

SEMICONDUCTOR MATERIAL AND DEVICE SIMULATIONS
INVOLVING HIGHLY MISMATCHED ALLOYS

by

Istvan Gulyas, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Physics
August 2019

Committee Members:

Mark Wistey, Chair

Luisa Scolfaro

Alex Zakhidov

COPYRIGHT

by

Istvan Gulyas

2019

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

This work is protected by the Copyright Laws of the United States (Public Law 94-553, section 107). Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgement. Use of this material for financial gain without the author's express written permission is not allowed.

Duplication Permission

As the copyright holder of this work I, Istvan Gulyas, authorize duplication of this work, in whole or in part, for educational or scholarly purposes only.

ACKNOWLEDGEMENTS

I would like to thank Dr. Mark A. Wistey, my graduate advisor, for his support, advice, and mentoring throughout my years as a graduate student. I would like to thank the other members of my thesis committee, Dr. Luisa Scolfaro and Dr. Alex Zakhidov, for their patience and valuable advice. I would also like to thank the members of the EPÉE Lab for the insightful discussions and detailed feedback.

I would like to thank my mother for encouraging me to follow my dreams of studying Physics. Without her support I would not have been able to finish this thesis.

This work used the Extreme Science and Engineering Discovery Environment (XSEDE) through allocation DMR140133, supported by National Science Foundation grant number ACI-1548562. It was supported in part by the National Science Foundation under Grants DMR-1508646 and CBET-1438608, and the Texas State University LEAP center.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT.....	x
CHAPTER	
I. INTRODUCTION	1
II. BACKGROUND.....	3
A. Highly Mismatched Alloys	3
B. BGaAs	5
C. Density Functional Theory.....	8
D. VASP	13
E. Tunnel Field-Effect Transistors.....	14
III. FIRST PRINCIPLES CALCULATIONS OF BGaAs ALLOYS	16
A. Method	16
1. Ab Initio Models of BGaAs Alloys	16
a. Supercells	17
b. Full Composition Range Calculations Using Small Supercells.....	19
c. Dilute Alloy Calculations Using Large Supercells	20
d. Effect of B-B Nearest Neighbor Distance	21
e. Random Alloys.....	22
2. VASP Workflow	23
a. Electronic Minimization and Ionic Relaxation	25
b. Band Structure Calculation and Unfolding.....	27
c. Computational Details.....	30
3. Convergence Testing	31
4. Analysis of VASP Results	37

B. Results and Discussion.....	38
1. Band Gap Over the Full Composition Range	38
2. Dilute Regime	41
3. Electron Effective Mass	43
4. Lattice Constant	44
5. Band Unfolding Artifacts.....	44
6. Small Supercell Calculations	48
C. Future Work	51
D. Summary	52
IV. OPTIMUM TFET PERFORMANCE WITH HIGHLY MISMATCHED ALLOYS.....	54
A. Method	54
1. TFET Model.....	54
2. Computational Details	61
B. Results and Discussion.....	63
1. Effective Mass	63
2. Non-parabolic Bands	67
3. Channel Length and Source Doping Density.....	68
4. Effective Band Gap.....	71
5. I_{60} and V_4	72
C. Future Work	74
D. Summary	75
V. SUMMARY	76
APPENDIX SECTION.....	78
REFERENCES	83

LIST OF TABLES

Table	Page
1. Properties of Supercells	18
2. DG TFET Simulation Parameters	63
3. DG TFET Optimum I_{60} Parameters	63

LIST OF FIGURES

Figure	Page
1. BGaAs Supercell with 16 Atoms.....	19
2. BGaAs Supercells with 54, 64, and 128 Atoms.....	21
3. 54-Atom BGaAs Supercells with 2 B Atoms in Different Arrangements.....	22
4. VASP Simulation Workflow	24
5. GaAs Band Structure Before and After Unfolding.....	29
6. Effect of K-Point Mesh Density on Computation Time	32
7. K-Point Convergence of Lattice Constant	33
8. K-Point Convergence of Band Gap	34
9. Estimation of Convergence Error for Large Supercells.....	35
10. Effect of ENCUT Parameter Value on Computation Time	36
11. ENCUT Convergence of Lattice Constant and Band Gap.....	37
12. Band Structures of Pure GaAs and BAs	39
13. BGaAs Band Gaps Over the Full Composition Range.....	40
14. Band Gaps of Dilute BGaAs Alloys at High-Symmetry Points	42
15. BGaAs Electron Effective Mass in the Direct Band Gap Region	43
16. BGaAs Lattice Constant Versus B Concentration	44
17. Conduction Band Energy Eigenvalues of Random Alloys.....	46
18. Unfolding Artifacts in Effective Band Structures of $B_8Ga_{24}As_{32}$ Supercells	47

19. Classifying Ambiguous Conduction Band Minima.....	48
20. Band Gaps of 16-Atom BGaAs Supercells at High-Symmetry Points.....	49
21. Lattice Constant of 16-Atom BGaAs Supercells	50
22. Schematic and Band Diagram of a DG TFET Device.....	55
23. I_{60} Versus Electron and Hole Effective Mass	64
24. I_{60} at Higher Drain Voltage and Large Electron and Hole Effective Mass	65
25. Effect of Electron Effective Mass on DG TFET Band Structure	66
26. Effect of Hole Effective Mass on DG TFET Band Structure	67
27. I_{60} Versus Non-Parabolicity Factors.	67
28. Effect of Band Non-Parabolicity on Tunneling.....	68
29. I_{60} Versus Source Doping Density and Channel Length	69
30. I_{60} Versus Source Doping Density and Electron and Hole Effective Mass.....	69
31. Effect of Source Doping Density on DG TFET Band Structure	70
32. Short-Channel Effects	71
33. I_{60} Versus Effective Band Gap.....	72
34. I_{60} and V_4 Comparison.....	74

ABSTRACT

In this thesis I investigate a specific highly mismatched alloy, $B_xGa_{1-x}As$, over the full composition range using first principles DFT simulations with HSE06 hybrid functionals in VASP. I find that at low boron percentages the direct band gap decreases slightly, then increases towards the large minimum direct gap of BAs as more boron is added. My results show that the effect of isolated boron atoms on the band gap is small (<5%) at concentrations below 13%. I estimate that BGaAs transitions from direct to indirect band gap at around 17% boron content. I calculate the electron effective masses in the direct band gap region and investigate the effect of B-B pairs in nearest-neighbor group III sites on band gap, conduction band dispersion, and total free energy. I find that the lattice constant of BGaAs follows Vegard's law and estimate that the boron concentration required to lattice match BGaAs to silicon is outside the direct gap regime. I then introduce TFETs as one possible application for highly mismatched alloys. Using the UCSD TFET model which I extended to include Kane's non-parabolic dispersion relation I find the optimal combination of material and device properties that maximize I_{60} . For low drain voltages ($V_d = 0.1$ V), the maximum $I_{60} = 39$ $\mu A/\mu m$ occurs at moderate effective masses for both electrons and holes, while at larger drain voltages ($V_d = 0.2, 0.5$ V) the I_{60} continues to increase up to at least $1.3 m_0$, which makes highly mismatched alloys good candidates for TFET applications due to their increased effective masses.

I. INTRODUCTION

In this thesis I investigate semiconductor materials and devices using computational modeling and simulation. The focus of my research is highly mismatched alloys (HMAs), which is a class of semiconductor alloys that is composed of elements with radically different atomic radii and electronegativity. HMAs promise to open up new avenues in device design due to the level of control they give over a range of material properties, such as lattice constant, band gap, and effective mass, which would not be possible with conventional semiconductor alloys.

The first part of this work is centered on investigating the material properties of boron gallium arsenide (BGaAs) alloy with first principles calculations using the Vienna Ab initio Simulation Package (VASP).^{1,2} Due to the difference in size and electronegativity between boron (B) and gallium (Ga), BGaAs is expected to show characteristics similar to other highly mismatched alloys like gallium nitride arsenide (GaNAs) and germanium-tin (GeSn), such as the band gap and lattice constant exhibiting a highly non-linear dependence on alloy composition. Though GaAs is a well-understood material that has seen widespread use in various device applications, the experimental and theoretical data available on boron arsenide (BAs) and BGaAs is still scarce and the results are often conflicting. I investigate the effect of B concentration on the lattice constant and the electronic structure over the full alloy composition range, from the highly dilute regime all the way to pure BAs and examine the effect of B-B distance on the band gap and effective mass. The method I chose for the calculations, HSE06 hybrid functionals,³⁻⁵ is one of the leading ab initio techniques that has been shown to

approximate the band gap better than the methods used in other existing first principles studies of BGaAs.⁶

In the second part, to examine the possible benefits of the broad control of electronic properties that semiconductor alloying promises, I simulate current-voltage characteristics of a double-gate tunnel field-effect transistor (DG TFET) that is built using hypothetical materials that allow an arbitrary combination of electron and hole effective masses and band offsets. I use a model and the corresponding MATLAB code that was developed by Taur et al.^{7,8} and modify the simulation to account for non-parabolic electronic bands using Kane's dispersion relation.^{9,10} To compare device performance, I adopt the I_{60} as a figure of merit,¹¹ which is defined as the drain current at which the subthreshold swing (SS) crosses 60 mV/decade, the thermodynamic limit of conventional MOSFETs. I find the combination of device and material parameters that yields the highest I_{60} and examine the overall trends that lead to increased device performance. I hope that the results and discussion presented in this work will help guide further research in semiconductor alloying and TFET device design.

II. BACKGROUND

In this chapter I give a brief overview of the current state of the art regarding the chosen topics of my research. I also describe the motivations that led me to investigate highly mismatched alloys and BGaAs in particular. I give justification to my choices of methods and software for the ab initio calculations and finally I present a brief introduction into the topic of TFETs.

A. Highly Mismatched Alloys

Most metal and ceramic alloys are designed to improve the mechanical properties of the material, while in case of semiconductors alloying is commonly done to tune electronic or optical properties, such as band gap or refractive index, or to change the lattice parameter of the crystal for strain engineering.

The electronic properties of conventional semiconductor alloys, which are made up of isoelectronic elements with relatively similar atomic radii and electronegativity, can often be predicted to a high accuracy using the virtual crystal approximation (VCA),¹² which is a model based on linear interpolation between the properties of the pure compounds that make up the alloy. For alloys with a slightly larger difference in atomic radii or electronegativity, the experimentally measured band gaps often do not match the values predicted by VCA and the introduction of a quadratic bowing parameter is necessary.^{13,14} In the case of highly mismatched alloys (HMA), the difference in the properties of the elements are so large that VCA can no longer adequately describe the alloy even with the use of the bowing parameter. The dramatic changes in electronic structure in many HMAs has been successfully explained by the band anticrossing (BAC)

model, which attributes the changes in band structure to the interaction between the extended states of the conduction band and the highly localized states of strongly electronegative impurity atoms.¹⁵ BAC has been shown to be in good agreement with experimental data for a wide range of semiconductor alloys,^{16,17} but appears to be less useful in describing alloys that consist of elements with small electronegativity, such as boron.¹⁸

For most conventional semiconductor alloys the lattice parameter has a linear dependence on the alloy composition, a relationship called Vegard's Law.¹⁹ The law only applies to random semiconductor alloys with no phase separation, made of materials with the same crystal structure. Deviations from this linear relationship are common for HMAAs,^{20,21} but also have been reported for some conventional alloys.²² Such deviations can be described by the introduction of a quadratic bowing parameter, similar to the one used in VCA to describe the band gap. Precise description of the relationship between lattice constant and material composition is important since the composition of an alloy is commonly estimated using the lattice parameter measured by X-ray diffraction (XRD). Controlling lattice constant is important for dislocation-free growth of epitaxial layers of different materials. When a material is grown on a substrate with a slightly different lattice parameter, the crystal structure will deform as the material stretches or compresses to match the lattice constant of the substrate. This deformation induces strain in the material. For a small lattice mismatch, growth can proceed up to a certain thickness, at which point the accumulated strain causes the material to relax, resulting in dislocations. These dislocations can severely degrade device performance by reducing mobility through scattering or introducing trap states that facilitate non-radiative recombination.

The strain caused by the lattice mismatch imposes an upper limit on the epitaxial layer thickness. Controlling the lattice parameter by alloying can push or eliminate this constraint and expand the possibilities of device design.

The strain arising from the lattice mismatch between layers of semiconducting materials can also be exploited to tune certain material properties. Strain introduces a shift in the energy levels of the band structure which is used in band gap engineering. Optical properties of the material, such as the refractive index may also change under strain, which can be useful for waveguides and lasers. By controlling the lattice parameter of the substrate through alloying, the strain induced on the adjacent epitaxial layer can be adjusted to yield the desired change in band gap or refractive index for a particular device application.

B. BGaAs

Group III-V semiconductor alloys have been the subject of extensive research for decades. Among them, GaNAs received considerable attention in recent years. Due to the small size and high electronegativity of nitrogen, it is classified as a highly mismatched alloy. Nitrogen alloying of GaAs reduces the band gap drastically even at very low N percentages^{23,24} and a large increase in electron effective mass have also been observed in N-containing III-V compounds,²⁵ behaviors that are consistent with other HMAs. GaNAs has been suggested as a promising material for multi-junction solar cell applications due to its variable direct band gap and high absorption coefficient.²⁶ However, higher N concentrations of have been difficult to achieve due to phase separation²⁷ that occurs between wurtzite GaN and zincblende GaAs, though significant improvements have been made in recent years using low temperature molecular beam epitaxy.²⁶ Boron, on the

other hand, has a higher predicted solubility limit in GaAs than that of nitrogen,²⁸ which suggests that a wider alloy composition range should be achievable. Compared to N, the electronegativity of B is much smaller, which implies that BGaAs alloys would have a strongly covalent character, while nitrides are highly ionic in nature. Thus, alloying GaAs with boron has been suggested as an alternative to nitrogen for band gap engineering and controlling lattice constant.

GaAs is a well understood material that has been used extensively in recent decades in microwave circuits used for mobile communications, infrared LEDs and lasers, solar cells, and electro-optic modulators. It forms a zincblende crystal structure with an equilibrium lattice constant of 5.65 Å²⁹ and has a direct band gap of 1.42 eV³⁰ which makes it ideal for optical applications due to the effective absorption and emission of photons. Its low electron effective mass of 0.063 m_0 ³⁰ contributes to a high electron mobility, which is necessary for the high-speed integrated circuits operating at microwave frequencies. The current record efficiency of 29.1% for single junction solar cell is held by a GaAs device created by Alta Devices.³¹

BA₃As₂, on the other hand, is one of the least investigated compounds among the III-V family of semiconductors, partly due to the difficulties involved in synthesizing high quality single crystals.^{32,33} BA₃As₂ started to receive more attention after prediction of exceptionally high thermal conductivity,³⁴ which has also been verified by experiment.^{33,35-37} Just like GaAs, BA₃As₂ forms a zincblende crystal structure with an experimentally measured lattice constant of 4.777 Å,^{38,39} but the existence of a B₁₂As₂ subarsenide phase has also been confirmed when grown under B-rich conditions or at high temperatures.^{38,40,41} There are conflicting reports on the band structure of BA₃As₂.

Among the few experimental studies, earlier reports measure an energy gap of 1.46 eV,^{39,42,43} though they disagree on whether BAs is a direct⁴² or indirect⁴³ gap material. A more recent study found the experimental gap to be 1.77 eV.⁴⁴ A wide range of energy gaps have been reported from theoretical calculations, though most seem to indicate the band gap is indirect, which agrees with more recent experimental results.⁴³

The difference in atomic radii between B and Ga implies that BGaAs would behave as a highly mismatched alloy, meaning its electronic properties would not be well described by a linear interpolation between GaAs and BAs, even with the introduction of a quadratic bowing parameter. However, theoretical and experimental studies paint a conflicting picture, much like in the case of BAs. Some reports are claiming an increase^{18,28,45–47} of band gap with B percentage, while several others suggest the opposite.^{39,48–50} There are reports of decreasing band gap in the dilute regime and then an increase as more boron is added, but there is a disagreement in the B concentrations that give the smallest band gap.^{51,52} Multiple experimental results show that B causes only a slight change in band gap^{18,46,51} and thus behaves as a conventional alloy, unlike GaNAs. The classification of BGaAs as a HMA has been questioned by theoretical calculations as well.^{28,48} It has been suggested, as a result of a theoretical study using a tight binding model, that isolated B atoms have a very small effect on the electronic structure, while B-B pairs and clusters cause a considerable decrease in band gap and a strong increase in electron effective mass.⁴⁸ Overall, most existing investigations are restricted to single-digit B percentages, with only one first principles study examining the full composition range,⁵² which makes it difficult to draw general conclusions on the effect of B on the band structure.

C. Density Functional Theory

The theory of semiconductors is built on the foundation of quantum mechanics. At the center of it is the Schrödinger equation, which describes a quantum mechanical system in terms of energy eigenvalues and the corresponding eigenfunctions called the wavefunctions, and the Hamiltonian operator, which is the sum of the kinetic and potential energy operators. The time-independent Schrödinger equation for a single particle in a potential V is

$$\left[\frac{-\hbar^2 \nabla^2}{2m} + V(\mathbf{r}) \right] \Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \quad (1)$$

where E is the energy, $\Psi(\mathbf{r})$ is the wavefunction, and the term in brackets is the Hamiltonian. In semiconductor physics we're interested in systems of many electrons, as in a crystal, and thus need to consider the many-body Schrödinger equation,

$$\left[\sum_1^N \left(\frac{-\hbar^2 \nabla_i^2}{2m} + V(\mathbf{r}_i) \right) + \sum_{i<j} U(\mathbf{r}_i, \mathbf{r}_j) \right] \Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = E\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad (2)$$

where other than replacing the Hamiltonian with a sum over the N electrons in the system we also have to introduce a new potential term $U(\mathbf{r}_i, \mathbf{r}_j)$ to describe the interaction between the electrons. In practice, the solution of the many-body Schrödinger equation in this exact form becomes computationally prohibitive for even a relatively small system, so there is a need to simplify the problem by introducing some approximations.

Density functional theory (DFT) is the most widely used approach to finding approximate solutions to the many-body Schrödinger equation. At its foundation are the theorems of Hohenberg and Kohn,⁵³ first published in 1964, which state that *the ground-state energy E is a unique functional of the electron density*, and that *the electron density*

that yields the minimum ground-state energy is the true ground-state electron density. In this context a functional is a mapping between a function and a number, a common example of which is the definite integral

$$F[f] = \int_{x_0}^{x_1} f(x) dx \quad (3)$$

where we can say that F is a functional of the function f , just like we say that f is a function of variable x .

We know that the electron density can be obtained directly from the wavefunctions as

$$n(\mathbf{r}) = 2 \sum_i \psi_i^*(\mathbf{r}) \psi_i(\mathbf{r}) \quad (4)$$

where the factor 2 is a result of the fact that two electrons of opposite spins can occupy an orbital. Thus, the ground state energy, in light of the Hohenberg-Kohn theorems, can be written as

$$E = E[\Psi] = E_T[\Psi] + E_U[\Psi] + E_V[\Psi] \quad (5)$$

where E_T , E_U , and E_V are functionals corresponding to the kinetic energy, the potential energy due to the interactions between electrons, and the potential energy due to the external potential, i.e. due to interaction between the electrons and the stationary ion cores.

Among these functionals, the external potential term is straightforward to calculate as

$$E_V[n] = \int d^3r n(\mathbf{r}) V(\mathbf{r}) \quad (6)$$

where we wrote E_V in terms of the charge density, but we could have written it in terms of the wavefunction using Eq. 4. The kinetic energy term can be written as

$$E_T = \frac{\hbar^2}{2m} \sum_i \int d^3 r \Psi_i^* \nabla^2 \Psi \quad (7)$$

but the situation is not this simple for the remaining electron-electron interaction term.

One can approximate E_U with the Coulomb-potential as

$$E_U \approx \frac{e^2}{2} \int d^3 r' \int d^3 r \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (8)$$

but this neglects any contribution arising from quantum mechanical effects, such as the electron-electron repulsion due to the Pauli exclusion principle. To mitigate this, all effects that cannot be described with the Coulomb interaction are lumped together into a term called the exchange-correlation functional E_{XC} . The specific treatment of this exchange-correlation functional has been one of the central areas of research within the topic of DFT, and various computational methods differ mainly on the approximation employed to describe this E_{XC} term.

Just expressing the problem in terms of density functionals is not enough, we have to minimize the energy to get the charge density corresponding to the system, which is still a prohibitively expensive computational task for a multi-electron interacting system. The Kohn-Sham (KS) equations,⁵⁴ published a year after the Hohenberg-Kohn theorem, reduce the problem to an iterative process of solving a set of single-electron Schrödinger equations that correspond to a fictitious system with orbitals which ultimately reproduce the same electron density as the original many-body system. The Kohn-Sham equations can be written as

$$\left[\frac{-\hbar^2 \nabla^2}{2m} + V(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r}) \right] \Psi_i(\mathbf{r}) = E_i \Psi_i(\mathbf{r}) \quad (9)$$

where the potential term V is the same external potential as in equations 1 and 2. V_H is the Hartree potential that describes the Coulomb interaction between the electron and the total charge density

$$V_H(\mathbf{r}) = e^2 \int d^3 r' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (10)$$

Since the electron in question also contributes to the total charge density, the Hartree potential includes a non-physical self-interaction term. This is corrected by the exchange-correlation potential V_{XC} , which is defined as the functional derivative of E_{XC} , and it also includes all other quantum mechanical effects discussed earlier.

The iterative process starts with a trial electron density, from which the Kohn-Sham equations are solved, yielding the wave functions of the fictitious non-interacting system, the Kohn-Sham orbitals. From these KS orbitals the electron density is calculated and compared to the original electron density. If the calculated electron density is different from the original, then the original electron density is somehow modified and the calculation is repeated until convergence, i.e. until the difference in electron density is below a certain threshold. This process is called the self-consistency cycle. Details such as the methods used for solving the Kohn-Sham equations and updating the charge density differ for the various implementations of DFT.

Over the decades since its initial formulation, DFT has become a widely used tool in quantum chemistry and solid-state physics. So much so, that Walter Kohn was awarded the Nobel prize in chemistry for his contributions to the development of DFT.⁵⁵ Since the first publications of Hohenberg, Kohn, and Sham, various approaches have been developed to improve the accuracy and computational efficiency of the calculations.

The most successful methods are often different for quantum chemistry and solid-state physics applications.

A diverse selection of functionals approximating the E_{XC} term are available. The least computationally expensive is the local density approximation (LDA), which uses the exchange-correlation energy of a homogeneous electron gas of the same density obtained using Quantum Monte-Carlo simulations. A more advanced functional is the generalized gradient approximation (GGA), which also considers the local gradient of the electron density. Perhaps the most commonly used GGA method in solid-state physics is the Perdew-Burke-Ernzerhof (PBE) functional.⁵⁶ The LDA and GGA functionals have been used extensively in both physics and chemistry, but they are not without issues. For example in solids both are known to severely underestimate the band gap.⁵⁷

Hybrid functionals use a mixture of the Fock exchange,⁵⁸ which is the basis of the Hartree-Fock⁵⁹ (HF) method used extensively in quantum chemistry, together with exchange and correlation terms from other functionals. The B3LYP functional, a mixture of the Fock⁵⁸ and Becke⁶⁰ (B) exchange functionals with the Lee, Yang, and Parr⁶¹ (LYP) and LDA correlation functionals, is perhaps the most common hybrid functional used in chemistry. The PBE0 functional⁶² uses a mixture of PBE and the Hartree-Fock functionals for exact exchange and PBE functionals for correlation. It is often used in solid-state physics and can predict band gaps with good accuracy, but at the expense of high computational cost.⁶ Another approach is to use the HF exchange more strongly at short range, and rely on the less computationally expensive PBE exchange potential for calculating long-range effects. This is implemented in the HSE06 hybrid functional,³ after Heyd, Scuseria, and Ernzerhof^{4,5}, which has been found to be the most accurate

among the three for semiconductor applications, while also being the most computationally efficient.⁶ Another notable set of functionals are called Meta-GGA,^{63,64} which rely on several parameters fitted to experimental data. Some Meta-GGA functionals also include the Hartree-Fock exact exchange and thus can be considered hybrid functionals. Comparisons between Meta-GGA and HSE06 found that for semiconductors the HSE06 functional can better approximate experimental band gaps.⁶⁵

When it comes to the basis function representation of the Kohn-Sham orbitals, linearized augmented plane waves (LAPW) and linear muffin-tin orbitals (LMTO) have been widely used in solid-state physics,⁶⁶ while Gaussian type orbitals⁶⁷ (GTO) are some of the most common methods used in quantum chemistry.

