

# The Intersubband Approach to Si-based Lasers

Greg Sun

*University of Massachusetts Boston, Massachusetts,  
U.S.A.*

## 1. Introduction

Silicon has been the miracle material for the electronics industry, and for the past twenty years, technology based on Si microelectronics has been the engine driving the digital revolution. For years, the rapid “Moore’s Law” miniaturization of device sizes has yielded an ever-increasing density of fast components integrated on Si chips: but during the time that the feature size was pushed down towards its ultimate physical limits, there has also been a tremendous effort to broaden the reach of Si technology by expanding its functionalities well beyond electronics. Si is now being increasingly investigated as a platform for building photonic devices. The field of Si photonics has seen impressive growth since early visions in the 1980s and 1990s [1,2]. The huge infrastructure of the global Si electronics industry is expected to benefit the fabrication of highly sophisticated Si photonic devices at costs that are lower than those currently required for compound semiconductors. Furthermore, the Si-based photonic devices make possible the monolithic integration of photonic devices with high speed Si electronics, thereby enabling an oncoming Si-based “optoelectronic revolution”.

Among the many photonic devices that make up a complete set of necessary components in Si photonics including light emitters, amplifiers, photodetectors, waveguides, modulators, couplers and switches, the most difficult challenge is the lack of an efficient light source. The reason for this striking absence is that bulk Si has an indirect band gap where the minimum of the conduction band and the maximum of the valence band do not occur at the same value of crystal momentum in wave vector space (Fig. 1). Since photons have negligible momentum compared with that of electrons, the recombination of an electron-hole pair will not be able to emit a photon without the simultaneous emission or absorption of a phonon in order to conserve the momentum. Such a radiative recombination is a second-order effect occurring with a small probability, which competes with nonradiative processes that take place at much faster rates. As a result, as marvelous as it has been for electronics, bulk Si has not been the material of choice for making light emitting devices including lasers.

Nevertheless, driven by its enormous payoff in technology advancement and commercialization, many research groups around the world have been seeking novel approaches to overcome the intrinsic problem of Si to develop efficient light sources based on Si. One interesting method is to use small Si nanocrystals dispersed in a dielectric matrix, often times SiO<sub>2</sub>. Such nano-scaled Si clusters are naturally formed by the thermal annealing of a Si-rich oxide thin film. Silicon nanocrystals situated in a much wider band gap SiO<sub>2</sub> can effectively localize electrons with quantum confinement, which improves the radiative recombination probability, shifts the emission spectrum toward shorter wavelengths, and

Source: *Advances in Lasers and Electro Optics*, Book edited by: Nelson Costa and Adolfo Cartaxo, ISBN 978-953-307-088-9, pp. 838, April 2010, INTECH, Croatia, downloaded from SCIYO.COM

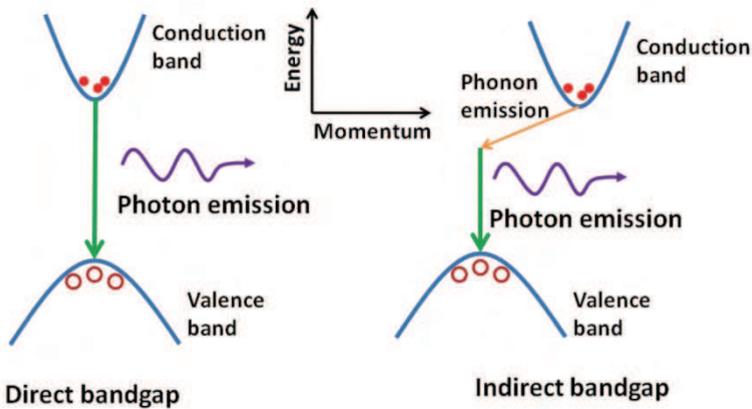


Fig. 1. Illustration of a photon emission process in (a) the direct and (b) the indirect band gap semiconductors.

decreases the free carrier absorption. Optical gain and stimulated emission have been observed from these Si nanocrystals by both optical pumping [3,4] and electrical injection [5], but the origin of the observed optical gain has not been fully understood as the experiments were not always reproducible – results were sensitive to the methods by which the samples were prepared. In addition, before Si-nanocrystal based lasers can be demonstrated, the active medium has to be immersed in a tightly confined optical waveguide or cavity.

Another approach is motivated by the light amplification in Er-doped optical fibers that utilize the radiative transitions in Er ions ( $\text{Er}^{3+}$ ) [6]. By incorporating  $\text{Er}^{3+}$  in Si, these ions can be excited by energy transfer from electrically injected electron-hole pairs in Si and will subsequently relax by emitting photons at the telecommunication wavelength of  $1.55 \mu\text{m}$ . However, the concentration of  $\text{Er}^{3+}$  ions that can be doped in Si is relatively low and there is a significant energy back-transfer from the  $\text{Er}^{3+}$  ions to the Si host due to the resonance with a defect level in Si. As a result, both efficiency and maximum power output have been extremely low [7,8]. To reduce the back transfer of energy,  $\text{SiO}_2$  with an enlarged band gap has been proposed as host to remove the resonance between the defect and the  $\text{Er}^{3+}$  energy levels [9]. Once again, Si-rich oxide is employed to form Si nanocrystals in close proximity to  $\text{Er}^{3+}$  ions. The idea is to excite  $\text{Er}^{3+}$  ions with the energy transfer from the nearby Si nanocrystals. Light emitting diodes (LEDs) with efficiencies of about 10% have been demonstrated [10] on par with commercial devices made of GaAs, but with power output only in tens of  $\mu\text{W}$ . While there have been proposals to develop lasers using doped Er in Si-based dielectric, the goal remains elusive.

The only approach so far that has led to the demonstration of lasing in Si exploited the effect of stimulated Raman scattering [11-13], analogous to that produced in fiber Raman amplifiers. With both the optical pumping and the Raman scattering below the band gap of Si, the indirectness of the Si band gap becomes irrelevant. Depending on whether it is a Stokes or anti-Stokes process, the Raman scattering either emits or absorbs an optical phonon. Such a nonlinear process requires optical pumping at very high intensities ( $\sim 100 \text{MW}/\text{cm}^2$ ) and the device lengths ( $\sim \text{cm}$ ) are too large to be integrated with other photonic and electronic devices in any type of Si VLSI-type circuit [14].

Meanwhile, the search for laser devices that can be integrated on Si chips has gone well beyond the monolithic approach to seek solutions using hybrid integration of III-V compounds with Si. A laser with an AlGaInAs quantum well (QW) active region bonded to a silicon waveguide cavity was demonstrated [15]. This fabrication technique allows for the optical waveguide to be defined by the CMOS compatible Si process while the optical gain is provided by III-V materials. Rare-earth doped visible-wavelength GaN lasers fabricated on Si substrates are also potentially compatible with the Si CMOS process [16]. Another effort produced InGaAs quantum dot lasers deposited directly on Si substrates with a thin GaAs buffer layer [17]. Although these hybrid approaches offer important alternatives, they do not represent the ultimate achievement of Si-based lasers monolithically integrated with Si electronics.

