PHYSICAL REVIEW B **91**, 045201 (2015)

Ultrafast scattering processes of hot electrons in InSb studied by time- and angle-resolved photoemission spectroscopy

H. Tanimura, J. Kanasaki, and K. Tanimura

The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

(Received 9 December 2014; revised manuscript received 20 December 2014; published 12 January 2015)

Ultrafast scattering processes of hot electrons photoinjected into the conduction band of InSb have been studied using time- and angle-resolved photoemission spectroscopy. The nascent distributions of hot-electron packets are captured directly in energy and momentum spaces, and their ultrafast scattering processes are traced at femtosecond temporal resolution on a state-resolved basis. Hot electrons injected in the Γ valley with excess energies above the minimum of the *L* valley show ultrafast intervalley scattering, with transition times of the order of 40 fs. The relaxation processes in the *L* valley are resolved in energy and momentum spaces, including their backscattering into the Γ valley during relaxation. In contrast, relaxation of hot electrons with excess energy below the minimum of the *L* valley is governed by the direct impact ionization (IMP). We reveal state-selective features of the IMP process, and we have determined the direct IMP rate to be 7×10^{12} s^{-1} for hot electrons with excess energy in the range of 0.35 to 0.6 eV. The direct IMP process results in a rapid increase, within 300 fs after excitation, of the electron density at the conduction band minimum (CBM), and phonon-assisted IMP by hot electrons scattered in the *L* valley and those backscattered into the Γ valley persistently enhances the electron density up to 8 ps after excitation. By analyzing correlations between the IMP rates of hot electrons and the electron densities near the CBM, an important role of a transient Auger recombination is proposed to quantify the yield of low-energy electrons generated in the IMP process.

DOI: 10.1103/PhysRevB.91.045201 PACS number(s): 78.47.db, 78.47.J−, 79.60.Bm

I. INTRODUCTION

Ultrafast scattering of energetic carriers in semiconductors constitutes the key process that determines the functional limits and properties of micro- and optoelectronics, and it has been an important field of study in both basic and applied researches during the past two decades [1,2]. A greater understanding of this process has now become crucial as fundamental quantum-mechanical carrier-phonon interactions over the whole Brillouin zone determine the properties of nanoscale devices, where the dimensions are comparable to carrier scattering lengths [3–5]. In addition, the determination of the fundamental mechanisms of gigantic structural dynamics in semiconductors induced by ultrashort laser excitation [6–11] requires precise knowledge of ultrafast dynamics of photogenerated hot carriers. Despite accumulating knowledge, a clear understanding of the physics involved in dynamic scattering processes still remains elusive due to the high level of complexity of the problem [1,2].

Transient electron distribution functions (EDFs) provide the key information necessary to elucidating incoherent scattering dynamics [2]. In fact, EDFs experimentally determined so far have provided deeper insight into processes such as electronic thermalization of hot electrons [12], transient electron velocity overshoot in nanostructured semiconductors [13], and intravalley hot-electron relaxation in Si [14]. However, studies based on directly determined transient EDFs are rare for several important ultrafast scattering processes. In particular, a fundamental understanding of processes with state-sensitive characteristics is still incomplete, partly because of the methodological limitation in determining the initial and final states in a scattering process in energy and momentum spaces at sufficient temporal resolution.

One such issue to be explored on the basis of state-resolved EDFs is that of impact ionization (IMP) in semiconductors. Impact ionization is a fundamental process induced by Coulomb interactions among electrons, which results in carrier multiplication through the decay of a hot electron into a lower-energy state while simultaneously generating an electron-hole pair [15,16]. It critically affects carrier transport under high-electric-field conditions [17], and it attracts current interest as a means to improve photovoltaic-device efficiency [18,19]. Extensive studies of the IMP process have been carried out using a variety of experimental and theoretical techniques [20–33].

Impact ionization is essentially a state-selective process, involving four electronic states precisely specified by conservation of energy and momentum. Traditional experimental studies have primarily measured the electric-field dependence of carrier multiplication [20]. Apart from the difficulty in correlating the electric field with the excess energy ε of carriers, information about state-selective characteristics is completely lost by the energy-integrating nature of this method. These difficulties can be reduced through the use of photoelectric techniques, in which the rate of IMP near the threshold may be derived from the wavelength dependence of the quantum yield [21,22]. Even this method, however, is still insufficient for determination of state-sensitive features of the process. In the soft-x-ray-induced core-level photoemission technique, the initial hot-electron state can be well specified. However, the only information obtained by this method is the photoemission-spectral width determined by all possible effects occurring during electron propagation from the interior of the semiconductor to the emitting surface. One must deconvolve all effects to obtain the rate of IMP by the state specified only by its energy. Consequently, these spectroscopic studies neither specify the pertinent energy and momentum states nor directly determine the absolute rate. It is essential to measure the quantum-mechanical dynamics of

IMP on a state-resolved basis in both energy and momentum spaces.

Hot-electron dynamics in InSb can provide a model that shows characteristic features of the dynamics in small band-gap semiconductors. InSb has a low band gap of 170 meV at room temperature [34]. The small band-gap energy suggests a low threshold for IMP and a strong interaction between valence and conduction band states [15]. Therefore, hot-electron relaxation is governed not only by electron-phonon (e-ph) interaction, which induces energy relaxation and intervalley scattering (IVS), but also by electron-electron (e-e) interaction, leading to IMP that is associated with carrier multiplications. It is crucial to resolve the competition and/or interplay between the energy dissipation by e-ph interaction and carrier multiplication by e-e interaction in InSb, not only for the basic understanding of hot-carrier dynamics in small-gap semiconductors, but for further development of technological applications of InSb, which has the highest electron mobility and saturation velocity among all known semiconductors. In addition, InSb has been a model system for the gigantic structural dynamics of nonthermal melting under short-pulse excitation [9–11]. Comprehensive understanding of ultrafast dynamics of hot-carrier relaxation provides a sound basis on which a mechanistic understanding of nonthermal melting may be established. Despite a few pioneering papers, the primary processes of ultrafast scattering of hot electrons in InSb have not yet been clarified [35–37].

Recently, we have demonstrated that transient EDFs, resolved in energy and momentum spaces at femtosecond temporal resolution, can be directly captured using time- and angle-resolved photoemission spectroscopy [38,39]. In particular, state-selective features of the IMP processes can be determined, and the hot-electron states involved in IMP and their ultrafast dynamics have been resolved in energy and momentum spaces at femtosecond temporal resolution in InSb [38]. In this paper, we report comprehensive studies of the entire scattering process of hot electrons with excess energies of up to 1.1 eV above the conduction-band minimum (CBM) in InSb using time- and angle-resolved photoemission spectroscopy. Not only the IMP processes, but also the IVS from the Γ to L valleys, which critically affects the IMP processes, has been thoroughly studied. We extensively discuss the physical processes involved in the ultrafast scattering processes in semiconductors with small band-gap energies, based on the detailed analysis of the newly obtained results and on results previously reported. A quantitative analysis of the relation between the decay of the impacting hot electrons and the increase of low-energy electrons near the CBM reveals a feature indicating that some ultrafast recombination processes coexist with the IMP process in InSb.

This paper is organized as follows. In the next section, we explain the experimental method used in the time- and angle-resolved photoemission spectroscopy. We then describe the primary scattering processes of hot electrons in order, presenting results and their associated quantitative analysis. First, we describe the specific features of the results obtained by this state-resolved spectroscopy. The nascent (as excited) distributions of hot electrons in the conduction band (CB) injected from the heavy-hole (HH), light-hole (LH), and split-off (SO) valence bands are directly captured in momentum and energy spaces. Next, we report the ultrafast dynamics of IVS of hot electrons from the Γ to L valley in detail. We then focus our attention on the IMP processes by analyzing the temporal evolution of both hot-electron distributions and population near the CBM. Two distinctive IMP processes are resolved, and characteristics of each process are described. Finally, we discuss the state-dependent rates of IMP of hot electrons in InSb with reference to the results of previous extensive theoretical studies. A short summary follows.

II. EXPERIMENTAL

Nondoped InSb (100) wafers with a thickness of 3 mm were cleaved under ultrahigh vacuum (UHV) conditions ($<5 \times 10^{-11}$ Torr). Surface structures were characterized *in situ* by a scanning tunneling microscope prior to photoemission measurements [40]. The (110) surfaces displayed well-ordered (1×1) structure with linear In and Sb rows, and with a surface-defect concentration less than 0.5%. We used two different femtosecond laser systems. A mode-locked Ti:Sapphire laser, operated at a 76 MHz repetition rate, was used to generate femtosecond laser pulses between 850 and 730 nm. The fundamental and its third harmonic, generated using betabarium borate crystals, were used as pump and probe pulses. The temporal widths of the pump and the probe pulses were 80 fs and 100 fs, respectively, giving a cross-correlation trace of 125 fs at the sample position in the UHV chamber. To generate laser pulses with shorter temporal widths and wider tunability, we also used a laser system consisting of a Ti:Sapphire laser oscillator, a regeneratively amplified Ti:Sapphire laser, and a tunable optical parametric amplifier operated at 250 kHz repetition rate. The optical parametric amplifier generated 50 fs laser pulses at photon energies tunable from 1.2 to 2.4 eV. A part of the amplified fundamental output at 790 nm was used to generate the third harmonic with a temporal width of 65 fs, giving a cross-correlation trace of 82 fs. Pump and probe pulses, with a preset time delay (Δt), were aligned coaxially and focused on the sample surfaces at 45° to normal. The fluence of a pump pulse was set less than 5 μJ/cm^2, resulting in excitation density at most 2×10^{18} cm^{-3}. The probe-pulse fluence was typically 5 nJ/cm^2.

For photoemission spectroscopy, a hemispherical electron analyzer, equipped with an angle-resolved lens mode and a two-dimensional image-type charge-coupled device (CCD) detector, served as the electron spectrometer. The entrance slit of the analyzer was placed on the optical plane (detection plane) defined by the incoming and reflected light. Two-dimensional images of photoelectrons were recorded as functions of energy and of emission angle θ with respect to the surface normal along the [001] crystallographic direction. The instrumental energy resolution with femtosecond probe light was 50 meV, while angle resolution was limited in the range of $\pm 1°$.

The ionization energy of InSb (110) is 4.77 eV [41]. Therefore, the third harmonic of the 790 nm fundamental beam, the photon energy of which is 4.71 eV, was mainly used to probe electron states populated in the CB. The probe photon energy does not directly ionize electronic states in the valence band. This discrimination is crucial to accurately capture

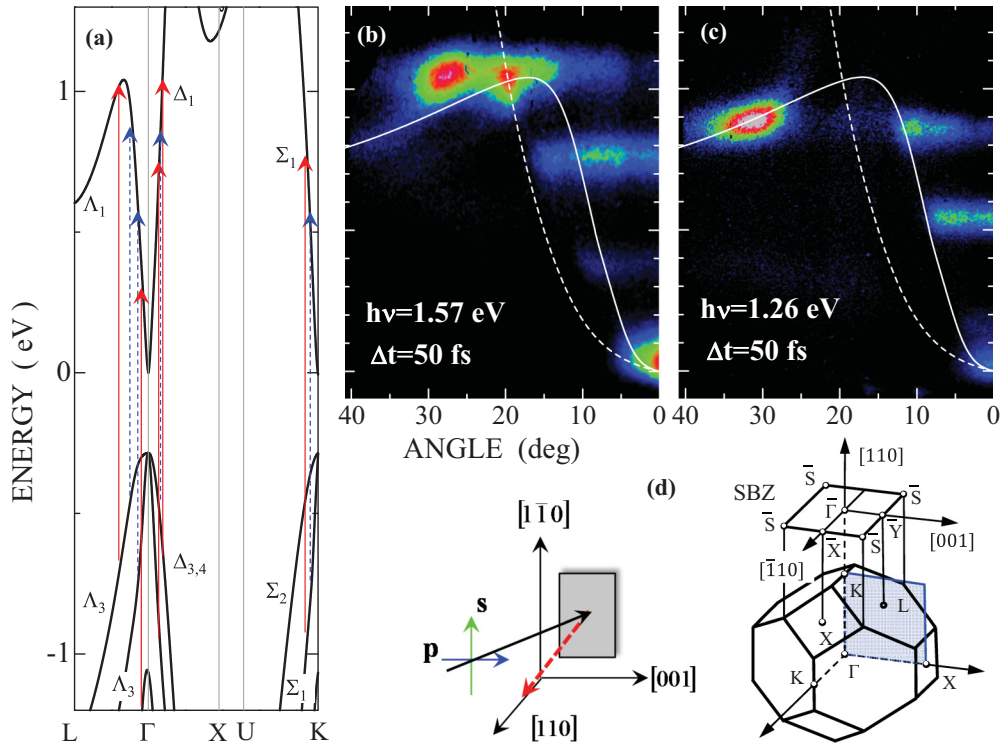

FIG. 1. (Color online) (a) A portion of the band structure of InSb [42]. Energy is referenced to the CBM. The symmetries of the spatial parts of wave functions are given for the points involved in the interband transitions [43]. Solid (broken) arrows show typical allowed optical transitions for p-polarized 1.57 eV (1.26 eV) photons. (b) The image of photoelectrons 50 fs after excitation with p-polarized 1.57 eV light pulses. (c) The image of photoelectrons 50 fs after excitation with p-polarized 1.26 eV light pulses. In (b) and (c), the color scale indicates the photoemission intensity. The solid and broken curves in (b) and (c) represent the dispersion of the conduction bands along the Γ-L and Γ-X directions, respectively. (d) The experimental geometry of photoemission measurements, and the relation between the SBZ and BBZ for InSb with (110) surface under this geometry. The shaded plane (shown by light blue color) is the projection plane.

the photoinjected hot electrons in the CB, while suppressing completely the very intense signals of direct photoionization from the valence band.

