Population inversion in a *p*-doped quantum well with reduced photon energy

Chi-Ken Lu,[1] Hsin-Fei Meng,[1,*] and Pin Han[2]
[1]*Institute of Physics, National Chiao Tung University, Hsinchu, Taiwan, Republic of China*
[2]*Institute of Precision Engineering, National Chung Hsing University, Taichung, Taiwan, Republic of China*
(Received 15 February 2005; revised manuscript received 4 May 2006; published 21 July 2006)

We study a multilayer silicon-germanium quantum well structure doped with acceptor impurities for resonant-state lasers capable of emitting photons of energy below 4 meV (1 THz). Unlike previous proposals on terahertz lasers in doped silicon-germanium quantum wells, the emitted photon energy does not need to exceed the acceptor binding energy, which is tens of meV. The energy constraint is relaxed by placing the acceptor impurity levels and the quantum well subband continuum in separate layers of different germanium compositions. We calculate the nonequilibrium behaviors of the holes in detail and demonstrate that population inversion between strain-split impurity levels can be built for sufficiently high acceptor densities under the application of a moderate dc electric field at about ten kelvins.

DOI: 10.1103/PhysRevB.74.035328 PACS number(s): 72.10.−d, 42.55.−f, 78.45.+h

I. INTRODUCTION

Electromagnetic radiation in the 1–10-THz range, corresponding to wavelengths of 30–300 μm, has important applications in optical communication, medical diagnosis, and radio astronomy. There has been significant progress in the realization of THz radiation sources.[1] For example, broadband THz radiation with output power up to 20 W is generated from subpicosecond electron bunches in an accelerator. It is applied to image the distribution of a specific protein or water in tissue and buried layers in semiconductors.[2] Despite the high power output this approach is not very convenient for compact applications. Owing to rapid advances in semiconductor technology, many earlier theoretical ideas based on solid-state THz sources can now be implemented. For instance, the Bloch oscillation (BO) is not possible for bulk crystals because of scatterings. The structure of superlattices makes it possible for BO to take place due to a shrinking zone boundary. By precise control of the period of the superlattice, emission at 1.7 THz was observed in a GaAs/AlGaAs superlattice due to BO.[3] However, BO has not been proved to be viable for a THz laser. In addition, there is growing interest in quantum cascade lasers (QCL's) where THz radiation results from intersubband electroluminescence.[4] QCL's have been demonstrated to emit cw THz radiation at liquid-nitrogen temperature in GaAs/AlGaAs.[5] However, to date there is no QCL able to emit radiation below 2.9 THz.[6]

One promising way to realize semiconductor THz lasers is resonant-state lasers[7,8] (RSL's) based on doped quantum well (QW) structures,[9,10] whose operation involves strain-induced resonant states and pumping by an electric field. A THz transition between higher and lower acceptor states has been observed.[9,10] In RSL's with one single QW, the two degenerate valence bands are split by symmetry-lowering external strain caused by external pressure or lattice mismatch. The splitting also removes the degeneracy of the hydrogen-like acceptor localized states, and therefore two localized states are formed with energy levels attached to each split band. As the strain is so strong that the energy splitting exceeds the binding energy of the acceptor, one of the two localized states becomes resonant with the band to which the other localized state is attached. The coincidence in energy leads to resonant scattering between the continuous and localized states. A population inversion between the two localized states can be achieved by resonant capture of the holes under an electric field.

In the previous approach to RSL's[7–10] there is a serious constraint on the emitted photon energy. In single QW's the acceptor level splitting needs to be greater than the impurity binding energy in order to have a resonant state. The photon energy therefore must be larger than the binding energy, which is several tens of meV [15 meV for Ge and 50 meV for Si (Ref. 11)] corresponding to more than 10 THz. In this paper we present a concept of silicon-germanium QW RSL's which is free of such a constraint. Instead of one single QW, in our structure the continuous and localized states are in different layers and the resonance can be controlled by independent strains in different layers. Therefore resonant scattering can occur even if the energy splitting is smaller than the acceptor binding energy. Silicon-germanium alloy is chosen as the material system for this concept because of its low absorption in the THz range and easy integration with Si electronics. We calculate the energies of the localized acceptor levels and continuous subband levels (indicated as "continuum" below) and give the relation between the emitted photon energy and the structure parameters. In order to show that population inversion can be realized under practical experimental conditions, we construct a comprehensive theoretical model for the nonequilibrium behaviors of holes in the QW structure and study in detail the the dynamical behaviors of the holes with an external pumping field. Our results indicate that emission as low as 1 THz can be obtained in the QW structure with reasonable germanium compositions under a dc electric field of about 100 V/cm at 10 K.

The paper is organized as follows. Section II introduces the QW structure featuring the flexible control of lasing frequency by germanium compositions. In particular we show how the QW structure is able to emit radiation of frequency at 1 THz. In Sec. III the wave functions and resonant transition between localized states and subband continuum are calculated. In Sec. IV the Boltzmann equation and rate equation

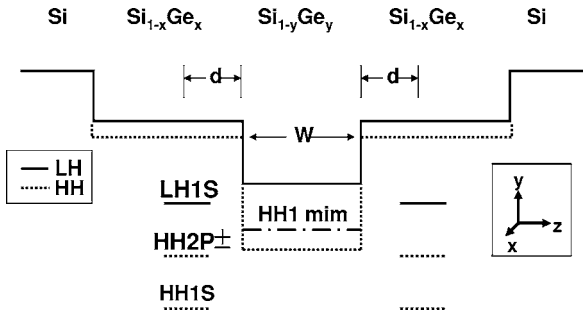

FIG. 1. The band edge profiles for light hole (LH, solid line) and heavy hole (HH, dashed line) of the proposed QW structure are shown. Both x and y directions are perpendicular to the crystal growth direction z. x, and y are the germanium compositions for the well and barrier layers, respectively. W is the well width. The energies of the strain-split acceptor levels LH1S, HH1S, and HH2P± relevant for the THz laser are also shown. d is the distance between the dopant and the boundary of the well. The HH1 minimum is indicated by the dash-dotted line. The inset defines the directions in the system.

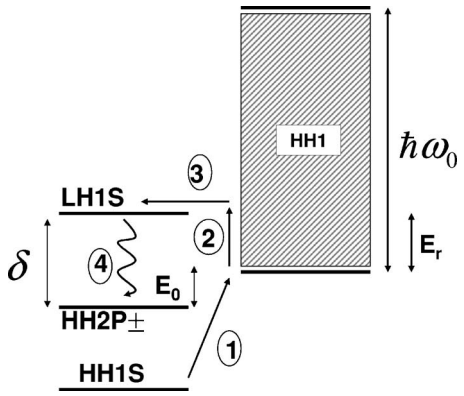

FIG. 2. Schematic four-level operation of the THz laser. The operation involves three acceptor states LH1S, HH1S, and HH2P± as well as the HH1 continuum (shaded region). The four major processes are also indicated. Process 1 indicates the field ionization of HH1S through the barrier to reach the HH1 minimum. The low-energy holes in HH1 are pumped toward resonant states by an electric field in process 2. Process 3 represents resonant capture of continuum holes of energy E_r to meta-stable LH1S. The radiative decay by stimulated emission into the lower localized state HH2P± is denoted by process 4.

are employed to study the population inversion. In Sec. V we present the final results on the conditions for the formation of a population inversion in the QW structure. Section VI draws the conclusion.

II. QUANTUM WELL STRUCTURE AND PUMPING MECHANISM

The profile of valence band edge diagram along crystal growth direction (z direction) of the proposed QW structure is shown in Fig. 1. For clarity the sign of the energy is reversed. The splitting of heavy-hole and light-hole band edges is due to strain caused by lattice mismatch between SiGe alloy and Si. The strain can be linearly related to the germanium compositions in the alloy. The two $Si_{1-x}Ge_x$ layers sandwiching the central $Si_{1-y}Ge_y$ layer have identical profiles and are δ doped with identical acceptor density n_a. The profile has been designed to be symmetric for simpler theoretical treatments.

As can be seen in the profile of the heavy-hole band edge in Fig. 1, holes are confined in the central layer in the z direction due to potential barriers constituted by the two identical $Si_{1-x}Ge_x$ layers at two sides. Series of subbands are formed due to the confinement. We label the energy minimum of the first heavy-hole subband (HH1), which is a function of the well width W, by the dash-dotted line in the central layer. In addition there is a series of localized acceptor levels attached to the heavy-hole band edge in each δ-doped layer. We focus on the low-lying heavy-hole $2p_{\pm 1}$ level (HH2P±), labeled by a short dashed line, and the light-hole acceptor $1s$ level (LH1S), labeled by a short solid line. LH1S and HH2P± have opposite parity and hence are expected to give the strongest intensity of radiation among all possible transitions. Besides the acceptor $1s$ level attached to the heavy-hole band edge is the very lowest state for holes in the system and is labeled by HH1S, which is shown below HH2P± in Fig. 1. With precise control of germanium compositions x and y in the QW structure, the heavy-hole and light-hole band edges as well as the localized acceptor levels can be adjusted to have the relative energies required for THz lasers.

Now we consider the pumping mechanism of the holes under an external electric field $\mathbf{F}$ (strength F) perpendicular to the z direction—say, the x direction. The physical picture is shown in Fig. 2. HH2P± is below LH1S and the minimum of HH1 by δ and E_0, respectively. Note that in our problem δ must be larger than E_0 to have resonance between HH1 and LH1S. At zero temperature all the holes stay in the low-lying HH1S state without field. When the external electric field is applied, some holes on HH1S are initially field ionized. Then more holes are excited to HH1 through impact ionization and acquire more kinetic energy until the occurrence of phonon scattering. The processes of field ionization and pumping of holes are denoted by process 1 and 2, respectively, in Fig. 2. Another channel for slowing down the holes in HH1 is the resonance capture by LH1S. The transition between heavy-hole and light-hole states, denoted by process 3, is facilitated by the off-diagonal matrix element[8] of the Luttinger-Kohn Hamiltonian to be discussed below. As the occupation of higher LH1S grows with increasing external field and the lower HH1S and HH2P± are gradually depleted by impact ionization, a population inversion is expected. Emission of THz photons, indicated by process 4, will take place due to the radiative decay of holes from LH1S to HH2P±.

The resonance of the localized state and the continuum is achieved by raising the strain of the lattice so that the localized state is lifted to immerse within the continuum. In the previous works on QW RSL's[9,10] this acceptor impurity is doped in the same layer as the continuous states, so E_0 is simply the binding energy. Apparently in such a case the strain splitting δ must exceed the binding energy, corresponding to a lower bound of photon energy. In this work we

spatially separate the acceptor impurity and quantum well, so the relative energy between the localized state and continuum has a much higher flexibility by adjusting the compositions x and y and the well width W. As a result no matter how small δ is we can always adjust the QW structure such that $E_0 < \delta$. However, the only lower bound for the emitted photon energy is the energy shift of the resonant state caused by the perturbation of the continuum as discussed in Sec. III A. Hence our proposed structure is able to emit photons of energy less than the binding energy which is usually several tens of meV (12 THz in the case of Si) and is expected to fulfill the needs of solid-state optical sources of several THz or even sub-THz range. Because the relative energies of the localized and continuous states are crucial to the laser operation, below we calculate the quantitative relations between the relevant levels in the QW structure and QW parameters like the width W and germanium compositions. Even though the acceptor levels are outside the central well, there is no difficulty for the holes in the central well to be resonantly captured by the acceptor as long as there is an overlap between the wave functions of the acceptor levels and HH1. In order for the above picture to be valid, it is important to choose an intermediate value for the distance between the dopants and quantum well. The distance should be neither so large relative to the acceptor Bohr radius that there is no overlap between the acceptor level and quantum well level nor so small that the acceptor level itself becomes heavily influenced by the well.