While progress is being made along these lines and debates continue about which method offers the best promise, yet another approach emerged that has received a great deal of attention in the past decade—an approach in which the lasing mechanism is based on intersubband transitions (ISTs) in semiconductor QWs. Such transitions take place between quantum confined states (subbands) of conduction or valence bands and do not cross the semiconductor band gap. Since carriers remain in the same energy band (either conduction or valence), optical transitions are always direct in momentum space rendering the indirectness of the Si band gap irrelevant. Developing lasers using ISTs therefore provides a promising alternative that completely circumvents the issue of indirectness in the Si band gap. In addition, this type of laser can be conveniently designed to employ electrical pumping – the so-called quantum cascade laser (QCL). The pursuit of Si-based QCLs might turn out to be a viable path to achieving electrically pumped Si-based coherent emitters that are suitable for monolithic integration with Si photonic and electronic devices.

In this chapter, lasing processes based on ISTs in QWs are explained by drawing a comparison to conventional band-to-band lasers. Approaches and results towards SiGe QCLs using ISTs in the valence band are overviewed, and the challenges and limitations of the SiGe valence-band QCLs are discussed with respect to materials and structures. In addition, ideas are proposed to develop conduction-band QCLs, among them a novel QCL structure that expands the material combination to SiGeSn. This is described in detail as a way to potentially overcome the difficulties that are encountered in the development of SiGe QCLs.

## 2. Lasers based on intersubband transitions

Research on quantum confined structures including semiconductor QWs and superlattices (SLs) was pioneered by Esaki and Tsu in 1970 [18]. Since then confined structures have been developed as the building blocks for a majority of modern-day semiconductor optoelectronic devices. QWs are formed by depositing a narrower band gap semiconductor with a layer thickness thinner than the deBroglie wavelength of the electron ( $\sim 10\text{nm}$ ) between two wider band gap semiconductors (Fig. 2(a)). The one-dimensional quantum confinement leads to quantized states (subbands) in the direction of growth  $z$  within both conduction and valence bands. The energy position of each subband depends on the band offset ( $\Delta E_C, \Delta E_V$ ) and the effective mass of the carrier. In directions perpendicular to  $z$  (in-plane), the carriers are unconfined and can thus propagate with an in-plane wave vector  $\mathbf{k}$  which gives an energy dispersion for each subband. (Fig. 2(b))

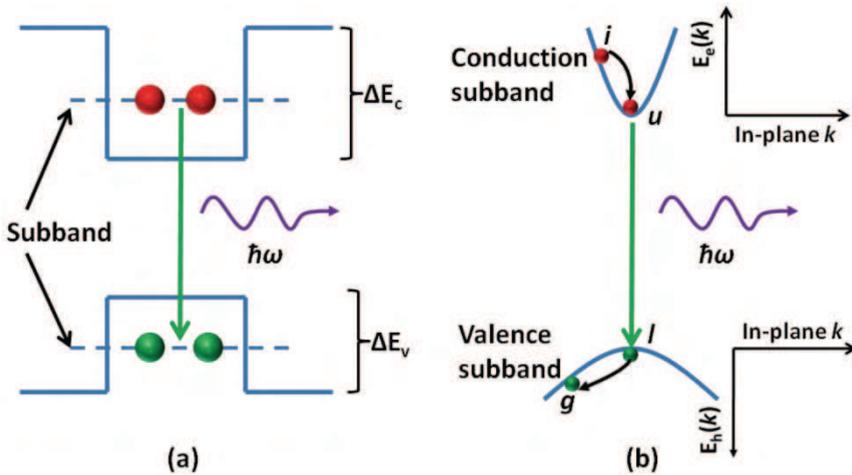


Fig. 2. Illustration of (a) conduction and valence subband formations in a semiconductor QW and (b) in-plane subband dispersions with optical transitions between conduction and valence subbands.

Obviously, if the band offset is large enough, there could be multiple subbands present within either conduction or valence band as shown in Fig. 3 where two subbands are confined within the conduction band. The electron wavefunctions (Fig. 3(a)) and energy dispersions (Fig. 3(b)) are illustrated for the two subbands. The concept of ISTs refers to the physical process of a carrier transition between these subbands within either the conduction or valence band as illustrated in Fig. 3. Carriers originally occupying a higher energy subband can make a radiative transition to a lower subband by emitting a photon. Coherent sources utilizing this type of transition as the origin of light emission are called intersubband lasers.

The original idea of creating light sources based on ISTs was proposed by Kazarinov and Suris [19] in 1971, but the first QCL was not demonstrated until 1994 by a group led by Capasso at Bell Laboratories [20]. In comparison with the conventional band-to-band lasers, lasers based on ISTs require much more complex design of the active region which consists of carefully arranged multiple QWs (MQWs). The reason for added complexity can be appreciated by comparing the very different band dispersions that are involved in these two types of lasers. In a conventional band-to-band laser, it appears that the laser states consist of two broad bands. But a closer look at the conduction and valence band dispersions (Fig. 2(b)) reveals a familiar four-level scheme where in addition to the upper laser states  $|u\rangle$ , located near the bottom of the conduction band and the lower laser states  $|l\rangle$ , near the top of the valence band, there are two other participating states - intermediate states  $|i\rangle$ , and ground states  $|g\rangle$ . The pumping process (either injection or optical) places electrons into the intermediate states,  $|i\rangle$ , from which they quickly relax toward the upper laser states  $|u\rangle$  by inelastic scattering intraband processes. This process is very fast, occurring on a sub-pico-second scale. But once they reach states  $|u\rangle$ , they tend to stay there for a much longer time determined by the band-to-band recombination rate which is on the order of nanoseconds. Electrons that went through lasing transitions to the lower laser states  $|l\rangle$  will quickly scatter into the lower energy states of the valence band - ground states  $|g\rangle$  --

by the same fast inelastic intraband processes. (A more conventional way to look at this is the relaxation of holes toward the top of the valence band.) The population inversion between  $|u\rangle$  and  $|l\rangle$  is therefore established mostly by the fundamental difference between the processes determining the lifetimes of upper and lower laser states. As a result, the lasing threshold can be reached when the whole population of the upper conduction band is only a tiny fraction of that of the lower valence band.

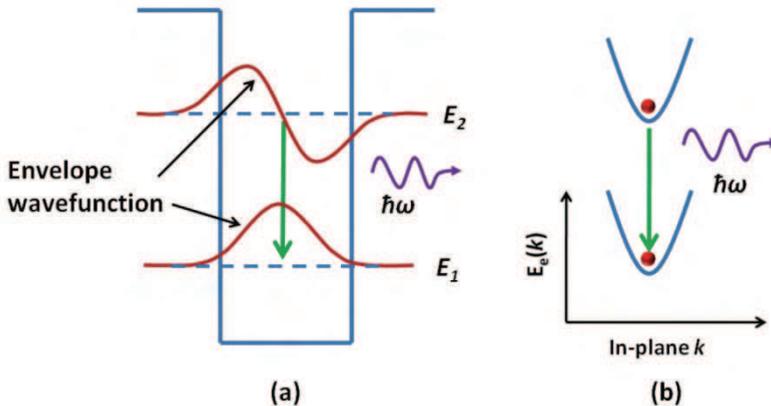


Fig. 3. (a) Two subbands formed within the conduction band confined in a QW and their electron envelope functions, (b) in-plane energy dispersions of the two subbands. Radiative intersubband transition between the two subbands is highlighted.