The computational cost of DFT calculations can be significantly reduced by the use of pseudopotentials. Electron orbitals close to the atom cores oscillate on short length scales, and to represent these orbitals in a plane-wave basis set we need very high energy plane waves, which is computationally expensive.⁶⁸ The solution is to use a pseudopotential instead of the rapidly oscillating electron density near the cores, which is called the frozen core method. Pseudopotentials are designed to give a smooth charge density near the core but approximate the physical properties of the real core electrons at a distance. Examples of commonly-used pseudopotentials include ultrasoft pseudopotentials⁶⁹ (USPP) and the projector augmented-wave method⁷⁰ (PAW).

D. VASP

The Vienna Ab initio Simulation Package^{1,2} (VASP) is a computer software capable of a wide range of materials modeling calculations. It implements various DFT methods discussed in section 2.C, including LDA, GGA, Meta-GGA, and Hartree-Fock

hybrid functionals. It can calculate electronic band structures, phonon dispersion, frequency dependent dielectric tensors, and many more. It represents orbitals, charge density, and local potential using plane-wave basis sets, and ships with USPP and PAW pseudopotentials.

Based on the total number of publications, it is by far the most widely used DFT simulation package. Currently it also offers the best computational performance compared to its competitors. Large-scale collaborations such as the Materials Genome Initiative⁷¹ have multiple projects that build heavily on VASP, such as the Materials Project⁷² or the AFLOW consortium⁷³.

A great number of software packages are available that can interact with VASP input and output files. Programs such as VESTA⁷⁴ can be used to view and edit crystal structures or visualize 3D charge densities. Libraries such as pymatgen⁷⁵, which is one of the key software components used in the Materials Project, allow programmatic processing of VASP data from structural analysis to plotting band diagrams.

E. Tunnel Field-Effect Transistors

Arguably the most important building block of modern integrated circuits is the metal-oxide-semiconductor field-effect transistor (MOSFET). With the widespread use of battery-powered electronics in recent decades, there is a growing need for chips that combine high performance with low power consumption. A major roadblock for developing low-power MOSFET devices is the thermal limit of the subthreshold swing (SS). The SS measures the change in gate voltage needed for a 10-fold increase in drain current in the sub-threshold region. The SS of a MOSFET can be calculated as

$$SS = \left(1 + \frac{C_D}{C_{ox}}\right) \frac{kT}{q} \ln 10 \quad (11)$$

where C_D is the surface depletion layer capacitance and C_{ox} is the oxide capacitance.

Thus, the minimum SS in MOSFETs is given by $(kT/q)\ln 10$ which has a value of approximately 60 mV/decade at room temperature.

Several novel device structures have been proposed to overcome the 60 mV/decade limitation of conventional MOSFETs. Among them, tunnel field-effect transistors (TFETs) are one of the most promising candidates. In a TFET device band-to-band tunneling is used as the fundamental switching mechanism, which results in extremely low OFF currents (I_{OFF}) and small SS. The device was first suggested by Chang and Esaki⁷⁶ in 1977. Since then several devices with SS below 60 mV/decade have been made in laboratories around the world,⁷⁷ but TFET technology is still in the experimental phase and far from commercial use.

The biggest challenge in TFET design has been to achieve high ON current (I_{ON}) while maintaining low I_{OFF} and SS below 60 mV/decade over a drain current range spanning at least four orders of magnitude.⁷⁸ Various TFET device structures have been proposed, including single-gate (SG), double-gate (DG) and gate-all-around (GAA) designs.⁷⁹ A wide range of materials are considered for TFET applications, from pure silicon⁸⁰ and SiGe⁸¹, through various group III-V alloys,^{82,83} to carbon nanotubes⁸⁴ and graphene nanoribbons.⁸⁵

III. FIRST PRINCIPLES CALCULATIONS OF BGaAs ALLOYS

The first part of my research focuses on investigating the material properties of BGaAs with ab initio simulations using VASP. In this chapter I outline the research method, present the results, and discuss their implications.

A. Method

First, I introduce the supercell method used for ab initio simulations of imperfect crystals, such as random alloys. I discuss the tradeoff between using larger supercells to simulate dilute alloy compositions and to better approximate the randomness of real alloys at the cost of increased computational time, versus using smaller supercells for a faster turnaround. I then give an overview of the VASP calculation workflow and describe the ionic relaxation and the band structure calculation steps in detail. I explain how I used convergence testing to optimize certain VASP parameters and describe the methods I used to analyze the simulation results.

1. Ab Initio Models of BGaAs Alloys

VASP takes advantage of periodic boundary conditions to simulate an infinite crystal with only a smaller volume of the system explicitly defined. The dimensions of this volume, along with the positions, velocities, and species of the contained atoms is given in the POSCAR input file. In the case of pure materials, this explicitly defined volume can be the primitive cell, consisting of the smallest number of atoms needed to represent the crystal. When we want to simulate alloys, such as BGaAs, we must define a larger volume with more atoms, since the random arrangement of atoms in real alloys lack the translational symmetry of perfect crystals. Due to the use of periodic boundary

conditions, the simulated crystal is still an infinite repetition of the defined volume, and thus has a translational symmetry that is not present in real alloys, but we can mitigate the effects of this periodicity by defining a larger volume, and thus increasing the period of repetition. However, because of the rapidly increasing computational cost corresponding to the number of electrons and ion cores contained in the defined volume, we must balance the tradeoff between speed, available composition range, and accurate modeling of the randomness of alloys found in Nature.

The goal of my research is to investigate the band structure of BGaAs alloys over the full composition range using ab initio methods. Since existing experimental results on BGaAs are of low B percentages, I designed the simulations to treat this dilute regime with increased detail. Other than looking at the effect of overall B concentration on the band structure and lattice constant, I examined the role of B-B distance in the narrowing of the band gap, which was suggested by earlier theoretical work on BGaAs.⁴⁸ During my research I ran VASP simulations of systems of various sizes. The relatively quick turnaround time of calculations involving a smaller number of atoms allowed me to validate the initial results and iteratively improve my methods. Calculations on larger systems made it possible to investigate a wider range of compositions and particular arrangements of atoms, and to better approximate the randomness of real alloys.

a. Supercells. In a crystal lattice, the cell with the smallest volume corresponding to a single lattice point is defined as the primitive cell. Cells corresponding to the same crystal lattice that contain more than one lattice point, but still have discrete translational symmetry, are called supercells. The simplest examples of supercells are simply repetitions of a unit cell along one or more primitive lattice vector. The unit cell being

repeated can be the primitive cell containing a single lattice point, or a conventional unit cell, containing 4 lattice points in case of the face-centered cubic (FCC) lattice. Note, that the FCC conventional unit cell is technically a supercell by itself.

If we want to generate isometric supercells for a crystal lattice, we can take a unit cell containing n_{unit} lattice points and repeat it along each crystal direction r times.

Using this method, we can create supercells where the total number of atoms is given by

$$n_{supercell} = n_{unit} \times r^3 \times n_{basis} \quad (12)$$

where n_{basis} is the number of atoms corresponding to each lattice point. The first few supercells that result from this method are summarized in Table 1. In this work I used isometric zincblende supercells that consist of 16, 54, 64, and 128 atoms.

Table 1. Properties of Supercells. The table shows the properties of the first few possible supercells created from isometric repetition of FCC and zincblende primitive and conventional unit cells.					
Unit cell	Latt. points per unit cell (n_{unit})	Repetitions (r)	Unit cells (r^3)	Atoms (FCC) ($n_{supercell}$)	Atoms (zincblende) ($n_{supercell}$)
primitive	1	1	1	1	2
primitive	1	2	8	8	16
primitive	1	3	27	27	54
primitive	1	4	64	64	128
conventional	4	1	1	4	8
conventional	4	2	8	32	64

It is possible to create supercells where the number of unit cell repetitions are different along the crystal directions. For such supercells the repetition distance, due to periodic boundary conditions, of any imperfection introduced into the lattice will also be

different depending on the crystal direction. There are circumstances when such asymmetry is acceptable or even desired, for example when simulating surfaces. In this work I investigate properties of the bulk material and thus I chose to only use supercells with equal repetitions along each crystal direction.

b. Full composition range calculations using small supercells. I use 16-atom supercells to simulate $B_xGa_{1-x}As$ alloys through the full composition range from pure GaAs ($x=0$) to BAs ($x=1$). The primary benefit of using a small supercell is the increased speed of the calculation, which allows for a quick turnaround time between successive iterations. 16-atom supercell calculations can finish in a matter of hours, as opposed to days or weeks, which is why I used a 16-atom BGaAs supercell with a single B atom, as shown in Figure 1, to develop the initial version of the calculation workflow and for all convergence testing.

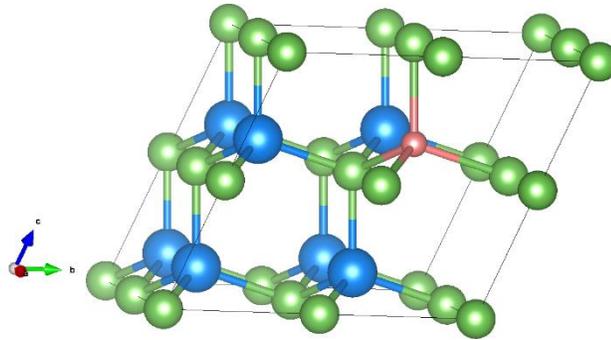


Figure 1. BGaAs Supercell with 16 Atoms. The largest atoms (in blue) are gallium, the medium-sized atoms (in green) are arsenic, and the smallest atom (in red) is boron. The supercell was created by repeating the zincblende primitive cell twice in each crystal direction, then replacing one Ga atom with B. Supercells of the same size were used to model the full composition range.

There are also drawbacks to using small supercells. The most obvious issue is the limited number of possible alloy compositions. In case of BGaAs we can only have 9 discrete data points for boron percentage, including pure GaAs (0%) and pure BAs

(100%). Another problem arises from the translational symmetry introduced by VASP's use of periodic boundary conditions. The solutions will reflect this additional periodicity and thus the resulting band structures may not be a good approximation of experimental data. The problem is not limited to just small supercells, but as the supercell size, and thus the repetition period increases, the effect of this additional symmetry on the ground-state electron wavefunctions diminishes.

c. Dilute alloy calculations using large supercells. The smallest non-zero B concentration that can be simulated with a 16-atom supercell is 12.5%, which corresponds to a single B atom. Since most of the existing data on BGaAs is of single digit B percentage or less,^{18,39,45-47,49-51} it was important to include the dilute regime in the calculations so that direct comparison can be made to experimental results. To simulate alloys with low B content, I increased the supercell size to 54, 64, and 128 atoms. Each of these supercells contain a single B atom, corresponding to BGaAs alloys with B concentrations of 3.7%, 3.1%, and 1.6%, respectively. The 54- and 128-atom supercells use the zincblende primitive cell repeated 3 and 4 times, respectively, along each primitive lattice vector direction. The 64-atom supercell is made from the conventional unit cell repeated twice along the x, y, and z directions. Ball-and stick models of these dilute alloy supercells are shown in Figure 2.

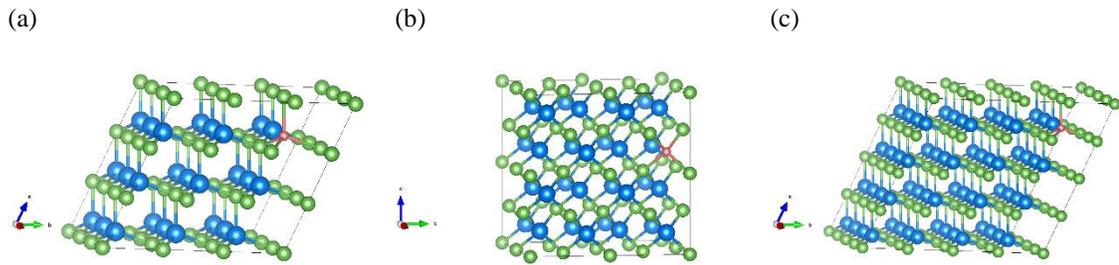


Figure 2. BGaAs Supercells with 54, 64, and 128 Atoms. The largest atoms (in blue) are gallium, the medium-sized atoms (in green) are arsenic, and the smallest atoms (in red) are boron. Supercells (a) and (c) were created by repeating the zincblende primitive cell 3 and 4 times, respectively, in each crystal direction. Supercell (b) was created by repeating the zincblende conventional unit cell twice in each crystal direction. Finally, one Ga atom was replaced with B.

d. Effect of B-B nearest neighbor distance. Existing theoretical research on BGaAs suggests that the formation of B-B pairs and clusters may have a strong effect on the band structure, causing a narrowing of the band gap and a strong increase in electron effective mass.⁴⁸ To investigate this claim I created 54-atom supercells with 2 B atoms ($x=0.074$) in three different configurations, as shown in Figure 3. In structure (a) the two B atoms are at nearest-neighbor group III sites, separated by only a single As atom, giving a smallest B-B distance of 3.82 Å. The structures (b) and (c) have the same nearest-neighbor B-B distance, 6.86 Å, the difference is that for structure (b) each B atom has 3 neighboring B atoms at the same shortest distance, while in structure (c) the atoms are shifted so that each B atom has only 2 neighboring B at the minimum distance of 6.86 Å, while the distance to the third B atom is 9.69 Å.

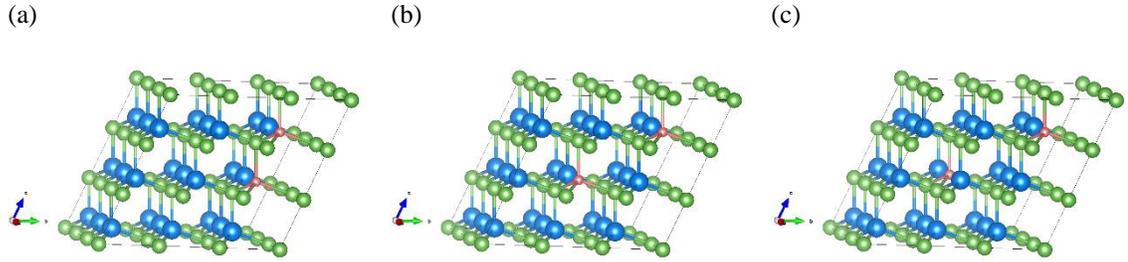


Figure 3. 54-Atom BGaAs Supercells with 2 B Atoms in Different Arrangements. The largest atoms (in blue) are gallium, the medium-sized atoms (in green) are arsenic, and the smallest atoms (in red) are boron. Supercell (a) shows two B atoms in nearest-neighbor (NN) group III sites, while supercells (b) and (c) the B atoms are farther from each other.

e. Random alloys. To approximate the random arrangement of atoms in real alloys, I created 64-atom supercells where the B atoms are substituted into group III sites in a random manner. The B concentrations of these random alloys are chosen to match the 9 possible B percentages of the 16-atom supercells. This allows direct comparison between the results and thus I can investigate the effect of the additional periodicity caused by the infinite repetition of the supercell on the band structure of the material.

To further emulate the randomness of alloys found in Nature, we can simulate multiple random configurations of atoms that all correspond to the same overall composition and then average the results. With this, the computational effort scales linearly with the number of configurations, as opposed to a roughly quadratic scaling that would result from increasing the supercell size. In this work I present results of simulations of two sets of 64-atom random alloy supercells.

2. VASP workflow

The calculation of electronic band structure using VASP is a multi-step process. Figure 4 shows a summary of all calculation steps. Before the electron energy levels can be determined, the simulated material undergoes ionic relaxation to determine the exact positions of the ion cores. This ionic relaxation is itself broken up into separate steps according to which degrees of freedom are allowed to change: ion core position, cell shape, and cell volume. The relaxation steps are first performed using the PBE method,⁵⁶ which gives very limited accuracy but converges quickly, and then continue using HSE06 hybrid functionals^{4,5,65} that yield highly accurate results but are orders of magnitude slower.⁶ Once the relaxation is finished we proceed to the band structure calculation, still using HSE06 for high accuracy, and post-processing steps such as band unfolding with the BandUP code.^{86,87} The band structure calculation steps are repeated multiple times to increase convergence.

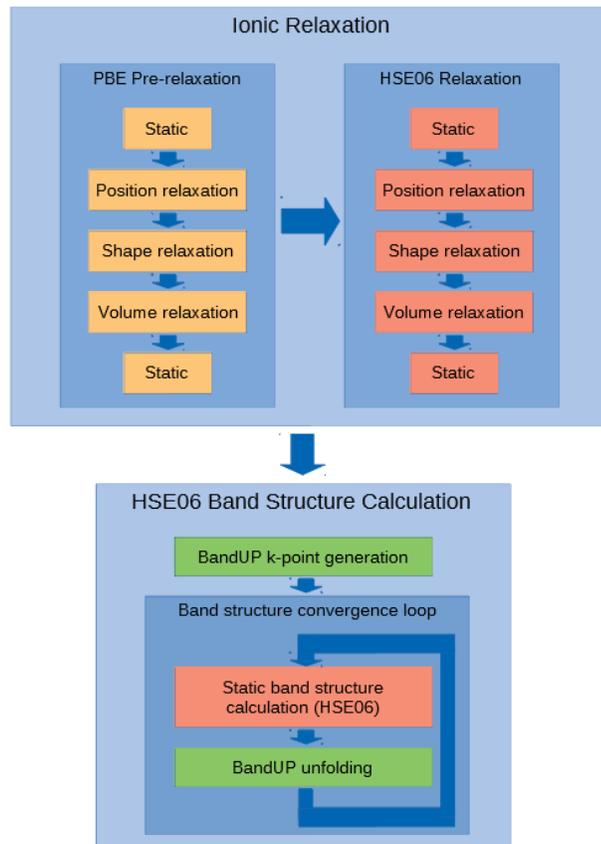


Figure 4. VASP Simulation Workflow. Orange and red rectangles represent VASP jobs, green rectangles are BandUP tasks.

VASP does not provide any functionality out of the box to automate this process. Each step requires at least 4 input files, which are often copied over from previous jobs. Some of these files need to be modified or renamed between successive steps. Doing this by hand is a tedious and error-prone process, especially if we want to carry out multiple simulations of materials with varying compositions.

To automate this process, I created a workflow script that handles the copying, renaming, and modification of the input and output files. Since the bulk of the work

consists of basic filesystem manipulations, I chose to implement the workflow as a bash script, relying on common Linux shell tools such as awk and sed for editing of input files between steps. Bash, or the Bourne Again SHell, is the default shell in most Linux distributions and MacOS, and since VASP offers no graphical user interface, basic knowledge of the Linux command line is essential for anyone using VASP. By writing the workflow as a bash script and using common Linux command line tools, I hope that I created a tool that is relatively easy to learn, use, and maintain.

a. Electronic minimization and ionic relaxation. The core concept behind DFT calculations in VASP is the electronic self-consistency loop, an implementation of the iterative minimization process developed by Kohn and Sham,⁵⁴ as described in section 2.3. VASP offers several minimization algorithms, including the blocked Davidson,^{88,89} RMM-DIIS,^{90,91} and a damped velocity friction algorithm⁹²⁻⁹⁴ highly optimized for Hartree-Fock calculations. In the workflow I use RMM-DIIS for all PBE calculations and the damped velocity friction algorithm for all HSE06 steps, following the recommendations of the VASP manual. At the end of each electronic minimization step the total free energy of the system and the band structure energy is calculated, and the values are compared to the energies from the previous step. If the difference for both is less than a given threshold value, then the orbitals are considered converged and the iteration stops. This energy difference threshold is the EDIFF parameter, with a default value of $10^{-4} eV$, and it essentially gives the number of significant figures for the total free energy. For all calculation steps I set EDIFF to $10^{-8} eV$ for increased convergence. The minimization also stops if the total number of steps reach a certain limit, in which case the result cannot be considered converged and the workflow should not proceed to

the next step. The solution might be as simple as increasing the step limit, or some other parameter needs to be adjusted to improve convergence speed.

For pure materials the ionic positions exactly match the coordinates defined by the crystal lattice. But any imperfection in the crystal will cause the atoms to slightly shift from their lattice sites, and the exact positions of the ion cores need to be determined, in a process called ionic relaxation, before we can proceed to calculating the band structure. To do this, we start with some initial atomic positions, e.g. the coordinates of the perfect lattice sites, then perform electronic minimization. When the electronic self-consistency cycle is finished VASP calculates the forces and the stress tensor, and updates the atomic positions corresponding to a predicted new energy minimum. The system then undergoes electronic minimization with the new ion coordinates, and the cycle repeats until the difference in total free energy between ionic relaxation steps is smaller, than the value given in the EDIFFG parameter, which is 10 times the value of EDIFF by default. Since convergence is not guaranteed, the loop also quits if the number of iteration steps reaches a set limit, like in the case of the electronic self-consistency loop. Troubleshooting ionic relaxation convergence issues can involve switching to a different relaxation algorithm.

The most commonly used algorithms for ionic relaxation of solids are RMM-DIIS^{90,91} and conjugate-gradient.⁹⁵ The RMM-DIIS method often leads to faster convergence, but it may fail if the initial coordinates are far from the energy minimum. The conjugate-gradient algorithm is slower, but more robust. Early versions of my VASP workflow used the RMM-DIIS algorithm for ionic relaxation, but after several calculations involving larger systems failed to converge, I switched over to the conjugate-gradient algorithm.

VASP allows for simultaneous relaxation of cell volume, cell shape, and individual ion positions, but in practice this can lead to calculations that fail to converge. For easier troubleshooting, I decided to split the ionic relaxation part of the workflow into three separate jobs. In the first calculation the atom positions are allowed to change, but the cell dimensions are held fixed. In the second step the cell shape is relaxed, but the overall cell volume and the fractional coordinates of the individual ions are held constant. In the final step the cell volume is allowed to change, ultimately determining the relaxed lattice constant. The PBE position pre-relaxation failed to converge for some random alloys with high B content because the GaAs lattice constant used for the initial dimensions of the supercell was too large. For those failed jobs I modified the pre-relaxation order so that volume relaxation is performed first, giving a better initial estimate for the lattice constant, but I preserved the original order, as seen in Figure 4, for the final HSE06 relaxation steps.

b. Band structure calculation and unfolding. Once the supercell dimensions and ion core positions for the fully relaxed material is determined, the next step is the band structure calculation. To determine the ground-state charge density and accurate forces and stress tensor, and thus calculate the relaxed atomic coordinates, VASP needs to sample the Brillouin zone with an evenly spaced k-point grid. However, when looking at the band structure of the material, we're usually interested in energy eigenvalues for k-points lying along specific symmetry lines. Thus, the charge density is first calculated using an even k-point grid in a self-consistent calculation, where the charge density is modified in each electrical minimization step using the updated wavefunctions. Once the ground-state charge density and the converged orbitals are determined, a continuation job

is started to calculate the energy eigenvalues at k-points along the symmetry lines of interest, where the previously calculated orbitals and charge density are loaded, but not allowed to change throughout the job (non-self-consistent calculation). The process is slightly different with hybrid functionals. In the current version of VASP, non-self-consistent calculations are not possible with HSE06. The suggested workaround is to run a self-consistent job to determine the band structure but set the weight of the k-points corresponding to the symmetry lines to zero, leaving the weights of the evenly spaced mesh untouched. This will cause VASP to only use the evenly spaced k-point mesh, with non-zero weights, for the electronic minimization process, but still calculate the energy eigenvalues corresponding to the k-points along the symmetry lines.

