**Time resolved imaging of carrier and thermal transport in silicon**

D. H. Hurley,1,a) O. B. Wright,2 O. Matsuda,2 and S. L. Shinde3

1Materials Science and Engineering, Idaho National Laboratory (INL), P.O. Box 1625, Idaho Falls, Idaho 83415-2209, USA
2Division of Applied Physics, Graduate School of Engineering, Hokkaido University, Sapporo 060-8628, Japan
3Sandia National Laboratories (SNL), New Mexico, P.O. Box 5800, Albuquerque, New Mexico 87185, USA

(Received 3 November 2009; accepted 12 November 2009; published online 27 January 2010)

We use ultrashort optical pulses to microscopically image carrier and thermal diffusion in two spatial dimensions in pristine and mechanically polished surfaces of crystalline silicon. By decomposing changes in reflectivity in the latter sample into a transient component that varies with delay time and a steady-state component that varies with pump chopping frequency, the influence of thermal diffusion is isolated from that of carrier diffusion and recombination. Additionally, studies using carrier injection density as a parameter are used to clearly identify the carrier recombination pathway. © 2010 American Institute of Physics. [doi:10.1063/1.3272827]

I. INTRODUCTION

Great potential exists for the use of nanostructured semiconductor materials in energy conversion systems, including advanced superlattice-based photovoltaics1 and nanoscale thermoelectric materials.2 In spite of this, there is still considerable debate regarding the detailed mechanisms governing thermal transport in these systems.3 Continued progress in this field requires the development of spatially and temporally resolved metrology tools with well-characterized experimental parameters.

Laser-based methods have emerged as a leading candidate for making precise thermal transport measurements due to their noncontact nature and well-defined optical coupling conditions. Laser excitation can be implemented in the frequency domain using amplitude modulated continuous wave (cw) laser heating4–6 or in the time domain using pulsed laser heating.7,8 The probe beam either reflects off the material surface, sampling the heated region (through thermoreflection), or propagates along the surface, sampling the heated gas just above the surface (through the mirage effect).9 For imaging applications, the highest lateral spatial resolution is obtained by scanning a probe beam spot at normal incidence. The spatial resolution depends on the thermal wavelength or the diffusion length, and is ultimately limited by the laser spot size.

For thermal transport measurements on semiconductors, a metallic transducer layer is typically used to avoid complicated data analysis owing to the presence of a diffusing photoexcited electron-hole plasma.10,11 However, the transducer layer introduces new experimental parameters. It is thus advantageous to use measurement schemes involving uncoated semiconductor materials. Such experiments require relating the changes in optical properties to the coevolution of the plasma and the temperature fields.

For pulsed-laser measurements, only transport properties in the depth direction are monitored when the pump optical spot size is much larger than the plasma or thermal diffusion lengths.12–14 In such an example closely related to the present study, Tanaka et al.15 have imaged plasma and thermal diffusion in an uncoated ion-irradiated crystalline Si sample. By changing the delay between the pump and probe pulses, they measured the relative contributions of the plasma and temperature fields to the change in probe reflectivity. However, they did not probe times greater than the carrier recombination time.

Alternatively, for pulsed-laser measurements, one can also monitor transport properties in the lateral direction when the pump optical spot size is on the order of or much smaller than the plasma or thermal diffusion lengths. For short laser pulse durations, τp<1 ns, this approach requires near-field optical focusing.16 Similarly, for chopped cw laser excitation, the chopping frequency can be varied to determine the relative contributions of the plasma and temperature fields;10 however, unlike pulsed-laser studies, the thermal and plasma waves cannot be easily decoupled.

Recently, a new approach7,17,18 that uses elements of both time and frequency domain methods has emerged. This hybrid approach uses a time-delayed probe pulse to laterally image transient as well as steady-state contributions to the change in the probe reflectivity. In this paper, we use this method to image plasma and thermal diffusion in a crystalline silicon sample damaged by mechanical polishing, thereby enabling lateral thermal transport to be completely decoupled from that of the plasma. For probe delays that are large in comparison to the carrier recombination time, it is shown that the optical reflectivity changes arise solely from thermal diffusion. In this way, the in-plane thermal properties of the substrate can be accurately probed.

