

# REPORT DOCUMENTATION PAGE

*Form Approved*  
*OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

<b>1. REPORT DATE (DD-MM-YYYY)</b> 15/12/2006			<b>2. REPORT TYPE</b> Final Technical Report		<b>3. DATES COVERED (From - To)</b> 10/01/2004 – 09/30/2006	
<b>4. TITLE AND SUBTITLE</b>  Semiconductor Device Synthesis					<b>5a. CONTRACT NUMBER</b>	
					<b>5b. GRANT NUMBER</b> N00014-05-1-0029	
					<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b>  Levi, A.F.J.					<b>5d. PROJECT NUMBER</b>	
					<b>5e. TASK NUMBER</b>	
					<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>  University of Southern California 3620 South Vermont Avenue, KAP 132 Los Angeles, CA 90089-2533					<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> DARPA/MTO Dennis Healy 3701 North Fairfax Drive Arlington, VA 22203-1714					<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>  ONR Chagaan Baatar, Program Officer Office of Naval Research, Code 312 800 N. Quincy St. Arlington, VA 22217	
					<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b>  Approved for public release: distribution is unlimited.						
<b>13. SUPPLEMENTARY NOTES</b>						
<b>14. ABSTRACT</b>  A systematic approach towards optimal design of optoelectronic semiconductor devices has been developed. The adaptive design of excitonic absorption in broken-symmetry quantum wells has been used as a prototype system with which to explore the device synthesis method. The device model used captured the basic exciton absorption physics but did not include RF design. Electro-absorption modulator designs were investigated experimentally and a modulator with 25 GHz small signal -3 dB response was measured near 1550 nm wavelength. This device has performance comparable to the best commercially available modulator from Oki electric. Further improvements to device performance could be achieved by improving RF design, in particular, incorporation of traveling wave electrodes. To become part of the semiconductor device synthesis procedure the physical model of the device should incorporate these and other aspects of RF design.						
<b>15. SUBJECT TERMS</b>						
<b>16. SECURITY CLASSIFICATION OF:</b>				<b>17. LIMITATION OF ABSTRACT</b>  UL	<b>18. NUMBER OF PAGES</b>  47	<b>19a. NAME OF RESPONSIBLE PERSON</b> Dennis Healy
<b>a. REPORT</b> Unclassified	<b>b. ABSTRACT</b> Unclassified	<b>c. THIS PAGE</b> Unclassified	<b>19b. TELEPHONE NUMBER (include area code)</b> 571-218-4330			

## **Final Technical Report**

Submission date: 12/15/06

## **Semiconductor Device Synthesis**

A. F. J. Levi  
University of Southern California  
3620 South Vermont Avenue, KAP 132  
Los Angeles, California 90089-2533  
(213) 740-7318 voice  
(213) 740-9280 fax  
alevi@usc.edu  
<http://www.usc.edu/alevi>

**Grant No. N00014-05-1-0029**

**USC account No. 53-4502-0029**

Performance period: 10/1/04 - 9/30/06

## **Funded by DARPA**

Dennis Healy, DARPA/MTO  
3701 North Fairfax Drive  
Arlington, Virginia 22203-1714  
(571) 218-4330 voice  
(703) 696-2206 fax  
dhealy@darpa.mil

## **Administered by the Office of Naval Research**

Chagaan Baatar, Program Officer -- Nanoelectronics  
Office of Naval Research, Code 312  
800 N. Quincy St., Arlington, VA 22217  
(703) 696-0483 voice  
(703) 696-2611 fax  
baataarc@onr.navy.mil

**20061215191**

# **Semiconductor Device Synthesis**

## **Abstract**

A systematic approach towards optimal design of optoelectronic semiconductor devices has been developed. The adaptive design of excitonic absorption in broken-symmetry quantum wells has been used as a prototype system with which to explore the device synthesis method.

The device model used captured the basic exciton absorption physics but did not include RF design. Electro-absorption modulator designs were investigated experimentally and a modulator with 25 GHz small signal  $-3$  dB response was measured near 1550 nm wavelength. This device has performance comparable to the best commercially available modulator from Oki electric. Further improvements to device performance could be achieved by improving RF design, in particular, incorporation of traveling wave electrodes. To become part of the semiconductor device synthesis procedure the physical model of the device should incorporate these and other aspects of RF design.

**Table of Contents:**

Cover Page	1
Abstract	2
Table of Contents	3
List of Figures	4
Summary	5
Results and Discussion	6
1. Introduction	6
2. Optimal Design	7
2.1 Adaptive quantum design	7
2.2 Adaptive design of excitonic absorption in broken-symmetry quantum wells	8
2.2.1 Physical model: Excitonic absorption in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ quantum wells	8
2.2.2 Frequency switch: Delocalized electron solution	11
2.2.3 Frequency switch: Ionized hole solution	13
2.3 Search: Motivation, terminology, and terrain	14
2.3.1 Adaptive quantum design search conditions	14
2.3.2 Parameters, spaces, and mapping	15
2.3.3 Assessing fitness	15
2.3.4 Dimensionality reduction	17
2.3.5 Set-valued objective functionals	17
2.4 Search: Methodology and implementation	18
2.4.1 Downhill methods	19
2.4.2 Simplex methods	19
2.4.3 Gradient methods	19
2.4.4 Annealing methods	20
2.4.5 Genetic algorithm methods	20
2.4.5.1 Encoding: Chromosomes	21
2.4.5.2 Selection schemes	21
2.4.5.3 Crossover	24
2.4.5.4 Mutation	24
2.4.6 Hybrid techniques	25
2.4.7 Reset techniques	25
3. Optimization Without an Objective Function	26
3.1 Quantum well electro-absorption modulator	26
3.2 Motivation and methodology	28
3.2.1 Defining utility	28
3.2.2 Classifying solutions	29
3.3 Double well electro-absorption modulator	32
3.4 Initially separated double well solution	33
4. Experimental Results and Analysis	35
4.1 Incorporating fabrication into the model	35
4.1.1 Poisson solver	36
4.1.2 Sensitivity analysis for inaccuracies in well thickness	37
4.2 Experimental results	39
4.2.1 MOCVD grown InGaAsP modulators	40
5. Conclusions	43
6. References	45

## List of Figures

- Figure 1.1:** Flow chart outlining the iterative optimal design processes.
- Figure 2.1:** Changes in excitonic absorption spectrum for a symmetric quantum well with increasing electric field.
- Figure 2.2:** Exciton binding energy calculated using the tight binding Hamiltonian as a function of quantum well width for  $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$  quantum well structures.
- Figure 2.3:** Broken-symmetry double quantum well obtained from numerical optimization of well width and depth parameters.
- Figure 2.4:** Broken-symmetry quantum well structure obtained from numerical optimization.
- Figure 2.5:** An illustration of a set-valued objective functional applied to a specific solution trajectory.
- Figure 3.1:** Broken-symmetry quantum well structure obtained by modifying the structure.
- Figure 3.2:** Band-edge potential profiles and the corresponding absorption peak paths selected by a machine-based search of configuration space without a human-specified objective function.
- Figure 3.3:** Various categories for exciton absorption peak paths exhaustively sampled for the case of quantum well structures generated from two equipotential regions ( $N=2$ ).
- Figure 3.4:** The paths of the electron and hole wave function peaks as a function of applied field,  $F$ , for a square well 26 monolayers wide.
- Figure 3.5:** The paths of the electron and hole wave function peaks as a function of applied field,  $F$ , for a step well 52 monolayers wide, with a 40 monolayer step of 15% Al.
- Figure 3.6:** Asymmetric double quantum well intensity modulator discovered by machine-based exploration without human specification of an objective function.
- Figure 3.7:** Electron and hole probability densities for the ISDW solution.
- Figure 3.8:** Absorption spectrum for the ISDW solution.
- Figure 4.1:** The conduction and valence band edge profiles for a 6 quantum well  $\text{AlGaAs}/\text{GaAs}$  electro-absorption modulator grown in a PIN layer structure.
- Figure 4.2:** Absorption spectra plotted at  $F = 0$  kV/cm for conventional  $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$  structures of varying well widths.
- Figure 4.3:** Insensitivity of ISDW structure to monolayer fluctuations, despite narrow well width.
- Figure 4.4:** (a) High speed test equipment for measuring the RF response of electro-absorption modulators.
- Figure 4.5:** Measured and simulated absorption coefficients at 1550 nm wavelength for a conventional square well device with integrated laser.
- Figure 4.6:** Measured absorption as a function of  $V_{\text{bias}}$  at 1530 nm wavelength for InGaAsP step-well.
- Figure 4.7:** Conduction and valence band profiles for InGaAsP ISDW structure.
- Figure 4.8:** Measured small-signal response of a 150  $\mu\text{m}$  long electro-absorption modulator.

## Summary

A systematic approach towards optimal design of optoelectronic semiconductor devices has been developed. The adaptive design of excitonic absorption in broken-symmetry quantum wells has been used as a prototype system with which to explore the device synthesis method.

The device model used captured the basic exciton absorption physics but did not include RF design. The model used a tight binding model in the z-direction and assumed a 2D hydrogenic wave function in the x-y plane. The calculation of absorption focused primarily on the lowest energy exciton peak, neglecting many other spectrum characteristics. These simplifications allowed a relatively fast forward-solve time (typically 10 s on a 2.8 GHz Intel Pentium machine), at the expense of some accuracy in the absorption spectrum. This approach is justified because inhomogeneous broadening due to fabrication inaccuracies is typically large. Improved growth capabilities may allow more complicated structures that require improvements in the physical model. Specific improvements to consider include multiple exciton states and the Sommerfeld enhancement factor in absorption.

Dimensionality reduction in low energy quantum systems such as the one investigated was found to occur naturally. Electron wave functions tend to spread out and average over monolayer fluctuations in quantum well thickness and alloy composition. These facts were exploited to reduce the number of variables used by genetic search algorithms. To avoid overly constraining the optimization procedure, set-valued objective functions were used to reveal families of solutions with similar characteristics.

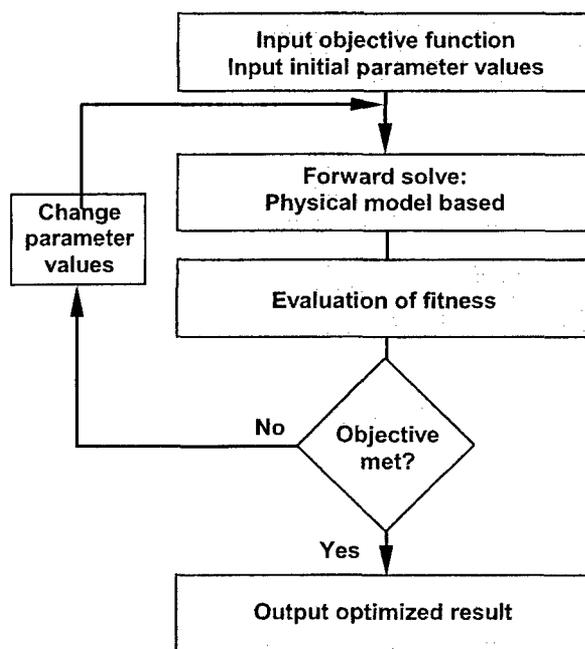
Electro-absorption modulator designs were investigated experimentally and basic aspects of the physical model validated. A modulator with 25 GHz small signal  $-3$  dB response was measured near 1550 nm wavelength. This device has performance comparable to the best commercially available modulator from Oki electric. Further improvements to device performance could be achieved by improving RF design, in particular, incorporation of traveling wave electrodes. To become part of the semiconductor device synthesis procedure the physical model of the device should incorporate these and other aspects of RF design.

## Results and Discussion

### 1. Introduction

By harnessing a combination of modern compute power, adaptive algorithms, and realistic physical models, it is possible to seek robust, manufacturable designs for semiconductor devices that meet previously unobtainable system specifications. In its simplest form, such semiconductor device synthesis algorithms solve an optimal design problem by identifying the best configuration to achieve a user-defined specification or objective function. Typically, this involves searching for a (broken-symmetry) spatial configuration of a semiconductor. It is anticipated that using such an approach it may be possible to efficiently discover completely new functionalities for both photonic and electronic semiconductor devices. The purpose of this report is not only to demonstrate that semiconductor device synthesis is feasible for a specific example of an electro-absorption modulator, but also to serve as a guide for others attempting to implement these methodologies in other physical systems.

A typical optimal design processes involves an iterative scheme where a desired objective is defined, and input parameters to a physical model are adjusted to find the best fit to the objective. This approach is outlined in Fig. 1.1.



**Figure 1.1:** Flow chart outlining the iterative optimal design processes. First an objective is defined, and an initial guess for physical input parameters is fed into the physical (forward) model. The results of the physical model are compared to the objective using a fitness value, and if the objective is not met, the input parameters to the physical model are changed and the evaluation process begins again.

This report is organized as follows: Section 2 opens with an introduction to the optimal design problem. Excitonic absorption in broken-symmetry quantum wells is then presented to serve as an example of how the design process is implemented. The

process of searching for an acceptable solution is then introduced, and various search algorithms are presented in the form of a guide to choosing the best method and parameters for a new design investigation. Section 3 focuses on the idea of performing a search for interesting or useful device behavior, without a human-specified objective, and Section 4 presents experimental results for device designs using the asymmetric quantum well solutions discovered during the search and classification described in Sections 2 and 3.

## **2. Optimal Design**

### **2.1 Adaptive quantum design**

In the near future, it will likely be possible to control the precise spatial positions of atoms and molecules using the experimental techniques now being developed by nanoscience [1, 2, 3, 4]. To complement these emerging capabilities it is clear that a new set of engineering tools has to be developed to assist in the exploration of a potentially vast number of atom configurations and a corresponding enormous range of physical properties.

Adaptive quantum design [5] sets out to address this challenging task. In contrast to classical systems, atomic scale devices exhibit quantum fluctuations and collective quantum phenomena caused by particle interactions. Besides offering an excellent testing ground for models of correlated electrons, they also force us to reconsider conventional paradigms of condensed matter physics, such as crystal symmetries that are imposed by nature. In some instances, such symmetries need to be explicitly broken to enable or optimize a desired system response. Consider for instance the quasiparticle density of states in tight-binding systems. In translationally invariant structures, i.e. crystals, it is well known that the spectral response function exhibits van-Hove singularities at positions of low dispersion, such as the band edges in a one-dimensional chain or the band center in a two-dimensional square lattice. These enhancements of the density of states can be very useful in amplifying system responses such as optical conductivity at specific incident energies. It is therefore important to be able to control the positions and shapes of such features by adaptive design techniques applied to models which capture the essential degrees of freedom of interacting atomic clusters.

Traditional ad-hoc methods for the design of nanoscale devices will likely miss many possible configurations. At the same time, it is unrealistic to expect individuals to manually explore the vast phase space of possibilities for a particular device function. A solution to this difficult design problem is to employ machine-based searches of configuration space that enable user-defined objective functions. Adaptive quantum design solves the optimal design problem by numerically identifying the best broken-symmetry spatial configuration of atoms and molecules that produce a desired objective function response.

The two major ingredients of adaptive quantum design are the physical model and the search algorithm that finds the global minimum in the parameter space of all possible

configurations. This problem is typically highly underdetermined in the sense that there can be several configurations that yield a system response very close to the desired objective function. Often, the associated landscape of solutions is shallow, non-convex, and has many nearly degenerate local minima.

