in Fig. 7(*b*) at the three peak wavelengths for *s* polarization. The polar plot of the emissivity corresponds the case when $a_1=0.5d_a$. It is important to note that the emissivity is intentionally plotted as a polar plot to clearly show the narrow angular lobe in the well-defined direction. However, if one considers the actual emission source with finite dimensions, due to the axial symmetry of the planar structure, coherent emission from the SiC–PC structure exhibits circular patterns [24], in contrast to the antenna shape for grating surfaces [11]. The patterns are very similar for *p* polarization at the corresponding peak wavelengths. The emissivity at each λ_c is confined in a very narrow angular region, although the angular spread corresponding to the peak at θ=0 deg is larger than the other two peaks.

The total thickness of the SiC–PC structure is less than 100 μm, which can be deposited using vacuum deposition techniques on a substrate such as silicon. In reality, however, the thickness of the deposited films may vary somewhat depending on deposition conditions. The effects of thickness variation on the emission characteristic can be examined by assuming a random Gaussian variation within ±150 nm (i.e., up to 10% of the original layer thickness). Note that many thin-film deposition techniques can control the layer thickness within few tens of nanometers. It turns out that very sharp spectral emission peak can still be achieved at slightly shifted wavelengths from the SiC–PC structure. Furthermore, the effects of loss in the dielectric layer are considered using a nonzero extinction coefficient κ. The peak value of emissivity reduces by less than 5% when $\kappa < 0.001$, as compared with the lossless case (κ=0). Hence, the effects of the thickness variation and loss in the constituent dielectrics of the 1D PC can be neglected under controlled deposition and for selected dielectric materials.

3.2 Region II: Cavity Resonance Mode. As seen from Fig. 6, the large emissivity due to cavity resonance can be found in the wavelength region between 12.5 μm and 13.5 μm for both polarizations. This region is outside of the SiC phonon absorption band and inside the stop band of the 1D PC. In general, the resonance like that in the Fabry–Pérot étalon can be found in the PC that has a defect layer inside it [15,18], and the resonance mode is called a defect or cavity mode [19]. For example, the 1D PC that has a single layer replaced by a different material can support a cavity mode. In this case, the photonic crystal containing a defect layer can be divided into three parts with respect to the defect layer. In other words, the structure is just like a defect layer sandwiched between two identical PCs, which act as a perfect mirror in the stop band. Consequently, the electromagnetic wave is confined inside the defect layer at the stop band due to high reflection from boundaries, resulting in a large absorption at the resonance condition if the defect layer is an absorbing medium. Recently, Celanovic et al. [20] used a PC on a dielectric cavity deposited on a metallic reflector to achieve coherent emission characteristics.

From the SiC–PC structure, a slightly different cavity resonance mode occurs compared to the above example because no defect layer is present inside the 1D PC. The cavity resonance mode associated with the proposed structure is closer to the simple resonator made of a semitransparent layer [39], which also demonstrates sharp spectral peaks and narrow angular lobes in its emissivity due to wave interference effects in the semitransparent slab. At the wavelength around 12.9 μm where a large emissivity peak appears, the refractive index of SiC is 8.95 and the extinction coefficient is 0.54 (see Fig. 4). The reflectivity at the interface between air and SiC is 0.64 for normal incidence. In addition, total reflection occurs from the 1D PC in the stop band. Therefore, the waves are confined in the SiC layer because of high reflection at both boundaries of the SiC slab. At the resonance condition when standing waves exist in the SiC layer, the reflectance of the SiC–PC structure reaches the minimum, resulting in a large emission because no energy can transmit through the PC in the stop band. Since phase shifts occur during the reflection from the boundaries of SiC due to the nonzero κ_s value of SiC and the total reflection from the 1D PC, $n_s d_s$ in the SiC layer at the resonance condition for θ=0 deg is not an exact multiple of the half wavelength as in the simple Fabry–Pérot étalon [27]. In essence, wave interference effects in the SiC layer are essential to enabling the cavity resonance mode in the SiC–PC structure, similar to the Fabry–Pérot resonator [15,18,20,39]. Thus, the cavity resonance mode may be excited by either polarization.

For the cavity resonance mode from the proposed structure, the

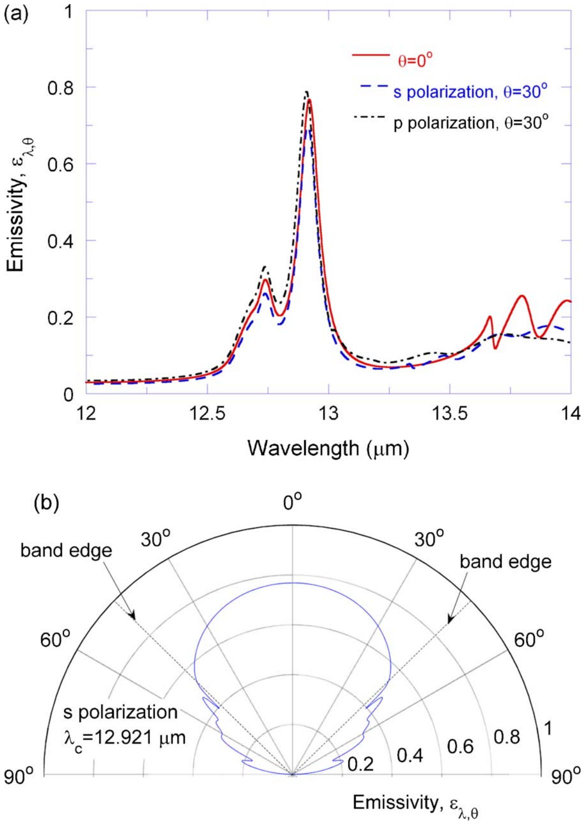


Fig. 8 The spectral–directional emissivity of the SiC–PC structure for the cavity resonance mode: (*a*) spectral dependence at θ=0 deg (solid line) and at θ=30 deg for *s* polarization (dashed line) and *p* polarization (dash-dot line); and (*b*) angular distributions at λ_c=12.921 μm for *s* polarization

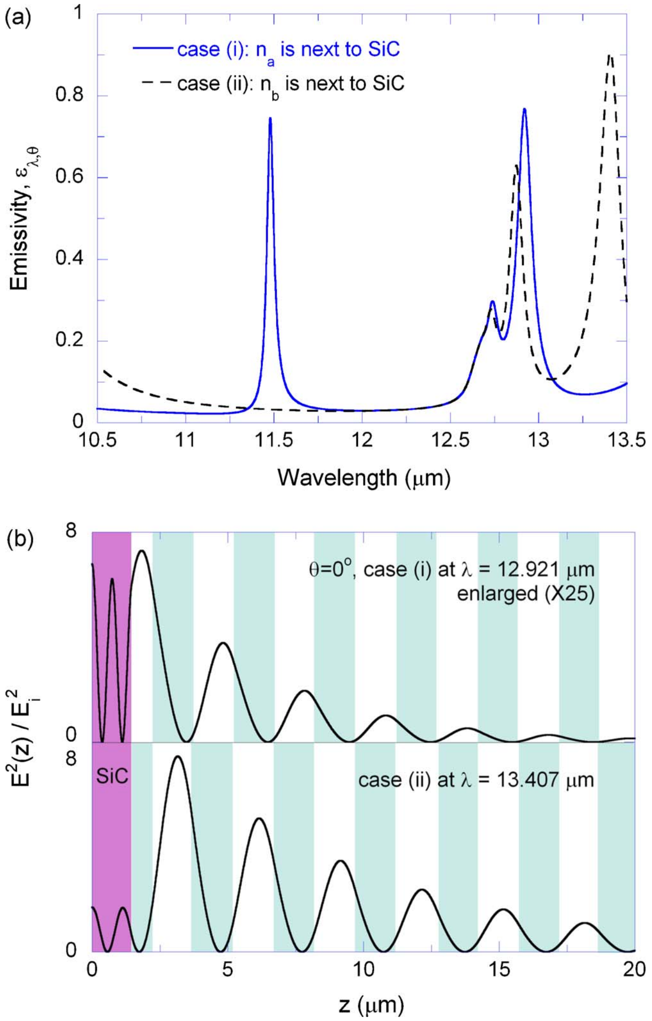


Fig. 9 Effects of the dielectric located next to the SiC film: (*a*) spectral–directional emissivity of the SiC–PC structure for two cases when (i) dielectric type *a* (solid line) or (ii) type *b* (dashed line) is next to the SiC layer; and (*b*) the square of the electric field, normalized by the incident, inside the SiC–PC structure for wavelengths corresponding to two cavity resonance modes. The upper panel of (*b*) is enlarged by 25 times.

emissivity values corresponding to emission angles of 0 deg and 30 deg are plotted for both polarizations in Fig. 8(*a*). The geometric parameters of the 1D PC are the same as in Fig. 5. In the cavity resonance mode, the wavelength position corresponding to the emissivity peaks does not change much with the emission angle. This is because the refractive index of SiC largely changes with the wavelength; hence, the effect of the polar angle change on the phase shift in the SiC layer is compensated for by the large change of the refractive index. As a result, the wavelength corresponding to the resonance condition changes little as the emission angle increases. Moreover, the calculated emissivity spectra reveal that the polarization states do not strongly affect the wavelength where the cavity resonance mode occurs. Similar to the surface wave, the emissivity peaks due to the cavity resonance mode are very sharp. The corresponding quality factors for the largest emissivity peak are 121 at θ=0 deg, and 127 or 121 at θ=30 deg for *s* or *p* polarizations, respectively. The angular distribution of the spectral–directional emissivity at λ=12.921 μm is plotted for s polarization in Fig. 8(*b*). Unlike in the case of surface waves where a narrow angular lobe is found in a certain direction, the diffuse characteristic of the emissivity is obtained from the cavity resonance mode. A similar trend of angular distribution is found for *p* polarization. Consequently, the SiC–PC structure can be used to produce a large emission either in a certain direction by the excitation of surface waves or in a wide range of the polar angles by the resonance mode, depending on the wavelength region of interest. Notice that the emissivity fluctuates outside the band gap of the PC because of interference effects in the PC.

Figure 9(*a*) shows the normal emissivity spectra for two cases when: (i) the dielectric type *a* (n_a=2.4) is next to the SiC film with $a_1 = d_a/2$ (identical to the structure in Fig. 5); and (ii) the dielectric type *b* (n_b=1.5) is next to the SiC by swapping type-*a* and type-*b* dielectrics in case (i). For case (i), the surface wave is excited near λ=11.5 μm, and two cavity resonance modes appear in 12.5 μm$<\lambda<$13.5 μm. On the other hand, no surface wave is excited for case (ii), and three cavity resonance modes are observed in 12.5 μm$<\lambda<$13.5 μm. This is because the excitation of surface waves strongly depends on the optical constants of the dielectric located next to SiC as well as other parameters, such as angle of incidence, order of the band gap, and polarization. It has been shown that surface waves can also be excited when the lower index dielectric is located at the surface of the PC for suitable parameters [35]. However, from the conditions given in the present study, excitation of the surface wave is not observable when the lower index dielectric is next to SiC. Unlike the surface wave excitation, cavity resonance modes appear in both cases, because the band structure of the 1D PC is the same for both cases.

When a surface wave is excited, the field decays exponentially into both the SiC layer and 1D PC from the interface as demonstrated in Ref. [24]. On the other hand, a standing wave exists in the SiC layer when the cavity resonance occurs, as illustrated in

Fig. 9(*b*). Here, $E(z)$ is the real part of the complex electric field vector. The upper panel is for case (i) at the resonance wavelength $\lambda_c = 12.921\ \mu\text{m}$; whereas the lower panel is for case (ii) at the resonance wavelength $\lambda_c = 13.407\ \mu\text{m}$. Notice that the field in the upper panel is enlarged 25 times for clear illustration. For instance, the actual value of the square of the electric field in the upper panel is $7.3/25 = 0.292$ at the maximum. While the emissivity peak for case (i) at $\lambda_c = 12.921\ \mu\text{m}$ is slightly smaller than that for case (ii) at $\lambda_c = 13.407\ \mu\text{m}$, the field strength in the upper panel is significantly lower than that in the lower panel. Because both peaks are within the stop band of the PC, the effective evanescent behavior of the electric field inside the PC is manifested by the oscillating amplitude that decays from left to right. It can be shown that the transmittance of the PC (with a total thickness of 90 μm for 30 periods) is nearly zero, indicating that no energy is transferred through the PC. Since SiC is the only material in the structure that can absorb the incident energy, it must be responsible for the thermal emission from the SiC–PC structure regardless of the modes.

3.3 Region III: The Brewster Mode. From Fig. 6(*b*), an emissivity enhancement is found for *p* polarization around $\lambda = 10.3\ \mu\text{m}$, which is distinguished from s polarization. In this case, the enhancement of emission is due to the Brewster mode. Figure 10(*a*) shows the emissivity spectra of the SiC–PC structure when the thickness of the SiC layer is 0.5 μm, 1.45 μm, and 2.5 μm at $\theta = 30$ deg for *p* polarization. For the 1D PC, the same parameters are used as in Fig. 5. The emissivity for *s* polarization when $d_s = 1.45\ \mu\text{m}$ is also plotted for comparison. It is clear that the emissivity enhancement due to the Brewster mode occurs only for *p* polarization, and the SiC layer thickness affects the emissivity peak values and their positions.

The conditions for the Brewster mode are more complicated than those for the cases of the surface wave excitation and cavity resonance. Most importantly, the Brewster mode occurs ideally when the reflection coefficient at the interface between SiC and 1D PC is zero [37]. When the reflection coefficient at the SiC–PC interface is zero, the reflectance of the SiC–PC structure is simply equal to the reflectivity at the air–SiC interface. At the same time, if the reflectivity at the air–SiC interface is small, then the reflectance of the SiC–PC structure becomes low. A reduction in the reflectance implies an increase in the emissivity since the transmittance is zero in the stop band. To further understand the Brewster mode, the case with $d_s = 1.45\ \mu\text{m}$ is considered in which the emissivity peak is located at $\lambda_c = 10.29\ \mu\text{m}$. At this wavelength, the refractive index of SiC is less than unity as seen from the inset of Fig. 4, and so that evanescent waves exist in the SiC layer for $\theta = 30$ deg. It is important to note that the reflection coefficient at the SiC–PC interface is not exactly equal to zero because of the nonzero κ_s value. Therefore, in the SiC layer, there exists a backward-decaying evanescent wave reflected from the SiC–PC interface. Consequently, the emissivity peak values and positions are strongly affected by the coupling effects between the forward and backward evanescent waves in the SiC. This is why the emissivity peak value and position depend strongly on the SiC layer thickness, as illustrated by Fig. 10(*a*).

The magnitude of the complex reflection coefficient between SiC and 1D PC is plotted in Fig. 10(*b*) for $\theta = 30$ deg and 45 deg. The complex reflection coefficient at the SiC–PC interface r_{12}, where the subscripts 1 and 2 represent SiC and 1D PC, respectively, can be calculated by using the equivalent layer method [23]. For illustration, $|r_{12}|$ for $\theta = 45$ deg is plotted on the left side in $10.15\ \mu\text{m} < \lambda < 10.23\ \mu\text{m}$, whereas $|r_{12}|$ for $\theta = 30$ deg is shown on the right side in $10.23\ \mu\text{m} < \lambda < 10.3\ \mu\text{m}$, separated by a vertical dash-dot line. The cases with $\Gamma = 0$ (no damping) in Eq. (10) are plotted as a solid line for comparison with the actual $|r_{12}|$ shown as a dashed line. When damping is neglected, the reflection coefficient at the SiC–PC interface goes down all the way to zero at the wavelength close to the emission peaks. On the other hand, a rather large $|r_{12}|$ exists if damping is considered, suggesting that the amplitude of the backward evanescent wave is indeed comparable to that of the forward evanescent wave. Consequently, the emissivity peak location depends on the thickness of the SiC layer

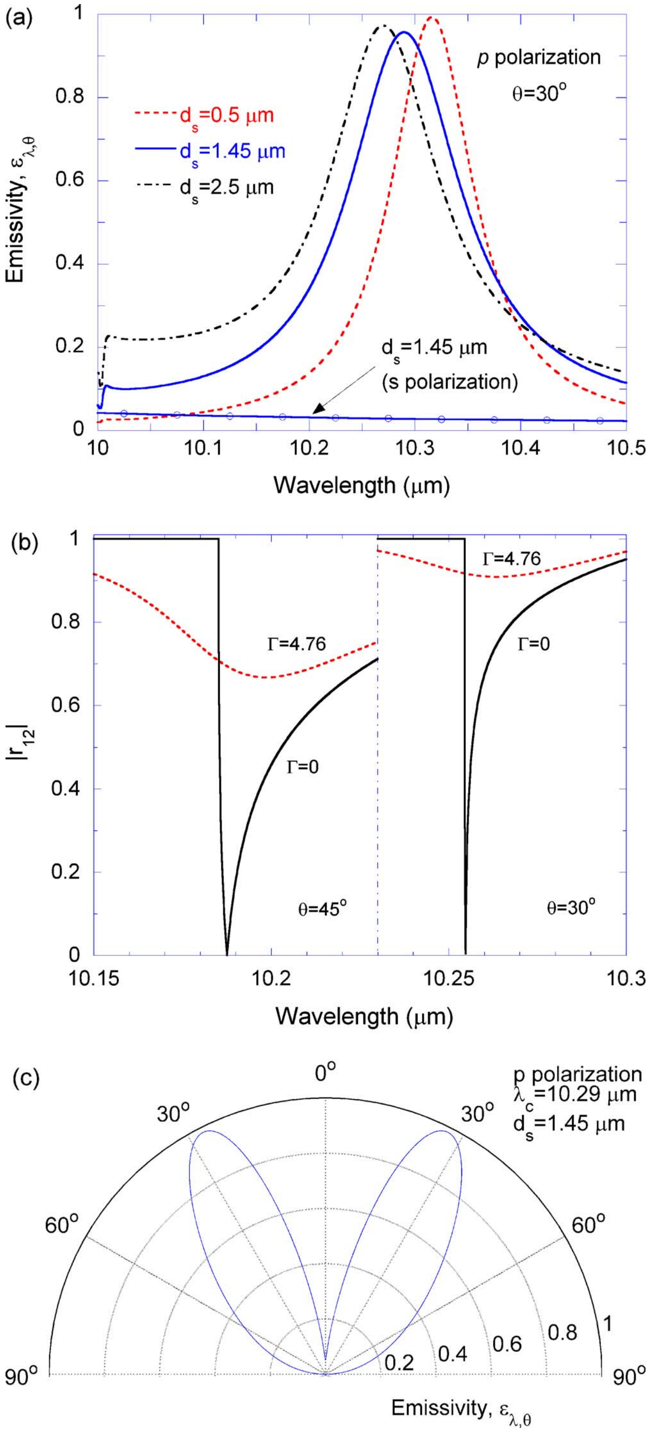


Fig. 10 The spectral–directional emissivity of the SiC–PC structure for the Brewster mode: (*a*) Spectral dependence when d_s is 0.5 μm, 1.45 μm, and 2.5 μm at $\theta = 30$ deg for *p* polarization. The solid line with circular marks is for the case when $d_s = 1.45\ \mu\text{m}$ for *s* polarization. (*b*) Magnitude of the complex reflection coefficient at the SiC–PC interface for *p* polarization with (dashed line) or without (solid line) damping. (*c*) Angular distributions at $\lambda_c = 10.29\ \mu\text{m}$ when $d_s = 1.45\ \mu\text{m}$ for *p* polarization.

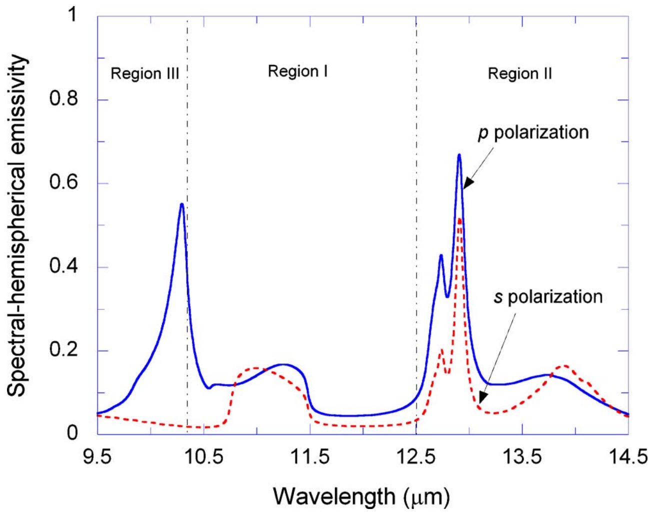


Fig. 11 The spectral–hemispherical emissivity of the SiC–PC structure with the same parameters as in Fig. 6

due to the coupling of evanescent waves.

The angular distribution of the emissivity at $\lambda_c = 10.29$ μm is plotted in Fig. 10(*c*) for the case of $d_s = 1.45$ μm. Emissivity values greater than 0.6 are found in the emission angles from approximately 15 deg to 45 deg. Compared to the surface wave excitation shown in Fig. 7(*b*), the angular lobe in the Brewster mode is much wider. On the other hand, the angular distribution of the emissivity from the Brewster mode is not as diffuse as that from the cavity resonance mode, which exhibits almost constant emissivity of 0.7 in the emission angles from 0 deg to 45 deg, as plotted in Fig. 8(*b*). As a result, the angular distribution of emission from the SiC–PC structure largely depends on which mode is excited. Again, the Brewster mode is distinguished from the surface wave excitation and cavity resonance, such that it is observable only for *p* polarization. Hence, the applicability of the emission source utilizing the Brewster mode is inevitably limited to *p* polarization, like grating structures.

The spectral–hemispherical emissivity of the SiC–PC structure is plotted in Fig. 11 for both polarizations. The spectral–hemispherical emissivity is obtained by integrating the spectral–directional emissivity $\epsilon_{\lambda,\theta}$ over the hemisphere [36]

$$\varepsilon_\lambda = 2\int_0^{\pi/2} \varepsilon_{\lambda,\theta}(\theta)\cos\theta\sin\theta\, d\theta \tag{11}$$

Due to the axial symmetry of the planar structure, the emission is independent of the azimuthal angle, unlike the case for 1D gratings. Notice that the electric field for *s* polarization is defined to be parallel to the unit direction vector of the azimuthal angle at any point on the surface. For *s* polarization, large values of the spectral–hemispherical emissivity are found in Regions I and II. The enhancement in Region I due to the excitation of surface waves is not as large as that in Region II at the cavity resonance mode. This is because cavity resonance enables a diffuse emitter, whereas surface waves induce a large emission into a narrow angular region. For *p* polarization, very similar trends are found in Regions I and II. However, a large enhancement of the hemispherical emissivity is also seen in Region III because of the Brewster mode, since the angular lobe of the spectral–directional emissivity from the Brewster mode is much wider than that due to surface waves.

4 Conclusions

Coherent thermal emission is demonstrated from a multilayer structure consisting of a one-dimensional photonic crystal coated with a SiC film. The calculated emissivity shows very sharp peaks in a narrow wavelength band and at a well-defined direction. A regime map is developed to identify different mechanisms or modes that are responsible for large emissivity values. By coating a SiC film on the 1D PC, surface waves can be directly coupled to propagating waves in air, resulting in coherent emission from the SiC–PC structure. This is the case when the wavelengths and emission angles are within the SiC phonon absorption band as well as the PC's stop band. When a surface wave is excited, it is shown that the emissivity from the proposed structure can be tuned by changing the thickness of the SiC layer as well as the thickness of the dielectric next to SiC. On the other hand, cavity resonance can occur from the SiC–PC structure when the wavelengths and incidence angles are in the stop band of the PC but outside of the SiC phonon absorption band. At the resonance condition, standing waves exist in the SiC layer, resulting in large emissivity values within a wide range of emission angles. Unlike the case with surface waves, diffuse emission within a narrow spectral band can be achieved from the cavity resonance mode. Furthermore, for *p* polarization, a region where the Brewster mode can be excited is also identified. The angular lobe of the emissivity from the Brewster mode exhibits a width between those of the surface wave excitation and cavity resonance mode. The proposed planar structure only involves dielectric films and can be fabricated with available vacuum deposition techniques. Further research will be performed to build the SiC–PC structure and measure the spectral–directional emissivity to experimentally demonstrate coherent emission characteristics.

Acknowledgment

This work was supported by the National Science Foundation (Grant No. CTS-0500113). The authors thank Keunhan Park for helpful discussions.

Nomenclature

c = speed of light in vacuum (2.998×10^8 m/s)
d = layer thickness (m)
$\mathbf{D}$ = dynamical matrix
$\boldsymbol{E}$ = electric field vector (V/m)
i = $\sqrt{-1}$
K = Bloch wave vector (1/m)
k = magnitude of wave vector (1/m)
k_x, k_y, or k_z = x, y, or z component of wave vector (1/m)
n = refractive index
$\mathbf{P}$ = propagation matrix
Q = quality factor
r = complex reflection coefficient
R = spectral–directional reflectance
T = spectral–directional transmittance

Greek Symbols

Γ = damping coefficient (1/m)
ε = relative electric permittivity
ε_0 = electric permittivity of vacuum (8.854×10^{-12} F/m)
ε_∞ = relative electric permittivity at high frequency
ε_λ = spectral–hemispherical emissivity
$\varepsilon_{\lambda,\theta}$ = spectral–directional emissivity
θ = polar angle (rad)
κ = extinction coefficient
Λ = period of one-dimensional photonic crystals (m)
λ = wavelength in air (m)
μ = relative magnetic permeability
μ_0 = magnetic permeability of vacuum ($4\pi\times10^{-7}$ N/A^2)
ν = wave number (1/m)

ν_{LO} = wave number of longitudinal optical phonon (1/m)

ν_{TO} = wave number of transverse optical phonon (1/m)

ω = angular frequency (rad/s)

References

[1] Perlmutter, M., and Howell, J. R., 1963, "A Strongly Directional Emitting and Absorbing Surface," J. Heat Transfer, **85**, pp. 282–283.

[2] Dimenna, R. A., and Buckius, R. O., 1994, "Electromagnetic Theory Predictions of the Directional Scattering from Triangular Surfaces," J. Heat Transfer, **116**, pp. 639–645.

[3] Chen, Y. B., Zhu, Q. Z., Wright, T. L., King, W. P., and Zhang, Z. M., 2004, "Bidirectional Reflection Measurements of Periodically Microstructured Silicon Surfaces," Int. J. Thermophys., **25**, pp. 1235–1252.

[4] Hesketh, P. J., Zemel, J. N., and Gebhart, B., 1986, "Organ Pipe Radiant Modes of Periodic Micromachined Silicon Surfaces," Nature (London), **324**, pp. 549–551.

[5] Hesketh, P. J., Gebhart, B., and Zemel, J. N., 1988, "Measurements of the Spectral and Directional Emission from Microgrooved Silicon Surfaces," J. Heat Transfer, **110**, pp. 680–686.

[6] Sai, H., Yugami, H., Kanamori, Y., and Hane, K., 2003, "Spectrally Selective Thermal Radiators and Absorbers with Periodic Microstructured Surface for High-Temperature Applications," Microscale Thermophys. Eng., **7**, pp. 101–115.

[7] Boueke, A., Kuhn, R., Fath, P., Willeke, G., and Bucher, E., 2001, "Latest Results on Semitransparent Power Silicon Solar Cells," Sol. Energy Mater. Sol. Cells, **65**, pp. 549–553.

[8] Coutts, T. J., 1999, "A Review of Progress in Thermophotovoltaic Generation of Electricity," Renewable Sustainable Energy Rev., **3**, pp. 77–184.

[9] Heinzel, A., Boerner, V., Gombert, A., Blasi, B., Wittwer, V., and Luther, J., 2000, "Radiation Filters and Emitters for the NIR Based on Periodically Structured Metal Surfaces," J. Mod. Opt., **47**, pp. 2399–2419.

[10] Kreiter, M., Oster, J., Sambles, R., Herminghaus, S., Mittler-Neher, S., and Knoll, W., 1999, "Thermally Induced Emission of Light from a Metallic Diffraction Grating, Mediated by Surface Plasmons," Opt. Commun., **168**, pp. 117–122.

[11] Greffet, J.-J., Carminati, R., Joulain, K., Mulet, J.-P., Mainguy, S., and Chen, Y., 2002, "Coherent Emission of Light by Thermal Sources," Nature (London), **416**, pp. 61–64.

[12] Enoch, S., Tayeb, G., Sabouroux, P., Guerin, N., and Vincent, P., 2002, "A Metamaterial for Directive Emission," Phys. Rev. Lett., **89**, p. 213902.

[13] Fu, C. J., Zhang, Z. M., and Tanner, D. B., 2005, "Planar Heterogeneous Structures for Coherent Emission of Radiation," Opt. Lett., **30**, pp. 1873–1875.

[14] Maruyama, S., Kashiwa, T., Yugami, H., and Esashi, M., 2001, "Thermal Radiation from Two-Dimensionally Confined Modes in Microcavities," Appl. Phys. Lett., **79**, pp. 1393–1395.

[15] Schubert, E. F., Hunt, N. E. J., Vredenberg, A. M., Harris, T. D., Poate, J. M., Jacobson, D. C., Wong, Y. H., and Zydzik, G. J., 1993, "Enhanced Photoluminescence by Resonant Absorption in Er-Doped $\mathrm{SiO_2/Si}$ Microcavities," Appl. Phys. Lett., **63**, pp. 2603–2605.

[16] Pralle, M. U., Moelders, N., McNeal, M. P., Puscasu, I., Greenwald, A. C., Daly, J. T., Johnson, E. A., George, T., Choi, D. S., El-Kady, I., and Biswas, R., 2002, "Photonic Crystal Enhanced Narrow-Band Infrared Emitters," Appl. Phys. Lett., **81**, pp. 4685–4687.

[17] Lin, S. Y., Moreno, J., and Fleming, J. G., 2003, "Three-Dimensional Photonic-Crystal Emitter for Thermal Photovoltaic Power Generation," Appl. Phys. Lett., **83**, pp. 380–382.

[18] Ben-Abdallah, P., and Ni, B., 2005, "Single-Defect Bragg Stacks for High-Power Narrow-Band Thermal Emission," J. Appl. Phys., **97**, p. 104910.

[19] Joannopoulos, J. D., Meade, R. D., and Winn, J. N., 1995, *Photonic Crystals*, Princeton University Press, Princeton, N.J.

[20] Celanovic, I., Perreault, D., and Kassakian, J., 2005, "Resonant-Cavity Enhanced Thermal Emission," Phys. Rev. B, **72**, p. 075127.

[21] Yeh, P., Yariv, A., and Hong, C. S., 1977, "Electromagnetic Propagation in Periodic Stratified Media. I. General Theory," J. Opt. Soc. Am., **67**, pp. 423–438.

[22] Robertson, W. M., and May, M. S., 1999, "Surface Electromagnetic Wave Excitation on One-Dimensional Photonic Band-Gap Arrays," Appl. Phys. Lett., **74**, pp. 1800–1802.

[23] Gaspar-Armenta, J. A., and Villa, F., 2003, "Photonic Surface-Wave Excitation: Photonic Crystal-Metal Interface," J. Opt. Soc. Am. B, **20**, pp. 2349–2354.

[24] Lee, B. J., Fu, C. J., and Zhang, Z. M., 2005, "Coherent Thermal Emission From One-Dimensional Photonic Crystals," Appl. Phys. Lett., **87**, p. 071904.

[25] Zhang, Z. M., Fu, C. J., and Zhu, Q. Z., 2003, "Optical and Thermal Radiative Properties of Semiconductors Related to Micro/Nanotechnology," Adv. Heat Transfer, **37**, pp. 179–296.

[26] Born, M., and Wolf, E., 1999, *Principles of Optics*, 7th ed., Cambridge University Press, Cambridge, UK.

[27] Yeh, P., 1988, *Optical Waves in Layered Media*, Wiley, New York.

[28] Ashcroft, N. W., and Mermin, N. D., 1976, *Solid State Physics*, Harcourt College Publishers, Fort Worth, TX.

[29] Hu, L., Schmidt, A., Narayanaswamy, A., and Chen, G., 2004, "Effects of Periodic Structures on the Coherence Properties of Blackbody Radiation," J. Heat Transfer, **126**, pp. 786–792.

[30] Ruan, X. L., and Kaviany, M., 2005, "Photon Localization and Electromagnetic Field Enhancement in Laser-Irradiated, Random Porous Media," Microscale Thermophys. Eng., **9**, pp. 63–84.

[31] Palik, E. D., 1998, *Handbook of Optical Constants of Solids II*, Academic, San Diego, CA.

[32] Raether, H., 1988, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer, Berlin, Germany.

[33] Fu, C. J., Zhang, Z. M., and Tanner, D. B., 2005, "Energy Transmission by Photon Tunneling in Multilayer Structures Including Negative Index Materials," J. Heat Transfer, **127**, pp. 1046–1052.

[34] Spitzer, W. G., Kleinman, D., and Walsh, D., 1959, "Infrared Properties of Hexagonal Silicon Carbide," Phys. Rev., **113**, pp. 127–132.

[35] Ramos-Mendieta, F., and Halevi, P., 1997, "Electromagnetic Surface Modes of a Dielectric Superlattice: The Supercell Method," J. Opt. Soc. Am. B, **14**, pp. 370–381.

[36] Siegel, R., and Howell, J. R., 2002, *Thermal Radiation Heat Transfer*, 4th ed., Taylor & Francis, New York.

[37] Burstein, E., Chen, W. P., Chen, Y. J., and Hartstein, A., 1974, "Surface Polaritons — Propagating Electromagnetic Modes at Interfaces," J. Vac. Sci. Technol., **11**, pp. 1004–1019.

[38] Robertson, W. M., 1999, "Experimental Measurement of the Effect of Termination on Surface Electromagnetic Waves in One-Dimensional Photonic Bandgap Arrays," J. Lightwave Technol., **17**, pp. 2013–2017.

[39] Kollyukh, O. G., Liptuga, A. I., Morozhenko, V., and Pipa, V. I., 2003, "Thermal Radiation of Plane-Parallel Semitransparent Layers," Opt. Commun., **225**, pp. 349–352.

Ramesh Chandrasekharan

Shaurya Prakash

Mark A. Shannon[1]
e-mail: mshannon@uiuc.edu

Department of Mechanical and Industrial Engineering,
University of Illinois at Urbana-Champaign,
1206 W. Green Street,
Urbana, IL 61801

R. I. Masel
Department of Chemical and Biomolecular Engineering,
University of Illinois at Urbana-Champaign,
600 South Mathews,
Urbana, IL 61801

Change in Radiative Optical Properties of Ta_2O_5 Thin Films due to High-Temperature Heat Treatment

*Thin films (*0.85 μm, 3 μm*) of* Ta_2O_5 *deposited on Si and* SiO_2 *were heated to* 900 °C. *Their reflectance in the infrared was measured using a Fourier transform infrared spectrometer equipped with a multiple angle reflectometer before and after exposure to the high-temperature heat treatment. An interfacial layer* ($TaSi_xO_y$) *formed by the diffusion of Si from the substrate into the deposited film was observed using Auger depth profiling, and the effect of this interfacial layer on the reflectance was measured. Using a least squares optimization technique coupled with an optical admittance algorithm, the multiple angle reflectance data were used to calculate the optical constants of the as deposited* Ta_2O_5 *film, crystalline* Ta_2O_5*, and the interfacial layer in the 1.6 to* 10 μm *range. The interfacial layer formed due to exposure to high temperature was found to be more absorptive than the crystalline* Ta_2O_5. [DOI: 10.1115/1.2401195]

Keywords: infrared reflectance, optical constants, interfacial layer, high temperature

Introduction

Several research groups have reported the construction and operation of microcombustion-based energy conversion devices [1,2]. The sustenance of combustion at small scales requires significant reduction of thermal and radical quenching, which are both strongly influenced by the temperature of the walls [3]. To reduce the heat loss via radiation, thin-film based reflective radiation shields can be employed. However, whenever homogenous combustion occurs in millimeter or submillimeter scales, the wall and flame temperatures tend to be very high (flame temperatures are approximately the stoichiometric adiabatic flame temperature and the wall temperatures are $\sim 1000°C$ [1]). High wall temperatures make vacuum packaged heat shields difficult to construct and thus any other thin-film based radiation shields must be able to endure oxidizing conditions. This rules out metallic radiation shields (like Ta that exhibits good high-temperature characteristics and high reflectivity in the infrared (IR), which is the relevant wavelength range for the dominant radiative heat transfer occurring). Hence certain dielectric materials like tantalum pentoxide that have good high-temperature structural properties [4], resistance to oxidation, and low optical absorption in the IR are better candidates for the construction of reflective thin film radiation shields. Frequently, to make the thin-film radiation shields as reflective as possible, multiple layers of dielectrics with alternating high and low refractive indices are used [5] (for e.g., Ta_2O_5, SiO_2). However, this study focuses on Ta_2O_5 deposited on Si and SiO_2.

The heat transfer analysis we desire to conduct needs knowledge of the radiative optical properties ($N=n-ik$) at the temperatures of operation. However, despite an extensive literature search, the radiative optical properties for Ta_2O_5 could not be found in the infrared at high temperatures. In fact, only the optical properties in the infrared for amorphous Ta_2O_5 at room temperature [6] were found. Given that the exposure of a-Ta_2O_5 to high temperatures converts it into its crystalline form [7], the optical properties of the crystalline form are required. Also, many authors have documented the diffusion of Si substrate into Ta_2O_5 thin films when exposed to temperatures above $500°C$ [7,8]. None of the many papers documenting such material changes report the optical properties of Ta_2O_5 or the changes in optical properties due to the material inhomogenities.

An added complication for the estimation of optical constants at high temperatures is the change in the material composition that occurs at elevated temperatures. Hence, the effects of temperature and composition changes on the radiative optical properties must be deconvolved. In this study, thin films of Ta_2O_5 on Si and SiO_2 were subjected to high temperatures ($900°C$). The changes in material properties and reflectance due to the high-temperature heat treatment are measured.

The paper first presents a brief description of the experimental techniques used to measure the changes in reflectance and also the changes in material properties. Then a generalized algorithm for calculating the optical properties from the reflectance is presented. The procedure to calculate the optical properties is developed for a stack of films with interfacial layers (formed by diffusion/reaction) between them. The material analysis, which confirmed the formation of a thick interfacial layer between the Ta_2O_5 and Si when heated to high temperatures, is discussed next. The reflectance measurements showing the increased absorption of films with interfacial layers between them and the substrate are shown next. Finally, the application of the admittance matrix based least square error (LSE) optimization to find the optical properties of the film and interfacial layer is discussed, showing that the interfacial layer is more optically absorbing than the film.

Experimental Techniques

Thin films of tantalum pentoxide were deposited by e-beam evaporation by using a Denton Infinity 22 vacuum deposition system (Denton Vaccum, NJ). The pressure for evaporation was maintained at 5×10^{-5} Torr or lower. The deposition was done

---

[1] Corresponding author.

Manuscript received October 14, 2005; final manuscript received May 3, 2006. Review conducted by Ranga Pitchumani. Paper presented at the 2005 ASME International Mechanical Engineering Congress (IMECE2005), Orlando, FL, November 5–11, 2005.

either on top of a bare Si (100) substrate (one side polished) or on top of a thermally grown SiO_2 layer (on 100 Si). The film thicknesses were measured using a Dektak-3 stylus profilometer.

Given the temperature of the flame and the wall, the wavelength range of the thermal radiation is from $\sim 2\ \mu m$ to $20\ \mu m$ (Wien's law [9]). The quarter wave optical thickness [10] for this wavelength range for Ta_2O_5 and SiO_2 in the infrared is approximately 3–4 μm. Hence, for investigating the effect of temperature on Ta_2O_5 films on Si, films of $\sim 3\ \mu m$ were deposited. The silica layers were grown using thermal oxidation and the Ta_2O_5 films were deposited on it using *e*-beam deposition. Given that very thick films of Ta_2O_5 could not be grown on silica without significant cracking (most probably due to thermal mismatch), the Ta_2O_5 layers on the silica were $\sim 1\ \mu m$ thick.

After the Ta_2O_5 films were deposited, all the films were heated to 900°C. The heating and cooling ramp rates were 200°C/h and the heating time at 900°C was 9 h. The ramp rates and the heating time were based on trial and error to prevent cracking of the films and to completely crystallize the films.

The material studies on the films were done using two different methods: x-ray diffraction (XRD) to detect the crystallinity of the films, and Auger electron depth profiling spectroscopy to detect the formation of reaction/diffusion interfacial layers between the thin films. The Auger spectroscopy (Perkin-Elmer PHI-660 scanning Auger microscope) provided information on the elemental composition of the film before and after exposure to high temperatures. Similarly, the XRD (Rigaku D-Max) was used to check the crystalline nature of the film after heat treatment. It must be pointed out that the Auger depth profiling gives the material composition as a function of sputtering time, i.e., the time for which the film was sputtered by the ion gun in the Auger setup before analysis of the Auger electrons. The sputtering time can be translated into actual depth values only if calibration charts are available for the material, in the form being used (crystalline versus amorphous). Since calibration data were not available for the materials studied here, the Auger plots have been presented as a function of sputtering time. However, the Auger data can be used to determine the ratio of the thicknesses of two different sections (depthwise) of the film as long as the two sections are similar in stoichiometry.

The reflectance in the IR was measured using a Nicolet IR-750 Fourier transform infrared (FTIR) with a multiple angle reflectometer accessory fitted with wire-grid polarizers (*Seagull* attachment, Harrick Scientific, Ossining, NY). The measurements were done at multiple angles (25–60 deg angle of incidence with steps 5 deg apart). All of these measurements were done using the mercury cadmium telluride (MCT) detector with a 2 cm^{-1} resolution and covered the mid-IR range of the MCT detector (1.6–16 μm). Given the reflective nature of the substrate, a polished gold mirror was used to collect the background spectra. The techniques and related issues for measuring the reflectance using a FTIR equipped with a *Seagull* reflectometer attachment are discussed in detail by Ford et al. [11], and Kawka [12].

Data Reduction: Extracting Optical Properties From Reflectance for Multilayer Films With Interfacial Layers

The mathematical equations governing the interaction of electromagnetic fields with thin films are well documented in standard textbooks like MacLeod [10] and Born and Wolf [13]. These equations, which are solutions of Maxwell's equations for the appropriate boundary conditions, allow the calculation of reflectance of a single thin-film or a multiple-film stack on a substrate. In this study, the optical admittance technique as described by MacLeod [10] is used to quantify the interaction of IR light with Ta_2O_5 and SiO_2 thin films. A brief summary of these equations is presented here. A relevant schematic is shown in Fig. 1.

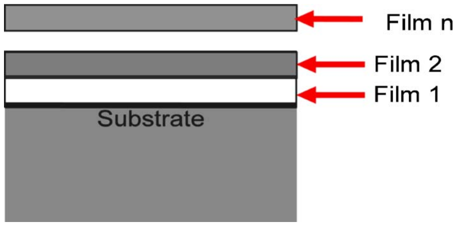


Fig. 1 Schematic of a multilayer thin film stack for Eqs. (1)–(5). Each thin film layer has its own admittance as given by Eq. (1) and the whole stack's reflectance is given by Eq. (5). The individual layers could be a homogenous film or be the interfacial layer between two different films.

The optical admittance, Y, of any layer (film or interfacial layer between films) is given by the matrix

$$Y_k = \begin{bmatrix} \cos \delta_k & i \sin \delta_k / \eta_k \\ i \sin \delta_k \eta_k & \cos \delta_k \end{bmatrix} \tag{1}$$

where

$$\delta_k = \frac{2\pi N_k d_k \cos \theta_k}{\lambda}, \quad N_k = n_k - ik_k \quad \text{and} \quad \eta_k = N_k \cos \theta_k \tag{2}$$

and the angles are measured from the surface normal.

It must be noted that the optical constant N_k, and hence the phase thickness δ_k and the admittances are complex quantities. Also, note that the definition of η_k given here is applicable only for *s*-polarization. All other equations and quantities are defined the same irrespective of polarization. The optical admittance of a stack of thin films is simply the product of the admittances of individual layers

$$Y = [\text{Thin-film } n]^{*} \ldots ^{*}[\text{Thin-film1}]^{*}[\text{Substrate}] \tag{3}$$

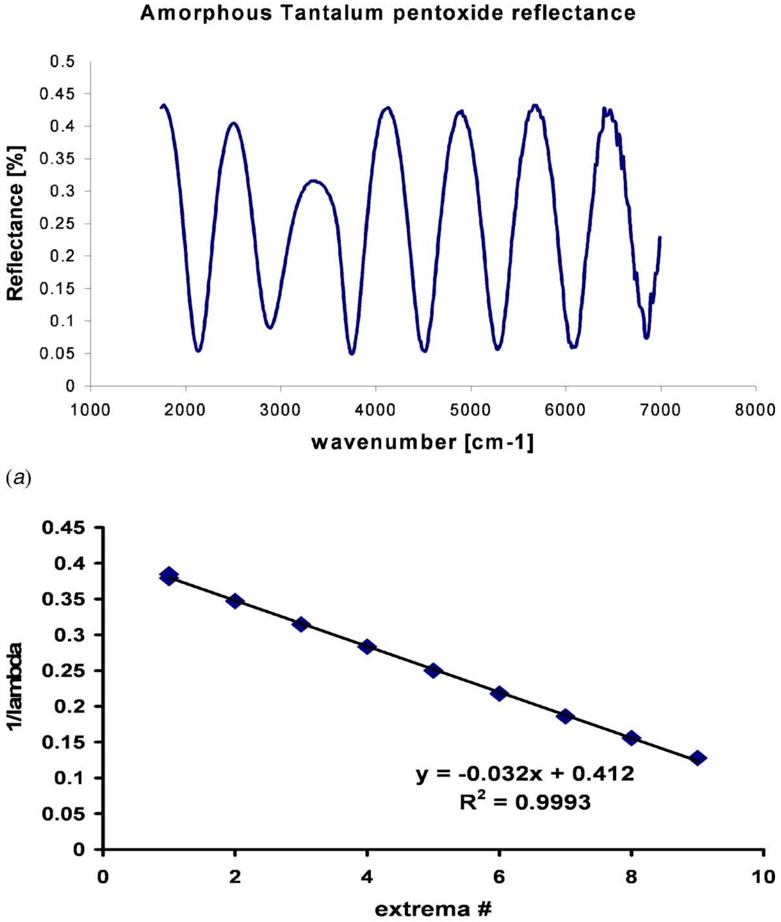
The reflectance is given by

$$R = \left(\frac{\eta_0 - Y}{\eta_0 + Y} \right) \left(\frac{\eta_0 - Y}{\eta_0 + Y} \right)^{*} \tag{4}$$

where η_0 is the admittance of air ($=N_0 \cos(\theta)$ for s polarization). The above equations are sufficient to calculate reflectance when the optical properties and the individual layer thicknesses are known for all relevant wavelengths. Such a model accounts for the multiple reflections within a thin film. However, these equations do not account for the roughness between two films or between film and substrate. In this study the roughness was assumed to be smaller than the wavelength of light, given that the focus is on infrared radiation (sufficiently long wavelength). The reflectance at a particular angle of incidence and at a particular wavelength is a function of the admittance for that angle and wavelength calculated using the optical properties and thickness of the films and substrate, which can be summarized as

$$\text{Reflectance}(\lambda, \theta) = f[Y(\lambda, \theta, d_{\text{films}}, N_{\text{films}}, N_{\text{sub}})] \tag{5}$$

Despite the fact that the above set of equations completely describe the relation between the reflectance and the optical properties and thickness, the extraction of refractive index from the reflectance is not trivial [14]. The complicating factors include the multiple solutions for a single value of reflectance and thickness. There are a variety of techniques cited in literature ranging from auxiliary data collected in transmission to use of Kramers–Kronig transform. Given the substrate in use, transmission studies are not feasible and Kramers–Kronig transforms tend to be approximate when the data are not available over the entire frequency range [15]. Also, given that the thicknesses of the films are known, that the films are not very absorbing (since it is a dielectric), and that the optical constants are approximately known (from the literature for amorphous films [6]), the use of a least square optimization technique combined with multiple reflection measurements was



(*a*)

(*b*)

Fig. 2 Example of reflectance curve (specular reflection, *s* polarization 45 deg). For clarity the plot shows reflection as a function of the wavenumber (instead of wavelength). The positions of the interference peaks and troughs from plot (*a*) are used to plot (*b*).

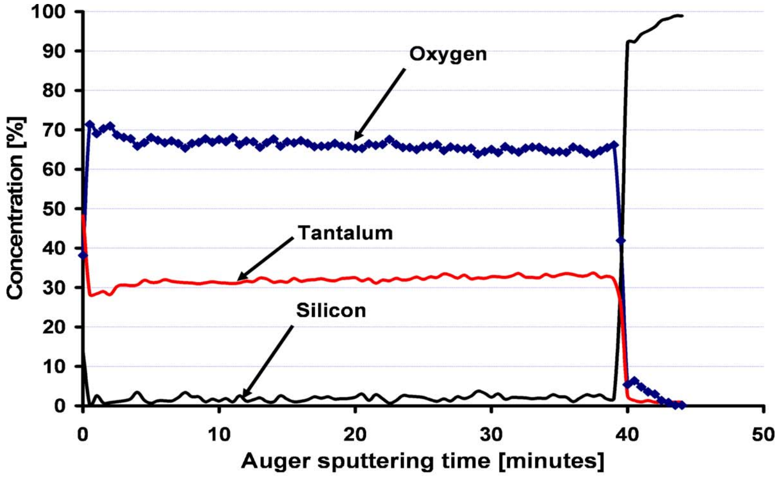
used in this study. The use of reflectance data collected at different angles of incidence for the same film provides additional information and increases the confidence in the solution obtained by a least squares technique. The equations used and the overall algorithm is presented in the equations below. Note that the procedure remains the same for any number of layers involved and hence the equations, applicability extends to all the cases considered in this paper; film on substrate, film on substrate with an interfacial layer, etc.

$$
\begin{aligned}
&\text{Duplicate } (\lambda = 1 \text{ to } \#\,\text{readings})\\
&\text{Reflectance } (\lambda,\theta_1) = f[Y(\lambda,\theta_1,d_{\text{film}},N_{\text{film}},N_{\text{sub}})]\\
&\text{Reflectance } (\lambda,\theta_2) = f[Y(\lambda,\theta_2,d_{\text{film}},N_{\text{film}},N_{\text{sub}})]\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\\
&\text{Reflectance } (\lambda,\theta_n) = f[Y(\lambda,\theta_n,d_{\text{film}},N_{\text{film}},N_{\text{sub}})]\\
&\text{LSE} = \sum_{\lambda} [\text{Ref}(\lambda,\theta_1) - \text{Ref}_{\text{measured}}(\lambda,\theta_1)]^2 + [\text{Ref}(\lambda,\theta_2)\,.\,.\,.\,.\,.\,]
\end{aligned}
\tag{6}
$$

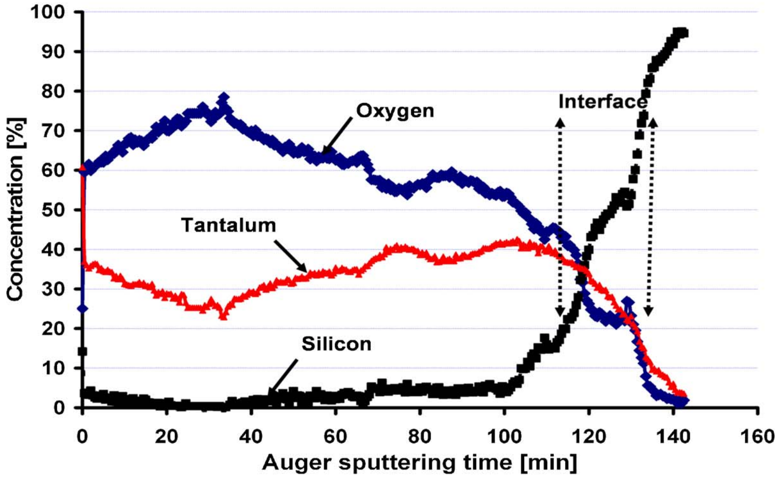
Since the above least square optimization is a multiple variable optimization, we use a quasi-Newton numerical method to solve them. In this paper, the optimization algorithm was implemented in Engineering Equation Solver (f-Chart solver, WI [16]). The variable metric technique [17] was used to optimize the least square error. This software platform was chosen given its ability to deal with complex numbers and its built-in optimization techniques.

Referring to Eq. (6), the number of variables involved increases $\sim n$ times by the introduction of n different angles. Such an increase in variables leads to computational resource issues. In this study, the multiple angle optimizations were done for two different angles at any one time and then repeated for a different set of angles. The final sets of results are averaged values over these different sets. Also, the introduction of films/interfacial layers whose optical properties and thickness are unknown introduces up to three unknown dependent variables to be optimized for each film/interfacial layer. Given that multiple solutions are possible, good initial guess values are needed to arrive at the physically correct solution [14]. Initial guess values for the thicknesses were obtained by combining the use of a stylus profilometer (Dektak-3) along with Auger depth profiles (used to estimate the ratio of interfacial layer to film thickness). The guess values for the optical properties are obtained using a plot of the peak positions as shown in Figs. 2(*a*) and 2(*b*).

The reflectance shows numerous peaks and troughs due to interference between the light reflected at the air–film and the film–substrate boundaries. The position of these peaks and the peak maxima and minima values provide important clues about the optical properties [14,18]. When a plot of the peak positions versus peak number is linear, the optical constant N can be assumed to be near constant (i.e., no anomalous dispersion etc.) in that



(*a*)



(*b*)

Fig. 3 (*a*) Auger depth profile of an amorphous Ta_2O_5 film on Si (as deposited). The Auger technique does not give exact concentration information and needs a standard sample for calibration. Here the calibration was done against a thermally grown Ta_2O_5 thin film sample whose composition was verified independently using XPS and XRD. Similarly, the Auger sputtering time is a rough indicator of the depth and needs a standard for calibration. Note the sharp transition from film to substrate and the homogeneous material composition with depth. (*b*) Auger depth profile of a Ta_2O_5 film on Si heated to 900°C. Note the formation of a $TaSi_xO_y$ interfacial layer between the Ta_2O_5 film and Si substrate. Also note the inhomogeneous nature of the film and the interfacial layer.

wavelength regime [19]. This linear result is seen in Fig. 2(*b*). Such a plot also provides guess values for the product of $n \times d$ (i.e., product of the real part of the refractive index and the physical thickness of the film) [19].

The departure of maxima values from the substrate reflectance value reveals the presence of inhomogeneous layers [18,20]. Similarly, the minima positions provide information on the value of the absorption coefficient [18]. This procedure will be clearer when applied to the reflectance of tantalum pentoxide heated to a high temperature in the next section.

3 Results and Discussion

The first sample tested was a 3 μm *e*-beam deposited sample of Ta_2O_5 on bare Si. Figure 3(*a*) shows the Auger depth profile of the as-deposited Ta_2O_5, while Fig. 3(*b*) shows the Auger depth profile of the sample after heating it to 900°C. Apart from the material inhomogenities arising out of diffusion and reactions, the overall film thickness was measured using a profilometer to be 2.81 μm (versus 3 μm for the unheated sample). This reduction in thickness could be due to the densification of the as-deposited film [21] via consolidation of pores. Further, from Fig. 3(*b*), the ratio of interfacial layer to film thickness was estimated, giving an initial guess value of 0.4 μm for the thickness of the interfacial layer. The reflectance measured using the FTIR is shown in Figs. 4(*a*) and 4(*b*) for the as-deposited film and the crystalline inhomogeneous Ta_2O_5 film, respectively.

The Auger plots show the inhomogenities caused by heating the sample while the reflectance plots show the effect of these inhomogenities on the reflectance in the IR (Fig. 4). The main inference from Fig. 4(*b*) is the increase in overall absorption upon heating the film.

Also, there are absorption peaks at higher wavelengths. These absorption peaks could correspond to crystalline Ta_2O_5 or the $TaSi_xO_y$ interfacial layer formed. The authors could not determine

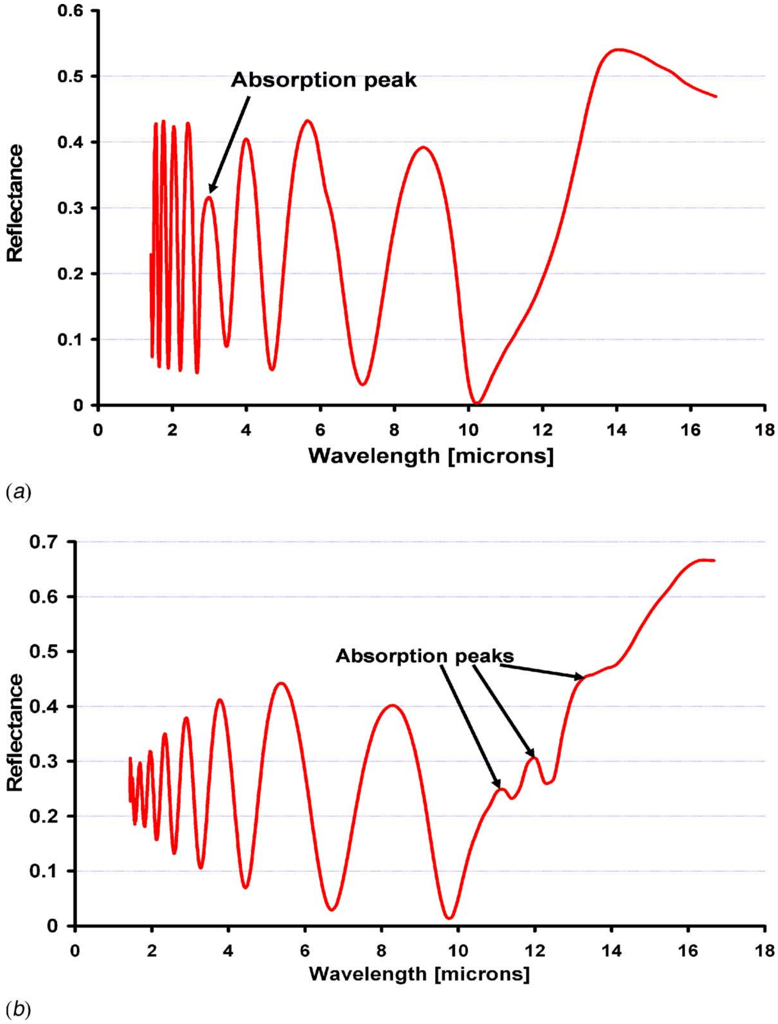


Fig. 4 **(*a*) Specular reflectance (*s* polarization, 45 deg) of the as-deposited Ta_2O_5 film on Si. Note the absorption peak at 2.82 μm (absorption peaks positions are confirmed from measurements at different angles). The reflectance maxima from 1.6 μm to 2.7 μm all equal reflectance of bare substrate. Also note the decreasing value of the minima from 4 μm onward (indicating changing *n* and *k*). Linear plots similar to Fig. 2 show an approximately constant value of n from 1.6 μm to 2.7 μm and then from 3 μm to 10 μm. (*b*) Specular reflectance (*s* polarization, 45 deg) of heated (900°C) Ta_2O_5 film on Si. Note the large change in reflectance values at lower wavelengths and the introduction of absorption peaks at higher wavelengths. The monotonic increase in the reflectance from 1.6 to ~7 μm indicates a constant value of *k*. Linear plots of peak positions similar to Fig. 2 show an approximately constant value of *n* from 1.6 μm to 10 μm wavelength.**

which of the two mechanisms is responsible from the FTIR peak positions alone, due to lack of the respective peak data in the literature. Also, an XRD analysis of the films before and after heating revealed that the as-deposited film was amorphous as expected and that the heated film had an orthogonal phase crystalline Ta_2O_5.

