Synthesis for semiconductor device design

Jason Thalken and Stephan Haas[a)]
Department of Physics and Astronomy, University of Southern California, 920 Bloom Walk, Seaver Science Center (SSC) 211A, Los Angeles, California 90089-0484

A. F. J. Levi
Department of Physics and Astronomy, University of Southern California, 920 Bloom Walk, Seaver Science Center (SSC) 211A, Los Angeles, California 90089-0484 and Department of Electrical Engineering, University of Southern California, 3620 South Vermont Avenue, Kaprilian Hall (KAP) 132, Los Angeles, California 90089-2533

(Received 28 February 2005; accepted 11 July 2005; published online 22 August 2005)

Synthesis of semiconductor device design requires access to realistic physical models and adaptive algorithms. To demonstrate that such synthesis is feasible we design elements of a quantum-confined Stark-effect modulator. Optimization with respect to a target function is achieved using a genetic algorithm. It is then shown how automated searches of configuration space may be performed without the need to input a specific target function. © *2005 American Institute of Physics*. [DOI: 10.1063/1.2014942]

INTRODUCTION

The design of both photonic and electronic devices has evolved over many years in an essentially *ad hoc* way. There is, for example, a long history of laser diode and photodiode designs that incrementally builds on previous designs and concepts.

We believe that now an opportunity exists to harness a combination of modern computer power, adaptive algorithms, and realistic physical models, to seek robust, manufacturable designs for components that meet previously unobtainable system specifications. In such a scheme, device design is synthesized. In its simplest form, semiconductor device synthesis software solves the inverse problem by identifying the best configuration to achieve a user-defined specification or target function. Typically, this involves searching for a (broken-symmetry) spatial configuration of a semiconductor structure that produces a desired target function response. Using such an approach it is possible to break the cycle of piecemeal design methods and discover functionalities for both photonic and electronic semiconductor devices. The purpose of this paper is to demonstrate that such device synthesis is, in fact, feasible. We aim to show this by synthesizing key parts of the design of a quantum-confined Stark-effect (QCSE) optical modulator.[1–3]

In the following we present the physical model used in simulations, the method used to perform optimization, and results for different target functions. We then explore the idea of automated, intelligent machine-based searches of configuration space for device functionality. Finally, we present the conclusions of our study.

PHYSICAL MODEL

A QCSE modulator makes use of the facts that optical absorption at near band-gap photon energies in direct band-gap semiconductor quantum wells is dominated by the presence of excitons and that application of an electric field can modify the absorption spectrum. Optical absorption strength in quantum wells can be greater than in bulk material, even in the presence of large external electric fields applied perpendicular to the plane of the quantum wells. The physics underlying exciton formation and mechanisms to control the QCSE are therefore of great interest for the design of modulators and detectors.[4]

The absorption of photons in a direct band-gap semiconductor involves the creation of electron-hole pairs. For uncorrelated electron-hole pairs, absorption occurs at photon energies greater than the band-gap energy of the material. However, the Coulomb interaction between an electron and hole can form a correlated exciton bound state with an absorption energy less than the band-gap energy. The binding energy of this exciton may be increased by confining both the electron and the hole within a two-dimensional quantum well structure.[5] The absorption spectrum that emerges from this quantum well potential profile shows a strong peak just below the band-gap energy. This fact has been exploited for use in many modern optoelectronic devices such as modulators and detectors.[1–3]

When a uniform electric field F is applied in the z direction, perpendicular to the plane of a quantum well, the band-edge potential profile is modified as illustrated in Fig. 1. In the figure, the conduction and valence band-edge profiles are shown as a function of position z along with the lowest-energy single-particle probability distribution for electrons, $|\psi_e|^2$, and for holes, $|\psi_h|^2$. With increasing applied electric field in the z direction, electron and hole tunneling results in a decrease of the exciton absorption peak energy. Shifts in the distribution of the electron and hole wave functions lead to a reduction in spatial overlap which in turn reduces the dipole matrix element, diminishing the peak absorption strength.

Specializing to the $Al_xGa_{1-x}As$ material system, the conduction- and valence-band potential profiles are calculated using the band-gap energy