Let us now turn our attention to the intersubband transition shown in Fig. 3(b). The in-plane dispersions of the upper  $|u\rangle$  and lower  $|l\rangle$  conduction subbands are almost identical when the band nonparabolicity can be neglected. For all practical purposes they can be considered as two discrete levels. Then, in order to achieve population inversion it is necessary to have the whole population of the upper subband exceed that of the lower subband. For this reason, a three- or four-subband scheme becomes necessary to reach the lasing threshold. Even then, since the relaxation rates between different subbands are determined by the same intraband processes, a complex multiple QW structure needs to be designed to engineer the lifetimes of involved subbands.

Still, intersubband lasers offer advantages in areas where the conventional band-to-band lasers simply cannot compete. In band-to-band lasers, lasing wavelengths are mostly determined by the intrinsic band gap of the semiconductors. There is very little room for tuning, accomplished by varying the structural parameters such as strain, alloy composition, and layer thickness. Especially for those applications in the mid-IR to far-IR range, there are no suitable semiconductors with the appropriate band gaps from which such lasers can be made. With the intersubband transitions, we are no longer limited by the availability of semiconductor materials to produce lasers in this long wavelength region. In addition, for ISTs between conduction subbands with parallel band dispersions, the intersubband lasers should therefore have a much narrower gain spectrum in comparison to the band-to-band lasers in which conduction and valence bands have opposite band curvatures.

A practical design that featured a four-level intersubband laser pumped optically was proposed by Sun and Khurgin [21,22] in the early 1990s. This work laid out a comprehensive

analysis of various intersubband processes that affect the lasing operation including scattering mechanisms that determine subband lifetimes, conditions for population inversion between two subbands, band engineering to achieve it, and optical gain sufficient to compensate for losses under realistic pumping intensity. The QCLs developed soon thereafter significantly expanded the design in order to accommodate electrical pumping by implementing a rather elaborate scheme of current injection with the use of a chirped SL as the injector region placed in between the active regions (Fig. 4). The QCL has a periodic structure with each period consisting of an active and an injector region. Both active and injector regions are composed of MQWs. By choosing combinations of layer thicknesses and material compositions, three subband levels with proper energy separations and wave function overlaps are obtained in the active region. The injector region, on the other hand, is designed with a sequence of QWs having decreasing well widths (chirped SL) such that they form a miniband under an electric bias that facilitates electron transport. The basic operating principle of a QCL is illustrated in Fig. 4. Electrons are first injected through a barrier into subband 3 (upper laser state) of the active region, they then undergo lasing transitions to subband 2 (lower laser state) by emitting photons, followed by fast depopulation into subband 1 via nonradiative processes. These electrons are subsequently transported through the injector region into the next active region where they repeat the process in a cascading manner, typically 20 to 100 times.

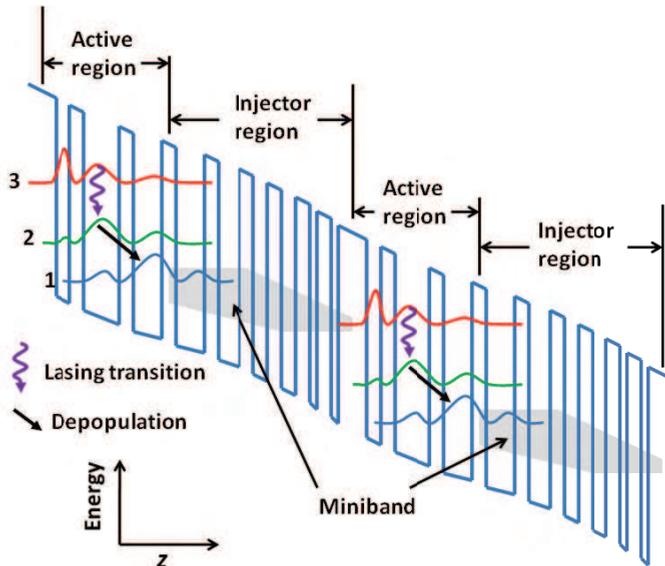


Fig. 4. Schematic band diagram of two periods of a QCL structure with each period consisting of an active and an injector region. Lasing transitions are between the states 3 and 2 in the active regions with rapid depopulation of lower state 2 into state 1 which couples strongly with the minibands formed in injector regions that transport carriers to state 3 in the next period. The magnitude-squared wavefunctions for the three subbands in active regions are illustrated.

Advances of QCLs since the first demonstration have resulted in dramatic performance improvement in spectral range, power and temperature. They have become the dominant mid-IR semiconductor laser sources covering the spectral range of  $3 < \lambda < 25 \mu\text{m}$  [23-25], many of them operating in the continuous-wave mode at room temperature with peak power reaching a few watts [26,27]. Meanwhile, QCLs have also penetrated deep into the THz regime loosely defined as the spectral region  $100\text{GHz} < f < 10 \text{ THz}$  or  $30 < \lambda < 3000 \mu\text{m}$ , bridging the gap between the far-IR and GHz microwaves. At present, spectral coverage from 0.84-5.0 THz has been demonstrated with operation in either the pulsed or continuous-wave mode at temperatures well above 100K [28].

### 3. Intersubband theory

In order to better explain the design considerations of intersubband lasers, it is necessary to introduce some basic physics that underlies the formation of subbands in QWs and their associated intersubband processes. The calculation procedures described here follows the envelope function approach based on the effective-mass approximation [29]. The  $\mathbf{k} \cdot \mathbf{p}$  method [30] was outlined to obtain in-plane subband dispersions in the valence band. Optical gain for transitions between subbands in conduction and valence bands is derived. Various scattering mechanisms that determine the subband lifetimes are discussed with an emphasis on the carrier-phonon scattering processes.

#### 3.1 Subbands and dispersions

Let us treat the conduction subbands first. It is well known in bulk material that near the band edge, the band dispersion with an isotropic effective mass follows a parabolic relationship. In a QW structure, along the in-plane direction ( $\mathbf{k} = k_x \hat{x} + k_y \hat{y}$ ) where electrons are unconfined, such curvature is preserved for a given subband  $i$ , assuming the nonparabolicity that describes the energy-dependent effective mass  $m_e^*$  can be neglected,

$$E_{i,\mathbf{k}} = E_i + \frac{\hbar^2 k^2}{2m_e^*} \quad (1)$$

where  $\hbar$  is the Planck constant and  $E_i$  is the minimum energy of subband  $i$  in a QW structure. This minimum energy can be calculated as one of the eigen values of the Schrödinger equation along the growth direction  $z$ ,

$$\left[ -\frac{\hbar^2}{2} \frac{d}{dz} \frac{1}{m_e^*(z)} \frac{d}{dz} + V_c(z) \right] \varphi_i(z) = E_i \varphi_i(z) \quad (2)$$

where the  $z$ -dependence of  $m_e^*$  allows for different effective masses in different layers and  $V_c(z)$  represents the conduction band edge along the growth direction  $z$ . The envelope function of subband  $i$ ,  $\varphi_i(z)$ , together with the electron Bloch function  $u_e(\mathbf{R})$  and the plane wave  $e^{j\mathbf{k}\cdot\mathbf{r}}$ , gives the electron wavefunction in the QW structure as

$$\Phi_i(\mathbf{r}, z) = \varphi_i(z) u_e(\mathbf{R}) e^{j\mathbf{k}\cdot\mathbf{r}} \quad (3)$$