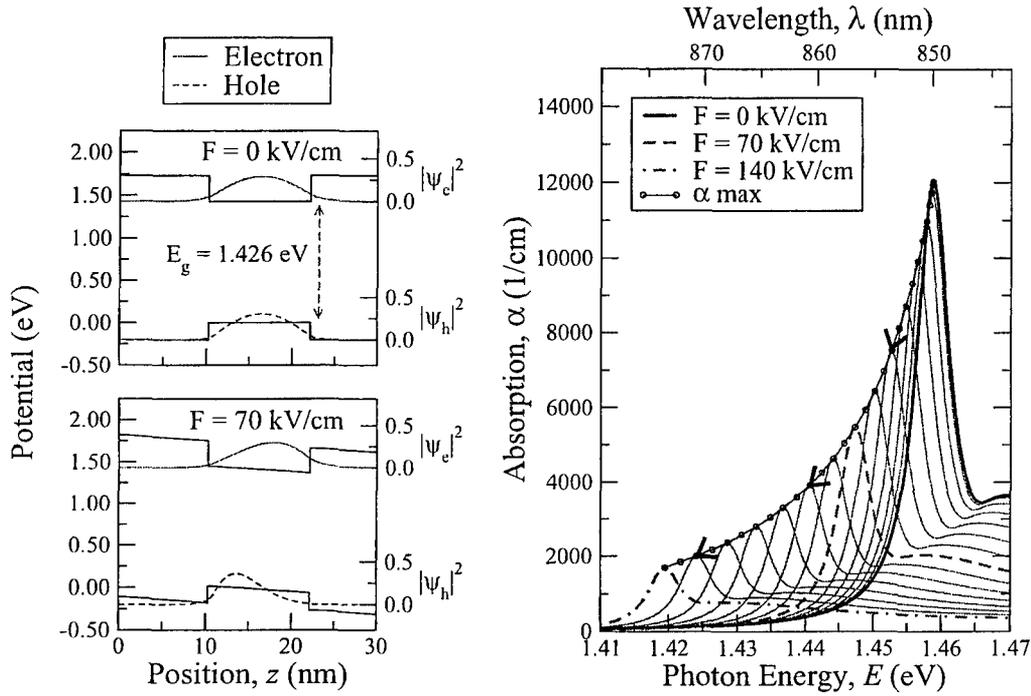
Adaptive quantum design therefore relies heavily on efficient algorithms that accurately model the interacting system and find its optimal configuration, i.e. the global minimum in the available parameter space that yields the best match to the desired objective function.

## **2.2 Adaptive design of excitonic absorption in broken-symmetry quantum wells**

Here, adaptive quantum design is used to identify broken-symmetry quantum-well potential profiles with optical response properties superior to previous ad-hoc solutions. This technique enables engineering of many-body excitonic wave functions and thus provides a new methodology to efficiently develop optimized quantum-confined Stark effect device structures.

### **2.2.1 Physical model: Excitonic absorption in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ quantum wells**

The absorption of photons in a direct band gap semiconductor involves 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 [6]. 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 [6, 7, 8].



**Figure 2.1:** 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 kV/cm 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.

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. 2.1. 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 that in turn reduces the dipole matrix element, diminishing the peak absorption strength.

Specializing to the  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  material system, the conduction and valence band potential profiles are calculated using the band gap energy

$$E_g = (1.426 + 1.247x) \text{ eV} \quad (2.7)$$

and a band gap offset ratio between conduction and valence band of 67/33 [16]. The effective electron mass in the conduction band is

$$m_c = (0.067 + 0.0835x)m_0 \quad (2.8)$$

and the effective hole mass in the valence band is

$$m_h = (0.34 + 0.42x)m_0 \quad (2.9)$$

To minimize computation time, 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

$$H = H_e + H_h = -t_e \sum_{\langle i, j \rangle} (c_{ei}^+ c_{ej} + c_{ej}^+ c_{ei}) + \sum_i \varepsilon_{ei} + t_h \sum_{\langle i, j \rangle} (c_{hi}^+ c_{hj} + c_{hj}^+ c_{hi}) + \sum_i \varepsilon_{hi}, \quad (2.10)$$

where  $t_e$  ( $t_h$ ) represents the electron (hole) hopping energy,  $\varepsilon_e$  ( $\varepsilon_h$ ) denotes the onsite 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 [9]. Following Refs. [10] and [9], a variational ansatz is used to obtain the separable 1s exciton wave function

$$\psi^{\text{ex}}(z_e, z_h, \rho) = \phi_e(z_e) \phi_h(z_h) \phi(\rho) \quad (2.11)$$

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

$$\phi(\rho) = \sqrt{\frac{2}{\pi\lambda^2}} e^{-\rho/\lambda} \quad (2.12)$$

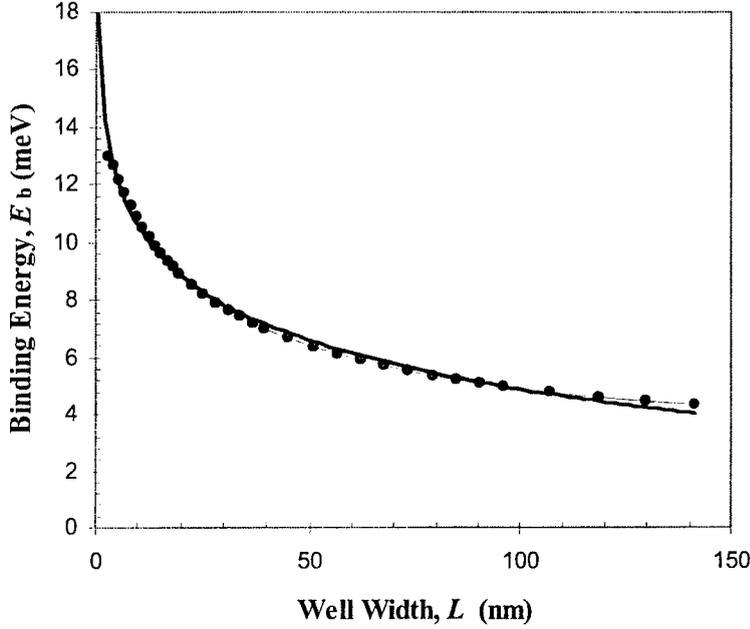
where the in-plane coordinate is  $\rho = \sqrt{x^2 + y^2}$ . The exciton wave function is found by varying the parameter  $\lambda$  to minimize the exciton binding energy using the Hamiltonian

$$H_{\text{ex}} = H_e + H_h - \frac{\hbar^2 \nabla_\rho^2}{2\mu} - \frac{e^2}{4\pi\varepsilon_0\varepsilon_r \sqrt{\rho^2 + (z_e - z_h)^2}} \quad (2.13)$$

where the first two terms are constant with respect to  $\lambda$ , 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,  $\mu$  is the reduced mass,  $\varepsilon_r$  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 [9]. The energy of the exciton is

$$E_{\text{ex}} = E_e + E_h - \frac{\hbar^2}{2\mu\lambda^2} - \left\langle \psi^{\text{ex}} \left| \frac{e^2}{4\pi\varepsilon_0\varepsilon_r \sqrt{\rho^2 + (z_e - z_h)^2}} \right| \psi^{\text{ex}} \right\rangle \quad (2.14)$$

where the coulomb integral is approximated with a piecewise polynomial found in Ref. [9]. 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  $3 \times \text{Ry}^*$  for structures with well width  $L = 2.825$  nm (10 monolayers), as can be seen in Fig. 2.2. The dependence of the binding energy on the width of the confining potential is approximately logarithmic and extrapolates to a value of  $4 \times \text{Ry}^*$  for  $L = 0.2825$  nm (1 monolayer).



**Figure 2.2:** Exciton binding energy calculated using the tight binding Hamiltonian as a function of quantum well width for  $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$  quantum well structures. The grey line represents a logarithmic fit to data that approximates the function away from the limits of large and small  $L$ .

The zero-temperature continuum absorption spectrum may be calculated as a function of photon energy [10] using

$$\alpha(\omega) = \frac{C_0 \omega}{L} \sum_m \sum_n |I_{nm}|^2 \sum_{E_i} \frac{M(E)\Gamma}{(E_g + E_m + E_n + E_i)^2 ((E_g + E_m + E_n + E_i - \hbar\omega)^2 + \Gamma^2)}, \quad (2.15)$$

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  $\Gamma$  is a half line width broadening term. Absorption from discrete transitions may be calculated separately with a similar formula [10] 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 [9, 11] of the QCSE while maintaining a low computational cost for calculating the absorption spectrum. As illustrated in Fig. 2.1, 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 [9].

### 2.2.2 Frequency switch: Delocalized electron solution

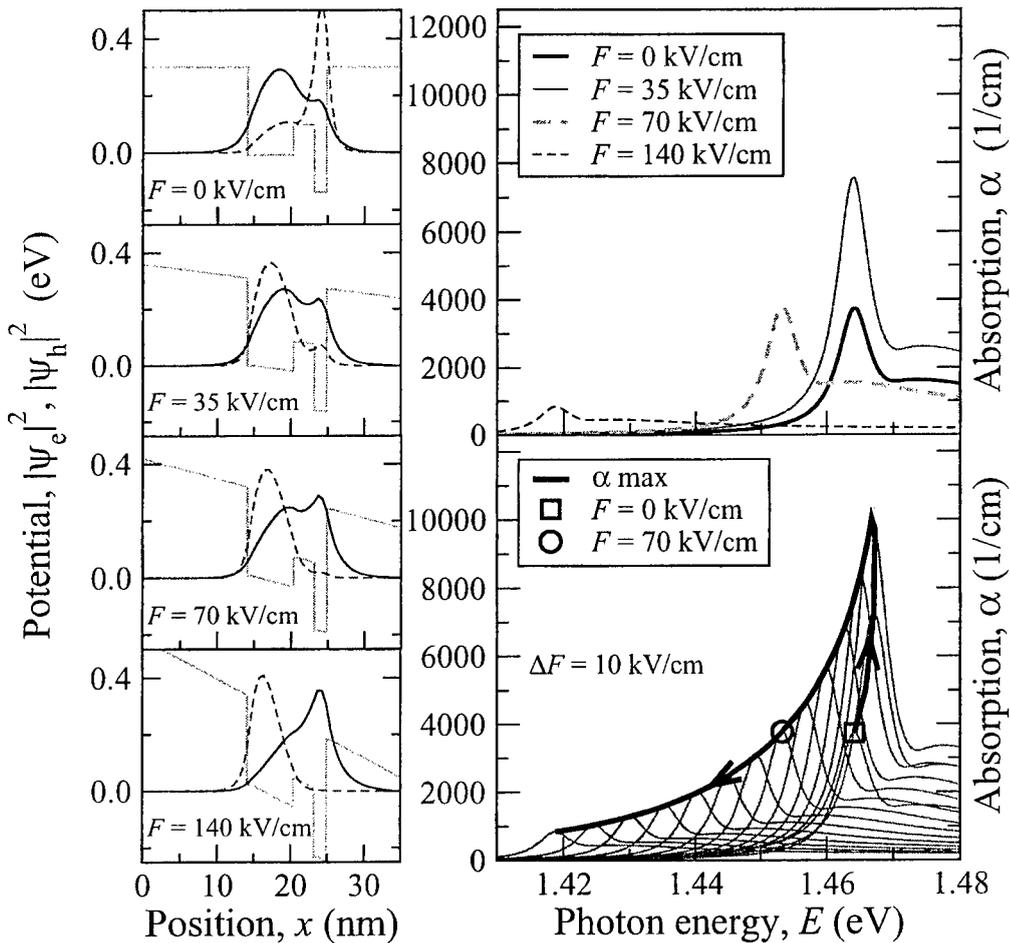
Consider the situation where the goal is to design a “frequency switch”, where the absorption peaks at 0 kV/cm and 70 kV/cm are required to have the same strength, but the energy of the absorbed photons is shifted by more than twice the width of the

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 fitness function for this search is

$$G = 2 \frac{|\alpha_0 - \alpha_{70}|}{\alpha_{70}} + 1 \times 10^4 \frac{1}{\alpha_0 + \alpha_{70}} + 45 \sqrt{0.02 - |E_0 - E_{70}|} \quad (2.16)$$

where  $\alpha_0$  and  $\alpha_{70}$  are the strengths of the absorption peak at 0 kV/cm 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.



**Figure 2.3:** Broken-symmetry double quantum well obtained from numerical optimization of well width and depth parameters. The objective response is an absorption spectrum with maximized, equal-absorption excitonic peaks at electric fields  $F = 0$  kV/cm and  $F = 70$  kV/cm which are separated in energy by 10 meV.

The best solution found by the adaptive quantum design method for this restricted search is shown in Fig. 2.3. It is reminiscent of structures investigated previously [13, 14] but obtained using an entirely different methodology. The optimized double well

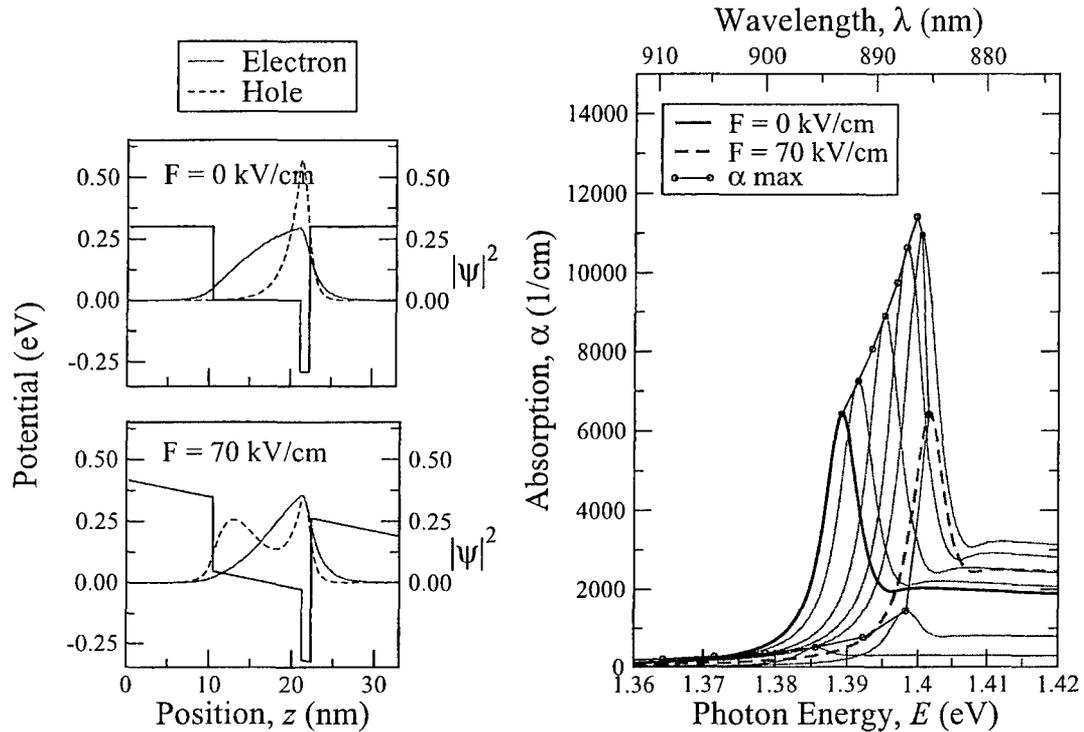
causes the ground state wave function of the hole (broken curve) to develop two maxima whose relative weight is shifted from left to right as the electric field is increased. Simultaneously, the center of the electron wave function (solid curve) moves from left to right, having a maximum spatial overlap with the right peak of the hole wave function at  $F = 0$  kV/cm and with the left peak at  $F = 70$  kV/cm. The resulting exciton peaks in the absorption spectrum, shown on the right, have the desired strength and separation. They are located on two sides of a maximum resonance that is reached at  $F = 20$  kV/cm. In this broken symmetry structure, the maximum strength of the excitonic absorption ( $\alpha$  max) initially increases with applied field, and then decreases. The corresponding shift of the resonant energy is also non-monotonic.

