Large nonlinear refraction in InSb at 10 μm and the effects of Auger recombination

V. Dubikovskiy, D. J. Hagan,* and E. W. Van Stryland

CREOL, The College of Optics and Photonics, and the Department of Physics, University of Central Florida, Orlando, Florida 32816-2700, USA
**Corresponding author: hagan@creol.ucf.edu*

Received October 15, 2007; revised December 11, 2007; accepted December 12, 2007; posted December 20, 2007 (Doc. ID 88576); published January 29, 2008

Narrow bandgap semiconductors exhibit very large optical nonlinearities in the infrared owing to large two-photon absorption that scales as the inverse cube of the bandgap energy and the large losses and refraction from two-photon generated free carriers. Except for extremely short pulses, the free-carrier effects dominate the nonlinear losses and nonlinear refraction. Here we develop a method for the calculation of the free-electron refraction cross section in InSb. We also calculate the Auger recombination coefficient in InSb and find it to be in good agreement with existing experimental data. In all the calculations we rely on Fermi–Dirac statistics and use a four-band k·p theory for band structure calculations. Experiments on the transmission of submicrosecond CO_2 laser pulses through InSb produce results consistent with the calculated parameters. © 2008 Optical Society of America

OCIS codes: 190.0190, 020.4180, 190.5970.

1. INTRODUCTION

The large optical nonlinearities of semiconductors make them attractive materials for optical switching, optical bistability, and other applications in nonlinear optics [1]. Specifically, the narrow bandgap semiconductor InSb is considered in this work due to its possible use for clamping the transmitted energy below a certain level in the infrared. In order to successfully utilize the nonlinear optical (NLO) properties of InSb without having considerable linear absorption, one needs to work in a region of photon energies below the direct bandgap energy, E_g, and above the cut-off energy for the allowed two-photon transitions, $E_g/2$ [2]. This corresponds to a wavelength range between about ~7 and 14 μm at room temperature.

The main interband transition mechanism for the wavelength of interest is two-photon absorption (2PA). Narrow bandgap semiconductors are very interesting from the point of view of 2PA, since their 2PA coefficient is ~3 orders of magnitude larger than that of large-gap semiconductors and dielectrics. This happens due to an inverse cubic dependence of the 2PA coefficient, β_2, on the bandgap energy, E_g [3,4]. When this scaling rule is applied to InSb at 300 K ($E_g = 0.18$ eV), it predicts β_2 ~7 cm/MW, which is reasonably close to the experimentally observed value of ~2 cm/MW [5–7].

Although large, 2PA alone is usually not the dominant contributor to nonlinear absorption for CO_2 laser pulses of nanosecond or longer durations. However, 2PA plays a critical role in reducing the transmittance by generating free carriers, which in turn cause strong free-carrier absorption and refraction. Absorption of a photon by a free electron in the conduction band requires simultaneous absorption or emission of a phonon (Fig. 1). Both acoustic and optical phonons can participate in this process. The free-electron absorption cross section is $\sigma_e \approx 2.3 \times 10^{-17}$ cm^2 [8]. Unlike the free-electron process, free-hole absorption is dominated by direct interband transitions from the heavy-hole to the light-hole band, resulting in a large free-hole absorption cross section of $\sigma_h \approx 8.6 \times 10^{-16}$ cm^2 [8]. Free-carrier transitions as well as other important effects in InSb are shown in Fig. 1. According to the Drude–Lorentz classical model, free-carrier absorption is accompanied by free-carrier refraction, which plays the dominant role in nonlinear beam propagation through InSb. The nonlinear refraction from the free carriers is always negative, i.e., self-defocusing, and is very strong [1]. It is so strong that light often misses the detector placed behind the sample in the experiments. This has probably caused some of the significant errors in measurements of β_2 due to this added "loss" causing overestimation of β_2 [7].

The effects reducing the overall nonlinear response include 2PA blocking [9] and Auger recombination [9]. As the number of electrons in the conduction band increases, the available states in the band fill up, resulting in the blocking of the 2PA. However, the main mechanism that limits the buildup of the excited carrier population in InSb for nanosecond and longer pulses is Auger recombination. Therefore, the precise knowledge of Auger rates is extremely important when analyzing the performance of a NLO device based on narrow bandgap semiconductors. In this respect, the well-known discrepancy [10] between calculated and measured values of Auger rates in InSb presents a major challenge for analysis. Section 2 of this paper addresses this problem. Nonlinear refraction due to the 2PA generated photoexcited carriers as the main nonlinear mechanism reducing the transmitted fluence in InSb is analyzed in Section 3. The experimental setup for

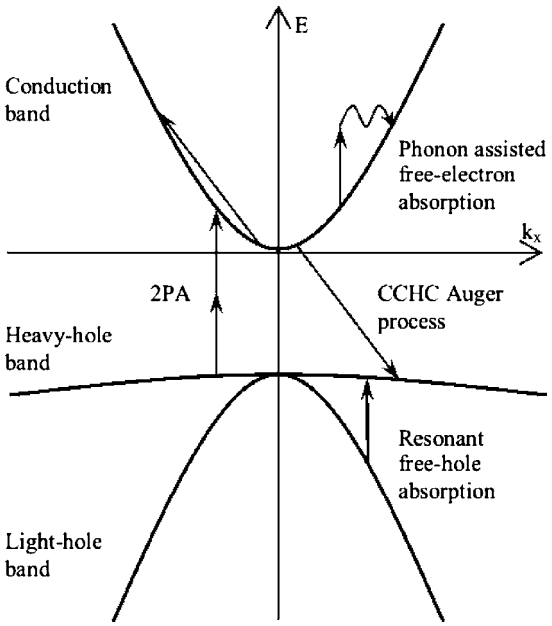

Fig. 1. Band structure and transitions in InSb.

analysis of nonlinear transmission of CO_2 pulses through InSb is described in Section 4. The discussion of the results is presented in Section 5.

2. AUGER RECOMBINATION

The Auger process is the main mechanism for carrier recombination in narrow bandgap semiconductors at high carrier densities [11]. In the following discussion we ignore other recombination processes such as trap, radiative, and surface recombination. Band-to-band Auger transitions in InSb are shown in Fig. 1. The electron–electron process (CCHC), depicted in Fig. 1, is the dominant process at room temperature, while the light-hole–heavy-hole process (CHLH) is also important [11,12]. The notation here is C for conduction band, H for heavy-hole band, S for split-off band, and L for light-hole band. Transitions involving the split-off band (CHSH) do not contribute significantly in InSb since the split-off energy is much larger than the bandgap energy in this material. In the case where the carrier excitation occurs through 2PA, the following relation can be used to describe carrier dynamics in InSb:

$$\frac{\mathrm{d}(\Delta n)}{\mathrm{d}t} = \frac{\beta_2 I^2}{2\hbar\omega} - C_{\mathrm{Auger}}\Delta n(n_0+\Delta n)(2n_0+\Delta n), \qquad (1)$$

where Δn is the excess carrier density, n_0 is the equilibrium (intrinsic/undoped) free-carrier density, β_2 is the 2PA coefficient, I is the irradiance, and C_{Auger} is the Auger coefficient. Strictly speaking, the second term on the right hand side of Eq. (1) represents a valid expression for Auger recombination only when Boltzmann (nondegenerate) statistics adequately describe equilibrium carrier densities in the bands. The cubic dependence of the Auger rate on excess carrier density no longer applies at room temperature when Boltzmann statistics fail, and the use of Fermi–Dirac (degenerate) statistics is necessary. Material properties (band structure) influence the actual functional dependence of Auger recombination on the carrier density when degenerate statistics apply [12]. Numerical analysis has to be performed for each material separately to determine the effect of degeneracy on the Auger process. We describe this in more detail at the end of this section.

Landsberg and Beattie [13] established the foundation for the theoretical analysis of Auger recombination in semiconductors. We follow their approach in this work and note some problematic areas in the literature.

Auger recombination can be considered as an electron–electron scattering process mediated by a screened Coulomb interaction. The presence of multiple mobile carriers around a certain electron results in a faster decrease of the potential of this electron with distance. This is usually referred to as screening of the Coulomb potential. It presents the simplest way to account for many-body effects on electron dynamics in solids and plasmas. Treating the screened Coulomb potential as a perturbation to the Hartree–Fock Hamiltonian and ignoring Umklapp-type transitions [14], Landsberg and Beattie arrived at the following expression for the matrix element of Auger transitions [11,12]:

$$M_{1234} = \frac{e^2}{\varepsilon\varepsilon_0 V}\sum_{s_i,s_f}\left(\frac{F_{14}(s_{1i},s_{1f})F_{23}(s_{2i},s_{2f})}{\lambda^2+(\mathbf{k}_1-\mathbf{k}_4)^2} - \frac{F_{24}(s_{1i},s_{1f})F_{13}(s_{2i},s_{2f})}{\lambda^2+(\mathbf{k}_2-\mathbf{k}_4)^2}\right)\delta_{\mathbf{k}_1+\mathbf{k}_2,\mathbf{k}_3+\mathbf{k}_4}, \qquad (2)$$

where overlap integrals (i.e., for states 1 and 3) are defined as follows:

$$F_{13}(s_i,s_f) = V_{\mathrm{cell}}^{-1}\int u_1^*(\mathbf{r},s_i)u_3(\mathbf{r},s_f)\mathrm{d}r, \qquad (3)$$

and the states are specified by wave vector $\mathbf{k}$. Spin variable s, $u_{nk}(\mathbf{r},s)$ is the cell-periodic part of the Bloch wave function normalized in the volume of a unit cell V_{cell}, and λ is the linear screening constant determined from

$$\lambda^2 = \frac{e^2}{\varepsilon\varepsilon_0}\left(\frac{\mathrm{d}n}{\mathrm{d}\mu_c} - \frac{\mathrm{d}p_h}{\mathrm{d}\mu_h} - \frac{\mathrm{d}p_l}{\mathrm{d}\mu_l}\right), \qquad (4)$$

where the derivatives determine the rates of carrier concentration in the respective bands with the change of quasi-Fermi energy in those bands (c—conduction band, h—heavy-hole band, and l—light-hole band).