III. LOCALIZED AND CONTINUOUS STATES

A. Subband and impurity wave functions

In this subsection we calculate the wave functions and energies of the relevant states. We first consider a perfect crystal. The wave functions for the heavy-hole and light-hole bands can be represented by the eigenfunctions of the Luttinger-Kohn Hamiltonian[12] H_{LK} in the Bloch function basis $\{u_{3/2}, u_{1/2}, u_{-1/2}, u_{-3/2}\}$, which is the periodic sum of the atomic orbitals with total angular momentum quantum number $j=\frac{3}{2}$. The subscripts stand for their z component j_z of total angular momentum j. The column vector Ψ formed by the envelope functions $\{\varphi_{3/2}(r), \varphi_{1/2}(r), \varphi_{-1/2}(r), \varphi_{-3/2}(r)\}$ is the eigenfunction of H_{LK}. The true wave function $\psi(r)$ of the state is given by $\psi(r)=\Sigma_\tau \varphi_\tau(r) u_\tau$. The Luttinger-Kohn Hamiltonian can be written as

$$H_{LK} = \frac{\hbar^2}{2m_0} \begin{pmatrix} \hat{a}_+ & \hat{b} & \hat{c} & 0 \\ \hat{b}^\dagger & \hat{a}_- & 0 & \hat{c} \\ \hat{c}^\dagger & 0 & \hat{a}_- & -\hat{b} \\ 0 & \hat{c}^\dagger & -\hat{b}^\dagger & \hat{a}_+ \end{pmatrix} \begin{matrix} j_z = \frac{3}{2} \\ j_z = \frac{1}{2} \\ j_z = -\frac{1}{2} \\ j_z = -\frac{3}{2} \end{matrix}, \tag{1}$$

and the matrix elements are

$$\hat{a}_+ = -\hat{k}_z(\gamma_1 - 2\gamma)\hat{k}_z - (\gamma_1 + \gamma)(\hat{k}_x^2 + \hat{k}_y^2), \tag{2}$$

$$\hat{a}_- = -\hat{k}_z(\gamma_1 + 2\gamma)\hat{k}_z - (\gamma_1 - \gamma)(\hat{k}_x^2 + \hat{k}_y^2), \tag{3}$$

$$\hat{b} = \sqrt{3}(\hat{k}_x - i\hat{k}_y)(\gamma\hat{k}_z + \hat{k}_z\gamma), \tag{4}$$

$$\hat{c} = \sqrt{3}\gamma(\hat{k}_x - i\hat{k}_y)^2. \tag{5}$$

m_0 is the free electron mass and $\hat{k}_i = i\frac{\partial}{\partial x_i}$, $i=x,y,z$, are operators for the envelope functions. γ_1, γ_2, and γ_3 are material-dependent Luttinger parameters, and $\gamma=(2\gamma_2+3\gamma_3)/5$. For crystals with translational invariance the envelope functions are all proportional to plane waves $e^{i\mathbf{k}\cdot\mathbf{r}}$ and the above operators turn into c numbers. Diagonalization of the matrix gives the spectrum $E_\pm(\mathbf{k})$ which possesses fourfold degeneracy at the band edge. The sign $\pm$ indicates that there are two branches: the heavy-hole and light-hole bands. The spectrum $E_\pm(\mathbf{k})$ is given by

$$E_\pm(\mathbf{k}) = \frac{\hbar^2}{m_0}\left[\gamma_1\frac{k^2}{2} \pm \sqrt{\gamma_2^2 k^4 + 3(\gamma_3^2 - \gamma_2^2)(k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2)}\right]. \tag{6}$$

When the perfect crystal is subject to a stress due to external strain or lattice mismatch the crystal symmetry is lowered and the fourfold degeneracy at the valence band edge is split into two twofold degeneracies. If the strain is along the [001] axis, which is parallel to the z direction, this effect is to add a strain term V_{st} to the Hamiltonian.[13] It can be represented by the diagonal matrix

$$V_{st} = \begin{pmatrix} \zeta & 0 & 0 & 0 \\ 0 & -\zeta & 0 & 0 \\ 0 & 0 & -\zeta & 0 \\ 0 & 0 & 0 & \zeta \end{pmatrix}. \tag{7}$$

The coincidence of heavy-hole and light-hole bands at the band edge is split by the strain factor ζ which is proportional to external force and dependent on the direction of strain. In QW's the strain results from the lattice mismatch between Si and SiGe alloy. In epitaxially grown SiGe QW structure on Si substrate, the lattice constant of the whole structure is fixed by the Si lattice constant. Because the natural lattice constant of SiGe alloy is different from Si, there must be a strain in the alloy to force the lattice constant to match Si. The relation between valence band splitting ζ due to strain and the germanium composition t in $Si_{1-t}Ge_t$ alloy was studied before.[14] The expression in eV is $\zeta(t)=0.01+0.2t-\frac{1}{4}\sqrt{0.0016+0.0074t+0.24t^2}$. In our proposed QW structure the germanium compositions vary in the z direction and the hence the strain factor ζ is a function of z.

For an acceptor in the stressed crystal we shall add the Coulomb potential V_I due to the charged center:

$$V_I(r) = v_I(r)I = \frac{1}{4\pi\epsilon}\frac{e}{r}I, \tag{8}$$

where ϵ is the dielectric constant. r is the distance from the acceptor. I represents the 4×4 identity matrix. In the high-strain limit the off-diagonal coupling $\hat{b}$ and $\hat{c}$ can be considered as perturbations and H_{LK} becomes approximately diagonal with twofold degeneracy for heavy and light holes. The

resultant localized states can also be divided into two subgroups like the band states.

After reviewing the bulk crystals we can extend the discussions to the states in QW structures shown in Fig. 1. Even without strain the valence band edge depends on the germanium compositions,[15] described by $V_b(z)=v_b(z)I$. For $Si_{1-t}Ge_t$ alloy grown on Si, the valence band offset in eV can be written as $v_b=0.84t$. The total band edge profile in Fig. 1 comes from the sum of $V_b(z)$ and $V_{st}(z)$. The Luttinger parameters have different values in different layers; hence, they are functions of z. The homogeneity of those parameters is assumed within each silicon-germanium layer, and their values are determined by linear interpolation between pure Si and pure Ge. The heavy-hole and light-hole subbands in the structure can be expressed by the total Hamiltonian H:

$$H = H_{LK} + V_b(z) + V_{st}(z). \tag{9}$$

Note that $z=0$ is at the center of well, so there is a parity symmetry with respect to $z \to -z$ in this problem. Here we separate H_{LK} into diagonal and off-diagonal parts, labeled by H^0_{LK} and H^1_{LK}, respectively. The wave functions for HH1 emerge from eigenfunctions of the diagonal parts $H^0=H^0_{LK}+V_b(z)+V_{st}(z)$ of the full Hamiltonian H. The off-diagonal heavy-hole–light-hole mixing H^1_{LK} will be considered later as a perturbation. The unperturbed Schrödinger equation can be written as

$$H^0\Psi = \epsilon\Psi. \tag{10}$$

We solve this to obtain the HH1 envelope functions Ψ of the wave functions $\psi_{\mathbf{k}}$ with eigenvalues $\varepsilon(\mathbf{k})$. On the other hand, the localized states ϕ^{LH}_{1s} and $\phi^{HH}_{2p_\pm}$ with respective eigenvalues E^{LH}_{1s} and $E^{HH}_{2p_\pm}$ are eigenstates of the Hamiltonian $H^0_{LK}+V_I(r)+[V_b(z)+V_{st}(z)]_{z=z_0^\pm}$. Here $z_0^\pm \equiv \pm\left(\frac{W}{2}+d\right)$ denote the positions of the acceptors. The implicit assumption is that the Coulomb potential V_I has little effect on the subband wave functions while the nonuniform strain is irrelevant to the localized state. The above approximations are justified by the condition that the distance between the dopant and quantum well boundary d as well as the thickness of outer $Si_{1-x}Ge_x$ layers be both larger than the acceptor Bohr radius. The equations turn out to be the typical one-dimensional potential well problem for the subband and hydrogen atom problem for the localized states. The energy spectra for the relevant states are shown in Fig. 3. The explicit wave functions for HH1 can be expressed as

$$\psi_{\mathbf{k}}(\rho,z) = \frac{1}{\sqrt{A}} g(z) e^{i\mathbf{k}\cdot\vec{\rho}} u_{\pm 3/2}, \tag{11}$$

where the $\pm$ sign in the wave functions reflects the twofold degeneracy guaranteed by time-reversal symmetry in the absence of a magnetic field and the envelope function $g(z)$ has even parity to yield the lowest energy of all subbands. A is the QW area. $\vec{\rho}=(x,y)$ is the in-plane coordinate. The acceptor wave functions localized at $z=z_0^\pm$ and $\vec{\rho}=0$ are of the form

$$\phi^{LH}_{1s}(\vec{\rho},z) = \varphi_{1s}[\vec{\rho}, z - z_0^\pm] u_{\pm 1/2}, \tag{12}$$

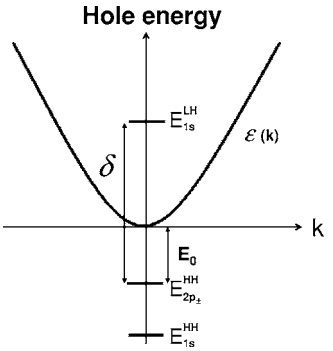

FIG. 3. Spectrum of the diagonal part H^0 of the full Hamiltonian H. The acceptor states of interest and the continuous HH1 are shown. The binding energy for HH2P± and the emitted photon energy are denoted by E_0 and δ, respectively. Their values and corresponding variational Bohr radius are shown in Table I. $\varepsilon(\mathbf{k})$ is the spectrum of the subband HH1.

$$\phi^{HH}_{2p_\pm}(\vec{\rho},z) = \varphi_{2p_\pm}[\vec{\rho}, z - z_0^\pm] u_{\pm 3/2}, \tag{13}$$

where $\pm$ stands for $z>0$ and $z<0$, respectively. We use the hydrogenic trial functions

$$\varphi_{1s}(\vec{\rho},z) = \frac{1}{\sqrt{\pi a^2 b}} \exp\left(-\sqrt{\frac{\rho^2}{a^2}+\frac{z^2}{b^2}}\right), \tag{14}$$

$$\varphi_{2p_\pm}(\vec{\rho},z) = \frac{1}{2\pi a^4 b} \rho e^{i\phi} \exp\left(-\sqrt{\frac{\rho^2}{a^2}+\frac{z^2}{b^2}}\right). \tag{15}$$

a (in-plane Bohr radius) and b (out-of-plane Bohr radius) are variation parameters for minimizing their energy, and ϕ is the polar angle in the x-y plane. ρ is the modulus $|\vec{\rho}|$. Variational calculations are performed to obtain the acceptor level splitting δ and the difference E_0 between HH2P± and HH1 minima. Variational calculations are performed to obtain the acceptor level binding energy. The resultant binding energies and the variational Bohr radius of the levels of interest are shown in Table I.