where the position vector is decomposed into in-plane and growth directions  $\mathbf{R} = \mathbf{r} + z\hat{z}$ . Since we are treating electron subbands, the Bloch function is approximately the same for all subbands and all  $\mathbf{k}$ -vectors. The electron envelope function can be given as a combination of

the forward and backward propagations in a given region  $l$  of the QW structure (either a QW or a barrier region),  $d_l < z < d_{l+1}$

$$\varphi_i(z) = A_l e^{jk_z(z-d_l)} + B_l e^{-jk_z(z-d_l)} \quad (4)$$

where  $A_l$  and  $B_l$  are constants that need to be fixed with the continuity conditions at each of the interfaces  $z = d_l$ ,

$$\varphi_i(z) \text{ and } \frac{1}{m_e^*(z)} \frac{d\varphi_i(z)}{dz} \text{ continuous} \quad (5)$$

in conjunction with the relationship between the subband minimum energy  $E_i$  and the quantized wave vector  $k_z$  in the  $z$ -direction

$$E_i = \frac{\hbar^2 k_z^2}{2m_e^*(z)} + V_c(z) \quad (6)$$

where  $k_z$  assumes either real or imaginary value depending on  $E_i - V_c(z)$ . The continuity conditions in Eq.(5) ensure continuous electron distribution and conservation of electron current across the interface.

In the presence of an electric field  $\mathbb{E}$  applied in the  $z$ -direction, the potential term  $V_c(z)$  in the Schrödinger equation Eq.(2) becomes tilted along the  $z$ -direction according to  $-e\mathbb{E}z$ . If the Coulomb effect due to the distribution of electrons in the subband needs to be taken into consideration, then the potential in region  $l$  of the QW structure with the conduction band edge  $V_{c,l}$  should be modified as

$$V_c(z) = V_{c,l} - e\mathbb{E}z - e\phi(z) \quad (7)$$

where  $e\phi(z)$  takes into account the potential due to electron distributions in all subbands and can be obtained by solving the Poisson equation

$$\frac{\partial^2}{\partial z^2} \phi(z) = -\frac{e}{\varepsilon_0 \varepsilon(z)} \left[ \sum_i n_i |\varphi_i(z)|^2 - N_d(z) \right] \quad (8)$$

consistently with Eq.(2), where  $e$  is the charge of a free electron,  $\varepsilon_0$  is the permittivity of free space,  $\varepsilon(z)$  is the  $z$ -dependent dielectric constant of the QW structure,  $n_i$  is the electron density of subband  $i$ , and  $N_d(z)$  is the  $n$ -type doping profile in the structure.

In comparison with the conduction band, the situation in the valence band is far more complex mostly because of the interactions between subbands of different effective masses that produce strong nonparabolicity. The in-plane dispersion of valence subbands and their associated envelope functions can be obtained in the framework of the effective mass approximation by applying the  $\mathbf{k} \cdot \mathbf{p}$  theory [30] to QWs [31] where, in the most general treatment, an  $8 \times 8$  Hamiltonian matrix is employed to describe the interactions between the conduction, heavy-hole (HH), light-hole (LH), and spin-orbit split off (SO) bands. Often times, for semiconductors in which the conduction band is separated far in energy from the valence band, the coupling of the conduction band can be ignored. For the group-IV semiconductors Si and Ge with indirect band gaps, this approximation is particularly adequate. In those structures where there is little strain such as GaAs/AlGaAs, the SO band coupling can also be ignored. The  $8 \times 8$  Hamiltonian matrix can then be reduced to a  $4 \times 4$  matrix. But for systems with appreciable lattice mismatch, strain induces strong coupling

between LH and SO bands. For the SiGe system with a large lattice mismatch, the SO band should be included and a 6×6 matrix Hamiltonian equation needs to be solved to come up with the dispersion relations and envelope functions. Such a 6×6 matrix Hamiltonian equation can be solved exactly in multiple QW structures under the bias of an electric field. A procedure based on the Luttinger-Kohn Hamiltonian [32,33] is outlined as follows. The 6×6 Luttinger-Kohn Hamiltonian matrix including the uniaxial stress along (001) is given in the HH ( $|\frac{3}{2}, \pm\frac{3}{2}\rangle$ ), LH ( $|\frac{3}{2}, \pm\frac{1}{2}\rangle$ ), and SO ( $|\frac{1}{2}, \pm\frac{1}{2}\rangle$ ) Bloch function space as

$$H = - \begin{matrix} & \begin{matrix} |\frac{3}{2}, \frac{3}{2}\rangle & |\frac{3}{2}, \frac{1}{2}\rangle & |\frac{3}{2}, -\frac{1}{2}\rangle & |\frac{3}{2}, -\frac{3}{2}\rangle & |\frac{1}{2}, \frac{1}{2}\rangle & |\frac{1}{2}, -\frac{1}{2}\rangle \end{matrix} \\ \left[ \begin{matrix} P + Q & -S & R & 0 & -\frac{1}{\sqrt{2}}S & \sqrt{2}R \\ -S^\dagger & P - Q & 0 & R & -\sqrt{2}Q & \sqrt{\frac{3}{2}}S \\ R^\dagger & 0 & P - Q & S & \sqrt{\frac{3}{2}}S^\dagger & \sqrt{2}Q \\ 0 & R^\dagger & S^\dagger & P + Q & -\sqrt{2}R^\dagger & -\frac{1}{\sqrt{2}}S^\dagger \\ -\frac{1}{\sqrt{2}}S^\dagger & -\sqrt{2}Q & \sqrt{\frac{3}{2}}S & -\sqrt{2}R & P + \Delta & 0 \\ \sqrt{2}R^\dagger & \sqrt{\frac{3}{2}}S^\dagger & \sqrt{2}Q & -\frac{1}{\sqrt{2}}S & 0 & P + \Delta \end{matrix} \right] + V_v(z) \end{matrix} \quad (9)$$

where  $V_v(z)$  is the valence band edge profile (degenerate for HH and LH bands) of the QW structure,

$$\begin{aligned}
 P &= \frac{\hbar^2}{2m_0} \gamma_1 (k_x^2 + k_y^2 + k_z^2) - a_v (\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}) \\
 Q &= \frac{\hbar^2}{2m_0} \gamma_2 (k_x^2 + k_y^2 - 2k_z^2) - \frac{b}{2} (\epsilon_{xx} + \epsilon_{yy} - 2\epsilon_{zz}) \\
 S &= \frac{\hbar^2}{2m_0} 2\sqrt{3}\gamma_3 (k_x - jk_y)k_z \\
 R &= \frac{\hbar^2}{2m_0} \sqrt{3} [-\gamma_2 (k_x^2 - k_y^2) + 2j\gamma_3 k_x k_y]
 \end{aligned} \quad (10)$$

in which  $m_0$  is the mass of a free electron,  $\gamma_1, \gamma_2, \gamma_3$  are the Luttinger parameters and  $a_v, b$  are the deformation potentials [34] with different values in QWs and barriers, and the lattice mismatch strain

$$\epsilon_{xx} = \epsilon_{yy} = \frac{a_0 - a}{a}, \quad \epsilon_{yy} = -\frac{2C_{12}}{C_{11}} \epsilon_{xx} \quad (11)$$

with  $a_0, a$  being the lattice constants of the substrate (or buffer) and the layer material, and  $C_{11}$  and  $C_{12}$  the stiffness constants.

