Journal of Nonlinear Optical Physics and Materials
Vol. 4, No. 1 (1995) 191–243

MULTIBAND FINITE ELEMENT MODELING OF WAVEFUNCTION-ENGINEERED ELECTRO-OPTICAL DEVICES

L. R. RAM-MOHAN
Departments of Physics, Electrical and Computer Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, USA

J. R. MEYER
Code 5613, Naval Research Laboratory, Washington, DC 20375, USA

Received 30 November 1994

Recent advances in the modeling of semiconductor heterostructures with complex geometries allow one to go beyond band-structure engineering to the more general concept of *wavefunction engineering*. In this work, we illustrate how tailoring the band mixing and spatial distribution of the carriers leads to an expanded degree of control over such properties as the dispersion relations, interband and intersubband transition matrix elements, nonlinear optical and electro-optical coefficients, and lifetimes. The computations are based on a multiband finite element method (FEM) approach which readily yields energy levels, electron and hole wavefunctions, and optical matrix elements for heterostructures with arbitrary layer thickness, material composition, and internal strain. Application of the FEM to laterally-patterned heterostructures is also discussed.

1. Introduction

It has been more than two decades since Esaki and Tsu[1] originally proposed the superlattice as an artificial substance whose properties could be tailored through a binary modulation of the material composition along one spatial axis. Once the fabrication of multi-layers with thicknesses less than 100 Å became feasible with the advents of molecular beam epitaxy (MBE) and metalorganic chemical vapor deposition (MOCVD),[2,3] that concept laid the groundwork for a vast array of optical and electronic devices based on quantum-mechanical phenomena in semiconductor heterostructures. Commercially-significant examples include the double-barrier resonant tunneling diode,[4] the quantum-well diode laser[5,6] and the high-electron-mobility transistor (HEMT).[7]

Although materials design through the adjustment of the layer thicknesses and compositions was a recurring theme from the outset, about ten years ago Capasso and co-workers significantly expanded the available degrees of freedom by

introducing the concept of 'bandgap engineering', which is essentially an arbitrary grading of the energy gap so as to achieve independent modulation of the field-dependent conduction and valence band profiles in the presence of an applied or internal electric field.[8,9] New devices resulting from this approach included avalanche photodiodes with enhanced ionization ratios and hence reduced noise, channeling diodes, high-speed photo-transistors, and bipolar transistors with graded-gap base regions.

Over the past decade, improved growth techniques (which now allow control down to the atomic scale) coupled with rapid advances in device-fabrication technologies (including the capability of inducing quantization along more than one spatial dimension) have encouraged a trend toward increasingly complex structural configurations. These employ not only a larger number of material constituents in the same device, but also more intricate substructures (often involving asymmetry) and an expanded menu of parameters which may be modulated. Although the term 'bandgap engineering' is sometimes still used, it is by now far too restrictive to encompass the enhanced flexibility which may currently be exploited in the design of advanced heterostructure materials. The same also applies to the alternatives 'band-structure engineering' and 'band engineering' which sometimes appear in the current literature.[10–12] We will instead base our discussion on the more general concept of '*wavefunction engineering*', which is intended to signify no less than a fundamental redesign of the electronic states at the quantum mechanical level, so as to maximize the desired band mixing and spatial distribution of the electron and hole wavefunctions.

Accurate theoretical modeling of the electronic dispersion relations and optical properties of the new multi-layered quantum heterostructures entails a serious computational investment. Whereas standard multiband $\mathbf{k} \cdot \mathbf{P}$ model calculations are adequate for treating simple, two-constituent quantum wells and superlattices, the boundary conditions become intractable whenever there is either a larger multiplicity of constituents or a substructure within each period. Here we discuss a multiband finite element implementation of the $\mathbf{k} \cdot \mathbf{P}$ model, which is particularly well suited for taking into account the details of the geometry and the complex boundary conditions imposed on the wavefunctions of the carriers. This computational package permits a straightforward and flexible construction of the geometrical model for the heterostructure. The energy levels and wavefunctions of the carriers can then be obtained in a routine manner. Excellent precision can be achieved by improving on the standard engineering finite element methods for application to quantum heterostructures.

The finite-element method (FEM) has long been a standard computational tool in such engineering and scientific fields as structural mechanics, fluid dynamics, and atmospheric modeling,[13] but has rarely been applied to quantum mechanics except in crude one-band approximations. This is probably due in large part to the unfortunate perception that the FEM is a clumsy method of last resort, which should be avoided whenever more elegant techniques are available. However, it will

become apparent in the following sections that the FEM is in fact ideally suited to the types of computation required here, i.e. finding solutions to complex systems of coupled nonlinear and partial differential equations with complicated boundary conditions.

The algorithm derives much of its power from a particularly effective manner of treating the boundary conditions, as well as from an optimization of the non-uniform inter-element boundary placements. This separates it from more elementary FEM calculations carried out previously for quantum systems, even in the limit where only one band is considered. Thus, besides enabling one to accurately calculate the electronic structure and optical properties of complex structures which would otherwise be intractable without severe approximations, our FEM algorithm is also remarkably efficient. We find that the computer time required to treat simple two-constituent quantum wells or superlattices is only marginally greater than that needed for a conventional 8-band $\mathbf{k} \cdot \mathbf{P}$ calculation. It should be noted that the tight-binding approach.[14,15] may be viewed as a limiting form of the FEM, in which each atom or plane of atoms functions as a separate element. However, while tight-binding calculations can in principle treat many of the same problems, excessive matrix sizes and computation times significantly limit their practicality when applied to complex systems of the type which are quite tractable within the FEM. Furthermore, the tight-binding parameters are not directly related to experimentally-observable quantities, as in $\mathbf{k} \cdot \mathbf{P}$, which adds another layer of uncertainty to the input parameters.

The implementation of this high-precision FEM model has several useful aspects. These include the ability to incorporate any III-V or II-VI direct gap semiconducting material in any geometry, and to include the effects of built-in strain and external perturbations such as electric or magnetic fields or hydrostatic pressure. One of the most valuable features is the ability to generate the carrier wavefunctions and then display them on the computer screen. This visualization should be viewed as an integral component of wavefunction engineering as it is ideally practiced, since in effect it allows one to manipulate the electron and hole spatial distributions dynamically to achieve a desired set of properties. The tailoring may be accomplished, e.g., by changing the geometrical placement of different materials in the heterostructure so as to maximize the desired localization or shape of the wavefunctions.

This has profound implications for both optoelectronics device design and the exploration of fundamental physical issues. By tailoring the wavefunctions we can control the optical selection rules, optical matrix elements, carrier lifetimes, overlap integrals, tunneling currents, electro-optical and nonlinear optical coefficients, and so on. The discovery that carriers with energies above the barrier height in superlattices are localized in the barrier layers and have quantum well-like spectra is leading to a new spectroscopy of above-barrier states. Other quantum phenomena that are currently being explored with the finite element method with very interesting results include surface quantum wells and their role in defining heterostructure optical properties, bound states and their symmetry properties in laterally-confined

systems, control over optical nonlinearities in checkerboard superlattices of quantum wires, modification of the density of states in Type-II laser diodes, and the suppression of saturation effects in intersubband second-harmonic generation devices.

In the following, we elaborate on the implications of wavefunction engineering and describe the integration of the 8-band $\mathbf{k}\cdot\mathbf{P}$ model for calculating band structures of composite quantum heterostructures into the finite element method. Following a summary of the multiband FEM theory in Sec. 2, we will provide in Sec. 3 an overview of the numerous mechanisms and design parameters which may be exploited in wavefunction engineering as it is currently practiced. We then consider several examples illustrating application of the FEM modeling and wavefunction engineering to quantum systems of practical and fundamental interest. Calculated wavefunctions, interband and intersubband optical matrix elements, and energy levels and dispersion relations will be discussed for both complex multi-layer 2D quantum wells (Sec. 4) and lateral superlattices of 1D quantum wires (Sec. 5).

2. Theory

We first consider planar layered semiconductor heterostructures with the modulation only along the growth direction z. The layers are taken to be composed of compound III-V or II-VI semiconductors with their conduction- and valence-band edges located at the Γ-point in the Brillouin zone (BZ). The periodic components of the Bloch functions, $u_{j,k=0}(\mathbf{r})$, with j being the band index, at the band edges are assumed not to differ much as we traverse the layer interfaces.[16–18] We assume that the original bulk crystal translational symmetry is maintained in the transverse direction.

We consider the zone-center bulk band structure of the constituent semiconductors, within the spirit of the $\mathbf{k}\cdot\mathbf{P}$ model. The usual eight-band model consists of the Γ_6 conduction band (c), the Γ_8 heavy-hole (hh) and light-hold (lh) bands and the Γ_7 spin-orbit split-off band (s.o.), with their spin degeneracies. The complete 8×8 Hamiltonian is displayed in Ref. 19. However, for the sake of clarity in the presentation, we will consider here the simplified example of a layered system with no external electric or magnetic fields or built-in strain, and in the limit of vanishing-in-plane wavevector $[k_{\|}\equiv(k_x^2+k_y^2)^{1/2}\rightarrow 0]$. We note that the following considerations also hold for the more general case, except that there the dimensions of the matrices correspond to the full 8-band model and with external perturbations or finite $k_{\|}$ the Kramers degeneracy of the bands is lifted.

With these simplifications, the problem reduces to a three-band model, with the hh band factoring out. Within the envelope function approximation, the problem then reduces to the solution of a set of three simultaneous second-order differential equations for the envelope functions of the constituent layers. We have

$$H_{ij}(k_{\|}=0,\,k_z)\;f_j(z)=Ef_i(z)\,, \tag{1}$$

where k_z has to be replaced by the differential operator $-i\partial/\partial z$.