III. RESULTS AND DISCUSSION

A. Imaging nascent hot-electron distributions in energy and momentum spaces

Figure 1(a) shows the band diagram of InSb [42]. Figures 1(b) and 1(c) show photoemission maps as a function of electron energy ε and emission angle θ, measured at $\Delta t = 50$ fs with p-polarized 1.57 eV and 1.26 eV pump pulses. In these figures, the energy is referenced to the CBM; ε represents the excess energy of hot electrons above the CBM. The location of the CBM in the photoemission spectra plotted as a function of measured electron kinetic energy can be precisely determined, as will be shown later. We denote the photoemission intensity as $I(\varepsilon, \theta, \Delta t)$ hereafter. Excitation with pump-photon energy ($h\nu_{\text{pump}}$) of 1.57 eV injects electrons at energies of 1.05 eV, 0.76 eV, and 0.38 eV above the CBM. Energetically, these peaks correspond to the transitions from the HH, LH, and SO valence bands, respectively, as shown by red solid arrows in Fig. 1(a). However, excitation with $h\nu_{\text{pump}} = 1.26$ eV injects electrons at 0.86 eV, 0.53 eV, and 0.16 eV. Considering the difference in pump-photon energies, these peaks correspond to the transitions from the HH, LH, and SO valence bands, similar to 1.57 eV excitation, as shown by blue broken arrows in Fig. 1(a).

As displayed in Fig. 1(d), the [110] crystal axis is aligned along the surface normal, and the [110] and [001] axes define the detection plane. The relation between the bulk Brillouin zone (BBZ) and the surface Brillouin zone (SBZ) is also shown under this geometry [44]. As the emission angle θ corresponds to electron momentum ($k_{//}$) parallel to the surface along the $\bar{\Gamma}$-$\bar{Y}$ direction of the SBZ, the measured photoemission image represents a one-dimensional cut, along the $\bar{\Gamma}$-$\bar{Y}$ direction of the SBZ, of the two-dimensional projection of three-dimensional electron distributions. In the projection, all states along the Γ-K direction in the BBZ are projected at $\bar{\Gamma}$, contributing to normal photoemission. However, the states along the Γ-X direction in the BBZ are projected on the $\bar{\Gamma}$-$\bar{Y}$ direction with $k_{//} = k_X$. Similarly, the states along the Γ-L in the BBZ are projected on the $\bar{\Gamma}$-$\bar{Y}$ direction with $k_{//}$ given by $k_{//} = k_L \cos(54.7^\circ)$. Here k_X and k_L are wave vectors along the Γ-X and Γ-L directions in the BBZ of InSb. The shaded plane (light blue) in Fig. 1(d) represents the projection plane. In order to correlate the photoemission images in Figs. 1(b) and 1(c) with the band structure, the dispersion of the CB along the Γ-L and Γ-X directions is plotted by the solid and broken curves, respectively, in the figures (see Appendix).

The polarization selection rule [43] predicts that the transition from the HH (Σ_2) or LH (Σ_1) band at Σ is selectively

allowed under s- or p-polarized excitation in this geometry. In fact, the intensity around $k_{//} = 0$ for the highest-energy peak injected from the HH band is weak for both $h\nu_{\rm pump}$ values. However, the intensity around $k_{//} = 0$ of the second-energy peak from the LH band is intense, as predicted by the selection rule. In the case of 1.57 eV excitation [Fig. 1(b)], the highest-energy peaks from the HH band show maxima at $\theta = 19^\circ$ and $\theta = 27.5^\circ$. Based on the conduction-band dispersion, we can identify that the peak with $\theta = 19^\circ$ is due to the transition at $k_X = 1.85 \times 10^7\,{\rm cm}^{-1}$ near Δ in the Γ valley. However, the peak at $\theta = 27.5^\circ$ results from packets formed outside the Γ valley, as the emission angle is greater than the border of the Γ valley, defined by the dispersion along the Γ-X direction [see Fig. 1(e)]. Because of rather poor agreement between the theoretical band calculation and experimental results, it is not conclusive whether they are injected directly by the optical transition in the L valley. However, as will be discussed in the next section, we have concluded that the photoemission peak at $\theta = 27.5^\circ$ is certainly due to the direct transition from the HH band to the CB near Λ in the L valley.

In the case of 1.26 eV excitation in Fig. 1(c), the excess energies of hot-electron packets are reduced by the amount predicted by the reduction of pump-photon energy. For the highest-energy peaks from the HH band, the most intense peak is observed at $\theta = 32^\circ$, i.e., larger than with 1.57 eV excitation. For $h\nu_{\rm pump} = 1.26\,{\rm eV}$, the band structure calculation predicts the optical transition from the HH band to the state with $\varepsilon = 0.85\,{\rm eV}$ within the Γ valley at $k_L = 1.99 \times 10^7\,{\rm cm}^{-1}$, which corresponds to an emission angle of 11.5°. In fact, Fig. 1(c) shows a clear peak at the point predicted by the theory. We therefore conclude that the intense peak at $\theta = 32^\circ$ in the L valley is not generated by direct optical transitions, but by efficient IVS of hot electrons generated in the Γ valley. This conclusion is proven by measuring the temporal changes in the population of the peaks, as shown in the next section.

Figure 2(a) displays the angle-integrated spectra of $I(\varepsilon, \theta, 50\,{\rm fs})$ measured at $\Delta t = 50\,{\rm fs}$ for p-polarized light with values for $h\nu_{\rm pump}$ ranging from 1.20 to 1.57 eV. The angle range for integration was limited from 0° to 12°; the hot-electron population only within the Γ valley is displayed. For all spectra, we clearly see three peaks, together with the peak from electrons at the CBM. In Fig. 2(b), the photoemission spectrum integrated with respect to θ from 0° to 35° is shown for $h\nu_{\rm pump} = 1.57\,{\rm eV}$. It is clear that the optical transitions from the HH band contribute the largest proportion of injecting hot electrons. However, the major fractions are populated outside of the Γ valley, even at values of Δt as short as 50 fs, leading to the relatively weak peak intensities of the highest-energy peak in the spectra in Fig. 2(a).

As seen in Fig. 2, as well as Figs. 1(b) and 1(c), an intense photoemission peak is detected near the CBM. A similar peak is observed even without pump pulses because of a significant population of electrons at the CBM thermally excited in InSb; the intrinsic carrier density in InSb has been reported as $2.0 \times 10^{16}\,{\rm cm}^{-3}$ [45]. Spectral-shape analysis using convolution procedure for the finite experimental energy resolution of 50 meV was used to determine the energy level of the CBM and the initial electron density before the incidence of pump-laser pulses (see Appendix). The observed photoemission spectrum was well fitted by the Fermi distribution with $E_F = 0.01\,{\rm eV}$

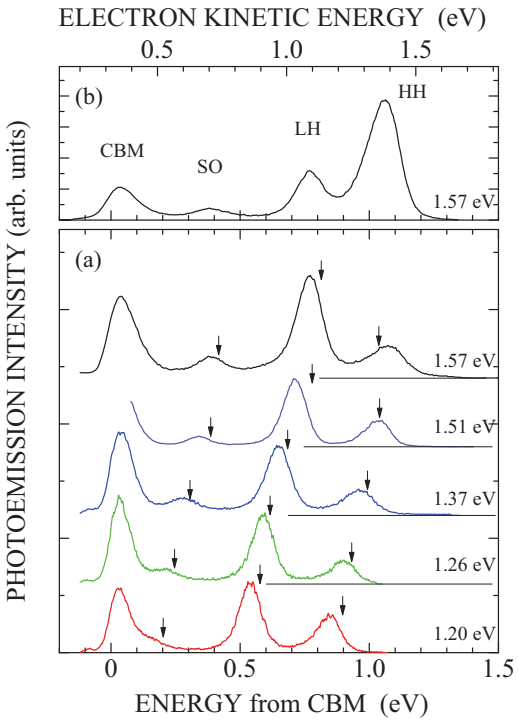

FIG. 2. (Color online) (a) Angle-integrated photoemission spectra, measured at $\Delta t = 50\,{\rm fs}$, excited by p-polarized light pulses with photon energies ranging from 1.20 to 1.57 eV. The angle range for integration is from 0° to +12°. (b) Angle-integrated photoemission spectrum, measured at $\Delta t = 50\,{\rm fs}$, excited by p-polarized light pulses with photon energy of 1.57 eV. The angle range for integration is from 0° to 35°. In (a) and (b), the upper scale represents the electron energy referenced to the work function of the analyzer. The small arrows in (a) near the peaks are the final-state energy of interband transitions predicted from band-structure calculation near the Λ point [42].

and $T_e = 293\,{\rm K}$, with the energy of the CBM at 0.32 eV in the electron kinetic energy referenced to the work function of our analyzer (upper scale in the figure). Using the CBM energy, energies ε values of hot electrons are referenced to the CBM to represent the excess energies above the CBM. The estimated electron density was $3.0 \times 10^{16}\,{\rm cm}^{-3}$ [46].

The arrows in Fig. 2(a) show the predicted final-state energies in the CB for optical transitions from HH, LH, and SO bands around the Λ point along the Γ-L direction. The observed peak energies agree reasonably well with the predicted energies [42]. Therefore, our photoemission spectroscopy directly captures the nascent distribution of photoinjected electrons in the bulk CB of InSb, resolved in energy and momentum spaces.

B. IVS of hot electrons

As shown in Fig. 1(c), two hot-electron packets are generated at $\varepsilon = 0.9\,{\rm eV}$ under $h\nu_{\rm pump} = 1.26\,{\rm eV}$; one at $k_L = 1.99 \times 10^7\,{\rm cm}^{-1}$ ($\theta = 11.5^\circ$) in the Γ valley, and the other at $k_L = 4.7 \times 10^7\,{\rm cm}^{-1}$ ($\theta = 32^\circ$) in the L valley. The intensity of electrons injected at the Δ region is much weaker than these two as seen in Fig. 1(c). We neglect this component in the analysis. In Fig. 3(a), we show the temporal changes

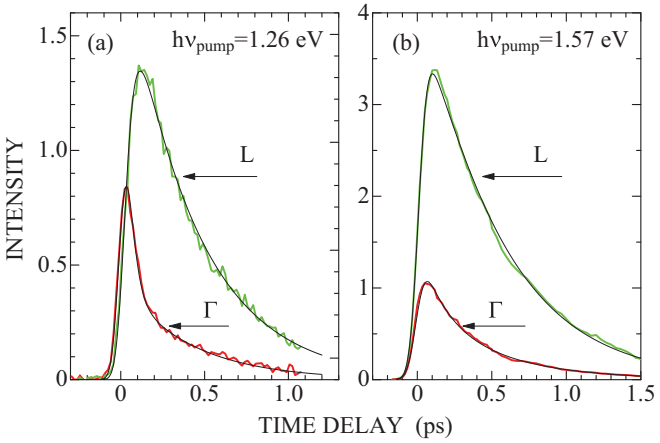

FIG. 3. (Color online) (a) Temporal changes in the population of hot electrons with excess energy of 0.85 ± 0.25 eV in the Γ and L valleys injected by 1.26 eV pump lasers. The thin solid curves show the best fit obtained by the set of rate equations. (b) Temporal changes in the population of hot electrons with excess energy of 1.05 ± 0.25 eV in the Γ and L valleys injected by 1.57 eV pump lasers. The thin solid curves show the best fit obtained using the set of rate equations.

in the hot-electron populations injected from the HH band at two different points in the momentum space. The intensities at $\varepsilon = 0.85$ eV, integrated with a width of ± 0.025 eV, are plotted. It is clear that generation in the Γ valley precedes that in the L valley, and that the fast decay of the population in the Γ valley is accompanied by continuous growth in the population in the L valley. This result demonstrates that the direct photoinjection of hot electrons occurs in the Γ valley first, and that photoinjected hot electrons are scattered into the L valley at a high rate. Interestingly, the decay characteristics of populations in both valleys are the same at $\Delta t > 300$ fs after the rapid decrease in the population in the Γ valley. Therefore, it is likely that the backscattering from the L to Γ valleys contributes to the slow-decaying part of the intensity in the Γ valley.

