Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Ganesh Hegde

Entitled Generation and Optimization of Tight Binding Parameters Using Genetic Algorithms and Their Validation Using NEMO-3D

For the degree of Master of Science in Electrical and Computer Engineering

Is approved by the final examining committee:

G. Klimeck
Chair

M. J. Mclennan

S. Datta

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): G. Klimeck

Approved by: M. R. Melloch 4/21/10
Head of the Graduate Program Date

Graduate School Form 20
(Revised 10/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL

Research Integrity and Copyright Disclaimer

Title of Thesis/Dissertation: Generation and Optimization of Tight Binding Parameters Using Genetic Algorithms and Their Validation Using NEMO-3D

For the degree of Master of Science in Electrical and Computer Engineering

I certify that in the preparation of this thesis, I have observed the provisions of *Purdue University Executive Memorandum No. C-22,* September 6, 1991, *Policy on Integrity in Research.**

Further, I certify that this work is free of plagiarism and all materials appearing in this thesis/dissertation have been properly quoted and attributed.

I certify that all copyrighted material incorporated into this thesis/dissertation is in compliance with the United States' copyright law and that I have received written permission from the copyright owners for my use of their work, which is beyond the scope of the law. I agree to indemnify and save harmless Purdue University from any and all claims that may be asserted or that may arise from any copyright violation.

Ganesh Hegde
Signature of Candidate

4/21/10
Date

*Located at http://www.purdue.edu/policies/pages/teach_res_outreach/c_22.html

GENERATION AND OPTIMIZATION OF TIGHT BINDING (TB) PARAMETERS USING GENETIC ALGORITHMS AND THEIR VALIDATION USING NEMO3D

A Thesis

Submitted to the Faculty

of

Purdue University

by

Ganesh Krishna Hegde

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Electrical and Computer Engineering

May 2010

Purdue University

West Lafayette, Indiana

This work is dedicated to my parents Hema Hegde and Krishna Hegde.

Your struggle has made my education possible.

ACKNOWLEDGMENTS

I would like to thank Prof. Klimeck for providing me the opportunity to work in his group. His guidance during the course of this project has been valuable. Prof. Timothy Boykin was especially helpful in answering questions I had about Tight Binding and guiding me in the validation of the TB parameters. Special thanks are due to Dr. Michael McLennan for his initial guidance and Prof. Supriyo Datta for agreeing to be a part of my thesis committee.

To SDCI and NCN, for funding me and providing me with computational resources through the duration of my Master's degree.

To my lab mates at EE-350 (now DLRC), thank you for the discussions, I have learned a lot from you.

To Smita and Bharati Hegde and Uttara Iyer, this would not have been possible without your support. Thank you.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Hegde, Ganesh Krishna M.S.E.C.E., Purdue University, May 2010. Generation and optimization of Tight Binding (TB) parameters using Genetic Algorithms and their validation using NEMO3D. Major Professor: Gerhard Klimeck.

It is often the case in engineering problems that a desired output from a particular process is known, but the inputs that are fed into the process to get this output are not. Optimization, commonly referred to as fitting, reverse engineering, or mathematical programming is a well developed class of problem solving techniques that is used to tackle the above scenario. This thesis intends to discuss the application of one such scheme, the Genetic Algorithm, to specific problems in nano-electronics.
We initially intend to motivate the thesis by outlining the main problem at hand: The Tight Binding (TB) parametrization of materials at low temperature (LT). We then show why the Genetic Algorithm is well suited to tackle this non-trivial problem effectively and discuss some of the unique features of Genetic Algorithms. Finally, we describe in detail the procedure used to generate, optimize and validate TB parameters and discuss some results we have for materials like InAs and GaAs. We also include a description of some of the attempts to create a general purpose optimization engine for the nanoHUB (www.nanoHUB.org)

1. THE NEED FOR OPTIMIZATION

In this chapter, we shall illustrate, through a scenario typically encountered in real-life engineering situations, the need for optimization.We shall elaborate on this example and successively simplify our engineering problem to its bare essentials until it leads to the actual problem we wish to solve. Since this chapter is of an introductory nature, the use of equations will be kept to a bare minimum and general references will be provided to papers and documents that discuss details.

1.1 A common engineering scenario

Semiconductor lasers (also called laser diodes) are ubiquitous nowadays. They can be found in everyday consumer electronic items such as laser pointers, bar code readers, laser printers and CD-ROM read drives to industrial applications like high power welding. Laser diodes emitting coherent light throughout the visible spectrum can be produced these days [1].

Imagine a scenario where one is asked to find a material/s for a laser diode that will operate at a peak wavelength 'x nm' in the visible spectrum. How would we go about finding a solution to this problem?

1.2 Initial Questions

We would initially go about asking some very general and obvious questions that illuminate the problem better on the following lines:
How does a semiconductor laser work?
What decides the wavelength?

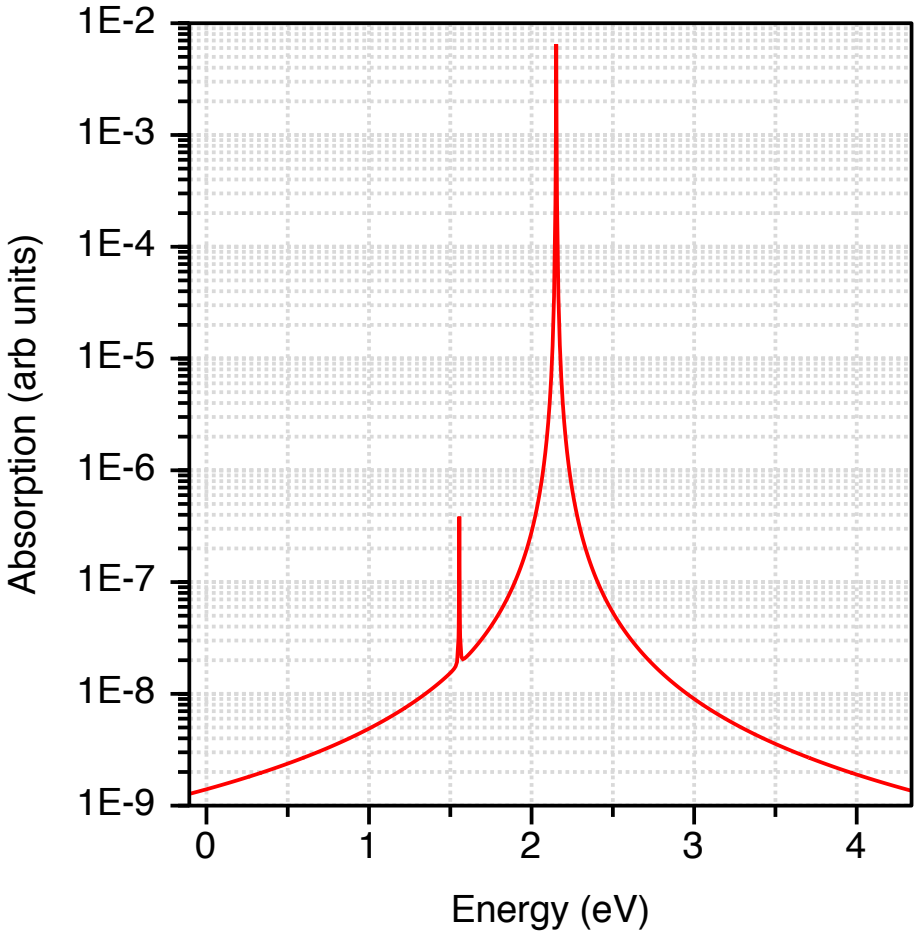


Fig. 1.1. Optical absorption plot with maximum absorption at 2.1 eV. Image obtained from [2]

1.3 Clues from Physics

From basic semiconductor laser theory [3], one knows that the wavelength of light emitted from a semiconductor laser depends upon the band gap of the material used. Fig 1.2 below indicates this from an elementary, discrete level perspective. Our initial questions would now be appropriately modified to the following: What material has a band gap of 'z eV' (that corresponds to the peak wavelength)?

One might think that the first question can easily be answered by looking up a standard materials handbook and finding a material with the required band gap. But what if the peak wavelength did not correspond to any of the standard band gaps?

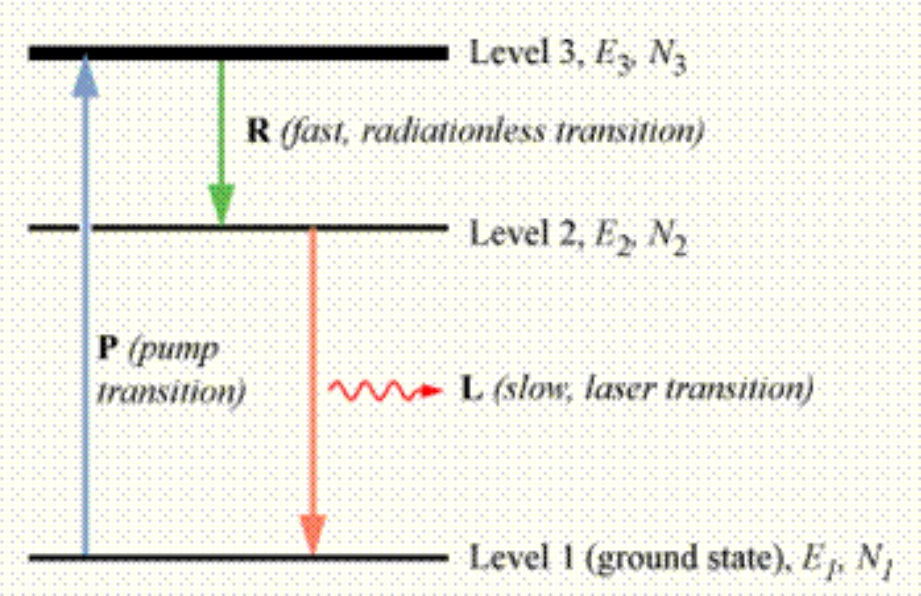


Fig. 1.2. A toy model for allowed transitions in a laser obtained from [4]. The wavelength is $\lambda = \frac{c}{\nu}$ where $\nu = \frac{(E_2 - E_1)}{h}$

(e.g. What if we require a band gap of 1.46 eV for our laser and the semiconductor that has the closest band gap of 1.424 eV at room temperature is GaAs?). It is a well known fact that limited straining of a semiconductor modifies the band gap of the semiconductor [5]. We then further modify our question to the following: Given the closest suitable material is GaAs, how do we strain GaAs to give us a band gap of 1.46 eV?

The answer to this question lies in a detailed study of the electronic band structure of GaAs under strain. The electronic structure of a crystalline material is a detailed description of the allowed energies that an electron takes in a material versus the electron's momentum. By studying the band structure of GaAs and the effect of strain on it, we can not only observe how the band gap changes, but also obtain directly the density of states of the material, which affect performance characteristics such as intensity of the laser.

Having gained some insight into the deciding factors for wavelength, we now have a problem that essentially requires determination of band structure of a material under various strain conditions and examining the gap variation with strain.

1.4 Determination of Electronic Structure and it's variation with strain

The determination of electronic band structure of materials is a well developed sub-field in condensed matter physics. The central problem of electronic structure determination is the solving the Schrodinger equation for realistic systems [6].The time-independent Schrodinger Equation essentially equates to an eigenvalue equation

$$[H]\psi = E\psi \tag{1.1}$$

Here [H] is a matrix called the Hamiltonian and E represents the eigenvalues of [H], i.e. the allowed energy levels that an electron can occupy in a material. ψ is a vector called the wave function of the electron and it has definite physical connotations (See [7], for a detailed explanation). For realistic devices like the laser diode under consideration, the Hamiltonian can be larger than 1×10^6 times 1×10^6 elements in size. Electronic structure methods essentially differ in the approximations involved in constructing [H]. While solving the Schrodinger equation for devices such as the laser diode, additional complexities including strain effects, defects, alloying, hetero-structure effects, a lack of crystal periodicity besides the aforementioned size of the Hamiltonian come into play and this complicates the solution further.

There are several competing electronic structure methods and each of these has its own merits (and demerits). The introductory nature of this chapter prevents us from going into details at this stage. One finds excellent descriptions in books devoted to electronic structure calculations [8] [9] [6]. Broadly speaking, however, the choice of a particular electronic structure calculation method generally comes down to an optimization of the following factors:

a) Speed of operation: Modeling realistic devices with strain, non-periodic structures in reasonable time is generally a prime consideration.

b) Accuracy: While time is critical, getting physically feasible results that match relevant experiments is another deciding factor.

The most well developed and accurate electronic structure calculations found in commercial quantum chemistry software are a category of methods called the ab-initio. As the name suggests, these methods try to adopt a 'first-principles' approach, i.e. no assumptions (in practice very few assumptions) are made while solving the Schrodinger equation. The Hamiltonian matrix elements are obtained by actually evaluating integrals representing these elements in real-space. The accuracy of these methods is offset, however, by their disadvantages. Commercial quantum chemistry codes can typically handle bulk crystalline materials well. However, they become prohibitively time consuming in case of real devices having millions of atoms and defects (see [10] for an explanation of the reasons). Other methods, such as semi-empirical methods and perturbation theory based methods [6], are usually preferred for applications such as the one we intend to study due to an acceptable tradeoff between the above deciding factors.

1.5 Focus of this thesis

This thesis will focus on the semi-empirical Tight Binding (TB) Method of electronic structure calculations [8]. A single line description of the TB method is that it treats Hamiltonian elements as unknown parameters that are fit to known experimental/theoretical solutions. Various theoretical and experimental methods are used to calculate energies and carrier effective masses of materials at certain critical points in the band structure. This knowledge is in turn used to determine the matrix elements that give rise to a band structure that matches these values. No attempt is made to match the band structure point by point against detailed calculations.

This implies that we can reformulate our laser diode optimization problem as a TB parameterization optimization problem. A little thought helps us understand better the details of this optimization problem. Since the Hamiltonian matrix is of the order

1×10^6 times 1×10^6, it would imply that we need to find this many unknowns. Here, physical insight comes to our rescue; we restrict the range of these unknowns to values that are physically realistic.

Experimental and theoretical energies and carrier effective masses are the known solutions to be fit by optimizing the matrix elements. The number of outputs and to be fit and the applicability of the parameters depends chiefly upon the intended application. For e.g. If we were studying the behavior of electrons in bulk GaAs intended for use in transistors, we would fit masses obtained from the band structure very tightly. The matrix elements could then be used to model electron transport through GaAs. On the other hand, if we intended to use GaAs for opto-electronic applications, we would fit energies at high symmetry points very tightly since energies and their relative difference are of chief importance in such applications. In the laser applications, one would also need to consider the variation of these energies with strain and that would increase the constraints on our optimization problem.

The vague initial problem statement that we had has now been transformed into a definite optimization problem as follows: Parameterize Hamiltonian matrix elements for use in optical applications given a certain set of known physical quantities such as energies and their variation with strain, effective masses, etc given the constraints of time and accuracy. In order to solve this optimization problem we need to find a suitable solution method and this shall be the focus of the next chapter.

2. SELECTING AN OPIMIZATION TECHNIQUE

Having defined our problem in the previous chapter, we shall discuss some of its salient features in this chapter. We shall see how these features help us choose a between a variety of optimization techniques that shall be qualitatively described.

2.1 Salient Features of the Tight Binding (TB) parametrization problem

The Tight Binding problem discussed briefly in the preceding chapter has the following salient features [8].

1) It is a linear algebraic problem that involves finding the eigenvalues of a matrix called the Hamiltonian [H].

2) The elements of H correspond to energies of overlap of different orbitals. This restricts the range of values that they can assume.

3) The elements of H are treated as unknown parameters to be fitted to experimental (and theoretical) energies, gaps and effective masses.

4) The size of H and its nature depend on the system under consideration. The number of unknown parameters in H depends on the TB model under consideration.

5) The elements of H are not unique. They are 'correct' insofar as they are fit as closely as possible to the energies and masses of interest.