A major side effect of using supercells larger than the primitive cell is that the energy bands in the resulting electronic structure fold over across the symmetry points separating the Brillouin zones of the supercell. This results in a band diagram that include bands that are not real, as shown in Figure 5. Band folding also makes it impossible to locate the k-point corresponding to the conduction band minimum, and thus classify a material as direct or indirect gap. To create a band diagram that resembles that of a primitive cell, the calculation results need to undergo a procedure known as band unfolding.

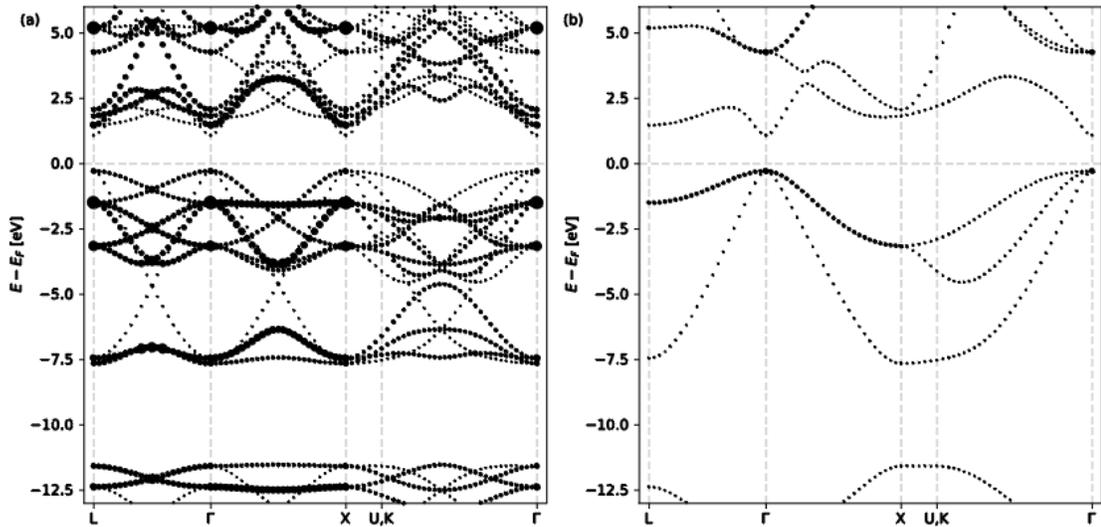


Figure 5. GaAs Band Structure Before and After Unfolding. The band diagram was calculated using a 16-atom supercell. The original, folded bands are visible in (a), while (b) shows the result of BandUP unfolding.

VASP does not provide any band unfolding functionality, but there are a few software packages available that can perform this operation and are compatible with VASP, such as BandUP,^{86,87} vasp_unfold,⁹⁶ fold2Bloch-VASP,^{97,98} or GPAW.^{99,100} I decided to use BandUP to unfold the band structures because it is relatively well-documented and it appears to be continuously maintained. Once the ionic relaxation is complete, the workflow runs one last electronic structure relaxation with an even k-point mesh, then it invokes BandUP to generate the k-point coordinates along the high-symmetry lines, which will become the horizontal axis of the band diagram. These k-point coordinates then are added, with zero weights, to the original evenly spaced k-point mesh, and the band structure calculation is then performed by VASP, whereby the energy eigenvalues along the specified k-points are determined. When the job completes, the

workflow calls BandUP one more time to generate an unfolded band structure from the VASP calculation results. These steps are repeated multiple times to improve the convergence of the wavefunctions, resulting in sharper lines in the band diagram and more accurate determination of band gap and effective mass.

c. Computational details. The simulations were run on the Texas State University LEAP High Performance Computing (HPC) Cluster. The LEAP cluster features computing nodes with 28 CPU cores each. Since the computational complexity scales with the number of atoms in the simulation, the number of CPUs used in parallel was increased for larger supercells. Rigorous investigation of optimal CPU count per job was out of the scope of this work, but as a simple rule of thumb, the number of nodes per job was chosen so that there is at least one CPU core per atom in the supercell. This means that most 16-, 54-, 64-, and 128-atom supercell calculations used 1, 2, 3, and 5 nodes, that is 28, 56, 84, and 140 CPUs, respectively. Occasionally some jobs would fail with errors due to insufficient memory and restarting the jobs, even with increased dedicated memory allocation, did not resolve the issue. After some troubleshooting, I found that the simplest solution was to increase the number of computing nodes for those failing jobs, so some results presented in this work were simulated using a larger number of CPUs. In certain other cases, such as the second set of 64-atom random alloys, the number of computing nodes per job was increased with hopes to expedite the calculation, but the improvement in computation speed was smaller than expected. A rigorous test of optimum job parameters on the LEAP cluster could be useful for future ab-initio studies using VASP at Texas State University.

3. Convergence Testing

Various parameters of the VASP calculation can be adjusted to make the results more accurate at the expense of increased execution time. Finding the optimum parameters is always a tradeoff between accuracy and speed. The process of exploring this tradeoff and finding parameters that give good accuracy while keeping the execution times reasonable is called convergence testing. During this process a sample job is executed multiple times while varying only a single parameter value, and the change in calculation results such as lattice constant, band gap, and effective mass is compared to the change in execution time

One of the parameters I varied during convergence testing is the density of the k-point mesh used by VASP to sample the Brillouin zone. This is defined in the KPOINTS input file and determines the k-point coordinates in reciprocal space at which the wavefunctions are calculated during the electronic minimization process. An evenly spaced k-point grid is needed for all calculation steps leading up to the band structure calculation, at which point more k-points are added along the symmetry lines of the Brillouin zone, which will become the horizontal axis of the E - k diagram. The evenly spaced mesh is defined by the number of divisions in the Brillouin zone along each reciprocal lattice vector. VASP uses the symmetry of the lattice to reduce the number of k-points where the wavefunction needs to be calculated. For a perfect zincblende crystal (e.g. pure GaAs) with a $3 \times 3 \times 3$ mesh only four out of the 27 k-points will be evaluated after symmetry reduction. For less symmetric supercells, such as some 64-atom random alloys, the same k-point grid is still reduced to 14 unique k-points. To test the convergence of the number of k-point divisions, I ran a series of calculations using a 16-

atom BGaAs supercell containing a single boron atom ($B_1Ga_7As_8$) with the k-point grid density varying from $3 \times 3 \times 3$ to $21 \times 21 \times 21$.

Increasing the k-point mesh density causes the computation time to grow exponentially, as illustrated by Figure 6. The run time of the final HSE06 ionic position relaxation, the most time-consuming step in the workflow, increases by about one order of magnitude when the number of k-point divisions is increased by two. This shows the importance of finding a good compromise between the increased accuracy that comes with denser k-point sampling of the Brillouin zone, and the exponentially growing computation time. For jobs above $7 \times 7 \times 7$ k-point grid density only PBE results are available, since the HSE06 part of those calculations did not finish due to various VASP errors, mostly related to insufficient memory.

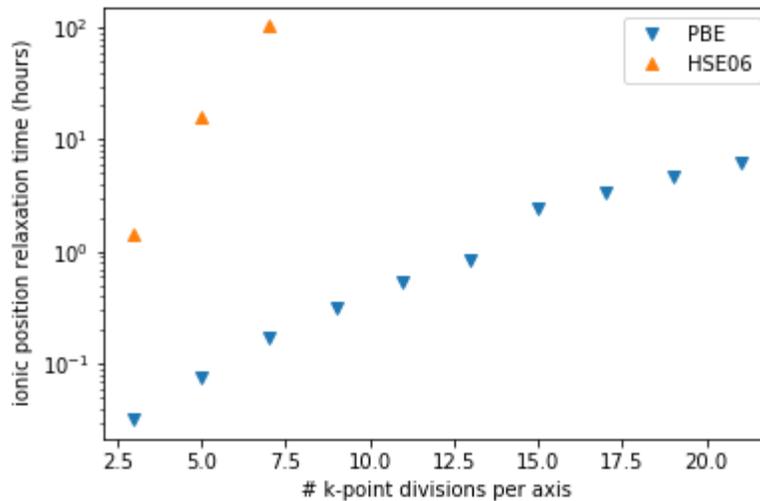


Figure 6. Effect of K-Point Mesh Density on Computation Time. Convergence testing results from repeated calculations with increasing k-point mesh density using a 16-atom BGaAs supercell containing a single boron atom ($B_1Ga_7As_8$). Computation times for both PBE pre-relaxation and final HSE06 are shown.

Figure 7 shows the convergence of the relaxed lattice constant with increasing k-point mesh density, for both the (a) PBE pre-relaxation and (b) final HSE06 volume relaxation steps. The lattice constant is increasing as the k-point mesh gets denser, but the overall change in the tested range is below 0.02% for the PBE pre-relaxation process and below 0.004% for the final HSE06 relaxation. Thus, lattice constant convergence by itself doesn't justify increasing the k-point density above $3 \times 3 \times 3$.

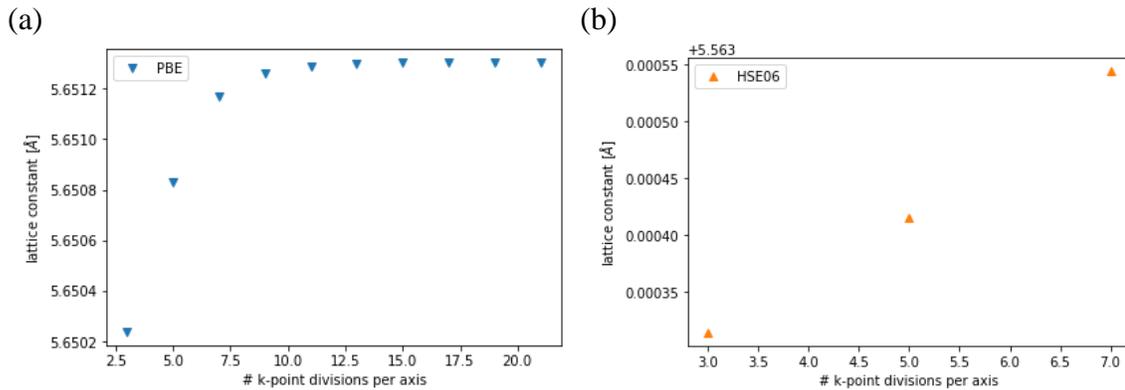


Figure 7. K-Point Convergence of Lattice Constant. Convergence testing results from repeated calculations with increasing k-point mesh density using a 16-atom BGaAs supercell containing a single boron atom ($B_1Ga_7As_8$). Increasing the density of the k-point mesh causes the fully relaxed lattice constant to increase, for both PBE pre-relaxation (a) and final HSE06 relaxation (b). However, the overall change in lattice constant is very small.

The convergence of the band gap with increasing k-point grid density is shown in Figure 8. The relative change is larger than in the case of lattice constant. Increasing the number of divisions along each direction from three to five causes a 16 meV reduction in the band gap energy, which is a 1.5% change. Further increasing the density to $7 \times 7 \times 7$ only decreases the band gap by about 2 meV, or 0.2%.

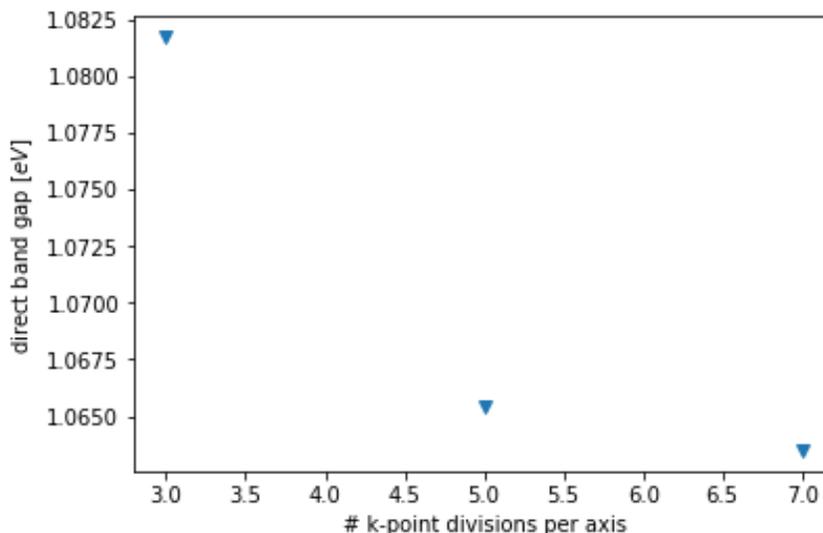


Figure 8. K-Point Convergence of Band Gap. Convergence testing results from repeated calculations with increasing k-point mesh density using a 16-atom BGaAs supercell containing a single boron atom ($B_1Ga_7As_8$).

All convergence testing was done with a 16-atom supercell due to its relatively fast turnaround time, but I can use the results to estimate the level of convergence of the calculations involving larger supercells with a simple back-of-the-envelope calculation. As the cell volume increases in real space, the size of the Brillouin zone in reciprocal space decreases proportionally. This means that the same number of k-point divisions generate a finer mesh for a larger supercell, and thus lead to better sampling of the Brillouin zone. For 54-, 64-, and 128-atom supercells the factor by which the volume is increased relative to a 16-atom supercell is 3.375, 4, and 8, respectively. Taking the cube root gives the factor along a single axis, yielding 1.5, 1.59, and 2. Thus, using a $3 \times 3 \times 3$ mesh for the larger supercells would yield a k-point density that is roughly equivalent of a 16-atom supercell with 4.5, 4.76, and 6 k-point divisions along each direction, for the 54-, 64-, and 128-atom supercells, respectively. So, while for 16-atom supercells using a $3 \times 3 \times 3$ may result in a 15-20 meV convergence error, this reduces

to the order of 2-3 meV for the larger supercells. assuming they follow a similar k-point convergence trend as seen in Figure 9.

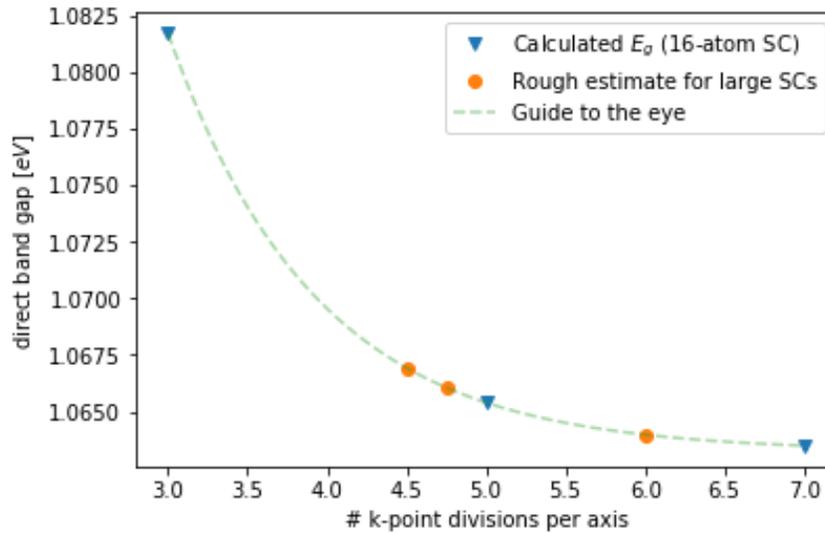


Figure 9. Estimation of Convergence Error for Large Supercells. Taking the k-point convergence testing results a 16-atom supercell (blue triangles) I can roughly estimate the convergence error of 54-, 64-, and 128-atom supercells (orange circles).

The entire workflow finished in a day or two for most 16-atom supercells, but the calculations slowed down significantly for larger supercells, even with increasing the number of CPU cores used in parallel. Most 54- and 64-atom supercells finished within a few weeks, while the 128-atom supercell simulation ran for several months total. As seen in Figure 6, increasing the number of k-point divisions from three to five comes with a roughly tenfold increase in computation time, which would make the larger supercell simulations prohibitively expensive. Thus, I decided to use a $3 \times 3 \times 3$ k-point mesh for all calculations.

Another parameter I was interested in is the energy cutoff of the planewave basis set (ENCUT parameter in the INCAR file). The default value for this parameter varies for each element, and it can be found in the corresponding pseudopotential file (POTCAR) as

the value for the ENMAX parameter. The VASP manual warns to increase the default value by 30% if we are performing volume relaxation due to errors introduced by Pulay stresses.¹⁰¹ I wanted to see the effect of this error on the equilibrium lattice constant, so for the same 16-atom B₁Ga₇As₈ supercell I also varied the ENCUT parameter from 150% to 250% of the default value, in steps of 25%.

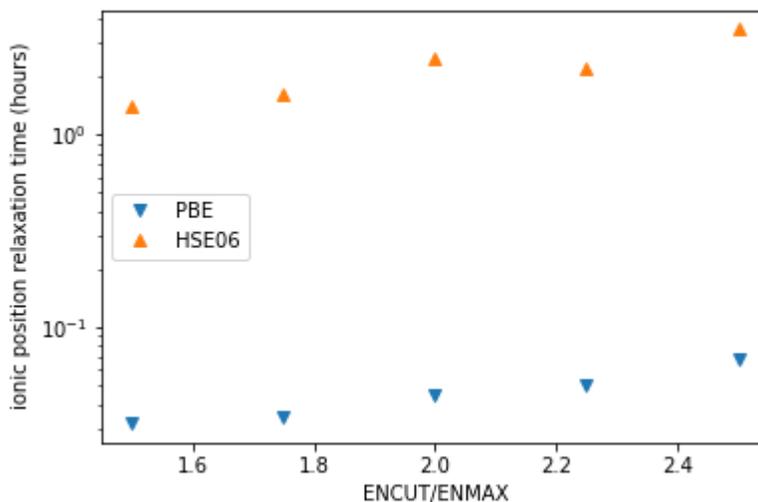


Figure 10. Effect of ENCUT Parameter Value on Computation Time. Convergence testing results from repeated calculations with increasing planewave energy cutoff (ENCUT parameter) using a 16-atom BGaAs supercell containing a single boron atom (B₁Ga₇As₈). The horizontal axis shows the factor of increase from the default value (ENMAX parameter in the pseudopotential file). The total computation time required for the ionic position relaxation step, both during PBE pre-relaxation and final HSE06 relaxation, shows a moderate increase.

A larger planewave energy cutoff results in a moderately increased execution time, as shown in Figure 10. However, it can be seen from Figure 11 that neither the lattice constant nor the band gap seems to converge with increasing ENCUT value. Thus, I decided to use a planewave cutoff energy of 475 eV, that is 50% larger than the ENMAX value given in the pseudopotential file for boron, which was the largest value out of the three elements.

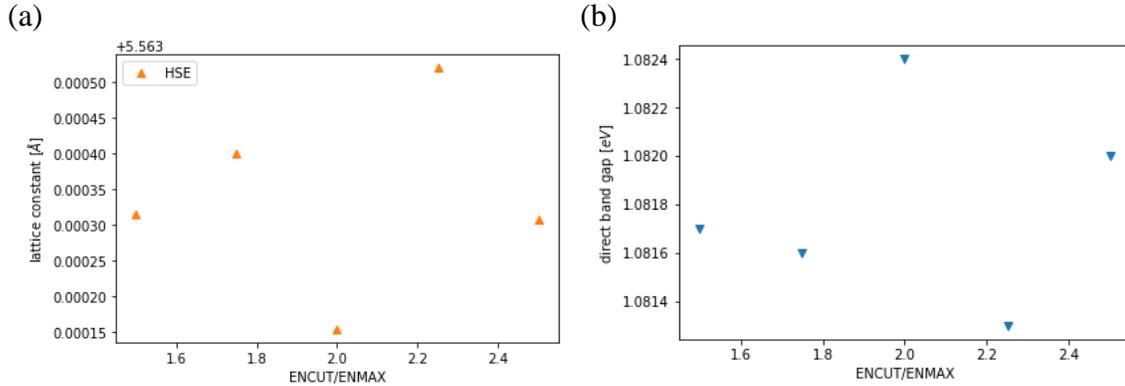


Figure 11. ENCUT Convergence of Lattice Constant and Band Gap. Convergence testing results from repeated calculations with increasing planewave energy cutoff (ENCUT parameter) using a 16-atom BGaAs supercell containing a single boron atom ($B_1Ga_7As_8$). The horizontal axis shows the factor of increase from the default value (ENMAX parameter in the pseudopotential file). Neither the (a) lattice constant nor the (b) band gap shows convergence with increasing ENCUT.

4. Analysis of VASP results

When the band structure calculation is complete, I extracted notable features from the unfolded band diagrams for comparison between supercells of different compositions. I located the valence band maximum (VBM) and conduction band minimum (CBM) and calculated the band gap energy. I extracted and tabulated the energy gap not just at the CBM, but along the Γ , L, and X symmetry points and examined the trends with regards to material composition. I also calculated electron effective masses in the direct band gap region by means of quadratic interpolation through the CBM at Γ and the lowest CB energy eigenvalue at the neighboring k-points in the direction towards the L, X, and K symmetry point, followed by taking the geometric mean of the three effective mass values.

Other than looking at the electronic properties of the material, I also analyzed the fully relaxed crystal structure. I examine the dependence of lattice constant on B content and the effect of B-B pairs on bond length. When simulating alloys with the same B

percentage, but different arrangement of atoms, I compare the total free energy of the structures to determine which configuration is more likely to form during crystal growth.

I did most of the analysis using the Python programming language, with libraries such as numpy¹⁰² and pandas¹⁰³ to help with calculation and tabulation of results. The pymatgen⁷⁵ library helped with extracting information from VASP output files, while matplotlib¹⁰⁴ was used for plotting the results. For structural analysis and ball-and-stick models I used the VESTA software.⁷⁴

B. Results and discussion

In this section I present the results of VASP simulations of BGaAs. All results discussed in sections 1 to 5 are from supercells of sizes 54, 64, and 128 atoms. For simplicity's sake I will refer to these supercell sizes collectively as large supercells. Results of simulations of 16-atom supercells, which I will refer to as small supercells, are discussed separately in section 6.

