uOttawa

L'Université canadienne
Canada's university

t

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

James George M. Hooper
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Chemistry)
GRADE / DEGREE

Department of Chemistry
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Using Evolutionary Inspired Search Methodologies to Explore Polymorphism or Defect Associations in Materials Design

TITRE DE LA THÈSE / TITLE OF THESIS

Tommy Woo
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

Anatoli Ianoul

Paul Mayer

René Fournier
York University

Alain St-Amant

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Using Evolutionary Inspired Search Methodologies To Explore Polymorphism or Defect Associations in Materials Design

James Hooper

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the Ph. D. degree in Chemistry

Candidate	Supervisor
James Hooper	Tom Woo

Ottawa-Carleton Chemistry Institute
Faculty of Science
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-79731-0
Our file Notre référence
ISBN: 978-0-494-79731-0

TABLE OF CONTENTS

ABSTRACT

The work presented in this thesis reinforces the notion of how powerful evolutionary search methodologies, specifically those inspired by the genetic algorithm (GA), are when applied to search out the chemical and/or configurational spaces of a number of chemical systems. Researchers in the field often sell the capability of GA search techniques to operate without constraints but, as is demonstrated in this thesis, sometimes the use of constraints can not only streamline computational costs but open up new avenues to explore in materials design.

As a starting point, this work first focuses on developing effective methodologies that map out nitrogen's high-pressure potential energy surface (PES) such that polymeric nitrogen, a relatively new material with a potentially promising outlook in future high-energy density materials applications,[1] could be explored. High pressure nitrogen (> 30 GPa) is an ideal candidate to explore theoretically since conclusive experimental characterizations are often beyond the reach of current laboratory technologies.[2] Theoretical and experimental assessments of polymorphism in high-pressure nitrogen have been reported in the recent literature,[3] and, upon the inception of this thesis, much of its potential energy surface remained unmapped. Using both established and novel search methodologies on nitrogen's high-pressure potential energy landscape, with GA-inspired methodologies included among them, a number of previously unreported allotropes of molecular and non-molecular nitrogen at high pressures are reported which relate well with experimental studies.

The second focal point of this thesis is to present a general methodology capable of mapping generic defect associations in doped metal oxides. Doped metal oxides have

many applications as substrates in catalysis, as gas sensors, and as next generation solid electrolytes in solid-oxide fuel cell technologies, just to name a few. The performance of these metal oxide materials can be enhanced (or, more generally, tuned) via doping of the parent oxide with other elements at varying concentrations; however, the optimization of the dopant composition is often performed by trial and error. By introducing appropriate constraints, GA-inspired routines can sample the configurational and chemical spaces of any defects introduced to a native oxide lattice. Thus, genetic algorithms are used herein in two capacities: 1) to find the lowest energy configurations of a specific metal oxide composition and 2) to search a doped metal oxide's chemical space in order to optimize a specific property of interest based on a given 'fitness' metric. This approach is validated by optimizing electronic mobility in doped zinc oxide, which is used in gas sensor technologies, and ionic mobility in Lanthanide doped ceria (LDC), which is used as a solid electrolyte in solid-oxide fuel cell technologies.

[1] M. I. Eremets, A. G. Gavriliuk, I. A. Trojan, D. A. Dzivenko, and R. Boehler, Nature Materials **3**, 558 (2004).

[2] E. Gregoryanz, C. Sanloup, R. Bini, J. Kreutz, H. J. Jodl, M. Somayazulu, H.-k. Mao, and R. J. Hemley, Journal of Chemical Physics **124**, 116102 (2006).

[3] W. D. Mattson, D. Sanchez-Portal, S. Chiesa, and R. M. Martin, Phys. Rev. Lett **93**, 125501 (2004).

TABLE OF FIGURES

TABLE OF TABLES

LIST OF ACRONYMS

BN: ***B***oron-***N***itride

DFT: ***D***ensity ***F***unctional ***T***heory

DOS: ***D***ensity ***O***f ***S***tates

GA: ***G***enetic ***A***lgorithm

GDC: ***G***adolinium-***D***oped ***C***eria

GGA: ***G***eneralized ***G***radient ***A***pproximation

HEDM: ***H***igh ***E***nergy-***D***ensity ***M***aterials

LDC: ***L***anthanide-***D***oped ***C***eria

LDOS: ***L***ocal (layered) ***D***ensity ***O***f ***S***tates

LuDC: ***L***utetium-***D***oped ***C***eria

lp: ***l***one ***p***air

NMR: ***N***uclear ***M***agnetic ***R***esonance

NN: ***N***earest ***N***eighbor

NNN: ***N***ext ***N***earest ***N***eighbor

PAW: ***P***rojector-***A***ugmented ***W***ave method

PBE: ***P***erdew, ***B***urke, ***E***rnzerhof density functional

PDOS: ***P***artial (site-decomposed) ***D***ensity ***O***f ***S***tates

PES: ***P***otential ***E***nergy ***S***urface

SC: ***S***imple ***C***ubic

SDC: ***S***amarium-***D***oped ***C***eria

SOFC: ***S***olid ***O***xide ***F***uel ***C***ell

SPM: ***S***canning ***P***robe ***M***icroscopy

USPEX: ***U***niversal ***S***tructure ***P***redictor- ***E***volutionary ***X***tallography

YDC: ***Y***ttrium-***D***oped ***C***eria

Simulation Packages

GULP: ***G***eneral ***U***tility ***L***attice ***P***rogram (classical mechanics)

MOPAC: ***M***olecular ***O***rbital ***PAC***kage (semi-empirical)

SIESTA: ***S***panish ***I***nitiative for ***E***lectronic ***S***imulations with ***T***housands of ***A***toms (DFT)

VASP: ***V***ienna ***A***b-Initio ***S***imulation ***P***ackage (DFT)

Phases of Polymeric Nitrogen

A7: *A*rsenic

azoN6: phase of polymeric nitrogen with N_6 structural motifs

BP: ***B***lack ***P***hosphorous

bstr8, bstr11: structures recovered from the systematic search method in Section 2.3

cg-N: *c*ubic *g*auche structure of polymeric *N*itrogen

ctCH: *c*is-*t*rans ***CH***ain-like phase, chains sport alternating cis/trans dihedrals

ct3CH: ***CH***ain-like phase, chains sport alternating *c*is/*t*rans/*t*rans/*t*rans dihedrals

CW: ***C***haired-***W***eb phase

gtCH: ***CH***ain-like phase, chains sport alternating *g*auche/*t*rans dihedrals

LB: ***L***ayered-***B***oat phase

NG: *N*itrogen *G*raphite phase

pN: *p*olymeric structure of *N*itrogen recovered from USPEX code

pN7: *p*olymeric structure of *N*itrogen with N_7 structural motifs

pN8: *p*olymeric structure of *N*itrogen with N_8 structural motifs

TW: ***T***wisted-***W***eb phase

zzCH: *z*ig*z*ag ***CH***ain-like phase

ACKNOWLEDGEMENTS

I studied for five years at the University of Ottawa; there were of course some rough patches here and there but there were plenty of good times filled with new research experiences, great companionship, opportunities to not only learn but to teach, tests of willpower and intellectual strength, and perhaps an overly heavy emphasis on honing my skills at racquet sports.

As such, my first thanks go out to my research group, the Woo Lab. Having worked with fellow Ph. D. student Chris Rowley almost my entire time here, my first thanks go out to him. I have come to really appreciate his dedication to his work and his remarkable, and independent, perspective on the world at large. In this regard, I would also like to thank Dr. Saman Alavi, who joined the lab in 2006 as a post-doctoral fellow. He set a great example in the lab by showing dedication and passion for his work. Concerning my own student learning experience, I would like to thank Dr. Federico Zahariev, Dr. Anguang Hu, Dr. Sergey Dudiy, and, in particular, Dr. Zongxian Yang for their help in getting me familiarized with linux systems, with the software packages I've used throughout my Ph. D., and for frank discussions about the early stages of my research project. And, of course, I would like to thank Dr. Tom Woo for giving me the opportunity to come to Ottawa and pursue a graduate degree in an area of research I'm very interested in. Tom is a respected scientist with a fiery resolve to pursue cutting-edge, applicable computational methodologies and apply them to mainstream, topical chemical projects.

I didn't have the chance to work with everyone that came into the Woo Lab, but the lab has always had a friendly, albeit sometimes unorthodox, atmosphere (it's a lab of theoretical chemists after all). It's safe to say that my love of squash came from the infamous "Woo Cup" that David Gerbasi, Peter Dornan, and I organized within and around the lab. I'd like to thank former and current group members Nick Mosey, Nick Trefiak, Peter Boyd, Andrew Sirjoosingh, Dr. Carlos Campana, Dr. Serge Gorelsky, Michael Nohr, Dave Vanrijswijk, Eric Beamish, Tianyi Zhang, Max Garcia-Melchor, and Tosaporn "Jae" Sattasathuchana for their support. Concerning this thesis, I'd like to add a special thanks to Peter Boyd and Nick Trefiak, who

helped me edit segments of my thesis.

I would like to thank the National Science and Engineering Research Council of Canada (NSERC), the Canada Research Chair program, the University of Ottawa, and the University of Western Ontario for financial support throughout my studies. I am also grateful to the SharcNet and HPCVL high-performance compute clusters, CFI, the Ontario Innovation Trust, and IBM Canada for providing computing resources.

One the most enjoyable and fulfilling aspects of my graduate studies was the assigned responsibilities of instructing and/or supervising other researchers. I won't name all the peers that I worked with here, but I had a great time training a number of undergraduates, graduates, and even a couple of post-doctoral fellows; the scope of these projects forced me to learn methods I would not have had to use over the course of my own research. Foremost, I would like to thank Effiette Sauer, who was finishing up her Ph. D in the department in Dr. Louis Barriault's lab while I was just getting started. It's hard to do this justice really... her great work ethic and general congeniality were infectious. In spite of my "instructor" role in the project, I must say it wasn't easy to stay one step ahead. Second, I'd like to thank Arif Ismail, who is currently working on his Master's degree in our lab. Arif and I started working together last summer and, together, we have made a lot of progress on the project over one short year and have set the stage for a research program that could endure.

Beyond my own research lab, I'd also like to thank the members of my 'adopted' lab, the Bryce lab, for their support. I've shared a few beers with Dr. David Bryce over the years and really appreciate his example, his friendship, and his words of support. As for the rest of his lab, it's certainly not a stretch to say that I have dropped by to visit their lab two or three times a day over my entire tenure in Ottawa and would like to thank all of them, past and present, for their support. In particular, I'd like to single out Becky Chapman and Cory Widdifield, who have always been up for hanging out, lending a hand, proofreading parts of my thesis, or giving some much needed advice when the going got rough. I'd also like to single out some other friends from the department, namely my ex-roommate and longtime friend Mark Perry, and Heather Foucault, Caitlin Ritz, Terri Clarke, Jessica Priem and Ian Mallov.

I would like to thank Dr. Javier Giorgi for his ongoing collaborations and helpful discussions. His timely insight and constructive criticisms throughout every stage of the project on doped ceria (Chapters 5 and 6) have been greatly appreciated. I would also like to thank Dr. Louis Barriault for his willingness to let Effie, his PhD student at the time, dive head over heels into running calculations with myself to try to gain insight into her research project. Most recently, his role in helping me tie up loose ends and publish some of our, at the time unpublished, results (even though Effie had long since graduated) was very much appreciated.

Most of all, I'd like to thank my family. I have always tried to follow the example set by my Mom and Dad, Dawn and David Hooper, and, of course, I know that my siblings, John, Rebecca, and Marcia, and my grandmother have always been there and will continue to support me; the sentiment is mutual. This of course extends to my extended family as well: my in-law, my cousins, my aunts, and my uncles have persistently kept themselves informed on my seemingly perpetual status in school and their regular messages and uplifting birthday or Easter cards were much appreciated.

And last, but not least, I'd like to thank you, the reader, for taking the time to read my Ph. D. thesis.

CHAPTER ONE

INTRODUCTION

This chapter sets the stage for the motivation behind the work presented throughout this thesis.

CHAPTER 1 – *KEY SUBSECTIONS*

1.1 – Modern Computational Approaches Toward Mapping Out Crystal Polymorphism

1.1.1 – Motivation

The potential energy surface (PES) for any chemical system consists of an intricate web of local minima and saddlepoints that describe all possible isomers and reaction coordinates the system may assume. In chemistry, more often than not, the research community is interested in studying the global minimum of the PES, since this point in configurational space represents the most thermodynamically stable isomer and can be used to characterize the crystal macroscopically should that isomer dominate its overall expression. Sometimes, however, higher-energy configurations are more relevant to a material's experimental applications, particularly in regimes with highly strained chemical environments. For example, density functional theory (DFT) calculations on the high-pressure ε phase of molecular oxygen, the ε-O_2 structure, suggest that it assumes a thermodynamically unstable crystal structure beyond 17 GPa pressures. [1, 2] The true ε-O_2 structure was identified from single-crystal x-ray diffraction experiments in a diamond anvil cell environment and its lattice has a higher enthalpy than other structural candidates proposed at the DFT level of theory.[1] The true coordinates were, in fact, not recovered in the DFT search and show that, when computationally investigating a hitherto unexplored potential energy surface, locating relevant local minima on a generic PES can be tedious and onerous on computational resources.[1, 2] Such difficulties persist, for the most part, even with small and seemingly trivial molecular crystals.

When considering small, isolated molecular systems, it is often straightforward to find the lowest energy isomer by first using one's chemical intuition to design all relevant configurational isomers and then singling out structures of interest via static energy

calculations on each conformer. However, the problem becomes increasingly difficult as the number of atoms in the chemical system increases. First, the number of rotable bonds will likely increase and lead to a higher number of local minima on the PES. Second, the probable existence of a number of thermoneutral local minima complicates the landscape of thermally accessible structures available to the chemical system and must be taken into account when probing macroscopic observables. Thus, relating static energy calculations to these macroscopic properties seen in experiment becomes increasingly difficult, necessitating the practice of extracting measurements from a pool of structures. The means of collecting such a pool of structures relies mainly on either static calculations guided by chemical intuition, deterministic molecular-dynamics simulations, or stochastic Monte-Carlo techniques.

In solids, local minima on the PES correspond to mechanically stable crystal configurations and, as with isolated gas-phase molecules, the global minimum corresponds to the most thermodynamically stable crystal structure of the system. When the potential energy barrier required to escape a local minima on the PES is large, as depicted in Figure 1.1, the system is effectively trapped in that configuration and is said to be "metastable". Metastable structures will propagate to the global minimum configuration under thermodynamic conditions, but often the transformations occur at such slow rates, as dictated by the magnitude of the potential energy barrier, that they effectively never come to pass. For example it is well known that diamond gradually transforms into graphite, the more thermodynamically stable form of carbon, under thermodynamic conditions, but such transitions are not seen due to diamond's inherent kinetic stabilities. Consider that with recent advances in synthesizing diamond directly

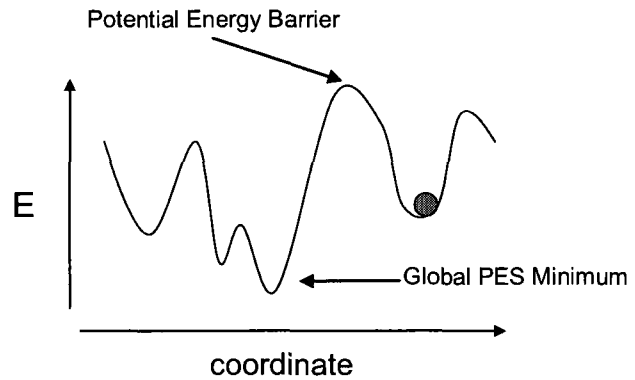


Figure 1.1. Illustration of metastability with a one-dimensional schematic of a potential energy surface with one degree of freedom. The energy (y axis) is plotted against the reaction coordinate (x axis), depicting a metastable state (marked by the filled circle) separated from the global minimum by a large energy barrier.

from chemical vapor deposition techniques,[3] thereby bypassing the need to compress graphite, it serves as an example of a chemical system in which the global minimum of its PES can be ignored in favor of, in carbon's case, its superhard diamond allotrope.

The existence of more than one possible kinetically stable crystal structure for a given chemical system leads to "polymorphism", the ability of a solid material to exist in more than one form or crystal structure. Each structure, or "polymorph", has unique constituent environments and can therefore exhibit different macroscopic properties. For example the rutile, anatase, and brookite structures of TiO_2 are polymorphs which exhibit different electronic structure and optical properties, the rutile structure is the most abundant and stable form[4] but the anatase phase in particular plays a prominent role in photochemical solar cells.[5] It stands to reason that for any new solid-state material, whose structure is ambiguous or unclear, it is often necessary to explore a number of relevant, metastable local minima in order to be able to predict/explain its inherent macroscopic chemical properties. For example, studies of the rutile TiO_2 phase alone would give an incomplete picture of its optical prowess.

From an experimental point of view, structure determination has been an active area of research since the latter half of the 20th century. There have been a number of elaborate and well-refined characterization techniques developed that will determine and/or refine more-or-less unambiguous crystal structures at standard experimental conditions, namely those based on X-ray diffraction,[6] Raman spectroscopy,[7] Solid State NMR spectroscopy,[8] or Scanning Probe Microscopy (SPM)[9] techniques. But from a modeling perspective, crystal-structure prediction has been a long withstanding problem in computational circles due to the high dimensionality of the PES as chemical systems grow in size. This increased dimensionality explains the field's need for the ever-expanding and evolving DFT-based simulation packages now readily available to the research community at large.[10-12] As such, effective methodologies that incorporate extensive or pseudo-"exhaustive" theoretical predictions of polymorphism on hitherto unexplored potential energy surfaces have only recently become feasible.

The goal of studying polymorphism in crystals from a computational point of view is to design an effective means of exploring the entire PES, ideally by mapping out all local minima and their relative enthalpies. This train of thought first leads to the most straight-forward conceivable search method, the brute-force approach. The brute-force approach entails generating every possible configuration of the chemical system by constructing an appropriate grid in configurational space and evaluating their respective energies in separate energy calculations. With these energies, one could extrapolate the entire PES and find the local minima. Such an approach leads to unreasonable computational demands, however, since the dimensionality of the PES increases to 3N as N, the number of constituents (ie. atoms) in the unit cell, increases and becomes

intractable even for relatively small simulation cells due to the combinatorial expansion of the search space. For example, it would take ~244 million energy evaluations to exhaustively search all configurations of the PES of two atoms in a 125 $Å^3$ cubic box (5 Å sides) if the search is carried out by placing atoms on a three-dimensional cartesian grid in the simulation cell such that each grid point is separated from its neighbor by 0.2 Å. Placing an additional i^{th} atom in the simulation cell will scale the cost of the systematic search by 15625-i energy evaluations. In addition, if one wishes to explore different choices of cell parameters then even more degrees of freedom will need to be introduced. Although these generous estimates could be reduced by screening symmetry-equivalent geometries or using a more physically intuitive choice of grid points, it can readily be seen that brute-force approaches for even the simplest worthwhile crystal systems are unfeasible and more intelligent methodologies are needed to streamline systematic searches of more complex potential energy surfaces.

1.1.2 – Current Computational Methodologies Dealing with Explorations of Polymorphism

Many methodologies have been developed to deal with the scope of mapping polymorphism in solid state materials, varying from those guided by rigorous mathematical treatments or transformations of configurational space, stemming from graph theory for example, [13, 14] to those which use stochastic or deterministic algorithms in the hopes of sampling relevant and representative regions of the potential energy surface.[15-18]

From the topological perspective, several methods have been developed which use algorithms that systematically explore all possible configurations of a crystal by

enumerating all possible ways to tile relevant topological surfaces. Each of these "enumerations" can then be related to 3D crystal structures in euclidean space using methods from combinatorial group theory,[13] most simply by having nodes and edges of a periodic graph represent atoms and bonds respectively,[14] and each real-space structure is processed to determine its physical relevance. The breadth of these "enumerations" is realized by finding all the possible ways to arrange regular geometrical shapes in some well-defined topological space such that the union of all their faces partitions the space into bound, connected regions or "tiles".[13, 14] The nature of these tiles is carefully chosen to reflect the local bonding environments of the atoms or constituent molecules in the crystal, thereby allowing the method to be used in a generic sense by defining particular shapes to reflect particular components of the crystal. Such methodologies have been applied to investigate allotropes of sp^2-carbon[14] or zeolites, aluminophosphates, nitrides, and ice.[13] The downside of these methods, however, is that a large number of physically irrelevant structures (particularly for large systems) must somehow be efficiently weeded out, the tiling concepts are complex and difficult to grasp, and the methodology is inefficient for screening crystals with specific local atomic environments and a small number of atoms per primitive cell.[14]

Generally, however, crystal structure predictions are still guided prominently by non-systematic search methods,[18] like simulated annealing techniques.[19] Annealing simulations incorporate successive molecular dynamics simulations from a relevant, local minimum on the PES. These simulations are designed to deterministically propagate the motion of the structure's atomic constituents in time while sequentially increasing and decreasing the simulation temperature. The goal behind such work is to give the system

enough kinetic energy to overcome potential energy barriers and hopefully fall into an unexplored local, or global, minimum. A recent extension of this methodology is metadynamics,[20] which propagates a system as an ordinary molecular dynamics simulation would but adds a bias potential to the PES over time that effectively prevents the simulation from exploring the same local minimum twice. Although physically intuitive, these methods often only explore a limited region of the PES or, in the case of metadynamics, are quite sensitive to the functional form of the added bias potential and are more suited for studying phase transitions than polymorphism.

Other less-physical methods, based on stochastic monte-carlo principles,[15, 16] have also been used extensively to search out favorable conformations by recursively generating new structures from applications of random perturbations to previously-generated structures. The drawback of monte-carlo based searches however is that their inherent markovian statistics[21] cause the algorithm to completely "forget" the structural information from prior monte-carlo steps, causing the search algorithm to spend significant time re-exploring high-energy or uninteresting parts of the PES.

Recently, "intelligent" search algorithms have been developed which are specifically designed to explore relevant regions of the PES with minimal redundant sampling of the same local minima. These methodologies are guided by storing pertinent structural information throughout the course of the entire simulation, namely by using principles inspired by intelligent design or a genetic algorithm.[1, 2, 22-29] Search procedures inspired by the genetic algorithm (GA) have historically been used extensively by the research community at large,[29, 30] particularly in specialized chemical studies. The fundamental innovation of these techniques is that a number of mating and mutation

operations are applied to a pool of suitable genetic representations of the system under study. Each genetic representation, or "gene", corresponds to a specific chemical structure the system can undertake and the mating and mutation routines are designed such that favorable structural motifs are promoted from one pool of structures to the next. When these routines are applied recursively, it is hoped they will effectively search out favored structures of the chemical system at large. Historically, however, such GA applications were typically tuned specifically to the problem at hand, diminishing their value as a generic tool for studying arbitrary crystal configurations.

The first reports that renewed general interest in applying generic GA procedures to crystal polymorphism were reported in the mid-1990's, when Ho *et. al* proposed defining the genetic representation of a crystal structure directly from its spatial atomic coordinates.[31] These unconstrained search methods were applied successfully to model carbon clusters (like C_{60}) with tight-binding model potentials[31] and Ar_nH_2 microclusters[22] or Lennard-Jones clusters[32] with lennard-jones pairwise potentials. Since this breakthrough, the advent of high-performance computational research facilities and the improved accuracy and scalability of electronic structure methods have enabled exhaustive GA-inspired searches of monatomic covalent crystals, like carbon, oxygen, sulfur, and nitrogen[2] at the DFT level of theory. Very recent studies of high-pressure $CaCO_3$,[23] metallic oxygen,[1] polymeric nitrogen,[25, 28] ionic boron,[33] and transparent sodium[27] have all incorporated GA-inspired protocols to recover previously undiscovered high-pressure allotropes that had not previously been characterized experimentally.

GA-inspired methodologies have also been applied successfully to a number of other solid-state chemistry related applications. For example, they have been used to

determine reconstructions of the Si(105) and Si(114) surfaces,[34, 35] to identify transition-metal complexes capable of mediating N_2 cleavage,[36] to determine interaction potentials of N_2O in *para*-hydrogen clusters,[37] and to predict new alloys relevant to materials design.[38] Furthermore since the inception of this thesis, GA-based methodologies have enjoyed continued and immense success in applications to a wide variety of chemical systems at the DFT level of theory. They have been used to determine organic crystal structures from electron diffraction data[39] and powder x-ray diffraction data,[40] and to lay the foundation for an inverse band structure method designed to optimize specific properties in semiconductor alloys.[41]

The work presented in this thesis reinforces the notion of how powerful these methodologies are and combines GA protocols with previously proposed approaches in materials design toward specific applications at the DFT level of theory. The focal point of this work is twofold: 1) to work toward mapping the nitrogen potential energy surface such that polymeric nitrogen, a relatively new material with potential uses as a high-energy density material, could be explored, and 2) to apply GA-inspired search procedures to broader and more application-specific research interests. Recent developments in polymeric nitrogen research circles are discussed in Section 1.2.

1.2 – Polymeric Nitrogen as a High Energy-Density Material

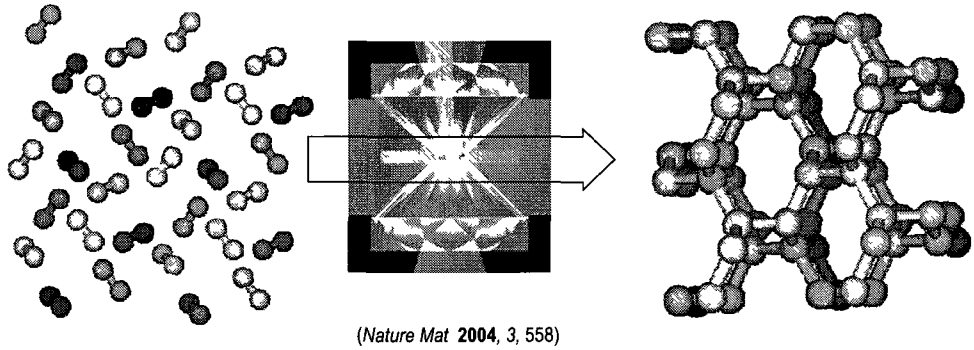
1.2.1 - Motivation

Recently, metastable allotropes of non-molecular nitrogen have been ardently pursued as potential future high energy-density materials.[42-48] High energy-density materials (HEDMs) containing high compositions of nitrogen are particularly promising candidates for energy storage, due to the large binding energy differences between triple-

and double- or single- bonded nitrogen. The 4.89 eV/atom molecular N_2 bond strength is well over three times the average single N-N bond energy of 1.42 eV/atom, making nitrogen possess, by far, the greatest such energy difference among all the covalent, high-pressure molecular solids. For example, the bond energy of a C≡C triple bond is only ~2.4 times that of a C-C single bond. Thus, the amount of energy released when polynitrogen species decompose into molecular N_2 overshadows even the strongest conventional HEDMs. For this reason, polymeric forms of nitrogen, networks of solid nitrogen connected via single or double bonds, are, in some research circles, considered the holy grail of HEDMs.

The scientific community has long sought to synthesize homoleptic molecular polynitrogen compounds, likely since the discovery of the N_3^- anion in the late 19th century. The N_5^+ cation was the first such compound synthesized in the lab in macroscopic quantities, achieved only recently just before the turn of the millenium.[49] Efforts have persisted since its breakthrough synthesis and, although larger pure nitrogen molecules remain elusive, further syntheses of molecular compounds with a high composition of nitrogen have recently been reported, such as the successful synthesis of CHN_7 [50] and the CN_7^- anion.[51] The theoretical literature on pure-nitrogen molecules is particularly exhaustive, covering an array of N_5, N_6, N_7, N_8, and larger all-nitrogen molecules,[52, 53] even the nitrogen "bucky ball" (N_{60}) which was theoretically designed from N_5 (N_{10}) building blocks.[54] The consistent inability to synthesize any of these in the lab however has prompted further study on high nitrogen content molecules, like the CN_7^- anion, which stands as the current focus of research efforts on molecular energetic materials.

In spite of substantial activity in ultra-high pressure research, non-molecular polymeric nitrogen itself remains elusive due to the extraordinary stability of its precursor, the triple-bonded molecular nitrogen phase. The first successful synthesis of crystalline polymeric nitrogen required pressures beyond 100 GPa, in a diamond anvil cell environment, and laser heating such that temperatures encroached near 2000 K in order to form the polymeric network,[45] see Figure 1.2.



(*Nature Mat* **2004**, *3*, 558)

Figure 1.2. Schematic depicting the observed molecular N_2 crystallization reported by Eremets *et al*: compression of molecular N_2 in a diamond anvil environment to create the cg-N[55] form of polymeric nitrogen

Regarding the pressure evolution of nitrogen itself, weak intermolecular N_2 bonding interactions, derived from weak van der Waals and quadrupole forces, instigate the formation of molecular, not polymeric,[44] crystals at low temperatures. The disordered ß-N_2 and ordered α-N_2 structures of molecular nitrogen are first formed as the pressure is increased from ambient conditions.[56] The weak, attractive, intermolecular interactions soften upon further pressurization until the electronic structure is largely directed by the intramolecular repulsion from the localized, high kinetic-energy electrons in the dinitrogen molecule.[57] This electronic instability is well-known to spark further type II phase transitions to the γ-N_2 and ε-N_2 phases of nitrogen.[58, 59] A number of further molecular phase transformations have been identified from subsequent Raman and X-ray

investigations, but the limits inherent to X-ray diffraction measurements of low-Z materials make precise structure characterizations elusive.

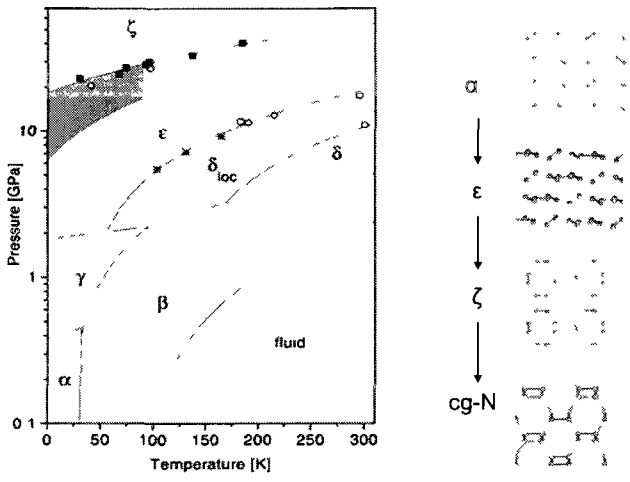
The first syntheses of non-molecular nitrogen from the compression of solid N_2 in diamond anvil cell environments were reported in 2001,[44, 48] wherein the reported transformations to amorphous nitrogen samples required pressures beyond 100 GPa and temperatures between 300 and 500 Kelvin. Only one of the samples were recovered at ambient conditions[44] and they were heralded as the mysterious η-N phase of nitrogen,[48] known only to be an amorphous, non-molecular solid nitrogen structure. Several crystalline polymeric nitrogen samples have since been recovered[42, 43] and were confirmed, through x-ray diffraction, to have the infamous cubic gauche, cg-N, structure of polymeric nitrogen.[55] However, efforts to recover the cg-N structure at ambient conditions failed, wherein the most promising sample was only maintained until 42 GPa. Remarkably, the successful characterization of the cg-N structure agreed with an earlier DFT study by Mailhot *and* McMahan (from a decade before),[55] which had predicted the cg-N structure to be the most thermodynamically stable form of nitrogen beyond 50 GPa. The authors also predicted that, like diamond, cg-N should remain metastable at ambient conditions but, to date, this has not been achieved or confirmed experimentally.

The promising experimental breakthroughs made in the early 21st century sparked substantial experimental and theoretical interest in polymeric nitrogen.[42, 44] The emergence of highly parallelized and accurate ab-initio plane wave codes allowed broader theoretical characterizations of the nitrogen landscape than ever before, and the emergence and improvement of diamond anvil cell techniques allowed the scientific community to aggressively pursue high-pressure covalently-bonded molecular solids as

HEDMs, like those of CO, CO_2, and N_2.[57] From these efforts, a broad literature for both molecular and polymeric nitrogen solids at high pressures has emerged and an overview of these works is discussed in Sections 1.2.2 and 1.2.3.

1.2.2 - Nitrogen Phase Diagram

Historically, the phase diagram of high pressure molecular nitrogen has been studied extensively and the subject of heavy debate. As the pressure is increased, molecular solids adopt crystal structures that are more densely packed. In 2004, the observed molecular-to- polymeric pathway started by freezing nitrogen at low pressures to obtain the solid, ordered, cubic molecular α-N_2 phase after passing through the intermediate hexagonal, disordered ß-N_2 phase.[45] As the pressure is increased, Raman and x-ray diffraction studies identified successive phase transformations through the ordered molecular γ-N_2, ε-N_2, and ζ-N_2 phases to the polymeric cg-N phase,[45] as shown in an abbreviated pathway shown in Figure 1.3. Note that the molecular to crystalline polymeric transition was seen at 120 GPa only if the sample was laser heated to temperatures beyond 2000 K.[42]



Bini et. al. *JCP* **112** (2000) 8522

Figure 1.3. The hypothesized pathway from molecular N_2 to cg-N in 2004.[45]

Since the initial ground-breaking synthesis by Eremets *et al*, several molecular to polymeric pathways have been observed. A second pathway uses the cubic, disordered δ-N_2 phase as a starting point and then passes through the ordered δ_{loc}-N_2, ε-N_2, ζ-N_2, and κ-N_2 (not shown in Figure 1.3) molecular phases to the amorphous, polymeric η-N phase at 160 GPa.[43] A final transformation is observed after subjecting the sample to high temperatures to obtain the final cg-N structure.[43] Meanwhile, the hysteresis-dependent ι-N_2 and θ-N_2 molecular crystal structures, which were identified experimentally by noting changes in the vibrational and lattice Raman modes of the nitrogen sample, have been seen after annealing the ε-N_2 and η-N_2 phases.[48, 60]

The current depiction of the phase diagram of nitrogen is shown in Figure 1.4.[61] Concerning the molecular N_2 solids, the atomic coordinates of the α-N_2, β-N_2, δ-N_2, and ε-N_2 phases have been solved explicitly from experiment. However, there is still no general consensus on the ζ-N_2, δ_{loc}-N_2, ι-N_2, and θ-N_2 structures, despite several proposed symmetries for each phase and a number of proposed atomic coordinates derived from symmetry arguments pertaining to the observed cell parameters and systematic extinctions in the diffraction pattern. Note that a similar debate raged in the literature over whether the ε-N_2 structure had $R\bar{3}c$ or R3c symmetry for the better part of a decade before finally being resolved as $R\bar{3}c$[62] from single crystal x-ray diffraction in the late 1990s.

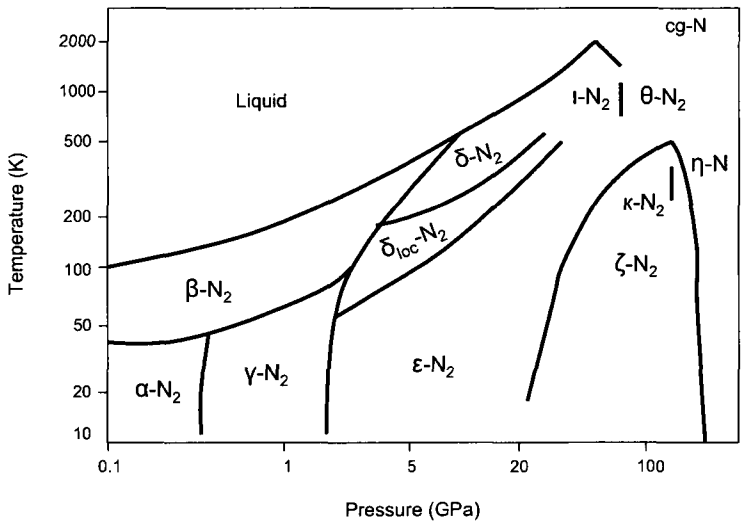


Figure 1.4. Current picture of the nitrogen phase diagram.[61]

1.2.3 – Allotropes of Polymeric Nitrogen

As shown in Figure 1.4, the only polymeric nitrogen structures observed from experiment are the crystalline cg–N and amorphous η-N forms of nitrogen. However, a number of theoretical studies on high-pressure nitrogen have recovered several other structural candidates for polymeric nitrogen which could prove relevant to the research community. Polymeric nitrogen itself is, after all, highly sought for its potential HEDM applications and these alternative phases would, conceivably, possess higher energy-density than the cg-N structure.

Prior to 2006, the ultra-high pressure PES of pure polymeric nitrogen had only been computationally explored either by creative design,[55] by exploring structures designed from chemical intuition,[63, 64] or by computational experiments.[65, 66] The early recovery of the cg-N structure from theoretical studies[55] best demonstrates how the theoretical tools at the disposal of the research community are effective in probing the

feasibility of polymeric nitrogen, since these calculations were performed well before the cg-N structure was isolated experimentally.

The proposed cg-N (*cubic gauche*) structure, shown in Figure 1.5a, was initially "designed from scratch" as a candidate polymeric nitrogen structure with exclusively *gauche* lp-N-N-lp (lp = lone pair) dihedral angles and proposed to be the most thermodynamically stable nitrogen allotrope when pressures exceed 50±15 GPa.[55] Note that it has since been determined that cg-N is in fact the only polymeric nitrogen phase that can feature exclusively gauche dihedral angles. By further relating the cg-N structure to the β-O_2[67] diatomic oxygen structure via a reaction coordinate derived from common symmetry subgroups, the cg-N structure was additionally proposed to be highly metastable even at ambient conditions. Although this result has yet to be realized experimentally, it nonetheless sparked substantial interest in polymeric nitrogen concerning its possible application as a HEDM and experimentalists redoubled their efforts to synthesize polymeric nitrogen. Shortly thereafter, Barbee *et. al* published theoretically calculated phonon dispersion curves of the cg-N structure which suggested and supported the claim that cg-N was indeed mechanically and dynamically stable at ambient pressures.[68]

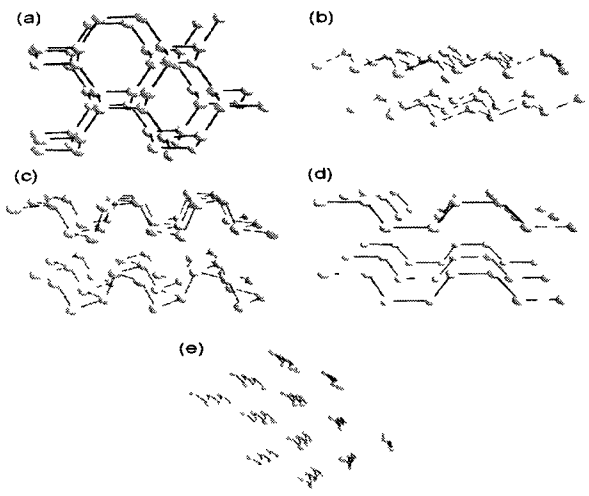


Figure 1.5. Structures of select polymeric nitrogen phases predicted from theoretical investigations, (a) cg-N, (b) A7, (c) BP, (d) ctCH, and (e) zzCH.

In further computational studies, prominently using DFT, Mailhiot *et. al* proposed another polymeric nitrogen structure by mimicking a structure from another known Group V monatomic covalent network, namely that of the layered black phosphorous (BP) structure adopted by phosphorous, shown Figure 1.5c.[55] The arsenic (A7) structure, shown in Figure 1.5b, had previously been proposed on similar premises and used in energy barrier estimates of the polymeric to molecular transition.[64] Although proposed to be stable, these phases were noted to be markedly higher in enthalpy than the cg-N structure below ~200 GPa. Mailhiot *and* McMahan also proposed a chain-like phase of polymeric nitrogen, consisting of two-fold coordinate nitrogen atoms arranged in a one-dimensional chain such that the intrachain N-N-N-N dihedral angles alternated between *cis* and *trans* along the length of the chain.[55] This chain-like phase, shown in Figure 1.5d, will herein be referred to as the ctCH phase.

More recent theoretical studies investigated the nitrogen landscape by computationally simulating annealing or by simulating an application of pressure on solved molecular or proposed polymeric structures.[65,66] Such explorations are typically

confined to limited, but hopefully relevant, regions of the PES, and carried out by either running tandem molecular dynamics simulations and geometry optimizations[65] or by applying controlled perturbations, like isotropic lattice rescalings, to the initial structures.[66] From the molecular dynamics simulations, the authors recovered a new chain-like phase composed of nitrogen atoms arranged in a zig-zag pattern along one dimension.[65] This new phase was heralded as being the lowest-energy form of non-molecular nitrogen below 20 GPa, supplanting even the cg-N structure. A year later, Martin *et. al* also recovered a phase reminiscent of this zigzag chain-like phase by homogeneously scaling the crystal lattice coordinates of the ε-N_2 phase to assume a volume of 6.4 $Å^3$/atom and reoptimizing the scaled structure.[66] The authors found, however, that the zig-zag chains aligned differently to create a distinct, lower-energy phase with *C*mcm symmetry, this phase is herein referred to as the zzCH phase and is shown in Figure 1.5e.

From these computational studies three classes of polymeric nitrogen structures can be identified: 1) the three-dimensional networks composed solely of three-coordinate, single-bonded nitrogen, 2) the layered-sheet structures sporting single-bonded nitrogen networks that spread, under periodic boundary conditions, to infinity in two dimensions, and 3) the chain-like structures with single and double bond character that span to infinity in only one dimension. The highly sought cg-N structure is a pure single-bonded three-dimensional network, the BP and A7 structures are layered two-dimensional structures, and the ctCH or zzCH chain-like structures can be considered their one-dimensional analogs. The chain-like structures are thought to be the immediate successors to the molecular N_2 phases, since their bonding character would fall between that of cg-N and

its molecular precursors, and generally assumed to play a key role in the ill-defined η-N phase.[66]

All of the discussed polymeric nitrogen phases become more thermodynamically stable with respect to solid molecular N_2 beyond 50 GPa. To gauge their mechanical stabilities, the cg-N phase was shown to be mechanically stable from its calculated phonon dispersion curves[68] and the zzCH was deemed to be stable based on long-running molecular dynamics simulations and random cartesian-coordinate perturbation tests designed to test the curvature of the PES.[66] The existence of these phases on the nitrogen PES shows that the landscape of polymeric nitrogen is likely as complicated as that of molecular nitrogen and, upon the inception of this thesis, remained largely unexplored, thus warranting further study. Given the extreme conditions needed to synthesize polynitrogen allotropes, and the high energy barriers that need to be overcome experimentally, other metastable phases of nitrogen may be accessible and prove relevant to the research community should they prove to be more stable at ambient conditions.

1.3 – Thesis Outline

As discussed in section 1.1.2, the genetic algorithm (GA) has proven to be an effective computational tool for probing the potential energy surface of new materials. In this work, PES searches of previously unexplored systems are carried out by using a computational package of python modules designed to interface evolutionary-inspired algorithms and electronic structure energy calculations. Most prominently, the GA-inspired procedures are used herein to help map out the PES of high-pressure nitrogen, which is an ideal candidate to explore theoretically since the harsh conditions in the diamond anvil cell environments and the weak scattering properties of nitrogen make

exhaustive and conclusive experimental characterizations beyond the reach of current laboratory technologies.

Given the relative novelty of polymeric nitrogen in the literature, a number of research avenues remained unexplored upon the inception of this thesis. The PES of polymeric nitrogen was preliminarily explored but still largely unmapped, allowing the possibility for other structures to be discovered, characterized, and reported which may prove relevant to polymeric nitrogen's applications as a HEDM. Furthermore, the phase diagram of polymeric nitrogen's precursor, molecular nitrogen, is also not well understood at high pressures. Although a consensus on the high-pressure ε-N_2 phase has been reached, the structure of cg-N's immediate precursor, the so-called ζ-N_2 phase, is still unresolved. With these goals in mind, the evolutionary search procedures developed in this work are used, among other methods, to search for alternative polymeric nitrogen allotropes to the cg-N structure and for structural candidates of the ζ-N_2 molecular phase.

In addition, the potential benefits and limitations of these GA-inspired search procedures in materials design are probed to promote the technology on a broader and more application-specific research scale. It is shown that by actually constraining the search space of the procedure itself, these GA-inspired search procedures can readily be used to explore metastable crystal structures, such that the procedure recovers structures consistent with experimentally-derived lattice parameters, and used in materials design to search doped metal-oxide materials for structures with either low association energies or optimal electronic and/or ionic properties at the DFT level of theory. The advantage of such practices over other search techniques is the ability of the GA to readily explore high-energy structures that show promise toward a specific application, thereby allowing

researchers to envision how to go about seeking optimal performances when synthesizing a particular material. A chapter-by-chapter breakdown of this thesis is given below.

In Chapter Two, the high-pressure nitrogen PES is explored with both systematic and non-systematic search methods, which are not based on the genetic algorithm, to illustrate the complexity of the problem at hand. Prior to the presented work, the PES of pure nitrogen had only been explored either by creative design, exploring structures designed from chemical intuition, or by computational experiments confined to limited regions of the PES. In the first section of this chapter, computational experiments are applied to the α-N_2 and ζ-N_2 molecular phases,[45, 58] which had not been investigated in the previous computational studies reported in the literature. New molecular and polymeric phases are recovered from the procedure, the "aligned" and layered boat (LB) phases respectively, and the "aligned" molecular phase is further related to the cg-N structure via a simple reaction coordinate. In the second section, it is shown that one can fashion relevant polymeric nitrogen phases by recognizing structural motifs preferably adopted by nitrogen in its local chemical environments. These motifs can readily be used as building blocks to explore other low energy structures or construct pathways between them. Another metastable polymeric-nitrogen phase is reported, the chaired-web (CW) phase, and a pathway between the ζ-N_2 and cg-N phases is discussed. Finally, in the third section, a novel systematic search method is reported that relates single-bonded nitrogen allotropes to Peierls-like distortions of a reference simple cubic structure. All of the previously reported polymeric nitrogen phases were recovered from the procedure along with two additional low-enthalpy allotropes, the bstr8 and bstr11 phases.

In Chapter Three, a detailed description of the genetic algorithm inspired search procedure and the nuances of the particular code developed 'in-house' for this work are discussed, highlighting the machinery behind structure generation and fitness evaluation. The code is first benchmarked against high-pressure nitrogen and the high-pressure post-aragonite phase of $CaCO_3$.[23] All of the relevant low-enthalpy phases of $CaCO_3$ and nitrogen are recovered, along with a number of new polymeric nitrogen phases similar to others discussed in the literature. One of the recovered nitrogen phases, to so-called nitrogen graphite (NG) phase, becomes thermodynamically stable to cg-N and zzCH at ambient pressures, supplanting zzCH as the most stable form of polymeric nitrogen at ambient pressures. The limits and optimal parameters of the GA search procedure are then discussed in more detail, namely by probing the flexibility of the mating/mutation and fitness components. And, finally, the GA protocols are applied to predict new high-pressure allotropes of hydrogen-capped nitrogen chains as potential HEDM materials.

In Chapter Four, the genetic algorithm is applied to study structures that are not thermodynamically favored on their PES. In the first section, the search procedure is modified to search for low-enthalpy structures of molecular nitrogen compatible with experimentally-identified lattice coordinates of the ζ-N_2 molecular phase in the 80-90 GPa pressure range. The findings shed light on the two conflicting experimental reports in the literature, identify several structural candidates, and stand as the first reported case where the genetic algorithm was used to explore a class of structures that were not thermodynamically favored at the given P-T conditions; polymeric nitrogen becomes the favored nitrogen allotrope beyond 50 GPa.[69] In the second section, further and complementary studies are presented to show the merit of using evolutionary algorithms

in tandem with powder x-ray diffraction data to obtain crystal structures for systems found to be difficult to resolve directly from the diffraction pattern. Such approaches were effective for the ζ-N_2 study discussed above and are additionally shown to efficiently recover the recently discovered ε-O_2, ζ-O_2, post-aragonite $CaCO_3$, and DyB_4C structures directly from crystallographic data. Finally, the constrained GA search techniques are applied to study low-pressure phases of polymeric nitrogen, which are otherwise unmapped by traditional GA searches at low pressures due to the extraordinary thermodynamic stability of molecular nitrogen; a number of novel polymeric phases are recovered from the procedure.

In Chapter Five, the evolutionary search procedures are adapted so they can be used as effective computational tools in materials design toward studying the tunability of metal oxides. In the first section, the genetic algorithm is used to explore association energies of lanthanide-doped ceria (LDC), a material which has direct applications in existing solid-oxide fuel-cell technologies due to its high ionic conductivity. The search is shown to reproduce the results of recent systematic searches of 3.2% LDC (low concentrations) at classical levels of theory, to reproduce the expected defect-association trends across the lanthanide series at 6.6% LDC concentrations, and replace systematic searches of concentrations higher than 3.2% LDC at an effective DFT level of theory using tandem classical and DFT energy evaluations. Finally, a number of searches are performed at even higher SDC concentrations, up to 20.0% SDC, to assess the nature of its increased conductivity at low concentrations and decreased conductivity beyond ~11% SDC thresholds. An "ordering" of intrinsic defects, characterized by neighboring oxygen vacancies, was noted beyond 12.5% SDC concentrations, thereby reinforcing the

proposed reasoning for SDC's diminished ionic conductivity beyond ~11% SDC. In addition, an introduction of more diverse dopant-vacancy interactions and a gradual removal of anisotropy in defect association complexes were noted as the concentration was increased below ~11% SDC. These findings would explain SDC's enhanced conductivity at low concentrations.

Finally, in Chapter Six, the evolutionary search procedures are shown to coalesce well with proposed "inverse optimization" schemes, methodologies which optimize structure configurations by a metric other than the structures' relative energies. Specifically, this is demonstrated by optimizing electronic mobility in doped zinc and tin metal oxides, an application which could have useful implications in helping develop improved gas sensor technologies, and by optimizing ionic mobility in LDC materials, an application which could have useful implications in shedding additional light on the role of the dopant atoms and screening libraries of possible co-dopants in next-generation solid electrolytes. The optimal arrangements from Hg-/Cd-/Ba-/Sr-/Ca-/Mg- doped ZnO and Pb-doped SnO_2 simulation cells are deduced from the procedure at the DFT level of theory and the results subsequently analyzed, and the LDC inverse optimization scheme is shown to be fundamentally sound in a "proof of principle" context on 10.3% LDC configurations (L = Sm, Gd, Lu, Y). It is noted that this work lays a solid foundation for studying optimal chemical processes in heterogeneous catalysis, namely that of a guest molecule adsorbed onto a doped metal oxide surface slab. When coupled with more "standard" inverse optimization schemes on the parent bulk metal-oxide itself, it has the potential to deliver a complete theoretical profile, both explanatory and predictive, of an entire catalytic process.

1.4 - References

1 Y. Ma, A. R. Oganov, and C. W. Glass, Phys. Rev. B **76**, 064101 (2007).
2 A. R. Oganov and C. W. Glass, J. Chem. Phys. **124**, 244704 (2006).
3 B. V. Deryagin and D. V. Fedoseev, Rus. Chem. Rev. **39**, 783 (1970).
4 N. N. Greenwood and A. Earnshaw, *Chemistry of the Elements* (Oxford: Pergamon, 1984).
5 S.-D. Mo and W. Y. Ching, Phys. Rev. B **51**, 13023 (1995).
6 L. V. Azaroff, R. Kaplow, N. Kato, R. J. Weiss, A. J. C. Wilson, and R. A. Young, *X-ray diffraction* (McGraw-Hill, 1974).
7 D. J. Gardiner, *Practical Raman Spectroscopy* (Springer-Verlag, 1989).
8 R. K. Harris, R. E. Wasylishen, and M. J. Duer, *NMR Crystallography* (The Encyclopedia of Magnetic Resonance, 2009).
9 F. J. Giessible, Reviews of Modern Physics **75**, 949 (2003).
10 W. Kohn and L. J. Sham, Physical Review **140**, A1133 (1965).
11 G. Kresse and J. Furthmuller, Computational Materials Science **6**, 15 (1996).
12 J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal, Journal of Physics: Condensed Matter **14**, 2745 (2002).
13 O. D. Friedrichs, A. W. M. Dress, D. H. Huson, J. Klinowski, and A. L. MacKay, Nature **400**, 644 (1999).
14 B. Winkler, C. J. Pickard, V. Milman, and G. Thim, Chemical Physics Letters **337**, 36 (2001).
15 R. Y. Robert and D. P. Kroese, *Simulation and the Monte Carlo Method* (New York: John Wiley and Sons, 2007).
16 N. Metropolis and S. Ulam, J. Am. Stat. Assoc. **44**, 335 (1949).
17 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, 1989).
18 A. Leach, *Molecular Modelling: Principles and Applications* (Prentice Hall, 2001).
19 S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1985).
20 A. Laio and M. Parrinello, Proceedings of the National Academy of Sciences **99**, 12562 (2002).
21 G. Bolch, S. Greiner, H. de Meer, and K. S. Trivedi, *Queuing Networks and Markov Chains* (John Wiley, 2006).
22 Y. Zeiri, Phys. Rev. E **51**, R2769 (1995).
23 A. R. Oganov, C. W. Glass, and S. Ono, Earth Planet.Sci.Lett. **241**, 95 (2006).
24 G. Gao, A. R. Oganov, A. Bergara, M. Martinez-Canales, T. Cui, T. Iitaka, Y. Ma, and G. Zou, Phys. Rev. Lett **101**, 107002 (2008).
25 Y. Yao, J. S. Tse, and K. Tanaka, Phys. Rev. B **77**, 052103 (2008).
26 Q. Li, M. Yanming, A. R. Oganov, H. Wang, H. Wang, Y. Xu, T. Cui, H. K. Mao, and G. Zou, Phys. Rev. Lett **102**, 175506 (2009).
27 Y. Ma, M. I. Eremets, A. R. Oganov, Y. Xie, I. A. Trojan, S. Medvedev, A. O. Lyakhov, M. Valle, and V. Prakapenka, Nature **458**, 182 (2009).
28 Y. Ma, A. R. Oganov, Z. Li, Y. Xie, and J. Kotakoski, Phys. Rev. Lett **102**, 065501 (2009).

[29] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, 1989).

[30] D. Ashlock, *Evolutionary Computation for Modeling and Optimization* (Springer, 2006).

[31] D. M. Deaven and K. M. Ho, Phys. Rev. Lett **75**, 288 (1995).

[32] D. M. Deaven, N. Tit, J. R. Morris, and K. M. Ho, Chem. Phys. Lett **256**, 195 (1996).

[33] A. R. Oganov, J. Chen, C. Gatti, Y. Ma, Y. Ma, C. W. Glass, Z. Liu, T. Yu, O. O. Kurakevych, and V. L. Solozhenko, Nature **457**, 863 (2009).

[34] F. C. Chuang, C. V. Ciobanu, V. B. Shenoy, C. Z. Wang, and K. M. Ho, Surface Science **573**, L375 (2004).

[35] F. C. Chuang, C. V. Ciobanu, C. Predescu, C. Z. Wang, and K. M. Ho, Surface Science **578**, 183 (2005).

[36] M. C. Durrant, Chemistry A European Journal **13**, 3406 (2007).

[37] H. Zhu and D. Xie, Journal of Computational Chemistry **30**, 841 (2009).

[38] G. H. Johannesson, T. Bligaard, A. V. Ruban, H. L. Skriver, K. W. Jacobsen, and J. K. Norskov, Phys. Rev. Lett. **88**, 255506 (2002).

[39] S. Habershon and A. H. Zewail, ChemPhysChem **7**, 353 (2006).

[40] K. D. M. Harris, Materials and Manufacturing Processes **24**, 293 (2009).

[41] K. Kim, P. A. Graf, and W. B. Jones, Journal of Computational Physics **208**, 735 (2005).

[42] M. I. Eremets, A. G. Gavriliuk, I. A. Trojan, D. A. Dzivenko, and R. Boehler, Nat. Materials **3**, 558 (2004).

[43] E. Gregoryanz, A. F. Goncharov, C. Sanloup, M. Somayazulu, H.-k. Mao, and R. J. Hemley, J. Chem. Phys. **126**, 184505 (2007).

[44] M. I. Eremets, R. J. Hemley, H. K. Mao, and E. Gregoryanz, Nature **411**, 170 (2001).

[45] M. I. Eremets, A. G. Gavriliuk, N. R. Serebryanaya, I. A. Trojan, D. A. Dzivenko, R. Boehler, H. K. Mao, and R. J. Hemley, J. Chem. Phys. **121**, 11296 (2004).

[46] M. J. Lipp, J. P. Klepeis, B. J. Baer, H. Cynn, W. J. Evans, V. Iota, and C. S. Yoo, Phys. Rev. B **76**, 014113 (2007).

[47] M. I. Eremets, A. G. Gavriliuk, and I. A. Trojan, Applied Physics Letters **90**, 171904 (2007).

[48] E. Gregoryanz, A. F. Goncharov, R. J. Hemley, and H. K. Mao, Phys. Rev. B **64**, 052103 (2001).

[49] K. O. Christe, W. W. Wilson, A. Sheehy, and J. A. Boatz, Angewandte Chemie International Edition **38**, 2004 (1999).

[50] G. Steinhauser, Sitzungsber. Abt. II **217**, 3 (2008).

[51] T. M. Klapotke and J. Stierstorfer, Journal of the American Chemistry Society **131**, 1122 (2009).

[52] M. Noyman, S. Zilberg, and Y. Haas, J. Phys. Chem. A **113**, 7376 (2009).

[53] L. Gagliardi, S. Evangelisti, P.-O. Widmark, and B. Roos, Theor. Chem. Acc. **97**, 136 (1997).

[54] M. R. Manaa, Chem. Phys. Lett **331**, 262 (2000).

[55] C. Mailhiot, L. H. Yang, and A. K. McMahan, Phys. Rev. B **46**, 14419 (1992).

[56] B. M. Powell, G. Dolling, and H. F. Nieman, J. Chem. Phys. **79**, 982 (1983).

[57] M. J. Lipp, W. J. Evans, B. J. Baer, and C. S. Yoo, Nat. Materials **4**, 211 (2005).
[58] A. F. Schuch and R. L. Mills, J. Chem. Phys. **52**, 6000 (1970).
[59] R. L. Mills, B. Olinger, and D. T. Cromer, J. Chem. Phys. **84**, 2837 (1985).
[60] K. Nordlund, A. Krasheninnikov, N. Juslin, J. Nord, and K. Albe, Europhysics Letters **65**, 400 (2004).
[61] H. Katzke and P. Toledano, Phys. Rev. B **78**, 064103 (2008).
[62] M. I. M. Scheerboom and J. A. Schouten, J. Chem. Phys. **105**, 2553 (1996).
[63] A. K. McMahan and R. LeSar, Phys. Rev. Lett **54**, 1929 (1985).
[64] R. M. Martin and R. J. Needs, Phys. Rev. B **34**, 5082 (1986).
[65] M. M. G. Alemany and J. L. Martins, Phys. Rev. B **68**, 024110 (2003).
[66] W. D. Mattson, D. Sanchez-Portal, S. Chiesa, and R. M. Martin, Phys. Rev. Lett **93**, 125501 (2004).
[67] C. A. English and J. A. Venables, Proc. R. Soc. London, Ser. A **340**, 57 (1974).
[68] T. W. Barbee III, Phys. Rev. B **48**, 9327 (1993).
[69] J. Kotakoski and K. Albe, Phys. Rev. B **77**, 144109 (2008).

CHAPTER TWO

DFT INVESTIGATIONS OF THE HIGH-PRESSURE POTENTIAL ENERGY SURFACE OF NITROGEN

This chapter focuses on illustrating the complexity of the high-pressure nitrogen potential energy landscape. Several computational methodologies (based on approaches adapted from both well-established and novel research practices) are applied to molecular and polymeric nitrogen solids.

CHAPTER 2 – *KEY SUBSECTIONS*

2.1 – Pressure-Induced Transformations of High-Pressure Molecular Nitrogen Phases

2.1.1 - Motivation

The first theoretical explorations of polymeric nitrogen relied on using the researcher's chemical intuition to design candidate nitrogen structures, either by mimicking structures of other Group V covalent crystals, like black phosphorous and arsenic, or solving crystal structures that optimize hyperconjugation and lone-pair repulsion interactions between the constituent nitrogen atoms.[1-3] The latter approach was the basis behind the original prediction of the lowest-enthalpy cubic gauche (cg-N) polymeric structure.[3] A more recent approach that successfully recovered several previously unexplored polymeric nitrogen phases was to computationally mimic the application of pressure. This was carried out by incrementally changing the pressure on the cell of known solid molecular structures in an attempt to view novel martensitic phase transitions.[4-6]

These "computational experiments" can be justified for high pressure nitrogen since most of its solid molecular-to-molecular phase transitions are associated with small volume changes[7-11] consistent with diffusionless transitions, characterized by coordinated molecular displacements over distances smaller than the interatomic distances in the parent phase.[12] For example, an early classical molecular dynamics study suggested the δ-N_2 → ε-N_2 transition occurred in three cascading steps, progressing from the disordered δ-N_2 parent phase, whose dinitrogen molecules are classified by disk-shaped or by sphere-shaped disordered molecular sites, to the ordered ε-N_2 phase. First, the disk-

shaped disordered molecules align themselves parallel to the [001] axis of δ-N_2, second, the sphere-shaped disordered molecules align along the trigonal direction of the cell until, finally, the ε-N_2 phase is recovered after a shear-dominated cell transformation.[13, 14] Similar volume-preserving transitions are characteristic of other solid dinitrogen phase transitions and are seen throughout the 2-120 GPa pressure range, suggesting the observed molecular nitrogen phases are closely related by first- or second- order transitions which preserve group-subgroup relationships.[12] By either applying or releasing pressure at regular increments on solved high-pressure nitrogen structures, or cascading the two treatments together, insight into the experimentally-observed pressure-induced phase transitions at high pressures should prove prolific.

The observed molecular-to-polymeric nitrogen pathways were discussed in section 1.2.2, namely the ß-N_2→ α-N_2→ γ-N_2→ ε-N_2→ ζ-N_2→ cg-N and δ-N_2 → δ_{loc}-N_2→ ε-N_2 → ζ-N_2 → κ-N_2 → η-N → cg-N routes. There have been reports that other molecular nitrogen phases, namely the unsolved ι-N_2 and θ-N_2 crystal structures,[11] have been recovered by annealing the ε-N_2 and ζ-N_2 phases at high temperatures and pressures, suggesting that other unexplored transition pathways remain unexplored on the nitrogen PES. Computational density-functional-theory (DFT) studies on the high pressure ε-N_2 phase have already been reported in the literature,[5, 6] and, in this section, the α-N_2, ε-N_2, and ζ-N_2 coordinates are used as starting points for further computational pressure-application studies. Note that the α-N_2 and ε-N_2 structures themselves are well-characterized, having been solved explicitly from experiment, but there is no general consensus on the ζ-N_2 phase, although ζ-N_2 coordinates have been proposed.[7] Therefore,

these calculations also shed light on the viability of the preliminarily proposed ζ-N_2 phase coordinates.

2.1.2 – Computational Details

The Spanish Initiative for Electronic Simulations with Thousands of Atoms (SIESTA)[15] programming suite was used in all the calculations presented in this section (2.1), employing Kohn-Sham DFT calculations[16, 17] with the gradient-corrected exchange-correlation functionals of Perdew, Burke, and Ernzerhof (PBE).[18, 19] Since DFT is a well established computational method used to calculate electronic structures of chemical systems, the details of DFT will not be elaborated upon in this thesis; the reader is referred to the comprehensive review of DFT and its applications, *Electronic Structure: Basic Theory and Practical Methods* by Richard M. Martin,[17] for further information. All of the SIESTA calculations used a Troullier-Martins norm-conserving pseudopotential[20-22] referencing a core $1s^2$ configuration with a 0.98 Å cutoff. The low cutoff was chosen to allow for greater transferability in the high pressure simulation environments. Custom SIESTA-type numerical doubled s, p_x, p_y, and p_z orbitals and d polarization orbitals[23] were used with a real-space mesh cut-off of 200.0 Ry. A 10 Å cutoff for k-point sampling was used to construct Monkhorst-Pack grids for Brillouin-zone integration. Similar DFT calculations on polymeric nitrogen with SIESTA have been recently reported.[6] The projector-augmented[24] planewave-based DFT program VASP[25] was used to verify the existence of all local minima on the PES recovered from the simulations.

Specifically, sixty-four atom supercells of all the structures, both molecular and polymeric, were optimized at all the presented pressures by allowing both the cell shape

and nuclear geometries to change under constant external pressure. Molecular dynamics simulations were used to probe the dynamical and mechanical stability of some of the phases reported in this section; the reader is referred to *Essentials of Computational Chemistry: Theories and Models* by Christopher J. Cramer [26] for further information regarding first-principles molecular dynamics studies. The isothermal-isobaric (NPT) molecular dynamics simulations reported in Sections 2.1.4 and 2.2.5 were performed using the Parinello-Raman variable cell method[27] with Nose-Hoover thermostating[28] and a 1 fs time step.

2.1.3 - Compression of the α-N_2 molecular phase

The α-N_2 molecular phase, pictured in Figure 2.1a, has orthorhombic $Pa\bar{3}$ symmetry with eight atoms occupying the cell's 8c special Wyckoff sites.[29] The α-N_2 phase atomic coordinates and cell lattice vectors were first optimized under variable cell conditions at 10 GPa, yielding a lattice constant of 4.52 Å. We note that, experimentally, the α-N_2 phase is only seen at low temperatures below ~0.5 GPa, but the Woo lab has found from prior benchmark calculations that successive pressurization of the zero-pressure unit cell showed negligible strain when compressed to 10 GPa. Additionally, since low-pressure molecular nitrogen is guided largely by weak intermolecular dispersion or quadrupole interactions, DFT's well known difficulties when modeling dispersion interactions[30] can be expected to be less of a factor in pressurized simulation environments where steric interactions play a larger role. Sixty-four atom supercells were used for all the succeeding optimizations, originating from the 2x2x2 simulation cell of the 10 GPa -optimized parent α-N_2 phase. The optimized 2x2x2 α-N_2 phase crystal structure coordinates were first reoptimized at 15 GPa under variable cell

conditions, the optimized coordinates were then extracted and reoptimized once more at 20 GPa. This reoptimization cycle was repeated by incrementally increasing the pressure on the simulation cell by 5 GPa between optimizations until a final pressure of 280 GPa was reached. A series of structural transformations were observed during this computational isotropic compression and are discussed in detail below.

First, a gradual phase transformation to a triclinic $P2_1/C$ symmetry structure was observed at 30 GPa, this phase's unit cell is shown in Figure 2.1b. The monoclinic structure has a four atom unit cell with its constituent atoms distributed among 4e special Wyckoff sites. The structure is formed when the α-N_2 phase molecules centered at 8c Wyckoff sites rotate off the body diagonals to realign themselves along the triclinic square diagonals of its adopted, sheared cell. The $P2_1/C$ phase is closely related, via a 15° shearing distortion of the ß lattice angle and an 18% dilation of the relative c/b lattice vector ratios, to the well-characterized tetragonal γ-N_2 phase, which also has four atoms in its unit cell distributed over 4f special Wyckoff sites as shown in Figure 2.1c. The triclinic distortion seen in the observed molecular phase, which is expected from nitrogen's phase diagram to be the γ-N_2 phase, is likely an artifact of DFT's problematic description of van der Waal's interactions, as suggested by the incorrect lattice parameters, or of the artificial, incrementally applied isotropic pressure on the α-N_2 phase coordinates. The sudden jumps in pressure effectively force the atomic constituents into unfavorable positions with each jump, inducing a pseudo-barrierless shearing strain on the cell which could otherwise be alleviated by a gradual continuous application of pressure. The observed transformation is an encouraging development nonetheless since the α-$N_2 \rightarrow \gamma$-N_2 phase transformation is expected from the literature,[21] albeit at much

lower pressures, manifesting itself when the packing of the ellipsoidal molecules on a tetragonal lattice overcomes the electric quadrupole-quadrupole interactions responsible for the α-N_2 phase crystal structure.[14] It would be more fundamentally problematic if no transformation was observed at all.