[a)]Electronic mail: shaas@usc.edu

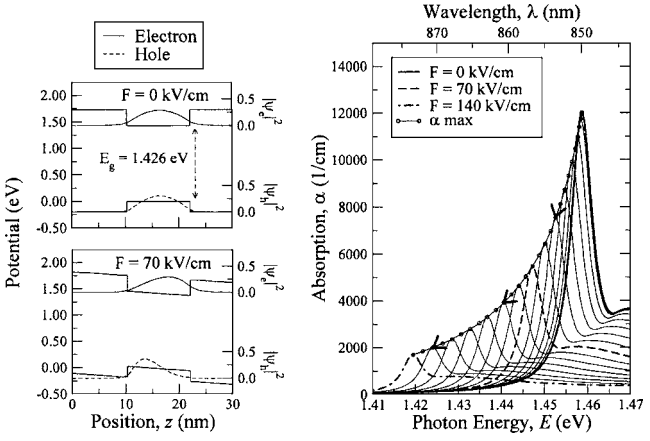

FIG. 1. Changes in excitonic absorption spectrum for a symmetric quantum well with increasing electric field. On the left, electron and hole ground-state probabilities are shown against the conduction and valence band-edge potential profiles at external field strengths of 0 and 70 kV/cm. Decreasing electron-hole overlap and tunneling into the energy gap result in loss of absorption strength and lowered absorption peak energy (shown right) characteristic of the QCSE. Arrows indicate the direction of increasing electric field. Absorption spectra are shown in increments of 10 kV/cm.

$$E_g = (1.426 + 1.247x)\ \text{eV} \tag{1}$$

and a band-gap offset ratio between conduction and valence band of 67/33.[6] The effective electron mass in the conduction band is

$$m_e = (0.067 + 0.0835x)m_0 \tag{2}$$

and the effective hole mass in the valence band is

$$m_h = (0.34 + 0.42x)m_0. \tag{3}$$

To minimize computation time, the discretization of the quantum well band-edge profile into atomic monolayers is incorporated by using a nearest-neighbor tight-binding Hamiltonian to find the single-particle eigenfunctions in the z direction. This Hamiltonian for the uncorrelated electrons and holes is given by

$$H = H_e + H_h = -t_e \sum_{\langle i,j \rangle} (c_{e,i}^{+} c_{e,j} + c_{e,j}^{+} c_{e,i}) + \sum_i \varepsilon_{e,i} + t_h \sum_{\langle i,j \rangle} (c_{h,i}^{+} c_{h,j} + c_{h,j}^{+} c_{h,i}) + \sum_i \varepsilon_{h,i}, \tag{4}$$

where $t_e(t_h)$ represents the electron (hole) hopping energy, $\varepsilon_e(\varepsilon_h)$ denotes the on-site electron (hole) energy, $c_e^{+}(c_h^{+})$ and $c_e(c_h)$ are the electron (hole) creation and annihilation operators, and $\langle i,j \rangle$ indicates a sum over nearest neighbors only. The single-particle eigenenergies, the electron eigenfunction $\phi_e(z_e)$, and the hole eigenfunction $\phi_h(z_h)$ reproduce previously published results.[7] Following Refs. 4 and 7, a variational ansatz[8] is used to obtain the separable $1s$ exciton wave function

$$\psi^{\text{ex}}(z_e, z_h, \rho) = \phi_e(z_e)\phi_h(z_h)\phi(\rho) \tag{5}$$

in which the in-plane exciton wave function takes the form

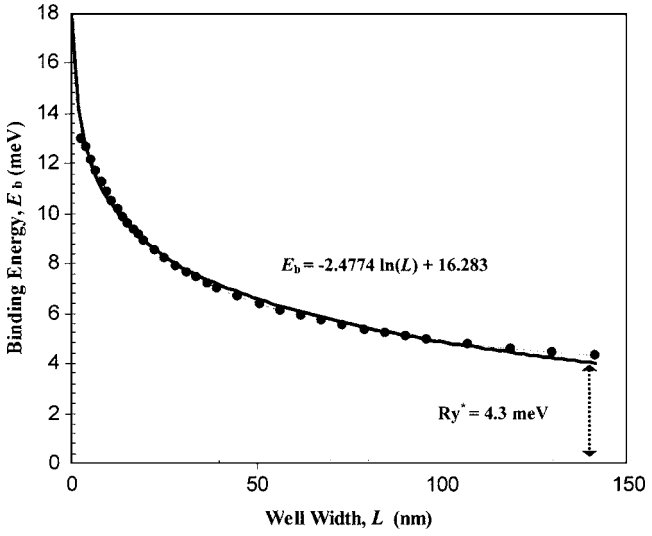

FIG. 2. Exciton binding energy calculated using the tight-binding Hamiltonian as a function of quantum well width for $Al_{0.3}Ga_{0.7}As/GaAs$ quantum well structures. The grey line represents a logarithmic fit to data which approximates the function away from the limits of large and small L.