The eigenvalues of the 3×3 matrix are given by the secular equation (in atomic units)

$$\begin{vmatrix} E_c + (F + \frac{1}{2})k_z^2 - E & -\sqrt{E_P/3}k_z & -\sqrt{E_P/6}k_z \\ -\sqrt{E_P/3}k_z & E_v - \frac{1}{2}(\gamma_1 + 2\gamma_2)k_z^2 - E & -\sqrt{2}\gamma_2 k_z^2 \\ -\sqrt{E_P/6}k_z & -\sqrt{2}\gamma_2 k_z^2 & E_s - \frac{1}{2}\gamma_1 k_z^2 - E \end{vmatrix} = 0 . \quad (2)$$

Here E_c, E_v and E_s are the band edge energies of the conduction, lh and s.o. bands. The 3 coupled second-order differential equations represented by Eq. (2) can be written as

$$\left[-\mathcal{A}_{ab} \frac{\partial^2}{\partial z^2} - i\mathcal{B}_{ab} \frac{\partial}{\partial z} + \mathcal{C}_{ab} \right] f_b(z) = E f_a(z) . \quad (3)$$

While the matrix coefficients $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ in Eq. (3) are usually assumed to be constant in each layer, there are cases for which each material parameter should more generally be considered to be a function of coordinate, e.g., in a graded-bandgap system or a structure for which the strain is a function of position within a given layer. In a heterostructure, the differences in the band edge energies give rise to the confining potentials experienced by the carriers. In the full 8-band formalism, Eq. (3) consists of 8 coupled Schrödinger equations as discussed in Ref. 19.

The finite element method[20–23] may be used to solve Eq. (3) for symmetric as well as asymmetric quantum wells, for superlattices with two or more than two constituents in each period, for resonant tunneling structures, etc. In fact, *this method can accommodate the possibility of every material parameter being a function of coordinate.* It has been demonstrated that the present FEM can be adapted to yield very accurate eigenvalues for bound state problems,[24,25] and for obtaining solutions in quantum semiconductor heterostructures with complex geometries.[26,27] One begins by writing out the appropriate symmetrized (hermitian) Lagrangian that would generate Eq. (3) through a variational procedure. The integral over the physical region of the Lagrangian density, the action integral, is then split up into a number of 'cells' or elements, in each of which the physical considerations of the problem hold. The wave functions are assumed to be given locally in each element by fifth-order Hermite interpolation polynomials, which have the property that the expansion coefficients correspond to the values of the wave function and its derivatives at select points, called nodes, in the element. The global wave functions $f_i(z)$ are constructed by joining the locally-defined interpolation functions and matching the function and its derivatives across the element boundary for each of the bands included in the analysis. The *heterointerface* boundary conditions consisting of continuity of the envelope functions and of the *probability current*, and the boundary conditions for the bound states as well as for scattering states[28] at $z = \pm\infty$ are readily incorporated into the FEM. It is useful to derive the probability current density that is conserved across the interfaces by employing a gauge-variational approach for the multi-component wavefunctions using an extension of the gauge variational method of Gell-Mann and Levy.[29] The spatial dependence of the wavefunctions, manifested

through the interpolation polynomials, is next integrated out, leaving the action integral dependent on the unknown nodal values of the wavefunction. The usual variational principle is then implemented as a variation of the nodal values of $f^*(z)$ under which the action integral is a minimum. This *nodal variational principle* leads to a 'Schrödinger equation' for the nodal values. The integration of the action integral is performed element by element, giving rise to element matrices which are then overlaid into a global matrix in a manner consistent with the boundary conditions. This results finally in a generalized eigenvalue problem,[20,30] which may be solved for the eigenenergies and wave functions with a standard diagonalizer on a computer workstation.

Use of the 8-band FEM with three elements per layer leads to very accurate quantum well and superlattice energies and eigenfunctions. In the limit of simple geometries the eigenvalues agree with those obtained from the eight-band transfer-matrix method[19,31-33] to within 10^{-6} eV, and double-precision accuracy can be obtained by employing more elements in the computation. As an example of a band structure calculation,[34] we show in Fig. 1 the conduction and valence bands for a 18 Å/10 Å GaAs/$Al_{0.3}Ga_{0.7}As$ superlattices. This energy band diagram was generated in about 3 minutes on a DEC3100-workstation.

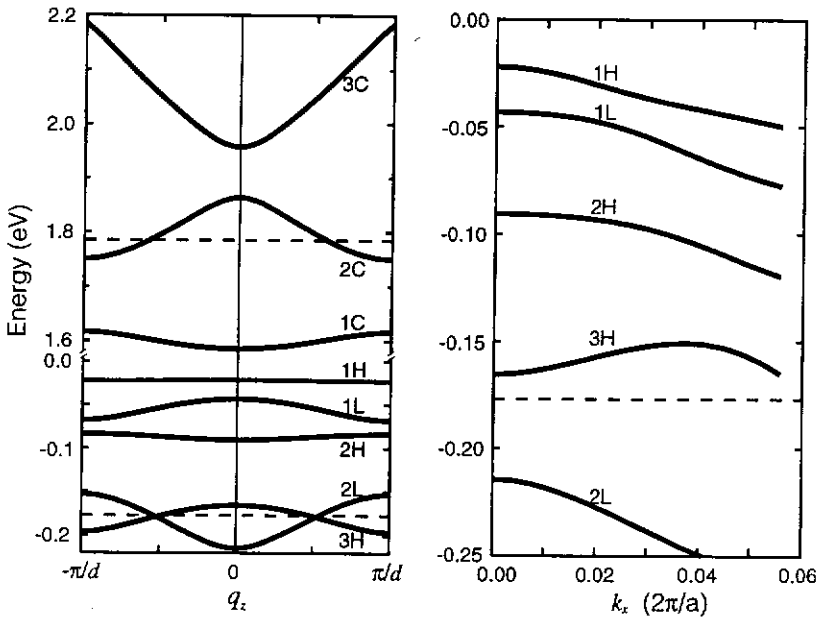

Fig. 1. Conduction and valence bands in the first Brillouin zone for a 18 Å/10 Å GaAs/$Al_{0.3}Ga_{0.7}As$ superlattice, as computed by the FEM. All energies are referenced to the GaAs valence band edge, and the dashed lines indicate the $Al_{0.3}Ga_{0.7}As$ band edges. The left panel is the superlattice dispersion along the growth direction, while the right panel is the in-plane dispersion (After Ref. 34).

The methods described above can be extended to the case of carrier confinement in two dimensions (quantum wires and other structures) and to three dimensions. The results based on two-dimensional considerations with just one energy band are already interesting and will be discussed in Sec. 5. The power of the finite element method becomes particularly apparent when applied to the two-dimensional Schrödinger equation, since the 'kitchen sink' potential experienced by a carrier in a quantum wire structure is non-separable. The FEM has revealed the interesting removal of 'accidental' degeneracy[26] in wires with a square cross-section. (This is relevant to the theory of dielectric microwave waveguides as well). A further example will be the calculation of wavefunctions, energy bands and optical nonlinearities in a bidirectionally-periodic stacking of quantum wires — the checkerboard superlattice.[27] Here again the potential is non-separable, and the FEM provides an elegant and natural approach to solving this problem.

3. Overview of Wavefunction Engineering

This section will briefly overview some of the numerous ways in which wavefunction engineering may be exploited, either to enhance the performance of optical devices or to probe new physical phenomena.

Even in the case of bulk semiconductors one can exercise a degree of control over the band structure, e.g., by varying the temperature or pressure, by applying an external electric or magnetic field, by varying the alloy composition or extrinsic doping level, or by employing optical excitation to generate non-equilibrium electrons and holes. However, a much richer range of phenomena becomes accessible when one introduces some additional type of modulation on a length scale small enough that the free carrier energy levels become quantized along at least one axis. Familiar examples include compositional modulation in layered heterostructures grown by MBE or MOCVD, band bending accompanied by charge redistribution in gated metal-insulator-semiconductor structures, carrier confinement induced by selective doping (e.g., in *n-i-p-i*s or modulation-doped structures), and strain-induced confinement resulting from a lattice-constant mismatch between the quantum-well layer and the substrate or buffer layer. In recent years, epitaxial growth technology has advanced to the point where even atomic-scale control over the interface characteristics has become a viable mechanism for band structure engineering (e.g., one can modulate the lateral island size for monolayer-scale fluctuations of the interface-boundary position,[35] or the valence band offset at the interface between two constituents having no common cation or anion[36,37]). Quantum wires, quantum dots, and other mesoscopic nanostructures can also be fabricated by a variety of techniques, including sub-micron etch lithography,[38] laterally-patterned gates,[38] laterally-selective doping or ion-beam damage,[39] laterally-modulated strain,[40] selective growth onto V-grooved substrates,[41] cleaved-edge overgrowth,[42] or vicinal growth onto misoriented substrates.[43]

With so many knobs now available for turning in a controllable and reproducible way, we find ourselves presented with numerous opportunities for fabricating new, artificial materials whose properties differ substantially from those of the bulk constituents. Optical and electronic devices based on the successful application of band structure engineering have been reviewed by a number of previous authors.[11,44–48] Here we will approach the topic from a slightly more fundamental standpoint by emphasizing the tools of the trade, i.e. the primary physical mechanisms which may be invoked in adapting the material properties to the needs of the various device classes.