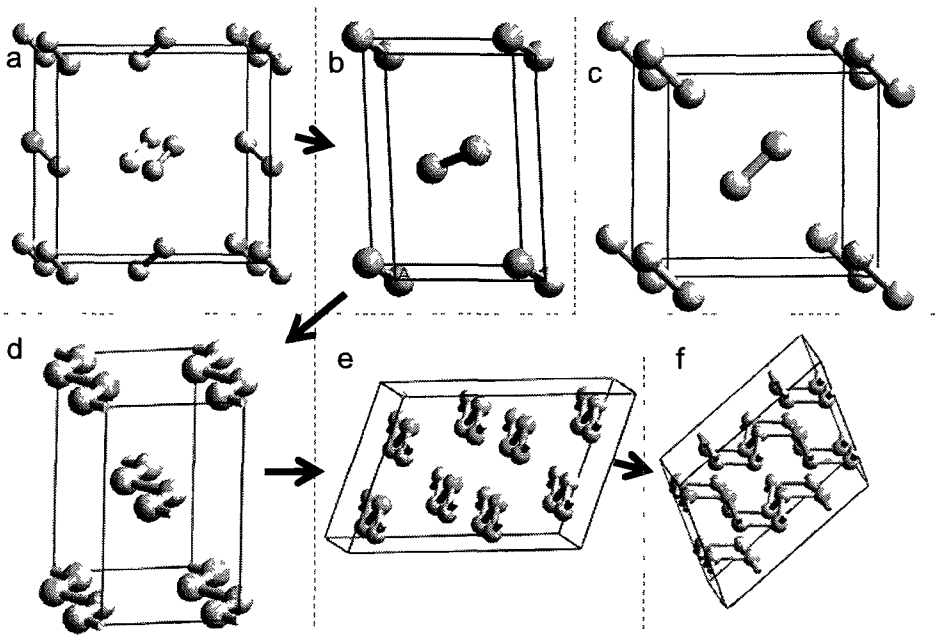


Figure 2.1. Unit cell representations of the structures observed when compressing a 2x2x2 supercell of the α-phase: a) α-N_2 phase, b) distorted γ-N_2 phase, c) γ-N_2 phase, d) zzCH phase, e) clustered zzCH phase observed at 250 GPa, and f) newly discovered LB phase. The arrows show the sequence of transformations observed from isotropic compressions.

As the pressure on this monoclinic γ-N_2 -like structure was increased, the ß lattice angle decreased from 115° to 105° until a transformation was observed at 170 GPa to a metallic chain-like phase reminiscent of the recently reported *C*mcm-symmetry zig-zag phase,[6] the zzCH phase, with atoms occupying general 8g Wyckoff sites of its *C*mcm symmetry cell. The transformation was accompanied by a volume drop of 12.3% and the zzCH-like structure is shown in Figure 2.1d. Note that the zzCH phase was initially recovered from similar computational experiments starting from the ε-N_2 phase.[6] As the

pressure was increased to 250 GPa, adjacent "zigzag" chains gradually aligned themselves in pairs and formed two-chain "clusters" along the [010] direction, as shown in Figure 2.1e. By 250 GPa the cell had sheared by 10° from the initial occurrence of the zzCH phase, as reflected by the increase in its ß lattice angle, while regressing to $P2_1/C$ symmetry. A final transformation was observed at 280 GPa, wherein the intramolecular bonds elongated and the intermolecular atomic distances moved within distances characteristic of a N-N single bond. This previously undiscovered, layered, single-bonded nitrogen structure, visualized in Figure 2.1f, is discussed in detail in the next section wherein it is referred to as the layered boat (LB) phase.

2.1.4 – New Layered Boat (LB) Phase of Polymeric Nitrogen

As the name of this newly discovered "layered boat" (LB) phase suggests, its structure is shown in Figure 2.2, this layered phase consists of fused six-membered rings in the classic boat conformation arranged in graphite-like sheets. Conversely, the other two layered polymeric nitrogen phases discussed in Section 1.2, the BP and A7 phases, are composed of layers of fused six-membered rings in the classic chair conformation. The LB crystal structure has a monoclinic $P2_1/m$-symmetry unit cell with its four nitrogen atoms distributed among 2e and 2e special Wyckoff sites. At 150 GPa, the lattice parameters are a=3.183, b=2.295, c=2.310, and ß=94.9°, and the atoms are generated from the N1=(0.00, 0.00, 0.00) and N2=(1.59, 0.93, 0.84) Cartesian coordinates of the lattice when a is projected along the x axis and b lies in the xy plane. The layers are stacked such that each layer is offset 1.90 Å from the layer beneath along the (001) plane, which holds each layer, as illustrated from the perspective shown in Figure 2.2c.

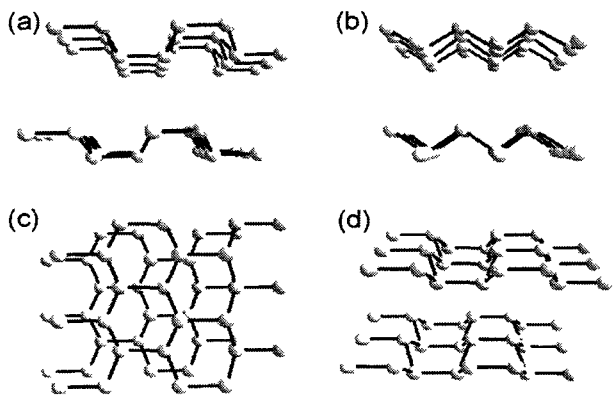


Figure 2.2. The crystal structure of LB from different perspectives: (a) along the [100] axis, (b) along the [010] axis, (c) perpendicular to the (001) plane, and (d) a perspective view to show the twisted boat N_6 structural motifs.

An enthalpy versus pressure phase diagram for LB and other selected phases of polymeric nitrogen between 10 and 200 GPa is shown in Figure 2.3, split into separate plots for the 0-80 and 80-200 GPa pressure ranges. The LB phase is higher in enthalpy than cg-N and zzCH throughout the entire pressure range and lies between the A7 and BP polymeric nitrogen phases up to 210 GPa. Beyond 210 GPa (not shown), LB becomes higher in enthalpy than A7 as well. The enthalpy curve for the ctCH chain-like phase, the *cis-trans* chainlike polymeric nitrogen phase labeled CH in Figure 2.3, crosses the LB curve at 60 GPa, becoming higher enthalpy at higher pressures.

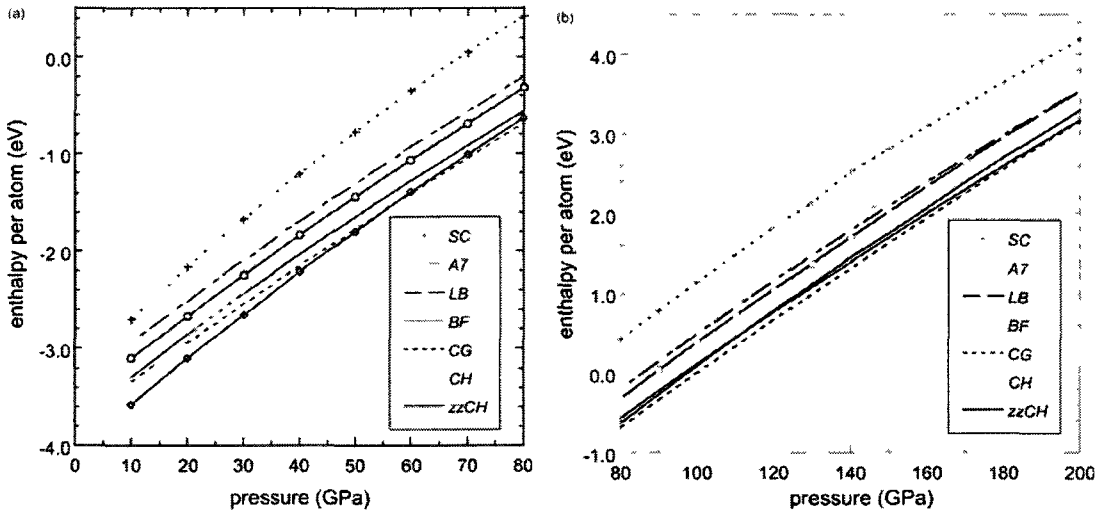


Figure 2.3. Phase diagram of select polymeric nitrogen phases. Enthalpies are given in enthalpy per atom relative the cg-N phase at 100 GPa. The curves are split into two pressure regions, a) 0–80 GPa and b) 80-200 GPa.

Further analysis of the layered polymeric nitrogen phases, the BP, A7, and LB structures, reveals an interesting relationship between the lp-N-N-lp dihedral angles (lp = lone pair) and their relative internal energies. Quantum mechanical calculations of small molecules of the form $R_2N\text{-}NR_2$ show the two minimum energy conformers correspond to the *gauche* and *trans* structures depicted in Figure 2.4. The *gauche* conformer, although it strays marginally from the conventional definition of a 60° "gauche" angle, adopts the ideal dihedral angle such that it minimizes the two-orbital/four-electron destabilizing interaction between the two lone-pair orbitals of adjacent N atoms. To illustrate the importance of this dihedral preference in polymeric nitrogen, note that the cg-N structure itself was formulated by designing a single-bonded solid nitrogen crystal such that all of its lp-N-N-lp dihedral angles were *gauche*. When the cg-N structure is optimized, these dihedrals measure ~107° and it has been repeatedly shown that cg-N is

indeed the lowest-enthalpy polymeric nitrogen phase in the literature between 50 GPa and 180 GPa.[3, 5, 6, 31, 32]

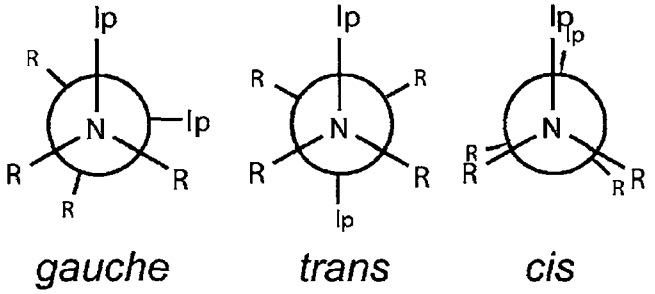


Figure 2.4. Gauche, trans, and cis dihedral angle depictions using Newman projections of the R_2N-NR_2 fragment (lp = lone pair).

The layered polymeric phases have six unique dihedral N-N angles in each N_6 ring. The *gauche/trans/cis* ratio among these six dihedral angles is 4:2:0 for the BP phase, 0:6:0 for the A7 phase, and 0:4:2 for the LB phase. From the observed dihedral trend for R_2N-NR_2 molecules, it was noted above that the *gauche* conformation is the most stable followed by the *trans* and *cis* conformations, respectively. It follows that the following order of internal energies is predicted for the layered, fused N_6–ring polymeric nitrogen phases: BP<A7<LB. The calculated internal energies of BP, A7, and LB agree with this trend at 20 GPa, specifically determined from first principles computations to possess energies of 0, 61, and 80 meV/atom respectively for BP, A7, and LB relative to the BP structure. Note that the pV term additionally contributes to the enthalpy at elevated pressures and bucks this trend, favoring LB over A7 at "low" pressures.

Ab-initio molecular dynamics were run to test the stability of the LB structure. At 300K, the structure was stable throughout 3 picoseconds of simulation time at 10 GPa but quickly disintegrated into smaller nitrogen molecular and zzCH chain-like fragments at 0 and 1 GPa pressures. When the temperature was reduced to 30K, the LB structure was

stable after 2 picoseconds of simulation time at all pressures. In addition, random atom-displacement tests were performed on the LB structure, a stability test in which the cartesian coordinates of randomly chosen atoms are displaced from their equilibrium positions by 1-8% of the N-N bond length and reoptimized.[6] From its 0 GPa optimized lattice coordinates, the LB structure successfully reoptimizes with 2% distortions but fails with higher distortions, indicating LB is metastable at low pressure only at low temperatures. Furthermore, phonon dispersion curves below 150 GPa for a structure similar to the LB structure in which the sheets are stacked directly on top of each other, further discussed in the Section 2.3 as bstr15, show negative-frequency phonon modes. This indicates that, following a shear deformation, there is a barrierless collective motion of atoms in the lattice that destabilize the structure and LB is not mechanically stable when the cell distorts in such a manner.

Given the highly strained environment native to diamond anvil cell experiments, such instabilities with respect to shear distortions limit the value of the LB phase as an alternative HEDM to cg-N polymeric nitrogen. Notably, however, the existence of this LB phase shows there are likely other polymeric forms of nitrogen, which are potentially mechanically stable, that remain unexplored. Should these phases have comparable enthalpies to cg-N, they would be relevant to the research community since these phases could be accessible via other reaction pathways, perhaps via high-pressure shock-wave or photoexcitation methods with high-pressure diamond anvil cells, and would be relevant should they possess high metastability at ambient conditions.

2.1.5 – Compression of the ε-N_2 molecular phase

In similar spirit to the computational compression of the α-N_2 molecular phase presented in section 2.1.4, the rhombohedral ε-N_2 molecular phase atomic coordinates and cell lattice vectors[7] were first optimized under variable cell conditions at 65 GPa. The experimentally derived/confirmed ε-N_2 phase crystal lattice places sixteen atoms in its rhombohedral unit cell representation and forty-eight atoms in its hexagonal representation, with 0.96 Å molecular N-N bond lengths. Large ionic forces were observed when the experimental ε-N_2 phase structure, in its hexagonal representation, was input into first-principles DFT calculations, exceeding 10 eV/Å^2 magnitudes directed along the bond of the N_2 molecules away from their centre of mass. Following full geometric relaxation using both projector augmented wave calculations of the ε-N_2 phase and all-electron calculations of isolated N_2 molecules,[32] the N-N bond lengths reoptimize to 1.106 Å. Presumably the repulsive coulombic interactions between the nitrogen atoms were too great to be reconciled by the pressure. Furthermore, the calculated vibrational frequencies of isolated 1.10Å and (constrained) 0.96Å N_2 molecules are 2340 and 3550 cm^{-1} respectively, as determined from the Gaussian '03 software package.[33] Experimentally, the observed Raman vibrational modes of the ε-N_2 molecular phase are measured at ~2300 cm^{-1},[7, 8, 11] showing that, based on the observed Raman vibrational modes and barring intermolecular vibrational coupling, 0.96Å N-N bonds lengths seem unlikely for the ε-N_2 structure. Regarding DFT's description of the ε-N_2 phase structure itself, its $R\bar{3}c$ symmetry is conserved upon optimization but two of the molecules fall slightly out of alignment, while remaining parallel, along the [110] projection.

Similar to those observed from compression of the α-N_2 molecular phase, structural transformations were observed when the pressure was increased incrementally by 5 GPa until 250 GPa. The first two transitions recovered the zzCH phase as previously reported by Mattson *et. al.*[6] Alternatively, when the ε-N_2 phase coordinates were instead recursively optimized at lower pressures, by decreasing the pressure on the native unit cell from 60 GPa in 5 GPa increments, an interesting transformation was seen at 45 GPa. The dinitrogen molecules adopt a close-packed structure by aligning themselves parallel to the same axis, roughly inducing a *C*mmm symmetry orthorhombic unit cell. The transformation is best described by referencing the 48-atom rhombohedral representation of the ε-N_2 unit cell, shown in Figure 2.5. The molecules generated from 36f general atomic sites, depicted in red in Figure 2.5, rotate such that they align themselves roughly parallel with the molecules generated from the 12c special atomic sites, depicted in blue. A similar structure was obtained from decompression of the ζ-N_2 molecular phase and will be discussed more thoroughly in Section 2.1.6.

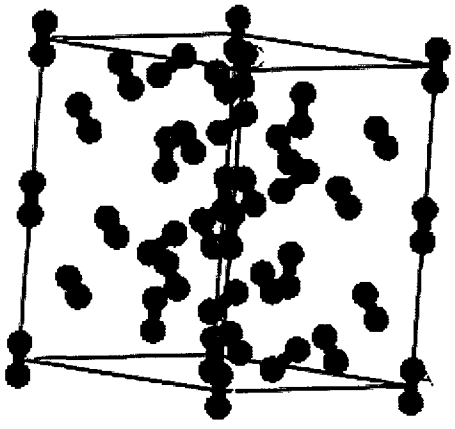


Figure 2.5. Rhomobedral represent of the ε-N_2 unit cell. Molecules composed of N atoms at 36f Wyckoff sites are coloured red and those from 12c special sites are coloured blue

2.1.6 – Compression of the ζ-N_2 molecular phase

The proposed lattice coordinates of Eremets *et. al.* for the orthorhombic ζ-N_2 molecular phase[7] were optimized under variable cell conditions at 80 GPa. The proposed ζ-N_2 phase coordinates, derived from the authors' interpretation of the x-ray diffraction pattern and symmetry elements common to the cg-N structure, contain eight atoms in its $P222_1$–symmetry unit cell distributed among 4e and 4e Wyckoff sites with 0.98 angstrom N-N bond lengths.

DFT electronic structure calculations on the proposed ζ-N_2 phase coordinates proved very difficult to converge, which was unexpected at high pressures and indicate an inherent problem with the structure. When allowed to optimize, the structure underwent significant relaxation by allowing the 2b and 2d site-centered N_2 molecules to align preferentially along the [100] axis in a disordered manner, crafting a seemingly amorphous *P*1 symmetry crystal structure. Further compression of this phase formed another seemingly amorphous phase at 195 GPa, composed of single dinitrogen molecules and elongated helical chains. Only through a controlled optimization scheme was the proposed ζ- N_2 structure recovered more-or-less intact at 80 GPa. First the cell was allowed to relax while the fractional atomic coordinates were kept fixed, this caused the cell to expand while increasing the N-N bond length to 1.09 Å, resulting in a $P222_1$–symmetry structure at 25 GPa. This was followed with a carefully controlled succession of fixed-volume optimizations, wherein the volume was decreased by ~4.0 $Å^3$ between each calculation by homogeneously scaling the atomic and cell coordinates, until the 80 GPa crystal structure was recovered.

The atomic coordinates of this distorted ζ-N_2 phase were more or less unchanged but the new cell parameters were found to be a=5.819, b=4.58, and c=5.039, deviating significantly from the native parameters proposed for the "parent" ζ-N_2 structure, reported to be a=4.159, b=2.765, and c=5.039. The large change in lattice parameters suggests either the proposed atomic coordinates were incorrect or gradient-corrected DFT calculations grossly misrepresent the high-pressure ζ-N_2 intermolecular interactions. Since the proposed ζ-N_2 coordinates have since been retracted due to inconsistencies with experiment,[34] the former seems to be the case. Interesting behaviors and transformations were seen nonetheless from these coordinates and are discussed below.

When the modified ζ-N_2 phase coordinates, pictured in Figure 2.6a, were optimized at lower pressures, namely by decreasing the pressure on the native unit cell from 80 GPa in 5 GPa increments, a new set of transformations were seen. At 45 GPa, the N_2 molecules adopt a close-packed structure, depicted in Figure 2.6b, and align themselves parallel to the [010] axis to induce a *C*mmm-symmetry orthorhombic unit cell similar to that observed from decompression of the ε-N_2 phase reported in Section 2.1.5. This phase, generated by occupying 4i Wyckoff sites, is closely related to previously studied $I4_2$/mmm symmetry[21] orthorhombic structures with only two molecules per unit cell. The authors originally proposed their $I4_2$/mmm structure as a strong candidate structure for the θ-N_2 molecular phase.[21] A similar structure was also published concurrently and independently of this work in ref 31, wherein the authors recovered a perfectly "aligned" structure[35] by optimizing a high-pressure nitrogen construct which mimicked the recently discovered high pressure ε-O_2 molecular phase of oxygen.[36, 37]

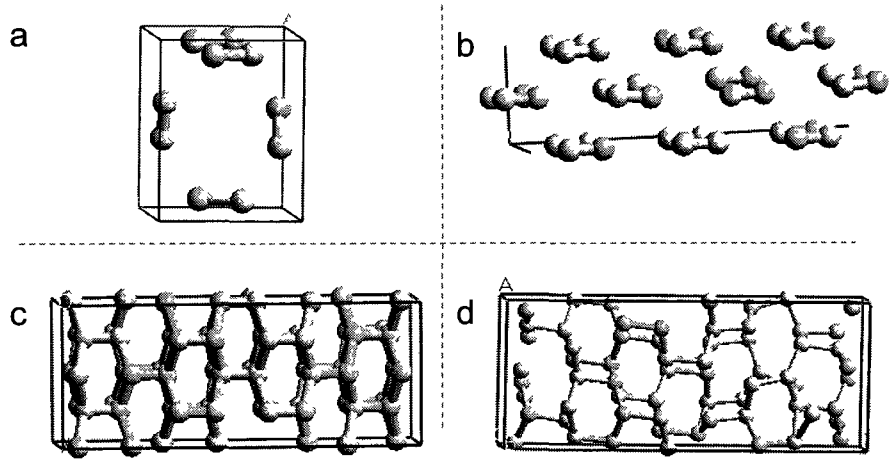


Figure 2.6. Simulation cell representations of the structures observed when compressing a 2x2x2 supercell of the ζ-phase: a) 1x1x1 cell of the ζ-N_2 phase, b) distorted Cmmm aligned N_2 phase, c) compressed phase seen at 225 GPa, and d) distorted cg-N structure observed at 280 GPa.

Upon compressing the intermediate "aligned phase" up to extreme pressures near 225 GPa, by again incrementally increasing the pressure in 5 GPa steps and reoptimizing the structure under variable cell conditions, another transformation to a new condensed nitrogen phase was observed. This hybrid molecular-polymeric condensed nitrogen structure forms such that the N_2 bond lengths increase to 1.26 Å while the intermolecular atomic distances reduce to 1.67 Å. This highly strained structure is shown in Figure 2.6c.

Upon further compression up to 280 GPa, instead of gradually succumbing to the metallic simple cubic structure typically observed for heavier Group V covalent crystals,[38] the alignment of N_2 molecules is distorted and trigonal pyramidal, three-fold, local bonding environments are observed. Although the structure itself was largely disordered and seemingly amorphous, as shown in Figure 2.6d, different structural motifs are evident in the structure. Surprisingly, the LB structural motif is present but, more interestingly, the cg-N structure is the dominant structural motif. Similarly amorphous-like, history-dependent structures have previously been reported[5] after the authors carried

out molecular dynamics simulations on sixty-four atom simulation cells of molecular nitrogen at temperatures approaching 10000 K and subsequently quenching the simulations. The authors proposed this was typical of the mysterious η-N phase of nitrogen, which was recovered in several experimental studies as an amorphous precursor to the cg-N structure.

More importantly, these results show that there is simple structure deformation of the aligned *I*mmm phase to the cg-N structure, namely a concurrent, concerted set of rotations of each molecule in the unit cell within its square diagonal planes to create the three-fold coordinate nitrogen environments. This transformation was also concurrently observed and discussed at length in ref. 31. Most importantly, these calculations also relate the proposed ε-N_2 and ζ- N_2 phase coordinates directly to the highly sought cg-N phase, albeit with defects, through a series of diffusionless transformations. The particular defects in the final polymeric structure likely stem from the numerical error associated with evaluation of the forces upon perfectly isotropic compression of the "aligned" orthorhombic *C*mmm-symmetry cell and, to a greater extent, the resulting imposed symmetry of the simulation cell. These collective effects prevent the necessary perturbations during the isotropic compression, causing the structure to fall into the pressure-induced, barrierless observed local minima instead of annealing to the cg-N structure. Simulated annealing would be required to remove the defects in the simulation cell and form a cg-N crystal; this observation could be used in an *ad hoc* manner to explain the experimental need for extreme temperatures to spark a crystalline polymeric nitrogen transition.

2.2 – Exploring Polymeric Nitrogen Using Structural Motifs as Building Blocks

2.2.1 – Introduction

In Section 2.1, novel nitrogen phases were recovered from compression of known molecular nitrogen phases. In an effort to design new polymeric nitrogen candidates from scratch, consider that a crystal is modeled as a solid-state material whose constituent atoms, molecules, or ions are arranged in an orderly repeating pattern extending in all three spatial dimensions. The nature of the interactions binding the constituents of a given crystal structure ultimately determine the physical properties of that crystal, all derived from its elastic and electronic interaction energies. By recognizing the structural motifs preferably adopted in the local chemical environments of a crystal's constituent components, one can readily use these motifs as building blocks to explore other low energy crystal structures. To quantify molecular crystals for example, one must understand the conformer distribution of the constituent molecule and the nature and effect of the intermolecular interactions on each constituent's stability.[39-41] The idea of incorporating preferred structural motifs in crystal structure prediction has been used extensively, dating back to the early 20th century.[42, 43]

Evidently, the constituent component of solid molecular nitrogen is the N_2 molecule itself, and the nature of the intermolecular interactions at different temperatures and pressures determines what polymorph is assumed. At low temperatures the N_2 solid structures are determined by weak quadrupole interactions or, as the pressure is increased, by simple close-packing arguments of the ellipsoidal N_2 molecules.[29] The electronic interactions from higher-pressure structures are dominated by the

intramolecular repulsion from highly localized electrons possessing high kinetic energy,[44] thus favoring delocalization/metallization of the crystal structure into polymeric or metallic polymorphs. In these polymorphic cases, the N_2 moiety is melded into the lattice and the constituent component of the crystal could be considered to be the nitrogen atom itself. As discussed in Section 2.1.4, one metric for ranking internal energies of polymeric nitrogen structures is to tabulate the distribution of *gauche/trans/cis* dihedral angles in the structure, the idea being to minimize destabilizing orbital interactions between lone pairs on adjacent nitrogen atoms (ie. to minimize destabilizing interaction between the crystal's constituent components).

In this section, the presented calculations reinforce the notion that in solid state chemistry it is often useful to investigate classes of polymorphs by exploiting common substructures, as both a means of predicting new structures and of finding reaction pathways between the polymorphs themselves. A precedent for this work is first set by further relating the layered polymeric nitrogen phases (BP, A7, and LB) to particular arrangements of the ctCH and zzCH chain-like phases. Building from these observations, it is then discussed how helical motifs play a role in determining the cg-N structure and how this concept relates the proposed ζ-N_2 molecular coordinates to the cg-N polymeric structure. Alternative helical reconstructions are then used to build another new phase of polymeric nitrogen, the Chaired Web (CW) phase.

2.2.2 – Computational Details

All the calculations reported in this section use Kohn-Sham DFT calculations[16, 17] with the PBE exhange-correlation functional.[18] The SIESTA programming suite[15] was used in all calculations derived from incorporating helical structural motifs into

polymeric nitrogen. The VASP simulation package[25] was used with the projector augmented wave (PAW) method[24] to treat the core states in all calculations with the ζ-N_2 molecular phase coordinates and in all the presented phonon calculations. As such, VASP was also used to verify all local minima recovered from the SIESTA simulations.

In all the VASP calculations, a plane-wave cutoff of 39 Ry was used and Brillouin-zone integration was performed using Monkhorst-Pack grids created from an 8x8x8 mesh. The phonon densities of states for the CW phase were calculated from the force constant matrix derived for a (64 atom) 2x2x2 supercell by the finite differences method.

All the SIESTA calculations used a Troullier-Martins norm-conserving pseudopotential[20-22] referencing a core $1s^2$ configuration with a 0.98 Å cutoff. The low cutoff was chosen to allow for greater transferability in the high pressure simulation environments. Custom SIESTA-type numerical doubled s, p_x, p_y, and p_z orbitals and d polarization orbitals[23] were used with a real-space mesh cut-off of 200.0 Ry. A 10 Å cutoff for k-point sampling was used to construct Monkhorst-Pack grids for Brillouin-zone integration. Similar polymeric nitrogen calculations with SIESTA have been reported.[6]

2.2.3 – Chain-like Building Blocks of the Layered Polymeric Nitrogen Phases

It was reported in Section 2.1.3 that during the compression of the α-N_2 phase, zzCH-like chains preceded the formation the LB phase (Section 2.1.4) as an intermediate between a distorted γ-N_2 phase and a "clustered" zzCH-like phase; the zzCH phase was part of the α-N_2 to LB transition. In this sense, one can consider the LB structure as a

forced alignment of zig-zag chains under high pressures to form inter-chain bonds. Specifically, the LB sheets could be characterized as a parallel arrangement of vertically aligned zig-zag chains that are tilted to the left and right in an alternating fashion. This construct is schematically represented on the left hand side of Figure 2.7, wherein the dotted lines represent the distortion needed to form the alleged inter-chain bonds.

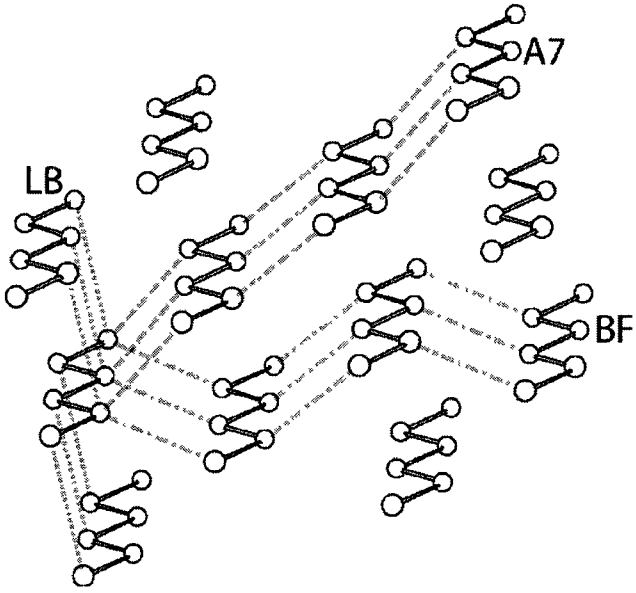


Figure 2.7 The zzCH phase with the possible inter-chain bonding drawn, BP (dashed-dotted lines), A7 (dashed lines), and A7 (dotted lines).

In this same spirit, the A7 and BP forms of polymeric nitrogen could also be portrayed as particular arrangements of zig-zag chains coming together and forming inter-chain bonds between them. The A7 phase could readily be classified as a series of zig-zag chains aligned parallel to each other, with all of them simultaneously tilted to one side. This schematic is illustrated with dashed lines along the diagonal of Figure 2.7, wherein the chains are depicted being tilted to the left. The BP phase can be recovered by arranging aligned zig-zag chains in a staggered zig-zag formation, as illustrated in the

bottom half of in Figure 2.7 by the chains connected via dashed/dotted lines, and then decreasing this multi-chain zig-zag formation's chain width.

These relationships by no means prove the layered phases are formed most favorably from the zzCH phase but, in a broader sense, one can rationalize that there are possible pathways to these polymeric structures from more elementary building blocks that do not require significant structural rearrangements. Gratifyingly, the BP, A7, and LB structures were all obtained by optimizing such high-pressure arrangements (~120 GPa) of zzCH chains at the GGA density functional level of theory. Consider that both the LB and BP phases could also be constructed using the *cis-trans* chains of the ctCH phase as building blocks instead, but those constructions will not be discussed here. Ideally, one would wish to identify such a building block for the cg-N polymeric nitrogen and use such structural motifs for further explorations of polymeric nitrogen. Although ctCH-like fragments can be rationalized from the cg-N lattice coordinates (not discussed here), the chain fragments are highly distorted in its optimized structure. Helical motifs, however, can readily be observed in cg-N and are discussed further in Section 2.2.4.

2.2.4 - Helical motifs in Polymeric Nitrogen

Helical motifs are a well-studied, recurring theme in a number of chemical systems. They are particularly prevalent in supramolecular chemistry for their importance in biological systems, perhaps most famously in DNA, and asymmetric catalysis.[45, 46] In such systems, these helical substructures are typically stabilized either by weak inter-strand bonding interactions, like hydrogen bonding, or by the steric demands of substituents on the backbone of the helix.[46] Helical motifs are prominent fixtures in monatomic covalent crystal structures as well; several high-pressure sulfur,

selenium, tellurium, and scandium covalent structures, for example, have recently been shown to favor helical motifs.[47-51] The trade-off of these helices is that they often introduce angle strain into covalent solids due to the geometrical requirements of the helix, thus the need for extreme pressures.[47]

With regards to nitrogen, one can rationalize that perfectly round helical structures will be unfavorable by considering the torsion and angle strain. However, distorted helical structures, similar to those observed in group-VIb elements,[49] are quite common in nitrogen allotropes and helical motifs can readily be seen in the cg-N structure, as shown in Figure 2.8. A left- or right-handed helix of nitrogen can readily be fashioned by constructing a single chain of atoms with all *gauche* lp-N-N-lp dihedral angles, this gauche preference in polymeric nitrogen has been discussed in Section 2.1. Of course, nitrogen cannot form a single stable helix on its own, the resulting unsaturated valences would evoke the formation of partial double bonds, change the hybridization at the nitrogen centers, and likely ultimately transform to one of the chainlike phases (either zzCH or ctCH).

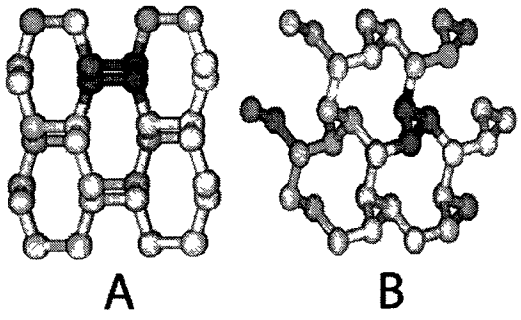


Figure 2.8. Highlighted helical motifs (coloured in red) in the cg-N structure, a) four-fold helix viewed along the [100] axis and b) three-fold helix viewed along the [111] axis.

To create pure nitrogen helical substructures, one has to either cap the unsaturated valences or bring adjacent helices close enough to create interhelical bonds. It follows

that the single-bonded cg-N structure, a single-bonded network of nitrogen atoms with all *gauche* dihedral angles, must contain these helical substructures. Indeed the high symmetry of cg-N includes several screw axis symmetry elements stemming from these helical chains; examples of four- and three- fold helices along the [100] and [111] projections are highlighted in Figure 2.8 and an eight-fold motif along the [100] projection (not shown) can also be observed along the [100] direction. It stands to reason that the reaction coordinate connecting the cg-N structure to its molecular precursor, the ζ-N_2 molecular phase[7], must generate this structural motif.

Connecting the cg-N and ζ-N2 polymeric and molecular phases

By analyzing the published ζ-N_2 molecular phase coordinates available at the time,[7] it was noted that, by rotating two of the molecules in the unit cell, one can fashion an eight-fold helix as shown in Figure 2.9A. This rotation about the two molecules' center of mass effectively "inverts" their orientation, setting up the desired connectivity for the helical motif. Using this "inverted" eight-fold helical construct, shown in Figure 2.9C as a transformation of Figure 2.9B, as a starting geometry for a variable-cell first-principles calculation at 80 GPa, the structure reoptimizes to a high-energy, primitive, monoclinic molecular phase with $P2_1/C$ symmetry. When optimized at 110 GPa, the pressure required to experimentally crystallize cg-N,[52] the helical construct created from the ζ-N_2 phase optimizes directly to the cg-N phase.

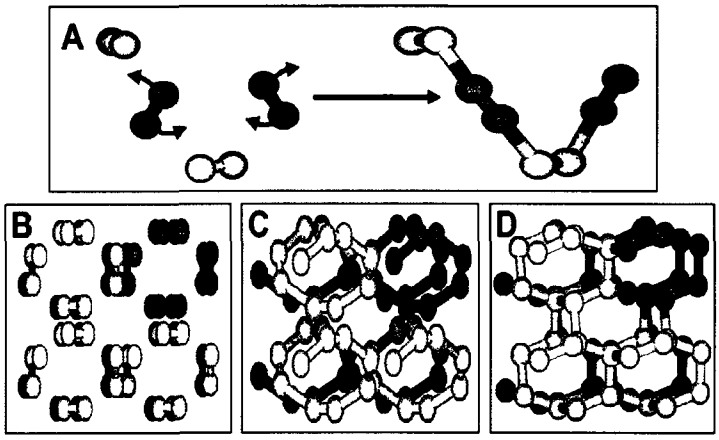


Figure 2.9. Relation of ζ-phase to cg-N by generation of helical motif. (A) Flipping or "inverting" two out of four N_2 molecules in the primitive unit cell of the ζ-N_2 phase (left) results in an eightfold helical chain structure (right). (B) 2x2x2 supercell of ζ-N_2, (C) 2x2x2 supercell of the "inverted" ζ-N_2, and (D) 2x2x2 supercell of cg-N structure. An 8-fold helical chain is highlighted in red.

The original motivation behind finding this physical connection to cg-N phase was to provide a viable reaction pathway connecting the ζ-N_2 molecular phase to the cg-N polymeric phase. Since the helical construct optimizes directly to cg-N, the molecular rotation used to create the phase constitutes such a pathway. Using the proposed ζ-N_2 phase coordinates and the aforementioned helical construct coordinates as endpoints, a rough reaction coordinate of the molecular to cg-N transformation was constructed by generating intermediate structures from a linear interpolation between the structures. This was done by rotating the molecules centered at the 2d special Wyckoff sites in 8° increments about the [010] axis, taking care to preserve the N-N bond lengths. Energy evaluations of these intermediate structures, shown in Figure 2.10, revealed a rough estimate of the energy barrier to formation to be ~1.04 eV/atom. The optimized cg-N structure lies 2.01 eV/atom below the maximum of the presented reaction coordinate.

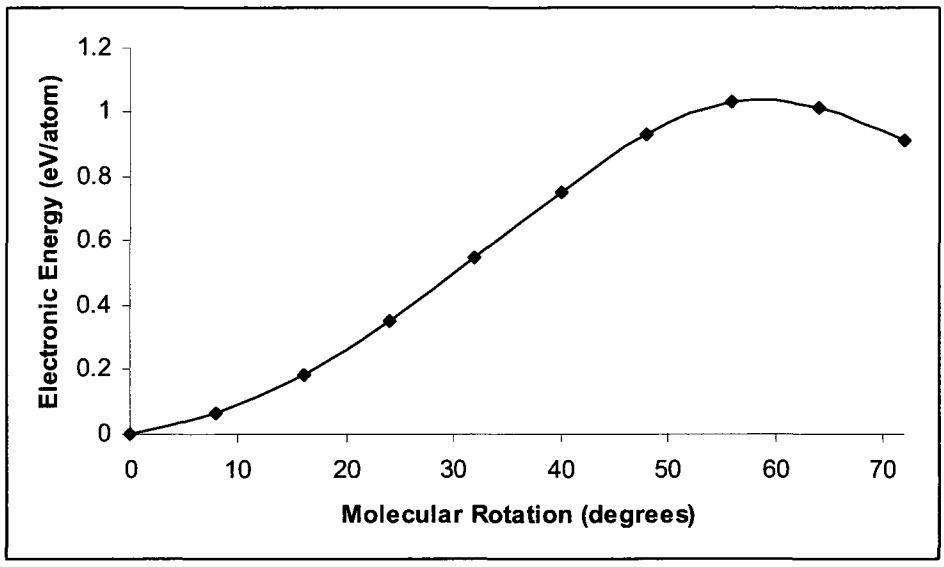


Figure 2.10. Plot of internal electronic energy of structures interpolated between ζ-N_2 and the "inverted" ζ framework. Structures were derived by rotating the 2d-centred N2 molecule about the [010] axis by 8°.

The 1.04 eV/atom energy barrier is markedly higher than the ~0.9 eV/atom barrier to the formation of cg-N calculated by relating the cg-N structure to other selected high pressure group V and VI allotropes through deformations of the simple cubic structure.[1, 3] When another pathway is constructed by simultaneously rotating the 2b site-centered molecules about the [100] axis in 3° increments along the proposed reaction coordinate, the desired helicity is preserved and the rotation barrier is reduced to 0.89 eV/atom, matching the barrier relating cg-N to β-O_2 through the simple cubic structure.[3]

Given the difficulties in modeling the proposed atomic coordinates for the ζ-N_2 molecular phase, however, and their subsequent retraction by the authors, further work using this reaction pathway was ceased. Fundamentally, this lays a possible foundation for relating molecular phases to polymeric phases through structural motifs and probing energy barriers when the true ζ- N_2 phase coordinates are known.

2.2.5 - Deriving new low-pressure phases from the helical motif

The successful optimization of an eight-fold structural motif, the "inverted" ζ-N_2 structure discussed in Section 2.2.4, to the highly sought cg-N structure suggests that other stable polymeric nitrogen phases could be found by reoptimizing other arrangements of helical nitrogen motifs. Encouragingly, geometry optimizations of properly arranged fourfold and threefold helical chains also result in the formation of cg-N at 120 GPa.

By reoptimizing adjacent six-fold helices, a symmetry element not seen in the cg-N structure, a new single-bonded polymeric nitrogen structure, pictured in Figure 2.11, is formed. This phase, herein referred to as the chaired-web (CW) phase, consists of a three dimensional network of interconnected six-membered rings which adopt the classic chair conformation. During the optimization, the six-membered rings are formed in the voids between the helices to accommodate the nitrogen lone pairs. The interplay of the helical-chain and N_6-ring structural motifs in CW is reminiscent of the balance of *cis-trans* chains and N_6-ring motifs in the layered BP, A7 and LB phases.[53]

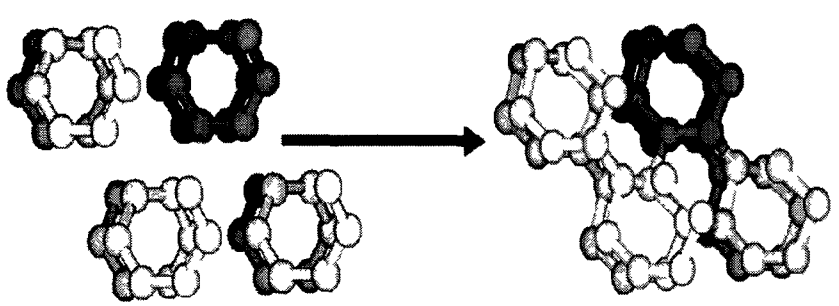


Figure 2.11. Depiction of the arrangement of six-fold helices (left) optimizing to the CW structure (right).

The CW structure, shown in Figure 2.12, can be represented by a rhomobohedral primitive unit cell. At 28 GPa its $R\bar{3}m$ symmetry unit cell has lattice parameters of a=3.5 Å and ß = 99.17°. The six atomic positions are distributed over 6c special Wyckoff sites, recovered from the (0.8756, 0.8756, 0.3206) fractional coordinates. There are two types of nitrogen bond distances in the crystal lattice, a shorter N-N distance within each N_6 ring and a slightly longer N-N distance between two adjacent N_6 rings. At 28 GPa these distances are 1.36 Å and 1.45 Å respectively. As expected from its construction, there is a preference for the gauche lp-N-N-lp dihedral angle in the CW phase, possessing a 2:1:0 ratio of *gauche-*, *trans-* and *cis-* dihedral angles in its crystal lattice.

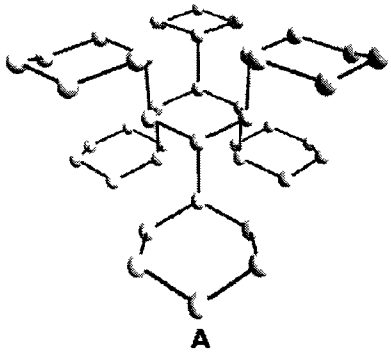


Figure 2.12. A perspective view of the CW phase.

A phase diagram portraying the relative enthalpies of the CW phase to the cg-N, BP, and A7 structures between 0 and 65 GPa is shown in Figure 2.13, demonstrating how close CW is to cg-N in terms of enthalpy. The most striking feature of this CW phase is its lower enthalpy, and respective higher thermodynamic stability, at ambient pressure compared to cg-N phase at low pressure. The structure of CW is, like other purely single-bonded polymeric nitrogen phases, governed to a large extent by the repulsion of the lone pairs. In CW, the inherent stability of the initial *gauche* helices and the newly

formed helices of reverse handedness is offset by the repulsion of the lone pairs from neighboring helices. Importantly, however, the lone pairs point into large void spaces, thereby reducing the repulsion. At high pressure these large voids are thermodynamically unfavorable because of the associated enthalpic penalty but, at low pressure, the balance between these two contradictory influences of the lone pairs is such that CW is thermodynamically comparable to the cg-N structure.

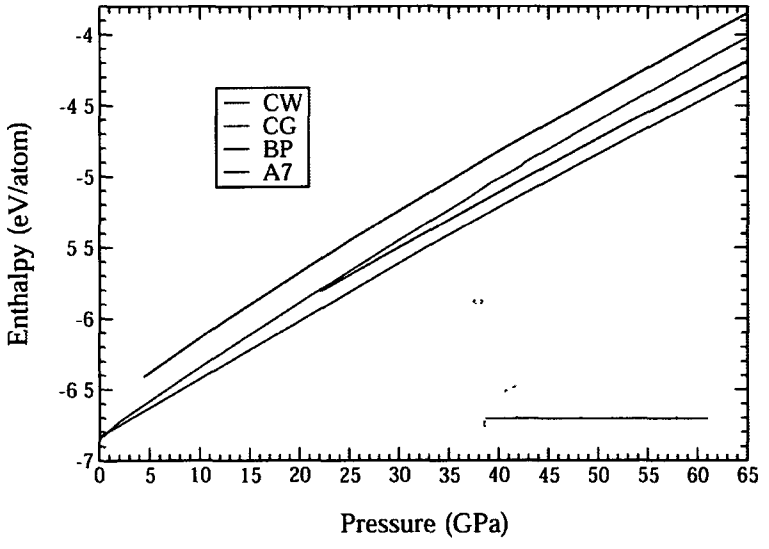


Figure 2.13. Phase diagram of select polymeric nitrogen phases. Enthalpies are given in enthalpy per atom relative the cg-N phase at 120 GPa. The inset in the lower right corner enlarges the low-pressure region (between 0 and 5 GPa).

The inset in Figure 2.13 shows the CW phase actually becomes enthalpically favored over cg-N below 0.75 GPa, however the structure also transforms at this threshold such that the six-membered rings flatten into benzene-like molecular structures. Each N_6 ring separates from its neighboring ring structures to resemble a recently proposed Na/N_6 molecular solid. Thus, the structure seems to gain its enthalpic advantage over cg-N at ambient conditions by becoming a molecular solid. Nonetheless,

when it exists as a polymeric structure the CW structure is still lower in enthalpy than the A7 phase and is competitive with the BP and cg-N phases at higher pressures. Phonon dispersion curves, shown in Figure 2.14A, show the phonon frequencies stay positive over all of q-space (the x-axis) which confirms the polymeric structure is indeed mechanically stable at ~15 GPa. Band structure calculations show the CW phase is an insulator at low pressures with a calculated band gap of ~5 eV, as shown in Figure 2.14B.

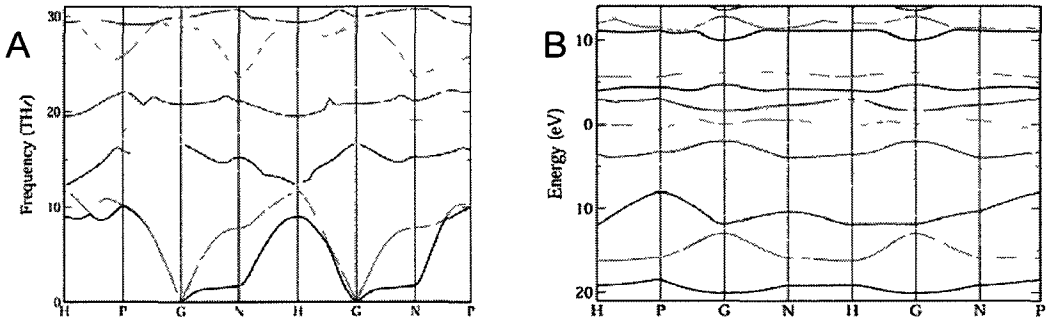


Figure 2.14. A) Phonon dispersion curves for CW at 15 GPa and B) electronic band structure referenced to the Fermi level at 6.4 eV.

It has been shown that structural motifs can be used to identify new, relevant polymeric nitrogen structures. The newly discussed CW phase was recovered by optimizing the geometries of closely packed six-fold helices and it can also be noted that the previously studied BP, A7, LB, and cg-N structures can all be recovered from optimizing particular arrangements of either zig-zag chains or helices. At the time of its discovery, the CW structure was the only single-bonded polymeric nitrogen structure that was enthalpically competitive with cg-N at 15 GPa, but it does transform at ambient pressure into a molecular solid. The CW phase may prove useful, however, in ultra-high pressure synthesis as a means of synthesizing polymeric nitrogen via alternative

pathways. For example, one may proceed by first synthesizing a polymeric phase such as CW from a symmetry-related molecular phase and annealing to an alternative polymeric structure like cg-N.

2.3 - Systematic Method to Explore Polymeric Nitrogen from Peierl's Distortions of the Simple Cubic Structure

2.3.1 – Introduction

Thus far, the search methods discussed for polymeric nitrogen have all involved propagations from a well-defined initial structure, be it a known solid molecular structure of a polymeric precursor (section 2.1), a mimic of another covalently bound elemental allotrope, or a structure designed under the premise of optimizing structural motifs of local nitrogen environments (section 2.2). Alternatively, in order to explore the PES without the bias inherited from chemical intuition, systematic methods need to be developed to explore the entire potential energy surface pertaining to high pressure nitrogen. In this section, such a systematic search method that effectively explores the most physically relevant structural candidates for nitrogen is discussed. This method deviates substantially from the more generic frameworks used to navigate configurational space, such as the "quotient graph" technique[54] or tiling surfaces of negative curvature methods[55] used to explore the carbon PES, and takes advantage of polymeric nitrogen's relation to a simple cubic (SC) reference structure.

All of the single-bonded metastable nitrogen phases reported prior to 2007 can be considered Peierls-like distortions of the SC structure,[3] and such relations have been used to estimate energy barriers for solid-nitrogen phase transitions. This convenient geometrical construct hints at a search method capable of systematically exploring all

single-bonded metastable candidate structures for polymeric nitrogen. In this section, such a method is described that employs modern first-principles calculations with the geometrical model of crystal structures developed by Wells[42], Adams[43], and Burdett.[56, 57] The method functions by exploiting the connection of high-pressure nitrogen phases to a SC reference structure, which has historically been used to study transformations of particular polymeric nitrogen phases into the SC structure and then, ultimately, to another nitrogen phase. The ingenuity behind the search method in this section is that the SC structure is instead used as a starting point to cycle though all the unique structural relations that generate generic structures consistent with polymeric nitrogen materials. The remainder of this discussion will go on to describe this systematic search method in more detail and discuss the results when the method is applied to an eight atom SC reference structure. It is shown the procedure successfully recovers all of the previously reported phases of polymeric nitrogen among a number of other candidate structures.

2.3.2 – Method

Curious as it may be to use a SC reference structure for nitrogen allotropes, this structure emerges as a natural reference structure at high pressures for other Group V elements as well. Specifically, both arsenic and phosphorous have been observed to transform to SC at high pressures,[38] and, equally important to this work, the reverse transformations are seen when the pressure is gradually released. The driving forces behind these reverse transformations are attributed to either Peierls/Jahn-Teller distortions and/or effects triggered by s-p orbital hybridization. In pure nitrogen, this lowering in energy of the SC structure could be achieved by changes in the local hybridization (to sp^3), symmetry breaking, and splitting of degeneracy at the Fermi level,

all are notably characteristic of a Peierls deformation. The physical change, of course, corresponds to the nitrogen atoms' assumption of trigonal pyramidal local geometries in three-fold coordination environments.

The systematic search method proposed herein stems from applying these Peierls-like deformations of a SC reference structure to generate single-bonded structures of nitrogen. Consider that each atom in a SC reference structure has six nearest neighbors that are equidistant from it, there are two along each cartesian coordinate (if the conventional expression of the SC structure is used), and is therefore in a six-fold coordination environment. To obtain relevant nitrogen structures from an SC lattice, one must apply a distortion that forms only three covalent bonds, thereby breaking three of the "bonds" in the SC structure. An appropriate distortion can be made by moving the atom along the body diagonal of the primitive SC cell, as illustrated in Figure 2.15b, thereby shifting the nitrogen atom to its native trigonal pyramidal environment. The distortion would cause the nitrogen atom to remain bonded with its three neighbors that border the primitive SC cell it moves into, while breaking the "bonds" with its three neighbors bordering other cell images. The nitrogen bonding of this distorted structure can be described in the original SC framework by labeling the SC unit cell edges as either "bonds" and "nonbonds" to describe the connectivity after the Peierl's distortion.[56] These labeling patterns are depicted in Figure 2.15a as solid lines for "bonds" and dashed lines for "nonbonds". By ensuring that such labels are consistent with local trigonal pyramidal environments for each nitrogen atom and then simultaneously applying the distortions to all atoms in the 8 atom SC, one can readily generate crystal structures with the desired three-fold connectivity. Figure 2.15c, for example, shows the structure

obtained from distorting the cell depicted in Figure 2.15a. In this particular example, the BP phase is generated.

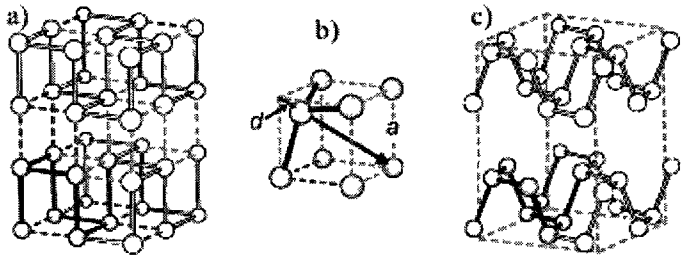


Figure 2.15. Depiction of relating SC reference structure to polymeric nitrogen structure. a) 2x2x2 SC structure with solid lines describing the desired three-fold connectivity of the nitrogen crystal. The thick black lines show the reference 8 atom cell. B) Shows the distortion performed on a single atom in the unit cell, and C) shows the structure obtained when all such distortions are applied to the reference structure.

A complete search of all pure single-bonded phases of polymeric nitrogen can be attained by mapping all possible SC distortions that promote three-fold connectivity of its constituent atoms. To demonstrate, it will be shown in this section that all the 8-atom unit cell polymeric nitrogen structures known upon inception of this work can be recovered from exhaustive searches of all possible Peierl's-like distortions on a 2x2x2 SC structure. First, one must map out all the possible ways to label the edges of the 2x2x2 SC simulation cell as "bond" or "nonbond". The complete set can be mapped using the combinatorial analysis procedure of Burdett and McLarnan,[56, 57] which itself is a generalization of Polya's enumeration theory.[58] This set, three of which are shown in Figure 2.16, includes connectivity patterns consistent with all the previously reported threefold-connected allotropes of nitrogen, namely the A7, BP, cg-N, and LB structures.

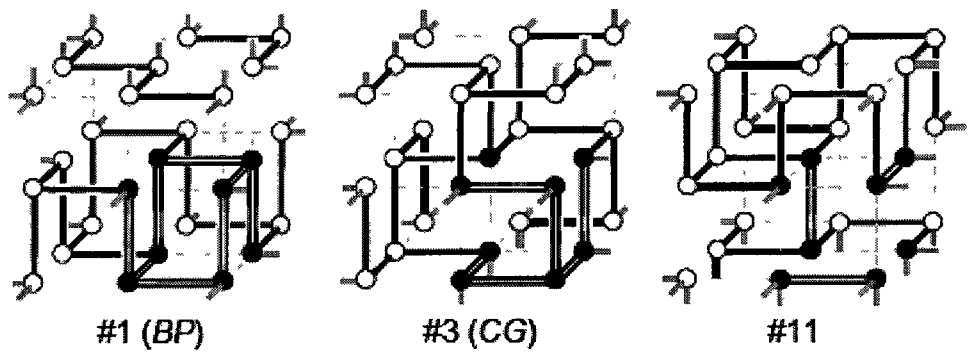


Figure 2.16. Connectivity diagrams of a SC reference structure consistent with threefold coordination. Shown are the diagrams corresponding to BP, cg-N, and a new low enthalpy phase, bstr11.

In total, there are 36 unique mappings (out of possible 4096 mappings) one can derive to describe the complete set of deformations for an eight-atom simple cubic cell of nitrogen, these are discussed at length by Burdett *et.al.*[56] By carefully scanning the distortion parameter, the distance *d* in Figure 2.15b that defines the distance each atom is perturbed along a primitive SC cell's body diagonal, of each Burdett mapping and incrementally optimizing the atomic and lattice coordinates from first-principles calculations, one can obtain the three-fold connected nitrogen crystal structure corresponding to each of the 36 connectivity diagrams, namely each "Burdett structure".

All the calculations used Kohn-Sham DFT[16, 17] with the PBE exhange-correlation functional.[18] The VASP simulation package[25] was used with PAW method[24] to treat the core states. A plane-wave cutoff of 39 Ry was used and Brillouin-zone integration was performed using Monkhorst-Pack grids created from a 8x8x8 mesh. The phonon densities of states were deduced from a force constant matrix derived from a (64 atom) 2x2x2 supercell, calculated from consecutive displacements on each atom in the unit cell by 0.02 Å. A larger supercell could not be used for the phonon calculations with the available computational resources, so care must taken that, in some cases, these periodic

effects may give rise to false, low-lying negative frequencies. However, the affirmative outcomes of the phonon dispersion tests should be treated as conclusive.

Specifically, the following computational procedure was used to produce a series of calculations to retrieve each Burdett structure. First, for a given Burdett mapping, the lattice constant is set to 4.1 Å and a sequence of single-point calculations are submitted to scan for the optimal global distortion parameter, d, which minimizes the calculated total energy when applied to all the atoms in the SC simulation cell. Once the optimal distortion parameter is determined, the cell is allowed to relax under the constraints of fixed volume and fixed fractional atomic coordinates. Finally, full lattice relaxation of the unit cell and atomic coordinates is allowed with a fixed volume constraint. The structure is then checked to confirm whether the structure is both fully optimized and possesses the desired three-fold connectivity. If such a structure is not recovered, then the procedure is repeated by scanning the lattice constant downwards in increments of 0.05 Å until an appropriate fully optimized structure is obtained. If the procedure fails to recover any polymeric nitrogen structures by the time the lattice constant reaches 3.0 Å, then the Burdett mapping is discarded.

2.3.3 – Results of the systematic search for single-bonded nitrogen allotropes derived from a 2x2x2 SC reference structure

In total twenty-four unique crystal structures were found; stable local minima could not be located for the remaining twelve Burdett mappings. For clarity, these structures will be referenced using the label "bstr", standing for "Burdett-derived structure". For example the nitrogen structure derived from the eleventh mapping will be labeled as the bstr11 structure, the eighth as bstr8, and so on. The relative enthalpies,

electronic energies, and volumes of all twenty-four phases are presented in Figure 2.17. As expected from the literature,[3, 5, 6, 32] the cg-N structure, recovered as the bstr3 structure, is found to be the lowest enthalpy single-bonded phase of nitrogen below 120 GPa. Encouragingly, all of the previously reported single-bonded phases that have been related to Peierls-like distortions of an eight atom SC unit cell were recovered from the procedure as well, namely the BP phase as bstr1, the A7 phase as bstr7, and a sheared LB phase as bstr15. In total, thirteen of the Burdett structures were at least equal or lower in enthalpy than the sheared LB phase at 100 GPa, with BP and A7 among them, and there are ten such structures comparable in enthalpy to A7 at 10 GPa. This confirms that a number of energetically-competitive nitrogen allotropes remained unexplored on nitrogen's PES. Two new particularly low-enthalpy phases emerge as alternative single-bonded nitrogen allotropes to cg-N, the bstr11 and bstr22 structures.

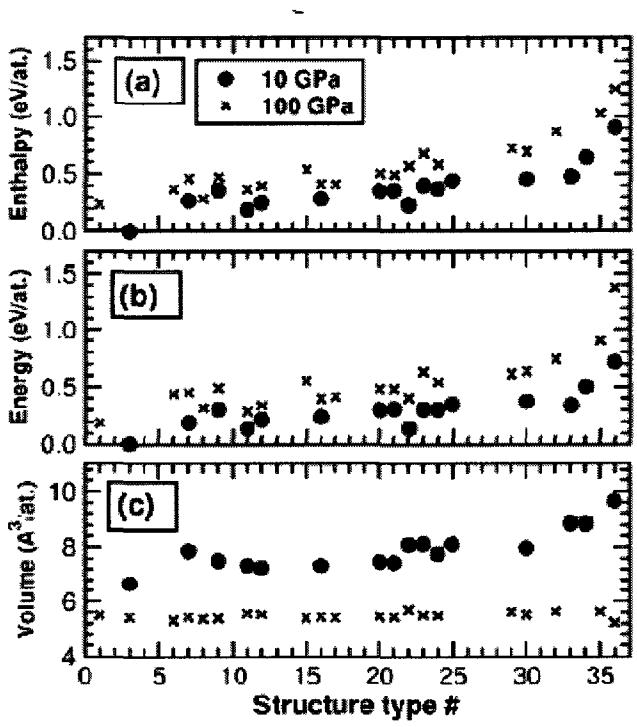


Figure 2.17. Calculated (a) enthalpy, (b) total energy, and (c) volume per atom vs. structure type # at pressures of 10 GPa and 100 GPa. The enthalpies and energies are given relative to the cg-N phase at the corresponding pressure. Structure #1,3,7, and 15 correspond to BP, cg-N, A7, and LB, respectively.

The calculated phonon density of states of all twenty-four structures, over a range of pressures, is summarized in Figure 2.18. Fifteen of the SC-derived phases have negative modes and are therefore not mechanically stable at the gradient-corrected DFT level of theory. Notably a sheared version of the layered LB phase, derived from Burdett mapping #15, is not mechanically stable at any of the calculated pressures. In this particular case, the LB layers are stacked directly on top of each other instead of being slightly offset as they are in the LB structure discussed in Section 2.1.4, whose layers were offset 1.90 Å along the plane. The results in Figure 2.18 also suggest the BP and A7 structures, corresponding to bstr1 and bstr7, are not mechanically stable. However, the faintly negative frequencies seen in BP at 60 GPa and A7 occur slightly above the

gamma point. Further phonon calculations on 3x3x3 supercells of the parent structures at 38 GPa and 13 GPa, respectively, revealed that these negative frequencies are likely the result of weak inter-sheet coupling and floating point rounding errors. Although a 2x2x2 supercell contains 64 atoms, the nature of the weak dispersive forces between the layers of the BP and A7 structures is delicate and must be handled with great care. The eigenvectors corresponding to the larger of the negative frequencies were found to correspond to a shearing motion of the layers across each other and should therefore be taken as inconclusive.

Of the remaining phases, bstr22, one of the lowest enthalpies structures at 10 GPa, was found to be mechanically unstable below 55 GPa. On the other hand, the bstr11 structure, the 2nd lowest enthalpy phase at 10 GPa beyond cg-N, is shown to be stable below 20 GPa. The bstr8 structure also shows remarkable stability, registering all-positive phonon modes between 25 and 200 GPa.

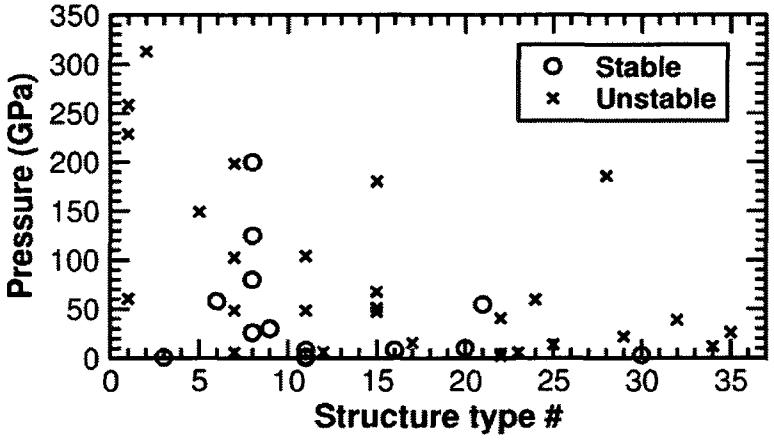


Figure 2.18. Results of the phonon spectra based mechanical stability tests. The mechanical stabilities of the 27 Burdett structures derived from the original mappings from 1-36 (x-axis) are shown to be stable (open circle) or unstable (x) with respect to pressure (y-axis).

2.3.4 – New Low Enthalpy Phases of Polymeric Nitrogen at High Pressure

As shown in section 2.3.3, several promising, new, low-enthalpy, threefold-connected nitrogen phases were recovered using the featured systematic search method that relates nitrogen to Peierl's distortions of SC reference structures. Subsequent stability tests using phonon dispersion curves confirmed that, among these candidates, only the high-symmetry bstr11 structure, shown in Figure 2.19a, and the low symmetry bstr8 structure, shown in Figure 2.19b, are particularly relevant, based on enthalpic properties and mechanical stability tests at low pressures.

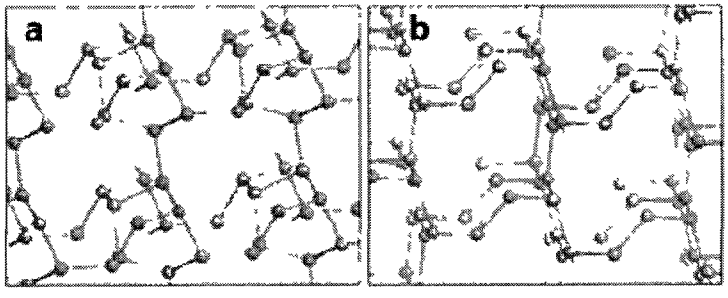


Figure 2.19. Extended periodic structures of two new low energy and metastable phases, (a) bstr11 and (b) bstr8.

The rhombohedral, *R*32-symmetry, unit cell of the bstr11 structure has lattice parameters of a=3.945 and α=93.9° at 0 GPa with its atoms distributed among 6f and 2c Wyckoff sites. The crystal structure is derived from N1=(0.8551, 0.1575, 0.3172) and N2=(0.1859, 0.1859, 0.1859) fractional cell coordinates, respectively, at zero pressure. The bstr11 structure also happens to possess the 2nd most *gauche* lp-N-N-lp dihedral angles of the proposed candidate structure, behind only the cg-N structure. Its curious structure can be described as intersecting, interconnected, orthogonal sets of ctCH chains directed along each of the three coordinate axes. The inter-chain bonding creates an

interconnected web of N_8 wedge-like motifs that are, incidentally, well studied in three-dimensional networks.[42]

The triclinic unit cell of the bstr8 phase has lattice parameters of a=3.63, b=3.65, c=3.66°, α=73.4°, ß=104.2°, and γ=98.8°, with its constituent atoms spread over four 2i general Wyckoff sites at 0 GPa. The zero-pressure crystal structure can be derived from N1=(0.9594, 0.05122, 0.7927), N2=(0.5736, 0.8882, 0.7051), N3=(0.6103, 0.5199, 0.6833), and N4=(0.9614, 0.4380, 0.6932) fractional cell coordinates. Like bstr11, the bstr8 can be considered as being similarly shaped by sets of ctCH chains but along just two out the three axes, hence the lower symmetry. The bstr8 phase could also be considered as an interconnected LB structure, best illustrated by the perspective shown in Figure 2.19b wherein the "layers" run along the diagonal of the figure.