1. Band gap over the full composition range

Figure 12 shows the simulated band structures of pure GaAs and BAs using 64-atom supercells. The calculations yielded a direct band gap of 1.40 eV for pure GaAs, which is slightly under the experimental value of 1.42 eV.³⁰ In the case of pure BAs, the band diagram shows an indirect band gap of 1.84 eV near the X symmetry point, which is above earlier experimental reports of 1.46 eV,^{39,43} but is in better agreement with a more recent measurement of 1.77 eV.⁴⁴ The minimum direct band gap is 4.12 eV. Comparing with existing theoretical data, these are larger than values obtained using local density approximation (LDA)^{28,105–109} and GGA^{107,109–112} methods, but those are known to

underestimate the band gap.¹¹³ My result falls within the range of band gaps from previous calculations using GW^{28,105,106} and is in good agreement with recent studies involving hybrid functionals.^{44,114–117}

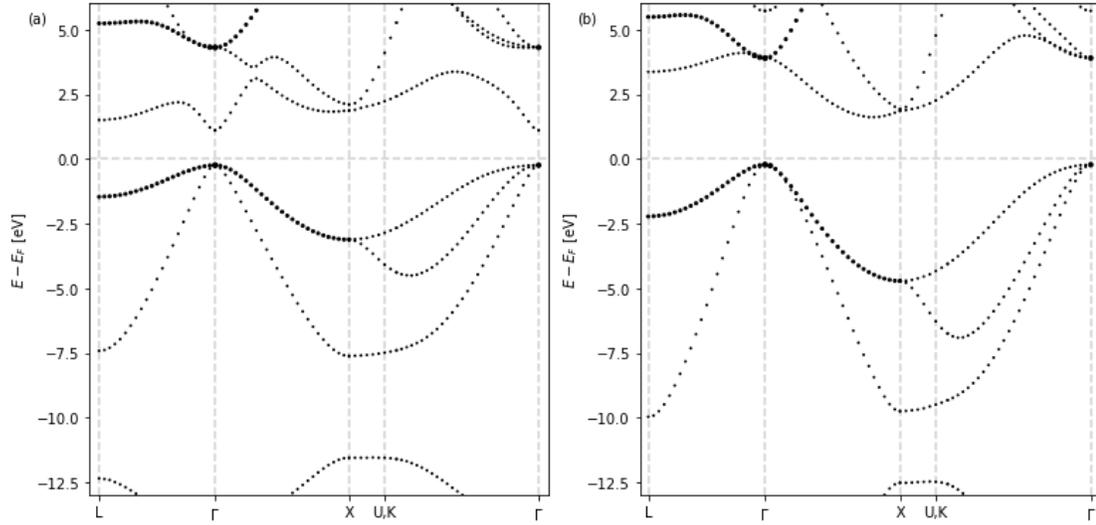


Figure 12. Band Structures of Pure GaAs and BAs. E-k diagrams of pure (a) GaAs and (b) BAs showing a direct band gap of 1.40 eV for GaAs and an indirect band gap of 1.84 eV near the X symmetry point for BAs. Both diagrams are from simulation results using 64-atom supercells.

Band gaps at the Γ , L, and X symmetry points of BGaAs alloys from large supercell simulations excluding 54-atom supercells with B atoms at nearest-neighbor group III sites are shown in Figure 13. The simulations indicate that in the dilute regime B does not have a large effect on the direct band gap. The band gap first decreases slightly to 1.34 eV at 7.4% B concentration, then starts to increase towards the 4.12 eV minimum direct gap of BAs. The actual lowest band gap result in the entire range is 1.30 eV at 12.5% B content, but the second random alloy with the same concentration yielded 1.40 eV, which gives a rough average of 1.35 eV. Similar trends of first decreasing, then

at higher concentrations increasing band gaps have been reported from both theory⁵² and experiment,⁵¹ but the B content corresponding to the band gap minimum differ from my result (30% and 2.5%, respectively). The L gap decreases in a roughly linear fashion up to 25% B concentration. To estimate the transition of BGaAs from direct to indirect gap I used simple quadratic fits on the L and Γ band gaps, as shown with the dashed and dash-dotted lines. I predict that the shift from direct to indirect band gap happens at around 17% B content. The conduction band (CB) minimum is in the L-valley for only the 25% B concentration random alloy, then shifts towards X as the L gap starts to increase at higher percentages.

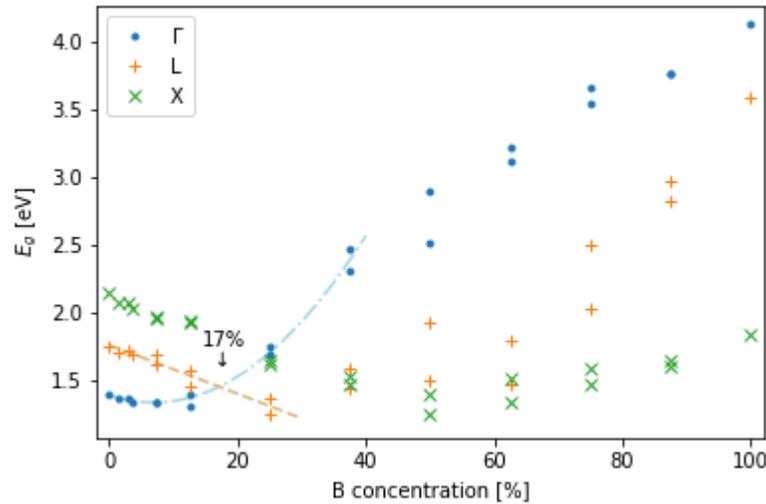


Figure 13. BGaAs Band Gaps Over the Full Composition Range. The figure shows band gaps at the Γ , L, and X symmetry points versus B concentration. All data points at or below 7.4% are from supercells of various sizes with isolated B atoms, while points at 12.5% and above are from 64-atom random alloy supercells. Using quadratic fits, shown with dashed and dash-dotted lines, I estimate the direct-indirect gap transition at around 17%.

2. Dilute regime

Figure 14 shows the effect of B on the band gaps at the Γ , L, and X symmetry points from all large supercell simulations in the dilute regime. The narrowest band gap observed among supercells with isolated B atoms is 1.33 eV, corresponding to a 3.7% B content. This means that the effect of lone B atoms on the direct band gap is small (within 5%) for B concentrations under 13%. The narrowing is larger for alloys with B-B pairs in nearest-neighbor (NN) group III sites. The 54-atom supercell with 2 B atoms in NN configuration have a direct gap of 1.23 eV as opposed to 1.34 eV for the two other supercells of the same size and composition. This means the presence of B-B pairs almost triples the narrowing of the band gap (12%) as opposed to the isolated B atoms (4.3%). The 54-atom supercell with a continuous line of B atoms, corresponding to an 11% B concentration, exhibit a direct gap of 1.06 eV, which is a 24% decrease from pure GaAs. This confirms earlier theoretical results that B-B pairs reduce the band gap more than isolated B atoms.⁴⁸

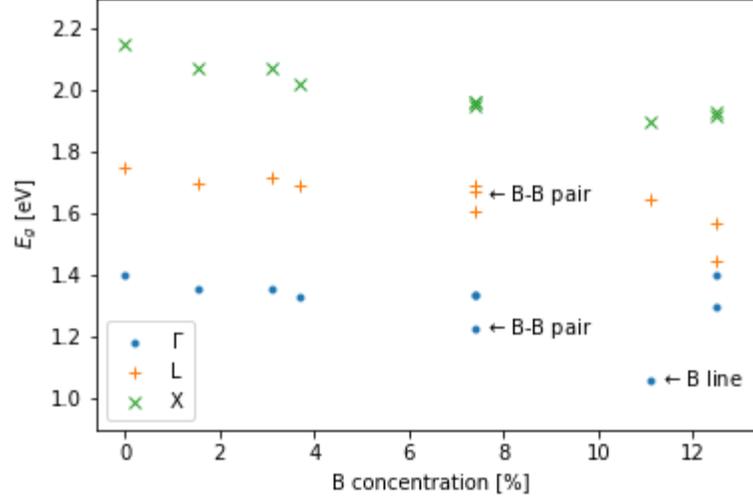


Figure 14. Band Gaps of Dilute BGaAs Alloys at High-Symmetry Points. The points at 7.4% are from 54-atom supercells with two B atoms, the marked points correspond to B-B pairs at nearest-neighbor group III sites (NN). The points at 11%, with the Γ -point marked with an arrow, correspond to a 54-atom supercell with 3 B atoms in NN sites forming a continuous line. Isolated B atoms do not have a significant effect on the band gap, while B-B pairs in NN sites cause a significant narrowing.

The total free energy of the 54-atom supercell with 2 B atoms at NN sites is -272.56 eV while the other two systems with the same composition but isolated atoms equally yielded -272.82 eV, so placing the B atoms at NN sites increased the total energy by 260 meV, or 4.8 meV per atom. This implies that B-B pairs are energetically unfavorable to evenly distributed B atoms, though the difference in energy is small.

The increased free energy can be explained with the stretching of bonds around B-B pairs at NN sites. Measuring all 4 bond lengths for both B atoms in the fully-relaxed supercells I found that the average B-As bond length in the structure with B-B pairs is 2.19 Å, as opposed to 2.17 Å for the supercells with isolated B atoms, which is an average stretching of 1.2%. In the NN configuration the longest bonds are along the B-As-B triangle with lengths of 2.24 Å, while the shortest bonds are in the direction

perpendicular to the plane defined by the B-As-B atoms, with a length of 2.15 Å. This is a 4.1% difference, while for the two supercells with isolated B atoms the difference between the longest and shortest bonds is 0.36% and 0.22%, respectively.

3. Electron effective mass

I extracted the electron effective masses from the band structures of BGaAs alloys in the direct band gap regime, as shown in Figure 15. The simulations yield an m_e^* of $0.068 m_0$ for GaAs, which is in excellent agreement with the known experimental value.¹¹⁸ The electron effective mass increases in a roughly linear fashion to $0.123 m_0$ at 12.5% B concentration. B-B pairs in NN sites do not seem to have a strong influence on effective mass. Among the three 52-atom supercells with 2 B atoms, the one with NN B-B pairs is in the middle with $0.094 m_0$, while the two structures with isolated B atoms yield a CB dispersion of $0.095 m_0$ and $0.088 m_0$. Thus I cannot confirm the strong change in electron effective mass that was reported by Lindsey et al.⁴⁸

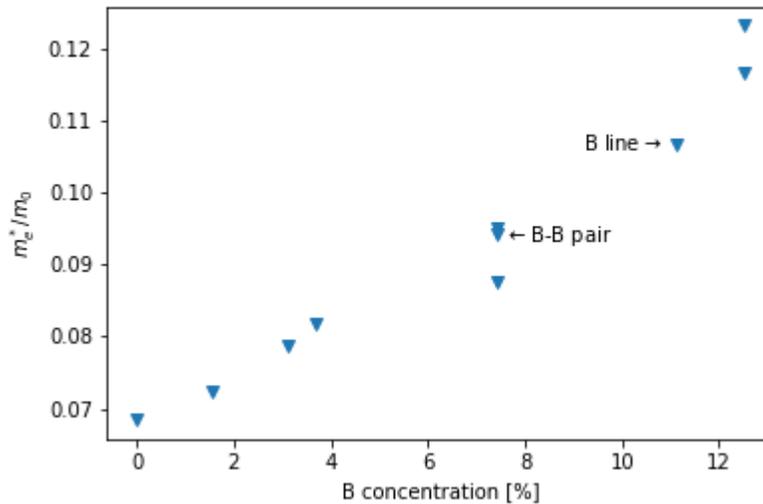


Figure 15. BGaAs Electron Effective Mass in the Direct Band Gap Region. The electron effective mass shows a roughly linear relationship with B concentration. B-B pairs don't seem to have a drastic effect on electron effective mass.

4. Lattice constant

Figure 16 shows the geometric mean lattice constant versus B concentration. The fully-relaxed lattice parameters for pure GaAs is 5.667 Å, a slight overestimate of the known value of 5.653 Å,³⁰ while for pure BAs it is 4.771 Å, which is very close to the 4.777 Å³⁸⁻⁴⁰ reported experimentally. The lattice parameter varies linearly with B content, which means BGaAs does not deviate from Vegard's Law, contrary to earlier reports.^{20,52} I calculated that the B percentage that would be required to match BGaAs to the silicon lattice constant of 5.431 Å³⁰ is 26.3%. This is above the 17% limit predicted for the direct-indirect gap transition, so it does not appear possible to achieve a Si lattice matched direct semiconductor by only adding B to GaAs.

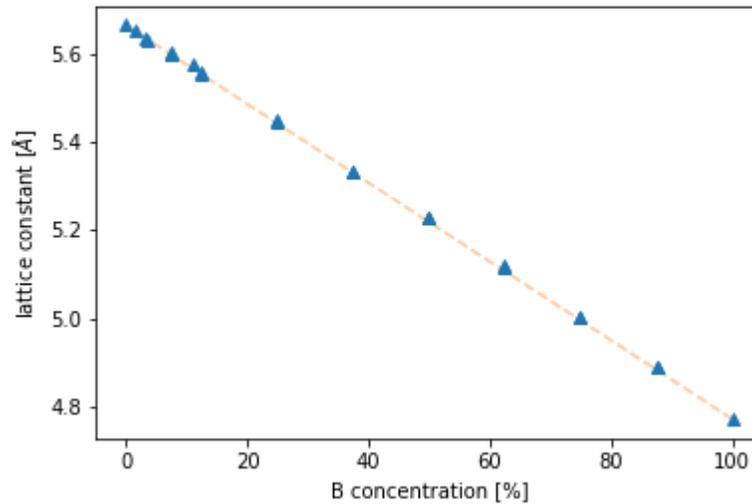


Figure 16. BGaAs Lattice Constant Versus B Concentration. The lattice constant shows perfect agreement with Vegard's Law, shown by the orange dashed line.

5. Band unfolding artifacts

All electronic properties presented in this work have been extracted from effective band structures unfolded using the BandUP code.^{86,87} In an unfolded band diagram, the

energy eigenvalues are represented with the spectral weight. When unfolding the band structure of a supercell made of a perfect crystal, such as the GaAs and BAs shown in Figure 12, the spectral weight is an integer number corresponding to the degeneracy of the energy level. Once imperfections are added to the crystal supercell, breaking symmetry, the spectral weight is no longer integer, and it can be interpreted as the amount of primitive cell Bloch character of an eigenstate.⁸⁷ In practice this means that faint images of the folded bands remain in the effective band structure, and every L and X energy level also shows up at Γ with a smaller spectral weight. Most of the time these shadow bands are faint enough that the effective band diagram can be cleaned up by simply discarding energy eigenvalues that have spectral weights below a certain threshold. In this work I filtered out spectral weights under 0.07. In some cases, the shadow bands have spectral weights comparable to the real bands, which can make it difficult to determine the exact location of the conduction band minima.

Figure 17 shows energy eigenvalues at the Γ , L, and X symmetry points extracted from unfolded effective band structures of the two sets of 64-atom random alloys used in this work. The area of the markers is proportional to the spectral weights. I filtered out eigenvalues with weights below 0.07, but the corresponding markers would be so small that they would not be visible on the plot either way. The separation of results into two sets is arbitrary but plotting them separately means that only one supercell result is plotted at each percentage value, which makes the interpretation of the figure easier. The only supercells where the determination of the conduction band minimum is not obvious is at 25% B concentration, in both sets of random alloys, shown at the bottom with overlapping markers of similar size.

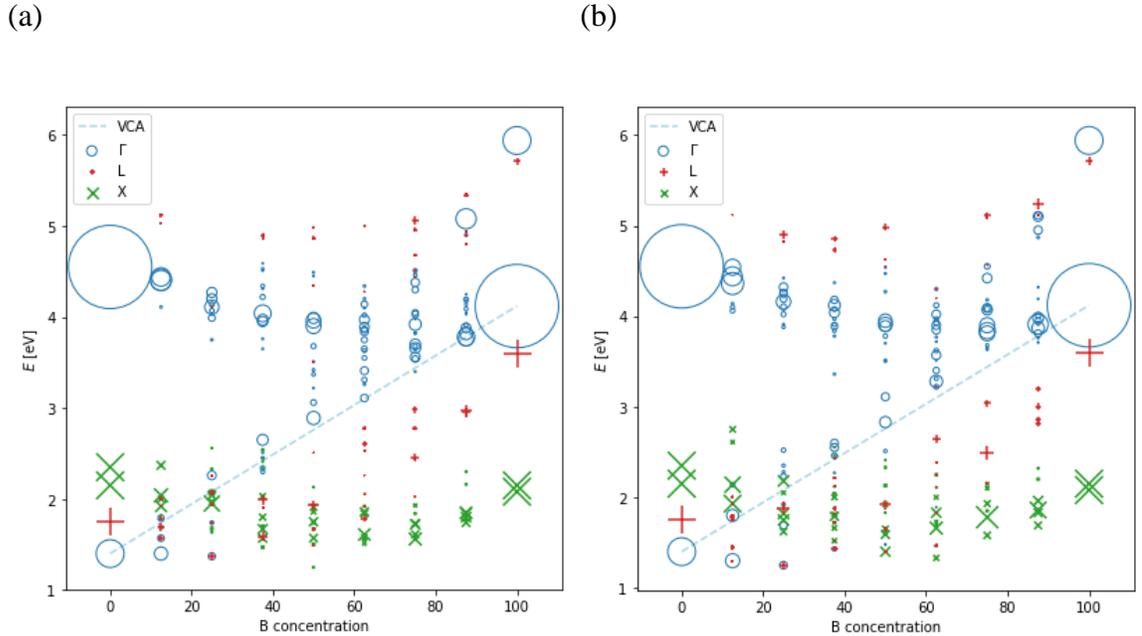


Figure 17. Conduction Band Energy Eigenvalues of Random Alloys. The energy levels, shown at the Γ , L, and X Symmetry Points, are measured relative to the valence band maximum, so the bottom markers represent the band gaps. The area of each marker corresponds to the spectral weight of the eigenvalue. The dashed line shows the minimum direct band gap as predicted by the virtual crystal approximation (VCA) without bowing. The separation of results into sets (a) and (b) is arbitrary and is only done to increase the clarity of the plots.

The effective band structure of the two random alloys with 25% B content is shown in Figure 18. A mirroring of both the valence and conduction bands is visible between the L and Γ symmetry points for alloy (a), while for alloy (b) the reflection of the valence band is fainter. I repeated the unfolding procedure at an increased resolution of 1 meV and found that the lowest CB energy levels at L and Γ are exactly equal. For alloy (a) the spectral weight of the lowest CB eigenvalue is about 30% larger in L, but for alloy (b) it is about 2% larger in Γ . I compared the eigenstates at neighboring k-points along the L- Γ line equidistant from the symmetry points and found that the energy levels are exactly equal, and the corresponding spectral weights are always larger for k-points

closer to L. It is very unlikely that a real band would have the exact same energy levels up to 4 significant figures and a perfectly matching dispersion at both L and Γ , so I classify the Γ portion as an unfolding artifact, a shadow of the real band at L. I used the same procedure to decide the CB level at L for one of the random alloys with 12.5% B concentration, seen in Figure 17 (a) as the second data point from the bottom.

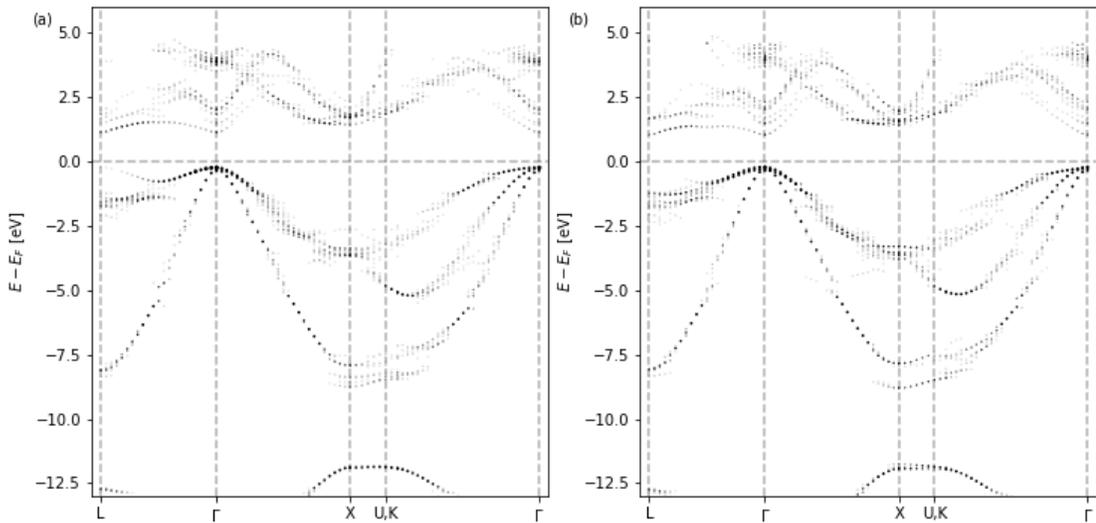


Figure 18. Unfolding Artifacts in Effective Band Structures of $B_8Ga_{24}As_{32}$ Supercells. The two plots show two different random arrangements of atoms with the same overall B concentration. The area of the dots corresponds to the spectral weights of the energy levels. Both structures show unfolding artifacts in the conduction band between the L and Γ symmetry points, while in structure (a) the reflection is also strongly visible in the valence band. The mirroring of the conduction bands seen here corresponds to the bottom data points in Figure 17 at 25% B concentration where two markers overlap.

Figure 19 shows band gaps at the Γ , L, and X symmetry points in the full composition range, excluding 54-atom supercells with B atoms at NN sites. In plot (a) the supercells with 25% B content, and ambiguous CMB k-point location, are classified as direct band gap, while plot (b) shows them classified as indirect band gap. If I interpret

the supercells in question as indirect band gap, then the Γ and L gaps align in a more continuous, linear fashion.

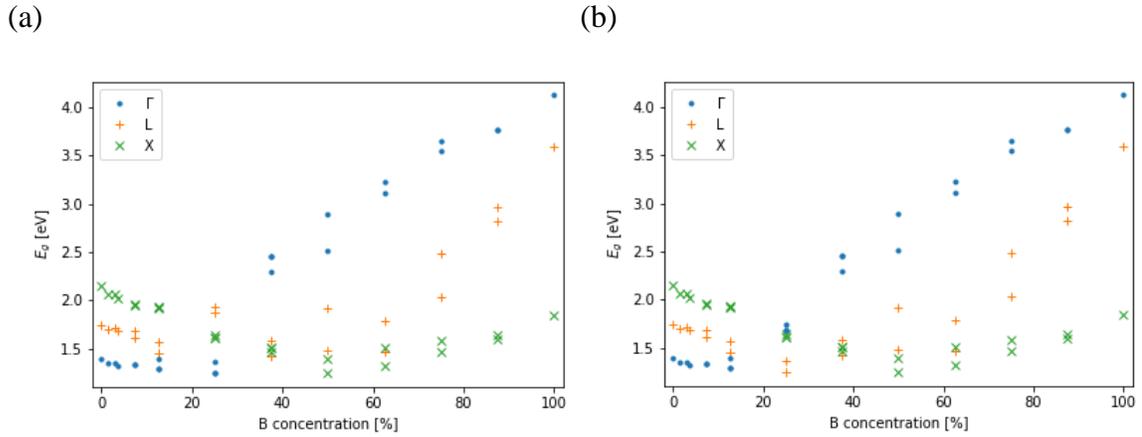


Figure 19. Classifying Ambiguous Conduction Band Minima. The figures show band gaps at the Γ , L, and X symmetry points versus B concentration if the CBM of the 25% random alloys is classified as (a) located at Γ or (b) located at L. Placing the CBM in L results in more even trends over the full composition range.

6. Small supercell calculations

I discuss the 16-atom supercell calculations separately because the results disagree with the 64-atom random alloys. Figure 20 shows the band gaps at the Γ , L, and X symmetry points versus B concentration for 16-atom supercell simulations. For BAs the direct band gap is 1.85 eV, which is in perfect agreement with the 1.84 eV result seen with the 64-atom supercell. For GaAs, however, the band gap shrinks to 1.37 eV from the 1.40 eV obtained using the larger supercell, which is a worse underestimation of the known 1.44 eV gap. The differences grow more significant for mixed alloys. There is a large narrowing of the band gap, even at relatively low B concentrations. At 12.5% B content, which is one B atom in the 16-atom supercell, the band gap shrinks to 1.08 eV, which is more, than a 20% decrease from the GaAs small-supercell direct band gap. In comparison, the band gaps for the two large random alloys with the same B concentration

were 1.30 eV, which is a 7% narrowing, and 1.40 eV which exactly matches the large-supercell GaAs band gap. For the 16-atom supercells the narrowing continues at higher B percentages, all the way down to a 0.65 eV indirect (X) band gap found at 62.5% B concentration. There are no clear observable trends in the band gaps at Γ , L, and X, unlike with the 64-atom random alloys. Unfolding artifacts with reflected bands were seen through the entire composition range, which also makes the interpretation of the results difficult.