2.2 Implications from an optimization perspective

Nature of optimization problem: The problem is constructed to be a global minimization problem. The inputs that minimize the difference between the calculated and desired outputs are to be obtained.

Nature of optimization landscape: The optimization landscape is discontinuous. There

may be combinations of inputs that give completely unphysical results.

Relation between input and output values: There is no direct relation between the input and the outputs. The outputs are derived from the band structure which is calculated by solving for eigenvalues of the Hamiltonian which has the unknown parameters as its elements.

Dimensionality of search space: Depending on the TB model used, the dimensionality of the search space changes. e.g. the number of unknowns is 10 in the sp3s* model but is 35 in the sp3d5s* model.

Range of input values: This is restricted by the physical interpretation of the matrix elements as overlap energies between orbitals. These energies are expectedly finite and are negative since bound electrons in solids must have negative energies. The search space is real and finite.

Number of output values: This is limited only by the number of experimental (or theoretical values when experimental values are not available) values available for a given semiconductor.

Output values like band energies and effective masses have a limited ranges of validity. e.g. From experiment, one knows that the band gap of unstrained bulk GaAs at 300K is 1.424 eV. Any input combination that gives us a gap of 6 eV is obviously incorrect since it gives us a result far from what is expected in reality. Restricting the output values in such a manner makes the search space discontinuous and places constraints on the search space and the type of optimization algorithm.

Fig 2.1 illustrates the essential features of the problem we wish to solve. Having broadly defined the features of our optimization problem, we now discuss some approaches that may be used to find a solution to our problem. We shall discuss their relative merits and demerits and zero in on a solution approach that handles all of the above constraints acceptably. A detailed discussion of these methods is beyond the scope of this thesis and references for detailed reading will be provided.

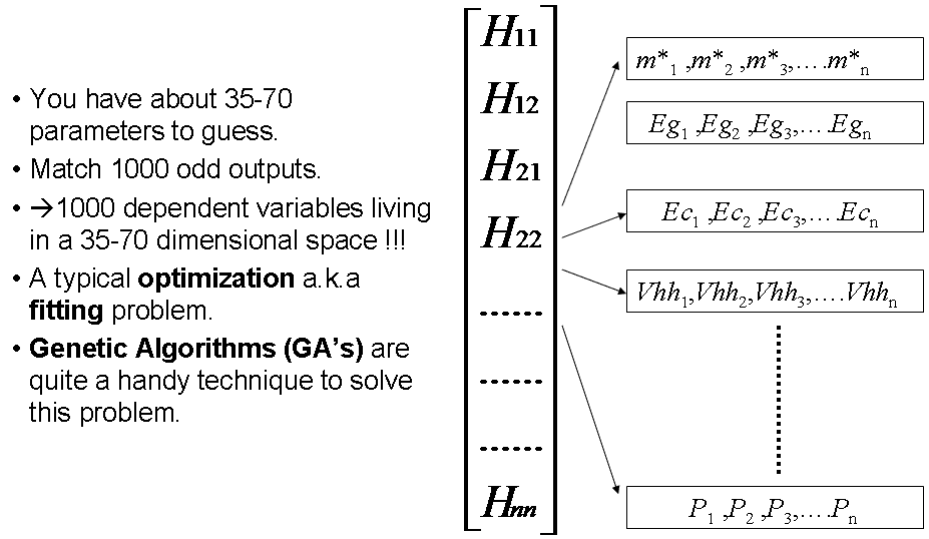


Fig. 2.1. TB optimization challeges - Multidimensional optimization

2.3 Optimization Techniques: A bird's eye view

2.3.1 Intuitive approaches

The most intuitive method one can think of is a brute force approach. One might be tempted to test all possible input combinations and find a particular combination that minimizes the output. As mentioned previously, the search space is real and to find combinations of 10 -100 real numbers fitting 100+ outputs is not a feasible task. One runs into problems with other intuitive approaches such as trial and error approaches and random sampling of search spaces due to the sheer size of the search space and the number of possible combinations of inputs. As we will see later, random searches are often combined with other techniques to solve optimization problems.

2.3.2 Gradient based approaches

Gradient based approaches attempt to find the minimum value of a smooth function, by calculating the gradient of the function with respect to the variables that constitute the function. Quoting from [11], we see how such an optimmization pro-

cedure might work:

Consider the minimization of a function $J(x)$ where x is an n dimensional vector. Suppose that $J(x)$ is a smooth function with first and second derivations defined by the gradient

$$g_i(x) = \frac{\partial J}{\partial x_i} \tag{2.1}$$

and the Hessian matrix

$$A_{ij}(x) = \frac{\partial^2 J}{\partial x_i \partial x_j} \tag{2.2}$$

Generally it pays to take advantage of the smooth dependence of J on x by using the available information on g and A. Suppose that there is a minimum at $x*$ with the value $J* = J(x*)$. Then

$$g(x*) = 0 \tag{2.3}$$

and in the neighborhood of $x*$, J can be approximated by the leading terms of a Taylor expansion as a quadratic form

$$J(x) = J* + \frac{1}{2}(x - x*)^T A(x - x*) \tag{2.4}$$

where A is evaluated at $x*$. The minimum could be approached by a sequence of steps in the negative gradient direction

$$x_{n+1} = x_n - \beta_n g_n \tag{2.5}$$

where β_n is chosen small enough to assure a decrease in J, or may be chosen by minimizing J with a line search in the direction defined by g_n.

One may however, reject this solution method on the basis of the fact that our optimization landscape is discontinuous and gradient based methods work for continuous spaces. One might still adapt the technique to work for discontinuous spaces by assigning arbitrarily high gradient values to points in the space where the outputs are undefined. Even so, this method suffers from a debilitating drawback - it is a local search technique. Successive evaluations of the gradient that may lead us to a local

minimum depend upon the proximity of the initial solution to the local minimum. Also, there is no guarantee that this local minimum is indeed the global minimum that we are looking for. Fig 2.2 shows an example of a function having several local minima. A local search technique starting at $x = 2$ would converge to a local minimum, not a global minimum.

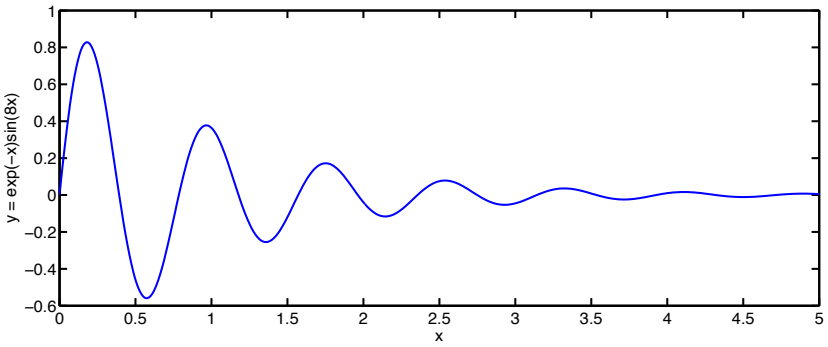


Fig. 2.2. Sample smooth function having several local optima

In order to be a global technique, a more elaborate scheme must be devised in which one divides the search space into smaller sub spaces, evaluates minima in these sub-spaces and compares them in order to obtain a global minimum. Besides the additional computational cost involved, this method is not, in general advisable for rapidly undulating optimization landscapes since sub-dividing the search space may lead to sub-spaces having several local minima ad infinitum. Another difficulty with this approach is the fact that it is iterative. Successive solutions depend on the quality of the initial solution. This slows the optimization process by limiting the region of search space that can be sampled. Gradient based approaches are not without their advantages. They are a preferred technique in certain convex optimization problems where the optimization landscape is smooth and solutions converge to global minima in a matter of a few gradient evaluations. [11]

2.3.3 Non-linear least squares optimization techniques

This set of techniques aim to find a solution that minimizes the square of the error between desired outputs and the calculated outputs. The Levenberg-Marquardt Algorithm [12] and the Gauss-Newton Algorithm [13] are examples of this class of optimization techniques. Imagine a parameter space consisting of n independent variables x_i and dependent variables y_i. Assume that y_i are only known by measurement. The optimization then consists of finding a set of m parameters β_j such that the residual error S shown in the equation below is minimized

$$S = \sum_{i=1}^{n} r_{i^2} \tag{2.6}$$

where $r_i = y_i - f(x_i, \beta)$ While there is no closed form solution to the non-linear least squares minimization problem, numerical algorithms to find βs that minimize S are used. An initial set of βs is assumed. Subsequent sets are found iteratively.

$$\beta_j^{k+1} = \beta_j^k + \Delta\beta_j \tag{2.7}$$

where k is the iteration number. The essence of the technique, then, is estimating successive solutions by finding linear extrapolations from the current solution subject to the conditions of the equations above. This method is widely used as a form of curve-fitting. It's usage in the problem we have defined is limited because it involves an additional set of matrix manipulations after our initial matrix-eigenvalue-band structure solution. This method is also iterative and though refinements to the original technique of Gauss have been created [13] to overcome dependence of the final solution on the initial solution, the method is still serial in nature.

2.3.4 Simulated Annealing (SA) and variants

SA is a name given to variants of a global optimization heuristic technique [14]. The technique aims to replicate algorithmically the process of annealing in metallurgy, a technique involving heating and controlled cooling of a material to improve

crystallization and reduce defects. The heat causes the atoms to become unstuck from their initial positions (a local minimum of the internal energy) and wander randomly through states of higher energy; the slow cooling gives them more chances of finding configurations with lower internal energy than the initial one. By analogy with this physical process, each step of the SA algorithm replaces the current solution by a random "nearby" solution, chosen with a probability that depends on the difference between the corresponding function values and on a global parameter T (called the temperature), that is gradually decreased during the process. The dependency is such that the current solution changes almost randomly when T is large, but increasingly "downhill" as T goes to zero. The allowance for "uphill" moves saves the method from becoming stuck at local minimawhich are the bane of greedier methods. This technique has been previously used in studying the electronic properties of atomic clusters with moderate success [15]. Parallel Simulated Annealing has been attempted by some groups [16] [17] by dividing the search space into multiple sub-spaces as mentioned before and facilitating information interchange between the multiple solutions so obtained. A variation of the thermal annealing technique is called quantum annealing [18]. Here, a solution can move through a classically forbidden area, such as a potential energy barrier and the resulting tunneling process can sample different areas of the search space instead of being stuck at local minima.

2.3.5 Evolutionary Algorithms and bio-inspired algorithms

Like Simulated Annealing, Evolutionary algorithms are heuristics. They aim to solve optimization problems by mimicking processes seen in nature. Their exploration of the search space is non-mathematical, i.e. they are void of gradients and simply rely on evaluation of fitness at different parts of the search space. Due to this feature, these algorithms lend themselves easily to a parallel implementation, which speeds up the search for an optimal solution. An introduction has been provided to two such schemes, while a detailed description of the third is given in the next

chapter. Heuristics often overcome potential problems with more mathematical techniques. They can work on real and discrete spaces, they are global in nature, they are parallel as mentioned before. Several of these techniques employ an initial stage of randomness in order to sample the search space effectively followed by the specifics of the algorithm. the dimensionality of the search space affects this initial sampling, but subject to constraints of memory, if well distributed initial samples can be obtained, then these techniques are essentially independent of the dimensionality of the search space.

Particle Swarm optimization (PSO)

The essential premise of PSO is that potential solutions in the search space must exchange information in order to converge quickly to global optima. This procedure is designed to mimic social-interaction between living beings, where a better solution to a problem found by one person/being is quickly adopted by other beings. A more detailed exposition can be found in [19].

Ant Colony optimization (ACO)

The behavior of ants in finding the shortest possible path to a source of food is mimicked in this algorithm. It has been observed that ants, in their search for food, leave a trail of pheromones along the path that they have traveled. Pheromones evaporate with time. This implies that a short path to the food source is likely to have a longer-lasting pheromone trail. As more and more ants traverse the path and leave trails, the pheromone concentration increases. By exploring several such trails, the entire population converges towards the shortest trail. [20] has a detailed description of the ACO process and some of its applications.

Genetic Algorithms

Briefly, GA's attempt to solve optimization problems by mimicking biological genetic reproduction and consequent natural selection amongst members of a species or across species. Genetic Algorithms share all of the features of heuristics discussed above, making them an attractive choice for solving the non-trivial problem we have posed i.e. they are global in nature, individual solutions are inherently independent of each other facilitating parallelism, their dependence on input dimensionality is subject to memory constraints while the number of outputs isn't an issue, they can handle real and discrete inputs, they can offer multiple solution choices to designers and their execution in global multi-dimensional problems is order of magnitudes better than local search techniques. [21] In addition to this, parallel implementations of the GA using MPI like PGAPACK [22] have been around for quite some time, reducing the development time for solutions considerably. The GA has been used to parametrize materials like Si, GaAs, InAs at 300K [23] and was hence an automatic choice of optimization technique.

It cannot be emphasized enough that our choice of GA is not unique. There are several optimization methods that can tackle the problem we have posed, each with their own merits and demerits. This chapter must therefore not be viewed as an exhaustive review of pertinent techniques, but as a review of techniques we considered while selecting a solution method. The fact that the GA clearly worked well in the past for similar problems weighed heavily in favor of continuing to use the GA as an approach. We shall explore the basic features of a GA in detail in the next chapter.

3. THE GENETIC ALGORITHM(GA)

Having selected the GA for its supposed merits, we now wish to see how the GA operates and how it achieves its objective of solving optimization problems by mimicking biological reproduction.

3.1 A brief biological background

In 1959, a naturalist named Charles Darwin published a document titled 'On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life' [24]. In this work, Darwin details how life-forms evolve and adapt to the conditions around them and how this adaptation to local conditions leads to specific features in creatures. Commonly referred to as 'the survival of the fittest', the principle of natural selection explains the chances of survival of 'strong' or 'fit' individuals in a population and increased chances of them producing offspring similar to themselves. The concept of biological inheritance in reproduction, first proposed by Austrian scientist (and friar) Gregor Mendel in the 1860's [25] received experimental verification in the early 1900's and a high-point was reached when the molecular structure of DNA was discovered by Watson and Crick in 1958 [26]. In order to understand the Genetic Algorithm, we need to first understand the biological foundation on which it is based. In doing so we first describe the basic units of reproduction in creatures, taken from [27]:

A gene is the basic unit of heredity in a living organism. All living things depend on genes. Genes hold the information to build and maintain an organism's cells and pass genetic traits to offspring. A modern working definition of a gene is "a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions, and or other functional sequence regions."

The total complement of genes in an organism or cell is known as its genome, which may be stored on one or more chromosomes; the region of the chromosome at which a particular gene is located is called its locus. A chromosome consists of a single, very long DNA helix on which thousands of genes are encoded. Figure 3.1 shows a schematic picture of a biological chromosome and it's component genes.