A phase diagram portraying the relative enthalpies of bstr11 [New(#8)], bstr8 [New(#11)], bstr1 (BP), and bstr7 (A7) to the cg-N (bstr3) phase from 0 to 300 GPa is shown in Figure 2.20. The cg-N phase , the zero on the y-axis, clearly remains the lowest enthalpy structure over the whole pressure range, from 0 to 300 GPa; when coupled with its pronounced mechanical stability from the phonon calculations shown in Figure 2.18, the reported stability of cg-N is similar to that observed elsewhere.[59] Beyond the cg-N structure, the phase diagram shows that the newly-discovered bstr8 and bstr11 structures have energies that are comparable or lower than those of the previously proposed structures bstr1 (BP), bstrr7 (A7), and bst15 (LB), as shown. In fact, both the bstr8 and bstr11 structures become lower enthalpy than the aforementioned phases as the curves approach ambient conditions.

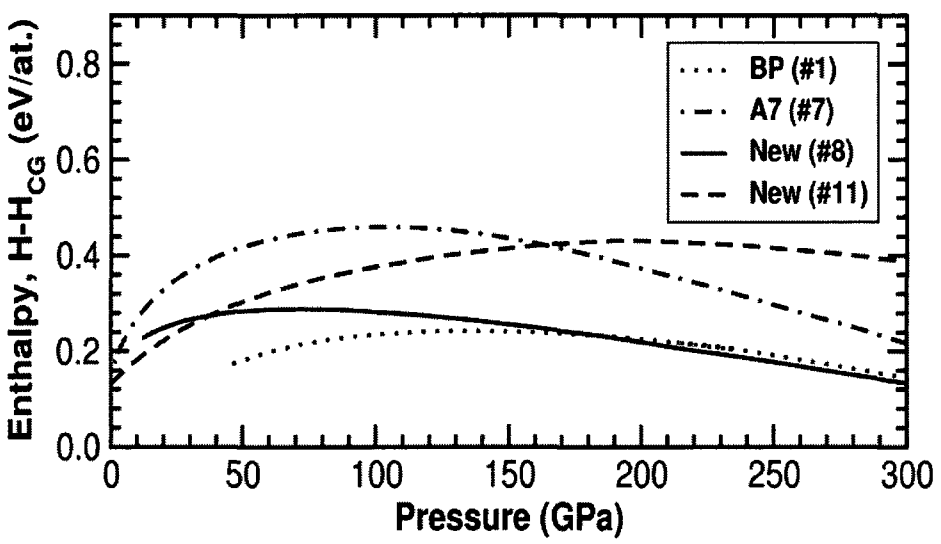


Figure 2.20. Phase diagram of select polymeric nitrogen phases. Enthalpies are given in enthalpy per atom relative to the cg-N phase.

A systematic method that recovers all pure single-bonded nitrogen allotropes that can be referenced to a distorted eight atom SC cell has been discussed. The method successfully recovered the cg-N, A7, BP, and LB polymeric nitrogen structures, and, among other candidate structures, two particularly promising novel nitrogen structures have been presented, labeled bstr8 and bstr11 in this work. In addition, note that the CW structure, discussed in Section 2.2.5 can also be obtained using this methodology by applying this systematic method to an appropriate reference SC structure. The eight-atom simple-cubic reference cell demonstrated here is simply incompatible with the six-atom CW unit cell; rather, a significantly expanded 48 atom (or larger) simple cubic reference structure would have to be employed.

2.4 – Conclusions

A number of computational approaches to investigate the high-pressure potential energy landscape of polymeric nitrogen have been presented. First, a thorough treatment

of computational pressure-application experiments on the α-N_2, ε-N_2, and ζ-N_2 solid molecular nitrogen phases has been discussed. Two novel nitrogen phases, namely the LB and Cmcm-symmetry "aligned" phases, have been recovered from the α-N_2 and ε-N_2/ζ- N_2 coordinates respectively. The "aligned" phase has been related directly to the cg-N structure, through slight and concerted molecular rotations that occur when it is compressed to high pressures, and therefore provides a hypothetical pathway from molecular nitrogen to its highly sought polymeric allotropes. It was noted that, when compressed, the resulting polymeric framework is best characterized as being amorphous due to its lack of symmetry, but clear structural motifs were apparent in its makeup. Specifically, the LB and cg-N motifs were prominently observed throughout its compression. These results reinforce the hypothesis proposed by Mattson *et. al*, who proposed that the amorphous η-N phase of polymeric nitrogen is a high-energy metastable framework that can be thought of as a collage of several differing local structural motifs.[5] The authors further suggest that the need for high temperatures during experiment stems from the need to anneal the η-N structure to a crystalline sample, like that of cg-N. The local order of these structural motifs in the η-N phase is thought to be influenced by the stable polymeric allotropes discussed in the literature,[5] like the LB and cg-N phases.

Second, a relationship between the layered BP, A7, and LB phases of polymeric nitrogen and particular arrangements of zzCH and ctCH chains has been discussed. Specifically, plausible pathways from arrangements of zig-zag chains to each layered phase, which do not require significant structural rearrangements, have been conceived and confirmed via subsequent geometry optimizations. These pathways are viable as

alternatives to the precedent SC -derived pathways discussed in the literature, but by no means do they prove the layered phases would be formed through such a mechanism. However, first-principles calculations of such chains and the experimental observation of contiguous phase transformations could support such claims.

Extending the methodology from chain-like building blocks as recurring structural motifs in polymeric nitrogen to helical motifs, a link between the proposed ζ-N_2 molecular coordinates and cg-N polymeric coordinates has been derived and used to construct a pathway between the two phases. The 0.89 eV/atom barrier derived from the pathway matches the energy barrier derived from a pathway reported in the literature which links a molecular form of nitrogen based on the β-O_2 phase with cg-N. Furthermore, arrangements of helical nitrogen motifs are shown to optimize directly to the cg-N and the newly discovered CW phase. This new phase of single-bonded polymeric nitrogen is shown to be metastable from high-quality phonon-dispersion calculations and was, at the time, the only single-bonded allotrope that was thermodynamically competitive with the cg-N structure at low pressure.

A systematic method for searching physically relevant single-bonded nitrogen allotropes has been discussed. The procedure recovers all the possible nitrogen allotropes that can be recovered from Peierls-like distortions of a SC reference structure. The link of polymeric nitrogen, and in fact all Group V covalent networks, to a SC framework is well-discussed in the literature and, specifically, the precedent cg-N, A7, BP, and LB structures have all been recovered from applying this methodology to an eight-atom SC reference simulation cell. A number of other candidate structures, derived from the 36 possible distortions of the SC reference cell, were recovered and, after extensive stability

and thermodynamic tests, two particularly promising novel nitrogen structures have been presented, labeled bstr8 and bstr11 in this work. The bstr11 structure, in particular, is metastable even at ambient conditions.

2.5 - References

1 A. K. McMahan and R. LeSar, Physical Review Letters **54**, 1929 (1985).

2 R. L. Mills, B. Olinger, and D. T. Cromer, Journal of Chemical Physics **84**, 2837 (1985).

3 C. Mailhiot, L. H. Yang, and A. K. McMahan, Physical Review B **46**, 14419 (1992).

4 K. Nordlund, A. Krasheninnikov, N. Juslin, J. Nord, and K. Albe, Europhysics Letters **65**, 400 (2004).

5 M. M. G. Alemany and J. L. Martins, Physical Review B **68**, 024110 (2003).

6 W. D. Mattson, D. Sanchez-Portal, S. Chiesa, and R. M. Martin, Physical Review Letters **93**, 125501 (2004).

7 M. I. Eremets, A. G. Gavriliuk, N. R. Serebryanaya, I. A. Trojan, D. A. Dzivenko, R. Boehler, H. K. Mao, and R. J. Hemley, Journal of Chemical Physics **121**, 11296 (2004).

8 E. Gregoryanz, A. F. Goncharov, C. Sanloup, M. Somayazulu, H.-k. Mao, and R. J. Hemley, Journal of Chemical Physics **126**, 184505 (2007).

9 H. E. Lorenzana, M. J. Lipp, and W. J. Evans, High Pressure Research **22**, 5 (2002).

10 H. Olijynx and A. Jephcoat, Physical Review Letters **83**, 332 (1999).

11 D. Schiferl, S. Buchsbaum, and R. Mills, Journal of Chemical Physics **89**, 2324 (1985).

12 H. Katzke and P. Toledano, Physical Review B **78**, 064103 (2008).

13 R. Etters, V. Chandrasekharan, E. Uzan, and K. Kobashi, Physical Review B **33**, 8615 (1986).

14 R. LeSar, Journal of Chemical Physics **81**, 5104 (1984).

15 J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal, Journal of Physics: Condensed Matter **14**, 2745 (2002).

16 W. Kohn and L. J. Sham, Physical Review **140**, A1133 (1965).

17 R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, New York, 2004).

18 J. P. Perdew, K. Burke, and M. Ernzerhof, Physical Review Letters **77**, 3865 (1996).

19 J. P. Perdew, K. Burke, and M. Ernzerhof, Physical Review Letters **78**, 1396(E) (1997).

20 N. Troullier and J. L. Martins, Physical Review B **43**, 1993 (1997).

21 D. R. Hamann, M. Schluter, and C. Chiang, Physical Review Letters **43**, 1494 (1979).

22 G. B. Bachelet, D. R. Hamann, and M. Schluter, Physical Review B **26**, 4199 (1982).

23 J. Junquera, O. Paz, D. Sanchez-Portal, and E. Artacho, Physical Review B **64**, 235111 (2001).

24 P. E. Blochl, Physical Review B **50**, 17953 (1994).

25 G. Kresse and J. Furthmuller, Computational Materials Science **6**, 15 (1996).

26 C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models* (John Wiley and Sons Ltd, Chichester, West Sussex, 2003).

27 M. Parrinello and A. Rahman, Physical Review Letters **45**, 1196 (1980).

28 S. Jose, Journal of Chemical Physics **81**, 511 (1984).

29 A. F. Schuch and R. L. Mills, Journal of Chemical Physics **52**, 6000 (1970).

30 E. R. Johnson, I. D. Mackie, and G. A. DiLabio, J. Phys. Org. Chem. **22**, 1127 (2009).

31 Y. Ma, A. R. Oganov, Z. Li, Y. Xie, and J. Kotakoski, Physical Review Letters **102**, 065501 (2009).

32 J. Kotakoski and K. Albe, Physical Review B **77**, 144109 (2008).

33 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. R. Cossi, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. E. Gomperts, O. Stratmann, A. J. Yazyev, R. Austin, C. Cammi, J. W. Pomelli, R. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, (Gaussian, Inc., Wallingford CT, 2009).

34 E. Gregoryanz, C. Sanloup, R. Bini, J. Kreutz, H. J. Jodl, M. Somayazulu, H.-k. Mao, and R. J. Hemley, Journal of Chemical Physics **124**, 116102 (2006).

35 R. Caracas and R. J. Hemley, Chemical Physics Letters **442**, 65 (2007).

36 L. F. Lundegaard, G. Weck, M. I. McMahan, S. Desgrenniers, and P. Loubeyre, Nature **443**, 201 (2006).

37 H. Fujihisa, Y. Akahama, H. Kawamura, Y. Ohishi, O. Shimomura, H. Yamawaki, M. Sakashita, Y. Gotoh, S. Takeya, and K. Honda, Physical Review Letters **97**, 085503/1 (2006).

38 Y. Akahama, H. Kawamura, S. Carlson, T. L. Bihan, and D. Hausermann, Physical Review B **61**, 3139 (2000).

39 A. Nangia, Accounts of Chemical Research **41**, 595 (2008).

40 C. Ouvrard and S. L. Price, Crystal Growth and Design **4**, 1119 (2004).

41 R. G. D. Valle and A. Girlando, Journal of Physical Chemistry A **112**, 6715 (2008).

42 A. F. Wells, *Three-Dimensional Nets and Polyhedra* (Wiley, New York, 1977).

43 D. M. Adams, *Inorganic Solids* (Wiley, New York, 1974).

[44] M. I. Eremets, R. J. Hemley, H. K. Mao, and E. Gregoryanz, Nature **411**, 170 (2001).
[45] E. J. Corey, C. L. Cymin, and M. C. Noe, Tetrahedron Letters **35**, 69 (1994).
[46] K. K. Nanda, A. W. Addison, E. Sinn, and L. K. Thompson, Inorganic Chemistry **35**, 5966 (1996).
[47] H. Fujihisa, Y. Akahama, H. Kawamura, H. Yamawaki, M. Sakashita, T. Yamada, H. Kazumasa, and T. Le Bihan, Phys. Rev. B **70**, 134106 (2004).
[48] R. J. Hemley, H. K. Mao, and V. Struzhkin, Journal of Synchotron Radiation **12**, 135 (2005).
[49] L. Crapanzano, W. A. Crichton, G. Monaco, R. Bellissent, and M. Mezouar, Nat. Materials **4**, 550 (2005).
[50] A. Decker, G. A. Landrum, and R. Dronowski, Z. Anorg. Allg. Chem. **628**, 295 (2002).
[51] Y. Akahama, H. Fujihisa, and H. Kawamura, Phys. Rev. Lett **94**, 195503 (2005).
[52] M. I. Eremets, A. G. Gavriliuk, and I. A. Trojan, Applied Physics Letters **90**, 171904 (2007).
[53] F. Zahariev, A. Hu, J. Hooper, F. Zhang, and T. K. Woo, Phys. Rev. B **72**, 214108 (2005).
[54] B. Winkler, C. J. Pickard, V. Milman, and G. Thim, Chemical Physics Letters **337**, 36 (2001).
[55] O. D. Friedrichs, A. W. M. Dress, D. H. Huson, J. Klinowski, and A. L. MacKay, Nature **400**, 644 (1999).
[56] J. K. Burdett and T. J. McLarnan, Journal of Chemical Physics **75**, 5764 (1981).
[57] J. K. Burdett and T. J. McLarnan, J. Chem. Phys. **75**, 5774 (1981).
[58] F. Hararay, *Graph Theory* (Addison-Wesley, Reading, MA, 1994).
[59] T. W. Barbee III, Phys. Rev. B **48**, 9327 (1993).

CHAPTER THREE

SEARCH ALGORITHMS INSPIRED BY THE GENETIC ALGORITHM

This chapter discusses the specifics behind the implementation of genetic algorithms into a program scripted in the Python programming language; this code is used throughout the work presented in the remainder of this thesis. As a follow-up to the work discussed in Chapter Two, the algorithms are applied to map out structures of polymeric nitrogen.

CHAPTER 3 – *KEY SUBSECTIONS*

3.1 – Computational Method

3.1.1 – The Basics

For the work presented in this thesis, a collection of python modules were coded from scratch to implement evolutionary search algorithms inspired by the genetic algorithm (GA), following an approach similar to that suggested by Ho *et al.*[1] In general, the building block of any GA-inspired procedure is the definition of its core representative unit, the gene. These "genes" can be used to assess phenotype variations in a generic body's genetic makeup, ie. its physical characteristics. These variations are expressed through differences between individuals' genotypes, giving rise to alleles in the genetic representation itself. Biologically, for example, a gene consists of a specific sequence of nucleotides encoded in a DNA, or RNA, strand. The correspondence between these nucleotides, the basic building blocks of genetic material, and amino acids, the basic building blocks of proteins, regulates crucial biological functions, eventually determining our body's physical traits, or phenotype expression.

In order to study polymorphism of a generic new material of interest to the solid-state-chemistry research community, GA-inspired searches of its potential energy surface (PES) can only be performed by defining a suitable genetic representation of the chemical system. Once defined, the phenotype variations of these genetic profiles, its polymorphs, can then be explored by manipulating the genotypes. Traditional genetic algorithms consist of binary descriptors, sequences of 1's and 0's, which serve as blueprints for defining the chemical structure, ideally spanning the set of all possible combinations on the studied PES. For example, a schematic describing the connectivity of a linear chain of atoms is shown in Figure 3.1, wherein "1" denotes a bonding

interaction between two adjacent atoms and "0" denotes a non-bonding interaction. In this context, a sequence of 1's would denote a linear chain of atoms "bound" to their neighbors whereas a sequence of 0's would denote a chain of "isolated" atoms. Note that the 101010 gene, shown at the bottom of Figure 3.1, would describe an effective Peierls-like deformation, with alternating pairs of atoms moving together.

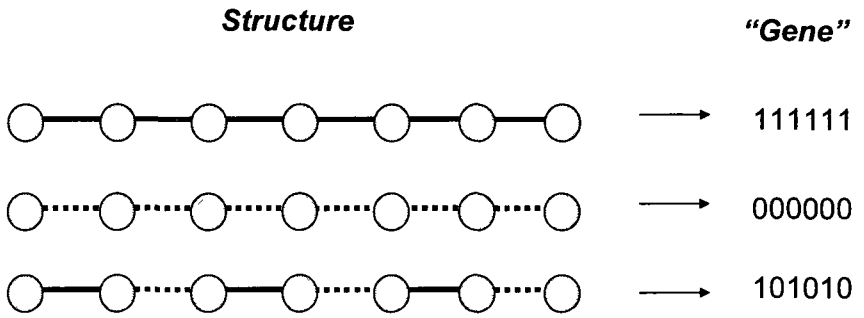


Figure 3.1. Structural and genetic representations of a chain of atoms. Adjacent atoms are considered "bonded" if connected by a solid line or "non-bonded" if connected by a dashed line. The genetic representation (the "Gene") of the structure is a binary sequence describing the connectivity, assuming 0 or 1 for non-bonded and bonded interactions respectively.

A "gene" for any chemical system in the context of this work refers to any representation of a specific crystal configuration. For nontrivial chemical systems, mapping the configurational space onto binary strings can be too ambiguous and ultimately too restrictive since, when unbound, each atom in the unit cell has three translational degrees of freedom. Alternatively, it makes more chemical sense to represent a structure by the Cartesian coordinates of the atoms themselves,[1] thus expressing the gene as a list of Cartesian coordinates; this concept is depicted in Figure 3.2a. When considering heteratomic chemical systems, a complementary list of atomic numbers must be stored in each structure's genetic representation as well, in order to keep track of which species occupies each site. A collection of genes, or a group of crystal structures, is herein referred to as a "population", as illustrated in Figure 3.2b.

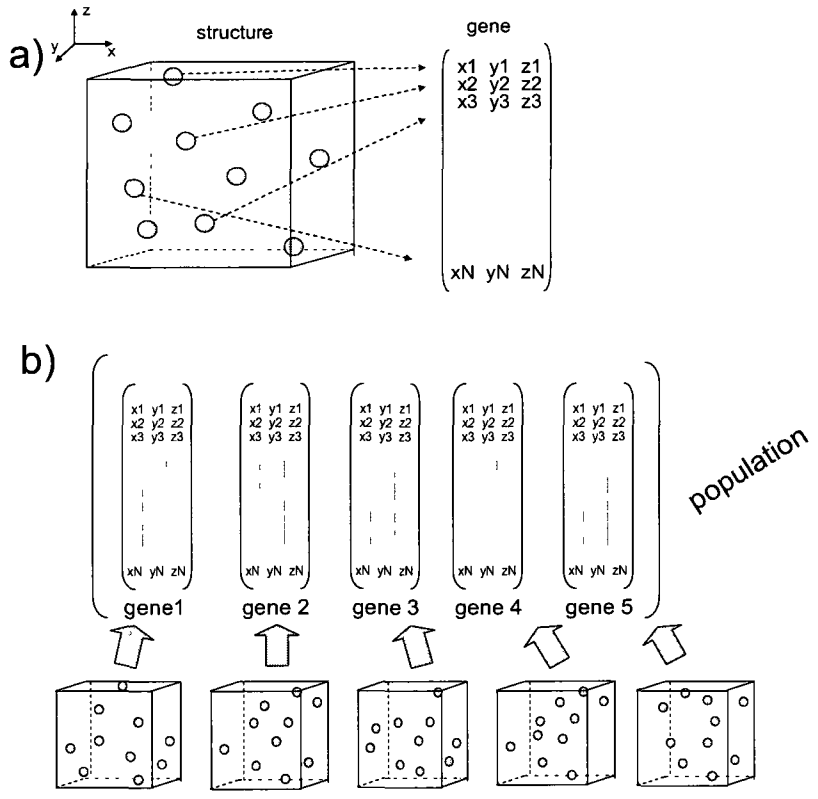


Figure 3.2. Illustrations of a) a "gene" for a generic crystal structure and b) a "population" of genes

The GA algorithm procedure, illustrated in Figure 3.3, starts by generating an initial population of candidate structures by randomly assigning each structure's cartesian atomic coordinates. These structures are stored in memory in the appropriate genetic representation, as lists of Cartesian atomic coordinates and numbers, and the "fitness" of each structure is evaluated. The fitness is a metric defined to relay how a structure fares compared to rest of the structures with regards to the desired property, thereby assigning "ranks" to each member of the population. For example, if the most thermodynamically stable conformer is sought then the total energy would be an appropriate fitness metric. A pool of new structures, a new population, is then fashioned from the genetic profile of the existing population by using evolutionary-based algorithms (mating and mutation

operations) to promote or combine favorable structural motifs of preferably fit members of the gene pool. The old population is then overwritten with the newly constructed population and the fitnesses are re-evaluated. The cycle continues until the algorithm converges such that no fitter structures are found from applying further mating/mutation operations.

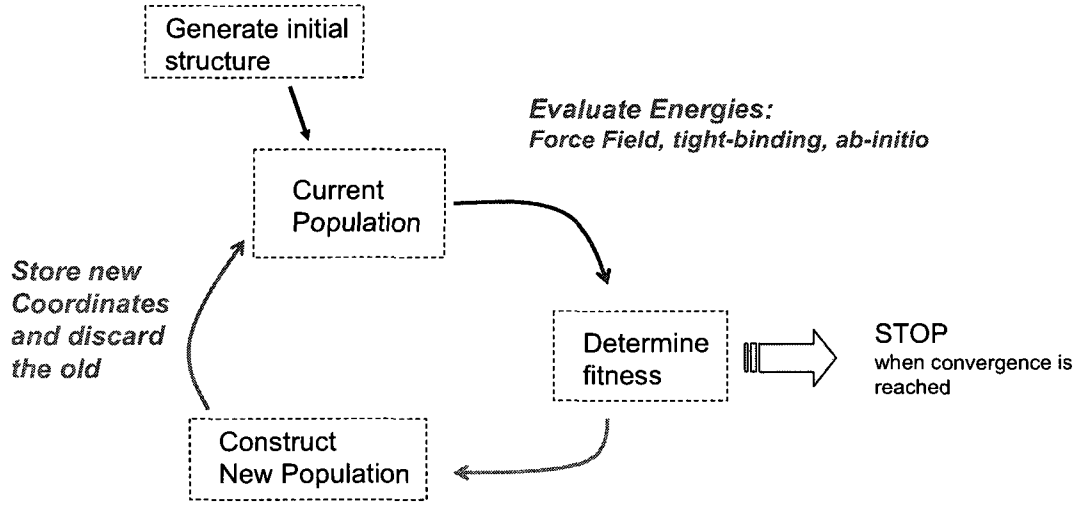


Figure 3.3. Flow chart of GA-inspired structure searches

From this generic setup for the GA, two main focal points emerge which, when specified, define the specific goal of the algorithm itself. First, the fitness metric must be carefully chosen to reflect the goals of the study while requiring reasonable computational resources. Second, the protocol for generating new structures, namely the procedure used to evolve of the gene pool throughout the simulation, must be designed to explore the PES efficiently. Ideally, an optimal GA procedure must effectively promote favorable structural straits from one generation to the next, mimicking a Darwinian "survival-of-the-fittest" approach, and still introduce enough diversity into the population to effectively explore other potential low-energy regions of the PES.

In the rest of this section (section 3.1) the machinery behind fitness evaluation and structure generation will be discussed in more detail. In Section 3.2, a straight-forward application of the encoded GA procedures toward high-pressure nitrogen, at a range of pressures, is presented to demonstrate the effectiveness of the algorithm in recovering all of the previously-discovered polymeric structures and several other polymeric nitrogen structures of interest. To validate the GA's mating/mutation procedures and probe its limits with heteronuclear systems, the code is then benchmarked extensively with a previous GA study on $CaCO_3$ [2] and additionally shown to successfully recover the cassiterite SnO_2 structure. Finally, the GA routines are used to study high-pressure N_xH_y polymeric structures as potential HEDM alternatives to polymeric nitrogen, wherein several interesting crystal structures are recovered.

3.1.2 -Fitness Evaluation

Since the immediate goal of the GA simulations presented in this chapter is to explore the electronic structure of the lowest enthalpy phases on a generic PES, it is appropriate to determine the fitness directly from electronic-structure energy calculations. Specifically, in all the GA studies intended to explore thermodynamically-favored structures on the PES throughout this thesis, the absolute fitness value of each candidate structure (or gene) is assigned as its relative enthalpy to the lowest-enthalpy structure of the population; this energy can be determined from any computational chemistry package deemed appropriate for an adequate description of the PES. Note that the general definition of the fitness is itself quite flexible, and can readily be recoded to better reflect another crystal property beyond thermodynamic stability. Regardless of its definition,

the choice of program to evaluate the fitness must be adapted to the problem at hand since the size, scope, composition, and nature of the system all play crucial roles in determining sustainable energy evaluations. For example, if one wished to study 90-atom simulation cells without constraining atomic positions to specific classes of crystals, then first-principles density-functional theory calculations are unfeasible. The scope of this problem will be further demonstrated throughout the work presented in Chapters 3-6. The results ultimately reinforce the notion that limits in computational resources should be carefully considered in this regard since overly-expensive fitness evaluations must be compensated for by restricting the effective search space.

To make the technology as robust as possible, a "complete" theoretical toolkit with first-principles, semi-empirical, and classical methodologies should all be available to the user. High-pressure nitrogen, for example, cannot effectively be described by a single robust, classical force field,[3] making electronic-structure methodologies necessary for such systems (as will be demonstrated in Sections 3.2 and 4.1). On the other hand, classical methods must be used to model the large simulations cell of ceria (100+ atoms)[4,5] this is discussed in Section 5.1. The GA program developed in this work is implemented and fully compatible with the first-principles VASP[6] and SIESTA[7] packages, the semi-empirical MOPAC package,[8] and the classical GULP[9] and DLPOLY[10] chemistry simulation packages.

It has been determined throughout the course of this work to be remarkably useful to probe a PES at an "effective" high level of theory by using lower levels of theory to screen for structures in assumed-to-be favorable regions of the PES and incorporate them into further GA trials at higher levels of theory, this is discussed further in Chapter 5. In

this regard, the implemented GA program allows user-specification of what specific simulation package, and therefore what level of theory, to use at each generation of the trial. There are also different "optimization schemes" one can use on each gene before evaluating its fitness. For example, the user can allow or disallow ionic or lattice relaxation before fitness evaluation. Examples of stand-alone and tandem uses of classical, semi-empirical, and density-functional theory derived fitness metrics, both energy based and otherwise, are reported in the remaining thesis chapters 3-6.

3.1.3 - Structure Generation

Beyond the fitness metric, the heart and soul and any GA procedure is the procedure used to generate a new population of structures from the population stored in memory. The goal is to use intelligent evolutionary-based algorithms designed to primarily incorporate favorable structural motifs from one generation into the next (survival-of-the-fittest), and concurrently introduce genetic diversity in the hopes of producing structures that are more "fit" than their predecessors (evolution). The algorithm works by applying three promotional schemes to generate a new population from the current population, or "gene pool", as shown in Figure 3.4: 1) Gene Promotions, 2) Random Generations, and 3) Mating and Mutation operations. These three promotional schemes will now be elaborated upon below on a point-by-point basis.

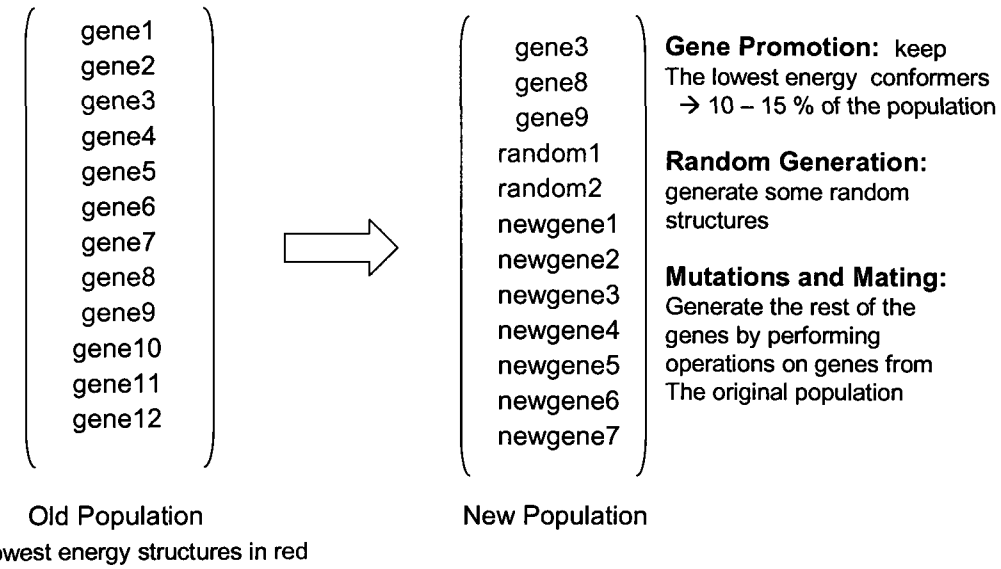


Figure 3.4. Schematic for population creation in the GA procedure

Gene Promotion (Elitism) Scheme

The Gene Promotion generation scheme simply copies the genetic representations of the structures with the best fitness values, or the so-called "elite" members of the population, directly into the next generation. This is partly how the algorithm "remembers" the lowest enthalpy regions of the PES it has already explored and builds upon those particular structural motifs through further mating/mutation applications on future populations.

A nuance of the direct gene promotion approach is the accidental promotion of identical genotypes. A schematic illustration of this is given in Figure 3.5, demonstrating how three "mutations" (discussed below) of state A on its fictitious one-dimensional PES can produce three new states, labeled B, C, and D. All four of these structures lie in the same local minimum on the PES and would fall back into the same minimum as state A upon geometry optimization. If the fitness of these four identical states happens to be low enough to be selected in the gene promotion stage, all four configurations would be

added to the genepool and subsequently used in the mating/mutation operations themselves. These new mating operations could again generate similar structures and the genepool would be heavily biased by state A and become "stagnant", rarely exploring the potentially more relevant parts of the one-dimensional PES (such as the global minimum depicted in Figure 3.5).

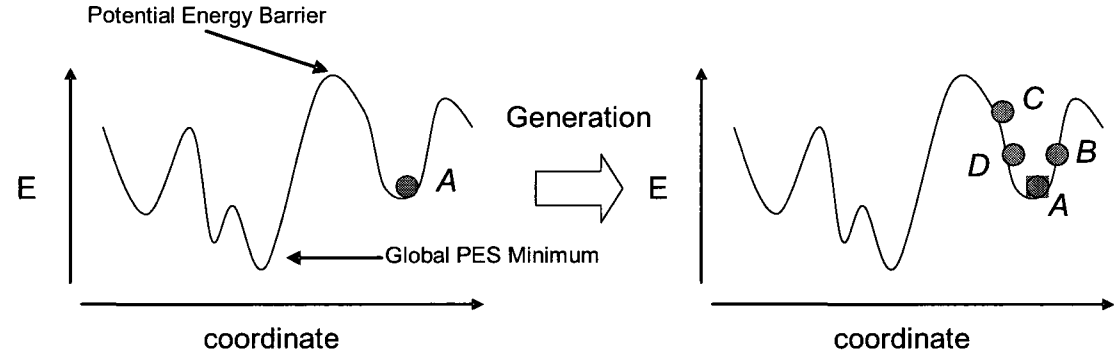


Figure 3.5. Schematic illustrating redundant structure generation in GA procedures on a fictitious 1D PES. A fit genotype on the PES (left) is mutated three times (right) such that all four genotypes fall in the same local minimum on the PES.

To remedy the promotion of redundant genotypes, an energy resolution is defined to screen out structures that have similar total electronic energies.[11] The downside of such a practice is that if there are energetically competitive configurations of the crystal then all but one of them will inevitably be screened out during Gene Promotion, requiring either that several redundant GA trials are run on each system or an additional metric is introduced and used simultaneously with the energy resolution to identify distinct structures. Therefore, an additional measure is taken in which each structure's radial distribution is compared with those already in the gene pool before screening out energetically competitive structures.

The optimal number of structures promoted during the elitism portion of the GA varies depending on the parameters of the simulation. For example, a large percentage of

the population would be needed when dealing with small population sizes, in order to preserve the recollection of distinct favorable motifs, but this same large percentage would prove detrimental when dealing with large population sizes, since structures with low or irrelevant fitness values would likely be promoted near the end of the scheme. In the program used in this work, the number of structures promoted directly to the next generation in the GA code can be specified by the user and the energy resolution was assigned a value of 1 meV/atom. If unspecified, the default value of the number of promoted structures was set to the ceiling function of 5 and 15% of the total population size.

Random Generation Scheme

The Random generation scheme generates structures with randomly-assigned cartesian coordinates and then adds them to the new population. Although the odds of generating a fit structure in the later stages of the GA trial are minimal, the motivation behind these additions is to introduce genetic diversity into the population. The stochastic nature of the GA allows a non-zero possibility of selecting these randomly-generated genes for mating/mutation operations. By the subsequent introduction of radical new structural motifs and mating them with the existing gene pool, relevant and previously unexplored parts of the PES are more efficiently searched. In this work, the number of randomly-generated structures is taken to be 5% of the total population.

Mutation and Mating Schemes

The remainder of each newly-generated population is filled out with genes generated from mutation and/or mating operations on candidate structures from the current generation. As discussed above, these schemes take structural information

embedded in the genetic representations of the current population, the "parent" structures, and manipulate their genes to create new, distinct structures for the next generation, the "offspring" structures. One parent structure is selected for mutation operations and two parents are selected for mating operations, as shown in Figure 3.6.

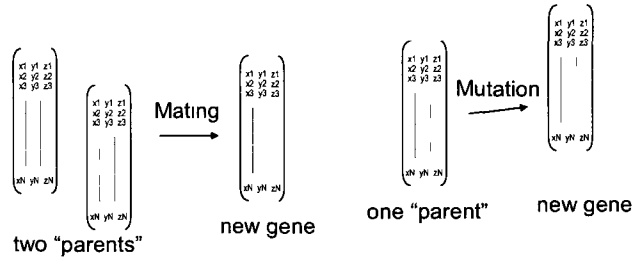


Figure 3.6 – distinction between mating and mutation operations. Mating operations read genetic information from two parents and mutation operations read from only one

The simplest way to generate new structures is to directly add an offspring created from a single mating or mutation operation to the new population. In this way, each structure-generation consists of one mating or mutation operation. However, a candidate structure can just as easily be created by applying consecutive (or cascading) mating or mutation operations to the same parent structure(s).

This brings about three different issues that need to be addressed to implement the mating/mutation schemes: 1) the selection criteria used to select the parent structures, 2) the nature of the mutation operations, and 3) the nature of the mating operations.

Selection Criteria

Any structure from the existing population has an equal chance to be selected as a candidate for mutation and mating operations. The likelihood of that candidate structure then being chosen as a parent for mutation/mating operations relates to a boltzmann-weighted probability distribution[1], shown in eq. [1].

$$P(i) = \exp(\frac{fit(i) - fit(1)}{kt}) \qquad [1]$$

Where P(i) is the probability of selecting the candidate structure, the i^{th} most fit structure in the population, as the parent of the current mating/mutation operation. The fitness of the i^{th} structure is denoted as fit(i), thus fit(1) is the fitness of the most fit structure. The kt term, eq. [2], is an arbitrary weighting factor chosen to allow at least a 25% chance to select the top five promoted structures.

$$kt = \frac{fit(5)}{-\ln(0.25)} \qquad [2]$$

Mutation Operations

Mutations are operations which select one parent structure and modify, or "mutate", its genetic representation to create a suitable offspring for the next generation. The mutation operations designed for exploring the PES of chemical systems in this study are: *i*) geometry perturbation, *ii*) atom swapping, and *iii*) gene coordinate-inversions.

i) **Geometry perturbation**: This is arguably the most conceptually simple mutation operation and it has been encoded into every GA-based search method reported in the literature.[1, 2] Its function is to simply perturb the Cartesian coordinates of one or more of the atoms in the parent structure's genetic representation, schematically illustrated in Figure 3.7 where the atoms colored red are perturbed. In its implementation in this work, one atom is randomly chosen from the structure and perturbed. For the remaining atoms, the probability of the associated cartesian coordinates being perturbed is determined by generating a number from a uniform probability distribution between 0 and 1. If the generated number is less than the reciprocal of the total number of atoms in

the system, then the atom is perturbed by a maximum of 0.4 angstroms in any or all of the x,y,z directions. A variant of this mutation is to allow cell perturbations as well, by allowing small, random perturbations to the lattice vectors themselves.

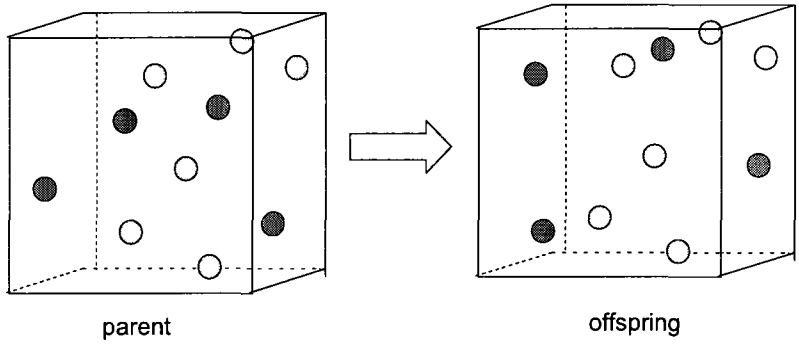


Figure 3.7- Schematic representation of Geometry Perturbation. Select atoms (red) in the parent structure (left) are perturbed to create the offspring (right).

ii) **Atom Swapping:** This mutation reconfigures the genetic information by swapping the cartesian coordinates of heteronuclear pairs of atoms, thus only creating a distinct offspring structure when swapping distinct atomic species. This operation is intended to help recover the preferred local atomic structures in the early stages of GA trials on heteronuclear systems, helping speed up their overall convergence.

iii) **Gene Coordinate-Inversions**: The least physical of the pure mutation operations is to preserve two of the cartesian coordinates for all the atoms but swap the third coordinate among a subset of the complete atom list. For example, Figure 3.8 shows its application when preserving the x and y coordinates of the subset of target atoms in the "parent" gene, atoms 12-16.

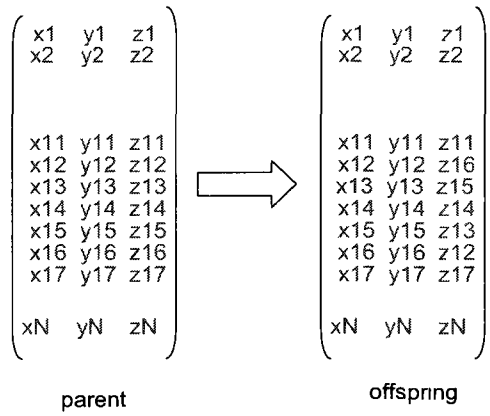


Figure 3.8 – Schematic representation of the Gene Coordinate-Inversion mutation operation, applied to atoms 12-16 along the z cartesian axis

Mating Operations

Mating operations are operations which select two parent structures, and combine, or "mate", their genetic representations to create a suitable offspring for the next generation. The mating operations designed for exploring the PES of chemical systems in this study are *i*) Cartesian Averaging, *ii*) Atom Splicing, *iii*) Random Mixing, and *iv*) Spatial Splicing.

i) **Cartesian Averaging** is an unphysical mating operation which takes the mathematical average, either the arithmetic or the geometric average, of each parent's corresponding atomic coordinates.

ii) **Atom Splicing** is a mating operation analogous to gene splicing. It selects a segment of one parent structure and "splices" that together with the complimentary segment of the other parent structure.

iii) **Random Mixing** is a mating operation which randomly chooses a subset of atoms from each parent and combines them together, ensuring that enough atoms are chosen to generate a physically meaningful offspring.

iv) **Spatial Splicing** is a mating operation which cuts a cubic subsection out of one parent structure and replaces that sub-structure with that of the other parent structure. This mating operation was designed to allow favorable local structural motifs to be readily transferred between generations.

After each mating/mutation operation, the offspring are checked, before being accepted, in order to ensure they still fit the general parameters of the study. For example, if the study is geared towards studying a particular crystal empirical formula, the number of atoms is monitored and corrected by adding or removing randomly selected atoms to obtain the correct number. The cell volume is similarly rescaled if the mating/mutation operation generates an unphysical cell, ie. if any atoms are too close to one another.

Each core mating/mutation operation is given an equal chance to be selected as the basis toward creating a new structure. Alternative operations are set up which consist of tandem applications of the core mating/mutation operations outlined above and given an equal chance to be selected to generate a new structure. For example, one such operation is Atom Swapping followed immediately by Spatial Splicing. In addition, all offspring, no matter how they were generated, are given a chance to mutate further by applying the geometry perturbation operation.

When enough structures are generated to create a new population, the old structures are discarded from memory and the fitnesses of all structures in the new population are re-evaluated. The cycle of population generation and fitness evaluation is continued until the most fit structures remain unchanged for several generations, the

precise number of generations varies depending on the complexity of the system. Note that although the genetic algorithm is guaranteed to progressively explore lower energy regions of the PES, there is no guarantee that a given run will find the global minimum of the problem at hand, only that the algorithm should explore relevant parts of the PES. This convergence will be discussed on a case-by-case basis throughout this thesis.

3.2 – GA applications to bulk systems

The genetic algorithm (GA) described above can readily be applied to study bulk crystal structures, and has repeatedly been used in the recent literature to navigate high-pressure monatomic solids.[12-15] A recent collaboration between the developers of the evolutionary code USPEX,[2] an implementation of the GA, and experimentalists resulted in the high-impact discovery and characterization of a new transparent phase of high-pressure sodium[14] and an ionic form of high-pressure boron.[12] Admittedly most of these studies operate in the ultra-high pressure regime, where the limits of x-ray diffraction and Raman characterization techniques prevent in-situ determinations of the atomic coordinates. Nonetheless, the procedure has been shown to be an effective tool for studying polymorphism and gaining insight into unknown crystal configurations. Its value at ambient pressures is simply lessened since experimental characterizations of crystal structures often suffice on their own.

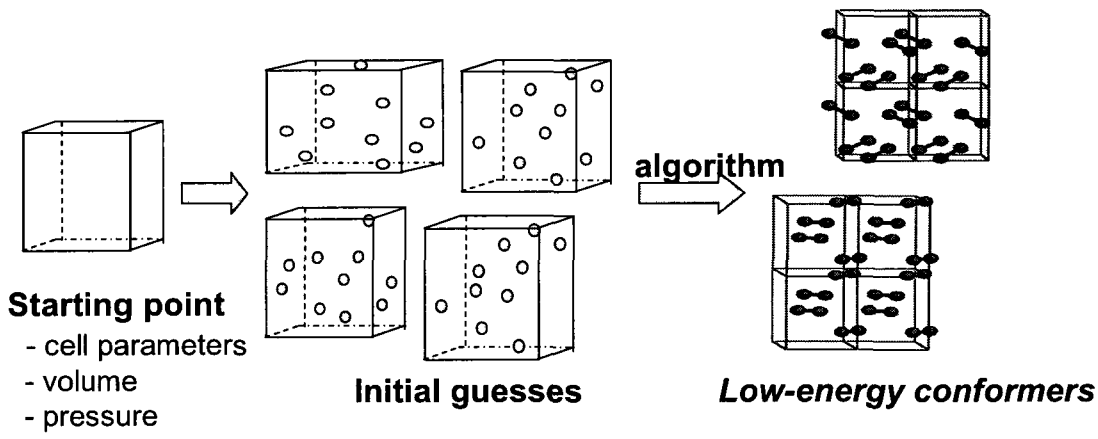


Figure 3.9. GA search procedure. Initial guesses are first constructed from insights gained in the lab (cell parameters, volume, pressure, etc...) and evolutionary protocols are run on these initial guesses to recover low-energy conformers.

The general implementation of these GA procedures is summarized in Figure 3.9. The GA trials start from random initial guesses of the crystal configurations. The nature of these initial guesses can be guided from experimental insights into the volume, the pressure, or the number of atoms in the unit cell, but these considerations are not strictly necessary. The evolutionary protocols discussed in Section 3.1 are then recursively applied to the pool of initial structures to generate favorable or relevant crystal configurations for the scope of the problem at hand. In the remainder of this chapter, the GA code presented in this work is benchmarked against high-pressure nitrogen,[13, 16] $CaCO_3$,[11] SnO_2, and N_xH_y.

The SIESTA[7] programming suite was used in all the calculations presented in Section 3.2.1. They employ Kohn-Sham density functional theory calculations with the gradient-corrected exchange-correlation functionals of Perdew, Burke, and Ernzerhof (PBE).[17, 18] A Troullier-Martins norm-conserving pseudopotential[19-21] is used to reference a nitrogen core $[He]2s^22p^3$ configuration with a 0.98Å cutoff. The low cutoff was chosen to allow for greater transferability in the high pressure simulation

environments. Custom SIESTA-type numerical doubled s, p_x, p_y, and p_z orbitals and d polarization orbitals[22] were used with a real-space mesh cut-off of 200.0 Ry. A 10 Å cutoff for k-point sampling was used to construct Monkhorst-Pack grids for Brillouin-zone integration. Similar DFT calculations on polymeric nitrogen with SIESTA have been recently reported.[23] In all the other sections, the projector augmented wave (PAW) method of Blöchl[24] was used, within the VASP[6] simulation package, to treat the core electronic states. The nitrogen, oxygen, and carbon 2s and 2p electrons, the calcium 3s, 3p, and 4s electrons, and tin 5s and 5p electrons were treated as valence electrons with a plane-wave cutoff of 430 eV for geometry optimizations and 650 eV for energy evaluations. A plane-wave cutoff of 39 Ry was used and the gradient-corrected exchange and correlation functionals of Perdew-Burke-Ernzerhof (PBE)[17] were used in all calculations. For energy evaluations within the genetic algorithm, the Brillouin-zone integrations were performed using 3x3x3 Monkhorst-Pack grids whereas 8x8x8 meshes were used in more refined calculations.

3.2.1 – GA Validation on Monatomic Nitrogen Systems

Since the unit cell volume of most high-pressure nitrogen crystals is relatively small, for example the unit cell of the cg-N structure occupies 54.9 $Å^3$ at 90 GPa, it is reasonable to use first-principles density-functional theory (DFT)[25, 26] to evaluate the fitness of the candidate structures. The fitness metrics were determined after allowing full ionic relaxation such that the largest atomic force was under 0.05 $eV/Å^2$ and the optimized coordinates and cell were read into the GA procedure to overwrite their initial unrelaxed genetic representations. All of the mating and mutation operations discussed in Section 3.1 except the "Atom Swapping" mutation routine were given an equal chance

of being selected for a given offspring generation; the "Atom Swapping" mutation was turned off because it has no effect on monatomic structures. Every offspring generated by the procedure was given a 20% chance to have its geometry further mutated by a "Geometry Perturbation" before being added to the new population. By first setting up the GA program to generate thirty 8-atom cubic simulation cells of nitrogen with 3.77 Å lattice parameters and keeping the cell fixed throughout the entire GA procedure, a number of low-enthalpy structures, pictured in Figure 3.10, were recovered after only ten generations.

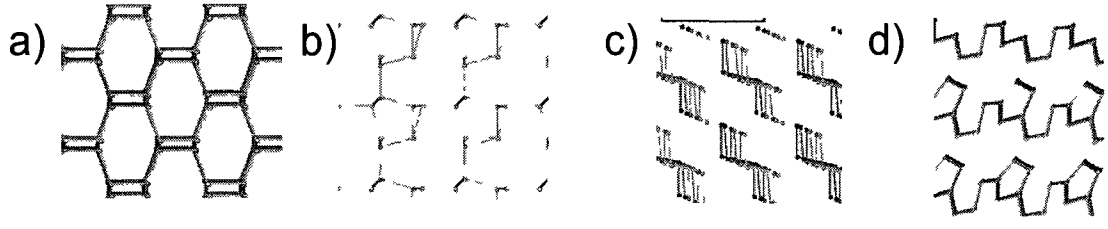


Figure 3.10. Low-enthalpy phases of polymeric nitrogen recovered from GA trials on 8-atom cubic cells of nitrogen with a=3.77Å. a) cg-N, b) bstr11, c) N_6-ring stripes, and d) compressed phases.

The cg-N allotrope, Figure 3.10a, was recovered as the lowest energy nitrogen crystal, along with bstr11 structure, Figure 3.10b, which was proposed in Chapter 2. Several chain-like structures were seen, both the zzCH and ctCH varieties, along with some clustered-chain structures, structures in which the chains are brought close enough together to form interchain bonds. An interesting variation of these chain-like phases is shown in Figure 3.10c, revealed to be the ctCH phase grouping together in interconnected pairs to form stripes of staggered N_6 rings in the chair conformation. Many other high-energy polymeric structures, an example is shown in Figure 3.10d, were also recovered. These phases typically possessed little or no symmetry, and although the atomic forces are converged to zero, chemical intuition suggests many of the phases were strictly

formed by artificially placing nitrogen atoms at random Cartesian coordinates in the unit cell and are not mechanically stable or enthalpically relevant on the nitrogen landscape.

To explore the potential-energy landscape of a chemical system which is largely unknown, as is the case for polymeric nitrogen, it is imperative to configure the GA to be unbiased to a particular set of lattice parameters. With this goal in mind, the GA procedure was modified to generate initial nitrogen structures with arbitrary orthorhombic unit cell parameters under the restriction that the volume of the generated cell remained between 50 and 60 $Å^3$. The volume restriction was only applied to ensure the polymeric regime of nitrogen was explored, as opposed to the lower-pressure molecular nitrogen phases. The fitness metrics were thus determined after allowing full ionic and lattice relaxation, such that the trace of the stress tensor reached the user-specified pressure. All of the mating and mutation routines were modified to read in and rescale the lattice cell vectors, if necessary, to ensure preservation of reasonable volumes in the simulation cells.

By implementing the GA program to generate 8-atom orthorhombic unit cells of nitrogen at 120 GPa, the expected low-enthalpy cg-N and bstr11 structures were again recovered after 10 generations using only 20 genes per population. Furthermore, the zzCH and BP phases were recovered along with a high-energy sheet-like phase reminiscent of LB, except the N_6 rings assume a twisted boat conformation (not shown).

Having successfully recovered the lowest-enthalpy, to date, cg-N polymeric nitrogen structure with Z=8 at 120 GPa, the GA procedure was retrofitted to optimize its gene population at 15 GPa instead of the prior 120 GPa pressures. As expected, these GA trials generated dinitrogen molecular solids. The expected α-N_2 and γ-N_2 phase

structures, the experimentally-observed molecular-solid phases in this pressure regime shown in Figure 3.11a and Figure 3.11b respectively, were recovered as the two lowest enthalpy phases, differing by less than 10 meV/atom in enthalpy. A monoclinic molecular phase resembling a slightly sheared γ-N_2 unit cell with Z=4 was also recovered with an enthalpy 15 meV/atom higher than γ-N_2, shown in Figure 3.11c. All three phases were later confirmed to be the same molecular phases discussed during the computational-compression experiments of the α-N_2 structure in Section 2.1.3.

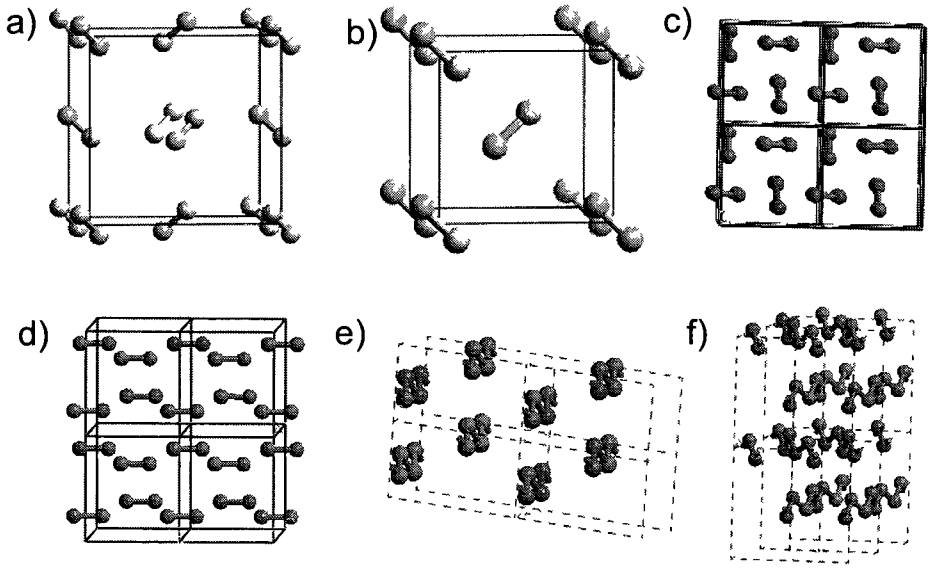


Figure 3.11. Low-enthalpy phases of high-pressure nitrogen recovered from GA trials on 8-atom cells. The a) α-N2, b) γ-N2, c) distorted γ-N2, and d) miscellaneous molecular phase were recovered at 15 GPa, while the e) zzCH phase and f) skewed zzCH phase were recovered at 80 GPa.

When rerun at 40 GPa, the GA search recovered the distorted γ-N_2 phase almost exclusively as the lowest enthalpy molecular phase. A triclinic phase with four atoms per unit cell, lying 0.038 eV/atom higher in enthalpy, possessing P_{-1} symmetry was also found to bear a striking resemblance to the aligned *I*mmm-symmetry phase reported in Chapter 2; this phase, which was also seen in the 15 GPa GA trials, is shown in Figure 3.11d. Beyond the 50 GPa threshold, molecular nitrogen phases become metastable to

polymeric nitrogen. Although only one GA trial was run at 40 GPa, several molecular/chain-like nitrogen allotrope mixtures were recovered from the procedure. These structures have been reported in several high-pressure nitrogen studies[23, 27] and come as no surprise at the threshold where molecular and polymeric nitrogen are more or less thermoneutral.

Four GA trials with eighty structures per population were run to generate nitrogen structures equilibrated to 75 GPa, to verify whether the GA algorithm would now once more recover the expected polymeric structures. Each trial was run for at least 15 generations, when convergence of the lowest three structures was maintained for three generations. Polymeric nitrogen structures were exclusively recovered from all four trials. The cg-N structure was recovered in two of the trials as the lowest enthalpy phase and the previously reported CW, BP,[28] ctCH,[29] and zzCH[23, 27] phases were recovered as well. Both arrangements of the zzCH phase reported in the literature were recovered, differing only by the relative stacking patterns of the chains. The lowest enthalpy zzCH structure is shown in Figure 3.11e. Several distorted or skewed zzCH structures were also produced, an example of such a structure is shown in Figure 3.11f. This particular phase consists of layered zigzag chains oriented such that chains on adjacent layers are nearly perpendicular to each other, their planes intersecting at 67° in the structure shown. Although the energy penalty is substantial, nearly 0.3 eV/atom, when compared with the the traditional zzCH structures,[23] it reinforces the notion of an evidently complex chain-like phase landscape and suggests other chain-like phases can likely be recovered by optimizing alternate stackings of the chain-like building blocks.

In concert with this work, evolutionary searches on polymeric nitrogen have been reported in the recent literature,[2, 16] wherein a low enthalpy Z=10 orthorhombic phase, the pN phase, and a Z=8 triclinic phase consisting of an interconnected web of N_8 rings, the pN8 phase, are reported. The GA searches presented in this chapter also recovered several novel polymeric nitrogen phases which are enthalpically competitive with other nitrogen structures reported in the literature, but pale in comparison to the cg-N structure. One such phase, herein referred to as the "pN7" structure whose primitive eight-atom unit cell is shown in Figure 3.12a, was recovered from one of the GA trials at 80 GPa. Its sixteen atom conventional cell representation, shown in Figure 3.12b, has *C*2 symmetry with its constituent ions spread over four 4c general Wyckoff sites. The cell is a C-centred monoclinic unit cell with lattice parameters of a=5.49 Å, b=6.06 Å, c=3.33 Å, and ß=88.5° at 40 GPa. Its structure consists of interconnected webs of N_7 rings, with the remaining N atom bridging the neighboring N_7 rings. This is similar to how the pN8 phase[16] can be thought of as N_6 rings interconnected by strings of the remaining two external N atoms. This train of thought also helps describe another polymeric nitrogen phase recovered from the GA trials as a web of ring structural motifs, shown in Figure 3.13b. This structure, herein referred to as the twisted-web (TW) structure, consists of interconnected N_6 rings in the twisted boat conformation. This structure, like pN8, has *P*1 symmetry with eight atoms distributed in its triclinic cell and assumes the following lattice parameters at 40 GPa: a=3.22 Å, b=3.73 Å, c=4.49 Å, α=113.4°, ß=89.8°, and γ=76.7°.

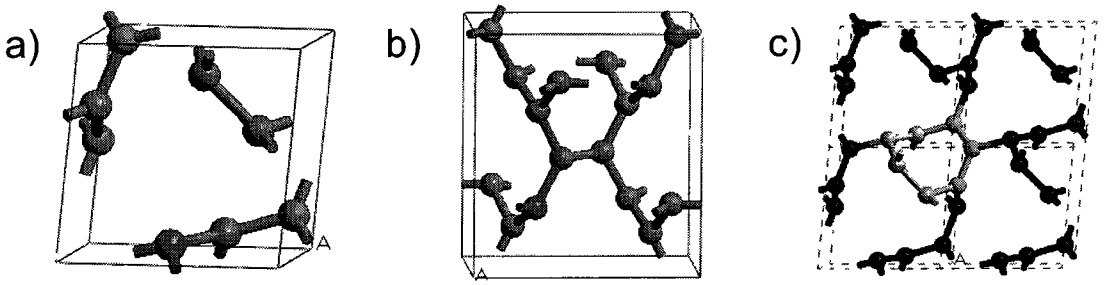


Figure 3.12. a) Eight-atom unit cell of pN7 phase, b) its sixteen-atom monoclinic represention, and c) extended 2x2x1 supercell highlighted the N_7-ring motif.

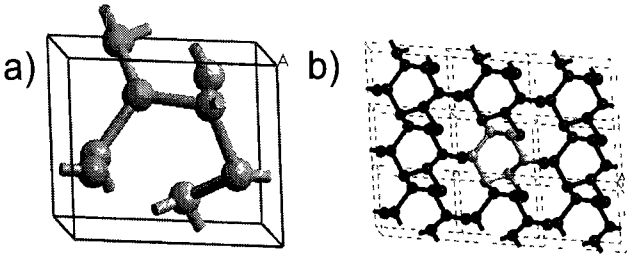


Figure 3.13. a) Eight-atom unit cell of the twisted-web (tw) phase and b) extended 3x3x1 supercell highlighted the N_6 twisted-boat motif.

Enthalpically, these structures are competitive with other reported phases of polymeric nitrogen, as shown in the relative enthalpy vs. pressure phase diagrams presented in Figure 3.14. The pN7 and TW structures lie 0.392 and 0.416 eV/atom higher than cg-N at 100 GPa, respectively, placing them between the A7 and bstr11 polymeric structures. Their relative enthalpies to the cg-N structures reduce to 0.291 and 0.213 eV/atom, respectively, when the pressures are lowered to 20 GPa. The TW phase then decomposes into ctCH chain-like polymeric structures by the time the pressure reaches 12 GPa, this is shown visually by the sharp and discontinuous decrease in enthalpy associated with the structural transformation. Note that the enthalpy diagram also shows the previously described decomposition of bstr8 at low pressures (section 2.3.4).

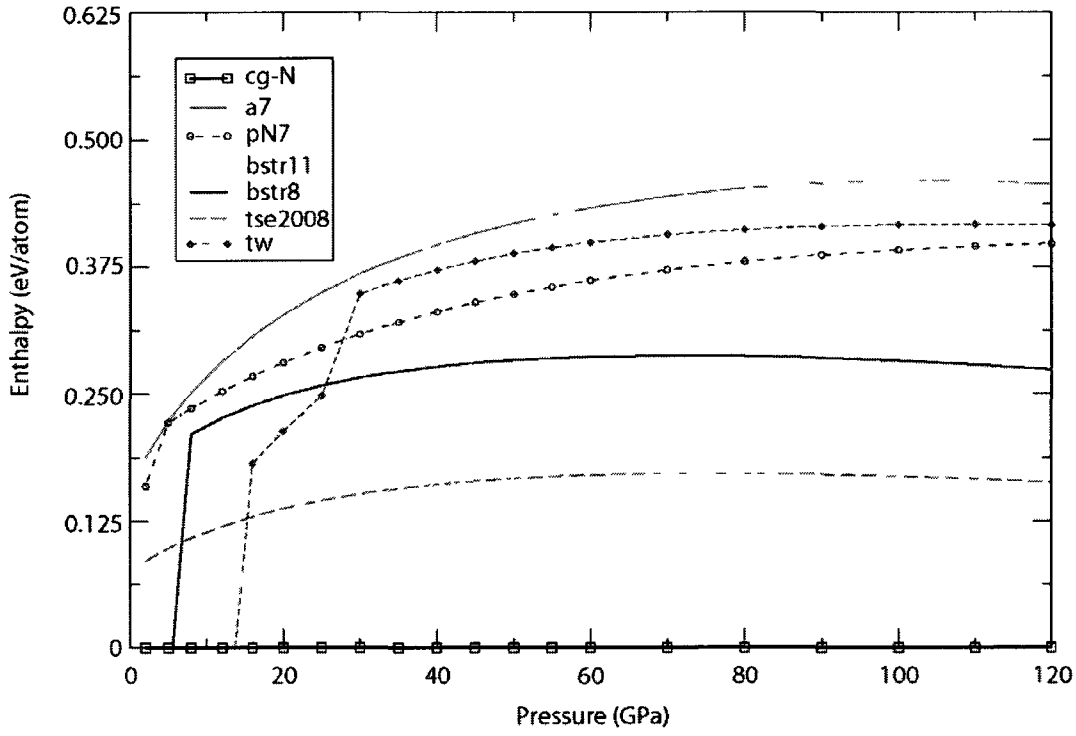


Figure 3.14. Relative enthalpy vs. pressure phase diagram of select polymeric nitrogen phases.

Finally, a particularly interesting layered polymeric nitrogen structure was recovered at 90 GPa. Its extended structure, shown in Figure 3.15b, consists of stacked planes of nitrogen atoms reminiscent of the well-known carbon graphite phase; each layer consists of fused 18-member star polygons arranged to tile all space in the plane, three of these regular-polygon constructs are shown in the overhead perspective in Figure 3.15c. Because of its similarity with its namesake, this phase is herein referred to as the "nitrogen graphite" (NG) phase. The 8-atom rhombohedral unit-cell representation of its *R*-3 symmetry lattice is shown in Figure 3.15a. At 90 GPa, its lattice parameters are a=4.73 Å and α=77.67° with nitrogen atoms distributed about 6*f* and 2*c* Wyckoff sites derived from the (2.03, 2.77, 4.42) and (1.16, 0.93, 0.78) cartesian atomic coordinates.

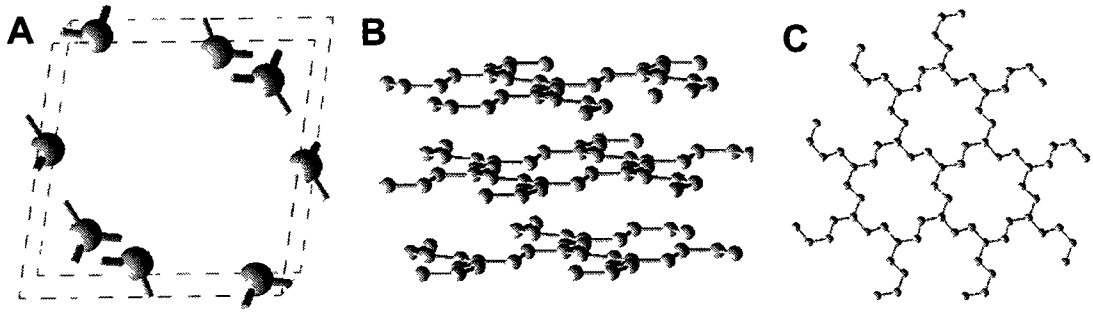


Figure 3.15. Structure of the layered NG (nitrogen graphite) phase: A) Rhombohedral unit cell representation (8 atoms), B) extended structure depicting the nitrogen "sheets", and C) an overhead perspective of each layer.

An enthalpy vs. pressure phase diagram of the NG, zzCH, and ctCH phases relative to the cg-N structure between 0 and 30 GPa is shown in Figure 3.16. Enthalpically, the NG phase is disfavored at 90 GPa (not shown), largely due to the large volume required to accommodate the high-order polygon shapes that make up its layers. At ambient pressures, however, the structure becomes enthalpically-competitive with the zzCH phase, the zzCH phase prominently stands as the most enthalpically-favored form of polymeric nitrogen at 0 GPa reported in the literature.[23] The inset in Figure 3.16 shows the crossover in enthalpies between the zzCH and NG structures at 1.5 GPa, making NG the lowest-enthalpy polymeric nitrogen structure at near-ambient pressures.

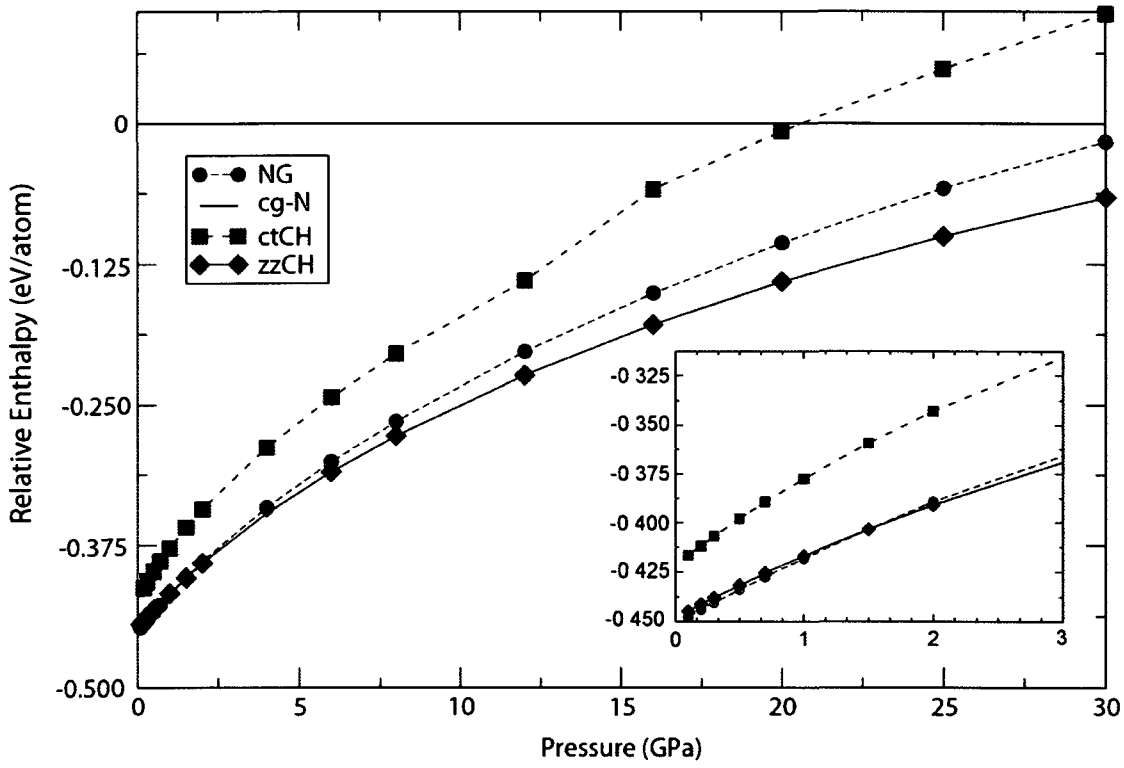


Figure 3.16. Calculated Enthalpy vs. Pressure diagram of the NG, cg-N, ctCH and zzCH phases. Enthalpies are given relative to the cg-N structure.