Equation (2) was first directly employed by Beattie [12] for the calculation of the matrix element for Auger transitions. In most earlier works a somewhat simpler expression was used [11,15], which allowed for the decoupling of spin and coordinate dependent parts of the wavefunctions. Strictly speaking, such an approximation is not applicable when the spin–orbit interaction is taken into account [12].

Fermi's golden rule gives the rate of the CCHC transition as follows:

$$r = \frac{2\pi}{\hbar}\frac{1}{V}\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4}|M_{1234}|^2 P_E \delta(E_1 + E_2 - E_3 - E_4), \tag{5}$$

where V is the crystal volume, P_E accounts for the occupation factors and includes the influence of the inverse process (impact ionization):

$$P_E = f_c(E_1)f_c(E_2)f_h(E_3)(1 - f_c(E_4)) - (1 - f_c(E_1))(1 - f_c(E_2))(1 - f_h(E_3))f_c(E_4), \tag{6}$$

where $f_c(E)$ and $f_h(E)$ are Fermi–Dirac factors for electrons and holes, respectively, and

$$f_c(E) = 1 \Big/ \left(1 + \exp\left(\frac{E - \mu_c}{k_b T}\right)\right), \tag{7}$$

$$f_h(E) = 1 \Big/ \left(1 + \exp\left(-\frac{E - \mu_h}{k_b T}\right)\right), \tag{8}$$

where μ_c and μ_h are electron and hole quasi-Fermi energies.

We can write the expression for the probability of Auger recombination, Eq. (5), in a more convenient form by going from summation to integration:

$$r = \frac{2\pi}{\hbar V}\left(\frac{V}{8\pi^3}\right)^3 \int |M_{1234}|^2 P_E \delta(E)\delta(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3 - \mathbf{k}_4) \times \mathrm{d}^3\mathbf{k}_1 \mathrm{d}^3\mathbf{k}_2 \mathrm{d}^3\mathbf{k}_3 \mathrm{d}^3\mathbf{k}_4, \tag{9}$$

where $E = E_4 + E_3 - E_2 - E_1$. The factor $V/8\pi^3$ represents the density of wave vectors and appears when we change a corresponding sum to an integral. This equation can be seen to have the correct dimensions, whereas the analogous equation in [16] included an extra density factor and is dimensionally incorrect. It is important to note that the cubic functional dependence of the Auger rate on the carrier density in Eq. (1) follows from Eq. (9) if Boltzmann statistics are used instead of Fermi–Dirac relations in Eqs. (7) and (8) [11,12]. In the degenerate case no such analytical simplification is possible.

The main difficulty in estimating the matrix element in Eq. (9) arises due to the overlap integrals of Eq. (3). The first attempts to relate the overlap integrals to known parameters were made using the effective mass sum rule [11]. The values for the matrix element obtained in this way were not accurate and resulted in a significant overestimation of the Auger recombination rate [17]. Similarly, an expression suggested by Takeshima [18] that relates the overlap integrals to the experimentally measurable momentum matrix element parameter P from the Kane theory [19] suffers from the same drawback. An excellent discussion on the applicability of various approximations in estimating the overlap integrals is given by Burt and co-workers [17]. Burt suggests using the 15-band full zone k·p theory [20] or the empirical nonlocal pseudopotential theory [21] as the only methods capable of producing correct wave functions to be used for the calculation of the overlap integrals in Eq. (3). Unfortunately the complexity of the problem formulated in this way renders all calculations of Auger recombination rates impractical. Scharoch and Abram [22] suggested using the standard four-band k·p method [19] with the effect of higher bands taken into account by the Lowdin procedure [23]. The applicability of this method was verified by comparing the overlap integrals with those obtained from the pseudopotential calculations. Beattie *et al.* [24] later developed this method further, obtaining the parameters for the k·p method by fitting the band structure from the pseudopotential method and using them in Auger rate calculations. In this way the anisotropy of the band structure was taken into account in the calculation of the overlap integrals. A further attempt to simplify the calculations was made by Beattie and White [25]. They introduced a simplified band structure with a nonparabolic isotropic conduction band and a flat valence band to obtain an analytical approximation for CCHC Auger rates. Unfortunately, this approximation does not yield the absolute value of the Auger rate, which has to be obtained by comparison with the full band result. Methods that differ in certain aspects from that used by Beattie and co-workers were suggested by other groups [16,26,27]. These methods differ in the way the integration in k space is performed and include certain approximations of realistic band structures, wave functions, and electron statistics.

It has to be noted that although the analysis of Auger recombination in semiconductors has dramatically evolved since the inception of the field by Landsberg and Beattie in 1959, the agreement between numerical calculations and experimental results in narrow band semiconductors is rarely observed. In fact, a 1 order of magnitude disagreement in the Auger rate is quoted by some authors as "common" [26] and is attributed to the inaccuracy of the models used for numerical calculations. However, as was discussed above, the numerical methods developed recently attained a degree of complication that includes the realistic band structures and, thus, should be expected to describe the Auger recombination more accurately. In this paper we find that the experimentally measured recombination rates agree well with the numerical calculations of InSb. We show that the previously known disagreement is largely due to the approximations made in the interpretation of the numerical results by experimental groups working in the field. Another important result of our work is that we find that it is not necessary to revert to pseudopotential [24,25] methods for band structure calculations, and that the simple four-band k·p theory [19] is sufficient to obtain accurate Auger recombination rates. Indeed, the need to include the influence of the higher bands on the band structure of InSb around the Γ point was emphasized based on a comparison of the angular dependence of the overlap integrals resulting from the exact band structure and from the four-band approximation. However, the comparison shows significant differences only for large wave numbers [22], from which the contribution to the total Auger rate is minimized due to the low occupation probability of large wave number states. Beattie also acknowledges [24] that the exact angular dependence of the overlap integrals has little effect on the lifetimes.

In order to obtain the overlap integrals in Eq. (3) we use the Kane wave functions [19], which have to be prop-

erly transformed to account for the direction of the wave vectors with respect to the crystallographic axes. In fact, the expressions for the overlap integrals that we obtain in this fashion differ from those shown by Beattie in Appendix 1 of [12]. Consequently our calculations of the Auger recombination rate differ from [12]. However, we find our results to be consistent with those presented by Beattie and co-authors in a later paper [25]. For practical purposes we find the presentation of the lengthy derivation of overlap integrals to be unnecessary as it provides little insight. We use numerical matrix multiplication for coordinate transformations.

We utilize the method of k-space integration suggested by Beattie [12]. The rate of recombination is obtained from Eq. (9) as follows:

$$r = \frac{4\pi e^4}{256\pi^8\hbar\varepsilon^2\varepsilon_0^2}\int_{k_{4\text{th}}}^{\infty} k_4^2 \mathrm{d}k_4 \int \left| \sum_{s_i,s_f} \left(\frac{F_{14}(s_{1i},s_{1f})F_{23}(s_{2i},s_{2f})}{\lambda^2 + (\mathbf{k}_1 - \mathbf{k}_4)^2} - \frac{F_{24}(s_{1i},s_{1f})F_{13}(s_{2i},s_{2f})}{\lambda^2 + (\mathbf{k}_2 - \mathbf{k}_4)^2} \right) \right|^2 P_E \delta(E) \mathrm{d}^3\mathbf{k}_1 \mathrm{d}^3\mathbf{k}_2, \quad (10)$$

where we used the properties of the delta function to carry out the integration over k_3 and used the spherical symmetry of the Kane band structure to integrate over the directions of k_4. The first integral is over the magnitude of k_4, and the second integral sign stands for the integration over a six-dimensional (6D) space of all possible wave vectors of initial states 1 and 2. The lower limit in the first integral corresponds to the smallest magnitude of k_4, for which the CCHC transition is possible. This value is referred to as the threshold wave number $k_{4\text{th}}$, and the corresponding energy of the final state 4 is known as the threshold energy $E_{4\text{th}}$ [11]. Equivalently, $E_{4\text{th}}$ is the lowest energy required for an electron to induce impact ionization.

Threshold energies for the CCHC and the CHLH transitions are given in the Appendix, where the computation of the integral in Eq. (10) is explained in considerable detail. The step by step explanation closely follows the discussion by Beattie [12]. The emphasis, however, is not on involved algebraic transformations, but rather on physical interpretations of intermediate results. This discussion in the Appendix uncovers the main reason for disagreements between theoretical and experimental groups with respect to Auger recombination in InSb.

In the literature on Auger recombination, considerable attention is devoted to finding the threshold energy $E_{4\text{th}}$ or threshold momentum $k_{4\text{th}}$ of a particular Auger process. Although the threshold energy is important when the impact ionization is considered, the same does not always hold for the Auger recombination. In fact, we find that in InSb at room temperature (and especially for CCHC transitions) it is not crucial to know the threshold precisely in order to obtain accurate values of the Auger recombination rates. As shown in the Appendix, transitions that occur well above the threshold constitute the main contribution to the Auger rate in InSb. Therefore, the analysis based on the assumption that Auger transitions predominantly take place at the threshold (dominant channel approximation) is incorrect [10]. Although Chazapis and co-authors use Beattie's treatment [12] as their starting point, the employment of the dominant channel approximation renders their estimations of the Auger rate invalid. Thus, it is not surprising that this analysis fails to predict the experimental results accurately. Unfortunately, some authors [26] attributed this disagreement not to the dominant channel approximation [10], but to the inaccuracy of Beattie's approach to overlap integral calculations.

