NEW METHODS, TOOLS AND TECHNIQUES FOR THE EVOLUTIONARY DETERMINATION OF ENSEMBLES OF PARETO EFFICIENT PARAMETERIZATIONS OF CLASSICAL POTENTIALS WITH APPLICATIONS

By

EUGENE RAGASA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2016

I dedicate this to everyone that helped revamp this template. Aliquam molestie sed urna quis convallis. Aenean nibh eros, aliquam non eros in, tempus lacinia justo. In magna sapien, blandit a faucibus ac, scelerisque nec purus. Praesent fermentum felis nec massa interdum, vel dapibus mi luctus. Cras id fringilla mauris. Ut molestie eros mi, ut hendrerit nulla tempor et. Pellentesque tortor quam, mattis a scelerisque nec, euismod et odio. Mauris rhoncus metus sit amet risus mattis, eu mattis sem interdum.

ACKNOWLEDGMENTS

Thanks to all the help I have received in writing and learning about this tutorial. Acknowledgments are required and must be written in paragraph form. This mandates at least three sentences.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS 4

LIST OF TABLES 8

LIST OF FIGURES 9

ABSTRACT 10

CHAPTER

1 INTRODUCTION AND OPENING REMARKS 12

- 1.1 The Section Command Text Should Be in Title Case 12
  - 1.1.1 Subsection Commands Are Also in Title Case 12
    - 1.1.1.1 Subsubsections are in sentence case 12
    - 1.1.1.2 If you divide a section, you must divide it into two, or more, parts 12
  - 1.1.2 I Need Another Second Level Heading in This Section 12
- 1.2 Image Handling in XeLaTeX 13
  - 1.2.1 The Traditional EPS Format 13
  - 1.2.2 Bitmapped Images Work As Well 13
  - 1.2.3 Not to Mention PDF 15
  - 1.2.4 JPG Is Absolutely Necessary 16
  - 1.2.5 PNGs Will Help Make Files Smaller 16
- 1.3 GIF, TIF, and Others 17

2 THE MANY BODY PROBLEM AND ATOMISTIC METHODS 20

- 2.1 The Many-Body Problem 20
- 2.2 Density Functional Theory 21
  - 2.2.1 The Exchange Correlation Term 21
- 2.3 Density Functional Theory for Solids 21
  - 2.3.1 Representation of an Infinite Solid 21
  - 2.3.2 Bloch's Theorem 21
  - 2.3.3 Plane Wave Formulation 21
  - 2.3.4 k-point sampling 21
- 2.4 Empirical Interatomic Potentials 21
  - 2.4.1 Pair Potentials 22
    - 2.4.1.1 Leonnard Jones 22
    - 2.4.1.2 Coulomb Potential 22
    - 2.4.1.3 Born-Mayer Potential 22
    - 2.4.1.4 Morse Potential 22
    - 2.4.1.5 Buckingham Potential 22
  - 2.4.2 Many Body Potentials 22

2.4.2.1 Tersoff Potentials . . . 22
2.4.2.2 Stillinger Weber Potentials . . . 22
2.4.2.3 Embedded Atom Method Potentials . . . 22
2.5 Interatomic Potentials . . . 23
2.6 Molecular Dynamics . . . 24
2.6.1 Role of the Interatomic Potential . . . 24
2.6.2 Numerical Integration . . . 24
2.6.3 Thermodynamic Ensembles . . . 25
2.7 Lattice Dynamics . . . 25
2.8 Calculation of Material Properties . . . 25
2.8.1 Structural Properties . . . 25
2.8.1.1 Minimization Techniques . . . 25
2.8.2 Greatest Descent . . . 25
2.8.3 Conjugate Gradient . . . 25
2.8.4 Phase Order Properties . . . 25
2.8.5 Point Defect Formation Energies . . . 25
2.8.6 Surface Energies . . . 25
2.8.7 Stacking Fault Energies . . . 25
2.9 Notation . . . 25
2.9.1 Simulation Cell . . . 25

3 A PARETO APPROACH TO PARAMETER OPTIMIZATION . . . 26

3.1 Multiobjective optimization . . . 26
3.1.1 Pareto optimality . . . 27
3.2 Parameter Estimation as a Multiobjective Optimization Problem . . . 27
3.2.1 Parameters . . . 27
3.2.2 Constraints on parameters . . . 28
3.2.3 Structure Property Relationships . . . 28
3.2.4 Constraints on structure property relationships . . . 28
3.2.5 Fitting Database . . . 28
3.2.6 Prediction Error function . . . 28
3.2.7 Parameter Optimization Problem as MOO formulation . . . 28
3.3 Pareto Front . . . 29
3.4 Surveys of Methods . . . 29
3.4.1 no preference methods . . . 30
3.4.2 *a priori* methods . . . 30
3.4.3 *a posteriori* methods . . . 30
3.4.4 Interactive methods . . . 30
3.5 Solution Methods . . . 30
3.5.1 Scalarization Methods . . . 31
3.5.1.1 Weighting Method . . . 31
3.5.2 Cost Function . . . 31
3.6 Optimization Methods . . . 32
3.6.1 Convex Optimization . . . 33

3.6.2 Global Approaches . . . 33
3.6.2.1 Genetic Algorithms . . . 34
3.7 Visualization . . . 36
3.8 Treatment . . . 36
4 PROBABILITY METHODS . . . 37
4.1 Probability . . . 37
4.2 Random Variable . . . 37
4.2.1 Expectation . . . 38
4.3 Estimation of a Distribution . . . 38
4.3.1 Non-parametric Estimation . . . 38
4.3.2 Kernel Density Estimation . . . 38
4.3.2.1 Choice of kernels . . . 39
4.3.2.2 Selection of bandwidth parameters . . . 39
4.4 Bayesian Estimation . . . 39
4.5 Sampling . . . 39
4.6 Markov Chain Monte Carlo Methods . . . 39
5 EVOLUTIONARY ALGORITHM FOR DETERMINING THE PARETO SET 40
5.1 Genetic Algorithm . . . 40
6 POTENTIAL DEVELOPMENT SOFTWARE . . . 42
7 APPLICATIONS TO IONIC SYSTEMS . . . 44
8 APPLICATIONS TO NICKEL EMBEDDED ATOM POTENTIALS . . . 45
9 TECHNIQUES FOR DIMENSIONALITY REDUCTION . . . 46
9.1 Principal Components Analysis . . . 46
10 SUMMARY OF WORK . . . 47
10.1 General Implications . . . 47
10.2 Future Work . . . 47
REFERENCES . . . 48
BIOGRAPHICAL SKETCH . . . 49

LIST OF TABLES

Table page

LIST OF FIGURES

Figure page

1-1 EPS format diagram. Note: no filetype is designated by adding an extension. . . 14

1-2 BMP format drawing. Note: no filetype is designated by adding an extension. . 14

1-3 PDF format graph. Note: no filetype is designated by adding an extension. . . . 15

1-4 JPG format image. Note: no filetype is designated by adding an extension. . . . 16

1-5 PNG format map. Note: no filetype is designated by adding an extension. . . . 17

1-6 Tom and Jerry 18

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

NEW METHODS, TOOLS AND TECHNIQUES FOR THE EVOLUTIONARY DETERMINATION OF ENSEMBLES OF PARETO EFFICIENT PARAMETERIZATIONS OF CLASSICAL POTENTIALS WITH APPLICATIONS

By

Eugene Ragasa

May 2016

Chair: Simon Phillpot
Major: Materials Science and Engineering

Abstracts should be less than 350 words. Any Greek letters or symbols not found on a standard computer keyboard will have to be spelled out in the electronic version so try to avoid them in the Abstract if possible. The best way to compile the document is to use the make_xelatex.bat file. If you are using Linux or Macintosh Operating Systems there are examples of make files for these systems in the Make Files Folder but they may be outdated and need to be modified for them to work properly. This document is the official tutorial outlining the use and implementation of the UF LaTeX2ϵ Template for use on thesis and dissertations. The tutorial will cover the basic files, commands, and syntax in order to properly implement the template. It should be made clear that this tutorial will not tell one how to use LaTeX2ϵ. It will be assumed that you will have had some previous knowledge or experience with LaTeX2ϵ, but, there are many aspects of publishing for the UF Graduate School that requires attention to some details that are normally not required in LaTeX2ϵ.