3.1. *Modulation of Energy Levels and Dispersion Relations*

The most obvious and direct consequence of quantum confinement is that one gains a high degree of control over the electron and hole energy levels. This enables us to modify the fundamental band gap which is the key to detectors, lasers, self electro-optic effect devices (SEED),[49] etc. However, we in fact have a much broader capability of altering the fundamental nature of the electron and hole dispersion relations $[E(k_{\|}, k_z)]$. For example, when the dimensionality is reduced from 3D in bulk to 2D in a single quantum well or interface layer, we 'flatten out' the dispersion along k_z and thereby increase the density of states near the bottom of the band. It is well known that this has important consequences for laser thresholds and gains,[3,6,48] electro-optical (EO) and nonlinear optical (NLO) modulation coefficients and saturation properties,[50] resonant-tunneling-diode (RTD) peak-to-valley ratios,[51] etc. In a superlattice, varying the barrier thickness leads to a continuous tuning of the growth-direction dispersion, from the 3D limit where the growth-direction mass (m_z) can be equal to the in-plane mass ($m_{\|}$) to the 2D limit where m_z is effectively infinite. The collection efficiency and tunneling noise for photovoltaic infrared detectors are governed by m_z.[52,53] Adjustment of the miniband width through tuning of the barrier thicknesses can be used to obtain mass filtering, i.e. a nearly infinite ratio between the vertical mobilities for light-mass carriers (usually electrons) versus that for heavier carriers (heavy holes), which can lead to large gains in photoconductive detectors.[44] Miniband transport plays an important role in some classes of QWIP designs.[54,55]

Heterostructure geometry also influences the in-plane effective masses. First, a shift of the electron and light hole masses automatically accompanies the confinement-induced modulation of the direct energy gap, since $m_{\|} \propto E_g$ from $\mathbf{k} \cdot \mathbf{P}$ theory.[56] In addition, the in-plane hole mass is particularly sensitive (by more than two orders of magnitude in the extreme case of HgTe-CdTe superlattices[57]) to whether the highest-energy valence subband has a primarily heavy-hole-like or light-hole-like character (it is important to remember that the heavy hole is heavy along the growth axis but light in the plane, whereas the light hole is light along the growth axis but heavy in the plane).[46] The ordering of the valence subbands is generally governed by a competition between quantum confinement (which tends

to push the light-hole-like subbands to lower energies) and strain (which pushes the light hole down and the heavy hole up when the strain is compressive but has the opposite effect when it is tensile). Strain therefore becomes a powerful design parameter for band structure engineering.[58] In laser applications, it is advantageous to design structures with small in-plane hole masses, not only because the population-inversion threshold is then reduced, but also because the non-radiative Auger process is suppressed.[10,46] Whereas only compressive strain makes the hole mass lighter, even tensile strain has a beneficial effect on the polarization selection rules, since it serves to match the symmetry of the carrier distribution with the uniaxial symmetry of the laser beam.[11] Strain tuning of the valence bands also leads to improved properties for p-type quantum well infrared photodetectors (QWIPs) operating at normal incidence.[59] When the hole mass is modulated, there is naturally a corresponding modulation in both the valence-band density of states and the in-plane hole mobility. These manipulations of the valence band properties should be viewed as wavefunction engineering through modulation of the band mix, which is an alternative to modulation of the wavefunction's spatial profiles.

3.2. *Modulation of Energy Resonances and Forbidden Gaps*

Closely connected with the tuning of a given band's energy levels is the capability of bringing bands into and out of resonance with one another (e.g., in RTDs[44] and EO devices employing injection via tunneling[60]) or of bringing the intersubband energy splittings into resonance with a given photon energy (QWIPs, intersubband EO and NLO modulators, intersubband lasers,[61,62] etc.). In NLO applications, one can use resonances to enhance the nonlinear susceptibilities.[63–65] In other cases one would like to remove a resonance, e.g., that between the energy gap and the split-off gap which is detrimental to Auger lifetimes in many IR laser materials[11,66] (see Sec. 4.1.). In 0D quantum boxes, the energy levels can in principle be placed such that *all* Auger processes are forbidden.[67,68] One sometimes wishes to induce optical transparency by causing the photon energy to fall within a forbidden gap between the electronic levels, which again in the 0D limit can occur even above the fundamental band edge. In the opposite extreme, photonic bandgaps which do not support the propagation of light in certain spectral regions have been demonstrated.[69]

3.3. *Carrier-Concentration Modulation*

The electron and hole densities in a quantum heterostructure may be varied in a number of ways, e.g., through controlled doping, a gate voltage, spatial charge transfer, electrical or optical injection, thermal activation from impurity states or across the band gap, etc. The resulting values for n and p then govern many of the properties most relevant to device operation, including absorption saturation and gain, optical or electrical modulation of the refractive index (e.g., through Kramers-Kronig and plasma contributions), the Fermi energy [the key to field-effect transistors (FETs), RTDs, quantum transport, etc.] quasi-Fermi energies (popu-

lation inversion, NLO saturation, etc.), space charge and band bending, electrical conductivity, and screening (important, e.g., in spatial light modulators[70]). Frequently, complex graded-gap or multi-constituent band profiles are employed to tailor the Fermi energy or quasi-Fermi energies as a function of vertical position in the structure. Modulation doping for enhancement of the electron and hole mobilities is a central feature of many heterostructure electronic devices (e.g., HEMTs[7] and microwave FETs[71]) as well as for optical devices in which narrow linewidths are essential.[72,73]

3.4. *Modulation of the Conduction Valley Minima* (Γ, *L and X*)

Because the energy levels in a quantum heterostructure are so readily tuned, one has much more flexibility than in bulk for reordering the relative positions of the Γ, L and X conduction band minima. This choice radically affects the density of states, effective mass, mobility, and selection rules for both interband and intersubband optical transitions. The Γ valley has a light, isotropic mass and strong interband coupling, but nearly-vanishing intersubband oscillator strength at normal incidence.[74] By contrast, the L and X valleys have much heavier and highly-anisotropic masses, weak interband coupling, and strong intersubband interactions at normal incidence[75] as long as the growth axis is misaligned with the symmetry axes of the effective mass ellipsoids. In many high-speed electronic applications, one would like to minimize the role of L and X states because intervalley scattering represents a significant parasitic process.[76] However, intersubband transitions involving L-valley or X-valley states in $Si_{1-x}Ge_x$,[77] $Al_xGa_{1-x}As$,[78] or $Ga_{1-x}Al_xSb$[79–81] quantum wells are currently being investigated for potential use in QWIPs,[75] EO modulators,[82] and second-harmonic generation (SHG) devices[83] operating at normal incidence. Also interesting is the possibility of establishing a competition between Γ-valley and L-valley or X-valley states at approximately the same energy (see Sec. 42). We have recently proposed that for an asymmetric double-well structure with one well having its conduction band minimum at the Γ point and the other having its minimum at the L-point, one can obtain nearly complete dynamic transfer of the electron population from Γ states to L states through the variation of an external electric field.[84] We also note that Γ-X intervalley transfer can dominate the resonant tunneling characteristics of GaAs(Γ)/AlAs(X) single-barrier and double-barrier structures.[85,86] Furthermore, one should view the dominant growth-direction transport process for GaAs/AlAs superlattices with thick barriers in terms of intervalley scattering rather than tunneling through the barriers.[87]

3.5. *Type-I or Type-II Band Alignment*

Whereas an L-point or X-point conduction-band minimum leads to an indirect gap in momentum space, Type-II heterostructures such as Si-$Si_{1-x}Ge_x$, GaAs-AlAs (with thin GaAs layers), InAs-GaSb and ZnTe-ZnSe offer the additional possibility of an indirect gap in real space,[88,89] a phenomenon for which there is no analog

in bulk. Since the electrons are confined primarily to one layer while the holes are localized in another, the wavefunction overlap which governs the interband optical coupling, band mixing and repulsion (which determines the electron and light hole masses),[90] carrier-carrier scattering,[91] and recombination lifetime tends to be weak. However, an applied bias can be used to dynamically tune the overlap integrals, since the electric field either pulls both electron and hole wavefunctions toward the same interface or pushes them both farther away. This leads to a strong EO modulation of the optical coupling observed in photoluminescence[92] and photoconductivity[93,94] experiments, and also significantly enhances the interband Stark shift. Under some conditions, an applied field can also be used to convert the heterostructure from Type-II to Type-I.[95,96] A Type-II alignment is advantageous in applications for which weak optical coupling is required in order to minimize the device insertion loss, since the interband absorption coefficient can then be low even for above-gap photon energies.[89,97] Recombination lifetimes on the order of microseconds or longer have been observed in GaAs-AlAs Type-II heterostructures, since in that system the energy gap is indirect in both real space and momentum space.[98,99]

One general consequence of a Type-II band alignment is that the heterostructure energy gap becomes smaller than the gaps of either of the constituent semiconductors. In the extreme case of InAs-GaSb (the first Type-II system to be fabricated[100]), the resulting material can actually be semimetallic because the valence band offset is such that the conduction band minimum of InAs lies somewhat below the valence band maximum of GaSb. That arrangement provides a fertile basis for imaginative device concepts, particularly when consideration is extended to more complex geometries involving three or more constituents (e.g., incorporating AlSb, $Ga_{1-x}Al_xSb$ or $Ga_{1-x}In_xSb$). An example is the development of high-performance RTDs based on interband tunneling resonances.[101–103] Furthermore, it has been demonstrated that strong interband optical coupling can be achieved despite the Type-II band alignment, simply by making the quantum wells for both electrons and holes very thin so as to maximize the wavefunction overlap. When strain is introduced in an $InAsGa_{1-x}In_xSb$ heterostructure to further increase the semimetallic offset,[104] strong absorption can be combined with small energy gap for longwave (8–14 μm) and very longwave (> 14 μm) infrared detectors.[105] The strong optical coupling and large band offsets are also favorable for midwave infrared lasers[106,107] (see Sec. 4.1). It has recently been demonstrated that layers of elemental Sb can be integrated into InAs-GaSb-AlSb-family heterostructures grown along the (111) axis, enabling the possibility for a much larger Type-II semimetallic overlap between InAs and Sb.[108]

3.6. *Wavefunction Localization*

A Type-II heterostructure may be viewed as one example of a more general class of engineered materials for which one has the ability to strongly modulate the spatial position of the electron and hole wavefunctions. Wavefunction localization

is key to a wide variety of device properties, including electrical confinement for lasers, increased gain in photodetectors,[44] transistors based on real-space transfer[109] and electrical isolation in numerous devices. Localization along one axis in a quantum well significantly enhances the exciton binding energy (crucial in SEED modulators[49,110,111]) and similarly enhances the binding energy for shallow impurities.[112] Wannier-Stark localization due to removal of the miniband dispersion by an electric field is the basis for a family of electro-optical modulators.[113,114] In barrier, reservoir, and quantum well electron transfer (BRAQWET) devices, electron wavefunctions localized in a reservoir at zero bias significantly overlap a quantum well once a reverse bias is applied.[115] Bragg confinement can localize the wavefunction in a central well surrounded by short-period superlattices.[116] Field-induced modulation of the wavefunction overlap between two states localized primarily in different spatial regions, e.g., electrons and holes at a Type-II interface or two electron subbands in an asymmetric double quantum well, can dramatically alter not only the matrix elements for optical transitions,[100] but also the electron-hole recombination rates[97] and intersubband relaxation times.[117,118]

Another class of wavefunction confinement which has attracted attention recently is the localization of above-barrier states in the barrier regions rather than the well regions of a superlattice.[119] This effect has been confirmed experimentally and has led to a new type of spectroscopy which will be discussed further in Sec. 4.4. The hybridization of well and barrier states has also been studied.[120] A related phenomenon is 'kinetic confinement',[121–123] which tends to localize carriers with large in-plane wavevectors in regions where the effective mass is larger, even in a superlattice or quantum well barrier if the barrier is not too high.