To confirm these qualitative conclusions, we analyzed the results in Fig. 3(a) using a set of rate equations and introducing the generation function $G(t,t_0,w)$, which describes the temporal change in the generation process of a given excited state (see Appendix). For the population $n_\Gamma(n_L)$ in the $\Gamma(L)$ valley,

$$\frac{dn_\Gamma}{dt} = G(t,t_0,w) - k_{\Gamma L} n_\Gamma + k_{L\Gamma} n_L, \quad (1)$$

$$\frac{dn_L}{dt} = k_{\Gamma L} n_\Gamma - (k_{L\Gamma} + k_L) n_L, \quad (2)$$

where $k_{\Gamma L}(k_{L\Gamma})$ is the IVS rate from the Γ to L (L to Γ) valleys, and k_L is the energy relaxation rate in the L valley. The experimental results in Fig. 3(a) are modeled well by the set of rate equations, with parameter values $k_{\Gamma L} = 22.7 \times 10^{12}\ s^{-1}$, $k_{L\Gamma} = 4.3 \times 10^{12}\ s^{-1}$, and $k_L = 2.9 \times 10^{12}\ s^{-1}$, as shown by the thin solid curves in the figure. Therefore, we conclude that hot electrons from the HH valence band are injected in the Γ valley by optical transitions, but scattered into the L valley at a high rate for the case of 1.26 eV excitation. The Γ-to-L IVS time is as short as 44 fs.

The rate of IVS from the Γ to L valley is fast, and the magnitude is similar to the rate in GaAs recently determined [39]. However, the rate of L to Γ backscattering is about 1/5 of $k_{L\Gamma}$. Assuming the same deformation potential interaction strength of Γ-to-L and L-to-Γ scattering for a given ε, the rate depends on the density of the final states of scattering at the specified energy [15]. The density of state effective mass in the L valley is 19 times larger than that in the Γ valley [34]. However, the nonparabolicity of the CB in the Γ valley is significant [47], resulting in a substantial enhancement of the density of state relative to that estimated simply by the effective mass at the CBM. In the literature, a value for the nonparabolicity parameter of $5.72\ eV^{-1}$ has been reported [47]. At an excess energy of $\varepsilon = 0.85$ eV, the density of states is calculated to be 27 times larger than that calculated from simple estimation based on the effective mass at the CBM. Then, the density of states in the Γ valley at $\varepsilon = 0.85$ eV becomes as high as 70% of that in the L valley at 0.28 eV from the minimum (L_1) of the L valley; the fast backscattering rate from the L to Γ valley may be justified.

Figure 3(b) displays similar temporal changes in the hot-electron populations injected from the HH band in the Γ and L valleys under 1.57 eV excitation. Intensities at $\varepsilon = 1.05$ eV, integrated with a width of ± 0.025 eV, are plotted. In contrast to 1.26 eV excitation, no time delays exist in the rise of the population in the L valley. Also, the peak intensity in the L valley, relative to that in the Γ valley, is much higher than for 1.26 eV excitation, as shown in Fig. 3(a). This result shows that hot electrons are injected in the L valley directly by optical transitions near the Λ for the case of 1.57 eV excitation. In fact, when we add another generation function for direct injection into the L valley, Eqs. (1) and (2) describe reasonably well the temporal changes in the hot-electron populations in both the L and Γ valley, using similar rates of L-to-Γ and Γ-to-L IVS to those for 1.26 eV excitation, as shown by thin solid curves in Fig. 3(b). Analysis predicts that the generation efficiency into the L valley is higher by a factor 1.8 than into the Γ valley. This may come from a strong singularity of the final state of optical transitions around Λ. A Γ-to-L IVS time of 39 fs is estimated from the analysis at $\varepsilon = 1.05$ eV.

The hot electrons injected from the HH band, after mutual repopulations between L and Γ valleys in a short temporal period of about 300 fs, are populated mostly in the L valley. In Figs. 4(a) and 4(b), we show the photoemission images measured at $\Delta t = 40$ fs and 1.0 ps for θ from 12° to 43°. The initial narrow wave packet of photoinjected hot electrons is redistributed towards wider energy and angular ranges at $\Delta t = 1.0$ ps during the intra-L-valley relaxation, accompanied by peak shifts toward larger angles. The population in the Γ valley becomes low around $\Delta t = 1.0$ ps, while the hot-electron population in the L valley is persistent even at $\Delta t > 1$ ps. Therefore, the largest proportions of hot electrons injected from the HH band are populated in the L valley 300 fs after excitation, and relaxing towards the minimum (L_1) in the L valley.

Figure 4(c) shows temporal changes in the energy-resolved populations of hot electrons in the L valley; the intensities integrated in the range of ± 25 meV for the indicated energy

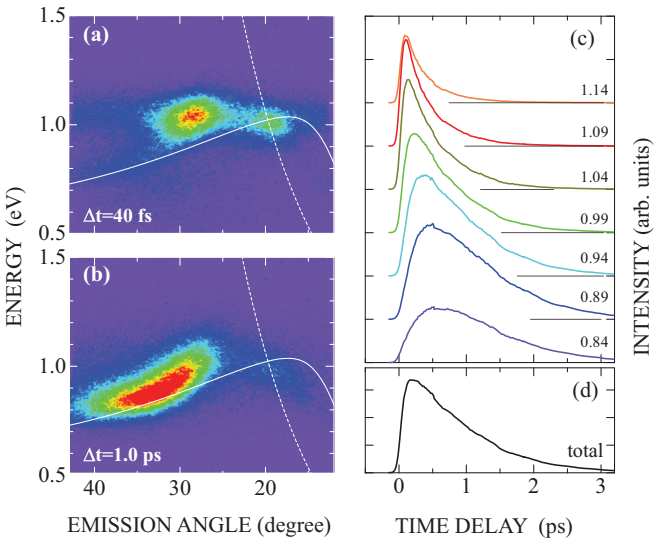

FIG. 4. (Color online) (a) Image of photoelectrons 40 fs after excitation with p-polarized 1.57 eV light pulses for the emission angle 12° to 43°. (b) Image of photoelectrons 1.0 ps after excitation with p-polarized 1.57 eV light pulses for the emission angle 12° to 43°. The solid and broken curves in (a) and (b) display dispersion of the CBs along the Γ-L and Γ-X directions, respectively, calculated as a function of emission angle. (c) Temporal evolution of the angle-integrated photoemission intensities from the hot electrons in the L valley at the excess energy shown by the values in the figure. The intensities within ±0.025 eV from the energy specified are integrated. (d) Temporal evolution of the total photoemission intensities from hot electrons in the L valley, integrated with respect to excess energy and emission angle ($\theta > 23°$).

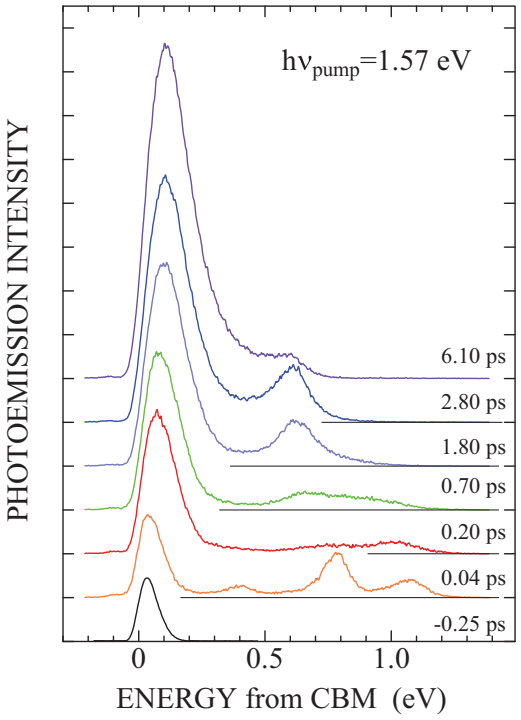

FIG. 5. (Color online) Angle-integrated photoemission spectra at typical time delays under p-polarized 1.57 eV light excitation. The range of angles for integration is from 0° to +12°. The energy scale is referenced to the CBM.

in the figure are plotted as a function of Δt. High-energy populations show peaks at shorter time delays, while lower-energy populations have peaks at longer Δt values. This feature demonstrates a direct consequence of cascade energy relaxation processes by e-ph interaction in the L valley. The hot electrons in the L valley are relaxed down to the L_1 with energy-dependent population decay times; the population decay time is 0.42 ps at 1.15 eV, while it is 0.78 ps at 0.85 eV. Figure 4(d) displays temporal changes in the total hot-electron population in the L valley, obtained by integrating $I(\varepsilon, \theta, \Delta t)$ with respect to ε and θ ($\theta > 22°$). The population is dissipated 3 ps after excitation, because of intra-L-valley relaxation towards larger angles; our detection window is limited to $\theta \sim 43°$. The hot electrons accumulated near the L_1 will be scattered back into the Γ valley at later times. This backscattering is in fact observed, and it is discussed in the next section.

C. Impact ionization processes

1. Overview of population increase near the CBM and hot-electron relaxation

Figure 5 displays angle-integrated photoemission spectra measured at typical pump-probe delays, Δt. Only the electron populations in the Γ valley are displayed. The peak from electrons at the CBM, with a substantial initial intensity due to thermal excitation, dramatically increases after photoexcitation and shows the following characteristics.

At $\Delta t = 40$ fs, the intensities of the hot electrons injected from the LH band show a maximum, and they decay almost completely at $\Delta t = 200$ fs. The intensity of the peak from the SO band shows similar ultrafast temporal behavior. In this short temporal domain, the population near the CBM increases significantly at $\Delta t < 200$ fs. The peak from the hot electrons injected in the Γ valley from the HH band loses it intensity while exhibiting a low-energy shift in peak energy. From the temporal shift of the peak energy, it is evaluated that the energy relaxation rate is 0.43 eV/ps. This rate is too slow to reach the CBM within 200 fs; the strong enhancement of the CBM population is irrelevant to the energy relaxation of hot electrons via the e-ph interaction. At $\Delta t = 700$ fs, a new peak appears around $\varepsilon = 0.6$ eV. Importantly, intensities of both the peak near the CBM and the new peak increase concurrently. The newly formed peak shows the maximum intensity 2.8 ps after excitation, and it loses intensity gradually, accompanied by further increase in the CBM population. However, the distinct peak structure survives until as long as 6 ps without merging with the peak from the CBM population. For the peak around $\varepsilon = 0.6$ eV, an energy relaxation rate of 0.03 eV/ps is evaluated. The rate is slow, so that energy relaxation via the e-ph interaction is not the mechanism inducing the continuous increase in the population near the CBM. The CBM population reaches a maximum around $\Delta t = 8$ ps, and then it starts to decrease.

The new peak at approximately $\varepsilon = 0.6$ eV is formed around $\Delta t = 1$ ps by the relaxation of hot electrons. As

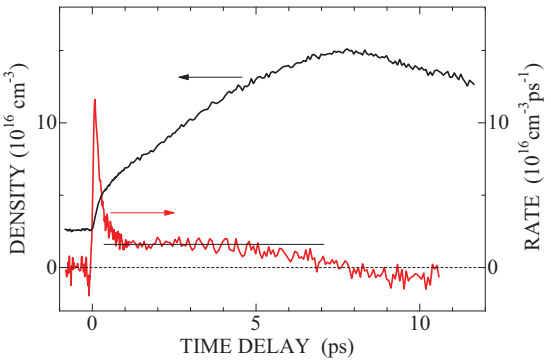

FIG. 6. (Color online) Temporal changes in the electron population near the CBM measured at 293 K, black line, under 1.57 eV excitation with an excitation density of 3.8×10^{17} cm^{-3}, and its first derivative with respect to time, shown by red (dark gray) line.