In Fig. 2 we compare the results of our calculations with the experiments of Chazapis and co-authors [10] (unconnected data points at intermediate carrier densities), and Almazov and co-workers [28] (a point at high carrier density). Contrary to previous claims in the literature [10,26] we find good agreement between the numerical calculations and the experimental data at carrier densities above $3\times10^{17}\ \mathrm{cm}^{-3}$. The agreement at higher carrier concentrations is important since it is known that Auger recombination is the dominant relaxation process under these conditions. We attribute the disagreement with the results of Chazapis [10] at low excess carrier densities to the possible significant contribution of surface recombination in the sample of 3.4 μm thickness used in the experiment [10]. Surface recombination could dominate the Auger processes at low excitation levels. Additional experiments with a thicker sample and/or detailed modeling of surface carrier dynamics in InSb is required in order to confirm this assumption.

As we mentioned above, our treatment of the band structure of InSb is not the most complete theoretical consideration attempted to date in Auger rate calculations [24,25]. However, a four-band Kane k·p model employed in this work turns out to be sufficient as shown below. We compare our calculations (squares in Fig. 3) with those of [24,25], which include the effects of higher bands and anisotropy of the heavy-hole band (solid curve in Fig. 3). The results of the total Auger recombination rate calculations by both models as well as the experimental data are shown in Fig. 3.

Although Beattie and co-authors [24,25] include the detailed band structure model in their calculations of the

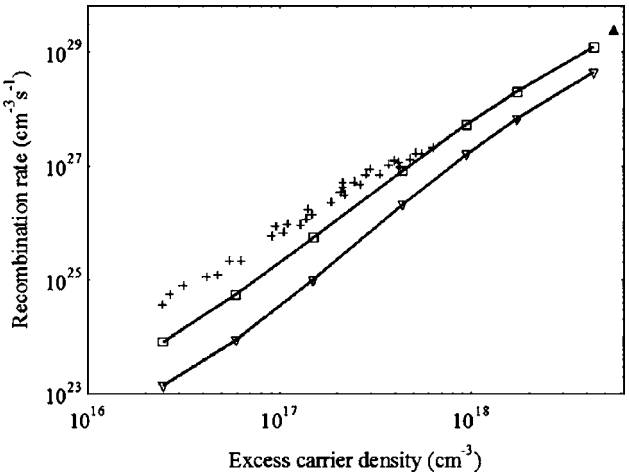

Fig. 2. Experimental and theoretical recombination rates versus excess carrier density. Experimental data is from [10] (plus signs) and [28] (closed triangle). Theoretical rates are calculated based on Eq. (10) for the CCHC Auger process (open squares) and for the CHLH process (open triangles).

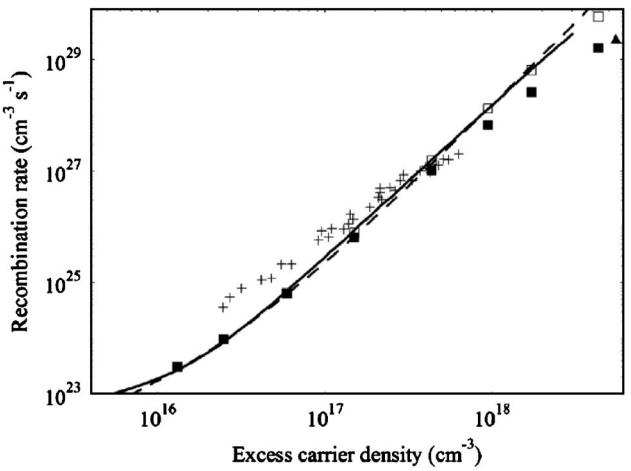

Fig. 3. Comparison of various models for calculation of the Auger recombination rate and experimental data. Detailed band structure calculations from [25] (solid curve), present work calculations with the simplified Kane band structure Eq. (10) without screening (open squares) and with static screening (closed squares), best fit to classical Auger process Eq. (1) (dashed curve), experimental data from [10] (plus signs) and [28] (closed triangle).

Auger rate (solid curve), at the same time they neglect the screening of the Coulomb potential. Our calculations that were based on a simple Kane model of the band structure and neglected the screening produced identical results (open squares), which confirms our assumption that a more detailed analysis of the band structure is not necessary for obtaining accurate recombination rates. Indeed, the question of whether screening needs to be included turns out to be more important. The calculations including screening (closed squares) appear to agree better with the experimental results of Almazov and co-workers at high carrier density [28] (closed triangle). The discussion on the validity of the static screening approximation employed here is given in [29]. Some authors suggest [17,29] that a dynamic rather than static screening model should be used to describe the free-carrier screening of the interaction in the Auger process. However, this issue still remains unresolved and requires additional experimental evidence in support of one or the other model.

The calculation of recombination rates in InSb at higher carrier concentrations may indeed require a more detailed analysis of the band structure. However, the inclusion of the higher and lower bands in the k·p formalism by perturbation theory proves not to be accurate enough at that point, since other effects such as alternation of the band shapes due to the electrostatic interaction of free carriers become important [17].

The most striking result of our calculations is that the simple classical theory of Eq. (1) based on Boltzmann statistics still works well for the excess carrier densities under consideration. The best fit of Eq. (1) to numerical results (dashed curve in Fig. 3) yields an Auger coefficient $C=1.8\times10^{-26}$ cm^6/s, which is roughly a factor of 1.7 smaller than obtained by Beattie [12] within the same framework. Although the source of this difference is not clear, we find that our results agree better with those using detailed band structure calculations [25]. The best fit with the classical expression Eq. (1) based on simple kinetics considerations agrees remarkably well with involved computations taking into account a detailed analysis of the band structure [24,25]. The agreement is better with calculations that neglect screening, producing a valid fit for excess carrier densities from 8×10^{15} to 3×10^{18} cm^{-3}. The same fit is appropriate for calculations including screening in the range of excess carrier densities from 5×10^{15} to 4×10^{17} cm^{-3}. The validity of the classical model of Eq. (1) in this range of carrier densities allows us to use it for the numerical modeling of our experimental results presented in Section 3.

It is important to emphasize that in all our calculations of Auger processes we used Fermi–Dirac (degenerate) statistics. There is no *prima facie* justification for assuming that using Boltzmann (nondegenerate) statistics from the outset in Eq. (6) could result in a meaningful estimation of an Auger coefficient [11–13]. However, if we use Eq. (1) based on Boltzmann statistics to fit experimental data and theoretical curves, we find it provides a reasonably good fit. Although unexpected, this establishes the applicability of using Eq. (1) in beam propagation modeling where more involved calculations would not be practical.

There appears to be no simple answer as to why the functional dependence of Auger recombination on excess carriers is still reasonably well described by the simple cubic law for InSb at room temperature. Unfortunately, it is impossible to analytically obtain a closed form solution of this dependence when using Fermi–Dirac statistics. Although numerical modeling is a means to extract the dependence, it does not provide as much insight as an analytical solution does. Thus, we can only hypothesize on the nature of this cubic dependence. Further studies are required to give a definitive answer.

An argument for a quadratic dependence based on simple kinetics goes as follows [11]. It is argued that the number of nondegenerate particles participating in an interaction determines the functional dependence of a process on carrier density. Since the Auger CHLH is a three-body process involving two holes (one light and one heavy) from a nondegenerate distribution and an electron from a degenerate distribution (it is always available), the process should be described by a quadratic dependence on carrier density. Similarly, the CCHC process is deemed to follow a linear dependence on carrier density. We find that these simplifying arguments are not valid where realistic degenerate distributions in InSb at room temperature are concerned. First, the Fermi–Dirac distribution is no longer a step function at room temperature. Second, we find that most Auger processes occur away from the bottom of the conduction band where electrons indeed fill up all available energies (see Figs. 13 and 14 of the Appendix). Thus, these processes are likely to involve electrons in the thermal tails of the Fermi–Dirac distributions. The nature of the distribution in the tails is more Boltzmann-type than step-like. It is reasonable to assume that the number of electrons available in the tails is proportional to the total excess carrier concentration. Hence the fact that detailed numerical calculations presented in this section yield a cubic dependence of the Auger rate on carrier density is not that surprising.

When considering the optical properties of InSb we can still use the classical model even at low carrier densities

where this model disagrees with the more detailed calculations of the Auger process. This is because at low excess carrier densities $\Delta n \ll n_0$, linear absorption dominates and nonlinear effects are not important.

3. NONLINEAR REFRACTION

Free-electron contributions to nonlinear refraction dominate bound-electronic effects in InSb for the pulsewidths and irradiances used in our and most other experiments we have found in the literature. Our calculations indicate that the only way to have the bound electronic refraction comparable with the carrier refraction is to use extremely short pulses, picoseconds to femtoseconds, where the irradiance is high with low energy so that few carriers are produced [30,31]. According to the Drude–Lorentz classical model, the free-electron refraction cross section is given by

$$\sigma_{\mathrm{ref}} = -\frac{e^2\lambda}{4\pi\varepsilon_0 m_c n_0 c^2}, \tag{11}$$

where e is the electron charge, c is speed of light in vacuum, λ is the wavelength, ε_0 is the permittivity of vacuum, m_c is the effective mass of conduction band electrons, and n_o is the linear refractive index. Due to nonparabolicity of the conduction band in InSb [19], electrons in different states are characterized by different values of effective mass according to

$$m_c = m_c^*\left(1+\frac{2E}{E_g}\right), \tag{12}$$

where m_c^* stands for the effective mass at the bottom of the conduction band. In order to account properly for the dependence of the effective mass on energy in optical electron transport, we need to use the average value of m_c in Eq. (11) [32]:

$$\langle m_c\rangle = \frac{\int_0^\infty m_c(E)N(E)f(E)\mathrm{d}E}{\int_0^\infty N(E)f(E)\mathrm{d}E}, \tag{13}$$

where $N(E)$ is the density of states function

$$N(E) = \frac{1}{2\pi^2}\left(\frac{2m_c^*}{\hbar^2}\right)^{3/2}\sqrt{E\left(1+\frac{E}{E_g}\right)}\left(1+\frac{2E}{E_g}\right) \tag{14}$$

and $f(E)$ is the Fermi–Dirac distribution of Eq. (7). We find that the values of effective mass obtained from $1/\langle m_c\rangle$ and $\langle 1/m_c\rangle$ calculations are almost identical.