3.7. *Wavefunction Manipulation*

By now it is well known that the dynamic modulation of quantum-well wavefunctions by an external electric field[124] is significantly enhanced in stepped wells and asymmetric double quantum wells (ADQWs).[125] This asymmetric distortion of the electron wavefunctions is being exploited in a wide variety of applications including SEED devices,[126] voltage-tunable QWIPs,[127,128] intersubband EO absorption and phase modulators,[125,129,84] IR bistable devices,[130] and intersubband SHG devices.[131,117,83] Furthermore, the trend in recent years has been toward quantum wells with complex internal structures (e.g., additional thin barriers or mini-wells within each well), which permit a more general tailoring of the wavefunctions. This is wavefunction engineering in its purest form, and it can be used to control the gaps between the minibands,[132,133] enhance the mobilities associated with optical phonon scattering[134] and ionized impurity scattering,[135] obtain dynamic 'wavefunction sweeping' from one side of the well to the other,[136] optimize the intersubband Stark shift,[137] enhance or suppress the dipole matrix elements for intersubband transitions,[12,138] enhance the responsivity for QWIPs,[139] shift the intersubband transition energies,[140–142] or shift the relative energies for interband

transitions vs subband index.[143] It has also been proposed that the incorporation of periodic δ-doping planes of impurities into a superlattice can be employed to create and annihilate bands,[144] significantly broaden the minibands,[145] or suppress the mobility.[146]

3.8. *Lateral Confinement*

Further reduction of the system dimensionality, from 2D to 1D, 0D, or more general mesoscopic configurations, will naturally induce additional modulation of most of the properties discussed above. Among the most prominently-affected are the energy levels and dispersion relations, density of states[147] (particularly important, e.g., in lasers[148] and RTDs[149,150]), exciton binding energies,[151] optical matrix elements and selection rules,[152] absorption and gain lineshapes,[47] relaxation rates and the possible presence of a 'phonon bottleneck',[153,154] transport,[155] quantum transport,[38] plasma and plasmon properties,[38] and Auger lifetimes.[67,68,156] Although nanostructure science currently remains primarily a research field which has completed relatively few transitions to commercial products, the rich possibilities for additional wavefunction engineering make it an attractive area for future device concepts.

4. Applications of Wavefunction Engineering – Layered Quantum Heterostructures

In this section and the next, we illustrate application of the FEM to the determination of energy levels, dispersion relations, wavefunctions and optical properties for several specific wavefunction-engineered materials. Layered heterostructures having complex compositional modulation along the growth axis will be considered here, whereas Sec. 5 will treat the effects of lateral confinement on the properties of quantum wires and quantum wire superlattices.

4.1. *Type-II Laser*

Virtually all of the quantum well (QW) lasers developed thus far have employed well and barrier constituents having a Type-I band alignment, since the achievement of gain requires strong optical coupling between the conduction and valence band states. However, it is known that Type-II InAs-$Ga_{1-x}In_xSb$ superlattices can display a large interband absorption coefficient as long as the layers are thin enough to allow significant interpenetration of the electron and hole wavefunctions.[104,105] Clearly, the same Type-II structures should produce substantial gain if the electrical confinement is adequate and the nonradiative lifetimes are long enough to establish a population inversion. In fact, it will be argued below that mid-wave infrared (MWIR) lasers based on the InAs-GaSb-AlSb family of Type-II heterostructures are not only feasible, but they may be expected to exhibit some significant advantages over Type-I III-V systems which have been investigated previously for this application.

Despite reports of dramatic improvements in the performance of diode lasers emitting at IR wavelengths out to about 2.5 μm,[157] the development of longer-wavelength diodes with high output powers and non-cryogenic operating temperatures has proven to be much more challenging.[158–162] Fundamental limitations include inadequate electrical confinement due to the small conduction and/or valence band offsets in most of the heterostructure systems considered thus far[163] (which leads to escape from the active region at higher temperatures) and the increasing predominance of Auger recombination when the energy gap is lowered and the temperature raised.[164] The nonradiative decay in most III-V systems currently under investigation tends to be dominated by the so-called CHHS Auger process, in which the conduction-to-heavy-hole (CH) recombination is accompanied by a heavy-to-split-off-hole (HS) transition. In InAs-rich alloys such as InAsSb, InAsSbP and InGaAsSb, this process easily conserves both momentum and energy because the energy gap is nearly equal to the split-off gap Δ_0.[158,165] It will be seen below that the Type-II band alignment effectively eliminates CHHS Auger transitions by removing the resonance between E_g and Δ_0, and also introduces large conduction and valence band offsets which make it straightforward to maintain electrical confinement. Grein *et al.* theoretically discussed the minimization of Auger rates in Type-II InAs-$Ga_{1-x}In_xSb$ superlattices for MWIR laser applications[106] and Miles *et al.* have quite recently achieved lasing in an optically-pumped InAs-$Ga_{0.75}In_{0.25}Sb$ superlattice device emitting at 3.2 μm.[107]

The conduction, valence, and split-off band profiles for the structure studied by Miles *et al.* are illustrated in Fig. 2. Also shown are the corresponding energy levels and wavefunctions calculated using the 8-band FEM algorithm. Note first that even though the electron wavefunctions (solid curves) have their maxima in the InAs layers and the hole wavefunctions (dashed curves) are centered on the $Ga_{1-x}In_xSb$ layers, there is significant overlap because each (particularly ψ_n) penetrates into the adjacent layers. Hence, the optical matrix element ($P_{CV} \approx 6.5$ eV Å) is nearly as large as values typically obtained for Type-I heterostructures. We also find that the resonance between E_g (the separation of E1 and H1) and Δ_0 (the difference between H1 and S1) is completely removed by the Type-II band alignment, even though it is present in bulk InAs and GaSb, and is potentially an issue in $Ga_{1-x}In_xSb$. Furthermore, Grein *et al.* pointed out that for these particular layer thicknesses, the energy gap does not resonate with any intervalence transitions involving H1 near its maximum (it falls approximately halfway between H1–H2 and H1–H3).[106] Hence, all multi-hole Auger processes are energetically unfavorable, and CCCH events [in which the CH recombination is accompanied by an electron transition to a higher-energy conduction-band state (CC)] will probably dominate the nonradiative lifetime. Moreover, even the CCCH rate is suppressed by the small in-plane effective mass for holes near the band extremum ($\approx 0.047m_0$). Youngdale *et al.* have recently demonstrated experimentally that at 77 K, InAs-$Ga_{1-x}In_xSb$ superlattices can display Auger lifetimes which are two orders of magnitude longer than those in $Hg_{1-x}Cd_xTe$ alloys with the same energy gap, which is consistent with

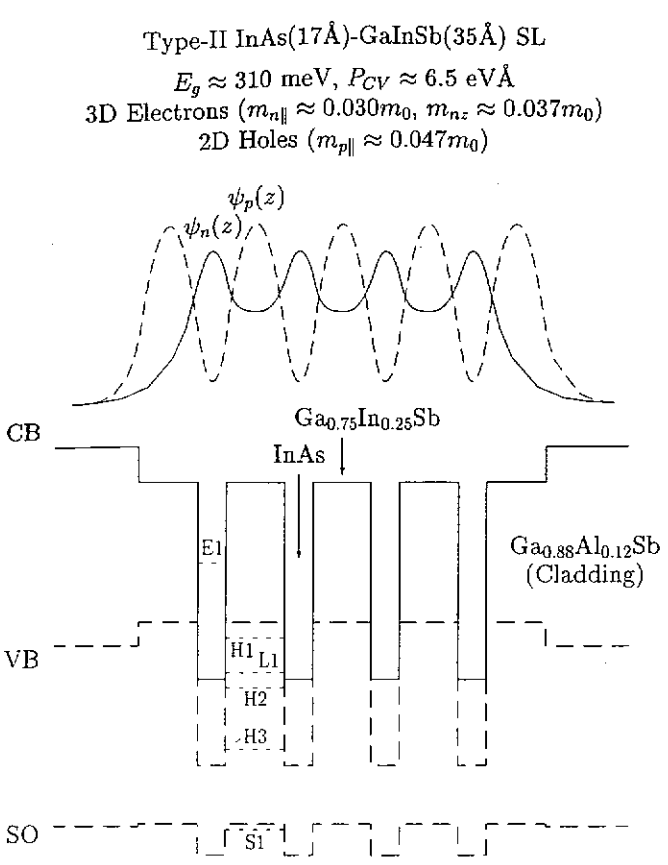

Fig. 2. Conduction, valence (heavy-hole), and split-off band profiles for the Type-II InAs-$Ga_{0.75}In_{0.25}Sb$ superlattice (clad by $Ga_{0.88}Al_{0.12}Sb$) studied experimentally by Miles *et al.*[107] Strain is included in the band alignments, assuming growth on a buffer layer whose lattice constant is matched to that of GaSb. Also shown are 8-band FEM results for the electron (solid) and hole (dashed) wavefunctions, along with energy minima for the various conduction and valence subbands. Note that the electrons have 3D dispersion with a nearly isotropic effective mass.

theoretical predictions.[166,167] The reduced density of states which accompanies the small in-plane hole mass also serves to decrease the threshold carrier density required to achieve population inversion[11] (Grein *et al.* have estimated a threshold concentration of $\approx 6 \times 10^{17}$ cm^{-3} at 300 K).