$$\phi(\rho) = \sqrt{\frac{2}{\pi\lambda^2}} e^{-\rho/\lambda}, \tag{6}$$

where the in-plane coordinate is $\rho = \sqrt{x^2 + y^2}$. The exciton wave function is found by varying the parameter λ to minimize the exciton binding energy using the Hamiltonian

$$H_{\text{ex}} = H_e + H_h - \frac{\hbar^2 \nabla_\rho^2}{2\mu} - \frac{e^2}{4\pi\varepsilon_0\varepsilon_r} \frac{1}{\sqrt{\rho^2 + (z_e - z_h)^2}}, \tag{7}$$

where the first two terms are constant with respect to λ, the third term represents the in-plane kinetic energy of the electron and hole about their center of mass, and the final term is the Coulomb potential energy. Here, μ is the reduced mass, ε_r is the relative dielectric permittivity, and $z_e(z_h)$ is the z position of the center of mass of the electron (hole). This variational method has been shown to accurately model excitonic absorption for quantum well profiles[7] and in the case of the asymmetric step well structure of Fig. 4, the predicted absorption spectrum peak at zero applied electric field matches closely with preliminary experimental data. The energy of the exciton is

$$E_{\text{ex}} = E_e + E_h - \frac{\hbar^2}{2\mu\lambda^2} - \langle \psi^{\text{ex}} | \frac{e^2}{4\pi\varepsilon_0\varepsilon_r} \frac{1}{\sqrt{\rho^2 + (z_e - z_h)^2}} | \psi^{\text{ex}} \rangle, \tag{8}$$

where the Coulomb integral is approximated with a piecewise polynomial found in Ref. 7. In bulk GaAs, the exciton Rydberg energy $\text{Ry}^* = E_{\text{ex}}$ is calculated to be 4.3 meV, but due to the confinement of the electron and hole in the z direction, exciton binding energies increase to values near 3Ry^* for structures with well width $L = 2.825$ nm (ten monolayers), as can be seen in Fig. 2. The dependence of the binding energy on the width of the confining potential is approximately logarithmic and extrapolates to a value of 4Ry^* for $L = 0.2825$ nm (one monolayer).

The zero-temperature continuum absorption spectrum may be calculated as a function of photon energy[4] using

$$\alpha(\omega) = \frac{C_0 \omega}{L} \sum_m \sum_n |I_{nm}|^2 \sum_{E_i} \frac{M(E)\Gamma}{(E_g + E_m + E_n + E_i)^2 [(E_g + E_m + E_n + E_i - \hbar\omega)^2 + \Gamma^2]}, \tag{9}$$

where L is the width of the quantum well, $M(E)$ is the energy- and polarization-dependant matrix element, $|I_{nm}|^2$ represents the electron-hole overlap integral, E_g is the band-gap energy, E_i is the energy of the state $|i\rangle$, C_0 is a constant factor, and Γ is a half linewidth broadening term. Absorption from discrete transitions may be calculated separately with a similar formula[4] and then combined with the continuum result to give the total absorption.

This approach accurately captures the main spectral features of other more detailed models[7,9] of the QCSE while maintaining a low computational cost for calculating the absorption spectrum. As illustrated in Fig. 1, for a rectangular quantum well potential profile, the expected loss in absorption strength and shift towards lower energy are obtained in agreement with previous calculations.[7]

As an initial definition of QCSE functionality we focus on the trajectory $\alpha_{\text{peak}}(E,F)$ that describes peak exciton absorption in the photon absorption spectrum $\alpha(E)$ as a function of applied electric field F.

We aim to demonstrate that it is possible to seek a physical quantum well potential profile which achieves the desired trajectory $\alpha_{\text{peak}}(E,F)$. An example of such a target function is a large variation in optical absorption at fixed photon energy E for a given range of applied electric field. Another example is a significant change in peak photon absorption energy for two fixed values of applied electric field. The configuration and hence the solution space for such targets are large. Obviously, searching the multidimensional configuration space is well suited to numerical methods that attempt to minimize the distance to the target function. Mathematically, this may be viewed as a constrained optimization problem in which one may define an objective or fitness function in terms of the difference between the result of any given configuration and the target function. The aim of the optimization procedure is to minimize the fitness function. In the Genetic Algorithms we discuss the use of a genetic algorithm to find optimal solutions for two types of optical absorption modulator target functions subject to physical and manufacturing constraints.

GENETIC ALGORITHMS

Genetic algorithms were first used as rudimentary models meant to simulate populations and keep track of changes in phenotype with each generation, but their utility outside the biological sciences was established when John Holland investigated their use for finding the optima of arbitrary mathematical functions.[10]

Unlike guided random walk and annealing methods,[11] which attempt to extract an underlying downhill trend from a noisy terrain, genetic algorithms use small groups of parameters called schemata to build potential solutions out of successful regions of parameter space, which enables efficient optimization even when overall trends do not lead towards a global minimum. This is possible because a genetic-algorithm-based search is not confined to a neighborhood (however large) around an existing solution, but rather, the search examines the relationship between current parameter values and builds a solution from these correlations, entirely independent of proximity in solution space. For complicated systems which may exhibit exponential sensitivity to minor changes in input parameters, such as the system under investigation here, methods based on neighborhood exploration are seldom able to construct a meaningful downhill trend, leaving neighborhood independent methods such as genetic algorithms as an efficient and reliable alternative.