The NG structure's phonon dispersion curves, shown in Figure 3.17, show all-positive phonon frequencies, thus confirming that the structure is mechanically stable at zero pressure and further suggesting the structure may be dynamically stable. Note, however, that the softest acoustic modes were calculated to have vibrational frequencies that dip as low as 50 cm^{-1} between the L and FA points in q-space. Nonetheless, DFT calculations indicate that the NG structure is both low-enthalpy and mechanically stable at ambient pressures, making it the only layered phase reported in the literature thus far that satisfies both criteria.

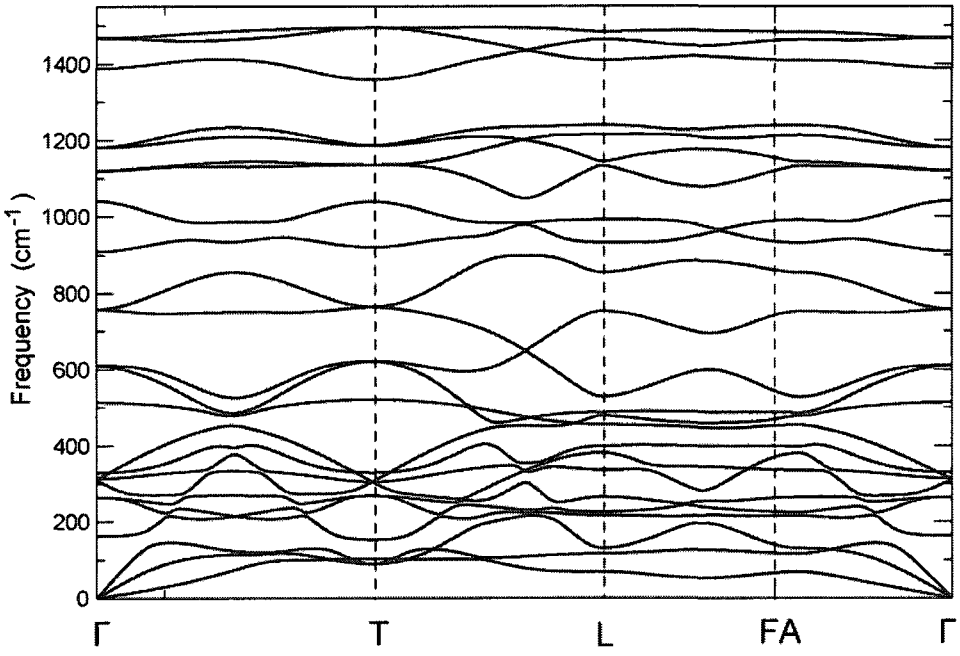


Figure 3.17. Phonon dispersion curves of the NG structure calculated along symmetry-irreducible paths in its R-3 symmetry brillouin zone.

Overall, the GA searches on nitrogen allotropes show the technology for recovering phases at specified pressure ranges behaves as expected, recovering molecular phases at "low" pressures and molecular/polymeric phases at higher pressures. The expected lowest-enthalpy nitrogen phases, within the constraints set by the simulation, were recovered and a number of higher enthalpy candidates were recovered. One of these higher-enthalpy candidate phases, the so-called NG phase, becomes lower enthalpy than both the cg-N and zzCH polymeric phases at ambient-like pressures, demonstrating, once again, the potential for further explorations on the polymeric nitrogen landscape.

3.2.2 – Validation on Heteronuclear Systems

The GA search entailed herein can readily be applied to much broader chemical systems than simple homonuclear, high-pressure solids. To test the applicability of the

GA algorithm on heteronuclear systems, the GA search procedure was applied to search out $CaCO_3$ and SnO_2 crystals. To accommodate this, the genetic representation of a crystal structure from Section 3.2.1 was modified to include a list atomic numbers, thereby mapping each set of atomic cartesian coordinates to an element of the periodic table. In this Section, the mapping was supplied by the user at the beginning of the GA search procedure and kept fixed throughout each trial.

The fundamentals of all the mutation operations for heteronuclear systems were left unchanged from those used for the homonuclear trials presented in Section 3.2.1. The only difference in their implementation was that the "Atom Swapping" mutation was turned "on" to allow the swapping of atomic label tags among parent structures without changing the Cartesian coordinates themselves. The "Cartesian Averaging", "Atom Splicing", and "Random Mixing" mating operations were left unchanged, ie. they were coded to ignore the atomic labels and the "Spatial Splicing" mating routine was modified to transfer the parent atomic labels along with their cartesian coordinates, additionally ensuring that the number of each specific atomic label in the offspring never exceeded that of the parents structures.

Post-aragonite $CaCO_3$

In 2006, Oganov *et. al* published a GA study of $CaCO_3$ in the hope of shedding light on its high pressure post-aragonite allotropes.[11] At ambient pressures, $CaCO_3$ assumes its well known calcite crystal structure, one of the most dominant carbonates in the Earth's crust.[30] Application of pressures typical of the planet's upper mantle induce a transformation to the aragonite form of $CaCO_3$, and further pressure application beyond 40 GPa induce a transformation to, what had been, a previously uncharacterized and

recently-observed post-aragonite form of $CaCO_3$.[31] Oganov *et. al*'s GA study of high-pressure $CaCO_3$ shed light on this post-aragonite form of calcium carbonate and discussed even further transformations of $CaCO_3$. In the hopes of validating the evolutionary GA protocols used in this work, similar $CaCO_3$ systems were run to help tune the specific mating/mutation routines of the GA algorithm.

i) Reproducing the Structure Distribution

In an effort to reproduce the results of Oganov *et. al*,[11] eight GA trials with twenty genes per population were run to generate $CaCO_3$ structures in an orthorhombic cell with fixed lattice parameters of a=4.101, b=4.561, and c=3.964 Å. Given the relatively small search space, the fitness metrics were extracted directly from single-point energy evaluations as opposed to full geometry optimizations. Each trial progressed for ten generations and, aside from the optimization scheme as discussed above, employed the same simulation parameters used for high-pressure nitrogen in section 3.2.1 with regards to mutation/mating rate, promotion scheme, energy resolution, etc. Since single-point energy calculations were used during the trial itself, it was necessary to implement screening subroutines of each gene's radial distribution to discard redundant or similar structures from the gene pool without relying on the energy resolution to successfully select potential copies first. Final geometries were fully optimized and select structures recovered from the GA trials are shown in Figure 3.18.

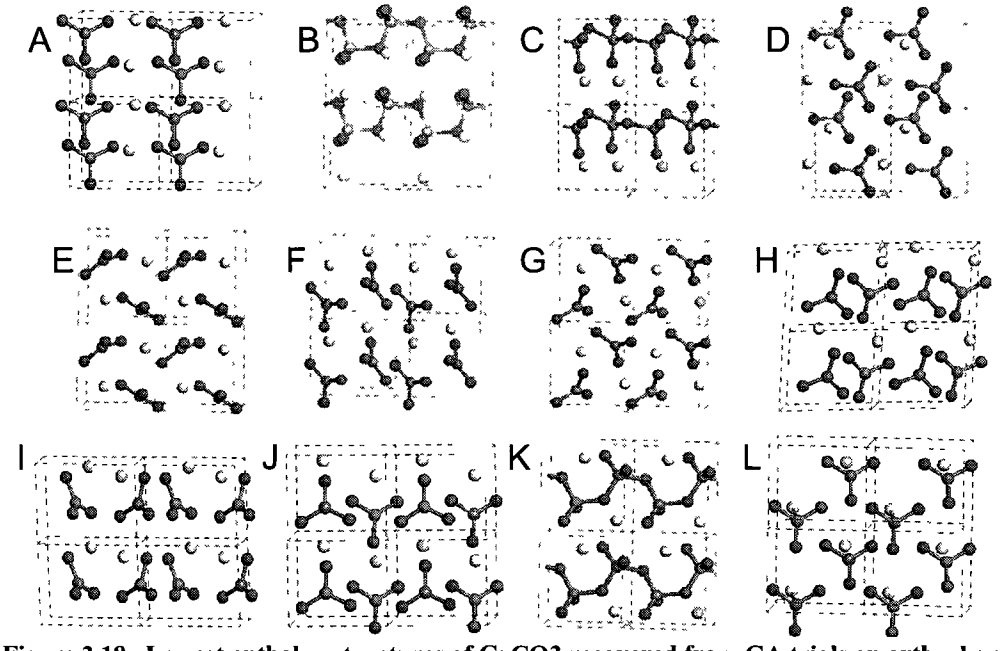


Figure 3.18. Lowest enthalpy structures of CaCO3 recovered from GA trials on orthorhombic cells with lattice parameters: a=4.101 Å, b=4.561 Å, and c=3.964 Å. Structure J corresponds to the post-aragonite structure proposed in ref. 2.

The lowest enthalpy structure J, the proposed structure of post-aragonite,[11] was recovered in seven of the eight GA trials. The relative enthalpies of the structures pictured in Figure 3.18 are presented in Table 3.1. The results agree with those reported by Oganov *et. al*,[11] showing there are many different structures competitive, but never lower, in enthalpy with structure J. All of the structures contain either CO_3^{2-} tetrahedra, like structures A and J, or CO_4^{4-} tetrahedra, as in structures B and C, an observation well noted in ref 11.

Table 3.1. Relative enthalpies of the most fit 12 structures from the GA trials on $CaCO_3$.

Structure	Enthalpy after VASP optimization (eV/atom)
A	0.0
B	0.001
C	0.165
D	0.160
E	0.085

F	0.062
G	0.083
H	-0.010
I	0.025
J	-0.087
K	0.010
L	0.183

Although a step in the right direction, the ten-atom unit cell of $CaCO_3$ is quite small and fails to demonstrate how powerful the GA evolutionary procedure would be to navigate a complicated PES. Four additional trials were done using 1x1x2 supercells of the original unit cell of post-aragonite as the GA simulation cell. The first four trials were done with forty, sixty, eighty, and one hundred genes per population, running for 30-40 steps before they were terminated.

The trials with forty and sixty genes per population all failed to recover any of the structures reported from the unit cell trials, building mostly amorphous-like structures with no symmetry. The trial with eighty genes per population recovered ordered phases resembling structures C, I, and G from the unit cell trials. The largest 1x1x2 unit cell GA trial with 100 genes per population only successfully recovered structures A and L. Incidently, structure A is the structure predicted to succeed post-aragonite $CaCO_3$.[11]

Further testing via three subsequent GA trials with 100 genes per population on a 1x1x2 simulation cell with twenty atoms revealed no promising low enthalpy candidates. The results indicate that using single point energies to evaluate unrestricted PES surfaces biases the algorithm to fall into a potential energy well, or, more specifically, become trapped in an area of the PES that optimizes to a common local minimum. This sets up a situation wherein the gene pool quickly becomes stagnant, since the further a search falls into a PES well, the more likely it becomes that the mating/mutation operations will

produce structures that optimize to identical structures. In this light, it can be rationalized that single-point energy GA optimization schemes only work for unconstrained PES surfaces when dealing with a small system, and even then care must be taken to screen out redundant structures to avoid replicas of the same gene in the population. Such screenings are ambiguous however when considering the continuum of structural states available even in confined regions of a PES, like a potential energy well, since the threshold between what is to be retained and what is to be discarded is ill-defined at best.

Furthermore, the complexity of mapping 20 atoms spanning three separate elements on the periodic table results in an inordinate number of degrees of freedom, thus most GA mating routines would inherently destroy the parents' structural motifs. More generally: the larger the system, the less likely it is that mating/mutation will preserve the structural motifs of the parent structures. Figure 3.19 shows the breakdown of successful mating/mutation operations that generated their generation's lowest energy structures throughout seven GA trials on the 1x1x2 $CaCO_3$ simulation cell with 100 genes per population. The pie chart shown in Figure 3.19a shows the successful operations from the first five generations and Figure 3.19b shows the outcomes from the remainder.

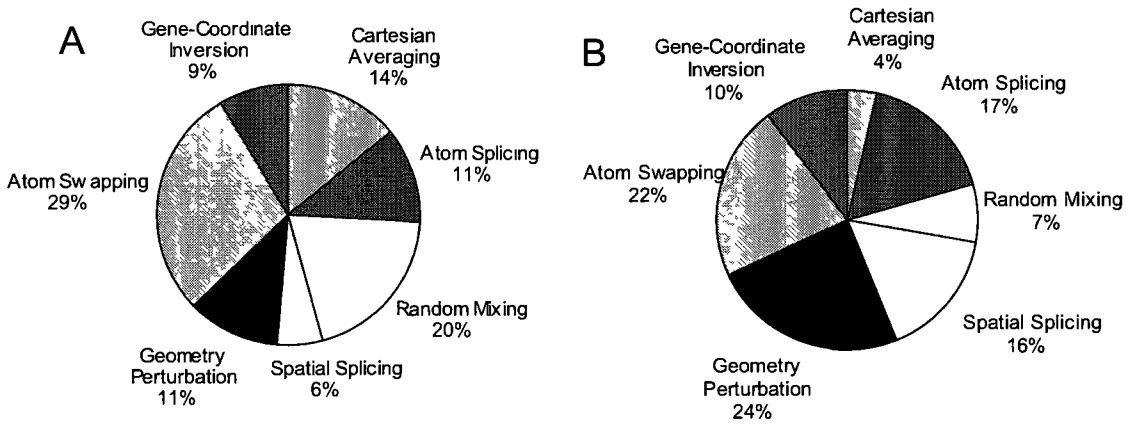


Figure 3.19. Pie charts showing the mating/mutation operations that successfully generated their generation's lowest energy structures throughout seven GA trials on the 1x1x2 $CaCO_3$ simulation cell. A) spans the first five generations, and B) shows the rest.

The successful outcomes of the first five generations of these test trials were dominated by the Atom Swapping and Random Mixing mutation and mating routines. Meanwhile, the later stages of the GA trials were more biased toward the Gene Perturbation, Atom Swapping, Atom Splicing, and Spatial Splicing routines. Incidentally, the most frequent operations from the latter stages of the trial are the routines that best preserve the physical ionic coordinates of the parent structures, inducing the fewest lattice distortions. The Spatial Splicing and Geometry Perturbation routines, which preserve most of the local constituent environments by design, show the largest net increase while, conversely, the Cartesian Averaging and Random Mixing operations, which are almost guaranteed to disrupt the local bonding environments of the constituent atoms, experience the most dramatic drop. From these results, it becomes apparent that less-physical mating routines are favored in the early stages of the trials, due the increased probability of sampling unexplored parts of the PES, and motif-preserving routines are favored in the latter stages, due to the increased likelihood of exploring relevant parts of the PES. The GA code was therefore modified to allow user-

specification of the chances of selecting each mating/mutation routine at each generation. Structure J was recovered in one of two GA trials n 1x1x2 $CaCO_3$ simulation cells which were set up such that later stages of the algorithm were three times more likely to select the Gene Perturbation, Atom Swapping, Atom Splicing, and Spatial Splicing routines.

Auxiliary trials designed to test the effect of the Boltzmann-temperature term used for selecting parent structures on the performance of the GA revealed an evident relationship between convergence speed and temperature. Ten additional GA trials on 1x1x2 simulation cells of $CaCO_3$ were run as discussed above for forty generations, where five trials were run to give the 3rd most fit structure in the gene pool a 25% chance to be selected as a parent (the low-temperature case study) and five trials were run such that the 6th most fit structure in the gene pool is given a 25% chance (the high-temperature case study). The four GA trials to recover the lowest-enthalpy $CaCO_3$ structures from 1x1x2 simulation cells were all high-temperature case studies, but note that none of these trials recovered structure J itself. A representative trial from both sets is shown in Figure 3.20. The solid line shows the relative energy of the lowest-energy gene to structure J at each generation, giving a measure of the absolute convergence of the GA trial since the expected lowest-energy structure is structure J. The dotted lines plot a metric that describes the variance in energies of the promoted structures, thereby reflecting the genetic diversity in the population at that particular generation. Specifically, they show the summed fitness values over the six lowest-energy structures.

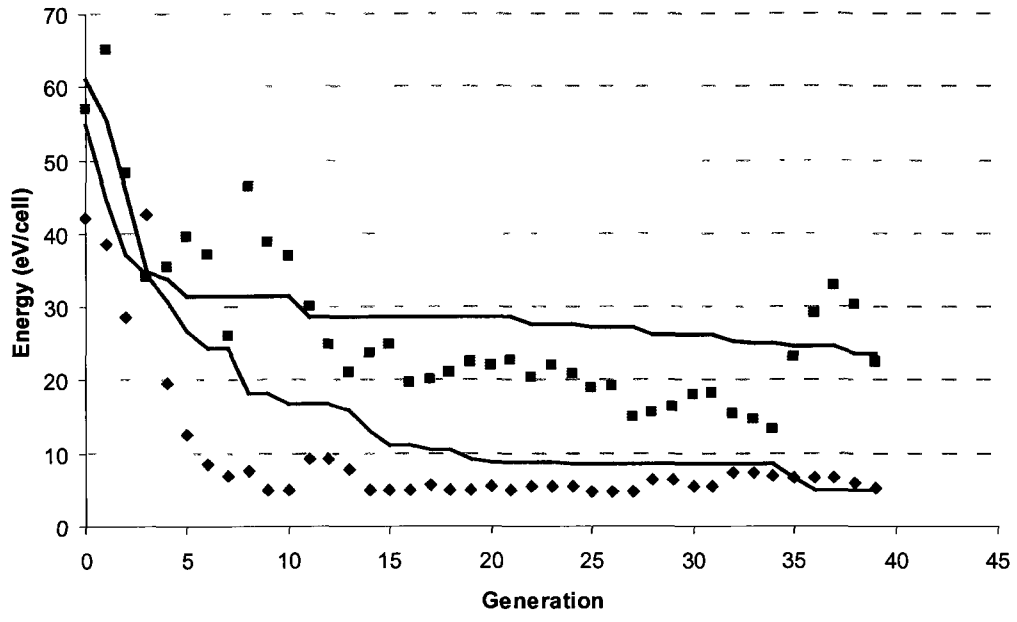


Figure 3.20. Plot showing the energy variance of the six most fit structures from two select GA trials on 1x1x2 $CaCO_3$ simulation cells, one with a lower (blue diamond) and one with a higher (pink squares) re-scaling temperature. The color-coded solid lines show the evolution the enthalpy of the fittest structure, relative to structure J.

As indicated by the relative energies of the six most fit structures over the course of the trial (ie. the dotted lines in Figure 3.20), the temperature influences the genetic diversity in the gene pool itself. The low temperature trials pick the lowest-enthalpy structures almost exclusively for mating/mutation operations, thus hindering their ability to generate radical new structural motifs and resulting in the addition or retention of energetically-similar structures in the genepool. Conversely, the high-temperature case studies are constantly sampling broader subpopulations and thus more consistently finding new and unique members to add to the populations. It was noted that when structures more fit than any others in its gene pool are recovered, their addition is typically accompanied by a substantial spike in the energy spread of the genepool, as shown in Figure 3.20, which diversifies the gene pool even further. There is a fine

balance of course as two subsequent GA trials which gave the 10th most fit structure in the gene pool a 25% chance to be selected as a parent fared poorer than the low-temperature trials. The GA code was modified to allow user-specification of the effective selection temperature, with the default being set to assign at least a 25% probability of selecting the top 15 percentile of the population.

Another way to improve efficiency of the GA is to eliminate as many degrees of freedom as possible. When four calcium atoms and four planar CO_3 molecular units were placed in the 1x1x2 simulation cell, thus replacing the twelve translational degrees of freedom of each CO_3 molecule's constituent atoms with the three translational and rotational degrees of freedom of the molecule, phases A, D, E, F, G, J, and L were recovered from two separate GA trials with 50 genes per population after propagating the simulations for only thirty generations. Such an implementation required removing the accounting of six degrees of freedom from the code and adding new structure-screening subroutines to ensure the van der Waals radii of the CO_3 molecules and Ca atoms never overlapped. Other higher enthalpy structures with native 1x1x2 unit cells were observed but not catalogued here since the goal of this particular work is to evaluate the effects of scaling the degrees of freedom on the GA search. However, note this practice of introducing additional constraints to the GA makes it such that none of the sp^3 hybridized carbon phases with CO_4^{4-} tetrahedra can be generated, like structures B and K from Figure 3.18, barring major geometric distortions from geometry optimization. Nonetheless, such mappings could prove effective in mapping organic crystals to study polymorphism where the molecular subunits are well-defined and, in fact, have been in the recent literature.[32]

ii) Fitness Evaluation

Alternative methods of fitness evaluation were then tested with further $CaCO_3$ GA trials. The method of evaluating energies thus far was to extract the enthalpy of candidate structures from GGA-DFT energy calculations, most effectively after loosely optimizing the ionic coordinates in fixed or variable cells. Given the nature of screening vast amounts of candidate structures, it is reasonable to attempt to save computational time by evaluating a gene's fitness with cheaper electronic structure methods, such as semi-empirical methodologies. Semi-empirical methods are built on a Hartree-Fock formalism but make several approximations when calculating the two-electron contributions to the Hamiltonian, thereby necessitating the use of parameters fit to empirical or ab-initio data in order to compensate for these approximations.[26, 33] The use of empirical parameters allows for the inclusion of electron correlation effects into the method beyond the electron exchange of Hartree Fock methods and, more importantly from an efficiency perspective, greatly enhances the compute time. In this spirit, the GA procedure was broadened to employ electronic structure calculations with the PM6[34] semi-empirical Hamiltonian, using the MOPAC (Molecular Orbital PACkage)[8] chemistry suite package. Five such GA trials were run using MOPAC to evaluate the fitness of all candidate structures in the 1x1x1 experimental simulation cell, using population sizes of fifty genes for forty generations. All the phases observed from the VASP trials except structure **I** were recovered from the trials, with structure **J** still being the lowest in energy.

As mentioned in Section 3.1, it could prove advantageous to use cheaper computational methods to screen structures for GA trials at higher levels of theory. The

motivation behind such practices would be to perform pseudo-"exhaustive" PES searches at an effective high level of theory that would otherwise be unfeasible. In this light, alternative fitness schemes to those discussed earlier in this section, which employ both VASP and MOPAC energy evaluations in the same GA trial, show encouraging results on $CaCO_3$. When full geometry optimizations are applied before determining each gene's fitness, MOPAC recovers all the structures that VASP recovers, therefore MOPAC could readily be used to screen low-energy structures and read them directly into the initial population of a VASP-based GA trial. It can readily be seen that by using MOPAC as a screening program to generate fit populations before turning the procedure over to VASP calculations, substantial computation time can be saved while still exploring the PES at an effective DFT level.

An even more effective cost-savings venture would be to employ MOPAC to evaluate only single-point energies of the structures as part of its screening procedure. To test such a scenario, five GA trials were set up with twenty or forty genes per population and run for fifty generations, the details of each trial is discussed below and their results are tabulated in Table 3.2. For trials 1, 2, and 4, the fitnesses were evaluated from PM6 single point calculations for the first 49 generations and the final generation was assessed from DFT geometry optimizations. Note the particular "optimization scheme", the formula which determines how the electronic structures are determined at each step, used in each trial is reflected in Table 3.2 under the column labeled "optimization scheme". Distinct low enthalpy structures were recovered from the trials within six hours from when the jobs were submitted over 20 processors. GA trials using VASP geometry optimizations to evaluate their fitness metrics, with 20 genes per

population, took, on average, 53 hours using the same computational resources as those reported in Table 3.2.

Table 3.2. Results of GA trials on $CaCO_3$ using "tandem" MOPAC/VASP energy evaluation.

Trial #	*Number of Generations*	*Genes per population*	*Optimization Scheme*	*Lowest MOPAC Enthalpy before Optimization*	*Enthalpy after VASP Optimzation*	*Structures Recovered*
1	50	20	49 mopac 1 vasp opt	-2005.8069	-67.75	J, A
2	50	20	49 mopac 1 vasp opt	-2013.408	-66.21	D, M, G
3	50	40	47 mopac 2 vasp 1 vasp opt	-2017.930	-65.98	K, I
4	50	40	49 mopac 1 vasp opt	-2015.959	-66.39	A, I
5*	50	40	49 mopac 1 vasp opt	-2025.950	-60.9	sp^3/sp^2C's

* one vasp step was inserted every 10^{th} generation from trial #5

For trial 3, the fitness metrics were evaluated from PM6 single point calculations for the first 47 generations, from DFT single point energy calculations for the 48^{th} and 49^{th} generations, and from DFT full geometry optimizations for the final step. Curiously, this resulted in poorer results even though lower-enthalpy structures were recovered during the MOPAC stage of the trial. Further inspection revealed all of the elite, or low enthalpy, structures at the PM6 level of theory were discarded as parent structures when DFT calculations were employed. The electronic energies calculated using VASP were higher in the structures created during the elitism promotion scheme then during the random or mating/mutation scheme. A similar optimization scheme to that employed in trial 1 was used in trial 5, but a single-point DFT evaluation was used every tenth step. This erroneously permitted the MOPAC PM6 procedure to converge to its lower enthalpy

structure yet, but the subsequent VASP optimizations converged to high-enthalpy structures.

These results suggest that overall span of the GA trial itself can be improved by periodically introducing major perturbations to the gene pool, as was the case for introducing VASP steps into trial 5 in Table 3.2. However, they also suggest extreme caution must be exercised when combining semi-empirical and GGA-DFT computational methodologies to static snapshots of the PES, as shown by the optimization of the seemingly promising static structures in trial 5 to nonsensical, amorphous-like crystal structures at the DFT level of theory. The take-home message is that tandem single-point energy evaluations should not be used when navigating the landscape of any PES beyond the simplest crystals. Tandem methods with full geometry-optimization fitness schemes show promise, however, and are further shown in Chapter 5 to be remarkably useful when used in materials design, specifically when applied to doped metal oxide materials.

Recovering the cassiterite form of SnO2

To further test our GA algorithm, and further validate our model for future studies of SnO_2 in Chapter 5, we also employed MOPAC in a series of GA trials to map out low-enthalpy structures of SnO_2. The trials were set up with the fixed lattice parameters of the well known 1x1x2 rutile cassiterite structure, specifically from its six-atom orthorhombic unit-cell representation. Five trials were run on the resulting twelve atom simulation cell using MOPAC single point energy calculations until the best structure remained unchanged for five generations, following by full GGA-DFT ionic and lattice relaxation. Samples of the six lowest enthalpy structures are shown in Figure 3.21.

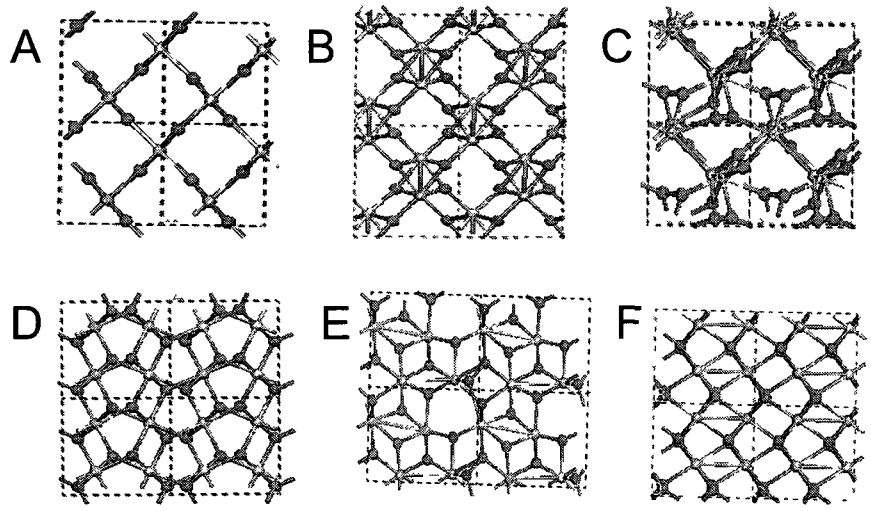


Figure 3.21. Low-enthalpy structures of SnO_2 recovered from GA-inspired search algorithms.

The expected low-pressure cassiterite structure, shown in Figure 3.21a, was recovered in all five trials. Encouragingly, structure A (cassiterite) stood out as the standout lowest-enthalpy structure of all the candidate genotypes, resting 0.120 eV/atom lower in energy than the 2nd lowest enthalpy SnO_2 structure, labeled E in Figure 3.21.

Table 3.3. Relative enthalpies of the select structures from the GA trials on SnO_2.

Structure	Enthalpy after VASP optimization (eV/atom)
A	0.0
B	0.125
C	0.281
D	0.158
E	0.120
F	0.339

3.2.3 – GA studies of Nitrogen/Hydrogen Allotropes

The difficulties in obtaining stable forms of low-pressure, pure-nitrogen allotropes as high energy-density materials are well-documented, as discussed in Chapter 1. As such, a number of research efforts devoted to designing molecular energetic materials

have shifted their focus toward developing molecules with high nitrogen content. A similar train of thought can be applied to designing high-pressure polymeric materials with high nitrogen content, for example hydronitrogen or perhaps carbon/nitrogen allotropes. Such materials would not be as effective as pure nitrogen allotropes in regards to high energy-densities, since they would almost certainly have nitrogen decomposition products other than dinitrogen, but the tradeoff is that they could be more stable at low pressures and perhaps more feasible to synthesize experimentally.

Given the success of the GA routines at recovering the coordinates of $CaCO_3$ and SnO_2, several GA trials were run on high-pressure simulation cells containing nitrogen and hydrogen in an effort to explore potential high-pressure allotropes. The initial structures were randomly generated by distributing four nitrogen atoms and four hydrogen atoms in a 30 $Å^3$ unit cell, a reasonable volume to expect for polymeric hydronitrogen allotropes at 90 GPa. The structures were optimized under variable cell conditions until the trace of the stress tensor was found to be 90 GPa. A number of low-enthalpy polymeric structures were recovered, but two particularly interesting phases, shown in Figure 3.22 and Figure 3.23, stood out among the rest. The lowest enthalpy phase, that shown in Figure 3.22, is reminiscent of a hydrogen capped zzCH phase. Its ctCH analog, the hydrogen-capped ctCH phase shown in Figure 3.23, was also recovered.

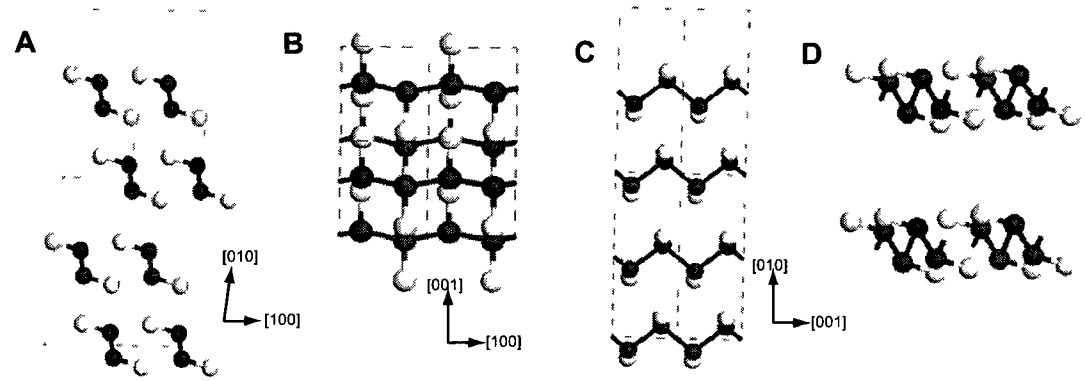


Figure 3.22. Structure of the H-capped zzCH phase. 2x2 projections normal to the lattice vectors of the structure are shown along A) [001], B) [010], and C) [100] axes. D) The extended zig-zag structure.

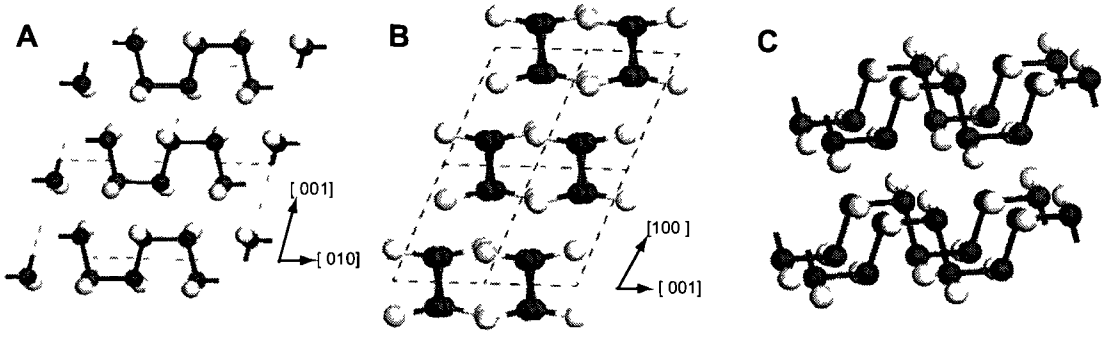


Figure 3.23. Structure of the H-capped ctCH phase. 2x2 projections normal to the lattice vectors of the structure are shown along A) [100] and B) [010] axes. C) The extended cis-trans structure.

At 90 GPa, the monoclinic $P2_1/m$-symmetry zzCH phase assumes a=8.59Å, b=2.32, c=3.00, and β=105.75° lattice parameters with its atoms distributed among N1 (6.43, 0.58, 1.99), N2 (5.54, 1.74, 1.72), H1 (6.45, 0.58, 3.04), and H2 (5.51, 1.74, 0.67) Cartesian coordinates when a is projected along the x axis and b lies in the xy plane. The triclinic P-1-symmetry ctCH phase assumes a=2.45Å, b=3.48, c=3.59, α=102.12°, β=109.94°, and γ=92.54° lattice parameters with its atoms distributed among N1 (-1.23, 1.57, -0.70), N2 (-2.31, 0.66, -2.59), H1 (-2.45,-0.04, -0.86), and H2 (-1.23, 2.54, -2.30) Cartesian coordinates.

The thermodynamic stability of the capped zzCH phase suggests it could be synthesized via ultra high-pressure synthesis techniques. However, the low-pressure NH-

structure regime must first be explored, before making such a claim, in order to understand how the overall DFT landscape relates to experimental results at those pressures and to propose a possible molecular precursor. Thus, to further unravel the structures expected at low pressures, the GA was rerun at 30 and 5 GPa. The lowest-enthalpy structures recovered from the GA were tetrazene, shown in Figure 3.24a, and ammonium azide, shown in Figure 3.24b.

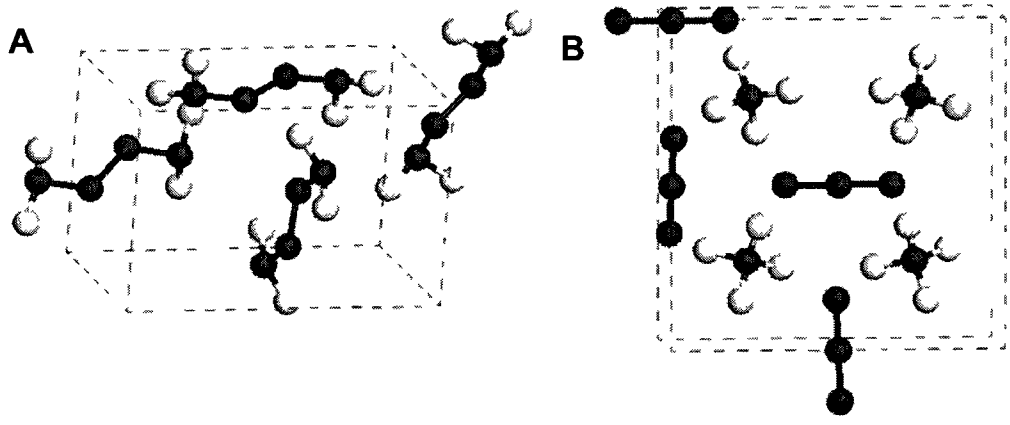


Figure 3.24. Stuctures of A) tetrazene and B) ammonium azide recovered from GA simulations of NH chemical systems.

In 1982, the crystal structure of tetrazene experimentally determined, from x-ray diffraction experiments, to have a four-molecule, triclinic unit cell.[35] The crystal recovered from the DFT GA calculations performed in this study also had four tetrazene molecules in a triclinic unit cell. Notably however, the space group was *P*1 instead of the expected *P*-1 symmetry, missing the inherent inversion-center entirely. Such discrepancies should be expected from DFT studies on this system: the *P*-1 structure was thought to arise from "weak, asymmetric hydrogen bridges forming a three dimensional network"[35] and it is well-known that density functional theory using the generalized gradient approximation (GGA) does not accurately describe weak dispersion interactions. Nonetheless, the relation of the GA results to an experimentally-determined crystal

structure is clear and further reinforces the validity of both the GA methodology itself and the notion to pursue the H-capped chainlike phases at high pressures.

3.3 - Conclusions

An overview of the motivation, theory, and implementation of the genetic algorithm (GA) into extensive explorations of the potential energy surface of select chemical systems at high pressures has been discussed. The battery of tests on nitrogen, $CaCO_3$, and SnO_2 chemical systems presented in this chapter show the technology behind the GA algorithm is in proper working order and performed comparably with other state of the art evolutionary search protocols used in solid state chemistry.[2]

The GA routines were first benchmarked on the high-pressure potential-energy landscape of nitrogen at high pressures. At high pressures, the routines were found to recover all the thermodynamically relevant phases of polymeric nitrogen derived from 8-atom unit cells proposed in the literature at the time. Furthermore, the routines were found to recover all the expected phases when the GA was used to sample a range of pressures, namely those of polymeric phases at high pressures (> 50 GPa) and molecular or chain-like structures at "low" pressures. From these searches, a previously unreported phase of polymeric nitrogen with an extended layered framework similar to graphite, aptly named the nitrogen graphite (NG) phase, was recovered as a relatively high-enthalpy phase relative to cg-N at high pressure (90 GPa). Its promise was found when its enthalpy was calculated at lower pressures, wherein it becomes ~0.44 eV/atom lower in enthalpy than cg-N and, in addition, becomes marginally lower than the zzCH structure. This made NG the lowest enthalpy structure of polymeric nitrogen at near-ambient pressures reported in the literature at the time.

The mechanics of the GA were further explored by applying it to study the post-aragonite structure of $CaCO_3$, which had recently been explored in the literature using USPEX,[11] a crystal structure prediction package which incorporates search algorithms inspired by the genetic algorithm.[2] The GA code developed in this work was found to replicate the results of Oganov *et al*[11] perfectly, even when using single-point energy evaluations to evaluate each structure's fitness. Furthermore, the $CaCO_3$ system was used to study the effect of the physical nature of the mating/mutation routines and the Boltzmann temperature, used to select preferably fit members of the gene pool for the said mating/mutation routines, on the performance of the GA. Mutation/mating routines which tend to preserve local structural environments were heavily favored to generate fit offspring near the end of the trial, while those which operated more stochastically and completely ignored local chemical environments were favored in the early stages. As a follow up, the pitfalls of high and low Boltzmann temperatures were discussed.

Finally, the GA routines were used in a predictive capacity to map high pressure allotropes of nitrogen/hydrogen solids as potential high energy-density alternatives to polymeric nitrogen. The lowest enthalpy candidate phases at high pressures are best described as H-capped arrangements of the aforementioned ctCH and zzCH phases of polymeric nitrogen (see Chapter 1.2 for further information regarding the ctCH and zzCH phases). To better rationalize the validity of suggesting these structures as relevant materials at high pressure, the GA routines were rerun at low pressures to establish a link between the GA search results and experiment. Encouragingly, the GA routines recovered a crystal structure of ammonium azide and tetrazine which relate well to those reported experimentally.

The GA algorithm as presented thus far is set up to be as general as possible, able to search the crystal space of any chemical system specified by the user and recover the lowest enthalpy phases as long as the computational means of evaluating the energies is reasonable. This brings up two trains of thought to be addressed in Chapters 4, 5, and 6: 1) Can the methodology be effectively used to study metastable chemical systems or optimize other electronic properties, and 2) How can the large expansion of the search space be accommodated when investigating larger systems? The remaining chemical systems discussed in Chapters 4, 5, and 6 address these concerns.

3.4 - References

1 D. M. Deaven and K. M. Ho, Phys. Rev. Lett **75**, 288 (1995).
2 A. R. Oganov and C. W. Glass, J. Chem. Phys. **124**, 244704 (2006).
3 P. Strak and S. Krukowski, J. Chem. Phys. **126**, 194501 (2007).
4 X. Wei, Pan, W., Cheng, L., Li, B., Solid State Ion. **180**, 13 (2009).
5 D. A. Andersson, Simak, S. I., Skorodumova, N. V., Abrikosov, I. A., Johansson, B., PNAS **103**, 3518 (2006).
6 G. Kresse and J. Furthmuller, Phys. Rev. B **54**, 11169 (1996).
7 J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal, J. Phys: Condens. Matter **14**, 2745 (2002).
8 J. J. P. Stewart, (Stewart Computational Chemistry, Colorado Springs, CO, USA, 2008).
9 J. D. Gale, J. Chem. Soc. Faraday Trans. **93**, 629 (1997).
10 W. Smith and T. Forester, Journal of Molecular Graphics **14**, 136 (1996).
11 A. R. Oganov, C. W. Glass, and S. Ono, Earth Planet.Sci.Lett. **241**, 95 (2006).
12 A. R. Oganov, J. Chen, C. Gatti, Y. Ma, Y. Ma, C. W. Glass, Z. Liu, T. Yu, O. O. Kurakevych, and V. L. Solozhenko, Nature **457**, 863 (2009).
13 Y. Ma, A. R. Oganov, Z. Li, Y. Xie, and J. Kotakoski, Phys. Rev. Lett **102**, 065501 (2009).
14 Y. Ma, M. I. Eremets, A. R. Oganov, Y. Xie, I. A. Trojan, S. Medvedev, A. O. Lyakhov, M. Valle, and V. Prakapenka, Nature **458**, 182 (2009).
15 Q. Li, M. Yanming, A. R. Oganov, H. Wang, H. Wang, Y. Xu, T. Cui, H. K. Mao, and G. Zou, Phys. Rev. Lett **102**, 175506 (2009).
16 Y. Yao, J. S. Tse, and K. Tanaka, Phys. Rev. B **77**, 052103 (2008).
17 J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett **77**, 3865 (1996).
18 J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett **78**, 1396(E) (1997).
19 N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1997).

[20] D. R. Hamann, M. Schluter, and C. Chiang, Phys. Rev. Lett **43**, 1494 (1979).
[21] G. B. Bachelet, D. R. Hamann, and M. Schluter, Phys. Rev. B **26**, 4199 (1982).
[22] J. Junquera, O. Paz, D. Sanchez-Portal, and E. Artacho, Phys. Rev. B **64**, 235111 (2001).
[23] W. D. Mattson, D. Sanchez-Portal, S. Chiesa, and R. M. Martin, Phys. Rev. Lett **93**, 125501 (2004).
[24] P. E. Blochl, Phys. Rev. B **50**, 17953 (1994).
[25] W. Kohn and L. J. Sham, Physical Review **140**, A1133 (1965).
[26] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models* (John Wiley and Sons Ltd, Chichester, West Sussex, 2003).
[27] M. M. G. Alemany and J. L. Martins, Phys. Rev. B **68**, 024110 (2003).
[28] A. K. McMahan and R. LeSar, Phys. Rev. Lett **54**, 1929 (1985).
[29] C. Mailhiot, L. H. Yang, and A. K. McMahan, Phys. Rev. B **46**, 14419 (1992).
[30] S. Ono, T. Kikegawa, Y. Ohishi, and J. Tsuchiya, American Mineralogist **90**, 667 (2005).
[31] S. Ono, T. Kikegawa, Y. Ohishi, and J. Tsuchiya, American Mineralogist **90**, 667 (2005).
[32] M. A. Neumann, F. J. J. Leusen, and J. Kendrick, Angew. Chem. Int. Ed. **47**, 2427 (2008).
[33] I. Levine, *Quantum Chemistry* (Prentice Hall, 4th edition, 1991).
[34] J. J. P. Stewart, Journal of Molecular Modeling **13**, 1173 (2007).
[35] V. M. Veith and G. Schlemmer, Z. anorg. allg. Chem. **494**, 7 (1982).

CHAPTER FOUR

SPECIALIZED GENETIC ALGORITHM APPLICATIONS WITH POWDER X-RAY DIFFRACTION

This chapter addresses the practice of incorporating experimentally-assigned symmetry elements into evolutionary-inspired structure searches such that experimentally-consistent results are obtained. In particular, it is emphasized that the use of additional constraints is sometimes required in order to account for deficiencies of the computational method or improper indexing of the powder x-ray diffraction patterns.

CHAPTER 4 – *KEY SUBSECTIONS*

4.1 - Introduction

Search algorithms inspired by the genetic algorithm (GA) were shown in Chapter 3 to be effective at mapping out the potential energy surface (PES) of several chemical systems, with high-pressure solid nitrogen among them. It is shown in this chapter that these algorithms can readily, and relevantly, be applied to study high-pressure molecular nitrogen as well, even though N_2 solids are only metastable beyond 50 GPa thresholds. Furthermore, this approach can be generalized further to tune the algorithm to search out any metastable class of crystal that is sought to be explored.

Consider that the characterization of such simple systems as homonuclear diatomic molecular solids remains a fundamental problem in the field of ultra-high pressure materials synthesis, since one often has limited X-ray diffraction data available due to the harsh experimental conditions and/or weak scattering properties of the atoms under study. High pressure phases of molecular oxygen and nitrogen, in particular, have been actively studied. This is highlighted by the recent identification of the structure of an elusive molecular oxygen phase using single-crystal X-ray diffraction data[1, 2] and an ongoing debate over the nature of high-pressure phases of molecular nitrogen.[3-11] These solid molecular structures are sought because they are typically intermediates to even more interesting high-pressure phases, such as the superconducting high pressure ζ-O_2 phase of oxygen[12] or the high energy-density polymeric form of nitrogen.[13]

These examples illustrate cases where adequate-sized single crystals cannot be synthesized in the lab, forcing the material's characterization to come from powder diffraction experiments. Unfortunately, traditional approaches toward structure solution with single-crystal x-ray diffraction, those derived from transforming the integrated peak

intensities of each diffraction minima in reciprocal space, are complicated when applied to powder-diffraction patterns by the increased peak overlaps from the grain orientations in the crystal.[14] For example, the Rietveld profile refinement techniques used to extract high-quality structure solutions from powder diffraction data cannot effectively be applied unless a reasonable structure, or class of candidate structures, is first proposed.[15, 16] Thus, precise atomic coordinates can be elusive in difficult systems. In spite of these difficulties, the lattice parameters and space group are still usually quite trustworthy even when dealing with harsh experimental environments. The lattice parameters of the unit cell can reliably be extracted from powder X-ray diffraction by studying the periodicity of the diffraction pattern in real space. Although somewhat less reliable, the symmetry operations in the crystal's unit cell can usually be determined from noting systematic extinctions of peaks in the diffraction pattern, thereby allowing the characterization of the crystal's space group from periodic absences of reflections in reciprocal space.

To assist efforts to solve the atomic coordinates in these difficult systems, a fitting theoretical approach toward structure determination or refinement would be to use GA-inspired search procedures to identify the most enthalpically-favored configurations that satisfy the lattice or symmetry requirements. This application actually falls right into line with the strengths and weaknesses of a GA; as discussed in Chapter 3, the biggest limitation of a GA is its demand on computational resources as the effective search space increases. Such limits could be mitigated in structure-refinement studies by introducing restrictions on the GA algorithm to only search out structures consistent with the experimentally-derived lattice parameters and space groups. Additional constraints, such as restricting the connectivity between the constituent components of the crystal or

biasing the procedure to search out specific material classes, could readily be employed to ensure that only thermodynamically relevant, or even metastable, structures are explored.

The idea of using experimentally-derived lattice parameters to aid in systematic searches of crystal structures has emerged as a staple of crystallography's direct-space approach toward structure determination, particularly when trying to identify the atomic coordinates of weak scatterers. The ever-expanding computational resources available to the research community have even allowed state-of-the-art systematic searches without lattice parameters.[17-19] However, as shown in this chapter, sometimes it is still necessary to introduce constraints into GA explorations in order to reproduce findings consistent with experiment. For example, constraints must be introduced to explore the elusive ζ–N_2 molecular phase, since it is metastable to polymeric nitrogen at the 80 GPa threshold it is observed at, and the high pressure ε–O_2 molecular phase, whose coordinates have never been recovered by theoretical means in spite of a number of attempts.[1]

The power of such modified GA-inspired searches is demonstrated in this chapter. First, a thorough GA study of the controversial and metastable ζ–N_2 phase[8, 10] is presented; note that since molecular nitrogen is metastable beyond 50 GPa,[20] this is the first GA-inspired PES search for metastable crystal structures. Second, the symmetry-restricted GA search technology is applied to the ε–O_2[1, 2] and ζ–O_2[21, 22] phases, becoming the first theoretical method to successfully recover the ε–O_2 phase coordinates. The restricted, but thorough, symmetry-adapted GA-procedure is then shown to successfully recover the recently-proposed post-aragonite $CaCO_3$ [23] and ternary Dy_2B_4C[24] structures from the experimental powder diffraction results reported in the literature. Finally, the

GA subroutines are used to "directly" explore the lowest enthalpy phases of polymeric nitrogen at low pressures for the first time, in regimes where molecular nitrogen allotropes are thermodynamically stable.

4.2 – GA Study of the High Pressure ζ-N_2 phase

4.2.1 - Motivation

The phase diagram of high pressure molecular nitrogen has been studied extensively, but attempts to elucidate the correct molecular crystal structures have proven challenging. Advances in X-ray diffraction techniques have allowed the scientific community to fully characterize the phase diagram up to pressures of ~30 GPa, but beyond this threshold the limits inherent to X-ray diffraction measurements of low-Z materials make data analysis difficult. It is well known that the ε-N_2 phase of molecular nitrogen, the highest pressure phase with a solved molecular crystal structure,[25] undergoes a structural transformation beyond 60 GPa to the so-called ζ–N_2 phase. The ζ–N_2 phase is proposed to be a precursor to the cubic gauche phase, cg-N, of solid polymeric nitrogen,[26] which, as discussed in Chapter 1, is a highly sought-after nitrogen allotrope for its potential use as a high energy density material at ambient pressures. The precise molecular structure of the ζ– N_2 phase is, to date, unconfirmed but could prove invaluable in understanding the limitations and feasibilities of the molecular to polymeric transition that is currently being intensely studied.

There have been a couple of different, and conflicting, analyses of the ε- N_2 to ζ-N_2 phase transition reported in the recent literature.[8, 10] In both studies the cells were indexed as primitive orthorhombic but they have markedly different lattice parameters. To simplify further discussion, the cell reported by Eremets *et al* [10] will herein be referred

to as Cell A and the cell reported by Hemley *et. al*[8] as Cell B. The *a*/c and *a/b* lattice vector ratios for the two cells are 0.83 and 1.50 versus 0.95 and 2.53 for Cells A and B, respectively. This constitutes a significant difference and suggests that two different phases with different crystal structures may be involved. Furthermore, the reported space groups consistent with the observed systematic extinctions in the diffraction patterns do not agree. Specifically, the reported space groups for Cell A are consistent with $P222_1$, $P2_12_12$, and $P2_12_12_1$ symmetries and the space group for Cell B has been assigned as *P*mma. Since the screw axes of the space group $P222_1$ are in the same crystallographic direction as those in the space group $I2_13$ of cg-N, the space group $P222_1$ was selected to be the best candidate for Cell A.[10] However, questions were raised if this space group could be one of candidates for the structure of the ζ-N_2 phase based on the number of measured Raman active vibronic and lattice modes.[7]

In this section, the discrepancy in the reported ζ-N_2 unit cells will be explored by performing an exhaustive GA search of molecular nitrogen's potential energy surface at the DFT level of theory.[27, 28] Structures compatible with each of the experimental cells are identified and plausible ζ- N_2 phase candidate structures are proposed, including stability analyses.

4.2.2 – Computational Method

As discussed in Chapter 3, applications of GA-inspired search procedures to high-pressure polymeric nitrogen have recently been reported, demonstrating the applicability of this algorithm to studying nitrogen systems.[29, 30] The specifications of the general GA algorithm in the context of solid-state structure searches was discussed at length in Chapter 3.1. Overall, these GA studies[27] traditionally represent a crystal by a list of N

atomic Cartesian coordinates, requiring no prior information about the system aside from the composition and volume of the unit cell. Furthermore, these algorithms will typically seek out the lowest enthalpy phases for whatever conditions are defined since they are designed to optimize an intrinsic fitness metric defined entirely by each crystal's relative enthalpy.

Given that polymeric phases of nitrogen are the lowest enthalpy phases by ~0.25 eV/atom at the pressures of interest (80 and 90 GPa) [20], the traditional GA procedure was modified for the work presented in this section to allow for an efficient mapping of the potential energy surface pertaining to molecular nitrogen. The crystal structures were represented by the Cartesian coordinates of each N_2 molecule's centre of mass, along with a corresponding set of spherical coordinates to define the remaining angular degrees of freedom. For structure evolution, the number of molecules was kept constant for the duration of the procedure wherein, at each step, randomly chosen mating operations were used to combine structural motifs of two randomly chosen "parent" genes. The mating operations were designed to introduce ample deviations for exploring un-sampled, low-enthalpy regions of the molecular PES. The mating operations used in this work were the following (building off the operations described in Chapter 3): i) Random Mixing: randomly choosing a subset of molecules from each parent to make the whole offspring, ii) Arithmetic Averaging: taking the arithmetic averages of the molecular coordinates iii) Geometric Averaging: taking the geometric averages of the molecular coordinates, iv) Gene Coordinate-Inversion: swapping individual Cartesian coordinates (x, y, or z) among a randomly chosen subset of molecules and v) Spatial Splicing: cutting out a cubic subsection from one parent structure and replacing that sub-structure with that of the

other parent structure. Only one randomly chosen mating operation was chosen to generate a single offspring. If the mating procedure changes the number of molecules in the system, then N_2 molecules are randomly added or removed to obtain the correct number. Often these mating procedures generated physically unreasonable crystal structures. Thus, offspring structures were only added to the next generation if the atoms between any two N_2 molecules were at least 0.85 Å apart, otherwise the same mating operation was repeated. The resulting offspring structures are given a twenty-five percent chance to mutate. The mutation used in this study involved randomly selecting between one and three N_2 molecules and randomly translating, rotating, or changing the N-N bond length to lie between 1.03 and 1.21 Å in the selected molecules.

The fitness of a given crystal structure is derived largely from its enthalpy as calculated from first-principles DFT calculations. The Vienna Ab-Initio Simulation Package (VASP)[31] was used for all enthalpy calculations. The projector augmented wave (PAW) method of Blöchl[32] was used to treat the core states and the nitrogen 2s and 2p electrons were treated as valence electrons with a plane-wave cutoff of 500 eV for geometry optimizations and 800 eV for energy evaluations. The gradient-corrected exchange and correlation functionals of Perdew-Burke-Ernzerhof (PBE)[33] were used in all calculations. For energy evaluations within the genetic algorithm, the Brillouin-zone integrations were performed using 3x3x3 Monkhorst-Pack grids whereas 10x10x10 meshes were used in more refined calculations for the phases selected for further study. Although costly, the density functional level of theory is necessary to adequately describe the high-pressure, solid, molecular-nitrogen electronic structure and have been shown to be sufficient for the arguably more complex high-pressure O_2 PES;[22] benchmark

calculations show that cheaper alternatives, such as force fields, are generally not transferable enough to be of value for the broad scope of the problem presented here.

With regards to mapping out molecular structures, the search algorithm was found to be more efficient if the fitness function was modified to penalize structures that were non-molecular by adding a correction term for every intermolecular bond distance that was less than 1.65 Å. The correction term was taken from a harmonic potential fit to give no correction for intermolecular distances of 1.65 angstroms and add a 1.25 eV/atom penalty for 1.41 angstrom distances, this being the N-N bond distance within the polymeric cubic gauche structure.

A population size of 100 individuals was utilized for each generation. The procedure for generating new populations remained, for the most part, unchanged from the previous GA method. The initial population was filled with randomly generated structures, ensuring that all configurations possessed inter-atomic distances greater than 0.85 Å. New generations were specifically created in the following manner. First, the ten lowest enthalpy structures were promoted, unperturbed, to the next generation. Second, five structures were generated with randomly assigned coordinates and added to the new population. To generate the rest of the population, low enthalpy structures were preferentially chosen as parent structures by selecting the parent structures from a Boltzmann weighted probability distribution.[27] The exponential term stems from the calculated enthalpies, scaled by an appropriate temperature to allow at least a 25% chance to select half of the promoted structures. Unphysical structures are screened out during the mating process and the remaining generated structures are loosely optimized, such that the magnitude of the largest atomic force is less than 0.05 eV/Å^2, before evaluating

their fitness. The genetic algorithm was halted when the energy of the four lowest energy structures remained unchanged for five generations, typically running for, on average twenty generations. At least five distinct production runs were employed resulting in at least 10000 structures being evaluated and partially optimized at the DFT level for each cell.

Phonon calculations, used to evaluate the mechanical stability of the phases generated from the genetic algorithms, were done with the finite differences method using VASP and the external package Fropho[34] and subsequently confirmed using forces calculated from SIESTA[35]. SIESTA calculations used a Troullier-Martins norm-conserving pseudopotential, referencing a $[He]2s^2 2p^3$ configuration with a 1.24Å cutoff. Custom numerical doubled s, p_x, p_y, and p_z orbitals and d polarization orbitals were used with a real-space mesh cut-off of 200.0 Ry. A 20 Å cutoff for k-point sampling was used to construct Monkhorst-Pack grids for Brillouin-zone integration. Simulated X-ray patterns were calculated using the Reflex module from the Materials Studio® simulation package version 3.2 from Accelrys, San Diego, CA, USA.[36]

4.2.3 – Results and Discussion

Given the discrepancy between the reported experimental unit cells of the ζ-N_2 phase, both Cell A and Cell B were investigated by running two distinct sets of evolutionary simulations under fixed cell conditions using each of the reported lattice parameters. From these simulations, further analyses and simulations were performed on the lowest enthalpy structures with the goal of elucidating a candidate ζ-N_2 phase structure from each of the reported lattice parameters.

GA Investigations of Cell A

For Cell A[10] the proposed lattice parameters were a=4.159, b=2.765, and c=5.039 at 80 GPa for an eight-atom orthorhombic unit cell with either $P222_1$, $P2_12_12$, or $P2_12_12_1$ symmetry. Using this cell, with four N_2 molecules, seven separate GA structure searches were performed at the DFT level with a population size of 100 individuals and run for, on average, 20 generations. The six lowest energy structures recovered from this constrained-cell search are shown in Table 4.1, labeled as structures **A1-A6**.

Table 4.1. Relative enthalpies and symmetry information of the six lowest enthalpy phases found from genetic algorithm based structure searches for Cell A.

	Fixed Cell[a]		*Variable Cell*[b]			
Structure	Enthalpy *(eV/atom)*	Symmetry	Enthalpy *(eV/atom)*	Symmetry	Lattice Ratios *(a/b, a/c)*	Lattice Angles (α,β,γ)
A1	0.0	$P2_1/C$	0.0	*P*bcn	1.59, 0.87	90,90,90
A2	0.185	$P2_12_12_1$	0.056	$P2_12_12_1$	1.44, 0.77	90,90,90
A3	0.330	*C*mca	0.051	*C*mca	1.24, 0.86	90,90,90
A4	0.451	*P*C	-0.001	$P2_1/C$	1.59, 0.80	90,69,90
A5	0.424	$P2_1/C$	0.042	*C*mcm	1.50, 0.75	82,90,90
A6	0.545	*P*-1	0.047	*P*-1	1.45, 0.74	79,97,89
Exp.[c]	-	-	-	-	1.50, 0.83	90,90,90

[a]Structural search using experimentally determined lattice parameters. [b]Results correspond to structures from the fixed cell structure search that have been optimized where the cell and ion positions are allowed to relax. [c]Ref. [10].

The enthalpy of each structure is reported relative to the lowest enthalpy structure **A1**, which was recovered from all seven GA structure searches. Structure **A1** is 0.185 eV/atom more stable than the next most stable structure, but its $P2_1/C$ symmetry does not correspond to any of the space groups assigned to cell A. The next lowest enthalpy

structure **A2**, however, possesses $P2_12_12_1$ symmetry, which is one of the candidate space groups proposed for cell A. Geometries of structures **A1** and **A2** are shown in Figure 4.1a and Figure 4.2a, respectively. The remainder of the low enthalpy structures **A3-A6** lie at least 0.330 eV/atom higher than **A1** and none have the symmetries that were assigned to cell A. The calculated powder x-ray diffraction patterns for structures **A1** and **A2** are shown in Figure 4.3a and Figure 4.3b, respectively, with the experimental peak positions and relative intensities also shown. The peak indices of both structures match up well with experiment, showing peaks at 9.6, 10, 10.5, 11.2, 11.5, 11.6, and 12.6 degrees. However the relative peak intensities are not as well-matched.

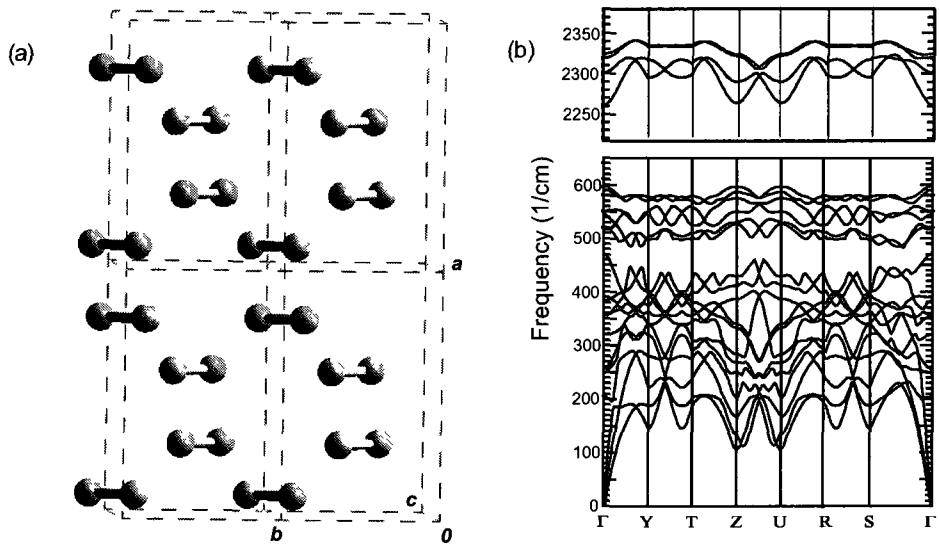


Figure 4.1. (a) 2x2x2 supercell of structure A1 following variable cell optimization. (b) phonon dispersion curves for structure A1 in the lattice and vibrational regions.

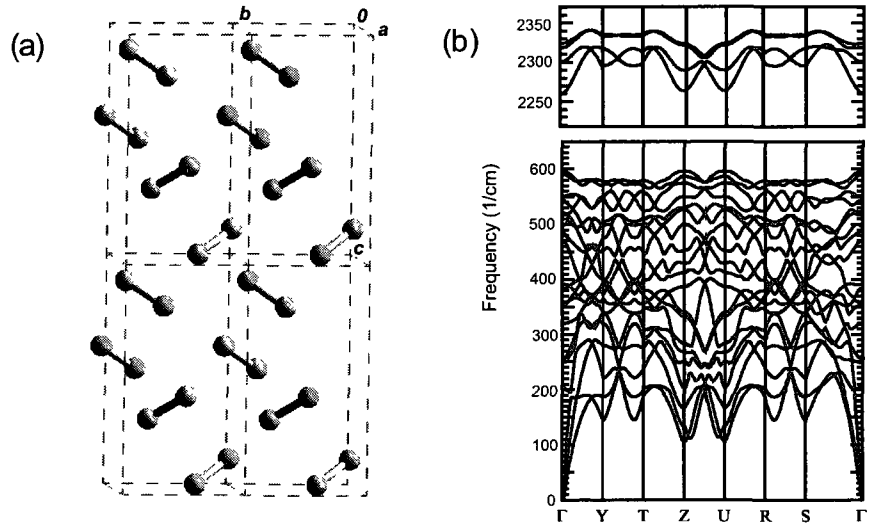


Figure 4.2. a) 2x2x2 supercell of structure **A2** following variable cell optimization. b) Phonon dispersion curves for structure **A2** in the lattice and vibrational regions

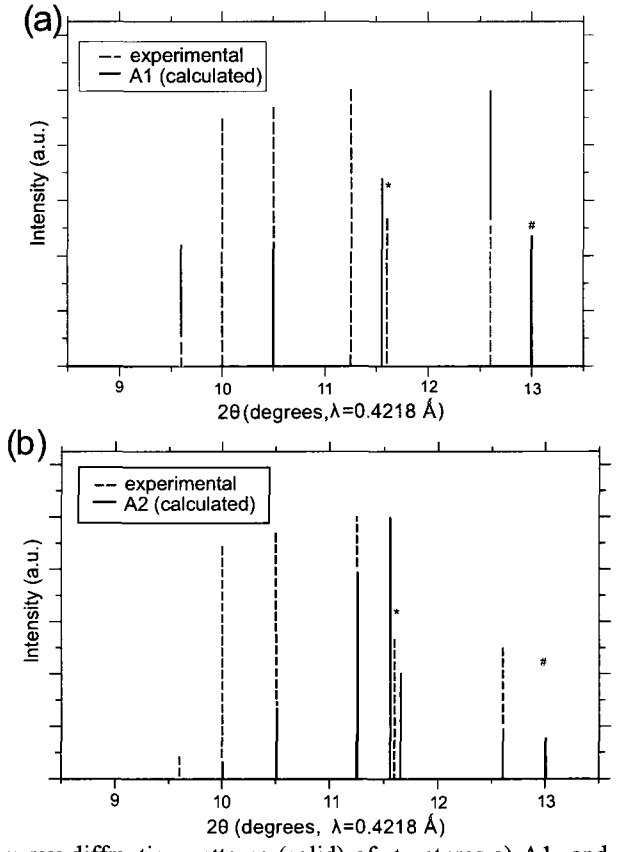


Figure 4.3. Calculated x-ray diffraction patterns (solid) of structures a) A1 and (b) A2 compared to the experimental diffraction pattern reported in ref.10 (dashed). Intensities, shown in arbitrary units, are scaled such that the peaks at 12.6° overlap fully. For the peaks labelled '*' only one peak is seen experimentally at ~11.6° with one Bragg angle position on either side. The peak labeled '#' at 13.0° is found in experiment.

Following the GA trials, structures **A1-A6** were allowed to fully relax with the cell constraints released during optimization, permitting the trace of the stress tensor to match the experimental pressure of 80 GPa. The enthalpies, lattice parameters and symmetry assignments of the relaxed structures are reported on the right-hand-side of Table 4.1. Following full relaxation, the spread in the enthalpy of structures **A1** to **A6** diminishes from 0.545 eV/atom to 0.056 eV/atom. Moreover, structure **A4** becomes the lowest enthalpy phase by a small margin over **A1**. However, following its full relaxation, **A4** also becomes monoclinic and is therefore not compatible with the symmetry analysis for Cell A. Structures **A5** and **A6** were noted to relax to non-orthorhombic unit cells and are therefore also not compatible with Cell A. Noting that molecular nitrogen itself is metastable at these ultra-high pressures, structures **A4-A6** will not be discussed further since their unit cells are not compatible with the experimental cell, making it unlikely that these are the structures of the identified ζ-N_2 phase in ref. 10. Upon full relaxation of the cell, structure **A3** lays only 0.05 eV/atom above structure **A1** and maintains its orthorhombic unit cell. However, the *C*mcm symmetry of structure **A3** does not have many common symmetry elements to the proposed experimental symmetries and therefore was not considered further.