Similar measurements were done for Ta_2O_5 on a 0.1 μm thermal oxide layer grown on a Si substrate. The Auger plot for the crystalline form (verified using XRD) obtained after heating is shown in Fig. 5, while the reflectance is shown in Fig. 6. The significant change from the first sample, where the deposition was on bare Si, is the lower extent of the inhomogenities formed due to exposure to elevated temperature.

4 Summary of Optical Constants

The first as-deposited sample is amorphous Ta_2O_5, which when heated becomes a crystalline Ta_2O_5 film with a significantly thick interfacial layer between film and substrate. In order to find the optical properties ($N=n-ik$) for the film before exposure to high temperature, it is modeled as a single layer of film on Si, which is treated as a semi-infinite substrate of known optical properties [22]. The thickness of the film is taken as a known quantity, having been obtained from profilometer measurements. For the case of the inhomogeneous crystalline film with an interfacial layer obtained after exposure to high temperature, the film is discretized into two layers on top of Si. The thickness of the interfacial layer and the film are only approximately known (from the Auger depth profile, the ratio of thicknesses is approximately known and the total thickness is known from profilometer measurements). Figures 7 and 8 show the values of *n* and *k* and also the comparison between fitted value (based on *n* and *k* calculated here) and measured value for the reflectance in the IR. As mentioned above, the

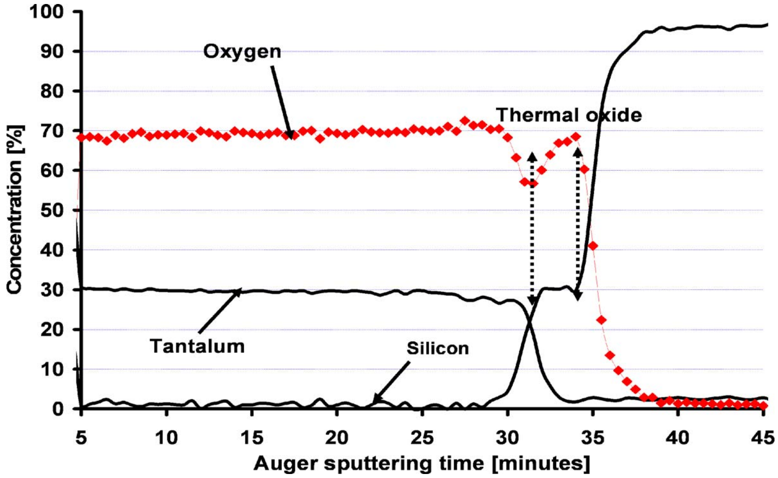
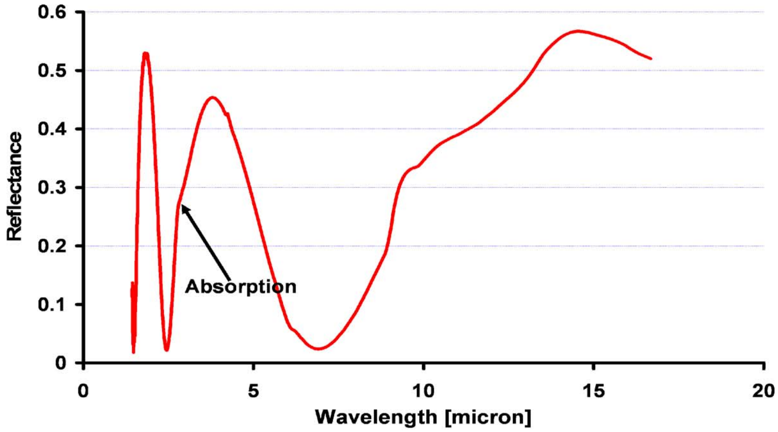
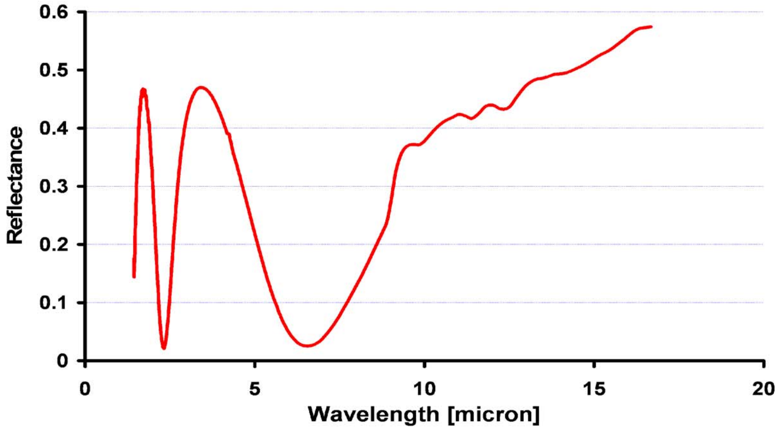


Fig. 5 Auger depth profile of a heated Ta_2O_5 film on SiO_2. Comparing with the heated Ta_2O_5 film on Si (Fig. 3(*b*)), note that the interfacial layer formed between the SiO_2 and the tantalum pentoxide is much thinner and also the film itself is more homogenous in material composition.

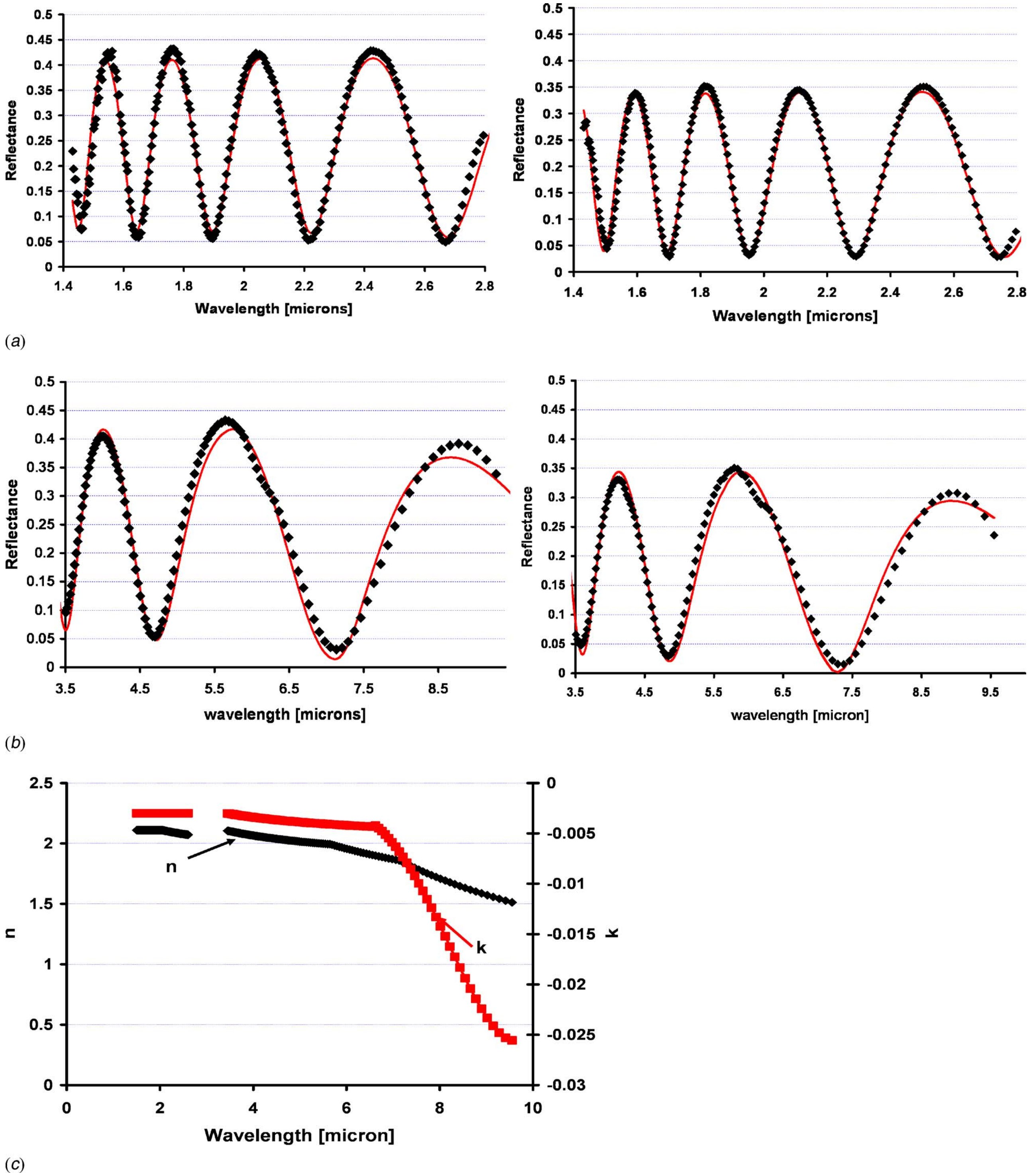


(*a*)



(*b*)

Fig. 6 (*a*) Specular reflectance (*s* polarization, 45 deg) of as-deposited Ta_2O_5 film on thermally grown SiO_2. Note the absorption peak at 2.82 μm (this is the same peak as seen in Fig. 4(*a*) for as-deposited Ta_2O_5 on Si). The values of the maxima do not correspond to the bare substrate given the thermal oxide layer present, giving rise to multilayer reflection enhancement. (*b*) Specular reflectance (*s* polarization, 45 deg) of heated (900°C) Ta_2O_5 film on thermal oxide. Note the change in reflectance values at lower wavelengths compared to the as-deposited Ta_2O_5.



(*a*)

(*b*)

(*c*)

Fig. 7 (*a*), (*b*) Comparison of the measured (dotted) and fitted (complete line) reflectance for the as-deposited Ta_2O_5, (3 μm thick, amorphous). The figures are shown for 30 deg (on the right) and 45 deg incident and reflected angles (*c*) *n* and *k* values as a function of wavelength for the as-deposited Ta_2O_5 film. Note that the *k* values are shown as negative to signify that the $N=n-ik$ notation was used in this study.

n and *k* values are obtained from experimental data at multiple angles. The values shown in the plots are average values and the variance (not shown on each plot) was always less than 4–5% for the *n* values and within 7% for the *k* values.

For the amorphous as-deposited film, the optical properties were found for the 1.6–10 μm wavelength range (Fig. 7(*c*)). In the portions of the reflectance spectrum, where the optical properties are not constant, which can be judged from the maxima and minima position values and their variation with wavelength [18], the nature of the variation in *n* and *k* was arrived at by trial and error until a best fit was achieved. A simple Cauchy dispersion relation fit did not produce a very good fit and was not used. The region up to 2.8 μm shows very little absorption ($k=0.003$). The film shows increasing absorption beyond 6.4 μm. Beyond 10 μm,

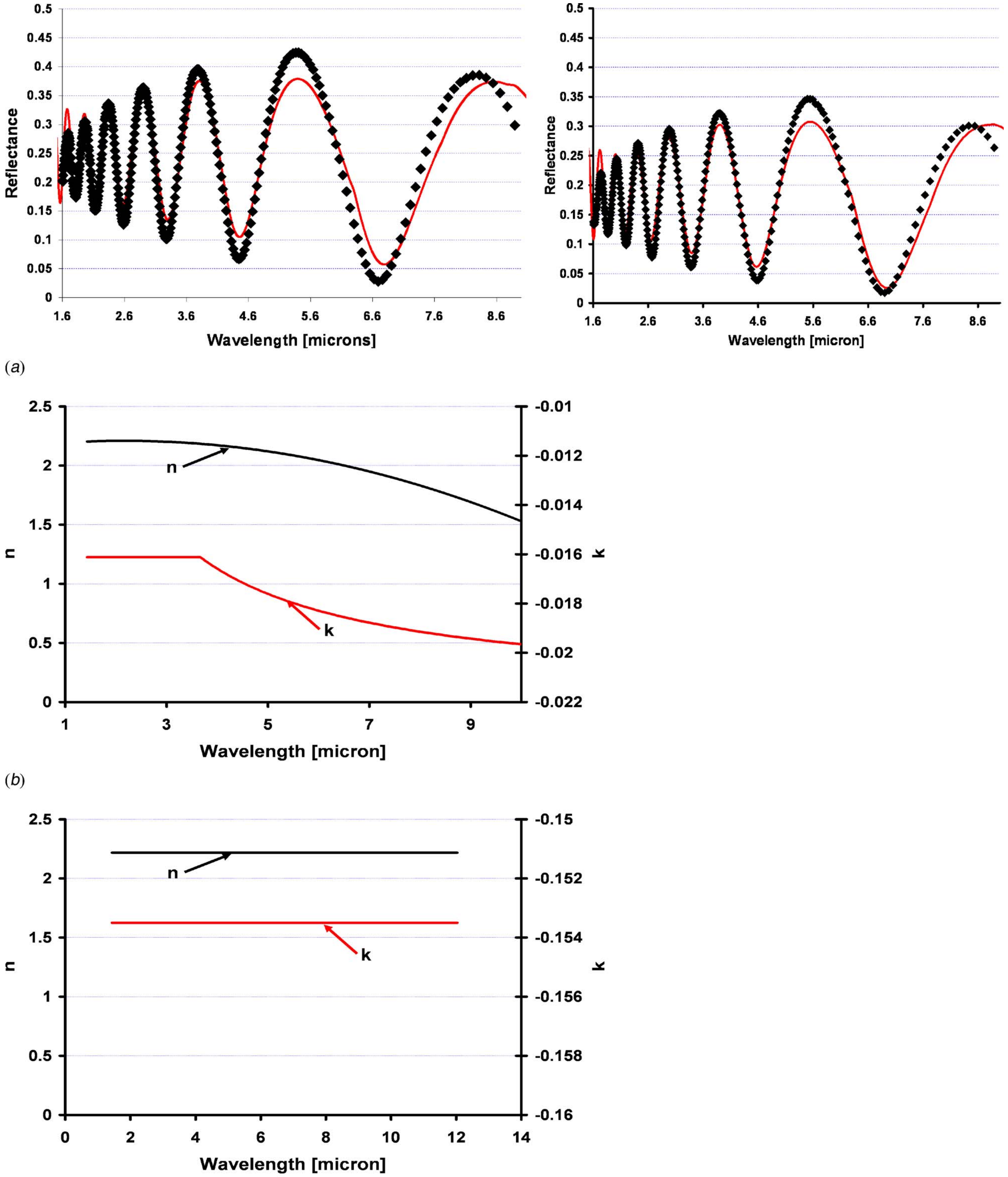


Fig. 8 (*a*) **Comparison of the measured (dotted) and fitted (complete line) reflectance for Ta_2O_5 heated to 900°C for 9 h, (overall film 2.81 μm thick, crystalline). Also, the optimization routine gave the thickness of the interfacial layer to be 0.41 μm. The figures are shown for 30 deg (on the right) and 45 deg incident and reflected angles (*b*) *n* and *k* values as a function of wavelength for the Ta_2O_5 film. (c) *n* and *k* values for the interfacial layer between film and substrate.**

the optical properties could no longer be calculated using the algorithm presented in this paper as the film became too absorbing and the optimization code would not converge for multiple angles. It might be possible to model the Ta_2O_5 as a Lorentz oscillator and thus find its optical properties using the reflectance data collected at a single angle [23]. However, such a study of the wavelengths where the absorption is very high does not add much to the study of the effect of interfacial layers formed as a result of high-temperature exposure and hence was not included in this study.

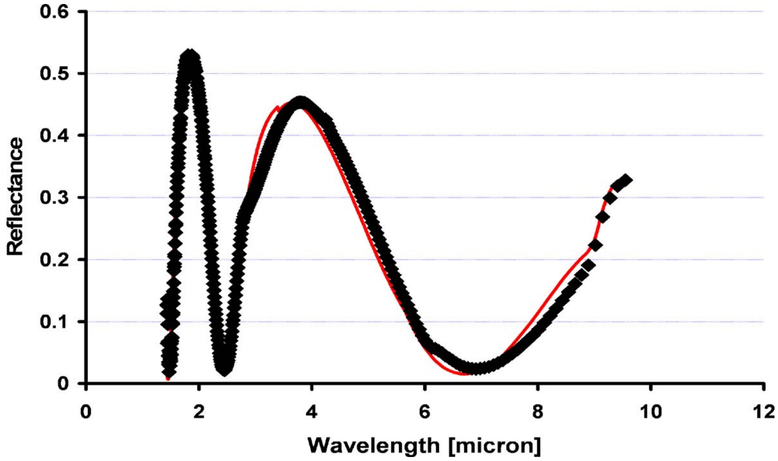


Fig. 9 Comparison of measured (dotted) and fitted (complete line) reflectance for a 0.85-μm-thick Ta_2O_5 film (as-deposited) on 0.1 μm SiO_2. The fitted reflectance was calculated from the *n* and *k* values shown in Fig. 7(*c*). This plot provides independent verification of the *n* and *k* values for the as-deposited film.

For calculating the properties of the films with interfacial layers, the best fit was obtained when the interfacial layer between film and substrate was assumed to have constant optical properties (i.e., spectrally independent) and the Ta_2O_5 to have varying optical properties. It is seen that the $TaSi_xO_y$ interfacial layer has an *n* value similar to the Ta_2O_5 film, but has much higher *k* values. This overall increase in absorption can be seen from the reflectance plot alone, but without calculating the optical properties, this increase could not be uniquely attributed to the formation of an interfacial layer, as it could be an effect of the conversion to crystalline form.

Once the optical properties of the first sample (3 μm when as-deposited) before and after heating were calculated, they were used to calculate the reflectance of films of dimensions equal to the second sample (0.85 μm when as-deposited). These films were on SiO_2, whose optical properties are well documented in the literature [22]. Hence the reflectance of the multilayer of SiO_2 and Ta_2O_5 (amorphous or crystalline) can be calculated. At this point, it must be mentioned again that the thickness of the Ta_2O_5 film on the thermal oxide is different from that on the Si. The effect of thickness of the film on the optical properties is beyond the scope of this study and hence the optical constants from the first sample were used to calculate the reflectance of the second sample. This was done in order to provide additional verification of the optical constants and also to verify their applicability to films on thermal oxide.

Figure 9 compares the measured reflectance of an as-deposited 0.85 μm Ta_2O_5 film with the reflectance calculated from the optical properties derived using a 3 μm Ta_2O_5 film and it shows good agreement ($|R_{\text{measured}}-R_{\text{fitted}}| < 0.02$). Figure 10 similarly compares the reflectance after the 0.85 μm film was heated to 900°C. Given that the Auger depth profile in Fig. 5 did not show a thick interfacial layer, it was assumed that the film on SiO_2 was simply a layer of crystalline Ta_2O_5 on top of SiO_2. The values of the crystalline Ta_2O_5 optical properties were those obtained from

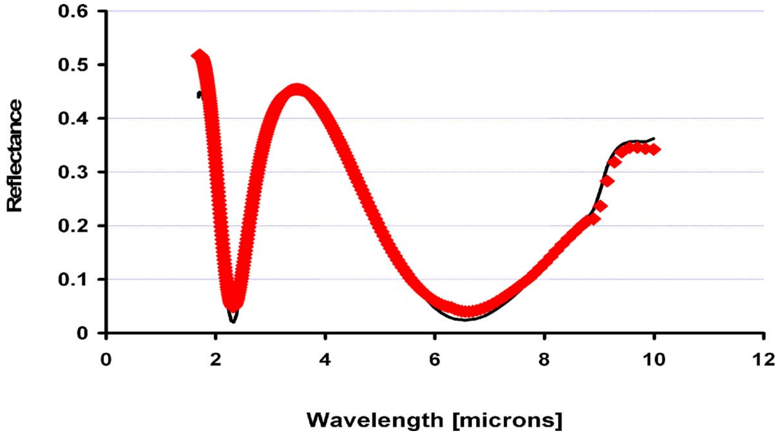


Fig. 10 Comparison of measured (dotted) and fitted (complete line) reflectance for a Ta_2O_5 film subjected to 900°C for 9 h on 0.1 μm SiO_2. The fitted reflectance was calculated from the *n* and *k* values shown in Fig. 9(*b*). Note that the interfacial layer formed between the Ta_2O_5 and SiO_2 films is not significantly thick (Fig. 5) and hence only the crystalline Ta_2O_5 film optical properties were substituted to calculate the fitted reflectance here. This plot provides independent verification of the *n* and *k* values for the crystalline Ta_2O_5 film (without the interfacial layer).

the prior tested sample (Fig. 8(*b*)), and the SiO_2 optical properties were obtained from the literature [22].Once again, the agreement between fitted and calculated values is good ($|R_{\text{measured}}\text{-}R_{\text{fitted}}| < 0.02$) at all wavelengths from 1.5 μm to 10 μm) confirming the crystalline Ta_2O_5 film optical properties and in effect also the optical properties of the interfacial layer between film and substrate.

5 Conclusions

Ta_2O_5 films on Si when heated to 900°C showed the formation of an interfacial layer between the tantala (Ta_2O_5) film and substrate. The reflectance measured before and after heating the sample showed decreased reflectance, especially at lower wavelengths. LSE optimization for the values of optical constants in the IR showed that the increased absorption was attributable to the interfacial layer formed. The interfacial layer thickness was seen to be much smaller for the films deposited on a thin SiO_2 layer, possibly due to the SiO_2 layer acting as a diffusion barrier to the Si. Finally, by using the optical constants calculated from the films on Si to calculate the reflectance of films on SiO_2, the optical constant values were verified to be correct.

Acknowledgment

The authors are grateful to Dr. Mauro Sardela and Mrs. Nancy Finnegan for their assistance in the use of x-ray diffractometry and Auger electron spectroscopy. The material analysis in this paper was carried out at the Center for Microanalysis of Materials, University of Illinois, which is partially supported by the U.S. Department of Energy under Grant No. DEFG02-91-ER45439. This work has been supported by Department of Defense Multi University Research Initiative (MURI) program under Grant No. DAAD19-01-1-0582 and The WaterCAMPWS, a Science and Technology Center under National Science Foundation Agreement No. CTS-0120978.

Nomenclature

d = physical thickness of film
FTIR = Fourier transform infrared (spectroscopy)
k = imaginary part of refractive index
LSE = least square error
MCT = mercury–cadmium–telluride (type of IR detector)
N = complex refractive index
n = real part of refractive index
XPS = x-ray photoelectron spectroscopy
XRD = x-ray diffraction
Y = optical admittance
δ = phase thickness of film
λ = wavelength
θ = angle measured from normal to surface of film

References

[1] Miesse, C. M., Masel, R. I., Jensen, C. D., Shannon, M. A., and Short, M., 2004, "Submillimeter-Scale Combustion," AIChE J., **50**(12), pp. 3206–3214.
[2] Fernandez-Pello, A. C., 2002, "Micropower Generation Using Combustion: Issues and Approaches," Proc. Combust. Inst., **29**, p. 883.
[3] Warnatz, J., Maas, U., and Dribble, R. W., 2001, *Combustion*, 3rd ed., Springer, New York.
[4] Chaneliere, C., Autran, J. L., Devine, R. A.B, and Balland, B., 1998, "Tantalum Pentoxide (Ta_2O_5) Thin Films for Advanced Dielectric Applications," Mater. Sci. Eng., **22**(4), pp. 269–322.
[5] Obata, Y., and Hashimoto, N., 2002, *Deposition of Multilayer IR Reflective Layers with Controlled Thickness on Lamp Bulbs*, Jpn. Kokai Tokkyo Koho, Tokyo, Japan.
[6] Franke, E., Trimble, C. L., Devries, M. J., Woollam, J. A., Schubert, M., and Frost, F., 2000, "Dielectric Function of Amorphous Tantalum Oxide from the Far Infrared to the Deep Ultraviolet Spectral Region Measured by Spectroscopic Ellipsometry," J. Appl. Phys., **88**(9), pp. 5166–5174.
[7] Atanassova, E., and Paskaleva, A., 2002, "Breakdown Fields and Conduction Mechanisms in Thin Ta_2O_5 Layers on Si for High Density DRAMs," Microelectron. Reliab., **42**, pp. 157–173.
[8] Zhang, J. Y., and Boyd, I. W., 2000, "Formation of Silicon Dioxide Layers During UV Annealing of Tantalum Pentoxide," Appl. Surf. Sci., **168**, pp. 312–315.
[9] Incropera, F. P., and DeWitt, D. P., 1990, *Fundamentals of Heat and Mass Transfer*, 3rd ed., Wiley, New York.
[10] Macleod, H. A., 1969, *Thin-Film Optical Filters*, Elsevier, New York.
[11] Ford, J. N., Tang, K., and Buckius, R. O., 1995, "Fourier Transform Infrared System Measurements of the Bi-directional Reflectivity of Diffuse and Grooved Surfaces," J. Heat Transfer, **117**, pp. 955–962.
[12] Kawka, P., 2000, "An Experimental Theoretical and Numerical Investigation of Directional Scattering from Configured Surfaces," M.S. thesis, University of Illinois, Urbana, IL.
[13] Born, M., and Wolf, E., 1970, *Principles of Optics*, 4th ed., Pergamon, New York.
[14] Abeles, F., 1963, "Methods for Determining Optical Parameters of Thin Films," Prog. Opt., **2**, pp. 251–288.
[15] Hopfe, V., Bussemer, P., Rlchter, E., and Klobes, P., 1992, "p- and s-Polarized FTIR Reflectance Spectroscopy at Oblique Incidence by Kramers–Kronig Transformation," J. Phys. D, **25**, 288–294.
[16] Klein, S., and Alvarado, F., 1991, "Engineering Equation Solver," F-Chart Software, Madison, WI.
[17] Davidon, W., 1959. "Variable metric method for minimization," SIAM J. Optim., **1**(1), pp. 1–17.
[18] Borgogno, J. P., Lazarides, B., and Pelletier, E., 1982, "Automatic Determination of the Optical Constant of Inhomogeneous Thin Films," Appl. Opt., **21**(22), pp. 4020–4029.
[19] Babu, S. V., David, M., and Patel, R. C., 1991, "Two-Step Regression Procedure for the Optical Characterization of Thin Films," Appl. Opt., **30**(7), pp. 839–846.
[20] Poitras, D., and Martinu, L., 2000, "Interphase in Plasma-Deposited Films on Plastics: Effect on the Spectral Properties of Optical Filters," Appl. Opt., **39**(7), pp. 1168–1173.
[21] Oehrlein, G. S., d'Heurle, F. M., and Reisman, A., 1984, "Some Properties of Crystallized Tantalum Pentoxide Thin Films on Silicon," J. Appl. Phys., **55**, pp. 3715–3725.
[22] Palik, E., 1992, *Handbook of Optical Constants*, Academic, New York.
[23] Brewster, Q., 1991, *Thermal Radiative Transfer and Properties*, Wiley, New York.

Obtaining Subwavelength Optical Spots Using Nanoscale Ridge Apertures

E. X. Jin
e-mail: xjin@purdue.edu

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

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

Concentrating light into a nanometer domain is needed for optically based materials processing at the nanoscale. Conventional nanometer-sized apertures suffer from low light transmission, therefore poor near-field radiation. It has been suggested that ridge apertures in various shapes can provide enhanced transmission while maintaining the subwavelength optical resolution. In this work, the near-field radiation from an H-shaped ridge nanoaperture fabricated in an aluminum thin film is experimentally characterized using near-field scanning optical microscopy. With the incident light polarized along the direction across the gap in the H aperture, the H aperture is capable of providing an optical spot of about 106 nm *by* 80 nm *in full-width half-maximum size, which is comparable to its gap size and substantially smaller than those obtained from the square and rectangular apertures of the same opening area. Finite different time domain simulations are used to explain the experimental results. Variations between the spot sizes obtained from a* 3×3 *array of H apertures are about 4–6%. The consistency and reliability of the near-field radiation from the H apertures show their potential as an efficient near-field light source for materials processing at the nanoscale.* [DOI: 10.1115/1.2401196]

1 Introduction

There has been a continuous need to increase the areal density of microelectronic devices. To further increase the areal density without relying on expensive light sources of shorter wavelength, one has to seek nonconventional approaches, for examples, solid immersion lens [1], and optical near-field techniques [2,3]. In the near-field optical approach, a subwavelength aperture in an opaque metal film is often used as a radiation transducer. It converts far-field radiation to spatially confined near-field radiation, and provides a subdiffraction-limited light source whose resolution depends on the size of the aperture instead of the wavelength of light. The re-radiated field from the subwavelength aperture is localized in the vicinity of the aperture, so the material to be processed is brought close to the metal film at a distance less than the wavelength of light. For example, a very small aperture laser (VSAL) has been proposed to achieve optical storage density exceeding tens of $\mathrm{Gb/in.}^2$ [4]. To reach higher areal density, the physical size of the aperture has to be reduced to much less than the diffraction limit. However, according to the standard aperture theory [5], the light transmission through a small hole decays as a function of the fourth power of the diameter of the aperture, which is attributed to the wavelength cutoff effect. Regular small apertures in circular and square shapes will suffer from extremely low power throughput that is insufficient for materials optical processing. Enhanced light transmission through small apertures is desired in order to be used as efficient transducers for radiation transfer.

Making a small aperture surrounded with periodic ring corrugations (bull's eye pattern) in noble metal films, the resonant transmission through the center aperture can be boosted by two orders of magnitude [6] relative to a bare aperture without any surrounding structures due to the resonant excitation of surface plasmon [7]. However, the high transmission efficiency of a hole smaller than 100 nm using this method has not been demonstrated yet. Another approach is to use a subwavelength aperture with a waveguide propagation mode supported in the aperture. In particular, ridge apertures in various shapes, i.e., *C* [8], *I* [9], or *H* [10,11], and bowtie [12,13] apertures, have been proposed to achieve nanoscale near-field radiation with enhanced optical transmission compared to regular small apertures. Ridge apertures have the specially designed geometry, which consists of a narrow gap connecting two side arms. The unique optical properties of ridge apertures benefit from the fundamental TE_{10} waveguide mode excited in the aperture channel, which has the transverse electric field confined within the nanometer-sized gap formed by the ridges and propagates through the aperture without experiencing much power loss [10].

Numerical simulations showed that *C* [8] and *H* [10] apertures made in a perfectly conducting plate provide enhanced power throughput of 2–3 orders of magnitude higher than that obtained by comparable conventional small square apertures. Detailed theories and numerical calculations of transmission enhancement and localization of radiated fields of ridged apertures can be found in Refs. [8–13]. The far-field radiation measurement of various apertures made in a gold film suggests *C* aperture shows an extraordinary transmission enhancement of about 10^6 over a conventional aperture of the same near-field resolution [14]. Furthermore, the confined optical near field of *C* aperture has also been experimentally demonstrated using near-field scanning optical microscope (NSOM) measurements [15]. This NSOM experiment [15] used an apertureless near-field probe, which scattered the radiated field from the *C* aperture.

In this paper, the near-field optical properties of an *H*-shaped ridge nanoaperture fabricated in an aluminum thin film are characterized using a specially designed aperture NSOM system. A cantilevered near-field probe with a nanoaperture at the end of the cantilever tip is used to directly map the near fields from the *H* aperture and conventional apertures of comparable sizes. The polarization dependence of near-field radiation from these apertures is investigated. The use of *H* apertures as highly efficient and reliable near-field transducers is demonstrated by imaging the optical near field from an array of *H* apertures. The experimental results are discussed and explained through finite difference time domain (FDTD) numerical computations together with the waveguide theory.

Contributed by the Heat Transfer Division of ASME for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received January 23, 2006; final manuscript received June 16, 2006. Review conducted by M. Pinar Menguc.

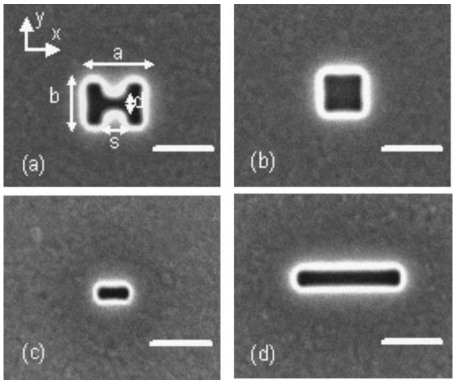


Fig. 1 SEM images of: (*a*) *H* aperture (*a*=210 nm, *b*=170 nm, *s*=100 nm, *d*=60 nm); (*b*) square aperture (160 nm by 160 nm); (*c*) small rectangular aperture (100 nm by 60 nm); and (*d*) large rectangular aperture (400 nm by 60 nm) fabricated in a 75 nm aluminum film on a quartz substrate. The scale bars are 200 nm.

2 Experimental Details

The *H* apertures and regular apertures of comparable sizes are fabricated in a thin aluminum film by focus ion beam (FIB) milling using a dual beam FIB machine (FEI Strata DB 235). The thickness of the aluminum film is 75 nm, which is deposited onto a quartz substrate by an *e*-beam evaporator. The scanning electron microscope (SEM) images of these fabricated apertures are shown in Fig. 1. The *H* aperture has an overall size of about 210 nm (a) × 170 nm (b) with a small gap of 100 nm (s) × 60 nm (d). The square and rectangular apertures are about 160 nm × 160 nm, and 400 nm × 60 nm in size, having roughly the same opening area as the *H* aperture. A 100 nm × 60 nm regular aperture, which is supposed to provide the near-field optical spot of similar size as the *H* aperture, is also made for the purpose of comparison.

In order to retrieve the near-field information from these subwavelength apertures, a NSOM technique [16] has been employed. As shown in the schematic view (Fig. 2) of our aperture NSOM setup, the aluminum sample with fabricated apertures is illuminated by an argon-ion laser of about 2 mW power output at 458 nm wavelength from the substrate side. The spot size of the laser beam on the sample is about 50 μm. Because the near-field radiation from the subwavelength apertures in the sample is localized in the vicinity of the apertures, a near-field probe with a nanometer-sized aperture at the end is used and brought to the close proximity of the sample surface. It collects both nonpropagating and propagating components of the near-field radiation, and converts them to the propagating radiation which can be detected in the far field by a photomultiplier tube (PMT). The incident laser beam is adjusted to be polarized either in the y direction or the x direction as indicated in Fig. 1(*a*), and focused on the metal film. A NSOM image is obtained by raster scanning the apertures in the sample and recording the optical signal from the PMT by photon counting. During the sample scanning, the probe is in soft contact with the surface, which is achieved by maintaining a small and constant normal force (the tip–sample separation is typically in the 0–10 nm range) based on the cantilever feedback control by monitoring the deflected diode laser beam on the position sensitive detector (PSD).

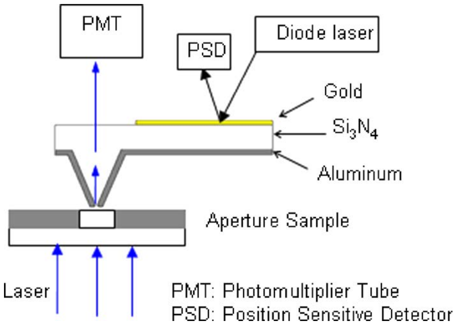


Fig. 2 Schematic view of the specially designed NSOM system using a cantilever aperture probe

The near-field aperture probe is the key element in this NSOM system. It determines the optical resolution that can be achieved. As shown in Fig. 3(*a*), the probe is in a pyramidal shape located near the end of a cantilever. It has a silicon nitride core coated with a thin aluminum film. An aperture of less than 100 nm in size at the apex of the probe is opened by FIB milling. To characterize the optical resolution of the NSOM measurements, the transmitted light from a pair of nanoholes in the same sample with known separation distances is collected by this probe. The obtained

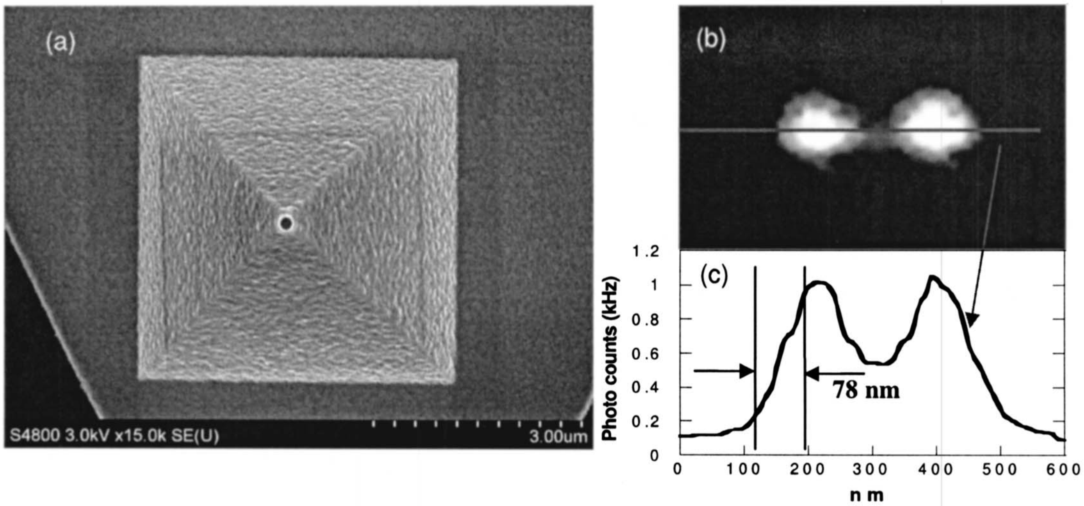


Fig. 3 (*a*) SEM image of cantilever aperture probe made by FIB milling; (*b*) NSOM image of a pair of nanoholes used to characterize the optical resolution the aperture probe; and (*c*) line scan on the NSOM image shows the 10–90% edge resolution of 78 nm

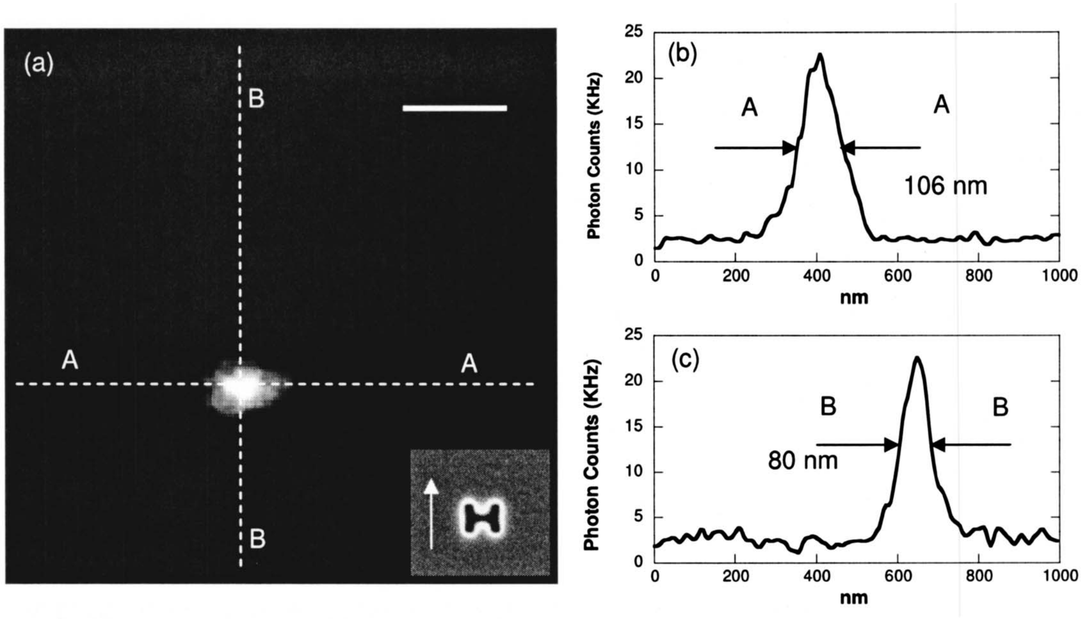


Fig. 4 (*a*) **NSOM image of the *H* aperture (210 nm × 170 nm outline with a 100 nm × 60 nm gap) and line scans in (*b*) A-A horizontal direction and (*c*) B-B vertical direction show the FWHM spot size is 106 nm by 80 nm. The illumination light is polarized along the *y* direction across the gap the *H* aperture as indicated in the inset of (*a*). The scale bar in (*a*) is 200 nm.**

NSOM image is shown in Fig. 3(*b*). Figure 3(*c*) show the 10–90% edge resolution is about 78 nm by plotting the line scan profile of the NSOM signal as indicated in Fig. 3(*b*). This optical resolution is about $\lambda/6$ of the illumination light, and much less than the diffraction limit as expected.

3 Results and Discussion

3.1 NSOM Measurements. The near-field radiation from the *H* aperture and comparable conventional apertures is measured using the previously described NSOM system and the aperture probe of 78 nm optical resolution. Figure 4(*a*) shows the NSOM image obtained from the *H* aperture with the *y* polarized laser illumination. Since the overall tip size (the aperture opening surrounded by aluminum coating) of the probe is larger than the opening area of the *H* aperture, the probe could not get inside the aperture during the acquisition of the near-field signal. Therefore, the obtained NSOM image can be considered as a NSOM image performed at the constant height mode with a short distance (0–10 nm) away from the exit plane of the *H* aperture. The line scan analysis along the *x* and *y* directions is conducted as indicated on the two-dimensional NSOM image and plotted in Figs. 4(*b*) and 4(*c*), respectively. The full-width half-maximum (FWHM) of the near-field light spot is found to be 106 nm along the *x* direction (A-A shown in Fig. 4(*b*)), and 80 nm along the *y* direction (B-B shown in Fig. 4(*c*)). The actual spot size should be smaller due to the convolution effect of the aperture probe. Considering the symmetries of the *H* aperture geometry and near-field light spot, the near-field light through the *H* aperture is expected to be localized in the central gap between the ridges.

The near-field radiation from the conventional apertures of regular shapes is also acquired under the same illumination conditions. Figures 5(*a*) and 5(*b*) show the NSOM images for the square aperture and large rectangular aperture, respectively. The measured FWHM spot sizes are 136 nm × 174 nm for the square aperture, and 212 nm × 120 nm for the rectangular aperture. Both light spots are substantially larger than that achieved by the *H* aperture. The NSOM measurement of the small rectangular aperture, which has the same size as the gap of the *H* aperture and is supposed to provide the near-field spot of similar size, was also attempted but failed to acquire sufficient light that can be detected by the measuring system. This indicates the expected low light transmission using the sub-100 nm regular apertures. Since these apertures are separated from each other by a distance of more than 10 μm, and thus the incident laser beam (diameter=50 μm) does not uniformly illuminate each aperture, the peak intensities of the NSOM spots from the apertures cannot be compared with each other. The relative transmission efficiencies through these apertures will be numerically investigated in the following section.

NSOM images of these apertures are also taken with the laser illumination polarized in the orthogonal direction or the *x* direction. It is found that both rectangular apertures do not produce enough light signals to form a NSOM image under this polarization. As shown in Fig. 6(*b*), a light spot of 167 nm × 132 nm in FWHM size is achieved by the square aperture. The square aperture results in similar near-field spots in both polarizations as expected, and the elongation of the near-field spot is along the direction of laser polarization. The *H* aperture results in a 170 nm × 188 nm light spot as shown in Fig. 6(*a*), which is significantly

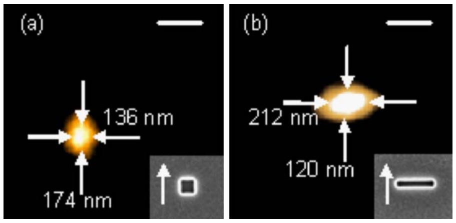


Fig. 5 NSOM images of: (*a*) the 160 nm × 160 nm square aperture and (*b*) the 400 nm × 60 nm rectangular aperture. The FWHM sizes of the near-field optical spots shown in the images are 174 nm × 136 nm and 212 nm × 120 nm for the square and rectangular apertures, respectively. The insets show the SEM pictures of the imaged apertures. The illumination light is polarized along the *y* direction as indicated in the insets. The scale bars are 200 nm.

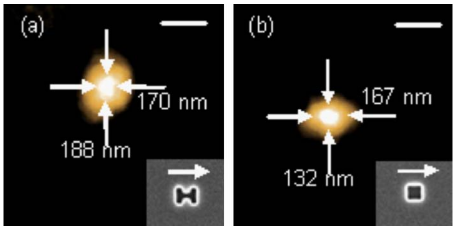


Fig. 6 NSOM images of: (*a*) the *H* aperture and (*b*) the 160 nm × 160 nm square aperture. The FWHM sizes of the near-field optical spots shown in the images are 170 nm × 188 nm and 167 nm × 132 nm for the *H* and square apertures, respectively. The insets show the SEM pictures of the imaged apertures. The illumination light is polarized along the *x* direction as indicated in the insets. The scale bars are 200 nm.

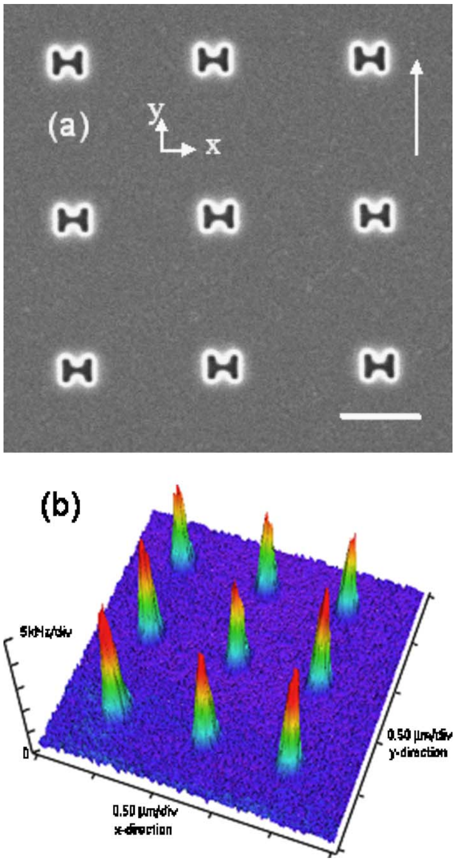


Fig. 7 (*a*) SEM image of an array of *H* apertures fabricated in an aluminum film and (*b*) three-dimensional NSOM image of the array. The illumination light is polarized along the *y* direction across the gap of the *H* apertures as indicated by the arrow in the inset of (*a*). The scale bar in (*a*) is 500 nm.

larger than that is obtained under the *y* polarized illumination. This light spot almost covers the whole opening area of the *H* aperture, therefore no field concentration is provided by the *H* aperture when it is illuminated by light linearly polarized across the two open arms.

In comparison with the conventional apertures, the NSOM measurements clearly show that the *H* aperture is able to achieve the smallest light spot with considerable enhanced optical transmission under the proper polarization. The spot size is essentially determined by the gap size of the *H* aperture, as will also be seen from the FDTD results. It is therefore expected that further reduction of the gap size can produce a smaller light spot [11]. This makes *H* apertures as well as other ridge apertures attractive for use as an efficient near-field transducer for optical processing.

To examine the consistency and reliability of the NSOM results obtained by the *H* aperture, the optical near field from a 3 × 3 array of *H* apertures is measured. As shown in the SEM image (Fig. 7(*a*)) of the *H* aperture array fabricated in the same aluminum sample, the nine *H* apertures are almost identical and uniformly arranged in the array, and the variation in the dimensions of these apertures is less than 5 nm measured from the SEM image. Figure 7(*b*) shows the three-dimensional NSOM image collected under the laser illumination polarized in the direction across the gap of the *H* apertures. An array of clean light spots is obtained in the same pattern as the *H* aperture array. The average FWHM size of these nine light spots is 107 nm × 84 nm. The standard deviations of FWHM sizes are 4.7 nm in the *x* direction and 5.2 nm in *y* direction, corresponding to 4.4% and 6.1% size variation (relative standard deviation), respectively. In fact, the variation of near-field spot sizes is in the same order of the scanning resolution. The peak intensity value of these light spots has a 14% variation. These results indicate the consistency of near-field radiation from the *H* apertures as well as the reliability of the NSOM measurements.

3.2 FDTD Computations. In order to explain the experimental results, FDTD simulations are carried out to calculate the optical near field of each aperture, and also to illustrate how the light propagates through each aperture. The detailed description of the FDTD numerical method can be found in the Refs. [17,18]. In the following FDTD computations, the dimensions of each aperture are chosen based on the measurement of the SEM images. Both the optical property of the aluminum at 458 nm [19] and the change of laser polarization are considered.

Figure 8 shows the FDTD results for the *H* aperture, square aperture, and large rectangular aperture under the *y* polarized illuminating light at 458 nm wavelength. The left column shows the near-field intensity distribution on the *xy* plane at 10 nm below the three apertures. Two hot spots can be seen for all three apertures, which were not observed in the NSOM measurements. The difference could arise from two possibilities. First, the apertures used in the calculations have sharp edges while the fabricated ones have round edges. In fact, the sharp edges induce the *z* component of the electric field in the FDTD calculations along the polarization direction due to the scattering effect [13]. Second, the optical resolution of the aperture probe is not likely enough to resolve the two hot spots. The NSOM spot is the convolution of the scanning probe transfer function with the actual near-field light spot transmitted through the measured aperture. It is only when the probe is a point source (representing a delta function) that the measured NSOM spot is exactly the same as the actual near-field spot assuming the probe does not disturb the radiating field. On the other hand, the numerical results clearly show that the *H* aperture confines the near-field radiation underneath the gap region and results in the smallest light spot in both lateral directions compared to the other two apertures. This is consistent with the NSOM measurements. The right column of Fig. 8 shows how light propagates through the three apertures. Both the *H* aperture and rectangular aperture have a propagating waveguide mode in the aperture channels (Figs. 8(*b*) and 8(*f*)), which provides the efficient transmission through these apertures. The existence of waveguide mode in these two apertures is because the incident wavelength is shorter than the cutoff wavelengths of the apertures [10]. By either elongating the small rectangular aperture (to become the large rectangular aperture) or adding two side arms to it (to become the *H* aperture), the cutoff wavelength of the aperture will be significantly increased. An evanescent mode with the field intensity decaying exponentially is found inside the square aperture as shown in Fig. 8(*d*). However, because the illumination wavelength of 458 nm is close to the cutoff wavelength of the

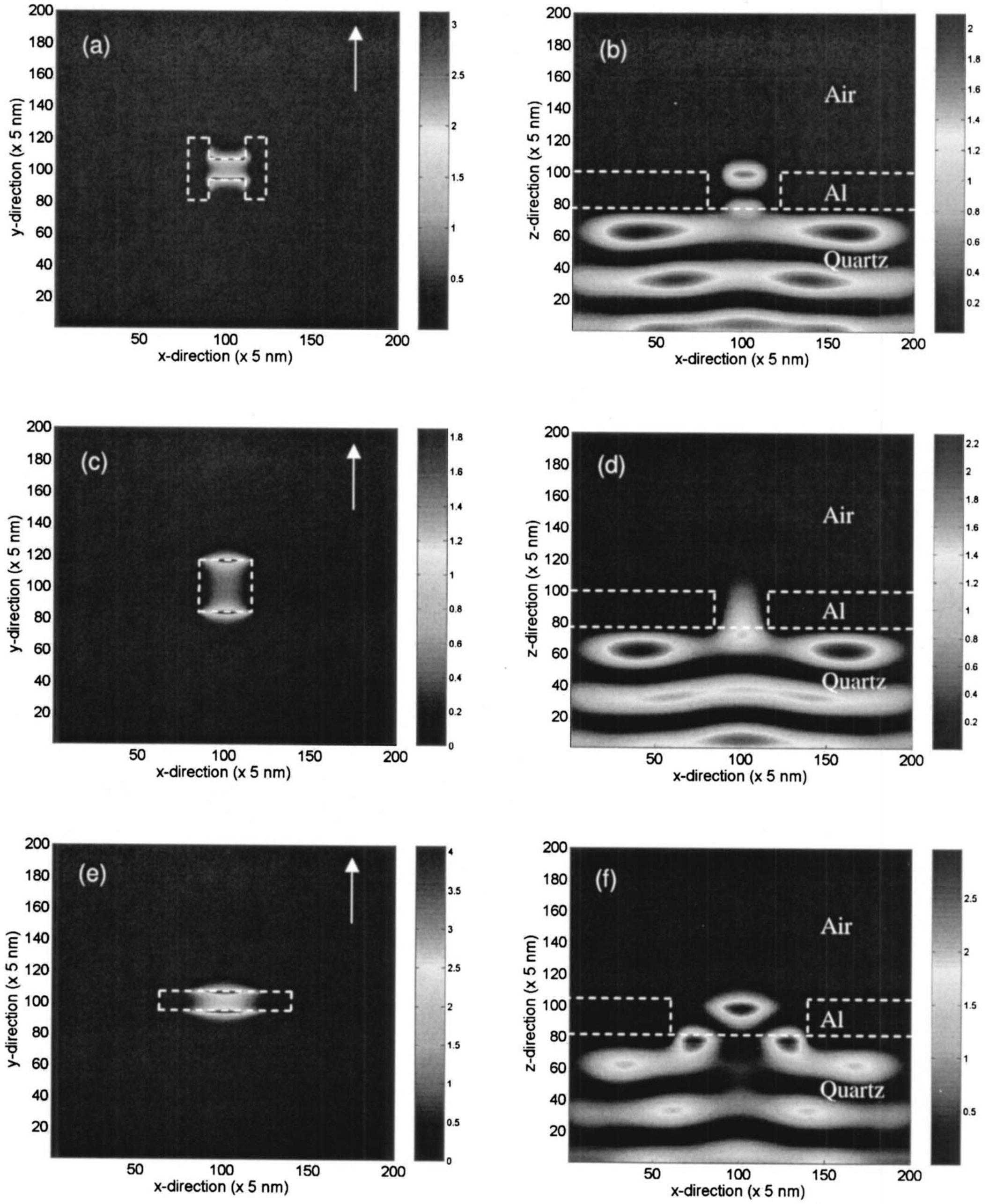


Fig. 8 **Left column: the computed near-field intensity distribution (normalized by incident intensity) on the *xy* plane at 10 nm below (*a*) the *H* aperture; (*c*) the square aperture; and (*e*) the rectangular aperture in a 75-nm-thick aluminum film on a quartz substrate. Right column: light propagation on the *xz* plane through the middle of (*b*) the *H* aperture; (*d*) the square aperture; and (*f*) the rectangular aperture. The *y* polarized plane wave at 458 nm wavelength is incident from the quartz side. The arrows indicate the polarization direction of the incident light.**

square aperture [20], a small attenuation of the transmitted field results. There are still enough photons tunneling through the subwavelength square aperture and resulting in signals that can be detected by the PMT. The integrated intensities over a 400 nm × 170 nm area at 10 nm below the *H*; square, large rectangular, and small rectangular apertures are 938, 610, 1761, and 45, respectively (the incident integrated intensity over the same area is 2720). Therefore, the *H* aperture is able to provide the smallest optical spot with the light transmission on the same order of those through the regular apertures of the same opening area, but 20 times that through the smallest rectangular aperture. According to the standard aperture theory [5], the transmission efficiency through the small rectangular aperture will be greatly decreased by further reducing its size in order to achieve a smaller optical spot. The *H* aperture, on the other hand, can maintain the transmission efficiency (on the order of unity) when the gap size is further reduced due to the existence of the propagating waveguide mode.