FITNESS FUNCTIONS

Genetic algorithms begin with the development of a fitness function. The fitness function is to be optimized by the algorithm and is analogous to a biological system's likelihood to reproduce. Fitness functions may be as simple as a linear function of a single discrete variable or as complicated as a subjective human response to a function of several continuous variables. While fitness functions may be either maximized or minimized, only instances of minimization will be addressed here, since a function $f(x)$ to be maximized is equal in form to a function $-f(x)$ to be minimized.

In science or engineering, fitness functions often include the results of a physical model or simulation. This model represents a map from the space of input parameters $x \in \Re^n$ to the space of the model's results $x' \in \Re^m$. Typically, this outcome is then reduced to a single number to be minimized.

TARGET FUNCTIONS AS A PART OF FITNESS FUNCTIONS

When using genetic algorithms for the purpose of design, it is useful to reduce the results of a physical model to a single number by use of a target function specified by human input. A target function represents the desired results of the physical model and is defined only over those regions for which the desired results are specified. A scalar value is then obtained from the model's results by summing up the Euclidian distance between those results and the target function over all space for which the target function is defined. The fitness value is a function of this scalar,

$$G_{x_1,x_2,\ldots x_n} = G\left(\sum_{i=1}^{m} |M_{i,x_1,x_2,\ldots x_n} - T_i|\right), \tag{10}$$

where $M_{i,x_1,x_2,\ldots,x_n}$ is the ith component of the physical model, T_i is the ith component of the target function, and the sum is performed only over those values of i for which T_i is defined. $G(y)$ is usually taken to be unity for the sake of simplicity, but monotonically increasing functions such as power laws can be used to increase selective pressure when necessary. If the design should call for results as far as possible from a specified target function, $G(y)$ may be a monotonically decreasing function of this difference instead.

When there are multiple independent targets required of an optimal solution, the final fitness function must be composed of several single-target fitness functions, such as

$$G = A_1 G_1 + A_2 G_2 + \cdots, \tag{11}$$

where $A_1, A_2, \ldots$ are weighting constants. These functions should be weighted to best represent the importance of each individual target, and may be adjusted after several trial optimizations to account for the relative range of variation in each fitness function.

FIRST EXAMPLE: THE ABSORPTION FREQUENCY SWITCH

A multiobjective target function with functionality different from the conventional QCSE is a "frequency switch," where the absorption peaks at 0 and 70 kV/cm are required to have the same strength, but the energy of the absorbed photons is shifted by more than twice the width of the absorption peak (in this case at least 10 meV). In addition, the strength of the absorption peak at each of these two points should be as large as possible. The target function for this frequency switch is only defined at two points, so the sums for each target have only two terms. The three-objective fitness function for this search is

$$G = 2\frac{|\alpha_0 - \alpha_{70}|}{\alpha_{70}} + 1 \times 10^4 \frac{1}{\alpha_0 + \alpha_{70}} + 45\sqrt{0.02 - |E_0 - E_{70}|}, \tag{12}$$

where α_0 and α_{70} are the strengths of the absorption peak at 0 and 70 kV/cm, respectively, E_0 and E_{70} denote the absorbed photon energies, and all coefficients are set such that on average, each term is of comparable weight. The first term in the fitness function represents the percent difference in strength between the two peaks and is compared to a target value of zero. The second represents the average peak strength, and is also compared to zero, but because this comparison is to be maximized instead of minimized, the function $G(y)$ in Eq. (10) becomes $1/y$. The final term considers the change in absorbed photon energy compared to a target value of 0.02 eV (twice the acceptable value of 10 meV in this case). In this case $G(y)$ is a square root in order to further increase the penalty for functions with vanishing energy shift.

PENALTY FUNCTIONS

A solution where each of the three terms we have been considering is well optimized could have the same fitness value as a solution with one poorly optimized term and two highly optimized terms. For example, a solution with the undesirable energy shift of $E_0 - E_{70} = 0$ but extremely high average peak strength and near zero difference in peak strength could match the fitness with a solution that had acceptable values for each of the three terms.

To avoid such unwanted solutions, an acceptable range is assigned to each term in the fitness function. Most iterative search methods gather information from current solutions to generate the next trial set, so deleting solutions with terms beyond the acceptable ranges often results in discarding important information about the location of an optimal solution. Instead, the fitness values for solutions lying outside the accepted range are multiplied by a penalty function. The penalty function is equal to unity over the acceptable range, and increases monotonically outside of that range. The strength and form of the penalty function are chosen to provide sufficient selective pressure to keep the majority of fit solutions within all acceptable ranges. Ultimately, the penalty function(s) must be developed by observing the effects of trial penalty functions on the average relative fitness of each term. In the present example, two separate penalty functions are employed for each term of the fitness function. Terms outside of the smaller primary range but inside the greater secondary range are multiplied by a linearly increasing penalty function, while terms outside of the secondary range are multiplied by a quadratically increasing function.