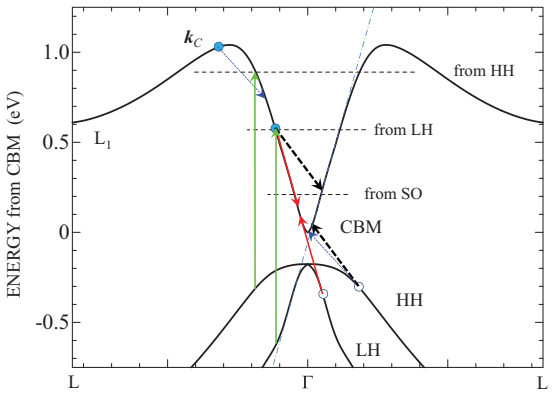

FIG. 7. (Color online) Dispersion of valence and CBs along the Γ-L direction (after Ref. [42]). The vertical arrows (green in color) indicate optical transitions from HH and LH valence bands to the CBs at excitation-photon energy of 1.26 eV. Red (dark gray) solid arrows show possible direct IMP process of hot electrons interacting with the LH valence-band states, and broken arrows show possible direct IMP process of hot electrons interacting with the HH valence-band states. The point labelled $\boldsymbol{k}_c$ in the L valley represents the critical point from which direct IMP via interaction with the HH band states is possible, and the dot arrows shows the direct IMP transition in this case. The chain line is inserted to emphasize the almost linear dispersion of the CB and the LH bands.

described in Sec. III B, the majority of hot electrons are populated in the L valley within 300 fs after excitation, and they are relaxed down to the L_1. They then induce L-to-Γ backscattering. As seen in Figs. 4 and 5, as the population in the L valley decreases, the peak around $\varepsilon = 0.6$ eV grows. The band calculation predicts that the minimum of the L valley is 0.61 eV above the CBM [42], which nearly coincides with the energy of the newly formed peak. Based on the theoretical prediction, together with the observed anticorrelation between the L-valley population and the peak intensity of the 0.6 eV peak, we conclude that the newly formed photoemission peak is due to backscattering of hot electrons from L_1 to the position $\Gamma(L_1)$ in the Γ valleys. Based on the spectral-shape analysis, we determined that the energy of the L_1 is 0.63 eV above the CBM at 293 K (see Appendix).

Figure 6 shows temporal changes in the electron population near the CBM under 1.57 eV excitation. The density n_{CBM} of electrons near the CBM was evaluated by integrating the photoemission intensities with respect to the angle and energy. The angle range for the integration was $\pm 10°$, based on the angle-resolved measurements as in Fig. 1(b). The energy range for the integration is $E_K < 0.65$ eV for $\Delta t < 0.5$ ps, while it is $E_K < 0.75$ eV for $\Delta t > 0.5$ ps, based on the spectral-shape changes seen in Fig. 5. The n_{CBM} increases rapidly within $\Delta t < 300$ fs, and it reaches a maximum at around $\Delta t = 8$ ps. Using the calibrated relation between the photoemission intensity and the electron density [46], we determine that the maximum corresponds to an electron density of 1.5×10^{17} cm^{-3} at $\Delta t = 7.8$ ps for the excitation density of 3.8×10^{17} cm^{-3}.

Base on the result shown in Fig. 6, we can distinguish two distinctive processes contributing to the growth of n_{CBM}: a fast process, at $\Delta t < 300$ fs, and a slow process, which follows it. In order to quantify the differences between the fast and slow processes, we plot the derivative of n_{CBM} with respect to Δt, which represents the rate of CBM-population growth. The rate shows the maximum value of 1.1×10^{17} cm^{-3} ps^{-1} at $\Delta t = 0.08$ ps, and it maintains a constant magnitude of 1.3×10^{16} cm^{-3} ps^{-1} at the later temporal range through $\Delta t = 5$ ps, after which it then decreases at $\Delta t > 5$ ps.

As seen in Fig. 5, hot electrons injected from the SO and LH bands decay within 300 fs postexcitation, just as with the fast process of CBM-population enhancement. Therefore, the ultrafast increase in the electron population near the CBM results from fast IMP due to hot electrons injected from SO and LH bands. Also, as mentioned above, the mechanism of CBM-population enhancement for the slow process is not due to the energy relaxation of hot electrons via the e-ph interaction towards the CBM, but rather by IMP of hot electrons. However, a significant difference in the rate shown in Fig. 6 implies that there are two different IMP mechanisms for the processes.

Under the present experimental conditions, we can exclude any role of the X valley, since the minimum of the X valley, 1.27 eV above the CBM [42], is too high to transfer the photogenerated hot electrons. Therefore, we consider only the Γ and L valleys of InSb. In Fig. 7, we show the band diagram from Ref. [42] for qualitative discussion of the possible differences in the IMP rates for the fast and slow processes. In a direct two-band IMP, the four electronic states involved have to satisfy energy- and momentum-conservation rules. Hot electrons in the Γ valley have finite probabilities to induce direct IMP by interacting with electrons at the HH valence band (the HH-band channel), as indicated by the broken arrows, and with those at the LH valence band (the LH-band channel), indicated by solid red arrows. The relative importance of the two channels will be discussed later. The electrons at the SO band are too deep for induced IMP by hot electrons, with excess energies of about 1 eV studied here.

On the other hand, hot electrons in the L valley cannot induce direct IMP through the LH-band channel, because of the large difference in the dispersion of the bands. For the HH-band channel, there may be a critical energy (E_C) and wave number ($\boldsymbol{k}_C$) that can satisfy the energy and momentum conservations for direct IMP because of the significant difference in the dispersions of the CB and HH bands. A possible IMP transition

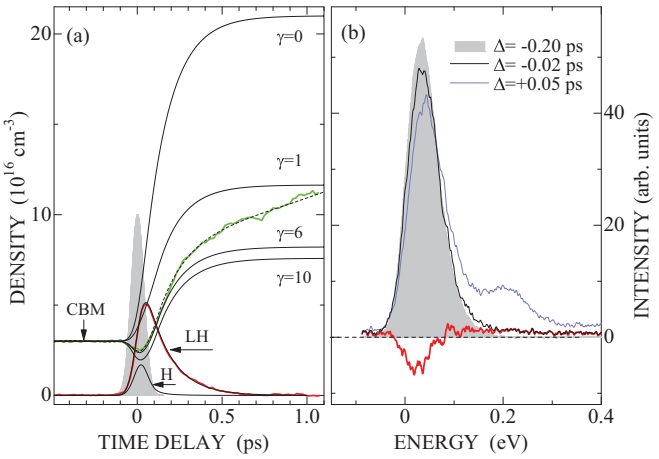

FIG. 8. (Color online) (a) Temporal changes in the electron population near the CBM, green (gray) curve, and in the hot-electron population, red (dark gray) curve, injected from the LH band under 1.26 eV excitation. The gray peak shows the cross-correlation trace of pump and probe pulses. The thin solid curves and the broken curves are the results of analysis using a rate equation model (see text). (b) Comparison of the photoemission spectral shapes measured at $\Delta t = -200$ fs, gray peak, -20 fs, black solid curve, and $+50$ fs, blue (gray) curve under 1.26 eV excitation. The thick red (dark gray) solid curve is the difference between the spectra measured at $\Delta t = -200$ fs and -30 fs.

near the critical point is shown by dot arrows. The critical values estimated for one-dimensional Γ-to-L direction in the band structure are $E_C = 1.0$ eV and $\boldsymbol{k}_C = 3.2 \times 10^7$ cm^{-1}. The magnitude may change when we consider all possible three-dimensional directions in the BBZ. For hot electrons with wave vectors larger than $\boldsymbol{k}_C$ in the L valley, direct IMP is no longer possible. Therefore, they can induce IMP only with assistance of phonons (phonon-assisted IMP) [16], or after intervalley backscattering into the Γ valley at $\Delta t > 1$ ps, as described above. Therefore, the majority of hot electrons transferred into the L valley can only contribute to the CBM-population enhancement at a longer temporal region after excitation, leading to the slow process of CBM-population enhancement in Fig. 6.

2. *The fast IMP process*

To gain deeper insight into the fast IMP process, we first focus our attention on the results under 1.26 eV excitation. In this circumstance, hot electrons injected from the SO band have excess energies of about 0.2 eV (see Fig. 2), below the threshold energy of IMP [33]. Also, electrons injected from the HH band are transferred in the L valley at a high rate to the states with k_L larger than $\boldsymbol{k}_C$, preventing direct IMP. On the other hand, the hot electrons from the LH band are injected into the Γ valley at states below the minimum energy of the L valley, but with sufficient excess energies to induce IMP. Therefore, we can analyze the simplest case where only one hot-electron wave packet induces the fast IMP.

In Fig. 8(a), we show temporal changes in the population of hot electrons injected from the LH band and in the electrons near the CBM under 1.26 eV excitation for the temporal domain of the fast process. As clearly seen in the figure, the electron density near the CBM is reduced first at approximately $\Delta t = 0$, and then it increases stepwise. The reduction is evident in Fig. 8(b), where photoemission spectra at $\Delta t = -200$ fs (before excitation) and $\Delta t = -20$ fs (the beginning of the pump-probe overlap) are compared. At $\Delta t = +50$ fs, the injection of hot electrons from the SO band induces a clear peak at $\varepsilon = 0.2$ eV, while the injected amount is still negligible at $\Delta t = -20$ fs. Therefore, the comparison of the spectra at $\Delta t = -200$ fs and $\Delta t = -20$ fs gives evidence of the initial reduction of the CBM population. It is clear that only the low-energy portion of the CBM peak, described by a Fermi distribution function with $E_F = 0.01$ eV, is predominantly reduced in this process.

Direct photoexcitation of the electrons near the CBM to higher states is not possible under 1.26 eV excitation [42]. In addition, since the position of the low-energy cutoff in the photoemission spectra remains the same at any value of Δt, energy shifts coming from excitation-induced surface band bending are excluded [36]. Band renormalization in the high-density excitation regime, if any, leads to the low-energy shift of the CBM position, resulting in the enhancement of the low-energy portion of the difference spectra, contrary to the experimental results. Therefore, the changes in the spectral shapes shown in Fig. 8(b) suggest that the reduction is associated with injection of holes into the initially occupied electronic states with Fermi distribution at the CBM.

Important roles of ultrafast recombination processes to reduce the CBM population are also suggested from the following features shown in Fig. 8(a). As shown in this figure, the hot-electron population injected from the LH band decays with a time constant of 130 fs. As the excess energy is lower than E_{L1} determined above, it is reasonable to assume that the decay is free from the IVS. The decay is entirely due to direct IMP, since radiative and phonon-assisted nonradiative recombinations occur at much slower rates [48]. We would then expect the enhancement of the CBM population shown by solid curve, labeled $\gamma = 0$, in Fig. 8(a). However, the observed enhancement is about one third of the expected magnitude, suggesting that some ultrafast recombination processes coexist with the fast process of IMP.

We presume that the initial reduction of the CBM population comes from some recombination processes induced by interactions of photogenerated carriers and the electrons preexisting at the CBM. A candidate for this is *a transient Auger process* induced by hot holes injected by light pulses. The direct Auger recombination is the reverse process of direct IMP, and energy and momentum conservation rules, similar to the case of the direct IMP, must be satisfied [15]. Because of the strong dispersion of the CB, the transient Auger recombination can be induced only by hot holes with a small wave number of less than $\sim 1 \times 10^7$ cm^{-1}. As shown in Fig. 7, hot holes injected into the HH band by optical transitions have $k_L > 2 \times 10^7$ cm^{-1} and thus can be ruled out. However, the holes injected in the LH band can satisfy the conditions to induce the transient Auger process, and both an electron and a hole can be excited to higher states. In particular, the LH band along the Γ-X direction shows an almost linear dispersion to far deeper states, providing available states for hole excitation [42].

By introducing transient Auger recombination as a hypothetical model for ultrafast recombination, we can analyze the fast process of IMP from hot electrons at 0.58 eV above the CBM. We use the following set of rate equations for the density, n_{HE}, of hot electrons generated from the LH band, the density n_h of holes injected into the LH valence band, and the electron density near the CBM n_{CBM}:

$$\frac{dn_{HE}}{dt} = G(t,t_0 w) - k_{IMP} n_{HE}, \tag{3}$$

$$\frac{dn_h}{dt} = G(t,t_0,w) - \frac{1}{\tau_h} n_h - \gamma n_h (n_{CBM})^2 + k_{IMP} n_{HE}, \tag{4}$$

$$\frac{dn_{CBM}}{dt} = 2k_{IMP} n_{HE} - \gamma n_h (n_{CBM})^2, \tag{5}$$

where $G(t,t_0 w)$ is the generation function defined previously, k_{IMP} is the rate of IMP, τ_h is the lifetime of holes in the LH band, and γ is the coefficient that characterizes the transient Auger recombination. In Eq. (4), holes in the LH valence band are generated at the same rate as the hot electrons from the LH band are, and also they are generated per each event of IMP. The quantities of $G(t,t_0 w)$ and k_{IMP} were already determined (see Appendix). Therefore, we essentially have two parameters, τ_h and γ, to determine in order to fit the temporal change in the CBM population in Fig. 8(a) [49].