One of the important consequences of taking the average effective mass for m_c in Eq. (11) is that the resulting cross section of free-electron refraction is significantly lower than obtained for the band edge value of m_c^*. Since the average effective mass grows with the increase of the electron concentration, the refractive cross section is reduced. The dependence of the free-electron refraction cross section on electron density at 10.6 μm is shown in Fig. 4 (solid curve). We should note that we neglect the free-hole contribution to nonlinear refraction. This approximation is based on the fact that the large mass of free holes (0.45 m_0 as opposed to 0.015 m_0 for free electrons) reduces the effect.

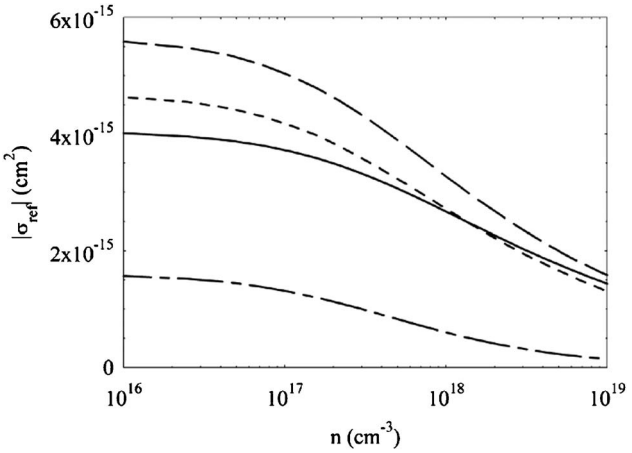

Fig. 4. Refraction cross section versus photoexcited carrier density. Absorption blocking contribution (dot-and-dash curve), free electron refraction contribution (solid curve), total refraction cross section (long dash curve), and 80% of total refraction cross section (short dash curve) producing the best fit to the time resolved measurements shown in Fig 11.

Another important effect that we take into account is the refractive index changes resulting from absorption blocking. The creation of free electrons in the conduction band and their rapid thermalization to the bottom of the band results in the blocking of linear absorption slightly above the band minimum. This absorption saturation is known as the Burnstein–Moss effect [33]. Through Kramers–Kronig relations, the change in absorption causes a change of the refractive index. The model of nonlinear refraction due to absorption blocking was first suggested by Miller *et al.* [1], developed further by Wherrett and co-authors [34], and used for the calculation of the index change due to the blocking of the linear transitions by two-photon excited electrons by Said and co-authors [35]. In all three papers the parabolic conduction band and Boltzmann statistics were assumed. As we pointed out earlier, these assumptions are not valid for InSb. Therefore, we proceed with Miller and co-authors' model [1] in a more general case taking into account the nonparabolicity of the conduction band and Fermi–Dirac statistics. We modify Miller and co-authors' expression for the direct interband absorption coefficient using the density of states expression for a nonparabolic band, Eq. (14). We obtain

$$\alpha(\hbar\omega) = \frac{\sqrt{2m}e^2}{3\pi\hbar^2 c n_0\varepsilon_0}\left(\frac{m_c^*}{m}\right)^{3/2}\frac{2mP^2}{\hbar^2}\sqrt{\frac{(\hbar\omega-E_g)}{E_g\hbar\omega}}\left(\frac{2\hbar\omega}{E_g}-1\right) \times\left(1-\frac{1}{1+\exp[(\hbar\omega-E_g-\mu)/kT]}\right), \tag{15}$$

where P is the momentum matrix element given, $\hbar\omega$ is the photon energy, m is the rest mass of electron, and μ is the Fermi energy. The following assumptions, consistent with the material properties of InSb, have been used in obtaining Eq. (15): we neglected the relatively weak absorption from the light-hole band, and since $m_h \gg m_c^*$, we assumed a flat heavy-hole band. As a consequence of the

second assumption, we neglected the filling of the states in the heavy-hole band. We also took the conduction band energy for the direct transition $E_c=\hbar\omega-E_g$, thus neglecting $1/m_h$ compared to $1/m_c^*$ in the expression for the reduced effective mass [1]. The validity of Eq. (15) has been verified by the comparison of the calculations with the experimental data available in the literature [36]. In order to obtain good agreement with the data, the dependence of the momentum matrix element on the transition energy has to be taken into account [37].

Absorption and refraction are related through Kramers–Kronig expressions. The change in the refractive index at photon energy $\hbar\omega$ is given by

$$\Delta n(\hbar\omega)=\frac{\hbar c}{\pi}\int_0^\infty \frac{\Delta\alpha(\hbar\omega')\mathrm{d}(\hbar\omega')}{(\hbar\omega')^2-(\hbar\omega)^2}, \tag{16}$$

The method of the refractive index change calculations based on Kramers–Kronig relations is very flexible as it allows finding the contribution to the nonlinear refraction of a certain absorption change with no relation to the specific physical process that produced this change in absorption. In such a way we are able to find the negative refraction contribution (at 10.6 μm) of linear absorption blocking induced by two-photon carrier excitation at 10.6 μm. Using the same formula, Eq. (16), we can estimate the positive change of the refractive index brought on by the appearance of new linear transitions due to the bandgap shrinkage with temperature when local optical heating of the material takes place. In a similar fashion, we can calculate the Kerr index (contribution of bound electrons) knowing the dependence of the nondegenerate coefficient on frequency. As verified by our calculations, contributions of bound electrons and thermal effects are much less significant than the absorption blocking at the wavelength of interest.

In order to identify how the change of refraction depends on carrier density we need to know the change of the absorption coefficient due to the creation of additional carriers. We introduce the notation: $x=\hbar\omega'-E_g/kT$, $a=\hbar\omega-E_g/kT$, $b=E_g/kT$, $g_0=\mu/kT$, and $g(n)=\mu(n)/kT$, where $\mu(n)$ is the electron Fermi energy depending on free-electron concentration, and μ is the Fermi energy at equilibrium at 300 K (no optical excitation of carriers). Then, using Eq. (15) we obtain

$$\Delta\alpha(x)=\frac{\sqrt{2m}e^2kT}{3\pi\hbar^2cn_0\varepsilon_0E_g^{3/2}}\left(\frac{m_c^*}{m}\right)^{3/2}\frac{2mP^2}{\hbar^2}\times\frac{\sqrt{x}(2x+b)[\exp(x-g(n))-\exp(x-g_0)]}{\sqrt{x+b}[1+\exp(x-g_0)][1+\exp(x-g(n))]}. \tag{17}$$

Substituting Eq. (17) into the Kramers–Kronig relation, Eq. (16), we obtain

$$\Delta n(a)=\frac{\sqrt{2m}e^2}{3\pi^2\hbar cn_0\varepsilon_0E_g^{3/2}}\left(\frac{m_c^*}{m}\right)^{3/2}\frac{2mP^2}{\hbar^2}\int_0^\infty\frac{\sqrt{x}(2x+b)[\exp(x-g(n))-\exp(x-g_0)]\mathrm{d}x}{\sqrt{x+b}(x-a)(x+a+2b)[1+\exp(x-g_0)][1+\exp(x-g(n))]}, \tag{18}$$

where the integration is performed for photon energies above the bandgap. The expression under the integral depends on the free-carrier density in such a way that carrier density cannot be taken out of the integral as in the case of Boltzmann statistics [34]. Therefore, we expect the refractive cross section to depend on electron density in a nontrivial way. The integral in Eq. (18) has to be evaluated numerically since no analytical simplification similar to that in [1,34,35] appears possible. The resulting refractive cross section is plotted in Fig. 4 as a function of carrier density (dot-and-dash curve) together with previously obtained results for free-electron refraction, (solid curve) as well as the total refraction cross section (long dash curve). The calculations include the correction for the variation of the momentum matrix element with energy. Figure 4 also shows the cross section (short dash curve) that produces the best fit to time resolved pulse transmission measurements described in Section 4.

4. EXPERIMENTAL PROCEDURE

A Laser Science Incorported PRF-150 grating-tuned transverse excitation atmospheric pressure (TEA) CO_2 laser was used to determine the effect of nonlinear refraction on the propagation of the laser pulse through InSb. The gas mixture, containing ~20% of CO_2 (active media), 15% of N_2 (to increase the excitation efficiency of CO_2), and 65% of He (ballast gas) is supplied to the main gas chamber at atmospheric pressure. Since the excitation occurs at atmospheric pressure, the laser line is significantly broadened. This results in multiple longitudinal modes present in the output laser pulse. Mode beating could significantly distort the response of the nonlinear material under investigation. To ensure single longitudinal mode operation, an additional intracavity low pressure pulsed gas cell is used, which narrows the linewidth (essentially serving as a narrow band seed pulse). Additionally, the cavity length is fine-tuned by adjusting the voltage applied to a piezoelectric aligner/translator (PZT) system on the output coupler. The absence of mode beating is verified by monitoring the laser output with a fast Au-doped Ge detector (see Fig. 5). The long tail of the pulse appears due to re-excitation of CO_2 molecules within the N_2–CO_2 gas system.

The laser cavity is formed by a coated Ge output coupler (80% reflectivity) and a 135 line/mm blazed grating. The Au coated grating on a copper substrate is used for tuning between different CO_2 lines. An aperture is placed

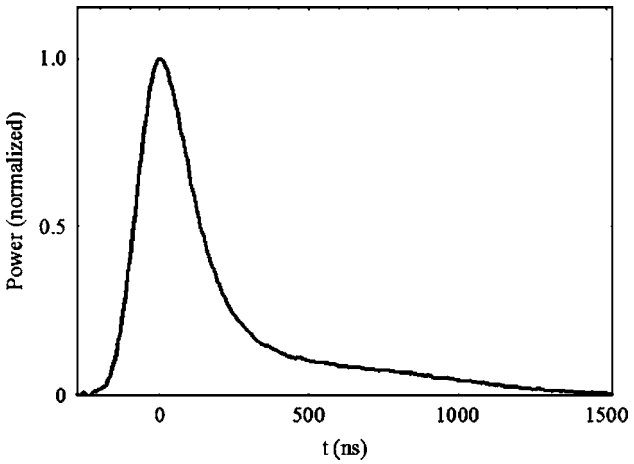

Fig. 5. Laser pulse temporal profile measured with a fast Au-doped Ge detector.

within the laser cavity in order to ensure the excitation of the fundamental spatial mode and suppress higher order transverse modes.