CHROMOSOMES: REPRESENTING A POTENTIAL SOLUTION WITH AN ARRAY

To use genetic algorithms for optimization, every trial solution must be represented as an array from which the fitness function may be evaluated. For one-dimensional mathematical functions, this array (referred to as a chromosome) often consists of single digits in the binary representation of a trial solution. In the case of physical models with floating point parameters each array element holds a single floating point value. Because epitaxial $Al_xGa_{1-x}As$ structures can be grown with monolayer precision, the chromosome used in the present example contains the potential energy and width of each region. This way, a potential profile composed of N regions is represented by an array of length $2N$, where every potential-energy element is followed by a corresponding width element.

SELECTION SCHEMES

As a genetic algorithm is implemented, each member of an initial population of randomly generated chromosomes is evaluated using the fitness function described above and then used to construct the next generation of solutions. This generation is established by combining members of an intermediate population composed primarily of the fittest members of the initial generation of solutions. The method for selecting this intermediate population varies greatly depending on

the specific implementation,[9,12,13] but only roulette wheel and standard deviation methods will be addressed here.

The roulette wheel method,[9] when used for minimization, randomly selects members of the older generation to be copied into the intermediate population with probability

$$1 - \frac{G_j}{\Sigma G_i}, \tag{13}$$

where G_j represents the fitness value of the solution being selected, and the sum is over all solutions in the population. This selection scheme is straightforward to implement and creates an intermediate population which favors more successful solutions, but when the algorithm begins to converge, the fitness value for each solution approaches

$$G_j = G_0 \pm \Delta, \tag{14}$$

where Δ is small compared to G_0. In this case, the selective power of the scheme diminishes and every member of the old generation passes on to the intermediate population with equal probability.

The standard deviation method assigns each solution a relative fitness value g_j. If the solution has a fitness value lower than the average, the relative fitness takes on the value of

$$g_j = \frac{\bar{G} - G_j}{\sigma} + 1, \tag{15}$$

where $\bar{G}$ is the average fitness value and σ is the standard deviation of the fitness values. This maps the fitness function onto the interval $[1, \infty]$ for solutions which are more desirable than the average. If the solution has a fitness value that lies above the average fitness, the relative fitness becomes

$$g_j = \frac{1}{\left(\frac{G_j - \bar{G}}{\sigma} + 1\right)^2} \tag{16}$$

which maps these fitness functions onto the interval [0 1]. Once the relative fitness function has been defined, the intermediate population is filled with one copy of each solution with a relative fitness greater than 1. As a copy is placed into the intermediate population, 1.0 is subtracted from the relative fitness value and the procedure is repeated until there are no longer any solutions with relative fitness greater than one. Each solution is then given a chance to copy itself into the intermediate population with probability g_j. If the population size reaches its maximum, the process is halted and the crossover stage begins. If the process completes without filling all vacancies, any empty slots are filled with copies of the best solution of the generation.

The standard deviation selection scheme has the advantage that it maintains selective pressure as the fitness values descend over orders of magnitude, but can also prematurely converge to a poorly optimized solution if improperly implemented. Premature convergence can happen if one solution is many standard deviations better than any other solution in its generation. In that case, the intermediate population will be dominated by this single solution, and there is a high probability that all subsequent generations will be dominated by this solution as well. This problem can be overcome by running all optimizations several times and being certain to have a population size big enough to avoid domination by an outlying solution. In general, if the ratio of the average fitness value in the first generation to the fitness value of an acceptable final solution is m, and the standard deviation of the first generation is σ, then a population size on the order of $2\ m/\sigma$ will be large enough to avoid domination by poorly optimized solutions. In this example, a population size of 40 is sufficient to avoid any problems with premature convergence.

CROSSING OVER

Once the intermediate population has been constructed, all intermediate solutions are randomly paired up with each other, and each pair is used to create two solutions to be placed in the final population. The two final solutions are created by performing crossover and mutation operations on the intermediate pair. During crossover, a single point along a chromosome is chosen at random.[14] Elements located prior to the crossover point are swapped with the solution's partner. Elements following the crossover point are left untouched. This allows the elements which contribute to one solution's success to combine with the elements which contribute to its partner's success. The underlying assumption here is that clusters of array elements, or schemata, can help to lower fitness values on average, and when two lowering schemata on different array locations end up on a single array, they combine constructively.[15] This assumption brings to light the importance of choosing a basis and an order for the values stored in the chromosome array. The most successful configuration of a chromosome results from picking a set of variables whose impact on the fitness value are as independent as possible, that is, a set which is nearly orthogonal, and then arranging those variables such that each variable's contribution to the fitness value is most closely tied to the value of its nearest neighbors. Potential energy and length of each region of the $Al_xGa_{1-x}As$ potential profile are stored in adjacent locations for this reason.