However, the structure shown in Fig. 2 is non-optimal in that the electron dispersion is effectively three-dimensional. It is well known that QW lasers (2D) usually significantly outperform[3,6,48] double heterostructure lasers (3D) once a given fabrication technology has matured, primarily because the more concentrated 2D density of states yields much higher gain per injected carrier at threshold.[168] While the holes in the Type-II superlattice shown in Fig. 2 have minimal dispersion along the growth axis (i.e. they are quasi-2D), the strong penetration of the electron wave-

functions into the thin $Ga_{1-x}In_xSb$ barriers leads to a nearly isotropic electron mass ($m_{nz}/m_{n_\parallel} \approx 1.2$). Thus, only a fraction of the hole population (which is distributed along the entire k_z axis) can participate in vertical stimulated emission processes since the thermal distribution of electrons is confined to small k_z.

The most straightforward approach to reducing the electron dimensionality is clearly to convert the superlattice into a multiple quantum well. Figure 3 illustrates the incorporation of additional $Ga_{1-x}Al_xSb$ layers, which serves as barriers for both electrons and holes, into each period of the structure. We emphasize here that while most multiband heterostructure $\mathbf{k} \cdot \mathbf{P}$ calculations can treat simple two-constituent superlattices such as that shown in Fig. 2, they are not easily adapted to more complicated geometries such as the three-constituent multiple quantum well shown in Fig. 3. In contrast, the multiband FEM algorithm readily yields band structures, wavefunctions, and optical properties for layered structures of arbitrary complexity.

It is not surprising that the FEM calculation for the structure in Fig. 3 yields 2D electron and hole dispersion relations (the electron miniband width is < 0.3 meV), since both wavefunctions fall nearly to zero in the barrier layers and the coupling between successive periods is quite weak. Unfortunately, a by-product of this particular asymmetric geometry is that the overlap is somewhat smaller than in the superlattice of Fig. 2. Electrons from a given InAs QW penetrate only about halfway into the adjacent $Ga_{1-x}In_xSb$ layer, and now there are no InAs electrons from the other side to overlap the other half of the hole wavefunction. As a consequence, the interband optical matrix element is somewhat smaller ($P_{CV} \approx 3.6$ eV Å) and the hole effective mass is more than twice as heavy ($0.11m_0$) as it was in the superlattice . The repulsion of the E1 and H1 dispersion relations is proportional to the overlap integral, hence both masses in this narrow-gap system are quite sensitive to the spatial distribution of the wavefunctions.[90]

Fortunately, we can largely eliminate these drawbacks by performing some further wavefunction engineering. Figure 4 illustrates the consequences of incorporating an additional InAs layer into each period, on the other side of the $Ga_{1-x}In_xSb$. The entire hole wavefunction for this four-constituent quantum structure is now seen to lie in a region of strong overlap with the electron wavefunction (as in the superlattice of Fig. 2). Hence, the in-plane hole mass is virtually equal to that in the superlattice ($0.048m_0$) and the interband optical matrix element is nearly as large ($P_{CV} \approx 5.3$ eV Å), yet the electrons and holes both have 2D dispersion due to the $Ga_{1-x}Al_xSb$ barriers. Because of the coupled-well nature of the conduction band profile, the electron states split into symmetric (E1S) and anti-symmetric (E1A) levels. However, only E1S will be populated at the laser operating temperatures of interest (≤ 300 K), since the energy separation between the two is 125 meV. The structure in Fig. 4 is illustrated with AlSb cladding layers, which provide excellent optical confinement and provide a nearly exact lattice match to the active quantum-well region (minor adjustments in the active-region lattice constant are easily made through composition variations in the $Ga_{1-x}In_xSb$ and $Ga_{1-x}Al_xSb$ layers). Note also that we still expect intervalence Auger processes to be weak because the energy

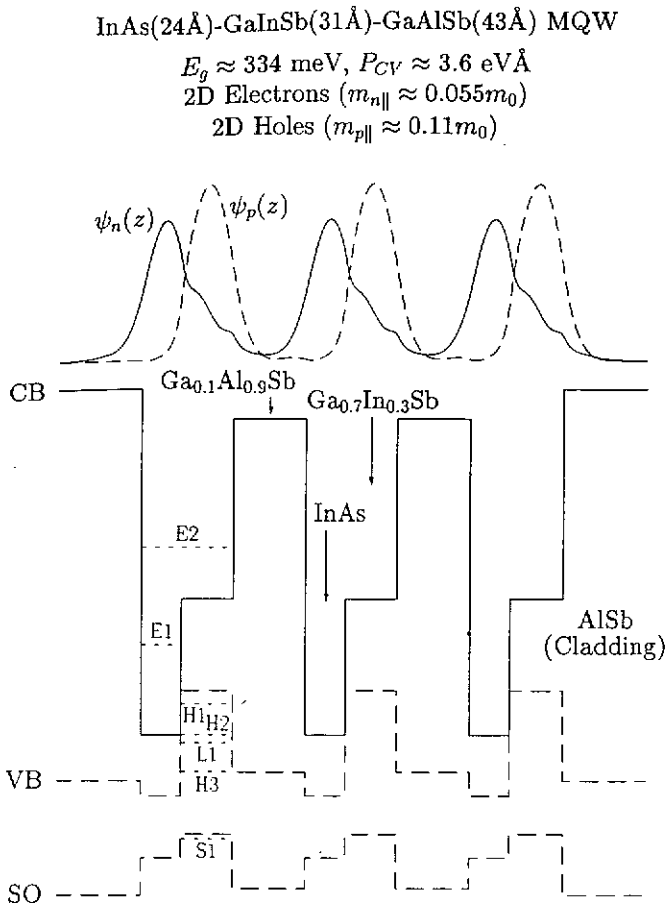

Fig. 3. Band profiles and FEM wavefunctions and energy levels for a Type-II InAs-$Ga_{0.70}In_{0.30}Sb$-$Ga_{0.1}Al_{0.9}Sb$ three-constituent multiple quantum well which is lattice-matched to the AlSb cladding layers. Although the electrons now have 2D dispersion due to the $Ga_{0.1}Al_{0.9}Sb$ barrier layers, the wavefunction overlap is reduced from that in Fig. 1.

gap between E1S and H1 again falls halfway between the gaps for H1-L1 and H1-H3 (and is of course far out of resonance with the split-off gap). We finally point out that electrical confinement ceases to be an issue in InAs-GaSb-AlSb-family Type-II laser structures because the AlSb or $Ga_{1-x}Al_xSb$ cladding layers provide large offsets for both the conduction and valence bands, in addition to having low refractive indices for optical confinement ($\Delta n > 0.2$). As in the intersubband quantum cascade laser,[62] the emission photon energy for the present Type-II laser is controlled almost entirely by quantum confinement rather than by the energy gaps of the constituents, and can in principle be tuned from zero to more than 1 eV.

We emphasize also that the quantum structural design considered here has quite literally taken the form of wavefunction engineering.

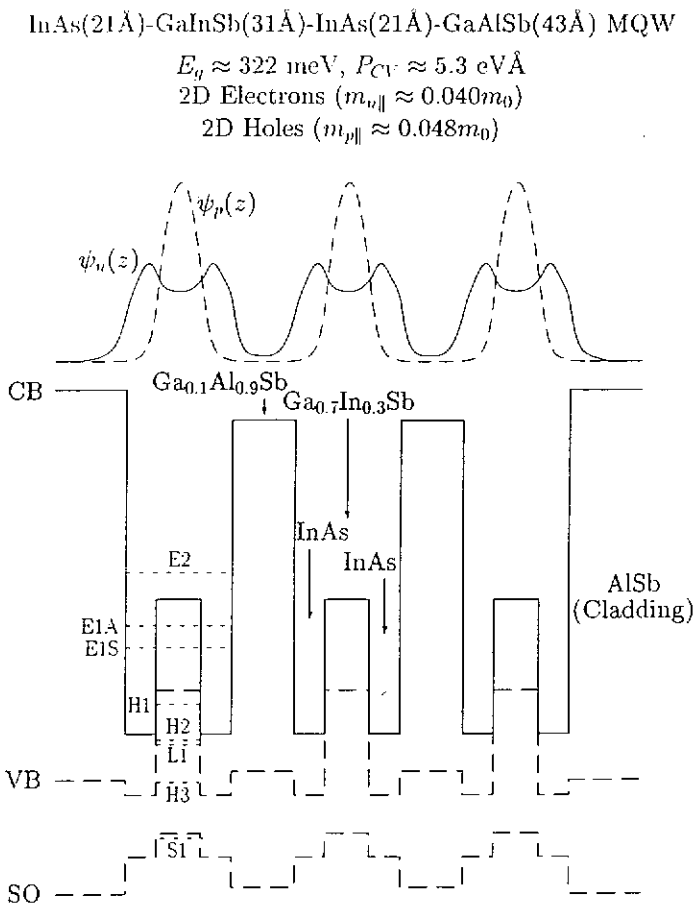

Fig. 4. Band profiles and FEM wavefunctions and energy levels for a Type-II InAs-$Ga_{0.70}In_{0.30}Sb$-InAs-$Ga_{0.1}Al_{0.9}Sb$ four-constituent multiple quantum well which is lattice-matched to the AlSb cladding layers. This structure combines both the 2D electron dispersion of Fig. 2 and the large wavefunction overlap of Fig. 1.