### 2.2.3 Frequency switch: Ionized hole solution

A second solution found using only two separate regions of constant potential,  $V(z)$ , is shown in Fig. 2.4. It resembles a conventional quantum well of 38 monolayers, except that near one edge of the well there is a 4 monolayer region with a much lower potential. At zero applied electric field, the hole is localized within this deeper sub-well 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 sub-well, 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 sub-well 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 decreases, 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.

Note, that this almost vertical drop of  $\alpha(E)$  in a very narrow range of electric fields indicates an exponential sensitivity that is highly desirable for the design of quantum well based modulators.

In summary, the adaptive quantum design methodology used reveals design options that have not been explored using conventional *ad hoc* approaches. The large number of possible solutions along with their exponential sensitivity to small parameter changes render the problem prohibitive to searches by hand [14, 15]. Instead, numerical searches identify optimized broken-symmetry structures that enable desired objective functionalities that are useful in the design of quantum-well based photonic switches and modulators. This new approach allows us to engineer many-body excitonic wave functions and thus illustrates a new paradigm in nanoscale design.



**Figure 2.4:** Broken-symmetry quantum well structure obtained from numerical optimization. The objective response is an absorption spectrum with strong, equal-strength absorption peaks at 0 kV/cm and 70 kV/cm, separated in energy 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 sub-well 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.

## 2.3 Search: Motivation, terminology, and terrain

A typical adaptive quantum design problem begins with a physical model that accurately simulates the physical system under investigation while remaining as computationally efficient as possible. The user will decide on a desired output, or “objective function”, and then employ advanced optimization methods to iteratively determine the set of input parameters which best approximates the desired result. While the previous example laid out the procedure involved in taking such an approach, the purpose of this section is to illuminate the nature and difficulty of adaptive quantum design problems in general.

### 2.3.1 Adaptive quantum design search conditions

An adaptive quantum design problem is characterized by a large and complicated search space. That is, the number of possible input parameters must be large enough and the computational cost of each function call must be high enough that an ad-hoc forward-only approach to the problem is not possible. Also, the complexity of the problem should be such that the results of the physical simulation cannot be approximated by linear interpolation or extraction of trends by random sampling. Having laid out the nature of the problem to be solved, the next point of interest is assessing the space and functions involved.

### 2.3.2 Parameters, spaces, and mapping

The parameter space, or search space,  $\mathbf{S}^n$ , is the  $n$ -dimensional space of all possible configurations that can be tested or input into the physical model. The size of this space depends on the number of input parameters,  $n$ , and the available range of each parameter. This set can either be discrete or continuous, depending on the details of the problem [12], but it is almost always bounded. The physical model acts as a map from the search space,  $\mathcal{S}^n$ , to a subset of the space of all possible observables,  $\mathcal{U}^m \subset \mathcal{U}^m$ . This mapping is usually not well-behaved, and  $\mathcal{U}^m$  may even be unbounded or disconnected, depending on the nature of the problem.

Once  $\mathcal{S}^n$  has been mapped onto  $\mathcal{U}^m$  by the physical model, it must be compared to an objective typically input by a human. This objective,  $\mathbf{T} \in \mathcal{U}^m$ , specifies a desired value for each of the  $m$  observables, and does not necessarily reside in the subset of obtainable solutions,  $\mathcal{U}^m$ . In those cases where  $m$  is large and each variable is similar in range, the objective can be considered a function of the index of its dimensionality. That is,  $\mathbf{T} = (t_1, t_2, t_3 \dots t_m)$  can be thought of as  $T(i)$ , with  $1 < i < m$ . This formulation arises from the fact that for many physical models, the observable variables actually represent discretized values of a function along an axis.

To compare a given set of observables  $\mathbf{P}_0 \in \mathcal{U}^m$  to a human-input objective  $\mathbf{T}$ , a cost or fitness function,  $F(\mathbf{X})$ , must be constructed for  $\mathbf{X} \in \mathcal{U}^m$ . The fitness function is a scalar field defined on the space of all possible observables,  $\mathcal{U}^m$ , and not just the space of all outputs from the physical model,  $\mathcal{U}^m$ , because  $\mathcal{U}^m$  cannot be known a priori. At the most basic level,  $F(\mathbf{X})$  represents the distance between the two points  $\mathbf{X}$  and  $\mathbf{T}$ , and can be visualized as a terrain in  $\mathcal{U}^m$  referred to as the "fitness landscape"[22]. This landscape is a useful concept because even though the physical model may be computationally expensive and impossible to explore exhaustively, the fitness landscape is often much simpler to compute, and can be used to evaluate the effectiveness of both the objective  $\mathbf{T}$  and the fitness function  $F(\mathbf{X})$  over the entire range of  $\mathcal{U}^m$ .

### 2.3.3 Assessing fitness

Choosing a fitness function is to a large degree arbitrary, but developing that fitness function into one which will provide the best realizable results during optimization is an involved and deliberate process. The fitness function may need to include various combinations of penalty functions, weights, and even re-casting of variables in order to improve the probabilities of a successful search. Much like the optimization process itself, the fitness function should begin with an initial guess. Assuming we would like to minimize the fitness (rather than maximize), a good initial guess for the fitness function would be

$$F(\mathbf{X}) = \sum_i D(x_i, t_i), \quad (2.17a)$$

where

$$D(x, t) = \begin{cases} -x & \text{if } t = \infty \\ x & \text{if } t = -\infty \\ |x - t| & \text{otherwise} \end{cases} \quad (2.17b)$$

Take, for instance, the fitness function for Sections 2.2.2 and 2.2.3. The observables in this case are

$$\mathbf{P} = (\Delta E, \Delta\alpha, \alpha_{ave}), \quad (2.18)$$

where  $\Delta E$  represents the change in energy of the exciton peak as the applied field changes from 0 to 70 kV/cm,  $\Delta\alpha$  represents the percent change in absorption at those same points, and  $\alpha_{ave}$  represents the average of the absorption strength at those two points. The objective in this case is

$$\mathbf{T} = (0.02, 0, \infty), \quad (2.19)$$

which implies that a good solution would have a very large change in energy (0.02 eV), minimal change in absorption, and a high average absorption. Applying the guess from Eq. 2.17 to this example, one obtains

$$F(\mathbf{X}) = (0.02 - \Delta E) + \Delta\alpha - \alpha_{ave} \quad (2.20)$$

After this, each term should be examined to be sure that the fitness landscape for that particular variable accurately mirrors the desirability of those values. For this example, to improve selective pressure, a square root was used for the  $\Delta E$  term. For  $\alpha_{ave}$ , it was determined that while stronger peaks were more desirable than weaker peaks, the primary goal was to exclude very weak peaks with a large amount of pressure, but not to spend too much computer time on raising peaks which were already of an acceptable height. For this reason, the  $\alpha_{ave}$  term was replaced with  $1/\alpha_{ave}$ . The next step is to properly weight each term according to their relative ranges and importance. In this case, each term is considered equally important and each term is weighted so that on average, when a random result from the physical model,  $\mathbf{P}_0$  is plugged in to  $F(\mathbf{X})$ , all three terms are of the same order of magnitude. The final fitness function used to find the solutions in these examples was

$$F(\mathbf{X}) = 2\Delta\alpha + 2 \times 10^4 \frac{1}{\alpha_{ave}} + 45\sqrt{0.02 - \Delta E}. \quad (2.21)$$

Any optimization scheme that uses a fitness function with more than one term, such as the one in Eq. 2.21, is in danger of falling into an undesirable extreme region of the fitness landscape. For example, a solution with the undesirable energy shift of  $\Delta E = 0$  but extremely high  $\alpha_{ave}$ , and near zero  $\Delta\alpha$  could match 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. For the fitness function in Eq. 2.21, two separate penalty functions are employed for each term of the fitness function. Terms outside the smaller primary range but inside the greater secondary range are multiplied by a linearly increasing penalty function, while terms outside the secondary range are multiplied by a quadratic increasing function.

### 2.3.4 Dimensionality reduction

For many physical models, small perturbations in parameter space result in negligible changes in the space of observables, and in some cases this low sensitivity is known to exist over all of parameter space. One such instance is the physical model in Sections 2.2.2 and 2.2.3, where the electron and hole ground state wave functions average out single-monolayer perturbations with energies less than the confinement energy. When this sort of averaging can be assumed, it is possible to adjust the step size of the investigation, or to re-cast the input parameters such that the dimensionality of the parameter space is reduced.

For the case of quantum well potential profiles in  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ , changes in peak absorption energy greater than half of the peak line width,  $\Gamma/2 = 0.007\text{eV}$  and changes greater than 5% in total absorption peak strength require uniform perturbations of 8 monolayers or more in a well 24 monolayers wide. This implies that allowing variation on the level of individual monolayers will not result in any functionality that is not also accessible by varying regions of uniform potential with a width of 8 monolayers or greater.

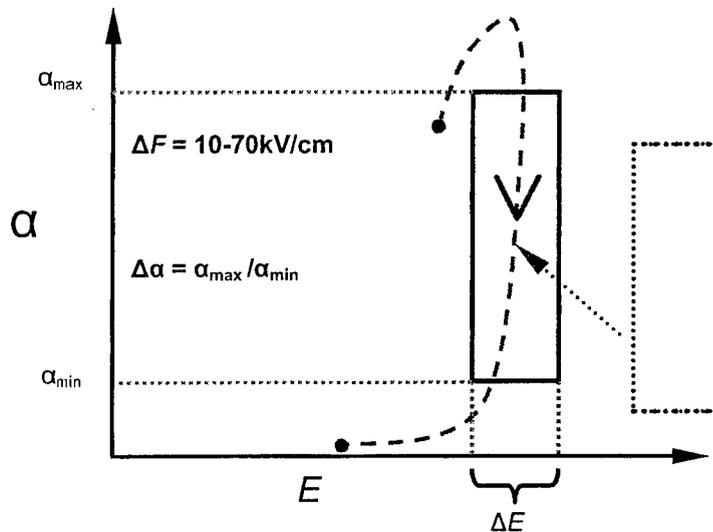
The result of this dimensionality reduction is a dramatic improvement in search results, due to the size of the search space decreasing by many orders of magnitude. If the well width is fixed at 24 monolayers, and aluminum concentration is discretized into steps of 1% with 30% aluminum barriers, there are  $30^{24} = 2.8 \times 10^{35}$  possible configurations in the search space. If, however, changes are made in segments of 8 or more monolayers, there are  $((24-8)-8)30^2 = 7200$  possible configurations in the search space, where  $((24-8)-8)$  is the total number of possible width combinations for the case of  $N = 2$  regions of uniform potential. The possibility of reducing the size of the search space by orders of magnitude makes it advisable to investigate the sensitivity of a given system to perturbations before any large-scale search efforts are launched.

### 2.3.5 Set-valued objective functionals

While for many problems, a single objective is sufficient to find an acceptable solution, it can be beneficial to use a set of objectives instead. Consider the search for an electro-absorption modulator, such as the one performed in Section 2.2.3. A conventional objective function would require high absorption values at the beginning of the trajectory, and low absorption values at the end, while varying only slightly in energy. The disadvantage to this fixed objective approach is that trajectories that start at low

absorption, rise first, and then fall again are excluded, as well as other unanticipated trajectories.

A more flexible approach defines a set of objectives based on relative boundaries in absorption and energy space, and then checks to see if any subsection of a given trajectory satisfies the boundary conditions of at least one of the objectives in the set. For electroabsorption modulators, a single objective is represented by a rectangular region bounded in energy by  $E$  and  $E+\Delta E$ , and bounded in absorption by  $\alpha_{\max}$  and  $\alpha_{\max}/\Delta\alpha$ , and the set of objectives used for optimization is the set of these regions for all possible  $\alpha_{\max}$  and  $E$ . Comparing a solution to an objective is not performed as a distance calculation here, but rather, the set of objectives is examined to determine if the solution's trajectory satisfies at least one of them. An illustration of this can be seen in Fig. 2.5. The implementation of a set-valued objective system such as this is computationally more taxing than conventional methods, but the fluidity in the method can be worth the effort for cases where desired functionality is less rigid than a point in the space of observables.



**Figure 2.5:** An illustration of a set-valued objective functional applied to a specific solution trajectory. In this example, a successful match occurs when the trajectory passes through both upper and lower  $\alpha$  boundaries of the rectangular region without crossing through either  $E$  boundary. This comparison is performed for all possible locations of the rectangular objective region.

## 2.4 Search: Methodology and implementation

Choosing the appropriate optimization scheme can be a difficult process of trial and error and often requires meticulous care and thorough sampling of the projection of input parameters onto the fitness landscape. The following serves as a reference guide for many of the more useful algorithms.

### 2.4.1 Downhill methods

The beauty in downhill methods lies in their simplicity. Because they are so easy to implement, it is recommended to try a downhill method before any other methods are considered. A typical downhill search proceeds as follows:

- 1) Select a random starting point.
- 2) Make a small random change to a single input parameter.
- 3) Compare the fitness value to the value before the change was made. If the fitness improved, keep the step, otherwise reject it.
- 4) Repeat steps 2 and 3 until solution converges.

One problem with this method is that if you have a large search space and choose a starting point far away from the minimum, it can take an extremely long time to converge. The primary disadvantage to this method, however, is that it is extremely likely to stop at a local minimum. Most of the following strategies are embellished versions of the downhill method developed specifically to get around the local minima.

### 2.4.2 Simplex methods

Not to be confused with the simplex method for linear programming, the downhill simplex method [23, 24, 25] explores  $n$ -dimensional parameter space using a simplex, or polytope with  $n+1$  vertices. Here is a sample simplex method algorithm:

- 1) Evaluate  $n+1$  random points ( $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{n+1}$ ) in parameter space.
- 2) Find the point with the highest fitness, and reflect it through the center of the simplex
- 3) If this new step is now the new minimum point, expand it even further away from the center of the simplex, if it is the new maximum point, contract it closer to the center of the simplex.
- 4) Repeat steps 2 and 3 until convergence is reached.

A good way to visualize this method spatially is to picture a triangle flipping over itself as it tumbles down a hill. While this method has the ability to avoid local minima which are smaller in size than the simplex itself, it is still a local method, and requires a smooth projection of parameters onto the fitness landscape on the order of the size of the simplex. This often results in a simplex too small to avoid local minima when used for nonlinear problems such as those approached for adaptive quantum design.

### 2.4.3 Gradient methods

There exist a wide variety of gradient methods for finding minima [24, 25], most of which become too computationally demanding for problems with more than four input parameters, and many of which provide little, if any improvement over downhill methods in terms of overall speed in finding a solution [12]. These methods (and in particular the Newton method) were designed before computers were used to solve problems iteratively, and therefore focus on reducing iterations in favor of extensive determinate calculation. For this reason, these methods are often favored those who are uncomfortable with stochastic processes, or who do not fully understand the complexity of the problem at hand. Here is an example gradient search algorithm:

- 1) Pick an initial guess for the solution.
- 2) Calculate the gradient at that point.

- 3) Take a step in the direction of steepest descent, using a step size which would lead to the minimum if the function were perfectly quadratic.
- 4) Repeat steps 2 and 3 until convergence is reached.