Next we consider the corrections to the impurity states resulting from the QW confinement potential as well as the off-diagonal couplings with the HH1 continuum. Such corrections are necessary for having a more precise prediction on the emitted photon energy. Here we focus on the corrections to the binding energy of LH1S, which is resonant with the continuum. Note that the binding energy is relative to the barrier, not the quantum well continuum. The expressions for the corrections ΔE_{1s} are given below, and the details of derivation are presented in Appendix A:

TABLE I. Binding energies and variational parameters of the localized acceptor states.

Level	Binding energy (meV)	Variational Bohr radius (nm)
HH1S	20.9	$a=3.6$, $b=3.0$
HH2P±	4.9	$a=4.4$, $b=3.9$
LH1S	21.0	$a=2.8$, $b=4.0$

$$\Delta E_{1s} = \Delta + P\frac{A}{(2\pi)^2}\int d\mathbf{k}\frac{|\alpha_{\mathbf{k}}|^2}{E_{1s}-\varepsilon_{\mathbf{k}}}. \quad (16)$$

Δ is the correction due to the confinement potential while the integral due to the coupling with the continuum. P stands for the Cauchy principal-value integration. $\alpha_{\mathbf{k}}$ and Δ are given by

$$\alpha_{\mathbf{k}} = \langle\varphi_{1s}|\hat{c}|\psi_{\mathbf{k}}\rangle, \quad (17)$$

$$\Delta = \langle\varphi_{1s}|[v_C(z) - v_C(z_0)]|\varphi_{1s}\rangle. \quad (18)$$

Note that only the off-diagonal elements involving k_x and k_y are considered because the resonance requires a large in-plane momentum. The confinement potential v_C is the diagonal element of V_b+V_{st} belonging to light-hole states. The correction due to the confinement potential Δ is negligible in the present case because a very small portion of the impurity wave function for the impurity falls on the QW region and the confinement potential is small compared to the impurity binding energy. In fact our calculation shows that this correction on E_{1s} is less than 0.1%. However, this effect for the case of smaller binding energy is important such as the shallow donors located in the barrier near the quantum well.[16] The second term resembles the formula for a second-order perturbation. Even though still only about 10% of E_{1s}, it provides significant corrections in case of a small emitted photon energy. The smallness of the corrections is reasonable since the light-hole localized states and the heavy-hole continuum have a small overlap and they can couple to each other only though the off-diagonal elements of H_{LK} which is treated as a perturbation in the high-strain limit.[8] The QW continuum and the HH2P± are assumed to be unaffected by the perturbation.

To be specific we consider the case which gives a radiation frequency of 1 THz (4 meV). We set δ=6 meV in Fig. 2. δ determines the germanium composition x in the $Si_{1-x}Ge_x$ layer in Fig. 1. The corresponding x is 0.088. Such an arrangement gives an unperturbed binding energy of LH1S 21 meV by a variational calculation. This value is shifted to 23 meV after the correction in Eq. (16) is put in. Consequently the real emitting photon energy is 4 meV as expected. Next we determine the width W and germanium composition y of the central layer such that the HH1 minimum lies equally between LH1S and HH2P±. In other words E_0 is set to 2 meV in Fig. 2. To meet the requirement we choose the composition y by coinciding the HH band profile in the $Si_{1-y}Ge_y$ layer (dashed line in Fig. 1) with HH2P± in the $Si_{1-x}Ge_x$ layers. This gives y=0.094. W is so determined such that the HH1 minimum (dash-dotted line in Fig. 1) is 2 meV above due to the spatial quantization. By solving the potential well problem with the barrier height given by the valence band offset, the well width W is 11.7 nm. The energy levels relevant to the laser operation are shown as functions of germanium composition x in $Si_{1-x}Ge_x$ layers for fixed well width W in Fig. 4. The valence band edges in Si layers are taken as zero. The parameters used in the calculation are summarized in Table II. The x dependence of the LH and HH band edges is due to the collective contributions from the x-dependent intrinsic band edge offset of SiGe alloy and x-dependent strain. The LH1S and HH2P± acceptor levels are downward shifted from the band edges by the binding energies calculated by the variational trial wave functions in Eqs. (14) and (15). The HH1 minimum in the central well depends on x through the QW barrier height. The lasing operation is possible only when the HH1 minimum lies between HH2P± and LH1S, as indicated by the arrow in Fig. 4. After having all the levels in the right order of energy, LH1S is immersed within the HH1 continuum as the resonant state.

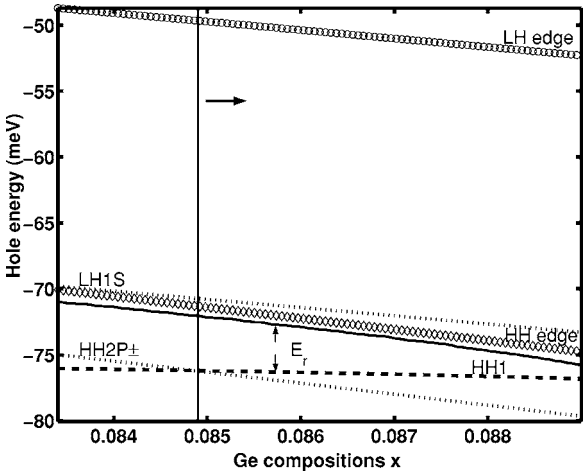

FIG. 4. Unperturbed LH1S (upper dotted line) and HH2P± (lower dotted line) energy levels in $Si_{1-x}Ge_x$ barrier layers are plotted versus germanium composition x. The solid line stands for the LH1S energy with the corrections in Eq. (16). The valence band edge in outer most Si layer is taken as zero energy. The HH1 minimum is also shown as a dashed line and the difference from the perturbed LH1S is denoted by resonance energy E_r. The heavy-hole (diamond) and light-hole (circle) band edges in barrier layers are also plotted for references. The QW width W is 11.7 nm. y=0.094. LH1S becomes a resonant state if the HH1 minimum lies between the two localized states HH2P± and LH1S. The arrow indicates that the resonance condition for the THz laser is satisfied in the region right to the vertical line.

TABLE II. Useful values in the calculation are listed with references. Luttinger parameters and optical phonon energy used here are obtained by interpolation with values between Si and Ge.

Parameter	Value	Description
$(\gamma_1, \gamma_2, \gamma_3)$	(4.22,0.39,1.44)[a]/ (13.38,5.69,4.24)[b]	Luttinger parameters in Si/Ge
$\hbar\omega_0$	53/37[c]	Optical phonon energy for Si/Ge (meV)
ν_A	5[b]	Average optical phonon emission rate (10^{12} s^{-1})
ϱ	2.328[c]	Mass density (g/cm^3)
c	9040[d]	Sound velocity (m/s)
Ξ	9[d]	Deformational potential (eV)

[a]Reference 17.
[b]Reference 7.
[c]Reference 11.
[d]Reference 18.

B. Resonant transition

The hybridization of the localized LH1S and HH1 continuum via the off-diagonal perturbation H_{LK}^{1} leads to a new set of resonant states $\{\Psi_E\}$ labeled by its complex energy $E+i\frac{\Gamma_E}{2}$. The imaginary part is given by

$$\frac{\Gamma_E}{2}=\pi\frac{A}{(2\pi)^2}\int d\mathbf{k}\,\delta(E-\varepsilon_{\mathbf{k}})|\alpha_{\mathbf{k}}|^2. \tag{19}$$

The nonzero imaginary energy Γ_E here represents that Ψ_E is a quasistationary state. More precisely speaking the HH1 holes of momentum $\mathbf{k}$ can be captured by LH1S with the transition rate $W_{\mathbf{k}}^{res}$ for a time interval $\hbar/\Gamma$. The transition rate and the time interval are determined in a self-consistent manner—that is,

$$W_{\mathbf{k}}^{res}=\frac{2}{\hbar}|\alpha_{\mathbf{k}}|^2\frac{\Gamma/2}{[\varepsilon(\mathbf{k})-E_{1s}]^2+\Gamma^2/4}, \tag{20}$$

$$\frac{\Gamma}{\hbar}=\sum_{\mathbf{k}}W_{\mathbf{k}}^{res}. \tag{21}$$

The center of the Lorentzian corresponds to E_{1s} because the resonant state $\Psi_{E_{1s}}$ contains the maximum component of the localized LH1S. For simplicity we regard the unknown Γ in Eq. (20) as close to zero and the Lorentzian is reduced to a δ function. As long as the resultant Γ from Eq. (21) is small compared to the resonance energy $E_r\equiv E_{1s}-\varepsilon(\mathbf{k}=0)$ (see Figs. 2 and 3) of LH1S, this method is self-consistent to obtain Γ.

Next we work out Γ in the small-Γ limit. In other words the resonant transition rate $W_{\mathbf{k}}^{res}$ can be given simply by the Fermi golden rule

$$W_{\mathbf{k}}^{res}=\frac{2\pi}{\hbar}|\alpha_{\mathbf{k}}|^2\delta[\varepsilon(\mathbf{k})-E_{1s}]. \tag{22}$$

In order to obtain an explicit expression for the transition rate we need to calculate the overlap integral $\langle\varphi_{1s}|\psi_{\mathbf{k}}\rangle$. Assuming that the main contribution to this integral comes from the region in the barrier—i.e., $|z|>\frac{W}{2}$—we arrive at

$$\begin{aligned}\langle\varphi_{1s}|\psi_{\mathbf{k}}\rangle&=\frac{1}{\sqrt{A}}\frac{1}{\sqrt{\pi a^2 b}}\int dz g(z)\int dxdy e^{i\vec{k}\cdot\vec{\rho}}\\&\quad\times\exp\left(-\sqrt{\frac{\rho^2}{a^2}+\frac{z^2}{b^2}}\right)\\&=\aleph\sqrt{\frac{16\pi ab^2}{A}}\frac{1}{\eta^2}\left\{e^{-\eta d/b}\frac{1}{\kappa b-\eta}\left[\frac{\kappa b-2\eta}{\eta(\kappa b-\eta)}+\frac{d}{b}\right]\right.\\&\quad\left.+e^{-\kappa d}\left[\frac{(2\eta-\kappa b)}{\eta(\kappa b-\eta)^2}+\frac{(2\eta+\kappa b)}{\eta(\kappa b+\eta)^2}\right]\right\}.\end{aligned} \tag{23}$$

The dimensionless quantity $\eta\equiv\sqrt{1+a^2k^2}$ is introduced. $\aleph$ is to normalize the envelope function $g(z)$ as $\int\aleph^2|g(z)|^2dz=1$. κ is the decay constant of $g(z)$ in the barriers.