4.2. *Intersubband Second-Harmonic Generation with Momentum-Space Reservoir*

The multiband FEM also provides an attractive means of calculating the intersubband optical, nonlinear optical and electro-optical properties for complex, multilayered structures. For example, intersubband-based second-harmonic generation has attracted a great deal of attention in recent years,[169,117,170] motivated in part by the need for coherent, high-intensity optical pulses at wavelengths which are not readily available from conventional sources. Intersubband $\chi^{(2)}_{2\omega}$ processes are especially attractive because they have narrow bandwidths and large oscillator strengths,

and because they are particularly adaptable to wavefunction engineering through modification of the structural design. The use of stepped wells and asymmetric double wells[171] to break the usual wavefunction symmetries has led to the observation of second-order nonlinearities which are more than 1000 times larger than those attainable from bulk GaAs.[117]

However, it has recently been shown[12] that saturation imposes a fundamental limit to the maximum second-harmonic conversion efficiencies that one might ultimately hope to attain from optimized quantum-well devices. A key figure of merit is the conversion efficiency at the onset of saturation, which may be reduced to the approximate form

$$\eta_s = \frac{\tau_\Gamma}{\tau_{12}} R_{12} , \tag{4}$$

where

$$R_{12} \equiv \left(\frac{z_{23} z_{13}}{z_{12}^2 + 2 z_{13}^2} \right)^2 , \tag{5}$$

z_{ij} is the dipole matrix element for transitions between subbands i and j, $\tau_\Gamma \equiv 1/\Gamma$ (Γ is the intersubband-transition broadening parameter), and τ_{12} is the intersubband relaxation time. Equation (4) is seen to contain no explicit dependences on most of the usual design parameters, including angle of incidence, fill factor, doping density, refractive index, or resonance frequency. Furthermore, one has limited flexibility in optimizing the parameters which do appear,[12] i.e. the broadening is ultimately limited by nonparabolicity, the intersubband relaxation must rely on intrinsic phonon processes, and interrelationships between the dipole matrix elements limit the maximum magnitude of R_{12}. Numerical solutions to the coupled propagation equations for the pump and second-harmonic beams, accounting for beam depletion and saturation of both the absorption coefficient and $\chi^{(2)}$, confirm that the maximum conversion efficiency $\eta_{\max}$ attainable from "conventional" semiconductor intersubband devices is fundamentally limited to $\approx 10\,\%$.[12]

This limit is governed primarily by saturation, which occurs when the ground-state level becomes depopulated by intersubband absorption. The relatively-rapid onset of saturation is in turn directly attributable to the inefficiency of the intersubband relaxation process, which unavoidably yields τ_{12} exceeding ≈ 1 ps for isolated 3-level subband systems with favorable $\chi^{(2)}$. While it would be difficult to increase the relaxation rate for the usual phonon mechanism, there are nonetheless artificial means by which we can attempt to reduce the effective τ_{12}. One proposal has been to use band structure engineering to place the subband system in contact with a high-density reservoir of optically-inactive states, whose function is to provide elastic scattering into and out of the active subband states.[12] Thus, whenever an absorption process excites an electron out of the $i = 1$ subband, that carrier is replaced by an electron from the reservoir on a time scale much faster than τ_{12}. Similarly, electrons excited to the $i = 2$ and $i = 3$ levels quickly scatter into reservoir states at those higher energies. Furthermore, saturation does not occur until the entire lower-energy portion of the reservoir is emptied.

While spatial reservoirs could be placed on each side of the active asymmetric region, the fill factor for such a device would be rather small due to the large ratio of reservoir-region thickness to active-region thickness. Probably a more attractive alternative is to introduce a reservoir in momentum space.[12] Figure 5 shows the band profiles, energy levels, and wavefunctions for a GaSb-AlSb-$Ga_{1-x}In_xSb$-$Ga_{1-x}In_xSb$-AlSb ADQW structure which embodies this concept. The key feature is that the Γ-point conduction-band minimum (labeled Γ_1 in the figure) is at only slightly lower energy than the ground state for the four degenerate L-point valleys (labeled L1). While the design of the asymmetric wells is such that the Γ-point subbands are doubly resonant ($E_{12}^{\Gamma} \approx E_{23}^{\Gamma} \approx \hbar\omega \approx 122$ meV in this example) and have favorable dipole matrix elements ($R_{12} \approx 0.33$, the L valleys are ideally suited to serving as a reservoir. Their 2D density of states is roughly 30 times that of the Γ valley, yet they can be made optically inactive by detuning the intersubband spacings to values far from resonance with the pump and second-harmonic photon energies. Measured L-Γ intervalley relaxation times tend to be extremely short, falling in the 50–200 fs range,[172,173] and it has been predicted theoretically that the Γ-valley intersubband relaxation rate can be increased by shunting through L-valley states.[174,175]

Since the $\mathbf{k} \cdot \mathbf{P}$ method is inapplicable to systems with ellipsoidal constant-energy surfaces, we derived L-valley subband energies and transition matrix elements from a 1-band version of the finite-element algorithm (the second-harmonic generation will be relatively insensitive to the details of the L-valley properties). That calculation employed L-point quantization masses for each constituent material [$m_z = 3m_t m_l/(m_t + 2m_l)$ for (100) growth, where m_t and m_l are the L-valley transverse and longitudinal masses]. Because the two lowest L-point subbands have energies only slightly above the Γ-point minimum (L_1 is 38 meV above Γ_1 and L_2 is 17 meV above L_1) and their densities of states are much higher, roughly 90 % of the electrons reside in the L-point reservoir in thermal equilibrium at 300 K. Nonetheless, since the intersubband splittings are far out of resonance with both $\hbar\omega$ and $2\hbar\omega$ (e.g., $E_{13}^{L} \approx 75$ meV and $E_{24}^{L} \approx 170$ meV), the pump-beam and second-harmonic-beam absorption coefficients due to L-point transitions are calculated to be less than 10 % of those for resonant Γ-valley processes. A broadening parameter of 10 meV was assumed for both Γ-valley and L-valley interactions.

Clearly, there are a number of adjustable design parameters which may be tuned in optimizing the structure shown in Fig. 5. For example, the splitting $E_{13}^{\Gamma} \approx 2\hbar\omega$ is determined primarily by the thickness of the $Ga_{1-x}In_xSb$ QW on the right, whereas the Γ_2 energy (and hence E_{12}^{Γ} and E_{23}^{Γ}) is governed by the thickness of the GaSb well on the left. Similarly, the Γ_1–L_1 energy difference (which is the key to the relative populations of the active system vs the reservoir system) is largely controlled by the alloy composition x of the $Ga_{1-x}In_xSb$ well (note from the figure that the minima for the Γ and L band profiles move in opposite directions when x is increased from 0 to 0.11). Finally, the matrix elements z_{12} and z_{23} are governed primarily by the

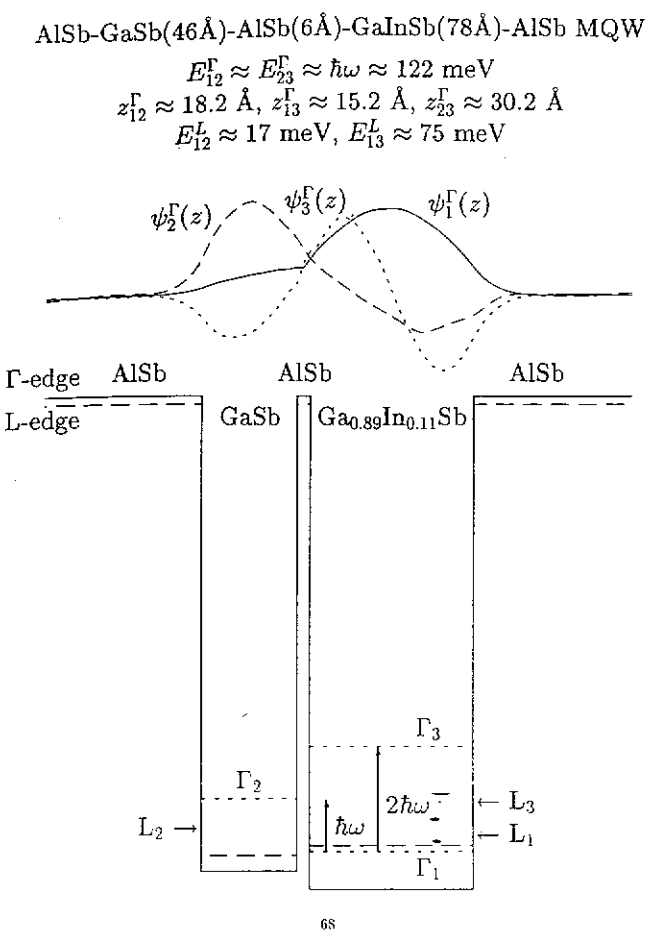

Fig. 5. Γ-point and L-point band profiles and FEM wavefunctions and energy levels for a GaSb-AlSb-$Ga_{0.89}In_{0.11}Sb$-AlSb asymmetric double quantum well. In second-harmonic generation, the Γ-valley states will be optically active while the L-valley sates function as a momentum-space reservoir.

thickness of the barrier separating the two wells. This barrier could be somewhat thicker if $Ga_{1-y}Al_ySb$ with relatively low y is employed instead of AlSb.

While the Γ_1 wavefunction (solid curve) is seen to be concentrated mostly in the right well and that for Γ_2 (dashed curve) mostly in the left, it is the leakage of each wavefunction into its opposite well which fixes the magnitude of the optical coupling. Naturally, z_{12} will become vanishingly small if the AlSb barrier is made too thick. Although ψ_3^{Γ} (dotted curve) is somewhat more evenly distributed between the two wells, it nonetheless has greater magnitude in the right well (it is well known that z_{13} and hence $\chi^{(2)}$ vanish for any system having inversion symmetry). Note also that while z_{12} and z_{13} should ideally be of comparable magnitude so as to equalize the depletion of the pump and second-harmonic beams, it is nearly always advantageous

to maximize z_{23}.[176] The present structure has a large value ($z_{23} \approx 30.2$ Å) because ψ_2^Γ and ψ_3^Γ have relatively large overlap.