Two primary faults with this method are that it is often no better at avoiding local minima than the simple downhill approach, and that gradient (and step size) calculation are often too expensive (if possible at all) to calculate in an iterative fashion. Having said this, it should be noted that the gradient information is still extremely useful to any search, so in those cases where this information happens to be inexpensive or when it can be calculated at little extra cost during the calculation of the forward problem [26], it is definitely recommended to use this in the search.

#### 2.4.4 Annealing methods

Simulated annealing methods [12, 27, 28] are modeled after thermal annealing, a physical process where a substance is heated and then slowly cooled in order to remove crystal defects. The physical process works by allowing ample opportunity for the constituent molecules to find the lowest energy state (a crystal) before the temperature becomes too low to allow for rearrangements. The computational equivalent is some variation on these steps:

- 1) Pick a starting point randomly
- 2) Pick a starting temperature,  $T = T_0$
- 3) Take a random trial step, as in the downhill method. If the step lowers the fitness, keep it. If it increases the fitness, accept the step with probability  $p = \exp(-\Delta F(\mathbf{X})/T)$ .
- 4) Repeat steps 2 and 3, periodically lowering the temperature in small increments, and stopping when convergence is reached.

The assumption behind this method is that larger barriers exist further away from the minimum, and that as the minimum is approached, the number and size of these barriers decreases. This assumption is valid for many physical systems in nature, but when simulated annealing methods are applied to those difficult or nonlinear problems for which this assumption is incorrect, little progress will be made towards a nonlocal minimum. Often, the best test for whether or not the annealing assumption applies is to run a simple simulated annealing search and look for any improvement over a downhill search.

Many variants exist, most of which find new ways to construct  $p$  or lower  $T$ . One variation, called "triggered annealing" [12], functions as a downhill method, and only switches to simulated annealing when minima are found, greatly decreasing the search time for problems which deviate from the annealing assumption. The best variant to use depends on the particular problem at hand and should be chosen by running trial searches.

#### 2.4.5 Genetic algorithm methods

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 [17, 30]. There exist many variations on genetic algorithms, with just as many names, so to disambiguate this section, the term "genetic algorithm" will be used to describe all such strategies.

Genetic algorithms selectively combine populations of trial solutions to build new solutions out of successful regions of parameter space, which enables efficient optimization even when local 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 new solution from these correlations, entirely independent of proximity in search space. For complicated systems that may exhibit exponential sensitivity to minor changes in input parameters, 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.

#### **2.4.5.1 Encoding: Chromosomes**

The simplest form of the genetic algorithm search is the binary search, which refers to the method used to encode the input parameters. This method encodes a single point in parameter space as a binary array, called a chromosome. The easiest problems to encode in this manner are those with rigid lattice constraints, where each lattice position is represented by a single bit in the array [29]. In these cases, a value of 1 would signify an occupied lattice site, and 0 would signify a vacancy. Most problems, however, are not constrained to a lattice, and have many real valued parameters. The chromosome can then no longer be treated as a binary array and instead contains a single real-valued parameter at each point in the array.

The encoding of the parameters is at the same time one of the most important aspects of the search and one of the most overlooked. The primary assumption underlying a genetic algorithm search is that variables or small combinations of variables contribute to the fitness of a solution with some degree of independence from other variables on the array. While the highly popular concept of schemata [17, 30] reinforces this assumption for the case of binary arrays optimizing a one-dimensional parameter space, the schemata explanation becomes misleading for complicated or real-valued problems, and will not be used here in favor of a more general variable-interdependence model that allows for scalability, nonlinearity, and chromosome reordering.

#### **2.4.5.2 Selection schemes**

The iterative process of optimizing with a genetic algorithm begins with a population of trial solutions, which are given a certain number of copies to contribute to the next generation. These copies are placed in an intermediate generation according to the selection scheme implemented and then operated on by crossover and mutation operators. The purpose of the selection scheme is to ensure that the solutions with the greatest fitness are allowed to have the greatest influence over the next generation, while still avoiding the premature convergence caused by domination of a single

solution over the population. Because of this, selection schemes are intimately interdependent on the size of the population, and both should be selected in tandem.

The most common method for small binary searches is the roulette wheel method [17, 31]. This method, when used for minimization, randomly selects members of the older generation to be copied into the intermediate population with probability

$$1 - \frac{F(\mathbf{X}_j)}{\sum F(\mathbf{X}_i)}, \quad (2.22)$$

where  $F(\mathbf{X}_j)$  represents the fitness value of the solution being selected, and the sum is over all solutions in the population. Because of this summation, it is obvious that the probability of selecting more than one copy of a successful solution depends directly on the spread of the fitness distribution and inversely on the number of solutions in the population. If this method is to be used, the population size should be selected such that for randomly generated populations, the best solution gets picked multiple times on average when plugged into Eq. 2.22. This size is generally around 20 (depending on the distribution) and generates between 2 and 4 copies of the best solution at each iteration.

The roulette wheel selection scheme is primarily used because straight forward to implement and is computationally inexpensive, but when the algorithm begins to converge the fitness value for each solution approaches

$$F(\mathbf{X}_j) = F_0 \pm \Delta, \quad (2.23)$$

where  $\Delta$  is small compared to  $F_0$ . 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. For this reason, roulette wheel selection is not recommended when the value of the fitness function changes by more than an order of magnitude during the search process.

Another commonly used method is the tournament selection method [22, 31], where two solutions are selected at random, and the more desirable solution is placed into the intermediate generation. This selection method has the advantage of maintaining its selective pressure as the fitness distribution converges, but like the roulette wheel method, it provides very few copies of good solutions, and can result in extremely slow convergence for populations of more than 20.

The standard deviation method [21, 22] assigns each solution a relative fitness value,  $f_j$ . If the solution has a fitness value lower than the average, the relative fitness takes on the value

$$f_j = \frac{\mu - F(\mathbf{X}_j)}{\sigma} + 1, \quad (2.24)$$

where  $\mu$  is the average fitness value and  $\sigma$  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 average. If the solution has a fitness value that lies above the average fitness (less desirable), the relative fitness becomes

$$f_j = \frac{1}{\left(\frac{F(\mathbf{X}_j) - \mu}{\sigma} + 1\right)^2}, \quad (2.25)$$

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 one. 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  $f_j$ . 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 also carries with it a much higher likelihood of premature convergence due to domination by extremely fit solutions. This sort of domination can occur if one solution is many standard deviations better than any other solution in its generation. In that case, the intermediate population will contain many copies of this single solution, and there is a high probability that subsequent generations will further contribute towards this trend of early convergence. For this reason, standard deviation methods require larger population sizes than other methods. 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  $\sigma$ , then a population size on the order of  $2m/\sigma$  will be large enough to avoid domination by poorly optimized solutions. It should also be noted that if there are factors limiting the improvement of the average fitness, such as a physical system where the effects of mutation are so strong as to return most solutions to near-random values, a standard deviation method will seldom avoid premature convergence, and other methods should be considered.

There are also many interesting modifications that can be made to a selection scheme to help improve the search. Elitism, or partial replacement, selects a certain portion of the population to remain untouched until the next generation. This practice blurs the lines between generations, and can be a useful tool when the fitness distribution changes drastically from generation to generation. In practice, however, it is often implemented as a result of either an uncomfotability with non-monotonically decreasing fitness functions, or as a crude fix to the problem of having either a population that is too large, or a selection scheme that is providing too little selective pressure. Another important modification is the incorporation of an artificial temperature, much like simulated annealing, and often in the form of a Boltzmann distribution [22]. This can be useful in cases where selective pressure is relatively unimportant at early stages of the search process, for instance, when the search space is large or varies rapidly compared to the number and sampling of a random population, but becomes increasingly more important as a single solution type is chosen.

While these options do not by any means exhaust the possibilities for selection, they should provide a broad enough sampling to help in the selection scheme development process. It is important to approach every optimization problem individually, and develop the selection scheme not only around the population size, but the rest of the optimization process as well.

### **2.4.5.3 Crossover**

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 new solutions to be placed in the final population. The two final solutions are created by performing crossover and mutation operations on the intermediate pair. The crossover operation randomly assigns elements of each parent chromosome to the two resulting offspring.

The most common crossover method is the single point crossover [17], where a point in the chromosome is chosen at random, and the offspring are generated by swapping all chromosome elements before that crossover point. If all elements of the chromosome have only finite-range interdependence on their neighbors, the single point crossover method is a good choice. This method preserves clusters of nearby parameters and passes them on with clusters from the other parent, in hopes that the two successful clusters will combine constructively. A variant of this method is the two-point crossover method, where all the information between two randomly chosen points is swapped. This method treats the chromosome as a connected ring, and is ideal for problems with periodic boundary conditions. If strong interdependence exists between elements which are not neighbors in the chromosomal array, either the chromosome should be re-ordered, the crossover should be customized to keep the interdependent variables together, or the parameters should be re-cast in a different basis. Otherwise, clusters of parameters that could add constructively to the fitness of a solution will often be broken, and the crossover operation will be reduced to a randomization of parameters.

In the case of binary encoding of occupied states on a two-dimensional lattice, the use of a two-dimensional chromosomal array results in a direct correlation between spatial and chromosomal location. Geometric crossover [29, 32] methods tend to work well in these situations. Geometric crossover methods extend the concept of a swapping point in one dimension to a swapping line in two dimensions or a swapping plane in three dimensions. This extension into higher dimensions, where a chromosomal array with  $n$  dimensions is cut by a hyperplane of dimension  $n-1$ , demonstrates again the intimate relationship between the encoding of parameter data and the crossover operation.

Another useful crossover method is the value-division method, where parameters are assumed to be related by value rather than position in the chromosomal array.

### **2.4.5.4 Mutation**

The mutation step is probably the most flexible step in the genetic algorithm process, and the most understood as well. The purpose of this step is to randomize some members of the population so that a larger portion of parameter space can be incorporated into the search. The mutations rates for various mutation types should be

set such that as a solution converges, a significant portion of the population retains a relatively undesirable fitness value. This guideline is based on the assumption that mutations are more likely to damage a solution than improve it, so a significant number of visibly damaged solutions implies a good chance of finding a small number of improved solutions. The ideal rate of mutation operations varies from problem to problem, and should be closely examined to provide a balance between sampling new regions of search space and destroying old solutions.

For binary encoded problems, the basic mutation operation is the point, or bit-flip mutation, which flips the value of a particular element from 1 to 0 or from 0 to 1. The real-valued equivalent is perturbing the value of a particular parameter by a random step-size, or alternately, reassigning the parameter with a new random value. Depending on the problem, the effect of this operation on the fitness value can vary greatly, and the mutation rate should be set accordingly, but typical values are on the order of 1% of the total number of bits or parameters.

More global mutation operations can be useful when the effects of point-mutation are relatively small. For binary systems, a bit-shift or group-shift mutation will perform a chromosomal translation operation on a 1 (or group of 1s) across an empty (0 valued) array location. Other variations can include rotations of segments in a 2 dimensional array or adding and removing elements. As with all other aspects of the genetic algorithm search method, the individual problem should be examined to determine all relevant and possible mutation types.

Once the population has been selected, placed into the intermediate generation, crossed over, and mutated, the newly generated population is then used as the initial population in the next iteration. This process continues until there is no longer any improvement on the best solution, but should always be examined afterwards by downhill exploration to make sure that the solution found is in fact a local minimum.

#### **2.4.6 Hybrid techniques**

It is often the case that searches of different types (i.e. downhill and genetic algorithm) can be combined to form hybrid techniques [33, 34], which can greatly improve results as well as the rate of convergence. The simplest construction of a hybrid method is to take several downhill steps after the mutation stage of a genetic algorithm, and can greatly improve the trajectory of the fitness distribution for a population. These methods are particularly effective when there are many local minima that are similar in magnitude but not equal, allowing for the minimum of each sampled region to be compared with each other. Depending on the problem, hybrid methods can reduce the randomization generated during mutation, so if a large number of downhill steps are taken, it is advisable to increase mutation rates accordingly, or invert the order and perform the downhill steps first.

#### **2.4.7 Reset techniques**

All of the above methods are capable of converging to a solution or neighborhood from which no further improvements can be made. Usually, such convergence is considered

a sign that a global minimum has been found, but this is not always the case. When convergence is reached, reset techniques [12] can be a helpful explorative tool for determining whether further optimization is still possible. The simplest reset technique to employ is the global reset, which is equivalent to starting the entire search over again from a new random starting point. When the ability to change one parameter is restricted by the values of other parameters, one helpful reset technique is particle replacement. This involves identifying the particle(s) that have moved the least in the optimization process and reassigning them to random locations. The optimization will then start over from this new configuration, and approach convergence once again, but with an improved chance of finding a more desirable solution than the solution before the particle was moved.

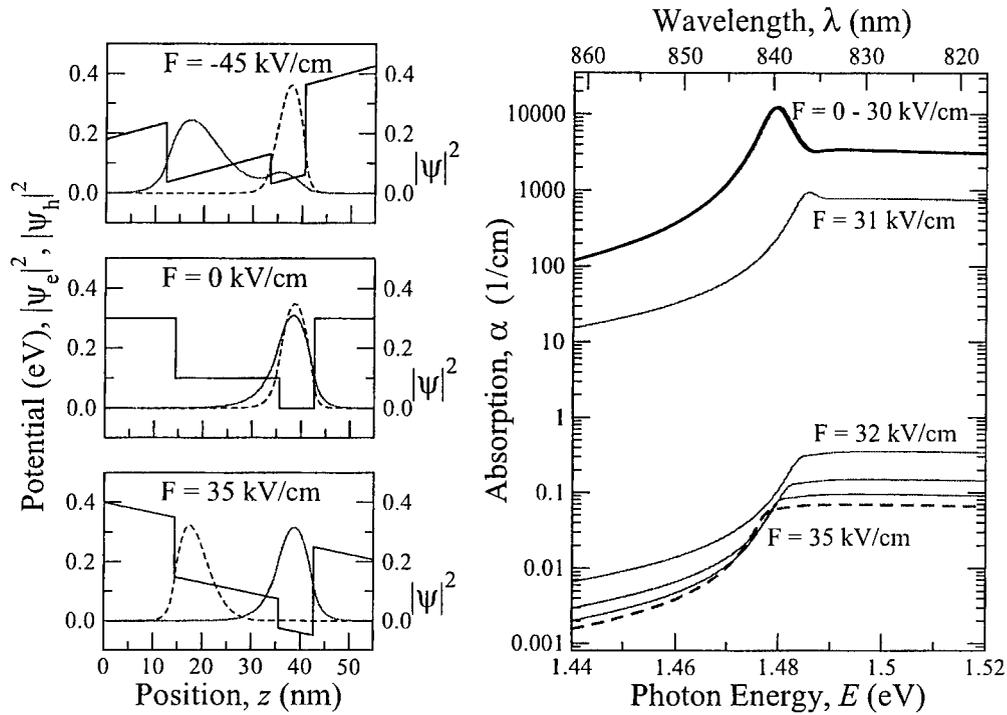
While the methods and techniques presented above do not cover all possible approaches or scenarios, they are to be considered the most likely candidates for a successful search when trying to solve a complex design problem such as those approached in the examples. If there is one resounding theme in these descriptions, it is that no one search parameter or scheme can be picked independently, but rather the search method should be developed in response to the projection of the parameters onto the fitness landscape for the problem at hand.