The FDTD results for the *x*-polarized illumination are shown in Fig. 9. Due to the symmetry of its geometry, the same result as for the *y* polarization is expected for the square aperture, therefore is not computed. It is clearly seen a greatly attenuated evanescent mode in the rectangular aperture in Fig. 9(*d*). This can be again explained by the waveguide cutoff calculation [20]: the illumina-

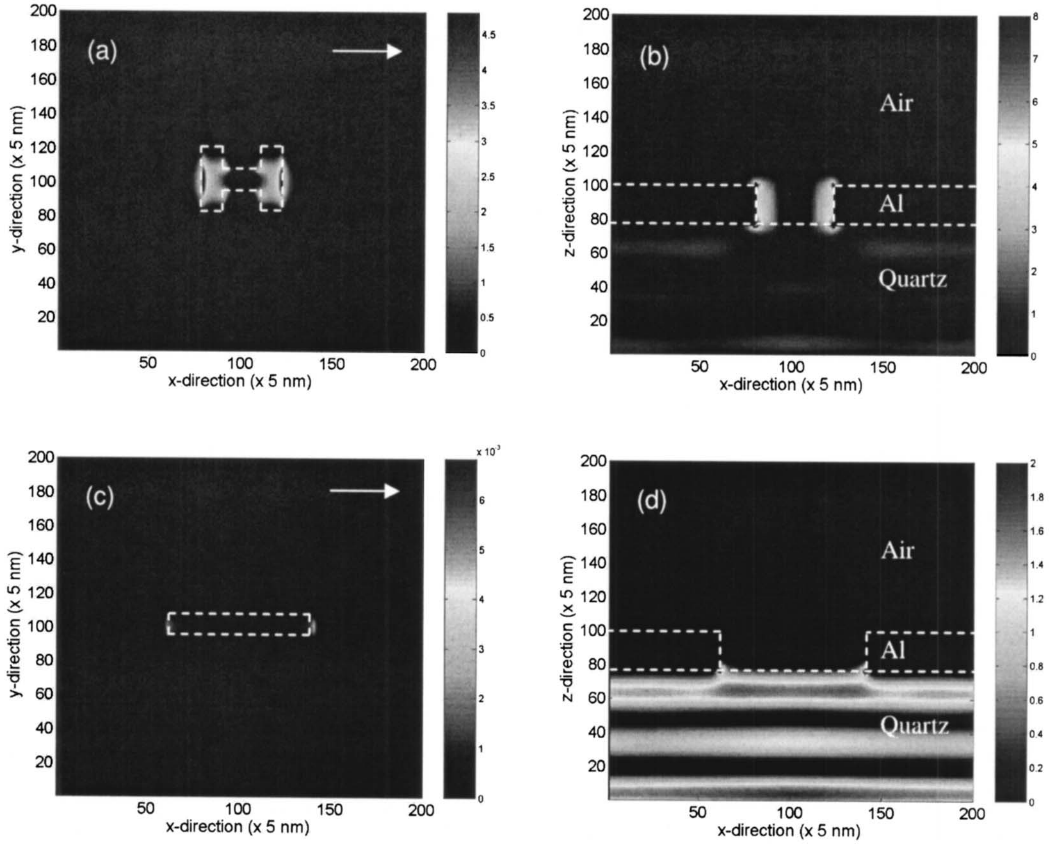


Fig. 9 Left column: the computed near-field intensity distribution (normalized by incident intensity) on the *xy* plane at 10 nm below (*a*); the *H* aperture, and (*c*) the square aperture in a 75-nm-thick aluminum film on a quartz substrate. Right column: light propagation on the *xz* plane through the middle of (*b*) the *H* aperture, and (*d*) the rectangular aperture. The *x* polarized plane wave at 458 nm wavelength is incident from the quartz side. The arrows indicate the polarization direction of the incident light.

tion wavelength is much larger than the cutoff wavelength of the rectangular aperture under the *x*-polarized illumination resulting in a large attenuation coefficient. In this case, the rectangular aperture looks like the other part of the aluminum film so that most of the incident photons are reflected back. The field intensity below the aperture is three orders less than that of the incident one as shown in Fig. 9(*c*) because only a few photons can tunnel through the short aperture channel. On the other hand, the *H* aperture is still able to provide a strong radiating field at the exit side whose intensity is in the same order of the incidence as shown in Fig. 9(*a*). This is because the two open arms of the *H* aperture act as propagating waveguides under this polarization. The surface waves are found on the side walls in the aperture channel, therefore allowing a large mount of light through the aperture as shown in Fig. 9(*b*). It is also found that the near-field spot of the *H* aperture under *x* polarization illumination is considerably larger than that in the *y* polarization illumination. These FDTD results at *x* polarization are all consistent with the NSOM observations.

4 Conclusions

In summary, the near-field radiation from the *H*-shaped apertures and conventional apertures of comparable sizes fabricated in a 75-nm-thick aluminum film is characterized using a specially designed aperture NSOM system with the optical resolution around 80 nm. FDTD numerical computations are used to explain the NSOM observations. It is found that the *H* aperture is able to provide a 106 nm × 80 nm optical spot when the incident light at 458 nm is polarized along the direction across the gap. A 4–6% FWHM size variation and a 14% intensity variation of near-field spots from a 3 × 3 array of *H* apertures are obtained, showing the consistency and reliability of the near-field radiation from the fabricated *H* apertures. The *H* aperture results in a larger spot under the *x*-polarized illumination light. Compared to the square aperture and the rectangular of the same opening area, the *H* aperture provides a substantially smaller near-field light spot without suffering the low light transmission, which allows many promising applications in near-field nanoscale optical engineering.

Acknowledgment

The financial support to this work by the Office of Naval Research and the National Science Foundation is gratefully acknowledged. Fabrications of aperture sample and aperture probe by FIB milling were carried out in the Center for Microanalysis of Materials, University of Illinois, which is partially supported by the U.S. Department of Energy.

Nomenclature

a = length of the *H* aperture
b = width of the *H* aperture
d = gap width of the *H* aperture
s = gap length of the *H* aperture
λ = wavelength of light/laser
E = electric field

References

[1] Mansfield, S. M., Studenmund, W. R., Kino, G. S., and Osato, K., 1993, "High Numerical-Aperture Lens System for Optical Data Storage," Opt. Lett., **18**(4),

pp. 305–307.
[2] Betzig, E., Trautman, J. K., Wolfe, R., Gyorgy, E. M., Finn, P. L., Kryder, M. H., and Chang, C. H., 1992, "Near-Field Magneto-Optics and High Density Data Storage," Appl. Phys. Lett., **61**, pp. 142–144.
[3] Peng, C., Mihalcea, C., Buchel, D., Challener, W. A., and Gage, E. C., 2005, "Near-Field Optical Recoding Using a Planar Solid Immersion Mirror," Appl. Phys. Lett., **87**, pp. 151105-1–151105-3.
[4] Partovi, A., Peale, D., Wuttig, M., Murray, C. A., Zydzik, G., Hopkins, L., Baldwin, K., Hobson, W. S., Wynn, J., Lopata, J., Dhar, L., Chichester, R., and Yeh, J. H.-J., 1999, "High-Power Laser Light Source for Near-Field Optics and its Application to High-Density Optical Data Storage," Appl. Phys. Lett., **75**, pp. 1515–1517.
[5] Bethe, H., 1944, "Theory of Diffraction by Small Holes," Phys. Rev., **66**, pp. 163–182.
[6] Thio, T., Pellerin, K. M., Linke, R. A., Lezec, H. J., and Ebbesen, T. W., 2001, "Enhanced Light Transmission Through a Single Subwavelength Aperture," Opt. Lett., **26**(24), pp. 1972–1974.
[7] Raether, H., 1988, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer, Berlin, Germany.
[8] Shi, X., and Hesselink, L., 2002, "Mechanisms for Enhancing Power Throughput from Planar Nano-Apertures for Near-Field Optical Data Storage," Jpn. J. Appl. Phys., Part 1, **41**(3B), pp. 1632–1635.
[9] Tanaka, K., and Tanaka, M., 2003, "Simulation of an Aperture in the Thick Metallic Screen that Gives High Intensity and Small Spot Size Using Surface Plasmon Polariton," J. Microsc., **210**, pp. 294–300.
[10] Jin, E. X., and Xu, X., 2004, "FDTD Studies on Optical Transmission Through Planar Nano-Apertures in a Metal Film," Jpn. J. Appl. Phys., Part 1, **43**(1), pp. 407–417.
[11] Jin, E. X., and Xu, X., 2005, "Radiation Transfer Through Nanoscale Apertures," J. Quant. Spectrosc. Radiat. Transf., **93**, pp. 163–173.
[12] Sendur, K., and Challener, W., 2003, "Near-Field Radiation of Bow-Tie Antennas and Apertures at Optical Frequencies," J. Microsc., **210**, pp. 279–283.
[13] Jin, E. X., and Xu, X., 2005, "Obtaining Super Resolution Light Spot Using Surface Plasmon Assisted Sharp Ridge Nanoaperture," Appl. Phys. Lett., **86**, pp. 111106-1–111106-3.
[14] Matteo, J. A., Fromm, D. P., Yuen, Y., Schuck, P. J., Moerner, W. E., and Hesselink, L., 2004, "Spectral Analysis of Strongly Enhanced Visible Light Transmission Through Single C-Shaped Nanoapertures," Appl. Phys. Lett., **85**, pp. 648–650.
[15] Chen, F., Itagi, A., Bain, J. A., Stancil, D. D., Schlesinger, T. E., Stebounova, L., Walker, G. C., and Akhremitchev, B. B., 2003, "Imaging of Optical Field Confinement in Ridge Waveguides Fabricated on Very-Small-Aperture Laser," Appl. Phys. Lett., **83**, pp. 3245–3247.
[16] Courjon, D., 2003, *Near-Field Microscopy and Near-Field Optics*, Imperial College Press, London, UK.
[17] Kunz, K., and Luebbers, R., 1996, *The Finite Difference Time Domain Method for Electromagnetics*, CRC, Boca Raton, FL.
[18] Taflove, A., 1995, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, Artech House, Boston, MA.
[19] Lide, D. R., 1996, *CRC Handbook of Chemistry and Physics*, CRC, Boca Raton, FL.
[20] Gordon, R., and Brolo, A. G., 2005, "Increased Cut-off Wavelength for a Subwavelength Hole in a Real Metal," Opt. Express, **13**(6), pp. 1933–1938.

Energy Transfer to Optical Microcavities With Waveguides

Zhixiong Guo[1]
Mem. ASME
e-mail: guo@jove.rutgers.edu

Haiyong Quan

Department of Mechanical and Aerospace Engineering,
Rutgers, The State University of New Jersey,
98 Brett Road,
Piscataway, NJ 08854

Micro/nanoscale radiation energy transfer is investigated in optical microcavity and waveguide coupling structures working on whispering-gallery mode optical resonances. The finite element method is employed for solving the Helmholtz equations that govern the energy transfer and time-harmonics electromagnetic (EM) wave propagation. The maximum element size concept is introduced for the numerically sensitive subdomains where local mesh refining is needed because of the presence of intensified EM fields. The results show that the energy storage capability of a resonant microcavity is predominantly determined by the cavity size. The stored energy in the 10 μm *diameter microcavity considered is several orders of magnitude larger than that in the* 2 μm *diameter microcavity. The gap between a microcavity and its light-delivery waveguide has a substantial effect on the energy coupling from the waveguide to the microcavity and consequently influences significantly energy storage in the microcavity. An optimal gap is found for maximum energy storage and most efficient energy coupling. This optimal gap dimension depends not only on the configurations of the microcavity and waveguide, but also on the resonance wavelength. With increasing gap the quality factor increases exponentially and quickly saturates as the gap approaches to one wavelength involved. The submicron/nanoscale gap is crucial for generating quality resonances as well as for efficient energy transfer and coupling.* [DOI: 10.1115/1.2401197]

Keywords: radiation energy transfer, energy storage, micro/nanoscale phenomena, optical microcavities, resonance quality, numerical simulation

Introduction

Optical microcavities store light in small volumes by resonant recirculation of photons [1,2]. Resonance quality is described by the cavity quality (Q) factor, which is proportional to the photon confinement time. $Q > 10^9$ has been observed at red and near-infrared wavelengths in fused-silica submillimeter particles [3,4]. Devices based on high-Q optical microcavities are already indispensable for a wide range of fundamental research and emerging applications [5]. For example, they are used in strong-coupling cavity quantum electrodynamics and applications to quantum information studies [6], in developing efficient microlasers that offer enhanced functionality [7], in optical communications [8,9], in high-resolution spectroscopy [10], and in miniature sensor technologies [11,12], to name a few.

Whispering-gallery mode (WGM) optical cavities are typically dielectric circular structures in which waves are confined by repeated total internal reflection (TIR) at the curvilinear boundary so that the electromagnetic (EM) field can close on itself at some specific frequencies, giving rise to resonances [2]. An ideal cavity would confine light indefinitely (i.e., without loss) and would have resonance frequencies at precise values (i.e., extremely narrow linewidth). In practice energy losses exist due to material absorption, scattering, or diffraction as a result of surface roughness or material inhomogeneity [2,13]. Ultimately the photon lifetime in a resonating cavity is limited, allowing the radiation eventually to leak out or be absorbed.

It is well known that WGM optical resonances are morphology dependent [1]. If the cavities are in the micrometer level, one obtains a very small mode volume and a high finesse. Microscale cavities ensure that the resonant frequencies are more sparsely distributed throughout the cavity-size-dependent resonant optical spectrum than they are in corresponding mesoscale cavities [5]. The demand of miniaturization in electronic and photonic devices has been rapidly increasing in recent years.

Different from Mie resonances in microspheres, which are observable by means of elastic and inelastic scattering of free-space beams, high-Q WGM resonances are not accessible by free-space beams and, therefore, require employment of near-field couplers. Numerous coupling devices, such as high-index prisms with frustrated total internal reflection [14], side-polished fiber couplers [15], optical fiber tapers [16], and waveguides [17,18], have been considered. The principle of all these devices is based on providing efficient energy transfer to the resonant EM waves in the resonator through the evanescent field of a guided wave or a TIR spot in the coupler. It is evident *a* priori that efficient coupling can be expected on fulfillment of two main conditions: phase synchronism and significant overlap of the two evanescent fields in the gap between the microcavity and the coupler. Thus, the nanometer to sub-micrometer gap that separates the microcavity and coupler is a critical parameter that affects the energy transfer and coupling efficiency.

However, little attention has been paid to the micro/nanoscale energy transfer in optical microcavity devices. It was difficult to control the small gap between a microcavity and a micro-coupler. For the sake of experimental measurement, many researchers [19] considered close contact between a resonator and its coupler. Recent advances in the technology of nanofabrication offer the possibility of manufacturing new optical devices with unprecedented control. It is now feasible to consider optical microcavities, light couplers, and coupling gaps having physical dimensions in nano- and micrometer levels. Dimensions of 100–200 nm are routinely achieved in the manufacturing of integrated circuits. Nanofabrication techniques allow the realization of semiconductor microcavities with evanescent wave coupling to microscale waveguides across nanoscale air gaps [10,18]. With high-quality etching, the scattering loss can be minimized to achieve simultaneously a high Q and a high finesse.

---

[1] Corresponding author.

Contributed by the Heat Transfer Division of ASME for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received February 6, 2006; final manuscript received July 21, 2006. Review conducted by M. Pinar Menguc. Paper presented at the 2005 ASME International Mechanical Engineering Congress (IMECE2005), November 5–11, 2005, Orlando, Florida, USA.

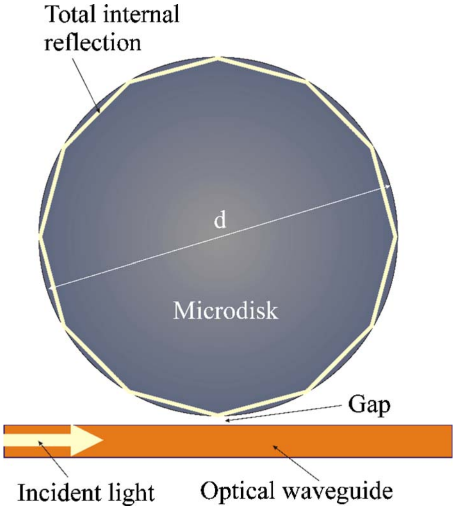


Fig. 1 Sketch of an optical microcavity coupled with a light-delivery waveguide

Radiative energy transfer, which involves photon transport and EM wave interaction with matter, is an important heat transfer mechanism. Thermal radiation transfer based on the radiative transfer equation is well summarized in comprehensive textbooks [20,21]. Several specific criteria have been established by some excellent reviews [22–25] to delineate the micro/nanoscale and the meso/macroscale radiation regimes. Maxwell's EM theory [26] is commonly utilized to describe radiation energy exchange in micro/nanometer levels [27]. The increasing demand for small structures and devices opens up many new opportunities and challenges in micro/nanoscale heat transfer [28–30].

In this manuscript, we will study the micro/nanoscale radiation energy transfer in optical microcavities coupled with light-delivery waveguides. Maxwell's equations that govern the energy exchange and EM wave propagation in the microstructures will be solved via the finite element method. Emphasis will be placed on investigating the characteristics of energy transfer and coupling from the waveguides to the microcavities, the energy storage capacity of resonant microcavities, and the quality of resonances.

Mathematical Formulation

Consider an optical microcavity coupled with a light-delivery waveguide as shown in Fig. 1. A small air gap separates the microcavity from the waveguide. A laser beam is focused into the microscale waveguide and the light will be confined inside the waveguide. If the refractive index of the microcavity is very close to that of the waveguide, however, photons will tunnel from the waveguide to the microcavity. The photons in the microcavity are confined by TIR at the curvilinear boundary when the refractive index of the microcavity is larger than that of the surrounding medium such as air. After repeated TIRs, if the EM wave returns to its starting point in phase, WGM resonance occurs. Under resonances, the energy transfer and coupling from the waveguide to the resonator become very powerful.

The microcavity-waveguide device is made of a dielectric material, which is assumed to be linear and isotropic. The EM wave propagation and photon tunneling in the device can be described by the time-dependent Maxwell equations [26] as

$$\nabla\cdot\bar{E}=\frac{\rho}{\varepsilon},\quad \nabla\times\bar{E}=-\mu\frac{\partial\bar{H}}{\partial t}$$
$$\nabla\cdot\bar{H}=0,\quad \nabla\times\bar{H}=\bar{J}+\varepsilon\frac{\partial\bar{E}}{\partial t}\tag{1}$$

For time-harmonic waves, $\bar{E}(\bar{r},t)=\bar{E}(\bar{r})e^{i\omega t}$. The Maxwell equations are then simplified to Helmholtz equations as follows:

$$\frac{1}{\mu}\nabla^2\bar{E}+\omega^2\varepsilon_c\bar{E}=0$$
$$\frac{1}{\mu}\nabla^2\bar{H}+\omega^2\varepsilon_c\bar{H}=0\tag{2}$$

in which the complex permittivity [21] is introduced as

$$\varepsilon_c=\varepsilon_{cr}\cdot\varepsilon_0=\varepsilon-i(\sigma/\omega)\tag{3}$$

It is customary to introduce the complex index of refraction, $m=n-ik$. The relationship between ε_{cr} and m is expressed by

$$\varepsilon_{cr}=m^2=n^2-k^2-i2nk\tag{4}$$

Under WGM resonances the EM field in a cylindrical microcavity typically consists of equatorial brilliant rings. The rings are located on the same plane as the waveguide. Further, the considered manufacturable microdisk/waveguide devices have a planar structure. So, it is feasible to use a two-dimensional theoretical model. In the present calculations we consider the in-plane transverse electric (TE) waves, where the electric field has only a z-component and it propagates in the x-y plane. Thus, the fields can be written as:

$$\bar{E}(x,y,t)=E_z(x,y)\bar{e}_z e^{i\omega t}$$
$$\bar{H}(x,y,t)=[H_x(x,y)\bar{e}_x+H_y(x,y)\bar{e}_y]e^{i\omega t}\tag{5}$$

At the interface and physical boundaries, the natural continuity condition is used, i.e.,

$$\bar{n}\times(\bar{E}_1-\bar{E}_2)=0\quad\text{and}\quad\bar{n}\times(\bar{H}_1-\bar{H}_2)=0\tag{6}$$

When solving for $\bar{E}$, $\bar{n}\times\bar{H}=0$ is the natural continuity condition for the tangential component of the magnetic field.

For the boundaries of the calculation domain, the low-reflecting boundary condition is adopted. The low-reflecting condition means that only a small part of the wave is reflected, and that the wave propagates through the boundary almost as if it were not present. This condition can be formulized as

$$\bar{e}_z\cdot\bar{n}\times\sqrt{\mu}\bar{H}+\sqrt{\varepsilon}E_z=0\tag{7}$$

The light source term E_{0z}, which propagates inwards through the entry of the waveguide, can be treated as an electrically low-reflecting boundary expressed by

$$\bar{e}_z\cdot\bar{n}\times\sqrt{\mu}\bar{H}+\sqrt{\varepsilon}E_z=2\sqrt{\varepsilon}E_{0z}\tag{8}$$

In the present computations, E_{0z} is assumed to be uniform and unity at the entry of the waveguide.

The quality factor Q is defined as a ratio of 2π stored energy to energy lost per cycle. It can be calculated by [26]

$$Q=\omega_0/\Delta\omega=\omega_0\tau\tag{9}$$

where ω_0 is the resonant central frequency, $\Delta\omega$ is the linewidth (full-width at half maximum) of the cavity resonance, and τ is the photon lifetime.

According to Poynting's theorem [26]:

$$\int_V\nabla\cdot\bar{S}\,dV=-\int_V\left(\bar{E}\cdot\frac{\partial\bar{D}}{\partial t}+\bar{H}\cdot\frac{\partial\bar{B}}{\partial t}\right)dV-\int_V\bar{J}\cdot\bar{E}\,dV\tag{10}$$

The quantity $\bar{S}=\bar{E}\times\bar{H}$ is called the Poynting vector, which has the same dimensions as the heat flux and a direction pointing to the direction of wave propagation. The term on the left-hand side of

Eq. (10) represents the radiative losses. The first integral on the right-hand side represents the rate of change in total energy. The second integral on the right-hand side represents the resistive losses that result in heat dissipation.

The total energy density can be denoted by

$$u = \frac{1}{2}(\bar{E} \cdot \bar{D} + \bar{B} \cdot \bar{H}) \tag{11}$$

For dielectric materials, however, $\bar{J}=0$. Since the volume V is arbitrary, Eq. (10) can be simplified and cast into a differential continuity equation [26]:

$$\frac{\partial u}{\partial t} + \nabla \cdot \bar{S} = 0 \tag{12}$$

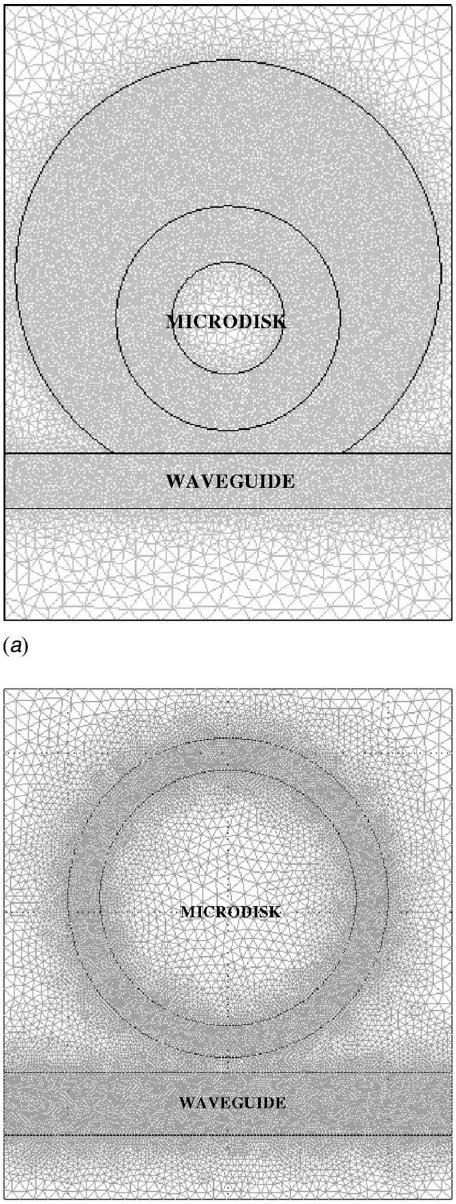
Numerical Models and Scheme

Versatile numerical approaches including the finite difference time-domain method [31] and the finite element method (FEM) can be employed for solving the above mathematical models. The FEM is very flexible in the treatment of irregular configurations such as the circular geometry of the microcavity. Thus, the in-plane TE waves application mode of the commercial FEMLAB package (version 3.0) is employed for the finite element analysis in the current simulations. The details of the numerical scheme are not repeated here. Interested readers should refer to the manual of FEMLAB or our recent publications [12,32].

In the present study, three sets of microdisk and waveguide coupling systems are considered: (1) a microdisk 2 μm in diameter is coupled with a waveguide 0.5 μm in width; (2) a microdisk 5 μm in diameter is coupled with a waveguide 1.0 μm in width; and (3) a microdisk 10 μm in diameter is coupled with a waveguide 2 μm in width. The waveguides are straight. All the microdisks and waveguides are made of the same material (silicon nitride) and are assumed to have a constant refractive index of 2.01 against the excitation wavelengths (600–850 nm) and to be lossless. The surrounding medium is air. The gap is defined as the narrowest width between the microdisk and the waveguide. The gap varies between 0 (in close contact) and 1000 nm for each set of optical microcavity systems in order to study the gap effects.

The cavity system is modeled as a rectangular simulation domain. The dimensions of the simulation domains are $4.0 \times 5.5\ \mu\text{m}^2$, $8.0 \times 8.7\ \mu\text{m}^2$, and $14 \times 16\ \mu\text{m}^2$ for the 2, 5, and 10 μm diameter microdisk systems, respectively. The simulation domains are meshed by many triangle elements generated automatically by the FEMLAB in which a mesh gradient approach is adopted to deal with abrupt changes in sensitive areas like the vicinity around the periphery of the microdisk, the small waveguide, and the gap region. However, such an approach was not very satisfactory because a strong resonant EM field exists inside a ring close to the cavity periphery. To meet with this requirement, we divide the microdisks into two regions and use hierarchical meshing to scale the cavity down to two different spatial levels. The meshes in the ring region inside the microdisk periphery are locally refined. We have used such a meshing approach in all our previous studies where the microcavity size was not smaller than 10 μm in diameter.

Exemplary meshes for the 2 and 10 μm diameter microdisk simulation domains are shown in Fig. 2. It is seen that an outer ring surrounding the 2 μm diameter microdisk is considered. This outer ring is not physical. It is for the purpose of local mesh refining. As the microcavity shrinks, its curvature becomes large. Hence, energy "leakage" from the cavity surface due to scattering and diffraction increases. It is necessary to add an outer ring for local mesh refining for small microdisks such as the 2 and 5 μm diameter microdisks. The results in Fig. 5 in the Results and Discussion section will confirm the necessity.



(b)

Fig. 2 Finite element meshes for the simulation domains of (*a*) the 2 μm diameter microdisk system and (*b*) the 10 μm diameter microdisk system, respectively

The numerical accuracy and sensitivity is examined using the 2 μm diameter microdisk system. The considered gap is fixed at 100 nm. The concept of the maximum element size $d_{\max}$ is introduced for the numerically sensitive subdomains including the inner ring, the outer ring, the photon tunneling gap, and the light-delivery waveguide. Local mesh refining in these subdomains is needed. The size of any meshed element in these sensitive subdomains should not be larger than $d_{\max}$. In general, the number of elements generated depends on the $d_{\max}$ value used. Smaller $d_{\max}$ creates more elements and requires larger computer memory and longer CPU time. As a reward, it leads to more accurate simulation results. However, there should be a trade-off among memory consumption, CPU time, and calculation accuracy. In the present calculations, we used a DELL PC equipped with one 2.8 GHz CPU and 2.0 GB memory. The FEMLAB (version 3.0) was found to be able to utilize up to 1.5 GB memory under the Windows XP operation system.

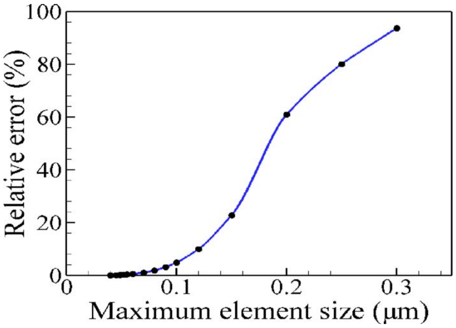


Fig. 3 The relative error of stored energy versus the maximum element size used in the computations

We found that when $d_{\max}$ equaled 0.04 μm, the FEMLAB could generate 104,460 elements for the simulation domain of the 2 μm diameter microdisk system, approaching to the memory limit of the PC used. The simulation results under this limit condition were the most accurate that we could obtain. In Fig. 3, the relative errors of the stored energy calculated using different $d_{\max}$ values are plotted as compared with the result in the limit condition. The 2 μm diameter microdisk is under the first-order resonance at a resonance mode with wavelength around 823 nm. It is evident that the computational error is less than 5% when $d_{\max}$ is below 0.1 μm ($\sim 1/8$ wavelength).

Figure 4 plots the curve of the normalized stored energy in the 2 μm diameter cavity versus the element number used in the simulation domain. It is seen that the normalized energy storage calculated using more than 30,000 elements converges to a constant value at unity. Thus, the calculation convergence is excellent.

CPU time consumption is basically not an issue for an individual run of simulation in the present computations because each run only took several minutes in the PC. In order to get a precise resonance frequency, however, dozens of simulation runs are needed. Such computational procedures must repeat for each set of the optical microcavity systems with each individual gap value and at each resonant mode. The whole simulation procedure is very time consuming.

Results and Discussion

Figure 5 illuminates the distributions of the electric field under a first-order WGM resonance for the three different optical microcavity systems. Figure 6 displays the corresponding energy density distributions. The electric fields are time-varying harmonic waves. The total energy densities are time averaged. The 2 μm diameter microdisk is coupled with a 0.5 μm wide waveguide through a 150 nm gap; the 5 μm diameter microdisk is coupled with a 1.0 μm wide waveguide through a 300 nm gap, while the 10 μm diameter microdisk is coupled with a 2.0 μm wide waveguide through a 500 nm gap. The resonance wavelengths are at 822.75, 807.735, and 801.1165 nm, respectively, for 2, 5, and 10 μm diameter microdisk devices.

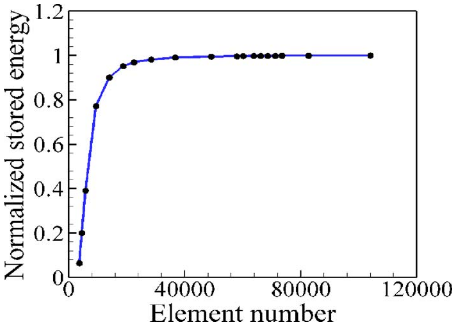


Fig. 4 The stored energy versus the element number used in the computations

It is seen that a bright resonant ring with intensified electric field or energy density is formed inside the periphery of the microdisks. Thus, enhanced photon tunneling from the waveguides to the microcavities occurred under resonances. The majority energy is concentrated in the resonant rings. The ratios of the energy storing in the microdisk to that passing through the waveguide are 1.88, 49.43, and 938.44 for the 2, 5, and 10 μm in-diameter microdisks, respectively. The energy storage capability soars with the increasing of the microcavity size because of increased volume and reduced energy leakage. One can see the obvious energy leakage from the 2 μm diameter microdisk in Fig. 5(*a*). The effect of diffraction grows as the cavity shrinks. Comparing Figs. 6(*a*)–6(*c*), it is found that the energy density in the resonant ring also increases as the cavity size increases. Under the considered conditions, the energy density in the resonant 10 μm diameter microdisk is about three orders of magnitude larger than that in the resonant 2 μm diameter microdisk.

Since the energy density is proportional to the square of the amplitude of the time-harmonic electric field, the contrast between the microcavity and the waveguide is stronger for energy density distributions than for electric field distributions. In order to see clearly the wave propagation and photon tunneling in the waveguide and gap, we use the harmonic E-field distributions to show the gap influence on the energy transfer in Fig. 7 for the 10 μm diameter microdisk system. The gaps are 250 nm in Fig. 7(*a*) and 750 nm in Fig. 7(*b*), respectively. The E-field for the same microdisk system with a 500 nm gap is already shown in Fig. 5(*c*). The resonant mode is at 801 nm wavelength. The propagation of the E-field in the waveguides in Figs. 7(*a*) and 7(*b*) is visible. The amplitude of the intensified electric field in the resonant rings depends on the gap value. It is seen that the EM field in the 10 μm diameter cavity with the 500 nm gap is much stronger than those with the 250 and 750 nm gaps. In particular, the electric field in the waveguide in Fig. 5(*c*) is extremely weak and nearly invisible. This means that the energy is almost completely coupled into the cavity. Thus, the gap dimension does substantially affect the energy transfer. An optimal gap for efficient energy transfer and coupling may exist!

The stored energy inside a microcavity is integrated as the main variable used to investigate energy transfer and coupling efficiency. By drawing a curve of the stored energy against the change of gap width, one can find an optimal gap for maximum energy transfer and most efficient energy coupling. Figures 8(*a*) and 8(*b*) show the profiles of the stored energy versus the gap variation for the 10 and 2 μm diameter microdisk systems, respectively. The simulated results are represented by the discrete symbols, while the continuous curves are fitted results. Two resonant modes for each cavity system are considered. For the 10 μm diameter microdisk system, the resonance modes are at 801 nm (near infrared light) and 608 nm (yellow light). For the 2 μm diameter microdisk system, the resonance modes are at 823 nm (near infrared light) and 605 nm (yellow light). As previously revealed, the capability of energy storage is predominantly determined by the cavity size. The stored energy in Fig. 8 is normalized by the respectively maximum stored energy for each individual resonant mode of the cavity. It is seen that the width of the curves of energy storage versus gap in Fig. 8(*a*) is narrower than that of the curves in Fig. 8(*b*). This suggests that the energy coupling in the 10 μm diameter microdisk system is more sensitive to the gap dimension as compared with the 2 μm diameter microdisk system.

From Fig. 8, it is clearly seen that an optimal gap dimension exists, which depends not only on the cavity size, but also on the resonance mode (wavelength). The optimal gap values can be estimated from the fitted curves in Fig. 8 and are listed in Table 1. Those optimal gap widths represent the most efficient energy cou-

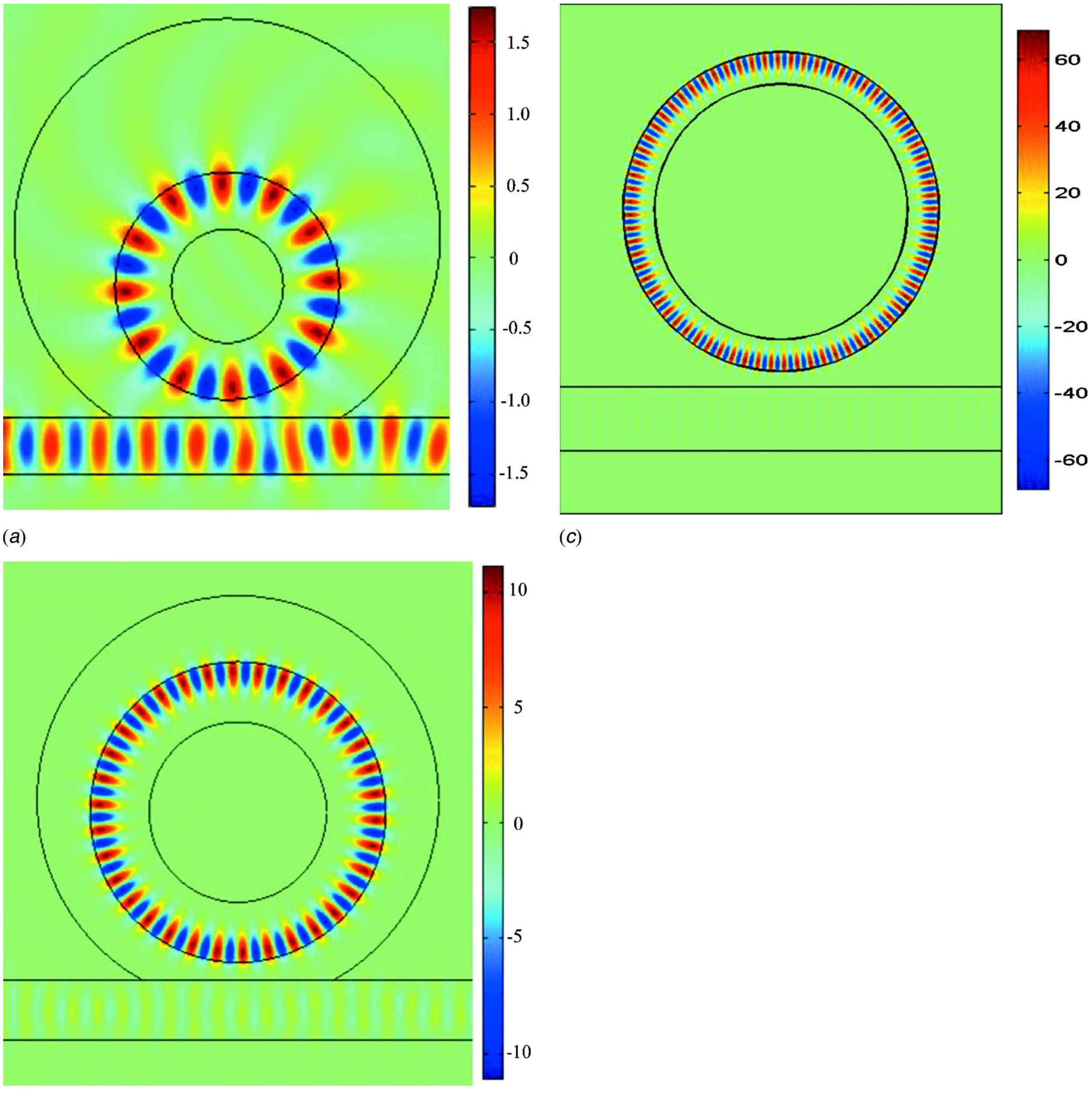
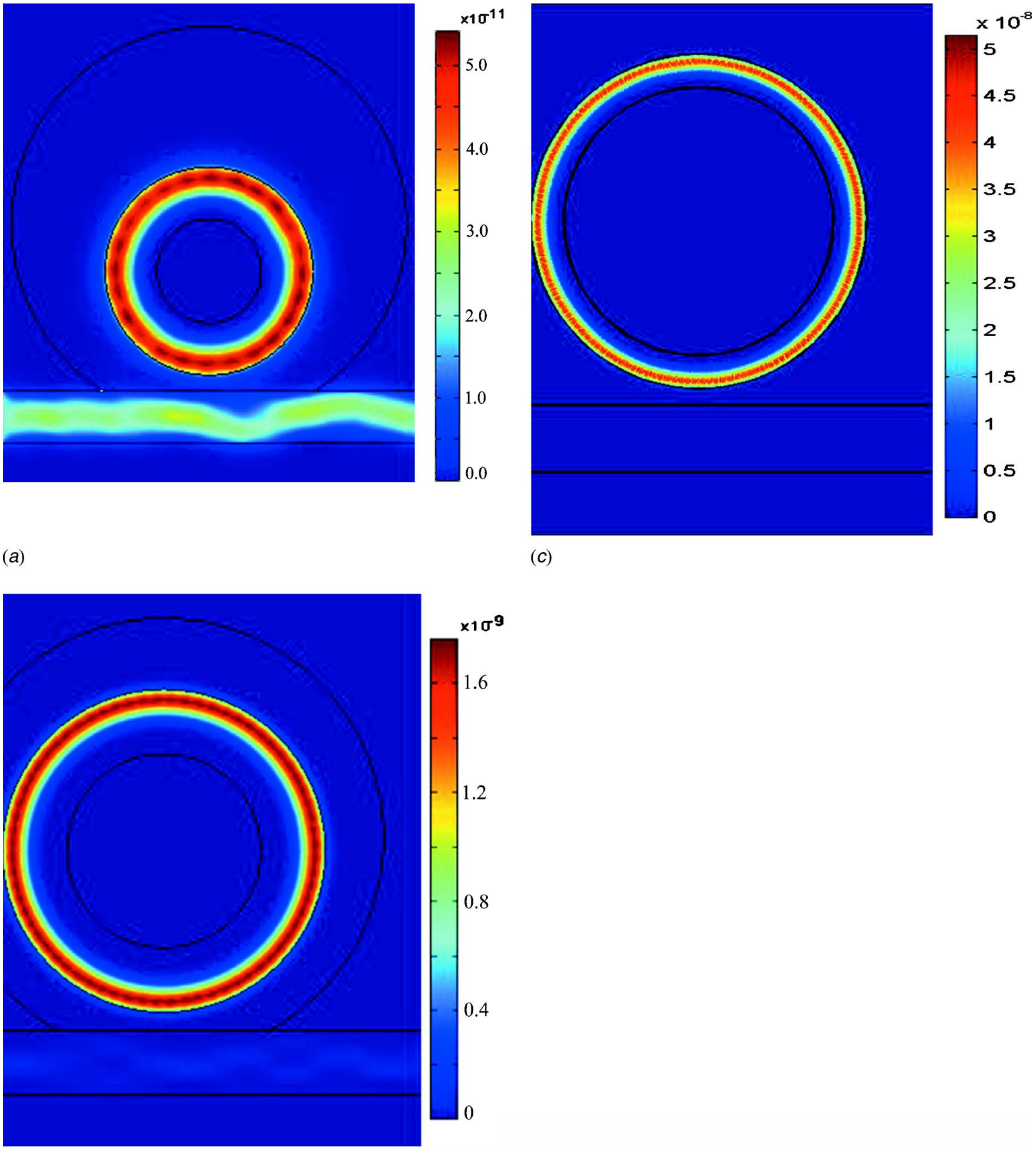


Fig. 5 The harmonic electric field distributions in the three optical microcavity systems: (*a*) the 2 μm diameter microdisk system (λ=822.735 nm, *g*=150 nm), (*b*) the 5 μm diameter microdisk system (λ=807.735 nm, *g*=300 nm), and (*c*) the 10 μm diameter microdisk system (λ=801.1165 nm, *g*=500 nm)

pling in the respective cavity/waveguide coupling systems at the respective resonance modes (represented by wavelength).

It is easy to understand the reduction of energy transfer and coupling efficiency at gaps larger than the optimal gap, because it is well known that the strength of the evanescent field from a surface decays exponentially as a function of the distance to the surface. The longer the distance, the weaker is the strength. How can we understand the reduction of the energy coupling efficiency for gaps smaller than the optimal gap? From the viewpoint of the evanescent strength, a smaller gap has a stronger strength and closer overlapping of the two evanescent fields in the gap from the resonator and the waveguide, respectively. A smaller gap affords more opportunities to photon tunneling. Indeed, because of the enhanced photon tunneling, photons confined in the cavity tunnel back to the waveguide. This reverse energy transfer tends to reduce the energy coupling from the waveguide to the cavity. Gorodetsky and Ilchenko [14] experimentally observed a dip in the output intensity of a prism coupler in a fused-silica sphere with increasing gap from in-close contact. This provided experimental evidence for the reverse energy transfer analysis. Therefore, an optimal gap for most efficient energy coupling exists as a result of bidirectional energy transfer.

The gap effects on the resonance linewidth and resonance quality are presented in Fig. 9 for the 2 and 10 μm diameter microdisk systems. The resonant mode is at 801 nm for the 10 μm diameter microdisk system and at 823 nm for the 2 μm diameter microdisk system. With increasing gap from zero (in close contact) to order of one optical wavelength of interest, the full-width at half maximum (FWHM) of the resonance band narrows and the Q factor increases. It is seen that the Q factor increases exponentially before the gap reaches to the optimal gap value, then slows down around the optimal gap range, and quickly reaches to an asymptotic limit (limit Q factor) when the gap approaches to the optical wavelength. As revealed in Fig. 8, however, the resonant energy in the cavity is extremely weak when the gap is around one



(*a*)

(*c*)

(*b*)

Fig. 6 The time-averaged energy density distributions in the three optical microcavity systems: the 2 μm diameter microdisk system (λ=822.735 nm, *g*=150 nm), (*b*) the 5 μm diameter microdisk system (λ=807.735 nm, *g*=300 nm), and (*c*) the 10 μm diameter microdisk system (λ=801.1165 nm, *g*=500 nm)

optical wavelength. Simultaneously considering Figs. 8 and 9, a trade-off exists between the requirements of energy coupling efficiency and Q factor. Depending on applications, an optimum gap may be selected between the optimal gap value for maximum energy coupling and one optical wavelength for maximum limit Q factor. Comparing Fig. 9(*a*) with Fig. 9(*b*), it is seen that the maximum Q value for the 10 μm diameter microdisk system is about four orders of magnitude higher than that for the 2 μm diameter microdisk system. Hence, the limit Q factor is predominantly determined by the cavity size. It falls rapidly as the cavity size shrinks.

Conclusions

The whispering-gallery mode optical resonances and energy transfer were simulated for the three dielectric microdisk/waveguide coupling planar structure systems with various submicron/nanoscale air gaps. The simulations were performed using the finite element analysis based on the FEMLAB. Accurate results are obtained when appropriate meshes are adopted. The computational error is found to be less than 5%, when the maximum element size is below 1/8 of the wavelength involved in the

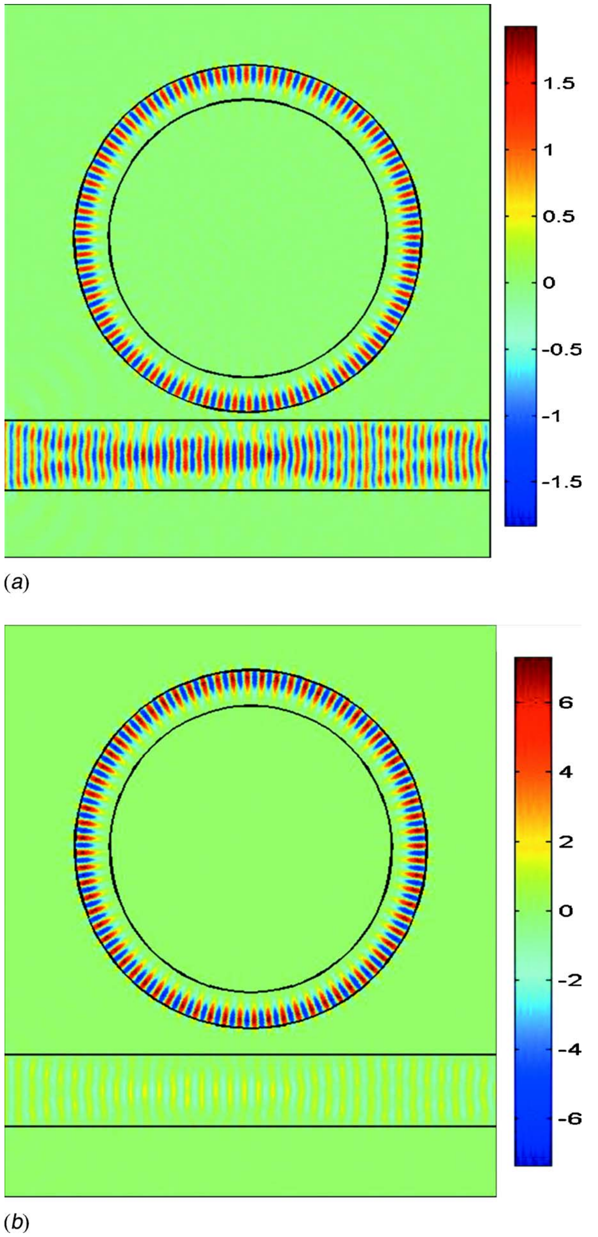


Fig. 7 The harmonic electric field distributions for the 10 μm diameter microdisk system resonating at the resonance mode of 801 nm with two different gap values: (*a*) 250 nm gap and (*b*) 750 nm gap

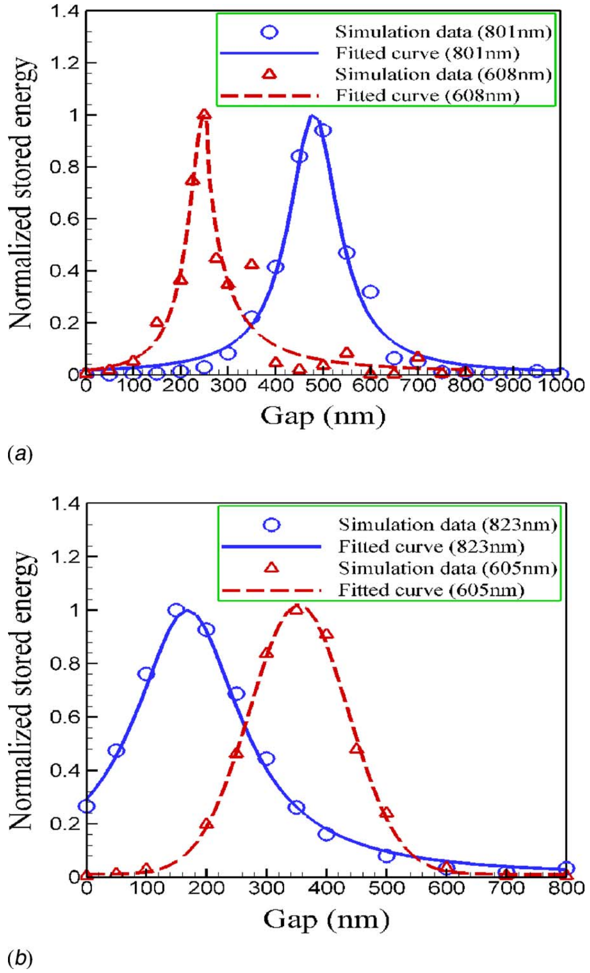


Fig. 8 The stored energy versus the gap variation: (*a*) the 10 μm diameter microdisk system at two different resonance modes and (*b*) the 2 μm diameter microdisk system at two different resonance modes

computationally sensitive regions such as the gap, the inner and/or outer rings close to the cavity periphery, and the waveguide. The convergence of the computations is excellent.

The energy transfer and coupling from the light-delivery waveguide to the microcavity are enhanced under resonances. A bright ring with intensified EM field and energy density is formed inside the periphery of the microdisk. The energy density in the resonant ring is a strong function of the cavity size. With increasing cavity size, both the energy storage capability and the limit Q factor soar rapidly. The stored energy as well as the limit Q factor is several orders of magnitude higher in the 10 μm diameter microdisk system than in the 2 μm diameter microdisk system.

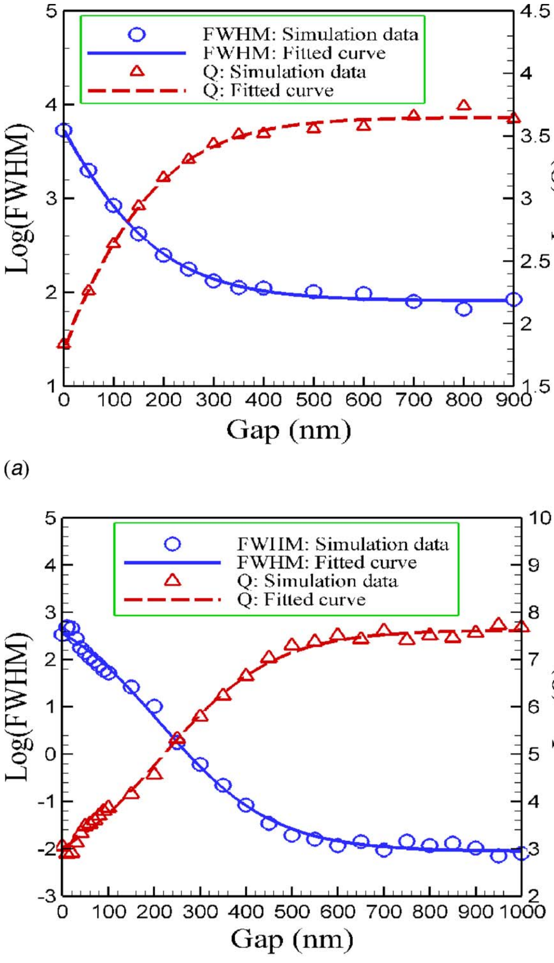
The plots of the stored energy inside the microcavities against the gap width variation were obtained. An optimal gap dimension in which the energy storage in the cavity is the maximum (and thus the energy coupling is the most efficient) can be obtained from the plot. The optimal gap is a complicated function of the microdisk/waveguide configurations and the mode of resonance. It is also found that the energy transfer and coupling in a larger microcavity is more sensitive to the gap dimension. With increasing gap, the Q factor increases, while the FWHM of the resonant band decreases. The Q factor increases exponentially as the gap width increases from zero to the optimal gap value, and then quickly approaches to an asymptotic limit (the maximum limit Q factor) with the gap further increasing to one optical wavelength of interest.

Acknowledgment

Z. Guo is grateful to the support of National Science Foundation Grant No. (CTS-0541585) and the Academic Excellence Fund Award from Rutgers University to the project.

Table 1 Optimal gap dimensions

Cavity diameter/ waveguide width	Resonance mode (wavelength) (nm)	Optimal gap width (nm)
2 μm/0.5 μm	605	450
	823	180
10 μm/2 μm	608	260
	801	480



(*a*)

(*b*)

Fig. 9 The gap effects on the FWHM (unit: GHz) of the resonance band and the Q factor: (*a*) the 2 μm diameter microdisk system (at 823 nm resonance mode)and (*b*) the 10 μm diameter microdisk system (at 801 nm resonance mode)

Nomenclature

$\bar{B}$ = magnetic induction
c = speed of light
d = diameter of microcavity
$\bar{D}$ = electric displacement
$\bar{e}$ = unit vector in a direction
$\bar{E}$ = electric field vector
g = gap
$\bar{H}$ = magnetic field vector
$\bar{J}$ = electric current density
k = absorption index
m = complex index of refraction
n = real part of the refractive index
$\bar{n}$ = unit vector in normal direction
Q = cavity quality factor
$\bar{S}$ = Poynting vector
t = time
u = total energy density
V = computational volume
x,y,z = coordinates
ε = permittivity
ε_c = complex permittivity
ε_{cr} = relative complex permittivity
λ = wavelength
μ = permeability
ρ = electric charge density
σ = electrical conductivity
τ = photon lifetime
ω = angular frequency
$\Delta\omega$ = resonance linewidth

Subscripts

0 = in vacuum or at central position

References

[1] Hill, S. C., and Benner, R. E., 1988, "Morphology-Dependent Resonances," in *Optical Effects Associated with Small Particles*, P. W. Barber and P. K. Chang, eds., World Scientific, Singapore, pp. 3–61.
[2] Arnold, S., 2001, "Microspheres, Photonic Atoms and the Physics of Nothing," Am. Sci., **89**(5), pp. 414–421.
[3] Collot, L., Lefèvre-Seguin, V., Brune, M., Raimond, J. M., and Haroche, S., 1993, "Very High-Q Whispering Gallery Modes Resonances Observed on Fused Silica Microspheres," Europhys. Lett., **23**(5), pp. 327–333.
[4] Gorodetsky, M. L., Savchenkov, A. A., and Ilchenko, V. S., 1996, "Ultimate Q of Optical Microsphere Resonators," Opt. Lett., **21**(7), pp. 453–455.
[5] Vahala, K. J., 2003, "Optical Microcavities," Nature (London), **424**(6950), pp. 839–846.
[6] Mabuchi, H., and Doherty, A. C., 2002, "Cavity Quantum Electrodynamics: Coherence in Context," Science, **298**(5597), pp. 1372–1377.
[7] Cai, M., Painter, Q., Vahala, K. J., and Sercel, P. C., 2000, "Fiber-Coupled Microsphere Laser," Opt. Lett., **25**(19), pp. 1430–1432.
[8] Blom, F. C., van Dijk, D. R., Hoekstra, H. J., Driessen, A., and Popma, T. J. A., 1997, "Experimental Study of Integrated-Optics Micro-Cavity Resonators: Toward an All-Optical Switching Device," Appl. Phys. Lett., **71**(6), pp. 747–749.
[9] Little, B. E., Chu, S. T., Haus, H. A., Foresi, J., and Laine, J. P., 1997, "Microring Resonator Channel Dropping Filters," J. Lightwave Technol., **15**(6), pp. 998–1005.
[10] Krioukov, E., Klunder, D. J. W., Driessen, A., Greve, J., and Otto, C., 2002, "Integrated Optical Microcavities for Enhanced Evanescent-Wave Spectroscopy," Opt. Lett., **27**(17), pp. 1504–1506.
[11] Vollmer, F., Braun, D., Libchaber, A., Khoshsima, M., Teraoka, I., and Arnold, S., 2002, "Protein Detection by Optical Shift of a Resonant Microcavity," Appl. Phys. Lett., **80**(21), pp. 4057–4059.
[12] Quan, H., and Guo, Z., 2005, "Simulation of Whispering-Gallery-Mode Resonance Shifts for Optical Miniature Biosensors," J. Quant. Spectrosc. Radiat. Transf., **93**(1–3), pp. 231–243.
[13] Dubreuil, N., Knight, J. C., Leventhal, D. K., Sandoghdar, V., Hare, J., and Lefèvre, V., 1995, "Eroded Monomode Optical Fiber for Whispering-Gallery Mode Excitation in Fused-Silica Microspheres," Opt. Lett., **20**(8), pp. 813–815.
[14] Gorodetsky, M. L., and Ilchenko, V. S., 1999, "Optical Microsphere Resonators: Optimal Coupling to High-Q Whispering-Gallery Modes," J. Opt. Soc. Am. B, **16**(1), pp. 147–154.
[15] Serpenguzel, A., Arnold, S., and Griffel, G., 1995, "Excitation of Resonances of Microspheres on an Optical Fiber," Opt. Lett., **20**(4), pp. 654–656.
[16] Knight, J. C., Cheung, G., Jacques, F., and Birks, T. A., 1997, "Phase-Matched Excitation of Whispering Gallery Mode Resonances Using a Fiber Taper," Opt. Lett., **22**(15), pp. 1129–1131.
[17] Klunder, D. J. W., Krioukov, E., Tan, F. S., van Der Veen, T., Bulthuis, H. F., Sengo, G., Otto, C., Hoekstra, H. J. W. M., and Driessen, A., 2001, "Vertically and Laterally Waveguide-Coupled Cylindrical Microresonators in Si_3N_4 on SiO_2 Technology," Appl. Phys. B, **73**(5–6), pp. 603–608.
[18] Guo, Z., Quan, H., and Pau, S., 2005, "Optical Resonance in Fabricated Whispering-Gallery Mode Microcavity," ASME J. Heat Transfer, **127**(8), p. 808.
[19] Arnold, S., 2005, private communications, Polytechnic University, Brooklyn, NY.
[20] Siegel, R., and Howell, J. R., 2001, *Thermal Radiation Heat Transfer*, 4th ed., Taylor & Francis, New York.
[21] Modest, M. F., 2003, *Radiative Heat Transfer*, 2nd ed., Academic Press, New York.
[22] Kumar, S., 1994, "Thermal Radiation Transport in Micro-Structures," Therm. Sci. Eng., **2**(2), pp. 149–157.
[23] Tien, C. L., Qiu, T. Q., and Norris, P. M., 1994, "Microscale Thermal Phenomena in Contemporary Technology," Therm. Sci. Eng., **2**(1), pp. 1–11.
[24] Chen, G., 1996, "Heat Transfer in Micro- and Nanoscale Photonic Devices," in *Annual Review of Heat Transfer*, C. L. Tien, ed., CRC, Boca Raton, FL, Vol. VII, pp. 1–57.
[25] Longtin, J. P., and Tien, C. L., 1997, "Microscale Radiation Phenomena," in *Miscroscale Energy Transfer*, C. L. Tien, A. Majumdar, and F. M. Gerner, eds., Taylor and Francis, Washington, DC, pp. 119–147.
[26] Jackson, J. D., 1998, *Classical Electrodynamics*, 3rd ed., Wiley, New York.
[27] Chen, G., 1997, "Wave Effects on Radiative Transfer in Absorbing and Emitting Thin-Film Media," Microscale Thermophys. Eng., **1**(3), pp. 215–224.