Pay particular attention to the section on references. NONE of the bibliography style files (.bst) are an assurance that your document's reference style will meet the Editorial Guidelines. You MUST get a .bst file that matches the style used by the journal you used as a guide for your references and citations. The files included in this document are examples only and are NOT to be used unless they match your sample article exactly!

You should have a .bib file (we have included several examples) that contains your reference sources. Place your .bib file in the bib folder and enter the name of the file in the list of bib files, or enter your reference information into one of our existing .bib files if you don't already have one. Just make sure to preserve the format of each kind of reference. Each time you cite a reference you enter the ”key” (the first field in the reference listing in the .bib file) associated with that reference. During the compilation process LaTeX will gather all the references, insert the correct method of citation and list the references in the correct location in the proper format for the reference style selected.

CHAPTER 1
INTRODUCTION AND OPENING REMARKS

We don't make the Chapter titles in All Caps Automatically because it is easier for you to type your Chapter Titles in uppercase than for those that need to have mixed case in their titles to find the correct command in the ufthesis.cls file and change it there. * We don't recommend that you change much of anything in the class file unless you're absolutely sure of what your are doing.[1]

1.1 The Section Command Text Should Be in Title Case

Title case is where all principal words are capitalized except prepositions, articles, and conjunctions. [1]

1.1.1 Subsection Commands Are Also in Title Case

The difference, of course, are the second level headings are left-aligned

1.1.1.1 Subsubsections are in sentence case

The third level subheadings are left-aligned but in sentence case. Only the first letter and any proper nouns are capitalized. [2]

1.1.1.2 If you divide a section, you must divide it into two, or more, parts

Paragraph headings. There is no official fourth level heading. Do not use the Paragraph heading feature in LaTeX, simply apply the bold characteristic to the first few words of a paragraph followed by a colon or period.

1.1.2 I Need Another Second Level Heading in This Section

Aliquam mi nisi, tristique at rhoncus quis, consectetur non mi. Phasellus blandit quam ligula, a viverra lacus commodo at. In iaculis nisl vel pretium sollicitudin. In

* an un-numbered footnote - this is how you tell the readers that this chapter was previously published and then cite the Journal where it was published

[1] and now we're back to normal footnote marking

efficitur massa vel elit sollicitudin, vel auctor sapien cursus. Proin feugiat sapien a mi tempus;

$X - X' = D + D'$

in consequat augue cursus. Nulla sed sagittis purus. Nunc eu consequat orci, eu laoreet enim. Ut euismod tincidunt sem, eget lacinia dui luctus eu. Aliquam mi augue, faucibus id semper vitae, porta ac ligula. Morbi sed ultrices odio. Mauris id luctus ex. Nulla ac libero dictum, interdum turpis lacinia, scelerisque leo. Praesent varius orci ac eros varius pharetra.

1.2 Image Handling in XeLaTeX

One of the biggest reasons for switching from the dvipdfm/dvipdfmx methods of compiling is the improved image handling capabilities. EPS, Bit-mapped, PDF, JPG, and PNG formats work well with the xelatex process.

1.2.1 The Traditional EPS Format

EPS format is the traditional format for LaTeX, but EPS files can be very large and many programs can't create or view these images. There are many programs that are used to interpret data and output the results as an EPS format image. It has been my experience that there are bounding box problems with these figures. On many occasions we have opened the image in Adobe Photoshop and, without making any changes, saved the document as a Photoshop EPS file, re-compiled the document, and the image worked correctly, so if you are having problems with an EPS image not showing in your document correctly, try this fix first.

Quisque malesuada a leo eget ullamcorper. Curabitur ut aliquam quam. Nam quis quam id mauris aliquam blandit porttitor sit amet quam. Donec ut erat eleifend turpis finibus pulvinar.

1.2.2 Bitmapped Images Work As Well

Bitmapped images are a standard file type on PCs, but these files are also usually very large so compressed images may be a better alternative.

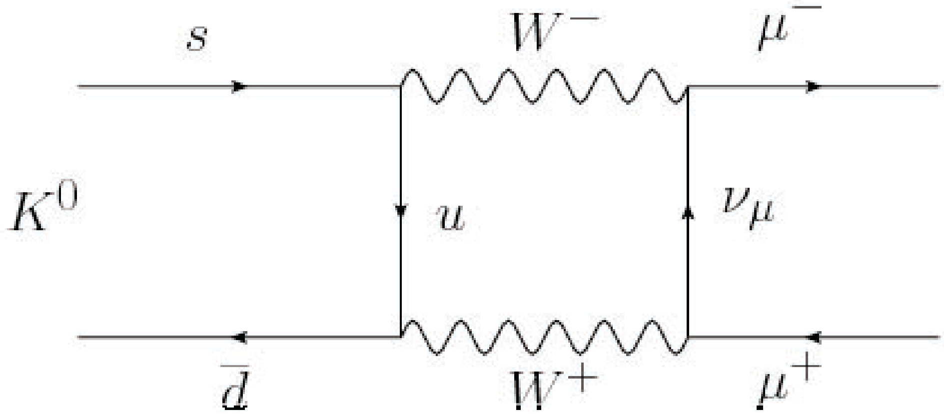

Figure 1-1. EPS format diagram. Note: no filetype is designated by adding an extension. The file type is determined and the correct procedure is automatically chosen by xelatex.

Figure 1-2. BMP format drawing. Note: no filetype is designated by adding an extension. The file type is determined and the correct procedure is automatically chosen by xelatex.

Morbi hendrerit risus nec quam posuere viverra. Donec quis tellus faucibus, molestie arcu sed, congue urna. Duis eget neque ac libero pulvinar porta eget et magna. Donec a magna eu eros suscipit cursus ac vitae nisl. Vivamus ligula purus, congue sed tortor blandit, ultrices egestas nisl.

1.2.3 Not to Mention PDF

It is often very handy to be able to include a pdf file as an image. By using XeLaTeX this is usually just matter of setting the size, or scale properties correctly.