MUTATION

While most genetic algorithms are capable of functioning without the mutation operation, it is important to keep the sampling of solution space from becoming too narrow. During mutation, each array element is given a small random chance to change its value. In the case of binary representation, this means flipping a single bit. In the case of floating point variables, this can either mean perturbing the current value by a small amount or setting the variable to a random value. Mutation probabilities typically lie somewhere between 0.001 and 0.1 and in some cases, these values change dynamically based on the homogeneity of the current generation. For the case of the frequency switch, the array is relatively short, so the probability of finding more than one mutation per array is nearly zero. In this case, it is computationally advantageous to combine the testing of each variable in the array into one test of the entire array. A mutation

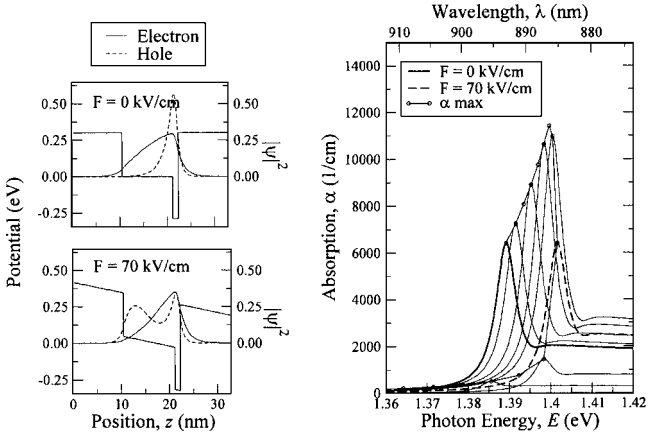

FIG. 3. Broken-symmetry quantum well structure obtained from numerical optimization. The target response is an absorption spectrum with strong, equal-strength absorption peaks at 0 and 70 kV/cm, separated by 10 meV. The solution involves two single-particle effects: first the electron shifts towards the localized hole, then the hole ionizes out of the deeper subwell and absorption is suppressed. Electron and hole ground-state probabilities are shown on the conduction band-edge profile for visualization purposes. Absorption spectra are shown in increments of 10 kV/cm.

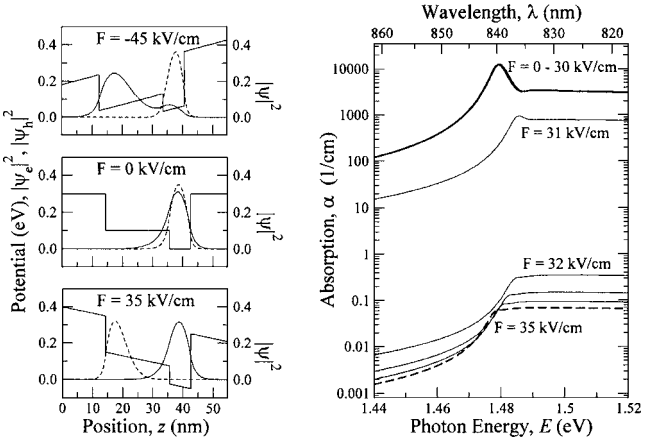

FIG. 4. Broken-symmetry quantum well structure obtained by modifying the structure from Fig. 3 to fit a target functionality: Rapid decrease of the absorption peak with increasing electric field F, small change in absorption peak energy, and stability with respect to crystal growth inaccuracies. The absorption peak diminishes with the application of either positive or negative electric field. Absorption spectra are shown in increments of 1 kV/cm.

probability of 0.2 is applied to each array, and those arrays which mutate have one element randomly selected and changed. The number of floating point operations avoided in this situation is of the order of the population size, 40.

Once the crossover and mutation operations are complete, the final generation replaces the initial generation and the process begins again. This continues until the fitness of the most fit solution in the population no longer shows significant improvement from one generation to the next.

RESULTS

One of the solutions found using this genetic-algorithm-based search is shown in Fig. 3. It resembles a conventional quantum well of 38 monolayers, except that near one edge of the well there is a four monolayer region of much lower potential. At zero applied field, the hole is localized within this deeper subwell region, but the electron is delocalized, leading to the initial absorption peak strength. As the field strength is increased, the electron wave function slowly shifts towards the deeper subwell, while the hole's position remains pinned. The absorption peak gradually increases until finally the applied electric field is sufficiently large for the hole to ionize out of the subwell by tunneling to the opposite region of the well. As this ionization process takes place, the electron-hole wave-function overlap and hence the absorption peak strength decrease, passing through the same value it held before the field was applied. In this way, the physics behind the trajectory of the absorption peak can be explained in terms of two distinct one-particle stages. While this serves as an acceptable solution to the design problem posed above, the rapid decrease in absorption due to field-induced ionization of the hole shows promise as a high contrast, low chirp, intensity modulator. We consider this next.