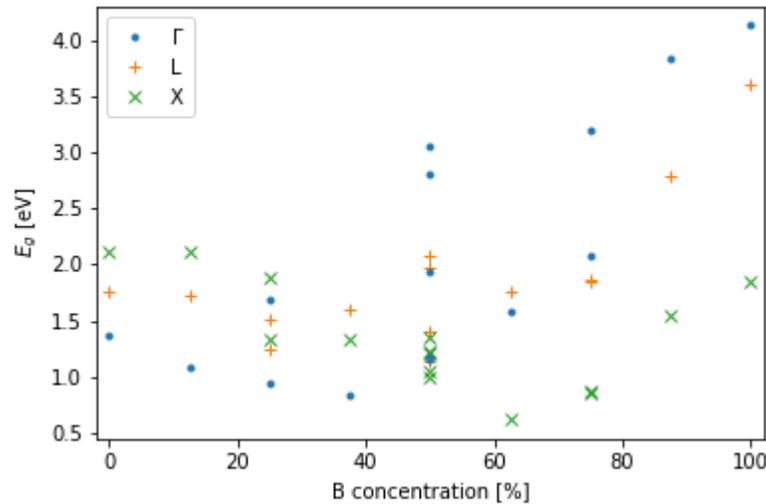


Figure 20. Band Gaps of 16-Atom BGaAs Supercells at High-Symmetry Points.

The figure shows band gaps at the Γ , L, and X symmetry points versus B concentration for small supercells. There is a significant narrowing of the band gap as B content approaches 50%, unlike the 64-atom random alloys.

The effect of B concentration on lattice constant also shows a deviation from Vegard's Law, as seen in Figure 21. The lattice parameters of pure GaAs and BAs are 5.667 Å and 4.771 Å, respectively, which is in perfect agreement with the 64-atom supercells. But at varying B concentrations, compared to the linear relationship seen for all large-supercell calculations, the lattice constant shows a noticeable negative bowing

with a quadratic bowing parameter value of -0.11 \AA . I extracted the bowing parameter value using least-squares fitting to the equation

$$a_{\text{B}_x\text{Ga}_{1-x}\text{As}} = xa_{\text{BAs}} + (1-x)a_{\text{GaAs}} - cx(1-x) \quad (13)$$

where a indicates the lattice constant, x is the fractional B content, and c is the quadratic bowing parameter.

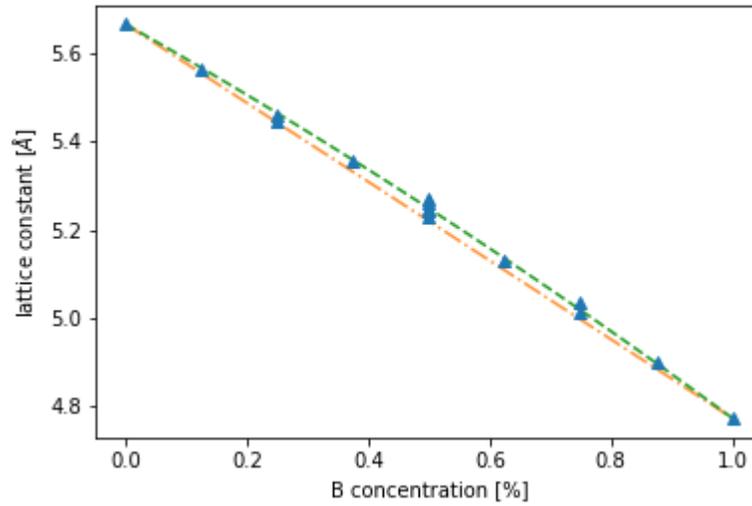


Figure 21. Lattice Constant of 16-Atom BGaAs Supercells. The lattice constant deviates from Vegard's law, shown with orange dash-dotted line. The extracted bowing parameter from a quadratic fit (green dashed line) is $c = -0.11 \text{ \AA}$.

Other than the supercell size and the arrangement of atoms, all VASP settings were identical for small and large supercell calculations. As I already discussed in the chapter on convergence testing, the effective k-point sampling of the Brillouin zone does get finer as the supercell size increases, even if the number of k-point divisions per axis stays the same, due to the inverse relationship between the volumes in real and reciprocal space. However, as seen in Figures 7 and 8, the effect of k-point mesh density on band gap and lattice constant is too small to explain the difference between the results obtained

with small and large supercells. Also, increasing k-point mesh density during convergence testing caused a narrowing of the band gap, which is contrary to the increased band gap seen in large-supercell results that correspond to a finer k-point sampling.

The most likely explanation for the significant differences seen with small supercells is the artificial periodicity introduced by the use of periodic boundary conditions. The wavefunctions calculated by VASP are solutions of a boundary value problem that is effectively an infinite repetition of the supercell. Any imperfection introduced to the supercell will be repeated with a period that is proportional to the supercell size. This introduces an additional translational symmetry into the system which distorts the solutions. The effect diminishes as supercell volume increases, provided that the arrangement of atoms in the supercells are sufficiently random. Since the larger supercells provide a better representation of the randomness of real alloys, and with the two sets of random alloys showing good agreement in band gap and lattice constant, I consider the results of the 16-atom supercells to be unrealistic, and conclude that the results from larger supercells are likely a better approximation of the properties of real BGaAs alloys.

C. Future Work

The calculations presented in this work ignore spin-orbit interaction. A follow-up study could explore the effect of spin-orbit coupling on the band gap, valence band dispersion, and the location of the split-off band relative to the valence band maximum, as well as the variation with B content. Examining band-decomposed partial charge densities could shed light on the localization of the states introduced by B atoms in dilute

alloys and may help explain the band gap narrowing caused by B-B pairs in nearest-neighbor group III sites. Future work may also include simulating optical properties, such as absorption coefficient, index of refraction, and electron transition rates. Considering the high thermal conductivity of pure BAs, the effect of B concentration on thermal conductivity could be investigated with phonon simulations. The probability of phase separation in BGaAs could be examined by comparing energetics of random alloys with mixed B-Ga distribution in group III sites to supercells with BAs-GaAs clusters. Additional work may include simulation of point defects, such as vacancies, split interstitials, and antisite substitution.

D. Summary

I calculated band structures of $B_x\text{Ga}_{1-x}\text{As}$ in the full composition range using HSE06 hybrid functionals. I found that as B concentration increases, the direct band gap first narrows slightly, then increases towards the large direct gap of BAs. The effect of B on the band gap is less than 7% for B concentrations under 13%. I estimate that BGaAs transitions from direct to indirect gap at around 17% B content. I found that B-B pairs in nearest neighbor group III sites cause a narrowing of the band gap that is 3-4 times larger than isolated B atoms with the same composition, but do not have a significant effect on the electron effective mass. My results show that the lattice constant of BGaAs does not deviate from Vegard's law and I estimate that the B concentration required to lattice match BGaAs to Si is 26.3%, at which point BGaAs is an indirect gap material. I found that results from small-supercell calculations show good agreement with large supercells for pure GaAs and BAs but significantly differ for mixed alloys, where I observed a large reduction in band gap and deviation from Vegard's law. I discussed that the difference

cannot be explained by too coarse k-point sampling and is likely caused by the distorting effect of the artificial periodicity introduced by the infinite repetition of the supercell, which diminishes as supercell size increases. Since large-supercell random alloys are a better representation of the randomness of real alloys, and since I have two sets of random alloys showing good agreement, I dismiss the small-supercell results as unrealistic and consider the large-supercell random alloys as the better approximation of real BGaAs alloys. I discussed band unfolding artifacts and showed that even though shadow bands can complicate the interpretation of unfolded effective band structures, the exact location of the conduction band minima can be determined with confidence by comparing the spectral weights and band dispersion at increased energy resolution.

IV. OPTIMUM TFET PERFORMANCE WITH HIGHLY MISMATCHED

ALLOYS

HMAAs can allow unprecedented control over material properties such as band gap or effective mass. In the second part of this thesis, to explore some of the new possibilities in device design opened up by such materials, I investigate the effect of various material and device parameters on the performance of a double-gate tunnel field-effect transistor (DG TFET) using computer simulations based on the UCSD TFET model.¹¹⁹ In this chapter I first give a summary of the derivations of the original model and discuss my modifications, then I present the parameters that yield optimum device performance and discuss the theoretical reasons behind the trends observed in the results.

A. Method

1. TFET model

My simulations are based on the UCSD TFET model,¹¹⁹ which is a semi-analytic model of tunneling in DG TFET devices, originally developed by Taur et al.,^{7,8} and which was also verified by numerical simulations.^{120,121} I further extended the model to include Kane's non-parabolic dispersion relation.^{9,10} The UCSD model has been revised by its authors several times,¹²⁰⁻¹²³ so I present a unified and detailed description below. A DG TFET device with the corresponding band structure is shown in Figure 22, where I also marked the variables used in this section.

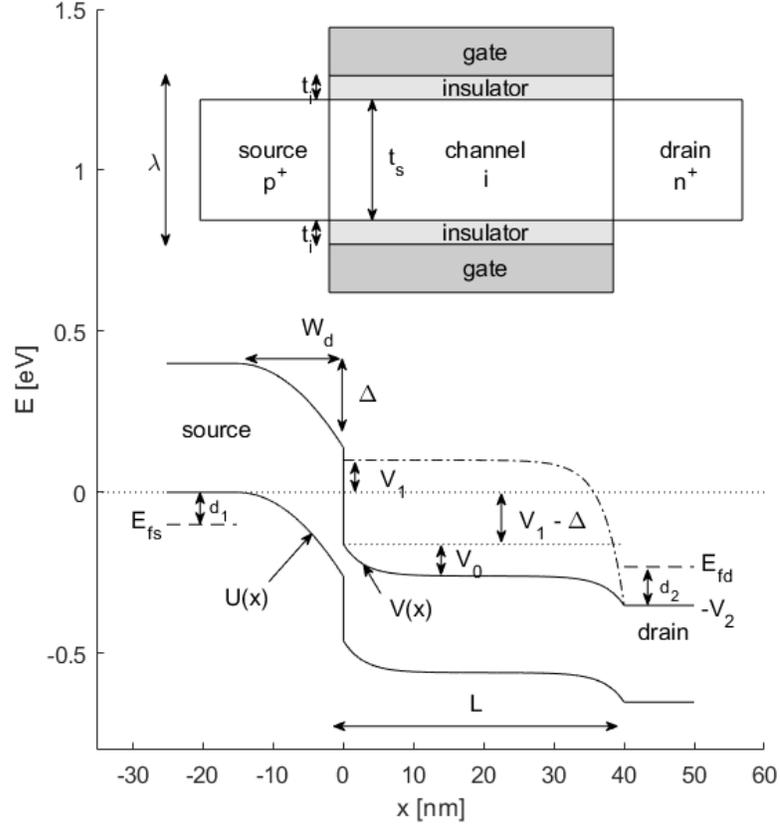


Figure 22. Schematic and Band Diagram of a DG TFET Device. Solid lines show the valence and conduction bands when the device is ON, the dash-dotted line is the channel conduction band in the OFF state.

First, to accurately calculate the curvature of the bands near the tunneling region, the 2D Poisson equation for a DG TFET of channel length L_g is solved analytically for the channel potential, written as a series of eigenfunctions

$$\psi(x, y) = V_g - \Delta\phi - \sum_{n=1}^{\infty} \left[\frac{b_n \sinh[\pi(L-x)/\lambda_n] + c_n \sinh(\pi x/\lambda_n)}{\sinh(\pi L/\lambda_n)} \sin\left(\frac{n\pi}{2} + \frac{\pi y}{\lambda_n}\right) \right] \quad (14)$$

where $\Delta\phi$ is the net gate work function and λ_n are the discrete eigenvalues, and the long-channel potential is represented by the constant term $V_g - \Delta\phi$. The coefficients b_n and c_n

are determined by source and drain boundary conditions,¹²⁴ while the eigenvalues depend on the material's dielectric constant and the film's thickness.

In case of a symmetric DG TFET, only the odd-order eigenfunctions remain. The eigenvalues λ_n are found using

$$\tan\left(\frac{\pi t_i}{\lambda_n}\right) \tan\left(\frac{\pi t_s}{2\lambda_n}\right) = \frac{\varepsilon_i}{\varepsilon_s} \quad (15)$$

with t_i and ε_i as the thickness and dielectric constant of the insulator, and t_s and ε_s are the same for the semiconductor. If $\varepsilon_i = \varepsilon_s$, Eq. 14 yields $\lambda_n = (t_s + 2t_i)/n$, with the first eigenvalue, $\lambda = t_s + 2t_i$ defined as the scale length. Note that λ is the distance separating the two gates.

The conduction band of the channel is found from Eq. 13, taking $n = 1$ and $\varepsilon_i = \varepsilon_s$ as

$$V(x) = V_0 \frac{\sinh[\pi(L-x)/\lambda]}{\sinh(\pi L/\lambda)} + V_1 - V_0 - \Delta + (V_0 + \Delta - V_2 - V_1) \frac{\sinh(\pi x/\lambda)}{\sinh(\pi L/\lambda)} \quad (16)$$

where the constant term $V_g - \Delta\phi$ in Eq. 13 was replaced by $V_1 - V_0 - \Delta$ and the coefficient values $b_1 = V_0$ and $c_1 = V_0 + \Delta - V_2 - V_1$ were substituted in to satisfy the boundary conditions. V_1 is the effective band gap between the valence band of the source and the conduction band of the channel in the off state. Δ is the potential variation in the source due to band bending and V_0 is the gate control of the conduction band in the channel. Thus, the first boundary condition is $V(0) = V_1 - \Delta$. The second boundary condition is $V(L) = -V_2$, which also defines the level of the conduction band of the drain. V_2 is expressed as

$$V_2 = qV_{ds} + d_1 + d_2 \quad (17)$$

where d_1 and d_2 are the degeneracies in the source and drain, respectively.

The original UCSD model used half-order Fermi-Dirac integrals to solve for the degeneracies assuming a parabolic dispersion relation. I accounted for the non-parabolic nature of the energy bands using a simplified form¹⁰ of Kane's non-parabolic approximation⁹

$$E(k)[1 + \alpha E(k)] = \frac{\hbar^2 k^2}{2m^*} \quad (18)$$

where α is the non-parabolicity factor. Note that for $\alpha = 0$, the expression reduces to the traditional parabolic dispersion relation. Using this, the density of states is

$$g(E) = \frac{m^*(1 + 2\alpha E)\sqrt{2m^*E(1 + \alpha E)}}{\pi^2 \hbar^3} \quad (19)$$

Assuming full ionizations of the dopants I find the degeneracies by setting

$$N_d = n = \int_{E_c}^{\infty} g_c(E) f(E, d_2) dE \quad (20)$$

$$N_a = p = \int_{-\infty}^{E_v} g_v(E) [1 - f(E, d_1)] dE$$

and solving numerically for d_1 and d_2 .

Assuming the depletion approximation and a uniform dielectric constant across the heterojunction

$$\begin{aligned} \frac{1}{q} \left| \frac{dV}{dx} \right|_{x=0} &= \left(\frac{\pi}{q\lambda} \right) \frac{V_0 \cosh(\pi L/\lambda) + V_2 - V_0 + V_1 - \Delta}{\sinh(\pi L/\lambda)} \\ &= \frac{2\Delta}{qW_d} = \sqrt{\frac{2N_a\Delta}{\epsilon_s}} \end{aligned} \quad (21)$$

where the depletion width is given as

$$W_d = \sqrt{2\varepsilon_s \Delta / (q^2 N_a)} \quad (22)$$

The gate voltage V_{gs} is defined to be zero just when the channel conduction band lines up with the source valence band. At this point $V_0 = V_1 - \Delta$. When $V_{ds} > V_{gs}$, mobile charge in the channel is negligible (saturation conditions), so Δ can be simply calculated from

$$qV_{gs} = V_0 - (V_1 - \Delta) \quad (23)$$

However, when $V_{ds} < V_{gs}$, the TFET is biased in the linear region, and the Fermi level of the drain is close to or above the conduction band. This causes an inversion charge that has a de-biasing effect^{8,125,126} and degrades V_{gs} by the potential difference across the insulator, Q_{inv}/C_{ox} . By solving the Poisson equation for a DG MOSFET¹²⁷ the inversion charge is obtained as

$$Q_{inv} = \frac{4kT\varepsilon_s}{qt_s} \beta \tan \beta \quad (24)$$

with β solved numerically from

$$\begin{aligned} \frac{q(V_{gs} - V_{ds} - d_1)}{2kT} - \ln \left[\frac{2}{t_s} \sqrt{\frac{2\varepsilon_s kT}{q^2 N_c}} \right] \\ = \ln \beta - \ln(\cos \beta) + \frac{2\varepsilon_s t_i}{\varepsilon_i t_s} \beta \tan \beta \end{aligned} \quad (25)$$

where N_c is the conduction band effective density of states. Once Q_{inv} is calculated, the gate bias is modified as $V_{gs} - Q_{inv}/C_{ox}$.

Expressing Δ from Eq. 22 and substituting into Eq. 20 yields

$$\left(\frac{\pi}{q\lambda} \right) \frac{V_0 \cosh(\pi L/\lambda) + V_2 - qV_{gs}}{\sinh(\pi L/\lambda)} = \sqrt{\frac{2N_a(qV_{gs} - V_1 - V_0)}{\varepsilon_s}} \quad (26)$$

which is solved for V_0 . With that, Δ and W_d are obtained from equations 21 and 22.

Thus, the valence band of the source can be expressed as

$$U(x) = -\frac{q^2 N_a}{2\epsilon_s} (x + W_d)^2 \quad (27)$$

The potential function for the depletion region in the drain is neglected here because it has a negligible effect on the tunneling current.

The band-to-band tunneling probability is calculated using WKB integrals. For a particular energy level $-E$, where $0 < E < \Delta$, we have a hole tunneling from the valence band of the source to the conduction band of the channel. A hole or an electron with a total energy E may have some kinetic energy component E_{\perp} in the direction perpendicular to the junction. We call this transverse kinetic energy component $E_{\perp v}$ in the valence band of the source and $E_{\perp c}$ in the conduction band of the channel. The tunneling probability is then calculated as the product of two separate tunneling events, a hole-like process on the source (VB) side followed by an electron-like process on the channel (CB) side

$$T(E, E_{\perp v}) = \exp \left\{ -2 \left[\int_{l_1}^0 k_{is}(E, E_{\perp v}) dx + \int_0^{l_2} k_{ig}(E, E_{\perp c}) dx \right] \right\} \quad (28)$$

where k_{is} and k_{ig} are the tunneling wave vectors in the source and the channel, respectively. Both integrals are evaluated numerically. The integration endpoints, which are the physical (x) coordinates of the start and end points of the tunneling event, are solved from $U(l_1) = -E + E_{\perp v}$ and $V(l_2) = -E - E_{\perp c}$.

To calculate the imaginary wave vectors, Kane's non-parabolic dispersion relation is used (Eq. 5), with the appropriate energy term for the barrier height. In the source, the potential barrier for the hole is determined by $U(x)$, calculated in Eq. 26, as

$$k_{is}(E, E_{\perp v}) = \frac{\sqrt{2m_v^*}}{\hbar} \sqrt{\mathcal{E}_s(1 + \alpha_v \mathcal{E}_s)} \quad (29)$$

$$\mathcal{E}_s = -U(x) - (E - E_{\perp v})$$

While on the other side of the junction, the barrier height is given by $V(x)$, found in Eq. 3 as

$$k_{ig}(E, E_{\perp c}) = \frac{\sqrt{2m_c^*}}{\hbar} \sqrt{\mathcal{E}_g(1 + \alpha_c \mathcal{E}_g)} \quad (30)$$

$$\mathcal{E}_g = -V(x) + E + E_{\perp c}$$

For both cases the energy of the particle with respect to the tunneling barrier height is reduced by the transverse kinetic energy component, since E_{\perp} does not contribute to the tunneling.

The transverse kinetic energy component in the conduction band, $E_{\perp c}$, is related to $E_{\perp v}$ in the valence band by conservation of momentum

$$m_c^* E_{\perp c} (1 + \alpha_c E_{\perp c}) = m_v^* E_{\perp v} (1 + \alpha_v E_{\perp v}) \quad (31)$$

where m_c^* and m_v^* are electron and hole effective masses and α_c and α_v are the non-parabolicity factors in the conduction and in the valence band, respectively.

Finally, the tunneling current is given by the Landauer equation¹²⁸

$$I_{ds} = \frac{qm_v^*}{2\pi^2 \hbar^3} \int_0^{V_2} (f_s - f_d) \left[\int_0^{E_{\perp max}} T(E, E_{\perp v}) dE_{\perp v} \right] dE \quad (32)$$

with $T(E, E_{\perp v})$ as the tunneling probability calculated with Eq. 26, and f_s and f_d are the source and the drain occupancies, respectively. The inner integration is done with respect to the transverse kinetic energy in the valence band, $E_{\perp v}$, and Eq. 29 is used to calculate $E_{\perp c}$ for the second WKB integral. The upper limit of the inner integral, $E_{\perp max}$, is either

the total kinetic energy in the valence band, $E_{\perp v} \leq E$, or in the conduction band, $E_{\perp c} \leq V_2 - E$, whichever is smaller.

To account for non-parabolic bands, I had to modify all parts of the calculation that are directly derived from the dispersion relation. This includes determining the degeneracies d_1 and d_2 (Equations 17-19), calculating the imaginary wave vectors k_{is} and k_{ig} (Equations 28-30), and the WKB integrals used to calculate the tunneling probability (Eq. 27). In the original model, the first WKB integral, corresponding to the source side of the junction, was carried out analytically, while the channel side was done using numerical integration. After modifying the model with Kane's dispersion relation, both WKB integrals are evaluated numerically.

2. Computational details

The simulation is implemented as a MATLAB program, originally created by Taur et al.¹¹⁹, which I modified as described in the previous section.

Once the tunneling current is calculated, I compute the subthreshold swing (SS). The SS is defined as

$$SS = \log(10) \left| I_{ds} \left(\frac{dI_{ds}}{dV_{gs}} \right)^{-1} \right| \quad (33)$$

Since the result of the simulation is a numerical array, the post-processing code must evaluate the derivative numerically. In the program I evaluate the SS as

$$SS = \frac{\Delta V_{ds}}{\Delta \log_{10}(I_{ds})} \quad (34)$$

where Δ means taking the difference of neighboring elements in the array. If we replace the Δ signs with differentials, then Eq. 34 is mathematically equivalent to the definition given in Eq. 33.

At this point the calculated SS is a numerical array, with each element corresponding to the mid-point values of the original array containing the gate voltage range. To find the exact gate voltage (V_{60}) at which the SS crosses 60 mV/decade, I convert it into a continuous function and interpolate using MATLAB's pchip function, which is an abbreviation for Piecewise Cubic Hermite Interpolating Polynomial. Then, to find the I_{60} , which is the current that corresponds to this gate voltage in the I-V curve, I use pchip again to turn the drain current into a continuous function of gate voltage and interpolate at V_{60} . Finally, I calculate the gate voltage swing needed to reduce the drain current by four orders of magnitude, directly below the I_{60} point, which I call V_4 . To do this, I divide I_{60} by 10000 and interpolate the reverse V-I curve with pchip to find the gate voltage at this point, then calculate V_4 as the difference between this voltage corresponding to $I_{60}/10000$ and V_{60} .