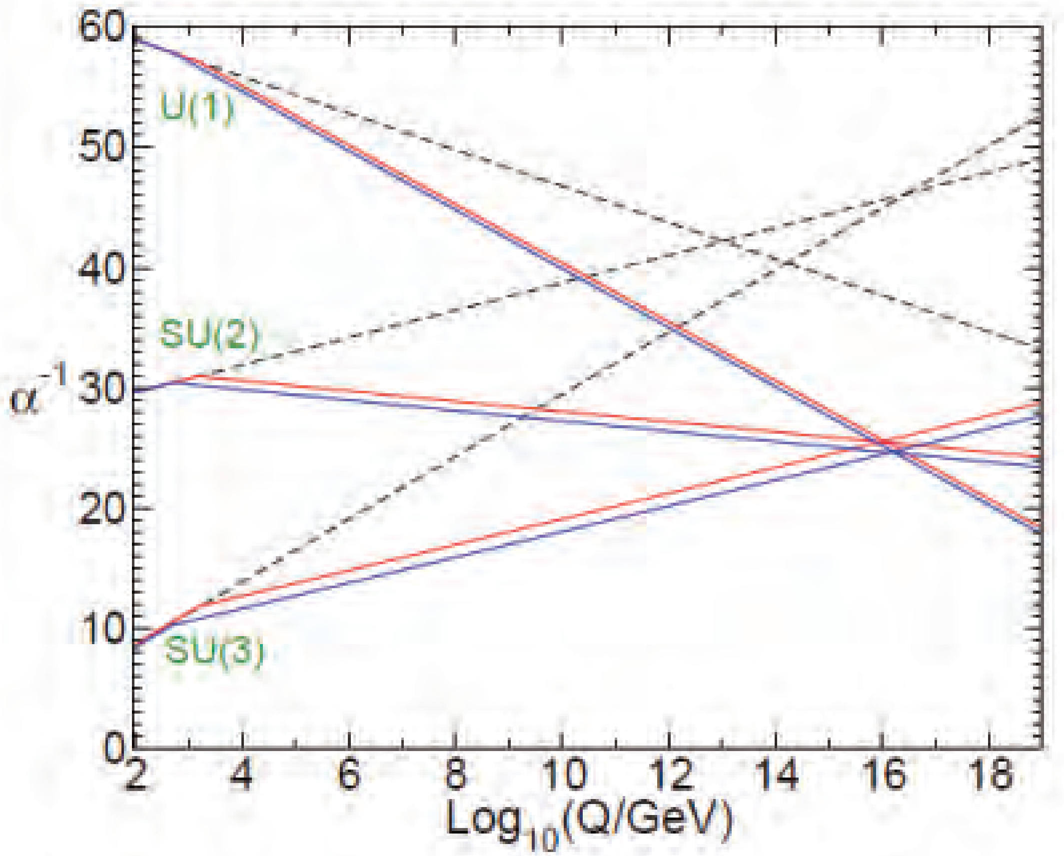

Figure 1-3. PDF format graph. Note: no filetype is designated by adding an extension. The file type is determined and the correct procedure is automatically chosen by xelatex.

Nulla mattis augue lacus. Nam non lectus dolor. Cras ac quam vel justo elementum vestibulum. Integer vulputate pulvinar lacus sit amet pulvinar.

1.2.4 JPG Is Absolutely Necessary

For photographs, JPG is the most common format. This format is a fraction of the size of Bit-mapped images and can deliver very good quality at a much smaller overhead. Vestibulum eu lectus vel orci dictum vehicula. Proin id maximus dolor. Integer augue ante, pulvinar ac erat vitae, porttitor ullamcorper libero. [3]

Figure 1-4. JPG format image. Note: no filetype is designated by adding an extension. The file type is determined and the correct procedure is automatically chosen by xelatex.

Nunc blandit scelerisque velit, ac facilisis dui finibus et. Sed facilisis tortor vel commodo luctus. Donec est felis, malesuada id nibh in, accumsan malesuada lectus. Sed lobortis volutpat felis, vitae aliquet augue congue id. Fusce ut odio tincidunt, condimentum nulla vel, pharetra arcu.

1.2.5 PNGs Will Help Make Files Smaller

PNG files are even smaller than JPGs and are very good when text and images are combined.

Figure 1-5. PNG format map. Note: no filetype is designated by adding an extension. The file type is determined and the correct procedure is automatically chosen by xelatex.

Aenean condimentum libero sed mi porta, tempus ullamcorper lectus venenatis. Aliquam in diam dolor. Maecenas tempus consectetur sem et pulvinar. Aenean aliquam at metus ut hendrerit. Vivamus molestie ac neque eu luctus. Nam convallis maximus quam non lobortis. Fusce sit amet lorem et massa convallis aliquet at sit amet nulla. Suspendisse nec ex elit. Aenean gravida, sapien vitae congue commodo, urna turpis ornare libero, at cursus risus libero in erat. [**?**]

1.3 GIF, TIF, and Others

Other file formats have not been successful, with or without file extensions. The tests have not been exhaustive so if you have a different type, give it a try. GIF, and TIF both do NOT work at this time. The next image demonstrates how to use multiple images as a single figure. Notice, there is a single caption for ALL figures and that caption starts with a discription of the ENTIRE figure before breaking off into the subfigure descriptions.

Aliquam mi nisi, tristique at rhoncus quis, consectetur non mi. Phasellus blandit quam ligula, a viverra lacus commodo at. In iaculis nisl vel pretium sollicitudin. In efficitur massa vel elit sollicitudin, vel auctor sapien cursus. Proin feugiat sapien a mi tempus, in consequat augue cursus. Nulla sed sagittis purus. Nunc eu consequat orci, eu

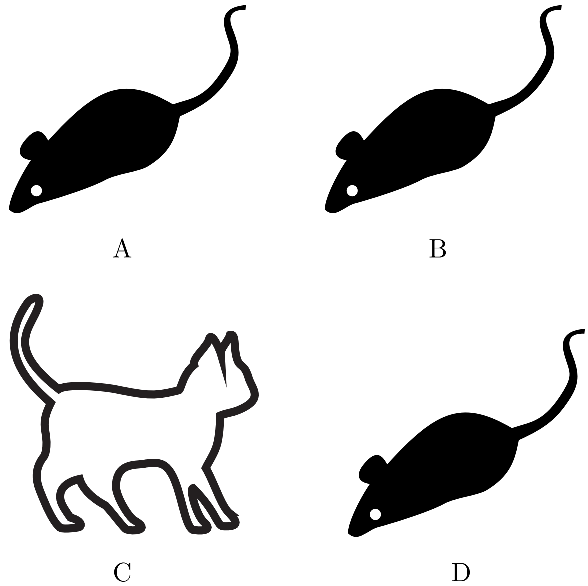

Figure 1-6. Tom and Jerries. This caption demonstrates how the sub-captions are left out of the List of Figures, but included in the figure itself. A) Tom the first; B) Tom the second; C) Jerry; D) Tom the third.

laoreet enim. Ut euismod tincidunt sem, eget lacinia dui luctus eu. Aliquam mi augue, faucibus id semper vitae, porta ac ligula. Morbi sed ultrices odio. Mauris id luctus ex. Nulla ac libero dictum, interdum turpis lacinia, scelerisque leo. Praesent varius orci ac eros varius pharetra.

Nunc blandit scelerisque velit, ac facilisis dui finibus et. Sed facilisis tortor vel commodo luctus. Donec est felis, malesuada id nibh in, accumsan malesuada lectus.

- WinEDT: This text editor is recommended for use editing TeX-files as it has many useful built in macros and is easy to use
- This program can be found and downloaded here: http://www.winedt.com/
- The GIMP (GNU Image Manipulation Program)
  - A freeware graphics editing program for picture editing and file conversions
  - Comparable to Adobe Photoshop
  - Can be downloaded here: http://www.gimp.org/
- A good reference of LaTeX2ϵ commands

- This should be included on the ETD website here: http://etd.helpdesk.ufl.edu/tex.php

Sed lobortis volutpat felis, vitae aliquet augue congue id. Fusce ut odio tincidunt, condimentum nulla vel, pharetra arcu. In ultricies libero diam, nec rutrum magna vehicula nec. Praesent dictum eros sit amet turpis ultricies, eleifend condimentum dui imperdiet. Donec congue urna ante, id rutrum mi commodo a. Vivamus id tincidunt nunc. Morbi id lacus ut augue ultricies convallis. Duis a lectus quis ante pretium scelerisque nec nec nisi. In id porta justo, at euismod diam. Suspendisse vel tempus arcu. Praesent vel cursus nisi, ac rhoncus odio.