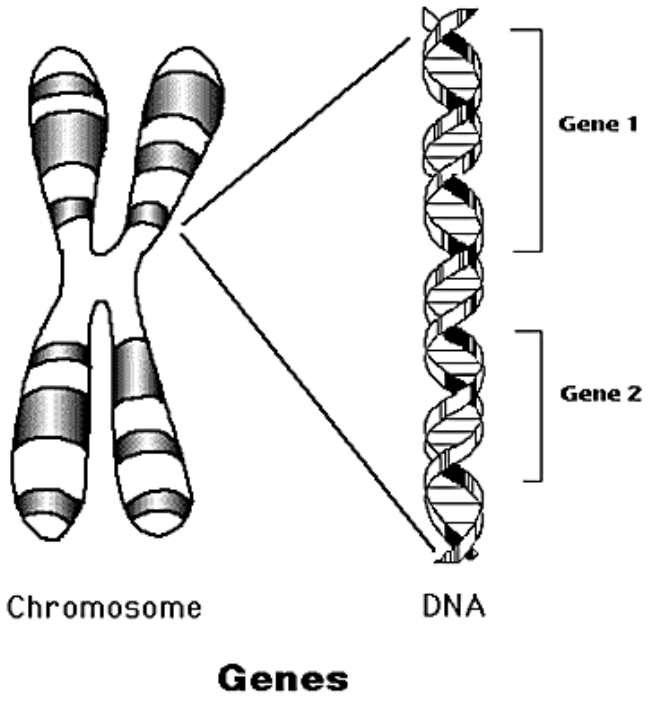


Fig. 3.1. Biological Chromosome and component genes

Simply (but somewhat incompletely) put, a gene is a strand of DNA representing a physical trait in an organism. Genes are stacked side-by-side to form a chromosome and thus chromosomes can be said to contain the sum total of all physical traits of an organism, i.e. the blueprint of an organism. In natural selection, mating between individuals puts into motion processes that combine chromosomes from each of the two individuals and produce new chromosomes that on the whole retain certain characteristics of the parent chromosomes through crossover while otherwise introducing completely new information through mutation. Thus characteristics that enable an

organism to survive in an environment are passed on to its offspring, while new characteristics that enable the offspring to adapt to changes in environment are developed in the process.

3.2 The algorithm proper

The credit for explicitly formulating the principles of GA's goes to John Holland, who, in his 1975 work titled 'Adaptation in Natural and Artificial Systems' [28] introduced a formal framework for predicting the quality of the successive generations of possible solutions to a given optimization problem. We now describe in detail the building blocks of a Genetic Algorithm. For the purpose of illustration, let us assume that our optimization problem consists of trying to find the minimum of the following function

$$z = f(x, y) = x \exp^{-x^2-y^2} \; where -2 \leq x, y \leq 2 \tag{3.1}$$

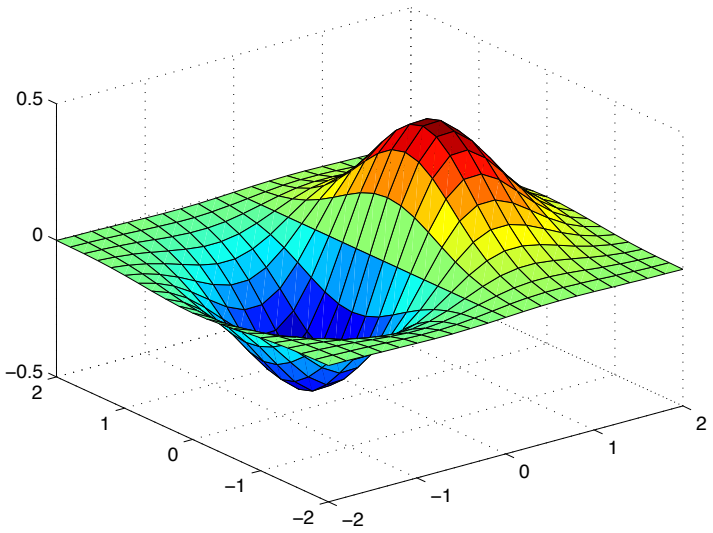


Fig. 3.2. $z = f(x,y) = x\exp^{-x^2-y^2} \; where -2 \leq x, y \leq 2$(3.3)

3.2.1 The Gene

The algorithmic gene is an encoded representation of an input variable. We could chose a floating point representation/real number representation for our problem, or an integral representation or a binary representation. For the toy problem we are considering, we use real encoding. We decide to encode each input variable as a real number in our sample problem $x = 0.123456$, $y = 0.17634$.

3.2.2 The Chromosome

The algorithmic chromosome is a collection of encoded genes. It is therefore an encoded representation of a potential solution to our optimization problem. Since in an algorithmic representation, there isn't any difference between an individual and a chromosome, we shall use the term individual and chromosome interchangeably. In our example, the real representation of X and Y collectively produces a chromosome of the following type :

$$[x, y] = [0.123456, 0.17634]; \tag{3.4}$$

3.2.3 Fitness of an individual

The fitness of an individual is an objective determination of how good a solution is to a given problem. In single-objective implementations it is taken to be a single number that indicates how close a potential solution is to the optimum. In the above problem, the fitness is simply the value of z at each [x,y] tuple. Fig 3.4 shows two chromosomes [x_1,y_1] = [0.1234,0.1425] and [x_2,y_2] = [0.1768, 0.1111]. The fitter individual will be that combination of x and y for which z is minimum; in this case [x_1,y_1] = [0.1234,0.1425] is the fitter individual of the two.

Having defined the building blocks of the algorithm, we now discuss the working of a GA in discrete steps. For a more detailed exposition on the subject, please refer to [29].

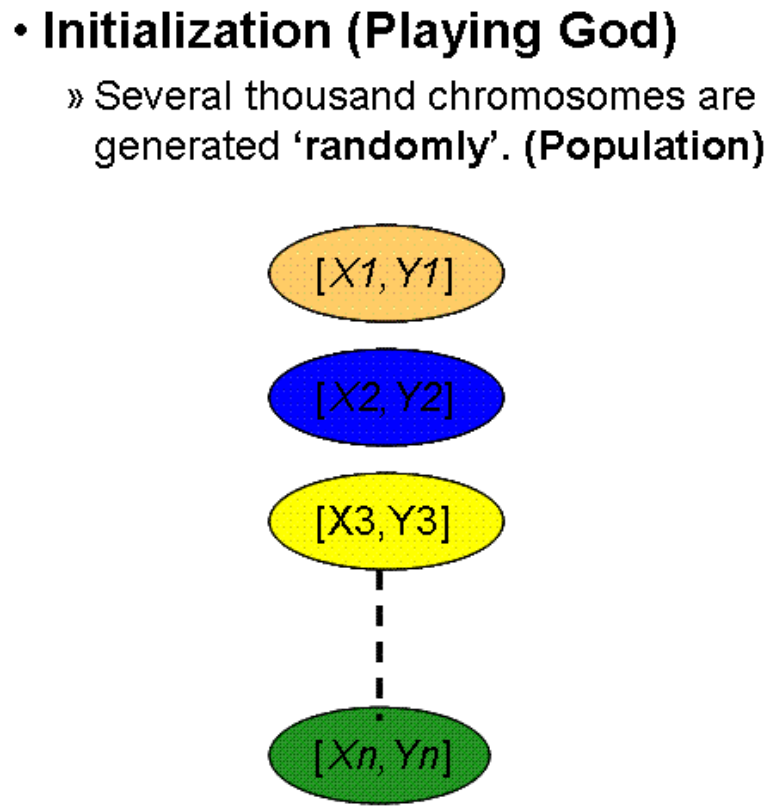


Fig. 3.3. The initialization process

3.2.4 Initialization

In the first step, we randomly generate a number of individuals. This set of individuals is termed as the first generation. The random number distribution can be uniform in case one has no idea about the possible location of a solution. Alternately, from the knowledge of the problem, we can draw these random solutions from a Gaussian distribution or a normal distribution in order to narrow down our search space considerably. The second case is desirable for quicker convergence. The initialization step is essential to sample the search space broadly since the optimization landscape for an n-dimensional problem is very complex. Conversely if the initialization is Gaussian, then it helps sample as many points close to the initial solution as possible.

3.2.5 Evaluation and Ranking

Here, the randomly generated individuals are evaluated for their fitness. A Fitness evaluation usually involves running the actual math/physics functions that need to be optimized and computing a number that represents the fitness of a chromosome. The individuals are then ranked according to their fitness value.

3.2.6 Selection

a) Tournament Selection: Here the fittest 'n' individuals find representation in subsequent generations. b) Proportional Selection: Here, the probability of an individual making it to the next generation is proportional to its fitness. This method does not guarantee that the fittest individual makes it to the next generation. Selected

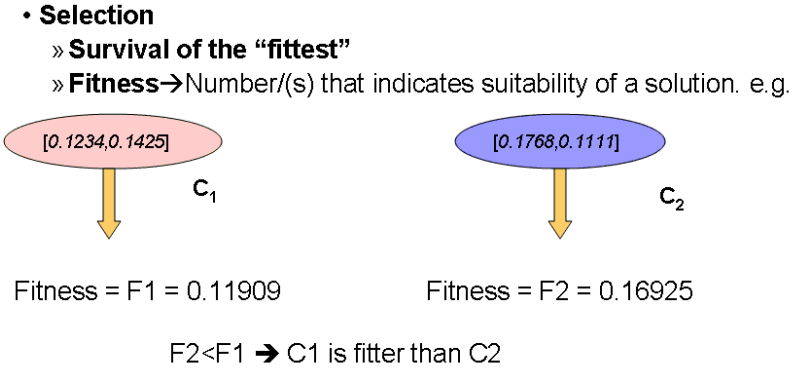


Fig. 3.4. The fitness of an individual

individuals are termed 'parents' and are 'mated' and modified in subsequent steps to form new populations. We shall discuss two primary mating operations, Crossover and Mutation that are central to a standard GA.

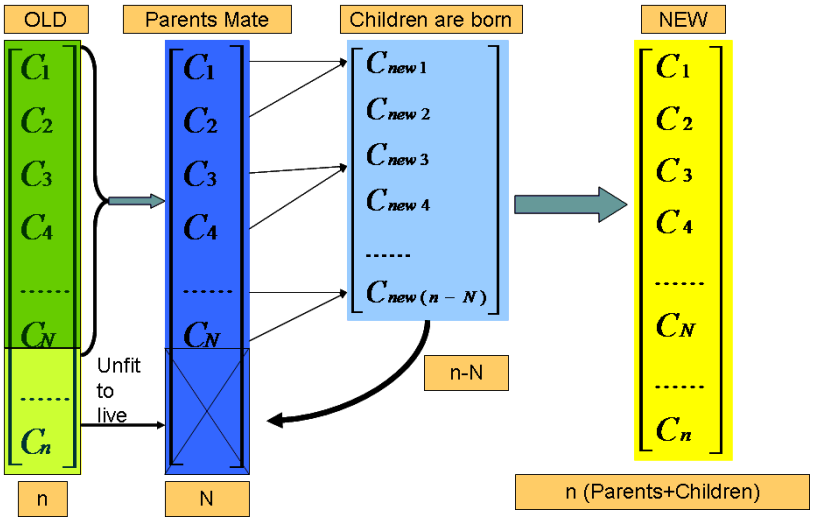


Fig. 3.5. The Mating process

3.2.7 Crossover

Parents selected from a previous population are mated to produce 'children'. The process is probabilistic (with the probability decided by the user depending on the nature of the problem), analogous to natural reproduction where mating parents may not necessarily produce children. However, this probability is not linked to the fitness of individuals as it is in nature. The simplest form of crossover is a single point crossover. Here, a gene is selected at random from a parent chromosome. The position of this gene in the chromosome is called the locus of crossover. Two parent chromosomes are aligned at their respective locii and the information to the right/left of each is swapped to form two new solutions. The essential power of the GA is manifested in this step. Fig 3.6 shows a schematic picture of the crossover process. Children solutions produced by crossover may or may not be fitter than parent solutions (In Fig 3.6 C_4 is fitter than C_2). In case they are fitter, they find representation in subsequent populations, while if they aren't they are likely to be weeded out by the selection step mentioned previously. It should be emphasized that the notion of schemata [28] proposed by Holland as an explanation to the improvement of fitness in a GA is disputed. It can be seen, however, that a crossover at worst

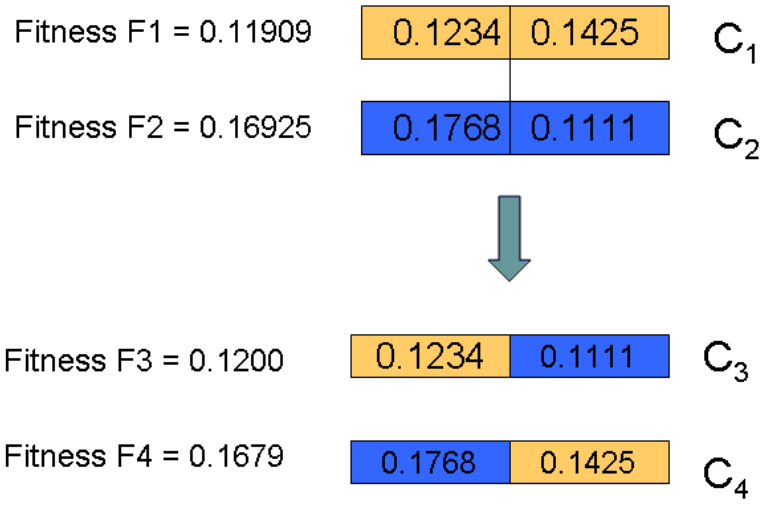


Fig. 3.6. An example of crossover in a real-encoded chromosome

produces inferior offspring which are discarded in subsequent generations, and at best produces superior offspring. Crossover is an exploitative operator, since it has a chance of exploiting strengths of parents to form fit children. Several varieties of the crossover operator exist and the reader is referred to [29] for details.

3.2.8 Mutation

When the above steps are repeated several times, one notices a certain homogeneity in the population owing to an improvement in average fitness. This leads to a lack of diversity amongst solutions and a cessation of exploration of the search space. In optimization parlance, the individuals are stuck at a local optimum that may or may not be a global optimum. Frequently, an operator called the mutation operator is used to remedy this situation by modifying a small number of individuals in the population. This is akin to biological mutation where diversity is introduced in the population by a mutation in a gene. While this mutation may or may not affect fitness considerably, if it does, it produces a solution that is better than existing solutions without crossover. If it degrades fitness it gets weeded out in subsequent selection. In Fig 3.7 we see how mutation can improve the fitness of a chromosome.

The x value of the chromosome is subtracted by 0.1 and it leads to a fitness that is one order of magnitude better than the fitness of the original chromosome. From the

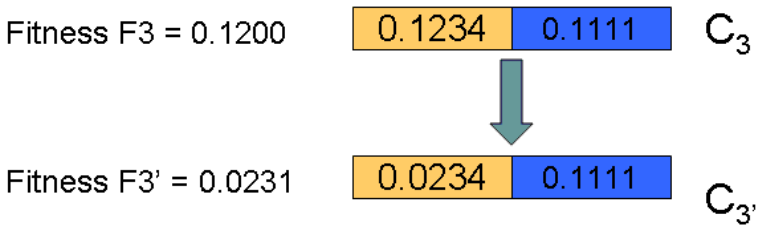


Fig. 3.7. An example of mutation in real-encoded chromosomes

above definition it can be seen that mutation is an exploratory operator.

When the above steps are repeated several times (to account for the probabilistic nature of the algorithm), one converges rapidly towards acceptable solutions. It may be argued (incorrectly) that random sampling of the search space produces similar results. There may be a chance of this happening in an odd trial, but on an average a GA always performs better than a random search. This is because a random search has no information storage and processing like a GA does and consequently, has no guarantee of improvement of fitness. The GA carries an implicit guarantee, if implemented correctly, that fitness at best improves and at worst remains constant.

Before ending this chapter, it might be useful to note the following:

1) Individuals in a generation are independent of each other. Their fitness can be evaluated independently of other individuals. Thus, a GA implicitly allows parallel execution. Parallel execution implies that one can evaluate a large number of potential solutions in the same time that an iterative process would evaluate just one.

2) The evaluation step represents a clear delineation between the algorithm and the physics/math of the problem. The calculation of the fitness is problem dependent.

3) The GA does not guarantee convergence, but it's features do give acceptable solutions in relatively short durations.

4) The operations enlisted form the Standard GA. This is the form that resembles

natural selection. Our central objective is to find a suitable solution and the definitions mentioned are not hard and fast rules. Several variations exist, but the essential philosophy of the GA remains the same regardless of a particular implementation i.e: 'Statistical Survival of fit individuals in a population and the increased probability of them mating to produce fitter individuals'.

Having seen the features of the GA, we now discuss its application to TB parameterization in the next chapter.