[28] Majumdar, A., 2004, “Thermoelectricity in Semiconductor Nanostructures,” Science, **303**(5659), pp. 777–778.

[29] Grigoropoulos, C. P., 1994, “Heat Transfer in Laser Processing of Thin Films,” *Annual Review of Heat Transfer*, Vol. V, C. L. Tien, ed., CRC, Boca Raton, FL, pp. 77–130.

[30] Chen, S. C., Cahill, D. G., and Grigoropoulos, C. P., 2000, “Melting and Surface Deformation in Pulsed Laser Surface Micro-Modification of NiP Disks,” ASME J. Heat Transfer, **122**(1), pp. 107–112.

[31] Taflove, A., and Hagness, S. C., 2000, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 2nd ed., Artech House, Norwood, MA.

[32] Guo, Z., Quan, H., and Pau, S., 2006, “Numerical Characterization of Whispering-Gallery Mode Optical Microcavities,” Appl. Opt., **45**(4), pp. 611–618.

Nanostructuring Borosilicate Glass With Near-Field Enhanced Energy Using a Femtosecond Laser Pulse

Alex Heltzel

Arvind Battula

J. R. Howell

Shaochen Chen

Department of Mechanical Engineering,
The University of Texas at Austin,
Austin, TX 78712

A model based on the evolution of electron density derived from the Fokker-Planck equation has been built to describe ablation of dielectrics during femtosecond laser pulses. The model is verified against an experimental investigation of borosilicate glass with a 200 fs *laser pulse centered at* 780 nm *wavelength in a range of laser energies. The ablation mechanisms in dielectrics include multi-photon ionization (MPI) and avalanche ionization. MPI dominates the ionization process during the first stages of the laser pulse, contributing seed electrons which supply avalanche ionization. The avalanche process initiates and becomes responsible for the majority of free-electron generation. The overall material removal is shown to be highly dependent upon the optical response of the dielectric as plasma is formed. The ablation model is employed to predict the response of borosilicate glass to an enhanced electromagnetic field due to the presence of microspheres on the substrate surface. It is shown that the diffraction limit can be broken, creating nanoscale surface modification. An experimental study accompanies the model, with AFM and SEM characterizations that are consistent with the predicted surface modifications.* [DOI: 10.1115/1.2360595]

Keywords: femtosecond ablation, microsphere irradiation, surface modification, near field effects

Introduction

In the field of optoelectronics and micro-electronics fused quartz and related silicate glasses are very important materials due to their high transmission properties in the UV to IR range, excellent thermal properties, good electrical insulation and high chemical stability [1]. Also, borosilicate glass has an excellent anodic bonding property and surface integrity which makes it the usual substrate for micro-electro mechanical systems (MEMS). Therefore, in order to build a communication interface, micro/nanoholes free of micro-cracks, with good edge and surface quality as well as high aspect ratios need to be formed on the glass substrate. Using traditional micro-machining process for a micro-hole with a diameter below 200 μm is difficult because of the extreme hardness, brittleness, corrosion resistance, and melting temperature of glass. Also the conventional thermal and/or chemical machining methods create an excessive heat-affected zone. Mechanical machining methods also have limitations in productivity and accuracy [2]. Hence, to solve this problem various machining methods have been proposed like the combination of micro electrical-discharge machining and micro ultrasonic vibration machining [3], micro-abrasive jet machining (AJM) [2], laser-induced plasma-assisted ablation (LIPAA) [4], direct laser write laser process using short pulses [5], combination of nanoindenter and hydrofluoric acid wet etching [6], and hologram encoding system with femtosecond laser pulses [7], etc. Due to the various working mechanisms of these methods, the results produced are distinct. However, high-quality holes in nanoscale were not obtained using these techniques.

Laser induced ablation has several advantages over the conventional mechanical and/or chemical machining; it is a single-step process with very high flexibility, noncontact process, direct patterning without the photoresist process, good material removal rate and does not require any etchants. Femtosecond regime offers advantages over the nanosecond regime, by depositing energy into a material in a very short time period, before any thermal diffusion can occur. In this paper we demonstrate a technique wherein, the optical near-field effect is utilized to overcome the diffraction limit to nano-pattern hard-to-machine borosilicate glass using a femtosecond laser. In the first part of the paper we discuss the theoretical background for the study with numerical results. The following part will describe the experimental study where a good agreement between the theoretical and experimental results can be seen.

Theoretical Development

The essentially nonthermal process of femtosecond laser ablation makes it very attractive as a machining tool due to the clean features and negligible heat-affected zone. As a nonthermal process, standard heat transfer models must be discarded in favor of ideas that describe the complex photon energy absorption as well as the mechanisms that transport the energy between the electronic system and lattice system. Several studies have provided empirical data that make a theoretical investigation of ultrafast ablation of dielectrics possible [8–13].

Due to the wide energy gap between the valence and conduction bands of dielectrics, the conduction band is occupied by low electron densities within a large range of electron temperatures. Energy will not be transported by phonons through the dielectric material during a sub-picosecond laser pulse because the free electron thermal relaxation times are generally larger than the pulse duration, leaving lattice temperatures essentially unchanged. For these reasons, the popular two-temperature model [14,15] used for femtosecond laser heating of metals is invalid for dielectric materials or any possessing a sufficiently wide bandgap. To model laser heating in fused silica and borosilicate glass, a rela-

---

Contributed by the Heat Transfer Division of ASME for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received January 24, 2006; final manuscript received May 26, 2006. Review conducted by M. Pinar Menguc.

tion derived by Stuart et al. [9] is used, describing the evolution of free electron density in a dielectric medium exposed to intense laser radiation

$$\frac{dn_e(r,z,t)}{dt} = a_i I(r,z,t) n_e(r,z,t) + \delta_N (I(r,z,t))^N \tag{1}$$

where I is the spatial and temporally dependent intensity of the laser pulse, n_e is the electron density, a_i is the avalanche ionization coefficient, and δ_N is the cross-section of N-photon absorption.

The first term on the right-hand side (rhs) of Eq. (1) denotes the contribution to free electron generation due to avalanche ionization, a nonlinear process where a single charged particle initiates the ionization of others around it. The "seed" ions needed for the avalanche process to progress are formed by multi-photon ionization (MPI) quantified along with impact ionization in the second term of Eq. (1). MPI dominates the production of ions (and hence conduction band "free" electrons) in the first stages of a fs laser pulse [8]. After several fs, the avalanche process dominates, which leads to a drastic change in the optical properties of the material and is instrumental in the final surface modification.

Lenzner et al. [8] have experimentally determined the coefficients for fused silica: $a_i=4\pm0.6$ cm^2/J, $\delta_N=6\times10^{8\pm0.9}$ cm^{-3} ps^{-1}(cm^2/TW)^6 with $N=6$, and barium aluminum borosilicate (BBS): $a_i=1.2\pm0.4$ cm^2/J, $\delta_N=7\times10^{17\pm0.5}$ cm^{-3} ps^{-1}(cm^2/TW)^3 with $N=3$. A model has been proposed by Jiang and Tsai [9,10] based on Eq. (1) which has given ablation depths and crater shapes in good agreement with experiments. The model used for the present calculations draws heavily from that of [10] with exceptions to be noted during the discussion. The laser intensity within the material is described by

$$I(r,z,t) = \frac{2\cdot F_{\text{peak}}}{\sqrt{\pi/\ln 2}\cdot t_p}[1-R(r,t)] \cdot \exp\left[-\frac{2\cdot r^2}{r_0^2} - (4\ln 2)\frac{t^2}{t_p^2} - \int_0^z \alpha(r,z,t)dz\right] \tag{2}$$

where t_p is the laser pulse duration, r_0 is the radius of the irradiated area, and α is the absorption coefficient of the material. Equation (2) provides for a Gaussian distribution of energy with space and time. The peak laser fluence, F_{peak}, is given by

$$F_{\text{peak}} = \frac{2\cdot E_{\text{pulse}}}{\pi\cdot r_0^2} \tag{3}$$

where E_{pulse} is the total pulse energy.

The reflectivity of the material, R, is calculated as the maximum of the inherent, or "steady-state" reflectivity of the glass and the reflectivity of the plasma: $R(r,t)=\max(R_{ss},R_p)$. Both values come from the complex refractive index of the material according to the Fresnel expression

$$R_p(r,t) = \frac{(n-1)^2+k^2}{(n+1)^2+k^2} \tag{4}$$

where n and k are the real and imaginary components of the refractive index, respectively. In the absence of plasma, the real refractive index for BBS of $n=1.47$ gives $R_{ss}=0.0362$. As the electron density evolves, the indices are calculated from the complex dielectric function of the material, $\varepsilon=\varepsilon_r+i\varepsilon_i$ by the relations

$$n = \sqrt{\frac{\varepsilon_r+\sqrt{\varepsilon_r^2+\varepsilon_i^2}}{2}} \tag{5}$$

$$k = \sqrt{\frac{-\varepsilon_r+\sqrt{\varepsilon_r^2+\varepsilon_i^2}}{2}} \tag{6}$$

The dielectric function of the plasma is calculated from Drude model [11]

$$\varepsilon_r = \left(1-\frac{\omega_p^2(n_e)\tau^2}{1+\omega^2\tau^2}\right) \tag{7}$$

$$\varepsilon_i = \left(\frac{\omega_p^2(n_e)\tau}{\omega(1+\omega^2\tau^2)}\right) \tag{8}$$

where the laser frequency, $\omega=2\pi c/\lambda$ with c as the speed of light in a vacuum and λ the laser wavelength. ω_p is the plasma frequency given by

$$\omega_p(n_e) = \sqrt{\frac{n_e(r,z,t)e^2}{m_e\varepsilon_0}} \tag{9}$$

and τ is the free electron relaxation time. The constant e is the electron charge, m_e is the electron mass, and ε_0 is the permittivity of free space.

The absorption coefficient in Eq. (2) has components due to free electron heating as well as ionization, given by

$$\alpha(r,z,t) = \alpha_H(r,z,t) + a_i n_e(r,z,t)U \tag{10}$$

Absorption due to free electron heating, α_H is

$$\alpha_H = \frac{4\pi k}{\lambda} \tag{11}$$

while U is the bandgap of the material. For fused silica, $U\approx 9$ eV. For BBS, $U\approx 4$ eV.

From [9], it is assumed that at the critical electron density, the plasma frequency is equal to the laser frequency. If a point in the material reaches the critical density, n_{cr}, a discrete amount of material is ablated

$$n_{\text{cr}} = \frac{4\pi^2c^2m_e\varepsilon_0}{\lambda^2e^2} \tag{12}$$

From Eqs. (4)–(9), it is evident that the reflectivity of the material is dependent on two variables, the electron relaxation time, τ, and the electron density. Jiang and Tsai proposed a variable calculation of the relaxation time based on quantum treatment of the electron specific heat and electron temperature [16–18]. An investigation was conducted on the dependence of both relaxation time and electron density in R. It was found that the surface reflectivity varies greatly with electron density, a consequence of plasma generation, which has strong reflection and absorption properties. In contrast, reflectivity was essentially constant throughout a wide range of free electron relaxation time τ, varying significantly only when τ fell below 10 fs.

Metals with very large conduction band densities can exhibit relaxation times on the order of 1–10 fs [19], however, a dielectric with minimal conduction band density would expect to show much larger periods between electron/ion collisions. A relaxation time of the semiconductor silicon has been measured at >200 fs for moderate carrier densities [20]. Therefore, due to the order of magnitude increase in computational expense for a variable τ calculation, and the weak dependency on ablated volume, a constant relaxation time of 100 fs was used. The potential loss in accuracy due to the constant value of τ would not be removed even with a fully quantum calculation, which requires ionization state energies. These energies are unknown in a multi-element material with constantly changing internal energy, requiring an approximation regardless.

The numerical procedure involves calculating an electron density and intensity at each spatial point using the most recently updated reflectivity and absorption coefficient. Equations (4)–(11) are performed, and the density and intensity are updated. The cycle repeats until the densities and intensities have converged, completing the time step. Time step independence was achieved with values of $\Delta t=0.01$ fs for a 100 fs pulse.

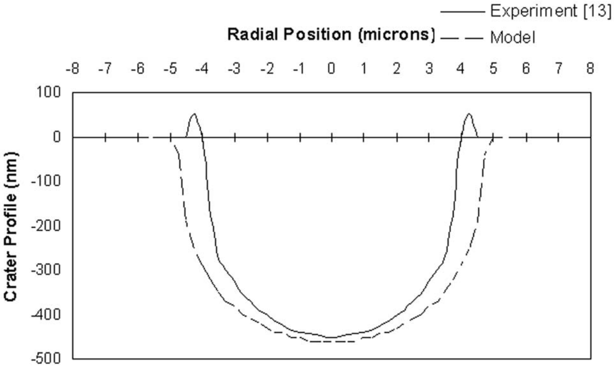


Fig. 1 Single shot ablated crater, $E_{\mathrm{pulse}}=13.8\ \mu J$

Theoretical Results

An experimental investigation of femtosecond ablation of borosilicate glass was conducted by Ben-Yakar and Byer [13] which offers excellent characterizations by which to verify the theoretical model. A quantitative evaluation of single shot ablated craters using three laser fluences with pulses of $\lambda=780$ nm and $t_p=200$ fs was carried out using an atomic force microscope. Figures 1 and 2 plot the experimental cross sections of the circular craters against the crater shapes predicted from the dielectric ablation model.

Figures 1 and 2 show a lip of material surrounding the experimental crater above the zero level of the substrate. This is assumed to be caused by a thermal after-effect of the excited electrons, which would heat the lattice and provoke both phonon and electron heat transfer after the pulse. If the lattice temperature is raised above the melting temperature of the material, a flow of molten glass would be subject to the local forces. The ablation model does not take into account fluid dynamics that may be the cause of the lip structure.

A notable result of the model is the successful prediction of a flat-bottom crater. Conventional thermal energy transport predicts shapes conforming to the Gaussian laser distribution, which result in round-bowl craters. Note the difference of scale in Figs. 1 and 2; depth in nanometers, radial position in micrometers.

The single shot ablation threshold of the borosilicate glass, or the minimum fluence at which material is removed, was found to be 2.6 J/cm^2 in air and 4.1 J/cm^2 in vacuum [13]. The dielectric ablation model, which does not consider the effect of a surrounding medium, predicts an ablation threshold of 3.6 J/cm^2.

The overall material removal during the pulse is highly dependent upon the optical response of the system. During the first few femtoseconds, multi-photon ablation dominates within the glass. The avalanche ablation effect increases nonlinearly as it becomes the dominant free-electron generation mechanism until the critical density is reached. At this point, plasma is formed at the surface as material is ejected. Figures 3 and 4 show how the surface reflectivity and absorption coefficient of the borosilicate glass transforms as the critical density is reached at approximately 50 fs. In the first stage of the pulse when the borosilicate is still transparent glass, the reflectivity retains its "standard" low value and the absorption coefficient is virtually zero. Near 50 fs, both values increase dramatically. The reflectivity levels to a nearly constant value where almost all energy is reflected, while the absorption of energy making it through the surface steadily increases as the plasma mass increases. Similar behavior was reported for fused silica [17].

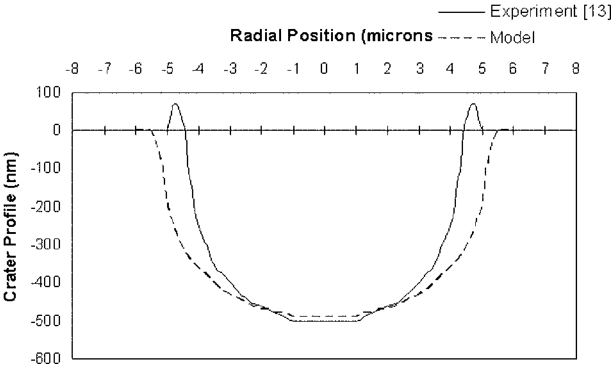


Fig. 2 Single shot ablated crater, $E_{\mathrm{pulse}}=18.8\ \mu J$

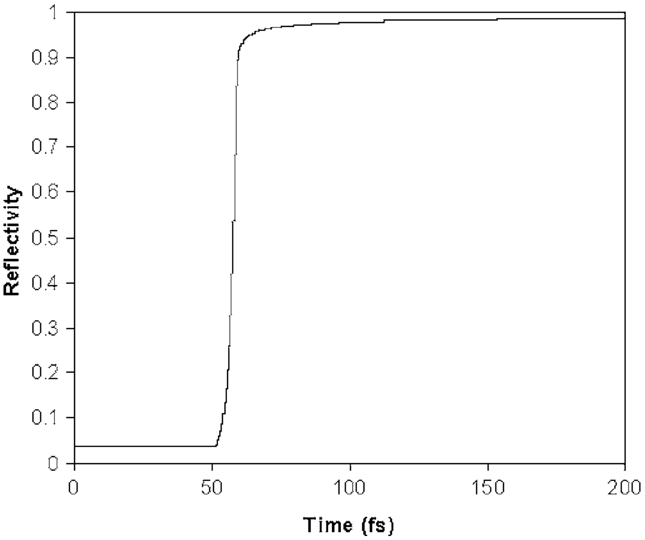


Fig. 3 Surface reflectivity at $r=0$ during 200 fs pulse

Near-Field Enhanced Laser Ablation

The presence of a dielectric sphere on a sbstrate surface can enhance the electromagnetic field near the sphere boundary substantially [21,22]. An earlier paper reports on a method of quantifying the enhancement to provide an energy input for the ablation model using Mie's theory [22]. In this work, the finite-

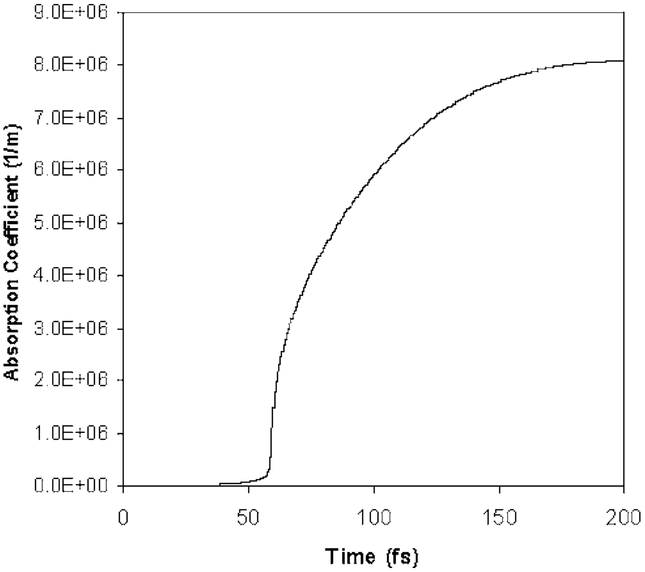


Fig. 4 Absorption coefficient at $r=0$, $z=0$ during 200 fs pulse

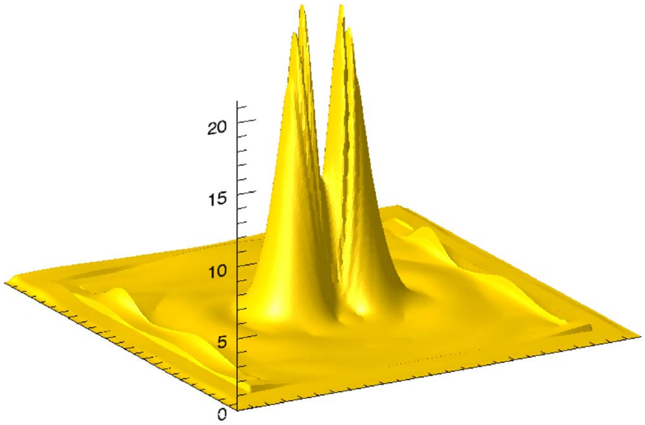


Fig. 5 Near-field enhanced laser intensity, λ=800 nm, *r* =880 nm

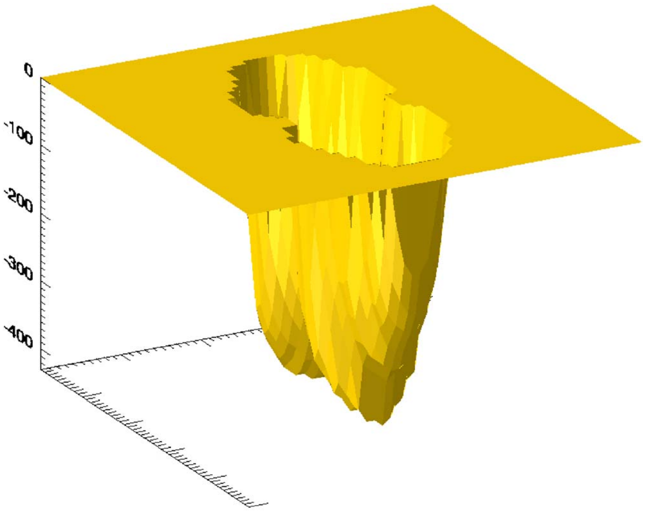


Fig. 6 Predicted ablation craters from near-field enhanced laser energy at three fluences

difference time-domain (FDTD) is employed in an in-house code to obtain the enhanced energy profile. The FDTD method is a widely popular technique of solving Maxwell's linear isotropic equations in complex environments where analytical solutions would be extremely difficult or impossible. While Mie's theory provides an exact solution for scattering of light from a sphere, the presence of the substrate and subsequent reflections add a level of complexity that is more easily tackled using FDTD. The Yee cell modeling geometry defines the mesh [23]. The incident radiation signal is generated using the total field/scattered field formulation [24], and the convolutional perfectly matched layer (CPML) absorbing boundary condition is employed at the boundary [25]. The details of the FDTD algorithm are covered in depth elsewhere [26,27] and, therefore, will not be addressed in this paper. The code was verified successfully against the analytical solution for a radiating dipole.

After initializing the property grids consistent with a vacuum, the sphere and substrate is placed in the environment by modifying the permittivity and permeability appropriately. The values used in this effort are $\mu=0.0005$ S/m and $\varepsilon=3.9^{*}\varepsilon_0$, where the permittivity of free space, $\varepsilon_0=8.8541878\times 10^{-12}$ F/m.

Figure 5 presents a surface plot of the 3D FDTD simulation results predicting an enhanced energy profile "seen" by the glass substrate beneath a 1.76 μm diameter sphere irradiated by a $\lambda=800$ nm, $t_p=100$ fs laser pulse. The plotting range is a 2.5 μm $\times$ 2.5 μm square beneath the sphere.

Figure 5 indicates a peak level of enhancement of roughly twenty times the incident laser energy. The peak enhancement theoretically lowers the threshold for identifiable damage to 200–300 mJ/cm^2. The central dip in enhancement is due to the presence of the substrate, ostensibly because of reflection and interference effects. Without the Pyrex substrate, simulations predict a continuous, single "hump" with a slightly lower peak level of enhancement. Other researchers have presented results with similar central dips in near-field enhancement [15]. Figure 5 also predicts an asymmetrical angular energy distribution, a consequence of polarization. An ideal ablation simulation would be three-dimensional to account for the ovular enhanced energy profile, however the nondiffusive ablation mechanisms contained in the model dictate lateral independence of energy transport. Therefore, several 2D calculations can take the place of a single 3D calculation to determine the damage profile.

With the enhanced energy input, Eq. (2) becomes modified as follows:

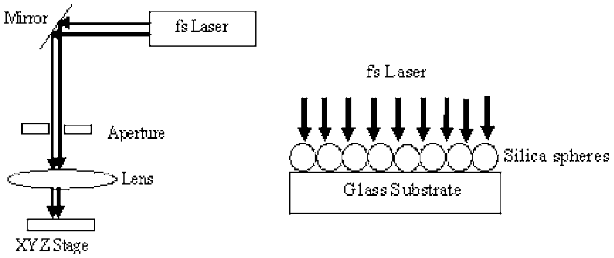
$$I(r,z,t)=\frac{2\cdot F_{\text{peak}}}{\sqrt{\pi/\ln 2}\cdot t_p}I_{\text{FDTD}}(r)[1-R(r,t)]\cdot\exp\left[-(4\ln 2)\frac{t^2}{t_p^2}-\int_0^z\alpha(r,z,t)dz\right] \tag{13}$$

where the additional $I_{\text{FDTD}}(r)$ term refers to the normalized level of enhancement relative to the incident plane wave prior to interaction with the sphere. The radial decay in the exponential argument is removed, as the spatial profile is accounted for in the FDTD results. Several pulses at a range of fluences were modeled with the enhanced laser input employed in Eq. (13). Figure 6 presents the damage profile predicted for a fluence of 950 mJ/cm^2. The oval-shaped crater has axes of length ~350 and ~700 nm, with a maximum depth of ~400 nm. The profile is in decent quantitative agreement with the experimental results to be presented, and qualitatively captures the ovular effect of polarization. A series of simulations conducted for the case of 330 mJ/cm^2 predicts crater dimensions of 160$\times$380 nm $\times$200 nm depth. A predicted threshold fluence of 240 mJ/cm^2 was calculated, which matches neatly with both the theoretical adjustment discussed above as well as experimental results to be presented.

It is clear from Fig. 6 that the energy required to damage the glass substrate has been significantly reduced due to the presence of the sphere. A notable distinction between the damage profiles occurs with the enhanced energy, enabling depth to width ratios greater than unity to be achieved. The minimum obtainable dimensions predicted are also much smaller than those obtained with identical spheres using nanosecond laser pulses [15]. Laser enhancement due to the presence of the spheres breaks the optical diffraction limit, allowing optical machining at the submicrometer level. In this computational work, the diffraction limit has been undercut by an entire order of magnitude.

Experimental Study

Experiments have been conducted corresponding to the modelling parameters. Figure 7 shows the schematic of the experimental setup for the study. A borosilicate glass wafer sample is used as the substrate and monodisperse silica (SiO_2) spherical particles with a diameter of 1.76 μm were used. These silica particles are transparent to the near infrared light (800 nm wavelength of the laser used). The particle suspension was diluted with de-ionized water (DI) water and deposited on the glass substrate using a



(*a*) (*b*)

Fig. 7 Schematic of (*a*) experimental setup, (*b*) irradiation of the silica spheres on borosilicate glass substrate

dispenser. The sample was then stored at room temperature over a period of few hours for all the water solvent to evaporate. A titanium-doped sapphire (Ti:Sapphire), solid state laser with pulse at full width half maximum (FWHM) around 100 femtosecond (fs) and wavelength around 800 nm is used as the light source. The polarization of this fs laser pulse is linear and horizontal to the surface of the substrate. As shown in Fig. 7(*b*), a fs laser is irradiated at a zero angle of incidence onto the monolayer of silica spheres deposited on the glass substrate. Different laser fluences were used to study the laser energy dependence of the nanostructures formed on the glass substrate. The laser energies used in the experiments varied from approximately 200 $\mathrm{mJ/cm^2}$ to 1 $\mathrm{J/cm^2}$. Zhou et al. [28] have recently used fs pulses to create subdiffraction limit modifications, though at much higher laser energies. The energy of the incident laser is varied by changing the delay time between the two pockel cells during the amplification using a synchronization and delay generator (SDG). Due to different delay times, the laser beam has to be compressed accordingly to maintain a constant pulse width and this is done by moving the compressor optics while monitoring the output laser beam pulse width with a single shot auto-correlator (SSA) and an oscilloscope. The glass substrate after laser nanostructuring was characterized by using a scanning electron microscope (SEM: Hitachi S4500) and the depths were measured by using an atomic force microscope (AFM: Nanoscope D3100). Figure 8 shows the SEM image of the monolayer of the silica spheres deposited on the borosilicate glass. To obtain good SEM images and in order to reduce the charging effects of the glass substrate a very thin layer of metal is deposited on the substrate.

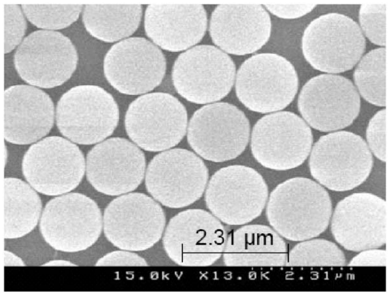
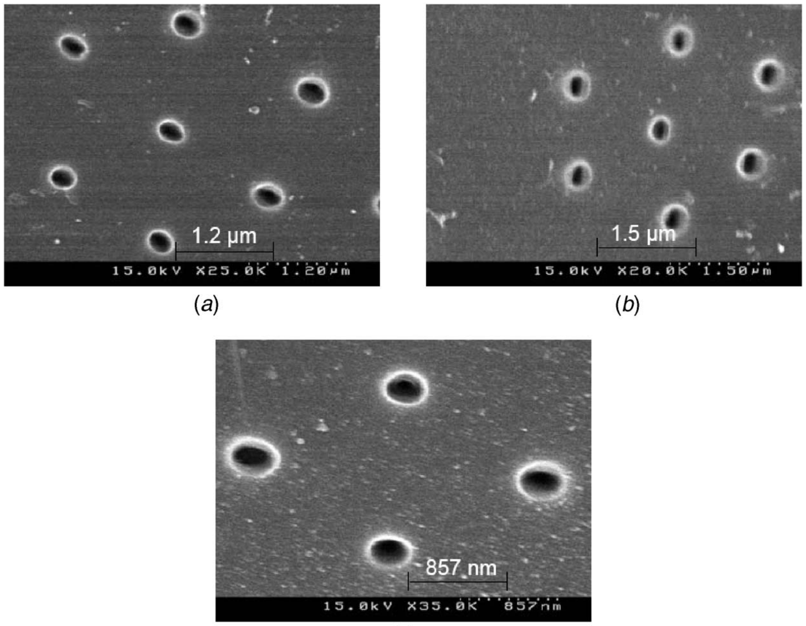


Fig. 8 SEM image of the monolayer of silica spheres with a diameter of 1.76 μm

Figure 9 shows the SEM image of the nanostructures formed on the borosilicate glass by using the optical near-field enhancement for different laser fluences and one pulse irradiation. The nanostructures formed seem to be elliptical in shape as predicted in Fig. 6. This is due to the laser polarization, which plays an important role in optical enhancement and hole formation on glass substrate. Also it was found that most of the silica particles are removed from the substrate surface. These particles are removed from the surface mainly due to the deformation force and/or the high ablative force exceeding the particle-surface adhesion force [29]. Figure 9 shows that as the laser fluence decreases the nanostructures formed did not change in shape, unlike in some studies where sombrero-shape nanodents are formed at high laser fluences and bowl-shape nanodents are formed at low laser fluences [21,30]. Also, from Fig. 9 it can be seen that for the incident laser fluences of 330, 550, and 765 $\mathrm{mJ/cm^2}$ the diameter of the holes along the major axis are approximately 250, 300, and 350 nm,



(*a*) (*b*)

(*c*)

Fig. 9 SEM micrograph of the features formed on the glass substrate using microspheres with a single laser pulse irradiation (λ=800 nm and FWHM =100 fs) having laser fluence of (*a*) 330 $\mathrm{mJ/cm^2}$, (*b*) 550 $\mathrm{mJ/cm^2}$, and (*c*) 765 $\mathrm{mJ/cm^2}$

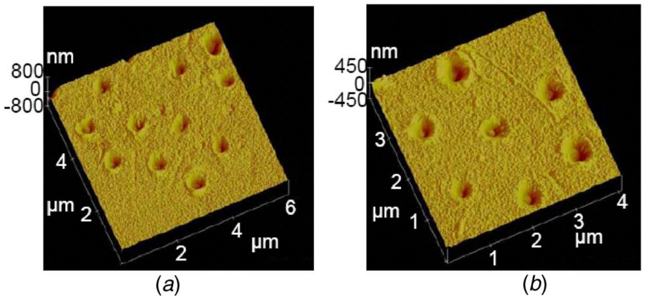


Fig. 10 AFM profiles of the features formed on the glass substrate using microspheres with a single laser pulse irradiation (λ=800 nm and FWHM=100 fs) having laser fluence of (*a*) 230 mJ/cm^2 and (*b*) 405 mJ/cm^2

respectively. Hence, the diameter of the nanostructures formed on the borosilicate glass increases with the laser fluence. Figure 10 shows the 3D AFM scanning profile of the nanostructure formed on the borosilicate glass and it can be seen that the structures formed have depths that are elliptical in shape which is in agreement with the SEM images (Fig. 9). Also the shapes of the nanostructures formed on the glass substrate remains the same for different laser fluences as seen earlier in Fig. 9.

Figure 11 shows the cross-sectional (or 2D) AFM profile of the nanostructures along their major axis and it can be seen clearly that for laser fluences of 230, 550, and 950 mJ/cm^2 the depth of the structures formed are approximately 180, 250, and 290 nm, respectively. These measured AFM cross-sectional profiles are in good agreement with the predicted profiles for similar incident laser energies, as shown in Fig. 6. From 2D AFM profiles we can observe that the bottom of the features along the major axis is flattened. The features can be explained by the above theoretical model predictions (Fig. 6) and also within the context of Mie's theory, which dictates strong forward scattering [31]. In general, the substrate modifications would be expected at the contact point due to the normal incidence of the laser pulse [32]. These nanostructures would then be driven by the intensity distribution in the surface plane.

Conclusion

In summary, a model has been built using the Fokker-Planck equation to describe the ablation mechanism of dielectrics during femtosecond laser pulses. The ablation model is then employed to predict the response of borosilicate glass to an enhanced near-field electromagnetic intensity due to the presence of microspheres on the substrate surface. An experimental study has been performed to observe the nanostructures created on the surface for lower energy fluences. The features formed on the surface have diameters and depths in the order of 300 and 250 nm, depending on the incident laser intensity. The experimentally observed nanostructures is in good agreement with the theoretical predictions of surface modifications. Therefore, it is shown that the surface features can be created below the diffraction limit and at the nanoscale in a massively parallel fashion. Applications of this unique process include nanostructuring optical materials, bio-polymers, and biological materials.

Acknowledgment

The authors wish to acknowledge support from the National Science Foundation under Grant No. CTS-0243160, and from the University of Texas Advanced Manufacturing Center. The femtosecond laser system was supported by a DURIP grant and Younger Investigator Award to S.C. from the US Office of Naval Research. The SEM and AFM analysis was conducted in the Center for Nano and Molecular Science and Technology (CNM) at UT-Austin. Valuable discussions with Professor L. Shi are greatly appreciated.

Nomenclature

α = total absorption coefficient
α_H = abs. coeff. due to free electron heating
δ_N = cross-section of N-photon absorption
ε_r = real component of dielectric function
ε_i = imag. component of dielectric function
ε_0 = permittivity of free space
τ = free electron relaxation time
ω = laser frequency
a_i = avalanche ionization coefficient
c = speed of light

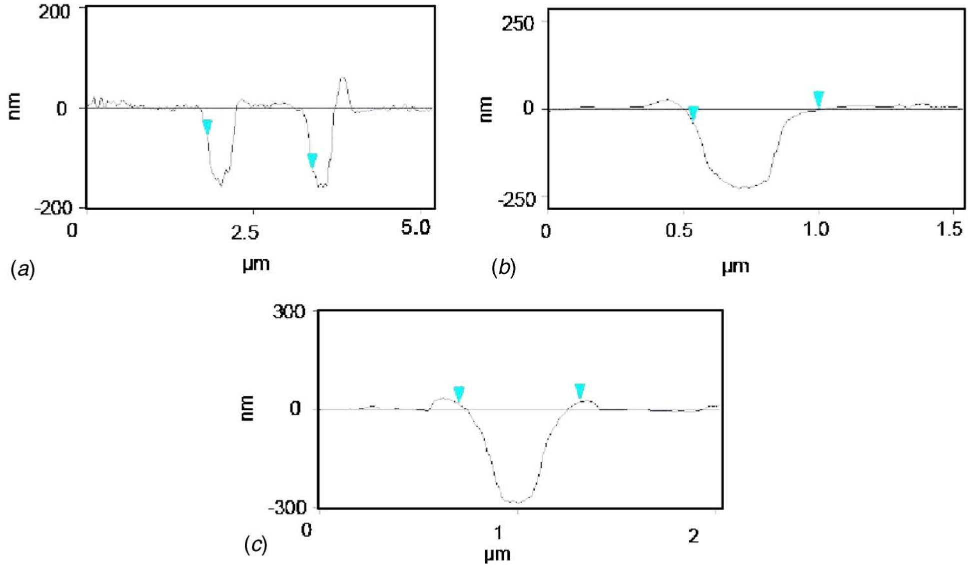


Fig. 11 AFM cross-section profiles of the features formed on the glass substrate using microspheres with a single laser pulse irradiation (λ=800 nm and FWHM=100 fs) having laser fluence of (*a*) 230 mJ/cm^2, (*b*) 550 mJ/cm^2, and (*c*) 950 mJ/cm^2

e = electron charge
F_{peak} = peak laser fluence
m_e = electron mass
n_e = free electron density
R = reflectivity
R_p = plasma reflectivity
R_{ss} = inherent reflectivity
t_p = laser pulse duration
U = bandgap

References

[1] Zhang, J., Sugioka, K., and Midorikawa, K., 1999, "High-Quality and High-Efficiency Machining of Glass Materials by Laser-Induced Plasma-Assisted Ablation Using Conventional Nanosecond UV, Visible, and Infrared Lasers," Appl. Phys. A, **69**, pp. 879–882.

[2] Park, D. S., Cho, M. W., Lee, H., and Cho, W. S., 2004, "Micro-Grooving of Glass Using Micro-Abrasive Jet Machining," J. Mater. Process. Technol., **146**, pp. 234–240.

[3] Yan, B. H., Wang, A. C., Huang, C. Y., and Hunag, F. Y., 2002, "Study of Precision Micro-Holes in Borosilicate Glass Using Micro EDM Combined with Micro Ultrasonic Vibration Machining," Int. J. Mach. Tools Manuf., **42**, pp. 1105–1112.

[4] Zhang, J., Sugioka, K., and Midorikawa, K., 1998, "High-Speed Machining of Glass Materials by Laser Induced Plasma Assisted Ablation Using a 532 nm Laser," Appl. Phys. A: Mater. Sci. Process., **67**, pp. 499–501.

[5] Nikumb, S., Chen, Q., Li, C., Reshef, H., Aheng, H. F., Qiu, H., and Low, D., 2005, "Precision Glass Machining, Drilling and Profile Cutting by Short Pulse Lasers," Thin Solid Films, **477**(1–2), pp. 216–221.

[6] Youn, S. W., and Kang, C. G., 2005, "Maskless Patterning of Borosilicate Glass Nanoindentation-Induced Etch-Hillock Surface Using Phenomena," J. Non-Cryst. Solids, **351**(37–39), pp. 3065–3074.

[7] Hirano, M., Kawamura, K. I., and Hosono, H., 2002, "Encoding of Holographic Grating and Periodic Nano Structure by Femtosecond Laser Pulse," Appl. Surf. Sci., **197**, pp. 688–698.

[8] Lenzner, M., Kruger, J., Sartania, S., Cheng, Z., Spielmann, Ch., Mourou, G., Kautek, W., and Krausz, F., 1998, "Femtosecond Optical Breakdown in Dielectrics," Phys. Rev. Lett., **80**, pp. 4076–4079.

[9] Stuart, B. C., Feit, M. D., Herman, S., Rubenchik, A. M., Shore, B. W., and Perry, M. D., 1995, "Laser-Induced Damage in Dielectrics With Nanosecond to Subpicosecond Pulses," Phys. Rev. Lett., **74**, pp. 2248–2251.

[10] 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.

[11] Campbell, S., Dear, F. C., Hand, D. P., and Reid, D. T., 2005, "Single-Pulse Femtosecond Laser Machining of Glass," J. Opt. A, Pure Appl. Opt., **7**, pp. 162–168.

[12] Ben-Yakar, A., Byer, R. L., Harkin, A., Ashmore, J., Stone, H., Shen, M., and Mazur, E., 2003, "Morphology of Femtosecond-Laser-Ablated Borosilicate Glass Surfaces," Appl. Phys. Lett., **83**(15), pp. 3030–3032.

[13] Ben-Yakar, A., and Byer, R. L., 2004, "Femtosecond Laser Ablation Properties of Borosilicate Glass," J. Appl. Phys., **96**, pp. 5316–5323.

[14] 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.

[15] Chimmalgi, A., Grigoropoulos, C. P., and Komvopolous, K., 2005, "Surface Nanostructuring by Nano-/Femtosecond Laser-Assisted Scanning Force Microscopy," J. Appl. Phys., **97**, p. 104319.

[16] Jiang, L., and Tsai, H. L., 2004, "Prediction of Crater Shape in Femtosecond Laser Ablation of Dielectrics," J. Phys. D, **37**, pp. 1492–1496.

[17] Jiang, L., and Tsai, H. L., 2005, "Energy Transport and Material Removal in Wide Bandgap Materials by a Femtosecond Laser Pulse," Int. J. Heat Mass Transfer, **48**, pp. 487–499.

[18] Kruer, W. L., 1987, *The Physics of Laser Plasma Interaction*, Addison Wesley, New York.

[19] Kittel, C., 1996, *Introduction to Solid State Physics*, Wiley, New York.

[20] Sjodin, T., Petek, H., and Dai, H.-L., 1998, "Ultrafast Carrier Dynamics in Silicon: A Two-Color Transient Reflection Grating Study on a (111) Surface," Phys. Rev. Lett., **81**, pp. 5664–5667.

[21] Lu, Y., Theppakuttai, S., and Chen, S. C., 2003, "Marangoni Effect in Nanosphere-Enhanced Laser Nanopatterning of Silicon," Appl. Phys. Lett., **82**, pp. 4143–4145.

[22] Heltzel, A., Theppakuttai, S., Chen, S. C., and Howell, J. R., 2005, "Analytical and Experimental Investigation of Laser Nanoscale Surface Modification," ASME J. Heat Transfer, **127**(11), pp. 1231–1235.

[23] Yee, K. S., 1966, "Numerical Solution of Inital Boundary Value Problems Involving Maxwell's Equations in Isotropic Media," IEEE Trans. Antennas Propag., **14**, pp. 302–307.

[24] Umashankar, K. R., and Taflove, A., 1982, "A Novel Method to Analyse Electromagnetic Scattering of Complex Objects," IEEE Trans. Electromagn. Compat., **24**, pp. 397–405.

[25] Roden, J. A., and Gedney, S. D., 2000, "Convolutional PML (CPML): An Efficient FDTD Implementation of the CFS-PML for Arbitrary Media," Microwave Opt. Technol. Lett., **27**, pp. 334–339.

[26] Taflove, A., and Hagness, S., 2005, *Computational Electrodynamics, The Finite-Difference Time-Domain Method*, 3rd ed., Artech House, Boston.

[27] Sullivan, D., 2000, *Electromagnetic Simulation Using the FDTD Method*, Wiley, New York.

[28] Zhou, Y., Hong, M. H., Fuh, JYH, Lu, L., Luk'yanchuk, B. S., Wang, Z. B., Shi, L. P., and Chong, T. C., 2006, "Direct Femtosecond Laser Nanopatterning of Glass Substrate by Particle-Assisted Near-Field Enhancement," Appl. Phys. Lett., **88**, p. 023110.

[29] Zheng, Y. W., Luk'yanchuk, B. S., Lu, Y. F., Song, W. D., and Mai, Z. H., 2001, "Dry Laser Cleaning of Particles from Solid Substrates: Experiments and Theory," J. Appl. Phys., **90**, pp. 2135–2142.

[30] Wang, Z. B., Hong, M. H., Luk'yanchuk, B. S., Huang, S. M., Wang, Q. F., Shi, L. P., and Chong, T. C., 2004, "Parallel Nanostructuring of GeSbTe Film With Particle Mask," Appl. Phys. A: Mater. Sci. Process., **79**, pp. 1603–1606.

[31] Ikawa, T., Mitsuoka, T., Hasegawa, M., Tsuchimori, M., Watanabe, O., and Kawana, Y., 2001, "Azobenzene Polymer Surface Deformation Due to the Gradient Force of the Optical Near Field of Monodispersed Polystyrene Spheres," Phys. Rev. B, **64**, p. 195408.

[32] Wang, Z. B., Hong, M. H., Luk'yanchuk, B. S., Lin, Y., Wang, Q. F., and Chong, T. C., 2004, "Angle Effect in Laser Nanopatterning With Particle-Mask," J. Appl. Phys., **96**(11), pp. 6845–6850.

Surface Plasmon Scattering by Gold Nanoparticles and Two-Dimensional Agglomerates

Pradeep Garudadri Venkata

Mustafa M. Aslan

M. Pinar Mengüç
e-mail: menguc@engr.uky.edu

Radiative Transfer Laboratory,
Mechanical Engineering Department,
University of Kentucky,
Lexington, KY 40506

Gorden Videen
U.S. Army Research Laboratory AMSRL-CI-EE,
2800 Powder Mill Road,
Adelphi, MD 20783

There has long been an interest in nanosized metallic particles for numerous novel applications, from the productions of colored glass in medieval times to the molecular-level sensors of today. These particles are known to display considerably different, and size-dependent, optical properties than those of their bulk counterparts. Yet it is very difficult to determine the size and structure of these particles in situ, such as monitoring the actual self-assembly process, because of their small size. In this paper, we present a methodology to predict the patterns of nanosized particles and agglomerates subjected to surface plasmon waves. For this characterization, the scattering patterns of different types of particles and agglomerates on or near the surface are needed. A combination of the T-matrix method, image theory, and a double interaction model are considered. The incident and scattered fields are expanded by employing spherical harmonic functions. The surface effects are incorporated using the Fresnel equations, in the incident-field expansion coefficients, and by including particle-surface interaction fields. The premise of the method is that the T-matrix is independent of incident and scattered fields and hence can be used effectively for cases involving incident surface waves. By obtaining the T-matrix for clusters or agglomerates of metallic particles, the scattering matrix elements (M_{11}, M_{12}, M_{33}, and M_{34}) of agglomerated structures on the surface are calculated using an additional T-matrix operation. The effect of size, shape, and orientation of gold nanosized particles on their scattering patterns are explored both in the visible spectrum and at resonance wavelengths. The results show that the normalized scattering matrix elements at certain observation angles and incident wavelengths provide significant information to monitor the structural change of gold nanosized particles on a gold substrate. [DOI: 10.1115/1.2401199]

Keywords: surface plasmon, scattering, scattering matrix, agglomerates, self assembly

Introduction

Scattering from small particles has been of significant interest in many areas of engineering and applied sciences because of their impact on the effective and apparent properties of media they are imbedded in, as well as for diagnosis of their size, shape, and structures. Understanding the scattering behavior of small (from micrometer down to nanometer size) particles is crucial for many applications, including atmospheric research on dust particles, astronomical studies on interstellar dust and radiation, optical diagnostics for industrial aerosol processes, combustion research, and detection of soot particles and precursors, as well as self-assembly of nanosized (1–100 nm) particles [1].

More than a century ago, the scattering of electromagnetic waves by a perfect sphere was studied by Lorenz, and later independently by Mie, to explain the effect of metallic spherical particles on the color of glass. The so-called Lorenz-Mie theory is an analytical solution of the Maxwell equations for this idealized geometry, and allows the computation of the scattering patterns for homogeneous spheres. Several extensions and generalizations of the method are available to treat inhomogeneous, spherically symmetric and axially symmetric nonspherical particles [2,3]. There are also several methods available for nonspherical and clusters of noninteracting particles, all of which are based on solving the Maxwell equations either analytically or numerically [4–12].

Of all the methods available to analyze scattering, the T-Matrix Method (TMM) has the advantage of being highly accurate and applicable to different kinds of particles [2,3]. The TMM was originally developed by Waterman [4] and since then has undergone several generalizations and modifications to be applicable to arbitrary clusters of nonspherical particles [5] and nonspherical chiral scatterers [6]. These methods are based on expanding the incident and scattered fields in terms of vector spherical harmonics. The expansion coefficients of the scattered field are then related to the coefficients of the incident field through a T-matrix, whose elements depend only on the shape, size parameter, refractive index, and orientation of the scattering particle. The expansion coefficients are independent of the incident and scattered fields. Knowing the incident-field expansion coefficients and the T-matrix allows for the calculation of the scattering properties for any configuration, though the T-matrix can become quite complicated for clusters of particles with no axial symmetry.

The problem of dependent scattering for a cluster of spheres have been investigated by many researchers, including Bruning and Lo [7], Tsang and Kang [8], Varadan et al. [9], Borghese et al. [10], Mackowski [11], and Mackowski and Mischenko [12], to name a few. The methods developed for this make use of addition theorems for vector spherical harmonics, and all have been used for particles isolated in a medium and illuminated by a plane electromagnetic (EM) wave. Potentially, these numerical models can be modified to study the characterization of particles on or near a substrate, which requires modeling the EM-wave interaction with both the surface and the particles. This problem has interesting applications in characterizing microstructures on flat substrates [13] and development of surface scanners [14].

Several methods, such as the extinction theorem, image theory, modified double interaction model, and ray tracing, have been

---

Contributed by the Heat Transfer Division of ASME for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received February 3, 2006; final manuscript received July 27, 2006. Review conducted by Walter W. Yuen.

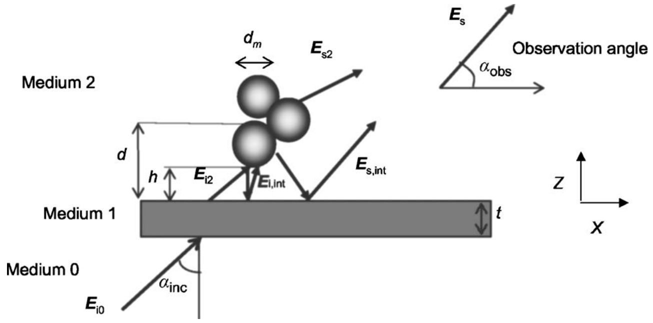


Fig. 1 The schematic of the scattering from a group of particles on (h=0) or near (h>0) a thin film. d is the distance between the surface of the film and the center point of the agglomerated structure.

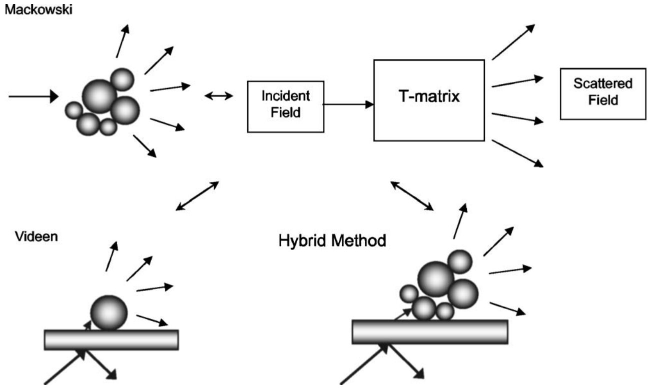


Fig. 2 Schematic for the hybrid method used

proposed to calculate scattering properties from particles on substrates [15,16]. However, much of the effort has been limited to characterizing single particles of sizes comparable to that of the incident radiation. Characterization of particles that are much smaller than the wavelength of incident radiation cannot be achieved using the above methods because they assume direct incidence of a planar wave on a particle. These methods deal with particles on the same side as the incident EM field; whereas the focus in the most interesting practical applications is on particles on the opposite surface of the interface. For such cases, it is possible to characterize them by methods that employ evanescent waves/surface plasmons incident on the particle. This characterization has been done for single particles on metallic substrates using a combination of the T-matrix method and image theory that includes interaction fields. The formulation for this case is presented in [17], and a detailed parametric study for a single nanosize particle on a gold film is outlined in [18].

This paper describes results from a methodology significantly more involved than those of the previous studies. The focus here is to characterize clusters or agglomerates of metallic particles on thin substrates, which is accomplished by using a combination of the T-matrix method and image theory. The method consists of calculating, using the appropriate Fresnel coefficients, the incident-field expansion coefficients after the incident field traverses through a thin metallic film. An outline of the general formulation of the problem is given below, combined with a discussion of how the governing equations are converted to working equations, and the details of the formulation for obtaining the expansion coefficients. By making use of the T-matrix of the particle system (calculated independently), the scattering properties of the system of particles are obtained. After that, a series of results are presented for different shaped structures comprised of nanoparticles. Based on these profiles, several suggestions are made about the potential use of these data for characterization of such structures from experiments.

Formulation of the Problem

The geometry of the model considered in this study is shown in Fig. 1. It is assumed that an electromagnetic wave is incident on a nanosized particle or agglomerate on or above a surface. The EM wave scattered by this configuration will be due to both the particle and the surface. The total scattered field then can be written as $\boldsymbol{E}_s=\boldsymbol{E}_{s2}+\boldsymbol{E}_{s,\mathrm{int}}$, where $\boldsymbol{E}_{i,\mathrm{int}}$ and $\boldsymbol{E}_{s,\mathrm{int}}$ are the electrical fields due to the particle-surface interactions. The incident and the scattered electric fields are expanded using vector spherical harmonics. When the particle-surface interaction is not considered, i.e., when $\boldsymbol{E}_{i,\mathrm{int}}$ and $\boldsymbol{E}_{s,\mathrm{int}}$ are negligible, the scattering coefficients are obtained directly from the incident-field expansion coefficients by making use of a response matrix called the B-matrix [17], the elements of which relate the scattered field expansion coefficients to the incident field. The B-matrix is indeed a T-matrix whose elements are reduced to the Lorenz Mie coefficients for the case of a single homogenous sphere [18]. This formulation is identical to the T-matrix formulation presented by Mackowski [11,12].

Here, we propose a new hybrid method to use a T-matrix obtained independently for a group of particles or particle system, and include the particle-surface interaction fields to obtain the scattering matrix for the configuration shown in Fig. 1. Note that while the model is *based* on the formulation given in [17,18], here we are modeling the scattering by agglomerated particles on or near the surface, which has not been done previously.

The concept of this hybrid method is presented in Fig. 2. The physical system consists of a thin metallic film (Medium 1) on a quartz layer (Medium 0), and the particles are assumed to be layered on it in Medium 2 (in our case, air). A laser beam is incident from below, at an angle (α_{inc}) to the quartz-gold interface. The laser beam yields surface plasmons/evanescent waves at the interface, and some of the energy is tunneled into the medium above the film (Medium 2). Though the system is three dimensional, the scattering matrix elements we present are obtained in a plane (2D) at an azimuthal angle of 0 deg, as shown in Fig. 1. The observation angle is the polar angle measured with respect to the horizontal axis (X-axis) parallel to the surface, in the anticlockwise direction.

This field incident to the particle-laden medium (M-2) above the thin film (M-1) is expressed as

$$E_{\mathrm{inc}}^2=\sum_{n=0}^{\infty}\sum_{m=-n}^{n} a_{nm}^{(1)}M_{nm,2}^{(1)}+a_{nm}^{(2)}N_{nm,2}^{(1)} \tag{1}$$

where the coefficients $M_{nm,2}^{(1)}$ and $N_{nm,2}^{(1)}$ are the vector spherical harmonics composed with spherical Bessel functions of the first kind [18,19] and $a_{nm}^{(1)}$ and $a_{nm}^{(2)}$ are the incident-field expansion coefficients. The incident-field expansion coefficients in M-2 were presented in [17]. The scattered fields in M-2 are given by

$$E_{\mathrm{sca}}^2=\sum_{n=0}^{\infty}\sum_{m=-n}^{n} b_{nm}^{(1)}M_{nm,2}^{(3)}+b_{nm}^{(2)}N_{nm,2}^{(3)} \tag{2}$$

Using Maxwell's equations and the appropriate boundary conditions, the relation between the incident and the scattered fields can be obtained in the spherical coordinate system as

$$\begin{pmatrix}E_{\mathrm{TM}}\\ E_{\mathrm{TE}}\end{pmatrix}_{\mathrm{sca}}=\frac{e^{ik(r-z)}}{-ikr}\begin{pmatrix}S_2 & S_3\\ S_4 & S_4\end{pmatrix}\begin{pmatrix}E_{\mathrm{TM}}\\ E_{\mathrm{TE}}\end{pmatrix}_{\mathrm{inc}} \tag{3}$$

E_{TM} and E_{TE} are the parallel (TM) and transverse (TE) components of the electric field, respectively. Subscript "inc" denotes the incident component of the field and "sca" for the scattered electric field. "**k**" is the wave vector for the wave and "**r**" is the distance between the scatterer and the observation plane. The scattered and incident fields are related by a simple matrix called the amplitude scattering matrix due to the linearity of Maxwell's equations. The elements S_i $(i=1,2,3,4)$ are the elements of the scattering amplitude matrix that are dependent on the scattering angle and the azimuthal angle.