CHAPTER 2
THE MANY BODY PROBLEM AND ATOMISTIC METHODS

The properties of a system may be obtained by solving the quantum mechanical (QM) wave equation which governs the system dynamics. For non-relativistic system, this equation is the Schrodinger's equation. For all but the simplest systems, this approach in an impossible task in practice; the resulting many body problem has only been solved for a limited number of system. Within this chapter we outline the many body problem, it's intractibility before considering the Hohenberg-Kohn-Sham formulation of density functional theory (DFT), particularly in it's formulation it's application for systems with periodic boundary conditions. This reformulates quantuim mechanics, using electron density as the fundamental parameter to solve, rather than the many-electron wavefuction. This takes the N-body problem and recasts it into N single-body problems; which is a dramatic simplification.

We then approach higher order models which reduces computational intensity by looking at classical empirical potentials and their role in both molecular dynamics and lattice dynamics.

2.1 The Many-Body Problem

Hartree and Hartree-Fock Methods

2.2 Density Functional Theory

2.2.1 The Exchange Correlation Term

2.3 Density Functional Theory for Solids

2.3.1 Representation of an Infinite Solid

2.3.2 Bloch's Theorem

2.3.3 Plane Wave Formulation

2.3.4 k-point sampling

2.4 Empirical Interatomic Potentials

The interatomic potential $U(\boldsymbol{R}_i)$ derived from the Born-Oppenheimer approximation is derived from a quantum-mechanical perspective. The computational cost of *ab initio* such as density-functional theory (DFT) can provide accurate structural energies and forces, but their computational cost limits approaches to compute $U(\boldsymbol{R}_i)$ makes the scientific inquiry of systems requiring longer simulation times or larger number of atoms to captures relevant feature sizes unreasonable.

An empirical interatomic potential $V(\boldsymbol{R}_i; \boldsymbol{\theta})$ is an analytical function parameterized by $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)$ which is meant to approximate $U(\boldsymbol{R}_i)$. The total energy of a potential of N atoms with an interaction described by the empirical potential, V, can be expanded in a many body expansion.

$$V(\boldsymbol{r}_1, ..., \boldsymbol{r}_N) = \sum_i V_1(\boldsymbol{r}_i) + \sum_i \sum_{i<j} V_2(\boldsymbol{r}_i, \boldsymbol{r}_j) + \sum_i \sum_{i<j} \sum_{j<k} V_3(\boldsymbol{r}_i, \boldsymbol{r}_j, \boldsymbol{r}_k) + ... \tag{2–1}$$

The first term V_1 is the one body term, due to an external field or boundary conditions, which is typically ignored in classical potentials. The second term V_2 is the pair potential, the interaction of the term is dependent upon the distance between $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$. The three-body term potential V_3 arises when the interaction interaction of a pair of atoms is modified by the presence of a third. Based upon this expansion, we can classify certain potentials into two classes: pair potentials when only V_2 is present and many-body potentials when V_3 and higher order terms are included.

Over the last few decades, a large number of potentials have been developed to descibe various bonding types and environments. To take representative examples, the Lennard Jones was developed for the van der Waals interactions of noble gases, pair potentials such as the Buckingham potential can be used for ionic solids, the embedded atom model (EAM) is developed for metallic systems, the Assisted Model Building with Energy Refinement (AMBER) for biomolecules, the tersoff potential for covalently bonded materials. To deal with bonding and chemical environments for heterogenous materials like metal/metal oxide interfaces have led to extensions such as MEAM, REBO, COMB, and ReaxFF.

More recently, potentials such as GAP and SNAP represent the atomic environment of an atom not from a collection of a vectors of atomic positions which feed into formulaic functional forms, but to calculate the bispectrum of the neighborhood of atoms. The bispectrum combined with an orthogonal expansion of components is dependent upon large amounts of density functional images to use in the fitting dataset to produce DFT fidelity reproductions of interatomic forces on an atom.

2.4.1 Pair Potentials

2.4.1.1 Leonnard Jones

2.4.1.2 Coulomb Potential

2.4.1.3 Born-Mayer Potential

2.4.1.4 Morse Potential

2.4.1.5 Buckingham Potential

2.4.2 Many Body Potentials

2.4.2.1 Tersoff Potentials

2.4.2.2 Stillinger Weber Potentials

2.4.2.3 Embedded Atom Method Potentials

This chapter reviews typical approaches to fitting empirical potentials.

2.5 Interatomic Potentials

The justification for the use of classical empirical potentials can be demonstrated from the Born-Oppenheimer approximation[**?**]. The Hamiltonian for a real material is defined by the presence of interacting nuclei and electrons:

$$H = \sum_i \frac{P_i}{2M_i} + \sum_\alpha \frac{p_\alpha}{2m} + \frac{1}{2}\sum_{ij} \frac{Z_i Z_j e^2}{r_{ij}} + \frac{1}{2}\sum_{\alpha\beta} \frac{e^2}{r_{\alpha\beta}} - \sum_{i\alpha} \frac{Z_i e^2}{r_{i\alpha}} \tag{2–2}$$

The first terms are kinetic energy terms, the latter terms are the nuclei-nuclei, electron-electron, and nuclei-electron interactions. Ideally, the solution of Schrödinger's equation, $H\Psi = E\Psi$ could be solved providing the total wavefunction $\Psi(\boldsymbol{r}_i, \boldsymbol{r_\alpha})$. Except for the simplest of systems, this approach is impossible computationally. The Born-Oppenheimer approximation [**?**] is ubiquitous in *ab initio* calculations, and forms the justification for classical empirical potentials. The kinetic energy is ignored since the heavy nulclei move more slowly than electrons. For the remaining interaction terms of the Hamiltonian, the nuclear positions are clamped at certain positions in space, the electron-nuclei interactions are not removed, since the electrons are still influenced by the Coulomb potential of the nuclei. This allows us to factor the wavefunction as

$$\Psi(\boldsymbol{R}_i, \boldsymbol{r}_\alpha) = \Xi(\boldsymbol{R}_i)\Phi(\boldsymbol{r}_\alpha; \boldsymbol{R}_i) \tag{2–3}$$

, where $\Xi(\boldsymbol{R}_i)$ describes the nuclei, and $\Phi(\boldsymbol{r}_\alpha; \boldsymbol{R}_i)$ describes the electrons parameterized by the clamped position of $\boldsymbol{R}_i$. In turn, the Hamiltonian is solve able as two Schrödinger's equations. The first equation contains the electronic degrees of freedom.

$$H_e\Phi(\boldsymbol{r}_\alpha; \boldsymbol{R}_i) = U(\boldsymbol{R}_i)\Phi(\boldsymbol{r}_\alpha; \boldsymbol{R}_i) \tag{2–4}$$

where

$$H_e = \sum_\alpha \frac{p_\alpha}{2m} + \frac{1}{2}\sum_{ij} \frac{Z_i Z_j e^2}{r_{ij}} + \frac{1}{2}\sum_{\alpha\beta} \frac{e^2}{r_{\alpha\beta}} - \sum_{i\alpha} \frac{Z_i e^2}{r_{i\alpha}} \tag{2–5}$$

Eqn. 2–4 gives the energy $U(\boldsymbol{R}_i)$ which depends on the clamped coodinates of $\boldsymbol{R}_i$. The electronic effects are contained in $U(\boldsymbol{R}_i)$ and is incorporated into the second equation which the motion of the nuclei

$$H_n \Xi(\boldsymbol{R}_i) = E\Xi(\boldsymbol{R}_i) \tag{2–6}$$

where

$$H_n = \sum_i \frac{P_i}{2m_i} + U(\boldsymbol{R_i}) \tag{2–7}$$

This later equation does not contain any electronic degrees of freedom, because all electronic effects are incorporated into $U(\boldsymbol{R}_i)$ which is the interatomic potential. For molecular dynamics, Schrödinger's equation is replaced with Newton's equation of motion.