4. TB PARAMETERIZATION USING GA'S - METHODOLOGY

Having seen the motivation behind the current work and the working of a GA, we are now in a position to discuss the application of the GA to our main problem: i.e. the TB parameterization of semiconductors. We briefly discuss the salient features of TB that make it attractive to model nanoscale devices. The methodology followed is essentially that mentioned in Klimeck et al. [23]

4.1 TB models: appeal, problems and challenges

4.1.1 Appeal

Nano-scale devices are characterized by material and potential variations on the length scale of a few atoms. The essential premise of the TB method is the representation of a single electron wave function as a linear combination of atomic (or atomic-like) orbitals. This indicates that the TB method may be able to capture the effects of nano-scale spatial variation in material nature and potential (and hence charge density) on an atomic scale rendering it useful to model nano-scale devices [9]

4.1.2 Problems

The appeal of the TB approach is offset by the fact that the basic building blocks of the method are not gaps and effective masses, but interaction energies of the various orbitals superposed to represent the eventual wave function. In addition to this, material parameters, such as band gaps, energies and carrier effective masses at high-symmetry points, are non-trivially related to the interaction energies. There is no quantitative way to tell how a particular parameter affects a particular physical

quantity, one can only solve for the eigenvalues of the material Hamiltonian and directly compute these quantities. In addition to this, nano-scale devices have atomic positions that vary drastically from their positions in a crystalline material, due to the presence of strain in these devices. This changes the coupling between the orbitals and hence the electronic structure.

4.1.3 Challenges

One must therefore not only be able to obtain the bulk band structure of a crystalline material reliably so that it matches experiment, one must also be able to include the effects of strain in these models to ensure that the parameters are transferable to realistic device models. In addition to this, the multi-dimensional nature of the optimization problem [Refer to chapter 2 and 3 for details] implies that a given parameterization is valid insofar as it is fit to represent certain parameters correctly. e.g. If we wish to model electron transport, the physical features of primary interest are the slope of the band structure near the conduction band minimum and the value of the conduction band minimum itself. On the other hand to model optical devices, we must ensure that energy gaps and valence and conduction band optima are fit within reasonable error margins.

The above challenges imply that in order to obtain a general, transferable set of parameters, we need a optimization method that can fit several outputs, handle arbitrary numbers of inputs and yet give acceptable fits in reasonable times. We have seen in previous chapters how the GA is an acceptable method in the light of these challenges. We now describe the parameterization process.

4.2 TB Parameterization - Overview

The TB parameterization process is divided into 4 distinct stages as follows:

1) Selecting a bulk TB model and its modification under strain.

2) Obtaining output values to be fit.

3) Optimizing TB parameters.

4) Validating obtained parameters.

4.2.1 Selecting a bulk TB model and its modification with strain

Zero strain

The TB method suffers from an inherent drawback that the electron wave function is constructed from an incomplete basis [9]. In order to obtain band structures that match experiment, and otherwise make sense physically, it is necessary to choose a sufficient number of orbitals to construct the basis. For tetrahedral semiconductors such as InAs, GaAs, AlAs and Si, it has been seen that choosing a sp3s* basis predicts that the effective mass at the X point is infinity [23]. This problem is alleviated by selecting a larger basis, including the 5 d-orbitals. This factor increases the computational complexity of the process, since greater number of orbital interactions must now be included. We use an sp3d5s* basis as outlined in [30]. The NanoElectronic MOdeling - 1D tool (NEMO) [31] that implements this model is used to calculate the band structure of the given semiconductor along a specified direction.

Finite Strain

Under the action of strain, the interactions between the orbitals are modified on account of a change in lattice spacing. The changes in the Tight Binding Hamiltonian with strain are outlined in references [32] and [30]. In realistic devices containing several millions of atoms, such as quantum dots and quantum wells, the strain need not be homogeneous and a parameterization that handles arbitrary strains is desired. The parameterization under strain mentioned in [32] handles arbitrary strains in the elastic limit.

4.2.2 Obtaining target values

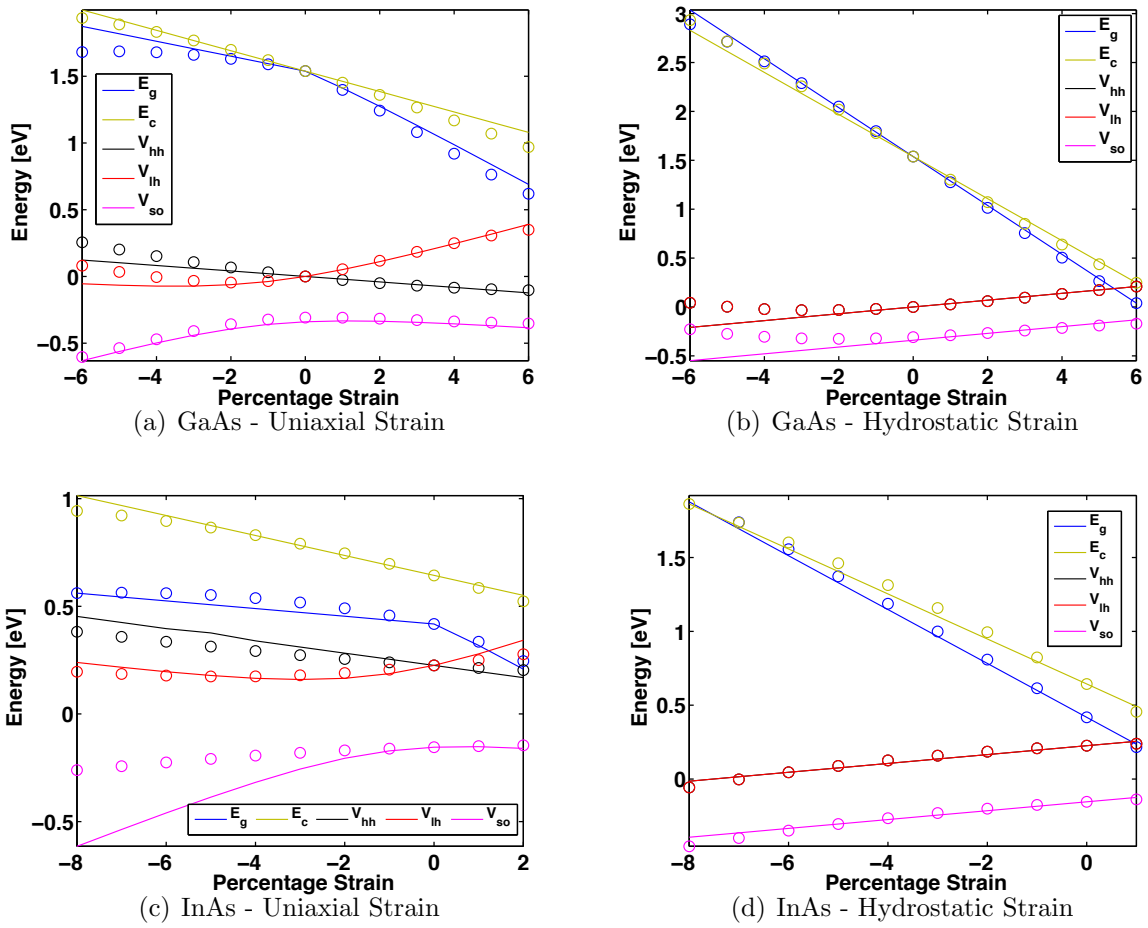
Zero strain

Where possible, experimental values are used as the outputs to be fit. Our principle source for experimental data is the exhaustive listing of zincblende material properties provided in Madelung [33], Vurgaftman et al. [34] and Adachi [35]. These references contain experimental values of conductivity effective masses, band edges at high symmetry points, band gaps at high symmetry points, variation in band gaps with temperature, variation of lattice constants with temperature and similar such useful data. Parameters such as band edges, k -values at high-symmetry points, band gaps that can directly be extracted from the calculated band gap are compared with these experimental values obtained from this handbook. The band structure effective masses at high symmetry points, that correspond to the slope of the bands at these points, are not obtained from experiment. We use the theoretical values provided in [34] as the masses to be fit in this case.

Finite Strain

It has been long known that strain modifies the electronic structure of semiconductors [5] besides modifying their structural properties. This effect has been used to create bulk, strained crystalline materials that have better device performance than their unstrained counterparts [36]. Semiconductors have also been observed to have in built strain when grown epitaxially on lattice mismatched substrates. This built in strain, beyond a threshold causes spontaneous formation of islands of the epitaxially grown semiconductor in the so called Self- assembly or Stranki-Krastanow process [37]. This process is the basis for growth of a variety of novel nanostructures such as self-assembled semiconductor quantum dots. Bulk electronic structures have historically been calculated assuming the valence band maximum to be the reference zero energy point. In semiconductor nanostructures, the presence of multiple ma-

terials complicates this process and the valence band maximum of one material is assumed to be the reference zero energy point and all other energies of technological interest are calculated offset to this reference. The theoretical values of these energy offsets are provided by the 'Model Solid Theory' described in [38]. This same reference also describes the effect of hydro-static strain and bi-axial strain on the conduction and valence band edges. We use this theory to find values of band edges and gaps under strain. Fig 4.1 is an example of the theoretical variation with strain obtained from van de Walle's parameterization and the fits we obtain using our GA process.



(a) GaAs - Uniaxial Strain

(b) GaAs - Hydrostatic Strain

(c) InAs - Uniaxial Strain

(d) InAs - Hydrostatic Strain

Fig. 4.1. Band gaps and band edges at Gamma for GaAs and InAs at 4K. Solid lines indicate theoretical fits. Circles are our calculated values.

4.2.3 Optimizing TB parameters

The optimization procedure itself is split up into several parts.

Choice of software

We make use of the Parallel Genetic Algorithm PACKage (PGAPACK) [22] as the software module that implements the Genetic Algorithm. PGAPack has several useful features such built-in libraries for parallel execution, a transparent interface that allows users to define custom data-types and GA operators such as crossover, mutation, etc. It also has a very useful User Guide that significantly reduces development time related to the GA. The interface between the band-structure code and the GA code ensures that the band-structure code is opaque to the GA. The only interaction involves the return of a floating point number representing the fitness of a string to the GA.

Initialization (done by GA)

We use a real number encoding scheme for the genes in our algorithm. Each gene represents a matrix element of the Hamiltonian or a strain-related parameter that modifies this matrix element. Each chromosome is simply a contiguous collection of these real-numbered genes. As mentioned in the previous chapter, the initialization step involves generating a large number of chromosomes/individuals to form an initial population of individuals that lives in the solution space. While this process can, in general be random, we find that restricting the range of numbers to physically meaningful values within a certain range helps in generating meaningful band structures. The number of initial individuals generated is user-defined and is typically a minimum of 20 times the number of input parameters to ensure sufficient sampling of the search space.

Evaluation (done by band structure code) and Selection (done by GA)

Each generated individual represents a unique band structure calculation. Once the band structure is calculated, gaps, edges, masses are extracted directly from the band structure. The errors between calculated and desired values defined in the material database are computed. The 'fitness' of an individual is then calculated as a weighted sum of the square of the errors. The smaller the error, the 'fitter' the individual. Our problem is a minimization problem. We wish to minimize the error between the calculated and desired values to as small a value as possible. This fitness value is then passed on through the interface to the GA which sorts and ranks the entire population of individuals in the ascending order of fitness. The fittest 'n' individuals are then selected through Tournament or Proportional Selection (described in [29]) and are used to form the next population. The calculation of fitness and the weights assigned to individual outputs (deciding their importance) are the most critical components of the entire GA process. Our GA process is a multi-objective optimization since we have several outputs to be fitted. However, we calculate just a single number which is a weighted sum of errors for each output. The assignment of weights makes the process extremely sensitive to weights. Though not implemented in PGAPACK, Pareto optimal fronts for an application can be coded for multi-objective optimization [39].

Crossover and mutation (done by GA)

Selected individuals form subsequent populations by crossover and mutation. One point, two point and uniform crossover for real values chromosomes is implemented in PGAPACK and we use these without modification. We explored two kinds of mutation operators with essentially the same net result. In the first mutation operator, we added or subtracted a very small fraction of the gene value to the gene randomly. In the second, we changed drastically the value of the gene by adding a large number but still keeping the gene value within physically sensible limits. These two opera-

tions produce an entirely new population and the steps from evaluation through to mutation are repeated until a stopping criterion is met.

Stoppage criteria

This can either be a maximum number of iterations, a minimum fitness, a condition of homogeneity in the entire population (i.e. when all chromosomes are very much alike) or a maximum number of evaluations.

The entire GA process described above is summarized in the following process diagram.

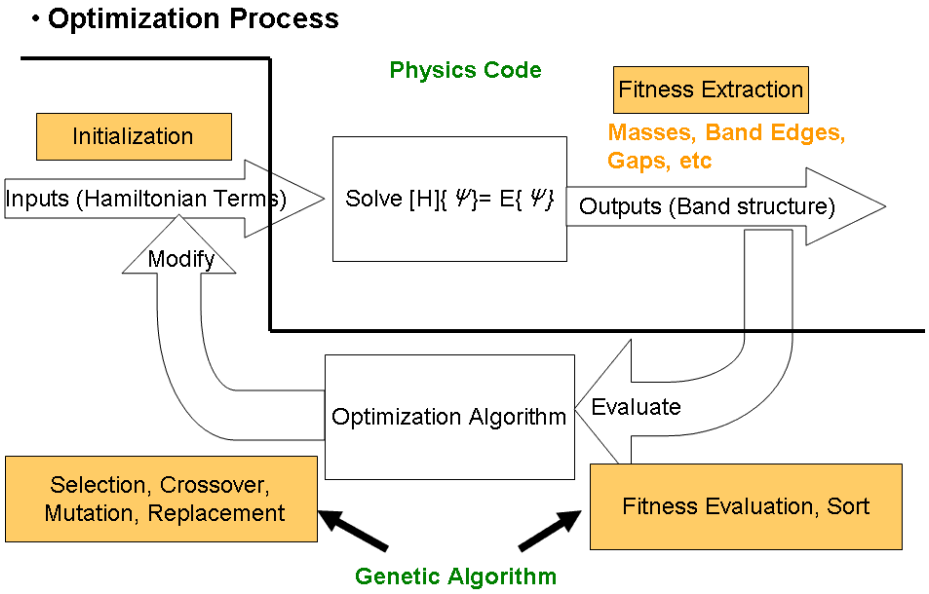


Fig. 4.2. TB parameterization with GA

5. RESULTS - TB PARAMETERIZATION BY GA-NEMO AND ITS VALIDATION IN NEMO-3D

Having seen how the GA is used to optimize TB parameters, we are now in a position to analyze some of the results we have obtained for low temperature TB parameterization of InAs and GaAs using this procedure. In this chapter, we also discuss the validation of these parameters by benchmarking simulations against actual experiment.

5.1 The InAs/GaAs quantum dot system and the need for low temperature TB parameters

Self-assembled InAs dots embedded in a GaAs substrate have been investigated with great interest due to their unique electronic properties that arise as a result of confinement of charge carriers in all three spatial dimensions. Due to these interesting properties, InAs/GaAs quantum dots have been studied in a variety of optical and electronic applications like photo-detectors [40], lasers [41], single electron transistors [42] and more recently, as potential solid state memory devices [43] and solid-state qubits [44] for quantum computing. In order to be accepted as potential replacements for currently existing technologies, these devices must exhibit robust room temperature operation. Prior TB studies of the electronic structure of quantum dots have therefore correctly focussed on a room temperature TB parameterization [30]. However, a number of experiments involving InAs/GaAs quantum dots have been carried out at low temperatures of about 4K. A number of these experiments have shown that the photo-luminescence spectrum of quantum dots is temperature sensitive and we cite Dai et al. [45] and Fry et al. [46] as just two examples among them. The activation of an increasing number of phonons with temperatures also affects the spin

coherence of electrons [47], a critical factor in qubits for quantum computation. The above examples indicate that a room temperature TB parameterization will give incorrect results when applied to model low temperature experiments. It is therefore essential that a new parameterization be derived to model the electronic structure of InAs/GaAs quantum dots at low temperature. In addition, prior detailed theoretical studies of the InAs/GaAs quantum dot system using pseudo-potentials by Williamson et al. [48] have been carried out at low temperature. Obtaining a TB parameterization at low temperature will help us compare our results against these theoretical studies.