Herein, structures **A1** and **A2** will be discussed in more detail since they remain amongst the lowest enthalpy phases and they experience the least change in their cell parameters upon full relaxation at 80 GPa. For example, their respective cells remain orthorhombic and their lattice vector ratios remain within 7% of the experimental parameters. Upon cell relaxation, the symmetry of structure **A2** remains $P2_12_12_1$ whereas the higher-symmetry **A1** structure adopts *P*bcn symmetry, which can be related to the

$P2_12_12$ space group by considering the added glide plane symmetry elements. The structural details of **A1** and **A2** are given in Table 4.2 (bottom of page 143). A calculated enthalpy (relative to the cg-N phase) vs. pressure phase diagram is shown in Figure 4.4 for structures **A1** and **A2**, along with the molecular α–N_2 and ε–N_2 phases. As expected, the cg-N polymeric nitrogen structure becomes thermodynamically favored above 50 GPa.[20] Structures **A1** and **A2** are both lower in enthalpy than either of the molecular α–N_2 or ε–N_2 phases in the 70-90 GPa pressure range. Below this range, the enthalpy-pressure curve for structure **A1** remains below the molecular phases until ~25 GPa while that of structure **A2** crosses that of the ε–N_2 phase near 68 GPa.

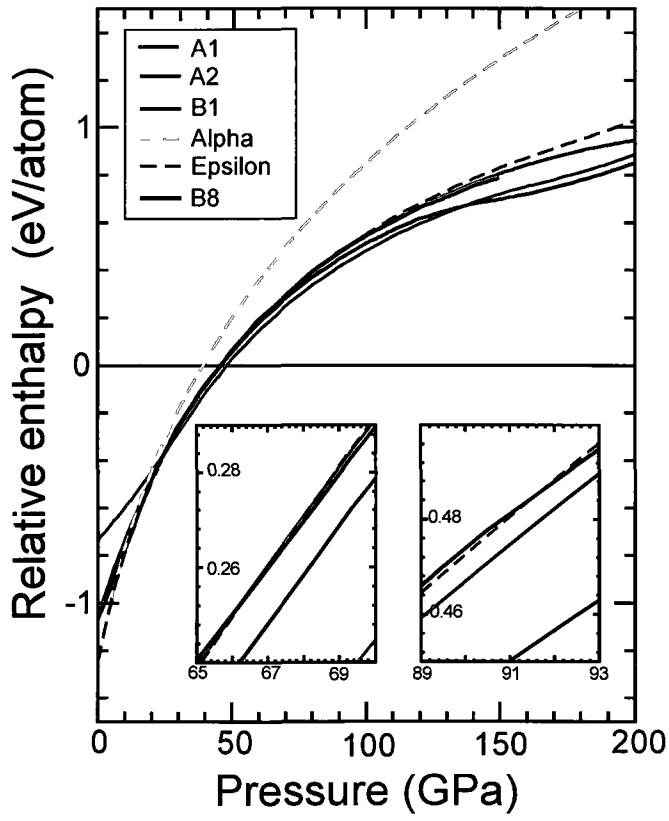


Figure 4.4. Pressure – relative enthalpy phase diagram for structures **A1**, **A2**, **B1**, and **B8**, and the α–N_2 and ε–N_2 phases of molecular nitrogen. Enthalpies are referenced to the cg-N structure. The two insets show the crossover of the ε phase with structures **A2** and **B8** between 65-69 GPa (left inset) and with **B8** at ~91 GPa (right inset).

Phonon dispersion curves for the **A1** and **A2** structures, optimized at 80 GPa, are given in Figure 4.1b and Figure 4.2b, respectively. These curves show that there are no negative modes and that both phases are mechanically stable at 80 GPa. Comparison with phonon dispersion curves at 30 and 50 GPa (not shown) show a softening of the lowest phonon with decreasing pressure, suggesting an instability in the structure as the pressure is relieved. This behavior is expected of high pressure molecular phases of nitrogen. The calculated equations of states for both structures and the cg-N structure are shown in Figure 4.5. At 110 GPa, the ζ-N_2 to cg-N transition is predicted to exhibit a large volume drop of 24.7 % and 23.4% for the **A2** and **A1** phases respectively, which correlates well with the experimentally observed volume drop of 22%.[10]

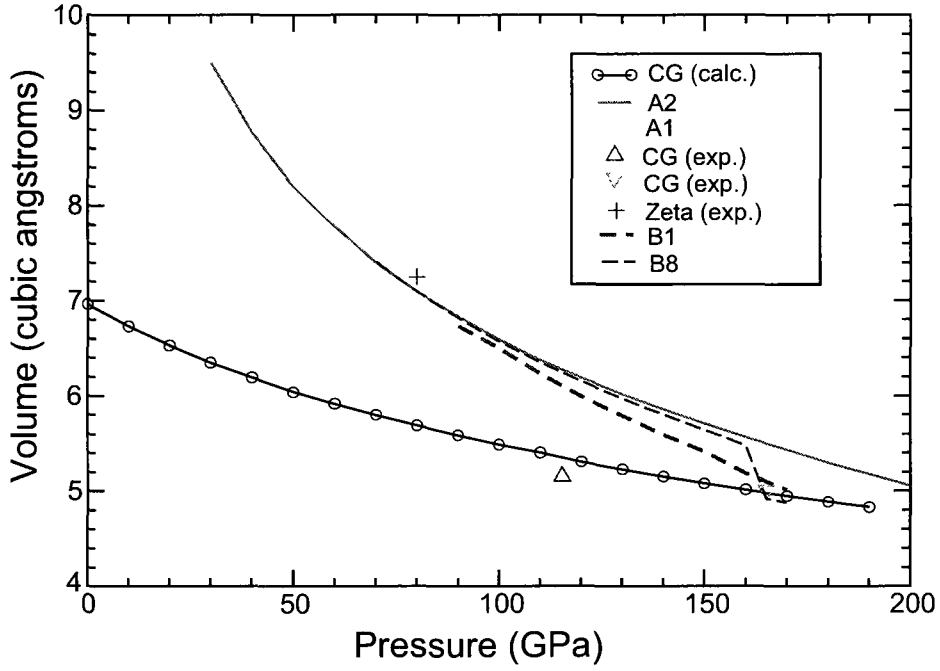


Figure 4.5. Calculated pressure-volume relations for cubic gauche (black line), structure A1 (green line), structure A2 (red line), structure B1 (violet dashed), and structure B8 (blue dashed). Experimental PV relations from ref. 13 (blue triangle), ref. 10 (black cross) and ref. 8 (pink triangle) are also shown.

Experimentally, Eremets et al.[10] report that the Raman modes were weak and difficult to interpret for Cell A. By visually inspecting the number of bands in the N-N

vibrational region of the calculated phonon spectra at the gamma point, one can obtain an upper bound on the number of available vibrational Raman modes. It can be seen that both structures **A1** and **A2** have four available vibrational modes, which can be further reduced to three if one considers frequencies within 10 cm^{-1} to be degenerate. There have been reports of high pressure molecular nitrogen in which only three active Raman modes were observed for an "anomalous" ζ-N_2 phase at 74 GPa.[5] Hemley and coworkers reported five Raman active N-N vibrational modes experimentally for cell B. From these results, it is proposed that either **A1** or **A2** is the ζ-N_2 phase characterized in Cell A[10], but, in light of the arguments given above, it is difficult to definitively assign one structure over the other. Given the better agreement of structure **A2**, however, with both the symmetry and vibrational analyses, **A2** is proposed to be the most probable candidate. Additional experiments and refinements may be required to obtain an unambiguous determination of the ζ-N_2 phase structure.

Table 4.2. The unit-cell parameters and atomic positions of selected structures resulting from the first principles genetic algorithm based structure search.

Structure	Space Group	a (Å)	b (Å)	c (Å)	Z[a]	Atoms	x	y	z
A1[b]	*P*bcn	2.680	4.249	4.908	4	N1	0.3662	0.8629	0.6640
A2[b]	$P2_12_12_1$	3.964	2.746	5.196	4	N1	0.7833	0.2502	0.178
						N2	0.6466	0.9586	0.0739
B1[c]	*I*mmm	3.450	2.965	2.636	2	N1	0.6614	0.0	0.0
B8[c]	*P*nma	5.249	7.857	2.640	8	N1	0.5795	0.5011	0.6377
						N2	0.1232	0.7500	0.6294
						N3	0.7822	0.7500	0.1471

[a]Number of N_2 molecules per unit cell
[b]Structures optimized at 80 GPa.
[c]Structures optimized at 90 GPa.

GA Investigations of Cell B

For Cell B[8] the reported lattice parameters are a=6.533, b=2.576, and c=6.844 at 90GPa for a sixteen atom orthorhombic unit cell. Unlike Cell A, the symmetry for Cell B was reported to be *P*mma[8] and one lattice vector in Cell B is substantially shorter than the other two. This would tend to favor crystal molecular structures that adopt molecular orientations perpendicular to the short lattice vector, otherwise the crystal would likely adopt a chain structure and effectively break the molecular triple bond in order to accommodate the intermolecular repulsion.

For Cell B, the genetic algorithm structure search was applied using the experimental cell parameters and setting Z=16. Examples of the lowest enthalpy molecular structures derived from Cell B are reported in Table 4.3, labeled **B1-B6**. Most of the structures that were generated from the GA structure search possessed no symmetry, with the notable exception of the lowest enthalpy structure **B1,** which possesses *P*mma symmetry, the same as that assigned experimentally. Structures **B2-B6** correspond to select examples of the lowest-enthalpy, low-symmetry molecular structures.

Table 4.3. Relative enthalpies and symmetry information of the low enthalpy phases generated from a genetic algorithm based structure searches for Cell B.

	fixed cell [a]		*variable cell* [b]			
structure[c]	enthalpy *(eV/atom)*	symmetry	enthalpy *(eV/atom)*	symmetry	lattice ratios *(a/b, a/c)*	lattice angles (α, β, γ)
B1	0.0	*P*mma	0.0	*I*mmm	1.11, 2.69	90, 90, 90
B2	0.085	*P*1	0.008	*P*1	1.02, 2.43	90, 92, 90
B3	0.154	*P*1	0.029	*P*1	1.11, 2.54	90, 90, 84
B4	0.169	*P*1	0.027	*P*1	1.01, 2.43	88, 93, 87
B5	0.227	*P*1	0.018	*P*1	1.04, 2.46	100, 84, 95
B6	0.315	*P*1	0.050	*P*1	1.11, 2.52	98, 94, 94
B7	0.363	*P*mma	-0.041	$P2_1/C$	1.41, 2.86	90, 96, 90
B8	0.485	*P*mma	0.019	*P*nma	1.50, 2.97	90, 90, 90
B9	0.398	*P*mma	0.368	*P*mma	1.07, 2.91	90, 90, 90
B10[d]	0.415	*P*mma	0.0[d]	*I*mmm	1.11, 2.69	90, 90, 90
Exp.[e]		*P*mma	-	-	1.05, 2.66	90, 90, 90

[a]Structural search using experimentally determined lattice parameters. [b]Results correspond to structures from the fixed cell structure search that have been optimized where the cell is also allowed to relax. [c]Structures B1-B6 are derived a symmetry unrestricted search, whereas structures B7-B10 are derived from a *P*mma space group restricted search as described in the text. [d]Upon cell relaxation, structure B10* optimized to structure B1. [e]Ref. 8.

The geometry of structure **B1** is shown in Figure 4.6a. The simulated x-ray powder diffraction pattern (not shown) for the **B1** structure shows sharp peaks at 10° and 12°. Although the peaks at 10° and 12 °, become the predominant peaks at high pressures[8] there are additional peaks observed experimentally. This could suggest that these peaks are perhaps indicative of a forced alignment of the highly strained molecules as the pressure is increased; nonetheless, the additional peaks seen in experiment below 10° and between 10° and 12° are notably absent in structure **B1**.

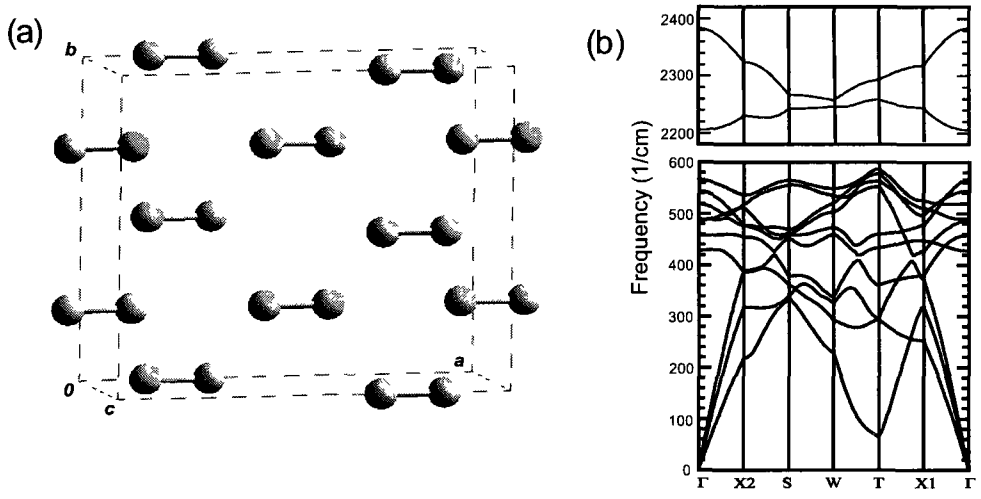


Figure 4.6. (a) 16 atom cell representation of structure **B1** following variable cell optimization. (b) Phonon dispersion curves for structure **B1** at 90GPa in the lattice and vibrational regions.

The atomic structures and cell vectors of structures **B1-B6** were then fully optimized. Upon cell relaxation, structures **B2-B6** still possessed no symmetry, and remained higher in enthalpy than structure **B1**. On one hand, structure **B1** remained orthorhombic after cell relaxation and the lattice ratios of the 16-atom simulation cell a/b and a/c remained within 5% of the experimental values. On the other hand, it was found that the unit cell of structure **B1** could be reduced to a 4-atom unit cell with *I*mmm symmetry. This contrasts sharply with the 16-atom unit cell and the lower *P*mma symmetry reported experimentally. Interestingly, the optimized **B1** structure was found to be closely related to a recently reported *I*mmm phase [37] of polynitrogen, which the authors obtained by optimizing coordinates of the recently solved ε-O_2 phase of high-pressure molecular oxygen. [1, 2].

The experimentally assigned symmetry of Cell B was clearly reported to be *P*mma,[8] which has noticeably less symmetry than *I*mmm. To explore this discrepancy

further and reinforce the notion that we had explored the region of the potential energy surface with *P*mma symmetry, the GA-inspired structure searching algorithm was modified to exclusively generate initial structures with *P*mma symmetry by placing the atoms on Wyckoff sites of the *P*mma space group in a 2/1/1 distribution; the necessity of such a Wyckoff site distribution to explain the observed Raman behavior of ζ-N_2 is explained in ref. 7.

With the symmetry restricted genetic algorithm search, structure **B1** again emerged as the lowest enthalpy structure using the experimental cell parameters. Four of the remaining lower enthalpy structures are presented in Table 4.3 as structures **B7-B10**. When constrained to assume the experimental cell parameters, structures **B7** to **B10** were at least 0.363 eV/atom higher in enthalpy than structure **B1**.

Upon full lattice optimization, wherein the cell parameters are allowed to relax at 90 GPa, structure **B10** optimizes to structure **B1**. This, and the fact that structure **B1** adopts *I*mmm symmetry from *P*mma symmetry upon cell relaxation, shows that the *I*mmm and *P*mma structures can be related by simply rotating some of the molecules out of the plane of the two largest lattice vectors. These distortions may manifest themselves within an imperfect crystal.

Structure **B7** emerged as the lowest enthalpy structure when the cell constraints were released during the optimization procedure. However the structure became monoclinic, deviating from the desired orthorhombic symmetry, and was subsequently not examined further. Although structure **B9** maintained an orthorhombic unit cell, the relaxed structure remained significantly higher in enthalpy than **B1**. Moreover, a phonon dispersion calculation of structure **B9** (following relaxation) revealed that it was not

mechanically stable at 90 GPa and therefore was not considered further. Although unspectacular when constrained to the fixed experimental cell, structure **B8**, shown in Figure 4.7(a), emerges as another low enthalpy phase. Encouragingly, it has an orthorhombic 16-atom unit cell with *P*nma symmetry, which is only slightly higher in symmetry than the experimentally assigned *P*mma symmetry.

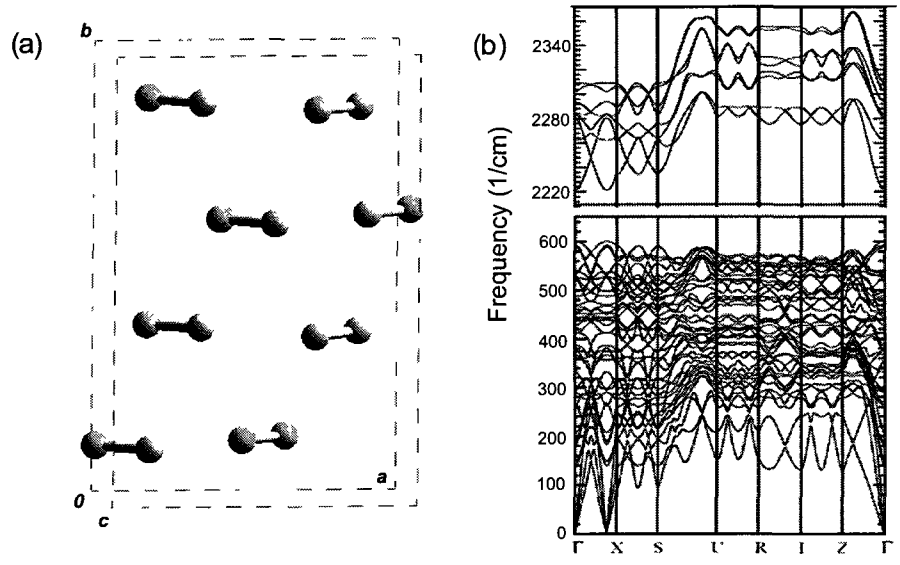


Figure 4.7. a) Unit cell representation of structure **B8** (following variable cell optimization). b) Phonon dispersion curves for structure **B8** at 90GPa in the lattice and vibrational regions

From the above results, structures **B1** and **B8** are considered the lead candidate structures for cell B. Figure 4.4 shows that structure **B1** is lower in enthalpy than both the α–N_2 and ε–N_2 molecular phases at 90 GPa while **B8** is only lower in enthalpy than the α–N_2 phase, becoming more enthalpically stable than the ε–N_2 phase at pressures between 67 and 76 GPa and pressures greater than ~91 GPa. Phonon dispersion curves for structures **B1** and **B8**, given in Figure 4.6b and Figure 4.7b, respectively, show that both phases are mechanically stable at 90 GPa. Although the phonon dispersion curves of

B8 show no negative frequency modes, Figure 4.7b reveals there is a very weak mode between the Γ and X symmetry points that dips as low as 6 cm^{-1}. Regarding its behavior at higher pressures, the calculated pressure versus volume curves in Figure 4.5 show structure **B8** undergoes a phase transformation at around 160 GPa accompanied by a notable decrease in volume; at this point structure **B8** transforms into a two dimensional polymeric chain-like phase. Although it is not as apparent as structure **B8**, structure **B1** also undergoes a phase transformation at around 150 GPa to a phase with *I4*mmm symmetry. Though its volume is close to that of cg-N polymeric nitrogen, the *I4*mmm structure is still a molecular phase. It is interesting to note that Hemley et. al[37] recently reported that a simple distortion of their proposed *I*mmm phase of polynitrogen can describe its transformation to the polymeric cg-N structure.[26]

Although the simplistic structure of **B1** is appealing, it clearly has a four atom unit cell which contrasts sharply with the Z=16 identified experimentally. Furthermore, the *I*mmm symmetry of structure **B1** has four atoms occupying 4*e* Wyckoff sites; this allows for only two Raman bands in the molecular vibrational region of the spectrum, whereas, for cell B, five Raman bands were clearly identified.[7] Alternatively, structure **B8** only lays 0.019 eV/atom higher in enthalpy than **B1** at 90 GPa, and has the desired sixteen atom unit cell. Furthermore, its *P*nma symmetry is significantly closer to the experimentally identified *P*mma symmetry. Structure **B8** is also more consistent than **B1** with the experimental Raman data of cell B. Its atoms are distributed among 8d and 4c Wyckoff sites, which allows for the desired five Raman stretching components; in addition, the 2/1/1 distribution of molecules over Wyckoff sites supports the redistribution of vibrational intensities at the ε-N_2 to ζ-N_2 phase transition.[7] On the other

hand, the a/b lattice vector ratio for structure **B8** is 44% larger than that determined experimentally. Despite the larger lattice vector ratio, structure **B8** is considered the best candidate structure from the presented DFT-based GA calculations on cell B because it is more consistent with the experimental symmetry and Raman data than structure **B1**.

4.2.4 – Final Comments Characterizing the ζ-N_2 phase

In this section, an exhaustive genetic algorithm search at the first principles DFT level of theory was performed on the high pressure ζ-N_2 phase of nitrogen using two sets of experimentally reported lattice parameters. For the lattice parameters reported by Eremets and coworkers,[10] determined at 80 GPa, two competitive structures, labeled **A1** and **A2**, were found to be most consistent with experiment. Their optimized cell parameters remain close to the experimental values and phonon dispersion calculations indicate that both structures are mechanically stable. Structure **A1** is the lowest enthalpy of the two aforementioned phases, but its *P*bcn symmetry does not match the proposed symmetry analysis.[10] However, it possesses many of the required symmetry elements and is closely related to the experimentally proposed candidate $P2_12_12$ space group. Conversely, structure **A2**, with its $P2_12_12_1$ symmetry, agrees with the proposed symmetry analysis and possesses a more common symmetry than *P*bcn, but it is higher in enthalpy.

For the lattice parameters of the molecular ζ-N_2 phase structure reported by Hemley and coworkers at 90 GPa,[8] two competitive structures were also recovered from the search and are labeled **B1** and **B8**. Both structures are calculated to be mechanically stable at 90 GPa. Although structure **B1** is 0.019 eV/atom lower in enthalpy than **B8** at 90 GPa, structure **B8** is considered to be a stronger candidate for the ζ-N_2 phase reported by Hemley and coworkers for several reasons. First, structure **B1** has only four atoms in

its unit cell, whereas, experimentally, the ζ-N_2 unit cell has 16 atoms, like structure **B8**. Furthermore, the *P*nma symmetry of structure **B8** is significantly closer to the experimentally identified *P*mma symmetry when compared to the *I*mmm symmetry of structure **B1**. Finally, structure **B8** is more consistent with experimental Raman data since both structure **B8** and ζ-N_2 are reported to have or allow five bands in the N-N stretching region. The distribution of molecules over special Wyckoff sites in structure **B1** only allows for two Raman bands in the vibrational region.

The conflicting experimental reports and the findings reported herein, having found dozens of local minima on the molecular potential energy surface, reinforce the notion that the phase diagram of nitrogen is quite rich and complex between 80 and 90 GPa. The results support the existence of multiple near-degenerate, history-dependent structures of high pressure molecular nitrogen, such as structures **A1** and **B1**. Since the samples are under tremendous strain experimentally, the "true" structures are likely highly distorted in the diamond anvil cell environment, this would explain the problems associated with characterizing the ζ-N_2 molecular phase structure. Being a supposed precursor to the cubic gauche phase, the true structure of the ζ-N_2 molecular phase, or an understanding of that region of the phase diagram, could prove invaluable in assessing the feasibility of polymeric nitrogen.

4.3 - Symmetry-Biased GA Searches

4.3.1 - Motivation

Recent single-crystal characterizations of the high pressure ε-O_2 phase of oxygen were recognized in the recent literature for their independent discoveries of antiferromagnetically-arranged O_2 tetramer subunits.[1, 2] Dated experimental studies from over a decade before had already postulated that the symmetry of the ε-O_2 phase was *C*2/m (or, more precisely, the isomorphic *A*2/m space group was reported),[38] and this was indeed verified from experiment. In their report, Lundegaard *et. al* further comment that none of the previous *ab-initio* studies on high pressure molecular oxygen had successfully recovered the ε–O_2 phase structure,[1] with GA-inspired search procedures, in particular, numbered among them.[17]

In this section, after showing their agreement with the unrestricted GA trials on Cell A for ζ–N_2 discussed above, symmetry-restricted GA trials are shown to effectively and efficiently recover the ε–O_2 atomic coordinates using only what was known to the scientific community in the early 1990s. This shows that not only are symmetry-restrictions a convenient tool for easing computational demands, but they are sometimes necessary to recover the correct crystal structure. Similar GA trials are run on the ζ–O_2 phase[12, 39, 40] to complement a recent GA-inspired investigation of the higher pressure, metallic ζ–O_2 structure.[22] Not limited just to homonuclear diatomics, these symmetry-restricted GA trials are also shown to be effective in recovering a recently identified phase of Dy_2B_4C,[24] with *I*mmm symmetry, and recovering the post-aragonite $CaCO_3$ structure[23] from a $P222_1$-restricted search, thereby addressing the symmetry-related inconsistencies between the experimental[41] and evolutionary theoretical studies.[23]

4.3.2 – High-Pressure Molecular Nitrogen

In Section 4.2, it was found the *P*mma-restricted GA structure search on cell B of ζ–N_2 [8] recovered the same low-enthalpy structure recovered from the unrestricted search, structure **B1**. But the symmetry-restricted GA search was also able to recover several other candidate structures as well, one of which is structure **B8**. In this case study, the symmetry-restricted search gives more information than its unrestricted analog, since the unrestricted searches became trapped navigating structures with *P*1 symmetry that have no correlation with experiment. These findings suggest that in some cases it may actually be a more powerful theoretical tool to use GA-inspired structure search in tandem with experimental powder x-ray diffraction studies to readily explore the relevant portions of the potential energy surface of new materials.

To reinforce this notion, four separate symmetry-restricted GA structure searches were set up to use the fixed lattice parameters and symmetry implementations described for cell A.[10] Initial structures were created such that they possessed either $P222_1$, $P2_12_12$, or $P2_12_12_1$ symmetry with forty structures being stored in a population at a given time. Each trial was run for 20-25 generations until the lowest-enthalpy structure remained unchanged for six generations. As expected, structure **A1** was recovered as the lowest-enthalpy structure in all four trials and structure **A2** was recovered as the 2nd lowest enthalpy structure in three of the trials. Three similar $R\bar{3}c$ symmetry-restricted GA trials were set up on the native eight-atom rhombohedral representation of the ε–N_2 unit cell[10] in the hopes of recovering its solved molecular structure from lattice information determined prior to its precise structure determination. The ε–N_2 phase structure, although slightly distorted as described in Section 2.1.5, was recovered in all three trials,

showing that these symmetry-biased GA procedures are effective in probing homonuclear diatomics when given reliable symmetry data.

4.3.3 - High-Pressure Molecular Oxygen

At ambient pressure and low temperature conditions, there are three well-known solid molecular-oxygen structures:[1] the cubic γ-O_2 phase with *P*m3n symmetry,[42] the rhombohedral ß-O_2 phase[43, 44] with $R\bar{3}$m symmetry, and the monoclinic α-O_2 phase[45, 46] with *C*2/m symmetry. At 3 GPa, the α-O_2 phase transforms into the orthorhombic δ-O_2 phase,[47, 48] assuming a higher *F*mmm symmetry, and, like the α-O_2 phase, is characterized as an insulating antiferromagnet.[48] The high-pressure ε-O_2 phase[1, 2, 49] is seen through a further phase transformation at 10 GPa and persists until pressures reach 96 GPa, when a final phase transformation to the metallic ζ-O_2 phase is observed.[40] The symmetries and crystal structures of both the ε-O_2 and ζ-O_2 phases have been heatedly debated in the literature,[17, 22, 38, 40, 50, 51] before being finally resolved through single crystal x-ray diffraction studies in 2006[1, 2] and 2009,[21] respectively.

GA Investigations of the ε-O_2 phase

The ε-O_2 molecular phase was determined in 2006 from single-crystal x-ray diffraction to have a C-centered, eight-molecule, monoclinic unit cell with *C*2/m symmetry.[1, 2] The unit cell of the ε-O_2 phase is shown in Figure 4.8a; the oxygen atoms are distributed over 8*j*, 4*i*, and 4*i* Wyckoff sites to create its distinguishing O_8 molecular clusters, its tetramer O_2 substructures. Before its solution, the ε-O_2 unit cell structure was well-known in the literature to retain the non-metallic character of its predecessors and was, similarly, presumed to exhibit the antiferromagnetic character of the δ-O_2 phase as

well. Various reports in the literature however disagreed with the magnetic characterization: magnetic frustration had been predicted by DFT calculations on triangular configurations of an O_2 lattice[52] and subsequent theoretical and Raman/ neutron experiments failed to reach a general consensus.[53-55] The uncertainty over the nature of its atomic positions, however, left the debate in limbo. Incidentally, the plot of the ε-O_2 cell magnetization density, the difference between the alpha and beta spin electronic densities, is shown in Figure 4.8b to agree with the original presumption, showing an evident antiferromagnetic ordering within each O_8 cluster.

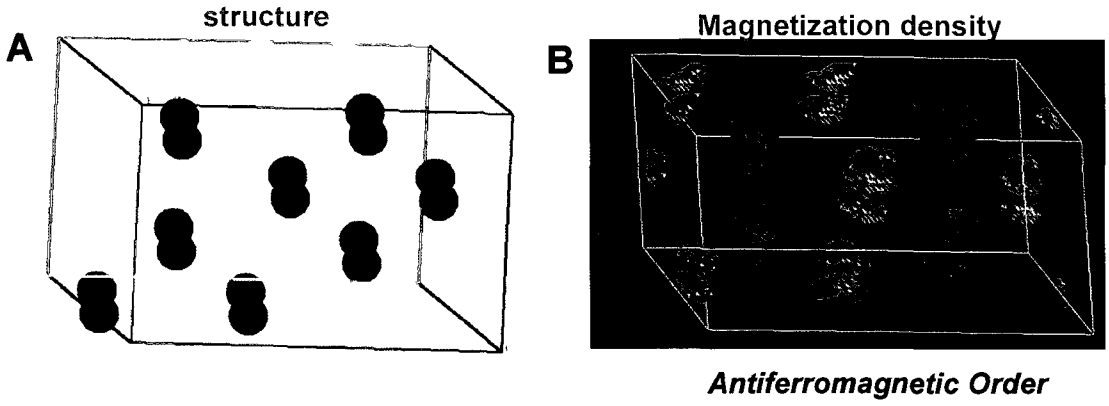


Figure 4.8. A) Unit cell of the experimentally verified ε-O_2 phase, and B) the magnetization density of the optimized ε-O_2 coordinates.

Despite substantial efforts, several ab-initio studies performed prior to its characterization had failed to predict the correct ε-O_2 structure.[17, 35, 52, 53] The symmetry of the ε-O_2 phase however had been postulated to have been *C*2/m or *A*2/m from independent studies in the early 1990s (note that *A*2/m symmetry is isomorphic to the *C*2/m space group since an A-face centered monoclinic cell can be reduced to a C-face centered monoclinic cell).[38, 40, 56] From 1995 through 2006, intermediary structural investigations both agreed and disagreed with the *A*2/m assignment but, nonetheless, the correct space group was known as of 1995.

Inspired by its similarity with high-pressure homonuclear diatomic nitrogen, the genetic algorithm was modified to search high-pressure oxygen phases in the hopes of quickly recovering the solved ε-O_2 phase structure. The GA procedure was set up identically to that used to search out the high-pressure ζ-N_2 phase with several key differences. First, the bond distance cut-off was redefined such that the algorithm generated and maintained O_2 bond lengths between 1.10 and 1.35 Å at all times, screening superfluous structures out of the population immediately when identified as described in Section 4.2.2. This bond-distance window was chosen to reflect the expected intramolecular O-O bond length in the solved ε-O_2 coordinates, which is 1.22 Å at 17.5 GPa. Second, the algorithm was modified to assign lattice parameters of a=7.76, b=5.50, c=3.67, and ß=116° when generating structures to fill the initial population or fill out subsequent populations.[38] Third, the structure generation was modified to generate structures with *C*2/m symmetry by distributing oxygen atoms over any combination of Wyckoff sites (8*j*, 4*e*-*i*, and 2*a*-*d*) such that 8 molecules are placed in the unit cell. For example, (8*j*, 8*j*) or (8*j*, 4*e*, 4*f*) structures could each be valid representations if the generated structure is molecular, which was enforced by monitoring the distribution of nearest and next-nearest neighbor oxygen atoms in the lattice.

Seven of these modified symmetry-restricted GA structure searches were set up and run using the lattice parameters and symmetry implementations described above. Only thirty structures were stored in a population at a given time, and the search was run for 10-15 generations in hopes of quickly recovering the ε-O_2 structure. The searches recovered several candidate structures and the enthalpies and symmetry information of the six lowest-enthalpy structures, labeled **E1**-**E6**, are presented in Table 4.4. Structures

E1 through **E6** are also shown collectively in Figure 4.9a-f, with the experimentally solved ε-O_2 structure appearing fourth on the list as structure **E4**. Structures **E1** and **E3** were recovered in all the GA searches, while all the remaining structures were recovered in at least two of the GA searches.

Table 4.4. Relative enthalpies and symmetry information of the low enthalpy phases generated from GA-based structure searches for ε-O_2 in the cell reported in ref. 39.

	Fixed Cell GA Search[a]		Relaxed Crystal Lattices[b]			
Structure	Enthalpy *(eV/atom)*	Symmetry	Enthalpy *(eV/atom)*	Symmetry	*a/b, a/c*	β
E1	0.0	*C*2/m	0.051	*C*2/m	1.40, 2.03	115.5
E2	0.013	*C*2/m	0.038	*C*2/m	1.38, 2.00	115.1
E3	0.062	*C*2/m	0.044	*C*2/m	1.66, 2.04	120.0
E4	0.082	*C*2/m	0.0	*C*2/m	1.39, 2.02	115.0
E5	0.238	*C*2/m	0.042	*C*2/m	1.41, 2.02	115.7
E6	0.313	*C*2/m	0.027	*C*2/m	1.70, 2.04	119.8
ε-O_2[c]	-	-	-	*C*2/m	1.41, 2.11	116.0

[a]Structural search using experimentally determined lattice parameters.
[b]Structures E1-E6 from the fixed cell GA structure search optimized under variable cell conditions at 17.5 GPa.
[c]Reported ε-O_2 lattice coordinates from ref. 39.

All six of the structures presented in Table 4.4 consist of stacked layers of O_2 molecules, stacking parallel to either the (010) plane to form structure **E1** or **E5**, to the (001) plane to form **E2** or **E4,** or to the (101) plane to form **E3** or **E6**. Two readily-identifiable classes emerge from these layered structures: those like **E1-E3** in which all the O_2 molecules are evenly distributed over each layer to ensure each O_2 has at least four equidistant neighboring molecules, or those like **E4-E6** in which the nearest neighbor distribution is skewed such that molecular "clusters" are formed on each layer. When these structures were further relaxed under variable cell conditions such that the trace of

the stress tensor reached 17.5 GPa, structure **E4** (the experimentally verified ε-O_2 coordinates) emerged as the lowest enthalpy phase, maintaining its *C*2/m symmetry.

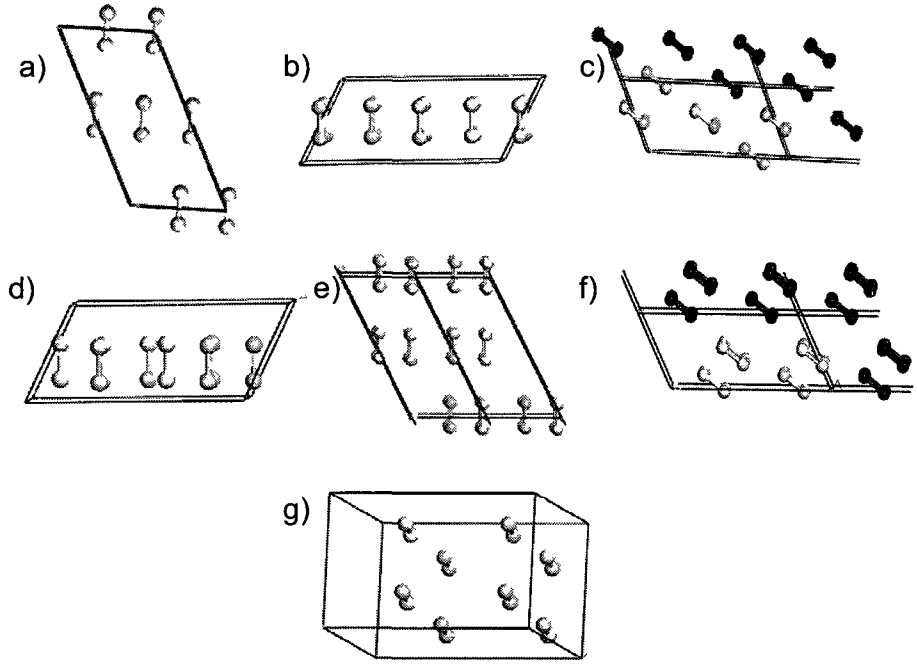


Figure 4.9. Structure of the lowest-enthalpy phases recovered from GA trials on the ε-O_2 cell from ref. [38]. Structures a) E1, b) E2, c) E3, d) E4 (ε-O_2), e) E5, and f) E6 are shown along with g) a chain-like phase detailed in the text.

A supplemental GA trial was run in which initial crystal structures were derived from randomly-generated, C-face centered, monoclinic unit cells and the structures optimized under variable cell conditions at 17.5 GPa before fitness evaluation. This trial differed from the GA trials reported in Table 4.4, which enforced fixed lattice parameters throughout their execution and thereby disallowed evolution of the lattice parameters themselves. To do so, the cell parameters were assigned such that $a > b > c$, $110.0^\circ < \beta < 125.0^\circ$, and the volume matched that of the lattice parameters used in the fixed cell GA searches, thereby agreeing more loosely with the monoclinic C-face centered monoclinic cell characterized in experiment. The population size of this "variable-cell" GA trial was

increased to hold 80 genes and the search was run for thirty generations before convergence of the four most fit structures was achieved for five generations. Structures **E4**, **E2**, and **E1** were recovered from the search with structure **E4,** instead of structure **E1**, emerging as the lowest enthalpy structure from the GA trial itself.

An interesting high-enthalpy layered phase, stacked similarly to structures **E2** and **E4**, with $P2_1/m$ symmetry was also recovered from the search and is shown in Figure 4.9g. Instead of forming the O_8 clusters reminiscent of the ε-O_2 phase, the O_2 molecules realign to form 2D chains along the layer-plane, causing metallic behavior. Although interesting from a theoretical viewpoint, since chain-like arrangements have been discussed in the literature, the plethora of O_2 phases will not be discussed further since the goal of this work was to study efficient theoretical means of recovering the ε-O_2 crystal coordinates and not to study alternative, higher-enthalpy allotropes of O_2. In this regard, the search methodology was successful.

GA Investigations of the ζ-O2 phase

The ζ-O_2 molecular phase of oxygen has just recently been confirmed,[21] after the work presented in this chapter was done, to have a C-centered, eight molecule, monoclinic unit cell with $C2/m$ symmetry. Prior to this, the best information regarding the ζ-O_2 phase was the result of a GA-inspired structure search with the USPEX program by Ma *et. al.*[22] It was determined that the lowest-enthalpy candidate phases for the ζ-O_2 structure between 100 and 250 GPa possessed either $C2/c$ or $C2/m$ symmetry with four and eight atoms in their unit cells respectively. From the single crystal x-ray diffraction study published in *Physical Review Letters* (ref. 21), the $C2/c$ candidate space group was ruled out based on careful analysis of the d-spacing from the x-ray diffraction results and

the allowed number of Raman active bands from spectroscopic investigations[21] and thus the *C*2/m structure was accepted as the correct structure.

To reiterate, at the time of this study the leading structural candidates were the proposed *C*2/c or *C*2/m candidates (the C2/m structure was later verified as the correct ζ-O_2 structure).[22] However, Ma *et. al* mention they had previously applied their GA search routines to study the ε-O_2 structure and recovered a *C*mcm chain-like phase instead of the true ε-O_2 *C*2/m-symmetry O_8 subunit structure. It was determined that the DFT level of theory within the generalized gradient approximation indeed predicted the *C*mcm phase to be 22 meV/atom more stable than the *C*2/m ε-O_2 structure at 25 GPa. The authors attributed this to a shortcoming within the method's description of the stretched intra-O_8-cluster molecular bonds, due to self-interaction error and DFT's description of the exchange-correlation hole.[22]

However, as reported earlier in this section, symmetry-restricted GA trials were able to recover the correct ε-O_2 structure by incorporating the lattice information available from experimental. In this light, it was prudent to further investigate the structure of the ζ-O_2 phase; these efforts were encouraged by one of the authors of ref. 21 (which characterized the ζ-O_2 phase experimentally)[21] since "something wasn't quite right" with the C2/m candidate structure.[57] Since the ζ-O_2 phase was also thought to possess a C-centered, monoclinic, *C*2/m unit cell, the same symmetry-biased GA search procedure was used on the high pressure, metallic ζ-O_2 phase of molecular oxygen. The algorithm was therefore modified to assign the experimentally determined lattice parameters of a=7.233, b=4.122, c=3.206, and ß=112.6° when generating structures to fill the initial population or pad subsequent populations.[57]

Eight separate GA structure searches were set up to use the fixed lattice parameters and symmetry implementations described above and enforce the cell parameters on each structure throughout their execution, thereby allowing only the atomic coordinates to relax freely. Fifty structures were stored in a population at a given time, and each search was again run for 8-12 generations. The results recovered a plethora of phases, but only two structures, labeled **Z1** and **Z2** in Table 4.5, were competitively low in enthalpy. Table 4.5 shows enthalpies of the **Z1** and **Z2** simulation cells, as well three other selected phases, the **Z3-Z5** structures, chosen to be representative of the remaining structures recovered from the algorithm. Only structures **Z1** and **Z2** were recovered in every GA trial.

Table 4.5. Relative enthalpies and symmetry information of the low enthalpy phases generated from GA-based structure searches for ζ-O_2 in the cell reported by ref. 58.

	Fixed Cell GA Search[a]		Relaxed Crystal Lattices[b]			
Structure	Enthalpy *(eV/atom)*	Symmetry	Enthalpy *(eV/atom)*	Symmetry	*a/b, a/c*	β
Z1	0.0	*C*2/m	0.0	C2/m	1.73, 2.23	112.1
Z2	0.005	*C*2/m	0.003	R-3m	1.72, 2.23	111.9
Z3	0.461	*C*2/m	-	-	-	-
Z4	1.629	*C*2/m	-	-	-	-
Z5	2.006	*C*2/m	-	-	-	-
ζ-O_2[c]	-	-	-	C2/m	1.75, 2.26	112.6

[a]Structural search using experimentally determined lattice parameters.
[b]Structures Z1-Z5 from the fixed cell GA structure search optimized under variable cell conditions at 105 GPa.
[c]Reported ζ-O_2 lattice coordinates from ref. 58.

As could perhaps be expected, the observed spread of phases very much resembled those obtained for the ε-O_2 phase. The lowest-enthalpy phases featured layered "sheets" of upright O_2 molecules, brought together such that the intermolecular

distances between nearest neighbors were all within 2.2 Å. Unlike the ε–O_2 phase however, the structures layered along the (001) plane, namely structures **Z1** and **Z2**, were heavily enthalpically favored due to the shortened c lattice vector. Collectively from all the GA searches, the energy spread between **Z1**/**Z2** and the remaining structures when constrained to the experimental cell was at least 0.42 eV/atom. The **Z1** structure is the same structure proposed by Ma *et. al.*,[22] and, as discussed, has been experimentally supported as the most likely ζ-O_2 structure.[21] The **Z1** structure is in fact the lowest-enthalpy structure, by a narrow margin, in both the experimental and relaxed cells, when the cells are allowed to relax under a pressure of 120 GPa. The symmetry information of the relaxed structures **Z1** and **Z2** is shown in Table 4.5, the symmetry information of structures **Z3**-**Z5** is not shown since they are not enthalpically competitive in the GA trials. The **Z2** structure adopts rhombohedral $R\bar{3}m$ symmetry with just one molecule in the rhombohedral representation of its unit cell when allowed to relax, while structure **Z1** retains its C2/m symmetry.

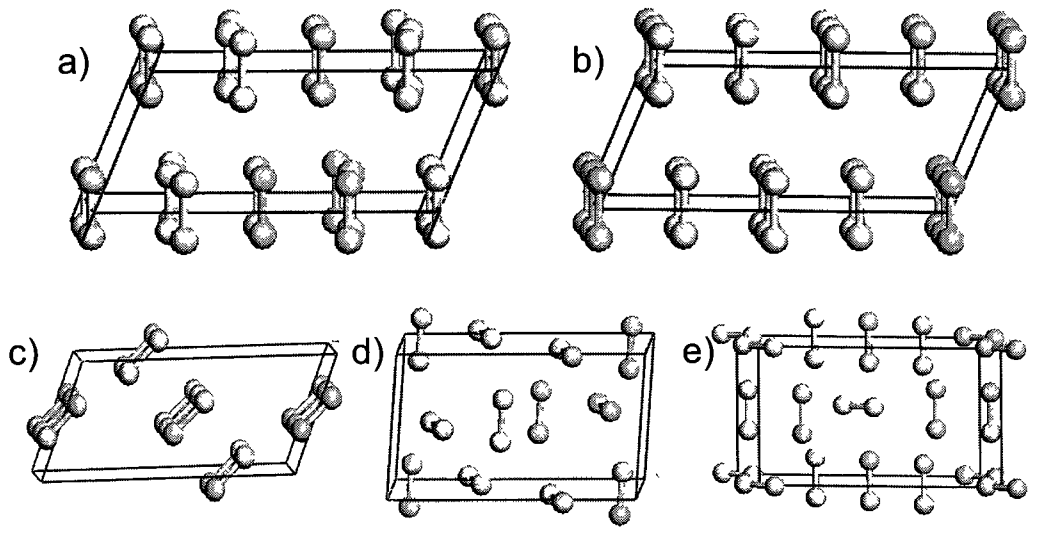


Figure 4.10. Structure of the candidate phases recovered from GA trials on the ζ-O_2 cell. Structures a) Z1 (ζ-O_2), b) Z2, c) Z3, d) Z4, and e) Z5 are shown.

Like **Z1**, the **Z2** phase does display mechanical properties of the ζ-O_2 phase which are consistent with experiment, in the sense it is non-insulating and magnetically disordered as shown in Figure 4.11, but its extraordinarily high symmetry dismisses it as a relevant allotrope for ζ-O_2. The **Z1** structure is successfully recovered with the symmetry-restricted GA search and is confirmed to be the only relevant C2/m symmetry structure from DFT-GGA calculations on the ζ-O_2 phase, due largely to its large enthalpic advantage over all other *C*2/m structures aside from **Z2**. Detailed arguments supporting **Z1** as the leading candidate are given in ref 21.

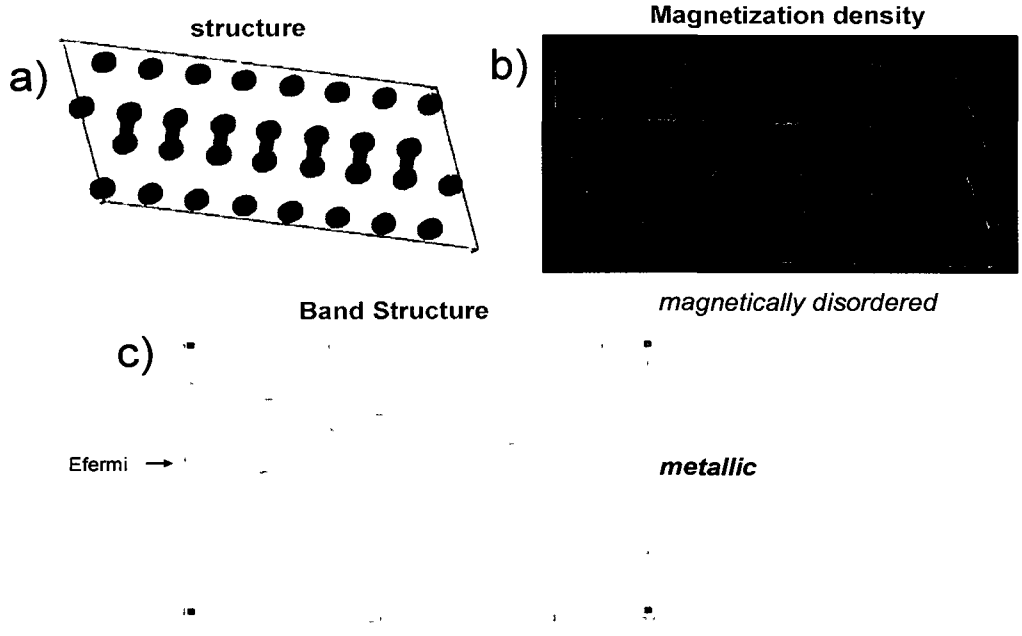


Figure 4.11. a) Extended representation of structure Z2 and b) its magnetization density and c) electronic band structure.

The results presented in this section show that experimentally derived lattice information can readily be incorporated into evolutionary algorithms to map out crystal structures that are difficult to characterize through ordinary x-ray diffraction or spectroscopic techniques. As demonstrated from the ε-O_2 and ζ-N_2 phases, the method is

particularly effective for mapping metastable crystal structures, wherein conventional evolutionary algorithms will tend to search out the thermodynamically preferred allotropes of the system under study instead of the metastable system of interest.

4.3.4 - Application to ternary chemical systems

Extending beyond homonuclear diatomic crystal structures, the symmetry-biased GA search routines discussed earlier in this section were applied to test-cases of heteronuclear systems. The two test systems were the (at the time) recently published Dy_2B_4C[24] and post-aragonite $CaCO_3$[23] phases. The first structure, the Dy_2B_4C crystal structure, was chosen because of its appearance in the recent literature; it was solved in 2007 and determined to have *I*mmm symmetry with 14 atoms in its unit cell.[24] The second structure, the post-aragonite form of $CaCO_3$, was discussed at length in Section 3.2.2. Its crystalline form was originally reported, from experiment, to have $P2_12_12$ symmetry,[41] but the crystal structure recovered from GA-inspired structure searches in 2006 was found to possess a higher *P*mmn symmetry.[23] Thus, GA-based structure searches of $P2_12_12$ symmetry biased $CaCO_3$ structures were run to attempt to link the experimentally- and theoretically- assigned symmetries together.

To set up the GA trials, the algorithm was first modified to generate chemically reasonable initial guesses, ie. structures with reasonable interatomic distances, with the desired symmetry extracted from the published experimental lattice parameters. All mating/mutation operations which broke the symmetry, namely by disrupting the ionic coordinates, were disabled, leaving only the Atom Swapping, Random Mixing, and Atom Splicing routines active. The post-mutation/mating geometry perturbations were left enabled to help increase the robustness of the GA algorithm and, in a similar spirit, the

number of random structures generated for each population was increased from 5% to 10% of the gene pool. Since the screening used to generate structures ensures physically-relevant structures are always created, it can be rationalized that randomly-generated structures can play a larger role in introducing worthwhile motifs to the gene pool. Full lattice and ionic relaxation of the genes, under 0 and 66.4 GPa external pressures respectively for Dy_2B_4C and $CaCO_3$, were run before evaluating each gene's fitness. The trials were set up to maintain forty genes in their populations and all trials generally converged after six generations for Dy_2B_4C and twenty generations for the 2x2x2 simulation cells of $CaCO_3$. Each set of GA calculations is further discussed, in kind, below.

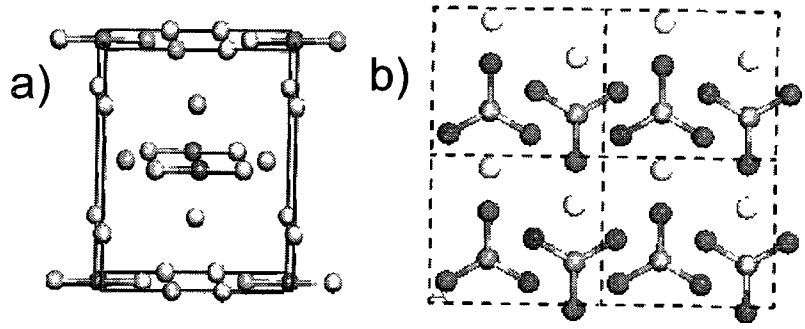


Figure 4.12. Crystal structures of a) Dy_2B_4C[24] and b) post-aragonite $CaCO_3$[23].

The lattice parameters of the Dy_2B_4C crystal structure were determined to be a=3.277, b=6.567, and c=7.542 Å and, together with its assigned *I*mmm symmetry, its lattice information was introduced into the GA code by distributing B atoms over 8*k*-8*n* special Wyckoff sites, B or Dy atoms over 4*e*-4*j* special sites, or B, Dy, or C atoms over 2*a*-2*d* sites such that the empirical formula Dy_2B_4C is preserved. The reported Dy_2B_4C coordinates were recovered in all five GA trials as the lowest enthalpy phase, retaining its *I*mmm symmetry. Further probing revealed that disabling the evolutionary algorithms,

thus generating each generation's new structures solely by randomly distributing atoms among *I*mmm Wyckoff sites of Dy_2B_4C's lattice, proved to be just as effective at recovering the crystal coordinates. This suggested the search space was not big enough to merit the use of evolutionary search algorithms; since the goal of this study was to test the evolutionary protocols, five additional GA trials were therefore run on larger 2x2x2 Dy_2B_4C simulation cells and were found to also recover the desired *I*mmm-symmetry crystal structures. GA trials on the 2x2x2 simulation cells without the evolutionary subroutines failed to converge to the correct structures even with 100-gene population sizes, showing the evolutionary protocols were responsible for the convergence of the trials on 2x2x2 simulation cells to the experimentally verified structure.

Five GA trials were then run on 1x1x1 and 2x2x2 simulation cells of $CaCO_3$, biased to generate $P2_12_12$ symmetry structures by distributing its atoms among 4*c*, 2*b*, and 2*a* Wyckoff sites to preserve the $CaCO_3$ empirical formula. The *P*mmn form gratifyingly emerged as the lowest enthalpy structure in all five 1x1x1 simulation cell trials and in four of the five 2x2x2 simulation cell trials, highlighting the link between the (theoretical-determined) *P*mmn-symmetry post-aragonite structure and the (experimentally-proposed) $P2_12_12$ space group. This suggests either that the supposed Pmmn symmetry of post-aragonite $CaCO_3$ is a failure of DFT to capture the specifics of the true $P2_12_12$ symmetry structure or that some systematic extinctions in the experimental diffraction pattern were unaccounted for in the original work. In either case, the establishment of a link between the two is encouraging.

Furthermore, the results presented in this section reinforce the notion that the scope of these GA-inspired evolutionary search techniques can be expanded by

introducing constraints on the search space. This was shown by successfully navigating the 2x2x2 simulation cells, which are too computationally demanding to explore without lattice or molecular constraints. This technology lends itself to further introduce and explore general defects in extended crystal lattices (like those of 2x2x2 or 3x3x3 simulation cells), thus exploring defect arrangements at low concentrations. Such an application is the study of defect associations in doped metal oxides, which is the focus of Chapters 5 and 6 in this thesis.

4.4 – GA Searches of Polymeric Nitrogen at Low Pressures: Searches of a Restricted Potential Energy Surface

In Section 4.2, several GA trial searches were carried out on thermodynamically-unstable molecular nitrogen structures by introducing a penalty function into the fitness criteria to penalize nitrogen structures with any of their constituent nitrogen atoms coordinated to more than one other nitrogen atom. Encouragingly, candidate structures that correlate well with recent experimental findings were recovered from the procedure. Another regime of interest for research efforts devoted to synthesizing ambient-pressure polymeric nitrogen structures is to study the lowest-enthalpy polymeric structures at low pressures. Though not directly related to any experimental findings due to the continued absence of any stable polymeric crystals at low pressure, knowledge of where the cg-N structure stands with respect to its stability over other solid nitrogen structures in these pressure regimes could prove insightful.

The incredible thermodynamic stability of polymeric nitrogen's precursors, dinitrogen molecular solids which are favored by ~1.2 eV/atom at ambient pressures, complicates direct structure searches of polymeric nitrogen at low pressures. In addition,

it can be expected that other molecular solids, like those made of N_5, N_6, or N_8 constituents, are also enthalpically favored over polymeric nitrogen at low pressure; polymeric structures possess covalently bound networks of nitrogen atoms which extend to infinity along at least one spatial dimension. Traditionally, this prompts an alternative stratagem of first performing structure searches at high pressure and following those up with simulated annealing or with subsequent calculations of the enthalpy vs. pressure equation of states to determine the stability of candidate phases at low pressures. This was the procedure used in Chapter 3 to recover the NG phase, a polymeric nitrogen phase which becomes lower in enthalpy than any other polymeric phase at low pressure, most notably lower than both the cg-N and zzCH phases. In this section however, it is shown that "direct" GA-inspired searches of low-pressure regimes can readily be used to study polymeric nitrogen by introducing a penalty to the fitness criteria that penalizes dinitrogen instead of promoting it.

The constraint-biased GA procedure used in Section 4.2 was therefore modified to allow for an efficient mapping of the potential energy surface pertaining to non-molecular nitrogen. The crystal structures were represented as they are in traditional GA's by lists of the constituent nitrogen atoms' Cartesian coordinates. The fitness of a given crystal structure is again derived from its enthalpy as calculated from first-principles DFT calculations. The Vienna Ab-Initio Simulation Package (VASP)[31] was used for all enthalpy calculations. The projector augmented wave (PAW) method of Blöchl[32] was used to treat the core states and the nitrogen 2s and 2p electrons were treated as valence electrons with a plane-wave cutoff of 500 eV for geometry optimizations and 800 eV for energy evaluations. The gradient-corrected exchange and correlation functionals of

Perdew-Burke-Ernzerhof (PBE)[33] were used in all calculations. For energy evaluations within the genetic algorithm, the Brillouin-zone integrations were performed using 4x4x4 Monkhorst-Pack grids whereas 10x10x10 meshes were used in more refined calculations for the phases selected for further study.

To address the thermodynamic instability of polymeric nitrogen at low pressure, the fitness function was modified to penalize structures that were molecular by adding a correction term for every nitrogen-nitrogen bond distance that was less than 1.18 Å. The correction term was taken from a harmonic potential fit to give no correction for intermolecular distances of 1.18 Å and add a 0.2 eV/atom penalty for 1.11 Å distances. This correction will thereby penalize N-N bond distances characteristic of dinitrogen (near 1.11 Å) while promoting N-N bond distances measuring 1.3 Å or beyond, which are characteristic of polymeric nitrogen. To further promote polymeric structures, a second penalty function was added to penalize terminal N atoms, atoms formally bound to only one other atom. The second correction term was taken from a harmonic potential fit to give no correction if an atom's second nearest nitrogen atom N-N distance was 1.55 Å and add a 0.2 eV/atom penalty for 1.75 Å, or greater, distances.

In this study a population size of 70 individuals was utilized for each generation. The genetic algorithm was halted when the energy of the four lowest energy structures remained unchanged for five generations, typically running for, on average fifteen generations. Two distinct production runs at each pressure were employed such that each structure was optimized under variable cell conditions at 0 GPa, 20 GPa, or 30 GPa, resulting in at least 5000 structures being evaluated and partially optimized at the DFT

level for each cell. The initial 8-atom unit cells assigned to each structure were either orthorhombic or monoclinic with volumes near 70 $Å^3$.

As expected from the previous studies, a number of structures were recovered from the GA trials. An assortment of the phases selected for further study is shown in Figure 4.13, the cg-N structure is also shown to reference the lowest enthalpy nitrogen structure at high pressures, and their crystal structures at 0 GPa are given in Table 4.6. The previously discussed zzCH and NG (Section 3.4) phases were recovered as the lowest-enthalpy candidates at 30 GPa. In addition, the higher-enthalpy ct3CH and gtCH chainlike phases were recovered, extended structures of each structure's chain-like constituent is shown in Figure 4.13d and Figure 4.13f respectively.

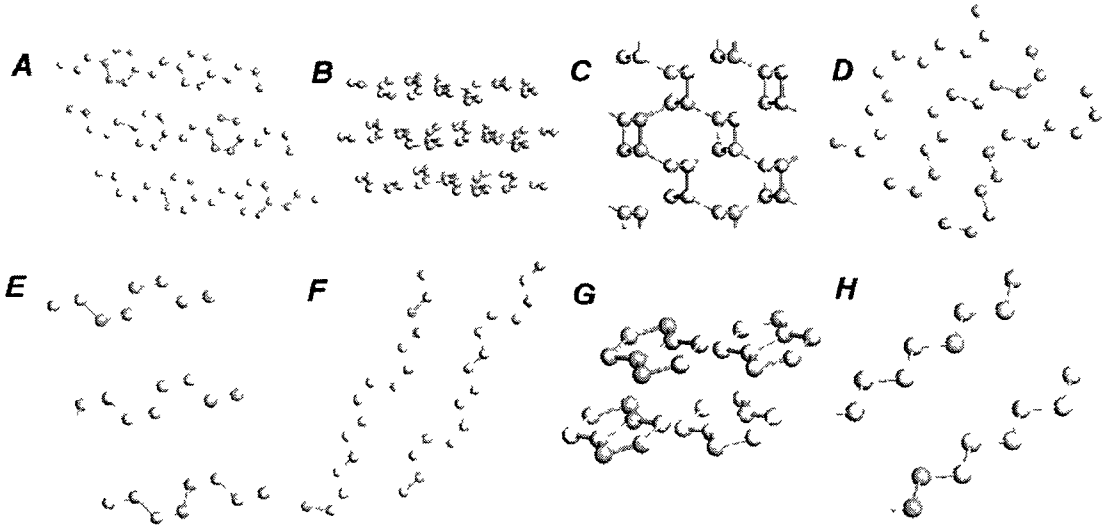


Figure 4.13. Extended structures of the A) azoN6, B) NG, C) cg-N, D) ct3CH, E) ctCH, F) gtCH, G) N8fused, and H) zzCH structures.

The gtCH and ct3CH structures can be related geometrically to the zzCH and ctCH by considering the sequence of dihedral angles along their chain constituents. The zzCH structure contains all *trans* dihedral angles and the ctCH structure contains alternating *cis* and *trans* dihedral angles. The ct3CH structure contains repeating

sequences of 1 *cis* and 3 consecutive *trans* dihedral angles while the gtCH structure shows a mix of *trans* and *gauche* dihedral angles.

At 20 GPa, a molecular solid, consisting of an N_8 fused-ring structure, was recovered as the lowest enthalpy phase, shown as the "N8fused" structure in Figure 4.13f. The zzCH was, once again, recovered along with a chainlike phase consisting of interconnected N6 rings, the so-called "azoN6" structure in Figure 4.13b. The NG structure was not recovered in either of the GA trials at 20 GPa but is noted to be almost

Table 4.6. Lattice coordinates of the azoN6, NG, ct3CH, ctCH, gtCH, N8fused and zzCH phases, recovered from the GA trials for polymeric nitrogen at 20 or 30 GPa. The cg-N is given as a reference.

Structure	*Symmetry*	*Wyckoff Sites*	*Cartesian Coordinates*	*Lattice Parameters* [a]
azoN6	C2	4c, 4c 4c, 4c	(3.429, -5.284, -6.064), (2.125, -5.848, -6.312) (4.027, -5.840, -4.907), (1.637, -1.322, -1.997)	a=10.91,b=4.03 c=5.78,β=117.91
NG	R-3	18f 6c	(-1.614, 1.951, 4.848) (0.0, 0.0, 1.642)	a=5.91 c=9.48
cg-N	$I2_13$	8a	(0.6259, 0.6260, 0.6259)	a=3.80
ct3CH	P-1	2i,2i 2i,2i	(0.293,3.169,-0.280),(-2.425,10.398,3.185) (-1.948, 6.461,2.025),(-1.948, 7.336, 1.098)	a=4.47, b=5.32 c=5.33,α=85.31 β=80.52,γ=50.52
ctCH	Cmcm	8f	(2.9541, -1.8667, 2.2620)	a=5.908,b=4.9156, c=3.8512
gtCH	P-1	2i,2i 2i,2i	(-4.237, 0.989, -7.881),(-2.395, -1.899, -6.068) (-2.413, -1.545, -2.180),(-2.302, -1.813, -3.383)	a=3.78,b=4.34 c=7.92,α=94.80 β=102.92,γ=112.34
N8fused	P-1	2i,2i 2i,2i	(-3.585, 1.159, 0.426),(-6.922, 2.594, 0.658) (-8.573, 7.771, -3.928),(-5.444, 6.011, -3.387)	a=4.84,b=5.10 c=5.15,α=82.38 β=110.99,γ= 85.61
zzCH	P-1	2i	(2.729, -0.472, -8.162)	a=2.16,b=3.50 c=3.51,α=65.35 β=89.98,γ= 89.50

[a] lattice is reconstructed from the cell parameters according to the Insight II standard (a is projected along the x axis and b lies in the xy plane)

thermoneutral with the azoN6 structure and therefore may have been screened out of the GA due to the energy resolution discussed in Chapter 3 or was simply never generated by the mating/mutation routines of these particular GAs.

At 0 GPa, only the N8fused and azoN6 structures polymeric structures were recovered among a number of other hybrid polymeric/molecular crystal structures with little to no symmetry. The azoN6 structure was only seen in the early stages, the first four generations, of the trial before being weeded out. Presumably, the dinitrogen and molecular structures are so thermodynamically favored at zero pressure, the penalty function does not selectively choose polymeric structures. The penalty function could simply be increased, but it was found that when the magnitude of the penalty function was scaled up by a factor of 10, the lion's share of optimized structures were still molecular and polymeric. Furthermore, since the geometry optimizations themselves are unconstrained and thereby not biased to mapping polymeric structures, such a practice seems inefficient.

The relative enthalpies of the azoN6, NG, ct3CH, ctCH, gtCH, N8fused, and zzCH phases to the cg-N structure between 0 and 30 GPa is shown in Figure 4.14. At 30 GPa, the zzCH was indeed the lowest-enthalpy polymeric structure recovered from the trials and is, in fact, thermodynamically competitive with a molecular solid at those pressures. Below this threshold, the molecular solid, the N8fused phase, perseveres as the lowest enthalpy structure, becoming favored by ~0.125 eV/atom at 0.1 GPa.

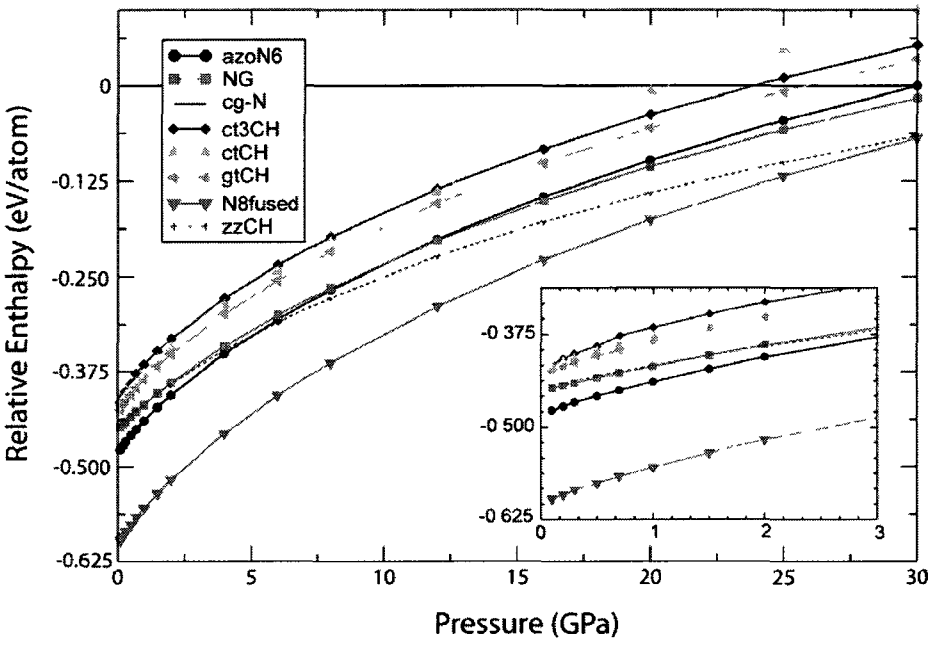


Figure 4.14. Calculated Enthalpy vs. Pressure diagram of the azoN6, NG, cg-N, ct3CH, ctCH, gtCH, N8fused and zzCH phases. Enthalpies are given relative to the cg-N structure.