The experimental setup shown in Fig. 6 included a 15 cm focal length lens used to focus the beam. The beam profile data were collected via transverse pinhole scans. Figure 7 shows the measured beam profile at the focus. Beam width measurements are shown in Fig. 8 along with a Gaussian propagation fitting:

$$w = w_0 \sqrt{1 + \frac{\lambda^2 z^2}{\pi^2 w_0^4}}, \tag{19}$$

where the width of the Gaussian beam w is defined as the half-width at $1/e^2$ of the maximum (HW1/e^2M) of fluence. A beam spot radius at focus of $w_0 = 120$ μm with $M^2 = 1.0$ produced the best fit to the measured profiles.

We found it critical to make sure that the beam profile is close to Gaussian, as even a small asymmetry in the input beam was observed to become highly pronounced after propagation through a nonlinear material at high input energies. Experimental evidence of this is shown in Fig. 9, where a slight asymmetry in the input beam is shown to be magnified after propagation through the antireflective (AR) coated InSb sample of 1 mm thickness placed at the focus. Although great care has been taken to ensure that the input beam is cylindrically symmetric, the transmitted beam is always found to be somewhat

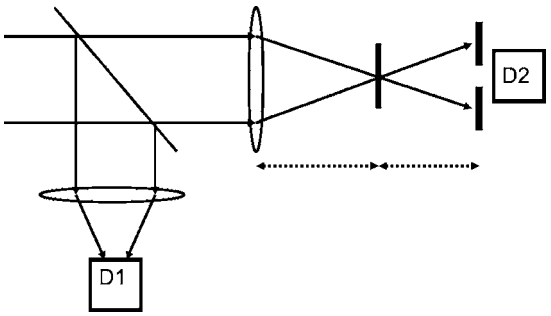

Fig. 6. Experimental setup. Squares indicate detectors (D1 reference, D2 sample transmittance), and arrows show distances from the lens to the sample and from the sample to an aperture placed in front of the transmission detector.

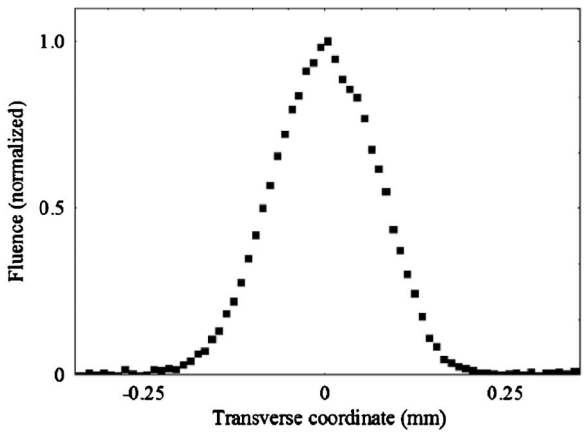

Fig. 7. Spatial beam profile at focus.

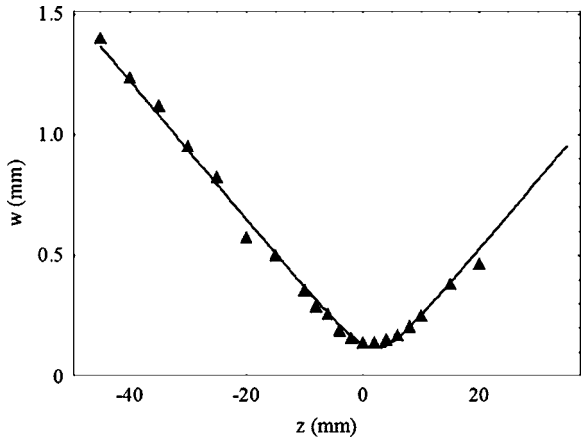

Fig. 8. Beam width measurements (closed triangles) and Gaussian fitting (solid curve) along the propagation path of the focused beam.

asymmetric, especially at high energy. We attribute this effect to a strong nonlinear refraction that magnifies any small imperfections in the input beam profile. A slight wedge in the sample could also lead to a similar result. However, that was not the case here as evidenced by no change in the profile after sample rotation. Small departures from a perfectly Gaussian input beam profile result in relatively small changes in the carrier density distribution, index change, and nonlinear phase. The change in the nonlinear phase shift accumulated upon propagation in the sample results in aberrations of the output beam wavefront that affect the propagation of the beam to the detector plane. Even though we tried to put the detector as close as possible to the sample, the effect of nonlinear aberrations was considerable, as evidenced in Fig. 9. It is worth noting that different parts of the laser pulse (see

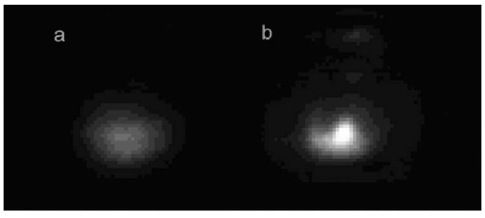

Fig. 9. Beam image obtained at the detector plane D2 with (a) no sample and (b) 1 mm thick InSb sample at the focus.

Fig. 5) experience different nonlinear phase shifts and distortions, thus affecting the temporal content of the pulse at any given spatial position differently after the sample. Since the collection of all angular components of the output beam by a fast detector is difficult to accomplish [37], partial aperturing of the output beam occurs. This often leads to observation of pulse-shaping [37]. Similar effects of strong nonlinear refraction have been observed in liquid crystals [38]. For a more detailed discussion of how nonlinear refraction can affect beam profile and rearrange temporal content of the laser pulse we refer the reader to [37].

5. RESULTS AND DISCUSSION

Comparison between the experimental data and the numerical analysis of CO_2 laser pulse propagation through InSb is complicated by nonlinear distortions of the output beam profile and temporal pulse content. Thus, we might expect the modeling based on our cylindrically symmetric nonlinear beam propagation method [39] to not perfectly fit the time-resolved measurements and experimental spatial beam profiles. However, the calculations of the integrated characteristics, such as the energy collected by an apertured detector, produce reasonable agreement with the experiment.

Here we use theoretically calculated values for the Auger recombination coefficient and the nonlinear refraction as well as the value of the 2PA coefficient obtained in picosecond experiments [6,9] for the numerical simulation of CO_2 laser beam propagation through InSb. The linear absorption cross section of the sample used in our numerical modeling was obtained by measuring the transmittance of the sample at low input energy ($\sigma_h = 4.8 \times 10^{-16}$ cm).

The results of the experiments are compared with the numerical modeling in Fig. 10 for a system with a 5 mm circular aperture (6.5 times the linear beam size) placed 13 mm after the AR coated InSb sample and before a pyroelectric detector used for the measurement of the transmitted energy (Fig. 6). Calculations show good agreement for the total energy transmitted through the sample, but underestimate the energy transmitted through the aperture. Still the agreement is quite reasonable considering the fact that the output beam is distorted, which could lead to the collection of angular components blocked by the aperture in our cylindrically symmetric analysis. A very good fit of the transmittance of the system with an aperture is obtained when the nonlinear refraction is reduced by 30% from the theoretically obtained values. Interestingly, the best fit of time resolved measurements is also obtained for lower nonlinear refraction. A 20% reduction of the calculated total refraction cross section (shown as a short dash curve in Fig. 4) produces the best fit of the pulse profiles collected by a 5 mm aperture, fast Au doped Ge detector placed 13 mm after the sample. The pulse profiles and the best obtained fits are shown in Fig. 11, where σ_{ref} is the fitting parameter (using the values from Fig. 4 by changing the percentage of the maximum value).

All energy was collected by the detector at low energies, therefore, the contribution of the nonlinear refraction was insignificant, and the low energy data is not shown here

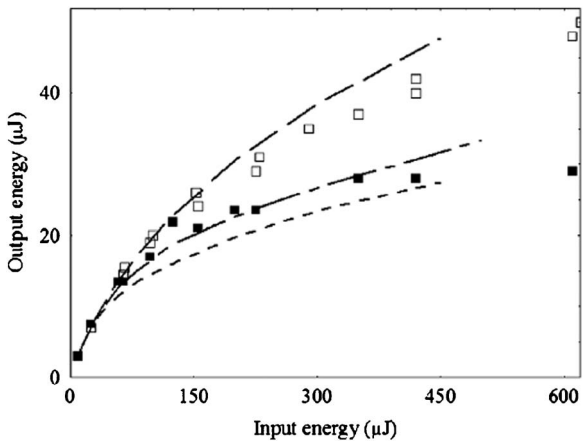

Fig. 10. Experimental and numerical results for output versus input energy. Open aperture Z-scan experimental results are shown as open squares and numerical simulation of beam propagation as a long dash curve. For closed aperture Z-scan data the detector is placed on axis 13 mm after the sample. The 5 mm aperture is 6.5 times larger than the beam FW1/e^2 M at this position in the linear regime. The closed aperture experimental results are shown as closed squares and the numerical simulation of the beam propagation as a short dash curve. Dot-and-dash curve shows numerical simulation results with 70% of the predicted theoretical refraction for comparison.

to make the graph more readable. Pulse profiles at low energies (9 μJ and 25 μJ) were used to obtain the value for the absorption cross section of $\sigma_h = 4.8 \times 10^{-16}$ cm^2, which is within the range of values $3.8-8.6 \times 10^{-16}$ cm^2 found in the literature [5,8,9]. The change of the output pulse shape at high input is a primary indicator of the nonlinear refraction effect. At the maximum of the pulse, where refraction is the strongest, the beam overfills the detector aperture, and only the central part of the beam is detected. This results in the temporal modulation of the output pulse.