The significant effect of the proposed transient Auger process in the dynamics of CBM-population enhancement is evident, as shown by thin solid curves obtained for different magnitudes of γ for a fixed parameter of $\tau_h =$ 40 fs. Therefore, the depression around $\Delta t = 0$, as well as the enhancement around $\Delta t = 500$ fs, becomes a sensitive measure for determination of reasonable values for γ. Taking into account the nonnegligible contribution of the slow process characterized by linear growth, the best fit was obtained for $\tau_h = 40$ fs and $\gamma = 6$, as shown by the solid curve; the broken curve is the result after adding the contribution from the slow process, which is determined by a later analysis for CBM-population enhancement in the larger Δt region. As seen in Fig. 8(a), the peak of estimated hole density in the LH valence band coincides with the minimum of the CBM population, consistent with our model that the initial depression of n_{CBM} is caused by the transient Auger effect by the photoinjected holes in the LH band.

In Fig. 9(a), we show temporal changes in the populations of hot electrons injected from the LH and SO bands and in the electrons near the CBM under 1.57 eV excitation for the temporal domain relevant to the fast process. In contrast to 1.26 eV excitation, an obvious initial reduction in the CBM population is not detected under 1.57 eV excitation. Although the magnitude of n_{CBM} obtained by integrating intensities for the range $\varepsilon < 0.3$ eV remains constant at $\Delta t < -30$ fs, the spectral shape of the photoemission peaks representing the CBM population changes significantly in this temporal domain. Figure 9(b) displays the photoemission spectra measured at $\Delta t = -200$ fs and -30 fs. The low-energy portion of the initial spectrum is reduced, while the high-energy region is enhanced; the net change by integration is almost zero. Therefore, the distribution of electrons at the CBM is significantly modified at this temporal domain. Similar to the

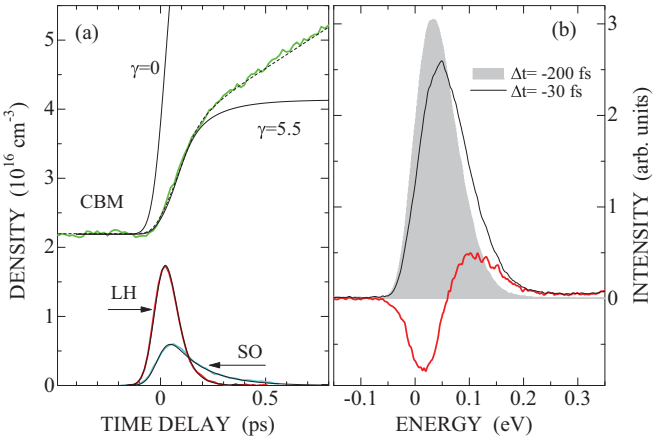

FIG. 9. (Color online) (a) Temporal changes in the electron population near the CBM, green (gray) curve, and the hot-electron populations injected from the LH band, red (dark gray) line, and from the SO band, light blue (gray) line, under 1.57 eV excitation. Thin solid curves and the broken curve are the results of analysis using a rate equation model (see text). (b) Comparison of the photoemission spectral shapes measured at $\Delta t = -200$ fs (gray peak) and -30 fs (solid curve) under 1.57 eV excitation. The thick red (dark gray) curve shows the difference between the two spectra.

case of 1.26 eV excitation, the low-energy part of the spectrum is selectively reduced. This result may indicate that the hole injection into the initial Fermi distribution is also true for the 1.57 eV excitation.

One reason of the different temporal response in n_{CBM} between the different excitation energies is that the temporal width of the pump pulse is wider for the 1.57 eV excitation than for 1.26 eV excitation; the width of the cross-correlation trace between pump and probe pulses is 125 fs for 1.57 eV excitation, while the width is 82 fs in the 1.26 eV excitation. This partly results in poorer temporal resolution in the measurement for the fast process. Also, the wider pump pulses make the peak density of induced holes relatively weaker, as the lifetime of holes is shorter than the pulse width. More importantly, fast IMP can be induced by hot electrons injected from the LH band, which show a decay time (50 fs) comparable to that (40 fs) estimated for holes in the analysis in Fig. 8. Therefore, a rapid rise in n_{CBM} due to IMP from the short-lived hot electrons could compensate for the effect by the transient Auger effect, as seen in Fig. 9(b).

We analyzed the temporal changes in populations of hot electrons from the LH and SO bands and of the electrons near the CBM obtained for the 1.57 eV excitation, using essentially the same model as for 1.26 eV excitation. However, two modifications in the rate equation model were necessary. First, an equation was added to describe the temporal changes in the hot electrons injected from the SO band, as they are capable of inducing the direct IMP. As each IMP event caused by the hot electrons generates a hole in the LH band, three generation sources are included in the rate equation [Eq. (4)] describing the holes in the LH band. Second, as decay of the hot electrons from the LH band includes both IVS and the direct IMP, we introduced a term for the fraction of IMP over the total decay rate. We used the fraction of 0.39,

based on the magnitude of decay times shown in Appendix, assuming simply that the IMP rate is the same as that of the hot electrons, with $\varepsilon = 0.54$ eV for the case of 1.26 eV excitation. Using similar values of τ_h and γ from the previous analysis, we obtain a reasonable fit for the temporal changes in the population of hot electrons from the LH and SO bands and of electrons near the CBM; $\tau_h = 40$ fs and $\gamma = 5.5$ give the best fit in this case. The inclusion of the nonnegligible contribution of CBM growth from the slow process describes almost perfectly the observed initial enhancement of the CBM population as shown in Fig. 9(a). Similar analysis was made for results obtained at different excitation densities ranging from 1.5×10^{17} to 1.2×10^{18} cm^{-3} under 1.57 eV excitation. The observed results were well fit only by changing the generation rates of hot electrons from the SO and LH bands, substantiating our analysis using the rate-equation model.

As described above, the rate equations become complicated as more parameters are included in the model to fit the results obtained under 1.57 eV excitation. Therefore, the reliability of the estimates may be less than the case of 1.26 eV excitation. Nevertheless, we can confirm similar effects for the ultrafast recombination process in the case of 1.57 eV excitation in Fig. 9. Without such a recombination process, we would expect the CBM-population enhancement upon the decay of hot electrons from the SO and LH bands to follow the thin solid curve labeled $\gamma = 0$, which leads to a density of 9.0×10^{17} cm^{-3}. However, the observed enhancement is less than half of the predicted amount, showing a significant role of the ultrafast recombination process under excitation by 1.57 eV photons. We presume that the excited electrons generated by the transient Auger effect may be redistributed in the high-excess-energy region with a broad distribution, and they contribute to the IMP later in the process.

The lifetime of hot holes in the LH band, estimated by the equation, is 40 fs. This magnitude may be reasonable, as the lifetime is governed by strong longitudinal optical (LO) and transverse optical (TO) phonon scattering to the HH band and/or by energy relaxation within the LH band [50–52]. In Eqs. (3)–(5), the unit of density is 10^{16} cm^{-3}, and the unit of time is picoseconds. Therefore, $\gamma = 6$ corresponds to the absolute value of the Auger coefficient of 6×10^{-20} cm^6 s^{-1}, which is six orders of magnitude greater than the reported magnitude (5×10^{-26} cm^6 s^{-1}) of the Auger coefficient under thermal equilibrium conditions in pure InSb [53]. The lifetime of minority carriers of 5×10^{-8} s is estimated under the equilibrium conditions using the Auger coefficient. The recombination process we are studying, however, takes place in temporal domain within 10^{-13} s. Therefore, such a large difference in the magnitude of γ is necessary when we use the same concept of "Auger recombination."

In contrast to the equilibrium cases, the present process within a few hundred femtoseconds after excitation is under strongly nonequilibrium conditions, where holes in the LH valence band are generated at a significant density as deep as 0.5 eV below the valence band maximum (VBM). Also, other factors that govern the rate of the Auger process, e.g., screening effects [54] and the statistics of carriers, may be significantly different from equilibrium cases. Therefore, a substantial difference in the magnitude of the Auger coefficient could be expected. However, the transient Auger process introduced here remains hypothetical until further experimental and theoretical studies demonstrate the physical origin of the high rate under strongly nonequilibrium conditions.

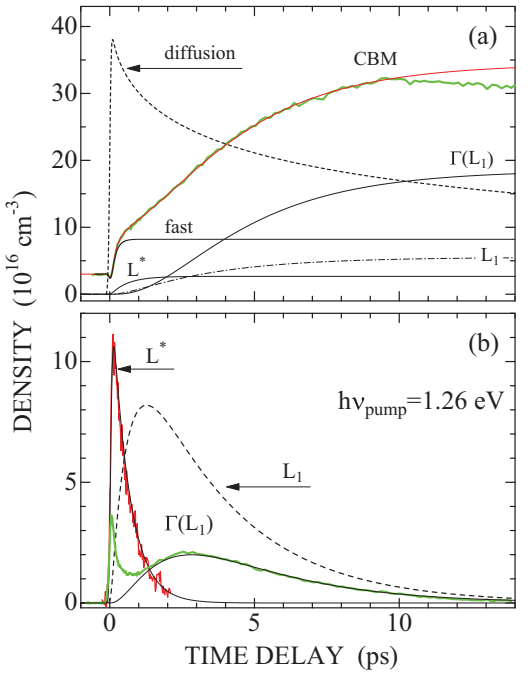

FIG. 10. (Color online) (a) Temporal changes in the electron population near the CBM, green (gray) curve, under 1.26 eV excitation. Thin solid curves are the results of analysis using a set of rate equations (see text), showing the contribution to the enhancement of the CBM population from the fast process, labeled fast, from the relaxing hot electrons in the L valley, L*, the hot electrons accumulated at the minimum of the L valley, L_1, and hot electrons backscattered from the L valley to the Γ valley, $\Gamma(L_1)$, respectively. The broken curve is the solution to the diffusion equation solved for 1.26 eV excitation. (b) Temporal changes in populations of hot electrons relaxing in the L valley, red (dark gray) curve labeled L*, and those populated at the $\Gamma(L_1)$ point in the Γ valley, green (gray) curve labeled $\Gamma(L_1)$. The thin solid curves are the results of the rate-equation analysis for temporal changes in the populations of relaxing electrons in the L valley and those at the $\Gamma(L_1)$. The broken curve is the predicted temporal change in the population at the L_1.

3. *The slow IMP process*

After the fast process of IMP, the populations near the CBM continue to increase to the maximum at ~10 ps after excitation. As clearly shown in Fig. 5, the energy relaxation of hot electrons by the e-ph interaction in the Γ valley is slow. Therefore, IMP still plays a dominant role in CBM-population enhancement in this slow process. For $\Delta t > 500$ fs, the major change in the hot-electron populations is the change in the L valley population injected directly and/or indirectly via efficient IVS. The decay of hot electrons in the L valley is associated with the formation of hot-electron packets in the Γ valley at the $\Gamma(L_1)$ as seen in Figs. 4 and 5.

Figure 10 shows the temporal changes in the population near the CBM (a), and in those of hot electrons in the L and Γ valleys (b), under 1.26 eV excitation. The hot electrons in the Γ valley are defined as electrons with $\varepsilon > 0.75$ eV. As the L

valley population decreases, hot electrons at $\Gamma(L_1)$ increase, and hot electrons at $\Gamma(L_1)$ decay within 10 ps. To quantitatively analyze the slow process, we introduce a set of rate equations for the densities n_L^* of relaxing hot electrons in the L valley, n_L at the minimum of the L valley, n_Γ at the $\Gamma(L_1)$, and n_{CBM} at the CBM:

$$\frac{dn_L^*}{dt} = G(t,t_0,w) - k_L^* n_L^*, \tag{6}$$

$$\frac{dn_L}{dt} = (1-\eta)k_L^* n_L^* - k_L n_L \tag{7}$$

$$\frac{dn_\Gamma}{dt} = (1-\varsigma)k_L n_L - k_\Gamma n_\Gamma, \tag{8}$$

$$\frac{dn_{CBM}}{dt} = 2\eta k_L^* n_L^* + 2\varsigma k_L n_L + 2k_\Gamma n_\Gamma, \tag{9}$$

where $G(t,t_0,w)$ is the generation function of hot electrons in the L valley, k_L^*, k_L, and k_Γ are the total decay rate of n_L^* n_L, and n_Γ, and η is the fraction of the IMP rate of the relaxing hot electrons in the L valley, and ς is the fraction of the IMP for hot electrons accumulated at the minimum of the L valley.