Γ as a function of resonance energy E_r is plotted in Fig. 5 for various acceptor in-plane Bohr radii a and separations d. The effect of coupling with the continuum can be investigated through Γ. For a distance d much larger than the Bohr radius, the coupling is diminished due to the decreasing overlap between the impurity state and the continuum. In such a case the formation of a resonant state is impossible. However, the dependence of the coupling on the Bohr radius is determined by two competing factors. Namely, in the z direction the envelope function $g(z)$ of the continuum has a larger overlap with the localized impurity state of larger Bohr radius, while in the x-y plane the continuum of higher kinetic energy can only be coupled to the impurity state of smaller Bohr radius because such a localized state has larger Fourier momentum components. For larger Bohr radius, it is shown in Fig. 5 that Γ is larger at lower E_r while it is smaller at higher E_r.

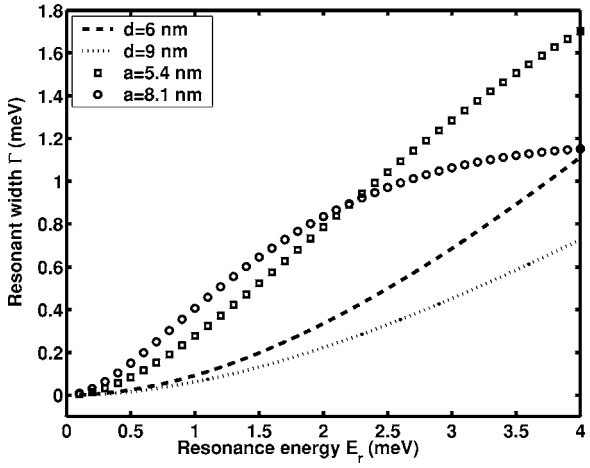

FIG. 5. The energy width Γ of the resonant state is shown as a function of the resonant-state energy E_r measured from the HH1 minimum. Symbol curves correspond to various in-plane Bohr radius a of LH1S orbital with fixed $d=6$ nm. Dashed and dotted lines correspond to various distances d between the acceptor and boundary of the central well with $a=2.7$ nm.

IV. HOLE DISTRIBUTION AND POPULATION INVERSION

So far there is no comprehensive theoretical model for the nonequilibrium behavior of acceptor levels interacting with a subband in QW's. In order to make quantitative predictions of the conditions for hole population inversion, below we construct a model which takes into account all of the relevant physical processes for such a system.

A. Hole statistics at equilibrium

The occupation probabilities of LH1S, HH2P±, HH1S, and HH1 states are indicated by f_1, f_2, f_g, and $f_{\mathbf{k}}$, respectively. In thermal equilibrium the occupation ratio of LH1S to HH2P± is given by the Boltzmann factor—i.e., $f_2/f_1=\exp(-\beta\delta)$. β is the inverse of the product of the Boltzmann constant k_B and temperature T. Moreover, at equilibrium the hole densities are determined by assigning each level its Boltzmann weighting. Note that all the holes are provided by the lowest localized level, and hence we have the following normalization of total holes:

$$n_a f_g + n_a f_1 + \frac{1}{A}\sum_{\mathbf{k}} f_{\mathbf{k}} + n_a f_2 = n_a. \tag{24}$$

When the electric field is turned on holes acquire kinetic energy from the external field and the distribution of holes deviates from the Boltzmann distribution. In order to give a quantitative account of how the nonequilibrium populations depend on the parameters (e.g., field strength F, temperature T, and acceptor density n_a), we need to study the microscopic kinetics governing the transitions among the states.

B. Boltzmann kinematic equation

The strategy for obtaining the nonequilibrium populations is as follows. First we neglect the low-lying HH2P± and HH1S temporarily and solve the kinetics of the subsystem containing HH1 and LH1S in order to obtain the relation between f_2 and $f_{\mathbf{k}}$, with considerations of phonon scattering within HH1 and the resonant transition between the continuous HH1 and the localized LH1S. This is justified because the resonant scattering is much faster than the decay through spontaneous emission from LH1S to HH2P±.[7] Afterwards the occupation probability f_1 of HH2P± is determined by the its balance with the nonequilibrium subband distribution $f_{\mathbf{k}}$ through impact ionization, thermal recombination, and their inverse processes Auger recombination and thermal excitation. Detailed calculations are given below.

For a given number of holes in the subsystem containing HH1 and LH1S, the nonequilibrium distribution $f_{\mathbf{k}}$ in HH1 and occupation of LH1S f_2 are studied by solving the Boltzmann kinetic equation numerically for various electric fields and acceptor densities. In the subsystem the holes in HH1 acquire kinetic energy from the constant electric field $\mathbf{F}$ applied along the x axis. For moderate electric field and low temperature, it is adequate to adopt the concept of *streaming motion*[19] in which the only significant scattering is due to optical phonons (energy $\hbar\omega_0$). This is implemented by introducing a particle drain in momentum space such that once a specific hole drifts with velocity $e\mathbf{F}/\hbar$ through the energy surface $\varepsilon=\hbar\omega_0$ (denoted by Π) in the momentum space, the hole will experience an optical phonon scattering and simultaneously reemerge as a hole of energy less than ϵ_0.[7,8] Hence $f_{\mathbf{k}}=0$ for $\varepsilon(\mathbf{k})\geqslant\hbar\omega_0$. The energy ϵ_0 is determined by the requirement that in the presence of constant electric field $\mathbf{F}$ the probability for a hole being able to drift beyond the constant energy surface $\varepsilon=\hbar\omega_0+\epsilon_0$ without emitting one optical phonon be negligibly small. The quantity ϵ_0 is equal to the product of the external force, eF, carrier velocity $\sqrt{2m^*\hbar\omega_0}/m^*$, and inverse of the average optical phonon emitting rate, ν_A. m^* stands for the effective mass. Note that the energy-independent optical-phonon-emitting rate is due to the constant density of states in two dimensions. Therefore the excess energy can be expressed as

$$\epsilon_0 = \frac{eF}{\nu_A}\sqrt{\frac{2\hbar\omega_0}{m^*}}. \tag{25}$$

The reemerging holes can be modeled as a particle source[7,8]

$$S(\mathbf{k},t) = \frac{\frac{e}{\hbar}\left[\int_{\Pi} f_{\mathbf{k}}(t)\mathbf{F}\cdot d\mathbf{S}\right]}{\left[\int \Theta(\epsilon_0-\varepsilon(\mathbf{k}'))d^2k'\right]}\Theta(\epsilon_0-\varepsilon(\mathbf{k})), \tag{26}$$

where Θ is the step function. The meaning of the above expression is that the hole reemerging rate is uniform for energy within ϵ_0 and the total reemergence rate must match the collection of the outward carrier flux $\frac{e\mathbf{F}}{\hbar}f_{\mathbf{k}}$ passing through the surface Π in momentum space.

In order to properly account for the temperature effects, we include the acoustic phonon scattering. The acoustic phonon scattering rate $W^{acu}_{\mathbf{k},\mathbf{k}'}$ is of the form[20]

$$W^{acu}_{\mathbf{k},\mathbf{k}'} = \frac{2\pi\Xi^2 q^2}{\varrho\omega_{\mathbf{q}}WA}\left(n_q+\frac{1}{2}\mp\frac{1}{2}\right)\delta(\varepsilon(\mathbf{k}')-\varepsilon(\mathbf{k})\mp\hbar\omega_{\mathbf{q}}), \tag{27}$$

where ϱ is the mass density of solid lattice and Ξ is the lattice deformation potential. The acoustic phonon involved in the transition has wave number $\mathbf{q}=\mathbf{k}'-\mathbf{k}$ and its dispersion is given by $\omega_{\mathbf{q}}=cq$ where c is the sound velocity in the solid. Emission and absorption of phonons in the processes correspond to $+$ and $-$, respectively. The product WA represents the QW volume.

We assume homogeneity in the x and y directions so that the distribution is a function of the variables k_x and k_y only. The set of kinetic equations can be written as

$$\frac{\partial f_{\mathbf{k}}}{\partial t} + \frac{e\mathbf{F}}{\hbar}\cdot\frac{\partial f_{\mathbf{k}}}{\partial \mathbf{k}} = S_{\mathbf{k}} - D_{\mathbf{k}} + C_1[f_{\mathbf{k}},f_2], \tag{28}$$

$$\frac{\partial f_2}{\partial t} = C_2[f_{\mathbf{k}},f_2]. \tag{29}$$

$C_i[f_{\mathbf{k}},f_2]$, $i=1,2$, represent the collision terms for the acoustic phonon and resonant scattering. They are functionals of the distribution functions. The explicit expressions for the collision terms are

$$C_1[f_{\mathbf{k}},f_2] = n_a A\{W^{res}_{\mathbf{k}}(f_2-f_{\mathbf{k}})\} + \sum_{\mathbf{k}'}\{W^{acu}_{\mathbf{k}'\mathbf{k}}f_{\mathbf{k}'} - W^{acu}_{\mathbf{k}\mathbf{k}'}f_{\mathbf{k}}\}, \tag{30}$$

$$C_2[f_{\mathbf{k}},f_2] = \sum_{\mathbf{k}} W^{res}_{\mathbf{k}}(f_{\mathbf{k}}-f_2). \tag{31}$$

The kinetic equations (28) and (29) are solved numerically by starting with the equilibrium distribution and then integrating forward in time until a steady state is reached. Note that the sum of densities, $n_a f_2+\frac{1}{A}\Sigma_{\mathbf{k}}f_{\mathbf{k}}$, is a conserved quantity in the time evolution, guaranteed by cancellation of collision terms and the boundary conditions at the surface Π. In this way not only the steady state but also the transient of the system can be modeled. The occupations of LH1S f_2 and the HH1 $f_{\mathbf{k}}$ are obtained up to an arbitrary total number of holes in the subsystem. In particular the relation between f_2 and $f_{\mathbf{k}}$ at steady state can be readily seen by setting the left-hand side of Eq. (29) equal to zero:

$$f_2 = \frac{\sum_{\mathbf{k}} W_{\mathbf{k}}^{res} f_{\mathbf{k}}}{\sum_{\mathbf{k}} W_{\mathbf{k}}^{res}} = \int d\varepsilon \, \delta[\varepsilon(\mathbf{k}) - E_r] f_{\mathbf{k}}. \tag{32}$$

Now we consider the special case with no electric field. The subsystem is in thermal equilibrium. The occupations of HH1 and LH1S obey the Boltzmann statistics guaranteed by the presence of a δ function in the expression for resonant scattering as well as the fact that the scattering between HH1 states $\mathbf{k}$ and $\mathbf{k}'$ due to acoustic phonon emission and absorption satisfies the relations

$$\frac{W_{\mathbf{k}'\mathbf{k}}^{acu}}{W_{\mathbf{k}\mathbf{k}'}^{acu}} = \frac{1+n_{\mathbf{q}}}{n_{\mathbf{q}}} = \exp\{-\beta[\varepsilon(\mathbf{k}') - \varepsilon(\mathbf{k})]\}. \tag{33}$$

$\varepsilon(\mathbf{k}') > \varepsilon(\mathbf{k})$ is assumed without loss of generality, and $\mathbf{q}$ is the wave vector of the phonon involved in the process. Therefore in equilibrium f_2 is given by

$$f_2 = \frac{N/A}{\frac{1}{A}\sum_{\mathbf{k}} e^{-\beta\varepsilon(\mathbf{k})} + n_a e^{-\beta E_r}} e^{-\beta E_r}, \tag{34}$$

where N represents the total number of holes in the subsystem.