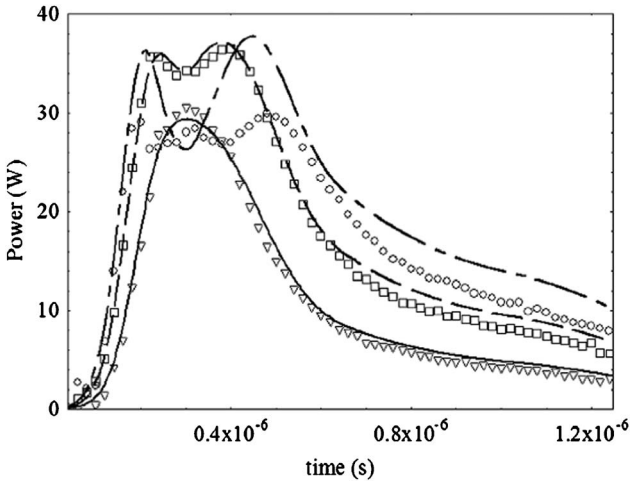

Fig. 11. Time-resolved measurements and the best fit with the numerical beam propagation. The refraction cross section σ_{ref} was the only fitting parameter used. The best fit is obtained by using 80% of the predicted theoretical refraction (Fig. 4). Input pulse energy: 58 μJ (experimental—triangles, numerical—solid curve), 125 μJ (experimental—squares, numerical—dashed curve), 200 μJ (experimental—circles, numerical—dot-and-dash curve).

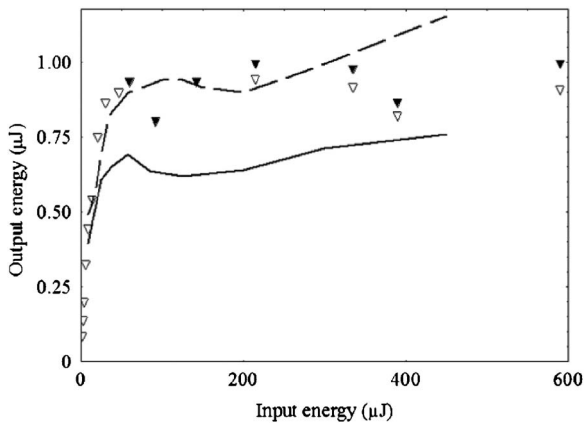

Fig. 12. Output versus input energy 13 mm after the sample with a 0.3 mm aperture on axis (open triangles) or at the maximum of transmittance (closed triangles if maximum is not on axis). Numerical simulation of beam propagation results are shown as a solid curve. Dashed curve shows numerical simulation results with 70% of the predicted theoretical refraction for comparison.

The calculated parameters were also used for comparison of the numerical modeling with an experiment for a system with a 300 μm diameter aperture (15% linear transmittance) 13 mm after the sample (Fig. 12).

Although we obtained good agreement of the experimental transmittance measurements for submicrosecond CO_2 laser pulses, the accurate determination of the nonlinear material parameters from such an experiment is a daunting task, as the number of parameters simultaneously affecting the pulse propagation is very large. The use of picosecond pulses is preferable for this purpose.

6. CONCLUSIONS

Previously, an order of magnitude disagreement between calculated and measured Auger recombination rates had been noted [10]. The inadequacy of the four-band Kane model of band structure had been widely thought to be the reason for this disagreement [17,26]. Our calculation of the Auger rate in a four-band model using Fermi–Dirac carrier statistics yielded good agreement with the experimental results available in the literature, proving it unnecessary to resort to more detailed band structure calculations. We attribute the previous inconsistency in the literature to unfortunate miscommunication between theoretical and experimental groups. Specifically, the neglect of surface recombination in the thin films used in [10] along with the incorrect application of the dominant channel approximation could have led to the incorrect conclusion of inconsistency of the theoretical calculations with the experiments. Although our calculated results agree well with the experiments at intermediate and high excess carrier densities, the low excitation experiments in thin samples yield higher relaxation rates. The importance of carrier diffusion and surface recombination is identified as a possible reason for this disagreement.

We analyzed the nonlinear refraction in InSb based on Kramers–Kronig relations. We refined the method initially described in [1] by taking into account Kane's band structure [19] and Fermi–Dirac carrier statistics.

We presented an accurate numerical simulation of the propagation of pulsed midinfrared radiation through InSb. Small deviations of numerical results from the experiment and distortions of the pulse-shape are attributed to slight beam asymmetry.

Our experiments confirmed the applicability of InSb for the purpose of clamping the transmitted energy of CO_2 laser pulses. One of the drawbacks of such a device is the damage of the InSb crystal itself at input pulse energies exceeding ~1 mJ, corresponding to a fluence of 0.44 J/cm^2. It might be possible to avoid this damage when using another material placed prior to the InSb sample and providing nonlinear defocusing at high inputs [40]. We suggest using CS_2, which exhibits thermal nonlinear defocusing at 10 μm initiated by its ~0.3 cm^{-1} linear absorption coefficient at this wavelength [41]. Use of wedge-shaped samples is also suggested to avoid hot-spot formation at the detector plane [42].

APPENDIX A

In this Appendix we present a step by step explanation of how the integral in Eq. (10) is calculated. We also establish which electron states are responsible for the majority of the Auger transitions in InSb.

The lower limit of the first integral in Eq. (10) is the threshold wavenumber $k_{4\text{th}}$, and the corresponding energy of the final state 4 is known as the threshold energy $E_{4\text{th}}$. At the threshold there is only one combination of initial and final states for which energy and momentum are conserved simultaneously in an Auger process. The energies and the momenta of the states 1–3 that correspond to this threshold transition are denoted as $E_{1\text{th}}$, $E_{2\text{th}}$, $E_{3\text{th}}$, and $\mathbf{k}_{1\text{th}}$, $\mathbf{k}_{2\text{th}}$, $\mathbf{k}_{3\text{th}}$ respectively. These energies and momenta should not be confused with the minimum energies and momenta possible for states 1–3. In order to avoid confusion, we find it important to emphasize that $E_{1\text{th}}$, $E_{2\text{th}}$, and $E_{3\text{th}}$ just stand for the energies of the carriers that combine to produce the lowest value of E_4. It is easy to show that $E_{1\text{th}}=E_{2\text{th}}$, and $\mathbf{k}_{1\text{th}}=k_{2\text{th}}$ for a CCHC transition [11]. We find that the threshold energy, $E_{4\text{th}}$, for a CCHC transition is determined as follows:

$$E_{4\text{th}} = E_g\left(1+\frac{2m_c}{m_h}\right), \tag{A1}$$

which corresponds to $k_{4\text{th}}$ given by

$$k_{4\text{th}} = \frac{2\sqrt{m_c E_g}}{\hbar}\left(1+\frac{3m_c}{2m_h}\right). \tag{A2}$$

For each value of k_4 above $k_{4\text{th}}$ the surface of allowed transitions in a 6D space $\mathbf{k}_1\otimes\mathbf{k}_2$ (direct product of wave vector spaces) is defined by the conservation of energy (the conservation of momentum is already ensured in Eq. (10) by taking $\mathbf{k}_3=\mathbf{k}_1+\mathbf{k}_2-\mathbf{k}_4$). Thus, the second integral in Eq. (10) presents the integration over this surface. Beattie has introduced a coordinate transformation in a combined wavevector space, which significantly simplifies the topology of the problem:

$$\mathbf{k}_1 = \mathbf{R} + \frac{m_c^{\mathrm{th}}}{\sqrt{m_h(m_h + 2m_c^{\mathrm{th}})}}\mathbf{S} + \frac{m_c^{\mathrm{th}}}{m_h + 2m_c^{\mathrm{th}}}\mathbf{k}_4,$$

$$\mathbf{k}_2 = -\frac{m_h + m_c^{\mathrm{th}}}{\sqrt{m_h(m_h + 2m_c^{\mathrm{th}})}}\mathbf{S} + \frac{m_c^{\mathrm{th}}}{m_h + 2m_c^{\mathrm{th}}}\mathbf{k}_4, \tag{A3}$$

where m_c^{th} is defined by

$$m_c^{\mathrm{th}} = \frac{\hbar^2 k_{1\mathrm{th}}^2}{2E_{1\mathrm{th}}}, \tag{A4}$$

and Cartesian coordinates of three-dimensional (3D) vectors $\mathbf{R}$ and $\mathbf{S}$ are expressed in a 6D generalized spherical coordinate system as follows:

$$\mathbf{R} = s(\cos\theta_1, \sin\theta_1\cos\theta_2, \sin\theta_1\sin\theta_2\cos\theta_3),$$

$$\mathbf{S} = s\sin\theta_1\sin\theta_2\sin\theta_3(\cos\theta_4, \sin\theta_4\cos\theta_5, \sin\theta_4\sin\theta_5), \tag{A5}$$

where $s^2 = R^2 + S^2$, $0 \leq \theta_i \leq \pi$ $(i = 1,2,3,4)$, and $0 \leq \theta_5 \leq 2\pi$.