As the lifetime of hot holes in the LH valence band has been estimated to be as short as 40 fs, we do not consider any effects of the transient Auger recombination in the slow process. The hot electrons in the L valley are generated by the IVS of hot electrons injected into the Γ valley, as described in the Sec. III B. The scattering time is as short as 44 fs. Therefore, we approximated the generation of the hot electrons in the L valley by introducing a generation function for simplicity.

Although Eqs. (6)–(9) include many parameters, we can determine the parameters almost uniquely, as the solutions must describe quantitatively not only the temporal changes but relative intensities of respective states. The width and time delay in the generation function are determined to obtain the best fit to the experimental results shown in Fig. 10(b). The decay of hot electrons in the L valley can be determined almost entirely by the decay characteristic of the population; it is characterized well by a single time constant of 590 fs. The temporal changes in the hot-electron population at the $\Gamma(L_1)$ is well fitted when we assume that $1/k_\Gamma = 1.2$ ps and $1/k_L = 3.8$ ps, as shown in Fig. 10(b). The broken curve shows the predicted changes in the population at the minimum of the L valley.

For the enhancement of the electrons near the CBM, the contribution from the third term in the right side of Eq. (9) is determined uniquely from the magnitude of k_Γ, and it gives the largest contribution to the CBM-population enhancement. Although the qualitative feature of the slow process can be described by this term only, quantitative agreement is not obtained when we assumed $\eta = \varsigma = 0$. The best fits of the contributions of each term in Eq. (9) are shown by thin solid curves in Fig. 10(a). The first term coming from the relaxing hot electrons in the L valley contributes to the enhancement at a temporal domain of less than 2 ps, while the second term gives a contribution in the latest temporal region due to the long decay time of electrons at the L_1. Adding the enhancement from the fast process previously determined gives an almost perfect fit to the experimental results for $\Delta t < 10$ ps, as shown by red curve in the figure. For the results obtained under 1.57 eV excitation, a similar analysis was conducted using Eqs. (6)–(9). We obtained a reasonable fit to experimental results, similar to those results for 1.26 eV excitation.

As seen in the results of the analysis shown in Fig. 10, the relaxing hot electrons in the L valley and those at the L_1 give a finite contribution to the CBM-electron enhancement via IMP. In the case of 1.26 eV excitation, hot electrons from the HH band are scattered into the L valley at the position with excess energy of 0.85 eV. This energy is lower than the energy ($\sim$1.0 eV) at the critical point for direct IMP via the HH-band channel shown in Fig. 7. Therefore, they are unable to induce direct IMP. However, hot electrons are populated in the L valley near the critical point under 1.57 eV excitation, as imaged in Fig. 1. Therefore, significant differences in the temporal changes in the CBM-population enhancement could be expected if direct IMP via the HH-band channel is effective. However, temporal changes in the fast processes are similar to each other, and they are well characterized using the same set of rate equations, which neglect any contribution of the HH-band channel of IMP. This result may indicate that the contribution of the HH-band channel of IMP by hot electrons in the L valley is not efficient, compared with that of the LH-band channel of IMP by hot electrons in the Γ valley. This may substantiate our quantitative analysis of the temporal evolution of n_{CBM} in the fast process using Eqs. (6)–(9).

In the IMP process of the relaxing hot electrons below the critical point in the L valley, and those accumulated near the minimum of the L valley, momentum conservation among electronic states can no longer be satisfied. Therefore, the IMP process is phonon-assisted [16], with smaller rates than for the direct IMP process. On the other hand, hot electrons accumulated at the $\Gamma(L_1)$ by backscattering from the L valley have a long lifetime of 1.2 ps, although they are placed at the momentum and energy positions similar to those of photoinjected hot electrons from the LH band, which have a high rate of IMP. As seen in Fig. 5, the photoemission peak from electrons near the CBM shows a maximum at 0.03 eV above the CBM before excitation. However, the peak moves to the higher-energy region at 0.07 eV with a strong enhancement of the intensity at $\Delta t = 300$ fs; originally empty states near the CBM are already occupied by electrons from the fast IMP process when hot electrons near the $\Gamma(L_1)$ are generated. Therefore, we presume that the slow IMP rate of hot electrons at the $\Gamma(L_1)$ is due to Pauli blocking for available states near the CBM in the IMP processes, although a quantitative analysis of the Pauli blocking factor requires a more precise spectral-shape analysis using a deconvolution procedure.

In the analysis described thus far, we completely ignored the possible effect of carrier diffusion. In InSb, the mobility of electrons near the CBM is high, and the absorption coefficient at 1.26 eV in InSb is estimated to be 5.8×10^5 cm^{-1} from compiled data of dielectric functions [55]. Under photoexcitation, carrier diffusion is governed by the ambipolar diffusion constant D_a. The diffusivity of carriers depends on the carrier density [56]. However, for excitation density less than 10^{19} cm^{-3} at room temperature, any high-density effects on D_a can be ignored. The reported diffusion constant D_e for electrons in the CB is $D_e \leqslant 2.0 \times 10^3$ cm^2 s^{-1}, while D_h for holes in the valence band is $D_h \leqslant 22$ cm^2 s^{-1} [53].

Therefore, D_a is $\leqslant 44\,\mathrm{cm}^2\,\mathrm{s}^{-1}$. To examine the possible effects of carrier diffusion from our probed region (∼10 nm from the surface) into the bulk, we calculated the diffusion rate, using an absorption coefficient and the ambipolar diffusion constant estimated above [57]. The result is shown by the broken curve, labeled diffusion, in Fig. 10(a). The maximum electron density injected by a 100 fs light pulse is reduced to 90% at 400 fs, and to 50% at 7 ps, showing sizable effects. In fact, it has been reported that the population loss of photogenerated electrons, when probed by photoemission technique, is governed by diffusion at Δt longer than several picoseconds after excitation [36].

However, the effects of carrier diffusion in the present case are complex. The diffusion constant for the electrons used in the diffusion equation reflects transport properties of electrons near the CBM. However, photoinjected hot electrons have distributions far from equilibrium, and hot electrons with different excess energies have different diffusivities. In particular, those in the L valley have a significantly smaller diffusion constant due to a heavier effective mass than those in the Γ valley [34]. Furthermore, electrons near the CBM are not instantaneously generated as with diffusion-equation calculation, but rather have time-dependent generation rates as shown in Fig. 6. Therefore, it is not currently possible to include the effects of carrier diffusion in the quantitative analysis of the IMP processes we discussed so far. For the fast process of IMP, possible errors due to ignoring carrier diffusion can be neglected, since the fast process is terminated within 400 fs. However, for the slow process, the determined magnitudes of physical parameters in Fig. 10 are regarded as those convolving possible effects of electron diffusion from our probed region into the bulk.

4. The rate of IMP in InSb

As described in the previous sections, hot-electron packets photoinjected into the CB from the three valence bands in InSb relax almost independently from each other along pathways that are sensitively dependent on the location specified by the momentum and energy. In Fig. 11, we summarize the relaxation rates, defined as the inverse of the decay times, of hot-electron packets injected by excitation with photon energies from 1.20 eV to 1.57 eV as a function of excess energy. It is clear that there is a critical energy of $\varepsilon = 0.63\,\mathrm{eV}$, the minimum energy of the L valley, above which relaxation rate becomes drastically higher with increasing excess energy. The results show that hot electrons injected above the minimum of the L valley undergo efficient IVS into the L valley.

The rate of IVS between valleys i and j by the deformation-potential interaction is given by

$$W_{ij}(\varepsilon) = \frac{\pi D_{ij}^2}{m\omega_{ij}}\{n(\omega_{ij})N(\varepsilon - \Delta E_{ij} + \hbar\omega_{ij}) + (n(\omega_{ij}) + 1)N(\varepsilon - \Delta E_{ij} - \hbar\omega_{ij})\} \quad (10)$$

where D_{ij} is an intervalley deformation potential, $\hbar\omega_{ij}$ is the intervalley phonon energy, $n(\omega_{ij})$ is the occupation number of intervalley phonons, ΔE_{ij} is the energy difference of the valley minima, $N(\varepsilon - \Delta E_{ij} \pm \hbar\omega_{ij})$ is the density of final states of scattering, and m is the mass density [15]. For the IVS from the Γ to the L valley, both LO and longitudinal acoustic (LA) phonons can contribute to the transition [15,58]. However, theoretical analysis predicts that LA phonons contribute predominantly to the deformation potential constant [58]. Although wave numbers of intervalley phonons depend on the excess energy of hot electrons, as seen in the images in Figs. 1(b) and 1(c), we assume for simplicity that the zone-edge phonons play the dominant role in the IVS. Then $\hbar\omega_{ij} = 16\,\mathrm{meV}$ [59], and $n(\omega_{ij}) = 1.13$ at 293 K. Under this assumption, we can calculate $N(\varepsilon - \Delta E_{ij} \pm \hbar\omega_{ij})$, using the density of state mass for the L valley [34], and we evaluate the IVS rate as a function of excess energy ε. The broken curve, labeled W_{IVS} in Fig. 11, shows thus evaluated IVS rate with material constants of InSb. The magnitude of $D_{ij} = 2.8\,\mathrm{eV/Å}$ was determined to fit the observed results of transition rates. The thin solid curve above $\varepsilon = 0.63\,\mathrm{eV}$ is a sum of the W_{IVS} and the decay rate of hot electrons having excess energy just below 0.63 eV. The almost perfect agreement of the energy-dependent relaxation rate above $\varepsilon > 0.63\,\mathrm{eV}$ substantiates our conclusion that hot electrons injected above

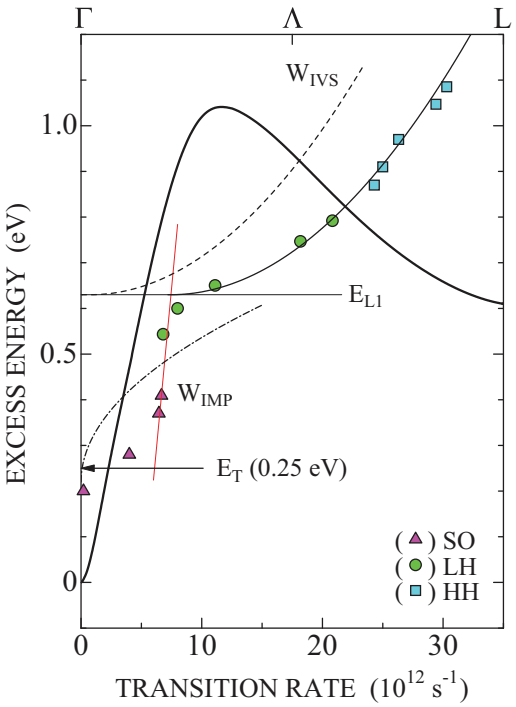

FIG. 11. (Color online) Relationship between excess energy and transition rate, defined by the inverse of the decay time, of hot electrons. Decay times were analyzed for several hot-electron packets injected into the Γ valley from the HH, LH, and SO valence bands, as plotted using different symbols. The solid thick curve is the dispersion of the CB along the Γ-L direction (after Ref. [42]). The broken curve, labeled W_{IVS}, shows the evaluated IVS rate. The thin solid curve above $\varepsilon = 0.63\,\mathrm{eV}$ is the sum of the W_{IVS} and the decay rate of hot electrons having excess energy just below 0.63 eV. The chain curve shows the quadratic relation of the IMP rate on the excess energy above the threshold energy; an appropriate constant is multiplied to obtain the magnitude in the transition-rate range.

the minimum of the L valley undergo efficient IVS into the L valley.

The analysis on the IVS rate given above confirms that the hot electrons injected below the L_1 are free from the IVS. These hot electrons show an almost constant decay rate of $7.0 \times 10^{12}\,\mathrm{s}^{-1}$ for ε from 0.35 to 0.6 eV. The hot electrons with the lowest excess energy of 0.28 eV, which are injected from the SO band by 1.37 eV photons, have a smaller rate of decay. This may be reasonable because the excess energy is very close to the threshold energy of 0.25 eV for IMP. Radiative and phonon-assisted nonradiative recombination occurs at much slower rates [48], and we have demonstrated above that decay of these hot electrons is responsible for the fast process of IMP. Therefore, these results show definitively that the direct IMP rate is $7.0 \times 10^{12}\,\mathrm{s}^{-1}$ for hot electrons with ε from 0.35 eV to 0.6 eV.