5.2 Distinct problems in the modeling of InAs/GaAs quantum dots

InAs/GaAs quantum dots grown by the Self-Assembly process [37] are characterized by spatial inhomogeneity on account of strain in the structure. This strain arises due to the large (approx 7 percent) mismatch between the lattice constants of InAs and GaAs. Any theoretical model of the structure can be realistic only if it captures this inhomogeneity accurately. The modeling of the electronic structure of InAs/GaAs in the TB approach consists, then, of 3 distinct problems as follows.

5.2.1 Determination of bulk electronic structure of InAs, GaAs and its variation with strain

Following the procedure outlined in the previous chapter, a bulk TB parameterization is first carried out. The InAs bulk band structure is offset to the GaAs valence band maximum during the parameterization. A comparison of the low temperature bandstructure with room temperature band structure is carried out as a first order sanity check. It is essential to ensure that the parameterization reproduces all bulk band symmetries and the correct magnitudes of energy gaps and edges. It must also be ensured that the parameters obtained are physically meaningful, implying that the

signs and relative magnitudes of these parameters must reflect the physical quantities they are supposed to model.

5.2.2 Determination of Atomic Structure of the dot

The Born-Oppenheimer approximation [7] allows the atomic degrees of freedom to enter electronic structure calculations as fixed parameters. However, the fact that the atomic positions affect the eventual electronic structure indicates that these positions must be determined to a high level of accuracy to model electronic structure accurately. In bulk materials, symmtery greatly reduces the error in determining atomic positons. However, in nanostructures such as quantum dots, the presence of inhomogeneous strain and the size of realistic dots (of the order of millions of atoms) complicates matters and it is important to model these effects correctly to determine eventual atomic structure.

5.2.3 Determination of electronic structure of the InAs/GaAs quantum dots

Having determined atomic structure of the quantum dot, we construct a TB Hamiltonian using the bulk parameterization. We then solve for its eigenvalues and eigen-vectors. This process is usually time consuming as one must solve a eigenvalue problem for a matrix of the order of millions. Having obtained the electronic structure of the dot, we then compare these results to experiments.

The results of each of these three stages for low temperature InAs/GaAs dots will now be discussed separately.

5.3 Bulk Electronic Structure of InAs and GaAs at 4K and variation with strain

In order to generate the TB parameters, an interface of GA and NEMO described in the previous chapter was used. Diagonal parameter shifts and scaling of two center integrals was obtained from the the theory of Boykin et al. [32] implemented in NEMO.

In Table 5.1 we list the on-site energies and the two center integrals in the Slater-Koster (SK) [49] notation for the sp3d5s* model [50]. The bulk band structure of InAs and GaAs at Low Temperature (4K) and Room Temperature (300K) generated from the parameters of table 5.1 and [32] are shown in Fig 5.1.

From this band structure we extract technologically relevant effective masses and band edges and present them in Table 5.2 along with target masses and edges obtained from experiment and theory [33] [34]. It is evident that the TB parameters thus generated reproduce these masses and band edges accurately. An examination of the band gaps (and band optima) in InAs and GaAs at Low Temperature and Room Temperature shows that the direct and indirect energy gaps at 4K are larger than their 300K counterparts as should be expected. At both 4K and 300K, InAs energies are offset to the valence band of GaAs, taken to be the zero energy reference.

In Table 5.3 we list parameters for strained systems at 4K. It must be noted at the outset that in generating these parameters, an assumption of elasticity has been made, even for higher strain values. While this assumption is not strictly valid, it has been used in the absence of a theory that handles arbitrarily large strains beyond the elastic limit. The SK two center integrals were scaled according to a generalized version of Harrison's $\eta = 2$ scaling law [51]

$$U = U_0(d_0/d)^\eta \tag{5.1}$$

Here U_0 is an ideal unstrained two center integral while d_0 and d are bond lengths in unstrained and strained materials respectively. The constants C representing the shift in onsite energies due to strain are also shown in the table.

Table 5.1
Low Temperature unstrained TB parameters

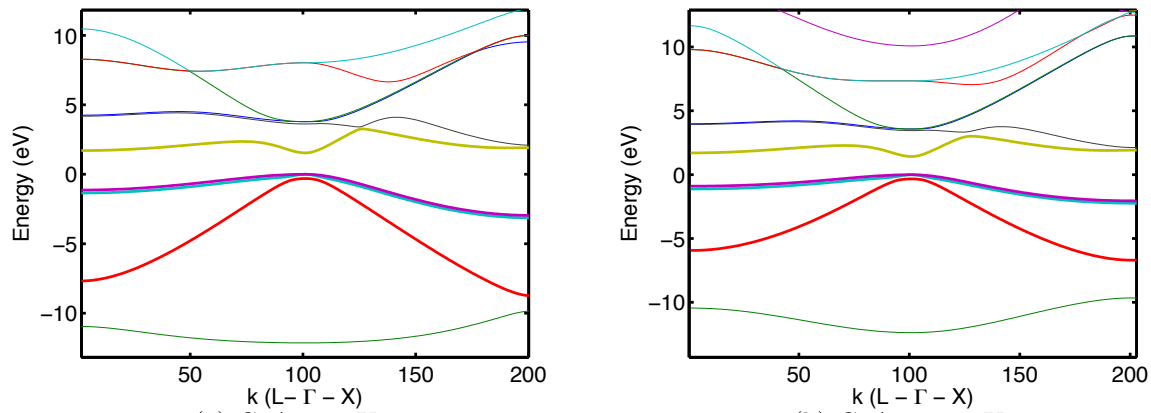
Parameter	InAs	GaAs
E_{cs}	-5.6462	-5.6462
E_{cp}	4.322525	4.322525
E_{css*}	21.17908	21.17908
E_{cdd}	13.21109	13.21109
λ_c	0.184481	0.184481
E_{ass*}	22.44819	20.95755
E_{as}	-1.11625	-0.19341
E_{ap}	5.988797	5.052347
E_{ad}	12.91644	11.89319
λ_a	0.117136	0.021002
$s_a s_c$	-1.77843	-1.77486
$s *_a s*_c$	-4.07846	-2.81242
$s *_a s_c$	-2.61516	-2.40899
$s_a s*_c$	-3.07826	-0.68952
$s_a p_c$	3.091671	2.945532
$s_c p_a$	2.463171	3.618029
$s *_a p_c$	2.129011	2.928687
$s *_c p_a$	0.429437	2.045608
$s_a d_c$	-2.7976	-1.26743
$s_c d_a$	-1.14575	-3.29464
$s *_a d_c$	-2.52751	-0.78685
$s *_c d_a$	3.153259	-0.36524
$pp\sigma$	4.095	4.214353
$pp\pi$	-1.53709	-1.53618
$p_a d_c \sigma$	-1.90317	-2.10917
$p_c d_a \sigma$	-1.167	-1.10345
$p_a d_c \pi$	2.364101	2.024289
$p_c d_a \pi$	2.578959	2.365867
$d_c d_a \sigma$	-2.1672	-1.74846
$d_c d_a \pi$	2.136146	2.35935
$dd\Delta$	-0.99874	-1.36161

Table 5.2
Energies, band gaps and k points at 4K

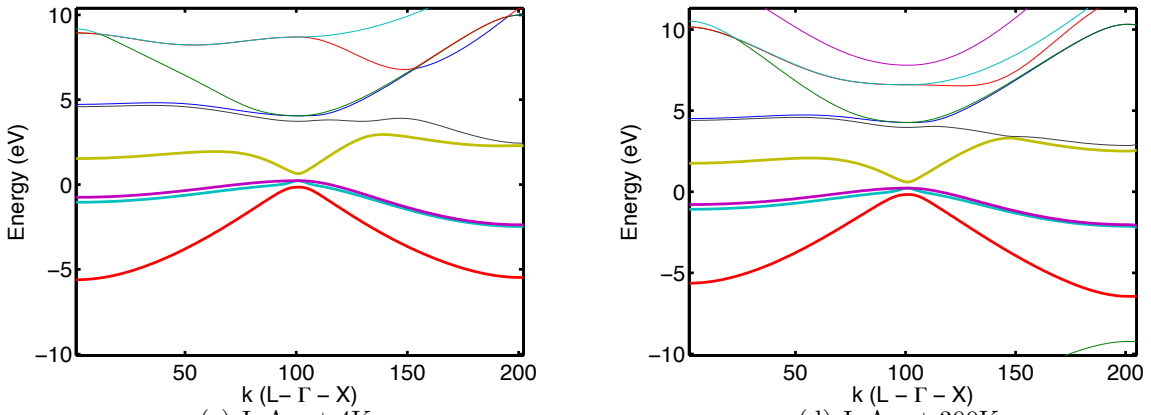
Quantity	GaAs calculated	GaAs Target	% deviation	InAs calculated	InAs Target	% deviation
Eg-gamma	1.5383092	1.5382	0.007099207	0.417766	0.418	0.055980861
Ec-gamma	1.5383087	1.5382	0.007066701	0.6435875	0.6437	0.017477086
Vhh	-0.0000006	0	–	0.2258215	0.2257	0.053832521
Vlh	-0.0000006	0	–	0.2258215	0.2257	0.053832521
Vso	-0.3082936	-0.34	9.325411765	-0.1543033	-0.1543	0.002138691
Delta-so	0.308293	0.34	9.325588235	0.3801247	0.38	0.032815789
E-X	1.899928	1.9	0.003789474	2.2799501	2.28	0.002188596
E-L	1.7079764	1.708	0.001381733	1.5300286	1.53	0.001869281
k-X	0.9	0.9	0	0.9	0.9	0
k-L	1	1	0	1	1	0
Electrons						
mstar-c-001	0.0658386	0.067	1.733432836	0.0229801	0.0239	3.848953975
mstar-X-long	1.3003649	1.3	0.028069231	1.1751791	1.3	9.601607692
mstar-X-trans	0.2288044	0.23	0.519826087	0.1603539	0.23	30.28091304
mstar-L-long	1.8996865	1.9	0.0165	1.6289335	1.9	14.26665789
mstar-L-trans	0.1102078	0.0754	46.16419098	0.0802688	0.0754	6.45729443
Holes						
mstar-lh-001	-0.0826915	-0.0871	5.061423651	-0.0281261	-0.0273	3.026007326
mstar-lh-011	-0.0731901	-0.0804	8.967537313	-0.0270091	-0.0264	2.30719697
mstar-lh-111	-0.0709123	-0.0786	9.780788804	-0.0266759	-0.0261	2.20651341
mstar-hh-001	-0.3106723	-0.403	22.91009926	-0.3258846	-0.3448	5.485904872
mstar-hh-011	-0.6059149	-0.66	8.194712121	-0.6236339	-0.6391	2.419981224
mstar-hh-111	-0.8233913	-0.813	1.278142681	-0.8766644	-0.8764	0.030168873
mstar-so-001	-0.1511486	-0.15	0.765733333	-0.0942714	-0.15	37.1524

Table 5.3
Low Temperature strain parameters

Parameter	InAs	GaAs
$\eta_{ss*\sigma}$	0	0
$\eta_{s*s*\sigma}$	2.122098	5
$\eta_{ss\sigma}$	0	0
$\eta_{sp\sigma}$	0	1.151641
$\eta_{pp\sigma}$	0.604445	4.338978
$\eta_{pp\pi}$	0	0
$\eta_{sd\sigma}$	0	0
$\eta_{s*p\sigma}$	2.490183	0.238984
$\eta_{pd\sigma}$	4.442049	3.100839
$\eta_{pd\pi}$	0.938528	0
η_{const}	1.507385	1.196034
$\eta_{dd\sigma}$	2.700428	2.410947
$\eta_{dd\pi}$	2.507279	0
$\eta_{dd\Delta}$	0	0
$\eta_{s*d\sigma}$	2.436776	5
C_{ss}	0	2.223464
C_{s*s*}	3.299733	5
$C_{s*_a s}$	4.867499	0
$C_{s_a s*}$	0	4.932335
$C_{s_a p_c}$	5	3.462799
$C_{p_a s_c}$	4.842725	0
$C_{s*_a p_c}$	5	0
$C_{p_a s*_c}$	3.525504	0
$C_{s_a d_c}$	0	0.646123
$C_{d_a s_c}$	5	2.032087
$C_{s*_a d_c}$	0	4.902844
$C_{d_a s*_c}$	5	0
C_{pp}	5	4.973038
$C_{p_a d_c}$	0	0.655428
$C_{d_a p_c}$	5	4.255733
C_{dd}	4.731473	0



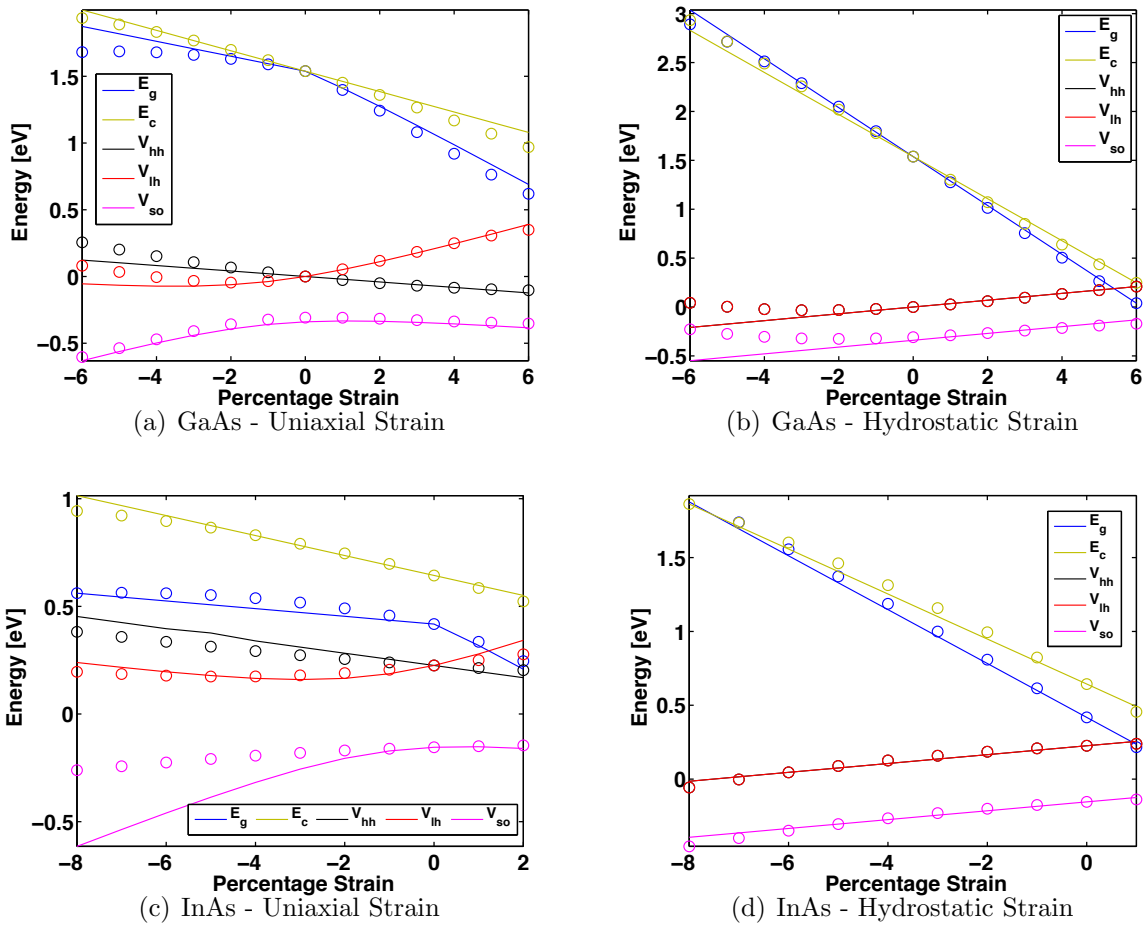
(a) GaAs at 4K. (b) GaAs at 300K.