In order to describe the effect of the electric field on the distribution, we define a dimensionless parameter $\lambda(F,T)$ by

$$\lambda(F,T) \equiv \frac{\frac{1}{A}\sum_{\mathbf{k}} f_{\mathbf{k}}}{n_a f_2 + \frac{1}{A}\sum_{\mathbf{k}} f_{\mathbf{k}}} = \frac{n_s}{n_2 + n_s}. \tag{35}$$

$\lambda(F,T)$ is the fraction of holes in HH1 for the subsystem. For low temperature at equilibrium virtually all holes stay near the HH1 minimum, so λ is close to unity. In the presence of the electric field the population of LH1S increases as a consequence of Eq. (32), since holes in HH1 acquire kinetic energy from the field, so the nonequilibrium distribution $f_{\mathbf{k}}$ has a larger value at $\varepsilon(\mathbf{k}) = E_r$. Therefore, for a given n_a, $\lambda(F,T)$ is expected to decrease as the electric field increases. An increase of the acceptor density n_a also raises f_2 because the distribution in HH1 becomes more concentrated on $\varepsilon(\mathbf{k}) \leqslant E_r$. This is because the stronger resonance scattering inhibits the holes from acquiring energy higher than the resonance energy E_r.

C. Impact ionization and thermal recombination rates

Next we turn to the interactions between HH1 and low-lying localized states including HH1S and HH2P±. The interactions are dominated by impact ionization and the thermal recombination as well as their inverse processes. In the impact ionization process one energetic hole in HH1 with momentum $\mathbf{k}$ scatters with one hole in the low-lying localized states ϕ_b in the barrier through the Coulomb interaction such that they both come out as free holes in HH1. The transition rate is given by

$$w^{ip}(\mathbf{k}) = \frac{2\pi}{\hbar} \sum_{\mathbf{k}_1,\mathbf{k}_2} \left| \langle \mathbf{k}_1, \mathbf{k}_2 | \frac{e^2}{r} | \mathbf{k}, b \rangle \right|^2 \times \delta(\varepsilon(\mathbf{k}) - E_b - \varepsilon(\mathbf{k}_1) - \varepsilon(\mathbf{k}_2)), \tag{36}$$

where r is the separation between the incident hole and localized hole. The summation is over all final two-particle Bloch states $(\mathbf{k}_1, \mathbf{k}_2)$. The first task is to evaluate the scattering matrix element $\langle \mathbf{k}_1, \mathbf{k}_2 | \frac{e^2}{r} | \mathbf{k}, b \rangle$. Substituting the explicit expressions for those localized wave functions and Coulomb potential into the scattering matrix element, it becomes

$$\int d^3\mathbf{r}_1 d^3\mathbf{r}_2 \frac{1}{\sqrt{A}} e^{-i\mathbf{k}_1\cdot\vec{\rho}_1} f^*(z_1) \frac{1}{\sqrt{A}} e^{-i\mathbf{k}_2\cdot\vec{\rho}_2} f^*(z_2) \times V(|\mathbf{r}_1 - \mathbf{r}_2|) \frac{1}{\sqrt{A}} e^{i\mathbf{k}_1\cdot\vec{\rho}_1} f(z_1) \phi_b(\mathbf{r}_2), \tag{37}$$

where the dummy coordinates $\mathbf{r}_i = (\vec{\rho}_i, z_i)$, $i = 1, 2$, are to be integrated out to obtain an impact ionization rate as a function of the momentum $\mathbf{k}$ of the incident hole. The integral is complicated by the entanglement of the dummy variables $\mathbf{r}_1$ and $\mathbf{r}_2$ but it can be eased by replacing the Coulomb interaction with its representation in Fourier expansions

$$\frac{1}{4\pi\epsilon} \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} = \frac{e^2}{\epsilon} \int \frac{d^3\mathbf{q}}{(2\pi)^3} \frac{1}{q^2} e^{i\mathbf{q}\cdot\mathbf{r}_1} e^{-i\mathbf{q}\cdot\mathbf{r}_2}. \tag{38}$$

Similar to Eq. (24), the overlap between the HH1 and LH1S, the major contributions to the matrix element come from $|z| > \frac{W}{2}$. After some algebra the scattering amplitude M arrives at the expression

$$M(\mathbf{k};\mathbf{k}_1,\mathbf{k}_2) = \langle \mathbf{k}_1, \mathbf{k}_2 | \frac{e^2}{r} | \mathbf{k}, b \rangle = \frac{1}{A^{3/2}} \frac{e^2}{\epsilon} \int \frac{dq_\perp}{2\pi} \frac{1}{q_\parallel^2 + q_\perp^2} \int dz_1 |f(z_1)|^2 e^{iq_\perp z_1} \times \int dz_2 f^*(z_2) I(z_2, q') e^{-iq_\perp z_2}. \tag{39}$$

$q_\parallel = |\mathbf{k} - \mathbf{k}_1|$, and the expression for $I(z)$ is given by

$$I(z,q') = \sqrt{\frac{4\pi a^2}{b}} \int_0^\infty d\rho h(\rho) \exp\left[-\left(\rho^2 + \frac{(z - z_0^\pm)^2}{b^2} \right)^{1/2} \right], \tag{40}$$

where the function $h(\rho)$ is $\rho J_0(aq'\rho)$ for the case of HH1S as initial state and is $\sqrt{\frac{1}{2}}\rho^2 J_1(aq'\rho)$ for the case of HH2P± as initial state. J stands for the Bessel functions. $q' = |\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2|$ and $\chi = \sqrt{1 + a^2 q'^2}$. The upper script ± is for $z > 0$ and $z < 0$, respectively. Note that $I(z,q')$ decreases with the momentum transfer q' as a consequence of localization of the initial acceptor state. The scattering amplitude is expected to decrease rapidly when the momentum transfer q' is larger than the inverse of the Bohr radius a of the localized orbital. Hence we simplify the expression, Eq. (36), as

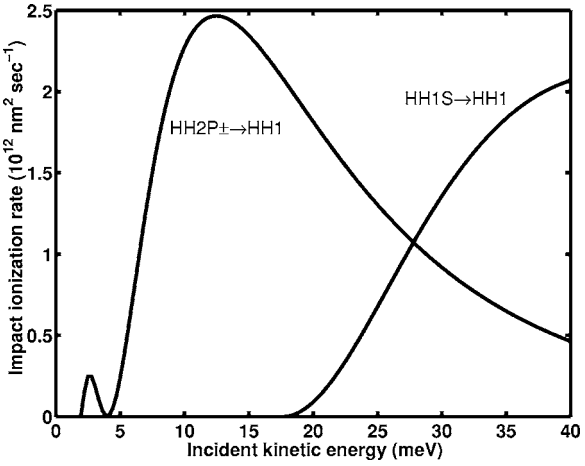

FIG. 6. Impact ionization rates w^{ip} as functions of kinetic energy ε of the incident subband hole for HH1S to HH1 and HH2P± to HH1 are respectively shown.

$$w^{ip}(\mathbf{k}) = \frac{2\pi}{\hbar}|\bar{M}|^2 \sum_{\mathbf{k}_1,\mathbf{k}_2} \Theta\left(\frac{1}{a} - |\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2|\right) \times \delta(\varepsilon(\mathbf{k}) - E_0 - \varepsilon(\mathbf{k}_1) - \varepsilon(\mathbf{k}_2)) = \frac{2\pi}{\hbar}|\bar{M}|^2\sigma(\mathbf{k}). \quad (41)$$

$\bar{M}$ stands for the maximum scattering amplitude which occurs when $k_1=k_2$, and the angle between $\mathbf{k}$ and $\mathbf{k}_1$ is equal to that between $\mathbf{k}$ and $\mathbf{k}_2$. The summation in the above expression gives the effective phase-space volume $\sigma(\mathbf{k})$ available for this scattering process given that the incident momentum is $\mathbf{k}$. Carrying out $\mathbf{k}_1$ and $\mathbf{k}_2$ integral one obtains

$$\sigma(\mathbf{k}) = A^2 \int \frac{d^2\mathbf{k}_1}{(2\pi)^2}\frac{d^2\mathbf{k}_2}{(2\pi)^2}\Theta\left(\frac{1}{a} - |\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2|\right) \times \delta(\varepsilon(\mathbf{k}) - E_0 - \varepsilon(\mathbf{k}_1) - \varepsilon(\mathbf{k}_2)) = \left(\frac{A}{(2\pi)^2}\right)^2 \int \frac{d^2\mathbf{u}}{2} d^2\mathbf{v}\Theta\left(\frac{1}{a} - u\right) \times \delta\left\{\frac{\hbar^2}{2m}\frac{1}{2}(|\mathbf{u}+\mathbf{k}|^2 + v^2) - [\varepsilon(\mathbf{k}) - E_0]\right\} = \frac{A^2}{2\pi}\frac{m}{\hbar^2}\int d^2\mathbf{u}\Theta\left(\frac{1}{a} - u\right) \times \Theta\left[-\frac{\hbar^2}{4m}|\mathbf{u}+\mathbf{k}|^2 + (\varepsilon(\mathbf{k}) - E_0)\right], \quad (42)$$

where the phase-space dummy variables $(\mathbf{k}_1,\mathbf{k}_2)$ were transformed into the new coordinates $(\mathbf{u},\mathbf{v})=(\mathbf{k}_1+\mathbf{k}_2,\mathbf{k}_1-\mathbf{k}_2)$ with the corresponding Jacobian equal to one-half. After integrating out the variable $\mathbf{v}$ the evaluation of $\sigma(\mathbf{k})$ can be obtained through counting the overlapping area of one circle centered at the origin with radius $1/a$ and another circle centered at $-k$ on the $\mathbf{x}$ axis with radius $\frac{\sqrt{4m[\varepsilon(\mathbf{k})-E_0]}}{\hbar}$. The resultant rate w^{ip} is plotted in Fig. 6 as a function of kinetic energy $\frac{\hbar^2k^2}{2m^*}$. The reverse process of impact ionization is Auger recombination, in which two HH1 holes collide and result in one localized hole and one HH1 hole with higher kinetic energy. The Auger process must be taken into account as well.

The holes impact-ionized to the HH1 can go back to the low-lying localized states by acoustic phonon emission—i.e., thermal recombination. The thermal recombination rate is given by[21]

$$w^{tr}(\mathbf{k}) = 2^{10}\pi\frac{c}{l_0}\frac{E_0^4 m^* c^2}{[\varepsilon(\mathbf{k})+E_0]^5}a^3|g(z_0^{\pm})|^2(N_{\mathbf{q}}+1), \quad (43)$$

where c is the sound velocity. $N_{\mathbf{q}}$ is the number of phonons involved in the scattering, and $\mathbf{q}$ is the wave vector of the phonon satisfying conservation of energy given by $q=[\varepsilon(\mathbf{k})+E_0]/\hbar c$. l_0 is the characteristic length for acoustic phonon scattering,

$$l_0 = \frac{\pi\hbar^4\varrho}{2m^{*3}\Xi^2}, \quad (44)$$

where ϱ and Ξ are mass density of the lattice and deformation potential as mentioned previously. The reverse process of thermal recombination is the thermal excitation of holes in the low-lying localized states by acoustic phonon absorption.