The Hamiltonian in Eq.(9) operates on wavefunctions that are combinations of six mutually orthogonal HH ( $|\frac{3}{2}, \pm\frac{3}{2}\rangle$ ), LH ( $|\frac{3}{2}, \pm\frac{1}{2}\rangle$ ), and SO ( $|\frac{1}{2}, \pm\frac{1}{2}\rangle$ ) Bloch functions

$$\psi_i(\mathbf{r}, z) = e^{j\mathbf{k}\cdot\mathbf{r}} \left[ \chi_1(z) \left| \frac{3}{2}, \frac{3}{2} \right\rangle + \chi_2(z) \left| \frac{3}{2}, \frac{1}{2} \right\rangle + \chi_3(z) \left| \frac{3}{2}, -\frac{1}{2} \right\rangle + \chi_4(z) \left| \frac{3}{2}, -\frac{3}{2} \right\rangle + \chi_5(z) \left| \frac{1}{2}, -\frac{1}{2} \right\rangle + \chi_6(z) \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \right] \tag{12}$$

where  $\chi_n(z), n = 1, 2, \dots, 6$  forms a six-component envelope-function vector  $\chi(z)$ . Each component in a given region  $l$  of the QW structure (either QW or barrier),  $d_l < z < d_{l+1}$ , is a superposition of the forward and backward propagations identical to Eq.(4) with constants  $A_{n,l}$  and  $B_{n,l}, n = 1, 2, \dots, 6$  that can be fixed by the continuity equations that require at each interface  $z = d_l$ ,

$$\chi(z) \text{ and } \begin{bmatrix} p+q & -s & 0 & 0 & -\frac{1}{\sqrt{2}}s & 0 \\ -s^\dagger & p-q & 0 & 0 & -\sqrt{2}q & \sqrt{\frac{3}{2}}s \\ 0 & 0 & p-q & s & \sqrt{\frac{3}{2}}s^\dagger & \sqrt{2}q \\ 0 & 0 & s^\dagger & p+q & 0 & -\frac{1}{\sqrt{2}}s^\dagger \\ -\frac{1}{\sqrt{2}}s^\dagger & -\sqrt{2}q & \sqrt{\frac{3}{2}}s & 0 & p & 0 \\ 0 & \sqrt{\frac{3}{2}}s^\dagger & \sqrt{2}q & -\frac{1}{\sqrt{2}}s & 0 & p \end{bmatrix} \chi(z) \text{ continuous} \tag{13}$$

where

$$\begin{aligned} p &= \gamma_1 \frac{\partial}{\partial z} \\ q &= -2\gamma_2 \frac{\partial}{\partial z} \\ s &= -\sqrt{3}j\gamma_3(k_x - jk_y) \end{aligned} \tag{14}$$

to maintain undisruptive carrier distribution and current across the interface.

It is important to point out that when the above described algorithms are used for the situation where an electric field is applied along the growth direction  $z$ , it is necessary to digitize the potential term  $V_c(z)$  and  $V_v(z)$ , i.e. the regions that are used in Eq.(4) are no longer defined by the QW and barrier boundaries; instead, there could be many regions within each QW or barrier depending on the number of digitization steps used to satisfy the accuracy requirement.

This procedure applied at each wave vector point ( $\mathbf{k} = k_x\hat{x} + k_y\hat{y}$ ) produces the in-plane dispersion for each subband. An example is illustrated in Fig. 5 for a  $70\text{\AA}/50\text{\AA}$

GaAs/Al<sub>0.3</sub>Ga<sub>0.7</sub>As SL [35]. In-plane dispersions of three subbands (two for HH and one for LH) are shown where strong nonparabolicity is demonstrated. It can be seen from Fig. 5 that the band nonparabolicity could be so severe that the LH subband maximum is no longer at the  $\Gamma$ -point which leads to useful valence QCL design applications in Section IV.

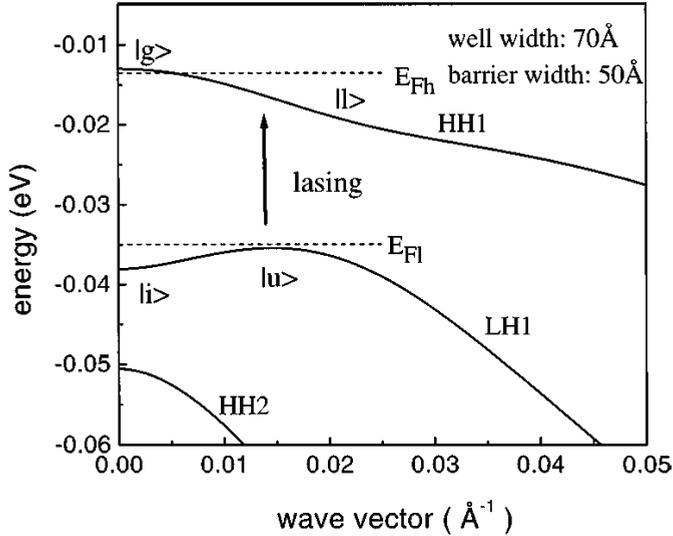


Fig. 5. In-plane dispersions of subbands HH1, LH1, and HH2 for a 70 Å / 50 Å GaAs/AlGaAs SL [35].

### 3.2 Optical gain

For lasing to occur between two subbands, it is necessary to induce stimulated emission between them. To sustain such emission of photons, there must be sufficient optical gain to compensate various losses in the laser structure. The intersubband optical gain can be obtained by analyzing transition rates between two subbands.

According to the Fermi Golden rule, the transition rate between two discrete states 1 and 2 that are coupled by a perturbation electro-magnetic (EM) field with a frequency of  $\omega$  is

$$g_{12} = \frac{2\pi}{\hbar} |H_m|^2 \delta(E_2 - E_1 - \hbar\omega) \tag{15}$$

where  $H_m = \langle 1|H_{ex}|2\rangle$  is the transition matrix element under the influence of a perturbation Hamiltonian  $H_{ex}$  between the two states with an exact transition energy  $E_2 - E_1$  in the absence of any broadening. In reality, the transition line  $E_2 - E_1$  is not infinitely sharp and is always broadened. As a result,  $E_2 - E_1$  is not known exactly, instead a probability for it to appear in the energy interval  $E \rightarrow E + dE$  is described. In the case of homogeneous broadening, this probability should be given as  $L(E)dE$  with the Lorentzian lineshape centered at some peak transition energy  $E_0$

$$L(E) = \frac{\Gamma/2\pi}{(E - E_0)^2 + \Gamma^2/4} \tag{16}$$

where  $\Gamma$  is the full width at half maximum (FWHM) that characterizes the broadening due to various homogeneous processes that include collisions and transitions. The transition rate in Eq.(15) should thus be modified by an integral that takes into account of this broadening as

$$g_{12} = \frac{2\pi}{\hbar} \int |H_m|^2 \delta(E - \hbar\omega) L(E) d = \frac{2\pi}{\hbar} |H_m|^2 \frac{\Gamma/2\pi}{(\hbar\omega - E_0)^2 + \Gamma^2/4} \quad (17)$$

essentially replacing the  $\delta$ -function in Eq.(15) with the Lorentzian lineshape Eq.(16). In the presence of an EM field with an optical potential vector  $\mathbf{A}$  in a medium with isotropic effective mass, the perturbation Hamiltonian  $H_{ex}$  that describe the interaction between the field and electron in isotropic subbands is

$$H_{ex} = \frac{e\mathbf{A} \cdot \mathbf{P}}{m_e^*} \quad (18)$$

where  $\mathbf{P}$  is the momentum operator.