Since the L-valley shunt is expected to result in faster intersubband relaxation, we estimate that saturation will be delayed nearly until the plasma-breakdown point. With the resulting enhancement of η_s, we obtain a maximum conversion efficiency of $\approx 20\,\%$, which is a factor of 2 larger than that calculated for the same Γ-valley subband system in the absence of the reservoir. Furthermore, the optimum sample thickness is decreased considerably, to $\approx 7\ \mu$m, which becomes quite feasible for growth by MBE.

We thus predict that the incorporation of a momentum-space reservoir will significantly enhance the performance of intersubband second-harmonic-generation devices. While we have illustrated this concept with the example of GaSb-AlSb-$Ga_{1-x}In_xSb$-AlSb ADQWs, similar behavior is expected for $Al_xGa_{1-x}As$-based structures with L-point or X-point reservoirs.

4.3. *Novel Localization of Carriers in Layered Heterostructures*

Recent investigations employing FEM-based computational methods have helped to clarify the concept of wavefunction localization, and have provided a visualization of that localization. Here we discuss several novel and perhaps counter-intuitive findings from those studies: that quasibound states in compositionally asymmetric quantum wells strongly participate in optical transitions, which may be exploited in nonlinear mechanisms;[177] that the asymmetric vacuum confinement in a heterostructure cap layer leads to surface-bound states;[178] and that states with energies above the barrier in a quantum well or superlattice localize in the barrier layers.

4.3.1. *Quasibound States*

In quantum wells constructed using semiconductor heterostructures, well-defined localization of the wavefunctions below a barrier leads to increased excitonic binding, which has been exploited in electro-optic devices based on the electric field modulation of the excitonic transition. While it is well known that symmetric quantum wells must have at least one bound state, this theorem does not hold in asymmetric QWs. In fact, the asymmetric well can 'push out' the wavefunctions such that there are no bound states at all in the well.

This feature of asymmetric wells has recently been verified experimentally.[179] Single asymmetric QWs made of GaAs were grown with unequal Al content in the $Al_xGa_{1-x}As$ barrier layers. This compositional asymmetry, shown in Fig. 6, yields barriers on the two sides with different heights V_1 and V_3, say, and leads to a continuum of energy levels with half-confined states in the energy range $V_1 \leq E \leq V_3$. In this continuum, there exist special states which are 'quasibound' in the sense that they have increased, or resonant, occupancy in the well layer. Such states are remnants of the bound states which would be present in the symmetric well if both barriers had height V_3, and closely correspond to such states in energy

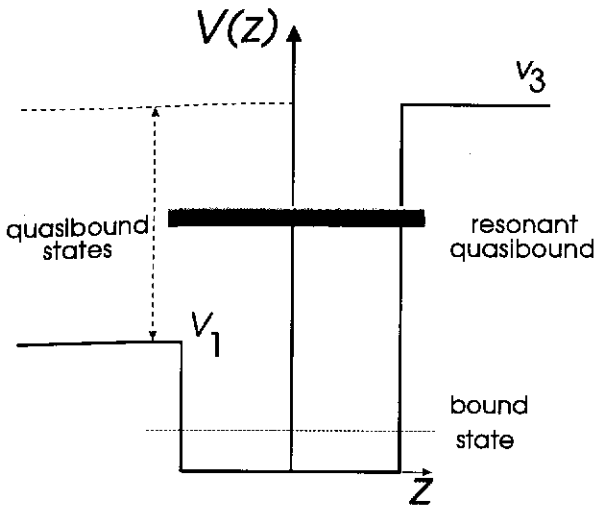

Fig. 6. The band edge profile for a compositionally asymmetric quantum well.

since the boundary condition on one side of the well is still the same as in the symmetric well; they occur at energies where an integral multiple of half the particle wavelength nearly fits within the well layer, and are associated with a maximum in the occupancy of the QW layer as defined by the probability density of the carrier.

In the FEM, we employ 3 elements per layer, and use standard boundary conditions for the vanishing of the wavefunctions at large distances into the barrier region on the right. Typically, 300–500 Å of the barrier layer is adequate to model the wavefunction decay. On the quasi-bound side, we rearrange the values of the wavefunction and its derivative at the first node to correspond to the incoming amplitude and the unknown reflected amplitude.[28] The set of simultaneous equations for the nodal values of the wavefunction that are obtained from the global matrix, after applying the boundary condition on the incoming wave, are solved using LINPACK routines. A reconstruction, by using the same interpolation polynomials that were used to represent the wavefunction together with the solutions for the nodal values, allows us to obtain the wavefunctions everywhere. With these wavefunctions we can obtain the occupancy of the carriers within the quantum well layer. The occupancy plots for the conduction and valence bands are shown for a 100 Å asymmetric well in Fig. 7. The peaks in the occupancy plot indicate the presence of two quasibound heavy hole states, along with one quasibound state each for the conduction and light hole bands.

These quasibound states are found to participate strongly in optical transitions.[177] Estimates for the second-order nonlinearity using virtual transitions involving such states are comparable to those with fully bound states in asymmetric 4-layer QW structures with a step potential within the well region. Interband transitions in a series of asymmetric wells of different width were observed experimentally in piezomodulated reflectivity and fitted theoretically using the FEM.

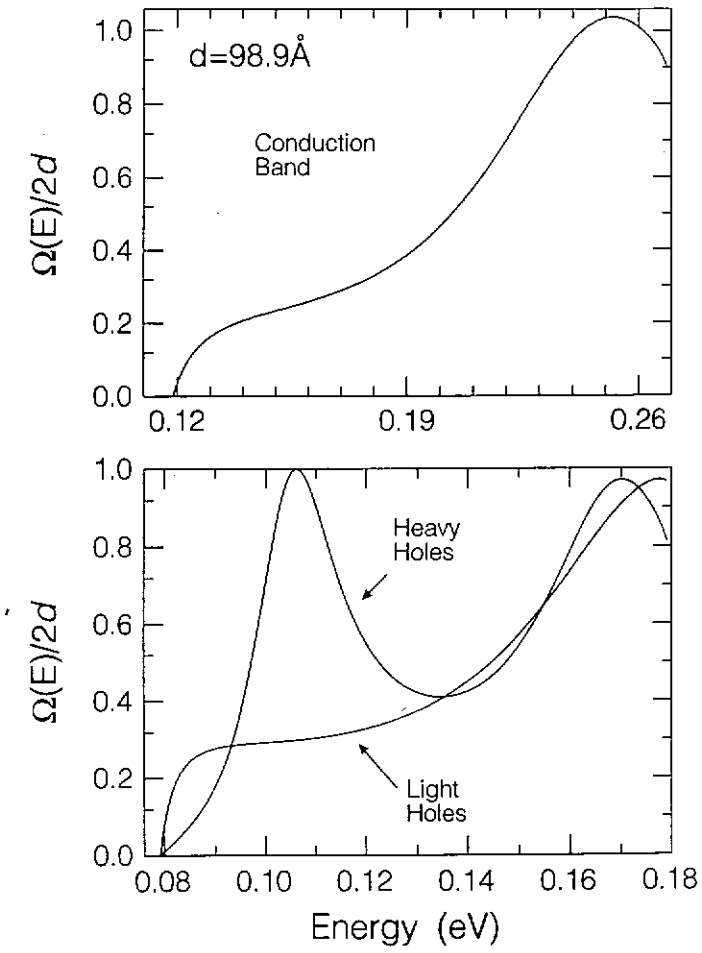

Fig. 7. Occupancy of the quasibound conduction, hh, and lh states in the well region versus energy for a 100 Å compositionally asymmetric quantum well (From Ref. 179).

As an illustration of wavefunction engineering, calculations were performed to predict the well width for which no bound states should be present in the asymmetric quantum well. Samples were MBE-grown to specifications, and experiments indeed verified the absence of a bound lh state in a 20 Å asymmetric well.[179]

4.3.2. *Surface Quantum Well Confinement and Above-Barrier Localization*

Here we consider the confinement of electronic states in the simplest of semiconductor heterostructures, *viz.*, a quantum well (QW) bounded by vacuum on one side and a quantum barrier on the other.[178] Figure 8 illustrates the conduction-band profile for vacuum/GaAs/$Al_xGa_{1-x}As$ QWs which were recently fabricated by MBE. The piezo- and electro-modulated reflectivity spectra for these surface quantum wells were found to exhibit an extraordinary number (as many as 69 in one sample) of

clear signatures for optical transitions between the quantum-confined states in the valence band to those in the conduction band.[178] Although the concept of the surface QWs and the occurrence of states confined in them had been contemplated in earlier reports,[180–183] the sensitivity and power of modulation spectroscopy[184,185] as exploited in Ref. 178 displayed their electronic structure with unprecedented clarity and richness. By comparison, photoluminescence experiments allow one to observe only optical transitions between the lowest quantum confined states in the conduction and valence bands.

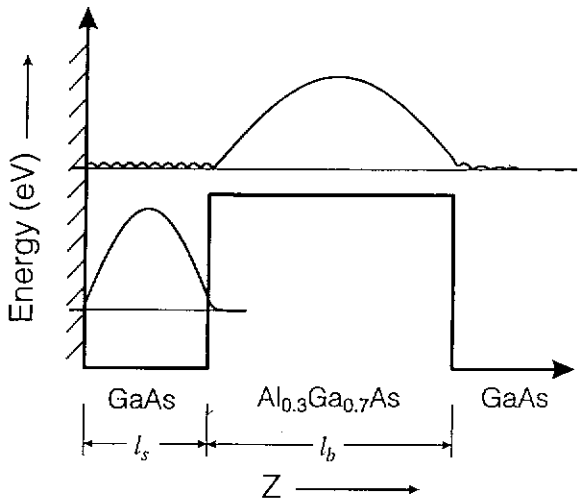

Fig. 8. Potential energy diagram for a surface quantum well (QW) and single quantum barrier (SQB). The surface QW states are localized within the GaAs layer, while the above-barrier resonant states are localized within the $Al_{0.3}Ga_{0.7}As$ layers. The probability density of the wavefunctions for the first surface QW and SQB energy levels are also shown (From Ref. 178).