SECOND EXAMPLE: INTENSITY MODULATOR

Optoelectronic intensity modulators made from square wells tend to have low extinction ratios (usually in the range of 4–6). In light of this, solutions were sought for a target function that has the highest possible absorption strength at zero applied field, and then with small applied electric field this absorption strength quickly decreases with negligible shift in frequency and low chirp. For solutions similar to that shown in Fig. 3, the rate at which the peak absorption decreases can be approximately measured by examining the fraction of $|\psi_h|^2$ which overlaps the deeper subwell as a function of applied electric field. This function, which can be approximated as a sigmoid, approaches a step function in an ideal switch and can be characterized by the slope of the best-fit line through the center of the step transition.

To aid in the fabrication of devices with this specified potential profile, an additional parameter is added to the search: the zero-field peak absorption energy should be robust with respect to monolayer errors in the potential profile. The conduction band-edge profile, shown in Fig. 4, has a subwell width that initially localizes both the electron and the hole. Robustness with respect to monolayer variations in the potential profile is achieved by making the subwell shallower and wider than the result illustrated in Fig. 3, while still maintaining the ability to localize both particles. This change in form adjusts the sensitivity of absorption peak energy to monolayer fabrication errors from 30 meV to a much more manageable 2 meV. The overall width of the well is also increased so that when the hole ionizes out of the subwell, there is very little overlap with the electron wave function that remains localized within the subwell. The ionization of the hole in this solution occurs at 30 kV/cm. The slope at the transition point is $\theta=0.27$ $(\mathrm{kV/cm})^{-1}$, compared to $\theta=0.04$ $(\mathrm{kV/cm})^{-1}$ for the solution in Fig. 3.

Since this solution begins with both the electron and the hole localized within the subwell, when an electric field is applied in the reverse direction, overlap between the two wave functions again falls rapidly to zero, but in this case it

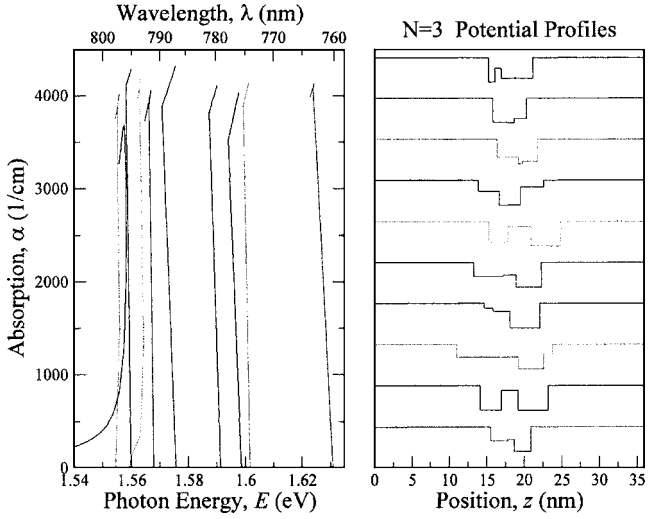

FIG. 5. Band-edge potential profiles and the corresponding absorption peak paths selected by a machine-based search of configuration space without a human-specified target function. Solutions are sorted from a random set by a binary test for properties which characterize interesting absorption peak paths. In this case, an interesting path is one with a large overall displacement in the E-α plane. Absorption peak paths are tracked as the external applied field increases from 0 to 125 kV/cm.

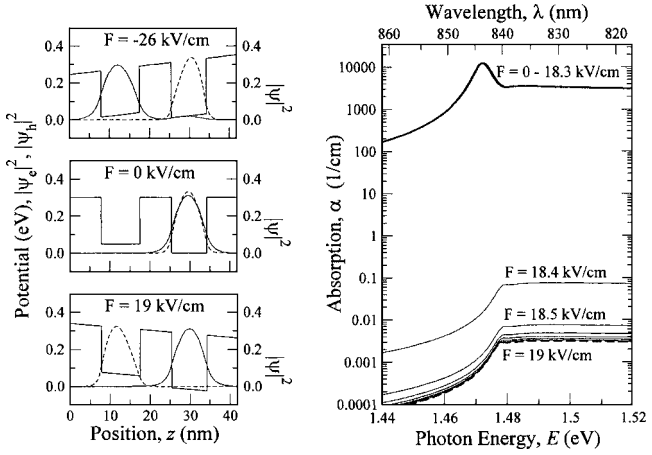

FIG. 6. Asymmetric double quantum well intensity modulator discovered by machine-based exploration without human specification of a target function. Absorption spectra are shown in increments of 0.1 kV/cm.

is due to the electron tunneling out of the well instead of the hole. This requires a slightly larger applied field (45 kV/cm compared to 35 kV/cm) but otherwise, the effect is nearly identical.