The rate of IMP has been studied theoretically using several different methods. An analytical formula has been obtained for the electrons near the threshold of IMP in a parabolic system:

$$W_{IMP} = W_0 \left(\frac{\varepsilon_0}{\varepsilon_\infty}\right)^2 \frac{m_c^*}{m_0} \frac{I_c^2 I_v^2}{(1+2\mu)^{3/2}} \left(\frac{\varepsilon - E_T}{E_g}\right)^2, \qquad (11)$$

where $W_0 = 4.14 \times 10^{16}\,\mathrm{s}^{-1}$, ε_0 is the free-space permittivity, ε_∞ is the high-frequency permittivity, m_c^* is the electron effective mass, $\mu = m_c^*/m_h^*$, m_h^* is the hole effective mass, I_c and I_v are the overlap integrals of cell-periodic parts of Bloch functions for e-e and electron-hole states, E_T is the threshold energy, and E_g is the band-gap energy [15,60]. In a parabolic system, E_T is given by $E_T = E_g(1+2\mu)/(1+\mu)$ [15]. Using m_h^* of the LH (HH) valence band in InSb [34], we estimate $E_T = 0.25\,(0.18)$ eV at 293 K. When we apply this formula to hot electrons with $\varepsilon = 0.5$ eV in InSb, we have $W_{IMP} = 1.1 \times 10^{12}\,(1.8 \times 10^{12})\,\mathrm{s}^{-1}$, even for $I_c^2 = I_v^2 = 1$, establishing an upper bound. Given that Eq. (11) neglects any screening effects [15,60], theory underestimates the rate significantly. We presume that this discrepancy comes partly from the parabolic-band approximation in Eq. (11). In fact, a Monte Carlo simulation has shown that inclusion of nonparabolicity dramatically enhances the IMP rate in InSb [23].

The rate of IMP in InSb has been evaluated using different theoretical approaches. Using a Monte Carlo simulation technique, Curby and Ferry [23] reported the rate of $1 \times 10^{13}\,\mathrm{s}^{-1}$ for hot electrons with an energy of 0.3 eV. The screening effect in electron collisions in IMP was taken into account using a Debye screening model. However, in their calculation, the overlap integral [I_v^2 in Eq. (11)] corresponding to the interband transition was evaluated for the CB and the HH valence band, based essentially on $\boldsymbol{k}.\boldsymbol{p}$ theory and f-sum rules. As discussed in details in Refs. [61,62], this method overestimates I_v^2 by a factor of 10^2. Therefore, the calculated rate of $1 \times 10^{13}\,\mathrm{s}^{-1}$ may be overestimated. Interestingly, recent Monte Carlo calculations [33] reported the rate of $1 \times 10^{10}\,\mathrm{s}^{-1}$ for electrons with energy 0.3 eV above E_T (=0.225 eV), although I_v^2 was evaluated by the same method as Ref. [23]. On the other hand, Beattie *et al.* [26] have evaluated the rate of IMP in InSb based on an accurate Bloch-function overlap integral for realistic band structures. They reported the rate of $4 \times 10^{10}\,\mathrm{s}^{-1}$ for hot electrons with energy of 0.3 eV above E_T (=0.268 eV). Thus, the rate obtained by previous theoretical studies is smaller by a factor of about 10^2 than the rate determined in this paper.

In these theoretical studies, the interaction between the impacting hot electron in the CB and the electrons at the HH valence band has been mainly considered to evaluate the rate of the HH-band channel shown in Fig. 7. Because of the symmetry, I_v^2 is small [61,62], although the density of state of the HH band is higher than that of the LH valence band. However, the LH-band channel can have I_v^2 close to unity [15,61,62]. More importantly, as seen in Fig. 7, the dispersion of the conduction band and LH band is not parabolic, but can be well approximated by a linear dispersion in the wide energy range from -0.5 to 1.0 eV, excluding the gap region, as shown by the chain line. Dispersions along the Γ-X and Γ-K directions show the same characteristic [42]. Therefore, momentum and energy conservations can be satisfied for the four quantum states involved in the LH-channel of the direct IMP process, without any additional limitations to the initial impacting electronic state except $\varepsilon > E_T$. These factors may enhance the IMP rate via the LH-band channel to contribute significantly to the direct IMP rate, despite the smaller density of state.

Apart from quantitative comparison between the theoretical results and the present experimental results of IMP, the quadratic dependence of the IVS rate on the excess energy has been generally assumed to predict the IMP rate. The energy dependence of the IMP rates calculated using a variety of methods gives similar quasiquadratic dependence on the excess energy near the threshold [23–25,30,31]. In Fig. 11, we show the predicted relation, with $E_T = 0.25$ eV, by the chain curve. It is evident that the theoretically predicted relation is qualitatively different from the experimental results at energies higher than E_T; the rate is more strongly enhanced than the quadratic relationship for the low ε region, and then it stays almost constant. The discrepancy may show that for excess energies of a few hundred millielectronvolts, the quadratic relationship between the IMP rate and excess energy is no longer valid. However, in a recent theoretical study on IMP rate using first principle calculations [32], it has been shown that the IMP rate in InAs, a narrow gap semiconductor like InSb, is much higher in the low-excess-energy region than calculated previously, which indicated quasiquadratic dependence on the excess energy. They ascribed this effect to the k-dependent matrix element, which was assumed to be constant in previous studies. The inclusion of such a low-energy contribution may change the energy dependence of the IMP rate substantially; our results may reflect this effect.

IV. SUMMARY

We have studied the ultrafast relaxation of hot electrons photogenerated in InSb, using time- and angle-resolved photoemission spectroscopy. The nascent distributions of hot-electron packets injected by the interband optical transitions were clearly resolved in energy and momentum spaces, and the relaxation processes were directly followed at femtosecond temporal resolution on a state-resolved basis. In particular,

it has been shown that hot electrons injected in the central Γ valley with energies above the minimum of the L valley (determined to be 0.63 eV in the present study) undergo ultrafast IVS into the L valley with transition times of the order of 40 fs, depending on the final density of state in the scattering. The relaxation processes of the hot electrons transferred into the L valley have been directly captured in momentum and energy spaces. They are relaxed within the L valley, at a finite backscattering rate into the Γ valley during relaxation, with a relaxation rate of 0.42 to 0.78 ps, depending on the excess energy.

Alternatively, the hot electrons injected in the Γ valley with an excess energy below the minimum of the L valley induce direct IMP at a rate of $7 \times 10^{12}\ s^{-1}$ at the excess energy range from 0.35 to 0.6 eV. The rate of direct IMP determined experimentally in this paper is significantly higher than previous theoretical predictions. The strong nonparabolicity of the CB in InSb has been emphasized as an important factor that enhances the rate. Also, the results suggest the crucial role of the interaction of hot electrons with the LH valence states in the direct IMP process. Importantly, it has been proposed that a transient Auger recombination between the electrons at the CBM and holes injected in the LH valence band significantly suppresses the yield of low-energy electrons generated in the IMP process. Further, we have shown that phonon-assisted IMP processes by hot electrons in the L valley significantly contribute to the enhancement of electron populations near the CBM, although the rate is smaller than for direct IMP processes.

Thus, state-selective features in IMP processes have been demonstrated for the first time. Our paper, based on direct measurements of energy- and momentum-resolved electron distributions, has provided deeper insight into the ultrafast scattering processes of hot electrons in semiconductors.

ACKNOWLEDGMENTS

This paper was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant No. 24000006.

APPENDIX

1. The dispersion of the CB as a function of emission angle of photoemission

In the photoemission processes from crystalline samples, $k_{//}$ is conserved [44]. Therefore, for a photoelectron with kinetic energy E_K, ionized from a state with energy ε referenced to the CBM by probe light with photon energy $h\nu$, the following equation holds:

$$k_{//} = (\sqrt{2m_0}/\hbar)\sqrt{\varepsilon + E_G + h\nu - \Phi_{vac}}\,\sin\theta, \qquad \text{(A1)}$$

where m_0 is the electron rest mass, $\hbar$ is Planck's constant divided by 2π, θ is the emission angle, Φ_{vac} is an ionization energy, and E_G is the band-gap energy. In Eq. (A1), E_G and Φ_{vac} are materials constants, and $h\nu$ is determined experimentally. Band structure calculations give ε as a function of the wave vector $k_X(k_L)$ along the Γ-X(Γ-L) direction. Therefore, the dispersion of the CB along the Γ-L and Γ-X directions can be calculated as a function θ. A constant ionization potential is assumed for any θ.

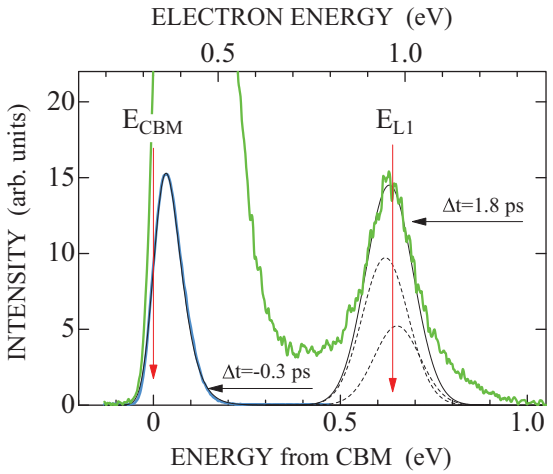

FIG. 12. (Color online) Photoemission spectra at −0.5 ps and at 1.8 ps after excitation with 1.57 eV laser pulses, and the results of spectral-shape analysis. The upper scale represents the kinetic energy of photoelectrons referenced to the work function of the analyzer, while the bottom scale shows the energy from the CBM. Thin solid and broken curves show the results of spectral-shape analysis. The vertical arrows indicate the position of the CBM, E_{CBM}, and the minimum of the L valley, E_{L1}.

2. Quantitative analysis of photoemission spectra with a finite energy resolution

a. Determination of the CBM level in the kinetic-energy spectra and the initial electron density

The intrinsic carrier density in InSb has been reported as $2.0 \times 10^{16}\ cm^{-3}$ [45]. Since the effective mass of the CBM is $0.014m_0$ [34], this density can bring the Fermi level E_F up to nearly the CBM. Therefore, the Fermi distribution function is more appropriate to describe electron distribution $\rho(e)$ in the parabolic band with an effective mass of $0.014m_0$, where ε is the excess energy with reference to the CBM. In the spectra plotted as a function of electron kinetic energy E_K referenced to the work function of our analyzer, $\rho(e)$ is given by

$$\rho(e) = A(m^*/m_0)^{3/2}\sqrt{\varepsilon}[1 + \exp\{(E_K - E_{CBM} - E_F)/k_B T_e\}]^{-1}, \qquad \text{(A2)}$$

where A is a constant, E_{CBM} is the energy of the CBM, T_e is the electron temperature, and k_B is the Boltzmann constant. The convolution of (A2) with a finite energy resolution of 50 meV describes the observed photoemission spectra, with parameters of E_{CBM} and E_F (T_e can be set at 293 K). In Fig. 12, the angle-integrated emission spectrum measured at $\Delta t = -0.5$ ps is well fit by the function with $E_F = 0.01$ eV and $E_{CBM} = 0.32$ eV.

b. Determination of the energy of the $\mathbf{L_1}$

To precisely determine the energy E_{L1} of the L_1 point, we analyzed the spectral shape of the peak around $\varepsilon = 0.6$ eV. For IVS from the L to Γ valleys, LA phonons contribute predominantly to the deformation potential constant [58].

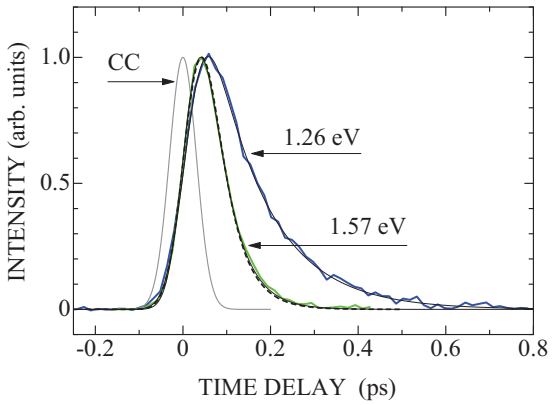

FIG. 13. (Color online) Temporal changes in the photoemission intensities from hot electrons injected at excess energies of 0.79 eV and 0.54 eV by 1.57 eV and 1.26 eV pump pulses, respectively. The gray curve, labeled CC, shows the cross-correlation trace between pump and probe pulses, which defines the zero time delay. The analysis for a whole temporal domain including rise time was made using the optical Bloch equation, and the results are shown by thin solid curves. The thick broken curve shows the results of simulation using a rate equation model with a generation function.