### **3. Optimization Without an Objective Function**

#### **3.1 Quantum well electro-absorption modulator**

The ionized hole solution of Section 2.2.3 met all the objective requirements, but displays an interesting behavior in addition to the frequency switching. When the hole ionizes out of the deeper sub-well, the electron-hole overlap diminishes exponentially, which results in a rapid decrease in absorption over a very small range of applied field. This effect shows promise as a high contrast, low chirp, intensity modulator. This section will focus on the development of this discovery into such a device.

Optoelectronic intensity modulators made from square wells tend to have low extinction ratios (usually in the range of 4 to 6). In light of this, solutions were sought for a new objective 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. 2.4, the rate at which the peak absorption decreases can be approximately measured by examining the fraction of  $|\psi_h|^2$  which overlaps the deeper sub well 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.



**Figure 3.1:** Broken-symmetry quantum well structure obtained by modifying the structure from Fig. 2.4 to fit a new objective 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.

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. 3.1, has a sub-well 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 sub-well shallower and wider than the result illustrated in Fig. 2.4, 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 sub-well, there is very little overlap with the electron wave function that remains localized within the sub-well. The ionization of the hole in this solution occurs at 30 kV/cm. The slope at the transition point is  $\theta = 0.27$  (kV/cm) $^{-1}$ , compared to  $\theta = 0.04$  (kV/cm) $^{-1}$  for the solution in Fig. 2.4.

Since this solution begins with both the electron and the hole localized within the sub-well, 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 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.

## 3.2 Motivation and methodology

During the search for an absorption frequency switch the functionality of an intensity modulator emerged without being deliberately sought [20, 21]. 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 compute power available today, an intelligent machine-based search may be developed to identify previously undiscovered useful solutions. To investigate this, a new exploration algorithm was implemented to sample solution space randomly and return a set of solutions it found "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.

The simplest implementation of this exploratory search is the binary test search, which examines random solutions and either saves them for human inspection, or ignores them, depending on whether or not the criteria for an interesting solution have been met. This has the advantage of being simple to implement, but in some cases can result in lists of interesting solutions that are too long to be examined individually by humans, and often will contain many similar or related interesting solutions. Another method for exploratory search is to construct a fitness function based on how interesting a solution is, and optimize to find the most interesting solutions possible. This search should be performed slightly differently from conventional searches like those listed in Section 2.4, however because in this case, while the lowest local minima is desired, other local minima should be saved for examination as well, because there may be more than one interesting functionality in the search space. For this reason, the results of many separate optimization efforts should be compounded, and random factors such as temperature or mutation rates should be slightly higher than when a single solution is sought.

### 3.2.1 Defining utility

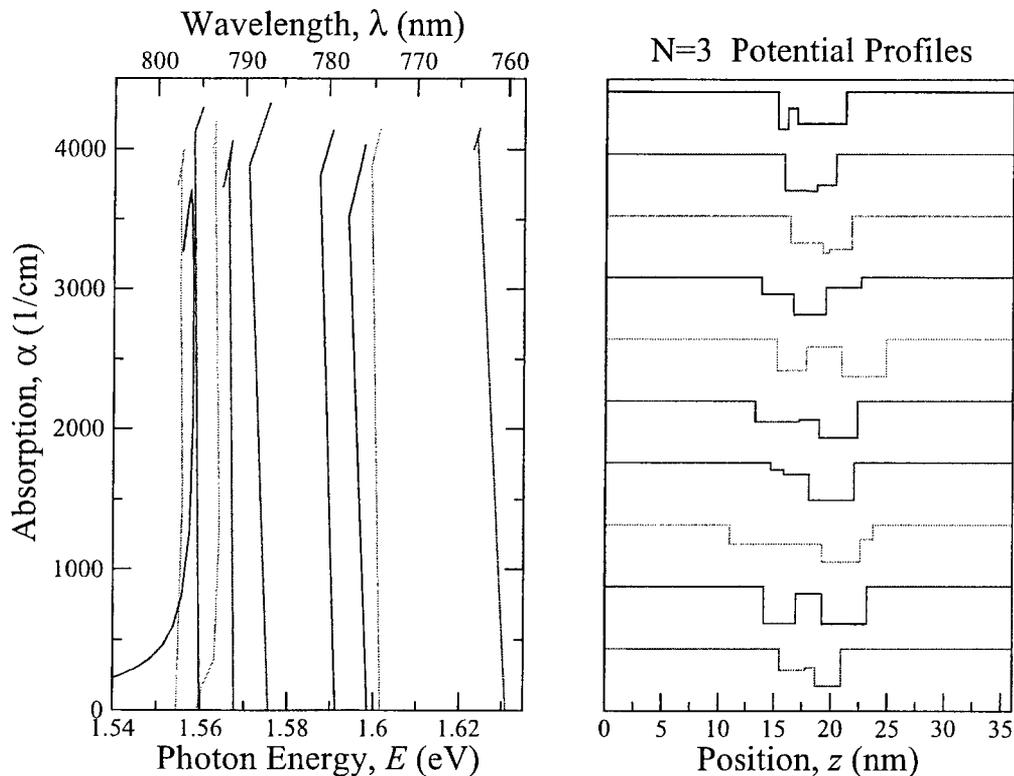
Before any objectiveless exploration can take place, the qualities that make a solution interesting must be enumerated in some way. While this process is particular to the physical model used, there are some commonalities to the recognition of interesting solutions, which can be addressed here.

In most cases, certain ranges of observables will make a solution useless for any functionality, and should not be considered. For the physical model used in Sections 2.2.2 and 2.2.3, any solution that has a low absorption for the entire range of applied field values is effectively transparent and uninteresting. For this reason, in order for a solution to be considered interesting in this scenario, it should have a high absorption for at least one applied field value.

When the physical model includes a control parameter in addition to the physical input parameters, such as the applied field in Sections 2.2.2 and 2.2.3, the amplitude of the response to the control parameter is often a good indication of the usefulness of a particular solution. If a solution is unresponsive to the control parameter, it is static and

presumable uninteresting, but if there is a large response to a small change in the control parameter, there is a significant chance that the responsive solution could function as a useful device.

An initial binary exploration was performed using this physical model where interesting solutions were defined to be those that traversed large distances in either the  $E$  or  $\alpha$  directions of the  $E - \alpha(E)$  plane. A minimum  $E$  distance of 0.001 eV (with nonvanishing absorption) or a minimum  $\alpha$  distance of 12,000 1/cm would suffice to label the solution "interesting". Reports of all such solutions (as in Fig. 3.2) were then stored for further



**Figure 3.2:** Band-edge potential profiles and the corresponding absorption peak paths selected by a machine-based search of configuration space without a human-specified objective function. Solutions are sorted from a random set by a binary test for properties that characterize interesting absorption peak paths. In this case, an interesting path is one with a large overall displacement in the  $E$ - $\alpha$  plane. Absorption peak paths are tracked as the external applied field increases from 0 kV/cm to 125 kV/cm.

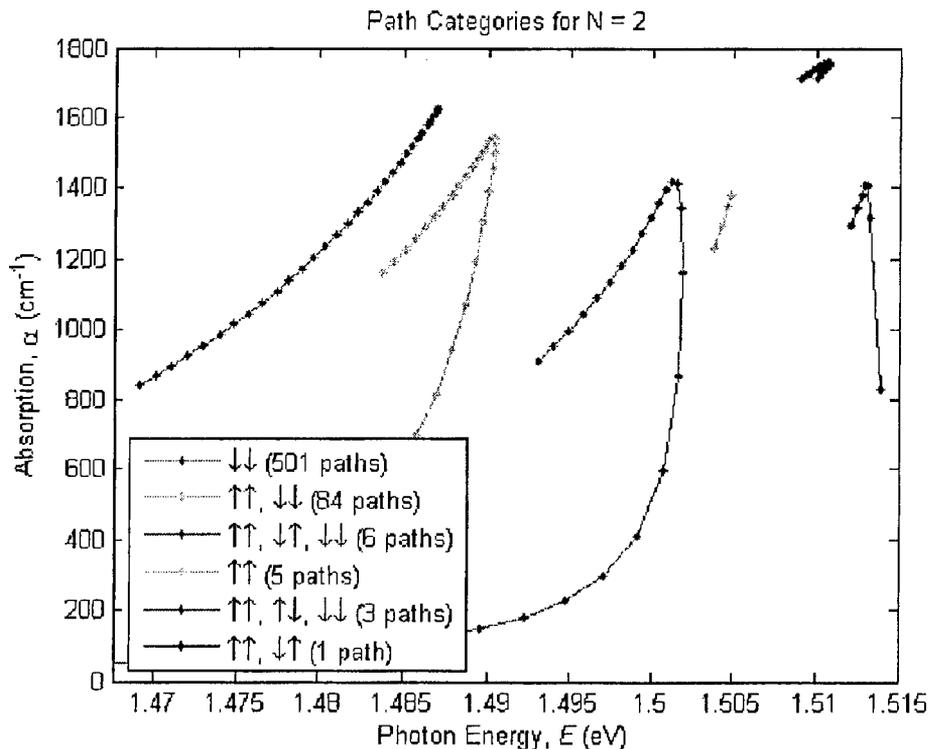
Some other characteristics to consider are symmetries, densities, and transitions within the observables.

### 3.2.2 Classifying solutions

A key problem faced by both binary and fitness-based exploratory searches is that of duplicate or near-duplicate solution reporting. This problem occurs when there are a large number of points in parameter space with similar results in the space of

observables, allowing for a long list of interesting solutions that differ from each other by only a small amount. Since the purpose of the search is to identify unique candidates for new devices, variations on a single functionality only serve to congest the list of solutions. It is for this reason that the classification of solutions by type is essential to keep the list of solutions manageable.

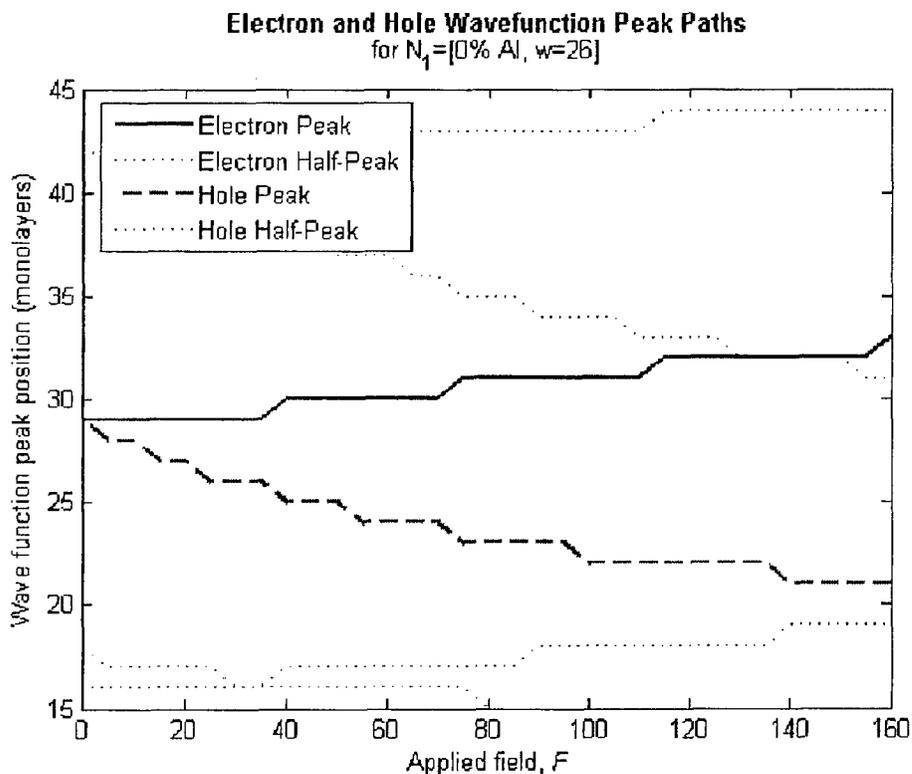
The simplest method of solution classification is to mirror the implementation of the definition of the utility of the solution. In the case of the quantum well structures with an applied electric field, a large change in exciton peak position or in exciton peak amplitude in response to a small change in applied field implied an interesting device. Classification of solutions in this physical model then naturally fell into groups based on the response of these two variables to the applied field, some samples of which can be seen in Fig. 3.3 for structures composed of two equi-potential regions ( $N = 2$ ).



**Figure 3.3:** Various categories for exciton absorption peak paths exhaustively sampled for the case of quantum well structures generated from two equipotential regions ( $N=2$ ). The first arrow in the legend indicates the sign of the first derivative of absorption,  $\alpha$ , with respect to applied field,  $F$ , and the second arrow indicated the sign of the first derivative of photon energy,  $E$ , with respect to  $F$ . Comma separated sets of arrows represent consecutive trajectory segments with different first derivatives. For  $N = 1$ , there is only 1 available category, for  $N = 2$ , there are 7, and for  $N = 3$ , there are 11.

Another method for categorizing solutions is to isolate similarities and differences in the physical quantities that the observables are constructed from. This is mathematically equivalent to the method illustrated above, but has the advantage of providing intuitive explanations for the differences between functionality categories. Keeping with the

absorption peak trajectories, the physical quantities which generate the absorption peak position for a given applied field value are the confinement energies for the electron and hole, the exciton binding energy, and the value of the electron and hole wave functions. An example plot of the electron and hole wave function peak positions for a conventional well structure ( $N = 1$ ) follows in Fig. 3.4.

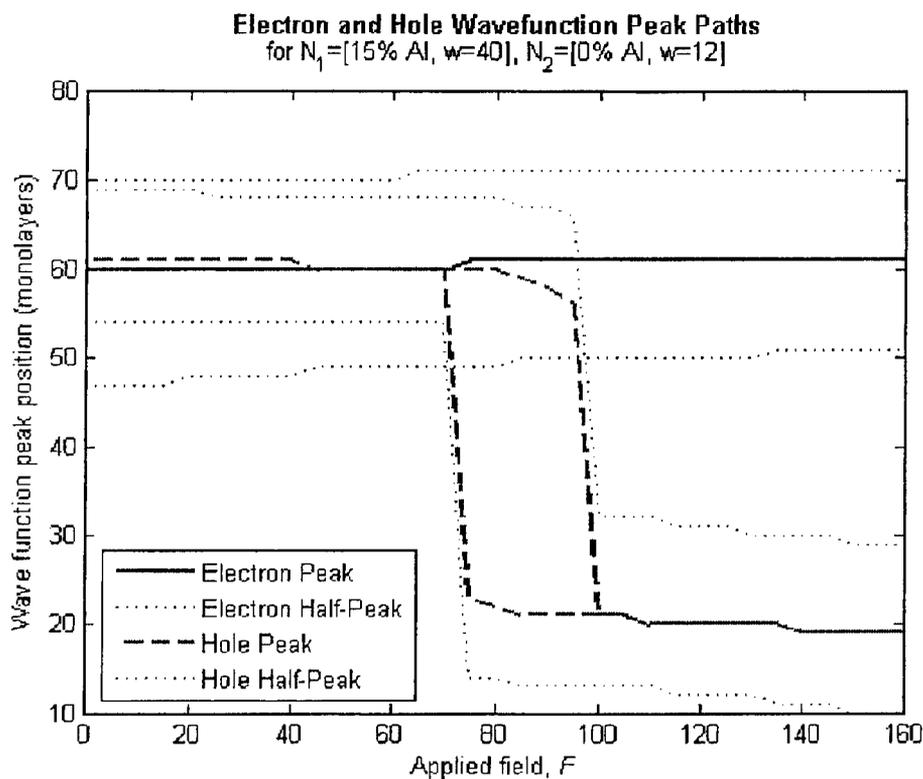


**Figure 3.4:** The paths of the electron and hole wave function peaks as a function of applied field,  $F$ , for a square well 26 monolayers wide. Dark lines represent the peak of the wave function, and thin lines represent the point at which the wave function is equal to half the value at the peak. For  $N = 1$ , all structures begin with the electron and hole located together and then they steadily separate as in this graph.