The coefficients $M^{(3)}_{nm,2}$ and $N^{(3)}_{nm,2}$ in Eqs. (1) and (2) are expanded using the spherical Hankel functions of the first kind. The scattering coefficients can be calculated using

$$b^{p}_{nm} = \sum_{n',m'} B^{(n',m',1)}_{nmp} a^{(1)}_{n'm'} + B^{(n',m',2)}_{nmp} a^{(2)}_{n'm'} \tag{4}$$

for the case when the particles are much smaller than the incident wavelength (i.e., when the particle-surface interaction fields are negligible). Here, $B^{(n',m',q)}_{nmp}$ are the B-matrix elements, which reduce to the Lorenz-Mie coefficients for the case of a homogeneous sphere.

We use the T-matrix elements calculated by Mackowski [11,12] for different spherical particles or corresponding agglomerates, and substitute them for the T-matrix elements shown above. In principle, this approach is solid, as T-matrix formulation gives the electromagnetic wave external to the particle (or cluster). The T-matrix method applied to a cluster of spheres is very similar, with the scattered field in this case given by [12]

$$E^{2}_{\text{sca}} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \sum_{p=1}^{2} \sum_{i=1}^{N_s} b^{i}_{nmp} h_{mnp}(r_i) \tag{5}$$

where N_s is the number of spheres in an agglomerate. The index p stands for the modes of scattered field; i.e., $p=1$ for TM mode and $p=2$ for TE mode. h stands for outgoing vector spherical harmonic function. The superscript i denotes the ith sphere in the cluster. b^{i}_{nmp} are the scattering coefficients for the ith sphere defined at the sphere origin, i.e., the scattered field for a single sphere is given by

$$E^{2}_{\text{sca},i} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} b^{i\text{TE}}_{nm} h_{mn}(r_i) + b^{i\text{TM}}_{nm} h_{mn}(r_i) \tag{6}$$

The scattered-field expansion coefficients for the ith sphere are related to the incident-field expansion coefficients by the T-matrix

$$b^{i}_{nmp} = \sum_{j=1}^{N_s} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} T^{ij}_{mnpkl1} a^{j\text{TM}}_{kl} + T^{ij}_{mnpkl2} a^{j\text{TE}}_{kl} \tag{7}$$

The sphere-centered T^{ij} matrix is converted to the cluster-centered T-matrix (based upon a single origin of the cluster). The transformation is given by

$$T_{nl} = \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} \sum_{n'=1}^{N_{o,i}} \sum_{l'=1}^{N_{o,i}} J^{0i}_{nn'} T^{ij}_{n'l'} J^{j0}_{l'l} \tag{8}$$

Matrices J^{0i} and J^{j0} are formed from the addition coefficients based on the spherical Bessel function. The subscripts n and l (and their primes) stand for order, degree, and mode. The scattering coefficients for the entire cluster origin are given from the cluster T-matrix as

$$b^{\text{TM}}_{mn} = \sum_{l=1}^{N_o} \sum_{k=-l}^{l} T_{mn1kl1} a^{\text{TM}}_{kl} + T_{mn1kl2} a^{\text{TE}}_{kl} \tag{9}$$

$$b^{\text{TE}}_{mn} = \sum_{l=1}^{N_o} \sum_{k=-l}^{l} T_{mn2kl1} a^{\text{TM}}_{kl} + T_{mn2kl2} a^{\text{TE}}_{kl} \tag{10}$$

The coefficients a_{kl} and b_{mn} are the incident-field expansion coefficients and scattering coefficients defined at the cluster origin. Since the T-matrix is independent of incident and scattered fields, it can be used to calculate any scattering coefficients provided the incident-field expansion coefficients are known.

We can obtain the scattering coefficients for the case of a thin film using the cluster T-matrix and the a-coefficients obtained for our case. Neglecting the particle-surface interactions, we have

$$b^{(1)\text{TM}}_{mn} = \sum_{l=1}^{N_o} \sum_{k=-l}^{l} T_{mn2kl1} a^{(2)\text{TM}}_{kl} + T_{mn2kl2} a^{(1)\text{TM}}_{kl} \tag{11}$$

$$b^{(1)\text{TE}}_{mn} = \sum_{l=1}^{N_o} \sum_{k=-l}^{l} T_{mn2kl1} a^{(2)\text{TE}}_{kl} + T_{mn2kl2} a^{(1)\text{TE}}_{kl} \tag{12}$$

$$b^{(2)\text{TM}}_{mn} = \sum_{l=1}^{N_o} \sum_{k=-l}^{l} T_{mn1kl1} a^{(2)\text{TM}}_{kl} + T_{mn1kl2} a^{(1)\text{TM}}_{kl} \tag{13}$$

$$b^{(2)\text{TE}}_{mn} = \sum_{l=1}^{N_o} \sum_{k=-l}^{l} T_{mn1kl1} a^{(2)\text{TE}}_{kl} + T_{mn1kl2} a^{(1)\text{TE}}_{kl} \tag{14}$$

The same T-matrix elements and incident expansion coefficients can be used to calculate the scattering coefficients by including the particle-surface interaction. Once the scattering coefficients are obtained, the rest of the scattering amplitudes can be calculated. The method and the equations are presented in [17]; following that formulation, the interaction fields between the surface and the particle are obtained as

$$E^{2}_{\text{int}} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} c^{(1)}_{nm} M^{(1)}_{nm,2} + c^{(2)}_{nm} N^{(1)}_{nm,2} \tag{15}$$

The interaction fields $c^{(j)}_{nm}$ are found by assuming that the fields scattered by the particle system reflecting off interface 0-1 are negligible; i.e., considering only the fields reflecting off interface 1-2 and by assuming the interaction field is approximately equal to the mirror image of the scattered fields multiplied by the Fresnel reflection coefficient at normal incidence. They are given as

$$c^{(1)}_{nm} = \sum_{n'=|m|}^{\infty} R_{2,1}(0)(-1)^{n'+m}[b^{(2)}_{n'm} B^{(n',m)}_{n} - b^{(1)}_{n'm} A^{(n',m)}_{n}] \tag{16}$$

$$c^{(2)}_{nm} = \sum_{n'=|m|}^{\infty} R_{2,1}(0)(-1)^{n'+m}[b^{(2)}_{n'm} A^{(n',m)}_{n} - b^{(1)}_{n'm} B^{(n',m)}_{n}] \tag{17}$$

The coefficients $A^{(n',m)}_{n}$, $B^{(n',m)}_{n}$ are vector translational coefficients. In the medium above the interface, the incident field is the sum of the incident fields (Eq. (1)) and the interaction fields (Eq. (15)). The scattering coefficients (b-coefficients) are calculated taking into account the interaction fields. Hence, Eq. (4) becomes

$$\begin{aligned} b^{(p)}_{nm} = \sum_{n',m'} B^{(n',m',1)}_{nmp} \Bigg\{ & a^{(1)}_{n',m'} + \sum_{q=|m'|}^{\infty} R_{2,1}(0)(-1)^{q+m'}[b^{(2)}_{qm'} B^{(q,m')}_{n'} \\ & - b^{(1)}_{qm'} A^{(q,m')}_{n'}] \Bigg\} + B^{(n',m',2)}_{nmp} \Bigg\{ a^{(2)}_{n',m'} + \sum_{q=|m'|}^{\infty} R_{2,1}(0) \\ & \times (-1)^{q+m'}[b^{(2)}_{qm'} A^{(q,m')}_{n'} - b^{(1)}_{qm'} B^{(q,m')}_{n'}] \Bigg\} \end{aligned} \tag{18}$$

Using the relationship between the normalized associated Legendre polynomials, the relationships between the incident-field expansion coefficients can be derived.

By substituting the B-matrix in Eq. (18) with the T-matrix calculated from Eq. (8) and using the following relations for scattered field coefficients involving surface interactions

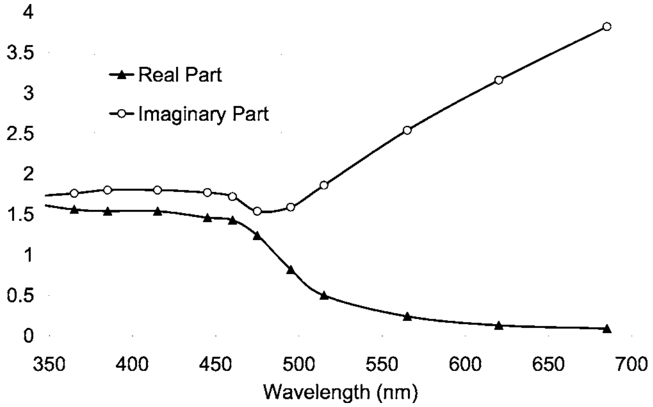


Fig. 3 Spectral variation of gold refractive index

$$
\begin{aligned}
b_{nm}^{(1)\mathrm{TE}} = \sum_{n',m'} T_{nm2n'm'2} & \left\{ a_{n'm'}^{(1)\mathrm{TE}} + \sum_{q=|m'|}^{\infty} R_{2,1}(0) \right. \\
& \left. \times(-1)^{q+m'}[b_{qm'}^{(2)\mathrm{TE}} B_{n'}^{(q,m')} - b_{qm'}^{(1)\mathrm{TE}} A_{n'}^{(q,m')}] \right\} \\
& + T_{nm2n'm'1} \left\{ a_{n'm'}^{(2)\mathrm{TE}} + \sum_{q=|m'|}^{\infty} R_{2,1}(0)(-1)^{q+m'}[b_{qm'}^{(2)\mathrm{TE}} A_{n'}^{(q,m')} \right. \\
& \left. - b_{qm'}^{(1)\mathrm{TE}} B_{n'}^{(q,m')}] \right\}
\end{aligned}
\tag{19}
$$

$$
\begin{aligned}
b_{nm}^{(2)\mathrm{TE}} = \sum_{n',m'} T_{nm1n'm'2} & \left\{ a_{n'm'}^{(1)\mathrm{TE}} + \sum_{q=|m'|}^{\infty} R_{2,1}(0) \right. \\
& \left. \times(-1)^{q+m'}[b_{qm'}^{(2)\mathrm{TE}} B_{n'}^{(q,m')} - b_{qm'}^{(1)\mathrm{TE}} A_{n'}^{(q,m')}] \right\} \\
& + T_{nm1n'm'1} \left\{ a_{n'm'}^{(2)\mathrm{TE}} + \sum_{q=|m'|}^{\infty} R_{2,1}(0)(-1)^{q+m'}[b_{qm'}^{(2)\mathrm{TE}} A_{n'}^{(q,m')} \right. \\
& \left. - b_{qm'}^{(1)\mathrm{TE}} B_{n'}^{(q,m')}] \right\}
\end{aligned}
\tag{20}
$$

the scattering coefficients and hence the M_{ij} can be calculated. The coefficients in (19) and (20) correspond only to the TE mode; the TM mode expressions are obtained in a similar fashion. As seen from these equations, the b-coefficients can be written in a matrix formulation that can be inverted (using a combination of LU decomposition and Gaussian inversion) to obtain the $b^{(1)}$ and $b^{(2)}$ coefficients. The number of terms required to calculate the b-coefficients depends on the particle size parameter. Spherical Hankel functions and Bessel functions are obtained using the routines developed earlier by Videen; the details of the computational formulation are presented in [17]. The T-matrix is independently obtained using the routines developed by Mackowski [11], whose programs calculate a sphere-centered T-matrix first through a matrix inversion method, which is followed by a contraction of this matrix into a cluster-centered T-matrix. Following these, we can obtain a relationship between the scattering coefficients and scattering matrix elements similar to those given in [19]. Here, only the theoretical framework is presented; details regarding the computational schemes and the formulation for T-matrix, B-matrix, and other functions used in the formulation are already available in the previous work [11,12,16–20]; therefore, they are not repeated here.

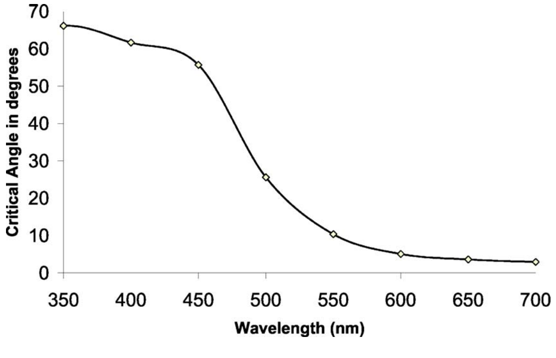


Fig. 4 Spectral variation of critical angle at quartz-gold interface (n-quart=1.75)

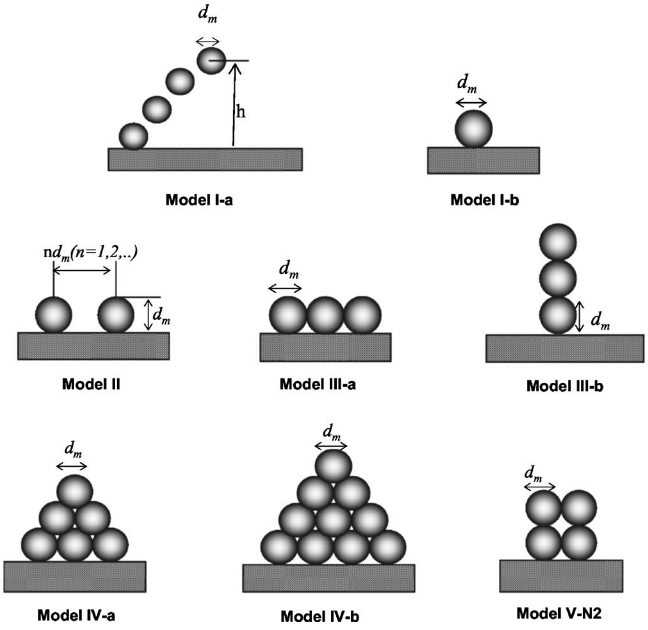


Fig. 5 Particle shape configurations and orientations considered

The addition theorem for spherical harmonics is used to translate the scattered fields in one coordinate system into the interaction fields centered at another coordinate system. In our case, the spherical Hankel functions are translated and expanded in terms of spherical Bessel functions. It is commonly stated that the expansion is valid in the region $r<d$, where d is the distance between the two coordinate systems. This is to avoid the pole at the origin about which the Hankel functions are expanded. In some of our simulations, the $r<d$ restriction is not upheld. However, this approximation is justified as the expansion is never used near the poles, which are located on the opposite side of the interface, and the arguments of the interaction fields are always $\ll\infty$.

Results and Discussion

The theoretical formulation presented in the previous section was implemented in a FORTRAN code. The original single-particle code was extended for multiple particles on the surface. The T-matrix for a cluster of spheres was obtained using the code by Mackowski [12]. In the present code, the cross-polarization terms are added to obtain accurate M_{ij} parameters. Below, we present the scattering profiles for the four M_{ij} parameters that are a normalized version of the four elements of the scattering matrix defined as

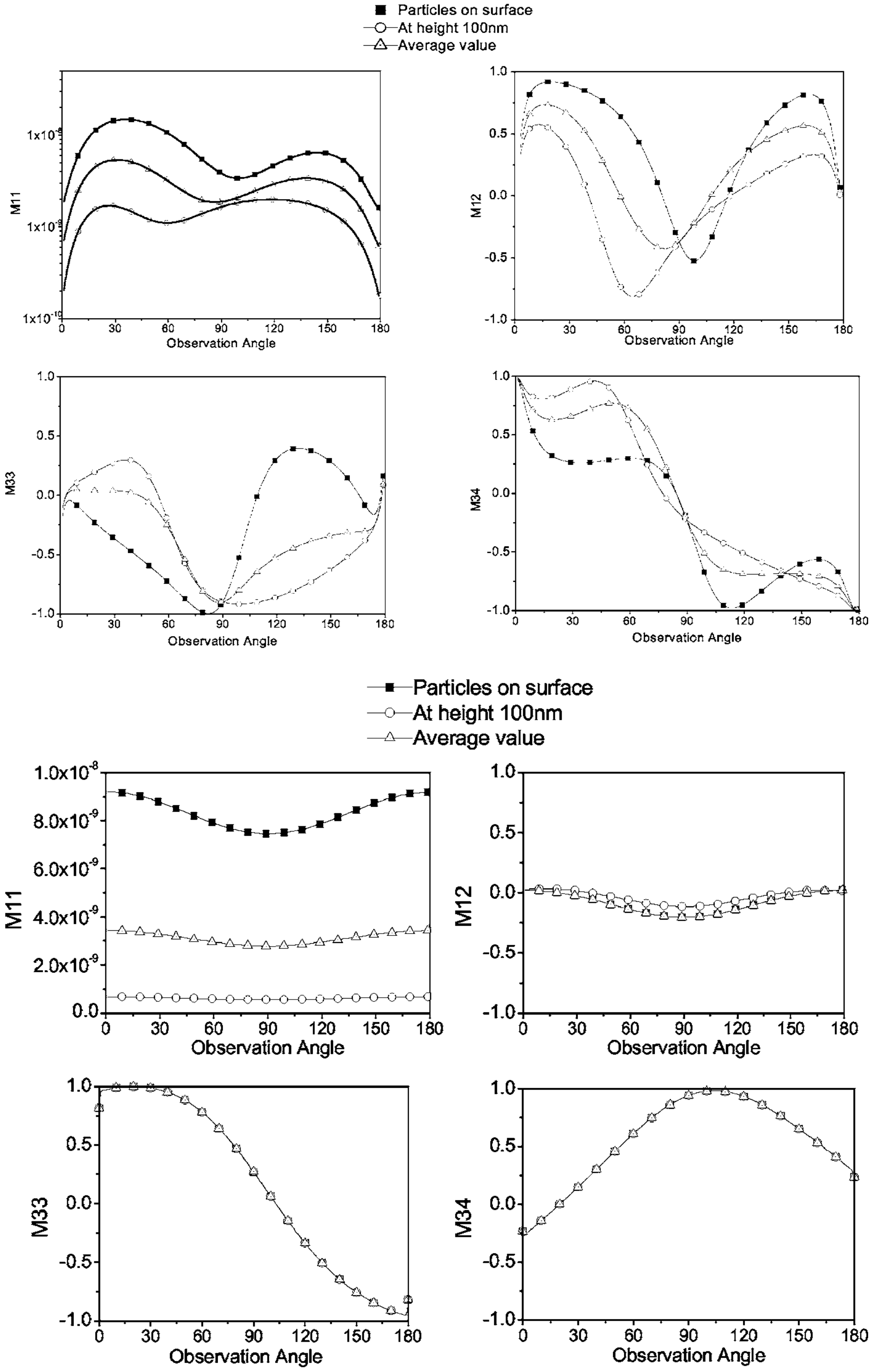


Fig. 6 10 nm diameter particle configurations (shown in Model I-a) on a gold substrate (top) and on a quartz substrate (bottom)

$$M_{11} = S_{11}$$

$$M_{12} = S_{12}/S_{11}$$

$$M_{33} = S_{33}/S_{11}$$

$$M_{34} = S_{34}/S_{11} \tag{20a}$$

The term "scattering profiles" refers to the scattering profiles for the four M_{ij} parameters described by Eqs. (20). The scattering profiles presented in this paper are generated for every observation angle between 0 and 180 deg; however, they are presented for only a few data points for the sake of clarity. The details of the scattering matrix and scattering matrix elements (M_{ij}) are given in our previous publications [17], and will not be repeated here. Note that S_{ij}-elements are related to amplitude scattering matrix elements, S_i (see Eq. (3)) via

$$S_{11} = \frac{1}{2}(|S_1|^2 + |S_2|^2 + |S_3|^2 + |S_4|^2) \tag{21}$$

$$S_{12} = \frac{1}{2}(|S_2|^2 - |S_1|^2 + |S_4|^2 - |S_3|^2) \tag{22}$$

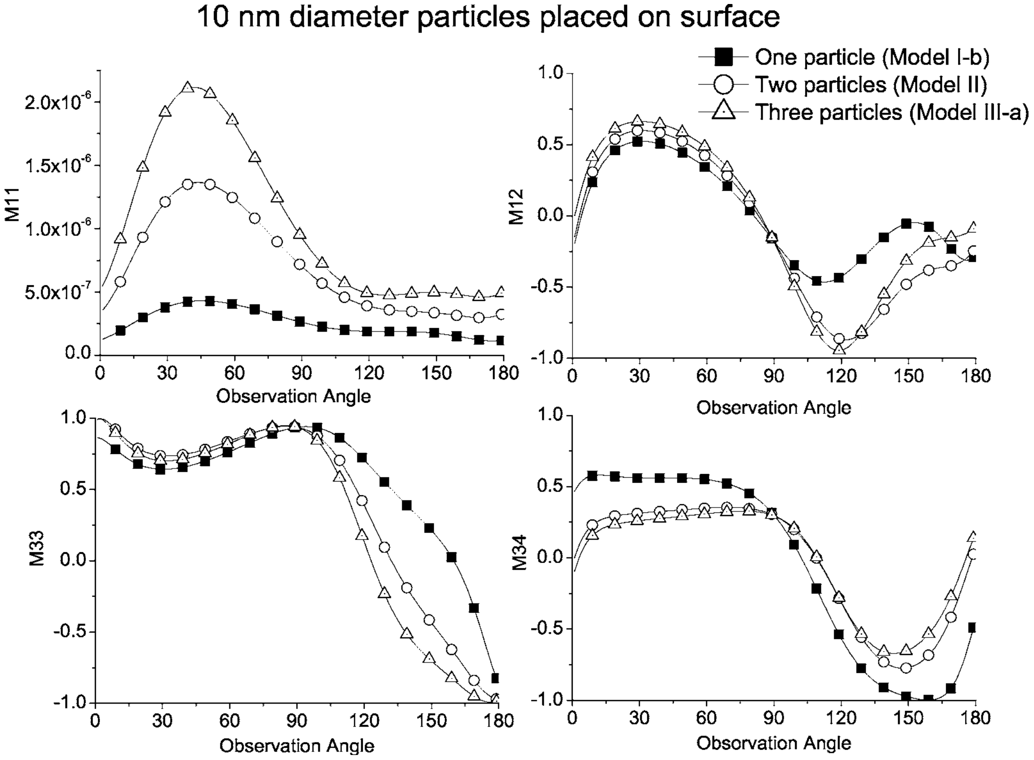


Fig. 7 Scattering profiles for particle configurations shown in Models I-b, II, and III-a, for a particle diameter of 10 nm

$$S_{33} = \mathrm{Re}(S_1 S_2^* + S_3 S_4^*) \tag{23}$$

$$S_{34} = \mathrm{Im}(S_2 S_1^* + S_4 S_3^*) \tag{24}$$

The present code was first evaluated against the standard results from the Lorenz-Mie formulation for a single homogenous spherical particle case; the agreement was found to be excellent (see [21] for details). Evaluation of the code for other particle shapes and structures was not possible due to the lack of benchmark results. However, convergence tests were conducted for different shapes, and the results were found to be consistent in all different scenarios considered. The computational time required for most cases was in the order of a few seconds. The memory requirements depended on the number of particles, the refractive index of the particles and the thin-film properties. For particle numbers less than 15, the code was fully functional on a 512 MB RAM personal computer. The memory requirements increased, as expected, when the number of particles was increased.

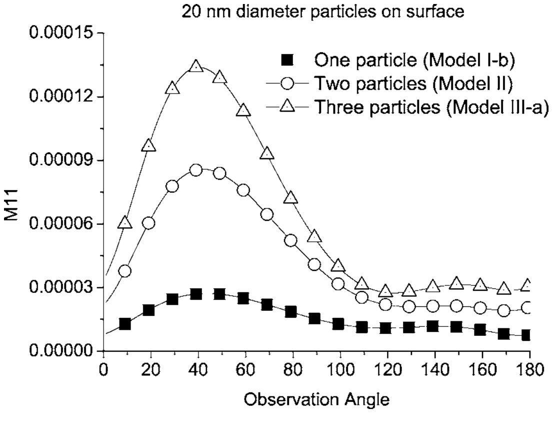


Fig. 8 M_{11} profile for particle configurations shown in Models I-b, II, and III-a, for a particle diameter of 20 nm

The main objective of the present research is to develop a predictive capability for experiments, where the effects of size and structures of gold nanosized particles within a surface-plasmon field on the scattering matrix elements can be quantified. Gold is one of the most widely used nano-size particles because of its desirable chemical and optical properties [22]. For that reason, simulations were carried out for gold particles. The refractive index of gold as a function of wavelength is shown in Fig. 3 [23]. For wavelengths greater than 480 nm, the absorption of gold increases rapidly. However, the real part of its refractive index decreases from 1.5 to 1.0.

We carried out a series of numerical experiments corresponding to different structures of gold nanoparticles. For diagnosis of such particles in actual experiments, different wavelength of light sources can be used. However, different wavelengths of a laser beam would have total internal reflection (TIR) at different angles. To explore the effect of the wavelength, the critical angle at the quartz-gold interface is determined as a function of the incident laser wavelength, which is depicted in Fig. 4. Based on these calculations, an incident angle of 70 deg at the quartz-gold interface was found to yield TIR at all wavelengths; for the sake of consistency we considered a fixed incident angle of 70 deg to ensure that total internal reflection at the interface is satisfied for all wavelengths. For the *x*-*y* plots, an incident wavelength of 515 nm was used, and for these cases an incident angle of 23.3 deg (at the interface) was assumed.

Scattering profiles were obtained for five different agglomerates shown in Fig. 5. These structures are a subset of many possible agglomerate structures that can be formed by nanoparticles on a substrate. Model I-b represents the single particle case, Model I-a has single monomers dispersed above the surface. Model II indicates two particles separated from each other by distances equal to

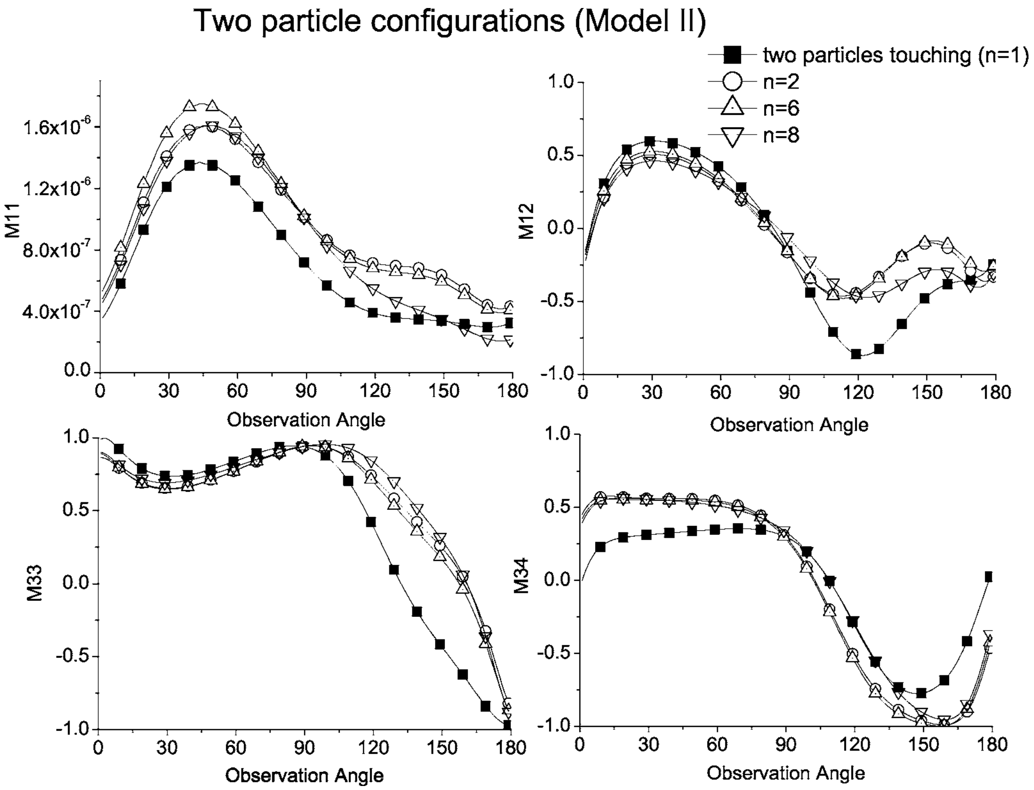


Fig. 9 Scattering profiles for particle configuration shown in Model II, for a particle diameter of 20 nm

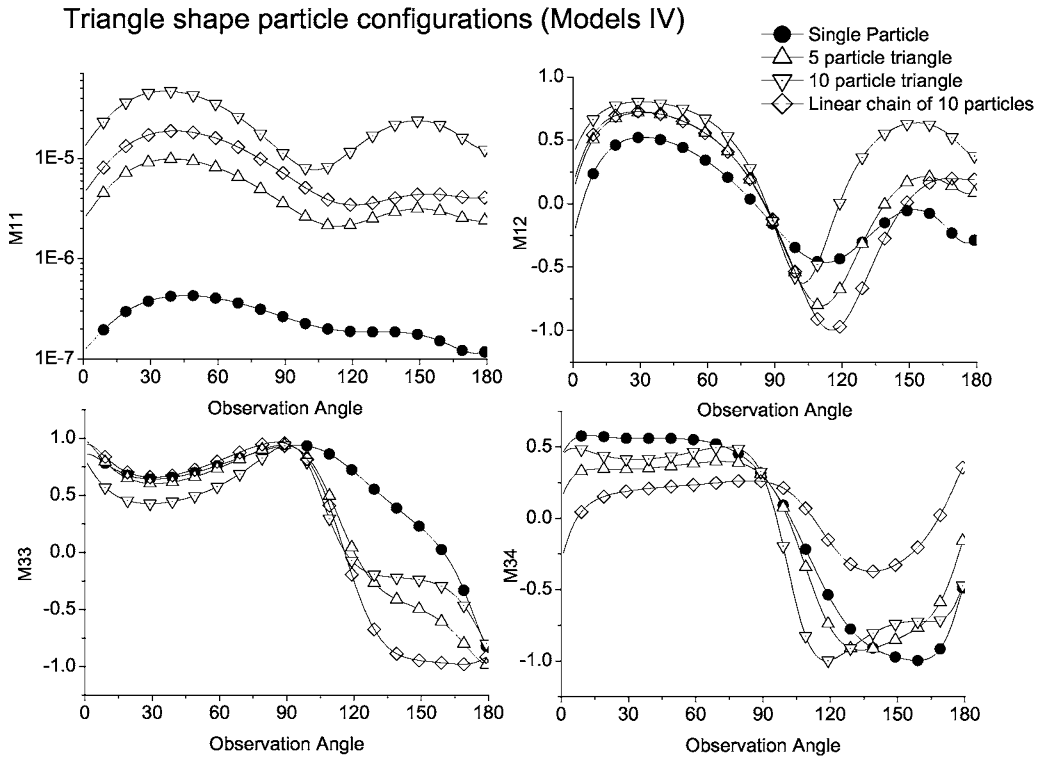


Fig. 10 Scattering profiles for particle configurations shown in Models IV-a and IV-b, for a particle diameter of 20 nm

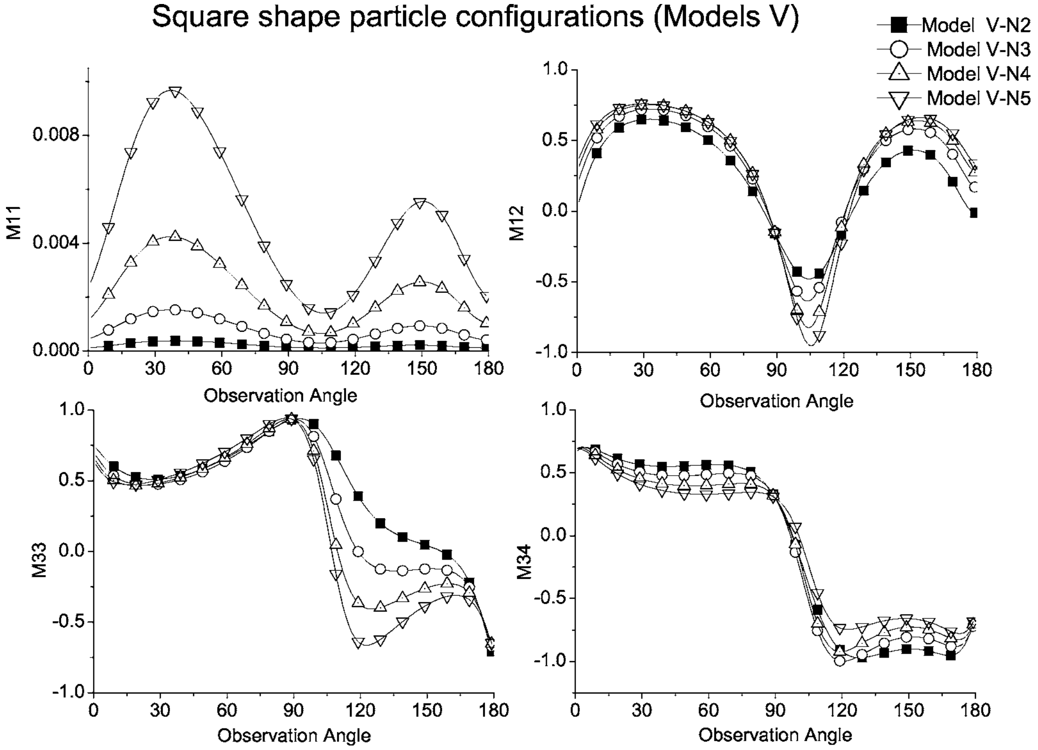


Fig. 11 Scattering profiles for particle configurations shown in Model V, for a particle diameter of 20 nm

integral multiples (n) of the particle diameter, which in this case is 20 nm. Models III-a and III-b are generated to observe the effects of orientation of particles. Models IV and V are to study the effects of triangle- and square-shaped particle configurations. Model V-N2 represents four spherical particles arranged in the form of a 2×2 square, Model V-N3, a 3×3 square consisting of nine spherical particles, and so on.

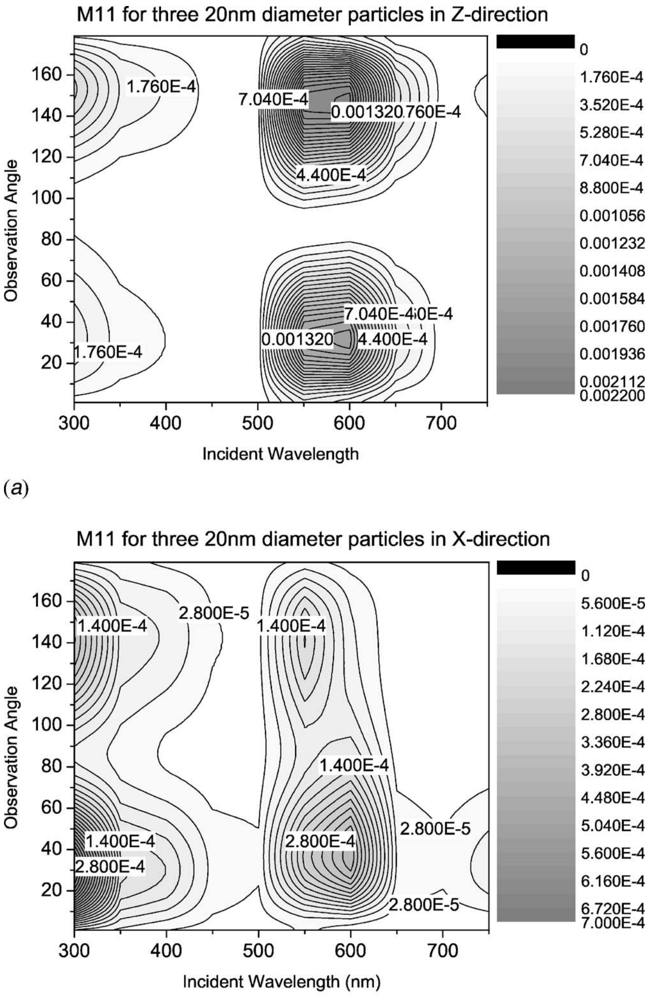
A series of results were obtained for a single gold particle at varying distances from the surface to explore the effect of particle-surface interaction. The effect of the surface is clearly observed in these cases when the gold film is absent. As seen from Figs. 6, there is a considerable difference between the results corresponding to two cases: (a) when the particles are on the surface or (b) when they are away (100 nm) from the surface; both the profile and amplitude of the M_{ij} parameters are sensitive to particle-surface separation. When the particle is placed at a distance of 100 nm from the surface, there is a decrease in the magnitude of M_{11} values. This decrease is owing to the decay of the surface wave from the surface. The M_{11} values follow a similar profile and vary in amplitude; however, the other M_{ij} elements show different profiles when the separation distance is altered. These differences are important and can be used for characterization purposes.

The average value depicted in Fig. 6 corresponds to an average of 10-nm particles placed at uniform separation distances of 10 nm each from the surface (Model I-a). It is clear that locations of particles near the film in the evanescent field affect the profiles of M_{11}, M_{12}, M_{33}, and M_{34}. Figures 6 also show the effect of substrate materials on scattering matrix elements. The choice of substrate has significant effect on all four M_{ij} parameters. The cross-polarization elements are not included in calculating the M_{ij} profiles as they are zero for this case of a single spherical particle.

Additional calculations were carried out by including a thin gold substrate and with particles touching or close to each other. Cross-polarization terms were included in these calculations to obtain accurate M_{ij} values. The centroid of the particle structure was used to calculate the distance of the structure from the surface. The results presented below were obtained for 10-, 20-, or 50-nm-diameter gold particles illuminated at an incident angle of 23.3 deg and at an incident wave length of 515 nm. Figure 7 presents comparisons of all the M_{ij} profiles for three particle configurations shown in models I-b, II, and III-a for a particle diameter of 10 nm. Except for M_{11}, the other M_{ij} profiles do not show any significant change. Hence only the M_{11} values are presented for a particle diameter of 20 nm in Fig. 8. The M_{11} profile is similar for a 50 nm particle diameter (not presented) with an increase of magnitude by about 100 times. These results show that the M_{11} values, though following a similar profile, vary greatly in amplitude for the cases considered. In general, they are more sensitive to the particle diameter than to the number of particles. For the angle of incidence considered here, M_{11} profile reach a peak at about 40 deg observation angle. It is, hence, possible to distinguish the size and structure of the particles from the experimental results by examining the amplitude of M_{11} at this angle.

Figure 9 depicts the scattering profiles for the particle configuration shown in Model II. Variation of the profiles though similar, is distinguishable for particle separations that are equal to the integer multiples of particle diameter (in this case, 20 nm). For agglomerates that have larger numbers of monomers, it is possible that the other M_{ij} values also vary greatly. We observe this for the triangle- and square-shaped configurations shown in Figs. 10 and 11, respectively. For triangle-shaped configurations, it is noted that a ten-particle triangle (Model IV-b) scatters about ten times higher intensity compared to a linear chain of ten particles and nearly 100 times more than a single particle. This configuration is observed to selectively scatter significantly at specific directions.

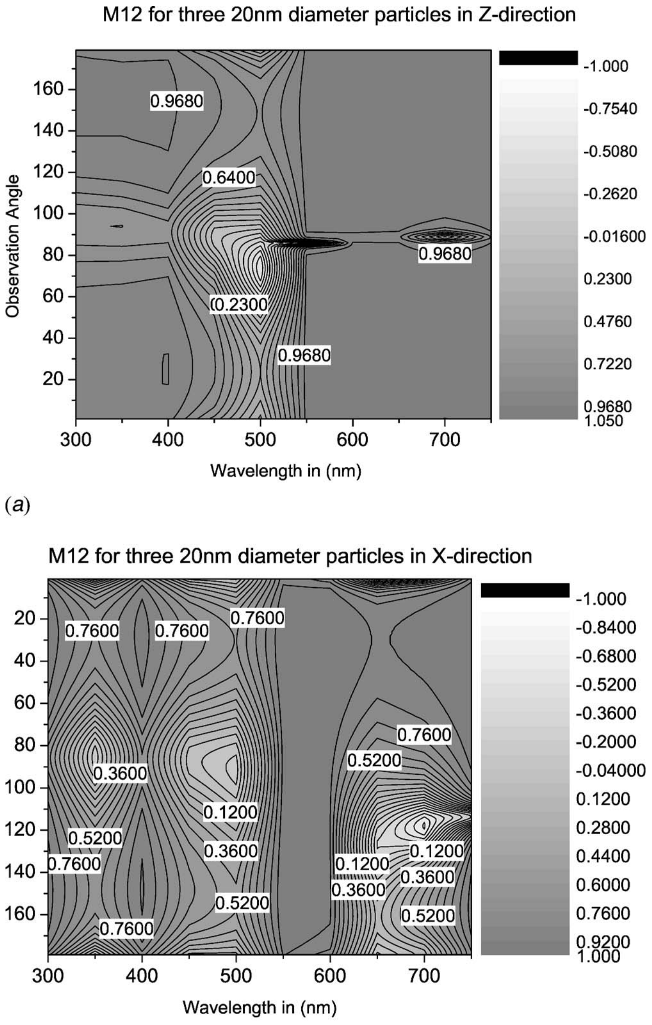
From Fig. 11, we observe that the increase in magnitude of M_{11} parameter, for configurations in Model V, is not large even though the number of particles increases to 25 (for Model-VN5) from 4 (for Model-VN2). This is not the case for particle configurations in Model IV, where there is an exponential increase in M_{11} values



(*a*)

(*b*)

Fig. 12 M_{11} **profiles for 20-nm-diameter particles placed on the surface for arrangements shown in Model III-a (*a*) and III-b (*b*)**



(*a*)

(*b*)

Fig. 13 M_{12} **profiles for 20-nm-diameter particles on the surface for arrangements shown in Model III-a (*a*) and III-b (*b*)**

when the number of particles increases from six to ten. A triangular configuration of particles seems to provide for higher scattering intensities.

In order to show the wavelength dependency of the profiles, additional results are presented as contour plots for different wavelengths; the variation of the M_{ij} parameters with wavelength for two different orientations (Model III-a and Model III-b) are depicted in Figs. 12–15. It is observed that the variation in results is most obvious in the red-wavelength region. We also observe a change of directionality for M_{12}, M_{33}, and M_{34} parameters for wavelengths greater than 600 nm. A gradual change can be observed between 500 and 600 nm. The scattering profiles are significantly different for 50 nm diameter particles, whereas they appear similar for 10-nm- and 20-nm-diameter particles. The contour plots for the two cases (Model III-a and III-b) are also different, providing information about the orientation of the structure. It can be seen that configurations in Model III-b have a strong second peak in the 150–170 deg range for the M_{11} parameter, with the magnitude also being an order high when compared to the configurations in Model III-a). Also note the change in signs of M_{12}, M_{33}, and M_{34} parameters for Model III-b predictions when the wavelength is longer than 550 nm.

The M_{ij} profiles obtained at a single wavelength may not be sufficient to obtain the structure information. The scattering response of 2D agglomerated structures requires calculation of M_{ij} at different wavelengths as well. Because sensitivity on M_{ij} profiles, due to the structure, varies with the wavelength of the incident light, and depends on the optical properties of both the surface and the particles, only bidirectional M_{11} profiles as a function wavelength is sensitive to size and shape. Therefore, M_{12}, M_{33}, and M_{34} parameters need to be considered simultaneously to obtain information about the structure of an agglomerate on the surface of a substrate. For example M_{33} at 120 deg (shown in Fig. 11) is very sensitive to the size of square agglomerates and can effectively be used for diagnostic purposes.

Conclusions

In this paper, we introduce a new hybrid methodology to explore the angular profiles of the scattering matrix elements of nano-sized particles and agglomerates on or near a metallic surface. The concept is a novel one wherein the T-matrix method is coupled with a transformation to include the effects of the surface. The theoretical framework is outlined and results are obtained for several different particle sizes and structures. The results show that characterization of gold nanosized particles on or near the surface is possible by using a combination of Mueller matrix element at certain wavelengths and certain observation angles. Using these windows of sensitivity of the M_{ij} elements, the size and structure of nanosized particles can be obtained. The accuracy of present results is limited with the summation assumption made when B-matrix is calculated and the approximation on the surface

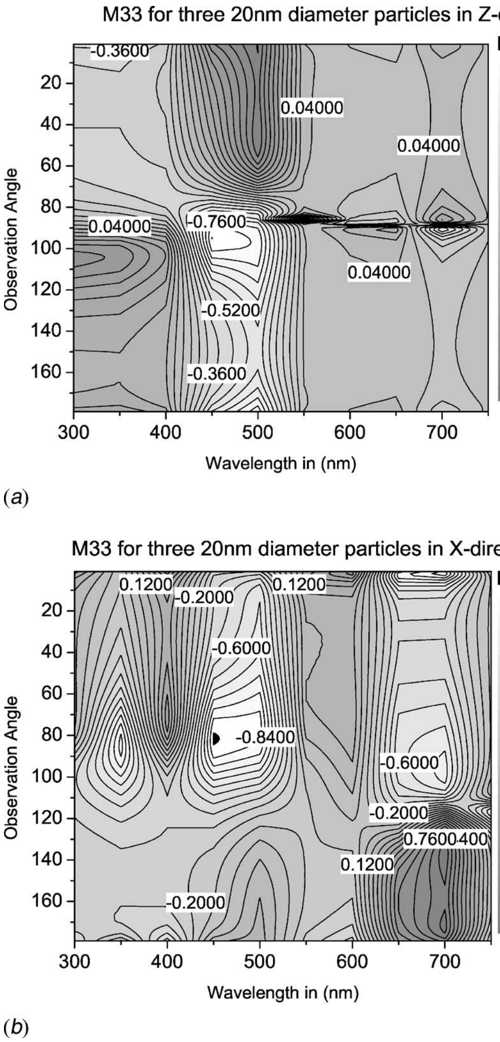


Fig. 14 M_{33} **profiles for 20-nm-diameter particles on the surface for arrangements shown in Model III-a (*a*) and III-b (*b*)**

interaction fields. This approach can be used effectively if the thickness of the film is greater than the location and size of the nanosized structures. In the future, the extension of this method to a larger group of agglomerates and three-dimensional structures will be considered, and the results will be reported in separate studies.

Acknowledgment

This work is partially sponsored by a National Science Foundation grant (NSF-NER DMI-0403703, 2004–2006). Additional funding is received for P. Garudadri Venkata from the Department of Mechanical Engineering, University of Kentucky. The authors also thank Dr. Mackowski for making his computer program available for this study.

References

[1] McMurry, P. H., and Naqwi, A. A., 1998, Proceedings of the Fifth International Congress on Particle Sizing, Minneapolis, MN.

[2] Mishchenko, M. I., Hovenier, J. W., and Travis, L. D., 2001, *Light Scattering by Non-Spherical Particles: Theory, Measurements, and Applications*, Academic Press, New York.

[3] Mishchenko, M., Videen, G., Babenko, V. A., Khlebtsov, N. G., and Wriedt, T., 2004, "T-matrix Theory of Electromagnetic Scattering by Particles and its Applications: A Comprehensive Reference Database," J. Quant. Spectrosc. Radiat. Transf., **88**, p. 357.

[4] Waterman, P. C., 1971, "Symmetry, Unitarity, and Geometry in Electromagnetic Scattering," Physica D, **3**, p. 825.

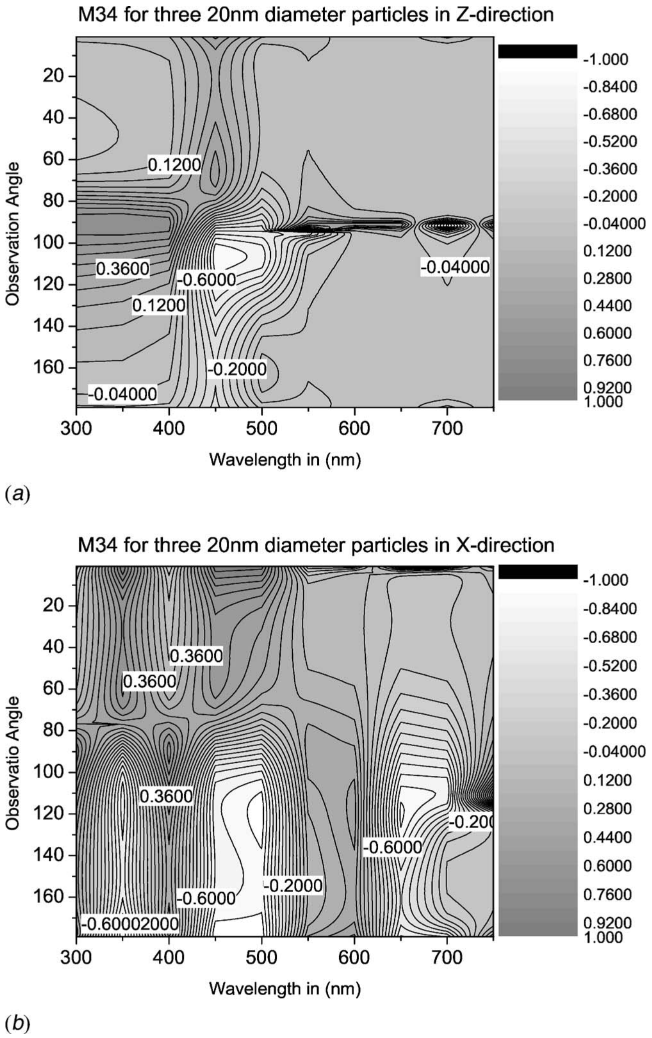


Fig. 15 M_{34} **profiles for 20-nm-diameter particles on the surface for arrangements shown in Model III-a (*a*) and III-b (*b*)**

[5] Peterson, B., and Ström, S., 1973, "*T* matrix for Electromagnetic Scattering From an Arbitrary Number of Scatterers and Representations of E(3)," Physica D, **8**, p. 3661.

[6] Lakhtakia, A., Varadan, V. K., and Varadan, V. V., 1985, "Scattering and Absorption Characteristics of Lossy Dielectric, Chiral, Nonspherical Objects," Appl. Opt., **24**, p. 4146.

[7] Bruning, J. H., and Lo, Y. T., 1971, "Multiple Scattering of E. M. Waves by Spheres. I. Multipole Expansion and Ray-Optical Solutions," IEEE Trans. Antennas Propag., **19**, p. 378.

[8] Tsang, L., and Kong, J. A., 2001, *Scattering of Electromagnetic Waves: Advanced Topics*, John Wiley & Sons, New York.

[9] Varadan, V. K., Bringi, V. N., and Varadan, V. V., 1979, "Coherent Electromagnetic Wave Propagation Through Randomly Distributed Dielectric Scatterers," Physica D, **19**, p. 2480.

[10] Borghese, F., Denti, P., Saija, R., and Toscano, G., 1987, "Optical Absorption Coefficient of a Dispersion of Clusters Composed of a Large Number of Spheres," Aerosol Sci. Technol., **6**, p. 173.

[11] Mackowski, D. W., 1994, "Calculation of Total Cross Sections of Multiple-Sphere Clusters," J. Opt. Soc. Am. A, **11**, p. 2851.

[12] Mackowski, D. W., and Mishchenko, M. I., 1996, "Calculation of the *T* Matrix and the Scattering Matrix for Ensembles of Spheres," J. Opt. Soc. Am. A, **13**, p. 2266.

[13] Gonzalez, F., Videen, G., Valle, P. J., Saiz, J. M., De la Peña, J. L., and Moreno, F., 2001, "Light Scattering Computational Methods for Particles on Substrates," J. Quant. Spectrosc. Radiat. Transf., **70**, p. 383.

[14] Wriedt, T., and Doicu, A., 1998, "Light Scattering From a Particle on or Near a Surface," Opt. Commun., **152**, p. 376.

[15] De la Peña, J. L., Gonzalez, F., Saiz, J. M., Moreno, F., and Valle, P. J., 1999, "Sizing Particles on Substrates: A General Method for Oblique Incidence," J. Appl. Phys., **85**, p. 432.

[16] Videen, G., Turner, M. G., Iafelice, V. J., Bickel, W. S., and Wolfe, L. W., 1993, "Scattering From a Small Sphere Near a Surface," J. Opt. Soc. Am. A,

10, p. 118.
[17] Videen, G., Aslan, M. M., and Mengüç, M. P., 2005, "Characterization of Metallic Nano-Particles via Surface Wave Scattering: A. Theoretical Framework and Formulation," J. Quant. Spectrosc. Radiat. Transf., **93**, p. 195.
[18] Aslan, M. M., Mengüç, M. P., and Videen, G., 2005, "Characterization of Metallic Nano-Particles via Surface Wave Scattering: B. Physical Concept and Numerical Experiments," J. Quant. Spectrosc. Radiat. Transf., **93**, p. 207.
[19] Bohren, C. F., and Huffman, D. R., 1983, *Absorption and Scattering by Small Particles*, John Wiley & Sons, New York.
[20] Videen, G., 1993, "Light Scattering From a Sphere Behind a Surface," J. Opt. Soc. Am. A, **10**(1), p. 110.
[21] Garudadri Venkata, P. K., 2006, "Surface Wave Scattering From Metallic Nano Particles and Agglomerates: Theoretical Framework and Numerical Analysis," Master's thesis, University of Kentucky, Lexington, KY.
[22] Feldheim, D. L., and Foss, C. A., 2002, *Metal Nanoparticles: Synthesis, Characterization, and Application*, Marcel Dekker, New York.
[23] Weaver, J. H., and Frederikse, H. P. R., 1978, *Optical Properties of Metals and Semi Conductors*, CRC Press, Boca Raton, FL.

Modeling the Radiative Properties of Microscale Random Roughness Surfaces

Kang Fu

Pei-feng Hsu[1]
e-mail: phsu@fit.edu

Department of Mechanical and Aerospace Engineering,
Florida Institute of Technology,
Melbourne, FL 32901

The radiative properties of engineering surfaces with microscale surface texture or topography (patterned or random roughness and coating or multi-layer) are of fundamental and practical importance. In the rapid thermal processing or arc/flash-assisted heating of silicon wafers, the control of thermal energy deposition through radiation and the surface temperature measurement using optical pyrometry require in-depth knowledge of the surface radiative properties. These properties are temperature, wavelength, doping level, and surface topography dependent. It is important that these properties can be modeled and predicted with high accuracy to meet very stringent temperature control and monitor requirements. This study solves the Maxwell equations that describe the electromagnetic wave reflection from the one-dimensional random roughness surfaces. The surface height conforms to the normal distribution, i.e., a Gaussian probability density function distribution. The numerical algorithm of Maxwell equations' solution is based on the well-developed finite difference time domain (FDTD) scheme and near-to-far-field transformation. Various computational modeling issues that affect the accuracy of the predicted properties are quantified and discussed. The model produces the bi-directional reflectivity and is in good agreement with the ray tracing and integral equation solutions. The predicted properties of a perfectly electric conductor and silicon surfaces are compared and discussed. [DOI: 10.1115/1.2401200]

Introduction

Electromagnetic (EM) wave scattering is an active, interdisciplinary area of research with myriad practical applications in fields ranging from atomic physics, energy conversion, and medical imaging to geosciences and remote sensing. In particular, the subject of wave scattering by random rough surfaces presents great theoretical challenges due to the large degrees of freedom in these systems and the need to include multiple scattering effects accurately. In the past three decades, considerable theoretical progress has been made in elucidating and understanding the scattering processes involved in such problems. Rapid advances in computer modeling have opened new approaches, for example, finite difference, finite element, and Monte Carlo simulations, in the numerical analysis of random media scattering. Numerical simulations allow us to solve the Maxwell equations exactly without the limitations of analytical approximations, whose regimes of validity are often difficult to assess [1].

The radiative properties of engineering surfaces with various surface topography (patterned or random roughness and coating) are of fundamental and practical importance. The properties are widely used in many diverse disciplines, for example, thermal systems design [2], remote sensing, computer graphics, composite materials, photonic devices, optical thin films, lasers, optical tomography, x-ray lithography, and semiconductor manufacturing [3]. In this study, the problem of interest is scattering of the incident wave with wavelengths in the ultraviolet to infrared range from surfaces with micron or submicron sized textures, i.e., the length scale is on the order of an incident electromagnetic source's wavelength.

Direct measurement of the surface property under well-defined conditions will be ideal. However, the measurement becomes increasingly difficult when the detected (or reflected) angle is close to the incident angle, high temperature sample measurement is performed, or partial coherence effects are important [4]. It is also not easy to produce well-defined, microscale, random roughness surfaces [5]. Reliable models for predicting and understanding the surface radiative properties are thus essential. There are various means to predict the surface radiative properties, among which the bi-directional reflectivity or bi-directional reflection distribution function (BRDF) is of particular interest in this study [6]. A commonly used modeling technique is ray tracing [7,8]. The method considers the multiple reflections at the surface as well as the shadowing effects. At each reflection, a Fresnel reflection from a local, optically smooth surface is assumed. In the case of thin film coating on the surface, a wave optics reflection [9] was used to account for the reflected wave interference. However, the latter approach is conceptually inconsistent with ray tracing (a geometric optics treatment). Such a hybrid method was examined in a recent work [10] by comparing with a rigorous EM solution. It is determined that the hybrid method is applicable for a wide range of one-dimensional (1D) rough surfaces.

This study uses a numerical scheme to solve the Maxwell's EM equations and examines the wave interference effects on the surface reflectivity without the simplifications assumed in the ray tracing and other methods. The solution scheme is the finite difference time domain (FDTD) method [11].