II. EXPERIMENTAL SETUP

The setup used for imaging changes in reflectivity is shown at the top of Fig. 1. A pump pulse is used to excite charge carriers (electrons) into the conduction band of a Si sample. The photoexcited carriers quickly relax to the conduction band edge, giving up their excess kinetic energy to
the lattice in the form of heat. The carrier plasma continues to expand under the influence of density-gradient-driven diffusion until the carriers recombine. Because Si is an indirect-bandgap semiconductor, carrier recombination provides a second heating source. A probe pulse is then used to monitor sample reflectivity changes caused by both the expanding carrier plasma and the diffusion of heat.

The imaging system is similar to the setup presented by Tachizaki et al., and involves translation at near-normal optical incidence of a microscopic probe spot relative to a fixed pump spot. A single probe pulse monitors changes in probe reflectivity. The probe beam is sent through a confocal lens pair (with focal length of 100 mm) and is focused onto the sample using a microscope objective. The first lens in the confocal system is attached to a stage that allows the lens and the probe beam to be translated in the $x$-$y$ plane. The probe beam propagation vector remains collinear with the lens axis. The second lens converts the $x$-$y$ motion of the beam into a change in entrance angle into the objective. The pump beam is guided along the optical axis of the objective using a dichroic beam splitter placed after the lens pair. The pump and probe beams, with wavelengths of 400 and 800 nm and incident pulse energies of 0.026 and 0.013 nJ, respectively, are derived from a Ti:sapphire laser with a pulse duration of $\sim$200 fs. The respective spot diameters of the pump and probe beams at the sample are 4.6 and 4.0 $\mu$m (at full width at half maximum intensity). The probe energy is half that of the pump, but the interrogation volume (defined by the spot diameter and the skin depth) of the probe is approximately two orders of magnitude larger than that of the pump. Thus, in our case, the probe can be considered to passively interrogate that sample.

FIG. 1. (Color) Top: experimental setup. The dashed box represents a dual stage system that allows the lens and the probe beam to be translated in the $x$-$y$ plane. A single objective is used for focusing light onto the sample. Bottom: Images of the amplitude and phase of the probe reflectivity change as a function of delay time, using a chopping frequency $f_0=50$ kHz. At early delay times, the transient component that varies with delay time is dominant. At large delay times, for the mechanically polished surface, the steady-state component is dominant.
The pump beam was chopped at a frequency in the range \(f_0=10–50\) kHz to enable measurements of diffusion processes at these frequencies. The maximum transient temperature rise caused by a single pump pulse is 5 K, \(^\text{20}\) and the steady-state temperature rise is \(\sim 0.5\) K. \(^\text{2}\) Typical values of the relative change in reflectivity of the probe beam \((\Delta R/R_0)\) ranged from \(\sim 10^{-4}\) to \(10^{-3}\).

The two samples studied are cleaved from a single boron-doped \((p\text{-type})\) \((100)\) Si wafer with a doping concentration of \(2 \times 10^{16} \text{ cm}^{-3}\), with two different surface preparations. One sample, referred to as the mechanically polished sample, is obtained by polishing the as-received \((100)\) surface using a \((3 \mu \text{m})\) diamond slurry and then \((0.05 \mu \text{m})\) colloidal silica. \(^\text{21}\) This process significantly damages the crystal lattice in the near-surface region \((\sim 1 \mu \text{m})\). \(^\text{22}\) This sample is then annealed at \(300\) °C to relieve the residual stress caused by polishing. The second sample is in the as-received state. In this case, the vendor used a dilute HF etch to remove any residual surface damage. In all cases investigated, the crystal [011] axis is aligned with the laboratory \(x\)-axis.

### III. RESULTS AND DISCUSSION

The series of images in Fig. 1 represent the modulus \(|\Delta R(t)|\) (left column) and phase lead \(\Delta \Phi(t)\) (right column) of the reflectivity variation measured by the lock-in amplifier at a chopping frequency \(f_0=50\) kHz at two delay times: \(t =2\) ps and \(t=1\) ns. We first consider the results for the mechanically polished surface. We decompose the signal into two components. The first is a transient component that depends on the delay time, and depends on a summation of the effects of thermal and plasma waves at harmonics of the laser pulse repetition frequency \((76\) MHz) \(^\text{23,24}\). The second is a steady-state component associated with a temperature oscillation at the pump chopping frequency. The relative magnitude of these components depends on \(f_0\) and \(t\). For these measurements at \(f_0=50\) kHz, the transient component is dominant at relatively short times on the order of a few picoseconds: at \(t=2\) ps, there is little lateral spatial variation in the phase over the excitation region, and the primary gradients in temperature and carrier density are in a direction perpendicular to the surface (over the pump optical skin depth of \(82\) nm). Thus, at \(t=2\) ps, \(\Delta \Phi(t)\) should primarily be influenced by such perpendicular gradients. \(\Delta R(t)\) is an averaged response over the probe skin depth \((\sim 10 \mu \text{m})\), which is considerably larger than the pump skin depth \((82\) nm) \((\text{Ref. 25})\) and the plasma and thermal diffusion lengths at 2 ps \((28\) and \(13\) nm, respectively). \(^\text{26}\)