For all conventional well structures ( $N = 1$ ), the wave function paths are similar. The electron and hole start at the same location, and then separate as the applied field is increased, as in Fig. 3.4. For  $N = 2$ , there are 7 available categories, 6 of which begin with a segment of increasing absorption rather than decreasing. This difference can be understood by examining the electron and hole paths for structures in these categories. As a result of the asymmetry in the well structure, the hole center of mass is on the lower potential side of the electron center of mass, and as the two peaks approach each other, absorption increases, much like for the structure in Fig. 2.4.

Another interesting property that arises from the asymmetry of  $N = 2$  structures is tunneling. This appears as a momentary double peak and a sudden change in position on the electron-hole path graph, but is not distinguished in the original first derivative trajectory classification system. Tunneling does not always imply new functionality,

though, because if the electron-hole overlap is low, and the tunneling process results in lower overlap, there is little or no effect on the absorption peak trajectory. The identification of effective tunneling processes can be incorporated by checking for a sudden increase or decrease in absorption, and doing so further subdivides the two largest functionality categories in two. An example structure can be seen in Fig. 3.5 which demonstrates asymmetric starting points and tunneling.

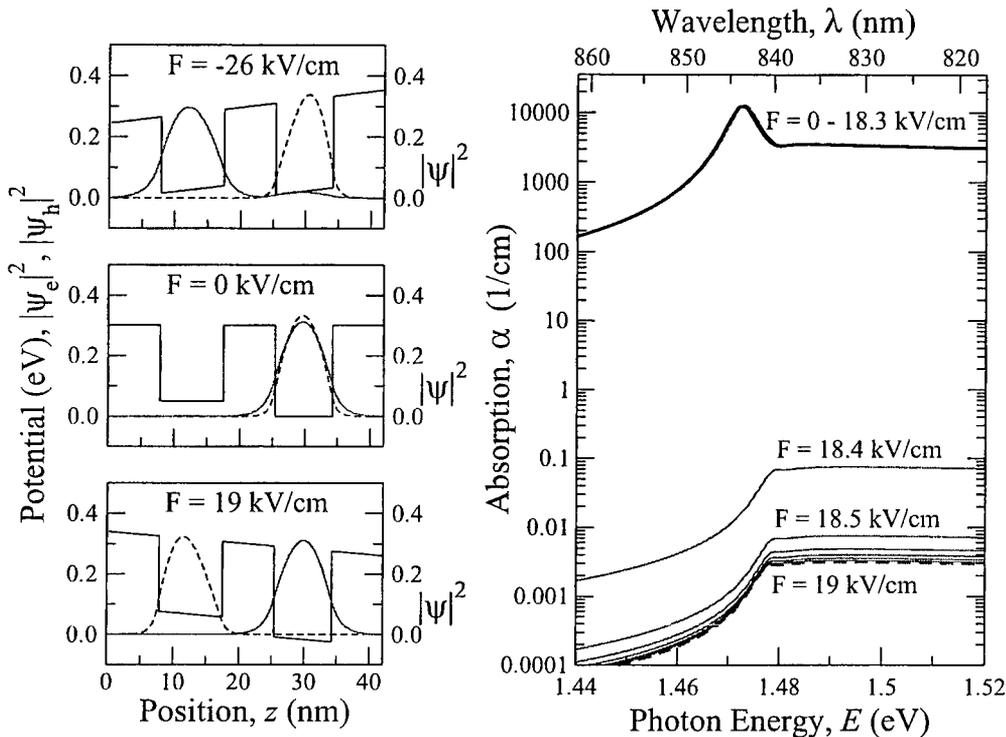


**Figure 3.5:** The paths of the electron and hole wave function peaks as a function of applied field,  $F$ , for a step well 52 monolayers wide, with a 40 monolayer step of 15% Al. The electron and hole peak positions are initially separated, but as the applied field increases, their two positions approach each other, and then the hole rapidly tunnels out of the sub-well and into the step region.

### 3.3 Double well electro-absorption modulator

Drawing from the results of the machine-based exploration in Fig. 3.2, an electro-absorption modulator was designed to better separate the electron and hole after ionization, and to better increase overlap before ionization. The solution in Fig. 3.6 consists of two narrow wells of varying potential with a narrow barrier between them. The extinction ratio of this solution is better than that of the well/sub-well solution in Fig. 2.14 by a factor of 100. While this solution looks quite different in form, it works on all the same principles as the well/sub-well solution, and can be generated from the previous solution by adding a small barrier at the well/sub-well 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.



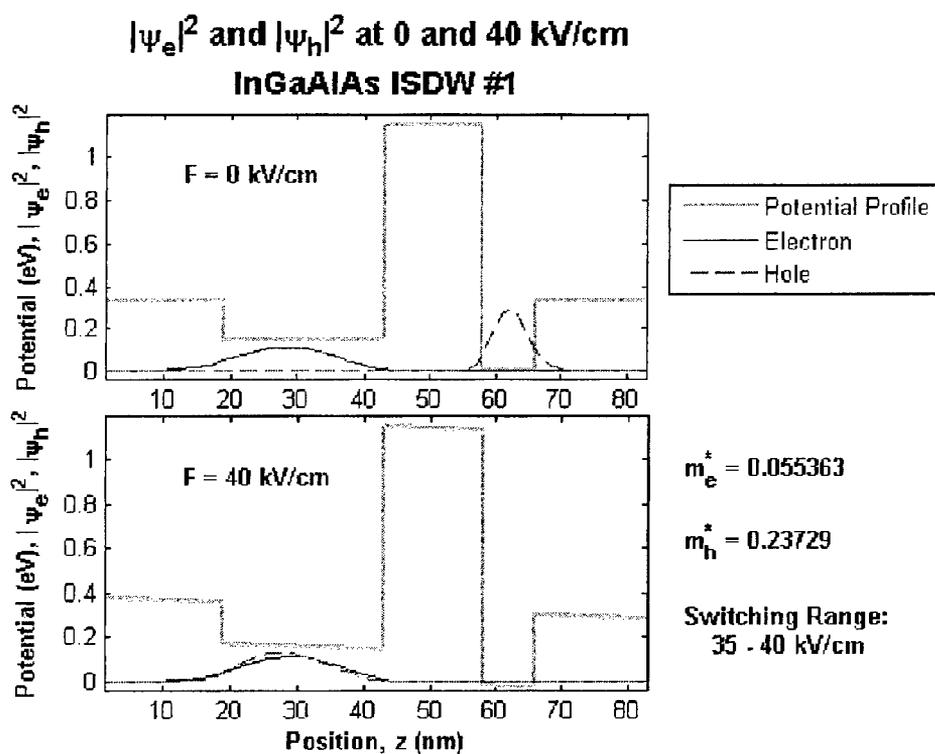
**Figure 3.6:** Asymmetric double quantum well intensity modulator discovered by machine-based exploration without human specification of an objective function. Absorption spectra are shown in increments of 0.1 kV/cm.

### 3.4 Initially separated double well solution

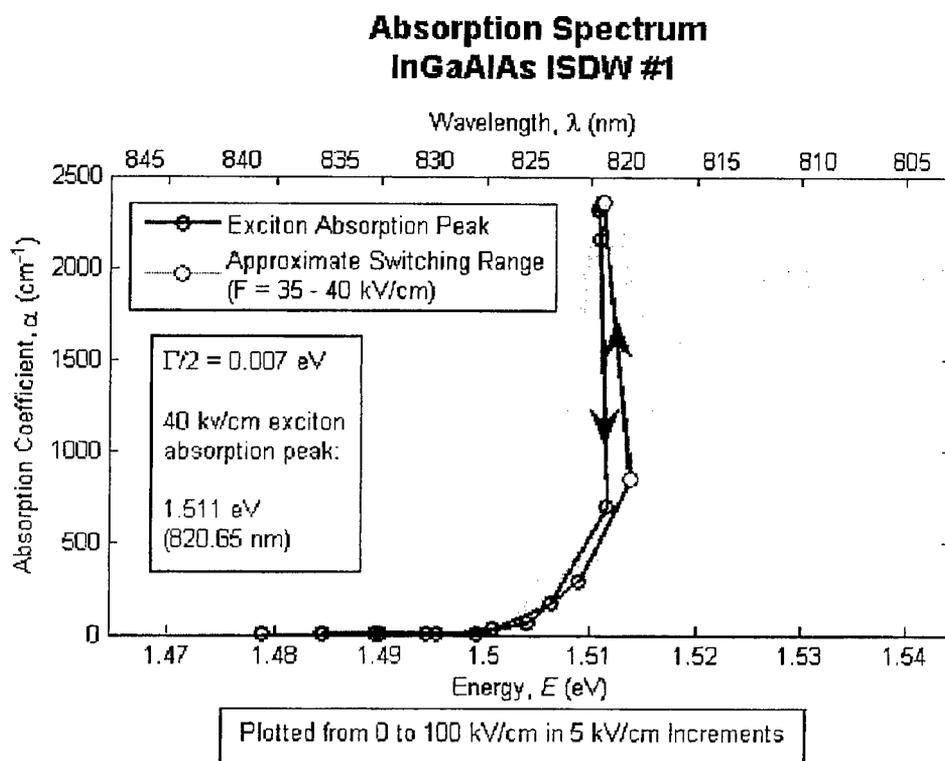
During the categorization of trajectories as in Fig. 3.3 – 3.5, it became apparent that since a high-speed electroabsorption modulator could be built from a trajectory with rapid decrease in absorption due to tunneling, a similar device could be designed from a trajectory which rapidly increases instead. In order to achieve this, the electron and hole wave functions must be initially localized in separate regions of the structure, but once a voltage is applied, the hole would tunnel into the same region as the electron, rapidly increasing the overlap of the two wave functions. This process would be advantageous over previous designs not only because the shifting of the continuum towards longer wavelengths will contribute to a high extinction ratio rather than detract from it, but also because carrier extraction is more efficient at high voltages, making it more beneficial to have high absorption coupled with high applied voltages.

Initial separation of the electron and hole wave functions requires at least one narrow sub-well region, typically less than 15 monolayers in an AlGaAs system, in order to exploit the asymmetry between the electron and hole effective masses. In order to isolate the electron and hole wave functions as much as possible before tunneling, a barrier region is also needed. These requirements can both be realized with a structure composed of only 3 equipotential regions ( $N=3$ ). The first region is a wide well where

the electron is localized, the second region is a barrier, and the third region is a narrow well with a lower potential than the first region, which localizes the hole, but which is also narrow enough that the lower potential in the region is much less than the energy required to confine the broad electron wave function. In order to accentuate this effect as much as possible,  $\text{In}_{0.27}\text{Ga}_{0.73}\text{As}$  was used for the 7 monolayer deeper sub-well region, which not only has a lower band gap (1.15 eV) than GaAs (1.42 eV), but is also a strained material, introducing further electron-hole asymmetry. For the 15 monolayer central barrier region, AlAs was used for its high band gap (2.17 eV), and for the wide well region, a conventional 24 monolayer GaAs well was used. The final initially separated double well (ISDW) design can be seen with the electron and hole probability distributions in Fig. 3.7, and the absorption spectrum can be seen in Fig. 3.8.



**Figure 3.7:** Electron and hole probability densities for the ISDW solution. At 0 kV/cm, the electron and hole wave functions are localized in separate wells, due to the difference in electron and hole effective masses. With the application of an applied field, the hole wave function tunnels into the wider well, rapidly increasing the exciton absorption peak strength, which is demonstrated in Fig. 3.8.



**Figure 3.8:** Absorption spectrum for the ISDW solution. At 0 kV/cm, the electron and hole are separated by an AlAs barrier, and there is no discernable exciton peak. As the hole tunnels into the GaAs barrier, the exciton peak rises, and at even higher applied field values, the electron tunnels into the  $\text{In}_{0.27}\text{Ga}_{0.73}\text{As}$  well, again reducing the exciton peak.

## 4. Experimental Results and Analysis

### 4.1 Incorporating fabrication into the model

This Section discusses experimental results for the electro-absorption modulator. While the model used is relatively simple, some pragmatic adjustments were required to better aide in the development of physically realizable devices. For instance, the relative strength of the exciton peak and the continuum varies with the smoothness of the interface between barrier and well layers [35, 36, 37], and that smoothness depends greatly on the growth conditions. For this reason the half-linewidth,  $\Gamma/2$ , and the relative strength of the exciton peak were fit to calibration data rather than calculated from material properties alone. It was also found that the Lorentzian distribution presented with the model in the literature does not fit experimental data as well as a Gaussian, so the model was adjusted to support both peak types.

The model was also modified to simulate not only AlGaAs systems, but any combination of the binary compounds GaAs, AlAs, InAs, GaP, and InP, with any binary used as the substrate. Including these new materials requires the use of interpolation of material properties [38, 39, 40, 41], as well as the inclusion of strain effects [42, 43], which are caused by a mismatch in lattice constants. Strain effects are difficult to measure precisely in a laboratory setting, however, and the best interpolative models have been

shown to provide inaccurate results in some cases [39], so the interpolations used in our model should not be trusted to provide accurate predictions without calibration.

#### 4.1.1 Poisson solver

The well structures in electroabsorption modulators are contained inside the intrinsic region of a PIN diode, which allows for the applied electric field to be adjusted linearly with bias voltage. Provided the background doping levels are low enough, the electric field will remain relatively constant across the multiple quantum well structures. The control parameter for this PIN device is the bias voltage,  $V_{\text{bias}}$ , which generates the applied field,  $F$ , that appears in the physical model. In order to relate our predicted behavior to measured modulation, it is necessary to include a Poisson equation solver for the entire PIN structure.