The FDTD method is well known and has been widely used over the last two decades. It has been traditionally used in EM wave scattering and radar cross section study [12,13]. This method is particularly suitable to determine the surface reflectivity of the microscale surface roughness at various incident wavelengths and roughnesses. The method is conceptually simple and systematic and can integrate with other physical processes, e.g., heat transfer, readily. The time stepping scheme is explicit but can be improved with implicit time stepping to go beyond the Courant-Friedrichs-Lewy (CFL) limit [14]. Some of the schemes frequently used in computational fluid dynamics and heat transfer, for example, the alternating-direction implicit (ADI) method, can be incorporated into FDTD to use various time step sizes. The finite volume approach can also replace the finite difference

---

[1] Corresponding author.

Contributed by the Heat Transfer Division of ASME for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received January 9, 2006; final manuscript received May 17, 2006. Review conducted by Zhuomin M. Zhang.

scheme to handle irregular geometries [15]. In this paper, a conventional FDTD scheme with perfectly matching layer (PML) boundary was used [16].

The surfaces used in this study were one-dimensional, i.e., the surface height (ζ) is a function of one spatial coordinate $\zeta = f(x)$. The surfaces under consideration have 1D random roughness structure. A random roughness can be described by a height distribution function, $f(x)$, and a height correlation function. Both functions are assumed to be Gaussian in this study. Two-dimensional surfaces with $\zeta = f(x,y)$ will be considered in a follow-up work. This is another advantage of the FDTD method—it can be readily extended to higher dimensional geometry. Since the FDTD scheme is based on the rectangular mesh, the surface slopes are represented by jagged edges of the discretized rectangular volume elements. The element size needs to be less than the incident wavelength. Therefore the mesh size effect on the solution accuracy has to be examined. The simulation time step size also needs to be controlled according to the CFL criterion, which is necessary for the explicit scheme. The total number of time steps also depends on the surface size used in the simulation and its effect will be discussed as well. Several other factors that impact modeling accuracy were also considered. Finally, the FDTD solutions are compared with the ray tracing solutions and the integral solutions of EM wave scattering. The predicted properties of a perfectly electric conductor and silicon surfaces were compared and discussed. The reflection characteristics from gold surfaces with random roughness will be considered in a follow-up paper. The gold surfaces have particularly interesting phenomena, e.g., surface plasmon, and there exist highly accurate experimental data for comparison with predictions.

Theoretical Development

One-Dimensional Random Roughness Surface. In this study, the random roughness surface is considered to be homogeneous and isotropic [1]. A statistically homogeneous surface means the distribution of heights is equally possible at any position within the surface. An isotropic surface is that the correlation function of surface heights is independent of the direction between two corresponding surface points' locations. A real wafer backside surface may not be isotropic, as shown in Lee et al. [17]. If the surface height ζ conforms to a normal (Gaussian) distribution, it follows a probability density distribution shown as the following:

$$f(\zeta) = \frac{1}{\delta\sqrt{2\pi}} e^{-\zeta^2/2\delta^2} \tag{1}$$

where δ is the standard deviation. In this distribution, the mean value of ζ is equal to 0 and δ is the rms roughness of a surface. Large δ value corresponds to large roughness. There are several other functions besides the normal distribution function that can describe the characteristics of surface roughness [18]. It is noted that Eq. (1) does not contain the correlation distance, but in a real surface the height at any position in the rough surface is correlated with the heights of the adjacent points. The correlation distance could be described by the correlation function $B(T)$, which has the following definition

$$B(T) = \lim_{L\to\infty} \frac{1}{2L} \int_{-L}^{L} \zeta(x)\zeta(x+T)dx \tag{2}$$

T is the distance between the two points. In some literature, e.g., Tsang et al. [1], T is expressed as a position vector. The normalized form of $B(T)$ could be written as

$$C(T) = \frac{B(T) - \langle\zeta\rangle^2}{D(\zeta)} = \frac{\langle\zeta_1\zeta_2\rangle}{\langle\zeta_1^2\rangle} \tag{3}$$

where $D(\zeta) = \langle\zeta^2\rangle$ is the variance [19] and the C value is between 0 and 1. It follows that, for a purely random surface, $\lim_{T\to 0} C(T) = 1$; for two independent points, $\lim_{T\to\infty} C(T) = 0$. The correlation distance τ is defined in such a way that $C(\tau) = e^{-1}$.

To generate a random surface, a specific form of the correlation function has to be used. Commonly used function forms are the Gaussian, exponential, etc. [1,18,19]. The Gaussian correlation function given below was used in this study

$$C(T) = e^{-T^2/\tau^2} \tag{4}$$

If the surface topographic or height data are available, the $C(T)$ function can be evaluated from the data. Furthermore, due to the surface anisotropy, the actual correlation function may depend on the direction of T (position vector) or azimuthal angle as shown in [10].

Finite Difference Time Domain Method. The direct numerical simulation of electromagnetic wave propagation can be carried out with the finite difference time domain method solution of the Maxwell equations. The rigorous solution allows one to understand the scattering process and compare with the experimental measurements with a few assumptions typically used in the analytical models. The equations take the following form for a linear isotropic material:

$$\varepsilon\frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{H} - \sigma\mathbf{E} - J_s \tag{5}$$

$$\mu\frac{\partial \mathbf{H}}{\partial t} = -\nabla \times \mathbf{E} - \sigma^*\mathbf{H} - M_s \tag{6}$$

where **E** and **H** represent the electric and magnetic fields, respectively; J_s and M_s are the electrical and magnetic field sources; and ε, μ, σ, and σ^* are the permittivity, permeability, electrical conductivity, and magnetic conductivity or equivalent magnetic loss of the material, respectively. In the current problem, J_s and M_s are zero. To demonstrate the application of Maxwell equations in the surface roughness geometry, a two-dimensional formulation is discussed below. Extension to three-dimensional geometry, e.g., patterned wafer, is straightforward. The 2D structure is assumed to be extended infinitely in the z direction so that all physical parameters and electromagnetic field are functions of x and y coordinates only. Maxwell equations can be decomposed into two independent sets of equations, the transverse electric (TE) polarization and the transverse magnetic (TM) polarization. By replacing the electric and magnetic fields with their vector components, the 2D TE set of equations is:

$$\varepsilon\frac{\partial E_z}{\partial t} = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - \sigma E_z \tag{7a}$$

$$\mu\frac{\partial H_x}{\partial t} = -\frac{\partial E_z}{\partial y} - \sigma^* H_x \tag{7b}$$

$$\mu\frac{\partial H_y}{\partial t} = \frac{\partial E_z}{\partial x} - \sigma^* H_y \tag{7c}$$

The 2D TM set of equations is similarly obtained for H_z, E_x, and E_y. In fact, the TM set is the dual form of the TE set. Therefore, solution to either set of equations will take the same mathematical form. The above equations are coupled, linear, first order partial differential equations. An equivalent form is two uncoupled, second order Helmholtz equations, i.e., wave equations, for both fields [6].

Equations (7a)–(7c) are expressed in the finite difference time domain equations based on the mesh shown in Fig. 1 [11,20]. The FDTD solution procedure of Eq. (7) is straightforward. However, several issues with FDTD need to be considered: (a) special treatment of the boundary condition is required to avoid mixing the reflected wave from the artificial computational domain boundary with the forward and backward (or reflected) propagation waves.

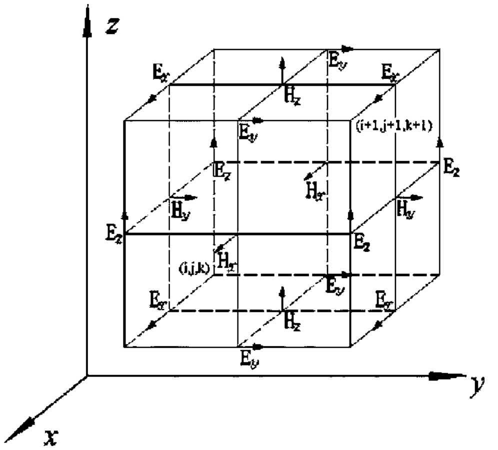


Fig. 1 Discretization and field vector components used in FDTD analysis

At the appropriate boundary surfaces, it is necessary to add the "perfectly matched layer" (PML) as the boundary condition [14,21]; (b) FDTD is not convenient to handle curved geometry; and (c) to avoid a checkerboard pattern in the solution, the staggered grid is used [20,22], i.e., **E** and **H** components are not solved at the same node points (Fig. 1). An alternative scheme will be the high order upwind scheme with finite volume mesh [13,23]. This can handle arbitrary geometry and avoid dispersion and diffusion errors usually associated with the hyperbolic wave equations [14].

For a two-dimensional geometry, the FDTD form of TE set, Eq. (7), can be written for the mid-plane in Fig. 1, i.e., the x-y plane that contains E_z, H_x, and H_y. The finite difference form of the TM equation set can be written for the top or bottom plane. The details are readily available in the original paper by Yee [20] or in any numerical electromagnetics text [11,24,25]. Therefore, the details are not repeated here and only the essential difference equations are given below. As shown in Fig 1, **E** and **H** are spatially shifted by 1/2 space increment, and the calculation of them is temporally shifted by 1/2 time increment, so that all unknown variables at any time step can be calculated based on the variables at the previous half-and full-time step status. With the second-order central difference scheme, Maxwell equations of Eq. (7) in 2D can be expressed as:

$$E_z^n(i,j) = \frac{\varepsilon - \sigma\Delta t/2}{\varepsilon + \sigma\Delta t/2} E_z^{n-1}(i,j) + \frac{\Delta t}{\varepsilon + \sigma\Delta t/2}\left\{ \frac{1}{\Delta x}[H_y^{n-1/2}(i+1/2,j) - H_y^{n-1/2}(i-1/2,j)] - \frac{1}{\Delta y}[H_x^{n-1/2}(i,j+1/2) - H_x^{n-1/2}(i,j-1/2)] \right\} \qquad (8a)$$

$$H_x^{n+1/2}(i,j+1/2) = \frac{\mu - \sigma^*\Delta t/2}{\mu + \sigma^*\Delta t/2} H_x^{n-1/2}(i,j+1/2) - \frac{\Delta t}{(\mu + \sigma^*\Delta t/2)\Delta y}[E_z^n(i,j+1) - E_z^n(i,j)] \qquad (8b)$$

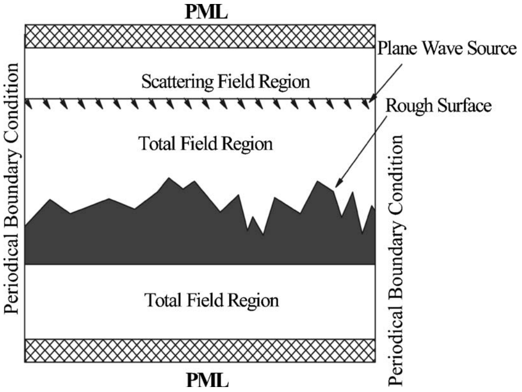


Fig. 2 The computation zones and boundary conditions used in the EM wave reflection calculation from a random roughness surface

$$H_y^{n+1/2}(i+1/2,j) = \frac{\mu - \sigma^*\Delta t/2}{\mu + \sigma^*\Delta t/2} H_y^{n-1/2}(i+1/2,j) + \frac{\Delta t}{(\mu + \sigma^*\Delta t/2)\Delta x}[E_z^n(i+1,j) - E_z^n(i,j)] \qquad (8c)$$

where the superscript of the field components stands for time step and the subscript represents the spatial node location. Note that in Eq. (8), σ and σ^* are set to zero for dielectric and nonmagnetic materials. According to Yee's notation, a spatial point in a Cartesian coordinate is written as $(i,j,k)=(i\Delta x, j\Delta y, k\Delta z)$ as shown in Fig. 1. Here, Δx, Δy, and Δz are the lattice space increments in the x, y, and z coordinate directions, and i, j, k are integers.

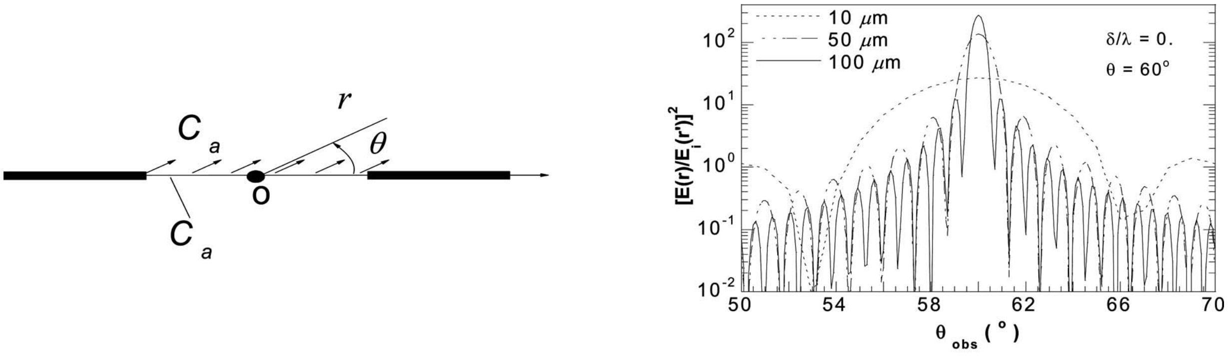
If the interface of different materials is parallel to one of coordinate axes, e.g., Eq. (7) applies to reflection from an interface plane parallel to the z axis in Fig. 1, continuity of tangential **E** and **H** is naturally maintained; the algorithm itself is divergence-free in the absence of electric and magnetic charge; the time-stepping algorithm is nondissipative, which means EM wave propagating within the mesh will not decay due to the algorithm itself. All these make Yee's algorithm a robust method in solving EM field issues.

Figure 2 illustrates the geometry and computational zone in the FDTD solution of the random rough surface scattering problem. The computational zone is divided into scattering field (SF) and total field (TF) by a virtual plane, at which a plane incident wave into the TF region is initiated at $t=0$ as the radiation source. In the SF, the **E** and **H** components in calculation belong only to the scattered wave, but in the TF region, they include contribution from both the scattered wave and incident wave. The whole region is surrounded by two PMLs (top and bottom) and two periodic boundary conditions on the left and right sides.

The FDTD method calculates the transient electromagnetic fields throughout the computational domain under the excitation of a monochromatic harmonic field. The results of interest in this study are the radiative properties. These are associated with the electromagnetic fields through the intensity. The intensity of the harmonic wave is defined as a mean value of the Poynting vector and its magnitude is expressed as

$$S(x,y) = I(x,y)d\Omega = \frac{1}{T^*}\int_0^{T^*} |\mathbf{E} \times \mathbf{H}| dt \qquad (9)$$

where $T^*=2\pi/\omega$ is the time period of the wave. The reflectivity is defined as a ratio of the reflected and incident intensities and the transmissivity is defined as a ratio of the transmitted and incident intensities. For the TE polarization, the reflectivity R_e is calculated at the far field (see below on the near field to far field transforma-



(a)

(b)

Fig. 3 (*a*) Virtual integral contour C_a used in NTFF transformation. (*b*) Diffraction patterns at the far field position after the transformation.

tion) above the rough surface and has the following form:

$$R_e = \frac{\int_0^{T^*} |\mathbf{E} \times \mathbf{H}| dt}{\int_0^{T^*} |\mathbf{E}_{in} \times \mathbf{H}_{in}| dt} \tag{10}$$

where the incident electric and magnetic fields $\mathbf{E}_{in}$ and $\mathbf{H}_{in}$ are input and $\mathbf{E}$ and $\mathbf{H}$ are obtained from the FDTD solution and far field transformation. For an unpolarized wave, its reflectivity is taken to be an arithmetic average of the reflectivities from TE and TM polarizations, that is,

$$R = 0.5(R_e + R_m) \tag{11}$$

where R_m is the TM mode polarization reflectivity. Other properties can be found similarly or using the energy balance relation. The relations given above are wavelength dependent. It is straightforward to obtain wavelength-integrated or total property.

Near Field to Far Field Transformation. The **E** and **H** components obtained from the FDTD solution are near field quantities. These need to be transformed to far field values to be related to the observable, macro-scale properties, e.g., Eq. (10). Taflove and Hagness [11] gave detailed derivation based on Green's theorem. The electromagnetic field at any location and any time is obtained from the FDTD scheme, which is composed of waves in different directions at various mesh points. Such results cannot be easily described by the Poynting vector. Thus, the near-to-far-field (NTFF) transformation is needed. The transformation is based on the harmonic property of EM wave equations, with which Green's theorem is applied. The NTFF transformation yields the far-field EM results from near-field values on a selected virtual contour in the SF region. The contour encloses the random roughness surface.

To apply NTFF transformation, the scalar values of the EM field need to be converted into complex form by Fourier transformation. The complex value contains both the amplitude and phase information. Assume V is the complex form (**E** or **H**), and $v(t)$ is its time-dependent scalar value, i.e., $|V|\cos(\omega t + \phi) = v(t)$, where ω is the angular velocity and ϕ is the phase. Thus, the Fourier transform of $v(t)$:

$$\begin{aligned} \int_t^{t+T} v(t) e^{i\omega t} dt &= \int_t^{t+T} |V| \cos(\omega t + \phi) e^{i\omega t} dt \\ &= |V| \left[\int_t^{t+T} \cos(\omega t + \phi)\cos(\omega t) dt \right. \\ &\quad \left. + i \int_t^{t+T} \cos(\omega t + \phi)\sin(\omega t) dt \right] \\ &= |V| \frac{1}{2} \cos(\phi) T - i |V| \frac{1}{2} \sin(\phi) T \end{aligned} \tag{12}$$

where T is the period, and $|V|$ and ϕ can be easily calculated using this equation with the given $v(t)$. With the complex form of $\breve{E} = |V| e^{i\omega t}$ the far field $\breve{E}_z$ can be calculated by

$$\begin{aligned} \lim_{k|\bar{r} - \bar{r}'| \to \infty} \breve{E}_z(\vec{r}) = \frac{e^{-ikr}}{\sqrt{r}} \frac{e^{i(\pi/4)}}{\sqrt{8\pi k}} \oint_{C_a} \{ \omega \mu_0 \hat{z}' \cdot [\hat{n}_a' \times \breve{\mathbf{H}}(\vec{r'})] \\ + k \hat{z}' \times [\hat{n}_{\prime}^{a} \times \breve{\mathbf{E}}(\vec{r'})] \cdot \hat{r} \} e^{ik\hat{r}\vec{r'}} dC'(\vec{r'}) \end{aligned} \tag{13}$$

where $\hat{z}'$ is a unit vector in the z direction, and $\hat{n}_a'$ is a unit vector normal to the integral contour C_a, $\bar{r}'$ is position vector on C', $\hat{r}$ is unit vector pointing to the far field position, and $\vec{r}$ is the far field position vector. C_a is an arbitrary shape curve and, in this case, can be assigned a rectangle that encloses the scattered field. Figure 3(*a*) shows a portion of a closed integration path C_a. Below is a simple test to verify the accuracy of near-to-far-field transformation of planar waves.

Let $\theta = 60°$, $\lambda = 1$ μm, amplitude of $E_z = 1000$ V/m, and the far field integration use angular resolution of 1 deg (interval between two neighboring angles in Fig. 3(*a*)). Three virtual surface lengths of 10, 50, and 100 μm were used in the path integral of Eq. (13). The results are plotted in Fig. 3(*b*). The results show that a small surface patch (10 μm) of planar wave generates a very distinctive Fraunhofer diffraction pattern. Initially, angular resolution of 1 deg was used for 50 and 100 μm surface integrations but was soon found to be insufficient. The diffraction pattern was totally missing in the larger surfaces. This is consistent with the observation of the optical diffraction—the diffraction is much less significant from larger light sources. Thus, the 0.1 deg resolution was used for larger surfaces. Then, the diffraction pattern appeared and the size of the pattern was enlarged as expected for the Fraunhofer diffraction (Fig. 3(*b*)).

Table 1 FDTD computed reflectivity of a thin film

Case	N_λ	S	$\tilde{v}_{axis}/c$	R
(a)	5.8377	0.06024	0.9472	0.61
(b)	14.5943	0.06024	0.9922	0.080
(c)	20.4320	0.06024	0.9960	0.482

Results and Discussion

The computations were carried out on a 2.2 GHz AMD Opteron 848 processor Linux server. The run time depends on the surface size. A calculation using 60 surfaces with each surface length equal to 50λ took about 10 h for 10,000 time steps. Several test cases were considered to examine various parameters relevant to the problem: mesh size, number of time steps for steady state results, resolution of surface geometry, and the numerical dispersion. In the end, BRDFs, of four random roughness perfectly electric conductive surfaces with various degrees of roughness are compared with those obtained by ray tracing and integral equation solutions of Maxwell's equations [26]. The BRDF of silicon surface is also given. The effect of large incident angle is discussed.

Numerical Dispersion. As in other numerical methods for solving wave equations, numerical dispersion needs to be considered when the finite difference approximation is applied in solving Maxwell equations. The apparent characteristics of numerical dispersion are (1) the phase velocity of a numerical wave differs from the physical propagating wave, and the amount of difference varies with wavelength, direction of propagation, and discretization; (2) nonphysical results, for example, a pseudo-reflection on the underside of a perfectly reflective surface; and (3) imprecise superstition of multiple scattered waves, etc. The numerical phase velocity propagating in vacuum along grid axes and that along grid diagonals can be derived as [11]:

$$\tilde{v}_{axial} = \pi c\left\{N_\lambda \sin^{-1}\left[\frac{1}{S}\sin\left(\frac{\pi S}{N_\lambda}\right)\right]\right\}^{-1} \tag{14a}$$

$$\tilde{v}_{diagonal} = \pi c\left\{N_\lambda \sqrt{2}\sin^{-1}\left[\frac{1}{S\sqrt{2}}\sin\left(\frac{\pi S}{N_\lambda}\right)\right]\right\}^{-1} \tag{14b}$$

where $N_\lambda = \lambda/(n\Delta\eta)$, $S = c\Delta t/(n\Delta\eta)$, $\Delta\eta$ is grid interval, Δt is the time step, and n is the medium's refractive index. S is the well-known CFL criterion [14]. From the above relations, the numerical phase velocity approaches to physical light speed when $N_\lambda \to \infty$, which means infinitely small mesh size. Furthermore, to maintain the stability of the FDTD algorithm, $S<1$ is required for 1D calculations. Similarly, $S\sqrt{2}<1$ for 2D and $S\sqrt{3}<1$ for 3D calculations are needed.

When the interference phenomenon is strong, the numerical dispersion may impact the final results significantly, especially when the wave goes through a long distance without dissipation. To test the effect of numerical dispersion in the current code, the application of the FDTD to determine the reflectivity of a single layer silicon thin film is illustrated. The film is placed in vacuum, with thickness $d=10\ \mu$m and refractive index $n=3.426$ (nonabsorbing). Since periodic boundary conditions are applied on both sides (Fig. 3), the calculation result is independent of the film length. The incident radiation source is normal to the thin film surface and in TE mode. The incident wavelength in vacuum is $\lambda=1.0\ \mu$m. According to the Fresnel equation, the theoretical reflectivity is $R=0.7098$. Three different cases are examined: (a) $\Delta x=\Delta y=1/20\ \mu$m, (b) $\Delta x=\Delta y=1/50\ \mu$m, and (c) $\Delta x=\Delta y=1/70\ \mu$m. The step time is chosen according to $\Delta t=[n\sqrt{(c/\Delta x)^2+(c/\Delta y)^2}]^{-1}$, which will ensure meeting the CFL requirement. The calculated reflectivity is shown in Table 1.

None of these cases produced the correct reflectivity R. The case (a) R value has only two decimal place precision due to the large mesh size. To verify if the errors were caused by the erroneous phases, one can modify these cases by replacing the thickness of thin film with $d'=d/(\tilde{v}_{axial}/c)$. This equation shows that the error is proportional to the thickness of the thin film (or, more precisely, the EM wave travel path length) and phase velocity error. Thus, for case (a) with modified film thickness of $d'=10.557\ \mu$m, the theoretical thin film wave optics [4] equation yields $R=0.6507$; for (b) $d'=10.079\ \mu$m, $R=0.0825$; and for (c) $d'=10.040\ \mu$m, $R=0.4715$. This clearly suggests the need to correct the phase error when it is significant.

Table 2 Effect of mesh size ($\Delta x=\Delta y$) on the solution accuracy

Mesh size	Angular resolution	Reflected energy/incident energy
λ/10	0.1	0.827
λ/20	0.1	0.940
λ/40	0.1	0.956
λ/80	0.01	0.973
λ/160	0.01	0.974

In the current problem, the phase error is not likely to be an issue when the coherence of reflected waves is very weak due to the random roughness. If the rms roughness is not very large in comparison with the incident wavelength and there is a certain degree of partial coherence, then the dispersion may become important. In general, if the coherence effect is relatively strong and the light beam path length is sufficiently large to cause spatial coherence, the dispersion error should be considered. For example, the wafer front side, which has periodic surface structure roughness and is nonabsorbing under the low temperature situation, is the case when phase error can be important. Although the current code does not correct the phase velocity induced dispersion error, it is determined that the effect on the random roughness surface calculation is minimal with reduced mesh size and is acceptable for the current problem. More discussion of mesh size effect on the solution accuracy is given below.

Mesh Size. The mesh size has the primary effect on the solution accuracy of the spatial central difference scheme. It also affects the magnitude of the dispersion error discussed above. Table 2 shows the computed result of the reflection from a flat surface using different mesh size. The conditions used for the calculation are given in Fig. 4. The incident light wavelength is 0.628 μm and the medium is a perfectly electric conductor (PEC). It is clear that the FDTD scheme slowly converges to exact value with smaller mesh size. The results indicate there is not much a small mesh can do to improve the solution accuracy beyond λ/80. To reduce the error without resorting to very small mesh size, the dispersion error mitigation methods should be used. To further demonstrate

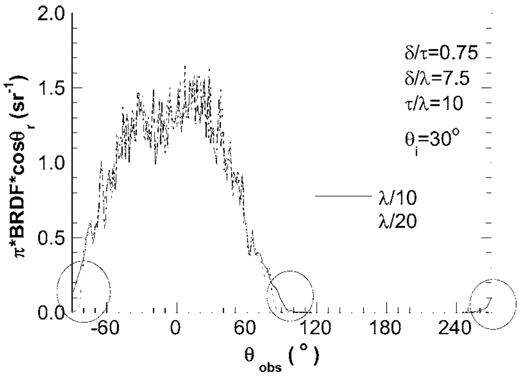


Fig. 4 Effect of mesh size on the rough surface FDTD solutions. The surface length is 50λ.

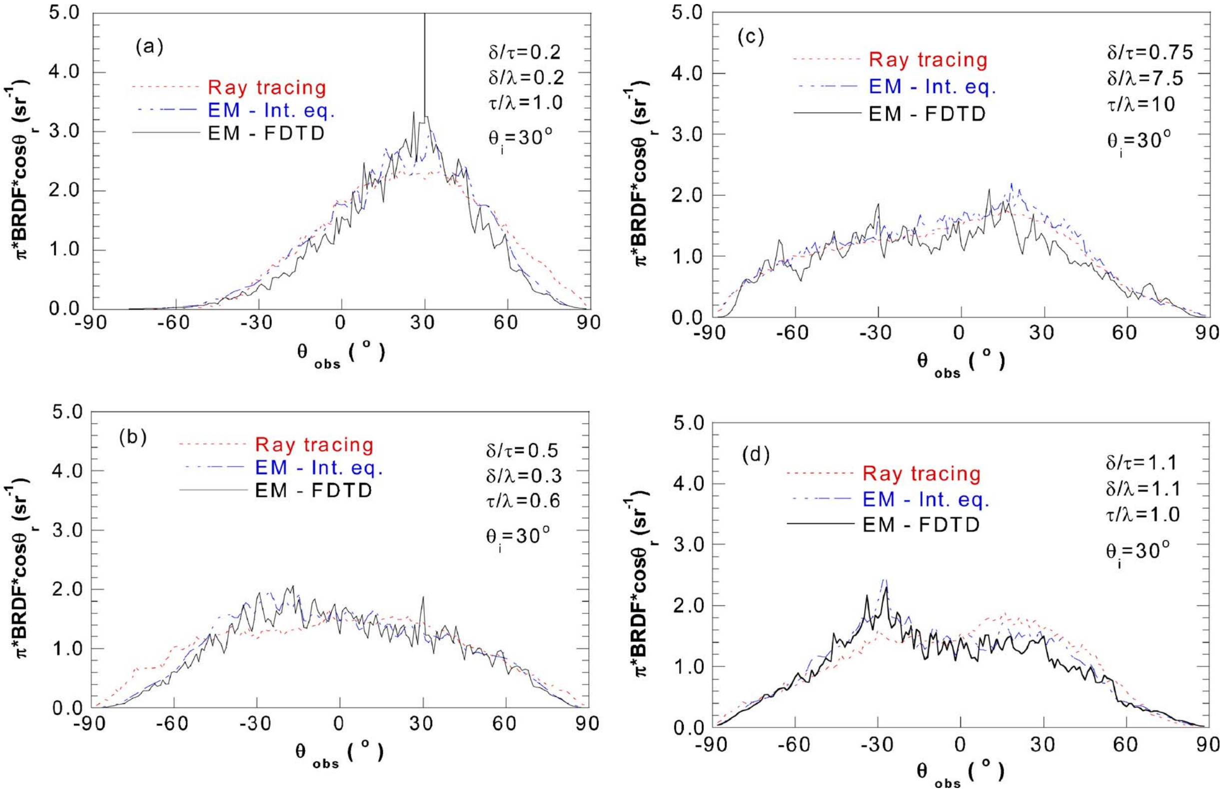


Fig. 5 Comparison of FDTD results with those obtained from ray tracing and integral equation methods [26] under different surface roughness

the mesh size's effect on the solution, a very rough surface was used. The results of two different mesh sizes ($\lambda/20$ and $\lambda/10$) were shown in Fig. 4. Three circled zones in the plot highlight the difference of the two solutions. Since reflection is expected only within $-90\ \text{deg} < \theta < 90\ \text{deg}$, the physically unrealistic nonzero reflection outside this range is visible in the $\lambda/10$ curve but disappears in the $\lambda/20$ mesh size curve. In this paper, $\lambda/20$ mesh size was considered sufficient and used for random roughness surface reflection calculation, unless specified otherwise. The mesh size is used as a compromise between the computational time and solution accuracy.

Resolution Used in the Surface Geometry. The resolution used in the generation of the random roughness surface profile (or geometry) should be considered. The BRDFs of surface profiles based on two resolutions ($\lambda/20$ and $\lambda/10$) were compared. The surfaces have very large roughness, which is considered to be a more stringent test to check the resolution. The surface roughness parameters are the same as those given in Fig. 4 but the result is not shown here. Although the overall BRDF patterns are similar in both cases, there are differences. The most significant one is the retro-reflection at $\theta = -30$ deg. The BRDF solution based on the surface generated with $\lambda/10$ resolution totally missed the retro-reflection peak. For the random roughness surface calculations in this study, the surfaces were generated with the $\lambda/20$ resolution.

Number of Time Steps. Generally speaking, the longer the physical path length traveled by the wave, the more time steps are needed to ensure there is sufficient time for all the scattered waves to interact and reach steady state results. Therefore, longer path length and larger surface roughness require a much larger number of time steps to reach a steady state BRDF value. In the rough surface computations, it is not possible to obtain a priori information about the required time step as the actual traversed path length is not known. In practice, several different time steps are used in FDTD computation on a rough surface and then the results are compared to determine the minimum required time steps. Once this information is known, the same time step number is applied in FDTD calculations on the remaining surfaces that are used in the result averaging. A time step adapting scheme can be easily added.

Surface Size. Ideally a very long surface length would have produced the most realistic result, but this would require a significant amount of memory and long computational time. It is thus not practical. It is also clear that a rough surface will require a longer surface length to produce equally accurate results in comparison with that of a smoother surface. An appropriate surface length needs to be determined.

For the problem of interest, a very long surface with random roughness was generated. The surface length is 3000λ. From this long surface, three different surface divisions were taken: 60 of 50λ length surfaces, 15 of 200λ length surfaces, and 3 of 1000λ length surfaces. The $60\times 50\lambda$ surfaces calculation result is given in Fig. 4, but the other two surfaces divisions are not shown since all three results are very close to one another. It is clear that all three different surface divisions produce similar BRDFs. However, the longer surface case ($3\times 1000\lambda$) used significantly much longer computing time to obtain the result. Therefore, the $60\times 50\lambda$ surfaces were used in the calculations. It should be noted that although the $60\times 50\lambda$ BRDF (Fig. 4) appears to have smaller oscillations, especially compared with the $3\times 1000\lambda$ BRDF, it is because more surfaces were used in the averaging process that produced smaller variance.

Even through the $60\times 50\lambda$ surface calculation is the most economical in terms of computation time and solution accuracy, it still took about 10 h run time. Since each surface calculation is independent from the other surfaces, the computation can be easily parallelized. In a large scale two-dimensional rough surface calculation, the large amount of surface data can be broken up into smaller pieces and parallel computing can be carried out on each

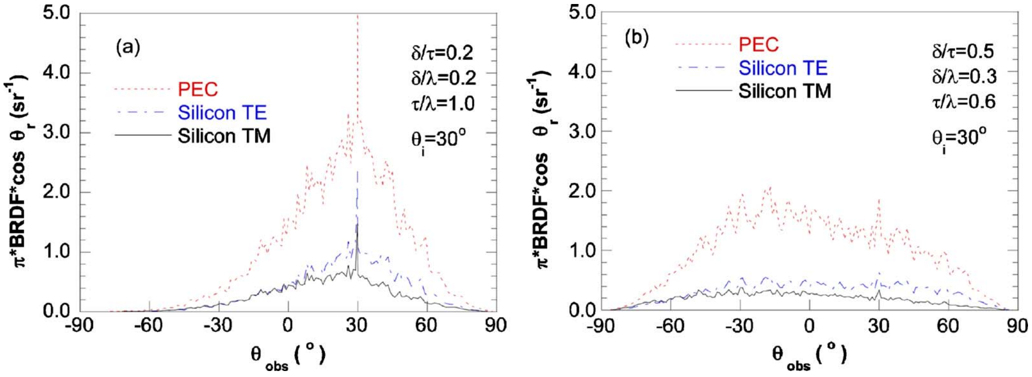


Fig. 6 Comparison of PEC and silicon surfaces

individual piece. It is expected that the parallelization efficiency will be very high due to the very small communication overhead among the independent surface calculations.

Scattering From Random Roughness Surfaces. Four PEC Gaussian surfaces with various degrees of roughness were considered. The BRDF of each surface was calculated and compared with the existing solution to validate the accuracy of the FDTD code. These results are all based on λ/20 mesh size and λ/20 surface generation resolution. There are existing solutions based on the geometric optic ray tracing and the integral equation solution of the EM wave equations. In general the difference between the two EM solutions is smaller than that between the ray tracing and either one of the two EM solutions.

Figure 5(*a*) shows that with the slightly rough surface the BRDF of the FDTD solution is somehow smaller than that of the integral equation solution. This is expected since Table 2 shows the overall reflected energy is 6% lower with the λ/20 mesh size. However, at the specular direction, $\theta_{\mathrm{obs}}=30$ deg, there is a sharp spike that is not seen in the other two solutions. It is not clear whether the spike is a numerical artifact. The clarification of this issue is beyond the scope of this paper. This will be treated in a follow-up study on metal surface reflection in which highly accurate experimental data are available for comparison. An earlier FDTD simulation did show a similar specular spike [27]. As the surface roughness increases, the magnitude of the spike reduces and eventually the spike disappears in the cases of δ/λ increases in Figs. 5(*c*) and 5(*d*). This behavior is physically reasonable and consistent with Tang [28]. Also, as the δ/τ increases, the retro-reflection becomes evident (see Fig. 5(*b*)). In Fig. 5(*c*), the agreement between the EM integral equation solution and ray tracing is better than that between the FDTD and ray tracing solutions. The retro-reflection peak, although small, is evident in the FDTD solution. Due to the relative shorter surface length versus the correlation length τ in this case, the FDTD solution exhibits larger oscillations than those in other cases. To ensure there is truly a retro-reflection peak at $\theta_{\mathrm{obs}}=-30$ deg and not simply a large oscillation occurring coincidently at that angle, twice the surface length as used in the FDTD calculation (results not shown). The results confirmed the retro-reflection peak. The integral equation solution has smaller oscillations in this case and a small peak at $\theta_{\mathrm{obs}}=-30$ deg, in comparison with FDTD results. At $\delta/\tau=1.1$ (Fig. 5(*d*)) the ray tracing starts to show a retro-reflection peak, although it is still smaller than those in the EM solutions.

Two different surface materials were examined on their impact to the BRDF: PEC and silicon, which has refractive index $n=3.426$. The overall trend of the BRDF is the same, especially the specular spike (Fig. 6). The silicon surface has two polarizations with TE mode reflectivity being higher than that of the TM mode. The silicons BRDF is about 1/3 of the PECs value due to the transmission into the material.

As the incident angle increases, the BRDF becomes more specular (Fig. 7). The large incident angle would effectively reduce the roughness. Thus, there is the appearance of a specular spike in the PEC surface of Fig. 7 as a result. A well-known phenomenon of the large incident angle is the off-specular peak described by Torrance and Sparrow [29]. An attempt was made to reproduce the off-specular peak with a dielectric surface, but it was not successful since the surface roughness description was incomplete in the original paper, in which only the rms roughness was given [29].

Conclusions

A finite difference time domain numerical scheme to solve the electromagnetic wave scattering from a micro-scale random roughness surface has been developed. The model applied the perfectly matching layer boundary condition and near-to-far field transformation to obtain the surface reflectivity property. Various numerical issues were considered and quantified to verify the method. These included dispersion error, mesh size, surface generation resolution, number of time steps needed to reach steady state result, and surface size. Bi-directional reflectivity of several random roughness surfaces was produced. Except for the strong specular spike in the small roughness surfaces not observed in the prior work, the finite difference time domain solutions agree well with the results obtained by ray tracing and integral equation solution of the Maxwell equations. The work will be continued in the near future to compare with existing experiment data and to consider the absorbing media, two-dimensional surface structures, as well as the revision of the code for parallel computation.

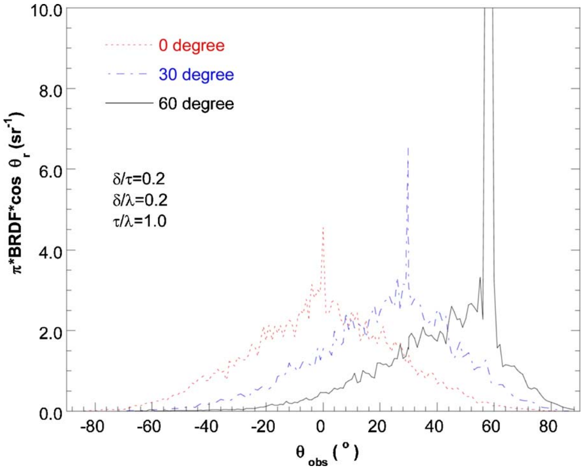


Fig. 7 BRDF at three different incident angles of a PEC surface

Acknowledgment

The work is supported by a grant from the National Science Foundation, Grant No. CTS-0331153. The ray tracing and integral equation solutions provided by Hyunjin Lee and Professor Zhuomin Zhang of Georgia Institute of Technology are much appreciated.

Nomenclature

c = light speed in vacuum, 3.0×10^8 m/s
d = period of a periodic structure μm
$\mathbf{E}$ = electric field N/C
$\mathbf{H}$ = magnetic field, C/m·s
i, j, k = grid index number along x, y, and z coordinates, respectively
i = imaginary unit, $\sqrt{-1}$
I = intensity, $\mathrm{W/m^2.sr.}\mu\mathrm{m}$
J_s = electrical field source, N/C·s
L = wafer thickness or depth of a PML medium, μm
m = time step
M_s = magnetic field source, $\mathrm{C/m\cdot s^2}$
n = refractive index
R = reflectivity from an unpolarized wave
R_e = reflectivity from a TE polarization wave
R_m = reflectivity from a TM polarization wave
$\mathbf{S}$ = Poynting vector, $\mathrm{W/m^2}$
t = time, s
T = distance, m
T^* = time period of a wave, s
v = velocity, m/s
x, y, z = physical coordinates, m

Greek Symbols

α = absorption coefficient, 1/m
δ = rms roughness, μm
ε = electrical permittivity, $\mathrm{C^2/N\cdot m^2}$
ζ = surface height, μm
η = grid spacing, μm
θ = angle, rad
λ = wavelength, μm
μ = magnetic permeability, $\mathrm{N.s^2/C^2}$
σ = electrical conductivity, $\mathrm{C^2/N\cdot m^2\cdot s=1/\Omega\cdot m}$
σ^* = magnetic conductivity, $\mathrm{N\cdot s/C^2}$
τ = correlation length, μm
ω = angular frequency, rad/s
Ω = solid angle, sr

References

[1] Tsang, L., Kong, J. A., and Ding, K.-H., 2000, *Scattering of Electromagnetic Waves—Vol. 1 Theories and Applications*, Wiley, New York.
[2] Drolen, B. L., 1992, "Bidirectional Reflectance and Secularity of Twelve Spacecraft Thermal Control Materials," J. Therm. Anal., **6**(4), pp. 672–679.
[3] Hebb, J. P., Jensen, K. F., and Thomas, J., 1998, "The Effect of Surface Roughness on the Radiative Properties of Patterned Silicon Wafers," IEEE Trans. Semicond. Manuf., **11**(4), pp. 607–614.
[4] Fu, K., Hsu, P.-f., and Zhang, Z. M., 2006, "Unified Analytical Formulations of Thin-Film Radiative Properties Including Partial Coherence," Appl. Opt., **45**(4), pp. 653–661.
[5] Knotts, M. E., and O'Donell, K. A., 1994, "Measurements of Light Scattering by a Series of Conducting Surfaces With One-Dimensional Roughness," J. Opt. Soc. Am. A, **11**(2), pp. 697–710.
[6] Siegel, R., and Howell, J. R., 2002, *Thermal Radiation Heat Transfer*, 4th ed., Taylor & Francis, New York.
[7] Tang, K., and Buckius, R. O., 1998, "The Geometric Optics Approximation for Reflection From Two-Dimensional Random Rough Surfaces," Int. J. Heat Mass Transfer, **41**, pp. 2037–2047.
[8] Zhou, Y. H., and Zhang, Z. M., 2003, "Radiative Properties of Semitransparent Silicon Wafers With Rough Surfaces," ASME J. Heat Transfer, **125**, pp. 462–470.
[9] Tang, K., Kawka, P. A., and Buckius, R. O., 1999, "Geometric Optics Applied to Rough Surfaces Coated With an Absorbing Thin Film," J. Thermophys. Heat Transfer, **13**, pp. 169–176.
[10] Zhu, Q. Z., Lee, H. J., and Zhang, Z. M., 2005, "Validity of Hybrid Models for the Bidirectional Reflectance of Coated Rough Surfaces," J. Thermophys. Heat Transfer, **19**(4), pp. 548–557.
[11] Taflove, A., and Hagness, S. C., 2005, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed., Artech House, Boston, MA.
[12] Fung, A. K., Shah, M. R., and Tjuatja, S., 1994, "Numerical Simulation of Scattering From Three-Dimensional Randomly Rough Surfaces," IEEE Trans. Geosci. Remote Sens., **Ge-32**(5), pp. 986–994.
[13] Wong, P. B., Tyler, G. L., Baron, J. E., Gurrola, E. M., and Simpson, R. A., 1996, "A Three-Wave FDTD Approach to Surface Scattering With Applications to Remote Sensing of Geophysical Surfaces," IEEE Trans. Antennas Propag., **44**(4), pp. 504–514.
[14] Tannehill, J. C., Anderson, D. A., and Pletcher, R. H., 1997, *Computational Fluid Mechanics and Heat Transfer*, 2nd ed., Taylor & Francis, Washington, DC.
[15] Wang, Z. J., Przekwas, A. J., and Liu, Y., 2002, "A FV-TD Electromagnetic Solver Using Adaptive Cartesian Grids," Comput. Phys. Commun., **148**, pp. 17–29.
[16] Berenger, J. P., 1994, "A Perfectly Matched Layer for the Absorption of Electromagnetic Waves," J. Comput. Phys., **114**, pp. 185–200.
[17] Lee, H. J., Zhu, Q. Z., Chen, Y. B., and Zhang, Z. M., 2005, "Radiative Properties of Anisotropic Microrough Silicon Surfaces," *Proceedings of the 38th AIAA Thermophysics Conference*, Toronto, Canada, June 6–9.
[18] Zhao, Y., Wang, G.-C., and Lu, T.-M., 2004, *Characterization of Amorphous and Crystalline Rough Surface: Principles and Applications*, Academic, San Diego, CA.
[19] Beckmann, P., and Spizzichino, A., 1987, *The Scattering of Electromagnetic Waves From Rough Surfaces*, Artech House, Norwood, MA, Chap. 5.
[20] Yee, K. S., 1966, "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media," IEEE Trans. Antennas Propag., **14**(3), pp. 302–307.
[21] Liu, J., Zhang, S. J., and Chen, Y. S., 2003, "Prediction of Radiative Properties of Patterned Silicon Wafers by Solving Maxwell's Equations in the Time Domain," Numer. Heat Transfer, Part B, **44**, pp. 329–345.
[22] Patankar, S. V., 1980, *Numerical Heat Transfer and Fluid Flow* (Hemisphere Series on Computational Methods in Mechanics and Thermal Science), Hemisphere Publishing Corp., Washington, DC.
[23] Liu, J., Zhang, S. J., and Chen, Y. S., 2004, "Rigorous Electromagnetic Modeling of Radiative Interaction with Microstructures Using the FVTD Algorithm," Int. J. Thermophys., **25**, pp. 1281–1297.
[24] Kunz, K. S., and Luebbers, R. J., 1993, *The Finite Difference Time Domain Method for Electromagnetics*, CRC, Boca Raton, FL.
[25] Sadiku, M. N. O., 2000, *Numerical Techniques in Electromagnetics*, 2nd ed., CRC, Boca Raton, FL.
[26] Lee, H. J., and Zhang, Z. M., 2005, private communications.
[27] Hastings, F. D., Schneider, J. B., and Broschat, S. L., 1995, "A Monte-Carlo FDTD Technique for Rough Surface Scattering," IEEE Trans. Antennas Propag., **43**(11), pp. 1183–1191.
[28] Tang, K., 1998, "Theory of Experiments of Scattering From Microscale Random Rough and Deterministic Surfaces," Ph.D. thesis, Univ. of Illinois, Urbana-Champaign, IL.
[29] Torrance, K. E., and Sparrow, E. M., 1967, "Theory for Off-Specular Reflection From Roughened Surfaces," J. Opt. Soc. Am., **57**(9), pp. 1105–1114.

Radiative Properties of Patterned Wafers With Nanoscale Linewidth

Y.-B. Chen

Z. M. Zhang[1]
Fellow ASME
e-mail: zzhang@me.gatech.edu

George W. Woodruff School of Mechanical Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332

P. J. Timans
Mattson Technology,
Fremont, CA 94538

Temperature nonuniformity is a critical problem in rapid thermal processing (RTP) of wafers because it leads to uneven diffusion of implanted dopants and introduces thermal stress. One cause of the problem is nonuniform absorption of thermal radiation, especially in patterned wafers, where the optical properties vary across the wafer surface. Recent developments in RTP have led to the use of millisecond-duration heating cycle, which is too short for thermal diffusion to even out the temperature distribution. The feature size is already below 100 nm *and is smaller than the wavelength* (200–1000 nm) *of the flash-lamp radiation. Little is known to the spectral distribution of the absorbed energy for different patterning structures. This paper presents a parametric study of the radiative properties of patterned wafers with the smallest feature dimension down to* 30 nm*, considering the effects of temperature, wavelength, polarization, and angle of incidence. The rigorous coupled wave analysis is employed to obtain numerical solutions of the Maxwell equations and to assess the applicability of the method of homogenization based on effective medium formulations.* [DOI: 10.1115/1.2401201]

Keywords: electromagnetic, microstructure, nanoscale, properties, radiation

1 Introduction

According to the International Technology Roadmap for Semiconductors [1], the gate length and junction depth will be 25 and 13.8 nm, respectively, for the 65-nm devices used in high-performance complementary metal oxide semiconductor (CMOS) technology. RTP currently provides the high-temperature annealing needed to create ultrashallow junctions. However, conventional RTP tools that use tungsten-halogen lamps for heating encounter a dilemma between limiting the junction depth and maximizing the degree of electrical activation of implanted dopants [2]. The ion implantation annealing time from 1 to 10 s above 1000 °C in conventional RTP is too long to confine ion diffusion and achieve the implanted doping distribution within the junction [3]. This difficulty can be overcome by using high-intensity flash lamps with millisecond optical pulses to raise the surface temperature of the wafer to approximately 1300 °C. The bulk wafer is rapidly preheated to around 700 °C to reduce the pulse energy requirement and thermal stress caused by millisecond pulse heating. The typical energy sources are Ar or Xe arc lamps, which mainly emit ultraviolet and visible radiation. In the flash-lamp heating, optical energy is absorbed at the wafer surface because of the small penetration depth resulting from the large absorption coefficient of Si within the lamp spectrum. It has been demonstrated that flash-lamp annealing can provide sufficiently low sheet resistivity and low junction leakage with negligible dopant diffusion to meet the requirements of advanced CMOS [2,4,5].

Because the energy is absorbed within milliseconds, thermal diffusion cannot distribute heat uniformly across the wafer surface. Therefore, temperature uniformity across the wafer is expected to be a critical issue. Temperature nonuniformity may cause uneven activation of the implants, as well as excessive thermal stresses that can introduce crystallographic defects [6]. Thermal stresses at the interfaces of Si, SiO_2, and poly-Si can be negligible because of their comparable thermal expansion coefficients (2.6, 2.3, and 2.8×10^{-6} K^{-1}, respectively) by selected formation process [7]. On the other hand, thermal stresses among differently patterned regions on the wafer surface can be significant if the temperature gradient is large [8]. A major reason for the temperature nonuniformity arises from the difference in the absorptance of various device patterns in different regions of the wafer surface.

A number of researchers have modeled the radiative properties of different patterned structures on wafers and obtained reasonable agreement with experimental results. Erofeev et al. [9] modeled the radiation interaction with 2D patterned wafers by solving the Maxwell equations using a finite element method coupled with the boundary integral equation. Hebb et al. [10] measured the reflectance of a memory die, logic die, and the backside of various multilayered wafers to assess the effectiveness of thin-film optics in providing approximations for the properties of patterned wafers. Tada et al. [11] evaluated the effects of thin SiO_2 film patterns on a Si wafer through the combination of coherent and incoherent approaches and compared the calculated results with experimental measurements. Liu et al. [12] predicted the radiative properties of patterned Si wafers through models based on the finite-difference time-domain (FDTD) and the finite-volume time-domain (FVTD) methods. The patterned structures were Si gratings, SiO_2 film on top of flat Si wafers, and SiO_2 films on top of Si gratings. However, earlier investigations are for relatively simple patterns with features on the order of micrometers and at wavelengths longer than 0.4 μm.

Radiative properties of nanostructures may be very different from those of microstructures made of similar materials [13]. In addition, patterned wafers can be highly absorbing in the ultraviolet region that is important to flash-lamp annealing. Little is known about the influence of nanoscale patterns on the radiation absorption and reflection during flash-lamp annealing. The objective of the present research is to model the radiative properties of periodically patterned wafers that include features expected in advanced CMOS device technologies. The effects of wafer temperature, wavelength, polarization, and angle of incidence on the spectral directional-hemispherical reflectance and directional absorptance are investigated for selected 2D patterned structures with multilayer 1D gratings. In the present study, the rigorous couple wave analysis (RCWA) is employed to obtain numerical solutions of the Maxwell equations [14–17]. For periodic structures, RCWA can produce accurate solutions much faster than

---

[1] Corresponding author.

Contributed by the Heat Transfer Division of ASME for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received January 31, 2006; final manuscript received June 8, 2006. Review conducted by Suresh V. Garimella. Paper presented at the 2005 ASME International Mechanical Engineering Congress (IMECE2005), November 5–11, 2005, Orlando, FL.

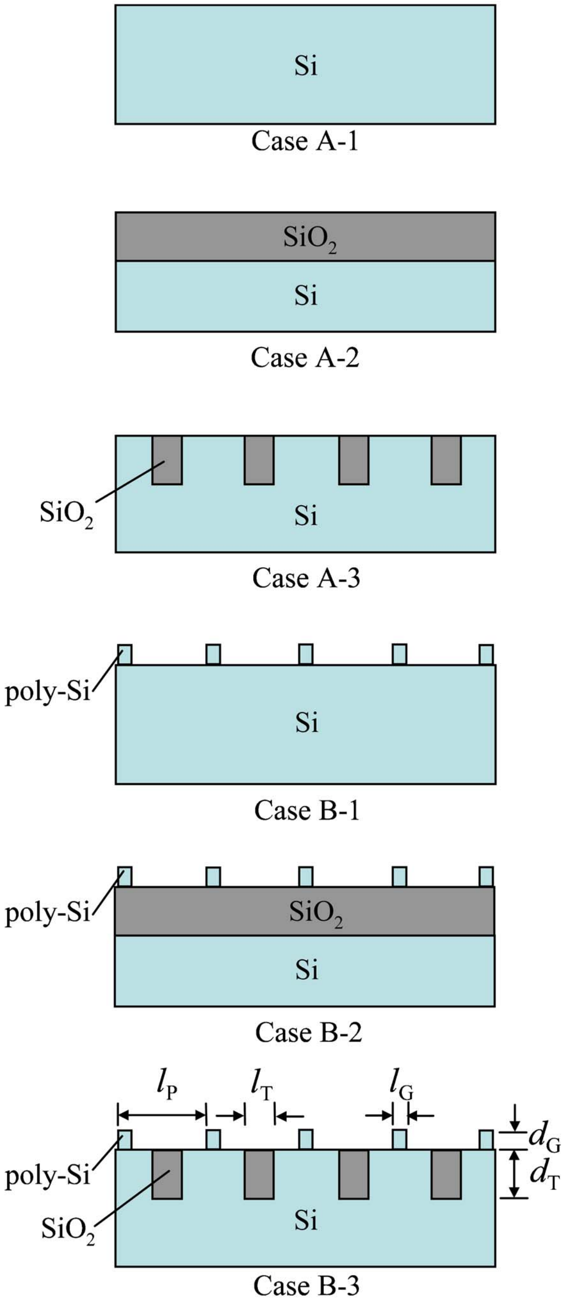


Fig. 1 Parameters of selected cases, where d_G and d_T are the depths of gate and trench, respectively, and l_G, l_P, and l_T are the lengths of gate, pitch, and trench

Table 1 Feature dimensions of selected cases (unit: nm)

Case No.	A-1	A-2	A-3	B-1	B-2	B-3
d_G	—	—	—	50	50	50
d_T	—	350	350	—	350	350
l_P	240	240	240	240	240	240
l_G	0	0	0	30	30	30
l_T	0	240	60	0	240	60

FDTD and FVTD. Some previous studies have used the method of homogenization, in which the grating region is treated as a homogeneous layer with an effective dielectric function. This way, the 2D patterned structures are simplified to planar multilayer structures, which requires much less computation time to solve the matrix equations based on thin-film optics [18,19]. To assess the applicability of the method of homogenization, different effective medium formulations are evaluated by comparison with the simulation results of RCWA.

2 Model Development and Numerical Methods

In order to understand the effect of nanostructure on the spectral absorptance, several model structures are considered, as illustrated in Fig. 1. Case A-1 is a bare Si wafer and case A-2 is a SiO_2 thin film on Si wafer. In case A-3, an array of SiO_2-filled trenches (350 nm deep) is formed in the Si substrate. Cases B-1, B-2, and B-3 involve polycrystalline silicon (poly-Si) gates (30 nm wide and 50 nm high) on top of the structure that forms a thin grating layer. Case B-3 illustrates a generic pattern of interest for CMOS device technology, in which periodic patterns are formed on a Si wafer that includes trenches filled with SiO_2. These trenches provide electrical insulation between active areas of the device where transistors are formed. In the middle of each active area there is a strip of poly-Si, representing the gate electrode. In reality, this gate is isolated from the Si substrate by a very thin dielectric film. However, this film is only about 1-nm thick for the 65-nm-node devices, and hence, it can be neglected in the optical models. The features are symmetrically located in a lateral period of $l_P = 240$ nm. Table 1 lists the dimensions of the trench and gate for structures of all cases. Various combinations of these structures can be expected in different parts of the wafer. It is hoped that the simplified structures will also reveal how individual periodic patterning of the isolation oxide and the poly-Si gate film influence the optical properties. While only 1D and 2D structures are studied, the results should be instructive to the more general 3D patterned structures. Similar approximations were also used in previous numerical simulations [9,12]. In the remaining of this section, the spectral distribution of the lamp and optical properties of materials are first discussed, followed by brief explanations of the RCWA and effective medium formulations used in the simulation.

2.1 Lamp Spectrum and Optical Properties of Materials. Arc lamps filled with Xe or Ar gases emit light by generating an electric arc across a gap between two electrodes. The peak wavelength corresponds to a blackbody temperature around 6000 K. However, the emission is confined to the spectral region for $200\ \text{nm} \leq \lambda \leq 1000\ \text{nm}$, where λ is the wavelength in vacuum. The actual emission spectrum is not a smooth curve due to the discrete plasma emission lines. The emission spectrum also varies with the discharge voltage, duration of the pulse, and the optical materials. Therefore, the reported spectral distributions by different manufacturers and researchers have a large variation. To study the total absorptance, the present work assumes a simple trapezoid spectral distribution of the lamp emission. The normalized spectrum distribution $G(\lambda)$ is 1 for $400\ \text{nm} \leq \lambda \leq 600\ \text{nm}$ and linearly reduces to 0 towards $\lambda = 200$ and 1000 nm.