The images of \(|\Delta R(t)|\) and \(\Delta \Phi(t)\) for the mechanically polished surface at \(t=1\) ns are markedly different. Unlike the behavior at \(t=2\) ps, \(\Delta \Phi\) decreases steadily with increasing distance from the source. Moreover, the profiles of \(|\Delta R(t)|\) and \(\Delta \Phi(t)\) no longer vary with increasing delay time, instead showing a steady-state behavior. The rapid decrease in the transient component is due to the short \((<1\) ns) recombination time, \(\tau_r\), associated with surface trap sites. \(^\text{27}\) After \(t =1\) ns, the transient component vanishes, leaving only the steady-state component.

For the as-received sample measured under the same experimental conditions, \(|\Delta R|\) at \(t=2\) ps (not shown here) is essentially the same as that for the mechanically polished surface. However, \(|\Delta R|\) and \(\Delta \Phi\) for \(t=1\) ns, as shown in Fig. 1, are qualitatively different from that of the mechanically polished surface: the transient component remains dominant, as evidenced by the flat \(\Delta \Phi\) spatial profile. This is understandable given that the plasma and thermal diffusion lengths at 1 ns, \(\sim 700\), and \(\sim 300\) nm, respectively, are considerably smaller than the laser spot sizes for this sample. Compared to the mechanically polished sample, surface passivation by etching results in a much slower surface recombination velocity and a persistence of the transient component. \(^\text{22}\)

Further analysis of the results for the mechanically polished sample is beneficial because, for delay times greater than the carrier recombination time, the experimental results can be analyzed using a simplified thermal wave model in three spatial dimensions (3D). This model, which has been discussed in detail in a previous paper, \(^\text{28}\) considers thermal waves produced by a sinusoidal heat input. Because of the relatively low chopping frequencies here, we do not consider the effect of thermal waves at frequencies corresponding to harmonics of the laser pulse repetition frequency. \(^\text{23,24}\) The volumetric thermal wave source associated with the chopped pump beam is taken to have a Gaussian profile in the lateral direction determined by the cross-correlation of the pump and probe spot sizes, \(^\text{29}\) and a decaying exponential profile in the depth direction (over the pump beam optical skin depth of \(82\) nm). While it is assumed that the damaged layer will have an appreciably smaller thermal conductivity than the pristine substrate, for the pump chopping frequencies considered it can be shown that the perturbation caused by the damaged layer on the thermal wave phase profile is relatively small. \(^\text{30}\) We will thus assume that the sample is homogeneous. The literature value of the optical skin depth for crystalline Si at the pump wavelength is used. \(^\text{31}\) Figure 2(a) shows the experimental profiles at \(t=1.5\) ns, at two different chopping frequencies \(f_0=10\) and \(50\) kHz, compared with \(\Delta \Phi(x)\) obtained from the thermal wave model, where \(x\) is the lateral scan distance. The thermal diffusivity used for the model results presented in Fig. 2(a) is obtained by fitting the predicted and experimentally measured thermal wave phase profiles \(\Delta \Phi(x)\). The best fit value for the thermal diffusivity \(D=(8.8 \pm 0.2) \times 10^{-5} \text{ m}^2\text{s}^{-1}\) is very close to the literature value \(D=8.71 \times 10^{-5} \text{ m}^2\text{s}^{-1}\) for pure crystalline Si. \(^\text{26,31,32}\) The good correspondence between experiment and simulation at \(f_0=10\) kHz (corresponding to a thermal diffusion length \(l=53\) \(\mu\text{m}\)) and \(50\) kHz \((l=24\) \(\mu\text{m}\)) confirms that after \(t=1.5\) ns the electrons have recombined and that the steady-state component is driven solely by thermal diffusion. Our value for \(D\) illustrates that the surface damaged layer does not significantly alter \(\Delta \Phi(x)\).