The Poisson solver [44, 45] calculates the band edge of the entire pin structure by iteratively finding a self consistent solution to the equation

$$\Delta^2 \Phi(z) = -\frac{\rho(\Phi(z), N_a(z), N_d(z))}{\epsilon_0 \epsilon_r}, \quad (4.1)$$

where  $\Phi(z)$  is the potential along the  $z$  direction,  $\rho(\Phi(z), N_a(z), N_d(z))$  is the charge density, and  $N_a$  and  $N_d$  are the doping concentrations. The charge density,  $\rho$ , can be solved numerically using a finite-difference Schrödinger equation solver [44], or by using a Taylor expansion of the exponential representation including the chemical potential at either end of the diode [45, 46],

$$\rho(z) = e(N_d(z) - N_a(z) - n_c(z) + p_v(z)), \quad (4.2)$$

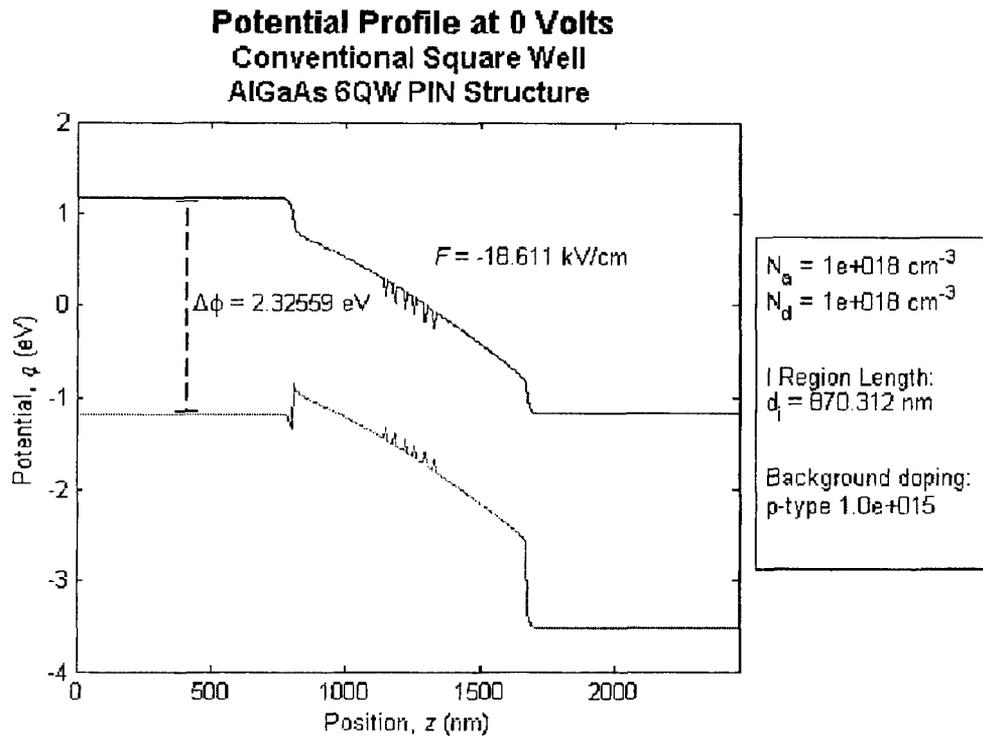
where  $e$  is the electron charge and

$$n_c(z) = N_d e^{-e(\Phi(z=z_{\text{max}}) - \Phi(z))/k_B T} \quad (4.3)$$

$$p_v(z) = N_a e^{-e(\Phi(z) - \Phi(z=0))/k_B T}$$

and  $z$  is defined such that the P side of the structure begins at  $z = 0$ .

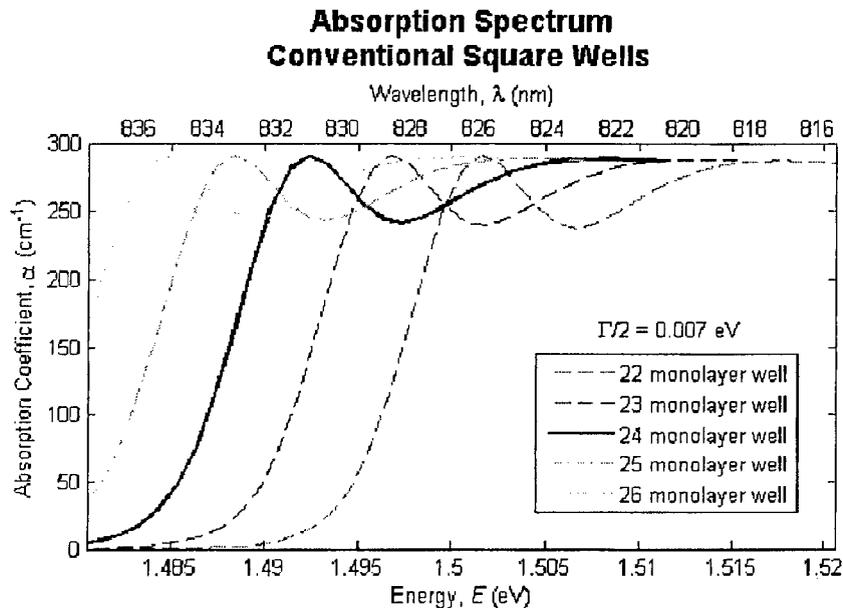
The final result of this simulation is that for a given structure, the inherent electric field when  $V_{\text{bias}} = 0$  V can be calculated, as well as the linear increase in electric field,  $F$ , as a function of  $V_{\text{bias}}$ . A sample plot of the conduction and valence band edges at  $V_{\text{bias}} = 0$  can be seen in Fig. 4.1.



**Figure 4.1:** The conduction and valence band edge profiles for a 6 quantum well AlGaAs/GaAs electro-absorption modulator grown in a PIN layer structure. At  $V_{\text{bias}} = 0\text{V}$ , the electric field across the quantum well structures is  $-18.6 \text{ kV/cm}$ , and for each volt applied, the field increases in magnitude by  $12 \text{ kV/cm}$ .

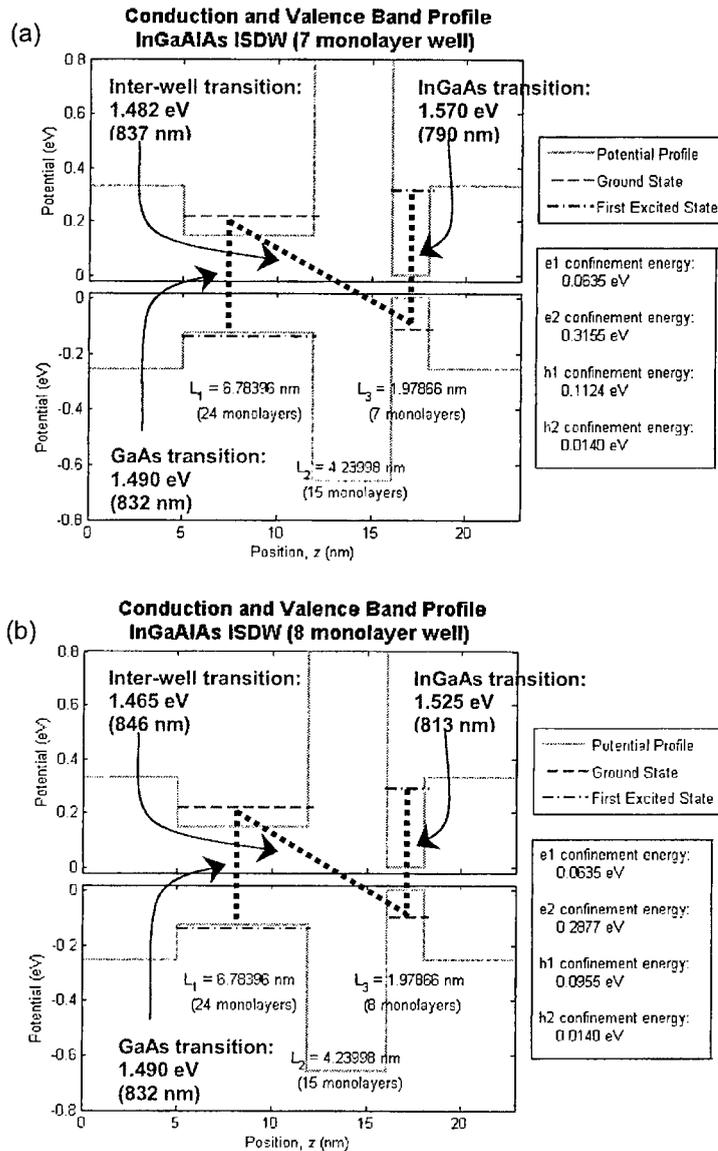
#### 4.1.2 Sensitivity analysis for inaccuracies in well thickness

While minor perturbations in potential and smoothness at well-barrier interfaces are included by broadening  $\Gamma/2$ , the half-linewidth, inaccuracies in quantum well thickness leads to a shifting in the energy of the exciton absorption peak, and must be considered separately. This occurs because of the shift in electron-hole confinement energies, and the magnitude of this shift decreases for wider well structures. A sample well-thickness analysis for a 24 monolayer  $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$  conventional well structure can be seen in Fig. 4.2.



**Figure 4.2:** Absorption spectra plotted at  $F = 0$  kV/cm for conventional  $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}/\text{GaAs}$  structures of varying well widths. A well fabricated two monolayers too thick shows a shift in absorption peak energy of  $\Delta E = -7$  meV, while a well fabricated two monolayers too thin shifts by  $\Delta E = +10$  meV.

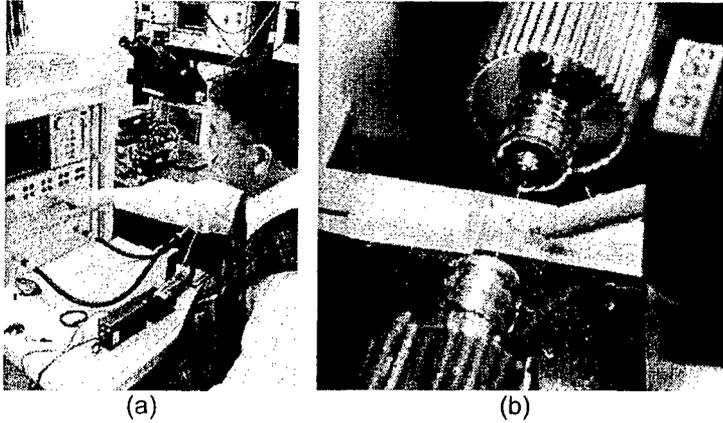
The well structures grown for experimental verification were all 24 monolayers or larger in size, in order to avoid large unexpected shifts in absorption energy, with the *exception* of the 7 monolayer narrow well in the ISDW structure of Section 3.4 and Fig. 3.7. At 7 monolayers, growing the structure only one monolayer too thick results in a transition energy shift of  $-45$  meV. The device still functions robustly, however, because ideally there is no absorption from transitions inside the narrow well. The positions of energy levels inside the 7 monolayer well dictate the applied field values necessary to switch the absorption on and off rather than the absorption energy, and appreciable absorption only occurs when both wave functions are localized in the wider GaAs well. In Fig. 4.3, a comparison between the transition energies can be seen for both 7 and 8 monolayer well structures, including the GaAs well transition which remains completely unchanged by the error.



**Figure 4.3:** Insensitivity of ISDW structure to monolayer fluctuations, despite narrow well width. The transition energies can be seen for 7 (a) and 8 (b) monolayer structures, but only the robust GaAs transition has a significant overlap integral. The energy levels in the narrow well control the voltage at which the exciton peak switches on, but otherwise have little bearing on the device.

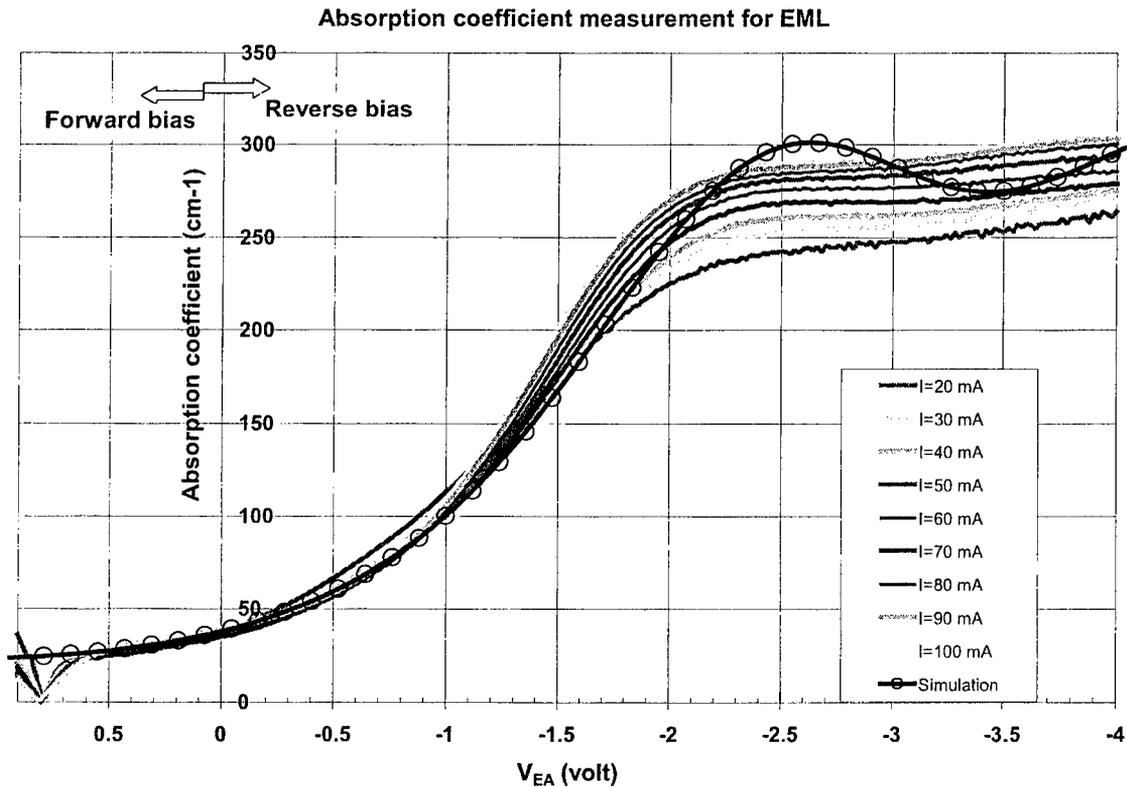
## 4.2 Experimental results

Experimental efforts included structure growth at three separate facilities, using both Metalorganic Chemical Vapor Deposition (MOCVD). Structures were grown on InP substrates for conventional, step-well, and initially separated double well (ISDW) designs. Measurements of these devices were performed using a tunable laser to sweep wavelength over the absorption spectrum near the band edge, or using a fixed wavelength integrated laser grown adjacent to the modulator section. Figure 4.4 shows a photograph of the measurement equipment.



**Figure 4.4:** (a) High speed test equipment for measuring the RF response of electro-absorption modulators. Calibrated small signal opto-electronic RF measurements can be performed up to frequencies in excess of 40 GHz. (b) Lensed single-mode fiber used to couple into and out of the modulator chip. High-speed RF probe is on the right-hand-side of the photograph.

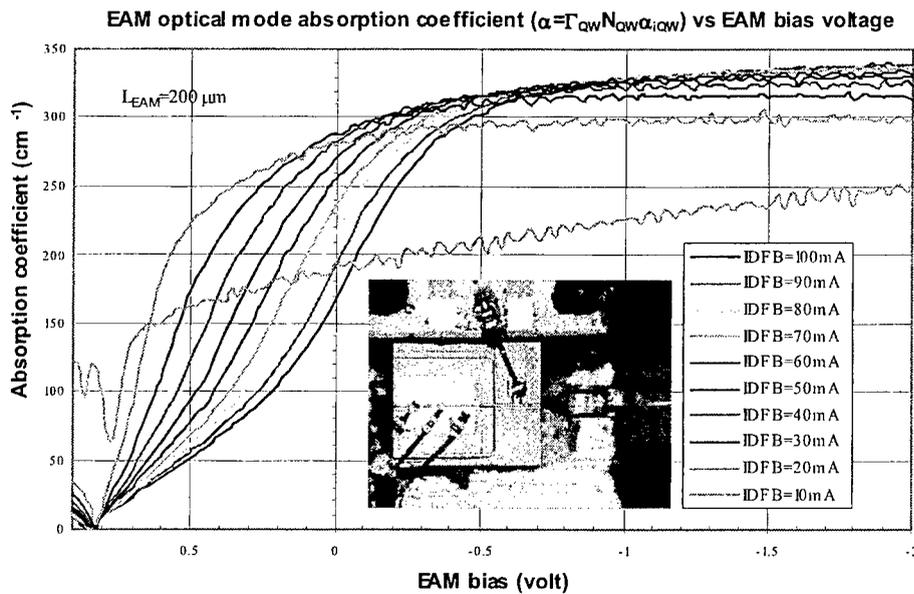
#### 4.2.1 MOCVD grown InGaAsP modulators



**Figure 4.5:** Measured and simulated absorption coefficients at 1550 nm wavelength for a conventional square well device with integrated laser. Material composition is  $\text{In}_{0.74}\text{Ga}_{0.26}\text{As}_{0.73}\text{P}_{0.27}$  for the well and  $\text{In}_{0.75}\text{Ga}_{0.25}\text{As}_{0.49}\text{P}_{0.51}$  for the barrier. The well width is 6.5 nm and the band gaps for these materials are 0.786 eV and 0.992 eV respectively, and the half-linewidth is  $\Gamma/2 = 0.024$  eV.