The Jacobean of the transformation into the coordinate system s, θ_i is given by

$$J(s,\theta_i) = \left(\frac{m_h + m_c^{\mathrm{th}}}{\sqrt{m_h(m_h + 2m_c^{\mathrm{th}})}}\right)^3 s^5 \sin^4\theta_1 \sin^3\theta_2 \sin^2\theta_3 \sin\theta_4. \tag{A6}$$

The origin of the new coordinate system corresponds to the threshold transition. For the values of $\mathbf{k}_4 > k_{4\mathrm{th}}$ all allowed transitions lie on a closed surface around the origin. For parabolic bands, the surface enclosing the origin is conveniently reduced to a sphere (in 6D). For a nonparabolic conduction band, the surface of integration is given by an equation $s = s(\theta_i)$, where s in each direction (in 6D) is obtained from the energy conservation considerations:

$$s^2 = \frac{2m_c^{\mathrm{th}} m_h}{\hbar^2(m_h + m_c^{\mathrm{th}})}(\mathbf{Z}(\mathbf{k}_4) + W(\mathbf{k}_4, \mathbf{R}, \mathbf{S})), \tag{A7}$$

where for a nonparabolic conduction band [19] and a parabolic and isotropic heavy-hole band (with a negative effective mass $-m_h$) the following expressions hold:

$$Z(k_4) = -\frac{3E_g}{2} + \sqrt{\frac{E_g^2}{4} + \frac{\hbar^2 k_4^2 E_g}{2m_c}} - \frac{\hbar^2 k_4^2}{2(m_h + 2m_c^{\mathrm{th}})}, \tag{A8}$$

$$W(k_4, \mathbf{R}, \mathbf{S}) = E_g - \sqrt{\frac{E_g^2}{4} + \frac{\hbar^2 k_1^2 E_g}{2m_c}} - \sqrt{\frac{E_g^2}{4} + \frac{\hbar^2 k_2^2 E_g}{2m_c}} + \frac{\hbar^2 k_1^2}{2m_c^{\mathrm{th}}} + \frac{\hbar^2 k_2^2}{2m_c^{\mathrm{th}}}, \tag{A9}$$

where expressions Eq. (A3) relate k_1 and k_2 to $\mathbf{R}$ and $\mathbf{S}$.

The resulting expression for the CCHC transition rate per unit volume is given by

$$r_{\mathrm{CCHC}} = \frac{1}{2\pi^2}\int_{k_{4\mathrm{th}}}^{\infty} k_4^2 \mathrm{d}k_4 \Omega_{\mathrm{CCHC}}(k_4), \tag{A10}$$

where $\Omega_{\mathrm{CCHC}}(k_4)$ is the rate of the CCHC Auger transition with a particular final state 4:

$$\Omega_{\mathrm{CCHC}} = \frac{e^4}{32\pi^5\hbar^3\varepsilon^2\varepsilon_0^2}\frac{m_c^{\mathrm{th}}(m_h + m_c^{\mathrm{th}})^2}{\sqrt{m_h(m_h + 2m_c^{\mathrm{th}})^3}} \times \int \frac{|M|^2 P s^4 \sin^4\theta_1 \sin^3\theta_2 \sin^2\theta_3 \sin\theta_4}{\left|1 - \frac{m_c^{\mathrm{th}} m_h}{\hbar^2(m_h + m_c^{\mathrm{th}})}\frac{1}{s}\frac{\partial W}{\partial s}\right|} \times \mathrm{d}\theta_1 \mathrm{d}\theta_2 \mathrm{d}\theta_3 \mathrm{d}\theta_4 \mathrm{d}\theta_5. \tag{A11}$$

A similar expression in [12] contains a misprint in the powers of the effective mass factor, which can be readily identified by dimensional analysis.

The numerical calculation of the recombination rate in Eq. (A11) proceeds as follows. First, quasi-Fermi energies for electrons E_f and holes E_h as well as the screening parameter λ are determined for a given excited carrier density. Then, a discrete set of values of $k_4 > k_{4\mathrm{th}}$ is chosen with a sufficient number of points for numerical one-dimensional (1D) integration in Eq. (A11). $\Omega_{\mathrm{CCHC}}(k_4)$ is computed using a Monte Carlo method in 6D. The "VEGAS" algorithm [43] is used; however, the adaptive mesh generation available in this algorithm has not been employed. Sufficient accuracy of the Monte Carlo procedure ($\sim 0.1\%$) is ensured by taking 10^6 different sets of generalized direction parameters θ_i $(i = 1,2,3,4,5)$. The evaluation of the integrand is complicated by the need to solve a transcendental equation for θ_i dependent parameters in Eq. (A7) each time a new set of θ_i is used. This is achieved by the Newtonian iteration method, which is found to converge for all $k_4 > k_{4\mathrm{th}}$. The accuracy of solving Eq. (A7) is chosen to accommodate the required accuracy of the Monte Carlo procedure. The overlap integrals in Eqs. (2) and (3) are calculated based on Kane wavefunctions [19], which have been suitably transformed to allow for arbitrary directions of wavevectors $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$ [19,20]. The scalar product of the wavefunctions resulting in the overlap integrals of Eq. (3) has been carried out numerically.

A similar procedure applies for the CHLH process. The threshold energy for the CHLH transition is given by

$$E_{4\mathrm{th}} = -E_g\left(2 + \frac{m_c}{m_h}\right). \tag{A12}$$

The origin of the energy scale is still at the bottom of the conduction band. Therefore, the threshold is $E_g(1 + m_c/m_h)$ below the top of the valence band, which corresponds to $k_{4\mathrm{th}}$ given by

$$k_{4\mathrm{th}} = \frac{2\sqrt{m_c E_g}}{\hbar}\left(1 + \frac{3m_c}{4m_h}\right). \tag{A13}$$

The coordinate transformation needed to place the CHLH threshold transition in the origin of the wavevector space is modified as follows:

$$\mathbf{k}_1 = \mathbf{R} + \frac{m_h}{\sqrt{m_c^{\text{th}}(2m_h + m_c^{\text{th}})}}\mathbf{S} + \frac{m_h}{2m_h + m_c^{\text{th}}}\mathbf{k}_4$$

$$\mathbf{k}_2 = -\frac{m_h + m_c^{\text{th}}}{\sqrt{m_c^{\text{th}}(2m_h + m_c^{\text{th}})}}\mathbf{S} + \frac{m_h}{2m_h + m_c^{\text{th}}}\mathbf{k}_4, \quad \text{(A14)}$$

where m_c^{th}, Z, and W are defined by

$$m_c^{\text{th}} = \frac{\hbar^2 k_{3\text{th}}^2}{2E_{3\text{th}}}, \quad \text{(A15)}$$

$$Z(k_4) = -\frac{3E_g}{2} + \sqrt{\frac{E_g^2}{4} + \frac{\hbar^2 k_4^2 E_g}{2m_l}} - \frac{\hbar^2 k_4^2}{2(2m_h + m_c^{\text{th}})}, \quad \text{(A16)}$$

$$W(k_4, \mathbf{R}, \mathbf{S}) = \frac{E_g}{2} - \sqrt{\frac{E_g^2}{4} + \frac{\hbar^2 k_3^2 E_g}{2m_c}} + \frac{\hbar^2 k_3^2}{2m_c^{\text{th}}}. \quad \text{(A17)}$$

The resulting expression for the CHLH transition rate per unit volume is given by

$$r_{\text{CHLH}} = \frac{1}{2\pi^2}\int_{k_{4\text{th}}}^{\infty} k_4^2 \mathrm{d}k_4 \Omega_{\text{CHLH}}(k_4), \quad \text{(A18)}$$

where $\Omega(k_4)$ is the rate of the CHLH Auger transition with a particular final state 4:

$$\Omega_{\text{CHLH}} = \frac{e^4}{32\pi^5\hbar^3\varepsilon^2\varepsilon_0^2}\frac{m_h(m_h + m_c^{\text{th}})^2}{\sqrt{m_c^{\text{th}}(2m_h + m_c^{\text{th}})^3}} \times \int \frac{|M|^2 P s^4 \sin^4\theta_1 \sin^3\theta_2 \sin^2\theta_3 \sin\theta_4}{\left|1 - \frac{m_c^{\text{th}} m_h}{\hbar^2(m_h + m_c^{\text{th}})}\frac{1}{s}\frac{\partial W}{\partial s}\right|} \times \mathrm{d}\theta_1 \mathrm{d}\theta_2 \mathrm{d}\theta_3 \mathrm{d}\theta_4 \mathrm{d}\theta_5. \quad \text{(A19)}$$

Ω_{CCHC} and Ω_{CHLH} are plotted as functions of the energy of the final state 4 in Figs. 13 and 14 for two different excess carrier densities. To make the comparison easier, the horizontal axis values correspond to E_4 for the CCHC transition and to $E_g - E_4$ for the CHLH transition. It is clear that for both carrier densities considered in Figs. 13 and 14, the CCHC transition dominates [one should keep in mind that integration in Eqs. (A10) and (A18) includes a geometrical factor k_4^2, therefore, the states with larger E_4 contribute more to the total recombination rate]. Another important observation one can make is that most transitions occur away from the threshold for the CCHC process. Near threshold transitions dominate only for the CHLH process at small excess carrier densities.

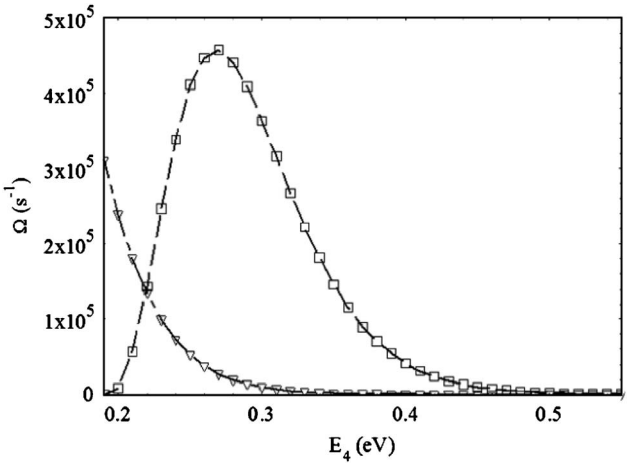

Fig. 13. Auger transition rate Ω versus the energy of the final state four for the excess carrier density $n = 2.47 \times 10^{16}$ cm^{-3}. The CCHC process results from Eq. (A11) (squares), and the CHLH process results from Eq. (A19) (triangles).