Between the two localized levels HH1S and HH2P±, the thermal capture/generation rates $t_{a/e}$ are given by

$$t_{a/e} = 2^{10}\frac{c}{l_0}\frac{mc^2}{\Delta\epsilon}\left(N_{\mathbf{q}} + \frac{1}{2} \mp \frac{1}{2}\right). \quad (45)$$

$\Delta\epsilon$ denotes the energy difference between the localized levels. The subscripts a and e indicate that these processes are accompanied by phonon absorption and phonon emission, respectively.

D. Hole population in subband and lower localized acceptor states

From Sec. IV B we are able to deal with the nonequilibrium occupations $f_{\mathbf{k}}$ and f_2 with normalization up to an arbitrary total number of holes. Using the impact ionization and phonon emission rates we are now able to deal with the occupations in the subsystem consisting of lower localized levels and HH1. To be precise, we adopt the normalization given by Eq. (24) where the total hole density of subsystem consisting of LH1S and HH1 is equal to the vacancy density in HH2P± and HH1S. Since the occupation probability f_2 is completely determined from $f_{\mathbf{k}}$, it is convenient to write the density of HH1 holes n_s as

$$n_s = \frac{n_s}{n_s + n_a f_2}(n_s + n_a f_2) = \lambda(F,T)n_a(1 - f_1 - f_g) \quad (46)$$

and the hole density of LH1S as

$$n_2 = [1 - \lambda(F,T)]n_a(1 - f_1 - f_g). \quad (47)$$

The dimensionless parameter $\lambda(F,T)$, given by Eq. (35), has values between zero and unity.

Once f_1 and f_g are known, f_2 can be determined from Eq. (47). f_1 and f_g can be calculated from the kinetics between

HH1 and the low-lying localized states. Impact ionization and thermal excitation processes cause the upward transitions while Auger recombination and thermal recombination processes cause the downward transitions. The respective downward Auger recombination rates from HH1 to HH2P± and HH1S are $r_{2p}^{ar}=A_{2p}(T)n_s^2(1-f_1)$ and $r_{1s}^{ar}=A_{1s}(T)n_s^2(1-f_g)$, where the coefficients A's are temperature and acceptor density dependent for the Auger recombination and the factors $(1-f_{1,g})$ account for the constraint that the process is forbidden when the lower acceptor state is filled with a hole. Note that holes in HH1 are not required to have threshold kinetic energy for the recombination process to take place, so we assume the coefficients A's have a negligible field dependence. The holes occupied the continuum can drop to the lower localized states, HH1S and HH2P±, by thermal recombination. In our case HH2P± is below the HH1 minimum by 2 meV, which is much smaller than the gap between HH1 and HH1S, 16 meV; here we neglect the latter recombination process since the rate is inversely proportional to the gap. This downward rate from HH1 to HH2P± is proportional to the hole density in HH1 and can be written as $r_{2p}^{tr}=C(F,T)n_s$. The coefficient $C(F,T)$, dependent on field and temperature, is taken as the average of Eq. (43) with respect to $f_{\mathbf{k}}$:

$$C(F,T)=\frac{\sum_{\mathbf{k}} w^{tr}(\varepsilon(\mathbf{k}))f_{\mathbf{k}}}{\sum_{\mathbf{k}} f_{\mathbf{k}}}. \tag{48}$$

We now consider upward transitions. The impact ionization rates for the respective processes, HH1S to HH1 and HH2P± to HH1, are of the expressions $r_{1s}^{ip}=B(F,T)n_s f_g$ and $r_{2p}^{ip}=B(F,T)n_s f_1$. Note that the factors f_1 and f_g in the expressions account for the requirement of an occupied initial localized acceptor state. The coefficients $B_i(F,T)$ can be written as the average

$$B_i(F,T)=\frac{\sum_{\mathbf{k}} w_i^{ip}(\varepsilon(\mathbf{k}))f_{\mathbf{k}}}{\sum_{\mathbf{k}} f_{\mathbf{k}}}. \tag{49}$$

The subscript of w_i^{ip} in Eq. (49) stands for different rates resulting from different initial localized states in the different collision processes in the present case. There exists a threshold of kinetic energy for the hole in HH1 for impact ionization, and consequently the coefficient B for low field and low temperature is negligibly small. Besides the upward transition caused by the inelastic collision, holes occupying the lower localized states can also be excited to the continuum through phonon emission. Here we also neglect the direct excitation of HH1S holes to HH1 because it requires absorption of phonons of much greater energy. Therefore we are left with the thermal excitation from HH2P± to HH1, and the rate can be expressed as $r^{te}=D(T)n_1$. The phonon absorption coefficient $D(T)$ is determined by detailed balance with r^{tr} at thermal equilibrium.

Since we have to consider two lower localized states in the kinetic problem, we are left with the transition between HH1S and HH2P±. For simplicity we only consider the thermal excitation and recombination. The upward and downward transitions among the two levels are given by $t_a n_g$ and $t_e n_1$. With all the necessary transitions at hand we are ready to write down the kinetic equations for the populations n_1 and n_g of the two localized states. Substituting all the formula into the relation we have

$$\begin{aligned}\frac{dn_1}{dt}-t_a n_g+t_e n_1 &= r_{2p}^{ar}-r_{2p}^{ip}+r^{tr}-r^{te}\\ &= A_{2p}n_s^2(1-f_1)-B_{2p}n_s f_1+Cn_s-Dn_1,\\ \frac{dn_g}{dt}+t_a n_g-t_e n_1 &= r_{1s}^{ar}-r_{1s}^{ip}=A_{1s}n_s^2(1-f_g)-B_{1s}n_s f_g.\end{aligned} \tag{50}$$

Now we are left with the determination of the coefficients A_{1s}, A_{2p}, and D which are assumed to be independent of the electric field. Since the occupations obtained from the rate equation must be restored to thermal equilibrium when the electric field is set to zero, the requirement of detailed balance at zero field gives A_{1s}, A_{2p}, and D using B_{1s}, B_{2p}, and C:

$$\begin{aligned}A_{2p}(T)(n_s^0)^2(1-f_1^0) &= B_{2p}(F=0,T)n_s^0 f_1^0,\\ A_{1s}(T)(n_s^0)^2(1-f_g^0) &= B_{1s}(F=0,T)n_s^0 f_g^0,\\ D(T)n_1^0 &= C(F=0,T)n_s^0.\end{aligned} \tag{51}$$

Note that the zeros as superscripts in f_g, f_1, n_1, and n_s stand for the equilibrium values.

V. RESULTS AND DISCUSSIONS

For a given F and T, Eq. (50) can be solved to give f_g and f_1. Then they can be substituted into Eq. (47) to give f_2. $f_2/f_1>1$ is the condition for population inversion. Putting everything together we are now able to obtain the nonequilibrium distribution of holes in all levels under electric field pumping. For the subsystem containing LH1S and HH1, the normalized subband distribution $\tilde{f}(\varepsilon)\equiv f(\varepsilon)\frac{k_p^2}{(2\pi)^2 n_s}$ versus hole kinetic energy for different acceptor densities is shown in Fig. 7 with applied electric field 1 kV/cm. k_p stands for the hole momentum corresponding to the kinetic energy of one optical phonon energy $\hbar\omega_0=40$ meV and the integration $\int_{|\mathbf{k}|<k_p}\tilde{f}(\varepsilon)d\varepsilon$ gives unity. For lower acceptor densities n_a holes in HH1 are more likely to be pumped to acquire energy exceeding the resonance energy E_r. This results in a lower occupation below E_r. Higher acceptor densities n_a lead to higher occupation probability at E_r—i.e., larger $\tilde{f}(E_r)$. This phenomenon results from strong resonant scattering for higher acceptor densities. From Eq. (32) the occupation f_2 of LH1S is consequently enhanced with increasing acceptor density. In other words, for the same hole density in HH1, higher acceptor densities n_a lead to higher LH1S occupation probabilities f_2. Therefore higher n_a is advantageous for building population inversion. The effect of electric fields is

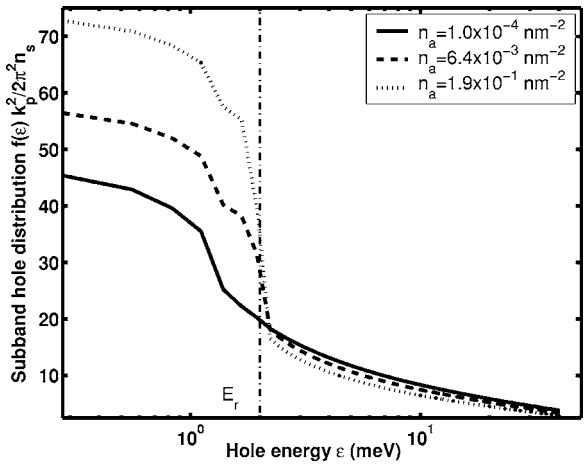

FIG. 7. Normalized subband hole distribution $\tilde{f}(\varepsilon)$ versus hole energy ε for electric field strength of $F=1000$ V/cm and $T=1$ K. The vertical line denotes the resonance energy E_r. The distribution is concentrated more in the $\varepsilon<E_r$ region as the acceptor density n_a increases.

shown in Fig. 8 by plotting the subband hole fraction λ. At low field, the occupation of LH1S compared to that of HH1 is suppressed by the Boltzmann factor and λ is near unity. As the field is turned on (between 10^{-2} and 10^{-1} V/cm), holes acquire kinetic energy by field pumping. Hence more holes accumulate in LH1S through resonant capture of holes in HH1 with kinetic energy $\varepsilon(\mathbf{k})=E_r$. As the field further increases, the fraction λ starts to increase because the field pumping overwhelms resonant capture and acoustic phonon scattering. In that case a large fraction of holes in HH1 acquire kinetic energy larger than E_r. The temperature effect diminishes in this regime as shown by the coincidence of the two curves in Fig. 8. Eventually the growth of λ in the high-field regime saturates when optical phonon scattering sets in.