The first structure tested was an InGaAsP conventional QCSE electro-absorption modulator (EAM) with a DFB laser placed adjacent to the modulator and integrated on the same chip. The absorption spectrum for this structure could not be measured as a function of wavelength, but the intensity of the modulated signal was monitored as a function of bias voltage at 1550 nm wavelength. The measured absorption coefficient and the simulation results fit to the data are shown in Fig. 4.5. As can be seen the agreement between calculation and experiment is good. Material composition is  $\text{In}_{0.74}\text{Ga}_{0.26}\text{As}_{0.73}\text{P}_{0.27}$  for the well and  $\text{In}_{0.75}\text{Ga}_{0.25}\text{As}_{0.49}\text{P}_{0.51}$  for the barrier. The well width is 6.5 nm and the band gaps for these materials are 0.786 eV and 0.992 eV respectively, and the half-linewidth is  $\Gamma/2 = 0.024$  eV.

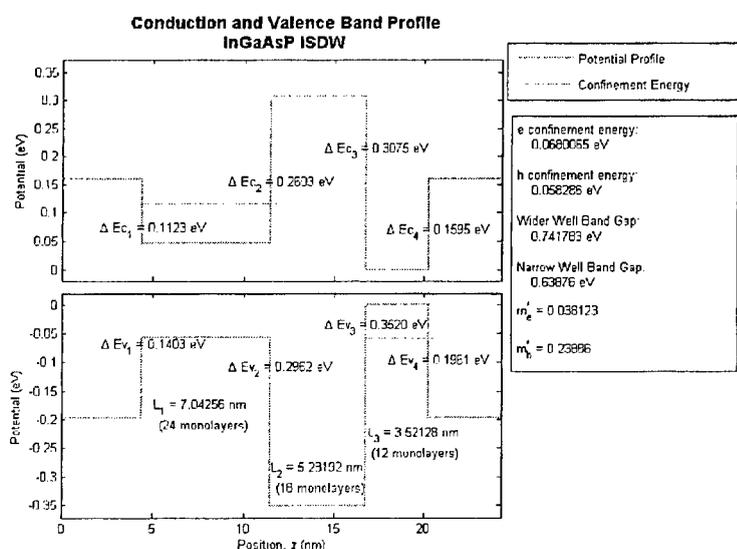
The next two devices grown were InGaAsP step-well structures similar to the AlGaAs structure in Fig. 3.1. However, because the modulator was integrated with a DFB laser, modulation could only be monitored at a single wavelength, rather than observing the entire spectrum. Figure 4.6 shows measured absorption as a function of  $V_{\text{bias}}$  at 1530 nm wavelength for the InGaAsP step-well EAM integrated with a DFB. The slope of the modulation curve is steep compared to the response of the device shown in Fig. 4.5. The inset in the figure is a photograph of the integrated DFB and EAM showing wire bonds and single-mode lensed fiber coupling.



**Figure 4.6:** Measured absorption as a function of  $V_{\text{bias}}$  at 1530 nm wavelength for InGaAsP step-well. The slope of the modulation curve is steep compared to the response of the device shown in Fig. 4.5. Inset is a photograph of the integrated DFB and EAM showing wire bonds and single-mode lensed fiber coupling.

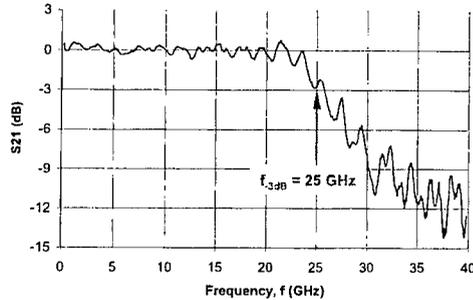
When  $V_{\text{bias}}$  is near 0 V, carriers cannot be efficiently extracted from the PIN structure, and when  $V_{\text{bias}}$  is in forward bias, carriers are actually being pumped in to the structure, filling low lying energy states, and making the creation of an electron-hole pair at the

band edge more difficult. In addition, the first excited state for the hole before ionization lies at the far end of the step, having only negligible overlap with the localized electron. For this reason, the next transition capable of contributing to the absorption before ionization is the second excited state, which has little effect when observing a band-edge wavelength. This explanation supports the measured data, but cannot be quantitatively evaluated because it is not possible to determine the degree to which the low-lying states have been filled inside the device. For these reasons, most of our subsequent efforts were spent on devices that switch from low absorption to high absorption and work with the effects of carrier extraction and continuum shifting rather than against them.



**Figure 4.7:** Conduction and valence band profiles for InGaAsP ISDW structure. Material composition is  $\text{In}_{0.75}\text{Ga}_{0.25}\text{As}_{0.49}\text{P}_{0.51}$  (tensile strain) for the outside barriers,  $\text{In}_{0.52}\text{Ga}_{0.48}\text{As}$  (unstrained) for the wider well,  $\text{In}_{0.97}\text{Ga}_{0.03}\text{As}_{0.06}\text{P}_{0.94}$  (unstrained) for the high inner barrier, and  $\text{In}_{0.67}\text{Ga}_{0.33}\text{As}$  (compressive strain) for the narrow well.

The next device grown and fabricated in the InGaAsP material system was an initially separated double well (ISDW) structure similar to the structure in Fig. 3.7. Because of the low potential barriers inherent in the InGaAsP system and the long minimum layer thicknesses required using MOCVD growth (4.4 nm), strain effects were exploited in addition to effective mass asymmetries to separate the electron and hole wave functions. High compressive strain in the narrow well shifts both the valence and conduction bands towards higher energy, localizing the hole in the narrow well and the electron in the wider well. The conduction band edge and valence band edge profiles are shown in Fig. 4.7. Care has to be taken with ISDW structures to manage strain in the quantum wells. The material composition is  $\text{In}_{0.75}\text{Ga}_{0.25}\text{As}_{0.49}\text{P}_{0.51}$  (tensile strain) for the outside barriers,  $\text{In}_{0.52}\text{Ga}_{0.48}\text{As}$  (unstrained) for the wider well,  $\text{In}_{0.97}\text{Ga}_{0.03}\text{As}_{0.06}\text{P}_{0.94}$  (unstrained) for the high inner barrier, and  $\text{In}_{0.67}\text{Ga}_{0.33}\text{As}$  (compressive strain) for the narrow well.



(a)

(b)

**Figure 4.8:** (a) Measured small-signal response of a 150  $\mu\text{m}$  long electro-absorption modulator. The  $-3$  dB frequency is measured to be  $f_{-3\text{dB}} = 25$  GHz and is limited by the RF design of the modulator chip. The input laser wavelength is  $\lambda_0 = 1545$  nm, optical input power is 6 dBm, RF input power is 0 dBm, and the measured DC optical extinction ratio of the device is 10 dB / V. (b) SEM image of processed devices prior to separation into single chips.

Figure 4.8(a) shows measured small signal response of a 150  $\mu\text{m}$  long buried heterostructure electro-absorption modulator device and Fig. 4.8(b) is a SEM image of processed devices prior to separation into single chips. The  $-3$  dB frequency is measured to be  $f_{-3\text{dB}} = 25$  GHz and is limited by the RF design of the modulator. The input laser wavelength is  $\lambda_0 = 1545$  nm, optical input power is 6 dBm, RF input power is 0 dBm, and the measured DC optical extinction ratio of the device is 10 dB with a drive voltage of 1  $V_{\text{pp}}$ . Improved RF design, including the use of a traveling wave structure would be expected to increase the value of  $f_{-3\text{dB}}$  and the extinction ratio. At present, the physical model of the device does not include these aspects of RF design and so were not part of the optimization procedure. The performance of this device is comparable to the best commercially available electro-absorption modulator from Oki Electric that has a 13 dB extinction ratio with a 2  $V_{\text{pp}}$  drive voltage and a measured small signal frequency response of  $f_{-3\text{dB}} = 33$  GHz.

## 5. Conclusions

A systematic approach towards optimal design of optoelectronic semiconductor devices has been developed. The adaptive design of excitonic absorption in broken-symmetry quantum wells has been used as a prototype system with which to explore the device synthesis method.

The device model used captured the basic exciton absorption physics but did not include RF design. The model used a tight binding model in the z-direction and assumed a 2D hydrogenic wave function in the x-y plane. The calculation of absorption focused primarily on the lowest energy exciton peak, neglecting many other spectrum characteristics. These simplifications allowed a relatively fast forward-solve time (typically 10 s on a 2.8 GHz Intel Pentium machine), at the expense of some accuracy

in the absorption spectrum. This approach is justified because inhomogeneous broadening due to fabrication inaccuracies is typically large. In the improved growth capabilities may allow more complicated structures that may require improvements in the physical model. Specific improvements to consider include multiple exciton states and the Sommerfeld enhancement factor in absorption.

Electro-absorption modulator designs were investigated experimentally and a modulator with 25 GHz small-signal  $-3$  dB response was measured near 1550 nm wavelength. This device has performance comparable to the best commercially available modulator from Oki electric. Further improvements to device performance could be achieved by improving RF design, in particular, incorporation of traveling wave electrodes. To become part of the semiconductor device synthesis procedure the physical model of the device should incorporate these and other aspects of RF design.

## References

- [1] T.M. Wallis, N. Nilius, and W. Ho, *Phys. Rev. Lett.* **89**, 236802 (2002).
- [2] L. Bartels, G. Meyer, and K.-H. Rieder, *Phys. Rev. Lett.* **79**, 697–700 (1997).
- [3] D. G. Grier, *Nature (London)* **424**, 810-816 (2003).
- [4] M. Yoshita, H. Akiyama, L. Pfeiffer, and K. West, *Appl. Phys. Lett.* **81**, 49 (2002).
- [5] J. Thalken, Y. Chen, A. F. J. Levi, and S. Haas, *Phys. Rev. B* **69**, 195410 (2004).
- [6] 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).
- [7] 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).
- [8] X. Huang, A. J. Seeds, J. S. Roberts, and A. P. Knights, *IEEE Photonics Technol. Lett.* **10**, 1697 (1998).
- [9] P. J. Mares and S. L. Chuang, *J. Appl. Phys.* **74**, 1388 (1993).
- [10] S. L. Chuang, "Physics of Optoelectronic Devices" (Wiley, New York, 1995).
- [11] C. Y-P. Chao and S. L. Chuang, *Phys. Rev. B* **48**, 8210 (1993).
- [12] J. Thalken, Y. Chen, A. F. J. Levi, and S. Haas, *Phys. Rev. B* **69**, 195410 (2004).
- [13] J. A. Trezza, J. S. Powell, and J. S. Harris, *IEEE Photonics Technol. Lett.* **9**, 330 (1997).
- [14] N. Susa and T. Nakahara, *Appl. Phys. Lett.* **60**, 2324 (1992).
- [15] W. Chen and T. G. Andersson, *Semicond. Sci. Technol.* **7**, 828 (1992).
- [16] G. Ji, D. Huang, U. K. Reddy, H. Unlu, T. S. Henderson, and H. Morkoç, *J. Vac. Sci. Technol. B* **5**, 1346 (1987).
- [17] David E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning", Addison-Wesley, Boston (1989).
- [18] Adam Prügel-Bennett and Jonathan L. Shapiro, *Phys. Rev. Lett.* **72**, 1305 (1994).
- [19] C. M. Fonseca and P. J. Fleming, "Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization," in *Genetic Algorithms: Proceedings of the Fifth Int. Conf.*, 416 (1993).
- [20] Jason Thalken, Weifei Li, Stephan Haas, A.F.J. Levi, *Appl. Phys. Lett.* **85**, 121 (2004).
- [21] Jason Thalken, Stephan Haas, and A. F. J. Levi, *J. Appl. Phys.* **98**, 44508 (2005)
- [22] Melanie Mitchell, "An Introduction to Genetic Algorithms", MIT Press, Cambridge, Massachusetts (1996).
- [23] J. A. Holland and R. Mead, *Comput. J.* **7**, 308 (1965).
- [24] Dimitri P. Bertsekas, "Nonlinear Programming", Athena Scientific, Belmont, Massachusetts (1995).
- [25] R. Fletcher, "Practical Methods of Optimization", Wiley, New York (1987).
- [26] Philip Seliger, et. al., to be published in *J. Appl. Phys.* (2006)
- [27] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).
- [28] R. A. Rutenbar, *IEEE Circuits Dev. Mag.*, pp. 19-26, Jan. (1989).
- [29] A. Håkansson and J. Sánchez-Dehesa, *Opt. Express* **13**, 5440-5449 (2005).
- [30] J. Holland, *SIAM J. Comput.* **2**, pp. 88-105 (1973).
- [31] D. E. Goldberg and K. Deb, "A comparative analysis of selection schemes used in genetic algorithms," in Morgan Kaufmann's "Foundations of Genetic Algorithms", pp. 69-93, San Mateo, CA (1991).
- [32] B. R. Moon, Y. S. Lee, C. K. Kim, *Journal of Intelligent Manufacturing* **9**, 401 (1998).
- [33] J. Renders and S. Flasse, *IEEE Trans. Syst. Man Cybern.* **26**, pp. 243, (1999).
- [34] Mohamed B. Trabia, *J. Mech. Des.* **126**, 969 (2004).
- [35] K. Leosson, J. R. Jensen, W. Langbein, and J. M. Hvam, *Phys. Rev. B* **61**, 10322–10329 (2000).
- [36] M. Sugawara, T. Fujii, S. Yamazaki and K. Nakajima, *Appl. Phys. Lett.* **54**, 1353 (1989).
- [37] M. Yoshita, H. Akiyama, L. Pfeiffer, and K. West, *Appl. Phys. Lett.* **81**, 49 (2002).
- [38] T. Ishikawa and J. E. Bowers, *IEEE J. Quantum Electron.* **30**, 562 (1994).
- [39] J. Minch, S. H. Park, T. Keating, and S. L. Chuang, *IEEE J. Quantum Electron.* **35**, 771 (1999).
- [40] S. Adachi, *J. Appl. Phys.* **53**, 8775 (1982).
- [41] P. Lawaetz, *Phys. Rev. B* **4**, 3460 (1971).
- [42] D. J. Arent, K. Deneffe, C. Van Hoof, J. DeBoeck, and G. Borghs, *J. Appl. Phys.* **66**, 1739 (1989).

- [43] K. S. Lee, W. S. Han, J. S. Kim, B. Lee, J. Korean Phys. Soc. **37**, 587 (2000).
- [44] I-H. Tan, G.L. Snider, L.D. Chang, and E.L. Hu, J. Appl. Phys. **68**, 4071 (1990).
- [45] F.R. Shapiro, IEEE Transactions on Education, **38**, 380 (1995).
- [46] N. W. Ashcroft and N. D. Mermin, "Solid State Physics", Thomson Learning, Inc. (1976).
- [47] D. A. B. Miller, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, Phys. Rev. B **32**, 1043 (1985).