For all calculations, the following values were held constant: the thickness of the semiconductor and the insulator ($t_s = 5$ nm, $t_i = 2$ nm), the doping density of the drain ($N_d = 3 \times 10^{19} \text{cm}^{-3}$), and the dielectric constants for the semiconductor and the insulator ($\epsilon_s = 16\epsilon_0$, $\epsilon_i = 14.6\epsilon_0$). The varied parameter ranges are shown in Table 2. The ranges have been extended in certain cases, as will be discussed later.

Name	Symbol	# of Points	Values	Units
Electron effective mass	m_c^*	40	from 0.01 to 1 (log scale)	m_0
Hole effective mass	m_v^*	40	from 0.01 to 1 (log scale)	m_0
Channel length	L_g	6	10, 12, 15, 20, 40, 60	nm
Source doping density	N_a	7	$3 \times 10^{17}, 1 \times 10^{18}, 2 \times 10^{18}, 4 \times 10^{18}, 1 \times 10^{19}, 3 \times 10^{19}, 1 \times 10^{20}$	cm^{-3}
Effective band gap	V_1	50	from 0.01 to 1 (log scale)	eV
Conduction band non-parabolicity factor	α_c	13	from 0 to 3 (linear scale)	eV^{-1}
Valence band non-parabolicity factor	α_v	13	from 0 to 3 (linear scale)	eV^{-1}

B. Results and discussion

For all the plots shown in this section, any parameter value not otherwise specified corresponds to the value, within the ranges described in Table 2, that yields the maximum I_{60} for the given drain voltage V_d . These values are summarized in Table 3. Except for section 2 of this chapter, which is discussing band non-parabolicity, all figures and tables show results obtained with fully parabolic bands ($\alpha_c = \alpha_v = 0$). All I_{60} values in contour plots are displayed in units of $\mu\text{A}/\mu\text{m}$ throughout the entire chapter.

V_d	m_v^*	m_c^*	N_a	L_g	V_1	I_{60}
0.1 V	$0.49 m_0$	$0.31 m_0$	$4 \times 10^{18} \text{ cm}^{-3}$	40 nm	0.01 V	$39 \mu\text{A}/\mu\text{m}$
0.2 V	m_0	m_0	$2 \times 10^{18} \text{ cm}^{-3}$	40 nm	0.01 V	$130 \mu\text{A}/\mu\text{m}$
0.5 V	m_0	m_0	$3 \times 10^{17} \text{ cm}^{-3}$	40 nm	0.01 V	$270 \mu\text{A}/\mu\text{m}$

1. Effective mass

Figure 23 shows the I_{60} versus electron and hole effective mass at different drain biases. At $V_d = 0.1\text{V}$ the maximum I_{60} is $39 \mu\text{A}/\mu\text{m}$, located at $m_c^* = 0.31m_0$ and

$m_v^* = 0.49m_0$. At higher drain voltages, the maximum I_{60} shifts to effective masses of $m_c^* = m_v^* = m_0$ and its value increases to $130 \mu\text{A}/\mu\text{m}$ and $270 \mu\text{A}/\mu\text{m}$ for V_d values of 0.2V and 0.5V , respectively.

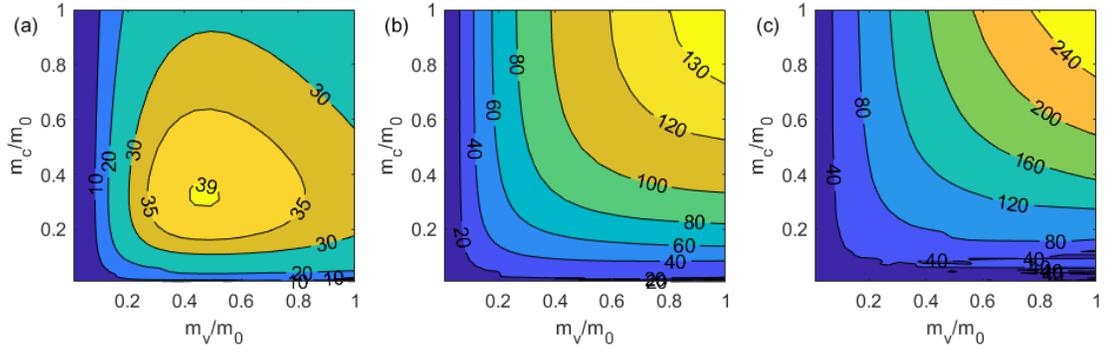


Figure 23. I_{60} Versus Electron and Hole Effective Mass. The three figures show different drain voltages: (a) $V_d = 0.1 \text{ V}$, (b) $V_d = 0.2 \text{ V}$, (c) $V_d = 0.5 \text{ V}$

Since at higher drain voltages the I_{60} reached its maximum value at the highest simulated electron and hole effective masses, I increased the range of m_v^* and m_c^* to find the actual peak I_{60} . Figure 24 shows I_{60} for drain voltages of 0.2 V and 0.5 V in an extended effective mass range. For $V_d = 0.2 \text{ V}$ the I_{60} reaches a maximum at $m_c^* = m_0$ and $m_v^* = 1.3m_0$, but its value doesn't increase appreciably above the $130 \mu\text{A}/\mu\text{m}$ seen at $m_c^* = m_v^* = m_0$ and starts to decrease for larger effective masses. For $V_d = 0.5 \text{ V}$ the I_{60} keeps rising as the electron and hole effective masses get heavier. Overall, the results show that for higher drain voltages heavier electron and hole effective masses lead to increased I_{60} .

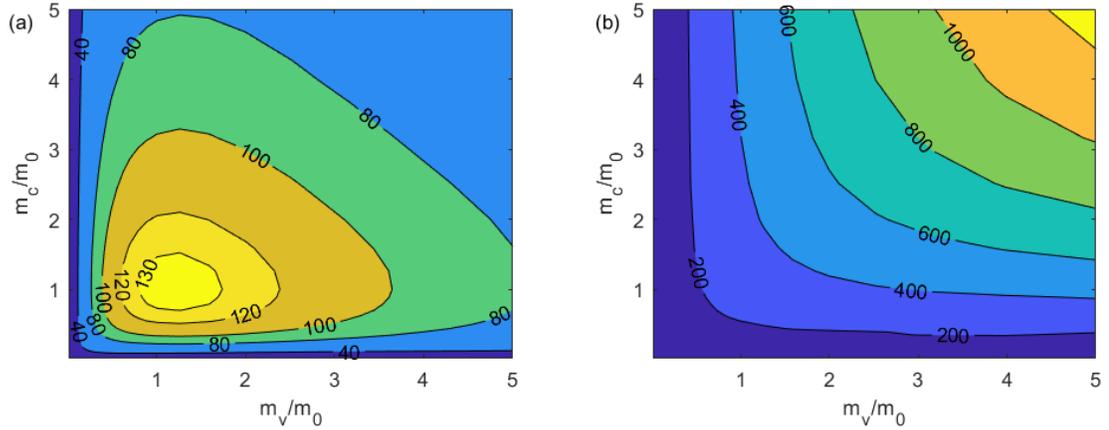


Figure 24. I_{60} at Higher Drain Voltage and Large Electron and Hole Effective Mass. Plot (a) shows $V_d = 0.2$ V while plot (b) $V_d = 0.5$ V

Electron and hole effective masses have a direct effect on the drain and source degeneracy through Equations 18 and 19. Since the degeneracies are used to determine V_2 through Eq. 16, which is used as one of the boundary conditions for solving the conduction band of the channel in Eq. 15, they have a strong effect on the overall shape of the bands in the channel. They also indirectly effect the band bending (Δ), calculated with Eq. 23. The optimal electron effective mass is a result of a tradeoff between band overlap and the potential barrier height, both determined by the amount of band bending. Figure 25 shows the calculated band structure of a DG TFET with low, optimum, and high electron effective mass. At too low m_c^* the amount of band bending increases in both the source and the channel, which allows for a larger overlap between the bands, allowing tunneling in a wider energy range, as seen from the increased thickness of the differential tunneling current peak shown with magenta dash-dotted line. However, the larger band bending also increases the height of the potential barrier in the forbidden region, which reduces tunneling thus and ultimately causes sub-optimal current. When the electron effective mass is too high, the band bending is reduced in both the source and the channel

and thus both the potential barrier height and the band overlap decreases, resulting in a tall but narrow differential tunneling current peak. The overall tunneling current is given by the area under the peak, which is maximum in the middle diagram.

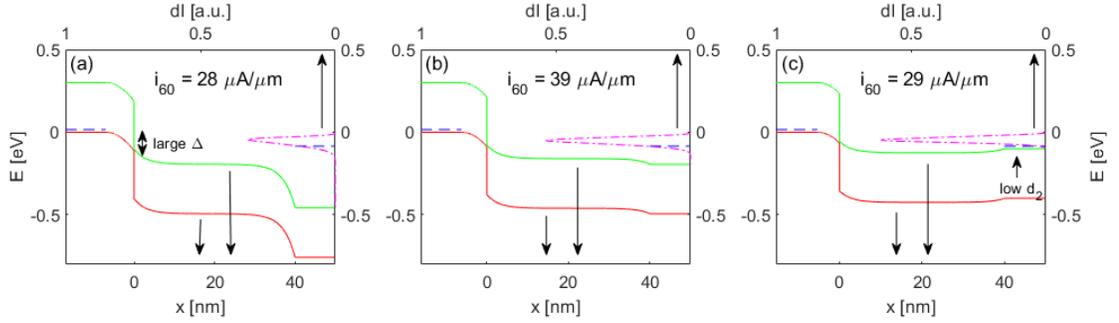


Figure 25. Effect of Electron Effective Mass on DG TFET Band Structure. The plots show (a) too low ($0.094 m_0$), (b) optimal ($0.30 m_0$), and (c) too high (m_0) electron effective mass (m_c^*). All plots show $V_d = 0.1$ V. The Fermi levels in the source and the drain are shown with blue dashed lines while the differential tunneling currents (dI) are shown with magenta dash-dotted lines.

Figure 26 shows the calculated band structure of a DG TFET with low, optimum, and high hole effective mass. The optimum tunneling current is once again determined by the amount of band bending, and the resulting opposing effects of band overlap and potential barrier height. The shifting of the Fermi level into the valence band in the source for low m_v^* also reduces the number of electrons available for tunneling. An additional effect is that m_v^* is also a prefactor for the overall tunneling current in the Landauer equation (Eq. 32), which also helps explain the reduced tunneling at lower hole effective mass.

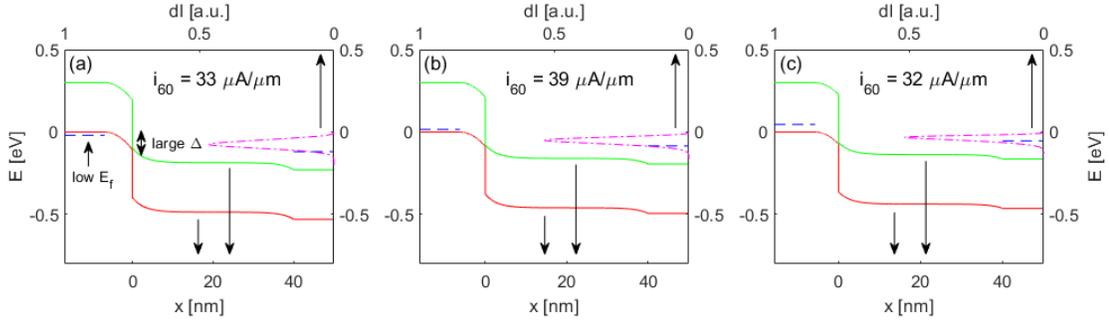


Figure 26. Effect of Hole Effective Mass on DG TFET Band Structure. The plots show (a) too low ($0.24 m_0$), (b) optimal ($0.49 m_0$), and (c) too high (m_0) hole effective mass (m_v^*). All plots show $V_d = 0.1$ V. The Fermi levels in the source and the drain are shown with blue dashed lines while the differential tunneling currents (dI) are shown with magenta dash-dotted lines.

2. Non-parabolic bands

The effect of non-parabolic bands on the I_{60} is very small, as shown in Figure 27. Across all parameter combinations, the highest I_{60} was always found at fully parabolic valence band ($\alpha_v = 0 \text{ eV}^{-1}$) and the maximum conduction band non-parabolicity factor in the simulated range ($\alpha_c = 3 \text{ eV}^{-1}$). However, the difference in I_{60} compared to fully parabolic bands is around 10% or less.

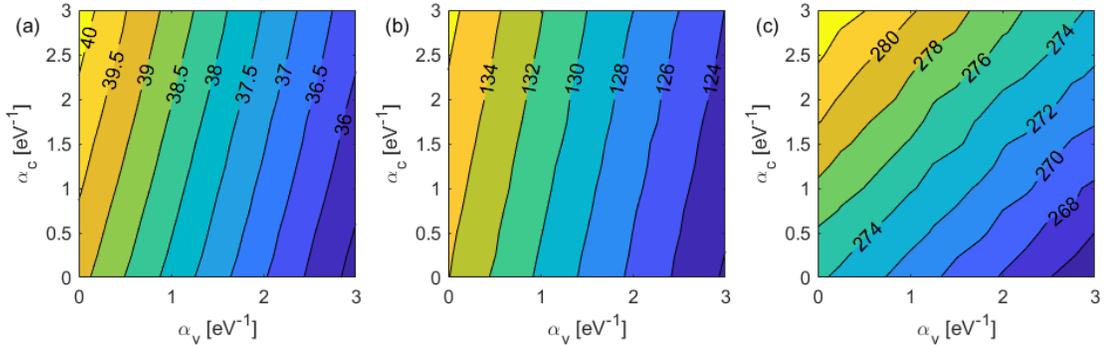


Figure 27. I_{60} Versus Non-Parabolicity Factors. The three figures show different drain voltages: (a) $V_d = 0.1$ V, (b) $V_d = 0.2$ V, (c) $V_d = 0.5$ V

Non-parabolicity has a very limited effect on the tunneling current mainly because the energy range where the majority of the tunneling happens is also very small, as shown in Figure 28. The typical energies involved in the tunneling are below 0.05 eV where the effect of non-parabolicity on the band dispersion is almost negligible.

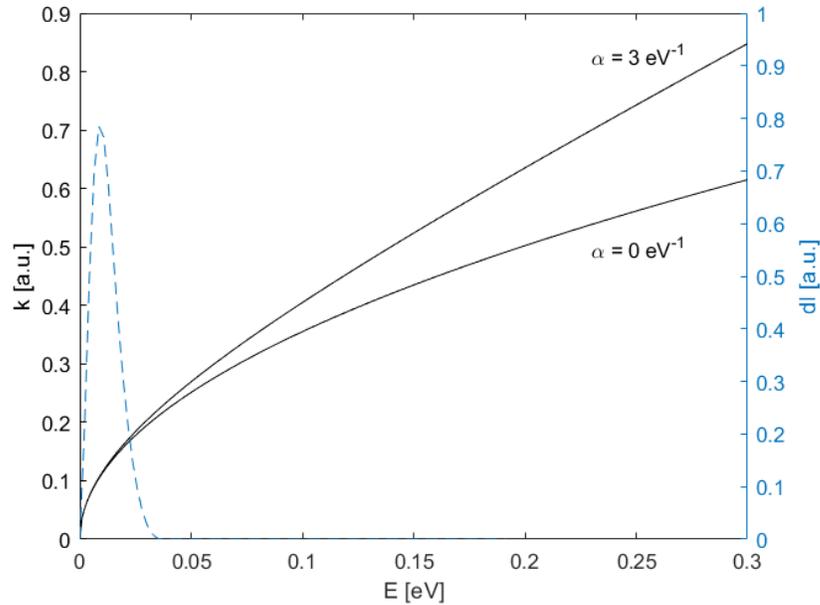


Figure 28. Effect of Band Non-Parabolicity on Tunneling. The figure shows the E - k relationship (with E as the horizontal axis) for parabolic ($\alpha = 0 \text{ eV}^{-1}$) and non-parabolic bands ($\alpha = 3 \text{ eV}^{-1}$) together with the differential tunneling current versus energy. The plot shows that at the typical energies involved in the tunneling ($E < 0.03 \text{ eV}$) the effect of non-parabolicity is negligible.

3. Channel length and source doping density

Figure 29 shows the I_{60} versus channel length and doping density at different drain biases. Independently of any other parameters, the I_{60} always decreases as L gets smaller, than 20 nm, but it's relatively constant for longer channel lengths. At $V_d = 0.1 \text{ V}$ the optimal source doping density is $4 \times 10^{18} \text{ cm}^{-3}$, but it shifts towards smaller concentrations at higher drain voltages.

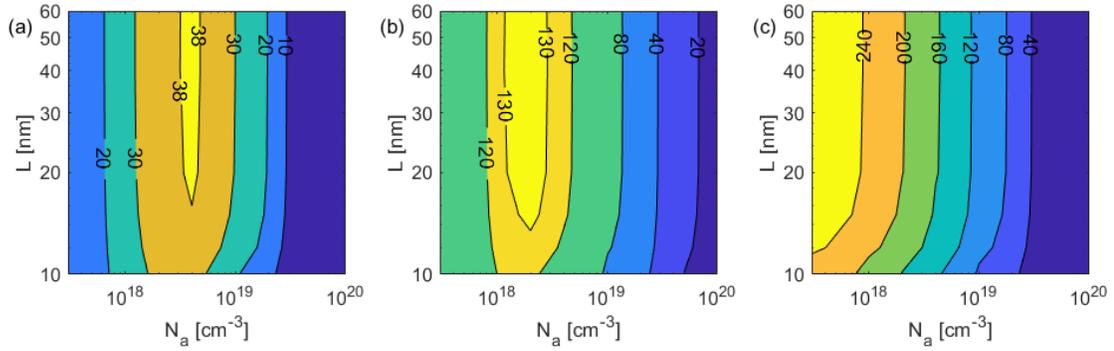


Figure 29. I_{60} Versus Source Doping Density and Channel Length. The three figures show different drain voltages: (a) $V_d = 0.1$ V, (b) $V_d = 0.2$ V, (c) $V_d = 0.5$ V

Figure 30 shows the variation of optimum doping density with electron and hole effective mass. Varying m_c^* doesn't affect the location of the maximum I_{60} , but changing m_v^* causes the optimum N_a to shift, with lighter hole effective mass values corresponding to smaller doping density.

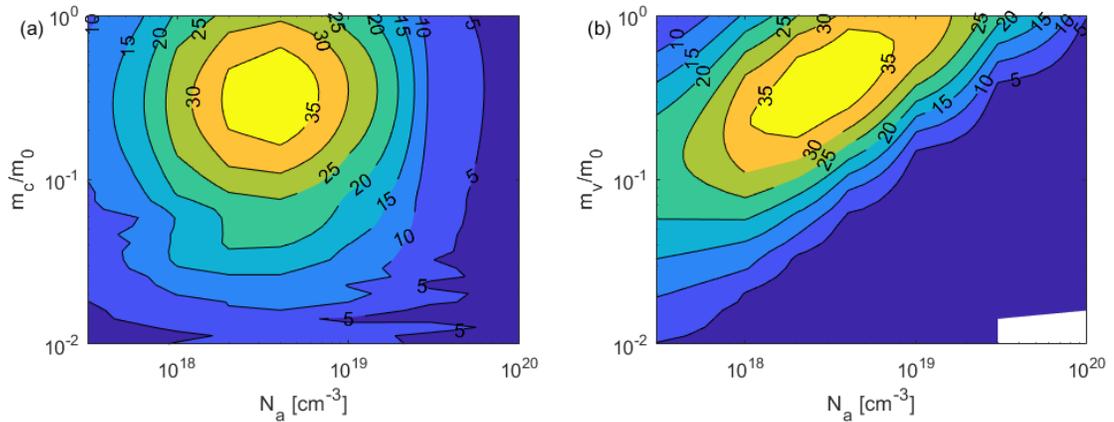


Figure 30. I_{60} Versus Source Doping Density and Electron and Hole Effective Mass. Plot (a) shows that electron effective mass does not affect the optimum doping density while plot (b) shows that the ideal value for N_a decreases for lighter hole mass. Both plots show $V_d = 0.1$ V but the overall trends are the same at higher drain voltages. The white part in plot (b) indicates that I_{60} does not exist for that parameter combination, i.e. the SS is always above 60 mV/decade.

The ideal source doping density is mainly determined by the depletion width (W_d), which is inversely proportional to N_a , as seen in Eq. 21, and the source degeneracy (d_1), calculated in Eq. 19, which determines the number of electrons available for tunneling. Figure 31 shows the calculated band structure of a DG TFET with low, optimum, and high doping. At too low N_a the depletion width increases, thus the carrier must tunnel through a longer forbidden region, which reduces tunneling probability. When the doping density is too high, W_d is reduced and the Fermi level shifts deep into the source valence band, which causes a large band bending in the source side of the channel. This shrinks the width of the forbidden region, which should allow for increased tunneling. However, the increased degeneracy in the source, caused by the low Fermi level, reduces the number of electrons available for tunneling at the band edge, which ultimately causes the tunneling current to diminish. The increased band bending in the channel increases the band overlap, and thus the energy range available for tunneling, but it also increases the height of the potential barrier. The optimum N_a is determined by the delicate balance of these opposing effects.

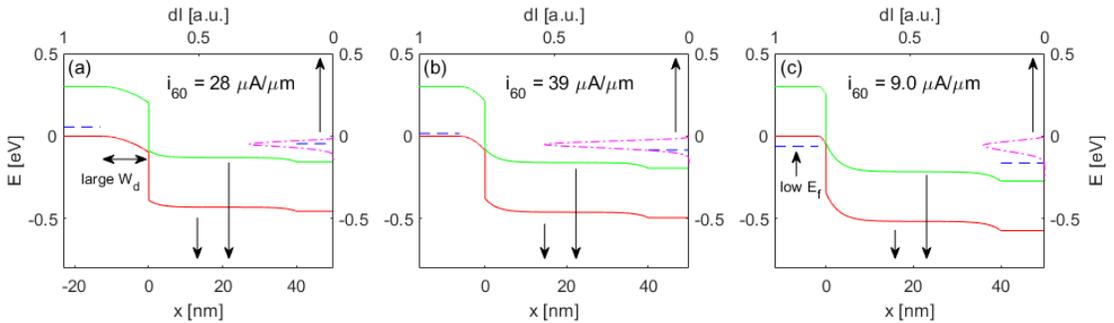


Figure 31. Effect of Source Doping Density on DG TFET Band Structure. The plots show (a) too low ($1 \times 10^{18} \text{ cm}^{-3}$), (b) optimal ($4 \times 10^{18} \text{ cm}^{-3}$), and (c) too high ($3 \times 10^{19} \text{ cm}^{-3}$) source doping densities (N_a). All plots show $V_d = 0.1 \text{ V}$. The Fermi levels in the source and the drain are shown with blue dashed lines while the differential tunneling currents (dI) are shown with magenta dash-dotted lines.