2.6 Molecular Dynamics

Molecular dyanmics (MD) is a simulation approach where the time evolution of aset of interacting atoms is followed by numerically solving their equations of motion. In MD, the behavior of atoms follow Newtonian mechanics [Landau and Lifshitz 1976]:

$$M\frac{d\boldsymbol{r}(t)}{dt} = F(\boldsymbol{r}(t)) = -\nabla V(\boldsymbol{r}(t)) \tag{2–8}$$

where t is time, $\boldsymbol{r}(t) = (\boldsymbol{r}_1(t), \boldsymbol{r}_2(t), ..., \boldsymbol{r}_N(t))$ represents the forces on the particles, and M is the mass matrix, which is a diagonal matrix with the mass, m_k, for $M_{k,k} = m_k$ for all diagonal entries. 1. Every object in a state of uniform motion tends to remain in motion unless an external force is applied to it.

2.6.1 Role of the Interatomic Potential

The total enegy is converged if the atoms follow Newton's equation of motion in a conservative force field, when forces can be written in terms of the spatial derivative of the potential field, while the kinetic energy and potential energy can themselves change values dynamically. Newton's equation of motion can be written in their Hamiltonian form (Allen,Tildesley, et al 1989)':

2.6.2 Numerical Integration

A dynamical simulation computes atomic positions as a function of time given their initial position $\boldsymbol{r}(t = 0)$ and velocities $\boldsymbol{v}(t = 0)$. Since Newton's equations of motion are

2nd order diffferential equations, an initial condition needs to specify both positions and velocities of all atoms at the initial condition.

2.6.3 Thermodynamic Ensembles

2.7 Lattice Dynamics

2.8 Calculation of Material Properties

2.8.1 Structural Properties

2.8.1.1 Minimization Techniques

2.8.2 Greatest Descent

2.8.3 Conjugate Gradient

2.8.4 Phase Order Properties

2.8.5 Point Defect Formation Energies

2.8.6 Surface Energies

2.8.7 Stacking Fault Energies

2.9 Notation

2.9.1 Simulation Cell

A simulation cell is defined by the lattice basis and the atomic basis. The lattice vectors which describes the periodic boundary conditions three lattice vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$Euclidean space which forms the basis for the crystallographic system when periodic boundary conditions are applied. The translational properties of a crystal allows the simulation of an infinite bulk material from a fixed volume. In traditional crystallography, the boundaries of the unit cell were defined as a, b, c corresponding to the length of each lattice vector and the angles α, β, γ. In computatonal materials, a more convenient representation

CHAPTER 3
A PARETO APPROACH TO PARAMETER OPTIMIZATION

Many decision and plannning problems involve multiple conflicting criteria which must be considered simultaneously. In the field of optimization, problems which have multiple criteria are deferred to as multiple critieria decision making problems (MCDM) and the algorithms used to solve them as multiple-objective optimization (MOO). Here the set of feasible solutons is not known in advance, but is restricted by constraint functions. We concentrate on nonlinear multiobjective optimization and ignore approaches designed for multiobjective linear programming.

Additionally, the approach described within this chapter is described briefly and is followed on by more detailed discussions and application in following chapters and appendices. The emphasis of the development of this methdology will be on the mathematical aspects of the subject and it's applications to the development of classical empirical potentials. The intent is to provide tools for a decision maker, rather than to convince which particular optimization to use.

Secondly, another purpose of the development of this methdology is the development of parameterizations which can be expressed as an ensemble of potentials described as a probability distribution function which can be used in UQ propogation.

3.1 Multiobjective optimization

The general multi-objection optimization problem using the notation of Marler and Arora[**?**]:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \boldsymbol{F}(\boldsymbol{x}) = [F_1(\boldsymbol{x}), F_2(\boldsymbol{x}), ..., F_k(\boldsymbol{x})]^T \tag{3–1}$$

$$\text{subject to} \quad g_j(\boldsymbol{x}) \leq 0, j = 1, 2, ..., m \tag{3–2}$$

$$h_l(\boldsymbol{x}) = 0, l =, 1, 2, ..., e \tag{3–3}$$

$$\boldsymbol{x} \in \boldsymbol{X} \tag{3–4}$$

where k is the number of objective functions, m is the number of inequality constraints, and e is the number of equality constraints. The vector $\boldsymbol{x} \in \boldsymbol{X} \subseteq \mathbb{R}^n$ is a vector design variables x_i, and X is feasible design space. $\boldsymbol{F}(\boldsymbol{x}) \in \mathbb{R}^k$ are called objectives, cost functions, or criteria. The feasible critereon space Z is defined as $\{\boldsymbol{F}(\boldsymbol{x})|\boldsymbol{x} \in \boldsymbol{X}\}$.

For MOOs, the objectives are generally conflicting, preventing simultaneous optimization.

3.1.1 Pareto optimality

If all functions are for minimization, a feasible solution $\boldsymbol{x}_1$ is said to dominate another feasible solution $\boldsymbol{x}_2$, denoted $\boldsymbol{F}(\boldsymbol{x}_1 \succ \boldsymbol{x}_2)$, if and only if $F_i(\boldsymbol{x}_1) \leq F_i(\boldsymbol{x}_2)$ for $i = 1, ..., k$ and $F_j(\boldsymbol{x}_1) < F_j(\boldsymbol{x}_2)$ for at least one objective function j. A solution is said to be Pareto optimal if it is not dominated by any other solution in the solution space. A Pareto optimal solution cannot be improved with respect to any objective without worsening at least one objective function. The set of all feasible non-dominated solutions in X is referred to as the Pareto optimal set, and for a given Pareto optimal set, the corresponding values in the objective space are called the Pareto Front.

3.2 Parameter Estimation as a Multiobjective Optimization Problem

It is rather trivial to state that parameter estimation is an MOO problem. However, while this is often stated occasionally in potential development literature. It is often stated within the context of the use of global optimization techniques. The purpose of this section is to provide a clear methodological approach to determining optimal parameters within the context of MOO, and eludicate the problems often encountered in potential development, specifically to the choice of the optimization techniques often employed in potential optimization.

3.2.1 Parameters

Let V be an empirical potential parameterized by P number of parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, ...\theta_P]$.

3.2.2 Constraints on parameters

3.2.3 Structure Property Relationships

3.2.4 Constraints on structure property relationships

3.2.5 Fitting Database

A fitting database is a collection of structure property functions q_i with an associated structures S_j. The notation of q comes from verification, validation, and uncertainty quantification literature where the term quantity of interest (QOI)

Lattice constant, bulk modulus, vacancy formation energy, or anything that can be defined from energy structures. In the fitting database, the structure proerty functions evaluated using an empirical potentials and compared to target reference values, with values either determined from experimental values or a a high-fidelity structure such as DFT.

The collection of structure property relationships, is denoted $\boldsymbol{q} = (q_1, q_2, ...q_N)$ for N structure property relationships. Usually accuracy and transferribility are tested against an external database.