The input material properties are the optical constants (i.e., refractive index n and extinction coefficient κ) of Si, SiO_2, and poly-Si in the wavelength range of the flash-lamp spectrum. The optical constants are related to the dielectric function $\epsilon = (n + i\kappa)^2$. In the present study, the optical constants of Si at room temperature are taken from the tabulated values in Ref. [20]. Jellison and Modine [21] measured the dielectric function of silicon from 260 to 840 nm from room temperature to 700°C. In another paper [22], these authors developed expressions of n and κ in terms of temperature and wavelength, applicable to the visible and near-infrared region. Optical constants at higher temperatures were measured by Sun et al. [23]. In the present work, the optical constants from 280 to 500 nm at 700 and 910°C are taken from Refs. [21,23], respectively. Because the expressions in Ref. [22] give good agreement with the measured data in Refs. [21,23] at wavelengths from 500 to1000 nm, the optical constants are calculated from the expressions in this spectral range. In the spectral

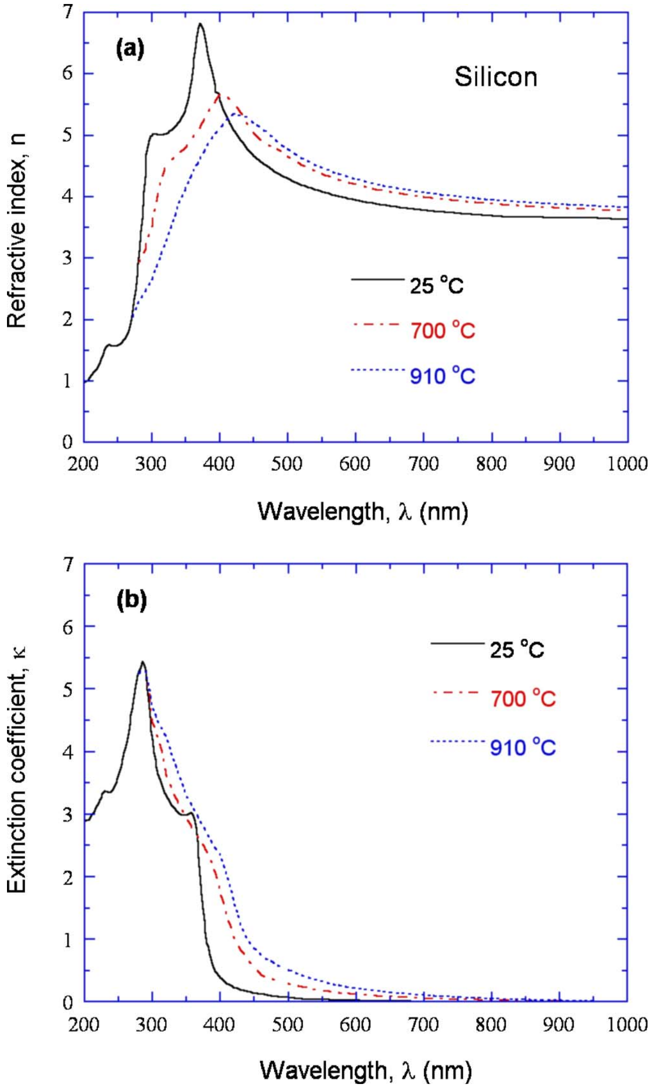


Fig. 2 Optical constants of silicon at 25, 700, and 910°C: (*a*) refractive index, and (*b*) extinction coefficient

region from 200 to about 280 nm, room temperature data are substituted. While the optical constants may be very different at high temperatures, the effect on the total absorbed energy should not be significant because it is near the source cutoff wavelength. The selected dielectric functions of Si at 25, 700, and 910°C are plotted in Fig. 2 for comparison. Some interpolations between the data are made to produce smooth curves in the calculation.

For SiO_2, the imaginary part of the dielectric function is negligibly small in the spectral range of interest and absorption can be neglected. Malitson [24] provided an equation as a fit for n of SiO_2 at 25°C. Because the relative error of four to seven significant digits in the equation is less than 0.01%, only four significant digits are selected as expressed in the following:

$$n_{\mathrm{SiO}_2} = \left(1 + \frac{0.6962\lambda^2}{\lambda^2 - 0.06840^2} + \frac{0.4079\lambda^2}{\lambda^2 - 0.1162^2} + \frac{0.8975\lambda^2}{\lambda^2 - 9.896^2}\right)^{1/2} \tag{1}$$

This equation is also believed to represent the values at higher temperatures due to the small temperature coefficient ($1.2 \times 10^{-5}\ \mathrm{K}^{-1}$) of the refractive index of SiO_2 [25].

The dielectric function of poly-Si is assumed to be the same as that for single-crystal silicon, although in reality differences are expected to arise from the presence of grain boundaries in the material and from variations in grain structure [4]. According to Ref. [25], the difference between the optical constants of poly-Si and Si is not so large in the red to infrared region. Note that doping effects of poly-Si is not taken into consideration here, and electrically active doping may strongly vary the optical properties in the infrared region. On the other hand, in the ultraviolet to visible wavelength region, the structural disorder in the poly-Si could change optical properties somewhat regardless of doping. Because poly-Si occupies a small volume fraction only, the use of the optical constants of silicon should not cause significant errors.

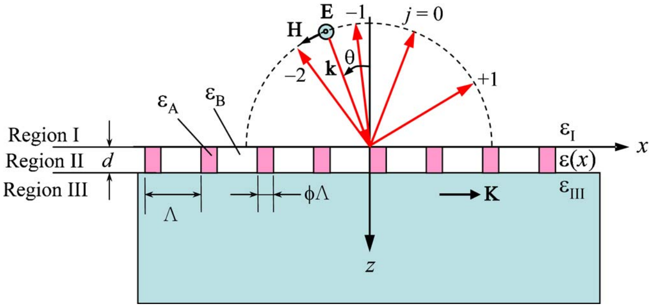


Fig. 3 Schematic drawing for the TE wave incidence on a grating layer, showing the reflected diffraction orders *j*=−2, −1, 0, and 1

2.2 RCWA. RCWA is one of the efficient tools for modeling radiative properties and analyzing diffraction efficiency of periodic gratings bounded by two semi-infinite media [14,15]. The basic principles of RCWA are summarized below using a single layer of rectangular grating shown in Fig. 3. A plane wave is incident on a one-dimensional grating surface from free space by neglecting the effect of gases in the RTP chamber. Region I is free space with $\epsilon_{\mathrm{I}} = n_{\mathrm{I}} = 1$ and $\kappa_{\mathrm{I}} = 0$. Region II is composed of materials A and B so that its dielectric function is a periodic function of x with a period Λ, which is the grating period. The filling ratio of material A is ϕ, and the lateral extension of the gratings is assumed to be infinite. Region III is the substrate with a dielectric function, ϵ_{III}. For cases B-2 and B-3 shown in Fig. 1, two grating regions can be used. Gratings of other shapes can be divided into planar thin slabs to approximate the grating profile [14].

The wave vector **k** defines the direction of incidence, and the angle between **k** and the surface normal $\hat{\mathbf{z}}$ is the angle of incidence θ, also called the polar angle. The grating vector **K** for the structure shown in Fig. 3 is defined in the positive x direction with a magnitude $K = 2\pi/\Lambda$. For simplicity, it is assumed that incidence wave vector is on the x-z plane, that is, the y component of **k** is zero. For a TE wave, the electric field **E** is perpendicular to the plane of incidence, i.e., parallel to the y direction and perpendicular to the grating vector **K**, as shown in Fig. 3. On the other hand, for a TM wave, the magnetic field **H** is perpendicular to the plane of incidence and the vector **K**. The following discussion is for the TE wave. The magnitude of the incident electric field, after normalization, can be expressed as $\exp(ik_x x + ik_z z - i\omega t)$, where k_x and k_z are, respectively, the x and z components of **k**, ω is the angular frequency, and t is time. For simplicity, the time harmonic term $\exp(-i\omega t)$ will be omitted hereafter. The magnitude of **k** in regions I and III can be expressed as

$$k_{\mathrm{I}} = \frac{2\pi n_{\mathrm{I}}}{\lambda} = \frac{2\pi}{\lambda} = k \quad \text{and} \quad k_{\mathrm{III}} = \frac{2\pi n_{\mathrm{III}}}{\lambda} = n_{\mathrm{III}} k \tag{2}$$

where n_{III} is the refractive index in region III. There exists a phase difference of $2\pi\Lambda \sin\theta/\lambda = k_x\Lambda$ between the incident wave at (x,z) and that at $(x+\Lambda, z)$ due to a path difference of $\Lambda \sin\theta$. This condition must also be satisfied by each diffracted wave, i.e., the magnitude of the jth order reflected wave can be written as

$R_j \exp(ik_{xj}x - ik_{\mathrm{I},zj}z)$, where R_j is the electric field reflection coefficient and k_{xj} is determined by the Bloch–Floquet condition

$$k_{xj} = \frac{2\pi}{\lambda}\sin\theta + \frac{2\pi}{\Lambda}j = k_x + Kj \tag{3a}$$

The above equation can be expressed in terms of the angle of reflection

$$\sin\theta_j = \sin\theta + \frac{j\lambda}{\Lambda} \tag{3b}$$

where $\theta_j = \sin^{-1}(k_{xj}/k)$ is the jth order diffraction angle for reflection and Eq. (3b) is the well-known grating equation. When $k_{xj} > k_{\mathrm{I}}$, $\sin\theta_j > 1$ and the jth order reflected wave decays exponentially towards the negative z direction. This is an evanescent wave that only exists near the surface within a distance on the order of wavelength. Note that the z component of **k** for the jth order reflected wave is

$$k_{\mathrm{I},zj} = \begin{cases} (k_{\mathrm{I}}^2 - k_{xj}^2)^{1/2}, & k_{\mathrm{I}} > k_{xj} \\ i(k_{xj}^2 - k_{\mathrm{I}}^2)^{1/2}, & k_{xj} > k_{\mathrm{I}} \end{cases} \tag{4}$$

Because k_{xj} must be the same in all media, similar criteria can be applied to the transmitted waves in region III to obtain $k_{\mathrm{III},zj}$ based on Eq. (2).

The electric field in region I is a superposition of the incident wave and the reflected waves; therefore

$$E_{\mathrm{I}}(x,z) = \exp(ik_x x + ik_z z) + \sum_j R_j \exp(ik_{xj}x - ik_{\mathrm{I},zj}z) \tag{5}$$

The electric field in region III can be obtained by superimposing all transmitted waves

$$E_{\mathrm{III}}(x,z) = \sum_j T_j \exp[ik_{xj}x + ik_{\mathrm{III},zj}(z-d)] \tag{6}$$

where T_j is the transmission coefficient for the jth order transmitted wave.

The electric field in region II can be expressed as

$$E_{\mathrm{II}}(x,z) = \sum_j \psi_j(z)\exp(ik_{xj}x) \tag{7}$$

where $\psi_j(z)$ is the amplitude of the jth space-harmonic component. Here, the order of j is matched with the diffraction order in regions I and III. Due to the periodic structure, the dielectric function of region II can be expanded in a Fourier series:

$$\epsilon(x) = \sum_m \epsilon_m \exp\left(i\frac{2m\pi}{\Lambda}x\right), \quad m = 0, \pm 1, \pm 2, \ldots \tag{8}$$

where ϵ_m is the mth coefficient that is given by

$$\epsilon_0 = \phi\epsilon_A + (1-\phi)\epsilon_B \quad \text{and} \quad \epsilon_m = \frac{(\epsilon_A - \epsilon_B)\sin(m\phi\pi)}{m\pi} \quad (m \neq 0) \tag{9}$$

for rectangular gratings shown in Fig. 3 with ϕ being the filling ratio of material A. It should be noted that each ϵ_m is not a physical property of the material and its imaginary part may be negative for a passive medium.

The coupled-wave formulation comes from the wave equation of the total electric field in region II, and can be expressed in the following for a TE wave incidence

$$\nabla^2 E_{\mathrm{II}}(x,z) + k^2\epsilon(x)E_{\mathrm{II}}(x,z) = 0 \tag{10}$$

A differential equation can be obtained by substitute Eqs. (7) and (8) into Eq. (10)

$$\sum_j \frac{d^2\psi_j}{dz^2}\exp(ik_{xj}x) - \sum_j k_{xj}^2\psi_j\exp(ik_{xj}x) + k^2\left[\sum_m \epsilon_m \exp\left(i\frac{2m\pi}{\Lambda}x\right)\right]\left[\sum_p \psi_p \exp(ik_{xp}x)\right] = 0 \tag{11}$$

Equation (11) can be rearranged in terms of the jth order in $\exp(ik_{xj}x)$ as follows:

$$\sum_j \left(\frac{d^2\psi_j}{dz^2} - k_{xj}^2\psi_j + k^2\sum_p \epsilon_{j-p}\psi_p\right)\exp(ik_{xj}x) = 0 \tag{12}$$

In order to satisfy this equation for any value of x, the coefficient of $\exp(ik_{xj}x)$ must be zero for all j. Hence, Eq. (12) is an infinite set of second-order coupled equations. Note that each space-harmonic term is coupled to other components through the harmonics of the grating. The numerical solution is obtained with sufficiently large number of diffraction orders. Suppose $j = 0, \pm 1, \pm 2, \ldots, \pm q$, there are $N = 2q+1$ diffraction orders to be used in Eq. (12) so that $p = 0, \pm 1, \pm 2, \ldots, \pm q$ will also have N terms and Eq. (12) can be represented by a $N \times N$ matrix. The Fourier expansion of the dielectric function will have $m = 0, \pm 1, \pm 2, \ldots, \pm 2q$ or $4q+1$ terms. The magnetic field can be obtained from Eq. (7) and expressed in terms of Ψ_j. The N unknown functions $\psi_j (j = 0, \pm 1, \pm 2, \ldots, \pm q)$ can be expressed assummations with eigenfunctions and $2N$ coefficients to be determined, which come from the forward and backward diffracted waves in region II due to coupling [15]. The $2N$ coefficients along with R_j and T_j $(j = 0, \pm 1, \pm 2, \ldots, \pm q)$ compose $4N$ unknowns. Boundary conditions require that the tangential components of the electric and magnetic fields be continuous at the boundaries between different regions. The field components are consistent for all x once the boundary conditions are satisfied at $x=0$ due to the Block–Floquet condition. Subsequently, $4N$ linear equations can be generated from Eqs. (5) to (7) at $x=0$ for $z=0$ and $z=d$. An enhanced, numerically stable transmittance matrix approach was developed and applied to the implementation of RCWA for surface-relief and multilevel gratings [14]. Detailed equations and solution procedure for both 2D and 3D problems can be found in Ref. [15]. The derivation of the TM wave is similar by using the corrected procedure proposed by Li [16].

Once the reflection and transmission coefficients are obtained, the reflectance and transmittance of each diffraction order can be computed from the time-average Poynting vector in the z direction. The directional-hemispherical reflectance ρ_λ is the summation of the reflectance of all orders. Furthermore, the absorptance is calculated by $\alpha_\lambda = 1 - \rho_\lambda$ since the silicon wafer is opaque in the spectral region of interest. The total absorptance can be calculated using the lamp spectral distribution $G(\lambda)$ by the following equation:

$$\alpha_{\mathrm{tot}} = \frac{\int \alpha_\lambda(\lambda)G(\lambda)d\lambda}{\int G(\lambda)d\lambda} \tag{13}$$

2.3 Effective Medium Formulations. When the grating period is much smaller than the wavelength, $\Lambda/\lambda < 1/(n_{\mathrm{III}} + \sin\theta)$, all the diffracted waves are evanescent wave except the zeroth order (specular direction). The reflection is similar to a smooth film with an effective uniform dielectric function. This approach is called the method of homogenization [18,19] and the underlying physics is based on the effective medium theory [26]. Effective medium formulations have been used widely to describe the optical properties of inhomogeneous media. The effective medium theory (EMT) was first postulated by Maxwell Garnett [27] to obtain the effective dielectric function of metal particles embedded in a dielectric medium. The general assumption is that the spacing separating the particles to be sufficiently large or the filling ratio of the particles to be small [28]. Bruggeman [29] assumed that two materials are both embedded in the effective medium and obtained an expression, which has been successfully applied to study the effect of porosity on the refractive index and absorption coefficient of different materials [13]. Bruggeman's expression is often called the effective medium approximation (EMA). The dielectric function of the effective medium ϵ_{EMA} is related to that of the two components by

$$\phi\frac{\epsilon_{\mathrm{EMA}}-\epsilon_A}{\epsilon_A+2\epsilon_{\mathrm{EMA}}}+(1-\phi)\frac{\epsilon_{\mathrm{EMA}}-\epsilon_B}{\epsilon_B+2\epsilon_{\mathrm{EMA}}}=0 \tag{14}$$

where ϕ is the volume fraction (filling ratio) of material A.

Rytov [30] first applied EMT for a periodic structure by treating a stratified medium as a homogeneous uniaxial crystal and obtained the effective permittivity and permeability tensors. The zeroth order is considered to be applicable when $\Lambda \ll \lambda$ and has been used for the design of surfaces with antireflection and selective radiative properties [19,31]. The expression has been extended to include higher-order terms for both one- and two-dimensional gratings [14,15]. The effective medium formulation for gratings depends on the polarization. The zeroth-order expressions are given below [18,19,31]

$$\epsilon_{\mathrm{TE},0}=\phi\epsilon_A+(1-\phi)\epsilon_B \tag{15a}$$

$$\text{and}\quad \epsilon_{\mathrm{TM},0}=\left(\frac{\phi}{\epsilon_A}+\frac{1-\phi}{\epsilon_B}\right)^{-1} \tag{15b}$$

where subscripts TE and TM indicate the polarization of the incidence light. The expressions including the second-order terms are [18,26]

$$\epsilon_{\mathrm{TE},2}=\epsilon_{\mathrm{TE},0}\left[1+\frac{\pi^2}{3}\left(\frac{\Lambda}{\lambda}\right)^2\phi^2(1-\phi)^2\frac{(\epsilon_A-\epsilon_B)^2}{\epsilon_{\mathrm{TE},0}}\right] \tag{16a}$$

$$\text{and}\quad \epsilon_{\mathrm{TM},2}=\epsilon_{\mathrm{TM},0}\left[1+\frac{\pi^2}{3}\left(\frac{\Lambda}{\lambda}\right)^2\phi^2(1-\phi)^2 \times(\epsilon_A-\epsilon_B)^2\epsilon_{\mathrm{TE},0}\left(\frac{\epsilon_{\mathrm{TM},0}}{\epsilon_A\epsilon_B}\right)^2\right] \tag{16b}$$

Notice that when $\Lambda \ll \lambda$ or in the case of small filling ratio, the second-order term drops out. It is hoped that the second-order correction may improve the useful range of the EMT. After the effective dielectric function is obtained, the grating region is treated as a homogeneous isotropic material for the given polarization and incidence. The transfer matrix method [13,32] for calculating radiative properties of multilayer thin films is then applied to obtain the reflectance. It should be noted that the condition $\Lambda/\lambda < 1/(n_{\mathrm{III}}+\sin\theta)$ is generally not satisfied for the gratings and spectral range studied in the present work. In addition, κ_{III} becomes larger than n_{III} for Si in the ultraviolet region. Because of the fast calculation speed using the method of homogenization, different effective medium formulations are examined to see whether there is an alternative way to evaluate the radiative properties with greatly reduced computational resources. To distinguish different formulations, in the present paper, EMA refers to Eq. (14), EMT-0 refers to Eq. (15) with zeroth order only, and EMT-2 refers to Eq. (16) with the second-order terms.

2.4 Validation of Computational Algorithm. The solution of RCWA converges without inherent numerical instabilities and it always satisfies the energy conservation. The solution is accurate to an arbitrary level of accuracy and depends only on the number of spatial-harmonic terms. The numerical code developed in this work is verified by comparison with the published results [15,17]. Some validations are also made by comparison with multilayer grating structures [14]. The convergence is checked by increasing the number of terms $N=2q+1$ from 41, 81, 121, to 201. Notice that while most of the orders are evanescent waves that do not contribute to the reflected energy, it is necessary to include them in the RCWA to represent the Fourier expansion of dielectric function in the grating region and to accurately describe the electromagnetic fields around the grating region. The convergence is the slowest for the TM wave at 200 nm wavelength. The reflectance variation is less than 0.001 between $N=81$ and 121. Therefore, 81 orders of diffraction $(j=0,\pm1,\pm2,\ldots,\pm40)$ are retained in all calculations.

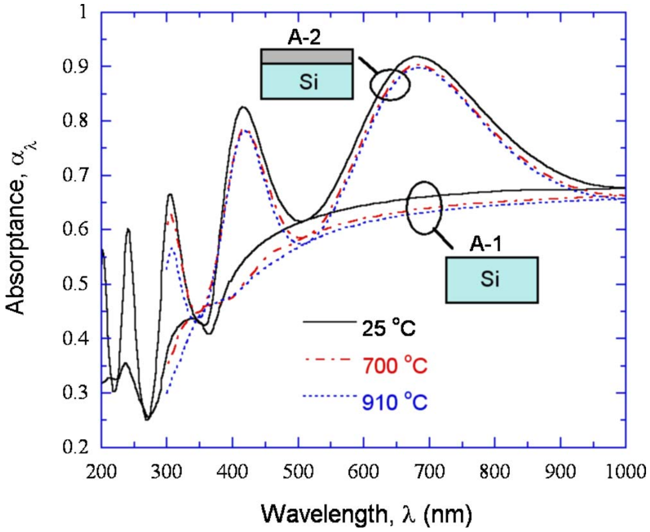


Fig. 4 Calculated normal, spectral absorptance for plain Si (A-1) and SiO$_2$-coated Si (A-2) at 25, 700, and 910°C

The radiative properties calculated from RCWA are successfully compared with the simulation results of patterned wafers from other numerical methods such as FDTD and FVTD [12]. The pattern is a 50-nm-thick Si$_3$N$_4$ grating on a Si substrate with a period of $\Lambda=2$ μm and a filling ratio $\phi=0.5$. In addition, when the filling ratio is 0 or 1, the grating becomes a homogeneous medium. The RCWA simulation gives the same results as these obtained from thin-film multilayer algorithm [32].

A Pentium 4 computer with a 3.2 GHz processor and 2 GB memory is used for computation. In order to compare the calculation efficiency of RCWA and EMT programs, case B-3 is selected due to its complexity. RCWA divides the structure into four regions with two grating layers. On the other hand, EMT approximates the case as two thin films on the Si substrate. The computation speed is 2000 data points per second for EMT and 2–3 data points per second for RCWA. That is to say that EMT algorithm is about 1000 faster than RCWA because of its large number of matrix elements. Additional complexity will arise in RCWA when the wave vector of the incident radiation and the grating vector are not perpendicular to each other. On the other hand, the effective medium formulation can easily be applied to compute the hemispherical properties when the grating region can be approximately isotropic and homogeneous. The validation of the EMT is therefore very important for practical applications.

3 Results and Discussion

Figure 4 shows the spectral absorptance for cases A-1 (plain Si) and A-2 (350-nm-thick SiO$_2$ on Si) at normal incidence. Thin-film optics formulation is sufficient for these calculations without using the RCWA or EMT algorithms. The results are shown in order to see the effect of temperature on the absorptance and to serve as benchmark results for comparison with grating structures later. For plain silicon, it can be seen that the absorptance increases towards longer wavelengths. A common misperception is that a large κ value will give a large absorptance and silicon will have a

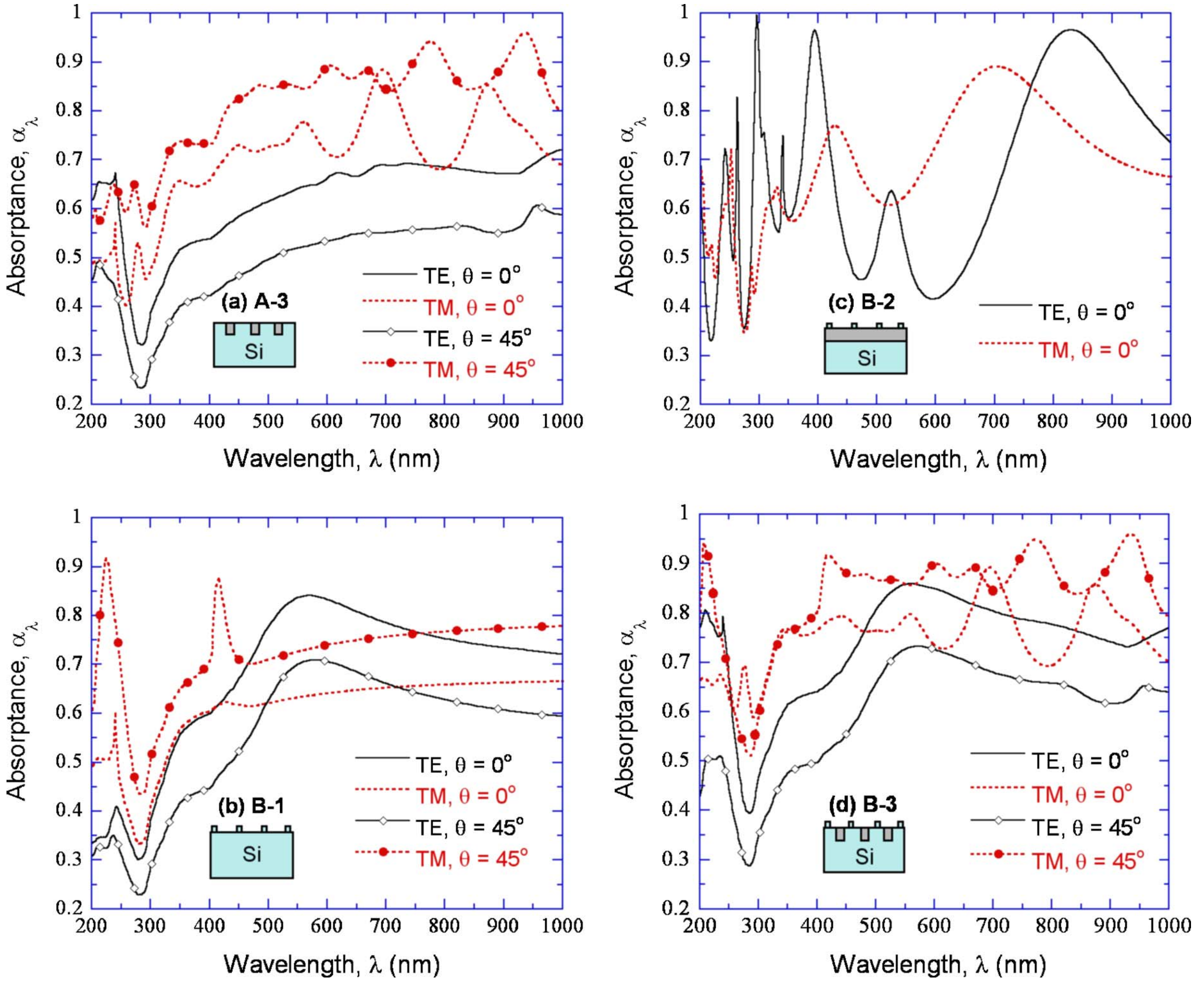


Fig. 5 Spectral absorptance predicted using RCWA for cases with gratings at 910°C at θ=0 and 45 deg: (*a*) case A-3; (*b*) case B-1; (*c*) case B-2; and (*d*) case B-3

large absorptance in the ultraviolet region. On the contrary, the absorptance is the smallest when κ is the greatest at around $\lambda = 280$ nm. This is because of the high reflectivity of Si in this region, similar to metals in the visible region. For normal incidence, the absorptance of plain silicon can be calculated from

$$\alpha_\lambda = \frac{4n_{\mathrm{Si}}}{(n_{\mathrm{Si}}+1)^2 + \kappa_{\mathrm{Si}}^2} \tag{17}$$

When the extinction coefficient is small, the absorptance decreases as the refractive index increases. Increasing the extinction coefficient will significantly reduce the absorptance as it becomes large. On the other hand, the extinction coefficient is directly related to the radiation penetration depth, $\delta_\lambda = \lambda/(4\pi\kappa)$. The calculated δ_λ increases from 4 nm for $\lambda = 200$ nm to ≈ 1 μm for $\lambda = 800$ nm at 910°C. The absorptance with $\mathrm{SiO_2}$ film oscillates due to interference effects. The phase shift in the $\mathrm{SiO_2}$ film is determined by $\beta = 2\pi d_{\mathrm{SiO_2}} n_{\mathrm{SiO_2}}/\lambda$ at normal incidence. The absorptance reaches a minimum close to that of plain Si when β/π is a positive integer. The corresponding wavelengths are $\lambda \approx 1000$, 500, 345, 262, and 214 nm. On the other hand, the absorptance is maximized when $\beta = \pi/2, 3\pi/2, 5\pi/2, \ldots$, corresponding to $\lambda \approx 680$, 420, 298, 236, and 200 nm. In general, $\mathrm{SiO_2}$ film behaves as an antireflection coating that enhances the absorptance [33], except near $\lambda \approx 214$ and 262 nm due to the lower refractive index of silicon at very short wavelengths.

The difference in absorptance at temperatures from 25 to 910°C is less than 0.05 except near $\lambda = 300$ nm, where the maximum difference is 0.08 for a plain Si and 0.12 for a $\mathrm{SiO_2}$ film on a Si substrate. The total absorptance for case A-1 is 0.59, 0.57, and 0.56 and for case A-2 is 0.72, 0.70, and 0.69 at 25, 700, and 910°C, respectively. This suggests that the temperature dependence of the optical constants may not affect the absorption significantly. One should be cautious about a generalization that the temperature dependence is weak to wavelengths longer than 1 μm, where the silicon wafer is semitransparent below 700°C, and consequently, the absorption depends strongly on temperature [13,25]. Because of the importance to high-temperature annealing, all calculations in the following are based on optical constants of Si at 910°C (except in the wavelength region from 200 to 300 nm where room temperature data are substituted).

Figure 5 shows the directional, spectral absorptance of cases A-3, B-1, B-2, and B-3 predicted by RCWA, with different polarizations at incidence angels of 0 and 45 deg. The absorptance for Case B-2 at $\theta = 45$ deg is not presented for the sake of clarity. Because of the gratings, the absorptance depends on the polarization even for normal incidence. When the normal absorptance for case A-3 shown in Fig. 5(*a*) is compared to that of plain silicon shown in Fig. 4, the absorptance for TE waves increases significantly at short wavelengths, by more than 0.3 from 200 to 240 nm, and the enhancement is less than 0.1 at $\lambda > 300$ nm. At wavelengths longer than 280 nm, the absorptance is in general greater for TM waves than for TE waves. Furthermore, the absorption spectrum for TM waves oscillates, but not so much for TE waves. On the contrary, for case B-1 as shown in Fig. 5(*b*), the

Table 2 Total absorptance for different cases at 910°C. For cases A-1 and A-2, thin-film multilayer formulation gives rigorous solution and there is no need of using RCWA or effective medium theory. For the other cases, RCWA gives the rigorous solution. Only the zeroth order EMT results are presented. The average is between results for TE and TM waves, and the difference is between the approximate model and rigorous solution.

Case No.	A-1	A-2	A-3	B-1	B-2	B-3
Angle of incidence $\theta=0$ deg						
TE-rigorous	0.557	0.688	0.614	0.711	0.634	0.747
TM-rigorous	0.557	0.688	0.715	0.615	0.712	0.765
Average	0.557	0.688	0.665	0.663	0.673	0.756
EMA	—	—	0.631	0.647	0.716	0.719
Difference			−0.034	−0.016	0.043	−0.037
TE-0	—	—	0.604	0.823	0.727	0.842
Difference			−0.010	0.112	0.093	0.095
TM-0	—	—	0.774	0.576	0.692	0.791
Difference			0.059	−0.039	−0.020	0.026
Average	—	—	0.689	0.700	0.710	0.817
Difference			0.024	0.037	0.037	0.061
Angle of incidence $\theta=45$ deg						
TE-rigorous	0.441	0.617	0.491	0.580	0.573	0.614
TM-rigorous	0.680	0.737	0.824	0.722	0.757	0.848
Average	0.561	0.677	0.658	0.651	0.665	0.731
EMA-TE	—	—	0.510	0.539	0.659	0.611
Difference			0.019	−0.041	0.086	−0.003
EMA-TM	—	—	0.753	0.723	0.764	0.790
Difference			−0.071	0.001	0.007	−0.058
Average	—	—	0.632	0.631	0.712	0.701
Difference			−0.026	−0.020	0.047	−0.030
TE-0	—	—	0.485	0.710	0.650	0.731
Difference			−0.006	0.130	0.077	0.117
TM-0	—	—	0.872	0.683	0.739	0.873
Difference			0.048	−0.039	−0.018	0.025
Average	—	—	0.679	0.697	0.695	0.802
Difference			0.021	0.046	0.030	0.071

normal absorptance is greater for TE waves than for TM waves at wavelengths longer than 450 nm, whereas the absorptance for TM waves is enhanced at short wavelengths with a peak at $\lambda=240$ nm. The effect of gates on the TE wave absorptance is rather surprising because the gates are only 50 nm high with a poly-Si filling ratio of 0.125. It is also interesting to notice the different effect on polarized absorptance between the gates and trenches.

The absorptance spectra for case B-2 are more complicated due to the coupling between the interference effect of the SiO_2 layer and the grating effect of the gates. Nevertheless, the gates have little effect on the absorptance for TM waves in the long wavelength region, as can be seen by comparison of Fig. 5(*c*) with the absorptance for case A-2 shown in Fig. 4. However, the absorptance for TE waves is dramatically different with and without poly-Si gates. Here again, the gates affect the TE wave absorptance spectra significantly. Interesting enough, the absorptance spectra for case B-3 exhibit the combined effects of gates and trenches, as can be clearly seen from Fig. 5(*d*). For TE waves, the absorptance spectra are similar to those of case A-3 at shorter wavelengths and largely modified by features in case B-1 at longer wavelengths. For TM waves, the absorptance spectra retain features for case B-1 at shorter wavelengths and resemble features shown for case A-3 at longer wavelengths.

The total absorptance calculated with different methods at 910°C for all six cases are listed in Table 2. The total absorptance of cases A-1 and A-2 calculated from thin-film optics is also presented for comparison. The total normal absorptance for plain Si is 0.56, and it increases to 0.61 for TE waves and 0.72 for TM waves in case A-3. On the other hand, the total normal absorptance is 0.71 for TE waves and 0.62 for TM waves in case B-1. Due to the combined effects of gates and trenches, the total normal absorptance of case B-3 is increased to 0.75 for TE waves and 0.77 for TM waves. At normal incidence, the total absorptance averaged over the two polarizations for case B-3 is 0.76, which is 0.20 higher than that of plain Si. The nanoscale gates give rise to the averaged total absorptance by about 0.1 from case A-1 to case B-1 and from case A-3 to case B-3. Similarly, the trench layer increases nearly 0.1 in the averaged total absorptance from case A-1 to case A-3 and from case B-1 to case B-3. On the other hand, the total normal absorptance of case B-2 for the TM wave is only 0.02 higher and that for the TE wave is 0.06 lower than that of case A-2. Therefore, the effect of gates on the total absorptance with SiO_2 film is very small. Note that the total absorptance is also a function of the source spectra.

Generally speaking, when the incidence angle increases from 0 and 45 deg, the absorptance for TM waves increases and that for TE waves decreases, as evidenced from Fig. 5 and Table 2. This is consistent with the Fresnel equation, which predicts that the reflectance of a plain Si for TM waves decreases, until the Brewster angle, and that for TE waves increases when the angle of incidence increases. It should be noticed that the oscillation for TM waves in cases A-3 and B-3 shifts phase when the angle of incidence is changed. While most of the absorptance spectra have a valley at 280 nm, the absorptance of case A-3 reaches a peak at $\theta=0$ and 45 deg and that of case B-3 reaches a peak only at $\theta=0$ deg for TM wave, see Figs. 5(*a*) and 5(*c*). The above discussions clearly demonstrate the complexity of wave interactions inside the grating regions. Nevertheless, some unique features associated with the grating structures can be identified and are elaborated in the next paragraph.

Absorptance peaks exist in the absorptance spectra at $\lambda = 240$ nm for normal incidence, as can be seen from Fig. 5 for both polarization. The features in Fig. 5(*c*) are too complicated to distinguish this peak. While there are shoulders in the optical constants of Si near $\lambda = 240$ nm and a peak in the absorptance of plain silicon, the peaks are much sharper and more prominent with gratings, especially for TM waves in cases A-3 and B-1. The very sharp peaks are caused by the well-known Wood's anomaly in the diffraction grating theory [34]. Wood's anomaly can cause an abrupt, asymmetric change in the absorptance spectrum when a diffraction order just appears at the grazing angle. Since the grating period $\Lambda = 240$ nm, for normal incidence at $\lambda = 240$ nm, the +1 and −1 diffraction order will be at a diffraction angle $\theta_{\pm 1} = \pm 90$ deg. According to the grating equation Eq. (3*b*), at 45 deg incidence, $\theta_{-1} = -90$ deg when $\lambda = 410$ nm. Wood's anomaly can be clearly seen for the TM wave at $\theta = 45$ deg in Fig. 5(*b*) and in Fig. 5(*d*) with a somewhat reduced and shifted peak due to the influence of the trench layer.

Figure 6 compares the normal absorptance spectra for cases A-3, B-1, and B-3 predicted by RCWA with different formulations of effective medium approaches. In the graphs, 0 and 2 refer to predictions based on EMT-0 and EMT-2, respectively. Note that for normal incidence, the predicted absorptance by EMA is the same for both polarizations. In Fig. 6(*a*), the results of both EMA and TE-0 agree well with those of RCWA at $\lambda \geqslant 250$ nm. While TE-2 follows the RCWA result closer near $\lambda = 1000$ nm, it underpredicts the absorptance at shorter wavelengths. The reason of disagreement is due to the second-order correction term of the dielectric function. Because the condition that $\lambda \gg \Lambda$ is not satisfied for the present study, the second-order term can be so large to produce a nonphysical dielectric function whose imaginary part is negative at $\lambda \leqslant 300$ nm for the trench region. In Fig. 6(*b*), the absorption spectra calculated from the effective medium formulations are not in phase with that from RCWA and none of them can predict the spectral absorptance well. On the other hand, if the absorptance is averaged over the spectral region or uses the source spectrum to obtain the total absorptance, one should not expect significant deviations from the RCWA if EMA and EMT-0 are used. The insignificant deviation comes from the spectrally periodic behavior though errors are not necessarily averaged out. This is indeed the case as can be seen from Table 2, where the total absorptance obtained from EMA and EMT-0 is compared with that from the rigorous solutions and the differences are also shown.

Figures 6(*c*) and 6(*d*) show absorptance of case B-1 for both TE and TM waves, respectively. The absorptance predicted by TE-0 and TE-2 are in good agreement with that using RCWA at $\lambda \geqslant 600$ nm. In addition, TE-0 can capture some of the features in RCWA results at short wavelengths, except that it largely overpredicts the absorptance. TE-2 again failed at shorter wavelengths completely. Compared with the RCWA prediction, EMA significantly overpredicts the absorptance at $\lambda \leqslant 400$ nm and underpredicts the absorptance at $\lambda \geqslant 400$ nm than that of RCWA. Figure 6(*d*) shows that all the effective medium formulations give good agreement with results from RCWA, except EMA overestimates the absorptance at $\lambda \leqslant 300$ nm. This behavior may arise because the absorptance at longer wavelengths changes a lot for the TE wave but very little for the TM wave when the gates are present. Similar observations can be made based on the absorptance spectra for case B-2, which is not shown. When comparing different models for case B-3, the spectra using TE-2 and TM-2 are not shown in Figs. 6(*e*) and 6(*f*), because the results are not as good as those of TE-0 and TM-0. The spectra from RCWA clearly show features that arise from the behavior seen in the absorptance for cases A-3 and B-1. Because the pattern of case B-3 includes two layers of periodic structure, none of the effective medium formulations can predict the complex absorptance spectra of the two-layer grating structures.

Although the effective medium formulations failed to describe the spectral behavior of the gratings studied here over the whole wavelength range, they may provide reasonable values for the total absorptance because the spectral absorptance oscillates and crosses one another in some regions. The absorptance from EMT-2 is not shown in Table 2 because adding the second-order term can only cause EMT to fail at short wavelengths. EMA gives good agreements with the total absorptance for most cases at both normal and oblique incidence. The average difference between two polarizations is less than 0.05. On the other hand, EMA predictions often have a larger disagreement for one polarization. For EMT-0, the averaged total absorptance has a slightly larger difference for cases B-1 and B-3 compared with EMA. Though EMT-0 deals different polarized wave separately, the average difference for case B-3 is 0.06 and 0.07 at $\theta = 0$ and 45 deg, respectively. In addition, the difference for the TE wave is greater than 0.1 for case B-1. It is interesting that EMT-0 always gives a higher estimate of the averaged total absorptance than RCWA. It can be concluded that EMA gives a better approximation for calculating the total absorptance for the studied cases. It can also be easily applied to compute the hemispherical properties. However, one must carefully assess the error introduced by EMA for different nanostructures, polarization, and source spectral distribution.

The effects of incidence angle on the absorptance are presented in Figs. 7 and 8, for all cases, at representative wavelengths, i.e., 200, 400, 600, and 800 nm. In the calculation, the RCWA algorithm is used for all cases with gratings. In Figs. 7(*a*) and 7(*b*), Si behaves like a dielectric in the long wavelength region, and the absorptance reaches unity for TM waves at the Brewster angle, where the reflectance is zero. However, the peak absorptance for TM waves at wavelengths of 200 and 400 nm does not reach one because of the large extinction coefficient of Si at these wavelengths (2.9 and 3.6, respectively). For lossy materials such as a metal or Si in the ultraviolet region, there exists a (nonzero) minimum when the reflectance for the TM wave is plotted against the incidence angle. The angle at which the ratio of the reflectance for the TM wave and TE wave is minimized is called the principal angle [33]. Therefore, the absorptance for the TM wave reaches a peak near the principal angle. As shown in Figs. 7(*c*) and 7(*d*), the absorptance for a thin SiO_2 film on top of a Si substrate exhibits a large variation with the angle, especially at short wavelengths. The absorptance oscillation can be explained by interference effects. The phase shift in the SiO_2 film is determined by $\beta = 2\pi d_{SiO_2}\sqrt{n_{SiO_2}^2 - \sin^2\theta}/\lambda$. At $\lambda = 200$ nm, absorptance minimum is around $\theta = 37$ deg that corresponds to $\beta = 5\pi$ and the maximum is around $\theta = 60$ deg that corresponds to $\beta = 9\pi/2$. The exact peak and valley locations depend on the polarization.

As can be seen from Figs. 7(*e*) and 7(*f*), the angular dependence of the absorptance for case A-3 is similar to that for case A-1. The Brewster angles and principle angles can be clearly seen. Case A-3 also exhibits some distinctly different features from case A-1. For example, the absorptance at 200 nm is greater than that at 400 nm for TE wave, and the absorptance of 600 nm is greater than that of 800 nm at $\theta < 30$ deg and $\theta > 70$ deg. Other significant difference can be seen in the abrupt changes of absorptance at 200 and 400 nm due to Wood's anomaly. When $\lambda = 200$ nm, the predicted $j = +1$ diffraction order from Eq. (3*b*) becomes $\theta_{+1} = 90$ deg at incidence angle $\theta = 9.6$ deg. The $j = -2$ diffraction order becomes $\theta_{-2} = 90$ deg at $\theta = 41.8$ deg. These anomalies are clearly seen in the angular dependent absorptance. Likewise, at $\theta = 41.8$ deg, the $j = -1$ diffraction order for $\lambda = 400$ nm also appears at the grazing angle.

Figure 8 shows the absorptance for cases with poly-Si gates. First of all, the Wood's anomalies also clearly show up for cases B-1 and B-3 because of the same grating periods used. However,

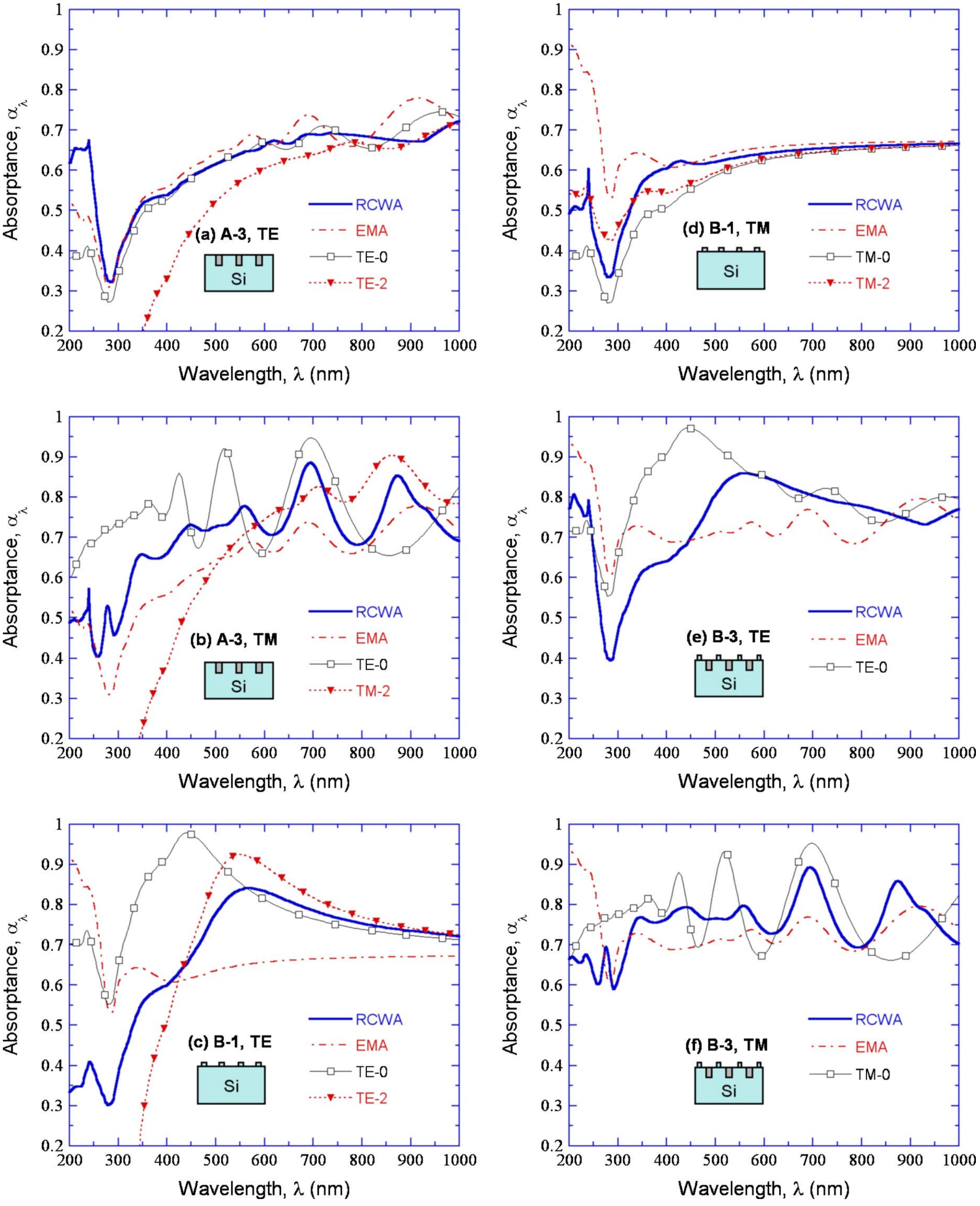


Fig. 6 Comparison of the absorptance predicted by different methods at 910°C: (*a*) TE wave for case A-3; (*b*) TM wave for case A-3; (*c*) TE wave for case B-1; (*d*) TM wave for case B-1; (*e*) TE wave for case B-3; and (*f*) TM wave for case B-3

it is difficult to identify Wood's anomaly for case B-2 because of the complex structures in the angular-dependent absorptance. Next, the B cases can be compared with the A cases to examine the effects of the periodic gates. When case B-1 is compared with case A-1, the absorptance for TE waves at all four wavelengths increases, especially at λ=600 nm, which exceeds that at λ=800 nm. The absorptance for TM waves does not reach unity at the Brewster angle for λ=600 and 800 nm and the principal angles become obscure for λ=200 and 400 nm. Figures 8(*c*) and 8(*d*) reveal the most complicated behaviors for case B-2, due to the coupling between the diffraction effects with the interference effect inside SiO_2. A peak absorptance close to unity occurs for the TE wave at $\theta \approx 40$ deg and λ=800 nm. On the other hand, the absorptance for the TE wave is the lowest at λ=600 nm. For the TM wave, the peak absorptance at λ=600 nm exists at a different angular position compared with that for cases B-1 and B-3. Although the structure is more complex than that of case B-2, the absorptance trends for case B-3 are much simpler than those for

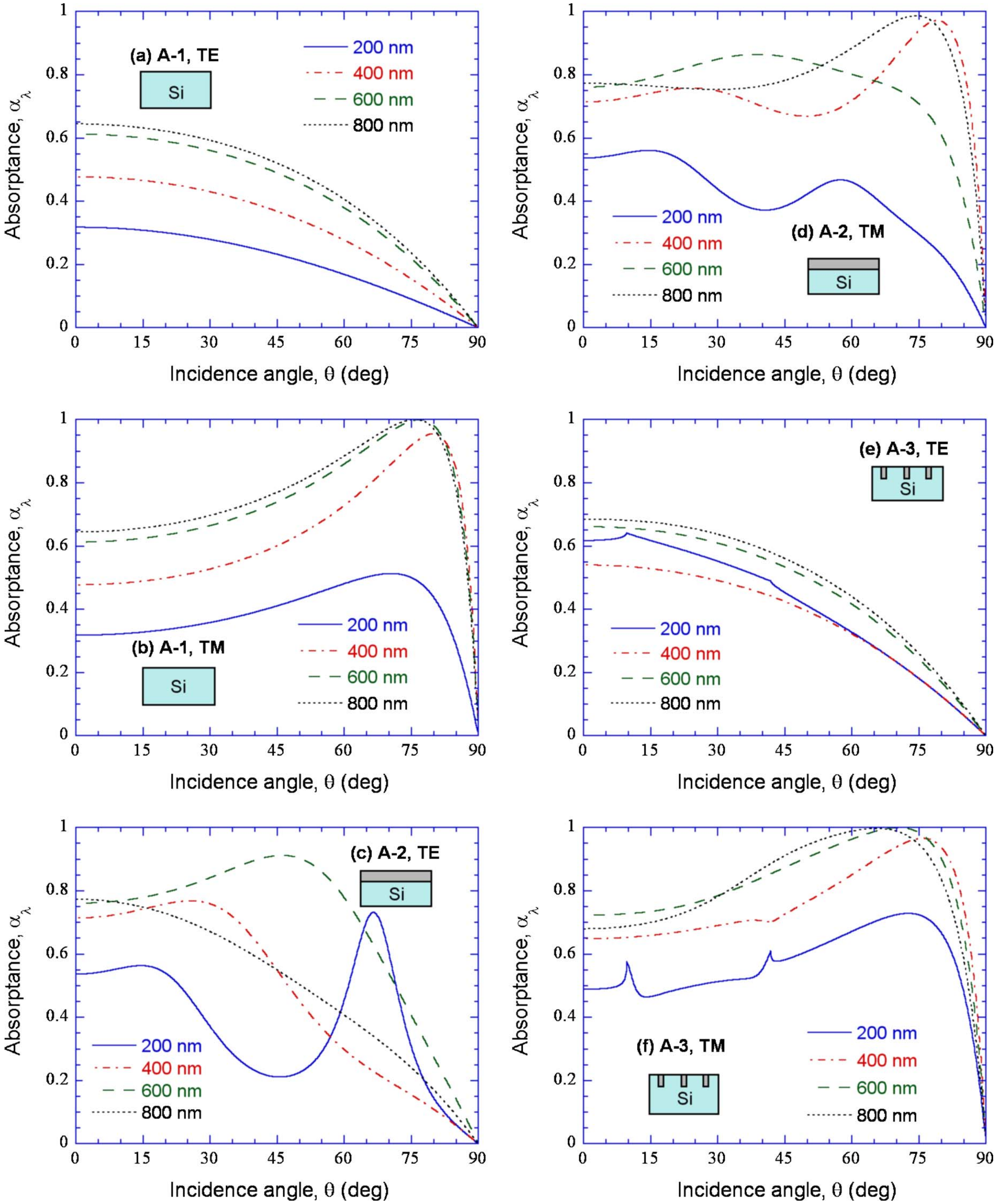


Fig. 7 Effect of incidence angle on the absorptance at 910°C for cases without gates: (*a*) TE wave for case A-1; (*b*) TM wave for case A-1; (*c*) TE wave for case A-2; (*d*) TM wave for case A-2; (*e*) TE wave for case A-3; and (*f*) TM wave for case A-3

case B-2, as can be seen from Figs. 8(*e*) and 8(*f*). In fact, except the much larger enhancement in the absorptance for case B-3, the features in case B-3 are comparable to those shown in cases A-3 and B-1.

4 Conclusions

This paper presents numerical predictions of the absorptance of selected nanoscale patterns that may be found in the device structures in advanced CMOS technology. The effects of temperature, wavelength, polarization, and angle of incidence are systematically studied. Comparison is made to different patterning structures to examine the effect of nanostructures on the radiative properties. Some major conclusions are summarized below.

1. In the cases studied, for wavelengths between 200 and 1000 nm, temperature does not affect the absorptance significantly when comparing absorptance at temperatures of 25, 700, and 910°C.

2. Periodic patterns, including arrays of gates and trenches lead

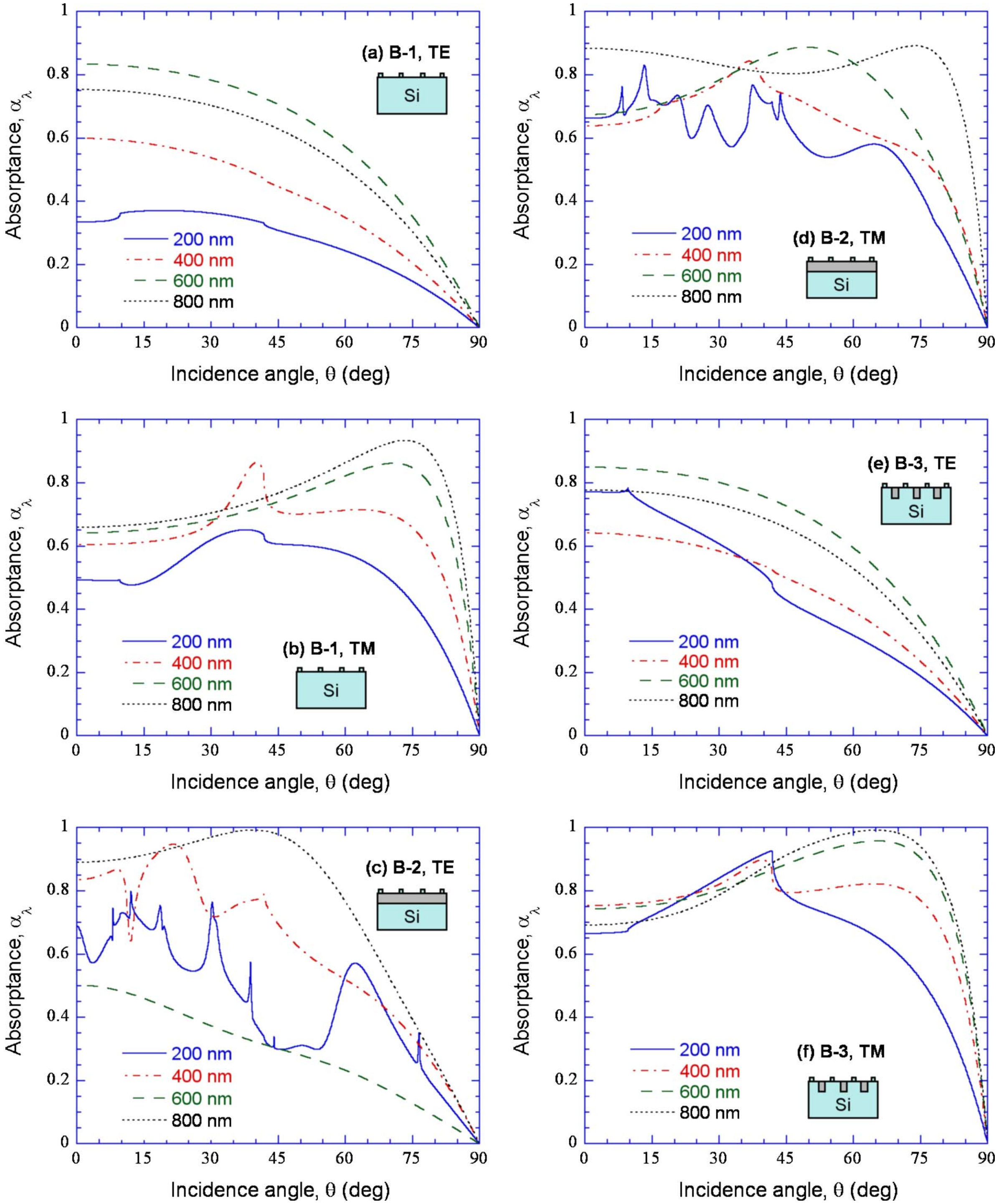


Fig. 8 Effect of incidence angle on the absorptance at 910°C for cases with gates: (*a*) TE wave for case B-1; (*b*) TM wave for case B-1; (*c*) TE wave for case B-2; (*d*) TM wave for case B-2; (*e*) TE wave for case B-3; and (*f*) TM wave for case B-3

to diffractions and, hence, affect the absorptance in a complex manner. Different polarizations of the incident radiation lead to different responses from lateral periodic structures. It is particularly notable that the gates affect the absorptance for the TE wave significantly but the trenches affect the absorptance for the TM wave significantly.