Short channel length decreases I_{60} by increasing the OFF current in the device, as shown in Figure 32. At L values at or below 15 nm, which is about twice the value of the scale length (λ), the width of the forbidden region between the valence band of the source and the bending conduction band on the drain side of the channel is short enough for carriers to tunnel through even when the device is turned off. Even though the ON current is mostly unaffected by a short channel length, the increased OFF current means that the TFET can no longer produce a SS below the 60 mV/decade limit at higher gate voltage levels. Short-channel effects of TFET devices have been investigated in more detail by the original authors of the UCSD TFET model.⁸

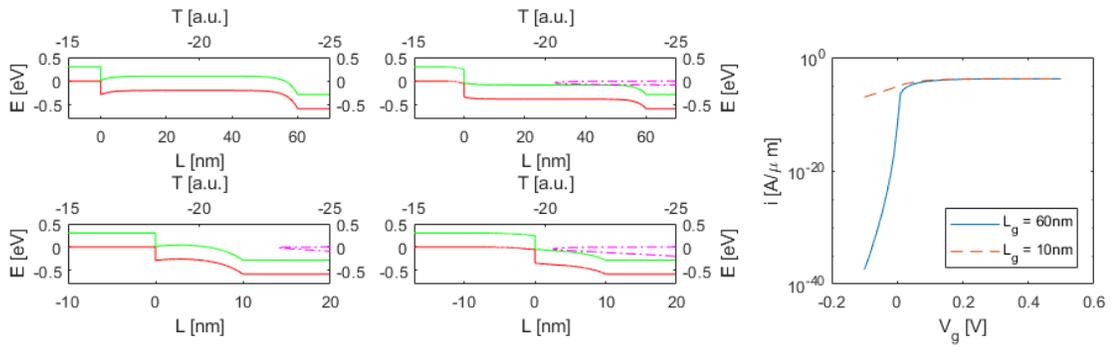


Figure 32. Short-Channel Effects. The plots show band structures and differential tunneling currents (magenta dash-dotted lines) with the device in the OFF (left) and ON (middle) state. The corresponding I - V curves are shown on the right.

4. Effective band gap

The overlap between the valence band of the source and the conduction band of the channel is most significantly affected by V_1 , i.e. the effective band gap. A near-broken band gap ($V_1 \rightarrow 0$) results in the biggest overlap and thus the highest tunneling current.

As shown in Figure 33 the lowest value of $V_1 = 0.01$ V yielded the best I_{60} , and this

trend is independent of other device parameters. This means that a heterojunction TFETs with a small V_1 are expected to perform significantly better than homojunction devices.

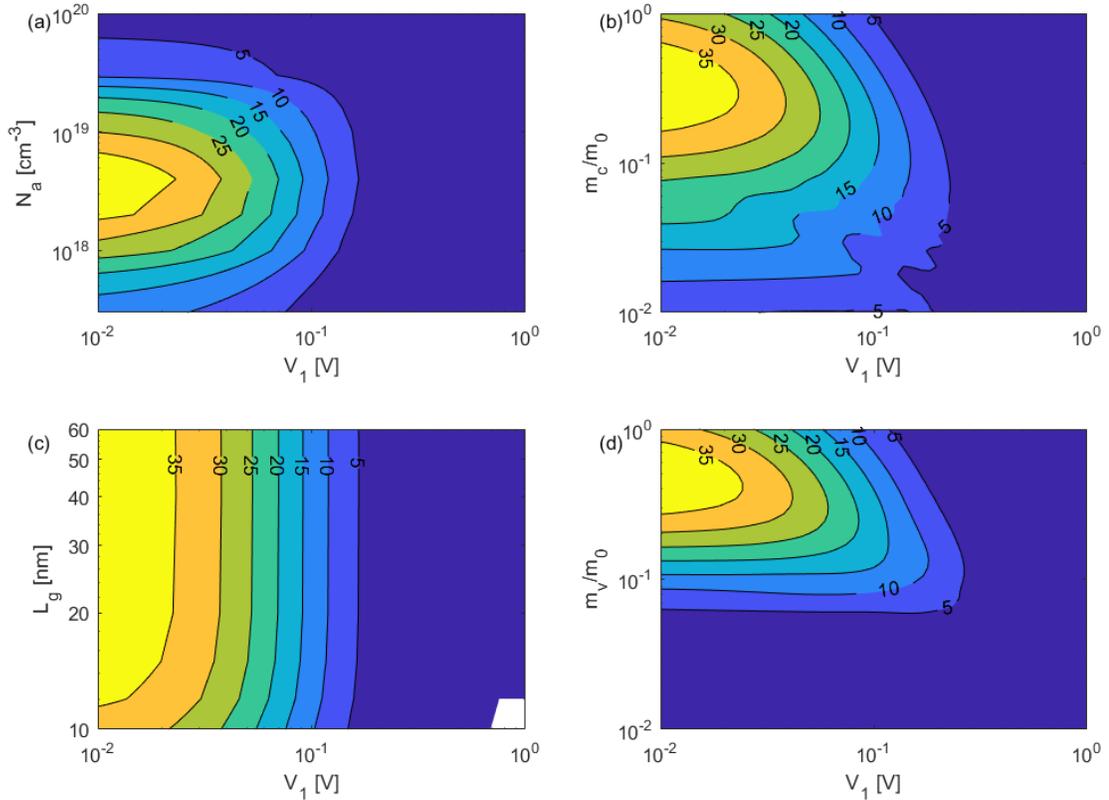


Figure 33. I_{60} Versus Effective Band Gap. The horizontal axis is V_1 in all plots, while the vertical axis is (a) source doping density, (b) electron effective mass, (c) channel length, and (d) hole effective mass. The best I_{60} always corresponds to the smallest effective band gap. All plots show $V_d = 0.1$ V but the overall trends are the same at higher drain voltages. The white part in plot (c) indicates that I_{60} does not exist for that parameter combination, i.e. the SS is always above 60 mV/decade.

5. I_{60} and V_4

I find that the highest I_{60} values generally correspond to a small V_4 value, as shown in Figure 34, meaning a low subthreshold swing is maintained in the practical

current range of the device. This reinforces that the I_{60} is a good figure of merit for TFET devices, better than the often used I_{ON}/I_{OFF} ratio, which can be misleading when the chosen ON and OFF current lies outside the range of low SS. Together with the I_{60} , which indicates the maximum practical drain current of the device (effectively I_{ON}), the V_4 shows the operating gate voltage range of the device and it indicates the steepness of the I - V curve throughout this range. The average subthreshold swing in the operating range can be easily calculated by dividing the V_4 by 4 decades. For the maximum I_{60} at $V_d = 0.1$ V, the resulting V_4 is 55 mV, which means the average SS in this range is 14 mV/decade.

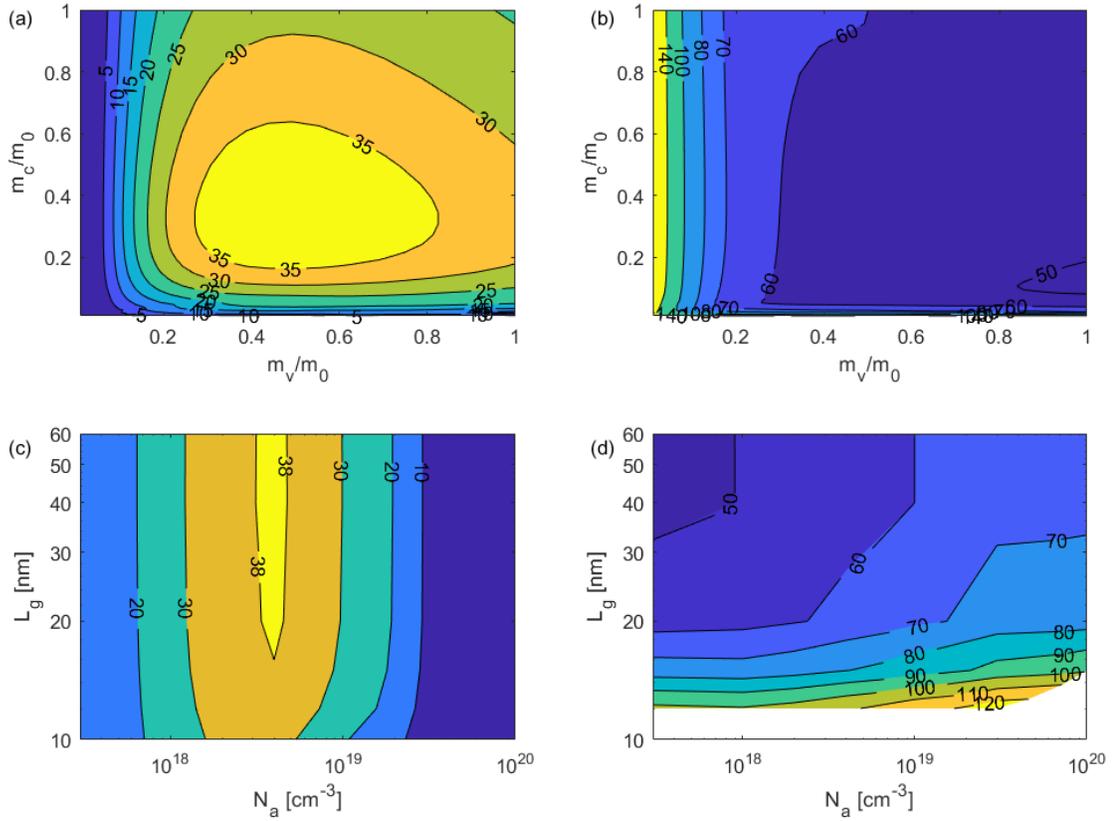


Figure 34. I_{60} and V_4 Comparison. The plots show (a) I_{60} and (b) V_4 versus electron and hole effective mass, and (c) I_{60} and (d) V_4 versus channel length and source doping density. All plots show $V_d = 0.1$ V but the overall trends are the same at higher drain voltages. The white part in plot (d) indicates that V_4 does not exist for that parameter combination, i.e. the SS does not stay below 60 mV/decade over a current range of four decades.

C. Future Work

The TFET simulations in this thesis were run with several parameters held constant. A possible follow-up study could investigate the effect of varying drain doping density and the dielectric constants and thicknesses of the semiconductor and the insulator. The simulations could be modified to incorporate more recent additions to the UCSD TFET model by Taur et al., such as using Franz’s two-band relation.^{7,123} Future

experimental work can be aimed at fabricating TFET devices using the optimum parameters presented in this work. My results show that large electron and hole effective masses are ideal for TFETs operating at higher drain voltages, which can drive future research into highly mismatched alloys with very flat band structures.

D. Summary

I used the UCSD TFET model, which I extended to use Kane's non-parabolic dispersion relation, to simulate I - V characteristics of a DG TFET device over a wide range of material and device properties. I extracted the SS from the I - V curves and identified the I_{60} and V_4 points. I found the combination of parameters that yield the highest I_{60} at three different drain voltages. I discussed that for most parameters the optimum I_{60} is determined by the band overlap and potential barrier height due to band bending in the source and the channel, as well as the position of the Fermi levels which determine the number of electrons available for tunneling. I discussed the effect of the depletion width on the optimum source doping density, and the reduction of I_{60} due to increased OFF current caused by short channel length. I showed that a nearly broken effective band gap results in the best I_{60} , and discussed V_4 as the operating gate voltage range and a measure of the average SS.

V. SUMMARY

In this thesis I investigated a specific highly mismatched alloy, BGaAs, using first principles DFT simulations with HSE06 hybrid functionals in VASP. I found excellent agreement with known values of band gap and lattice constant for the binary endpoints GaAs and BAs, as well as convergence to better than 1% with respect to number of simulation k-points and planewave energy cutoff. These validate my method and results for the intermediate alloys, which had not previously been studied in such detail. My results showed that the direct band gap in BGaAs first decreases at low B concentrations, then increases towards the large minimum direct band gap of BAs. I estimated that BGaAs transitions from direct to indirect band gap at around 17% B content. I showed that while B-B pairs at nearest-neighbor group III sites cause a narrowing of the band gap, this arrangement is somewhat unfavorable (+260 meV) and it does not significantly change electron effective mass. I found that the lattice constant of BGaAs closely follows Vegard's law. My results showed a disagreement between large and small supercells, with the small supercells showing a significant narrowing of the band gap and deviation from Vegard's law, which I attribute to the artificial periodicity introduced by the infinite repetition of the small supercell. It appears that large supercells are a better approximation of real BGaAs alloys and thus I dismissed the small-supercell results as unrealistic. This may also explain varying results in literature. Finally, I showed detailed examples of unfolding artifacts arising from the folding of the Brillouin zone and discussed methods of classifying the band gaps when the effective band structures appear to be ambiguous.

I then introduced TFETs as one possible application for highly mismatched alloys and sought to find the optimal combination of material and device properties to maximize I_{60} in TFETs. For low drain voltages ($V_d = 0.1$ V), I found the maximum $I_{60} = 39 \mu\text{A}/\mu\text{m}$ occurs at moderate effective masses for both electrons and holes. For larger drain voltages ($V_d = 0.2, 0.5$ V), I_{60} continued to increase up to at least $1.3 m_0$. Therefore, optimal TFET design depends strongly on circuit constraints such as drain voltage. Also, the increase in effective mass provided by highly mismatched alloys offers a significantly increased I_{60} over conventional group IV or group III-V semiconductors. Furthermore, in all cases, I_{60} was increased by reducing V_1 , i.e. the effective band gap or band offset from source conduction band to channel valence band, showing the necessity of a heterojunction that is nearly broken gap. Finally, I modified the UCSD TFET model to include Kane's non-parabolic dispersion relation but found only a 10% increase in I_{60} for non-parabolicity factors up to $\alpha = 3 \text{ eV}^{-1}$.

APPENDIX SECTION

This section contains the MATLAB code used for the TFET simulations presented in this work. The code was published by Taur et al. as the UCSD TFET model¹¹⁹ and was modified to include Kane's non-parabolic dispersion relation. Major modifications are highlighted with bold typeface.

```
% =====
% Original code Copyright (C) 2015 All Rights Reserved by
% ECE Department, University of California, San Diego
% La Jolla, CA 92093.
% Code modifications are Copyright (C) 2019 Mark Wistey & Istvan Gulyas
% All Rights Reserved, and are open sourced under the BSD 2-Clause License
% (FreeBSD/Simplified).
% -----
% Author      : Jianzhi Wu
% Supervisor  : Prof. Yuan Taur
% Program     : A UCSD analytic TFET model
% Date       : Nov-18-2015
% Reference  :
% [1] An analytic model for heterojunction tunnel FETs with exponential barrier, Yuan
Taur, Jianzhi Wu and Jie Min,
% IEEE Trans. Electron Devices, vol. 62, pp. 1399-1404, May 2015.
% [2] "Short channel effects in tunnel FETs" Jianzhi Wu, Jie Min and Yuan Taur,
% IEEE Trans. Electron Devices, vol. 62, pp. 3019-3024, Sept. 2015.
% [3] An analytic model for heterojunction and homojunction tunnel FETs with 3D density
of states,
% Jianzhi Wu, Jie Min, Jingwei Ji, and Yuan Taur, 73rd Device Research Conference, Ohio
State Univ. June 2015.
% [4] Analysis of source doping effect in tunnel FETs with staggered bandgap,
% Jie Min, Jianzhi Wu and Yuan Taur, IEEE Electron Device Letters, vol. 36, pp. 1094-
1096, Oct. 2015.
% =====

%% Define general parameters
q = 1.6e-19;           % [C] electron charge
h = 6.63e-34;         % [Js] Planck's constant
hbar = h/2/pi;        %
kB = 8.63e-5;         % [eV*K^-1] Boltzmann constant
T = 300;              % [K] Temperature
kT = kB*T;           % [eV]
eps0 = 8.85e-14;      % [F/cm]
eps_0 = 8.85e-12;     % [F/m]

%% Define Device related parameters
m0=9.1e-31;           % [kg] electron rest mass
mc_r = 0.31;          % relative electron eff. mass
mv_r = 0.49;          % relative hole eff. mass
mv = mv_r * m0;
mc = mc_r * m0;      % convert relative mass to absolute mass units

T_multi = -2*sqrt(2*q)*1e-9/hbar; % pre-factor for WKB integral
i_coeff = (mv*q)/(2*(pi^2)*(hbar^3)); % pre-factor for current

%% Material related parameters

eps_s = 16*eps_0;    % [F/m] GeC
eps_film = 16*eps0;  % [F/cm] GeC
t_film = 5e-7;       % [cm] body thickness
t_ox = 2e-7;         % [cm] oxide thickness
eps_ox = 14.6*eps0;  % [F/cm] oxide permittivity
```

```

Cox = eps_ox/t_ox; % [F/cm2]
lambda = (t_film + 2*t_ox) * 1e7; % [nm] Scale length

%% Initialize key parameters

VgAry = [-0.1:0.01:0.1, 0.15:0.05:0.5]; % [V] Gate bias

L = 40; % [nm] channel length
Na_cm = 4e18; % [/cm^3] source doping
Na = Na_cm*1e6; % [/m^3] source doping
Nd = 3e19*1e6; % [/m^3] drain doping
V1 = 0.01; % [eV] staggered bandgap
Vds = 0.1; % [V] drain bias
alpha_v = 0; % [/eV] VB non-parabolicity factor
alpha_c = 3; % [/eV] CB non-parabolicity factor

%% Start

Nc_film = 2.5094E19 * (mc/m0)^1.5*(T/300)^1.5; % Eff Ec DOS assuming direct

mv_np = mv_r;
mc_np = mc_r;
iAry = NaN(size(VgAry));
lim = 200;
% Replaced degeneracy calculations with Kane's non-par. dispersion rel.
d1 = -fzero(@(d1) numHoleNonParFF(0, d1, mv_np, alpha_v, T, lim) - Na_cm, 0);
d2 = fzero(@(d2) numElecNonPar(0, d2, mc_np, alpha_c, T, lim) - Nd*1e-6, 0);
V2 = Vds + d1 + d2; % conduction band of drain
Vm = V2; % integral uplimit of E
E_slices = 90;
Etr_slices = 90;
dE = abs(Vm-0)/E_slices;
EAry = 0+dE/1000000:dE:Vm-dE/1000000;
Integrand_iAry = zeros(size(EAry));
xlimAry2 = 0:L/1000:L;

% Constant prefactors moved outside the loop
debias_log = log(2/t_film*sqrt(2*eps_film*kT/q/Nc_film));
debias_pref = 2*eps_film.*t_ox/eps_ox/t_film;
V0_a = (pi/(q*lambda*1e-9));
cosh_term = cosh(pi*L/lambda);
sinh_term = sinh(pi*L/lambda);

for idxVg=1:length(VgAry)
    Vgs_ND = VgAry(idxVg); % Non-debiased Vgs

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Debias model %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    myfunBeta = @(beta) log(beta)-log(cos(beta))+...
        debias_pref.*beta.*tan(beta)-(Vgs_ND-Vds-d1)/2/kT+debias_log;
    try
        BETA = fzero(@(beta) myfunBeta(beta), [pi/1e12 pi/2-pi/1e12]);
        Qinv = eps_film*4*kT*BETA/t_film*tan(BETA);
        Vg = Vgs_ND-Qinv/Cox;
    catch
        Vg = Vgs_ND;
    end
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% end of Debias %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    V0_temp = fzero(@(V0) (V0_a.*(q*(V0*cosh_term+V2-Vg))/...
        sinh_term-sqrt(max(0, 2*Na*(q*(Vg+V1-V0))/eps_s))), 0) ;
    delta_temp = Vg - V0_temp + V1; % [eV] band bending in source
    if (delta_temp < 0)
        delta = 0 ;
        V0 = Vg + V1;
    else
        delta = delta_temp;
        V0 = V0_temp;
    end
    end
    W = (eps_s/(q*Na))*sqrt(max(0, (2*Na*q*delta/eps_s))) ; % [m] Source depletion width
    U = zeros(1, 1001);
    if W > 0

```

```

x_U = -W/W/1000:0;
U = - q * Na / (2 * eps_s) * (x_U + W).^2; % [eV] source VB
end
Vx_find2 = (V0.*sinh(pi.*(L-xlimAry2)./lambda)./sinh_term-V0+V1-delta...
            -(V2-V0+V1-delta) ...
            .*sinh(pi.*xlimAry2./lambda)./sinh_term); % [eV] with depletion
for idxE = 1:length(EAry) %For each E from src VB to drain CB
    E = EAry(idxE);
    % Transverse KE conversion using Kane's non-par. disp. rel.
    maxEtr = convertEtr(V2 - E, mc, mv, alpha_c, alpha_v);
    Etrmax = min(E, maxEtr); % integral upper limit of Etr
    dEtr = Etrmax/Etr_slices;
    EtrAry = 0:dEtr:Etrmax-dEtr/100000;
    TEtr_Ary = zeros(size(EtrAry));
    for idxEtr = 1:length(EtrAry)
        Etr=EtrAry(idxEtr);
        Etr_c = convertEtr(Etr, mv, mc, alpha_v, alpha_c);
        T_int_s = 0;
        % Replaced analytic integral for U(x) with numeric int.
        if W > 0 && E - Etr <= delta
            find_l1 = U + E - Etr;
            idxtemp_l1 = find(find_l1<0, 1, 'first');
            l1 = x_U(idxtemp_l1);
            dx_s = abs(l1)/200;
            xAry_s = (dx_s/10000):dx_s:(abs(l1)-dx_s/10000);
            Ux_Ary = q * Na / (2 * eps_s) * xAry_s.^2; % [eV] source VB
            E_diff_s = Ux_Ary - E + Etr;
            k_s_np = sqrt(max(mv .* E_diff_s .* (1 + alpha_v .* E_diff_s), 0));
            T_int_s = sum(k_s_np)*dx_s*1e9;
        end
        find_l2 = Vx_find2 + E + Etr_c; % [eV]
        idxtemp_l2 = find(find_l2<0, 1, 'first');
        l2 = xlimAry2(idxtemp_l2); % End tunneling, start real space
        dx_g = abs(l2-0)/200;
        % Calculate channel CB profile V(x):
        xAry_g = (dx_g/10000):dx_g:(l2-dx_g/10000); % [nm]
        Vx_Ary=(V0.*sinh(pi.*(L-xAry_g)./lambda)./sinh_term-V0+V1-delta...
            -(V2-V0+V1-delta).*sinh(pi.*xAry_g./lambda)./sinh_term); % [eV]
        E_diff_g = Vx_Ary + E + Etr_c;
        % Updated WKB integral with Kane's non-par. disp. rel.
        VarA_2_np = mc .* E_diff_g .* (1 + alpha_c .* E_diff_g);
        VarA_np = sqrt(abs(VarA_2_np));
        T_partEin_np = sum(VarA_np)*(dx_g);
        T_partE_np = exp(T_multi .* ...
            (T_partEin_np.*(E>=delta-V1-Etr_c) + ...
            T_int_s.*(E<=delta+Etr)));
        if (T_partE_np)
            TEtr_Ary(idxEtr) = T_partE_np;
        end
    end
end
T2D = sum(TEtr_Ary).*(q.*dEtr); % [J]
% Defining Fermi Level
Efs = d1; % [eV] Source Fermi Level
Efd = Vds + d1; % [eV] Drain Fermi Level
% Define quasi-Fermi level for source
fs = 1/(1+exp((Efs-E)/kT));
% Define quasi-Fermi level for drain
fd = 1/(1+exp((Efd-E)/kT));
% Calculate each element of the integral with a given E
Integrand_iAry(idxE) = (fs-fd).*T2D;
end %idxE
%calculate the integral
i = i_coeff.*sum(Integrand_iAry)*(dE*q);
iAry(idxVg) = i*5e-15; % unit of A/um, with body thickness of 5 nm
end %Vgs

%I 60 and V 4 calculation
VgAry_diff = 1000 * diff(VgAry);
VgAry_mid = VgAry(1:length(VgAry) - 1) + diff(VgAry) / 2;
iOFF = min(iAry);
current_diff = diff(log10(iAry));