To better understand the plasma dynamics for this mechanically polished sample, we examine an intermediate delay time: the right-hand plot in Fig. 2(b) shows \(|\Delta R|\) for \(f_0 =50\) kHz and \(t=200\) ps, at which delay the magnitude of the transient and steady-state components are comparable. The reflectivity change associated with the steady-state component is proportional to the surface temperature oscillation,
of amplitude $\Delta \tilde{N}_{tr}(x,t)$, and to the surface carrier density oscillation, of amplitude $\Delta \tilde{N}_{ss}(x)$, of the thermal and plasma waves, respectively, at the chopping frequency. Because the carrier recombination time (approximated here to be less than 1 ns) is much smaller than the pump chopping period, $\Delta \tilde{N}_{ss}(x)$ can be neglected. The reflectivity change associated with the transient component is proportional to the transient change in surface temperature, $\Delta \tilde{T}_{tr}(x,t)$, and to the transient change in surface carrier density, $\Delta \tilde{N}_{tr}(x,t)$. For early times before a significant fraction of the carriers have recombined, the transient component is dominated by the diffusing plasma. A simple model that reproduces $|\Delta R(x)|$ in the mechanically polished sample for a specific delay time is given by

$$\frac{|\Delta R(x)|}{R} = |A(t)[\Delta \tilde{N}_{tr}(x,t)e^{i\theta_{tr}}] + B[\Delta \tilde{N}_{ss}(x)e^{i\theta_{ss}}]|,$$

where $A(t)$ and $B$ are, respectively, the relative contributions of the plasma component and steady-state component, and $\theta_{ss}(x)$ and $\theta_{tr}$ are the respective phases of the thermal wave and transient component. $\Delta \tilde{T}_{ss}(x)$ and $\Delta \tilde{N}_{ss}(x)$ are taken from the 3D thermal wave simulation described above. The coefficient $B$ is not a function of time, and is determined from the measurements of $|\Delta R|$ for $f_0 = 50$ kHz and $t = 1$ ns, a time for which $|\Delta R|$ is solely associated with the thermal wave. $\Delta \tilde{N}_{tr}(x,t)$ is assumed to mimic the Gaussian spatial profile of the autocorrelation of the pump and probe pulses at $t = 2$ ps. The flat phase profile of the transient component is determined by the constant value of $\theta_{tr}$ equal to the difference between $\Delta \Phi$ at $t = 1$ ns (dominated by the thermal wave) and $\Delta \Phi$ at $t = 2$ ps (dominated by the plasma wave) at $x = 0$ (the spot center). This phase difference, illustrated in Fig. 1, is $\sim 180^\circ$, indicating that these two components have opposite signs. This model, which neglects the diffusing nature of the plasma (i.e., it replaces the plasma wave with a static amplitude profile and phase), is reasonable since the plasma diffusion length (700 nm) for a 1 ns recombination time is only a small fraction of the pump and probe spot sizes. The reflectance variations depending on $\Delta \tilde{T}_{ss}(x)$ and $\Delta \tilde{N}_{tr}(x,t) = 2$ ps, labeled thermal wave and Gaussian, respectively, are shown on the left-hand side of Fig. 2(b). Equation (1) can be fitted to $|\Delta R(x)|$ for any given delay using $A/B$ as a fitting parameter. The result of this procedure for $t = 200$ ps is shown on the right-hand side in Fig. 2(b). In Fig. 2(c), we plot $A/B$ versus $ln(t)$.

FIG. 2. (Color) (a) Experimental phase lead profile $\Delta \Phi$ (dots) measured at $t = 1.5$ ns and a 3D simulation based on thermal waves (solid lines) (b) Right: experimental $|\Delta R|$ (dots) and corresponding theoretical profile according to Eq. (1) (solid line). Left: theoretical profile for $|\Delta R|$ arising from the thermal waves. The Gaussian represents the cross correlation of the pump and probe pulses. (c) Amplitude ratio $A/B$ vs delay time. (d) Amplitude ratio $A/B$ vs plasma injection density for a delay time of 250 ps. The dashed lines in (c) and (d) are guides to the eye. The chopping frequency corresponding to (b)–(d) is $f_0 = 50$ kHz.