All of the layered or chainlike polymeric nitrogen structures possess enthalpies that lay between -0.4 and -0.5 eV/atom relative to the cg-N structure at 0.1 GPa. The α-N_2 molecular phase (not shown), the experimentally confirmed crystal structure of molecular nitrogen at near-ambient pressures, lies ~1.2 eV/atom lower than cg-N, remaining over 0.7 eV/atom lower in energy than the most stable polymeric nitrogen recovered in this study. This, in itself, demonstrates that the layered and chainlike forms of polymeric nitrogen still possess high energy-density and the cg-N structure is highly thermodynamically unstable not only with respect to molecular solids, but to a number of other forms of polynitrogen solids as well.

The zzCH chain-like phase was the only chain-like phase (among the zzCH, ctCH, ct3CH, gtCH, and azoN6 candidates) that was found to be mechanically stable at zero pressure, as indicated from the all-positive phonon frequencies of its phonon

dispersion curves. Its phonon dispersion spectrum (not shown) was calculated using force constants derived from the finite differences method on 128-atom 3x3x3 supercells. The ct3CH, gtCH, and ctCH structures either had negative phonon vibrational modes or decomposed into smaller nitrogen fragments within 5 ps of simulation time when modeled with ab-initio molecular dynamics at 300 K using a Nose thermostat with 1.0 fs timesteps. The lower enthalpy azoN6 structure was also found to be mechanically unstable from the observation of negative phonon vibrational modes.

The most interesting new polymeric nitrogen structure, the layered NG structure discussed in Chapter 3, was recovered at 30 GPa. This structure was recovered as a high-enthalpy candidate from alternative GA trials wherein nitrogen structures were optimized under variable cell conditions at 90 GPa (these are reported in Chapter 3.2). This phase was found to be mechanically stable at low pressures and its recovery was expected since it was known to thermoneutral with the zzCH phase at low pressures. This phase was discussed at length in Chapter 3.2.

In summary, the GA procedure used in this section was able to recover the previously reported zzCH and ctCH nitrogen allotropes at low pressures. Furthermore, the newly discovered gtCH and ct3CH phases were found to be thermoneutral with the previously proposed ctCH phase. Finally, in a more interesting development, the GA trials were able to recover another low-enthalpy polymeric nitrogen structure, the azoN6 phase, in addition to the previously discussed NG phase such that both phases becomes enthalpically favored to the zzCH structure, the most stable phase reported in the literature thus far, at zero pressure.

4.5 – Conclusions

In this chapter, the GA-inspired search procedures outlined in Chapter Three were applied to study chemical systems wherein the most thermodynamically stable configurations are not the sought-to-be-explored class of materials. Specifically, the GA procedure was used to map out experimentally-relevant candidate structures for the contested ζ-N_2 phase and for the elusive ε-O_2 coordinates from a theoretical search procedure for the first time. Building on these efforts, it was then used to map out low-pressure polymeric nitrogen structures lower in enthalpy than any other reported in the literature thus far.

By biasing the fitness metric of the GA to efficiently study metastable, molecular solids of nitrogen, several candidate structures for the high-pressure ζ-N_2 phase were proposed; all of the proposed structures are compatible with one of the two sets of experimental lattice parameters reported in the literature.[8, 10] The discrepancy between the cells is addressed as two distinct molecular phases and, due to the absence of a low-enthalpy *P*mma-symmetry candidate for Cell B, the second set of lattice parameters,[8] the results further suggest either an erroneous symmetry characterization of Cell B or an inability of DFT under the generalized gradient approximation to describe molecular nitrogen solids at high pressures which deviate substantially from cubic symmetry (which is the distinguishing characteristic of Cell B among the two cells).

The experimentally-derived symmetry for cell B was tested by applying symmetry-restricted GA trials. These trials were successful in recovering the same low-enthalpy candidates recovered from unrestricted searches on cell A,[10] the first set of lattice parameters, but went one step further for cell B by mapping out a number of

structures beyond the single lowest-enthalpy structure recovered from the unrestricted search. These symmetry-restricted GA procedures were further tested by searching for and successfully recovering the ε-O_2, ζ-O_2, post-aragonite $CaCO_3$, and Dy_2B_4C structures directly from experimentally-proposed lattice information derived from powder x-ray diffraction studies. The searches not only saved substantial computation time over unrestricted searches, but ultimately proved to be necessary in order to recover the experimentally characterized ε-O_2 phase structure.

Finally, the "constrained" GA searches were applied to study low-pressure structures of polymeric nitrogen (< 30 GPa) by effectively reversing the bias added to the fitness metric when high-pressure molecular nitrogen was studied. In doing so, the fitness metric was biased to effectively search out non-molecular solid structures and successfully recovered the phases expected from the literature, namely the zzCH structure, along with several enthalpically favored candidate structures, notably the NG and azoN6 structures.

Overall, the work presented in this chapter shows that GA-based structure searches can be remarkably useful computational tools for studying thermodynamically unstable chemical materials at high levels of theory. This leads one to consider that perhaps GA-inspired procedures can play more prominent roles in materials design, to screen doped materials under the guise of optimizing some intrinsic material property at a high level of theory. Extensions of existing evolutionary-based methodologies toward such applications are the focus of Chapters Five and Six in this thesis.

4.6 - References

1 L. F. Lundegaard, G. Weck, M. I. McMahan, S. Desgrenniers, and P. Loubeyre, Nature **443**, 201 (2006).

2 H. Fujihisa, Y. Akahama, H. Kawamura, Y. Ohishi, O. Shimomura, H. Yamawaki, M. Sakashita, Y. Gotoh, S. Takeya, and K. Honda, Phys. Rev. Lett **97**, 085503/1 (2006).

3 M. I. M. Scheerboom and J. A. Schouten, J. Chem. Phys. **105**, 2553 (1996).

4 R. Bini, M. Jordan, L. Ulivi, and H. J. Jodl, Journal of Chemical Physics **108**, 6849 (1998).

5 H. E. Lorenzana, M. J. Lipp, and W. J. Evans, High Pressure Research **22**, 5 (2002).

6 M. J. Lipp, J. P. Klepeis, B. J. Baer, H. Cynn, W. J. Evans, V. Iota, and C. S. Yoo, Phys. Rev. B **76**, 014113 (2007).

7 E. Gregoryanz, C. Sanloup, R. Bini, J. Kreutz, H. J. Jodl, M. Somayazulu, H.-k. Mao, and R. J. Hemley, J. Chem. Phys. **124**, 116102 (2006).

8 E. Gregoryanz, A. F. Goncharov, C. Sanloup, M. Somayazulu, H.-k. Mao, and R. J. Hemley, J. Chem. Phys. **126**, 184505 (2007).

9 H. Katzke and P. Toledano, Phys. Rev. B **78**, 064103 (2008).

10 M. I. Eremets, A. G. Gavriliuk, N. R. Serebryanaya, I. A. Trojan, D. A. Dzivenko, R. Boehler, H. K. Mao, and R. J. Hemley, J. Chem. Phys. **121**, 11296 (2004).

11 R. Bini, L. Ulivi, J. Kreutz, and H. J. Jodl, J. Chem. Phys. **112**, 8522 (2000).

12 K. Shimizu, K. Suhara, M. Ikumo, M. I. Eremets, and K. Amaya, Nature **393**, 767 (1998).

13 M. I. Eremets, A. G. Gavriliuk, I. A. Trojan, D. A. Dzivenko, and R. Boehler, Nat. Materials **3**, 558 (2004).

14 K. D. M. Harris, Materials and Manufacturing Processes **24**, 293 (2009).

15 H. M. Rietveld, Journal of Applied Crystallography **2**, 65 (1969).

16 L. B. McCusker, R. B. Von Dreele, D. E. Cox, C. Louer, and P. Scardi, Journal of Applied Crystallography **32**, 36 (1999).

17 A. R. Oganov and C. W. Glass, J. Chem. Phys. **124**, 244704 (2006).

18 A. R. Oganov, J. Chen, C. Gatti, Y. Ma, C. W. Glass, Z. Liu, T. Yu, O. O. Kurakevych, and V. L. Solozhenko, Nature **457**, 863 (2009).

19 Y. Ma, M. I. Eremets, A. R. Oganov, Y. Xie, I. A. Trojan, S. Medvedev, A. O. Lyakhov, M. Valle, and V. Prakapenka, Nature **458**, 182 (2009).

20 J. Kotakoski and K. Albe, Phys. Rev. B **77**, 144109 (2008).

21 G. Weck, S. Desgrenniers, P. Loubeyre, and M. Mezouar, Phys. Rev. Lett **102**, 255503 (2009).

22 Y. Ma, A. R. Oganov, and C. W. Glass, Phys. Rev. B **76**, 064101 (2007).

23 A. R. Oganov, C. W. Glass, and S. Ono, Earth Planet.Sci.Lett. **241**, 95 (2006).

24 V. Babizhetskyy, C. Zheng, H. Mattausch, and A. Simon, Journal of Solid State Chemistry **180**, 3515 (2007).

25 T. Westerhoff, A. Wittig, and R. Feile, Phys. Rev. B **54**, 14 (1996).

26 C. Mailhot, L. H. Yang, and A. K. McMahan, Phys. Rev. B **46**, 14419 (1993).

27 D. M. Deaven and K. M. Ho, Phys. Rev. Lett **75**, 288 (1995).

28 Y. Zeiri, Phys. Rev. E **51**, R2769 (1995).

[29] Y. Yao, J. S. Tse, and K. Tanaka, Phys. Rev. B **77**, 052103 (2008).
[30] Y. Ma, A. R. Oganov, Z. Li, Y. Xie, and J. Kotakoski, Phys. Rev. Lett **102**, 065501 (2009).
[31] G. Kresse and J. Furthmuller, Phys. Rev. B **54**, 11169 (1996).
[32] P. E. Blochl, Phys. Rev. B **50**, 17953 (1994).
[33] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett **77**, 3865 (1996).
[34] A. Togo, F. Oba, and I. Tanaka, Phys. Rev. B **78**, 134106 (2008).
[35] J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal, J. Phys: Condens. Matter **14**, 2745 (2002).
[36] *Materials Studio Modelling Environment, v. 3.2* (Accelyrs, San Diego, 2005).
[37] R. Caracas and R. J. Hemley, Chem. Phys. Lett **442**, 65 (2007).
[38] D. Schiferl, S. W. Johnson, and A. S. Zinn, High Pressure Research **4**, 293 (1990).
[39] S. Serra, G. Chiarotti, S. Scandolo, and E. Tosatti, Phys. Rev. Lett **80**, 5160 (1998).
[40] Y. Akahama, H. Kawamura, D. Hausermann, M. Hanfland, and O. Shimomura, Phys. Rev. Lett **74**, 4690 (1995).
[41] S. Ono, T. Kikegawa, Y. Ohishi, and J. Tsuchiya, American Mineralogist **90**, 667 (2005).
[42] Y. A. Freiman and H. J. Jodl, Phys. Rep. **401**, 1 (2004).
[43] H. d'Amour, W. B. Holzapfel, and M. Nicol, J. Phys. Chem **85**, 130 (1981).
[44] D. Schiferl and e. al., Acta Crystallographica, Section B: Structural Science **39**, 153 (1983).
[45] R. A. Alikhanov, Sov. Phys. JETP **18**, 556 (1964).
[46] C. S. Barrett, L. Meyer, and J. Wasserman, J. Chem. Phys. **47**, 592 (1967).
[47] D. Schiferl, D. Cromer, L. Schwalbe, and R. Mills, Acta Crystallographica, Section B: Structural Science **39**, 153 (1983).
[48] F. A. Gorelli, M. Santoro, L. Ulivi, and M. Hanfland, Phys. Rev. B **65**, 172106 (2002).
[49] M. Nicol, K. R. Hirsch, and W. B. Holzapfel, Chem. Phys. Lett **68**, 49 (1979).
[50] S. W. Johnson, M. Nicol, and D. Schiferl, Journal of Applied Crystallography **26**, 320 (1993).
[51] F. A. Gorelli, L. Ulivi, M. Santoro, and R. Bini, Phys. Rev. Lett **83**, 4093 (1999).
[52] R. Gebauer, S. Serra, G. Chiarotti, S. Scandolo, S. Baroni, and E. Tosatti, Phys. Rev. B **61**, 6145 (2000).
[53] J. B. Neaton and N. W. Ashcroft, Phys. Rev. Lett **88**, 205503 (2002).
[54] F. A. Gorelli and e. al., Phys. Rev. B **63**, 104110 (2001).
[55] Y. Akahama and H. Kawamura, Phys. Rev. B **61**, 8801 (2000).
[56] Y. Akahama, H. Kawamura, D. Hausermann, M. Hanfland, M. Kunz, and O. Shimomura, Review of High Pressure Science Technology **7**, 781 (1998).
[57] S. Desgrenniers, (personal communication).

CHAPTER FIVE

USING A GENETIC ALGORITHM TO ASSESS DEFECT ASSOCIATIONS IN DOPED METAL OXIDES

In this chapter, genetic algorithms are used to map out low energy configurations of doped metal oxide materials. The search algorithms are tested on lanthanide-doped ceria (L = Sm, Gd, Lu) with various dopant concentrations. Most prominently, the methodology is used to navigate low-energy configurations of samarium-doped ceria (SDC) at a number of concentrations such that fresh insight into samarium's role in enhancing ceria's ionic conductivity is reported

CHAPTER 5 – *KEY SUBSECTIONS*

5.1 - Introduction

Metal oxides are prized as versatile catalysts, gas sensors, optoelectronic converters, and transparent electrodes, among other applications, with direct implications to the automotive, industrial, medicinal, and security sectors. Although metal oxide materials often suffice on their own, many of their application-specific and fundamental electronic or ionic properties can be improved upon by doping their lattice. Ceria (CeO_2), for example, is well suited as a solid electrolyte in solid oxide fuel cell (SOFC) technology due to its high ionic conductivity,[1] but it becomes even more effective when doped with a trivalent ion. Doped ceria's ionic conductivity improves when intrinsic defects, characterized by oxygen vacancies, are created in its doped fluorite lattice and it's the increased mobility of oxygen ions over these vacancies that is thought to be responsible for doped ceria's enhanced performance.[2] In recent studies, Samarium was found to be the best trivalent dopant for ceria in this regard[3, 4] and has since become a focal point of fuel-cell research intent on verifying and shedding light on the underlying principles responsible for this enhanced oxygen mobility.[3, 5, 6]

From a theoretical perspective, doped metal oxide materials are typically studied by, first, ascertaining how defects distribute themselves or "associate" within the oxide's lattice and, second, cataloging how these "defect associates" minimize the lattice energy or affect relevant reaction coordinates. For example, defect associations have been studied in Samarium-doped ceria (SDC) from independent theoretical works on 3.2% SDC ($[Sm_2O_3]_1[CeO_2]_{30}$) and 2.2% SDC ($[Sm_2O_3]_1[CeO_2]_{46}$) by quantifying association energies or oxygen migration barriers via classical polarizable force fields and first-principles DFT calculations.[3, 5] The 3.2% SDC model is particularly well-studied

because only one vacancy needs to be placed in a 2x2x2 ceria simulation cell, making it feasible to perform a systematic screening of every possible configuration. The theoretical difficulties for modeling higher concentrations stem from navigating the number of ways in which more than one vacancy can be placed in the simulation cell and speaks of a much broader issue in which systematic screenings of any doped metal-oxide become unfeasible when dealing with a large number of defect-defect interactions in the simulation cell. For example, a systematic search of all possible ways to model 6.6% SDC, by distributing four dopant atoms and two vacancies in a 2x2x2 ceria simulation cell, would involve screening approximately 3.5 billion possible configurations, ignoring the symmetry-equivalent geometries. Many applications of doped metal oxides use even higher defect concentrations, at which both electrical and ionic conductivity are notably optimal for SDC for example,[2, 4, 7] and the number of possible configurations for such materials increases in kind. Thus, it is necessary to reliably investigate more concentrated chemical systems in order to develop a better atomistic level understanding of the target material's inherent catalysis.

The standard approach towards developing sound theoretical models of doped metal-oxide materials still often relies on using one's chemical intuition to manually place dopants or defects about the simulation cell such that all "relevant" configurations are studied.[8, 9] As discussed in Chapters 3 and 4, a number of innovative methodologies have been developed to search out the most stable crystal structures of a system at given P-T conditions, notably those based on the genetic algorithm (GA),[10, 11] but they have only recently been successfully applied to study clusters, surfaces, and high-pressure solids at the DFT level of theory[12-14]. A recent GA-inspired experimental screening of

doped ceria materials was also reported as part of an elaborate search for selective hydrogen oxidation catalysts.[15] To reiterate, these methodologies recursively generate lower energy structures through a number of predefined operations intended to pass on favorable structural traits from generation to generation until the overall health of the population converges and have proven to be effective at navigating non-trivial potential energy surfaces. In materials design, inverse GA algorithms, algorithms designed to search out crystal structures that optimize the property in question, have been used to identify well-known or novel binary and ternary alloys with optimal heats of formation,[12] band gap features,[16] or target structure-energy relationships.[17] A cohesive methodology that unambiguously determines ground-state structures of unknown metal oxide materials remains undeveloped.[18]

In this chapter, it will be demonstrated that GA-inspired algorithms can readily be used to investigate doped metal-oxide crystal structures. Specifically, they are incorporated into the GA procedures discussed earlier in this thesis to study lanthanide-doped ceria (LDC), focusing particularly on larger simulation cells of SDC materials, and, in Chapter 6, to study doped ZnO and SnO_2 materials. The mating and mutation routines have thus been designed to accommodate both dopant and vacancy defects in a metal oxide lattice. In addition to SDC, gadolinium- and lutetium-doped ceria are investigated to compare the search algorithm's performance with several, distinct potential energy surfaces extracted from systematic, exhaustive searches of 3.2% lanthanide doped ceria reported in the literature. The routines reproduce the results of all the systematic searches and agree qualitatively when applied to systems of similar concentrations in larger 3x3x3 ceria supercells using classical potential-energy

evaluations. Note that, at the DFT level of theory, even the >100 atom 2x2x2 ceria simulation cells make purely DFT-based GA trials too computationally demanding beyond 3.2% LDC concentrations. Thus, in Section 5.3.2 the GA-inspired routines make use of tandem classical potential and first-principles DFT calculations to find the low-energy structures of SDC with and without intra-cell vacancy-vacancy interactions. Building on these efforts, subsequent GA trials are presented in Section 5.3.3 which probe even higher SDC concentrations, up to 20% SDC, derived from a 3x3x3 ceria simulation cell and evaluated at the classical level of theory. In doing so, a number of trends in each concentration's association energies were extracted from the trials that correlate well with the experimentally observed threshold in SDC ionic conductivity near 11% SDC concentrations. Furthermore, this work navigated potential energy surfaces with as many as 54 defects spread about a 324-atom simulation cell, extending well beyond the previous standard of 3-6 defects in a 96-atom simulation cell.

As a corollary to the work in this chapter, the methodology is shown, in Chapter 6, to function remarkably well within the framework of the "inverse" band structure techniques proposed in the literature to help tune electronic properties of ZnO and SnO_2, metal oxides used in gas-sensor technology materials. The reader is encouraged to read Chapter 6 as a follow up to this chapter since it is a closely related, but importantly distinct, application to the work presented here.

5.2 - Computational Method

The genetic algorithm (GA) inspired search algorithms used in this chapter build on the approach detailed in Chapter 3, Section 3.1, and apply them to investigate doped metal-oxide materials. GA studies traditionally represent a crystal by a list of N atomic

Cartesian coordinates, requiring no prior information about the system aside from the composition and volume of the unit cell. Since the target doped metal oxide materials in this work all represent more-or-less minor distortions of the native oxide's lattice coordinates, the crystal structures were similarly represented by the Cartesian coordinates of each lattice site[12] in the undoped metal oxide lattice. Concerning lanthanide doped ceria, the native undoped simulation cell is the *F*mcm-symmetry ceria (CeO_2) lattice, wherein the Ce and O lattice positions are each generated from their own single Wyckoff sites. A vacancy is stoichiometrically created in lanthanide-doped ceria (LDC) materials for every two dopant atoms, meaning that vacancies must also be incorporated into the allowed genetic representations. The genetic representation of a generic LDC crystal is therefore constructed by assigning N dopant atoms to distinct metal lattice sites and N/2 vacancies to distinct oxygen sites. A collection of doped structures, or "genes", is herein referred to as a "population". The initial populations are constructed by substituting dopants and/or oxygen vacancies at randomly selected Ce and O sites on the undoped lattice. For example, the 3.2% and 6.6% L_2O_3 concentrations in LDC materials are built by introducing one and two oxygen vacancies, respectively, to a 96-atom CeO_2 2x2x2 supercell.

For structure evolution, the number of atoms was kept constant for the duration of the procedure wherein, at each step, randomly chosen mutation or mating operations were used to combine structural motifs of one or two randomly chosen "parent" genes. The mating/mutation operations were designed to accommodate both dopant- and vacancy-type defects and preserve the lanthanide coordinates in the original *F*m-3m symmetry lattice. The mating and mutation operations used in this study were the following:

i) Geometry Perturbation, randomly perturbing the cartesian coordinates of a chosen subset of atoms from the parent to make the offspring, ii) Dopant Mutation, swapping individual cartesian coordinates of a randomly chosen subset of dopant atoms with those of a randomly chosen subset of target metal atoms, iii) Vacancy Mutation, swapping individual cartesian coordinates of a randomly chosen subset of oxygen atoms with those of a randomly chosen subset of oxygen vacancies, iv) Defect Swapping, swapping the Ce/L lattice sites of one parent structure with those of the other parent structure, and v) Defect Swapping and Mutation, a tandem application of the Defect Swapping mating routine with either a Dopant or Vacancy Mutation. Only one randomly chosen mating/mutation operation was chosen to generate a single offspring. Often these mating procedures generated physically unreasonable crystal structures, thus offspring structures were only added to the next generation if all atomic pairs were at least 1.0 Å apart, otherwise the same mating operation was repeated. The resulting offspring structures are given a twenty-five percent chance to mutate further through geometry perturbations. This perturbation involves randomly selecting between one and one half of the physically relevant lattice sites in the gene and randomly perturbing each of their x, y, and z Cartesian coordinates by a maximum of 0.2 Å.

The fitness of a candidate structure was determined from its relative lattice energy to the lowest-energy structure in its population, as derived from classical polarizable force fields or first principles DFT calculations. The population of subsequent generations was created in the following manner. First, the fittest structures of the top ~10th percentile are promoted, unperturbed, to the next generation. Second, a number of structures (~5%) are generated with random defect distributions and added to the new

population. To generate the rest of the population, fit structures were preferentially chosen as parent structures by selecting them from a Boltzmann weighted probability distribution.[10] The exponential term stems from the relative lattice energies, scaled by an appropriate temperature to allow at least a 25% chance to select half of the promoted structures. Unphysical structures are screened out during the mating process and the offspring structures are loosely optimized before evaluating their fitness.[16] The genetic algorithm was halted at the user's discretions, but, by default, when the energy of the four lowest energy structures remained unchanged for five generations for the 96-atom simulation cells and for twenty generations for the 324-atom simulation cells.

The GULP simulation package[19] was used to perform all the classical molecular mechanics calculations. The Buckingham potential, Equation [1], was used to model short range dispersion pair-wise interactions, Ewald's method[20] was used to sum the long-range Coulombic interactions, and the shell model[21] was used to account for the polarizability of the O^{2-} ions. The oxygen shell had a charge of -2.08 e and was tied to the core by a 27.29 eV/Å^2 spring constant. The Buckingham parameters, shown in Table 5.1, were assembled from the force fields reported by Balducci et al.[22] and Senyshyn et al.[23] The functions were evaluated between 0.5 and 10 Å for Ce-O and O-O pairs, and between 0.5 and 6 Å for Sm-O pairs. The elevation of the lower cutoff from 0.0 Å to 0.5 Å was employed to help efficiently eliminate unphysical structures from our genetic algorithm, it should in no way affect the energies of the low-energy structures. The specific force field parameters presented in Table 5.1 have not previously been used to study samarium-doped ceria, so they were benchmarked against other force fields that were previously used for doped CeO_2 materials.[5] The relative lattice energies of a pool

of randomly generated structures matched qualitatively but there were key differences, these are discussed further in Section 5.3. The force field was also benchmarked and validated with first principles DFT calculations by verifying that the order (by fitness) of selected structures was similar with both methods.

Table 5.1. Buckingham parameters for interatomic potential calculation

Species	*A (eV)*	*ρ (Å)*	*C (eV Å^6)*
O^{2-} - O^{2-} [ref. 22]	22764.3	0.149	27.89
Ce^{4+} - O^{2-} [ref. 22]	1986.8	0.3511	20.40
Sm^{3+} - O^{2-} [ref. 23]	4040.9	0.3034	0.0

$$S_{ij} = Ae^{\left(-r_{ij}/\rho\right)} - \frac{C}{r_{ij}^6} \qquad [1]$$

First principles DFT calculations were performed with the Vienna ab-initio Simulation Package (VASP).[24, 25] The projector augmented wave (PAW) method of Blöchl[26] was used to treat the core states, with the oxygen $2s^2$ and $2p^4$, the cerium $5s^2$, $5p^6$, $6s^2$, $5d^1$, and $4f^1$, and the samarium $5s^2$, $5p^6$, $6s^2$, and $5d^1$ electrons being treated as valence electrons. The Hubbard (U-term) correction to DFT was not employed because such parameterization has yet to be completed for lanthanide ions. A plane-wave cutoff of 520 eV was used in all geometry optimizations and single-point energy evaluations used the gradient-corrected exchange and correlation functionals of Perdew-Burke-Ernzerhof (PBE)[27]. For energy evaluations within the genetic algorithm, the Brillouin-zone was sampled only at the Γ-point, whereas Monkhorst-Pack grids were constructed from 3x3x3 meshes in more refined calculations of the phases selected for further study. Benchmark calculations showed that Γ-point calculations are sufficient for recovering

accurate geometries and qualitatively describing relative energies of 3.2% SDC materials. Variable cell geometry optimizations at zero pressure were used to evaluate the fitness of all structures, as derived from the GULP and VASP simulation packages.

5.3 - Using a Genetic Algorithm to Investigate Structural Properties of Lanthanide-Doped Ceria

In this section, the GA-inspired evolutionary algorithms described in section 5.2 are used to assess defect associations in lanthanide-doped ceria (LDC), focusing particularly on samarium-doped ceria (SDC), in order to ascertain the limitations and benefits of using GA technologies on metal oxide frameworks. The discussion begins by comparing results from GA trials on 3.2% SDC with a systematic search of 3.2% SDC materials at classical and DFT levels of theory to establish the effectiveness of the algorithm.[3, 5] The 3.2% LDC concentrations are constructed by introducing 2 dopant atoms and one vacancy to a 96-atom 2x2x2 CeO_2 simulation cell. There are ~64,000 ways to distribute two dopants and one vacancy in these cells, but, when symmetry is taken into account, the total number of unique 3.2% LDC structures totals only 35. This makes systematic screenings of the PES feasible but still allows for a large enough overall search space to test the GA-inspired search procedures. These findings are then supported with genetic algorithm (GA) searches of 3.2% gadolinium doped ceria (GDC) and lutetium doped ceria (LuDC) structures in order to test alternative potential energy surfaces and reproduce the findings of Wei *et al.*[5]

To illustrate the algorithms' effectiveness for larger systems, 1.9% and 3.8% SDC in 3x3x3 ceria simulation cells are investigated in order to compare study larger simulation cells with a concentration comparable to the well-studied 3.2% SDC.

Ignoring symmetry, there are $5.97*10^{12}$ and $3.00*10^{25}$ possible ways to distribute Sm atoms and oxygen vacancies about Ce and O lattice sites in the 1.9% and 3.8% SDC simulation cells respectively, making them a suitable demonstration of the method's applicability. The GA search routines are then used to explore the previously unexplored 6.6% SDC in a 2x2x2 ceria supercell at an effective DFT level of theory, by introducing two vacancies into the 2x2x2 ceria simulation cell and using tandem classical and density-functional-theory potential energy evaluations as part of a "tandem", or hybrid,[28] optimization scheme. The classical evaluations are used to effectively screen high-energy structures out the GA trial's gene pool and increase its overall efficiency by requiring fewer expensive DFT calculations to reach convergence. And finally, further GA trials are performed at the classical level of theory on large 3x3x3 simulation cells of ceria to investigate intermittent concentrations up to ~20% SDC, in order to recover an apparent "ordering" of vacancies beyond the optimal 11% SDC concentrations. This ambitious undertaking required mapping a configurational space that places as many as 54 defects (36 samarium atoms and 18 oxygen vacancies) in the ceria simulation cell.

5.3.1 - Explorations of 3.2% LDC materials

A number of GA-inspired structure searches were first run on 3.2% SDC in the hopes of validating the evolutionary protocols in the code. Using the undoped ceria 2x2x2 supercell to generate initial structures, seven separate GA structure searches were performed with classical energy evaluations. It was observed that when doped with Sm, the oxygen atoms neighboring the vacancy relax and the volume of the simulation cell expands by 3.2% to assume a slightly increased lattice constant of 5.47 Å.[7] In light of these expansions, the fitness of each structure was determined directly from lattice

energies calculated after full ionic and lattice relaxation of each structure and the algorithm was set up to read in the optimized cell parameters and incorporate the lattice information into the genetic representation of the structure itself.[16] With a population size of 20 individuals, the trials ran for, on average, twenty generations. First-principles DFT optimizations were then run on the best ten structures from the GA trials, starting from their classically-derived cells and atomic positions.

The six lowest energy structures recovered from the DFT optimizations are shown in Table 5.2, labeled as structures **S1**-**S6**. The lattice energies are reported relative to the lowest energy structure recovered from the GA search procedure. For the remainder of this section, LDC materials will be classified from the relative distribution of dopant atoms about vacancies in the lattice, specifically by referencing the dopant-vacancy distances as nearest neighbor (NN), next-nearest neighbor (NNN), and so on. There are two such dopant-vacancy distances to account for in 3.2% LDC materials constructed from 2x2x2 ceria simulation cells. For example, the SDC structure with a vacancy positioned NN to both Samarium atoms will be labeled NN-NN.

Table 5.2. Relative lattice energies and relevant structural information of 3.2% SDC structure recovered from GA search.

Structure	E_{GULP}[†] *(eV/cell)*	E_{VASP}[±] *(eV/cell)*	*Sm-Sm distance (Å)*	*Sm-vac distance*[‡] *(Å)*
S1	0.000	0.000	6.7	NN-NNN
S2	0.042	0.007	3.9	NN-NN
S3	0.098	0.023	6.6	NNN-NNN
S4	0.103	0.028	7.7	NNN-NNN
S5	0.158	0.107	3.9	NNN-NNN
S6	0.216	0.042	3.8	NN-NNN

[†] GULP constant pressure optimization with the potentials from refs 21 and 22.
[‡] NN denotes nearest neighbour (~2.4Å) and NNN denotes next-nearest neighbour (~4.5Å).

[±] Relative total electronic energies from VASP.

Prior first-principles calculations have found that in 3.2% SDC materials the dopant atoms prefer to occupy the NN site to the oxygen vacancy[3] to create a NN-NN structure, **S2**. The results presented in Table 5.2 agree with the DFT study of Andersson *et. al.* after taking into account that a limited number of configurations were considered in that work (namely only structures **S2**, **S5**, and **S6**). More recent DFT and classical studies claim a vacancy preference at the NNN site in SDC materials.[5, 6] The NN-NNN SDC structure, structure **S1**, was recovered as the lowest energy structure from the GA trials discussed in this work, separating the dopant Samarium atoms by 6.7 Å. In a systematic screening of all possible structures of 3.2% SDC at both classical and DFT levels of theory, the same optimal Sm-Sm distances and NN-NNN dopant positions with respect to the vacancy were noted in the lowest energy structure. The second, third, and fourth most fit structures from the GA trials, structures **S2-S4** in Table 5.2, were also recovered in the same energetic order within the systematic search. The systematic search results therefore parallel the GA results, implying the evolutionary protocols in the GA search procedure are sufficient for exhaustively exploring the PES of 3.2% SDC materials. Structures **S1** and **S2** are shown in Figure 5.1.

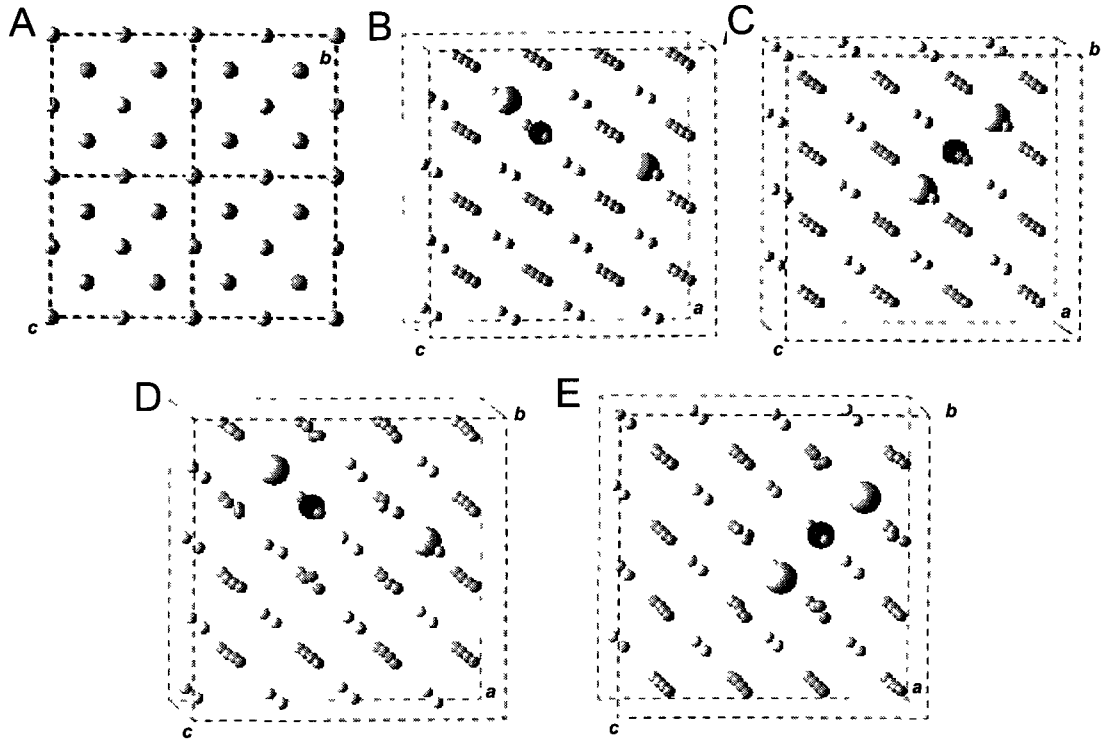


Figure 5.1. Extended structures of SDC. a) 2x2x2 supercell of CeO_2 crystal structure viewed along the [001] axis, the unrelaxed lattice of b) structure S1 and c) structure S2, and the relaxed geometries of d) structure S1 and e) structure S2. The Sm (large green balls), vacancy (medium black balls), cerium (small white balls), and oxygen (small gray balls) lattice positions are shown in all figures.

The fact that no site preference is observed in the lowest enthalpy structure (**S1**) at this concentration is not surprising. In accordance with Wei *et al.* and Nakayama *and* Martin,[5, 6] a crossover from NNN to NN dopant site preference with respect to the vacancy across the lanthanide series occurs at Gd^{3+}, and for Gd-doped ceria itself there is no preference. Samarium is positioned just to the left of Gadolinium (excluding Eu, whose trivalent ion is less favoured than the divalent ion), suggesting that the NNN site is not heavily favored over the NN site. This "hybrid" structure **S1** shows that the NN-NN and NN-NNN configurations have effectively the same association energies, with the NN-NNN structure being slightly favored at the DFT level of theory.

As expected, the oxygen atoms surrounding the vacancies in structures **S1** and **S2** show significant relaxation from the ideal CeO_2 lattice positions, as dictated by their relative positions to the Sm atoms. Consider that each vacancy, when formed, has six neighboring oxygen atoms, two along each lattice vector. In structure **S2**, each oxygen neighboring the vacancy along the [010] and [001] axes are in identical local chemical environments and move ~0.22 Å along their respective axis toward the vacancy. Along the [100] direction, both Samarium atoms lie between the vacancy and one of its neighboring oxygen atoms such that they rest on opposite sides of the axis. This neighboring oxygen atom is therefore restricted by the Sm dopant atoms and moves only 0.10 Å closer to the vacancy while the oxygen opposite the vacancy moves 0.27 Å. The Sm dopants thus cause anisotropic oxygen relaxations, as noted by Anderson *et. al,* by lowering the symmetry of the local chemical environment along the [100] axis.

Similar trends are observed from the oxygen relaxations of all the presented structures. For instance, in structure **S1**, the two oxygen atoms neighboring the vacancy along either the [100] and [010] axes relax ~0.28 Å and ~0.17 Å toward the vacancy, depending whether or not an Sm atom lies NN to the vacancy and the oxygen. Along the [001] axis, a larger split, 0.32 Å versus 0.18 Å, between the two neighboring oxygen atoms is observed. In this case one of the oxygen atoms neighboring the vacancy lies between a Sm dopant and the vacancy, thus relaxing further along the axis toward the vacancy. Asymmetric effects are noted to affect migration barriers as well as association energies, the degree of these effects on the migration barriers in structure **S1** remain unstudied.

Encouragingly, Table 5.2 shows the relative DFT energies of the lowest energy structures from the GA trials agree well with those predicted by the polarizable force field assembled from refs 22 and 23. However, Wei *et al.* reported 3.2% SDC favors structures that place the vacancy NNN to the dopant atoms,[5] as in structure **S4**. To address this discrepancy, further GA trials and a systematic search were run on 3.2% SDC using the force field parameters reported by Wei *et al.* The lowest energy structures from both search methods are given in Table 5.3.

Table 5.3. Relative lattice energies and relevant structural information of 3.2% SDC structure recovered from GA search using the force field parameters from ref. 5.

Structure	E_{GULP}[†] *(eV/cell)*	*Sm-Sm distance (Å)*	*Sm-vac distance*[‡] *(Å)*	*Structure symbol*[±]
S4	0.000	7.7	NNN-NNN	4_2_2
S3	0.001	6.6	NNN-NNN	3_2_2
S5*	0.016	3.8	NNN-NNN	1_2_2
S1	0.030	6.7	NN-NNN	3_1_2
S5*	0.090	3.8	NNN-NNN	1_2_2
S5*	0.091	3.8	NNN-NNN	1_2_2
S2	0.125	4.1	NN-NN	1_1_1
S6	0.138	3.8	NN-NNN	1_1_2

*structures were found that had the same Sm-Sm and Sm-vacancy distances but were marginally different due to in-plane oxygen relaxations around the vacancy

[†] GULP constant pressure optimization with the potentials from ref. 5.

[‡] NN denotes nearest neighbour (~2.4Å) and NNN denotes next-nearest neighbour (~4.5Å).

[±] Refers to the naming convention used in ref. 5.

The GA results match perfectly with those from the systematic search. While a systematic search with these force field parameters was reported recently in ref. 5, the authors used constant volume defect calculations as opposed to our constant pressure optimizations. The only discrepancy between these results and those of Wei *et. al*[5] is with structure **S5**, where a variety of energies were observed for seemingly identical (by

defect symmetry) NNN-NNN structures. This difference in energy is attributed to the fact that, between the unrelaxed structures, the two dopants form a different plane with the vacancy. As a result, oxygen atoms around the vacancy relax differently. Finally, since the GA results so closely reflect the results of the systematic search, any difference between these results and those of Wei. *et. al* can be attributed to the different simulation parameters.

To further demonstrate the effectiveness of the genetic algorithm with alternative potentials, GA investigations of the relative lattice energies of two other doped ceria materials were performed and compared to systematic searches in the literature.[5, 6] Since, in LDC materials, the dopant's preferred site with respect to the vacancy crosses over from NNN to NN near gadolinium as one progresses across the lanthanide series,[5, 6] the 3.2% Gd-doped ceria (GDC) and Lu-doped ceria (LuDC) materials were studied to complement the 3.2% SDC results discussed in Table 5.3. Two GA trials with fifteen genes in each population were run on each material and allowed to run for, on average, fifteen generations, until the fittest structure remained unchanged for five generations. The lowest energy LuDC and GDC structures extracted from the trials are shown in Table 5.4 and Table 5.5, labeled **L1** through **L8** and **G1** through **G6** respectively. For ease of comparison, the last columns of Table 5.4 and Table 5.5 refer to the structure notation used in reference 5.

Table 5.4. Relative lattice energies and relevant structural information of 3.2% LuDC structures recovered from GA search using the force field parameters from ref. 5.

Structure	E_{GULP}[†] *(eV/cell)*	*Sm-Sm distance (Å)*	*Sm-vacancy distance*[‡] *(Å)*	*Structure symbol* [±]
L1	0.000	4.1	NN-NN	1_1_1
L2	0.229	5.3	NN-NNN	2_1_2
L3	0.234	6.6	NN-NNN	3_1_2
L4	0.241	9.1	NN-NNNNN	5_1_4
L5	0.242	6.6	NN-NNNN	3_1_3
L6	0.267	7.6	NN-NNNN	4_1_3
L7	0.270	6.4	NN-NNNNN	3_1_4
L8	0.309	3.8	NN-NNN	1_1_2

[†] GULP constant pressure optimization with the potentials from ref. 5.
[‡] NN denotes nearest neighbour (~2.4Å), NNN denotes next-nearest neighbour (~4.5Å), and so forth.
[±] Refers to the naming convention used in ref. 5.

Table 5.5. Relative lattice energies and relevant structural information of 3.2% GDC structures recovered from GA search using the force field parameters from ref. 5.

Structure	E_{GULP}[†] *(eV/cell)*	*Sm-Sm distance (Å)*	*Sm-vacancy distance*[‡] *(Å)*	*Structure symbol*[±]
G1	0.000	6.8	NN-NNN	3_1_2
G2	0.020	4.1	NN-NN	1_1_1
G3	0.029	6.6	NNN-NNN	3_2_2
G4	0.029	7.6	NNN-NNN	4_2_2
G5	0.048	3.9	NNN-NNN	1_2_2
G6	0.097	3.8	NN-NNN	1_1_2

[†] GULP constant pressure optimization with the potentials from ref. 5.
[‡] NN denotes nearest neighbour (~2.4Å), NNN denotes next-nearest neighbour (~4.5Å), and so forth.
[±] Refers to the naming convention used in ref. 5.

The recovered low-energy structures agree with independent systematic searches of 3.2% LuDC and 3.2% GDC with Wei *et al.*'s potentials. Overall, the NN-NN structure is favored strongly (by 0.229 eV/cell) in LuDC, not favored (thermoneutral with other structures) in GDC, and disfavored in SDC by 0.125 eV/cell. Therefore, the GA

trials also successfully reproduce the trends obtained from systematic screenings of such LDC materials in ref. 5 and reinforce the notion that the methodology used in this work effectively explores low energy 3.2% LDC materials. Specifically, these results show the evolutionary protocols in the GA search algorithms effectively navigate the relevant areas of the 3.2% LDC PES. In the next sections, these routines are used to replace systematic searches of larger systems, like those of 1.9% and 3.8% SDC in 3x3x3 ceria simulation cells or 6.6% LDC materials in 2x2x2 simulation cells.

5.3.2 - Explorations of SDC up to 3.8% concentrations derived from 3x3x3 simulation cells of ceria

To test the scalability of the GA-inspired doping procedures, several GA trials were run on Sm-doped 3x3x3 ceria supercells in order to simulate concentrations comparable with the 3.2% SDC materials discussed above. When one, two, or four vacancy defects are introduced to the 3x3x3 ceria lattice, this simulates 0.9%, 1.9%, and 3.8% SDC concentrations, respectively. Initially, all the GA trials were set up to use 100-gene population sizes with classically-derived fitness metrics from the force field parameters assembled from refs. 22 and 23. The trials were run until the fittest structure remained unchanged for 60 generations in order to ensure sufficient convergence of the trial within the larger search spaces. When two such GA trials were run on 0.9% SDC, they converged after ten generations and recovered the NN-NNN structure as the lowest-energy structure. The results were analogous to those reported from placing one vacancy in a 2x2x2 ceria simulation cell.

Several GA trials were then run on 1.9% SDC in order to determine how large a population is needed to map out the system's potential energy surface. Four sets of three

trials were set up and run such that each trial in a given set stored either 100, 175, 250, or 325 structures, respectively, in its populations and was propagated for, on average, 250 generations. All the trials with 175 or more genes per population converged after, at most, 100 generations to the same low-energy structure while all but one of the smaller 100-gene trials converged to higher-energy structures. The twenty lowest-energy structures recovered from these trials were then read into the initial population of a further GA trial, which was run to test for lower energy structures. Although several lower energy intermediate structures were located, no alternative to the lowest-enthalpy structure was recovered from the extra trial. The lowest energy structures from the 1.9 % SDC structure search, labeled **B1-B7**, are shown in Table 5.6. An additional GA trial was set up to store 500 genes per population and recovered no additional relevant information but did converge after only 30 generations.

Table 5.6. Relative energies of most thermodynamically favoured structures of 1.9% SDC in a 3x3x3 ceria simulation cell.

Rank in opt GA	E_{GULP} *(opt) (eV/cell)*	*Sm-vac distances*[†]	*Sm-Sm distances*[†]	*vac-vac distance (Å)*
B1	0.000	6 NNN, 1 NN	2 NNN, 2 NN	~6.2
B2	0.056	6 NNN, 0 NN	4 NNN, 0 NN	~6.2
B3	0.071	6 NNN, 1 NN	2 NNN, 2 NN	~6.5
B4	0.089	5 NNN, 1 NN	3 NNN, 1 NN	~6.8
B5	0.106	6 NNN, 1 NN	3 NNN, 2 NN	~6.2
B6	0.144	6 NNN, 0 NN	1 NNN, 1 NN	~6.6
B7	0.205	6 NNN, 0 NN	2 NNN, 1 NN	~6.6
B8*	1.500	0 NNN, 2 NN	4 NNN, 1 NN	~6.1

[†] Sm-vacancy distance columns show the number of Sm with that particular distance followed by a letter code for the distance value (NN is on average 2.4 Å and NNN is 4.5 Å).

* a select high-energy structure (B8) is shown for comparison

The results show that several energetically-competitive structures were recovered from the GA trials, namely structures **B1-B5.** Since kT=~0.08 eV at 700 K, a typical

operating temperature in IT-SOFC (intermediate temperature SOFC) technologies, such configurations would contribute significantly to the overall makeup of the material if one assumes a Boltzmann distribution of configurations. Herein the structures derived from 3x3x3 ceria simulation cells are quantified by considering the number of dopant atoms in the first three coordination shells relative to each vacancy, note that only the structural information from the first two coordination shells is shown in Table 5.6. Figure 5.2 shows, graphically, the sum of the number of dopant Sm atoms in the first three coordination shells of each vacancy in structures **B1-B8**. The results portray multiple energetically-competitive configurations of SDC with dopant atoms predominantly positioned NNN to the vacancies. This type of behavior could not be captured from the 3.2% SDC materials in 2x2x2 simulation cells since there are only two dopant-vacancy interactions in the simulation cell. There are either no NN vacancy-dopant interactions,

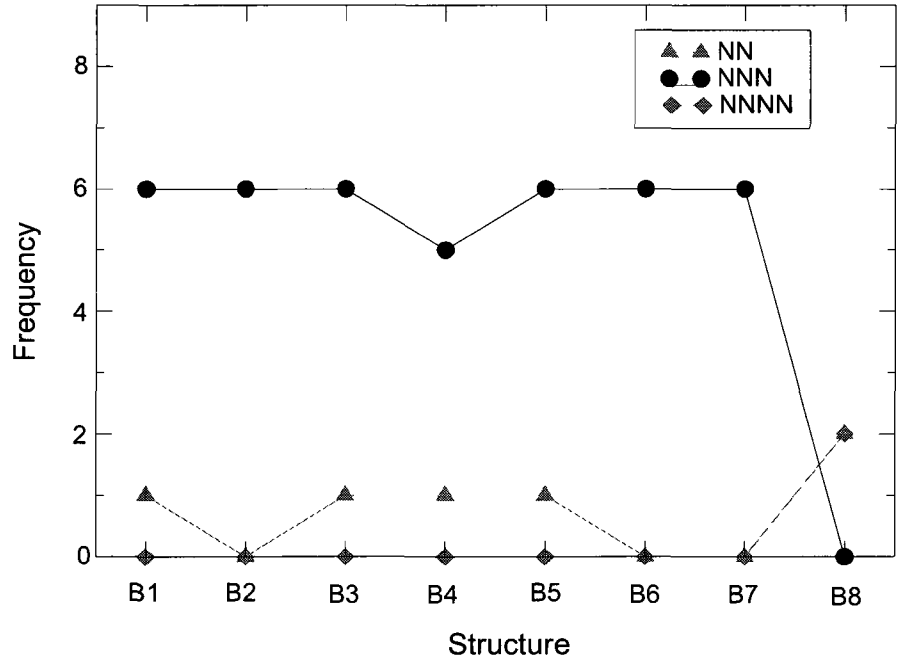


Figure 5.2. Plot showing the number of dopant Sm atoms in the first three coordination shells about each vacancy in structures B1-B8 of 1.9% SDC in a 3x3x3 ceria simulation cell.

as in structures **B2** or **B6**, or a singleton NN vacancy-dopant interaction, as in structures **B1** or **B3**. Structure **B1** was recovered in all of the trials except two of those which stored 100 genes per population, structures **B6** and **B7** were recovered as the lowest-energy candidate structures in those trials. A high-energy structure, structure **B8**, is shown in Table 5.6 and Figure 5.2 as a reference to a structure without NNN dopant/vacancy interactions.

Several GA trials were then run using classical energy-evaluations on systems with four vacancy defects in a 3x3x3 ceria simulation cell, thereby modeling 3.8% SDC materials and allowing a concentration comparable to the 3.2% SDC simulation cells discussed above. For the first three trials, three hundred structures were stored in each population and the trials were propagated until the fittest member of the population remained unchanged for sixty generations. Each trial converged to a different energy after running for, on average, 200 generations. The best four structures from each of these first three trials were then read into the initial population of a fourth GA trial which was propagated for an additional 200 generations. A new low-energy structure was found after only 50 generations and remained converged for the duration of the simulation. This structure was then read into the initial population of a fifth GA trial, in addition to the structures read in for the fourth trial, and persevered as the most fit gene throughout 250 generations. Finally, two more sets of traditional GA trials (starting from randomly generated populations) were set up with either 400 or 500 genes per population and two-of-five and four-of-five trials respectively converged to the same low-energy structure recovered above after 200 generations.

As expected, a number of structures were recovered from these 3.8% SDC trials.

To navigate the increased complexities of these large simulation cells, each of the three or four lowest-energy structures from a representative pool of the trials discussed above are presented in Figure 5.3, the lowest energy structure is of course included among them. The sum of the distribution of dopant Sm atoms over the first three coordination shells of each vacancy is shown for each structure and the plot is arranged such that the energy of the structures increases from left to right.

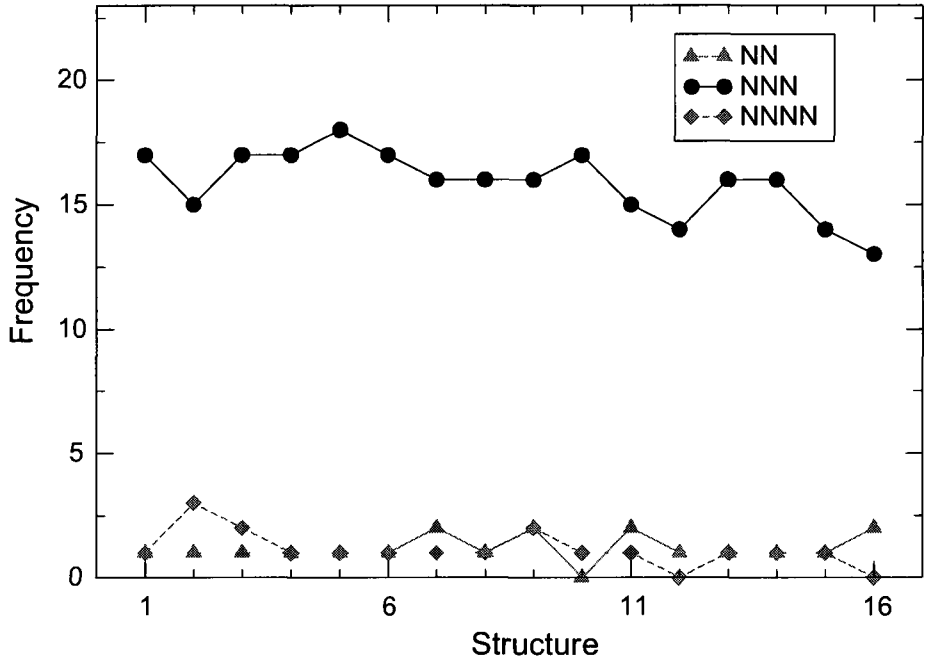


Figure 5.3. Plot showing the number of dopant Sm atoms in the first three coordination shells about each vacancy in 3.8% SDC structures. Structures increase in energy from left to right (see text for more details)

Energetically, all of the 3.8% SDC structures presented in Figure 5.3 fall within 0.324 eV/cell. The lowest energy structure, featuring a single NN dopant/vacancy association, lies 0.186 eV/cell lower in energy than the lowest recovered structure with no NN associations (the tenth structure from Figure 5.3). The observed trend is reminiscent of that observed in 1.9% SDC simulation cells, favoring configurations with dominant, but not exclusive, NNN dopant/vacancy interactions. In comparison with the

3.2% SDC results, these favored mixed site-preference 3.8% SDC configurations relate well, in an abstract sense, to the lowest energy NN-NNN structure recovered from 3.2% SDC simulation cells (structure S1). However, it is clear that configurations with exclusive NNN interactions are favored over those with exclusive NN interactions in 3.8% SDC materials, this was not the cases for the 3.2% SDC materials (note structure S2 in Table 5.2). Furthermore, the 3.8% SDC results show a much larger spread of low-energy, energetically-competitive configurations with mixed site preferences, this trend could not be recovered from the 3.2% SDC simulation cells due the imposed symmetry of the smaller simulation cell. In summary, although the lowest-energy structures of ~3.5% SDC in each simulation cell show mixed NN/NNN site preferences, the NN site preference implied by the results in Table 5.2 in 2x2x2 simulation cells is completely reversed to the expected NNN site preference in 3x3x3 simulation cells, further demonstrating the need for efficient evolutionary search algorithms.

Regarding the reliability of the search procedure itself, the overall convergence of each trial of the large 1.9% and 3.8% SDC simulation cells to their respective low-energy structure best demonstrates the power of these evolutionary procedures. On average, the lowest-energy structures from the 1.9% SDC and 3.8% SDC GA trials were recovered after only 10,000 and 80,500 energy evaluations with 250 and 400 genes, respectively, per population. Since the same energy was converged upon in several independent trials and, in addition, no alternative lower energy structures were found when the most fit candidate structures were read directly into the initial gene pools of subsequent trials, it can be reasoned that the encoded mating and mutation routines effectively sample LDC's configurational space. The larger system sizes however were noted to prolong the rates

of convergence, requiring at least 175- and 500- member populations to reliably recover the respective lowest energy 1.9% and 3.8% SDC structures within 200 generations. This suggests that consistent or "absolute" convergence can likely not be expected in systems with larger search spaces. Encouragingly, smaller-scale GA trials (with 300- gene population sizes) which build off the results of a previous trial were shown to function remarkably well in this regard on 3.8% SDC structures, providing an alternative stratagem. Given the stochastic nature of the algorithm and considering the results presented in Figure 5.3, it is reasonable to assume trials on larger systems can, nonetheless, readily be used to assess optimal defect/defect interactions even if absolute convergence cannot be guaranteed, since favored structural motifs will dominate the recovered structures' overall expressions.

The next two sections further illustrate the power of the methodology, wherein, first, tandem classical/DFT GA trials are used to perform an exhaustive search of the more-concentrated 6.6% SDC in a 2x2x2 ceria simulation cell at an effective DFT level of theory and, second, classical GA methodologies are used once more to recover the experimentally observed peak in ionic conductivity at higher concentration, near 11% SDC. The latter set of calculations requires navigating a PES with as many as 54 defects in the simulation cell.

5.3.2 - Tandem classical/DFT GA Investigations of 6.6 % LDC Materials

The added complications from studying dopant concentrations higher than 3.2% in LDC materials derived from 2x2x2 ceria simulation cells stem from parsing the additional vacancy-vacancy interactions introduced into the simulation cell. The larger

number of possible defect-defect interactions lead to a marked increase in the number of possible structures the simulation cell may assume. A systematic screening of these configurations at the DFT level of theory is unfeasible even at 6.6% LDC concentrations, the next level of complexity beyond 3.2% LDC, but these interactions are important to consider since they may adversely affect the observed dopant-vacancy association energies and higher concentrations of LDC materials are more relevant to their experimental applications.[7] It was demonstrated in Section 5.3.1 that GA-inspired algorithms effectively search the potential energy surface of ~3.2% SDC, GDC, and LuDC materials, recovering the lowest energy structures predicted from several distinct polarizable force fields or DFT calculations. Their success suggests that they can readily be used to replace systematic searches of higher concentrations, namely those of 6.6% LDC materials which are created by introducing one more vacancy, and two more dopants, to the simulation cell. Moreover, the SDC force field parameters assembled from refs. 22 and 23 reproduce the results of the DFT calculations on 3.2% SDC. Since DFT calculations are far more onerous and time-consuming, a large-scale GA structure search is used in this section to recursively scan the potential energy surface of 6.6% LDC materials using a tandem classical/DFT approach as outlined below.

First, it will be demonstrated that a tandem classical/DFT GA trial can reproduce the results obtained for 3.2% SDC materials. Specifically, a modified GA trial was run on 3.2% SDC, with fifteen genes per population, such that the first twenty generations used classically-derived fitness values[22, 23] before switching to DFT energy evaluations for the next five generations. As expected, this "tandem" GA trial finds no structure lower in energy than **S1**. Moreover, it was noted that the mating routines from the

classical trial recovered structure **S1** after five generations and the routines from the DFT portion of the trial found no structures that displaced the five lowest energy structures from the classical portion, implying that the lowest-energy structures were recovered after evaluating only one generation at the DFT level of theory. Note that the purpose of using DFT in the algorithm is to eliminate the possibility that a true low-energy structure is removed from the gene pool or a true high-energy is retained in the gene pool by the classical portion of the trial; clearly, this was not a problem at the 3.2% SDC concentration.

The basis for using tandem classical/DFT GA trials for 6.6% SDC materials assumes that the "screening" potential, the classical component, describes an adequate PES that agrees qualitatively to some degree with the "refined" potential, the DFT component. Considering that a systematic screening of all possible configurations at the DFT level of theory in unfeasible, it is impossible to conclude whether the classical potential agrees with the DFT potential and one must ensure that the dual-level tandem GA method would suffice if the screening potential did a poor job describing the refined potential. To ascertain whether the methodology would be sufficient even if a poor screening potential is used, a tandem classical/classical GA trial is presented on 6.6% LDC materials for which the complete profile of both potential energy surfaces can be independently determined. In this framework, independent GA trials are used to explore the potential energy surfaces of the screening and refined potentials, since classical energy evaluations are fast and can be rigorously checked. To check the "dual-level" protocol itself, the low-energy structures recovered from the screening potential are then used as a starting point for a further GA trial with the refined potential, in order to check

whether it converges to the expected low-energy structures. Specifically in this study, the LuDC potential of Wei *et. al*[5] was used to screen structures for the final SDC potential of Wei *et. al*,[5] thereby setting up a situation where the two potentials favor different dopant sites with respect to the vacancy (NN vs. NNN). This would simulate a situation where there is a mismatch between the classical and DFT potentials.

To reiterate, separate traditional GA trials were first run on 6.6% LuDC and SDC materials in order to understand the precise potential energy landscape of the LuDC and SDC classical potentials in 6.6% LDC simulation cells. To explore 6.6% LDC materials, four dopant atoms and two vacancies are placed in a 2x2x2 ceria simulation cell. With 120-gene population sizes, all of the 6.6% LuDC and SDC GA trials converged to the same low-energy structure after 20 generations. Larger GA trials, which were set up to store 150-200 individuals per population and run for 100 generations, recovered the same structures. A series of further trials were then run with smaller population sizes on 6.6% SDC, spanning 80, 50, 40, and 25 genes per population. All five 80-gene and four of the five 50-gene GA trials recovered the lowest-energy structure within 40 generations. However, only three of the five 40-gene and one of the five 25-gene trials recovered the low-energy 6.6% SDC structures within 50 generations, requiring ~100 generations to reach convergence instead.

As expected, the lowest energy 6.6% LuDC structure had two distinct NN-NN Lu/vacancy associates to create four overall NN Lu-vacancy interactions. The twenty lowest-energy LuDC structures all have at least three and at most five NN dopant-vacancy interactions. On the other hand, all of the 6.6% SDC trials converged to a low-energy structure with exclusive NNN dopant-vacancy interactions such that the vacancies

are separated by approximately six angstroms. This structure was 0.07 eV/cell lower in energy than the second lowest-energy structure, which also featured exclusive NNN dopant-vacancy interactions such that the vacancies rested slightly farther apart near seven angstroms. Both of the two most fit structures were recovered in all four 120-gene GA trials performed on 6.6% SDC and, in addition, each of the twenty lowest-energy structures were noted to heavily favor NNN interactions, with only one of the structures possessing more than one NN interaction.

With the goal of recovering the lowest energy 6.6% SDC structure, four "tandem" classical/classical GA trials with 50 genes per population were then set up for 6.6% SDC by reading in the 16 lowest-energy LDC structures from the 6.6% LuDC search into their initial gene pools and padding the rest of the population with randomly generated genes. The population size was chosen under the premise that since this trial represents the second, more computationally expensive, portion of the tandem GA trial, the smallest population size that reliably reproduces the lowest-energy structure, as discussed for the classical, standalone GA trials on 6.6% SDC, is desired. Encouragingly, all four "tandem" GA trials recovered the two expected low-energy SDC structures after forty generations. The longer search time for the latter potential (forty vs. twenty generations from the stand alone trials) is expected since the population size is smaller and the initial structures read into the gene pool from the LuDC trials are, on average, more unfavorable on the 6.6% SDC PES than randomly-generated structures. An additional set of tandem trials were then set up by reading in either the 25 or 40 lowest-energy structures from the LuDC trials and neglecting to introduce any randomly-generated genes into their initial populations. Only one of the four GA trials with 25 genes per populations recovered the

expected lowest-energy structure, within 50 generations, and two of the four GA trials with 40 genes per population recovered the two lowest-energy structures. These findings suggest that care must be taken when using a screening potential that does not correlate well with the final potential, since convergence to the final structure can be hampered by starting from a heavily-biased high-energy gene pool. Even so, the parameters used in the first tandem classical/classical GA trials did successfully reproduce the expected results even with an insufficient screening potential and should suffice for tandem classical/DFT GA investigations of 6.6% SDC, since these trials use a classical potential that correlates well with DFT calculations.

In light of these findings, the force field assembled from refs. 22 and 23 was used to screen 6.6% SDC structures for tandem classical/DFT GA trials. Seven classical GA trials, using classically-derived fitness metrics,[22, 23] and three subsequent DFT GA trials, using DFT-derived fitness metrics, were performed on 6.6% SDC. The five classical GA trials used population sizes of 80 or 120 individuals and were run for 40-50 generations. The six lowest energy structures recovered collectively from the classical trials are shown in Table 5.7, labeled as structures **A1-A6**. Structure **A2** was recovered in all five of the trials, structure A1 was recovered in three of the trials, and all the remaining presented structures were recovered in at least three of the trials. The DFT-based GA trials were initially set up with a collection of low-energy structures from the classical GA trials already in their gene pools, setting up a tandem GA trial as discussed above. Specifically, the top 16 structures recovered from the classical GA trials and 34 randomly generated structures (for a total of 50 genes) were read into the initial population, after which, upon the GA's execution, mating routines from the DFT portion of the trials

found no structures to displace the five lowest-energy structures carried over from the classical portion over their collective 25 generations. Each of the benchmark classical-based GA trials on 6.6% SDC converged after such a standard was met.

Table 5.7. Relative lattice energies (GULP) and relative electronic energies (VASP) of the top six 6.6% SDC structures from the "optimization" genetic algorithm.

Structure	E_{GULP} *(eV/cell)*	E_{VASP} *(eV/cell)*	*Vac-vac distance (Å)*	*Sm-vac1 distance*[†]	*Sm-vac2 distance*[†]	*Closest Sm-Sm distance*[‡]
A1	0.000	0.128	6.7	4 NNN	4 NNN	3.8 (2)
A2	0.048	0.000	6.1	3 NNN, 1 NN	3 NNN, 1 NN	3.9 (2)
A3	0.102	0.020	6.0	4 NNN	4 NNN	3.9 (2)
A4	0.154	0.020	6.0	4 NNN	3 NNN, 1 NN	3.9 (2)
A5	0.161	0.030	6.1	4 NNN	3 NNN, 1 NN	3.8 (1), 3.9 (2)
A6	0.211	0.041	6.0	4 NNN	2 NNN, 2 NN	3.9 (1), 4.1 (1)

[†] Sm-vacancy distance columns show the number of Sm with that particular distance followed by a letter code for the distance value (NN is on average 2.4 Å and NNN is 4.5 Å).

[‡] The closest Sm-Sm distances and, in parentheses, the number of times they appear in the cell.

First-principles DFT calculations do, however, disagree with our selected classical force field for 6.6 % SDC materials by penalizing structure **A1** in favor of all the other structures shown in Table 5.7. In general, the NNN vacancy positions with respect to the Sm atoms are still heavily favored, but some NN interactions are seen as well. The lowest energy structure **A2**, shown in Figure 5.4 along with structure **A1**, has one such NN dopant/vacancy interaction at each vacancy. Note that when all these structures were input directly into the gene pool of GA trials employing DFT calculations, thereby applying further mating/mutation operations at the DFT level of theory, structure **A2** persevered as the lowest energy structure throughout the entirety of each trial. It should

be emphasized that although the classical components of the GA trials recovered a seemingly low-energy structure which is strongly disfavored at the DFT level (structure **A1**), they did successfully recover the remaining five lowest energy structures in the same order as seen in the DFT portion of the GA trials. Importantly, this means the classical screening did effectively recover energetically-favorable candidates. No examples of structures favored at the DFT level of theory but strongly disfavored at the classical level of theory were found.