3. Because the pattern of case B-3 is composed of patterns in cases A-3 and B-1, its absorptance spectra exhibit features shown in both cases. The total absorptance averaged over two polarizations for case B-3 is 0.2 greater than that of plain Si. On the other hand, the average total absorptance for case A-3 and for case B-1 is about 0.1 greater than that of plain Si.

4. While case B-2 is not the most complicated structure, the coupling of diffraction and thin-film effects result in very complicated features in the absorptance spectra and angular distributions. On the other hand, the average total absorptance for case B-2 is similar to case A-2 that does not contain the gates.

5. Wood's anomalies introduce abrupt changes in absorptance at

short wavelengths and can be seen clearly in several absorptance spectra and angular distribution curves. The appearance of such anomalies can be predicted by the well-known grating equation.

6. Three effective medium formulations are compared with the RCWA to evaluate their applicability to predict the absorptance for the studied nanostructures. EMT-2 does not give good results and gives nonphysical values at short wavelengths. EMA and EMT-0 cannot reasonably predict the absorptance in the whole spectral range for both polarizations. On the other hand, the averaged total absorptance of EMA and that of EMT-0 are in reasonable agreement with that predicted by RCWA, within a difference of 0.05 for EMA and 0.07 for EMT-0. Although in some cases approximation formulations may be used to compute the average total absorptance, they cannot substitute the rigorous solutions for the spectral and polarized radiative properties.

In the future, this research may be extended to the study of the locally absorbed energy distribution, which may then be coupled with a heat conduction model to examine the effect of temperature nonuniformity on the dynamic stress field. Attention will also be paid to the 3D effect, when the incident wave vector has a nonzero y component.

Acknowledgment

The work at Georgia Institute of Technology was supported by the National Science Foundation (CTS-0500113). The authors thank Dr. Qunzhi Zhu for helping develop the RCWA computer program and Bong Jae Lee for valuable comments.

Nomenclature

d = depth, m
$\mathbf{E}$ = electric field vector, V/m
G = spectral irradiation, $\mathrm{W/m^2\,nm}$
$\mathbf{H}$ = magnetic field vector, C/m s
j = diffraction order
$\mathbf{K}$ = grating vector, $\mathrm{m^{-1}}$
$\mathbf{k}$ = wave vector, $\mathrm{m^{-1}}$
l = length, m
n = refractive index
R = reflection coefficient
T = transmission coefficient

Greek Symbols

α = absorptance
β = phase shift, rad
ϵ = dielectric function
θ = polar angle, rad
κ = extinction coefficient
Λ = grating period, m
λ = wavelength in vacuum, m
ϕ = filling ratio
ψ = amplitude of the electric field

References

[1] Semiconductor Industry Association, 2004, International Technology Roadmap for Semiconductors 2004 Update (http://public.otrs/net).

[2] Gelpey, J. C., Elliott, K., Camm, D., McCoy, S., Ross, J., Downey, D. F., and Arevalo, E., 2002, "Advanced Annealing for Sub-130 nm Junction Formation," *Rapid Thermal and Other Short-Time Processing Technologies III*, P. J. Timans, E. Gusev, F. Roozeboom, M. C. Öztürk, and D.-L. Kwong, eds., The Electrochemical Society, Pennington, NJ, pp. 313–324.

[3] Skorupa, W., Yankov, R. A., Anwand, W., Voelskow, M., Gebel, T., Downey, D. F., and Arevalo, E. A., 2004, "Ultra-Shallow Junctions Produced by Plasma Doping and Flash Lamp Annealing," Mater. Sci. Eng., B, **114**, pp. 358–361.

[4] Lindsay, R., Pawlak, B. J., Henson, K., Satta, A., Severi, S., Lauwers, A., Surdeanu, R., McCoy, S., Gelpey, J., Pages, X., and Maex, K., 2004, "Integration of Low and High Temperature Junction Anneals for 45 nm CMOS," *Advanced Short-Time Thermal Processing for Si-Based CMOS Devices II*, M. C. Öztürk, E. P. Gusev, L. J. Chen, D.-L. Kwong, P. J. Timans, G. Miner, and F. Roozeboom, eds., The Electrochemical Society, Pennington, NJ, pp. 145–156.

[5] Nishinohara, K. T., Ito, T., and Suguro, K., 2004, "Improvement of Performance Deviation and Productivity of MOSFETs With Gate Length Below 30 nm by Flash Lamp Annealing," IEEE Trans. Semicond. Manuf., **17**, pp. 286–291.

[6] Bentini, G. G., and Correra, L., 1983, "Analysis of Thermal Stresses Induced in Silicon During Xenon Arc Lamp Flash Annealing," J. Appl. Phys., **54**, pp. 2057–2062.

[7] Madou, M. J., 1997, *Fundamentals of Microfabrication*, CRC Press, Boca Raton, FL.

[8] Gu, L. X., Fu, C. J., Niess, J., Timans, P. J., and Zhang, Z. M., 2003, "Modeling the Temperature and Stress of the Wafer Around the Support Pin During Rapid Thermal Processing," *Proc. 6th ASME-JSME Thermal Eng. Joint Conf.*, Paper No. AJ03–587.

[9] Erofeev, A. F., Kolpakov, A. V., Makhviladze, T. M., Martjushenko, A. V., Panjukhin, A. V., Volchek, O. S., and Orlowski, M., 1995, "Comprehensive RTP Modeling and Simulation," *Proc. 3rd Int. Rapid Thermal Processing Conf.*, pp. 181–197.

[10] Hebb, J. P., and Jensen, K. F., 1998, "The Effect of Patterns on Thermal Stress During Rapid Thermal Processing of Silicon Wafers," IEEE Trans. Semicond. Manuf., **11**, pp. 99–107.

[11] Tada, H., Abramson, A. R., Mann, S. E., Miaoulis, I. N., and Wong, P. Y., 2000, "Evaluating the Effects of Thin Film Patterns on the Temperature Distribution of Silicon Wafers During Radiant Processing," Opt. Eng. (Bellingham), **39**, pp. 2296–2304.

[12] Liu, J., Zhang, S. J., and Chen, Y. S., 2004, "Rigorous Electromagnetic Modeling of Radiative Interactions With Microstructures Using the Finite Volume Time-Domain Method," Int. J. Thermophys., **25**, pp. 1281–1297.

[13] Zhang, Z. M., Fu, C. J., and Zhu, Q. Z., 2003, "Optical and Thermal Radiative Properties of Semiconductors Related to Micro/Nanotechnology," Adv. Heat Transfer, **37**, pp. 179–296.

[14] Moharam, M. G., Pommet, D. A., Grann, E. B., and Gaylord, T. K., 1995, "Stable Implementation of the Rigorous Coupled-Wave Analysis for Surface-Relief Gratings: Enhanced Transmittance Matrix Approach," J. Opt. Soc. Am. A, **12**, pp. 1077–1086.

[15] Moharam, M. G., Grann, E. B., Pommet, D. A., and Gaylord, T. K., 1995, "Formulation for Stable and Efficient Implementation of the Rigorous Coupled-Wave Analysis of Binary Gratings," J. Opt. Soc. Am. A, **12**, pp. 1068–1076.

[16] Li, L. F., 1996, "Use of Fourier Series in the Analysis of Discontinuous Periodic Structures," J. Opt. Soc. Am. A, **13**, pp. 1870–1876.

[17] Auslender, M., and Hava, S., 1995, "Zero Infrared Reflectance Anomaly in Doped Silicon Lamellar Gratings. I. From Antireflection to Total Absorption," Infrared Phys. Technol., **36**, pp. 1077–1088.

[18] Raguin, D. H., and Morris, G. M., 1993, "Antireflection Structured Surfaces for the Infrared Spectral Region," Appl. Opt., **32**, pp. 1154–1167.

[19] Sentenac, A., and Greffet, J.-J., 1994, "Design of Surface Microrelief with Selective Radiative Properties," Int. J. Heat Mass Transfer, **37**, pp. 553–558.

[20] Edwards, D. F., 1998, "Silicon(Si)," *Handbook of Optical Constants of Solids*, E. D. Palik, ed., Academic, San Diego, CA, pp. 547–569.

[21] Jellison, G. E., Jr., and Modine, F. A., 1983, "Optical Functions of Silicon Between 1.7 and 4.7 eV at Elevated Temperatures," Phys. Rev. B, **27**, pp. 7466–7472.

[22] Jellison, G. E., Jr., and Modine, F. A., 1994, "Optical Functions of Silicon at Elevated Temperatures," J. Appl. Phys., **76**, pp. 3758–3761.

[23] Sun, B. K., Zhang, X., and Grigoropoulos, C. P., 1997, "Spectral Optical Functions of Silicon in the Range of 1.13–4.96 eV at Elevated Temperatures," Int. J. Heat Mass Transfer, **40**, pp. 1591–1600.

[24] Malitson, I. H., 1965, "Interspecimen Comparison of the Refractive Index of Fused Silica," J. Opt. Soc. Am., **55**, pp. 1205–1209.

[25] Timans, P. J., 1996, *Advances in Rapid Thermal and Integrated Processing*, F. Rooseboom, ed., Kluwer Academic, Dordrecht, Chap. 2.

[26] Lalanne, P., and Lemercier-Lalanne, D., 1996, "On the Effective Medium Theory of Subwavelength Periodic Structures," J. Mod. Opt., **43**, pp. 2063–2085.

[27] Maxwell Garnett, J. C., 1904, "Colours in Metal Glasses and in Metallic Films," Philos. Trans. R. Soc. London, Ser. A, **203**, pp. 385–420.

[28] Granqvist, C. G., and Hunderi, O., 1977, "Optical Properties of Ultrafine Gold Particles," Phys. Rev. B, **16**, pp. 3513–3538.

[29] Bruggeman, D. A. G., 1935, "Calculation of Various Physics Constants in Heterogenous Substances I. Dielectricity Constants and Conductivity of Mixed Bodies From Isotropic Substances," Ann. Phys., **24**, pp. 636–664.

[30] Rytov, S. M., 1956, "Electromagnetic Properties of a Finely Stratified Medium," Sov. Phys. JETP, **2**, pp. 466–475.

[31] Glytsis, E. N., and Gaylord, T. K., 1992, "High-Spatial-Frequency Binary and Multilevel Stairstep Gratings: Polarization-Selective Mirrors and Broadband Antireflection Surfaces," Appl. Opt., **31**, pp. 4459–4469.

[32] Lee, B. J., Zhang, Z. M., Early, E. A., DeWitt, D. P., and Tsai, B. K., 2005, "Modeling Radiative Properties of Silicon With Coatings and Comparison With Reflectance Measurements," J. Thermophys. Heat Transfer, **19**, pp. 558–569.

[33] Born, M., and Wolf, E., 1999, *Principles of Optics*, 7th ed., Cambridge University Press, Cambridge.

[34] Hessel, A., and Oliner, A. A., 1965, "A New Theory of Wood's Anomalies on Optical Gratings," Appl. Opt., **4**, pp. 1275–1297.

Journal of Heat Transfer

Technical Briefs

Infrared Microscopy Thermal Characterization of Opposing Carbon Nanotube Arrays

X. Jack Hu[1]
e-mail: jack.hu@intel.com

Matthew A. Panzer

Kenneth E. Goodson

Mechanical Engineering Department,
Stanford University,
440 Escondido Mall,
Stanford, CA 94305

Carbon nanotubes (CNTs) have received much recent research interest for thermal management applications due to their extremely high thermal conductivity. An advanced thermal interface structure made of two opposing, partially overlapped CNT arrays is designed for thermally connecting two contact surfaces. The performance of this interface structure is thermally characterized using diffraction-limited infrared microscopy. Significant temperature discontinuities are found at the CNT-CNT contact region, which indicates a large thermal resistance between CNTs. Due to this intertube resistance, the thermal performance of the CNT-based interface structure is far below expectation (with a thermal resistance value about 3.8×10^{-4} K m^2/W*).*

[DOI: 10.1115/1.2401202]

Keywords: contact thermal resistance, carbon nanotube, thermal interface material, thermal infrared microscopy

Introduction

Carbon nanotubes (CNTs), a man-made material first reported by Iijima in 1991 [1], are promising for advanced thermal management because of their extremely high thermal conductivity. Hone et al. [2] found that the thermal conductivity of aligned single-walled nanotube (SWNT) ropes is about 250 W/m K at 300 K, and that the thermal conductivity of an individual SWNT in the longitude direction ranges from 1750 to 5800 W/m K. The thermal conductivities of individual CNTs (multi-walled [3] or single-walled [4]) have also been experimentally tested: at least 3000 W/m K at room temperature, which is about one order higher than that of the high thermal conductivity materials commonly used for thermal management (for example, copper). Theoretical studies even predict much higher thermal conductivities, such as 6600 W/m K at room temperature reported by Berber et al. [5].

The technique of using CNTs for thermal management, however, is not straightforward due to the nanoscale nature of CNTs. Early attempts used CNTs as fillers to form high thermal conductivity fluids or composites. Choi et al. [6] measured the effective thermal conductivity of nanotube-in-oil suspensions, and found that with only 1 vol.% of nanotubes, the effective thermal conductivity can be 2.5 times the value of the base fluid. Such an increase in thermal conductivity has never been found previously with any other particles. Biercuk et al. [7] also found that epoxy filled with 1 wt.% of CNTs showed a 70% increase in thermal conductivity at 40 K and 125% at room temperature. Hu et al. [8] proposed the combined use of CNTs and traditional heat conductive fillers for thermal interface materials (TIMs), achieving a thermal conductivity value seven times that of the base fluid, and almost doubling the thermal conductivity of the corresponding TIM composed of only traditional fillers. However, two problems with CNT-based TIMs lead to low thermal conduction efficiencies. One problem is that CNTs are randomly dispersed, and thus only a small portion of CNTs are effectively contributing to heat conduction. The other problem is that heat is not directly conducted from one side to the other through CNTs. CNTs are discontinued by other fillers or the base fluid. The low thermal conductivity of the interstitial media, as well as the contact resistance between those and the CNTs [9–11], degrades the thermal performance of CNT composites.

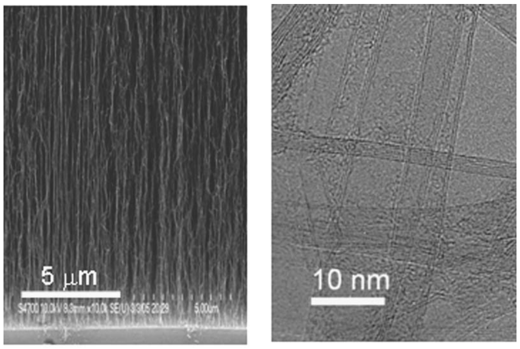
Most recent interest focuses on growing CNTs directly on a silicon or copper substrate, with the CNTs oriented in the direction of heat conduction (i.e., perpendicular to the substrate). These CNTs are mechanically like an elastic cushion due to their high aspect ratio and mechanical strength [12], and therefore can be used between a thermal expansion mismatched interface (for example, the interface between a CPU die and its heat spreader, or the interface between a heat spreader and a heat sink) as a highly heat conductive interface structure [13–17]. Hu et al. [18] measured the effective thermal conductivity of this CNT layer, including the effects of voids between CNTs, to be about 80 W/m K, which is one order higher than that of thermal greases (the most widely used TIMs). Unfortunately, the over all thermal resistance across the CNT layer, including CNT contacts, is still too large due to the thermal bottleneck where the CNTs contact the other interface.

The understanding of heat conduction at CNT contacts is very limited. Part of the problem is the lack of effective thermal characterization method capable of resolving temperature distributions across CNT contacts at small scale. In this paper, we attempted to use the diffraction-limited infrared microscopy to measure the heat conduction in a CNT-based thermal interface structure. This CNT structure is formed by growing two opposing CNT arrays from two facing surfaces. The purpose of growing CNT from both surfaces is to avoid the poor CNT-surface contact. A similar struc-

---

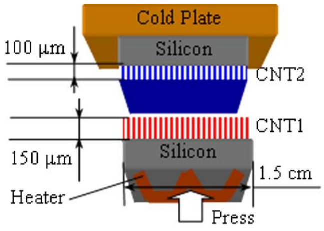
[1] Current address: Intel Corporation, 5000 W. Chandler Blvd, CH5-157, Chandler, AZ 85226.

Contributed by the Heat Transfer Division of ASME for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received January 16, 2006; final manuscript received June 25, 2006. Review conducted by Zhuomin M. Zhang. Paper presented at the 2005 ASME International Mechanical Engineering Congress (IMECE2005), November 5–11, 2005, Orlando, Florida, USA.

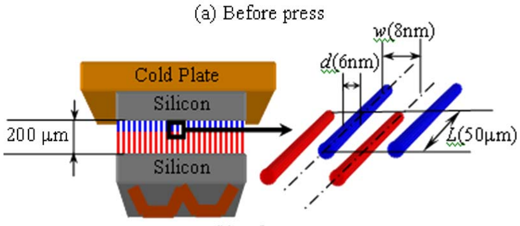


(a) SEM image (b) TEM image

Fig. 1 Scanning electron microscopy and transmission electron microscopy images of vertically oriented CNTs grown on silicon substrates (Courtesy of Molecular Nanosystems Inc.)



(a) Before press



(b) After press

Fig. 2 Schematic of a CNT interface structure with two partially overlapped opposing CNT arrays

ture has been made by [19]. This work addresses the remaining question concerning the thermal performance of the CNT-CNT contact region, in which CNTs are discontinued and heat is forced to transport from one CNT to another.

Infrared Microscopy Measurement

Sample Preparation. An ideal interface structure with partial CNT overlaps is difficult to grow directly due to the limitations of the CNT growth process. To grow vertically oriented CNTs, chemical vapor deposition (CVD) or plasma enhanced chemical vapor deposition are generally used. CNTs are assembled with carbon atoms from hydrocarbon gases though a well-controlled oxidation process with proper catalyst. When the CNTs are grown from two facing surfaces close to each other, the CNTs will block the gas flow, shutting themselves off from the carbon source, and thereby prohibiting the growth necessary to form an interface structure with CNT overlaps.

In this study, we make the CNT interface structure using two separate CNT samples (see Fig. 1 for scanning electron microscope (SEM) and transmission electron microscope (TEM) images). These CNT samples were grown on the rough side of 15 mm × 15 mm single-side-polished silicon substrates using thermal CVD. On the back side of one sample, a tungsten heater was integrated. The length of the CNTs on the heater substrate (CNT1) is about 150 μm and the length of the CNTs on the other chip (CNT2) is about 100 μm, estimated by regular optical microscope. All these CNTs are oriented perpendicular to the silicon substrates, and on average, the diameter of the tubes (d) is about 6 ± 2 nm, and the half-distance between the axes of two tubes (w) is about 8 nm, based on high-magnification SEM and TEM images. Most of the tubes are double walled.

The CNT interface structure is made by pressing the two pieces of the CNT samples together, as illustrated in Fig. 2. After removing the pressure, the two samples remain mechanically bonded to each other due to the strong van der Waals force between CNTs. The final distance between the two substrates, without any press, is about 200 μm, and therefore the CNT overlap region L is estimated to be around 50 μm.

Experiment Setup and Method. The experiment system is given in Fig. 3. The CNT interface structure is attached to a copper cold plate using thermal grease. The cold plate is cooled by a thermoelectrical cooler and a water-cooling heat sink. The integrated heater on the other side of the CNT sample is connected using two probes. This whole structure, as shown in part (a), is then mounted on a stage with precisely controlled translational and rotational movement, and placed under an infrared microscope, as shown in part (b). The IR microscope features a 256 × 256 InSb focal plane array, with a detection wavelength ranging from 3 to 5 μm and a temperature sensitivity of 0.1 K. The spatial resolution of the IR microscope is about 2 μm, which is close to the diffraction limit at the given wavelength. The surfaces exposed to the IR microscope, including the CNT sample to be examined, the copper cold plate, and the thermal grease in between, are coated with a thin layer of amorphous carbon to achieve relatively uniform emissivity.

The experimental system needs a two-step calibration process prior to the measurement. The first step is to map the radiance intensity, including the effects of radiance from ambient, with absolute temperatures. This step is done by testing a standard blackbody at given temperatures. The second step is to map out the emissivity of the carbon coating, which may not be exactly uniform. During this calibration step, the thermo-electrical cooler is operated in its heating mode. The whole structure, including the copper plate and the CNT sample, is uniformly heated and reaches a stable final temperature that can be measured by the thermocouples attached to the backside of the copper plate. The temperature uniformity of the test surface is checked under this isothermal condition, and the calibrated radiance is used as a reference to determine the emissivity of the carbon coating, including the effect of surrounding radiance. The relationship between the reference radiance and the effective emissivity has been given in [20].

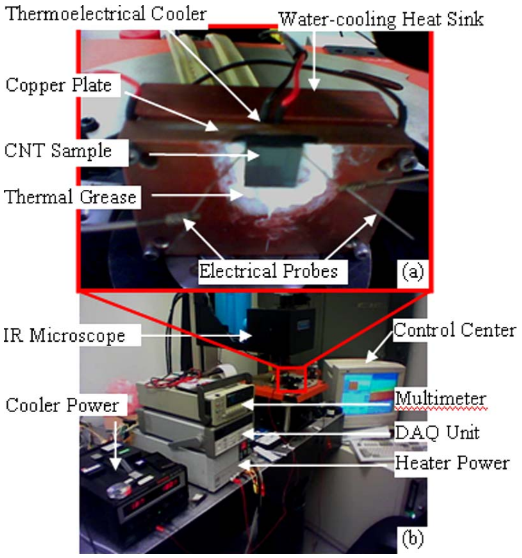


Fig. 3 Experimental system for measuring the temperature profile along a CNT interface structure subject to a constant heat flux

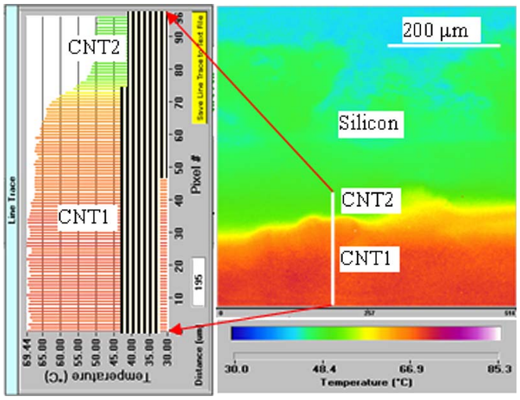


Fig. 4 Measured temperature profile across a CNT interface structure at a heat flux of 10 W/(15 mm×15 mm)

Results and Discussion

The temperature field across the thermal interface structure at a constant heating power, i.e., $P=10$ W, is given in Fig. 4. On the right side of the figure is a two-dimensional thermal image taken by the IR microscope, and on the left side is the temperature profile along the white line across the CNT layers (parallel to the CNT orientation). Apparently, there are two temperature regions: a hot region ranging from the integrated heater until the end of the CNT1, with a temperature T_1 of about 340 K, and a cold region, ranging from the end of CNT1, including CNT2, until the silicon substrate, with a temperature T_2 of about 323 K. Within both the hot region and the cold region, the temperature distributions are uniform, indicating that aligned CNTs are good heat conductors along the axial direction, and that the contact resistance at the CNT growth surface is small. However, an apparent temperature drop is found at the end of the CNT-CNT overlap region. This temperature drop implies a significant CNT intertube thermal resistance in the CNT overlap region, which is estimated to be about 1.4 K m^2/W, based on an effective tube heat exchange surface area $A_e=\pi dL$, where d and L are the tube diameter and the length of tube overlap, respectively. The equation for estimating this intertube thermal resistance is

$$R_t=\frac{T_1-T_2}{(P/\phi A_s)(A_t/A_e)} \tag{1}$$

where T_1 and T_2 are the temperatures of CNT1 and CNT2, P is the heater power, ϕ is the tube volume fraction, A_s is the area of the CNT interface structure, and $A_t=\pi d^2/4$ is the cross-sectional area of a single tube.

Due to the large intertube thermal resistance, the thermal performance of the CNT interface structure is far below expectation. The data indicate that the thermal resistance of the interface structure (R_s) are about 3.8×10^{-4} K m^2/W according to

$$R_s=\frac{T_1-T_2}{P/A_s} \tag{2}$$

Conclusions

An interface structure made by pressing two pieces of vertically oriented CNTs against each other is characterized using diffraction-limited IR microscopy. It is found that the thermal performance of the CNT interface structure is far below expectation. The problem is the large intertube contact resistance between the CNTs, which results in a significant temperature drop at the end of the CNT-CNT contact region when the CNT interface structure is subject to a constant out-of-plane heat flux. Many experimental and modeling efforts are still needed to understand the mechanism of thermal coupling between CNTs.

Acknowledgment

Authors thank Molecular Nanosystems Inc. for providing the CNT samples. This research receives financial support from SRC 2003-NJ-1064 and DARPA N6001-04-1-8916.

Nomenclature

A_e = effective heat transfer area per tube, nm^2
A_t = cross-sectional area of a single tube, nm^2
A_s = area of the sample, cm^2
d = diameter of a nanotube, nm
L = length of the tube overlap region, μm
P = heater power, W
R_t = thermal resistance between CNTs, K m^2/W
T_1 = temperature of CNTs on the hot side, K
T_2 = temperature of CNTs on the cold side, K
w = distance between two adjacent nanotubes, nm

References

[1] Iijima, S., 1991, "Helical Microtubules of Graphitic Carbon," Nature (London), **354**, pp. 56–58.
[2] Hone, J., Whitney, M., Piskoti, C., and Zettl, A., 1999, "Thermal Conductivity of Single-Walled Nanotubes," Phys. Rev. B, **59**(4), pp. R2514–R2516.
[3] Kim, P., Shi, L., Majumdar, A., and McEuen, P. L., 2001, "Thermal Transport Measurements of Individual Multiwall Nanotubes," Phys. Rev. Lett., **87**, pp. 215502 (1-4).
[4] Yu, C., Jiang, W., Hanrath, T., Kim, D., Yao, Z., Korgel, B., Shi, L., Wang, Z. L., Li, D., and Majumdar, A., 2003, "Thermal and Thermoelectric Measurements of Low Dimensional Nanostructures," Proc. ASME Summer Heat Transfer Conference, HT2003-47263, pp. 1–6.
[5] Berber, S., Kwon, Y. K., and Tomanek, D., 2000, "Unusually High Thermal Conductivity of Carbon Nanotubes," Phys. Rev. Lett., **84**(20), pp. 4613–4617.
[6] Choi, S. U. S., Zhang, Z. G., Yu, W., Lockwood, F. E., and Grulke, E. A., 2001, "Anomalous Thermal Conductivity Enhancement in Nanotube Suspensions," Appl. Phys. Lett., **79**(14), pp. 2252–2254.
[7] Biercuk, M. J., Llaguno, M. C., Radosavljevic, M., Hyun, J. K., Johnsond, A. T., and Fischer, J. E., 2002, "Carbon Nanotube Composites for Thermal Management," Appl. Phys. Lett., **80**(15), pp. 2667–2769.
[8] Hu, X., Jiang, L., and Goodson, K. E., 2004, "Thermal Conductance Enhancement of Particle-Filled Thermal Interface Materials Using Carbon Nanotube Inclusions," *Proc. 9th Intersociety Conf on Thermal and Thermo-Mechanical Phenomena in Electronic System*, Vol. 1, pp. 63–69.
[9] Huxtable, S. T., Cahill, D. G., Shenogin, S., Xue, L. P., Ozisik, R., Barone, P., Usrey, M., Strano, M. S., Siddons, G., Shim, M., and Keblinski, P., 2003, "Interfacial Heat Flow in Carbon Nanotube Suspensions," Nat. Mater., **2**(11), pp. 731–734.
[10] Nan, C. W., Liu, G., Lin, Y. H., and Li, M., 2004, "Interface Effect on Thermal Conductivity of Carbon Nanotube Composites," Appl. Phys. Lett., **85**(16), pp. 3549–3551.
[11] Shenogin, S., Xue, L. P., Ozisik, R., Keblinski, P., and Cahill, D. G., 2004, "Role of Thermal Boundary Resistance on The Heat Flow in Carbon-Nanotube Composites," J. Appl. Phys., **95**(12), pp. 8136–8144.
[12] Cao, A., Dickrell, P. L., Sawyer, W. G., Ghasemi-Nejhad, M. N., and Ajayan, P. M., 2005, "Super-Compressible Foamlike Carbon Nanotube Films," Science, **310**, pp. 1307–1310.
[13] Montgometry, S. W., and Holalkere, V. R., 2003, "Carbon Nanotube Thermal Interface Structures," US Patent Publication # US20030117770.
[14] Xu, J., and Fisher, T. S., 2006, "Enhanced Thermal Contact Conductance Using Carbon Nanotube Arrays," IEEE Trans. Compon. Packag. Technol., **29**(2), pp. 261–267.
[15] Ngo, Q. X., Cruden, B. A., Cassell, A. M., Sims, G., Li, J., Meyyappan, M., and Yang, C. Y., 2004, "Nano-Engineered Multiwall Carbon Nanotube-Copper Composite Thermal Interface Material for Efficient Heat Conduction," J. Miner., Met., Mater., **56**(11), pp. 153–154.
[16] Chuang, H. F., Cooper, S. M., Meyyappan, M., and Cruden, B. A., 2004, "Improvement of Thermal Contact Resistance by Carbon Nanotubes and Nanofibers," J. Nanosci. Nanotechnol., **4**(8), pp. 964–967.
[17] Sample, J. L., Rebello, K. J., Saffarian, H., and Osiander, R., 2004, "Carbon Nanotube Coatings for Thermal Control," *Proc. 9th Intersociety Conf on Thermal and Thermomechanical Phenomena in Electronic System*, Las Vegas, NV.
[18] Hu, X. J., Padilla, A. A., Xu, J., Fisher, T. S., and Goodson, K. E., 2006, "3-Omega Measurements of Vertically Oriented Carbon Nanotubes on Silicon," ASME J. Heat Transfer, **128**, pp. 1109–1113.
[19] Tong, T., Zhao, Y., Delzeit, L., Kashani, A., and Majumdar, A., 2004, "Multiwalled Carbon Nanotube/nanofiber Arrays as Conductive and Dry Adhesive Interface Materials," Proc. Integrated Nanosystems:Design, Synthesis & Applications Conf.
[20] Hu, X., Jiang, L., and Goodson, K. E., 2004, "Thermal Characterization of Eutectic Alloy Thermal Interface Materials with Void-like Inclusions," *Proc. 20th IEEE Semiconductor Thermal Measurement and Management Symposium*, San Jose, CA, pp. 98–103.

Thermal Radiative Transport Enhancement via Electromagnetic Surface Modes in Microscale Spherical Regions Bounded by Silicon Carbide

James S. Hammonds, Jr.
Department of Mechanical Engineering,
The City College of New York,
New York, NY 10031
e-mail: hammonds@ccny.cuny.edu

A Green function approach is used with the fluctuation-dissipation theorem to develop a qualitative theoretical model of radiation heat transfer across an evacuated microscale spherical geometry bounded by silicon carbide. The appropriate scalar Green function is presented by employing an impedance boundary condition to describe the electromagnetic spherical interface condition and thus capture the surface modes. This work shows that the spherical boundary can result in spectral conditions for surface mode excitation that depend not only on the dielectric function, but on the sphere radius as well. The surface modes are shown to enhance the radiation significantly and are attributed to surface phonon polariton modes excited at the interface, and surface modes excited by the mechanism of total internal reflection.

[DOI: 10.1115/1.2401203]

Introduction

For energy conversion technologies such as ceramic burners, and other technologies in which thermal gradients and fluxes can significantly impact device operation, microporous materials may be employed to facilitate proper thermal management. In high-temperature applications, radiative transport can be an important mode of heat transfer. Radiation transport across pores is often not considered in radiative analysis of the porous materials however, this work demonstrates that for specific pore diameters and ceramic optical properties, radiative exchange across pores may be important. For microscale pores, thermal transport within the pore and its vicinity can be dominated by near-field transport effects. Microporous SiC is an important thermal management material because of its ability to withstand high temperatures and corrosive environments, and because it has a high mechanical strength, thus, it will be the focus of this work to demonstrate potentially broadly applicable results. Recent work has demonstrated fabrication techniques that will help in realizing the potential of microporous SiC as effective membranes for gas and liquid mixture separation at high temperatures [1], catalyst supports for hydrocarbon fuel reforming [2], and porous ceramic burners [3]. This work examines the conditions that result in the excitation of surface modes inside an evacuated pore bounded by SiC, and shows that the resulting radiation can be changed substantially, requiring analysis that seeks to model thermal management to be adjusted accordingly.

In the technologies noted in the previous paragraph, significant thermal gradients could be present across a porous ceramic such as SiC. When the temperatures are high (close to 1000 K), thermal radiation can dominate heat transfer. In cases wherein thermal radiation dominates, microscale pores are obstructions to the radiative transport because their respective characteristic length scales are of the same order. More specifically, the radiation can be scattered away from its principle propagation direction, and in general this scattering can result in the attenuation of the radiation across the transport domain. The study in this work, however, explores cases of radiation transport in porous materials in which the scattering mechanism actually excites surface modes. To accomplish this, the simplified problem of thermal radiation scattered by an evacuated spherical pore located in a surface active material is studied. It is expected that the results of this work will allow generalizations about problems in which there are many more pores, and cases wherein the pore is filled by other dielectrics.

This paper analyzes transport across an evacuated spherical pore in SiC whose diameter is on the order of micrometers. A temperature gradient in the SiC region drives radiation transfer across the pore. Because the pore diameter is on the order of the radiation wavelength, the radiation must be analyzed using electromagnetic (EM) theory. A previously developed Green function approach is used to describe the EM field, and this approach requires that the radiation be characterized by a source that generates the radiation. This radiation source is approximated as one that produces plane waves. Energy balance analysis that does not take into account wave effects is sufficient when analyzing radiation heat transfer through a medium with a length scale L that is much greater than the characteristic wavelength of the radiation $\lambda_{\max}$ where the radiation is regarded as incoherent, and where $\lambda_{\max}$ is the wavelength of maximum irradiance given by Wien's displacement law [4]. It has been demonstrated that for L on the order of $\lambda_{\max}$, however, small scale spacing effects must be considered [5]. For many macroscale materials with $L \gg \lambda_{\max}$ and with nanoscale pores where the pore diameter $d_{\text{pore}} \leq \lambda_{\max}$, it is often not necessary to consider surface modes, and employment of an energy balance that does not take wave effects into account is sufficient.

However, as will be discussed below, small scale effects such as radiation enhancement and coherence effects can be extended, and thus may require consideration even in domains for which $L \gg \lambda_{\max}$ [6]. The work of Ref. [6] investigates a nonphysical two-dimensional pore, while the work presented in this report studies the impact of a three-dimensional sphere on surface mode excitation and discusses the spectral characteristics of *surface phonon polariton* (SPP) wave modes. As discussed in previous work, SPP mode excitation is possible on flat surfaces for materials with a complex dielectric function, ϵ, for which $\mathrm{Re}(\epsilon) < 0$, and conditions on ϵ specify the spectral location of the SPP mode [7]. For example, for a flat surface active material bounded by a vacuum, the SPP mode occurs when $\mathrm{Re}(\epsilon(\omega)) < -1$, giving the value for ω at which the SPP modes occur. It is also important to note that for the flat surface, the SPP mode is restricted to propagation parallel to the ceramic interface. For the curved boundary, however the SPP mode propagates *away* from the surface. The mechanism associated with surface wave radiation from curved surfaces is discussed in this work. In addition the spherical boundary conditions used in this report demonstrate radiation enhancement by surface modes excited via total internal reflection, which can, in fact, dominate over the SPP modes at certain spherical radii. The specific surface mode dominance is shown to be strongly dependent on radius.

The general theoretical approach that applies the Green function, the fluctuation-dissipation theorem, and an impedance boundary condition (IBC) to the analysis of an evacuated spherical pore bounded by a ceramic material is developed in the next section. Similar approaches have been employed for the analysis of flat geometries [8], however, the work presented here uses the IBC to capture the curved interface conditions. The limits of ac-

---

Contributed by the Heat Transfer Division of ASME for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received January 15, 2006; final manuscript received July 24, 2006. Review conducted by M. Pinar Menguc. Paper presented at the 2005 ASME International Mechanical Engineering Congress (IMECE2005), November 5–11, 2005, Orlando, Florida.

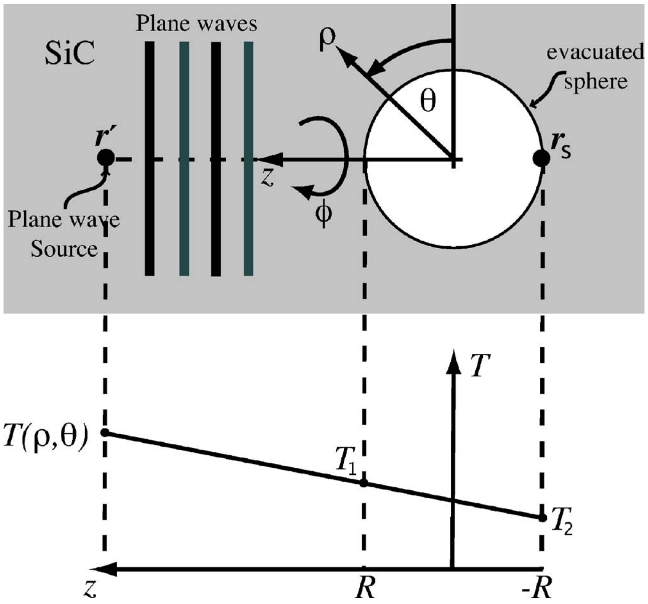


Fig. 1 The evacuated sphere shown in this figure is irradiated by plane waves that originate from the thermal radiation source located at r'. This analysis assumes a linear temperature distribution, as illustrated in the lower part of the figure. For a given temperature of T_2 at r_s ($z=-R$), and T_1 at ($z=R$), Eq. (7) gives the temperature at all other points r'.

curacy of the IBC are also discussed below. The theoretical approach is applied to the specific case of a spherical pore in SiC. Results are then presented that show that the impact of SPP modes on thermal radiation across the pore can be substantial.

Theory

The interest of this work is the component of the heat flux normal to the inner surface of the sphere, at the point $\boldsymbol{r}_s$, as shown in Fig. 1. The radiative flux is the energy carried by electromagnetic (EM) waves. The EM plane waves are modeled as originating from a source located at $\boldsymbol{r}'$, and this analysis examines the impact of the energy radiated from this source on the point $\boldsymbol{r}_s$. Note that $\boldsymbol{r}'$ lies in the ceramic region, outside the evacuated sphere, as shown in Fig. 1. The surface mode of interest in this work will be associated with transverse magnetic (TM) waves, which have an electric field vector that is polarized parallel to the plane of incidence defined by $(\rho,0,\phi)$, or the (ϕ,z) plane of Fig. 1. The area across which heat flows is defined by the sphere inner surface whose normal points in the ρ direction, requiring the calculation of the ρ component of the flux at $\boldsymbol{r}_s$ by

$$Q_{\mathrm{TM},\rho}=\mathrm{Re}\langle E_\theta H_\phi^*\rangle, \qquad (1)$$

where E_θ is the θ component of the electric field and H_ϕ^* is the complex conjugate of the ϕ component of the magnetic field, and where the brackets denote statistical averaging. A spectral form of the flux defined by Eq. (1) is given by Ref. [6] as

$$\frac{\partial Q_{\mathrm{TM},\rho}}{\partial\omega}=\frac{\mu_0}{(2\pi)^{3/2}}\int_{r'}G_{\mathrm{TM}}\frac{\partial G_{\mathrm{TM}}^*}{\partial\rho}\frac{\omega^3\,\hbar\;\mathrm{Im}(\epsilon_1)}{c^2(e^{\hbar\omega/k_{\mathrm{B}}T}-1)}d^3\boldsymbol{r}' \qquad (2)$$

where ϵ_1 is the complex dielectric function of the SiC. The Green function that satisfies Eq. (2) for spherical geometry shown in Fig. 1, is given by [9] as

$$G_{\mathrm{TM}}^E=e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}-\sum_{m=0}^{\infty}\sum_{n=m}^{\infty}\Gamma_n\eta_m i^{n+1}\frac{(2n+1)(n-m)!}{(n+m)!}$$
$$\times\cos m(\phi-\phi')P_n^m(\cos\theta)P_n^m(\cos\theta')h_n^{(1)}(k\rho)h_n^{(1)}(k\rho') \qquad (3)$$

where $\eta_m=1$ for $m=0$ and $\eta_m=2$ for $m\geq1$. In Eq. (3), $h_n^{(1)}$ is the first-order spherical Hankel function, P_n^m is the Legendre polynomial and the nth order reflection coefficient Γ_n is given by

$$\Gamma_n=\frac{j_n'(kR)+iZj_n(kR)}{h_n'^{(1)}(kR)+iZh_n^{(1)}(kR)} \qquad (4)$$

where j_n is the spherical Bessel function, the primes denote the derivative with respect to the argument, and where the relative impedance of SiC is given by

$$Z_1=\sqrt{\frac{1}{\epsilon_1(\omega)}} \qquad (5)$$

As discussed in Ref. [10], the IBC is valid for curved boundaries for which $\mathrm{Im}(\sqrt{\epsilon_1(\omega)})\geq2.3/kR$, where $k=\omega/c$. Additionally the IBC is typically only valid for $Z_1\ll1$, but Ref. [10] demonstrates qualitative accuracy can be obtained for $\epsilon_1(\omega)$ near surface wave resonance frequency. The Legendre polynomial term $P_n^m(\cos\theta')$, and the cosine term $\cos m(\phi-\phi')$ capture the plane wave propagation originating at angles (ϕ',θ'), and the impact of wave propagation on the angle defined by θ. The argument of P_n^m is limited to $[-1,1]$ thus the Legendre polynomial terms are nonsingular, while the zeros of the reflection coefficient Γ_n are associated with the surface modes. Since the surface modes are expected to dominate the final solution, and because the Legendre polynomial and cosine term do not depend on the frequency of the incoming light, they do not contribute any information about the qualitative spectral properties of the dominating surface mode. Therefore, as it is the goal of this work to characterize the spectral properties of thermal radiation exchange that result from surface mode domination, qualitative accuracy is obtained by limiting this analysis to cases wherein the source location is defined by $\theta'=0$ and $\phi'=\phi$, which leaves only the $m=0$ mode and reduces Eq. (3) to [11]

$$G_{\mathrm{TM}}^E=e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}-\sum_{n=0}^{\infty}\Gamma_n(2n+1)i^{n+1}P_n(\cos\theta)h_n^{(1)}(k\rho)h_n^{(1)}(k\rho') \qquad (6)$$

where the dependencies on θ, ρ, and ρ' are maintained. The first term on the right-hand side of Eq. (6) can be interpreted as the incident plane wave, while the second term is the scattered field. The normalized spectral flux can now be calculated for a material with a dielectric function and temperature T, defined in the next section.

Results

The theoretical analysis developed in the previous section will now be applied for the case of SiC as the bounding material. Investigations of thin insulating ceramic materials used in high-temperature environments have shown nearly linear temperature variations over microscale distances in applications such as thermal barrier coatings for turbine blades [12]. Therefore a linear temperature distribution is used to demonstrate near-field effects as shown in Fig. 1. The temperature variation results in a temperature gradient T that drives heat exchange across the evacuated space, which is described by

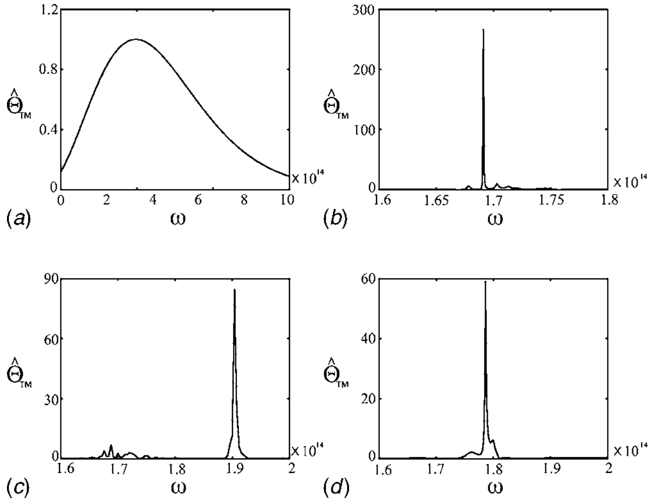


Fig. 2 The normalized spectral flux, $\hat{\Theta}_{\mathrm{TM}}$, at point r_s of the spectral region shown in Fig. 1 for sphere radii (*a*) $R=1$ mm, (*b*) $R=90$ μm, (*c*) $R=100$ μm, and (*d*) $R=10$ μm. The small sphere results are characterized by spectral selectivity and a significantly enhanced intensity. These characteristics are attributed to surface mode excitation. The spectral location of the peaks shift with radius, due to radial dependence of the surface wave dispersion relation.

$$T(\rho,\theta) = \frac{1}{2}\left(\frac{T_1 - T_2}{R}\rho\cos\theta + T_2 + T_1\right) \tag{7}$$

where T_1 and T_2 are a high and low temperature, respectively, and which specifies the temperature difference across the pore. The high temperature T_1 is at the point $(R,0,\phi')$, while T_2 is located at point (R,π,ϕ'), as shown in Fig. 1. Evaluation of the flux using Eqs. (2) and (6) would only give the flux from $\boldsymbol{r}'$ *to* $\boldsymbol{r}_s$. To obtain the net flux, it is necessary to include the energy radiated away from $\boldsymbol{r}_s$. A normalized version of an appropriate equation is given as

$$\hat{\Theta}_{\mathrm{TM}} = \int_{r'} G_{\mathrm{TM}}\frac{\partial G^*_{\mathrm{TM}}}{\partial\rho}\frac{\omega^3}{\bar{\omega}^3}\left(\frac{1}{e^{\hbar\omega/k_{\mathrm{B}}T(\rho,\theta)}-1} - \frac{1}{e^{\hbar\omega/k_{\mathrm{B}}T_2}-1}\right)d^3\boldsymbol{r}' \tag{8}$$

which is defined via Eq. (2) with $\hat{\Theta}_{\mathrm{TM}}=[\partial Q_{\mathrm{TM},\rho}/\partial\omega]W^{-1}$, where

$$W = \frac{\mu_0}{(2\pi)^{3/2}}\frac{\bar{\omega}^3}{c^2}\hbar\ \mathrm{Im}(\epsilon_1) \tag{9}$$

where $\bar{\omega}$ is a factor arbitrarily chosen to allow the maximum normal spectral emission to be unity, while maintaining its spectral characteristics. In Eq. (8), $T(\rho,\theta)$ is the temperature at $\boldsymbol{r}'$ and T_2 is the temperature at $\boldsymbol{r}_s$. The net flux can now be calculated from Eq. (8) using the dielectric properties of SiC

$$\epsilon_1(\omega) = \epsilon_\infty\frac{(\omega_L^2 - \omega^2 - i\gamma\omega)}{(\omega_T^2 - \omega^2 - i\gamma\omega)} \tag{10}$$

where the optical dielectric constant $\epsilon_\infty=6.7$, the longitudinal optical frequency $\omega_L=18.3\times10^{13}$ s^{-1}, the transverse optical frequency $\omega_T=15.0\times10^{13}$ s^{-1}, and the damping constant $\gamma=1.0\times10^{12}$ s^{-1} [13], with $T_1=1000$ K and $T_2=970$ K. The results of this analysis are shown in Fig. 2. To establish relevant length scales it is convenient to refer to a blackbody at the equilibrium temperature T_1. Wien's displacement law gives the wavelength corresponding to the maximum amount of thermally emitted energy as $\lambda_{\max}=3$ μm. Figure 2 shows the normalized spectral flux, $\hat{\Theta}_{\mathrm{TM}}$, for a large sphere, $R=1$ mm$\gg\lambda_{\max}$ and for mesoscale spheres where $R=90$ μm and $R=100$ $\mu\mathrm{m}>\lambda_{\max}$. Figure 2 also shows results for the microscale case $R=1$ μm.

The large sphere solution was chosen to demonstrate that the SPP modes are unimportant in large geometries. It should be cautioned that for cases in which $R<\lambda_{\max}$, the limits of validity of the IBC render the analysis inaccurate for the very near field; it is also expected that for $R<\lambda_{\max}$, the surface modes under investigation in this report become less important, as quasi-static modes begin to dominate the heat flux [14]. The quasi-static regime is characterized by transport dominated by other evanescent modes, which result in a substantial increase in heat flux as the radius is reduced, as reported for plane surfaces in close proximity reported by other authors [5,8,15–17]. Furthermore, for radii smaller than $\lambda_{\max}$, it is expected that the temperature gradient, and thus the thermal radiative exchange, cannot be maintained. The results of Fig. 2(*a*) gives the results for the large sphere, in which a relatively broadband flux is shown and where the radiation is regarded as incoherent. The mesoscale pore case of Fig. 2(*b*), however, is characterized by spectral peaks and an intensity that is enhanced substantially. The surface modes associated with the peaks of Fig. 2 will be discussed below.

Discussion

An important result of this work is shown as Fig. 2. The peaks of the curve in Figs. 2(*b*)–2(*d*) correspond to those spectral locations where the denominator of the reflection coefficient, given by Eq. (4), equals zero. Thus, the denominator expression set to zero gives the surface wave dispersion curves. The tallest peaks of Figs. 2(*b*) and 2(*d*) occur at frequencies for which $\mathrm{Re}(\epsilon_1)<-1$ and thus give the SPP mode [7]. Early work has shown that because of the curvature associated with the spherical regions, the SPP modes shown in Figs. 2(*b*) and 2(*d*) are radiative [18]. Thus, the SPP mode may be considered an additional transport mechanism in porous materials with micro- to mesoscale pores. For flat surfaces that support surface modes, the SPP modes are restricted to the interface and cannot participate in energy exchange unless they are *coupled* by an additional transport mechanism. An example of SPP coupling at flat surfaces employs surface microstructure that scatter SPP modes into modes that radiate energy away from the surface [19]. For a qualitative physical explanation of SPP radiation at curved surfaces, the radiating mechanism may be thought of as being analogous to the mechanism of radiation loss at a curved wave guide. The wavefront velocity of the surface wave at the spherical interface might be given by $v_s=\Delta s_s/\Delta t$, where Δs_s gives a small distance along the spherical surface that the wavefront travels over a time period Δt. While the SPP mode does decay in the ρ direction, the field extends out to a distance comparable to $\rho\le\lambda_{\max}$. Thus, for increasing ρ, the portion of the surface wave that extends away from the spherical surface will have a different wavefront velocity given by $v=\Delta s/\Delta t$ where $\Delta s>\Delta s_s$ for the same Δt, and thus $v>v_s$. The physical limit on v is the speed of light inside the medium (SiC in this case). For increasing ρ, when v exceeds this physical limit, the surface wave must compensate by radiating energy away in the ρ direction [20]. Thermal radiation with a frequency range that includes the frequency band $\omega_T<\omega<\omega_L$, for which $\mathrm{Re}(\epsilon_1)<-1$, can strongly couple to crystal lattice vibrations that propagate parallel to the irradiated surface. Due to charge displacements associated with lattice vibrations, an electromagnetic field is established at the crystal surface. This coupled lattice wave and EM wave at the interface is the SPP.

The tallest peak of Fig. 2(*c*) occurs when the sphere radius is changed slightly from that used for Fig. 2(*b*), and does not fall within the frequency range $\omega_T<\omega<\omega_L$; thus, it is not an SPP mode. The peak apparently dominates over the SPP modes and occurs because of a zero in the denominator of Γ_n made possible by negative values of the Hankel function $h_n^{(1)}$. The order n of $h_n^{(1)}$ can be thought of as representing particular directions (and phases) of a plane wave incident on the sphere. Hence the peak of Fig. 2(*c*) can be said to result from those plane waves oriented such that they are totally and internally reflected at the spherical surface exciting a surface mode. Figures 2(*b*)–2(*d*) also show that the spectral location of the tallest peaks shift with radius. This is evident by the dispersion relation for surface modes given by the zeros of the denominator of Γ_n, which is not only dependent on the dielectric properties as given by Z, but also on the radius.

The work of Ref. [19] shows that the radiative SPP mode can have a much longer coherence length than that typically associated with thermal radiation. This long coherence length could effect heat transport dramatically. The propagation length and relatively long coherence length means that the potential for EM wave interference, and other confinement effects, may be important over length scales much larger than the characteristic wavelength, $\lambda_{\max}$. Thus, when the radiative transport dominates and the pore size, temperature distribution, and dielectric properties are such that SPP modes are important, coherence effects could significantly effect radiation heat transfer. This result can have important implications for high temperature applications, in which porous ceramics are used, where radiation can be an important heat transfer mechanism.

Conclusion

Results of this report show that for surface active porous materials containing meso- to microscale pores in which radiation dominates, transport via surface phonon polariton modes may be an important thermal exchange mechanism. Furthermore the spherical boundary may allow for radiation enhancement by surface mode excitation outside the $\omega_L < \omega < \omega_T$ range via total internal reflection, and spectral shifting of the enhancement peak. Surface phonon polariton modes at curved surfaces are radiating modes characterized by significantly enhanced flux, and extended coherence lengths that may cause effects such as interference to become important and affect radiation transport dramatically. While the theoretical approach utilized in this work has been applied by several researchers to qualitatively describe radiative transport in very small regimes, more research is required to asses the qualitative accuracy of these approaches.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. CTS-0421661.

References

[1] Suwanmethanond, V., Goo, E., Liu, P., Sahimi, G., and Tsotsis, T., 2000, "Porous Silicon Carbide Sintered Substrates for High-Temperature Membranes," Ind. Eng. Chem. Res., **39**, pp. 3264–3271.

[2] Sung, I.-K., Christian, Mitchell, M., Kim, D.-P., and Kenis, P., 2005, "Tailored Macroporous SiCN and SiC Structures for High-Temperature Fuel Reforming," Adv. Funct. Mater., **15**, pp. 1336–1342.

[3] Wang, H., Zheng, S.-Y., Li, X.-D., and Kim, D.-P., 2005, "Preparation of Three-Dimensional Ordered Macroporous SiCN Ceramic Using Sacrificing Template Method," Microporous Mesoporous Mater., **80**, 357–362, and references therein.

[4] Siegel, R., and Howell, J., 1992, *Thermal Radiation Heat Transfer*, 3rd ed., Hemisphere, Washington, DC.

[5] Cravalho, E., Tien, C., and Caren, R., 1967, "Effect of Small Spacings on Radiative Transfer Between Two Dielectrics," J. Heat Transfer, **89C**, pp. 351–357.

[6] Hammonds, J., 2006, "Thermal Transport via Surface Phonon Polaritons Across a Two-Dimensional Pore," Appl. Phys. Lett., **88**(4), p. 041912.

[7] Raether, H., 1988, *Surface Plasmons*, Springer-Verlag, Berlin.

[8] Polder, D., and Hove, M. V., 1971, "Theory of Radiative Heat Transfer Between Closely Spaced Bodies," Phys. Rev. B, **4**(5), pp. 3303–3314.

[9] Felsen, L., and Marcuvitz, N., 1973, *Radiation and Scattering of Waves*, Prentice-Hall, Englewood Cliffs, NJ.

[10] Wang, D.-S., 1987, "Limits and Validity of the Impedance Boundary Condition on Penetrable Surfaces," IEEE Trans. Antennas Propag., **AP-35**(4), pp. 453–457.

[11] Lax, M., and Feshbach, H., 1948, "Absorption and Scattering for Impedance Boundary Conditions of Spheres and Circular Cylinders," J. Acoust. Soc. Am., **20**(2), pp. 108–124.

[12] Padture, N., Gell, M., and Jordan, E., 2002, "Thermal Barrier Coatings for Gas-Turbine Engine Applications," Science, **296**(5566), pp. 280–284.

[13] Shchegrov, A., Joulain, K., Carminati, R., and Greffet, J.-J., 2000, "Near-Field Spectral Effects Due to Electromagnetic Surface Excitations," Phys. Rev. Lett., **85**(7), pp. 1548–1551.

[14] Henkel, C., Joulain, K., Carminati, R., and Greffet, J.-J., 2000, "Spatial Coherence of Thermal Near Fields," Opt. Commun., **186**, pp. 57–67.

[15] Caren, R., 1974, "Thermal Radiation Between Closely Spaced Metal Surfaces at Low Temperature Due to Traveling and Quasi-Stationary Components of the Radiation Field," Int. J. Heat Mass Transfer, **17**(7), pp. 755–765.

[16] Xu, J.-B., Läuger, K., Möller, R., Dransfeld, K., and Wilson, I., 1994, "Heat Transfer Between Two Metallic Surfaces at Small Distances," J. Appl. Phys., **76**(11), pp. 7209–7216.

[17] Park, C., Haus, H., and Weinberg, M., 2002, "Proximity-Enhanced Thermal Radiation," J. Phys. D, **35**, pp. 2857–2863.

[18] Pfeiffer, C., Economou, E., and Ngai, K., 1974, "Surface Polaritons in a Circularly Cylindrical Interface: Surface Plasmons," Phys. Rev. B, **10**(8), pp. 3038–3051.

[19] Greffet, J.-J., Carminati, R., Joulain, K., Mulet, J.-P., Mainguy, S., and Chen, Y., 2002, "Coherent Emission of Light by Thermal Sources," Nature (London), **416**, pp. 61–64.

[20] Marcuse, D., 1982, *Light Transmission Optics*, 2nd ed., Van Nostrand Reinhold, New York.

Journal of
Heat Transfer

Erratum

Erratum: “Assessment of Overall Cooling Performance in Thermal Design of Electronics Based on Thermodynamics” [Journal of Heat Transfer, 2001, 123(5), pp. 999–1005]

Ken Ogiso

In the description of Experimental Data on p. 1001, the unit for “the total heat transfer area” should be “m^2.”