The vanishing of the transient component at long delay times further illustrates that the photoexcited carriers in the mechanically polished sample have recombined after 1 ns. Carrier-density-dependent Auger recombination can be neglected because the recombination time connected with this pathway is estimated to be in excess of 20 ns. Rather, our data are consistent with fast recombination associated with the Shockley–Read–Hall (SRH) recombination pathway. SRH recombination sites arise from lattice defects in highly damaged surface layers. In our mechanically polished sample, this layer extends to approximately 1 $\mu$m below the surface, which is larger than the plasma diffusion length of 700 nm at 1 ns. The diffusing photoexcited carriers therefore cannot escape from the damaged layer before re-
combination. Although not investigated here, the temporal dependence of the coefficient $A$ and hence $A/B$ is indicative of the coevolution of carrier and temperature distributions mediated by electron-phonon coupling.

The experimentally derived functional dependence of $A/B$ on the average injected carrier density $N_0$ sampled by the probe spot is shown in Fig. 2(d) for a delay of 250 ps. $A/B$ increases linearly with increasing fluence and hence with increasing injection density $N_0$. Previous studies on polycrystalline Si have shown that SRH recombination centers saturate at high $N_0$. This behavior is consistent with the observed increase in the transient component with increasing $N_0$. Moreover, the screening of SRH sites slows recombination, which causes $N_0$ to decrease more slowly with time. Because the transient component is governed by $N_0$, this in turn leads to the observed persistence of the transient component.

IV. CONCLUSION

In conclusion, we use ultrashort optical pulses to microscopically image carrier and thermal diffusion in two spatial dimensions in pristine and mechanically polished surfaces of crystalline silicon. By decomposing changes in optical reflectivity into a transient component that varies with delay time and a steady-state oscillating component at the pump chopping frequency, the respective influences of carrier recombination and thermal diffusion are identified. We demonstrate that for probe delays that are large in comparison to the carrier recombination time in the mechanically polished sample, thermal and plasma diffusion can be completely decoupled. By means of a 3D heat diffusion model for the oscillating temperature at the pump chopping frequency, we derive the thermal diffusivity of the polished sample. Furthermore, studies as a function of carrier injection density are used to confirm the recombination pathway in this sample. This study provides a useful starting point for the further investigation of the 3D diffusion of thermal and plasma waves in a variety of treated semiconductor surfaces.

ACKNOWLEDGMENTS

This work was supported by an INL and SNL LDRD under DOE Contract No. DE-AC07-05ID14517.

21. The colloidal silica contains dilute sodium peroxide. Thus, there is some chemical polising in addition to mechanical polishing.
25. The index of refraction used in this calculation is for crystalline Si and is taken from Handbook of Optical Constants of Solids, edited by E. Palik (Elsevier, Amsterdam, 1998); it has been observed that the index of refraction for a mechanically polished sample can differ from that of a passivated surface, M. Ohldal, I. Ohldal, and F. Lakes, Surf. Sci. 55, 467 (1976).
26. The plasma and thermal diffusion lengths are given by $L = (D t)^{1/2}$, where $D$ is either the ambipolar diffusion coefficient or the thermal diffusivity, and $t$ is the pump-probe delay time. The literature value of the carrier-density-dependent ambipolar diffusion coefficient is taken from C.-M. Li, T. Sjodin, and H.-L. Dai, Phys. Rev. B 56, 15252 (1997); the value of the thermal diffusivity is taken from H. R. Shanks, Phys. Rev. 130, 1743 (1963).
30. The error incurred by assuming the sample to be homogeneous can be estimated using a simple expression for the thermal wave dispersion taken from A. A. Maznev, J. Hartmann, and M. Reichling, J. Appl. Phys. 78, 5266 (1995). For a $1\mu m$ damaged layer with a thermal conductivity that is one half that of single crystal Si, the difference in the thermal wave phase over a $20\mu m$ scan distance is 0.02°.
31. The best fit was obtained by minimizing the standard deviation. The uncertainty corresponds to a value of $D$ that has an associated standard deviation that is 30% larger than the minimum.
33. The Auger recombination time is given by $\tau_A = 1/CN$, where $N$ is the carrier density and the Auger recombination coefficient ($C = 4 \times 10^{21} cm^2 s^{-1}$) is taken from V. E. Gusev and A. A. Karabutov, Laser Optoacoustics (AIP, New York, 1993).
37. The average injected carrier density as measured by the probe spot is taken as the maximum carrier density at the centre of the pump spot reduced by a factor of $1/(1+w_{probe}/w_{pump})$, where $w_{pump}$ and $w_{probe}$ are the respective spot diameters of the pump and probe beams. 
38. $B$ was assumed to be proportional to the pump fluence.