Next we consider the subsystem consisting of HH1 and the lower localized states. At low temperature and equilibrium, most of the holes are bound by the acceptors and occupy the lowest HH1S. There are very few holes in HH1 and even fewer holes with enough kinetic energy to inelastically collide with the localized holes. Therefore the process of impact ionization is negligible and the so is the Auger recombination because in such a dilute case the average distance between the free holes is so large that the probability of collision is extremely small. Hence the populations of these levels are dominated by thermal processes and the statistics obey the Boltzmann distribution. When the electric field is turned on, holes can acquire more kinetic energy and impact ionization of the low-lying localized state is possible through inelastic collisions with energetic holes. The subsequent distribution of holes is balanced by those upward and downward transitions, as illustrated by Eq. (50). In order to have a quantitative understanding of how an electric field change the steady-state distribution of holes as the impact ionization rates increase, it is easier to consider the subsystem as HH1 and one single localized state, which is below HH1 minimum by e_g. The rate equation can be written in a similar manner—that is,

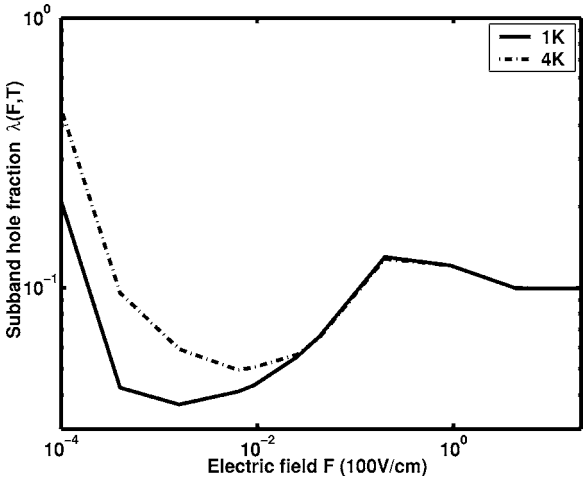

FIG. 8. The subband hole fractions $\lambda(F,T)$ as a function of electric field F for $T=1$ K (solid line) and $T=4$ K (dashed line) are shown. The coincidence for different temperatures at higher electric fields suggests that k_BT becomes irrelevant compared with the scale of E_r and optical phonon energy $\hbar\omega_0$.

$$\tilde{A}n_a\lambda^2\xi^3+\tilde{B}\lambda\xi^2+\left(\tilde{C}-\tilde{B}+\frac{\tilde{D}}{\lambda}\right)\lambda\xi-D=0. \tag{52}$$

The variable $\xi=1-\tilde{f}$ and $\tilde{f}$ stands for the population in the localized state. The capital letters with tildes represent the effective coefficients for the corresponding processes. Now we first focus on the limit of low temperature and low field. In such a case the occupation of lower localized levels is close to unity ($\xi\ll 1$) and the impact ionization coefficient $\tilde{B}$ is nearly zero. So it is a good approximation to neglect the term of highest power in ξ in the rate equation (52). The solution is given by

$$\xi=\frac{(\tilde{B}-\tilde{C})\lambda+\sqrt{(\tilde{B}-\tilde{C})^2\lambda^2+4\lambda\tilde{B}\tilde{D}}}{2\lambda\tilde{B}}, \tag{53}$$

where the term for thermal excitation $\tilde{D}/\lambda$ is dropped in the parentheses of Eq. (52) because the thermal excitation process is much weaker than the thermal recombination process $\left(\tilde{C}\gg\frac{\tilde{D}}{\lambda}\right)$ at low temperature. Note $\lambda\simeq 1$ at low field. In order to illustrate how an electric field affects the solution ξ through the impact ionization coefficient $\tilde{B}$, we set $\lambda=1$ and define the relative coefficients for impact ionization, $b_c\equiv\frac{\tilde{B}}{\tilde{C}}$, and thermal excitation, $d_c\equiv\frac{\tilde{D}}{\tilde{C}}$, to the thermal recombination coefficient C. The solution can be rewritten as

$$\xi=\frac{(b_c-1)+\sqrt{(b_c-1)^2+4b_cd_c}}{2b_c}. \tag{54}$$

$d_c\simeq e^{-\beta e_g}$ is a temperature-dependent parameter in the expression as suggested by Eq. (51). For $b_c=1$, ξ is $\sqrt{d_c}\ll 1$, justifying the omission of the ξ^3 term in Eq. (52) in the regime of discussion. The relation between ξ and the relative impact ionization coefficient b_c is plotted in Fig. 9 for temperature from 1 K to 4 K. In the limit of small b_c the solution can be approximated as $\xi=d_c$ which is nothing but the

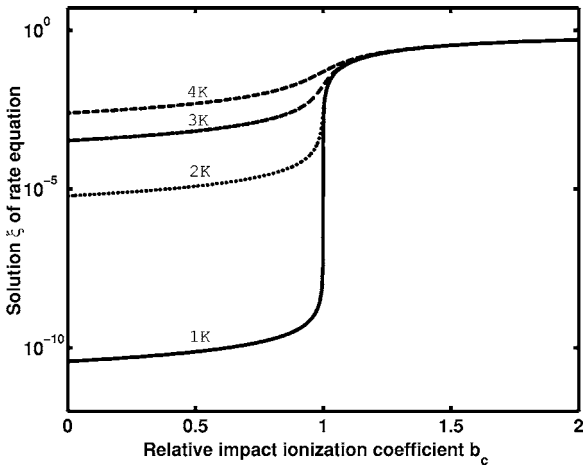

FIG. 9. The solution ξ of Eq. (52) at low field and low temperature is shown as a function of the relative impact ionization coefficient b_c. The abrupt jump around $b_c=1$ is due to the depletion of the lower localized levels by impact ionization.

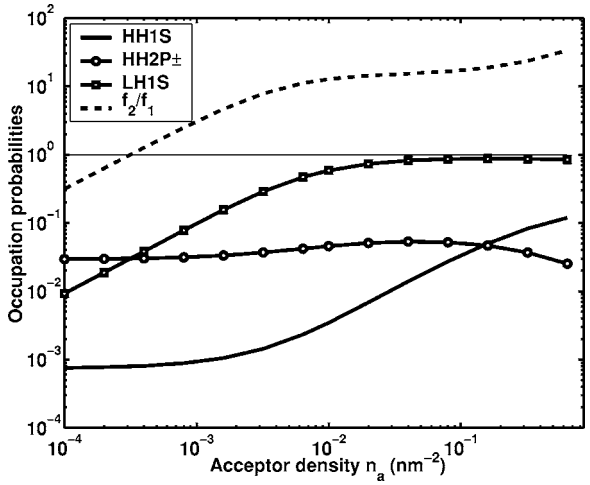

FIG. 10. Occupation probabilities f_1 (HH2P±), f_2 (LH1S), and f_g (HH1S) for $T=4$ K are plotted as functions of the acceptor density n_a at fixed electric field $F=100$ V/cm. The population ratio f_2/f_1 is shown as a solid line. The horizontal line at $f_2/f_1=1$ denotes that population inversion is built when n_a exceeds some threshold acceptor density.

thermal equilibrium. Such a case corresponds to the low-field situation in which impact ionization is not yet activated. As the electric field increases, b_c grows towards unity because more holes in HH1 acquire enough kinetic energy from the field. In the crossover regime where the term (b_c-1) in Eq. (54) turns positive from negative, ξ grows rapidly as both the population and average kinetic energy of holes in HH1 increase. As b_c gets larger and larger than 1 the solution approaches $1-\frac{1}{b_c}$. In the crossover there is competition between the two terms $(b_c-1)^2$ and $b_c d_c$ in the square root of Eq. (54). Consequently the size of the crossover is determined by $\sqrt{d_c}$. Since the impact ionization parameter b_c is strongly field dependent, this crossover corresponds to the variation of field δF as

$$\delta F \sim e^{-\beta e_g/2}\left(\frac{\partial b_c}{\partial F}\right)^{-1}. \tag{55}$$

This δF characterizes how sensitive pumping is to electric field. The dramatic jump of ξ at $b_c \simeq 1$ is due to the dominance of upward impact ionization over the downward thermal recombination. The depletion of the lower localized levels when $b_c>1$ is critical for the realization of the hole population inversion.

After combining the two subsystems, we are able to obtain the occupation of each level in the system. The occupation probabilities f_g, f_1, and f_2 for the strain-split acceptor levels and the ratio f_2/f_1 at 4 K are shown in Fig. 10. By definition a population inversion is established if $f_2/f_1>1$. There is a threshold acceptor density n_a of about 10^{-3} nm^{-2} when the applied field is 100 V/cm. The threshold acceptor density reflects the fact that the resonance scattering is necessary for building the population inversion. As n_a increases further, it becomes harder for HH1 holes to acquire higher energy, which is shown in Fig. 7, and this effect leads to a suppression of the impact ionization processes from the lower levels. Even though the upward transitions get suppressed due to more resonant scattering, the population f_2 remains at fixed values due to the increase of λ with increasing n_a. However, this effect leads to the fact that the population ratio f_g/f_1 is getting closer to its equilibrium value. For $T<4$ K the result is the same because acoustic phonon scattering is irrelevant for low temperature and higher field. The behaviors of the system differ for low-temperature $(k_BT<\delta)$ and high-temperature $(k_BT>\delta)$ regimes. At low temperature $(T<10$ K) population inversion can be realized for only a moderate electric field (100 V/cm) because there is almost no acoustic phonon scattering and the hole distribution in HH1 can be easily distorted by the field. At high temperature, the distribution is stabilized by the strong acoustic phonon scattering. Therefore population inversion is impossible even for a stronger field.

In Fig. 11 the populations of localized levels versus field strength F for $T=10$ K are shown. As the field is turning on and increasing toward 20 V/cm, holes in HH1 become more and more energetic. Consequently more and more free holes are generated due to the increase of the coefficients B_{1s} and B_{2p}. Note that presently the resulting upward transition is mainly from HH1S to HH1 because the upper level HH2P± is empty and the population f_1 mainly results from the combined processes, impact ionization HH1S to HH1 plus the thermal recombination from HH1 to HH2P± is empty and the population f_1. As F continues to increase, the populations f_1 and f_2 grow significantly and the lowest HH1S begins to be depleted due to the fact that the intracenter recombination from HH2Pñ to HH1S is quite slow. Now the upward transition is contributed more by HH2P± is empty and the population f_1 than HH1S. When the field exceeds the threshold field, 20 V/cm in our case, the lowest HH1S is almost empty and the pumping process is mainly controlled by the transitions between HH1 and HH2P± is empty and the population f_1. The abruptness of the growth of f_2 is inversely proportional to the temperature according to Eq. (55). However, the population f_2 comes to a fixed value for the field F

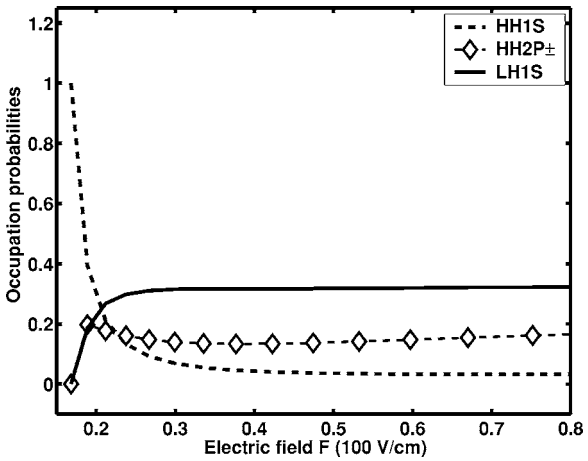

FIG. 11. Occupation probabilities f_g of HH1S, f_1 of HH2P±, and f_2 of LH1S versus electric field F at $T=10$ K are shown. Population inversion is achieved when the electric field is larger than 20 V/cm. As the temperature further increases, population inversion is not possible because the electric field cannot drive the hole distribution significantly away from equilibrium. Therefore $T=10$ K is the critical temperature for population inversion in our case.

>30 V/cm. This saturation is indicative of the fact that impact ionization rates have a upper bound. If we further increase the temperature, acoustic phonon scattering and thermal recombination become important and k_BT comes back as a relevant energy scale. When the thermal energy dominates the transport process, the electric field is no longer able to significantly push the distribution away from equilibrium. Based on these results we predict that the optimal conditions for hole population inversion ratio f_2/f_1 are $n_a=10^{-3}$ nm^{-2}, $F=100$ V/cm, and temperature below tens of kelvins. These conditions in a QW structure specified in Sec. III A, with central width $W=11.7$ nm and germanium compositions $x=0.088$ and $y=0.094$, are well within the range of experimental implementation.