3.2.6 Prediction Error function

In order to assess the prediction errors of the structure property functions, we denote the $\hat{\boldsymbol{q}}(\boldsymbol{\theta}) = (\hat{q}_1(\boldsymbol{\theta}), \hat{q}_2(\boldsymbol{\theta}), ..., \hat{q}_N(\boldsymbol{\theta}))$ as the predicted material properties

The difference between the prediction values and target values of the QOIs produces a vector of error functions, $\boldsymbol{\epsilon}(\boldsymbol{\theta}) = (\hat{q}_1(\boldsymbol{\theta}) - q_1, \hat{q}_2(\boldsymbol{\theta}) - q_2, ..., \hat{q}_N(\boldsymbol{\theta}) - q_N)$,

3.2.7 Parameter Optimization Problem as MOO formulation

DEFINITION OF CONFIGURATION SPACE

The incorporation of first-principles data in the fitting database significantly improves the reliability of semi-empirical potentials by sampling a larger area of configuration space[21-28]

Ercolessi F, Adams JB. Europhys Lett 1994;26:583. Mishin Y, Farkas D, Mehl MJ, Papaconstantopoulos DA. Phys Rev B 1999;59:3393 Baskes MI, Asta M, Srinivasan SG.

Philos Mag A 2001;81:991 Mishin Y, Mehl MJ, Papaconstantopoulos DA, Voter AF, Kress JD. Phys Rev B 2001;63:224106 Mishin Y, Mehl MJ, Papaconstantopoulos DA. Phys Rev B 2002;65:224114 Li Y, Siegel DJ, Adams JB, Liu XY. Phys Rev B 2003;67:125101 Zope RR, Mishin Y. Phys Rev B 2003;68:024102 Mishin Y. Acta Mater 2004;52:1451

3.3 Pareto Front

In multiobjective optimization problems, it is characteristic that no unique solution exists, but a set of mathematically equally good solutions can be identified. These solutions are known as nondominated, efficient, noninferior or Pareto optimal solutions. In MCDM literature, these terms are synomous.

In MCDM literature, the idea of solving a multiobjective optimization problem is understood as helping a human decision maker (DM) in understanding the multiple objectives simultaneously and finding a Pareto optimal solution. Thus, the solution process requires some interaction with the DM in the form of specifying preference information and the final solution is determined by these preferences.

In potential development, the preferences of potential developer likewise influences are particular parameterization, which has results in the development of empirical potentials as somewhat of a black art. In the end, empirical potentials are simplified models which predict structure property relationships.

In classical potential optimization, the identification of an optimal parameterization is determined by the minimization of a cost function which couples multiple objective functions, usually a weighted sum of squares, and different weights are used in an interactive fashion until an acceptable parameterization is determined.

3.4 Surveys of Methods

Chankong and Haimes 1983 Hwang and Masud 1979 Marler and Arora 2004 Miettinen 1999 Sawaragi et al 1985 Steuer 1987 Vincke 1992

We start our review of methods using Hwang and Masud 1979 and Miettinen 199, to classify the different classes of approaches by methological approach rather than technical techniques.

3.4.1 no preference methods

The task is to find some neutral compromise solution without any additional information. This means instead of asking the DM for preference information, some assumption are made about what a reasonable compromise could be like.

3.4.2 *a priori* methods

In *a priori*, the DM first articulates preference information and the solution tries to find a Pareto optimal solution satisfying them as well as possible.

3.4.3 *a posteriori* methods

A representation of a set of Pareto optimal solution is first generated and then the DM is supposed to select the most preferred one among them. This approach gives the DM an overview of different solutions available but if there are more than two objectives in the problem, it may be difficult for the DM to analyze the large amount of information.

3.4.4 Interactive methods

After each iteration, some information is provided to the DM in order to specify preference information. What is noteworthy is that the DM can specify and adjust one's preferences between each iteration and at the same time learn about interdepencies between each iteration and at the same time learn about interdependencies in the problem as well as one's own preferences.

3.5 Solution Methods

MOO solution methods fall under the category of scalarization or non-scalarization methods. Scalarization is the primary method for MOO problems [Miettinen 1999]. Scalarization converts the MOO problem into a paramterized single-objective problem which can be solved using using well-established single-objective optimization methods.

3.5.1 Scalarization Methods

3.5.1.1 Weighting Method

3.5.2 Cost Function

$$C(\boldsymbol{\theta}) = \sum w_i(\hat{q}_i(\boldsymbol{\theta}) - q_i)^2 \tag{3–5}$$

Gass and Saaty 1955 Zadeh 1963

For a interatomic potential being fit with respect to k quantities of interest,

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{minimize}} \quad & \sum_{i=1}^{k} w_i \varepsilon_i^2(\boldsymbol{\theta}) \\ \text{subject to} \quad & \boldsymbol{\theta} \in \boldsymbol{\Theta} \end{aligned} \tag{3–6}$$

where $w_i \geq 0$ for $i = 1, ..., k$ Weakly Pareto optimal.

In the development of interatomic potentials, the DM is asked to specify weights in which case the method is used as an *a priori* method.

Algorithms for multiobjective optimization should produce Pareto optimal solutions, and that any Pareto optimal solution can be found. Censor1977 discusses the conditions which the whole Pareto set can be generated by the weighting metho when positive weights are presented. In this respect, the weighting method has a serious shortcoming. Any Pareto optimal solution can be found by altering weights only if the problem is convex. Some Pareto optimal solutions of nonconvex problems cannot be found regardless of how the weights are selected.

The problems of the weighting schemes have ben explored by the classical potential development community. The method may jump from one vertex to anohter vertex leaving intermediate solutions undetected with relatively small changes in the weighting schemes.

Scaling of the objective functions.

The weighting method can be used as an *a posteriori* method where different weight can be used to generate different Pareto optimal solutions, and then the DM selects the most satisfactory solution. Systemic methods of perturbing the weights to obatain different Pareto optimal solutions are suggested (Chankong and Haimes 1983), but Das and Dennis 1997 illustrates that an evenly distributed set of weights does not necessarily produce an evenly distributed representation of the Pareto optimal set, even when the problem is convex.

When the weighting scheme is used as an *a priori* method, the DM is expected to represent his/her preferences in the form of weights. Roy and Mousseau (1996) suggests that the role of weights in expressing preferences maybe mis leading. Although the relative importance of weights show the relative importance of the objective functions it is not clear what underlies this notion. The relative importance of objective functions is usually understood globally, for the entire decision problem, while many practical applications show that the importannce typically varies for different objective function values, that is, the concept is only meaningful locally. (Podinovsky 1994).

Weights that produce a certain Pareto optimal solution are not necessarily unique, and different weights may produce similar solutions. On the other hand, a small change in weights may cause big differences in the objective function. It is not easy for the potential developer to control the solution process because weights behave in an indirect way.
The solution process then becomes an interactive one where the DM trues to guess such weights that would produce a satisfactory solution, and this is not desirable because the DM cannot be properly suppported which leads to frustation complications in potential development. Instead, in such cases it is advisable to use real interactive methods where the DM can better control the solution process with more intuitive preference information.

The weighting method is also difficult

3.6 Optimization Methods

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0 \quad h_j(x) = 0 \quad x \in X \end{aligned} \tag{3–7}$$

here x is the optimization variable, f is the objective function, g_i are inequality constraints, and h_j are equality constraint functions.

3.6.1 Convex Optimization

Numerical algorithms make heavy use of scalarization results, and most papers in the field of MOO and economics deal with non-linear programming problems, corresponding duality theorems, and the repeated application of the simplex method.

However, within the literature of potential development approaches focus upon local minimization techniques and global optimization techniques.

objective function is concave. constraint set is convex. KKT requirements for uniqueness.

3.6.2 Global Approaches

Genetic algorithms are a popular meta-heuristic that is particularly well-suited for this class of problems. Traditional GA are customized to accomodate multi-objective problems by using specialized fitness functions and introducing methods to promote solution diversity.