(c) InAs at 4K. (d) InAs at 300K.

Fig. 5.1. Bulk Zero-strain Band Structures for InAs and GaAs generated by the parameters in Table 5.1 and 5.2

Band edges (direct and indirect) and gaps and their variation with strain are of prime importance in the modeling of InAs/GaAs quantum dots and these were given large weights in the fitting of parameters. Figure 5.2 shows the variation of the conduction band minimum at Γ the valence band maxima for the heavy hole, light hole and split off bands with hydrostatic and uniaxial [001] strain. Our results follow van de Walle's theoretical parameterization quite closely both qualitatively and quantitatively.



(a) GaAs - Uniaxial Strain

(b) GaAs - Hydrostatic Strain

(c) InAs - Uniaxial Strain

(d) InAs - Hydrostatic Strain

Fig. 5.2. Band gaps and band edges at Gamma for GaAs and InAs at 4K. Solid lines indicate theoretical fits. Circles are our calculated values.

It is interesting to note that the variation of the direct band gap with uniaxial strain is quite small in magnitude as compared to its corresponding variation with hydrostatic strain. This observation directly affects our study of InAs/GaAs quantum dots, since it must be expected that apart from confinement, the fact that the InAs dot is strained plays an important role in determining the optical gap of these dots. In addition to these results, Fig. 5.3 shows the variation of the bulk band gap with temperature for InAs and GaAs as obtained with our TB parameterization. We also obtained three independent parameterizations to fit Varshni's relation [52] as

shown in Fig. 5.3. We treated these parameterizations independently of each other, i.e. no attempt was made to find a pattern in the variation of TB parameters with temperature. While the gaps at 77K and 4K were chosen to fit the Varshni relation quite closely, the previous parameterization of Boykin et al. [32] chooses the gap at 300K to be 0.37 eV. Since we used the unstrained parameters mentioned in this work, our calculated gap at 300K is about 20meV higher than that expected from the Varshni relation.

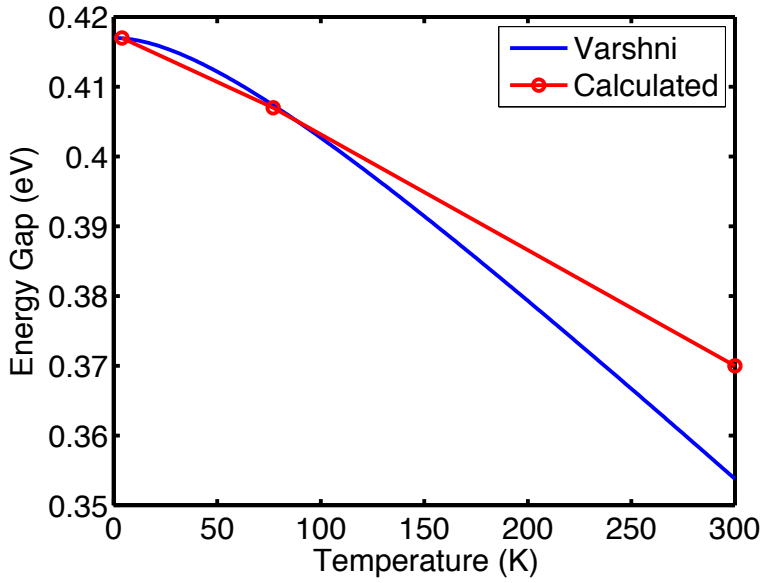


Fig. 5.3. Energy gaps at Γ: Varshni relation vs. our calculated values

To ensure that our TB parameterization behaved correctly under strain and confinement and produced larger band gaps than room temperature at all strains, we simulated free standing, rectangular InAs cubes of edge 5 nm in NEMO-3D at 4K and 300K. We manually adjusted the lattice constants of the InAs box to correspond to hydrostatic strain, which was varied from 0 to -8 percent. Figure 5.4 shows the variation of the band gap of this free-standing box versus strain and temperature. The optical gap (difference between ground state electron and hole levels in the box) shows a linear increase with increasing compressive strain at both 4K and 300K.

The optical gap at 4K is uniformly larger than the 300K value. Since the spatial confinement is the same in all 3 cases, the variation with hydrostatic strain should be expected to qualitatively follow bulk values (not quantitatively, since in bulk we assume periodic boundary conditions, while we assume closed boundary conditions in the box).

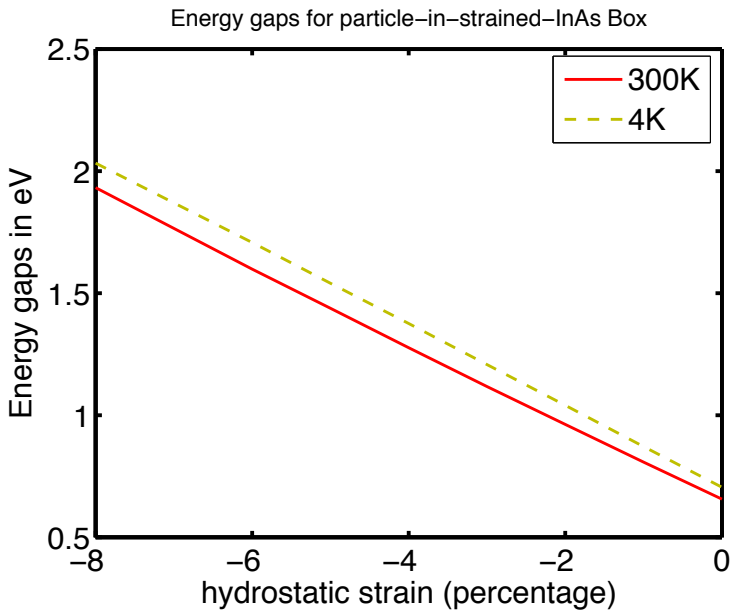
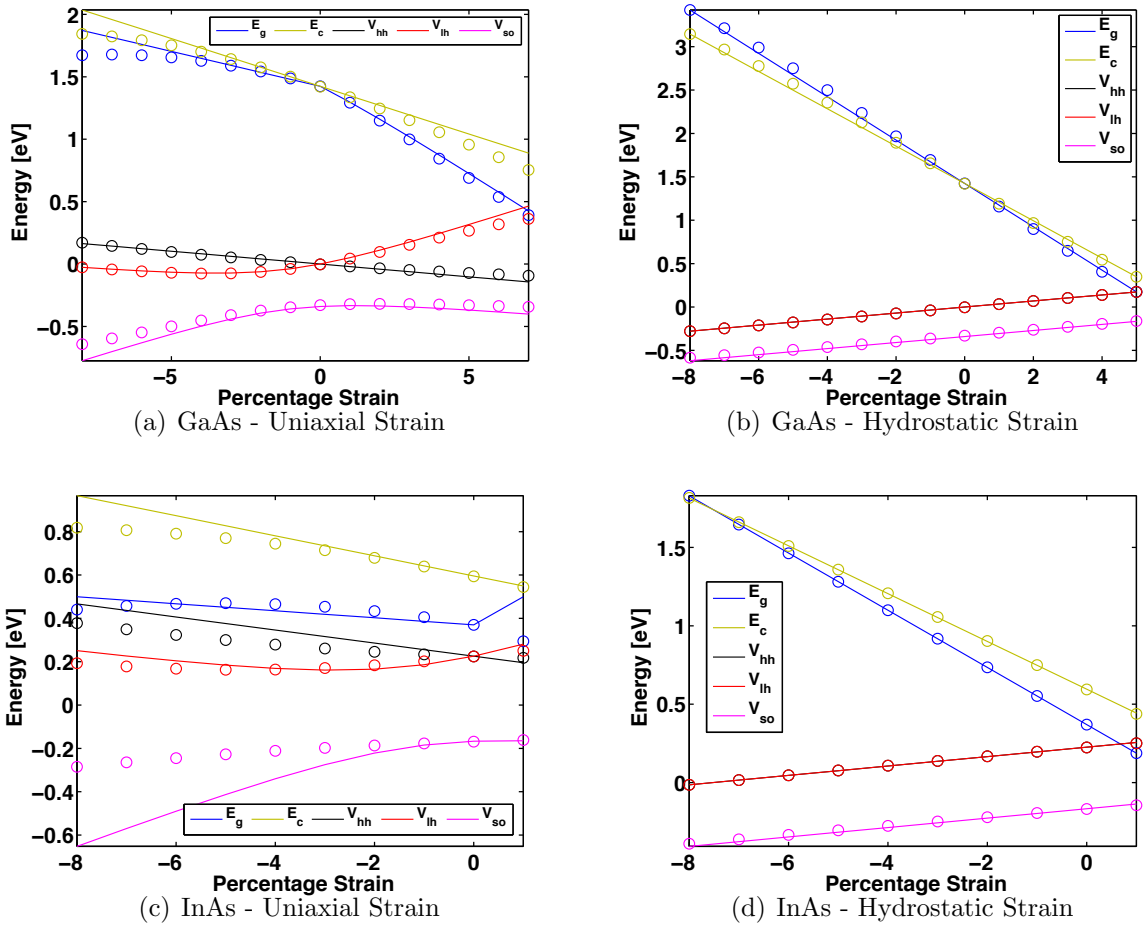


Fig. 5.4. Optical gaps (difference between ground state electron and hole levels) in a free standing cubic InAs box of side 5 nm

In the course of this study, it was found that the 300K parameterization reported in [32] overestimated optical gaps at hydrostatic strains greater than 5 percent. The strain parameters ηs and Cs were then modified to ensure correct behavior at all strains keeping the unstrained two center integrals and on-site energies the same. The corrected values of these parameters at 300K is presented in Table 5.4 It was ensured that the signs of the TB parameters and their relative magnitudes correspond to physically feasible values by restricting the range that these parameters could take in the optimization to physically meaningful values. In Fig 5.5 we show the variation of gaps and energies at Γ obtained from the 300K TB strain parameterization for uniaxial [001] and hydrostatic strain.



(a) GaAs - Uniaxial Strain

(b) GaAs - Hydrostatic Strain

(c) InAs - Uniaxial Strain

(d) InAs - Hydrostatic Strain

Fig. 5.5. Band gaps and band edges at Gamma for GaAs and InAs at 300K. Solid lines indicate theoretical fits. Circles are our calculated values.

Having ensured (to the extent possible) the correctness of the TB parameters thus generated, we now investigate the atomic structure of realistic InAs/GaAs quantum dots in NEMO-3D.

5.4 Atomic Structure of InAs/GaAs quantum dots

Before beginning a discussion on atomic structure it is imperative to say a few words about the simulator NEMO-3D that is used for these calculations. NEMO-3D is

Table 5.4
Room temperature strain parameters

Parameter	InAs	GaAs
$\eta_{ss*\sigma}$	3.413554	3.061963
$\eta_{s*s*\sigma}$	2.313446	3.611876
$\eta_{ss\sigma}$	3.998445	1.204155
$\eta_{sp\sigma}$	0.934251	4.8
$\eta_{pp\sigma}$	0	2.560341
$\eta_{pp\pi}$	0	1.2
$\eta_{sd\sigma}$	0	1.832811
$\eta_{s*p\sigma}$	1.23561	4.8
$\eta_{pd\sigma}$	3.690212	4.704001
$\eta_{pd\pi}$	0.853037	4.159215
η_{const}	1.858506	1.843388
$\eta_{dd\sigma}$	5.024782	2.589812
$\eta_{dd\pi}$	0	4.8
$\eta_{dd\Delta}$	0.501759	1.2
$\eta_{s*d\sigma}$	5	4.8
C_{ss}	3.06707	2.264256
C_{s*s*}	0	0
$C_{s*_a s}$	0.371833	2.421205
$C_{s_a s*}$	1.881896	0
$C_{s_a p_c}$	0	1.411449
$C_{p_a s_c}$	4.389994	2.995678
$C_{s*_a p_c}$	1.180655	0
$C_{p_a s*_c}$	3.243818	1.828038
$C_{s_a d_c}$	2.897887	0
$C_{d_a s_c}$	3.559295	0.364774
$C_{s*_a d_c}$	0	2.007091
$C_{d_a s*_c}$	4.791748	4
C_{pp}	3.925621	2.154247
$C_{p_a d_c}$	0.898993	0
$C_{d_a p_c}$	3	3
C_{dd}	2.936632	0

an atomistic, 3-D, nano-scale simulator. It can calculate the electronic structure, carrier wavefunctions, phonon dispersions, strain distribution and number of such physically useful quantities for a variety of nano-structures. It utilizes two strain models (to be described in this section) for atomic structure calculations and a Rayleigh-Ritz or Lanczos solver for electronic structure calculations. It incorporates a variety of boundary conditions for both electronic structure and atomic structure calculations. The simulator is described in detail in Klimeck et al. [30] Realistic InAs/GaAs quantum dots contain several millions of atoms. While it is desirable to simulate the growth of these dots to match experimental techniques and capture inhomogeneous strain effects, the enormous number of atoms in the system makes such a calculation unfeasible. NEMO-3D uses two strain models - a quasi-harmonic Valence Force Field (VFF) - Keating model [53] [54] and an anharmonic VFF model [55] [56] to describe the strain energy in nanostructures. We now describe the two strain models.

5.4.1 The quasi-harmonic Keating model

The quasi-harmonic Keating model calculates the local strain energy at atomic position i as:

$$E_i = \frac{3}{16}\sum_j \left[\frac{\alpha_{ij}^2}{d_{ij}^2}(R_{ij}^2 - d_{ij}^2)^2 + \sum_{k>j}^{n} \frac{\sqrt{\beta_{ij}\beta_{ik}}}{d_{ij}d_{ik}}(\mathbf{R}_{ij}.\mathbf{R}_{ik} - \mathbf{d}_{ij}.\mathbf{d}_{ik})^2\right] \tag{5.2}$$

where the sum is over neighbours j of atom i. Here d_{ij} and R_{ij} are the equilibrium and actual distances between atoms i and j respectively. The parameters α and β represent force constants for bond length and bond angle distortions respectively. In the absence of Coulomb corrections, they are related to the bulk elastic moduli by

$$C_{11} + 2C_{12} = \frac{\sqrt{3}}{4d_{ij}}.(3\alpha_{ij} + \beta_{ij}) \tag{5.3}$$

$$C_{11} - C_{12} = \frac{\sqrt{3}}{d_{ij}}.\beta_{ij} \tag{5.4}$$

$$C_{44} = \frac{\sqrt{3}}{4d_{ij}}.\frac{4\alpha_{ij}\beta_{ij}}{\alpha_{ij} + \beta_{ij}} \tag{5.5}$$

In zincblende materials, modified relations to account for Coulomb effects due to unequal charge distribution between anion and cation must be used. α and β obtained by Martin [57] are used in this work.