From Eq.(18), it is not difficult to see that the selection rules for intersubband transitions in the conduction band are such that only those EM fields that are polarized in the growth direction ( $z$ ) can induce optical transitions. The transition matrix element can then be given as

$$H_m = \frac{eA}{m_e^*} P_{12} \quad (19)$$

where the momentum matrix element

$$P_{12} = \left\langle \varphi_1 \left| -j\hbar \frac{\partial}{\partial z} \right| \varphi_2 \right\rangle \quad (20)$$

is evaluated as the envelope function overlap between the two subbands, which is related to the dipole matrix element [36]

$$z_{12} = \langle \varphi_1 | z | \varphi_2 \rangle = \frac{i\hbar}{m_0 E_{12}} P_{12} \quad (21)$$

and to the oscillator strength [37]

$$f_{12} = \frac{2m_0}{m_e^* E_{12}} |P_{12}|^2. \quad (22)$$

It is not difficult to see from Eq.(17) that the transition rate induced by an EM field between two eigen states is the same for upward and downward transitions. Now let us apply Eq.(17) to intersubband transitions between the upper subband 2 and lower subband 1 in the conduction band (Fig. 3). Since momentums associated with photons are negligible, all photon-induced transitions are vertical in  $\mathbf{k}$ -space. It is therefore possible to obtain a net downward transition rate (in the units of number of transitions per unit time per unit sample area) between the two subbands by evaluating the following integral

$$g_{\text{net}} = \int g_{12} \{ f_2(E_{2,k}) [1 - f_1(E_{1,k})] - f_1(E_{1,k}) [1 - f_2(E_{2,k})] \} \rho_r(E_{2,k} - E_{1,k}) d(E_{2,k} - E_{1,k}) \quad (23)$$

where  $f_1(E_{1,k})$  and  $f_2(E_{2,k})$  are the electron occupation probabilities of those states at the same  $\mathbf{k}$  in subbands 1 and 2, respectively, and  $\rho_r(E_{2,k} - E_{1,k})$  is the reduced density of states (DOS) between  $E_{1,k}$  and  $E_{2,k}$ , which is equal to DOS of subbands 1 ( $\rho_1$ ) and 2 ( $\rho_2$ ) when they are parallel,  $\rho_r = \rho_1 = \rho_2 = m_e^*/\pi\hbar^2$ . Since the Lorentzian lineshape in Eq.(23) should be much broader than the spread of the energy transitions between the two parallel subbands which can be approximated as sharply centered at the subband separation at their energy minima  $E_{12} = E_2 - E_1$ . Thus,

$$g_{net} = \frac{2\pi}{\hbar} |H_m|^2 \frac{\Gamma/2\pi}{(\hbar\omega - E_{12})^2 + \Gamma^2/4} \left[ \int f_2(E_{2,k}) \rho_2 dE_{2,k} - \int f_1(E_{1,k}) \rho_1 dE_{1,k} \right] \tag{24}$$

$$= \frac{2\pi}{\hbar} |H_m|^2 \frac{\Gamma/2\pi}{(\hbar\omega - E_{12})^2 + \Gamma^2/4} (N_2 - N_1)$$

where  $N_1$  and  $N_2$  are the total electron densities in subband 1 and 2 per unit area, respectively.

The optical gain coefficient  $\gamma$  that describes the increase of the EM field intensity,  $I$ , as  $\gamma = I^{-1}dI/dz$  can be defined as power increase per unit volume divided by the intensity, which in turn can be expressed in terms of the net downward transition rate Eq.(24) using the momentum and dipole matrix element relation

$$\gamma(\omega) = \frac{g_{net}\hbar\omega}{IL_p} \tag{25}$$

where  $L_p$  is the length of the QW structure that is equal to the length of one period in case of QCLs. In order to relate the EM field intensity  $I$  that propagates in in-plane with the optical potential  $\mathbb{A}$  polarized along  $z$ , a real expression for the potential  $\mathbb{A}$  has to be used

$$\mathbb{A} = A_0 \cos(\boldsymbol{\beta} \cdot \mathbf{r} - \omega t) \hat{\mathbf{z}} = \frac{1}{2} A_0 \hat{\mathbf{z}} [e^{j(\boldsymbol{\beta} \cdot \mathbf{r} - \omega t)} + e^{-j(\boldsymbol{\beta} \cdot \mathbf{r} - \omega t)}] \tag{26}$$

where  $\boldsymbol{\beta}$  is the in-plane propagation wave vector of the EM field. It is easy to see that only one of the two terms on the right side of Eq.(26) couples with subbands 1 and 2,  $E_2 - E_1 = \hbar\omega$ . Thus, the optical potential that participates in the transition matrix Eq.(19) is only half of its real amplitude,  $A = A_0/2$ . Since the EM field intensity  $I$  is related to the optical potential amplitude  $A_0$  as  $I = \epsilon_0 c n_{eff} A_0^2 \omega^2 / 2$ , Eq.(25) can be written as

$$\gamma(\omega) = \frac{e^2 |P_{12}|^2}{2\epsilon_0 c n_{eff} m_e^{*2} \omega L_p} \frac{\Gamma}{(\hbar\omega - E_{12})^2 + \Gamma^2/4} (N_2 - N_1) \tag{27}$$

$$= \frac{e^2 m_0^2 \omega z_{12}^2}{2\epsilon_0 c n_{eff} m_e^{*2} L_p} \frac{\Gamma}{(\hbar\omega - E_{12})^2 + \Gamma^2/4} (N_2 - N_1)$$

where  $c$  is the speed of light in free space and  $n_{eff}$  is the effective index of refraction of the QCL dielectric medium. The population inversion  $N_2 - N_1 > 0$  is clearly necessary in order to achieve positive gain which peaks at the frequency  $\omega_0 = E_{12}/\hbar$  with a value of

$$\gamma(\omega_0) = \frac{2e^2 m_0^2 \omega z_{12}^2}{\epsilon_0 c n_{eff} m_e^{*2} \Gamma L_p} (N_2 - N_1) \tag{28}$$

For transitions between valence subbands with nonparallel dispersions and strong mixing between HH, LH, and SO bands, we have to re-examine the intersubband transition rate.