Zone-edge LA phonons have energy $\hbar\omega_{LA} = 16\,\text{meV}$ [59]; the phonon occupation number $n(\omega_{LA}) = 1.13$ at 293 K. Then, the final states of backscattering in the Γ valley have a peak energy of E_{L1}-$\hbar\omega_{LA}$ with intensity proportional to $[1 + n(\omega_{LA})]$ for the phonon-emitting transition, and peak energy of $E_{L1} + \hbar\omega_{LA}$ with intensity proportional to $n(\omega_{LA})$ in the phonon-absorbing process. Because of the finite energy resolution of 50 meV in the analysis, each component was broadened using Gaussian bands of an appropriate width, but the relative intensities were fixed to the ratio determined by $n(\omega_{LA})$ and $[1 + n(\omega_{LA})]$. In Fig. 12, the photoemission spectrum measured at $\Delta t = 1.8\,\text{ps}$ is displayed, and the spectral shapes of the 0.6 eV peak are compared with the results of spectral-shape analysis. The best fit is obtained when we choose $E_{L1} = 0.63\,\text{eV}$, as seen in the figure.

3. Method of analysis of ultrafast temporal changes in hot-electron populations

As mentioned in the Experimental methods section, the temporal resolution of the present study is limited by finite temporal widths of pump and probe pulses. In the analysis of ultrafast temporal changes characterized by time constants comparable to (or even shorter than) the cross-correlation widths, we used the optical Bloch equations for a two-level system to consider the effects of a finite polarization dephasing time T_2 [38,39,63]. In Fig. 13, we show the temporal changes in the hot-electron populations injected from the LH bands under 1.26 eV and 1.57 eV excitation, and the results of analysis using the optical Bloch equation.

The relaxation of a photogenerated nonequilibrium electron distribution is a complicated many-body problem. In most cases, temporal changes of a given state are not determined simply by the population decay time of the state, but are also governed by changes in other states. Therefore, we need to analyze temporal changes of a few different states simultaneously; the optical Bloch equations cannot be used in such cases. To make quantitative analysis possible for such cascade processes, we use the set of rate equations by introducing a generation function $G(t,t_0,w)$, which can simulate the solution of optical Bloch equations including effects of polarization dephasing. In the case of the results shown in Fig. 13, the solutions of the optical Bloch equation for the hot electrons injected from the LH bands are well simulated by the following rate equation;

$$\begin{aligned} \frac{d\rho_{22}}{dt} &= G(t,t_0,w) - \frac{\rho_{22}}{T_1}; \\ G(t,t_0,w) &= I_{\text{pump}} \exp\{-(t-t_0)^2/w^2\}, \end{aligned} \tag{A3}$$

using appropriately chosen parameters of w and t_0. The broken curve in Fig. 13 is the solution of this equation for the case of 1.57 eV excitation, with pulse width 11% wider than the experimentally determined one and an additional time delay $t_0 = 14\,\text{fs}$. This approximation is rather artificial, but it allows us to use a set of rate equations to correlate temporal changes in several states while keeping zero delay the same.

[1] J. Shah, *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures*, 2nd ed. (Springer, Berlin, 1999).
[2] F. Rossi and T. Kuhn, Rev. Mod. Phys. **74**, 895 (2002).
[3] R. P. Prasankumar, P. C. Upadhya, and A. J. Taylor, Phys. Stat. Sol. B **246**, 1973 (2009).
[4] B. Fischer and K. R. Hofmann, Appl. Phys. Lett. **76**, 583 (2000).
[5] M. Hase, M. Kitajima, A. M. Constantinescu, and H. Petek, Nature **426**, 51 (2003).
[6] G. C. W. Cho, W. Kütt, and H. Kurz, Phys. Rev. Lett. **65**, 764 (1990).
[7] T. Pfeifer, W. Kutt, H. Kurz, and R. Scholz, Phys. Rev. Lett. **69**, 3248 (1992).
[8] H. J. Zeiger, J. Vidal, T. K. Cheng, E. P. Ippen, G. Dresselhaus, and M. S. Dresselhaus, Phys. Rev. B **45**, 768 (1992).
[9] A. Rousse, C. Rischel, S. Fourmaux, I. Uschmann, S. Sebban, G. Grillon, Ph. Balcou, E. Förster, J. P. Geindre, P. Audebert, J. C. Gauthier, and D. Hulin, Nature **410**, 65 (2001).
[10] A. M. Lindenberg, J. Larsson, K. Sokolowski-Tinten, K. J. Gaffney, C. Blome, O. Synnergren, J. Sheppard, C. Caleman, A. G. MacPhee, D. Weinstein, D. P. Lowney, T. K. Allison, T. Matthews, R. W. Falcone, A. L. Cavalieri, D. M. Fritz, S. H. Lee, P. H. Bucksbaum, D. A. Reis, J. Rudati *et al.*, Science **308**, 392 (2005).
[11] P. B. Hillyard, K. J. Gaffney, A. M. Lindenberg, S. Engemann, R. A. Akre, J. Arthur, C. Blome, P. H. Bucksbaum, A. L. Cavalieri, A. Deb, R. W. Falcone, D. M. Fritz, P. H. Fuoss, J. Hajdu, P. Krejcik, J. Larsson, S. H. Lee, D. A. Meyer, A. J. Nelson, R. Pahl *et al.*, Phys. Rev. Lett. **98**, 125501 (2007).
[12] D. W. Snoke, W. W. Rühle, Y.-C. Lu, and E. Bauser, Phys. Rev. B **45**, 10979 (1992).
[13] E. D. Grann, K. T. Tsen, O. F. Sankey, D. K. Ferry, A. Salvador, A. Botcharev, and H. Morkoç, Appl. Phys. Lett. **67**, 1760 (1995).
[14] T. Ichibayashi and K. Tanimura, Phys. Rev. Lett. **102**, 087403 (2009).

[15] B. K. Ridley, *Quantum Processes in Semiconductors* (Oxford University Press, Oxford, 1999).
[16] D. J. Robbins, Phys. Stat. Sol. B **97**, 9 (1980).
[17] D. K. Ferry, *Semiconductor Transport* (Taylor & Francis, London, 2000).
[18] P. Würfel, Solar Ener. Mat. Solar Cells **46**, 43 (1997).
[19] R. D. Schaller and V. I. Klimov, Phys. Rev. Lett. **92**, 186601 (2004).
[20] B. Ancker-Johnson, in *Plasmas in Semiconductors and Semimetals*, edited by R. C. Willardson and A. C. Beer (New York, Academic Press, 1967), Vol. I.
[21] W. Shockley, Solid-State Electron. **2**, 35 (1961).
[22] E. A. Eklund, P. D. Kirchner, D. K. Shuh, F. R. McFeely, and E. Cartier, Phys. Rev. Lett. **68**, 831 (1992).
[23] R. C. Curby and D. K. Ferry, Phys. Stat. Sol. A **15**, 319 (1973).
[24] R. C. Alig, S. Bloom, and C. W. Struk, Phys. Rev. B **22**, 5565 (1980).
[25] J. Geist and W. K. Gladden, Phys. Rev. B **27**, 4833 (1983).
[26] A. R. Beattie, R. A. Abram, and P. Scharoch, Semicond. Sci. Technol. **5**, 738 (1990).
[27] H. Wen, M. Wiczer, and A. M. Lindenberg, Phys. Rev. B **78**, 125203 (2008).
[28] M. C. Hoffmann, J. Hebling, H. Y. Hwang, K.-L. Yeh, and K. A. Nelson, Phys. Rev. B **79**, 161201(R) (2009).
[29] I-Chen Ho and X.-C. Zhanga, Appl. Phys. Lett. **98**, 241908 (2011)
[30] N. Sano and A. Yoshii, J. Appl. Phys. **77**, 2020 (1995).
[31] S. Picozzi, R. Asahi, C. B. Geller, A. Continenza, and A. J. Freeman, Phys. Rev. B **65**, 113206 (2002).
[32] T. Kotani and M. van Schilfgaarde, Phys. Rev. B **81**, 125201 (2010).
[33] S. Asmontaš, R. Raguotis, and S. Bumelienė, Semicond. Sci. Technol. **28**, 025019 (2013).
[34] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, J. Appl. Phys. **89**, 5815 (2001).
[35] I. L. Shumay and U. Höfer, Phys. Rev. B **53**, 15878 (1996).
[36] H. S. Karlsson and U. O. Karlsson, Phys. Rev. B **59**, 10796 (1999).
[37] A. Lobad and L. A. (Vern) Schlie, J. Appl. Phys. **95**, 97 (2004).
[38] H. Tanimura, J. Kanasaki, and K. Tanimura, Sci. Rep. **4**, 6849 (2014).
[39] J. Kanasaki, H. Tanimura, and K. Tanimura, Phys. Rev. Lett. **113**, 237401 (2014).
[40] J. Kanasaki, E. Inami, and K. Tanimura, Surf. Sci. **601**, 2367 (2007).
[41] G. W. Gobeli and F. G. Allen, Phys. Rev. **137**, A245 (1965).
[42] B. D. Malone and M. L. Cohen, J. Phys.: Condens. Matter **25** 105503 (2013).
[43] D. W. Niles, D. Rioux, and H. Höchst, Phys. Rev. B **46**, 12547 (1992).
[44] E. W. Plummer and W. Eberhardt, Adv. Chem. Phys. **49**, 533 (1982).
[45] R. W. Cunningham and J. B. Gruber, J. Appl. Phys. **41**, 1804 (1970).
[46] Because of the inhomogeneity in the samples, initial electron densities fluctuated depending on the position of samples; the smallest density is $2.1 \times 10^{16}\,\mathrm{cm}^{-3}$ and the largest is $3.1 \times 10^{16}\,\mathrm{cm}^{-3}$, for an average of $2.6 \pm 0.5 \times 10^{16}\,\mathrm{cm}^{-3}$ for the initial electron densities. The density we determined is a little larger than the literature value. One reason may come from the experimental conditions using a repetition rate of 76 MHz for pump laser pulses. The pump-pulse interval is 13 ns, which is not long enough for the electronic system photoexcited by a preceding pulse to recover completely to a thermally equilibrated state. The initial density before pump was evaluated by the spectral-shape analysis, and this value was used in correlating the photoemission intensity to the density of electrons for quantitative analysis. Throughout the paper, the same scaling factor to convert the integrated photoemission intensity into the electron density was assumed for hot-electron states in both the L and Γ valleys.
[47] M. Lundstorm, *Fundamentals of Carrier Transport* (Cambridge University Press, London, 2002).
[48] R. D. Grober, H. D. Drew, G. L. Burdge, and B. S. Bennett, J. Appl. Phys. **71**, 5140 (1992).
[49] In the rate equations, we do not include any contribution of IMP of the HH-band channel, which involves electron interaction with the HH band state. One reason is, as discussed above, that the photoinjected holes in the HH band are not relevant to the transient Auger effect. Another reason for this exclusion is to simplify the analysis. Clearly, the analysis could be made arbitrarily complex, including all possible effects, but a more fruitful approach is to begin with simple models, which include the most essential factors, and add complexity as needed. By neglecting the possible effects coming from the hot-electron interaction with the HH states, we can reduce the number of parameters to only two in the fitting procedure.
[50] W. Potz and P. Kocevar, Phys. Rev. B **28**, 7040 (1983).
[51] M. A. Osman and D. K. Ferry, Phys. Rev. B **36**, 6018 (1987).
[52] C. J. Stanton, D. W. Bailey, and K. Hess, Phys. Rev. Lett. **65**, 231 (1990).
[53] The Auger recombination coefficient, and diffusion constants of electrons and holes in InSb were taken from the electronic archive of Ioffe Physico-Technical Institute, http://www.ioffe.rssi.ru/SVA/NSM/.
[54] R. Huber, F. Tauser, A. Brodschelm, M. Bichler, G. Abstreiter, and A. Leitenstorfer, Nature **414**, 286 (2001).
[55] S. Adachi, J. Appl. Phys. **61**, 4869 (1987).
[56] J. F. Young and H. M. van Driel, Phys. Rev. B **26**, 2147 (1982).
[57] For InSb, with well-ordered surfaces obtained by UHV cleavage, surface-recombination velocity is small, as discussed in Ref. [36]. Also, the Auger recombination rate is slow enough ($<10^{10}\,\mathrm{s}^{-1}$) for the excitation densities less than $10^{18}\,\mathrm{cm}^{-3}$. We neglected both effects in the diffusion-equation calculation.
[58] S. Zollner, S. Gopalan, and M. Cordona, Appl. Phys. Lett. **54**, 614 (1989).
[59] D. L. Price, J. M. Rowe, and R. M. Nicklow, Phys. Rev. B **3**, 1268 (1971).
[60] B. K. Ridley, J. Phys.: Condens. Matter **10**, L607 (1998).
[61] S. Brand and R. A. Abram, J. Phys. C: Solid State Phys. **17**, L201 (1984).
[62] M. G. Burt, S. Brand, C. Smith, and R. A. Abram, J. Phys. C: Solid State Phys. **17**, 6385 (1984).
[63] T. Hertel, E. Knoesel, M. Wolf, and G. Ertl, Phys. Rev. Lett. **76**, 535 (1996).