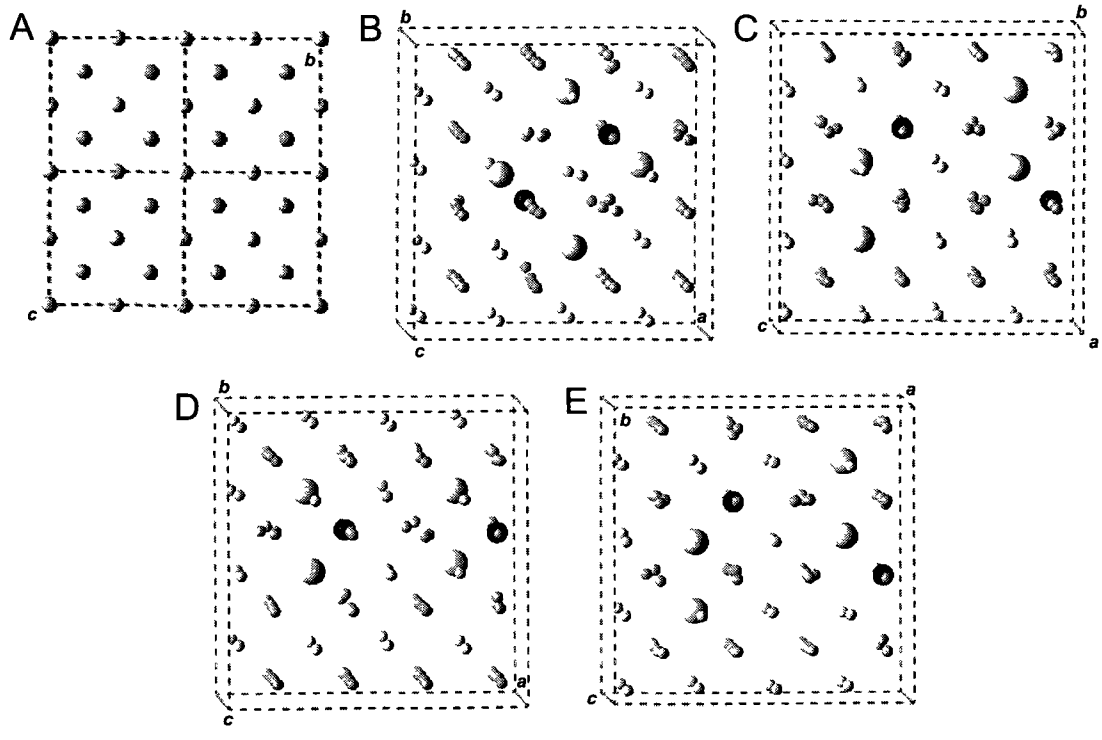


Figure 5.4. 6.6% SDC structures. a) 2x2x2 ceria supercell. Two views of structure A1 are shown projected along the b) [001] axis and c) [100] axis. Two views of structure A2 are shown projected along the d) [001] axis and e) [010] axis.

The difference between the classical and DFT methodologies is best captured by comparing structures **A1** and **A3**, which differ only by the relative positions of the vacancies. Classically, the short-range dispersion and coulombic interactions are indeed

lower in energy in structure **A3**, by 0.1687 and 0.1129 eV/cell respectively. The long-range coulombic interactions, however, heavily favor structure **A1** (by 0.3833 eV/cell), thereby tilting the total energy to favor structure **A1**. Thus, the force field assembled for this work seems to overestimate long-range vacancy coulombic interactions. However, as discussed above, it correctly ranks the remaining low energy structures and provides a quick, manageable way to screen the PES of 6.6% SDC before using DFT calculations. Given that a genetic algorithm using only first principles methods is unfeasible due to limited computational resources, unless the PES is extremely small, it can be stated that classical methods are necessary but not sufficient for "exhaustive" first-principles evaluations.

The results presented for 6.6% SDC parallel that of the 3.2%, 1.9%, and 3.8% SDC studies discussed in Section 5.3.1, finding that configurations with dopants positioned at both the NNN and NN positions to the vacancies, such that the NNN positions are favored, are the most thermodynamically stable configurations of SDC. The pool of low-energy, near-degenerate, structures clearly show the NNN site is preferred in SDC materials, agreeing qualitatively with recent classical and DFT investigations of less concentrated SDC in the literature.[5, 6]

In summary, the GA-inspired search algorithms of doped metal oxides implemented in this work was able to automate the search process of 6.6% LDC materials. Such concentrations have not been previously systematically investigated due to the sheer number of possible configurations. It was shown that by automating an evolutionary search process, one is able to quickly screen out unfavorable candidate structures with classical methods and then apply these same evolutionary procedures with

higher-theory DFT calculations. Note that genetic diversity is preserved in the DFT-based components of the GA by introducing a number of random structures into the gene pool and continuing the mating and mutation routines, this is particularly important when using an inadequate screening potential.

5.4 - Classical GA Investigations of SDC Materials: theoretical insights into trends in optimal ionic conductivities

There is a general consensus that doped ceria's ionic conductivity can be determined from its relation to the activation energy of oxygen vacancies diffusing through the lattice through a conventional Arrhenius relationship. The activation energy is parsed into two contributions: 1) the migration barrier of vacancies hopping between adjacent oxygen lattice sites, and 2) the association energy of the dopants and vacancies in the bulk material. The migration barriers have recently been studied extensively from cluster models derived from 3.2% SDC materials,[3, 6] and were found to have a linear relationship across the lanthanide series, failing to single out SDC or GDC as the most promising candidates seen experimentally.[4]

The association energy of the dopant/vacancy aggregates evidently depends on the relative dopant/vacancy distribution in the native ceria lattice and has itself been partitioned into several different components, such as the electronic or elastic contributions to the energy,[3] but no particular component was shown to dominate the overall trends seen in the relative energies. Association energies are often quantified between LDC materials by referencing configurations to a carefully chosen reference state, be it the highest-energy state[5] or a construction with non-interacting defects[3] (modeled by separating the dopants and vacancies as much as possible in the simulation

cell). Wei. *et. al.* suggest that association energies become increasingly important at low temperature since the activation energy of an oxygen to an adjacent vacancy site will depend on both the height of the energy barrier and on the relative energies of the initial and final states. For example, on the left-hand reaction coordinate in Figure 5.5, both the forward and reverse directions would have to overcome the same energy barrier. However on the right-hand reaction coordinate, the forward direction has to overcome a much larger barrier than the reverse direction, thereby introducing an asymmetry to the reaction profile that would favor the lower-energy minimum on the potential energy surface and perhaps trap it there. The crossover, across the lanthanide series, from NNN to NN dopant site preference[5] with respect to the vacancy at gadolinium suggests the association energies play a central role in defining LDC's improved ionic conductivity near the center of the lanthanide series.

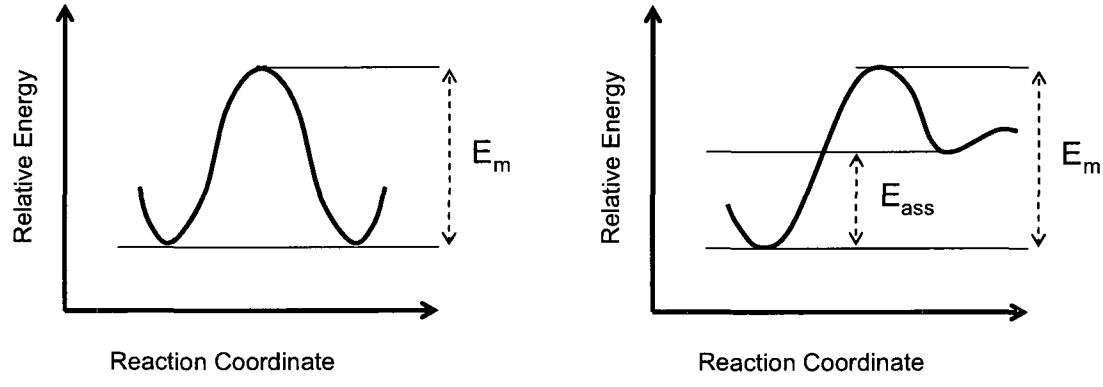


Figure 5.5. Illustration of the effect of relative association energies, E_{ass}, on migration barriers, E_m, at low temperatures. Higher relative association energies (right) lead to more asymmetric reaction profiles.

From Sections 5.3.1 and 5.3.2, the relative energies of different 3.8% and 6.6% LDC configurations to the lowest energy structure can be used to quantify relative association energies within a specific LDC material, specifically for 6.6% LuDC and GDC materials or 3.8% and 6.6% SDC materials. Previous LDC studies focused solely

on 3.2% LDC simulation cells, neglecting intra-cell vacancy-vacancy interactions and thereby restricting the overall configurational space such that the vacancies are distributed homogeneously throughout the crystal; the vacancies were separated by the lattice constant of the cell. Using the force field parameters from ref. 5, it was found that specific configurations with exclusive NN or NNN interactions were heavily thermodynamically favored in 6.6% LuDC and SDC materials, respectively, by 0.04 eV/cell in LuDC and by 0.07 eV/cell in SDC. Consider that, under these conditions, should a vacancy migrate such that it shifts from NNN to NN sites with respect to a dopant atom, it would result in an asymmetric reaction profile which would heavily favor one direction over the other. Identical GA trials on 6.6% GDC, however, show that neither interaction is particularly favored, the most favored structures that lie within 0.02 eV/cell have both NN and NNN dopant/vacancy interactions. The thermodynamic arguments discussed above would indicate that GDC configurations have lower relative association energies and would therefore have improved ionic conductivity, since vacancies shifting between NN and NNN positions relative to the dopant atoms would have near-equivalent association energies and more symmetric reaction profiles. There would then be no significant energy penalty invoked when a vacancy migrates from NNN to NN, or vice-versa, and promote diffusion of the vacancy throughout the lattice.

In this section, further evolutionary structure searches are used to study even higher concentrations of doped ceria, up to 20.0% SDC, at the classical level of theory. The optimal ionic conductivities of Samarium-doped ceria are typically measured near ~11 % SDC,[2] and, until this threshold, the conductively increases with Sm concentration. These trends reinforce the notion that explorations of association energies at high SDC

concentrations are needed to properly explain its functionality, most notably to explain SDC's similarity with GDC. In this regard, the reported trends of enhanced conductivity as the concentration is increased at low concentrations and of decreased conductivity beyond ~11% SDC thresholds are explored in this section by linking these trends to physical defect observables from the GA studies. Within each GA trial, the relative energies of different LDC configurations at a given concentration to its lowest energy configuration are once more used to quantify the structures' relative association energies. The trends in defect-defect associations observed over the range of SDC concentrations are then used to assess the trends in SDC's overall conductivity. Furthermore, since these calculations use inexpensive classical force fields, the structures are derived from 324-atom 3x3x3 ceria simulation cells instead of the standard 96-atom 2x2x2 cell, thereby allowing more flexibility in the way defects are distributed about the cell.

The specifications of the evolutionary-inspired configuration search methodologies used in this study have been described in Section 5.2. These methodologies were shown to be effective at navigating non-trivial potential energy surfaces of doped ceria. Specifically, two GA trials for each concentration were set up such that each trial stored 500 structures per population. Each trial was propagated for 250 generations. An extra trial was introduced with a population size of 800 for the two highest concentrations and each was propagated for 500 generations. Although each trial converged to a different energy and absolute convergence of the structures cannot be guaranteed for such large systems, it is reasonable to assume that the trials progressed far enough such that optimal defect/defect interactions dominate the overall structure, given

the frequency at which the mating/mutation rates produced lower energy structures at the end of the trials was near one in every 50 generations.

The GULP simulation package[19] was used to perform all the classical molecular mechanics calculations. The Buckingham potential was used to model short range dispersion pair-wise interactions, Ewald's method[20] was used to sum the long-range Coulombic interactions, and the shell model[21] was used to account for the polarizability of the O^{2-} ions. The force field parameters were taken to be the same as those reported in Table 5.1, wherein the Buckingham and shell parameters were assembled from the force fields reported by Balducci et al.[22] and Senyshyn et al.[23] Note this force field was found to agree with first principles DFT calculations at 3.2% SDC concentrations.

Consider that, experimentally, both SDC and GDC are noted to have high ionic conductivities.[7] There is a general consensus in the literature that SDC materials however favor NNN dopant/vacancy interactions while GDC does not, suggesting SDC's conductivity would be limited by the asymmetric reaction profile of vacancies migrating from NNN to NN sites. In Section 5.3.1, the GA search procedures were shown to effectively probe low dopant concentrations of SDC in 3x3x3 ceria simulation cells, up to 3.8% SDC, reinforcing that SDC materials heavily favor structures that preferably place the dopant Sm atoms NNN to the vacancies, which agrees with recent experimental EXAFS and theoretical studies.[5-7] In addition, the 6.6% SDC and 3.8% SDC results presented in Sections 5.3.1 and 5.3.2 suggest that, at the DFT level of theory, there is no heavily favored distribution with exclusive NN or NNN dopant/vacancy associates in SDC, as supported by the recovery of several configurations with minimal, near-equivalent association energies that sport both NN and NNN vacancy/dopant interactions.

At these concentrations, this notably puts SDC materials on a more level playing field with GDC materials in the sense that there are a number of energetically-accessible "hybrid" configurations (possessing both NN and NNN vacancy-dopant interactions) with near-degenerate association energies.

One could conceive that as the concentration of Sm is increased, more NN vacancy-dopant interactions would be introduced as the defects fill more of the simulation cell and allow an even larger pool of low-energy and near-degenerate defect configurations. This claim can readily be investigated with the GA methodologies described above by setting up a series of classical GA trials to incrementally investigate higher concentrations. With this goal in mind, seven sets of two GA trials were run on SDC materials derived from 3x3x3 ceria simulation cells such that 6, 8, 10, 12, 14, 16, or 18 vacancies are distributed over the lattice within each set. This simulates 5.8% (11.1 mol %), 8.0% (14.8 mol %), 10.2% (18.5 mol %), 12.5% (22.2 mol %), 14.9% (25.9 mol %), 17.4% (29.6 mol %), and 20.0% (33.3 mol %) SDC materials, respectively. The 12.5% and 17.4% SDC concentrations were chosen to match the experimental data points reported at and beyond the peak in SDC ionic conductivity from ref. 2 (shown *vide infra* in Figure 5.9), which correspond to ~20 and ~29 mol % Sm concentrations respectively.

As expected, it was noted that as the dopant concentrations were increased, the most favored dopant/vacancy distributions took on more NN dopant-vacancy interactions per simulation cell. To illustrate this emergence of NN interactions, the number of NN interactions in the three lowest-energy dopant-vacancy associates recovered at each concentration is shown in Figure 5.6. These findings further reinforce the notion that as

the concentration is increased, SDC behaves similarly to GDC with regards to mimicking the latter's preference of NN/NNN hybrid configurations.

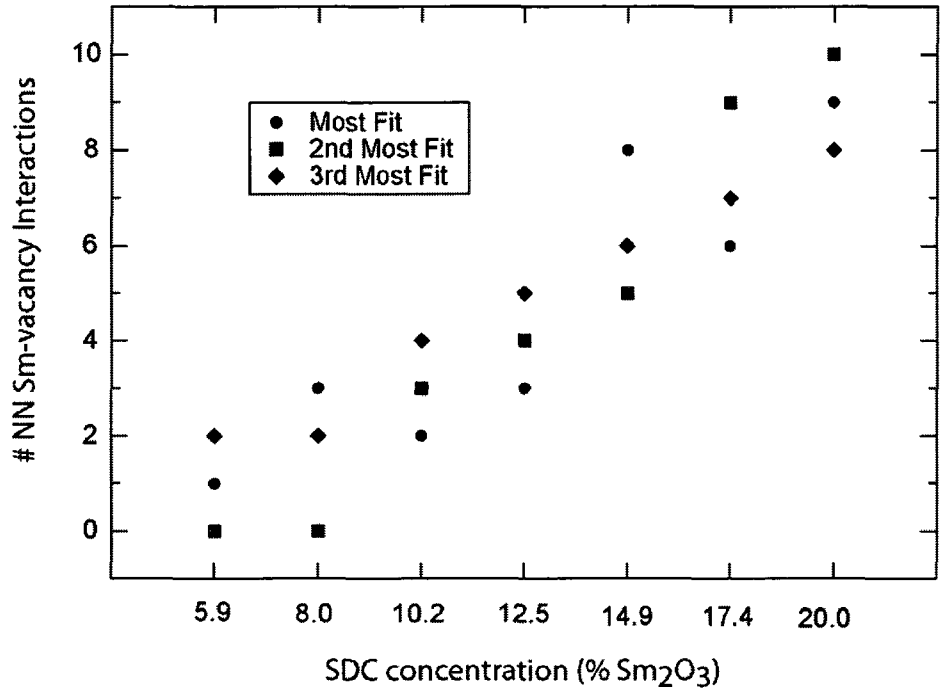


Figure 5.6. Plot of the number of NN Sm-vacancy interactions in the 1st, 2nd, and 3rd most fit structures recovered from the high SDC concentration GA trials on 3x3x3 CeO_2 simulation cells.

In addition, we noted an additional nuance in the low concentration SDC structures. In all of the low energy configurations, the defects would cluster together within the simulation cell and create pockets of undoped ceria in the periodic lattice. The most extreme uniaxial projection in this regard within the lowest energy 3.8% SDC structure is shown in Figure 5.7a, wherein the defects aggregate into layer-like structural motifs. This anisotropy is quantified at each concentration by showing the number of oxygen planes, those which lay perpendicular to a lattice vector, that contain no oxygen vacancies. As shown in Figure 5.8, the spread of defects about the simulation cell indeed becomes more isotropic in SDC materials at higher concentrations; the number of oxygen

planes with no defects approaches and converges to zero as the concentration is increased beyond 12.5% SDC concentrations. For a qualitative comparison with 3.8% SDC, the most anisotropic uniaxial projection of the lowest energy 12.5% SDC structure is shown in Figure 5.7b. To ensure this behavior was not simply an artifact of periodic boundary conditions at low concentrations, we ran a further GA trial on 3.2% SDC derived from a 576-atom 4x4x3 ceria simulation cell and noted similar defect-defect associations, shown by a gradual convergence within the most fit structures to boast vacancy-vacancy associates with inter-defect distances near 6.1Å.

The aforementioned anisotropy in low energy configurations at low concentrations would undoubtedly introduce an anisotropy into the crystal's overall ionic conductance and could thus be used to explain both the lower average conductivity at low dopant concentration (see discussion of Figure 5.9 below) and the increased conductivity of SDC at higher concentrations (from a more isotropic distribution of defects and a mixed NN/NNN preference of the vacancy sites).

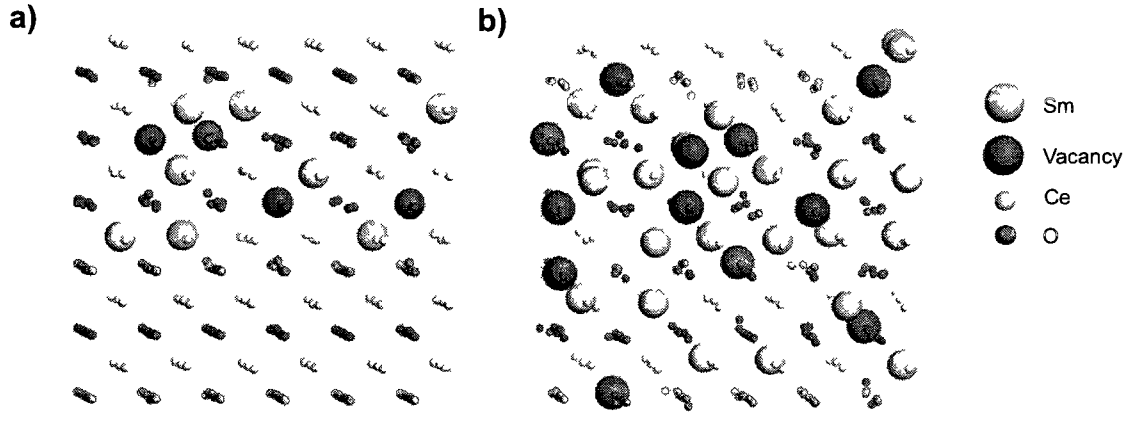


Figure 5.7. Perspective views of most the fit structure recovered from GA-based structure searches of a) 3.8% SDC and b) 12.5% SDC. The ceria and oxygen atoms are depicted as small balls and the samarium and oxygen vacancies are depicted as larger balls.

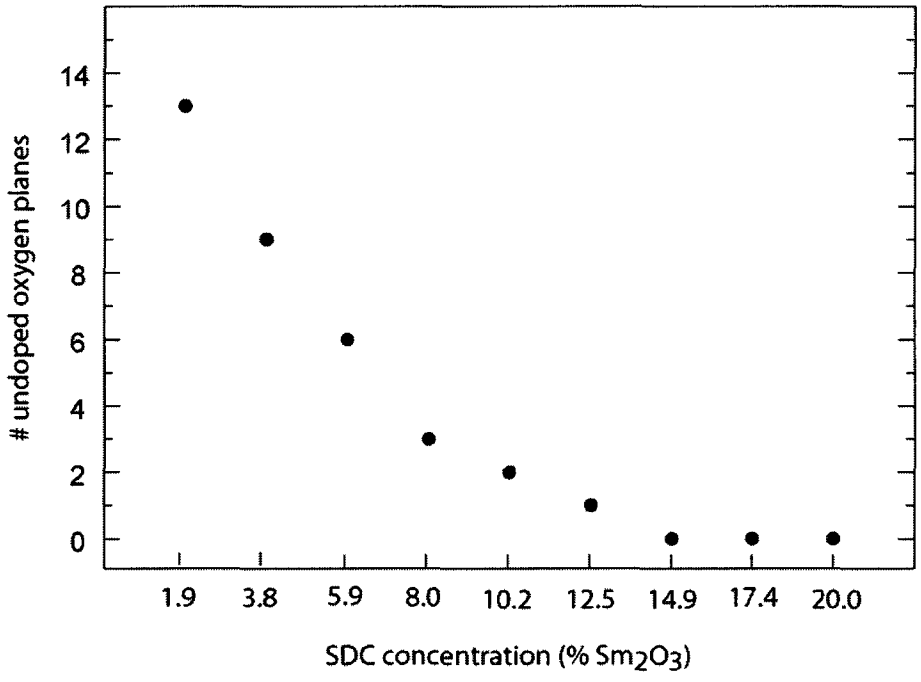


Figure 5.8. Plot of the sum of the number of undoped oxygen planes perpendicular to each lattice vector vs. SDC concentration in the lowest-energy structure recovered at each concentration.

The trends reported in Figures 5.6 and 5.8 can be used to explain an enhancement in ionic conductivity as the Sm concentration is increased. Although this trend is observed experimentally at low concentrations, the ionic conductivity is known to peak at a Sm_2O_3 concentration of ~11% after which it decreases with increasing dopant concentration,[2] as shown in Figure 5.9 wherein the data points from the experiment are shown by the inverted triangles. This suggests there must be a competing structural feature of SDC that reduces the ionic conductivity as the Sm concentration increases. The decrease in ionic conductivity beyond ~11% was proposed to be due to the ordering of oxygen vacancies in the crystal lattice.[2] More recently, a DFT study[6] of doped ceria at low concentration calculated the association energy of vacancy pairs in a ceria lattice. It was concluded that neighboring vacancy-vacancy associates disrupt diffusion pathways

available to the oxygen atoms to migrate throughout the lattice, thereby decreasing the material's overall ionic conductivity. If short vacancy-vacancy distances are linked to SDC's diminished ionic conductivity beyond the peak at ~11%, it should be observable in the GA-inspired structure searches at higher concentrations.

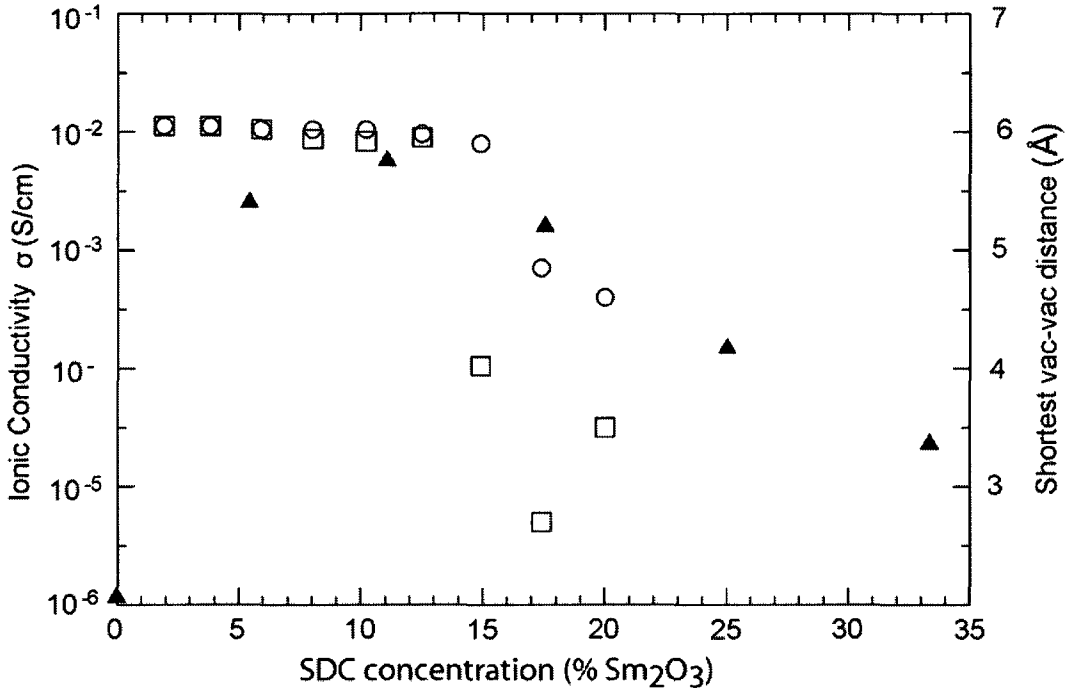


Figure 5.9. Plot showing the dependence of Ionic Conductivity on dopant concentration of SDC materials at 500 °C temperatures (black triangles/left axis).[2] The closest vacancy-vacancy distances from the most fit structure (open circles) and the collective pool of the four most fit structures (open square) recovered from the GA trials vs. concentration is plotted with unfilled symbols (right axis).

To assess the nature of vacancy-vacancy interactions in concentrated SDC structures, the shortest vacancy-vacancy distance in the lowest energy structure recovered at each concentration is plotted as an open circle in Fig. 5.9, corresponding to the axis on the right of the plot. At first glance, it can be seen that vacancy-vacancy interactions measuring near 6.0 Å were consistently observed in the lower concentration regimes (< 8.0% SDC), once more suggesting that the vacancies cluster in the simulation cell and

further reinforcing the notion that there is an anisotropic distribution of vacancies at lower concentrations.

Encouragingly, all of the lowest energy structures from the GA trials at concentrations up to 14.9% SDC kept all the vacancies at least 6 Å apart from one another. Conversely, all the trials beyond 14.9% SDC concentrations converged to structures with at least one vacancy-vacancy distance which measured less than 6 Å. This observed drop in vacancy-vacancy distances correlates well with the experimentally-observed drop in conductivities beyond ~11% SDC. It is assumed that such associations would persist at even higher concentrations as well, but higher concentrations were not studied here since 20.0% SDC already represents a concentration which exceeds the solubility limit of samarium in doped ceria.[2]

Note that it is important to consider not only the lowest energy structures at each concentration, but other low energy structures as well in order to assess the defects' potential impacts on vacancies migrating from one configuration to another. To get a better overview of a representative pool of low energy structures at each concentration, the shortest vacancy-vacancy distance observed in the four lowest-energy structures from each concentration trial is plotted as an unfilled square in Fig. 5.9. The observed trend agrees very well with that of the single lowest-energy structure except the distances drop off after 12.5% SDC concentrations instead of after 14.9% SDC, forcing some vacancy pairs to even assume NN positions with respect to one another. These results, therefore, provide further evidence that oxygen vacancy 'ordering' or, more specifically, vacancy-vacancy associates are responsible for the decreased ionic conductivity at high dopant concentrations. In this discussion the four lowest energy structures were discussed

because this corresponds to the two most 'fit' structures recovered from each GA run at each concentration. However, it is important to note that if the eight lowest energy structures are analyzed in the same manner, there is still a dramatic drop in the shortest vacancy-vacancy distance at 12.5% SDC. As such, the forced introduction of neighboring vacancy-vacancy interactions observed in low energy structures beyond 12.5% SDC indeed appears to be responsible for the material's decreased conductivity.

To further justify the role of the aforementioned neighboring vacancy-vacancy associates on the GA trials, a rough estimate of vacancy-vacancy association energies with the force field used in this study in 1.9% SDC structures is shown in Figure 5.10. The estimates were calculated by setting up two thousand sets of calculations in which four Sm atoms and two vacancies are distributed about the simulation cell. In each set, the four Sm atoms are placed at random Ce lattice sites and kept constant while the two

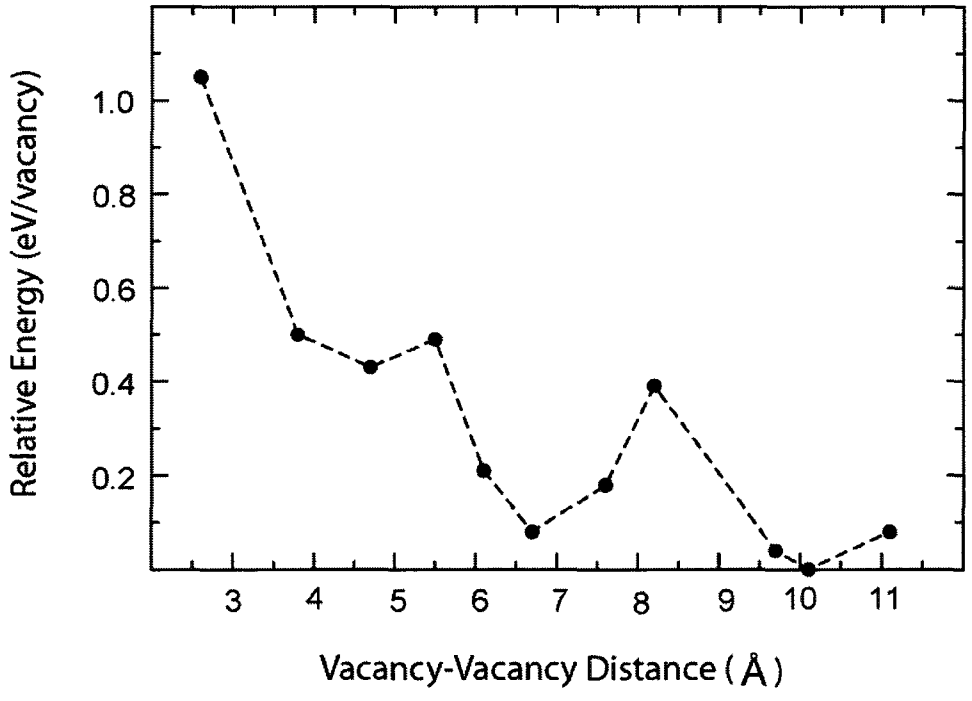


Figure 5.10. Plot of vacancy-vacancy association energies as a function of vacancy-vacancy distance in 1.9% SDC simulation cells.

vacancies are placed at predefined oxygen sites such that they sample the vacancy-vacancy distance range shown in Figure 5.10. Although the long-range electrostatic Sm-vacancy interactions are likely not consistent within each set, their effects are assumed to balance out over the series' entirety. Overall, the results do suggest a general repelling of the oxygen vacancies as they are brought closer together, becoming particularly repulsive below 6 Å, and sport a locally favored vacancy-vacancy distance between 6 and 8 Å. It can be rationalized from this plot that vacancy-vacancy distances between 6 and 8 Å are stabilized over their periphery by Sm-vacancy interactions. It is noted that such vacancy-vacancy distances are in fact observed in all the low-energy configurations recovered at low concentrations, see Figure 5.9, and are characteristic of the aforementioned anisotropic defect distributions. Coupled with the high vacancy-vacancy association energies observed below 6 Å, these observations correlate well with the presumed "ordering" of oxygen vacancies assumed to be responsible for the decrease in ionic conductivities beyond 11% SDC.

In conclusion, it has been shown genetic-algorithm -inspired search procedures for Samarium doped ceria (SDC) can shed light on the underlying principles responsible for Samarium's role in enhancing the ionic conductivity of doped ceria. The methodology successfully reproduced the expected trend of neighboring vacancy-vacancy interactions beyond 12.5% SDC, reinforcing the proposed notion that SDC's decreased ionic conductivity beyond ~11% SDC is caused by the "ordering" of oxygen vacancies. Furthermore, it was shown that its increased conductance properties at low concentrations can be attributed to two factors. First, although SDC clearly favors NNN Sm-vacancy interactions to NN interactions, its most favored configurations feature

"hybrid" interactions that incorporate both NN and NNN interactions such that NNN is favored strongly but not exclusively. This opens up a pool of energetically-favored and energy-degenerate structures that would ease a vacancy's migration through the lattice. Importantly, this also puts SDC on a more level playing field with GDC materials in this regard, which is an important connection since the most promising conductive substrates in LDC-based electrolytes for solid-oxide fuel cell technology is both SDC and GDC. Second, we noted that at low concentration the defects tend to aggregate in the simulation cells at low concentrations, thereby introducing an anisotropy into its conductivity profile. The distribution becomes more isotropic at higher concentrations and would henceforth improve its overall conductivity as well, explaining the rise in conductivity seen in experiment.

Furthermore, these calculations further demonstrate that the GA-inspired search procedures for doped metal-oxide materials presented in this work can even navigate 324-atom and 576-atom simulation cells with as many as 54 defects, making for incredibly large search spaces that are inaccessible through other search methods even when symmetry-equivalent geometries are taken into consideration.

5.5 - Conclusions

In summary, a specialized genetic-algorithm (GA) -inspired search procedure for doped metal oxides has been presented and tested in this chapter; the procedure allows the study of generic defect association complexes in metal oxide materials at experimentally-relevant dopant concentrations. The key aspect behind these search techniques was the incorporation of mating and mutation routines that preserve the

identity of the parent metal oxide itself; note that the specific mating operations applied herein are readily transferable to any doped metal oxide framework.

To demonstrate the effectiveness of these GA-inspired search procedures on doped metal oxides, it has been shown such structure searches of doped ceria reproduce the results of classical, systematic searches for 3.2% samarium-, gadolinium-, and lutetium-doped ceria. A number of distinct potentials were tested and the GA protocols successfully recovered the well-documented crossover across the lanthanide series from NNN to NN dopant site-preference to the vacancy at Gd^{3+}. By using tandem classical and density functional theory energy calculations, the GA-inspired search procedures successfully mapped out 6.6% SDC at the DFT level of theory for the first time. The use of a classical optimization scheme in the early stages of the GA greatly accelerated the search process and proved to be an excellent screen for the more rigorous DFT optimizations that followed.

Specifically, it was found that the NN-NNN structure was the most stable 3.2% SDC structure from first-principles density-functional theory calculations, representing a balance of the competing effects of dopant-vacancy and dopant-dopant interactions. At higher concentrations, the NNN site tends to be clearly preferred over the NN site, as expected from the literature. It was interesting to observe these NNN interactions in the larger 3x3x3 ceria cells at a similar concentration to those studied in previous works on 2x2x2 simulation cells (3.8% vs. 3.2% SDC). In 3.8% SDC, representing the larger 3x3x3 simulation cell, there is a single NN dopant/vacancy association in the six most stable configurations, implying there are multiple energetically-competitive such configurations of SDC since, by symmetry, there are a number of ways to construct

crystal structures with such defect-defect associations. These trends were again seen at the DFT level of theory in 6.6% SDC simulation cells; note that such observations could not be extracted from the 3.2% SDC simulation cells due to the imposed symmetry of the smaller 2x2x2 CeO_2 supercells. It is noted that these energetically-competitive structures would undoubtedly impact the material's oxygen mobility, since their existence provides an abundance of SDC structures with relative association energies near zero and eases the energetic profile of diffusion pathways throughout the lattice. Given the linear dependence of the migration barrier on trivalent lanthanide dopants as one progresses along the lanthanide series, this would then single out SDC, along with GDC, as being a promising material for minimizing the asymmetric thermodynamic profile of migration barriers, as discussed by Wei *et al.* and Nakayama *et al.*

Building on these efforts, the GA-inspired search procedures were shown to function remarkably well even with unprecedented large search spaces. Specifically, they were applied to systems with high Sm concentrations in 3x3x3 simulation cells such that the experimentally-observed optimal concentration of ~11% SDC could be assessed. This involved navigating simulation cells with up to 18 vacancies (and 36 dopant atoms) to reach a peak concentration of ~20.0% SDC. The observed trends were reminiscent of the NNN-favoring structures recovered from the 6.6% and 3.8% SDC structure searches and, furthermore, the experimentally-expected ordering of the oxygen vacancies in the lowest energy structures was observed beyond ~11% SDC concentrations. In addition, an anisotropy was noted in defect associations at lower SDC concentrations which would help further explain the nature of increased ionic conductivity as the concentration is increased in this regime.

The methodology presented in this work was therefore shown to allow the study of generic defect association complexes in metal oxide materials. Knowledge of the atomic positions of SDC provided herein, for example, could make possible a computational study of the catalysis of this material. Since SDC and GDC are so prominently used in solid oxide fuel cells, and their mechanism of action remains unknown, these results provide a basis for investigating this mechanism through appropriately-designed surface models. Regarding further possible applications of GA-inspired search methodologies toward other metal oxides, it is noted that, when it becomes feasible to do so, the GA algorithms presented here could readily be used to sample the free energy instead of the enthalpy to assess scenarios where the temperature and oxygen chemical potential are known to play key roles. Furthermore, this methodology not only allows general studies of defect association energies, but lays a further foundation to readily be incorporated into proposed "inverse optimization" schemes[16] in order to optimize a particular property of an oxide material. This particular application is discussed further in Chapter 6.

5.6 - References

1 H. Inaba and H. Tagawa, Solid State Ion. **83**, 1 (1996).

2 H. Yahiro, Y. Eguchi, K. Eguichi, and H. Arai, J. Appl. Electrochem. **18**, 527 (1988).

3 D. A. Andersson, S. I. Simak, N. V. Skorodumova, I. A. Abrikosov, and B. Johansson, PNAS **103**, 3518 (2006).

4 K. Muthukkumaran, Bokalawela, R., Mathews, T., Selladurai, S., J. Mater. Sci. **42**, 7461 (2007).

5 X. Wei, W. Pan, L. Cheng, and B. Li, Solid State Ion. **180**, 13 (2009).

6 M. Nakayama and M. Martin, PCCP **11**, 3241 (2009).

7 Y.-P. Fu, S.-B. Wen, and C.-H. Lu, J. Am. Ceram. Soc. **91**, 127 (2008).

8 Z. Yang, G. Lao, Z. Lu, and T. K. Woo, J. Phys.: Condens. Matter **20**, 035210/1 (2008).

9 Z. Yang, T. K. Woo, and K. Hermansson, J. Chem. Phys. **124**, 224704/1 (2006).

[10] D. M. Deaven and K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995).
[11] Y. Zeiri, Phys. Rev. E **51**, R2769 (1995).
[12] G. H. Johannesson, T. Bligaard, A. V. Ruban, H. L. Skriver, K. W. Jacobsen, and J. K. Norskov, Phys. Rev. Lett. **88**, 255506 (2002).
[13] A. R. Oganov, Glass, C. W., J. Chem. Phys. **124**, 244704 (2006).
[14] G. Trimarchi and A. Zunger, Phys. Rev. B **75**, 104113 (2007).
[15] J. Beckers, F. Clerc, J. H. Blank, and G. Rothenberg, Adv. Synth. Catal. **350**, 2237 (2008).
[16] K. Kim, P. A. Graf, and W. B. Jones, J. Comp. Phys. **208**, 735 (2005).
[17] S. V. Dudiy and A. Zunger, Phys. Rev. Lett. **97**, 046401 (2006).
[18] J. K. Norskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen, Nature Chemistry **1**, 37 (2009).
[19] J. D. Gale, J. Chem. Soc. Faraday Trans. **93**, 629 (1997).
[20] P. P. Ewald, Ann. Phys. **64**, 523 (1921).
[21] J. D. Gale, Phil. Mag. B **73**, 3 (1996).
[22] G. Balducci, Kaspar, J., Fornasiero, P., Graziani, M., Islam, M. S., J. Phys. Chem. B **102**, 557 (1998).
[23] A. Senyshyn, Oganov, A. R., Vasylechko, L., Ehrenberg, H., Bismayer, U., Berkowski, M., Matkovskii, A., J. Phys.: Condens. Matter **16**, 253 (2004).
[24] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).
[25] G. Kresse and J. Furthmuller, Comput. Mater. Sci. **6**, 15 (1996).
[26] P. E. Blochl, Phys. Rev. B **50**, 17953 (1994).
[27] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett **77**, 3865 (1996).
[28] D. Balamurugan, W. Yang, and D. N. Beratan, J. Chem. Phys. **129**, 174105 (2008).

CHAPTER SIX

INCORPORATION OF DFT EVOLUTIONARY SEARCH METHODOLOGIES INTO A GENERIC INVERSE OPTIMIZATION SCHEME: TUNING ELECTRON OR ION MOBILITY IN DOPED METAL OXIDES

This chapter focuses on showing that evolutionary search procedures of defect associations in doped metal oxides coalesce well with proposed "inverse optimization" schemes from the literature, methodologies which optimize structure configurations by a metric other than the structures' relative energies. The methodology is shown to be capable of tuning electronic and/or ionic properties of doped metal oxides by optimizing electronic mobility in doped zinc and tin metal oxides and ionic conductivity in doped ceria.

CHAPTER 6 – *KEY SUBSECTIONS*

6.1 - Introduction

In Chapter 5, it was shown that search algorithms inspired by the genetic algorithm (GA) can be used to efficiently search out low-energy defect-associate configurations in doped metal-oxide substrates. These techniques fully warrant further development on their own merits, since they can be used to solve a long withstanding problem in solid state chemistry as to how to go about distributing defects about a suitably sized simulation cell. Without methodologies like these, current computational probes into defect association energies of doped metal oxides have been limited to either: 1) low concentration scenarios, where it is hoped that the documented trends persist at higher and more experimentally relevant concentrations, or 2) configurations dictated by chemical intuition or dynamical trajectories, which, despite the best intentions of the researcher(s), could lead to erroneous conclusions if relevant configurations are overlooked.

However, the scope of these GA methodologies toward research efforts concerning doped metal oxides extends well beyond searching out configurations with low association energies. Often, experimentalists and theoreticians alike are interested in doping metal oxides to improve one particular aspect of its fundamental ionic or electronic behavior. The idea that follows this line of thinking is to incorporate a descriptor of a specified property into the search algorithm itself in order to better focus on designing a material engineered specifically for the problem at hand. This concept contrasts sharply to the alternative scenario wherein one performs a number of different configurational searches of low-energy conformers from a variety a dopant profiles and, through further analysis, hopes some of them exhibit the desired behaviors.

A number of these so-called "inverse" optimization schemes, methodologies that recover dopant configurations which optimize a particular aspect of the starting material, have been proposed in the literature as a precedent for their application in band-structure computation methods to optimize electronic structure properties in semiconductor alloys.[1-4] In this chapter, it is shown the GA-inspired search procedures discussed in this thesis, specifically those used in chapter 5 to study doped ceria, can be readily applied in these inverse algorithm constructs to further study doped metal-oxide materials with the aim of improving the fundamental inherent electronic or ionic properties which suit their technological applications.

For example, it is well-known that cassiterite (SnO_2) is used extensively in gas-sensor technologies due to an observed spike in its resistivity when a guest molecule is adsorbed onto its surface at high temperatures.[5, 6] The nature of the anomaly in the material's surface conductivity can be used to distinguish between different guests, but such distinctions or selective characterizations are often difficult to parse or characterize. Several recent experimental studies have focused on doping SnO_2 materials to enhance their gas-sensing behaviors, wherein different dopants, such as Ce or Ru,[7, 8] are found to promote more selective behavior for specific guest molecules but the underlying mechanisms responsible for the selectivity are unknown.

Although a material's electronic conductivity can be affected by several contributing factors, the effective electronic masses and electronic concentrations generally play fundamental roles in its formulation.[9] In doped metal oxides, the modified bulk behaviors observed from the doping impurities are thought to stem from the dopants' effect on the concentration and mobility of charge carriers. In some cases, the

dopant raises the energy of the states just below the Fermi level, thereby narrowing the band gap and increasing the concentration of free electrons from the upward shift of the Fermi level.[6] When designing gas sensors, care must be taken in this regard not to enhance the electronic conductivity to the extent that the signal associated with guest adsorption on the surface is unobservable; this is a generalization of why semi-conductors, and not metals, are used in sensor technologies. It has also been proposed that highly-dispersed conduction bands are important towards promoting high electronic mobility, ie. low effective electronic masses (m*), due to the incidental higher-energy inter-band transitions from the valence and conduction bands alike with crystal momentum.[10] It is in fact well known that, through semi-classical theory of semiconductors, the electronic conductivity, σ, is related to the current density, J, which, in turn, is related to the electronic mobility, μ. The electronic mobility is, through equations 6.1 and 6.2, proportional to the dispersion of the conduction band at the Γ-point[10] (ie. the second derivative of the energy (E) with respect to the wavevector).

$$\mu = \frac{e\tau}{m^*} \qquad [6.1]$$

$$m^* = \hbar^2 \left(\frac{d^2E}{dk^2} \right)^{-1} \qquad [6.2]$$

In equations 6.1 and 6.2: μ represents the electronic mobility, m* represents the semi-classical electronic effective mass, e represents the electronic charge, τ relates to electronic collision probabilities, and ħ is Planck's constant.

Therefore, in order to enhance the electron mobility of a given oxide material, it should prove advantageous to introduce dopant atoms that promote dispersion, ie. a sharper curvature, of the metal oxide's conduction band. This would enhance the oxide's electronic mobility and, on a grander scale, could possibly have beneficial repercussions toward its role in gas-sensor technologies.

Building on these concepts, the GA-inspired methodologies are used in this chapter to study the effect of dopants on the band-gap, a rough measure of the carrier concentration, and on the curvature of the conduction band at the gamma point, a rough measure of the effective electronic masses, in ZnO and SnO_2 materials. ZnO and SnO_2 were selected since they both represent metal-oxides that are used extensively in current transparent conducting-oxide technologies.[9] The "inverse" optimization protocols can readily be introduced into GA frameworks by changing the nature of the GA's fitness metric to reflect the dispersion (ie. the curvature) of the conduction band and the band gap. Obviously, the energies of the structures themselves should not simply be ignored, but, given that the scope of this particular study is to probe the effectiveness of these inverse optimization schemes, they are not yet considered in this study. Note that meaningful conclusions can still be extracted from such a set up, for example it can readily be assessed whether the manner in which the defects are spread about the cell does indeed affect the property in question. This would be useful in helping experimentalists determine the best manner to go about synthesizing the material, eg. whether to perform the synthesis under thermodynamic or kinetic control. The syntheses of metal oxides are typically done at high temperatures and the manner in which they are cooled ultimately controls the oxides' configurations. For example, the oxides can be

gradually annealed or flash-frozen to emulate the thermodynamic or kinetic conditions, respectively. Should the defect configuration itself prove to be difficult to control, the results of inverse GA's can still be used to assess how reproducible the results can be expected to be.

Inverse GA methodologies also have the potential to optimize ionic properties of doped metal oxides, like their ionic conductivities. In Chapter 5.4, GA-inspired structure searches were run on high concentrations of Samarium-doped ceria (SDC), up to 20% Sm_2O_3 concentrations, and it was proposed that three factors govern the nature of SDC's ionic conductivity profile. The first concerns an anisotropic distribution of defects about the simulation cell at low concentrations. The second stems from an observation that although SDC clearly favors NNN Sm-vacancy interactions to NN interactions with the simulation parameters used herein, its most favored configurations feature "hybrid" interactions that incorporate both NN and NNN interactions such that NNN is favored strongly but not exclusively. The third concerns the introduction of neighboring vacancy-vacancy interactions at high concentrations, which is responsible for SDC's diminished performance beyond 12.5% SDC. As such, it stands to conjecture that these three proposed factors could themselves be used as fitness metrics in an inverse GA procedure to optimize the dopant identities in lanthanide-doped ceria (LDC) materials. A preliminary assessment of this strategy is reported for 10.3% LDC materials with L = Sm, Gd, Lu, or Y. In this case study, unlike the metric proposed above to tune electron mobility in ZnO and SnO_2 materials, the relative energies of each possible LDC structure is considered as well. The relative energies are crucial elements when considering a

material's ionic conductivity profile at low temperatures due to the thermodynamics that governs the inherent mobility of the oxygen ions.

Specifically in this chapter, traditional GA searches, those which use each structure's relative enthalpy as its fitness metric, were first run on Hg-doped ZnO and Pb-doped SnO_2 simulation cells at a number of different concentrations. Mercury and Lead were chosen as dopant atoms since they are heavier elements in the same group as the metal in the parent oxide (group 12 for Zn/Hg and group 14 for Sn/Pb) and the additional screening of the nuclear charge by the core electrons could be used to rationalize higher valence electron mobility. These preliminary GA searches were used in order to assess: 1) how many dopant atoms can be introduced before collapsing the band gap and 2) how the curvature of the conduction band is affected by the dopant atoms. These results were then used to define, implement, and test an alternative fitness metric for the GA protocols that reflects the curvature of a given structure's conduction band. In doing so, the GA program was modified to allow a flexible assignment of numbers and identities of user-specified dopant atoms and successfully reproduced the results from the systematic energy-based GA trials. The important point to consider here is that these inverse GA protocols now sample the oxide's chemical space in addition to its configurational space. The newly-implemented inverse band optimization scheme was run on Hg-/Cd-/Ba-/Sr-/Ca-/Mg- and then Hg-/Cd- doped ZnO simulation cells to demonstrate how appropriate the protocols are for screening a vast array of candidate doped metal-oxide frameworks at a range of unspecified, but naturally evolving, concentrations. Finally, an inverse GA optimization procedure intended to optimize ionic conductivity in 10.3% LDC materials (L = Y, La, Nd, Sm, Gd, Er, Yb, or Lu) is implemented in Section 6.4 in a "proof of

principle" context. A single GA trial with a fitness metric based on both the relative energies and the number of NN dopant/vacancy interactions was found to correctly rank 10.3% SDC and GDC structures as the most fit candidate materials. This correlates well with the well-known assertion that SDC and GDC are the most promising singly-doped substrates to use as solid electrolytes in solid oxide fuel-cell technologies.

6.2 – Computational Method

The specifications of the GA machinery used in this work have been described at length in Chapter 5. Modifications were made directly to this framework in order to first implement a methodology capable of tuning the dispersive behavior of the conduction band in ZnO/SnO_2 materials. The most significant changes were pertinent to the fitness evaluation and mating/mutation stages and will be described in detail in Section 6.2.2. A recap of the GA-inspired search methodologies employed in Chapter 5 to study doped metal-oxide frameworks is first given in Section 6.2.1 to remind the reader of the methodology's structure and make this chapter self consistent. The reader is advised to skip Section 6.2.1 if he/she is already familiar with the methodology outlined in Chapter 5, section 5.2.

6.2.1 – Recap of GA-Inspired Search Algorithms

Genetic algorithm studies traditionally represent a crystal by a list of N atomic Cartesian coordinates, requiring no prior information about the system aside from the composition and volume of the unit cell. Since the target doped metal oxide materials all represent more-or-less minor distortions of the native oxide's lattice coordinates, the crystal structures in this study were similarly represented by the Cartesian coordinates of

each lattice site[1] in the undoped metal oxide lattice. The genetic representation of a generic doped metal oxide crystal is therefore constructed by assigning N dopant atoms to distinct metal lattice sites and, if specified, M vacancies to distinct oxygen sites. A collection of doped structures, or "genes", is herein referred to as a "population". The initial populations are constructed by substituting dopant atoms and/or oxygen vacancies at randomly selected metal or oxygen sites on the undoped lattice.

For structure evolution, the number of atoms was kept constant for the duration of the procedure wherein, at each step, randomly chosen mutation or mating operations were used to combine structural motifs of one or two randomly chosen "parent" genes. The mating/mutation operations were designed to accommodate both dopant- and vacancy-type defects and preserve the metal and oxygen coordinates in the original undoped lattice. The mating and mutation operations used in this study were the following: i) Geometry Perturbation, randomly perturbing the cartesian coordinates of a chosen subset of atoms from the parent to make the offspring, ii) Dopant Mutation, swapping individual cartesian coordinates of a randomly chosen subset of dopant atoms with those of a randomly chosen subset of target metal atoms, iii) Vacancy Mutation, swapping individual cartesian coordinates of a randomly chosen subset of oxygen atoms with those of a randomly chosen subset of oxygen vacancies, iv) Defect Swapping, swapping the metal lattice sites of one parent structure with those of the other parent structure, and v) v) Defect Swapping and Mutation, a tandem application of the Defect Swapping mating routine with either a Dopant or Vacancy Mutation. Only one randomly selected mating/mutation operation was used to generate a single offspring. Often these mating procedures generated physically unreasonable crystal structures, thus offspring structures

were only added to the next generation if all atomic pairs were at least 1.0 Å apart, otherwise the same mating operation was repeated. The resulting offspring structures are given a twenty-five percent chance to mutate further through geometry perturbations. This perturbation involves randomly selecting between one and one half of the physically relevant lattice sites in the gene and randomly perturbing each of their x, y, and z Cartesian coordinates by a maximum of 0.2Å.

The fitness of a candidate structure was determined from its relative lattice energy to the lowest-energy structure in its population, as derived from classical polarizable force fields or first principles DFT calculations. The population of subsequent generations was created in the following manner. First, a pool of the fittest structures were promoted, unperturbed, to the next generation. The number of promoted genes can be specified by the user but, in this study, the default was taken to be the ceiling function of 15 and 15% of the total population size. To generate the rest of the population, fit structures were preferentially chosen as parent structures by selecting them from a Boltzmann weighted probability distribution.[9] The exponential term stems from the relative lattice energies, scaled by an appropriate temperature to allow at least a 25% chance to select half of the promoted structures. Unphysical structures are screened out during the mating process and the offspring structures are loosely optimized before evaluating their fitness.[2] The genetic algorithm is typically halted when the energy of a user-specified pool of the lowest energy structures remains unchanged for a number of generations. These convergence criteria are very dependent on the properties of the system being studied and are chosen carefully at the discretion of the user.

Within this framework, each structure was optimized at the DFT level of theory. All of the first principles DFT calculations were performed with the Vienna ab-initio Simulation Package (VASP).[11, 12] The projector augmented wave (PAW) method of Blöchl[13] was used to treat the core states, with the oxygen $2s^2$ and $2p^4$, the Zn, Cd, and Hg valence s^2 and d^{10}, the Ca, Sr, and Ba subvalence s^2 and p^6 and valence s^2, and the Mg $3s^2$ electrons being treated as valence electrons. A plane-wave cutoff of 520 eV was used in all geometry optimizations and single-point energy evaluations with the gradient-corrected exchange and correlation functionals of Perdew-Burke-Ernzerhof (PBE)[14]. For energy evaluations within the genetic algorithm, the Brillouin-zone was sampled only at the Γ-point, whereas Monkhorst-Pack grids were constructed from 5x5x5 meshes in refined calculations of the phases selected for further study. Variable cell geometry optimizations at zero pressure were used to evaluate the fitness of all structures, as derived from the VASP simulation package.

6.2.2 - Establishing an Appropriate Fitness Metric

First, traditional GA-inspired searches, as outlined in Section 6.2.1, were run on 13.9 mol% Hg-doped ZnO materials generated from 3x3x2 simulation cells of ZnO. The 3x3x2 replica of the four-atom $P6_3mc$-symmetry ZnO unit cell, generated from Zn *2a* and O *2b* Wyckoff sites, was chosen to create a 72-atom simulation cell, a reasonable choice to balance the computational demands and the flexibility in allowed dopant concentrations for subsequent calculations. The 13.9 mol% concentration corresponds to placing 5 Hg atoms at Zn sites in the simulation cell. Three separate GA trials were run with 50 genes per population and were propagated for 15 generations. All three trials converged to the same lowest-energy structures; their quick convergence was expected

due to the high symmetry of the ZnO lattice and the equivalent local chemical environments of every Zn atom in its native lattice. The band structures of the conduction, valence, and sub-valence (valence-1) bands along three high-symmetry paths in the native ZnO simulation cell's k-space are shown in Figure 6.1 for the 1st, 2nd, and 8th lowest energy structures from the GA searches. The results suggest that, in this system, the manner in which the dopant 5 Hg atoms are distributed in the simulation cell has a nominal effect on the curvature of the conduction band and a minimal effect on the material's band gap. From an experimental point of view, such behavior is desired since this would mean that one would not have to be careful to make a specific configuration when synthesizing the doped metal oxide. It will be shown later, however, that this is not the case with most of the configurations discussed in this section.

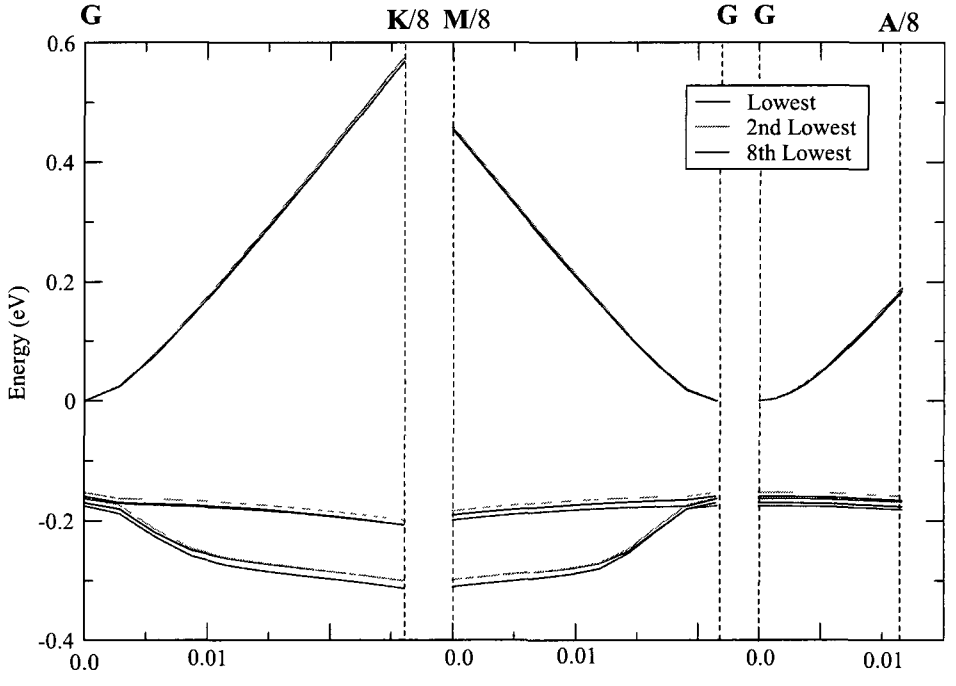


Figure 6.1. Band structures of the conduction band (top), valence band (middle), and valence-1 band (bottom) along high-symmetry paths in reciprocal space of the native ZnO 3x3x2 simulation cell. The band structures are color-coded for the 1st (black), 2nd (pink), and 8th (blue) lowest energy structures of 13.9% Hg-doped ZnO materials. (G = Gamma point)

Further sets of GA trials were then performed on ZnO 3x3x2 simulation cells doped with 1, 2, 3, 4, 5, 6, and 7 Hg atoms, spanning concentrations from 0 mol% to 19.4 mol% Hg. This range was chosen to encompass all the possibilities in which the doped oxide remains a semiconductor; the band gap was found to collapse in the lowest-energy structures, at the DFT level of theory, when ZnO was doped with 7 Hg atoms. Note that the Hubbard (U-term) correction to DFT was not employed for a similar reason as to why it was not applied in the aforementioned doped ceria calculations. The primary purpose of this work is to evaluate the robustness of a GA inspired structural search algorithm when applied to doped metal oxides. In this work, a number of different dopants are examined, which, like the Hubbard correction, would change the nature of the conduction band. The U parameter would rigorously need to be fit to some reliable metric, like the band gap or lattice constant, for each material.

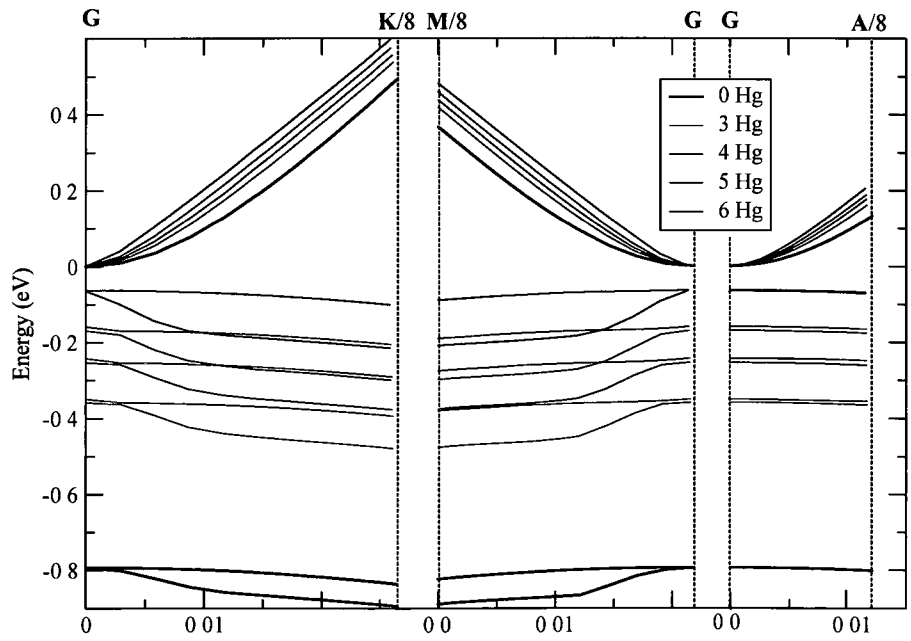


Figure 6.2. Band structures of the conduction band (top), valence band (middle), and valence-1 band (bottom) along high-symmetry paths in reciprocal space of the native ZnO 3x3x2 simulation cell. The band structures are color-coded for the simulation cells doped with 0 Hg(black), 3 Hg (pink), 4 Hg (orange), 5 Hg (blue), and 6 Hg (green) atoms. (G = Gamma point)

The band structures of the conduction, valence, and sub-valence bands along three high-symmetry paths in the native ZnO simulation cell's k-space are shown in Figure 6.2 for the lowest energy structures from the 0% (0 Hg), 8.3% (3 Hg), 11.1% (4 Hg), 13.9% (5 Hg), and 16.7% (6 Hg) Hg-doped ZnO GA searches. The plot shows the trends desired from doping ZnO to supposedly improve its sensor functionality, namely a concerted band-gap closure and a steeper energy-rise (higher dispersion) in the conduction band near the Γ-point as the dopant concentration is increased.

A similar series of traditional GA trials on 72-atom 2x2x3 simulation cells of the rutile SnO_2 structure demonstrates similar trends, as shown in Figure 6.3. In Figure 6.3, the band structures of the conduction, valence, and sub-valence bands of the lowest-energy structures of 0% (0 Pb), 12.5% (3 Pb), 16.7% (4Pb) Pb-doped SnO_2 along three high-symmetry paths in the native SnO_2 simulation cell's k-space are shown The doped SnO_2 material's band gap collapses when doped with five or more Pb atoms are introduced and although higher dispersion in the conduction band is noted with increasing dopant concentration, its effect is far less pronounced than that observed in Hg-doped ZnO materials at the DFT level of theory.

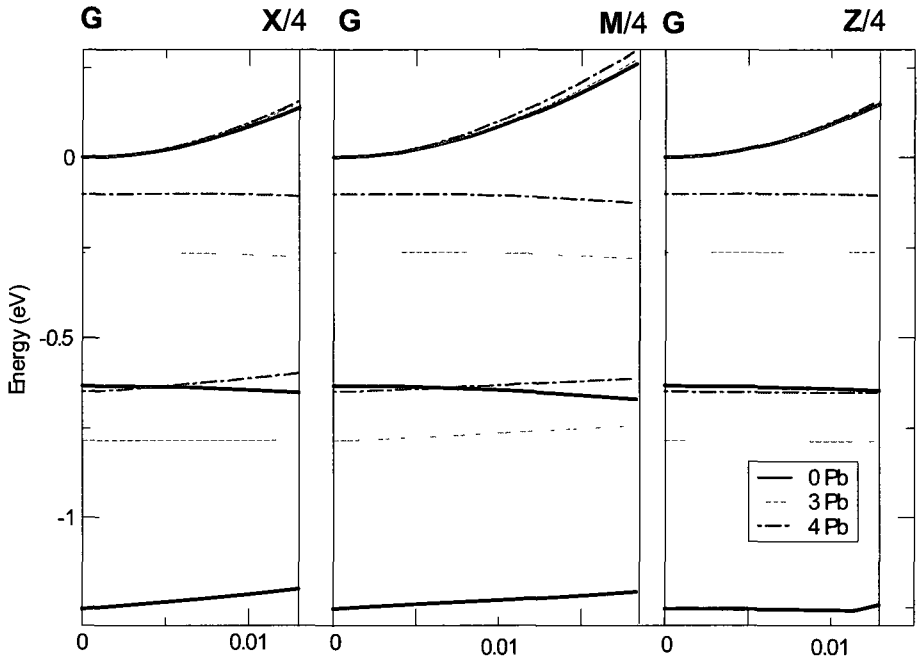


Figure 6.3. Band structures of the conduction band (top), valence band (middle), and valence-1 band (bottom) along high-symmetry paths in reciprocal space of the native rutile SnO_2 2x2x3 simulation cell. The band structures are color-coded for the simulation cells doped with 0 Pb (black), 3 Pb (dotted red), and 4 Pb (dashed-dotted blue) atoms. (G = Gamma point)

In light of these results, ZnO was chosen to help create, test, and benchmark a modification to the fitness metric of the GA procedures such that they are tuned to reflect the goals of the study to search out doped metal-oxide frameworks with sharper curvatures in their conduction bands at the gamma point, thereby increasing the materials' electronic mobility. The fitness metric was implemented by first checking the band gap of the material to ensure a direct band gap was still present in its overall band structure, otherwise a fitness of zero is automatically assigned to its genetic representation. Such an assignment is justified since the goal of this study is to tune the dispersive behavior of the conduction band, and the conduction band is ill-defined when there is no band gap. Furthermore, as mentioned in the introduction, semi-conductors are the target materials in sensor technologies since one must ensure that the signal associated with guest adsorption can be observed over the background conductivity of the

native oxide; metals perform poorly in this regard. This was implemented by ensuring the band gap at the Γ-point was at least 0.05 eV and that the valence band maximum and conduction band minimum were both located at the Γ-point. Since conventional DFT calculations (not employing DFT+U corrections) are being used for this particular study, its well-documented underestimation of band gaps in metal oxides ensures only semiconductors are assigned favorable fitness metrics. As such, this study falls into line with other recent DFT studies which use the "scissors" approach on doped SnO_2 to discuss its gas-sensing behaviors;[8] the "scissors" method simply shifts the unoccupied bands of a semi-conductor up by a fixed energy constant.