Consider the intersubband transition in Fig. 5 from the upper state  $|u\rangle$  in subband LH1 to the lower state  $|l\rangle$  in subband HH1, if the spread of intersubband transitions is wide enough compared to the homogeneous broadening, the Lorentzian lineshape in the net downward transition rate Eq.(23) can be approximated as a  $\delta$ -function yielding

$$g_{net}^{(v)} = \frac{2\pi}{\hbar} \left| H_m^{(v)} \right|^2 \rho_r(E_l - E_h) [f_{LH}(E_l) - f_{HH}(E_h)] |_{E_l - E_h = \hbar\omega} \tag{29}$$

where  $\rho_r(E_l - E_h)$  is the reduced DOS for the transition between subbands LH1 and HH1,  $f_{LH}(E_l)$  and  $f_{HH}(E_h)$  are hole occupation probabilities at states with energy  $E_l$  and  $E_h$  in subband LH1 and HH1, respectively, at the same in-plane wave vector  $\mathbf{k}$  separated by a photon energy  $\hbar\omega$ , and the optical transition matrix between LH1 and HH1 taking into account of the mixing

$$H_m^{(v)} = \sum_{n=1}^6 \frac{eA}{m_n^*} \left\langle \chi_n^{(l)} \left| -j\hbar \frac{\partial}{\partial z} \right| \chi_n^{(h)} \right\rangle \tag{30}$$

where  $\chi_n^{(l)}$  and  $\chi_n^{(h)}$  are respectively the  $n$ -th component of the envelope function vectors for subband LH1 and HH1 as defined in Eq.(12), and  $m_n^*$  are the corresponding hole effective mass in  $z$ -direction with  $m_{1,4}^* = m_0/(\gamma_1 - 2\gamma_2)$  for HH,  $m_{2,3}^* = m_0/(\gamma_1 + 2\gamma_2)$  for LH, and  $m_{5,6}^* = m_0/\gamma_1$  for SO. The optical gain can then be expressed

$$\begin{aligned} \gamma^{(v)} &= \frac{\pi e^2}{\epsilon_0 c n_{eff} \omega L_p} \left| \sum_{n=1}^6 \frac{P_n^{(lh)}}{m_n^*} \right|^2 \rho_r(E_l - E_h) [f_{LH}(E_l) - f_{HH}(E_h)] |_{E_l - E_h = \hbar\omega} \\ &= \frac{\pi e^2 m_0^2 \omega}{\epsilon_0 c n_{eff} L_p} \left| \sum_{n=1}^6 \frac{z_n^{(lh)}}{m_n^*} \right|^2 \rho_r(E_l - E_h) [f_{LH}(E_l) - f_{HH}(E_h)] |_{E_l - E_h = \hbar\omega} \end{aligned} \tag{31}$$

in terms of momentum matrix  $P_n^{(lh)} = \left\langle \chi_n^{(l)} \left| -j\hbar \frac{\partial}{\partial z} \right| \chi_n^{(h)} \right\rangle$  as well as dipole matrix elements  $z_n^{(lh)} = \left\langle \chi_n^{(l)} \left| z \right| \chi_n^{(h)} \right\rangle$  between the same  $n$ -th component of the envelope function vectors of the two valence subbands.

In comparison with the optical gain Eq.(28) for the conduction subbands, we can see that it is not necessary to have total population inversion,  $N_l - N_h > 0$ , in order to have positive gain between the valence subbands. Instead, all we need is local population inversion  $[f_{LH}(E_l) - f_{HH}(E_h)] |_{E_l - E_h = \hbar\omega} > 0$  in the region where the intersubband transition takes place (those states near  $|u\rangle$  and  $|l\rangle$  in Fig. 5).

### 3.3 Intersubband lifetimes

It has been established in Eqs.(27) and (28) that the population inversion between the upper (2) and lower (1) subbands,  $N_2 - N_1 > 0$ , is necessary in order to obtain optical gain. But what determines the population inversion? This question is answered with the analysis of lifetimes of these subbands as a result of various intersubband relaxation mechanisms including carrier-phonon, carrier-carrier, impurity, and interface roughness scattering processes. Among them, phonon scattering is the dominant process, especially when the energy separation between the two subbands exceeds that of an optical phonon, in which

case the transitions from upper to lower subband are highly efficient with the emission of optical phonons. Different from the optical transitions, these scattering processes do not necessarily occur as vertical transitions in  $k$ -space. In the case of phonon scattering, the conservation of in-plane momentum can be satisfied by a wide range of momentum of involved phonons as shown in Fig. 6(a) where intersubband as well as intrasubband transitions due to phonon scattering are illustrated.

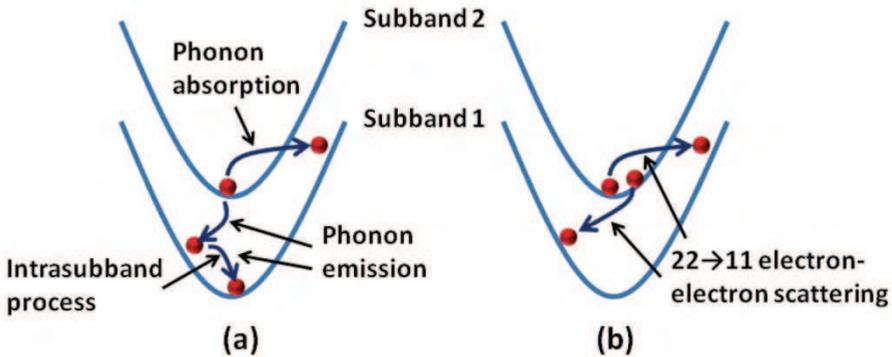


Fig. 6. (a) Intersubband and intrasubband transitions due to electron-phonon scattering (b) the 22→11 transition induced by the electron-electron scattering.

Up to now, practically all approaches in developing Si-based QCLs are based on materials from group-IV, mostly Si, Ge, SiGe alloy, and more recently, SiGeSn alloy. Different from the polar III-V and II-VI semiconductors, group-IV materials are nonpolar. The carrier scatterings by nonpolar optical phonons are much slower than those due to polar optical phonons [38]. Starting from Fermi Gold rule Eq.(15), the scattering rate for a carrier in subband 2 with the in-plane wave vector  $k$  to subband 1 with  $k'$  by a phonon with an energy  $\hbar\omega_Q$  and wave vector,  $Q = q + q_z\hat{z}$ , can be expressed as an integral over all the participating phonon states

$$\frac{1}{\tau_{12}} = \frac{2\pi}{\hbar} \int |H_{ep}|^2 \delta(E_{2,k} - E_{1,k'} \mp \hbar\omega_Q) dN_f \tag{32}$$

where  $H_{ep}$  is the electron-phonon interaction matrix element, the carrier energies  $E_{1,k'}$  and  $E_{2,k}$  are given by Eq.(1) for conduction subbands, but for valence subbands, they need to be obtained by the  $k \cdot p$  method described above. We will proceed with the following approximations: 1) all phonons are treated to be bulk-like by neglecting the phonon confinement effect in QW structures, 2) energies of acoustic phonons are negligible  $\hbar\omega_Q \approx 0$ , and 3) optical phonon energies are taken as a constant  $\hbar\omega_Q \approx \hbar\omega_0$ . The matrix element of carrier-phonon interaction for different type of phonons can be written as [39,40]

$$|H_{ep}|^2 = \begin{cases} \frac{\Xi^2 K_B T}{2c_L \Omega} \delta_{q,\pm(k'-k)} |G_{12}(q_z)|^2, & \text{acoustic phonon} \\ \frac{\hbar D^2}{2\rho\omega_0 \Omega} \delta_{q,\pm(k'-k)} \left[ n(\omega_0) + \frac{1}{2} \mp \frac{1}{2} \right] |G_{12}(q_z)|^2, & \text{nonpolar optical phonon} \end{cases} \tag{33}$$

where the upper sign is for absorption and lower for emission of one phonon,  $K_B$  is the Boltzmann constant,  $\Omega$  is the volume of the lattice mode cavity,  $c_L$  is the elastic constant for acoustic mode,  $\mathcal{E}$  and  $D$  are the acoustic and optical deformation potential, respectively, and  $n(\omega_0)$  is the number of optical phonons at temperature  $T$ ,

$$n(\omega_0) = \frac{1}{\exp(\hbar\omega_0/K_B T) - 1}. \tag{34}$$

The wavefunction interference effect between conduction subbands is

$$G_{12}(q_z) = \langle \varphi_1 | e^{jq_z z} | \varphi_2 \rangle \tag{35}$$

and between valence subbands is

$$G_{12}(q_z) = \sum_{n=1}^6 \langle \chi_n^{(1)} | e^{jq_z z} | \chi_n^{(2)} \rangle. \tag{36}$$