5.4.2 The anharmonic Keating model

The anharmonic Keating model introduces corrections to the force constants α and β used in the VFF model. In the anharmonic model, α and β are no longer constants, but are dependent on strain as follows:

$$\alpha_{ij} = \alpha_0^{ij}\left[1 - A\frac{(R_{ij}^2 - d_{ij}^2)^2}{d_{ij}^2}\right] \tag{5.6}$$

$$\beta_{ijk} = \beta_0^{ijk}\left[1 - B(cos\theta^{ijk} - cos\theta_0^{ijk})\right] \bullet \left[1 - C\frac{(r_{ij}r_{ik} - d_{ij}d_{ik})}{d_{ij}d_{ik}}\right] \tag{5.7}$$

Here α_0^{ij} and $\beta_0^{ijk} = \sqrt{\beta_0^{ij}\beta_0^{ik}}$ are the same force constants as in the harmonic VFF model. $cos\theta^{ijk}$ and $cos\theta_0^{ijk}$ are the angles between bonds formed by atoms having indices i,j,k in the strained and unstrained cases respectively. Constants A and C describe the dependence of α and β on hydrostatic strain, while B describes the dependence of beta on bi-axial strain. The values of A, B and C are taken from Lazarenkova et al. [55]

A schematic representation of the strain energy versus bond length distortions for the two models in shown in Figure 5.5 adapted from Lazarenkova et al. [56]. This picture, however, is inadequate as it does not describe the effect of bond-angle distortions. A more complete profile of strain energy versus arbitrary bond length and bond angle distortions is shown in the figure 5.5 as a coloured contour plot. This more complete picture serves to illustrate the essential difference between the two models. The quasi-harmonic strain energy shows a parabolic variation with both bond angle and bond length distortions in both compressive and tensile regimes. The anharmonic model, on the other hand shows a parabolic variation with compressive bond length distortions but is not parabolic versus bond angle variations. In addition,

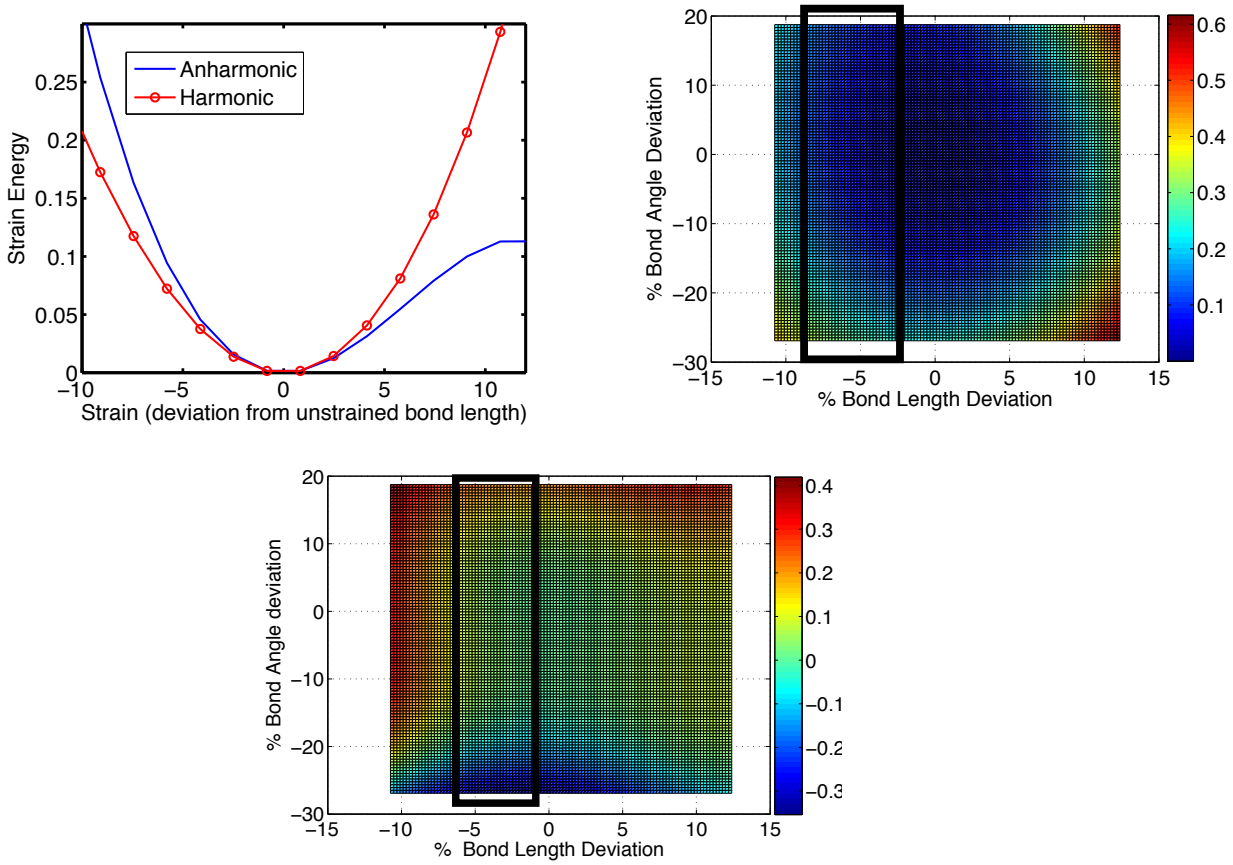


Fig. 5.6. Schematic sketch of Strain energy versus (a)bond length distortions (b) bond length and bond angle distortions in the harmonic model and(c) bond length and bond angle distortions in the anharmonic model. The black rectangle indicates our region of interest, as will be explained later on in this chapter.

the curvature of the parabolic variation with bond length distortions is greater in the harmonic model than the anharmonic model.

In NEMO-3D the device is initially setup with both InAs and GaAs set to the lattice constant of GaAs. This implies that the InAs is lattice mismatched by about 7 percent and is strained severely as a result. The total strain energy in the structure is calculated according to the above models and is then minimized by using a conjugate gradient energy minimization at each atomic position in the structure. The relaxed atomic structure so obtained is then used for electronic structure calculations. The

electronic structure of the InAs/GaAs dots obtained after using relaxed atomic positions obtained from the two different models differs in terms of energy offsets and absolute energy gaps. In the following section we discuss these results in details and offer explanations for the differences in results.

5.5 Electronic Structure of InAs/GaAs quantum dots

Once the relaxed atomic positions are known, an atomistic Tight Binding Hamiltonian is constructed for the quantum dot. NEMO-3D allows for a variety of boundary conditions. For the purpose of this study we used periodic boundary conditions, implying that our system is essentially not a single quantum dot but a 3-D periodic array of quantum dots. The eigenvalues and eigenvectors of this Hamiltonian are then calculated. NEMO-3D has a variety of parallel eigenvalue solvers as described in [30] and we use a parallel Lanczos solver for our calculations.

Before describing our results in detail, it must be noted that there are several experimental uncertainties that prevent absolute matching of results. The primary experimental uncertainty is that the amount of Ga that has diffused into the dot during nucleation (formation of the dot)and overgrowth (surrounding InAs with GaAs) is unknown experimentally [48]. The diffusion of Ga into the dot has the effect of increasing the optical band gap and altering the overall electronic structure. Another experimental uncertainty is the size of the dot. We have varied the sizes of simulated dots in order to make sure we get results that have correct qualitative trends. Finally, we compute strain in our models by initially setting the lattice constant of the InAs to equal that of GaAs. The initial stages of growth of an InAs quantum dot (before surrounding the dot with GaAs) are volume conserving steps [58]. This implies that the in-plane compressive strain in InAs must be accompanied by a out-of-plane tensile strain. While deposition of large quantities of GaAs after the growth of the dot forces the InAs to strain compressively on an average, it must be expected that there will some tensile strain in the dot due to the initial relaxation out of the plane. However,

in NEMO-3D, relaxation from an initially large compressive strain in InAs ensures that no InAs bonds will be tensile in nature on relaxation, since they are surrounded and outnumbered by GaAs several times over.

NEMO-3D enables the user to choose a variety of outputs mentioned previously. Our chief interest was in calculating the excitonic gaps for our structures. The excitonic gap calculated by NEMO-3D is essentially the difference between electron and hole ground states calculated assuming a single electron TB model. In order to model the excitonic gaps correctly, an additional correction due to excitonic Coulomb interactions must be subtracted from the NEMO-3D result. In the dots we simulated this correction is between 30-50 meV as reported in another theoretical study [48].

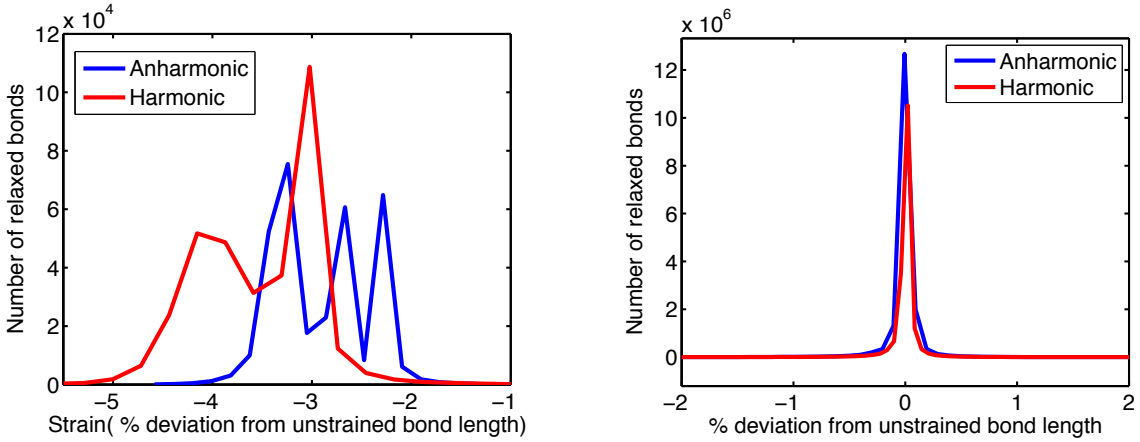


Fig. 5.7. Distribution of bonds in the relaxed structure for (a) InAs (b) GaAs at 4K

We simulated a variety of quantum dots using the low temperature TB parameters and compared their excitonic gaps with experimental and theoretical results. When the harmonic Keating model was used to model strain, the resulting excitonic gaps were higher than experiment by about 100 meV, while the anharmonic model always under-estimated the band gap by about 50 meV. To understand why this may be happening, we conducted a numerical experiment. We set up the same device structures having dimensions given in Table 5.5 and allowed the atoms in the structure to relax. On relaxation, we computed In-As bond lengths and plotted the districation

of InAs bonds versus their precentage deviation from the unstrained bond length as shown in Fig. 5.7. It can be seen from this figure that the anharmonic model results in a strain distribution that is shifted towards lower strain values as compared to the harmonic model. It can also be seen, as was previously hypothesized, that all In-As bonds are strained compressively and a majority are between 2-5 percent compressively strained. Also, GaAs is strained compressively and tensile-ly within 1 percent of its unstrained value. From the schematic picture of the strain energy versus bond length distortion, one can conclude that this shift towards lower total strain is due to the larger curvature of the anharmonic model in the compressive regime. A larger curvature and near parabolic nature in this regime implies an α that is much larger than that calculated from the harmonic model. Comparing this situation to a compressed spring with the same initial distortion and larger spring constant, one can see why the eventual distortion in the anharmonic model will always be lesser than the harmonic model.

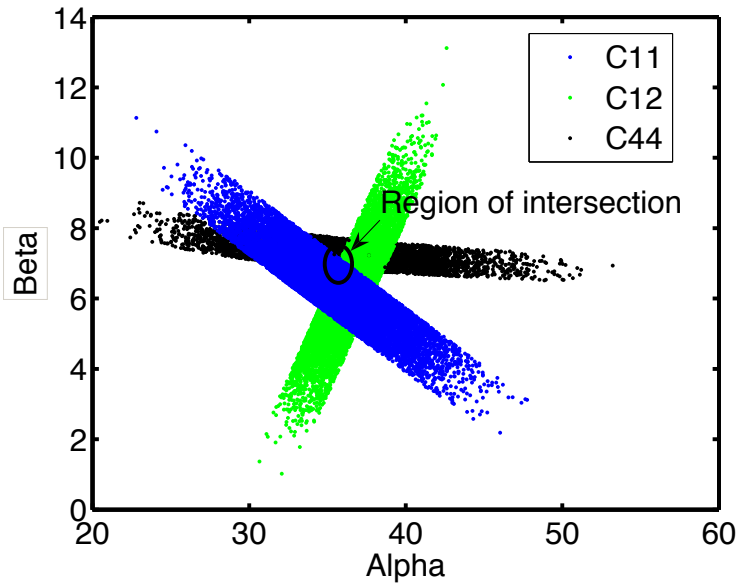


Fig. 5.8. Orthogonal projections of C_{11}, C_{12}, C_{44} versus α and β

Table 5.5
Optical gaps at 4K: calculated values versus experiment

Optical Gap (meV)				
Dot dimensions (Base nm X width nm)	Keating - Harmonic	Keating - Anharmonic	Keating - Anharmonic (Fit to Harmonic)	Experiment/Theory From [48] [59]
15 X 2.5	1238	1133	1207	1078
20 X 6	1184	1000	1100	1098
25.2 X 3.5	1164	1040	1132	1032
25.2 X 2.5	1178	1059	1153	1131
27.5 X 3.5	1154	1020	1125	1016

In the absence of detailed information about Ga diffusion into the InAs dot, we assume pure InAs dots. This is a worst case scenario in reality given the procedures of thermal annealing that are carried out to manufacture these dots. The fact that the anharmonic model under-estimates the excitonic gap even in the absence of the Coulombic correction implies that we need to tune the α and β in the anharmonic model to have a smaller curvature and thereby lead to larger eventual strains in InAs and larger optical gaps. It must be ensured, however that the relations of equation (5.3) through (5.5) are fit as closely as possible. Using αs and βs that do not fit elastic constants implies that we are calculating strains for a different material altogether. Fig. 5.8 shows the result of such a fitting procedure done using a GA. The figure is an orthogonal projection of the values of C_{11}, C_{12} and C_{44} versus α and β. Only those αs and βs that resulted in elastic constants within 5 percent of their experimental values at 4K were used. The region of intersection of the surfaces corresponds to those αs and βs for which all three elastic constants are simultaneously within 5 percent of experiment. Choosing the new values of α and β to be 35.78 and 6.94 respectively, we simulated the same dots once again and we found significant increase in the optical gap towards experimental values. These results are presented in table 5.5. We also varied the size of these dots in order to compare our results with previous theoretical results. In general, we find good agreement between our results and the results of Williamson et al [48]. On varying the size of the dots, our results show the same trends and are numerically in close agreement with experiment and theory.

The next chapter discusses other projects attempted and future directions for the work described in this thesis.

6. ATTEMPTED AND FUTURE WORK

The GA is an extremely versatile optimization algorithm. In this chapter we discuss some other optimization projects done using the GA in the course of this work. We also discuss some of the deficiencies in the current implementation of PGAPack and suggest improvements that can be undertaken.

6.1 Optimization engine for the nanoHUB

The nanoHUB is a web-based resource for research, education and collaboration in nanotechnology [60]. The nanoHUB contains several research-quality simulation tools in several sub-fields in nanotechnology and requires nothing but a Java enabled web-browser in order to use these tools. It was envisaged to create an optimization engine for the simulation tools on the nanoHUB keeping in mind the frequent need for optimization mentioned previously. The following are some of the desired features of an optimization toolbox for a multi-tool environment like the nanoHUB:

1) The toolbox must interface with Rappture [61], the architecture driving the tools on the nanoHUB.

2) The toolbox must be modular. Rather than create a specific optimization toolbox for each tool, it is preferable to keep the optimization toolbox separate and invoke it from each tool through a Rappture interface.

3) If point (2) holds it implies that there should be a unique and transparent way of specifying fitness for each tool without being tool-dependent.

4) Since tools have a varying number of parameters, it is desired that the optimization toolbox introduce parallelism in its execution to speed up the execution process. This involves two discrete steps: the first step involves the selection of an optimization method that can be parallelized, while the second step requires Rappture to be

able to execute individual tools in parallel. The subtle difference between the two steps is further explained below.