```

```

current_diff(current_diff < 0) = 0;
SS = VgAry_diff ./ current_diff;
[minSS, minidx] = min(SS);
if minSS > 60 || isnan(minSS) || minidx >= length(SS) || minSS < 0
    i60 = NaN;
    V60 = NaN;
    V4 = NaN;
else
    span = (minidx:length(SS));
    V60 = pchip(SS(span), VgAry_mid(span), 60);
    i60 = pchip(VgAry, iAry, V60);
    i60b4 = i60 / 1e4;
    if i60b4 > iOFF
        norepeat = iAry ~= iOFF;
        V4ip = pchip(iAry(norepeat), VgAry(norepeat), i60b4);
        V4 = V60 - V4ip;
        if V4 < 0
            V4 = NaN;
        end
    else
        V4 = NaN;
    end
end

semilogy(VgAry, iAry);
xlabel('V_g');
ylabel('I_d');

fprintf('I_60: %f uA/um, V_4: %f V\n', i60 * 1e6, V4*1e3);

% Calculate total number of electrons
function n = numElecNonPar(Ec, Ef, me, alpha, T, lim)
    k = 8.63e-5; % [eV/K] Boltzmann constant
    Etop = Ec + lim * k * T;
    E = linspace(Ec, Etop, 10000);
    n = trapz(E, electronDensityNonPar(E, Ec, Ef, me, alpha, T));
end

% Hole density calculation
function n_E = electronDensityNonPar(E, Ec, Ef, me, alpha, T)
    n_E = dosCNonPar(E, Ec, me, alpha) .* fermi(E, Ef, T);
end

% CB DOS calculation
function g = dosCNonPar(E, Ec, me, alpha)
    g = dosNonPar(E - Ec, me, alpha);
end

% Fermi distribution
function f = fermi(E, Ef, T)
    k = 8.63e-5; % [eV/K] Boltzmann constant
    f = 1 ./ (1 + exp((E - Ef) ./ (k * T)));
end

% Calculate total number of holes
function p = numHoleNonParFF(Ev, Ef, mh, alpha, T, lim)
    k = 8.63e-5; % [eV/K] Boltzmann constant
    Ebottom = Ev - lim * k * T;
    E = linspace(Ebottom, Ev, 10000);
    p = trapz(E, holeDensityNonParFF(E, Ev, Ef, mh, alpha, T));
end

% Hole density calculation
function p_E = holeDensityNonParFF(E, Ev, Ef, mh, alpha, T)
    p_E = dosVNonPar(E, Ev, mh, alpha) .* fermiFlip(E, Ef, T);
end

% VB DOS calculation
function g = dosVNonPar(E, Ev, mh, alpha)
    g = dosNonPar(Ev - E, mh, alpha);
end

```

```

% Equivalent to [1-f] but avoids floating point errors
function ff = fermiFlip(E, Ef, T)
    k = 8.63e-5; % [eV/K] Boltzmann constant
    ff = 1 ./ (1 + exp((Ef - E) ./ (k * T)));
end

% Generic DOS calculation using Kane's non-parabolic dispersion relation
function g = dosNonPar(E, mr, alpha)
    h_bar = 1.054571800e-34; % [J*s] reduced Plack constant
    m0 = 9.1e-31; % [kg] electron rest mass
    q = 1.6e-19; % [C] electron charge
    E_J = E * q; % convert [eV] to [J]
    m = mr * m0; % convert relative mass to [kg]
    g = real(m * (1 + 2 * alpha * E) ...
        .* sqrt(2 * m * E_J .* (1 + alpha * E)) ...
        / (pi^2 * h_bar^3));
    g(g<0) = 0;
    g = g * q * 1e-6; % convert [1/J/m^3] to [1/eV/cm^3]
end

% Transverse kinetic energy conversion using conservation of momentum
function toEtr = convertEtr(fromEtr, fromMass, toMass, fromAlpha, toAlpha)
    if toAlpha == 0
        toEtr = (fromMass / toMass) * fromEtr * (1 + fromAlpha * fromEtr);
    else
        toEtr = (-sqrt(toMass) + ...
            sqrt(4 * fromAlpha * toAlpha * fromMass * fromEtr^2 ...
                + 4 * toAlpha * fromMass * fromEtr + toMass)) / ...
            (2 * toAlpha * sqrt(toMass));
    end
end
end

```

REFERENCES

- ¹ G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
- ² G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
- ³ A. V Krukau, O.A. Vydrov, A.F. Izmaylov, and G.E. Scuseria, J. Chem. Phys. **125**, 224106 (2006).
- ⁴ J. Heyd, G.E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003).
- ⁵ J. Heyd, G.E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **124**, 219906 (2006).
- ⁶ A.J. Garza and G.E. Scuseria, J. Phys. Chem. Lett **7**, (2016).
- ⁷ J. Wu and Y. Taur, IEEE Trans. Electron Devices **63**, 841 (2016).
- ⁸ J. Wu, J. Min, and Y. Taur, IEEE Trans. Electron Devices **62**, 3019 (2015).
- ⁹ E.O. Kane, J. Phys. Chem. Solids **1**, 249 (1957).
- ¹⁰ E.M. Conwell and M.O. Vassell, Phys. Rev. **166**, 797 (1968).
- ¹¹ W.G. Vandenberghe, A.S. Verhulst, B. Sorée, W. Magnus, G. Groeseneken, Q. Smets, M. Heyns, and M. V Fischetti, Appl. Phys. Lett. **102**, 013510 (2013).
- ¹² L. Nordheim, Ann. Phys. **401**, 607 (1931).
- ¹³ D. Richardson, J. Phys. C Solid State Phys. **4**, L289 (1971).
- ¹⁴ J.A. Van Vechten and T.K. Bergstresser, Phys. Rev. B **1**, 3351 (1970).
- ¹⁵ W. Shan, W. Walukiewicz, J.W. Ager, E.E. Haller, J.F. Geisz, D.J. Friedman, J.M. Olson, and S.R. Kurtz, Phys. Rev. Lett. **82**, 1221 (1999).
- ¹⁶ J. Wu, W. Shan, and W. Walukiewicz, Semicond. Sci. Technol. **17**, (2002).
- ¹⁷ W. Walukiewicz, W. Shan, K.M. Yu, J.W. Ager, E.E. Haller, I. Miotkowski, M.J. Seong, H. Alawadhi, and A.K. Ramdas, Phys. Rev. Lett. **85**, 1552 (2000).

- ¹⁸ W. Shan, W. Walukiewicz, J. Wu, K.M. Yu, J.W. Ager, S.X. Li, E.E. Haller, J.F. Geisz, D.J. Friedman, and S.R. Kurtz, *J. Appl. Phys.* **93**, 2696 (2003).
- ¹⁹ L. Vegard, *Zeitschrift Für Phys.* **5**, 17 (1921).
- ²⁰ S.T. Murphy, A. Chroneos, C. Jiang, U. Schwingenschlögl, and R.W. Grimes, *Phys. Rev. B* **82**, 073201 (2010).
- ²¹ C. Xu, C.L. Senaratne, R.J. Culbertson, J. Kouvetakis, and J. Menéndez, *J. Appl. Phys.* **122**, 125702 (2017).
- ²² Z.R. Wasilewski, M.M. Dion, D.J. Lockwood, P. Poole, R.W. Streater, and A.J. SpringThorpe, *J. Appl. Phys.* **81**, 1683 (1997).
- ²³ M. Weyers, M. Sato, and H. Ando, *Jpn. J. Appl. Phys.* **31**, L853 (1992).
- ²⁴ K. Uesugi, N. Morooka, and I. Suemune, *Appl. Phys. Lett.* **74**, 1254 (1999).
- ²⁵ C. Skierbiszewski, P. Perlin, P. Wisniewski, W. Knap, T. Suski, W. Walukiewicz, W. Shan, K.M. Yu, J.W. Ager, E.E. Haller, J.F. Geisz, and J.M. Olson, *Appl. Phys. Lett.* **76**, 2409 (2000).
- ²⁶ K.M. Yu, S. V. Novikov, R. Broesler, A.X. Levander, Z. Liliental-Weber, F. Luckert, R.W. Martin, O. Dubon, J. Wu, W. Walukiewicz, and C.T. Foxon, *Phys. Status Solidi Curr. Top. Solid State Phys.* **8**, 2503 (2011).
- ²⁷ R. Kuroiwa, H. Asahi, K. Asami, S.-J. Kim, K. Iwata, and S. Gonda, *Appl. Phys. Lett.* **73**, 2630 (1998).
- ²⁸ G.L.W. Hart and A. Zunger, *Phys. Rev. B* **62**, 13522 (2000).
- ²⁹ T.E. Schlesinger, in *Encycl. Mater. Sci. Technol.* (Elsevier, 2001), pp. 3431–3435.
- ³⁰ S.M. Sze and K.K. Ng, *Physics of Semiconductor Devices: Third Edition* (2006).

- ³¹ www.altadevices.com/solar-world-record-nasa-selects-alta-devices, Alta Devices, Inc. (2018).
- ³² D.J. Stukel, *Phys. Rev. B* **1**, 3458 (1970).
- ³³ F. Tian, B. Song, X. Chen, N.K. Ravichandran, Y. Lv, K. Chen, S. Sullivan, J. Kim, Y. Zhou, T.-H. Liu, M. Goni, Z. Ding, J. Sun, G.A.G. Udalamatta Gamage, H. Sun, H. Ziyacee, S. Huyan, L. Deng, J. Zhou, A.J. Schmidt, S. Chen, C.-W. Chu, P.Y. Huang, D. Broido, L. Shi, G. Chen, and Z. Ren, *Science* (80-.). **361**, 582 (2018).
- ³⁴ L. Lindsay, D.A. Broido, and T.L. Reinecke, *Phys. Rev. Lett.* **111**, 025901 (2013).
- ³⁵ S. Li, Q. Zheng, Y. Lv, X. Liu, X. Wang, P.Y. Huang, D.G. Cahill, and B. Lv, *Science* **361**, 579 (2018).
- ³⁶ J.S. Kang, M. Li, H. Wu, H. Nguyen, and Y. Hu, *Science* (80-.). **361**, 575 (2018).
- ³⁷ C. Dames, *Science* (80-.). **361**, 549 (2018).
- ³⁸ J.A. Perri, S. La Placa, and B. Post, *Acta Crystallogr.* **11**, 310 (1958).
- ³⁹ S.M. Ku, *J. Electrochem. Soc.* **113**, 813 (1966).
- ⁴⁰ F. V Williams and R.A. Ruehrwein, *J. Am. Chem. Soc.* **82**, 1330 (1960).
- ⁴¹ J.W. Pomeroy, M. Kuball, H. Hubel, N.W.A. van Uden, D.J. Dunstan, R. Nagarajan, and J.H. Edgar, *J. Appl. Phys.* **96**, 910 (2004).
- ⁴² T.L. Chu and A.E. Hyslop, *J. Electrochem. Soc.* **121**, 412 (1974).
- ⁴³ S. Wang, S.F. Swingle, H. Ye, F.R.F. Fan, A.H. Cowley, and A.J. Bard, *J. Am. Chem. Soc.* **134**, 11056 (2012).
- ⁴⁴ J.L. Lyons, J.B. Varley, E.R. Glaser, J.A. Freitas, J.C. Culbertson, F. Tian, G.A. Gamage, H. Sun, H. Ziyacee, and Z. Ren, *Appl. Phys. Lett.* **113**, 251902 (2018).

- ⁴⁵ J.F. Geisz, D.J. Friedman, J.M. Olson, S.R. Kurtz, R.C. Reedy, A.B. Swartzlander, B.M. Keyes, and A.G. Norman, *Appl. Phys. Lett.* **76**, 1443 (2000).
- ⁴⁶ V. Gottschalch, G. Leibiger, and G. Benndorf, *J. Cryst. Growth* **248**, 468 (2002).
- ⁴⁷ V.K. Gupta, M.W. Koch, N.J. Watkins, Y. Gao, and G.W. Wicks, *J. Electron. Mater.* **29**, 1387 (2000).
- ⁴⁸ A. Lindsay and E.P. O'Reilly, *Phys. Status Solidi* **5**, 454 (2008).
- ⁴⁹ S. Ilahi, F. Saidi, R. Hamila, N. Yacoubi, H. Maaref, and L. Auvray, *Phys. B Condens. Matter* **421**, 105 (2013).
- ⁵⁰ R. Hamila, F. Saidi, A. Fouzri, L. Auvray, Y. Monteil, and H. Maaref, *J. Lumin.* **129**, 1010 (2009).
- ⁵¹ F. Saidi, F. Hassen, H. Maaref, H. Dumont, and Y. Monteil, *Mater. Sci. Eng. C* **26**, 236 (2006).
- ⁵² M. Guemou, B. Bouhafis, A. Abdiche, R. Khenata, Y. Al Douri, and S. Bin Omran, *Phys. B Condens. Matter* **407**, 1292 (2012).
- ⁵³ P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- ⁵⁴ W. Kohn and L.J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- ⁵⁵ www.nobelprize.org/prizes/chemistry/1998/kohn/facts/, Nobel Media AB (2019).
- ⁵⁶ J.P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- ⁵⁷ P. Mori-Sánchez, A.J. Cohen, and W. Yang, (2008).
- ⁵⁸ V. Fock, *Zeitschrift Für Phys.* **62**, 795 (1930).
- ⁵⁹ D.R. Hartree and W. Hartree, *Proc. R. Soc. London. Ser. A - Math. Phys. Sci.* **150**, 9 (1935).
- ⁶⁰ A.D. Becke, *Phys. Rev. A* **38**, 3098 (1988).

- ⁶¹ C. Lee, W. Yang, and R.G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- ⁶² J.P. Perdew, M. Ernzerhof, and K. Burke, *J. Chem. Phys.* **105**, 9982 (1996).
- ⁶³ J.P. Perdew, J. Tao, V.N. Staroverov, and G.E. Scuseria, *J. Chem. Phys.* **120**, 6898 (2004).
- ⁶⁴ Y. Zhao, D.G. Truhlar, Y. Zhao, and D.G. Truhlar, *Theor Chem Acc.* **120**, 215 (2008).
- ⁶⁵ B.G. Janesko, T.M. Henderson, and G.E. Scuseria, *Phys. Chem. Chem. Phys.* **11**, 443 (2009).
- ⁶⁶ O.K. Andersen, *Phys. Rev. B* **12**, 3060 (1975).
- ⁶⁷ S.F. Boys, *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **200**, 542 (1950).
- ⁶⁸ G. Kresse and D. Joubert, *Phys. Rev. B* **59**, 1758 (1999).
- ⁶⁹ D. Vanderbilt, *Phys. Rev. B* **41**, 7892 (1990).
- ⁷⁰ P.E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994).
- ⁷¹ J.P. Holdren, J. Morrow, T. Kalil, R. Paris, C. Wadia, H. Kung, L. Locascio, L. Horton, J. Warren, M. Drosback, J. Christodoulou, M. Galvin, D. Hardy, L. Horton, W. Joost, B. Morreale, H. Partridge, and C. Ward, *Materials Genome Initiative Strategic Plan* (2014).
- ⁷² A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K.A. Persson, *APL Mater.* **1**, 011002 (2013).
- ⁷³ S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnatek, R. V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M.J. Mehl, H.T. Stokes, D.O. Demchenko, and D. Morgan, (2012).
- ⁷⁴ K. Momma and F. Izumi, *J. Appl. Crystallogr.* **41**, 653 (2008).
- ⁷⁵ S. Ping Ong, W. Davidson Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, and G. Ceder, *Comput. Mater. Sci.* **68**, 314 (2013).

- ⁷⁶ L.L. Chang and L. Esaki, *Appl. Phys. Lett.* **31**, 687 (1977).
- ⁷⁷ H. Lu and A. Seabaugh, *IEEE J. Electron Devices Soc.* **2**, 44 (2014).
- ⁷⁸ A.M. Ionescu and H. Riel, *Nature* **479**, 329 (2011).
- ⁷⁹ A.S. Verhulst, B. Sorée, D. Leonelli, W.G. Vandenberghe, and G. Groeseneken, *J. Appl. Phys.* **107**, 024518 (2010).
- ⁸⁰ K. Boucart, A.M. Ionescu, and W. Riess, *IEEE Electron Device Lett.* **30**, 656 (2009).
- ⁸¹ F. Mayer, C. Le Royer, J.F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, *Tech. Dig. - Int. Electron Devices Meet. IEDM* **4**, 1 (2008).
- ⁸² H. Zhao, Y. Chen, Y. Wang, F. Zhou, F. Xue, and J. Lee, *IEEE Trans. Electron Devices* **58**, 2990 (2011).
- ⁸³ L. Wang, E. Yu, Y. Taur, and P. Asbeck, *IEEE Electron Device Lett.* **31**, 431 (2010).
- ⁸⁴ J. Appenzeller, Y.-M. Lin, J. Knoch, and P. Avouris, *Phys. Rev. Lett.* **93**, 196805 (2004).
- ⁸⁵ G. Fiori and G. Iannaccone, *IEEE Electron Device Lett.* **30**, 1096 (2009).
- ⁸⁶ P.V.C. Medeiros, S. Stafström, and J. Björk, *Phys. Rev. B - Condens. Matter Mater. Phys.* **89**, (2014).
- ⁸⁷ P.V.C. Medeiros, S.S. Tsirkin, S. Stafström, and J. Björk, *Phys. Rev. B - Condens. Matter Mater. Phys.* **91**, 1 (2015).
- ⁸⁸ B. Liu, *Numer. Algorithms Chem. Algebr. Methods*, Tech. Rep. LBL-8158, Lawrence Berkeley Lab. 49 (1978).
- ⁸⁹ E.R. Davidson, *J. Comput. Phys.* **17**, 87 (1975).
- ⁹⁰ P. Pulay, *Chem. Phys. Lett.* **73**, 393 (1980).
- ⁹¹ D.M. Wood and A. Zunger, *J. Phys. A. Math. Gen.* **18**, 1343 (1985).

- ⁹² I. Štich, R. Car, M. Parrinello, and S. Baroni, *Phys. Rev. B* **39**, 4997 (1989).
- ⁹³ M.J. Gillan, *J. Phys. Condens. Matter* **1**, 689 (1989).
- ⁹⁴ T.A. Arias, M.C. Payne, and J.D. Joannopoulos, *Phys. Rev. Lett.* **69**, 1077 (1992).
- ⁹⁵ W.T. Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, *Numerical Recipes* (Cambridge University Press, New York, 1986).
- ⁹⁶ M. Tomić, H.O. Jeschke, and R. Valentí, *Phys. Rev. B* **90**, 195121 (2014).
- ⁹⁷ O. Rubel, A. Bokhanchuk, S.J. Ahmed, and E. Assmann, *Phys. Rev. B* **90**, 115202 (2014).
- ⁹⁸ V. Popescu and A. Zunger, *Phys. Rev. Lett.* **104**, 236403 (2010).
- ⁹⁹ J.J. Mortensen, L.B. Hansen, and K.W. Jacobsen, *Phys. Rev. B* **71**, 035109 (2005).
- ¹⁰⁰ J. Enkovaara, C. Rostgaard, J.J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H.A. Hansen, H.H. Kristoffersen, M. Kuisma, A.H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P.G. Moses, J. Ojanen, T. Olsen, V. Petzold, N.A. Romero, J. Stausholm-Møller, M. Strange, G.A. Tritsarlis, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G.K.H. Madsen, R.M. Nieminen, J.K. Nørskov, M. Puska, T.T. Rantala, J. Schiøtz, K.S. Thygesen, and K.W. Jacobsen, *J. Phys. Condens. Matter* **22**, 253202 (2010).
- ¹⁰¹ G.P. Francis and M.C. Payne, *J. Phys. Condens. Matter* **2**, 4395 (1990).
- ¹⁰² S. Van Der Walt, S.C. Colbert, and G. Varoquaux, *Comput. Sci. Eng.* **13**, 22 (2011).
- ¹⁰³ W. McKinney, *Data Structures for Statistical Computing in Python* (2010).
- ¹⁰⁴ J.D. Hunter, *Comput. Sci. Eng.* **9**, 99 (2007).
- ¹⁰⁵ M.P. Surh, S.G. Louie, and M.L. Cohen, *Phys. Rev. B* **43**, 9126 (1991).
- ¹⁰⁶ K. Bushick, K. Mengle, N. Sanders, and E. Kioupakis, *Appl. Phys. Lett.* **114**, (2019).

- ¹⁰⁷ H. Meradji, S. Drablia, S. Ghemid, H. Belkhir, B. Bouhafis, and A. Tadjer, *Phys. Status Solidi* **241**, 2881 (2004).
- ¹⁰⁸ R.M. Wentzcovitch and M.L. Cohen, *J. Phys. C Solid State Phys.* **19**, 6791 (1986).
- ¹⁰⁹ N.N. Anua, R. Ahmed, A. Shaari, M.A. Saeed, B. Ul Haq, and S. Goumri-Said, *Semicond. Sci. Technol.* **28**, 105015 (2013).
- ¹¹⁰ A. Zaoui and F.E.H. Hassan, *J. Phys. Condens. Matter* **13**, 253 (2001).
- ¹¹¹ S. Cui, W. Feng, H. Hu, Z. Feng, and Y. Wang, *Comput. Mater. Sci.* **44**, 1386 (2009).
- ¹¹² A. Boudjemline, M.M. Islam, L. Louail, and B. Diawara, *Phys. B Phys. Condens. Matter* **406**, 4272 (2011).
- ¹¹³ M. Marsman, J. Paier, A. Stroppa, and G. Kresse, *J. Phys. Condens. Matter* **20**, 064201 (2008).
- ¹¹⁴ Y. Ge, W. Wan, X. Guo, and Y. Liu, *The Direct and Indirect Optical Absorptions of Cubic BAs and BSb* (2019).
- ¹¹⁵ J. Buckeridge and D.O. Scanlon, *Phys. Rev. Mater.* **3**, (2019).
- ¹¹⁶ S. Chae, K. Mengle, J.T. Heron, and E. Kioupakis, *Appl. Phys. Lett.* **113**, 212101 (2018).
- ¹¹⁷ I. Bravić and B. Monserrat, *Finite Temperature Optoelectronic Properties of BAs from First Principles* (2019).
- ¹¹⁸ W. Nakwaski, *Effective Masses of Electrons and Heavy Holes in GaAs, InAs, AlAs and Their Ternary Compounds* (1995).
- ¹¹⁹ W. Jianzhi and T. Yuan, *A UCSD Analytic TFET Model* (2015).
- ¹²⁰ J. Wu, J. Min, J. Ji, and Y. Taur, in *2015 73rd Annu. Device Res. Conf.* (IEEE, 2015), pp. 249–250.

- ¹²¹ Y. Taur, Jianzhi Wu, and Jie Min, *IEEE Trans. Electron Devices* **62**, 1399 (2015).
- ¹²² J. Min, J. Wu, and Y. Taur, *IEEE Electron Device Lett.* **36**, 1094 (2015).
- ¹²³ Y. Taur and J. Wu, *IEEE Trans. Electron Devices* **63**, 869 (2016).
- ¹²⁴ X. Liang and Y. Taur, *IEEE Trans. Electron Devices* **51**, 1385 (2004).
- ¹²⁵ P.M. Solomon, D.J. Frank, and S.O. Koswatta, in *69th Device Res. Conf.* (IEEE, 2011), pp. 197–198.
- ¹²⁶ B. Rajamohanam, D. Mohata, A. Ali, and S. Datta, *Appl. Phys. Lett.* **102**, 092105 (2013).
- ¹²⁷ Y. Taur, X. Liang, W. Wang, and H. Lu, *IEEE Electron Device Lett.* **25**, 107 (2004).
- ¹²⁸ E.O. Kane, *J. Appl. Phys.* **32**, 83 (1961).