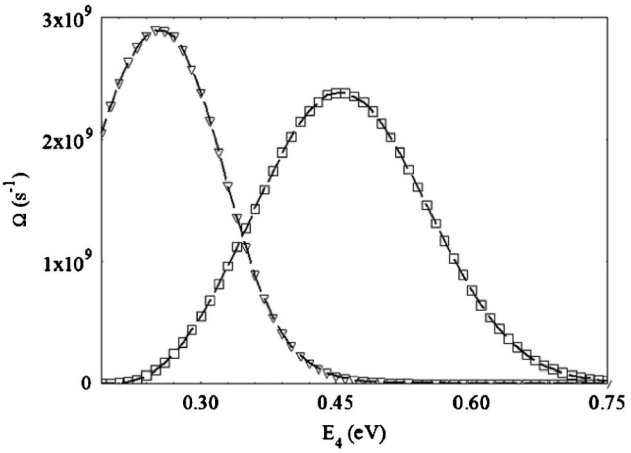

Fig. 14. Auger transition rate Ω versus the energy of the final state four for the excess carrier density $n = 1.72 \times 10^{18}$ cm^{-3}. The CCHC process results from Eq. (A11) (squares), and the CHLH process results from Eq. (A19) (triangles).

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support of the NSF, ECS 0524533, and AFOSR under contract FA95500410200.

REFERENCES AND NOTES

1. D. A. Miller, C. T. Seaton, M. E. Prise, and S. D. Smith, "Band-gap-resonant nonlinear refraction in III-V semiconductors," Phys. Rev. Lett. **47**, 197–200 (1981).
2. E. W. Van Stryland, M. A. Woodall, H. Vanherzeele, and M. J. Soileau, "Energy band-gap dependence of two-photon absorption," Opt. Lett. **10**, 490–492 (1985).
3. B. S. Wherrett, "Scaling rules for multiphoton interband absorption in semiconductors," J. Opt. Soc. Am. B **1**, 67–72 (1984).
4. E. W. Van Stryland, H. Vanherzeele, M. A. Woodall, M. J. Soileau, A. L. Smirl, S. Guha, and T. F. Boggess, "Two photon absorption, nonlinear refraction, and optical limiting in semiconductors," Opt. Eng. (Bellingham) **24**, 613–623 (1985).
5. M. Sheik-Bahae, P. Mukherjee, and H. S. Kwok, "Two-photon and three-photon absorption coefficients of InSb," J. Opt. Soc. Am. B **3**, 379–385 (1986).
6. M. Sheik-Bahae, T. Rossi, and H. S. Kwok, "Frequency dependence of the two-photon absorption coefficient in InSb: tunneling effects," J. Opt. Soc. Am. B **4**, 1964–1969 (1987).
7. E. Van Stryland and L. Chase, "Two photon absorption: inorganic materials," in *Handbook of Laser Science and Technology; Supplement 2: Optical Materials*, Sec. 8, M. Weber, ed. (CRC Press, 1994), pp. 299–328.

8. S. W. Kurnick and J. M. Powell, "Optical absorption in pure single crystal InSb at 298 and 78 K," Phys. Rev. **116**, 597–604 (1959).
9. M. P. Hasselbeck, E. W. Van Stryland, and M. Sheik-Bahae, "Dynamic band unblocking and leakage two-photon absorption in InSb," Phys. Rev. B **56**, 7395–7403 (1997).
10. V. Chazapis, H. A. Blom, K. L. Vodopyanov, A. G. Norman, and C. C. Phillips, "Midinfrared picosecond spectroscopy studies of Auger recombination in InSb," Phys. Rev. B **52**, 2516–2521 (1995).
11. P. T. Landsberg, *Recombination in Semiconductors* (Cambridge U. Press, 1991).
12. A. R. Beattie, "Auger transitions in semiconductors and their computation," J. Phys. C **18**, 6501–6515 (1985).
13. P. T. Landsberg and A. R. Beattie, "Auger effect in semiconductors," J. Phys. Chem. Solids **8**, 73–75 (1959).
14. C. Kittel, *Introduction to Solid State Physics*, 7th ed. (Wiley, 1996).
15. A. Haug, "Carrier density dependence of Auger recombination," Solid-State Electron. **21**, 1281–1284 (1978).
16. M. Combescot and R. Combescot, "Auger recombination in direct-gap semiconductors: effect of anisotropy and warping," Phys. Rev. B **37**, 8781–8790 (1988).
17. M. G. Burt, S. Brand, C. Smith, and R. A. Abram, "Overlap integrals for Auger recombination in direct-bandgap semiconductors: calculation for conduction and havy-hole bands in GaAs and InP," J. Phys. C **17**, 6385–6401 (1984).
18. M. Takeshima, "Auger recombination in InAs, GaSb, InP, and GaAs," J. Appl. Phys. **43**, 4114–4119 (1972).
19. E. O. Kane, "Band structure of indium antimonide," J. Phys. Chem. Solids **1**, 249–261 (1957).
20. M. Cardona and F. H. Pollak, "Energy-band structure of germanium and silicon: the k·p method," Phys. Rev. **142**, 530–543 (1966).
21. J. R. Chelikowski and M. L. Cohen, "Nonlocal pseudopotential calculations of the electronic structure of eleven diamond and zinc-blend semiconductors," Phys. Rev. B **14**, 556–582 (1976).
22. P. Scharoch and R. A. Abram, "A method for determining the overlap integrals used in calculations of Auger transition rates in semiconductors," Semicond. Sci. Technol. **3**, 973–978 (1988).
23. P. O. Lowdin, "A note on the quantum-mechanical perturbation theory," J. Chem. Phys. **19**, 1396–1401 (1951).
24. A. R. Beattie, R. A. Abram, and P. Scharoch, "Realistic evaluation of impact ionisation and Auger recombination rates for ccch transition in InSb and InGaAsP," Semicond. Sci. Technol. **5**, 738–744 (1990).
25. A. R. Beattie and A. M. White, "An analytic approximation with a wide range of applicability for electron initiated Auger transitions in narrow-gap semiconductors," J. Appl. Phys. **79**, 802–813 (1996).
26. M. E. Flatte, C. H. Grein, T. C. Hasenberg, S. A. Anson, D. J. Jang, J. T. Olesberg, and T. F. Boggess, "Carrier recombination rates in narrow-gap InAs/$Ga_{1-x}In_xSb$-based superlattices," Phys. Rev. B **59**, 5745–5750 (1999).
27. A. Haug, "Auger recombination in direct-gap semiconductors: band-structure effects," J. Phys. C **16**, 4159–4172 (1983).
28. L. A. Almazov, A. I. Liptuga, V. K. Malyutenko, and L. L. Fedorenko, Fiz. Tekh. Poluprovodn. (S.-Peterburg) **14**, 1940 (1980) [Sov. Phys. Semicond. **14**, 1154 (1980)].
29. D. Yevick and W. Bardyszewski, "An introduction to non-equilibrium many-body analyses of optical processes in III-IV semiconductors," in *Semiconductors and Semimetals*, R. K. Willardson and A. C. Beer, eds. (Academic, 1993), Vol. 39, pp. 318–388.
30. M. Sheik-bahae, D. J. Hagan, and E. W. Van Stryland, "Dispersion and band-gap scaling of the electronic Kerr effect in solids associated with two-photon absorption," Phys. Rev. Lett. **65**, 96–99 (1989).
31. M. Sheik-Bahae, D. C. Hutchings, D. J. Hagan, and E. W. Van Stryland, "Dispersion of bound electronic nonlinear refraction in solids," IEEE J. Quantum Electron. **QE-27**, 1296–1309 (1991).
32. W. Zawadzki, "Electron transport phenomena in small-gap semiconductors," Adv. Phys. **23**, 435–522 (1974).
33. K. Seeger, *Semiconductor Physics. An Introduction*, 3rd ed. (Springer–Verlag, 1985).
34. B. S. Werrett, A. C. Walker, and F. A. P. Tooley, *Optical Nonlinearities and Instabilities in Semiconductors* (Academic, 1988), pp. 239–272.
35. A. A. Said, M. Sheik-Bahae, D. J. Hagan, T. H. Wei, J. Wang, J. Young, and E. W. Van Stryland, "Determination of bound-electronic and free-carrier nonlinearities in ZnSe, GaAs, CdTe, and ZnTe," J. Opt. Soc. Am. B **9**, 405–414 (1992).
36. E. J. Johnson, "Absorption near the fundamental edge," in *Semiconductors and Semimetals*, R. K. Willardson and A. C. Beer, eds. (Academic, 1967), Vol. 3, pp. 153–258.
37. V. Dubikovskiy, "Optical limiting: numerical modeling and experiment," Ph.D. dissertation (University of Central Florida, 2003).
38. N. V. Tabiryan, B. Ya. Zel'dovich, M. Kreuzer, T. Vogeler, and T. Tschudi, "Higher-dimensionality caustics owing to competing reorientation of a liquid crystal by laser beams," J. Opt. Soc. Am. B **13**, 1426–1969 (1996).
39. D. I. Kovsh, S. Yang, D. J. Hagan, and E. W. Van Stryland, "Nonlinear optical beam propagation for optical limiting," Appl. Opt. **38**, 5168–5180 (1999).
40. F. E. Hernández, S. Yang, E. W. Van Stryland, and D. J. Hagan, "High dynamic range cascaded-focus optical limiter," Opt. Lett. **25**, 1180–1182 (2000).
41. M. Mohebi, P. F. Aiello, G. Reali, M. J. Soileau, and E. W. Van Stryland, "Self-focusing in CS_2 at 10.6 μm," Opt. Lett. **10**, 396 (1985).
42. The authors recently became aware of a recent publication by S. Krishnamurthy, Z. G. Yu, L. P. Gonzalez, and S. Guha, "Accurate evaluation of nonlinear absorption coefficients in InAs, InSb, and HgCdTe alloys," J. Appl. Phys. **101**, 113104 (2007); these results are consistent with those reported herein.
43. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C. The Art of Scientific Computing*, 2nd ed. (Cambridge U. Press, 1992).