Our previous experience with the inherent parallelism of the GA made it an automatic choice for implementing the core optimization algorithm. In order to interface the GA core to the Rappture interface, the Rappture optimization API (Application Programming Interface) was created with the above requirements in mind, on which subsequent modifications were made and additional wiring was added for greater functionality in terms of control provided to end-users of the optimization toolbox. Features (2) and (3) above presented themselves as unique challenges never implemented elsewhere. The Rappture interface is currently able to launch a single tool at a time. In order to have a parallel optimization operation, it would be necessary to launch several runs of a single tool in parallel, each run containing different input parameters for the same tool (corresponding to individuals/chromosomes). Some efforts have been made in this direction. However, this exercise is considerably time consuming from the point of view of development and it was decided to designate this task as work that could be undertaken in the future. Problem (2) is also designated as a task that can be undertaken in the future. A serial execution of Genetic Algorithms that interfaces with Rappture has been demonstrated for the Quantum Dot Lab tool on the nanoHUB.

In addition to Genetic Algorithms other parallel optimization approaches including some of the approaches mentioned in earlier chapters could be integrated into the Rappture API and the API has a provision for this purpose.

6.2 TB parameterization of Antimonides

The lack of a suitable strain parameterization better than the work of Jancu et al. [50] for Antimonide materials led to an attempt at parameterizing InSb, GaSb and AlSb. As of now, we have 300K TB parameters for each of these semiconductors

when they are bulk, unstrained materials. Strain parameterization is underway and the results are awaited.

6.3 Force Field Optimization

In place of time consuming ab-initio calculations for materials, it is desirable to model the forces between atoms in a material by using force fields and then obtain strained relaxed energy structures using the force fields so parameterized. We are currently working on a parallel GA immplementation of a ForceField optimization for Ni-Ni and Ni-Ti alloys. Serial GA runs have so far failed to yield minimum strain, energy and force structures and a parallel implementation is therefore desired.

6.4 Deficiences in current implementation of the GA and scope for future work

The main deficiency in the current implementation of the GA by PGAPack is the sensitivity of the material parameterization to the weights assigned to various outputs in the fitting process. It was frequently observed that minor variation in weights often led to entirely unpredictable results. Also, the assignment of weights in the process is arbitrary and can never be considered as something that can be adopted across the board for all materials. This senstivity affects the amount of time required to obtain an accurate material parameterization. As mentioned in chapter 5, the GA treats the parameterization as an optimization problem. In order to obtain physically meaningful values, it is necessary for the designer to carefully assign ranges of allowed values for parameters. It is therefore desirable that the parameters be derived from a more detailed and physically meaningful ab-initio calculation.

7. SUMMARY AND CONCLUSIONS

In chapter 1 of this thesis we motivated the need for optimization in nano-electronics. Chapter 2 was an overview of some optimization techniques and their suitability to the problem of TB parameterization. This led to the motivation for the selection of the GA as our optimization algorthm. Chapter 3 discussed the working of the Genetic Algorithm. We saw how features such as crossover and mutation helped towards getting optimum solution. Chapter 4 discussed some of the motivations behind the use of the TB approximation and the methodology of generating and optimizing these parameters using the GA. Chapter 5 presented the main results of this work, parameter sets for InAs and GaAs at 4K and strain parameter sets for InAs and GaAs at 300K. In this regard, some of the main conclusions are as follows:

1. The parameterizations obtained for InAs and GaAs at 4K are physically meaningful. The band energies and effective masses at high symmetry points match experimental and theoretical values quite well.

2. New 300K strain parameters were obtained to correct deficiencies in earlier parameterization.

3. The 4K and 300K strain parameterizations match the theoretical values of van de Walle [38] quite closely.

4. No effort was made to find a pattern in the variation of the TB parameters with temperature. Both parameterizations were treated as independent problems.

5. Atomic structure effects were found to be particularly important during the validation of the TB parameterization. In particular, effects of parabolicity of the strain energy were studied. It was found that in almost all cases, the harmonic model over-estimated the optical gap of InAs/GaAs quantum dots,

while the anharmonic model significantly underestimated the optical gap, in spite of not adding corrections for excitonic Coulomb interactions.

6. New values of α and β were obtained for the anharmonic model by fitting it to the harmonic model in the compressive regime. It was found that this brings the optical gap closer to experiment.

7. Further investigation into the role of atomic structure is needed, although, eventually, experimental uncertainties like dot composition and size prevent us from modeling these dots with absolute accuracy.

Eventually, we discussed some other work undertaken as part of our work with the GA and some scope for further work in this area.

LIST OF REFERENCES

LIST OF REFERENCES

[1] R. Steele, *Consumer applications boost laser sales 10 percent*, 2005.

[2] G. Klimeck, M. Mannino, M. McLennan, W. Qiao, D. Ebert, X. Wang, and B. Haley, "Quantum dot lab," Nov 2005.

[3] W. W. Chow, S. W. Koch, and M. Sargent, III, *Semiconductor-laser physics*. New York, NY, USA: Springer-Verlag New York, Inc., 1994.

[4]

[5] G. Bir and G. Pikus, *Symmetry and strain-induced effects in semiconductors*. Wiley New York, 1974.

[6] W. Harrison, *Electronic structure and the properties of solids: the physics of the chemical bond*. Dover Publications New York, 1989.

[7] C. Cohen-Tannoudji, B. Diu, and F. Laloe, *Quantum mechanics. vol. 1-2*. New York.

[8] E. Kaxiras, *Atomic and electronic structure of solids*. Cambridge Univ Pr, 2003.

[9] R. Martin, *Electronic structure: basic theory and practical methods*. Cambridge Univ Pr, 2004.

[10] E. Gross and R. Dreizler, *Density functional theory*. Springer, 1995.

[11] A. Jameson, *Gradient Based optimization methods*, 1995.

[12] J. Mor, "The Levenberg-Marquardt algorithm: implementation and theory," *Lecture notes in mathematics*, vol. 630, pp. 105–116, 1977.

[13] A. e. a. Bjoerck, *Numerical methods for least squares problems*. Society for Industrial Mathematics, 1996.

[14] S. Kirkpatrick, "Optimization by simulated annealing: Quantitative studies," *Journal of Statistical Physics*, vol. 34, no. 5, pp. 975–986, 1984.

[15] M. R. Lemes, C. R. Zacharias, and A. Dal Pino, "Generalized simulated annealing: Application to silicon clusters," *Phys. Rev. B*, vol. 56, pp. 9279–9281, Oct 1997.

[16] D. Greening, "Parallel simulated annealing techniques," *Physica D Nonlinear Phenomena*, vol. 42, pp. 293–306, 1990.

[17] R. Schneider, R. Weiss, and M. Schmid, "Parallel Simulated Annealing," 2007.

[18] B. Chakrabarti and A. Das, *Quantum annealing and related optimization methods*. Springer.

[19] J. Kennedy, R. Eberhart, *et al.*, "Particle swarm optimization," in *Proceedings of IEEE international conference on neural networks*, vol. 4, pp. 1942–1948, Piscataway, NJ: IEEE, 1995.

[20] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 26, no. 1, pp. 29–41, 1996.

[21] C. Fonseca, P. Fleming, *et al.*, "Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization," in *Proceedings of the fifth international conference on genetic algorithms*, pp. 416–423, Citeseer, 1993.

[22] D. Levine, "Users guide to the PGAPack parallel genetic algorithm library," *Argonne National Laboratory*, vol. 95, no. 18, pp. 1–77, 1996.

[23] G. Klimeck, R. Bowen, T. Boykin, C. Salazar-Lazaro, T. Cwik, and A. Stoica, "Si tight-binding parameters from genetic algorithm fitting," *Superlattices and Microstructures*, 1999.

[24] C. Darwin, "On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life," *New York: D. Appleton*, 1859.

[25] G. Mendel, *Experiments in plant hybridisation*. COSIMO CLASSICS, 2008.

[26] J. Watson and F. Crick, "Genetical implications of the structure of deoxyribonucleic acid," *Inspiring science: Jim Watson and the age of DNA*, p. 74, 2003.

[27] H. Pearson, "Genetics: what is a gene?," *Nature*, vol. 441, pp. 398–401, 2006.

[28] J. H. Holland, *Adaptation in natural and artificial systems*. Cambridge, MA, USA: MIT Press, 1992.

[29] D. Goldberg and J. Holland, "Genetic algorithms and machine learning," *Machine Learning*, vol. 3, no. 2, pp. 95–99, 1988.

[30] G. Klimeck, F. Oyafuso, T. Boykin, R. Bowen, and P. von Allmen, "Development of a nanoelectronic 3-D (NEMO 3-D) simulator for multimillion atom simulations and its application to alloyed quantum dots," *CMES-Computer Modeling in Engineering and Sciences(1526-1492)*, vol. 3, no. 5, pp. 601–642, 2002.

[31] D. Blanks, G. Klimeck, R. Lake, D. Jovanovic, R. Bowen, C. Fernando, W. Frensley, and M. Leng, "NEMO: General release of a new comprehensive quantum device simulator," in *CONFERENCE SERIES-INSTITUTE OF PHYSICS*, vol. 156, pp. 639–642, IOP PUBLISHING LTD, 1998.

[32] T. Boykin, G. Klimeck, R. Bowen, and F. Oyafuso, "Diagonal parameter shifts due to nearest-neighbor displacements in empirical tight-binding theory," *Physical Review B*, vol. 66, no. 12, p. 125207, 2002.

[33] O. Madelung, *Semiconductors: Data Handbook*. Springer Verlag, 2004.

[34] I. Vurgaftman, J. Meyer, and L. Ram-Mohan, "Band parameters for III–V compound semiconductors and their alloys," *Journal of applied physics*, vol. 89, p. 5815, 2001.

[35] S. Adachi, *Physical Properties of III-V Semiconductor Compounds: InP, InAs, GaAs, GaP, InGaAs, and InGaAsP*. Wiley-Interscience, 1992.

[36] K. Rim, R. Anderson, D. Boyd, F. Cardone, K. Chan, H. Chen, S. Christansen, J. Chu, K. Jenkins, T. Kanarsky, *et al.*, "Strained Si CMOS (SS CMOS) technology: opportunities and challenges," *Solid State Electronics*, vol. 47, no. 7, pp. 1133–1139, 2003.

[37] C.-h. Chiu, Z. Huang, and C. T. Poh, "Formation of nanostructures by the activated stranski-krastanow transition method," *Phys. Rev. Lett.*, vol. 93, p. 136105, Sep 2004.

[38] C. Van de Walle, "Band lineups and deformation potentials in the model-solid theory," *Physical Review B*, vol. 39, no. 3, pp. 1871–1883, 1989.

[39] J. Horn, N. Nafpliotis, D. Goldberg, *et al.*, "A niched Pareto genetic algorithm for multiobjective optimization," in *Proceedings of the first IEEE conference on evolutionary computation*, vol. 1, p. 418, Citeseer, 1994.

[40] S. Tang, S. Lin, and S. Lee, "Near-room-temperature operation of an InAs/GaAs quantum-dot infrared photodetector," *Applied Physics Letters*, vol. 78, p. 2428, 2001.

[41] O. Schmidt, T. Kirstaedter, N. Ledentsov, D. Bimberg, V. Ustinov, A. Egorov, A. Zhukov, M. Maximov, P. Kop'ev, and Z. Alferov, "InAs/GaAs quantum dot lasers," in *Indium Phosphide and Related Materials, 1996. IPRM'96., Eighth International Conference on*, pp. 727–730, 1996.

[42] K. Osborn, M. W. Keller, and R. Mirin, "Single-electron transistor spectroscopy of InGaAs self-assembled quantum dots," *Physica E: Low-dimensional Systems and Nanostructures*, vol. 21, no. 2-4, pp. 501–505, 2004.

[43] S. Cortez, O. Krebs, S. Laurent, M. Senes, X. Marie, P. Voisin, R. Ferreira, G. Bastard, J. Gerard, and T. Amand, "Optically driven spin memory in n-doped InAs-GaAs quantum dots," *Physical review letters*, vol. 89, no. 20, p. 207401, 2002.

[44] S. Li, J. Xia, J. Liu, F. Yang, Z. Niu, S. Feng, and H. Zheng, "InAs/GaAs single-electron quantum dot qubit," *Journal of Applied Physics*, vol. 90, p. 6151, 2001.

[45] Y. Dai, J. Fan, Y. Chen, R. Lin, S. Lee, and H. Lin, "Temperature dependence of photoluminescence spectra in InAs/GaAs quantum dot superlattices with large thicknesses," *Journal of Applied Physics*, vol. 82, p. 4489, 1997.

[46] P. Fry, I. Itskevich, S. Parnell, J. Finley, L. Wilson, K. Schumacher, D. Mowbray, M. Skolnick, M. Al-Khafaji, A. Cullis, *et al.*, "Photocurrent spectroscopy of InAs/GaAs self-assembled quantum dots," *Physical Review B*, vol. 62, no. 24, pp. 16784–16791, 2000.

[47] V. Golovach, A. Khaetskii, and D. Loss, "Phonon-induced decay of the electron spin in quantum dots," *Physical review letters*, vol. 93, no. 1, p. 16601, 2004.

[48] A. Williamson, L. Wang, and A. Zunger, "Theoretical interpretation of the experimental electronic structure of lens-shaped self-assembled InAs/GaAs quantum dots," *Physical Review B*, vol. 62, no. 19, pp. 12963–12977, 2000.

[49] J. Slater and G. Koster, "Simplified LCAO method for the periodic potential problem," *Physical Review*, vol. 94, no. 6, pp. 1498–1524, 1954.

[50] J. Jancu, R. Scholz, F. Beltram, and F. Bassani, "Empirical spdsˆ{*} tight-binding calculation for cubic semiconductors: General method and material parameters," *Physical Review B*, vol. 57, no. 11, pp. 6493–6507, 1998.

[51] W. Harrison, *Elementary electronic structure*. World Scientific Pub Co Inc, 1999.

[52] Y. Varshni, "Temperature dependence of the energy gap in semiconductors," *Physica*, vol. 34, no. 1, pp. 149–154, 1967.

[53] P. N. Keating, "Effect of invariance requirements on the elastic strain energy of crystals with application to the diamond structure," *Phys. Rev.*, vol. 145, pp. 637–645, May 1966.

[54] C. Pryor, J. Kim, L. Wang, A. Williamson, and A. Zunger, "Comparison of two methods for describing the strain profiles in quantum dots," *Journal of Applied Physics*, vol. 83, p. 2548, 1998.

[55] O. Lazarenkova, P. von Allmen, F. Oyafuso, S. Lee, and G. Klimeck, "Effect of anharmonicity of the strain energy on band offsets in semiconductor nanostructures," *Applied Physics Letters*, vol. 85, p. 4193, 2004.

[56] O. Lazarenkova, P. von Allmen, F. Oyafuso, S. Lee, and G. Klimeck, "Effect of anharmonicity of interatomic potential on strain distribution in semiconductor nanostructures," in *2004 Materials Research Societies 9th Spring Meeting, San Francisco, CA, April 12-16, 2004.*, Pasadena, CA: Jet Propulsion Laboratory, National Aeronautics and Space Administration, 2004, 2004.

[57] R. Martin, "Elastic properties of ZnS structure semiconductors," *Physical Review B*, vol. 1, no. 10, pp. 4005–4011, 1970.

[58] M. Grundmann, O. Stier, and D. Bimberg, "InAs/GaAs pyramidal quantum dots: Strain distribution, optical phonons, and electronic structure," *Physical Review B*, vol. 52, no. 16, pp. 11969–11981, 1995.

[59] E. Clarke, P. Spencer, E. Harbord, P. Howe, and R. Murray, "Growth, optical properties and device characterisation of InAs/GaAs quantum dot bilayers," in *Journal of Physics: Conference Series*, vol. 107, p. 012003, Institute of Physics Publishing, 2008.

[60] G. Klimeck, M. McLennan, S. Brophy, G. Adams III, and M. Lundstrom, "nanohub. org: Advancing education and research in nanotechnology," *Computing in Science & Engineering*, vol. 10, no. 5, pp. 17–23, 2008.

[61] M. McLennan, "The Rappture Toolkit," 2004.