The second general approach is to determine an entire Pareto optimal solution set or a representative subset. A Pareto optimal set is a set of solutions that are nondominated with respect to each other. While moving from one Pareto solution to another, there is always a certain amount of sacrifice in one objective(s) to achieve a certain amount of gain in the other(s). Pareto optimal solution sets are often preferred to single solutions because they can be practical when considering real-life problems since the final solution of the decision-maker is always a trade-off. Pareto optimal sets can be of varied sizes, but the size of the Pareto set usually increases with the increase in the number of objectives.

The ultimate goal of a multi-objective optimization algorithm is to identify solution in the Pareto optimal set. However, identifiying the entire Pareto optimal set, for multi-objective problems, is impossible to its size. Proof of solution optimality is computationally infeasible. Therefore, a practical approach is achieve successively better approximations of the Pareto surface that represent the Pareto set as well as possible.

A multi-objective optimization approach should achieve the following conflicting goals as described by Zitzler *et al*[**?**]: (1) the best known Pareto front should be as close as possible to the true Pareto front. Ideally, the best-known Pareto set should be a subset of the Pareto set, (2) solutions in the best known Pareto set should be uniformly distributed and diverse over the Pareto front in order to provide the decision-maker a true picture of trade-offs, and (3) the best-known Pareto front should capture the whole spectrum of the Pareto front at the extreme ends of the spectrum. While the first two goals are important for multi-objective optimization, the last goal is erroneous. In general, when developing potentials, the DM is interested in compromise solutions and a parameterization with high fidelity with respect to one material property at the expense of a loss of fidelity with respect to all other prediction would be a pathological parameterization.

3.6.2.1 Genetic Algorithms

The method which will be proposed in chapter 5 is not a genetic algorithm, but has many similarities as Genetic Algorithms but tailored to create an ensemble of Pareto optimal parameters. However, it is a genetic solution and the iterative approach of generating new populations is akin to previous solutions. As a result, the section of review in this section is necessarily incomplete but refer to an introductory review by Konak *et al*[**?**] as well as the book by Deb[**?**]

The concept of genetic algorithms were inspired by evolutionist theories explaining the origin of species[**?**]. In nature, weak and unfit speicies within their environment are faced with extinction by natural selection, while strong ones pass on their genes to future generations through reproduction. In the long run, species carrying the correct combinatioon in their genes become dominant in their population.

In GA terminology, a solution vector $\boldsymbol{x} \in \boldsymbol{X}$ is called an individual or a chromosome. Chromosomes are made of descrete units called genes. Each gene controls on or more features of the chromosome. Normally, a chromosome corresponds to a unique solution $\boldsymbol{x}$ in the solution space. This requires a mapping mechanism between the solution space and chromosome. GA operates with a collection of chromosomes, called a population. As the search evolves, the poulation includes fitter and fitter positions, eventually it converges, meaning that it is dominated by a single solution. Two operators are defined crossover and mutation. In the crossover operator, two parent solutions are combined togehter to form offspring. The mutation operator introduces random changes into the population.

The first multi-objective GA, called vector evaluated GA (or VEGA), was proposed by Schaffer [5]. Afterwards, several multi-objective evolutionary algorithms were developed including Multi-objective Genetic Algorithm (MOGA) [6], Niched Pareto Genetic Algorithm (NPGA) [7], Weight-based Genetic Algorithm (WBGA) [8], Random Weighted Genetic Algorithm (RWGA)[9], Nondominated Sorting Genetic Algorithm (NSGA) [10], Strength Pareto Evolutionary Algorithm (SPEA) [11], improved SPEA (SPEA2) [12], Pareto-Archived Evolution Strategy (PAES) [13], Pareto Envelope-based Selection Algorithm (PESA) [14], Region-based Selection in Evolutionary Multiobjective Optimization (PESA-II) [15], Fast Non-dominated Sorting Genetic Algorithm (NSGA-II) [16], Multi-objective Evolutionary Algorithm (MEA) [17], Micro-GA [18], Rank-Density Based Genetic Algorithm (RDGA) [19], and Dynamic Multi-objective Evolutionary Algorithm (DMOEA) [20]. Note that although there are many variations of multi-objective GA in the literature, these cited GA are well-known and credible algorithms that have been used in many applications and their performances were tested in several comparative studies.

Vector Evaluated Genetic Algorithm (VEGA). Schaffer proposed VEGA for finding multiple solutions to multiobjective problems. He created VEGA to find and maintain multiple classification rules in a set covering problem. VEGA tried to achieve this goal by selecting a fraction of the next generation using one of the objective functions.

Fitness Sharing encourage the search in unexplored section of a Pareto front by artificially thinning solutions in densely populated area. To achieve this goal, densely populated areas are identified and a penalty method is used to penalize the solutions located in such areas. This approach was recommended by Goldberg and Richardson[**?**] and used by Fonseca and Fleming[**?**] to penalize clustered solutions.

$$dz(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sqrt{\sum_{k=1}^{K} \left(\frac{z_k(\boldsymbol{x_1}) - z_k(\boldsymbol{x_2})}{z_k^{max} - z_k^{min}} \right)^2} \tag{3–8}$$

based on these distances, calculate a niche count for each solution $\boldsymbol{x} \in \boldsymbol{X}$ as

$$nc(\boldsymbol{x}_1, t) = \sum_{\boldsymbol{x}_2 \in \boldsymbol{X}, r(\boldsymbol{x}_2, t) = r(\boldsymbol{x}_1, t)} \max \left\{ \frac{\sigma_{share} - dz(\boldsymbol{x}_1, \boldsymbol{x}_2)}{\sigma_{share}}, 0 \right\} \tag{3–9}$$

where σ_{share} is the niche size by defining a neighborhood of solutions in the objective space. Solutions in the same neighborhood contribute to each other's nich count. Therefore, a solution in a crowded neighborhood will have a higher niche count, reducing the probability of selecting that solution from being culled from the survivor set.

3.7 Visualization

This problem is dealt with in discussions about visualization and and analysis of the large amounts of data generated from a posteriori approaches to solving these problems. Edgeworth 1881 Koopmans 1951 Kuhn Tucker 1951 Pareto 1896, 1906

3.8 Treatment

Our treatment of the mapping of the empirical potential is treated as a bijective mapping into two measure spaces.

Let us define parameter space with the probability measure space $(\Theta, \mathcal{F}(\Theta), \mathbb{P})$.

Then we define the error space of the structure property relationships with the probability measure space $(\mathcal{E}, \mathcal{F}(\mathcal{E})), \mathbb{Q})$.

CHAPTER 4
PROBABILITY METHODS

4.1 Probability

A random variable X is a variable whose possible values are outcomes of a random phenemon. As a function, a random variable is required to be measurable, which rules out pathological issues.

The underlying foundation of any probability distribution is the sample space, which is the set of all probable outcomes denoted as Ω. The realization of an outcome is denoted $\omega \in \Omega$.

The events for the measure space are defined in such a way that a probability measure can be assigned. This allows to assign probability measures on complex events to characterize groups of outcomes. The collection of all such events is a σ-algebra $\mathcal{F}$ of subsets of Ω. Not every subset of the sample space Ω must be considered an event, the σ-algebra restricts $\mathcal{F}$ to subsets of Ω for which $\mathbb{P}$ can be asssigned.

The probability measure, $\mathbb{P}$, a function which maps $\mathbb{P} : \mathcal{F} \to [0, 1]$. A probability is a real number between zero and one. Within this work we do not disguish the difference between impossible events which have probability zero, and probability-zero events which are not necessarily impossible. Events of probability one is an event that happens almost surely, with almost total certainty.