The Kronecker symbol  $\delta_{q,\pm(k'-k)}$  in the matrix element Eq.(33) represents the in-plane momentum conservation  $\mathbf{k}' = \mathbf{k} \pm \mathbf{q}$ .

Since phonon modes have density of states  $\Omega/(2\pi)^3$ , the participating phonon states in the integral Eq.(32) can be expressed as

$$dN_f = \frac{\Omega}{(2\pi)^3} q dq d\theta dq_z \tag{37}$$

where  $\theta$  is the angle between  $\mathbf{k}$  and  $\mathbf{q}$ . For conduction subbands with a parabolic dispersion Eq.(1), the phonon scattering rate Eq.(32) can be evaluated analytically

$$\frac{1}{\tau_{12}} = \begin{cases} \frac{\mathcal{E}^2 K_B T m_e^*}{4\pi c_L \hbar^3} \int |G_{12}(q_z)|^2 dq_z, & \text{acoustic} \\ \frac{D^2 m_e^* \left[ n(\omega_0) + \frac{1}{2} \mp \frac{1}{2} \right]}{4\pi \rho \hbar^2 \omega_0} \int |G_{12}(q_z)|^2 dq_z, & \text{nonpolar optical.} \end{cases} \tag{38}$$

But for valence subbands where there is a strong nonparabolicity, Eq.(32) can no longer be integrated analytically. However, if we take the wave vector of the initial state in subband 2 to be at the  $\Gamma$ -point  $\mathbf{k} = 0$ , then the phonon wave vector  $\mathbf{q} = \mathbf{k}'$ , Eq.(38) can still be used to evaluate the phonon scattering rate between valence subbands by substituting the effective mass with some average effective mass in the final subband 1.

The phonon scattering rate in Eq.(38) has been used to compare the lifetimes of two similar three-level systems, SiGe/Si and GaAs/AlGaAs, as shown in Fig. 7(a) [38]. The lifetime difference between the upper (3) and lower (2) subband is calculated as the function of the transition energy  $E_3 - E_2$  which is varied by changing the barrier width between the two QWs that host the two subbands. The main result in Fig. 7(b) is that the lifetimes in the SiGe system can be an order of magnitude longer than in the GaAs/AlGaAs system because of SiGe's lack of polar optical phonons. This property can potentially lead to significantly reduced lasing threshold for the SiGe system. The sudden drops in the lifetimes have to do with the shifting of subband energy separations  $E_2 - E_1$  and  $E_3 - E_2$ , to either below or above the optical phonon energy.

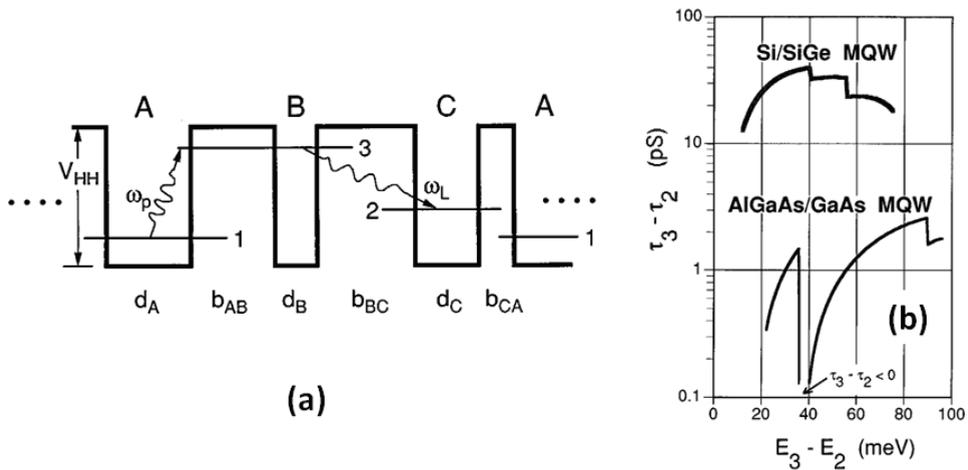


Fig. 7. (a) HH valence band diagram of one period of a SiGe/Si SL with hole energy increase in the upward direction, and (b) comparison of Lifetime difference ( $\tau_3 - \tau_2$ ) between the SiGe/Si and GaAs/AlGaAs SL (in a similar three scheme) as a function of the transition energy ( $E_3 - E_2$ ) [38].

Among the different phonon scattering processes – emission and absorption of acoustic and optical phonons, the emission of an optical phonon is by far the fastest process. But in far-IR QCLs when the subband energy separation is less than the optical phonon energy and the emission of an optical phonon is forbidden, phonon scattering may no longer be the dominant relaxation mechanism. Other scattering mechanisms need to be taken into consideration, such as the carrier-carrier [41], impurity [42], and interface roughness scatterings [43], all of which are elastic processes. The carrier-carrier scattering is a two-carrier process that is particularly important when carrier concentration is high which increases the possibility of two carriers interacting with each other. There are many possible outcomes as a result of this interaction in inducing intersubband as well as intrasubband transitions. Among them, the  $22 \rightarrow 11$  process where both carriers originally in subband 2 end up in subband 1 is the most efficient one in terms of inducing intersubband transitions (Fig. 6(b)). It has been reported in experiment that the intersubband transition times on the order of tens of ps have been observed for carrier densities of  $10^9 \sim 10^{11}/\text{cm}^2$  in GaAs/AlGaAs QWs [44]. In QCLs, since doping is mostly introduced away from the active region where optical transitions take place, impurity scattering does not seem to play a major role in determining the lifetimes of laser subbands, however, its influence on carrier transport in the injection region can be rather important. Interface roughness depends strongly on the process of structural growth, its impact on scattering should be more significant on narrow QWs, particularly for those transitions between two wavefunctions that are localized in MQWs that span across several interfaces.

#### 4. Valence band SiGe QCLs

Up to now, all of the demonstrated QCLs are based on epitaxially grown III-V semiconductor heterostructures such as GaInAs/AlInAs, GaAs/AlGaAs, and InAs/AlSb,

## Thank You for previewing this eBook

You can read the full version of this eBook in different formats:

- HTML (Free /Available to everyone)
- PDF / TXT (Available to V.I.P. members. Free Standard members can access up to 5 PDF/TXT eBooks per month each month)
- Epub & Mobipocket (Exclusive to V.I.P. members)

To download this full book, simply select the format you desire below