INTELLIGENT MACHINE-BASED SEARCHES OF CONFIGURATION SPACE

During the search for an absorption frequency switch the functionality of an intensity modulator emerged without being deliberately sought.[16] This scenario of unintentional discovery prompts the question of how many interesting and useful functionalities exist which have not yet been discovered simply because they have yet to be considered. By harnessing the computer power available today, an intelligent machine-based search may be developed to identify previously undiscovered useful solutions. To investigate this, an exploration algorithm was implemented to sample solution space randomly and return only those solutions it finds "interesting." The properties that categorize an interesting solution must be specified prior to the exploration, and should be as general as possible without allowing "uninteresting" solutions into the reported results.

Interesting solutions were defined to be those that traversed large distances in either the E or α directions of the $E-\alpha(E)$ plane. A minimum E distance of 0.001 eV (with nonvanishing absorption) or a minimum α distance of 12 000 1/cm would suffice to label the solution interesting. Reports of all such solutions (as in Fig. 5) were then stored for further examination of these interesting features.

During large exploratory searches such as this one, it is important to avoid losing an interesting solution among numerous perturbations about some other unrelated solution. For this reason, it is advantageous to categorize solutions by type and limit the number of solutions returned in each category. For the absorption peak path search, solutions which exhibit a monochromatic absorption falloff are reported separate from solutions with significant variation in peak energy. It is also possible to implement a ranking system in place of a binary test for interesting solutions, but again, in order to avoid populating the ranked list with solutions of only one type, categorization limits must be enforced.

Drawing from the results of the machine-based exploration, a solution was designed to better separate the electron and hole after ionization, and to better increase overlap before ionization. The solution in Fig. 6 consists of two narrow wells of varying height with a narrow barrier between them. The extinction ratio of this solution is better than that of the well/subwell solution by a factor of 100. While this solution looks quite different in form, it works on all the same principles as the well/subwell solution, and can be created from the previous solution by adding a small barrier at the well/subwell interface, and then allowing that barrier to grow in energy to match the potential value of the surrounding material. The narrow wells allow large overlap when the two particles are localized within the same well, and the barrier keeps overlap to a minimum while the two particles are separated. Because of this barrier, the overall structure is smaller in the z direction than the previous solution.

CONCLUSIONS

A realistic physical model and adaptive search algorithm can be used to design the elements of semiconductor devices such as a QCSE modulator. It was demonstrated that the physics of device operation in these optimized structures can differ from that governing the behavior of conventional devices. Automated searches of configuration space may be performed without the need to input a specific target function. The results of preliminary automated searches were presented and show improvement on a design optimized with a human-input target function. It is anticipated that these methods, as a component of device synthesis, will be of great use both for the design of devices and for the discovery of device functionality.

[1]D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Weigmann, T. H. Wood, and C. A. Burrus, Phys. Rev. Lett. **53**, 2173 (1984).
[2]D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, Appl. Phys. Lett. **45**, 13 (1984).
[3]X. Huang, A. J. Seeds, J. S. Roberts, and A. P. Knights, IEEE Photonics Technol. Lett. **10**, 1697 (1998).
[4]S. L. Chuang, *Physics of Optoelectronic Devices* (Wiley, New York, 1995).
[5]D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, Phys. Rev. Lett. **53**, 2173 (1984).
[6]G. Ji, D. Huang, U. K. Reddy, H. Unlu, T. S. Henderson, and H. Morkoç, J. Vac. Sci. Technol. B **5**, 1346 (1987).
[7]P. J. Mares and S. L. Chuang, J. Appl. Phys. **74**, 1388 (1993).
[8]The variational method has been shown to accurately model excitonic absorption for quantum well profiles (Ref. 7) and in the case of the asymmetric step well structure of Fig. 4, the predicted absorption spectrum peak at zero applied electric field matches closely with preliminary experimental data.
[9]C. Y.-P. Chao and S. L. Chuang, Phys. Rev. B **48**, 8210 (1993).
[10]David E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Boston, 1989).
[11]J. Thalken, Y. Chen, A. F. J. Levi, and Stephan Haas, Phys. Rev. B **69**, 195410 (2004).
[12]A. Prügel-Bennett and J. L. Shapiro, Phys. Rev. Lett. **72**, 1305 (1994).
[13]C. M. Fonseca and P. J. Fleming, *Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization*, Proceedings of the Fifth International Conference on Genetic Algorithms, Urbana, IL, July 1993, edited by S. Forrest (Morgan Kauffman, San Mateo, CA, 1993), 416.
[14]There are several variations in which multiple crossover points are selected, but there has been little work done to show that this added level of complexity yields any significant benefit.
[15]It is also possible to combine the values at each position along the array with a function such as a weighted average, but such approaches are not implemented here, because they reintroduce the assumption that a downhill trend can be easily extracted from the terrain of the fitness function.
[16]J. Thalken, W. Li, S. Haas, and A. F. J. Levi, Appl. Phys. Lett. **85**, 121 (2004).