Another unique property of a structure such as the one illustrated in Fig. 8 is the presence of states localized above the single quantum barrier (SQB) in the $Al_xGa_{1-x}As$ layer. These states are actually part of the continuum and not bound states as in a single QW. However, their wavefunctions and energy-level dependence on barrier thickness are remarkably similar to those for their bound counterparts in single quantum wells. In Fig. 9, we show the FEM wavefunctions for a 400 Å surface quantum well with a 800 Å barrier. Both the below-barrier well-confined states and the above-barrier localized states are displayed here. The calculations were carried out with a very high confining barrier to the right of the structure in order to generate real wavefunctions. The barrier-localized states are found to occur at energies where an integral multiple of half the particle wavelength nearly fits within the barrier layer.[177,179] In the modulated-reflectivity spectra of SQB's a large number of above barrier states are observed in these structures. Figure 10 shows the electromodulated reflectivity spectra of two surface QWs with $l_w = 400$ Å and $l_w = 150$ Å, both samples having $l_b = 800$ Å. The features between the E_0 and

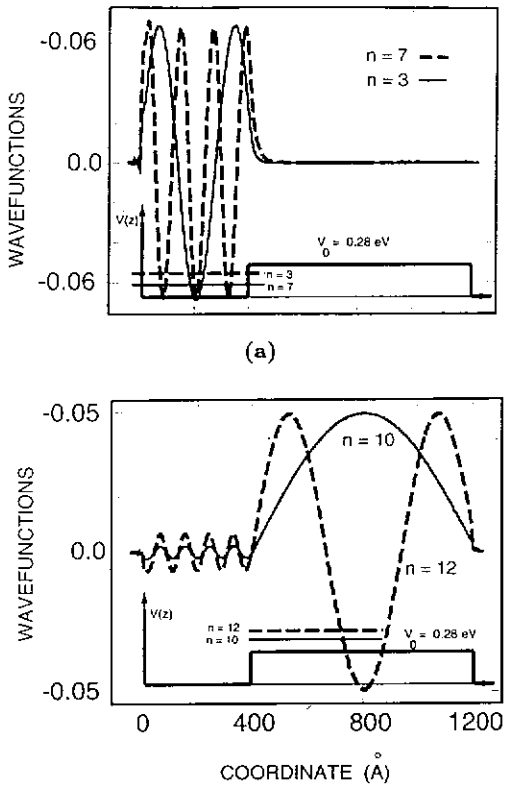

Fig. 9. The wavefunctions of (a) some states localized in the surface quantum well, and (b) above-barrier localized states in the vacuum/GaAs/AlGaAs/GaAs structure.

$E_0 + \Delta_0$ signatures of GaAs may be attributed to the surface QW states. As expected, noticeably fewer transitions below the barrier energy are observed in the sample with $l_w = 150$ Å as compared to that with $l_w = 400$ Å. These features arise from transitions between the quantum heavy-hole valence band and well-confined conduction band states. They are labeled $nm(H, L)$, where H and L denote heavy-hole and light-hole, respectively, and the transitions are between the nth conduction and mth valence band states. For such wide surface QWs the light-hole transitions are more difficult to observe in electromodulated reflectivity because they are three times weaker than the heavy-hole transitions, although for $l_w = 150$ Å Fig. 10 shows evidence for the first light-hole transition 11L. Several transitions between surface QW states in the conduction band and those in the spin-orbit split-off band are also observed between $E_0 + \Delta_0$ of GaAs and E_0 of $Al_{0.3}Ga_{0.7}As$. In Fig. 10, transitions involving SQB states in the 800 Å thick $Al_{0.3}Ga_{0.7}As$ layer appear in both samples above the E_0 feature of the barrier layer. The optical transitions between such states are very strong and, since both samples have $l_b = 800$ Å, the spectra are strikingly

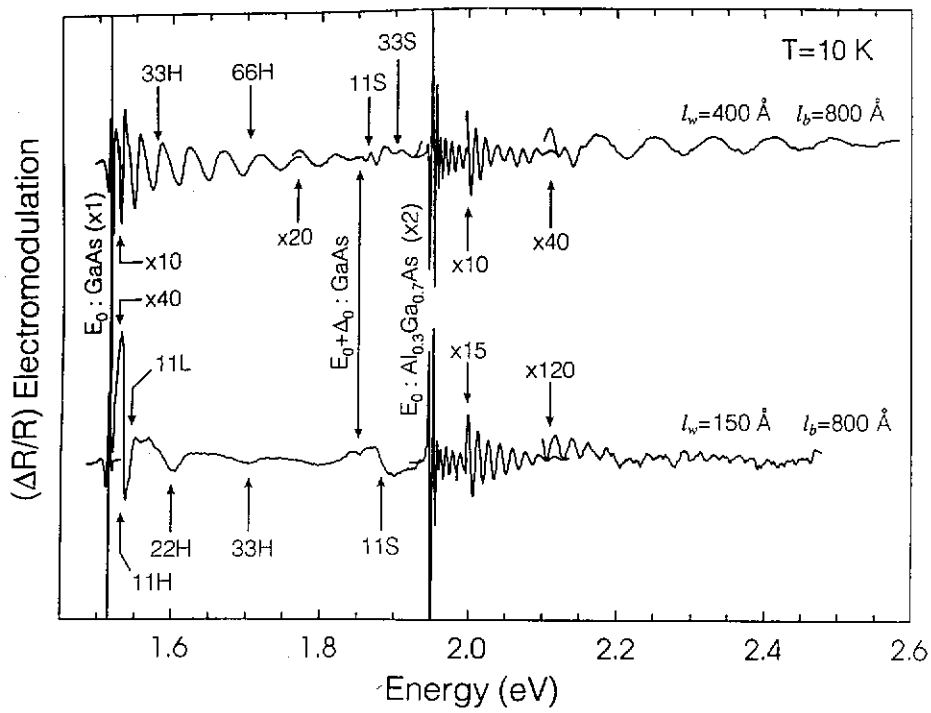

Fig. 10. The experimental electromodulated reflectivity spectra of two surface QWs. Both spectra are for samples with l_b = 800 Å and show the above-barrier quasi-bound state resonances. The top spectrum is for l_w = 400 Å and the bottom is for l_w = 150 Å. The spectra were recorded at T = 10 K (From Ref. 178).

similar above 1.943 eV. This shows that these states are not strongly influenced by the thickness of the GaAs cap-layer, and confirms the interpretation[119,186–188] that such states are strongly localized within the barrier layer. Above 2.1 eV a change in the spacing of the transition energies can be seen in the sample with l_w = 400 Å. In the sample with l_w = 150 Å this change does not occur; instead, the intensity of the SQB transitions confined predominantly in the barrier layer approaches zero. Given that the only difference between these two samples is the change in l_w, we attribute the extra transitions above 2.1 eV for l_w = 400 Å to a continuation of the surface QW states above the barrier energy. The surface QW states in the sample with l_w = 150 Å are not observed to continue above the barrier energy, presumably because their intensity decreases more rapidly with increasing energy than in the l_w = 400 Å sample. Furthermore, SQB transitions in a sample with a wide surface QW (l_w = 1650 Å) were not readily observed because localization of the wavefunctions in the barrier was marred by its thinness (l_b = 120 Å). While the spectra given in Fig. 10 were taken at 10 K, the surface QW states were also easily observable in the modulated-reflectivity spectra for room

temperature, and piezomodulated reflectivity gave results which were quite similar to the electromodulated spectra.

In order to establish the origin of the optical transitions as originating from electronic states confined in the surface QW as opposed to those localized above the SQB, we applied the finite element method to solve Schrödinger's equation for the envelope functions in the heterostructure with a one-band model as well as with an empirical two-band model.[32] The continuity of the envelope function and its 'mass-derivative' was applied at the layer interfaces. The confining surface potential was taken to be much larger than the E_0 of GaAs, while the material parameters in the layers were taken from Ref. 189. For each state, a calculation was made to determine where the wavefunction was localized. The overlap integral was calculated between conduction and valence states for all possible transitions; the ones with the largest overlap values were selected as the theoretical transition energies. After applying a correction for the excitonic binding energy as estimated by Nelson *et al.*,[190] the derived theoretical values were found to be in agreement with those obtained from the experiment. The strong localization of SQB states in the barrier layer was verified to be insensitive to the thickness of the GaAs cap-layer. The density of states (DOS) were also calculated for the conduction and valence bands.[191] The peaks in the DOS were used to re-confirm the above transition energies. Consistent with the experiment, the envelope of the peaks in the DOS showed a fall-off with energy, as well as oscillations whose maxima corresponded well with the observed transitions at high energies above 2 eV.

Figure 11 displays the optical transition energies between the surface QW states and the above-barrier SQB states as a function of transition index n for three different samples. The data points obtained from the electromodulation spectra are fitted with curves from the theoretical calculation, with l_w and l_b as the only adjustable parameters. The E_0 gap of the barriers was determined from the experimental data in Fig. 10 to be 1.943 eV in all samples. The departure of the theoretical calculation and experiment at higher energies is due to the neglect of non-parabolicity in the DOS calculation. Again, the data points for the SQB states from Fig. 10 virtually overlap because both samples have *identical* barrier heights and widths. We have plotted the $l_w = 400$ Å surface QW states in Fig. 11 above 2.1 eV as a continuation of the surface QW states from below the barrier energy. Strictly speaking, the DOS was calculated for the combined surface QW and SQB structure. However, the quantum numbers in Fig. 11 were still assigned above 2.1 eV as though the states in the $l_w = 400$ Å sample are a continuation of the surface QW states. Between the E_0 of $Al_{0.3}Ga_{0.7}As$ and 2.1 eV in Fig. 10, the theory correctly predicts the SQB states to be highly localized in the barrier layer, the corresponding transitions being more intense than the surface QW transitions below 2.1 eV.

The structure in Fig. 8 is not the only type which should exhibit surface QW states. Virtually any structure with a GaAs cap-layer of 200 Å or more can, in principle, exhibit observable surface QW states which may co-exist with features associated with the underlying heterostructure.