The triplet $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability measure space.

4.2 Random Variable

A random variable has a probability distribution, which specificies the probability falls in. Specifically, $X : \Omega \to \mathbb{R}$. If a random variable $X : \Omega \to \mathbb{R}$ is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the probability of an event A occuring is $\{\omega : X(\omega) = A\}$ which is denoted as $\mathbb{P}(X = A)$.

A probability density function for a random variable X has a density f_X, where f_X is a non-negative function:

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f_X(x)dx \tag{4–1}$$

4.2.1 Expectation

If X is a random variable with a finite number of outcomes, $\{x_1, x_2, ..., x_k\}$ occurs with the probabilities $\{p_1, p_2, ...p_k\}$. Then the expectation of X is defined as

$$\mathbb{E}[X] = \sum_{i=1}^{k} x_i p_i \tag{4–2}$$

If X is a continuous random variable, then

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) \tag{4–3}$$

If X admits a density $f(x)$, then the expected value is defined as

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x)dx \tag{4–4}$$

4.3 Estimation of a Distribution

4.3.1 Non-parametric Estimation

Fusce eget tempus lectus, non porttitor tellus. Aliquam molestie sed urna quis convallis. Aenean nibh eros, aliquam non eros in, tempus lacinia justo. In magna sapien, blandit a faucibus ac, scelerisque nec purus. Praesent fermentum felis nec massa interdum, vel dapibus mi luctus. Cras id fringilla mauris. Ut molestie eros mi, ut hendrerit nulla tempor et. Pellentesque tortor quam, mattis a scelerisque nec, euismod et odio. Mauris rhoncus metus sit amet risus mattis, eu mattis sem interdum.

4.3.2 Kernel Density Estimation

Let $(x_1, x_2, ...x_n)$ be a univariate and identically distributed sample drawn from some distribution with an unknown density f. The goal is to estimate the shape of this function

f. The kernel density estimator is

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{4–5}$$

where K is the kernel and $h > 0$ is a smoothing parameter called the badwidth. The kernel function satisfies the condition

$$\int_{-\infty}^{+\infty} K(x)dx = 1 \tag{4–6}$$

4.3.2.1 Choice of kernels

Popular kernels: Epanachnikov, Bi-weight, Triangular, Gaussian, Rectangular.

For the kernel method, we adopt the gaussian kernel ϕ,

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{t^2}{2\sigma^2}\right) \tag{4–7}$$

4.3.2.2 Selection of bandwidth parameters

The chosen bandwith is important because it has a strong influenjce on the boundary of the density curve. The curve boundary has poor smoothness quality when bandwidth takes small value; while as the increasing bandwidth, the smoothness improves, but the fitness of the curves becomes poor. The accuracy of kernel estimation is dependent on suitable bandwidth.

Scott' Method

Silverman Method

Chiu Method

4.4 Bayesian Estimation

4.5 Sampling

4.6 Markov Chain Monte Carlo Methods

CHAPTER 5
EVOLUTIONARY ALGORITHM FOR DETERMINING THE PARETO SET

Genetic algorithms are a popular meta-heuristic that is particularly well-suited for this class of problems. Traditional GA are customized to accomodate multi-objective problems by using specialized fitness functions and introducing methods to promote solution diversity.

We scale the objective function in

The use of evolutionary algorithms in the development of classical potentials is not new, and numerous optimization approaches such as gradient-based approaches, genetic algorithms, and neural networks have been developed.

However, in previous literature genetic algorithms are used to optimize potentials.

Our algorthm has the following goals: (1) to identify the strengths and weaknesses of solution of the Pareto optimal solutions, (2) to generate estimates of the Pareto optimal front in a serious of iteratively better approximations, and (3) to describe the candidate parameterizations through the use of a distribution function and use MCMC sampling, but updating the distribution using culling of the Pareto distribution.

5.1 Genetic Algorithm

Step 1: Set $t = 1$. Randomly generate N soluitions to form the first population, P_1. Evaluate the fitness of solutions in P_1

Step 2: Crossover

Step 3: Mutation

Step 4: Fitness assessment

Step 5: Selection. Select N solution from Q_t based on their fitness and copy them to P_{t+1}

Step 6: If the stopping criterion is satisfied, terminate the search and return to the current population, else set $t = t + 1$ and go to step 2.

In our case, the initial development of solutions requires

Constraint handling. Discard infeasible solutions (”death penalty”) Coello.

Elitism. When an external list is used to store elist solutions, several issues must be addressed. The first issue is which solutiojns are going to be stored in elitist list. Most multi-objective GA store non-dominat4ed solutions identified in the search and the list is updated by removing elitist solutions dominated by a new solution or adding the new solution if it is not dominated by any existing elitist solution.

CHAPTER 6
POTENTIAL DEVELOPMENT SOFTWARE

This software is a collection of software libaries which can be scripted together to evaluate software.

This project required the development of software for the development of Pareto frontier.

In order to support the largest number of classical potentials, software was written in python using object oriented techniques so that new types of simulations, new potentials, new quantities of interest, and new simulation software.

Evaluation of interatomic potentials. This software subprocesses either serial version of LAMMPS or GULP to calculate properties of interest. Parallelization is batched processed across iterations, which each processor rank being given a unique directory space to prevent IO conflicts.

Supports Buckingham, Tersoff, and EAM potentials. PYPOSPACK was largely developed to support LAMMPS simulations, but has limited support for GULP as well.

Support for atomic strucutures. PYPOSPACK defines structures in terms of the lattice basis of the simulation cell, and the atomic basis of the simulation cell. Typically, the software uses the POSCAR format for atomic structures used in VASP software. However, support for creating surfaces, interfaces, and other file formats from integration with the ASE toolkit.

PyposmatEngine. The purpose of this class is to evaluate a given parameterization for a QOI set by managing. This class allows potential optimization to be done through any software framework where the algorithm provides a set of parameters to be evaluated, and can return the result of structure property estimation or their predictive errors with respect to a predefined training set. Monitoring of spawned subprocesses and timeouts for poor parameterizations allows a rank to recover from structural minimization routines of unstable parameterizations.

QoiManager

TaskManager

CHAPTER 7
APPLICATIONS TO IONIC SYSTEMS

CHAPTER 8
APPLICATIONS TO NICKEL EMBEDDED ATOM POTENTIALS

The embedded atom methods

CHAPTER 9
TECHNIQUES FOR DIMENSIONALITY REDUCTION

9.1 Principal Components Analysis

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values linearly uncorrelated variables called principal compoents.

Pearson[1] and Hotelling[2]

CHAPTER 10
SUMMARY OF WORK

10.1 General Implications

10.2 Future Work

REFERENCES

[1] K. L. Green, "A wrinkle in time," *comiXology*, 2008.

[2] H. D. Strickler, P. S. Rosenberg, S. S. Devesa, J. Hertel, J. F. Fraumeni Jr, and J. J. Goedert, "Contamination of poliovirus vaccines with simian virus 40 (1955-1963) and subsequent cancer rates," *Jama*, vol. 279, no. 4, pp. 292–295, 1998.

[3] M. L'engle, *A Wrinkle in Time: 50th Anniversary Commemorative Edition.* Macmillan, 2012, vol. 1.

[4] D. Garfinkle, G. T. Horowitz, and A. Strominger, "Charged black holes in string theory," *Physical Review D*, vol. 43, no. 10, p. 3140, 1991.

BIOGRAPHICAL SKETCH

This section is where your biographical sketch is typed in the bio.tex file. It should be in third person, past tense. Do not put personal details such as your birthday in the file. Again, to make a full paragraph you must write at least three sentences.