We would like to revisit the fundamental issue of the relative time scales and verify its self-consistency in this problem. In general population inversion of a laser requires that the upper-level formation time (resonant scattering time in our case) be shorter than the the spontaneous emission time. The formation time is $\hbar$ divided by the resonance energy width Γ in Fig. 5. As expected Γ decreases with the distance d between the acceptor and the well due to reduced wavefunction overlap. However, Γ is always of the order of meV for the range of d relevant to this work, consistent with the previous work on impurities outside the well.[16] The corresponding formation time is of the order of picoseconds. This value is not far from the case for acceptors in the bulk.[7,8] On the other hand, the spontaneous emission time is as long as microseconds as discussed in Sec. IV B and previous work.[7] The population inversion condition is therefore easily satisfied. Another key time scale is the transient time required for the lower accepter level to move to the free hole levels by impact ionization. The impact ionization time is the inverse of the rate in Fig. 6 times the free carrier area density which is of the order of 0.01 nm^{-2} according to Fig. 10. The result is about 10 ps, also much shorter than the spontaneous emission time. Because of the large difference in the time scale, the neglect of spontaneous emission in the rate equation is self-consistent. Population inversion is partly made possible by the much faster resonant scattering time.

Before drawing the conclusion, some nonideal effects have to be remarked on here. In real structures there are always compensating donors present. We assume that the donor density in the δ-doped region is n_d. With the presence of the compensating donors, the density of holes is reduced to $n_a^*=n_a-n_d$ in the number normalization equation (24). The presence of donors reduces the total number of holes and increases the number of unoccupied impurity levels. Since the governing equations for the distribution of holes are nonlinear due to impact ionization and Auger recombination as shown in Eq. (50), the impact ionization threshold and many other properties depend on the total number of holes and therefore the compensation ratio n_d/n_a. The immediate consequence is an increase of the threshold electric field for impact ionization due to fewer energetic holes generated than in the ideal case without compensating donors. However, the laser threshold electric field will not be increased too much with the inevitable donors in reality. In addition the two-level impurity interacting with a continuum employed in our model can lead to an S-shaped current-field dependence and bistability when compensating donors are present. This nonlinear behavior has been studied extensively in *p*-Ge (Ref. 22) and for recombination-generation models in semiconductors in general.[23] It allows low- and high-current states for a given field. In order for our quantum well structure to achieve population inversion, the system must be in the high-current state. Actually the S-shaped dependence comes from multiple solutions of the steady-state rate equation for the occupations and is independent of carrier species and systems. Hence the S-shaped dependence is expected to occur as well in our subband system if the compensation ratio n_d/n_a is too large. In summary we estimated that the effects of donors are negligible if n_d/n_a is much smaller than 0.01.[22]

VI. CONCLUSION

We propose a QW structure with a resonant state and show that the relative energies of the strain-split localized states and the continuous states enable laser operation with photon frequencies as low as 1 THz. The hole distributions are studied in detail with consideration of all the related microscopic physical processes. Calculations of the occupation probabilities of the localized states reveal that there are thresholds for the external field and acceptor density in order to achieve population inversion. For 1 THz lasing, the required field is 100 V/cm at a temperature below 10 K. These conditions can be easily realized in experiments. This work leads to a new and practical direction for semiconductor THz lasers with arbitrarily small radiation frequency.

ACKNOWLEDGMENT

This work is supported by the National Science Council of Taiwan, R.O.C., under Grant No. NSC93-2112-M-009-006.

APPENDIX

In this appendix we give the details of the binding energy corrections to the LH1S impurity state in the presence of the QW and off-diagonal couplings with the HH1 continuum. For simplicity the envelope functions for the localized impurity state and the QW continuum are denoted by ϕ and $\psi_{\mathbf{k}}$, respectively. The unperturbed states satisfy the equations

$$[H_{LK}^0 + V_I(r) + V_C(z_0)]\phi u_{3/2} = E_{1s}\phi u_{3/2},$$

$$[H_{LK}^0 + V_C(z)]\psi_{\mathbf{k}} u_{1/2} = \varepsilon_{\mathbf{k}}\psi_{\mathbf{k}} u_{1/2}. \tag{A1}$$

Considering the full Hamiltonian, the eigenstates are a superposition of the form

$$\Psi_p = \sum_{m=\pm 1/2} a_p^{(m)}(\phi u_m) + \sum_{m=\pm 3/2} \Sigma_{\mathbf{k}} b_{p\mathbf{k}}^{(m)}(\psi_{\mathbf{k}} u_m). \tag{A2}$$

Here the index p denotes the label for the hybridized states and it runs through the total number of continuous states plus one. Substituting the hybridized states into the equation $H\Psi_p = \epsilon_p \Psi_p$, a set of algebra equations for the coefficients a_p and $b_{p\mathbf{k}}$ is obtained.

$$a_p^{(1/2)}(\epsilon_p - E_{1s} - \Delta) = \sum_{\mathbf{k}} b_{p\mathbf{k}}^{(-3/2)}\alpha_{\mathbf{k}},$$

$$b_{p\mathbf{k}}^{(-3/2)} - \Sigma_{\mathbf{k}'} b_{p\mathbf{k}'}^{(-3/2)}\beta_{\mathbf{k}\mathbf{k}'} = a_p^{(1/2)}\alpha_{\mathbf{k}}^*, \tag{A3}$$

where

$$\beta_{\mathbf{k}\mathbf{k}'} = \langle \psi_{\mathbf{k}} | v_I(r) | \psi_{\mathbf{k}'} \rangle. \tag{A4}$$

$\beta_{\mathbf{k}\mathbf{k}'}$ stands for the intraband transition due to Coulomb interaction from the impurity center. Note that we omit another set of equations for $m=-1/2$ and $m=3/2$ because they are identical to Eq. (A3). The eigenvalue ϵ_p can be solved by iterative substitution from Eq. (A3), and the leading terms are

$$\epsilon_p = E_{1s} + \Delta + \sum_{\mathbf{k}} \frac{|\alpha_{\mathbf{k}}|^2}{\epsilon_p - \varepsilon_{\mathbf{k}}} + \sum_{\mathbf{k}_1\mathbf{k}_2} \frac{\alpha_{\mathbf{k}_1}^* \beta_{\mathbf{k}_1\mathbf{k}_2}\alpha_{\mathbf{k}_2}}{(\epsilon_p - \varepsilon_{\mathbf{k}_1})(\epsilon_p - \varepsilon_{\mathbf{k}_2})}. \tag{A5}$$

The perturbed energy for LH1S can be obtained by directly substituting the unperturbed energy E_{1s} for ϵ_p on the right-hand side of Eq. (A5). Aided by the equality $\frac{1}{x+i} = P\frac{1}{x} - i\pi\delta(x)$ to avoid the singularity and neglecting the higher-order terms, expressions for the shift of binding energy ΔE_{1s} and the corresponding imaginary part energy are obtained:

$$\Delta E_{1s} = \Delta + P\frac{A}{(2\pi)^2}\int d\mathbf{k}\frac{|\alpha_{\mathbf{k}}|^2}{E_{1s} - \varepsilon_{\mathbf{k}}}, \tag{A6}$$

$$\frac{\Gamma_{E_{1s}}}{2} = \pi\frac{A}{(2\pi)^2}\int d\mathbf{k}\,\delta(E_{1s} - \varepsilon_{\mathbf{k}})|\alpha_{\mathbf{k}}|^2. \tag{A7}$$

*Corresponding author. Electronic address: meng@mail.nctu.edu.tw

[1] V. Krozer, B. Leone, H. G. Roskos, T. Loffler, G. Loata, G. H. Dohler, F. Renner, S. Eckardt, S. Malzer, A. Schwanhauber, T. O. Klaassen, A. Adam, P. Lugli, A. Di Carlo, M. Manenti, G. Scamarcio, M. S. Vitiello, and M. Feiginov, Proc. SPIE **5466**, 178 (2004).

[2] G. L. Carr, M. C. Martin, W. R. McKinney, K. Jordan, G. R. Neil, and G. P. Williams, Nature (London) **420**, 153 (2002).

[3] Y. Shimada, K. Hirakawa, M. Odnoblioudov, and K. A. Chao, Phys. Rev. Lett. **90**, 046806 (2003).

[4] G. Dehlinger, L. Diehl, U. Gennser, H. Sigg, J. Faist, K. Ensslin, D. Grützmacher, and E. Müller, Science **290**, 2277 (2000).

[5] S. Kumar, B. S. Williams, S. Kohen, Q. Hu, and J. L. Reno, Appl. Phys. Lett. **84**, 2494 (2004).

[6] S. Barbieri, J. Alton, H. E. Beere, J. Fowler, E. H. Linfield, and D. A. Ritchie, Appl. Phys. Lett. **85**, 1674 (2004).

[7] M. A. Odnoblyudov, I. N. Yassievich, M. S. Kagan, Yu. M. Galperin, and K. A. Chao, Phys. Rev. Lett. **83**, 644 (1999).

[8] M. A. Odnoblyudov, I. N. Yassievich, V. M. Chistyakov, and K. A. Chao, Phys. Rev. B **62**, 2486 (2000); M. A. Odnoblyudov, A. A. Prokofiev, I. N. Yassievich, and K. A. Chao, *ibid.* **70**, 115209 (2004).

[9] I. V. Altukhov, E. G. Chirkova, V. P. Sinis, M. S. Kagan, Yu. P. Gousev, S. G. Thomas, K. L. Wang, M. A. Odnoblyudov, and I. N. Yassievich, Appl. Phys. Lett. **79**, 3909 (2001).

[10] M. S. Kagan and I. N. Yassievich, Physica E (Amsterdam) **13**, 916 (2002).

[11] S. M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981).

[12] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).

[13] G. L. Bir and G. E. Pikus, *Symmtery and Strain Effects in Semiconductors* (Wiley, New York, 1974).

[14] C. G. Van de Walle, Phys. Rev. B **39**, 1871 (1989).

[15] C. G. Van de Walle and R. M. Martin, Phys. Rev. B **34**, 5621 (1986).

[16] A. Blom, M. A. Odnoblyudov, I. N. Yassievich, and K. A. Chao, Phys. Rev. B **65**, 155302 (2002).

[17] P. Lawaetz, Phys. Rev. B **4**, 3460 (1971).

[18] K. Tomizawa, *Numerical Simulation of Submicron Semiconductor Devices* (Artech House, Boston, 1993).

[19] A. A. Andronov, Sov. Phys. Semicond. **21**, 701 (1987).

[20] S. Datta, *Quantum Phenomena* (Addison-Wesley, New York, 1989).

[21] V. N. Abakumov, V. I. Perer, and I. N. Yassievich, in *Nonradiative Recombination in Semiconductors*, edited by V. M. Agranovich and A. A. Maradudin, *Modern Problems in Condensed Matter Sciences*, Vol. 33 (North-Holland, Amsterdam, 1991).

[22] W. Quade, G. Hupper, E. Schöll, and T. Kuhn, Phys. Rev. B **49**, 13408 (1994).

[23] E. Schöll, *Nonequilibrium Phase Transitions in Semiconductors* (Springer, Berlin, 1987).