Next, the program was modified to read in a user-specified set of k-points along which the curvature of the conduction band is to be measured. Care must be taken to at least sample all the symmetry-inequivalent directions in reciprocal space. This was tested by supplying an 8-point three-dimensional cubic grid of points, including the Γ-point. The grid is set up in k-space such that the points are homogeneously spread over the first quadrant in reciprocal space and each point is ~0.005 $Å^{-1}$ from the Γ-point along a low-index Miller axis in the native ZnO simulation cell. The curvature of the conduction band is then estimated by averaging the rise in energies of all the k-points surrounding the Γ-point divided by their distance from the Γ-point. As such, the fitness of each structure is determined as an averaged fitness over all specified k-points, the formula for the metric at each k-point is expressed in equation 6.3:

$$\text{fitness}(i, k_j) = \begin{cases} \dfrac{E_{CB,i}(k_j) - E_{CB,i}(\Gamma)}{\| k_j - \Gamma \|} & , \text{if } E_{CB,i}(\Gamma) - E_{VB,i}(\Gamma) > 0.05 \text{ eV} \\ 0 & , \text{if } E_{CB,i}(\Gamma) - E_{VB,i}(\Gamma) < 0.05 \text{ eV} \end{cases} \qquad [6.3]$$

where i denotes the *i*th metal oxide structure, j denotes the *j*th k-point in the grid supplied by the user, $E_{CB,I}(x)$ refers to the energy of the conduction band in the *i*th metal oxide structure at the point *x* in reciprocal space, and $E_{VB,I}(x)$ refers to the energy of the valence band in the *i*th metal oxide structure at the point *x* in reciprocal space.

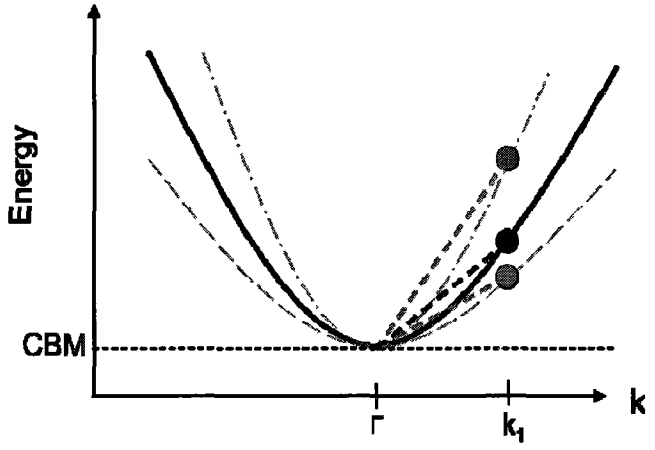


Figure 6.4. Illustration of the fitness metric to tune the dispersive behavior of the conduction band. A generic conduction band is shown near the Γ-point as a solid black line; CBM= conduction band minimum. Examples of more dispersed (thin dashed-dotted line) and less dispersed (thin dashed line) bands are shown and can be distinguished by the slopes of the thick dashed lines.

This metric is then used directly as the fitness of the candidate doped metal-oxide framework. In this way, structures which show a larger spread in energies around the Γ-point (ie. a larger slope) are assumed to have higher dispersion and assigned the better fitness values. A depiction of this metric for a simple conduction band in a one-dimensional k-space is shown in Figure 6.4, depicting how the dispersion of the three bands is assessed in this work by the slope of the line connecting the dots to the Γ-point.

The two most prominent concerns about this particular fitness metric stem from the treatment of the multidimensional curvature of the conduction band in reciprocal space. Specifically, two points will be addressed here: 1) the practice of sampling points in the first quadrant, and 2) the equal weighting assigned to each k-point when averaging

the curvature over the grid in reciprocal space. Since the introduction of dopant atoms lowers the symmetry of ZnO's hexagonal lattice, one would, ideally, like to sample the entire Brillouin zone. However, in order to accommodate limited computational resources and adhere to reasonable compute times, these studies are limited to sampling ~10 k-points when using a 72-atom simulation cell of ZnO. The first quadrant was explicitly chosen in this regard because the irreducible Brillouin zone (which contains the range of curvatures over the entire Brillouin zone) can often be encompassed within a single quadrant (this, of course, is not always the case) and it is shown later in this section that such a practice recovers the expected behaviors of Hg-doped ZnO simulation cells with increasing concentration. The equal weighting of each k-point was assigned simply because it is the most intuitive way to gauge the potentially complex behavior the band structure over momentum space and alternative weighting schemes can not be justified without further incorporation of the symmetry of the Brillouin zone into the calculations.

To evaluate this metric, the fitness evaluation stage of the GA consists of three stages: 1) a gamma-point DFT geometry optimization of the overall structure, 2) a single-point DFT energy evaluation with a 3x3x3 Monkhorst-Pack grid to extract the charge density, and 3) a single-point DFT energy evaluation, using the user-specified k-points and the charge density from the previous step, to extract the band structure behavior near the Γ-point as detailed above.

Concerning other GA concerns for the new methodology, the energy resolution used thus far in the GA procedure to weed out redundant genotypes, based on the structures' relative energies, is left unchanged and used in concert with the curvature-related fitness metric and defect-defect radial distributions to identify redundant

genotypes. The GA procedure was also modified to allow a flexible number of dopant atoms over each population within a user-defined threshold, ie. the user defines a maximum concentration. Finally, the GA was allowed to choose the dopant from a user-specified pool of atomic species, necessitating a further GA parameter to specify the maximum number of elemental species allowed in a given doped structure (ie. specify whether the generated doped MOs can be binary, ternary, quaternary, etc.). The last two changes required the introduction of three new mutation routines and one new mating routine: 1) the "Change Number of Dopants" mutation which, as its name suggests, adds or deletes a randomly chosen dopant atom to/from the ZnO lattice, 2) the "Change Identity of Dopants" mutation which leaves the number of dopants unchanged but changes the identity of one or more dopant atoms, 3) the "Change Number and Identity of Dopants" mutation which simply applies both the former mutations, and 4) the "Swap Dopant Types" mating routine which combines subsets of each parent's dopant atoms into one structure. The new routines were implemented such that their generated offspring always satisfy the parameters of the GA trial, namely to adhere to the maximum number of allowed dopant atoms and types of dopant atoms.

A single GA trial, using the fitness metric and modifications described above, was then run on Hg-doped 3x3x2 ZnO simulation cells such that between 0 and 8 Hg dopants atoms are distributed among the lattice, thus sampling both the concentration of Hg and its configurational space. Using these parameters, the optimal doped ZnO structures are expected to have 6 Hg atoms. With a twenty-five gene population size, the GA trial recovered the expected 16.7% Hg-doped ZnO after twelve generations, after which the most fit structure was converged for four generations. A supplemental GA trial also

recovered the same structure. The fitness metrics of an array of structures recovered from the trials is shown in Figure 6.5.

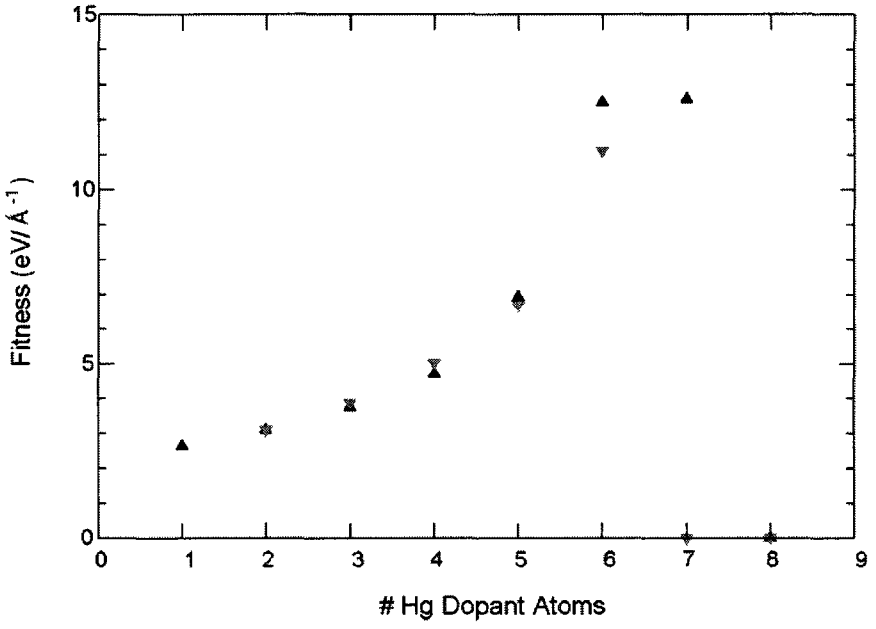


Figure 6.5. Summary of GA fitness values (reflecting the dispersion of the conduction band at the Γ-point) vs. the number of dopant Hg atoms in a 3x3x2 ZnO simulation cell. The symbols (triangles, inverted triangles, and circles) denote different configurations recovered from the GA trials.

As noted, the trials converged to recover the expected 16.7% Hg-doped ZnO materials after twelve generations, correctly ranking a subset of the remaining concentrations in the expected order (19.4% < 0% < 2.7% <5.56% <8.3% <11.1% <13.9% <16.7% Hg-doped ZnO shown by the inverted red triangles). However, the trials also recovered a couple of 19.4% Hg-doped ZnO materials with elite fitness metrics, on par with the ~13 eV/Å^{-1} fitness metrics of 16.7% Hg-doped ZnO. It was verified that these particular 19.4% Hg-doped ZnO structures had band gaps between 0.05 and 0.06 eV, thereby satisfying the band-gap criteria implemented for the fitness metric, which screens out structures with band gaps less than 0.05 eV. Furthermore, the results suggest that as the band gap narrows, the distribution of dopants about the lattice does indeed

affect the dispersion of the conduction band at the Γ-point, contrary to what was initially assumed from a systematic study of 13.9% Hg-doped ZnO and 12.5% Pb-doped SnO_2. Encouragingly, all of the 22.2% Hg-doped ZnO structures (8 dopant atoms) were assigned 0 $eV/Å^{-1}$ fitness metrics, correctly reflecting the collapse of their band gaps.

The fitness metric implemented in the featured GA-inspired inverse optimization methodology proved to be successful in reproducing the observed trend in more dispersive conduction-band behavior as Hg atoms are introduced to a ZnO simulation cell. Moreover, it is noted that 16.7% mol Hg concentrations produce the optimal fitness metrics, and hence the highest electron mobilities, but show more variation with respect to how the dopant atoms are distributed about the cell than the next lowest concentration, at 13.9 mol% Hg. These variations would need to be carefully considered when interpreting these results to design optimal experiments. To elaborate, consider that it may be desired to prepare a sample which shows the most consistent mobility behavior, independent of its configuration, in order to facilitate reproducible results in the lab. In this regard, it was found that 13.9 mol% Hg-doped ZnO would exhibit the most reproducible, optimal electron mobilities. Moreover, the lowest energy 16.7% Hg-doped ZnO registered a fitness metric near 11 $eV/Å^{-1}$, slightly lower than the optimal value of ~14 $eV/Å^{-1}$ at that concentration. This would imply that care must be taken when synthesizing 16.7% Hg-doped ZnO if one wishes to achieve optimal conductivity. However, since the presented results on Hg-doped ZnO were merely meant to probe the validity of the fitness metric, these insights will not be elaborated upon further here.

6.3 – Using GA-inspired Inverse Optimization Schemes to Screen a Large Number of Candidate Doped Metal-Oxide Structures at the DFT Level of Theory

The motivation behind devising these inverse optimization schemes is to use them to quickly screen a library of candidate dopant atoms to find the optimal identities and distributions of the dopant atoms. To demonstrate the validity of such schemes with the GA protocols implemented above, the ZnO framework is doped, in this section, with a flexible number of 1) Hg, Cd, Ba, Sr, Ca, and/or Mg dopant atoms and, as a refinement on the larger-scale search, 2) Hg and/or Cd dopant atoms. The GA trials performed in this section were performed on 2x2x2 ZnO simulation cells instead of the 3x3x2 simulation cells presented above only in the interest of saving time while benchmarking the procedure. The increased compute times for these systems stem from addressing the larger search space involved with placing distinct dopant species about the lattice.

First, to reference the Hg-doped ZnO GA trials presented in Figure 6.5, the inverse-optimization GA procedure was rerun on Hg-doped 2x2x2 ZnO simulation cells. In this light, two distinct GA trials were run on 32-atom 2x2x2 ZnO simulation cells which again allowed between 0 and 8 Hg dopant atoms to be distributed over the Zn lattice sites. Both trials used twenty-five gene population sizes and were propagated for 7-8 generations before convergence was achieved for four generations. The most fit structures are reported in Figure 6.6, labeled by their dopant distributions and fitness metrics.

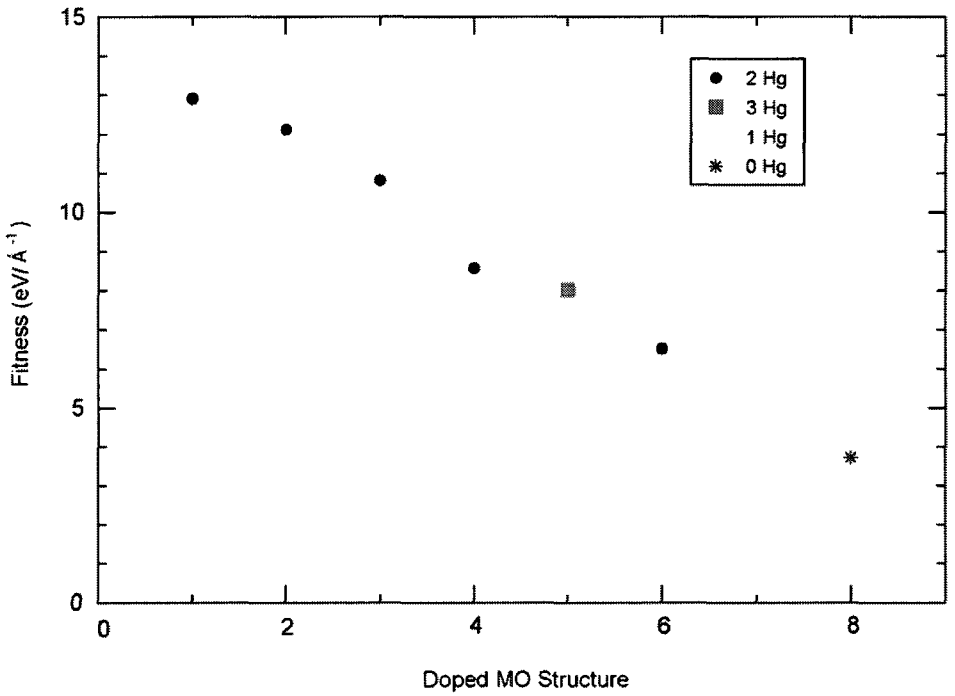


Figure 6.6. Fitness metrics of the 7 most-fit structures (ranked from 1 though 7) recovered from the GA trials optimizing the dispersive behavior of the conduction band in Hg -doped ZnO materials constructed from a 2x2x2 simulation cell. The undoped ZnO structure is shown as a reference (structure 8).

The results show that the cells doped with 2 Hg atoms (corresponding to 12.5% Hg-doped ZnO) emerged as the most fit structures, reflecting near-equal fitness metrics to the fitness metrics of the most fit Hg-doped ZnO structures constructed from 3x3x2 simulation cells. All structures with concentrations greater than or equal to 25 mol% Hg (> 4 Hg atoms) were assigned zero fitness metrics. The 18.75 mol% Hg (3 Hg atoms) structures behaved analogously to the 19.4% mol% Hg structures observed from 3x3x2 simulation cells, possessing band gaps near 0.05 eV that fluctuated depending how the dopants were distributed among the lattice. Unlike the 3x3x2 simulation cells however, no elite 18.8 mol% Hg structure was recovered in a 2x2x2 simulation cell, this can likely be attributed to the symmetry constraints of the smaller simulation cell. The 6.25 mol%

Hg-doped ZnO structures (1 Hg) fare the poorest among structures with non-zero fitness metrics, as expected from the results from the 3x3x2 simulation cells. The fitness metric of the undoped 2x2x2 ZnO simulation cell is also shown in Figure 6.6, to further demonstrate how the Hg dopants, in general, improve the dispersive behavior of the conduction band.

Overall, the results from the 2x2x2 ZnO simulation cells correlate well with those reported for the 3x3x2 simulation cell, suggesting that the optimal dopant concentration lies between 12.5-18.75 mol% and 16.7-19.4 mol%, respectively, for the two cells. It should be noted that the symmetry-constraints of the 2x2x2 simulation cell result in a larger window of uncertainty regarding the concentrations. One striking feature of the results from Figure 6.6 is the variations in the fitness metrics among 12.5 mol% Hg-doped ZnO structures, ranging from ~12.9 to ~6.5 $eV/Å^{-1}$. As shown in Table 6.1 however, each structure does indeed boast dissimilar Hg-Hg radial distributions and reinforce the notion that the structure-screening protocols in the GA can not rely on the fitness metric or atomic configuration alone.

Table 6.1. Structural characterizations of the Hg-Doped ZnO Structures from a 2x2x2 simulation cell

GA Rank	*# Dopants*	*Shortest Hg-Hg Distance (Å)*
1	2	6.81
2	2	3.30
3	2	4.63
4	2	6.30
5	3	3.30
6	2	3.48
7	1	-

Once the relation between the 3x3x2 and 2x2x2 simulation cells was established, the smaller cell was used to assess how scalable these GA protocols are to screen a library of dopant atoms. The GA-inspired search procedures were therefore run on Hg-/Cd-/Ba-/Sr-/Ca-/Mg- doped 2x2x2 ZnO simulation cells such that at most three distinct dopant species are introduced to the lattice in a given structure. Specifically, one single GA trial was run on a 2x2x2 ZnO simulation cell, allowing between 0 and 8 Hg, Cd, Ba, Sr, Ca, and/or Mg dopant atoms to be distributed over the Zn lattice. The trial consisted of seventy genes per population and was propagated for 15 generations. Absolute convergence was likely not achieved (the most fit structure had not stayed constant for more than two generations) but the trial had progressed far enough to draw meaningful conclusions. The most fit structures recovered from the trial are reported at the top of Figure 6.7, referenced by their dopant distributions and fitness metrics.

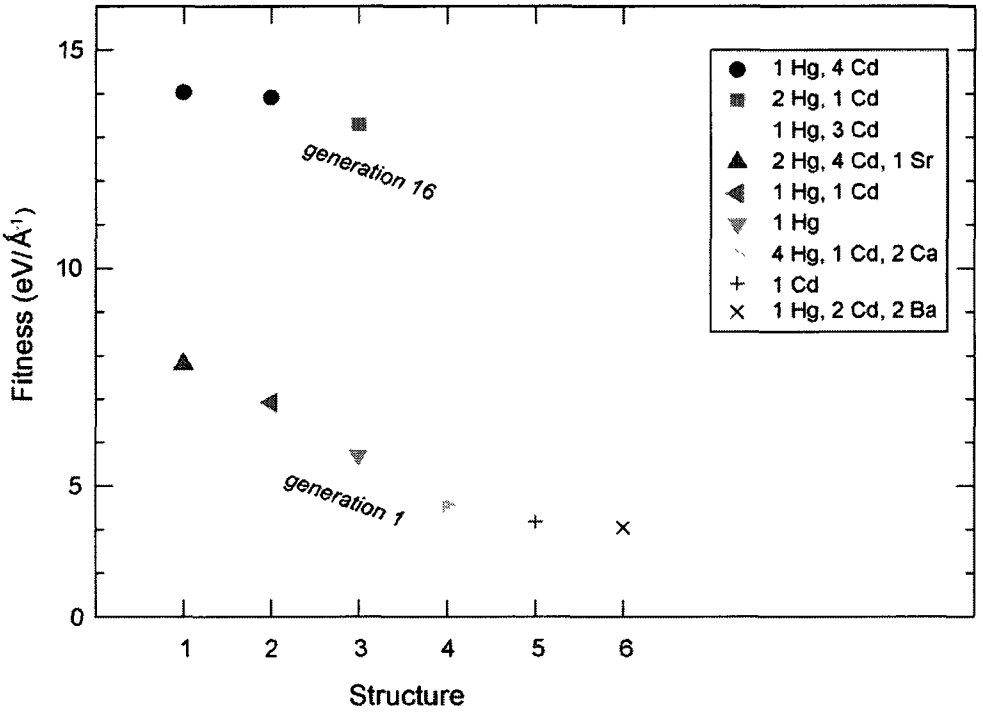


Figure 6.7. Fitness metrics of the most-fit structures recovered from the 1st and 16th generations of a GA trial optimizing the dispersive behavior of the conduction band in Hg-/Cd-/Ba-/Sr-/Ca-/Mg-doped ZnO materials derived from a 2x2x2 simulation cell.

The results reflect those obtained from the GA trials run on Hg- doped ZnO simulation cells, converging to structures with similar fitness values (~13 $eV/Å^{-1}$), and were noted to exclusively promote structures with Hg and Cd dopant atoms. This suggests that the Ba, Sr, Ca, and Mg dopant atoms do not contribute significantly to the conduction band minimum of the metal oxide, since such an effect would be reflected in the dispersion-based fitness metric. The genepool that was generated at random for the first generation was confirmed to be unbiased toward Hg-/Cd- doped materials and the six most fit structures promoted from that first generation are shown in Figure 6.7. The Hg/Cd selectivity can readily be seen even at this stage of the trial, wherein all the promoted structures boasted either Hg or Cd dopant atoms. To help investigate/reaffirm

the origin of this behavior, the density of states of pure ZnO and 12.5% Hg-, Cd-, and Ba-doped ZnO is shown in Figure 6.8. The 12.5 mol% concentration was chosen to mimic the most fit 12.5 mol% Hg-doped ZnO structure recovered from the GA trials, making it suitable to represent other promising candidate materials and, particularly, to study the effects of Hg itself.

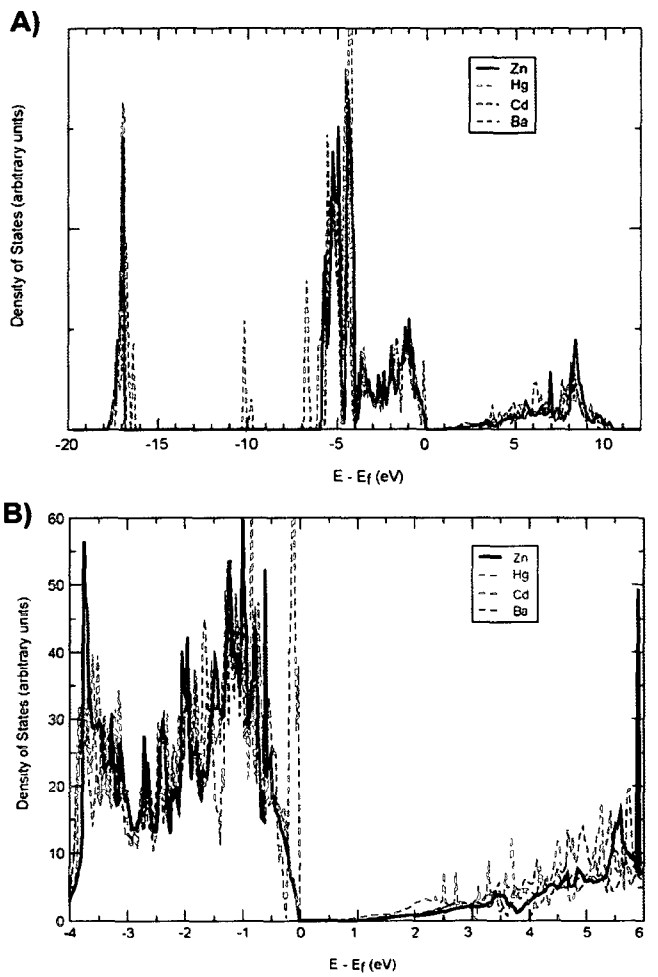


Figure 6.8. A) The calculated total density of states of pure ZnO (solid black line) and 12.5 mol% Hg-, Cd-, and Ba- doped ZnO materials (dashed lines) derived from a 2x2x2 ZnO simulation cell, and B) a close-up of the valence/conduction band region. All energies are given relative to the Fermi level.

All of the aforementioned 12.5% doped and undoped ZnO materials show similar DOS behaviors but there are significant deviations in the valence and conduction band regions, wherein the crucial electronic properties which define its sensor-technology applications lie. It can readily be seen that a number of states are introduced in the valence band near the Fermi level in Ba-doped ZnO materials, however, as speculated, this would not be reflected in the fitness metric used in this study since the metric depends solely on the energy states near the conduction band minimum, not the valence band maximum. In this regard, the Hg-doped ZnO material is observed in Figure 6.8b to introduce the most states near the conduction band minimum, between 0 and 3 eV relative to the Fermi level. Meanwhile, the Ba-doped ZnO material matches the pure ZnO DOS in this region and the Cd-doped DOS lies between the two extremes but closer to that of pure ZnO.

To investigate further, the summed site-projected density of states (PDOS) projected onto the two dopant atoms in each 12.5 mol% doped ZnO material is shown in Figure 6.9a. For pure ZnO, the summed PDOS is shown for the two Zn atoms occupying the equivalent dopant positions from the 12.5 mol% case studies. It is again clearly seen in Figure 6.9a that the Hg atoms contribute substantially more to the conduction band minimum than the Zn, Cd, or Ba atoms. A further projection of the Hg-doped ZnO total density of states onto angular momentum dependent spherical harmonics (not shown) suggests that these energy states arise from the valence s-character orbitals of the Hg atoms.

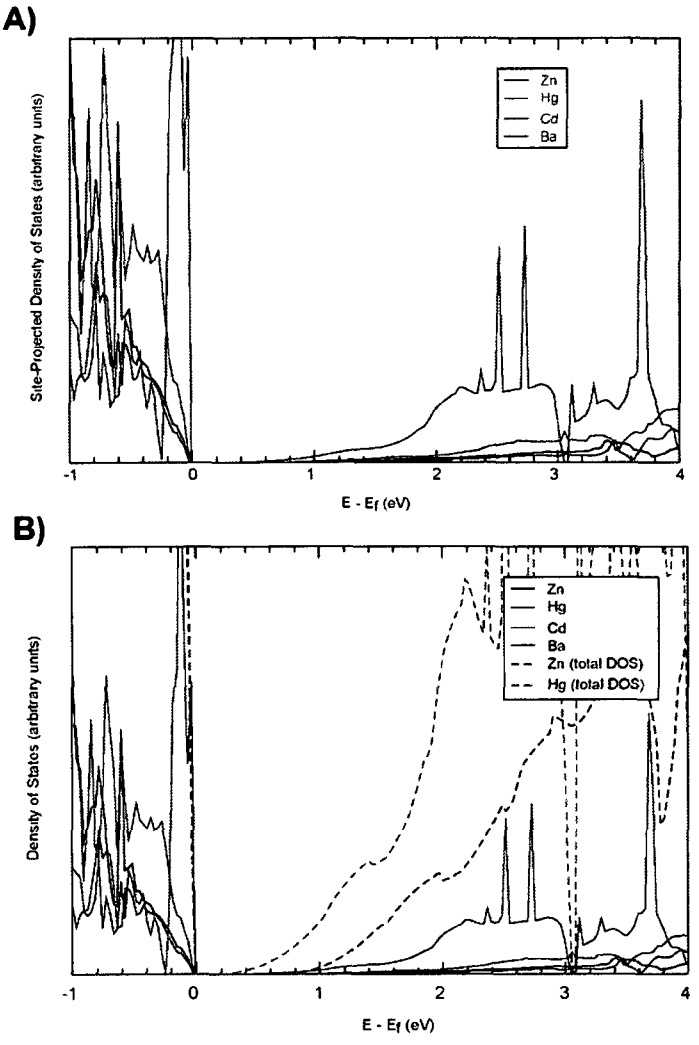


Figure 6.9. A) The calculated site-projected density of states (PDOS) onto the dopant atoms in pure ZnO (solid black line) and 12.5 mol% Hg-, Cd-, and Ba- doped ZnO materials (dashed lines) derived from a 2x2x2 ZnO simulation cell; the PDOS for the symmetry-equivalent Zn atoms are shown for pure ZnO (solid black line). B) A close-up of the valence/conduction band region. All energies are given relative to the Fermi level and the total DOS of Hg-doped and pure ZnO (dashed lines) are overlayed on the PDOS plots in B).

However, the most significant contribution of the dopant Hg atoms toward the conduction band minimum is not only the peak seen between 2 and 3 eV, but the long tail in the conduction band that stretches below 1 eV. To elaborate, consider that the curvature of the conduction band at the Γ-point in the band structure of materials with

direct band gaps, like ZnO, would also be reflected in the curvature of the density of states at the conduction band minimum. When the dispersion of the conduction band at the Γ-point is enhanced, by doping for example, this implies that the energy rises faster with respect to crystal momentum (the phase factor) and, as a consequence, the curvature is sharpened. This increased dispersion would result in fewer states within an infinitesimal energy interval near the conduction band minimum, thereby resulting in a smaller density of states at the corresponding energy. This would indeed be the scenario created in doped ZnO frameworks when dopant atoms are introduced which contribute states below the native conduction band minimum of the material, states that do not overlap with those from the original Zn atoms.

In Figure 6.9b, the total DOS plots of pure and Hg-doped ZnO are superimposed over the PDOS plots to better illustrate the net effect of these conduction band "tails" on the total DOS. Note that the total density of states plots are not normalized and reflect arbitrary units, so they should not be used for quantitative assessments of the Hg/Zn contributions. However, it can be seen that the DOS of Hg-doped ZnO is shifted closer to the Fermi level than that of the undoped ZnO and the curvature near the conduction band minimum is smoother and more pronounced. This further suggests that an alternative fitness metric could be introduced that monitors the density of states instead of the band structure, such a metric would prove useful in cases where more meticulous sampling of the Brillouin zone is required. Such a metric would fail, however, when higher virtual bands contribute to the band structure near the conduction band minimum. Importantly, Cd was noted to introduce a similar tail, to a much smaller degree, and Ba did not contribute any new states in these regions beyond those introduced by the Zn

atoms, reinforcing the expected notion that Hg>Cd>Ba when Hg, Cd, or Ba are introduced to enhance conduction band dispersion in ZnO materials.

To explain why the (1 Hg, 4 Cd) structures consistently emerged as the most promising doped ZnO material, consider that the limiting factor towards increasing Hg-dopant concentrations in ZnO was noted to be the collapse of the band gap beyond 12.5 mol% concentrations in 2x2x2 simulation cells. Although Cd is less effective at promoting dispersion, it also has less impact on closing the band gap. As such, the (1 Hg, 4Cd) structure strikes a fine balance between the two influences.

The GA-inspired search procedures were then run on Hg-/Cd- doped 2x2x2 ZnO simulation cells, thereby allowing the evolutionary protocols to concentrate solely on the favored Hg/Cd configurations. Three distinct GA trials were run on 2x2x2 ZnO simulation cells, allowing between 0 and 8 Hg and/or Cd dopant atoms to be distributed over the Zn atom lattice. The trials used twenty, thirty-five, and fifty gene population sizes and were propagated for 16, 14, and 12 generations, respectively, before convergence of the top three structures was achieved for four generations. The most fit structures recovered from the collective trials are reported in Figure 6.10, referenced by their dopant distributions and fitness metrics.

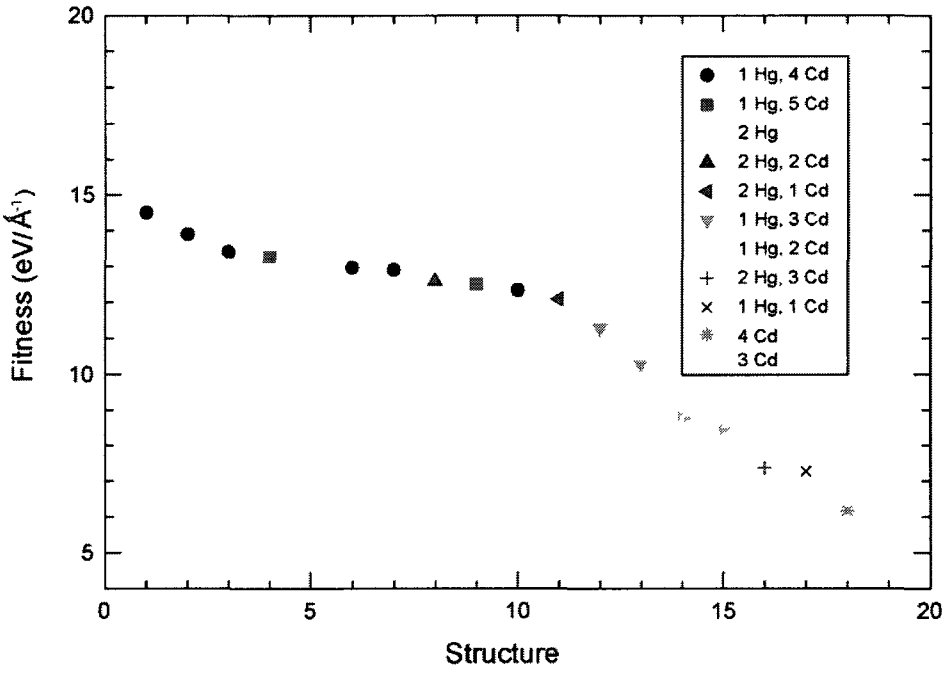


Figure 6.10. Fitness metrics of the most-fit structures recovered from the GA trials optimizing the dispersive behavior of the conduction band in Hg-/Cd- doped ZnO materials.

The results again reinforce those obtained from the GA trials run on Hg-/Cd-/Ba-/Sr-/Ca-/Mg- doped ZnO simulation cells, converging once more to the (1Hg, 4Cd) structure as the most fit structure. As expected, the dispersive behavior of the conduction bands increases with dopant concentration until the band gap collapses at the DFT level of theory. The most fit 12.5% mol Hg (2 Hg) structure from the Hg-doped trials, notably absent from the Hg-/Cd-/Ba-/Sr-/Ca-/Mg- doped trials, was recovered in the third GA trial, the trial with a 50-gene population size. It was noted that it was not recovered in the smaller trials however, wherein the structures promoted between generations became overrun by (1 Hg, 4 Cd) and (1 Hg, 5 Cd) doped structures. In these cases, the (2 Hg) structure was simply never recovered by any of the mating/mutation operations after the (1 Hg, 4 Cd) and (1 Hg, 5 Cd) structures established their elitism, due to the sheer

magnitude of the mutation needed to generate such a (2 Hg) structure from the Cd-influenced genepool. In the large trial however, it was recovered in the early stages of the trial alongside the (1 Hg, 4 Cd) structures, specifically from a mutation operation on a (1 Hg, 1Cd) structure, and persevered throughout the trial's execution.

All three trials converged on the (1 Hg, 4 Cd) structures as the most fit Hg-/Cd-doped ZnO material. Additionally, as Figure 6.10 shows, a number of distinct (1 Hg, 4 Cd) distributions were recovered among the most elite genes, the most fit gene registers a fitness value ~1.4 $eV/Å^{-1}$ higher than the (2 Hg) structure. The main difference between these trials and the Hg-doped ZnO GA trials is reflected in the number of dopant atoms that can be introduced before the band gap collapses. Only 2 or 3 Hg atoms could be introduced in Hg-doped ZnO but, as Figure 6.10 shows, up to 6 dopant atoms can be introduced to Hg-/Cd- doped ZnO, perhaps hinting once more the value of Cd in these simulations is its less-pronounced effect towards closing the band gap. Overall on an atom-per-atom basis, the (2 Hg) structure outshines the (2 Cd), or even the (4 Cd), structure, but, collectively, the two dopant atoms can promote higher dispersion of the conduction band at the Γ-point (at the DFT level of theory).

Finally, a stand-alone GA trial on Hg-/Cd-/Ba-/Sr-/Ca-/Mg- doped ZnO materials constructed from 72-atom 3x3x2 ZnO simulation cells was run to assess whether this system would also converge to Hg-/Cd- doped materials. As shown in Figure 6.10, the promoted structures all contained the heaviest Hg and Cd concentrations even after one generation, reinforcing the trends seen from GA trials on 2x2x2 ZnO simulation cells.

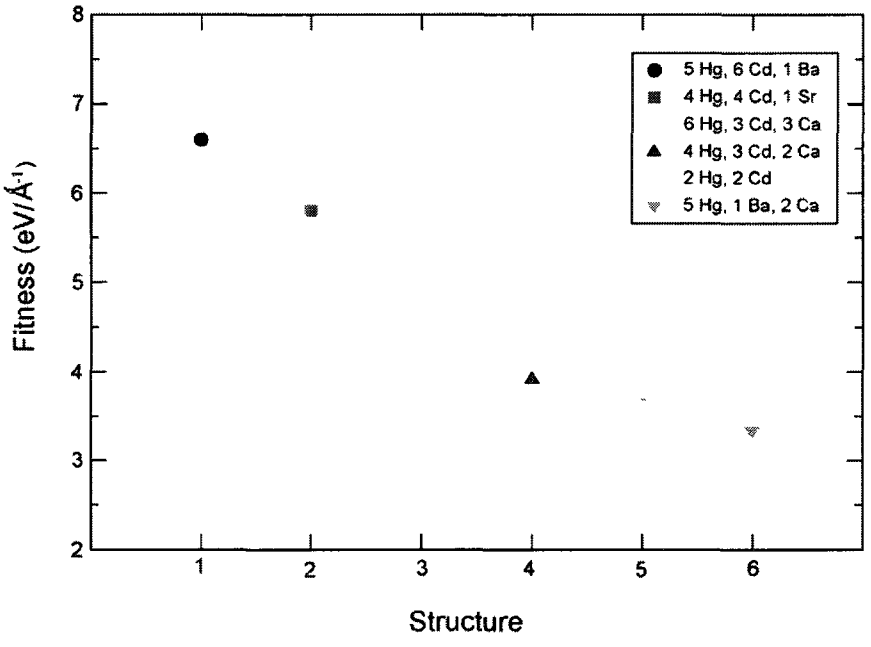


Figure 6.11. Fitness metrics of the most-fit structures recovered from the 1st generation of a GA trial optimizing the dispersive behavior of the conduction band in Hg-/Cd-/Ba-/Sr-/Ca-/Mg- doped, 72-atom 3x3x2 ZnO supercell materials.

Overall, the inverse optimization GA protocols presented in this section were shown to function remarkably well towards optimizing the dispersive behavior of the conduction band in ZnO and SnO_2 materials. It is noted that an in-depth discussion, and interpretation, of the physical significance of the trends discussed in this section would likely require the implementation, or at least consideration, of DFT+U methods[15-17] in order to properly account for the self-correlation electronic effects of the electrons possessing high angular-momentum. These corrections are well-known to account for the shortcoming of DFT to consistently underestimate band gaps, which would be a crucial element for the applications described here, but come with a price as they are typically twice as computationally expensive as conventional DFT calculations. Note that a cruder approach to correcting DFT band gaps, which is still used in the recent literature, is the scissors approach,[8] which simply shifts the unoccupied bands of a semi-

conductor up by a fixed energy constant. Such an approach has been used on doped SnO_2 to discuss its gas-sensing behaviors[8] and would fall into line with the work presented here. Furthermore, the results presented here lay a solid foundation for optimizing electronic effects in metal-oxides that benefit from dispersive behavior in its band structure and, as supported in the recent literature, could have a number of applications in designing or improving upon metal-oxide based technologies.

6.4 – Using an Inverse GA to Optimize Ionic Conductivity in 6.6% LDC materials: Proof of Principle

To further demonstrate how inverse GA procedures can be applied to doped metal oxides and follow up the work presented on concentrated SDC in Chapter 5.4, the inverse GA-inspired structure searches outlined above were run on 10.3% LDC materials in an effort to recover the optimal ionic conductivity trends seen across the lanthanide series. When ceria is doped with a trivalent lanthanide cation, its ionic conductivity increases sharply; it is well known that gadolinium and samarium are the best dopants for singly-doped ceria in this regard.[18] Recent modeling studies have reported that the optimal dopant-vacancy associations in low-energy configurations of LDC shift from next nearest neighbor (NNN) to nearest neighbor (NN) across the lanthanide series;[19, 20] Lanthanum favors configurations which place the dopant NNN to the vacancies while Lutetium favors NN placements with the force field reported by Wei. *et. al.*[19] It was reported in Chapter 5.4 that an increase in the number of NN dopant-vacancy interactions was observed as the concentration, and more importantly the ionic conductivity, of SDC was increased. Its physical consequence was to introduce more balance with respect to the number of NN and NNN dopant-vacancy interactions in the simulation cell. This is a

promising connection since it draws a physical link between SDC and GDC, which are the most promising singly-doped LDC materials. Since this correlation between NN interactions and ionic conductivity seems promising enough to warrant further study, the fitness metric was modified in this "proof of principle" case study to reflect the number of NN interactions in 10.3% LDC materials. The goal is to use a single inverse GA trial to optimize the chemical and configurational space of LDC such that SDC and GDC are recovered from the search.

The specifications of the GA machinery used in this case study were described at length in Section 6.2 and the necessary modifications to the genetic representation, the mating/mutation routines, and the fitness were made directly to this framework. In this light, the code was modified to once more introduce one vacancy into the simulation cell for every two trivalent dopant atoms in order to maintain charge neutrality. Concerning the mating and mutation operations, the "Swap Vacancies" mutation was reactivated and, since the number of dopant atoms was kept constant in this particular case, the "Change Number of Dopants" mutation was deactivated.

The fitness evaluation of each structure was carried out by first submitting a single variable-cell geometry optimization, evaluated at the classical mechanics level of theory, to evaluate each structure's electronic energy. The number of NN dopant-vacancy interactions in each structure was then extracted from the optimized lattice coordinates and the fitness metric was assigned as the absolute value of the difference between the number of NN interactions and the number of vacancies in the simulation cell. In the case of 10.3% LDC materials, this metric would then preferably seek out structures with fitness metrics of zero or, equivalently, three NN interactions in the

simulation cell (one per vacancy). Such a setup establishes the aforementioned "balance" across the series, since Lanthanum and Lutetium favor low-energy structures with zero and six dopant-vacancy interactions, respectively, in the 2x2x2 simulation cells used to construct 10.3% LDC models.

Each structure in the population was then divided into groups depending on the identity of the dopant atom used to construct it, for example all of the SDC structures were grouped together and so on. The relative energies of each structure within each LDC group were then used to penalize structures that were high in energy relative to other structures with their particular genetic makeups. To do so, the fitness metrics of the three lowest energy structures in each group were left unchanged while a correction factor totaling 1000 times the relative energy was added to the fitness metric of each remaining structure. Admittedly, the nature of this energy correction is not ideal in the sense that it adds a large discontinuity to the fitness profile, but it is shown below to suffice in this proof-of-principle context. Future work will focus on developing a more intuitive and smooth functional form for the energy correction. Care was taken to ensure all the structures possessing the three lowest, distinct energy values were preserved untouched within each LDC group; if two structures were found to be degenerate in energy, they were collectively considered as only one of the three lowest energies not penalized during the fitness evaluation stage. It is noted that such a practice may seem to encourage the promotion of redundant structures, but such structures are still screened out at a later stage in the algorithm.

The GULP simulation package[21] was used to perform all the classical molecular mechanics calculations. The Buckingham potential was used to model short range

dispersion pair-wise interactions, Ewald's method[22] was used to sum the long-range Coulombic interactions, and the shell model[23] was used to account for the polarizability of the O^{2-} and Ce^{4+} ions. The force field parameters were taken from a recent systematic study of low-energy 3.2% LDC configurations.[19]

A single GA trial, using the modifications detailed above, was first run on 6.6% LDC simulation cells doped with either L = Sm, Gd, Lu, or Y. The inverse GA will seek out LDC materials with low energy structures that have two dopant-vacancy interactions (and, thus, zero fitness metrics). With such a set up, it is expected from the work presented in Chapter 5 and in the recent literature[19, 20, 24] that Gd and Sm should emerge as the preferred dopant atoms. The trial used a population size of 200 genes within each generation and promoted the top 16 structures from one generation to the next. The GA trial was propagated for thirty generations, wherein convergence of the ten most fit structures was achieved for five generations. The fitness metrics of the twelve most fit structures recovered from the procedure are given in Figure 6.12.

SDC and GDC emerged as the only materials with low-energy structures to have fitness metrics less than two, meaning they had 1-3 NN dopant-vacancy interactions. Since SDC and GDC are the most promising singly doped LDC materials identified in ionic conductivity studies,[18, 25] their recovery is indeed a reassuring sign. The appearance of LuDC and YDC structures in the top ten further confirms that the energy-based corrections to the fitness metrics are in proper working order, since it is these corrections that ensure the fitness metrics of only three structures within each LDC material is unpenalized. The 4^{th} and 5^{th} ranked structures, which are both SDC, were found to be

degenerate in energy, thus they were both promoted collectively as one to give four SDC structures in the top ten.

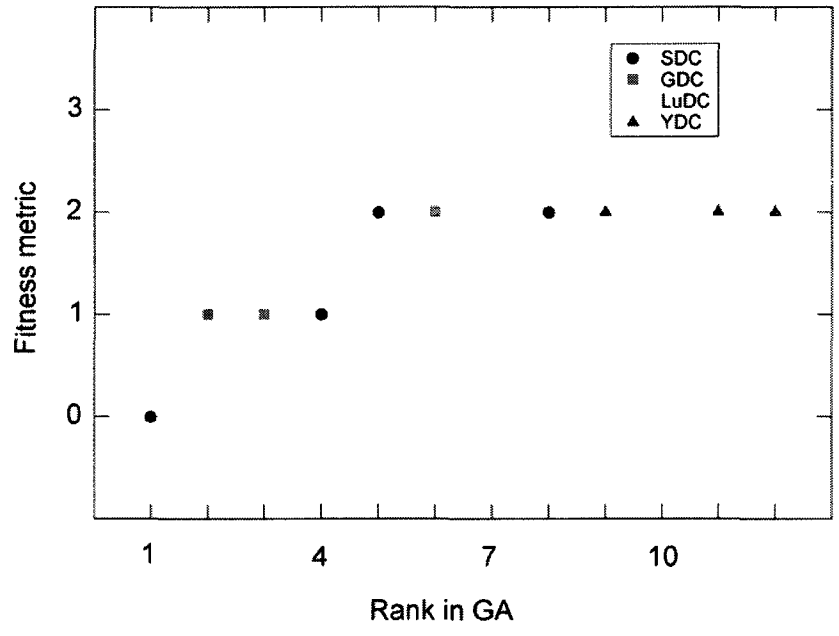


Figure 6.12. Fitness metrics of the most-fit structures recovered a single GA trial optimizing the number of NN dopant-vacancy interactions and relative energies in 6.6% LDC (L = Sm, Gd, Lu, Y) materials derived from a 96-atom 2x2x2 CeO_2 supercell. A fitness metric of zero corresponds to one NN dopant-vacancy interaction per vacancy in the simulation cell.

Furthermore, the 8th and 10th ranked structures are the lowest energy SDC and LuDC structures respectively, with SDC having four NNN dopant-vacancy interactions and LuDC having four NN interactions. These low-energy LuDC and SDC structures were confirmed to match the lowest energy structures recovered from traditional GA searches of each material's individual PES performed in Chapter 5.

A complementary GA trial was run on 6.6% LDC simulation cells to sample a larger library of candidate dopant atoms, namely such that L = Y, La, Nd, Sm, Gd, Er, Yb, or Lu are explored. The trial used a population size of 400 genes within each generation and promoted the top 24 structures from one generation to the next. The

number of genes selected for the elitism, or promotion, stage was chosen to ensure that all of the energetically unpenalized structures within each possible LDC material were promoted from one generation to the next. It was found from independent GA trials with smaller numbers of promoted genes (not shown) that such a set-up leads to spurious "false positives", ie. the promotion of structures with optimal fitness values but high relative energies. The GA trial was propagated for fifty generations, wherein convergence of the four most fit structures was achieved for ten generations. The fitness metrics of the twenty most fit structures recovered from the procedure are given in Figure 6.13. SDC and GDC once more emerged as the only materials with low-energy structures to have fitness metrics less than two, and the lowest energy structures of each material were recovered as structures 6 and 4, respectively.

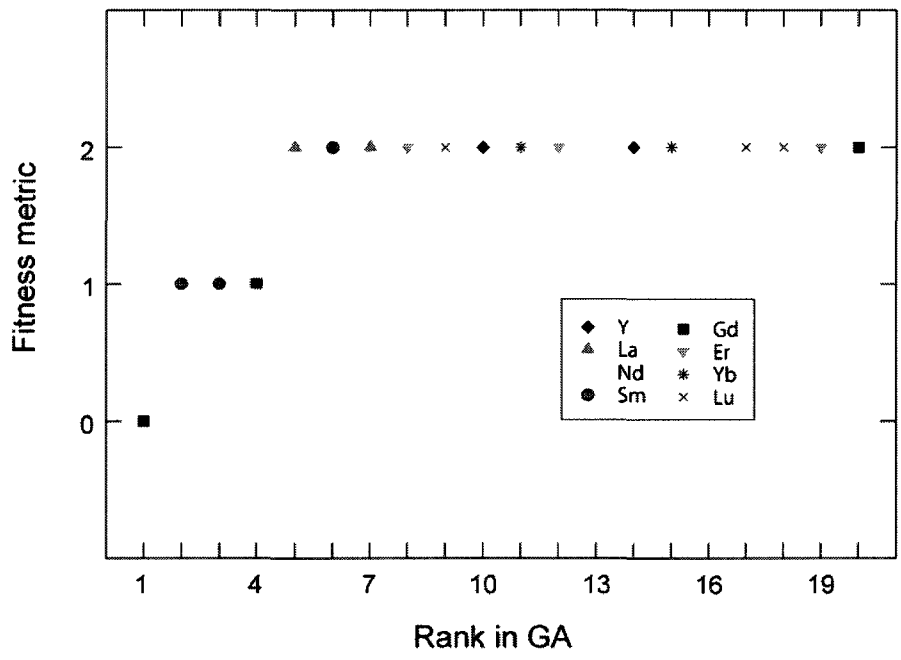


Figure 6.13. Fitness metrics of the most-fit structures recovered a single GA trial optimizing the number of NN dopant-vacancy interactions and relative energies in 6.6% LDC (L = Y, La, Nd, Sm, Gd, Er, Yb, Lu) materials derived from a 96-atom 2x2x2 CeO_2 supercell. A fitness metric of zero corresponds to one NN dopant-vacancy interaction per vacancy in the simulation cell.

Finally, a GA trial was run on 10.3% LDC simulation cells to sample both the larger library of candidate dopant atoms (L = Y, La, Nd, Sm, Gd, Er, Yb, or Lu) and a higher concentration. The trial used a population size of 400 genes within each generation and promoted the top 28 structures from one generation to the next. The GA trial was run for one hundred generations, wherein convergence of the five most fit structures was achieved for ten generations. The fitness metrics of the twenty most fit structures recovered from the trial are given in Figure 6.14. SDC and GDC once more emerged as the only materials with low-energy structures to have fitness metrics less than two. The lowest energy SDC structure was recovered (corresponding to structure 6) but note that, in this particular trial, the lowest energy GDC structure (which was identified from independent conventional GA trials on 10.3% GDC to have 2 NN dopant-vacancy interactions and would therefore be assigned a metric of one) was not recovered.

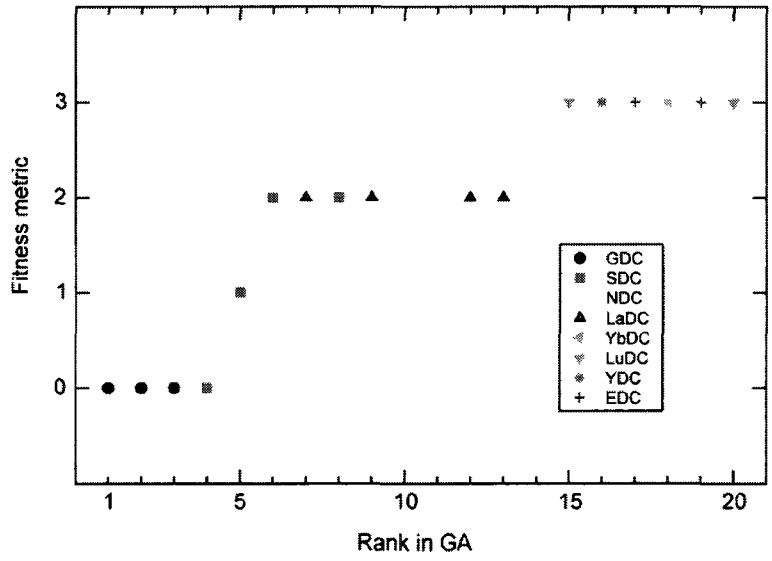


Figure 6.14. Fitness metrics of the most-fit structures recovered a single GA trial optimizing the number of NN dopant-vacancy interactions and relative energies in 10.3% LDC (L = Y, La, Nd, Sm, Gd, Er, Yb, Lu) materials derived from a 96-atom 2x2x2 CeO_2 supercell. A fitness metric of zero corresponds to one NN dopant-vacancy interaction per vacancy in the simulation cell.

In summary, an inverse GA optimization procedure has been shown to reproduce the expected trends seen across the lanthanide series with respect to the number of NN dopant-vacancy interactions in low-energy 10.3% LDC materials. As such, the procedure can further be used in future work to assess variable concentrations and incorporate the other two factors proposed in Chapter 5.4 which were thought to be responsible for SDC's ionic conductivity profile at higher concentrations, namely the spread of defects about the simulation cell and the closest vacancy-vacancy interactions. In this way, the procedure should be able to capture trends in ionic conductivity with respect to both the dopant atom's identity and its concentration. With such a model, co-dopants[18, 26, 27] can then be introduced to the simulation cell and the procedure can be used in a predictive capacity in addition to its current role.

6.5 - Conclusions

Overall, the GA-inspired search methodology presented in this thesis has been shown to readily allow for the study of generic defect association complexes in metal oxide materials. Extending beyond navigating only low-energy configurations, the methodology is shown in this chapter to fruitfully fuse with existing inverse optimization schemes[1-3] to screen large numbers of doped crystal frameworks for specific electronic and/or ionic behaviors.

The application of this method to optimizing dispersive band-structure behaviors of doped metal-oxides can easily be elaborated upon and used to support current research efforts geared toward enhancing gas-sensor technologies which use SnO_2 or ZnO materials. This was demonstrated by using the GA to find optimal dopant arrangements

which increase the dispersion of the native ZnO material's conduction band at the Γ-point at the DFT level of theory; the dispersion of the conduction band has been reported to be proportional to the oxide's electronic mobility.[10] It was observed that 16.7 mol% Hg-doped and 6.25/16.0 mol% Hg-/Cd- doped ZnO materials show the most promise for increasing ZnO's electronic mobility. However, it was also found that 13.9 mol% Hg-doped ZnO would exhibit the most reproducible, optimal electron mobilities since the conductivity was found to be more or less independent of the defect configuration. Moreover, the lowest energy 16.7% Hg-doped ZnO registered a fitness metric almost 3 $eV/Å^{-1}$ lower than the optimal value recovered at that concentration, suggesting care must be taken when synthesizing 16.7% Hg-doped ZnO if one wishes to achieve optimal conductivity. The lowest energy 6.25/16.0 mol% Hg-/Cd- doped ZnO configuration registered a fitness metric within 1 $eV/Å^{-1}$ of its concentration's optimal value.

To further illustrate how these inverse optimization schemes can be incorporated into materials research, they were used in a "proof of principle" concept to optimize ionic conductivity in 6.6% LDC materials. To probe this potential application, the fitness metric was modified to reflect the balance of NN and NNN dopant-vacancy interactions; this built on conclusions drawn from GA searches of SDC concerning its ionic conductivity profile at a range of concentrations (as presented in Chapter 5.4). Considering the notion that, at low temperatures, the mobility of the oxygen ions would be governed by the relative energies of the defect configurations, the relative energy of each structure within each LDC material (eg. LuDC, SDC, GDC, or YDC) was explicitly incorporated into the fitness metric. Using the force field parameters reported by Wei *et.*

al,[19] it was found that the procedure successfully singled out SDC and GDC as the most fit candidate materials, as expected from the literature.[20, 24]

Although there is value in expelling high energy configurations from the algorithm, since there is a limit on how high the energies can be allowed to go before they become irrelevant to experiment, note that this ability to exclude the energies from the fitness metric also gives these GA-inspired search methodologies a unique advantage over dynamical explorations of a given metal oxide system. Such a setup allows a more complete theoretical profile of a given oxide's PES.

Given that the purpose of the calculations presented in this chapter was to lay the groundwork for further studies of doped metal oxides, there are a number of research directions that can be explored further. For example, the inverse protocols used to reproduce the trends in ionic conductivity of LDC seen across the lanthanide series can be used to study co-doped ceria as well, like Nd-/Gd- doped systems. In this light, the scope of the calculations presented here can be extended from its current expositive role to assume a prognostic capacity. Furthermore, the methodology can be used to treat an entire surface slab as the chemical system. Since metal oxide substrates play fundamental roles in heterogeneous catalysis, such studies could readily complement mechanistic studies of surface reactions with the aim of finding an optimal doping arrangement of the slab that promotes its activity.

6.6 – References

1 G. H. Johannesson, T. Bligaard, A. V. Ruban, H. L. Skriver, K. W. Jacobsen, and J. K. Norskov, Phys. Rev. Lett **88**, 255506 (2002).

2 K. Kim, P. A. Graf, and W. B. Jones, Journal of Computational Physics **208**, 735 (2005).

3 S. V. Dudiy and A. Zunger, Phys. Rev. Lett **97**, 046401 (2006).

4 D. Balamurugan, W. Yang, and D. Beratan, J. Chem. Phys. **129**, 174105 (2008).

5 M. E. Franke, T. J. DKoplin, and U. Simon, Nanoparticles in Sensor Technology **2**, 36 (2006).

6 M. Rumyantseva and A. M. Gaskov, Russian Chemical Bulletin **57**, 1106 (2008).

7 C. Li, Z. Yu, S. Fang, S. Wu, Y. Gui, and R. Chen, Journal of Physics: Conference Series **152**, 012033 (2009).

8 Z. Zhu, R. Deka, A. Chutia, R. Sahnoun, H. Tsuboi, M. Koyama, N. Hatakeyama, A. Endou, H. Takaba, C. Del Carpio, M. Kubo, and A. Miyamoto, Journal of Physics and Chemistry of Solids, doi: 10.1016/j.jpcs.2009.07.012 (2009).

9 J. E. Medvedev, Europhysics Letters **78**, 57004 (2007).

10 J. E. Medvedev and A. J. Freeman, Europhysics Letters **69**, 583 (2005).

11 G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).

12 G. Kresse and J. Furthmuller, Comput. Mater. Sci. **6**, 15 (1996).

13 P. E. Blochl, Phys. Rev. B **50**, 17953 (1994).

14 J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett **77**, 3865 (1996).

15 V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).

16 V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **48**, 16929 (1993).

17 J. L. F. Da Silva, M. V. Ganduglia-Pirovano, J. Sauer, V. Bayer, and G. Kresse, Phys. Rev. B **75**, 089901 (2007).

18 F.-Y. Wang, B.-Z. Wan, and S. Cheng, J. Solid State Electrochem **9**, 168 (2005).

19 X. Wei, W. Pan, L. Cheng, and B. Li, Solid State Ion. **180**, 13 (2009).

20 M. Nakayama and M. Martin, PCCP **11**, 3241 (2009).

21 J. D. Gale, J. Chem. Soc. Faraday Trans. **93**, 629 (1997).

22 P. P. Ewald, Ann. Phys. **64**, 523 (1921).

23 J. D. Gale, Phil. Mag. B **73**, 3 (1996).

24 D. A. Andersson, S. I. Simak, N. V. Skorodumova, I. A. Abrikosov, and B. Johansson, PNAS **103**, 3518 (2006).

25 K. Muthukkumaran, Bokalawela, R., Mathews, T., Selladurai, S., J. Mater. Sci. **42**, 7461 (2007).

26 X. Guan, H. Zhou, Z. Liu, Y. Wang, and J. Zhang, Materials Research Bulletin **43**, 1046 (2008).

27 X. Sha, Z. Lu, X. Huang, J. Miao, Z. Ding, X. Xin, and W. Su, Journal of Alloys and Compounds **428**, 59 (2007).

CHAPTER SEVEN

CONCLUSIONS AND OUTLOOK

CHAPTER 7 – *KEY SUBSECTIONS*

7.1 Summary

The work presented in this thesis reinforces the notion of how powerful evolutionary search methodologies are when applied to search out the chemical and/or configurational space of chemical systems. In this work, this is, in part, shown by combining protocols inspired by the genetic algorithm (GA)[1, 2] with previously proposed approaches in materials design[3, 4] and gearing the methodology toward specific applications at the DFT level of theory.

This work first focused on developing effective methodologies that map out nitrogen's high-pressure potential energy surface (PES) such that polymeric nitrogen, a relatively new material with a promising outlook in future high-energy density materials

applications,[5, 6] could be explored. Theoretical and experimental assessments of polymorphism in high-pressure nitrogen have been reported in the recent literature[6-8] and further motivation and developments in polymeric nitrogen research circles is discussed in detail in Chapter 1.2. To summarize, high pressure nitrogen is an ideal candidate to explore theoretically since the harsh conditions in the diamond anvil cell environments and the weak scattering properties of nitrogen make exhaustive and conclusive experimental characterizations beyond the reach of current laboratory technologies.

Given the relative novelty of polymeric nitrogen in the literature, a number of research avenues remained unexplored upon the inception of this thesis; the PES of polymeric nitrogen was still largely unmapped. As such, the work presented in Chapter 2 focused on applying and developing a number of computational modeling techniques to further investigate the high-pressure potential energy landscape of polymeric nitrogen. First, a thorough treatment of computational pressure-application experiments was run on the α-N_2, ε-N_2, and ζ-N_2 solid molecular nitrogen phases; molecular nitrogen solids were chosen as the starting point of these simulations since they are the precursors to polymeric nitrogen. Second, the configurational space of polymeric nitrogen was explored by identifying structural motifs preferably adopted in high-pressure covalent solids and using these motifs to construct new allotropes of polymeric nitrogen or pathways between them. From this work, three novel polymeric nitrogen phases, the polymeric LB[9] and CW[10] structures and the molecular Cmcm-symmetry "aligned" phase,[11] were recovered. The "aligned" phase was further related to the cg-N structure, the lowest enthalpy form of polymeric nitrogen below 200 GPa, through concerted molecular rotations; this connection thereby provided a hypothetical pathway from

molecular nitrogen to its highly sought polymeric allotropes. It was noted that when the aligned phase was compressed, the resulting polymeric framework was best characterized as being amorphous due to its lack of symmetry, but clear cg-N structural motifs were apparent in its makeup.

A systematic method for searching physically relevant single-bonded nitrogen allotropes was then discussed at the end of Chapter 2. The procedure searches out all the possible nitrogen allotropes that can be recovered from Peierls-like distortions of a simple cubic (SC) reference structure;[12] the connection of polymeric nitrogen to a SC framework may seem unorthodox, but, as a matter of fact, other Group V covalent networks have been observed to adopt a SC crystal structure at extreme pressures and the connection is well-discussed in the literature. The cg-N, A7, BP, and LB polymeric structures were all recovered from applying the methodology to an eight-atom SC reference simulation cell. A number of other candidate structures, derived from the 36 possible distortions of the SC reference cell, were recovered and, after extensive stability and thermodynamic tests, two particularly promising and novel nitrogen structures were also presented, labeled bstr8 and bstr11 in this work. The phonon dispersion curves for the bstr11 structure, in particular, suggest it is metastable even at ambient conditions.

From the aforementioned studies on polymeric nitrogen, it was clear that its potential energy landscape was complex and should be explored further using a more reliable search method that is more representative of a brute force systematic approach. In this light, the genetic algorithm (GA) has proven to be an effective computational tool for probing the potential energy surface of new materials.[1] An overview of the motivation, theory, and implementation of the GA into extensive explorations of the PES

pertaining to select chemical systems at high pressures has been discussed in Chapters 1 and 3. In Chapter 3, the GA was used to complement the studies presented in Chapter 2 and further explore high-pressure candidate structures for polymeric nitrogen. For the rest of the studies presented in this thesis, PES searches of select chemical systems were carried out by using Python modules, written as a part of this thesis, to interface evolutionary-inspired (GA-inspired) algorithms with various electronic structure simulation packages, such as VASP. The code was first tested on $CaCO_3$, and SnO_2 chemical systems to show the technology behind the GA algorithm was in proper working order and performed comparably with other state of the art evolutionary search protocols used in solid state chemistry.[1, 13]

Concerning nitrogen, the GA routines were first benchmarked on the high-pressure potential-energy landscape of nitrogen at pressures where polymeric nitrogen is the most thermodynamically stable form of nitrogen.[7, 14] At high pressures, the routines were found to recover all the thermodynamically relevant phases of polymeric nitrogen derived from 8-atom unit cells that had been proposed in the literature at the time. From these searches, a previously unreported phase of polymeric nitrogen with an extended layered framework similar to graphite, aptly named the nitrogen graphite (NG) phase, was recovered as a relatively high-enthalpy phase relative to cg-N at 90 GPa pressures. Its promise, however, was found when its enthalpy was calculated at lower pressures, wherein it becomes ~0.44 eV/atom lower in enthalpy than cg-N at zero pressure and, in addition, becomes marginally lower than the zzCH structure. This made NG the lowest enthalpy structure of polymeric nitrogen at near-ambient pressures reported in the literature at the time.

Finally, the GA routines were used in a predictive capacity at the end of Chapter 3 to map high pressure allotropes of nitrogen/hydrogen solids as potential high energy-density alternatives to polymeric nitrogen. The lowest enthalpy candidate phases at high pressures are best described as H-capped arrangements of the aforementioned ctCH and zzCH phases of polymeric nitrogen (see Chapter 1.2 for further information regarding the ctCH and zzCH phases). To better rationalize the validity of suggesting these structures as relevant materials at high pressure, the GA routines were rerun at low pressures to establish a link between the GA search results and experiment. Encouragingly, the GA routines recovered a crystal structure of ammonium azide and tetrazine which relate well to those reported experimentally.

Building on these efforts, it was noted that polymeric nitrogen is not the only nitrogen allotrope at high pressures that is difficult to characterize experimentally. The phase diagram of polymeric nitrogen's precursor, molecular nitrogen, is also not well understood at high pressures.[8, 15] Although a consensus on the high-pressure α-N_2, β-N_2, δ-N_2, and ε-N_2 molecular solid phases have been reached (the ε-N_2 phase was only recently solved at the turn of the millennium), the structure of cg-N's immediate precursor, the so-called ζ-N_2 phase, is still unresolved along with a number of other molecular solids reported at pressures beyond 50 GPa.

Considering the ambiguity surrounding the ζ-N_2 phase, the evolutionary search procedures developed in this work were used in Chapter 4 to search for structural candidates of the ζ-N_2 molecular phase. This was done by biasing the fitness metric of the GA to efficiently study metastable, molecular solids of nitrogen by adding correction terms to the metric that penalized non-molecular structures.[16] Several candidate structures

for the high-pressure ζ-N_2 phase were proposed; all four of the proposed structures were compatible with one of the two sets of experimental lattice parameters reported in the literature, which have been labeled Cell A and Cell B in this thesis.[8, 15] The discrepancy between the cells was addressed as two distinct molecular phases and, due to the absence of a low-enthalpy *P*mma-symmetry candidate for Cell B,[8] the results further suggest either an erroneous symmetry characterization of Cell B or an inability of DFT under the generalized gradient approximation to describe molecular nitrogen solids at high pressures which deviate substantially from cubic symmetry (which is the distinguishing characteristic of Cell B among the two cells).

The "constrained" GA search technology (biased to search out a particular "class" of crystal structures) that was used to navigate high-pressure structures on the metastable molecular nitrogen landscape was further applied at the end of Chapter 4 to study low-pressure structures of, the now metastable, polymeric nitrogen (< 30 GPa). This was implemented by effectively reversing the bias added to the fitness metric when high-pressure molecular nitrogen was studied. In doing so, the fitness metric was instead biased to effectively search out non-molecular solid structures and successfully recovered the phases expected from the literature, namely the zzCH structure, along with several enthalpically favored candidate structures, notably the NG (which was also recovered from conventional GA trials at high pressures as discussed above) and azoN6 structures.

Concerning the generality of GA applications, researchers in the field often sell the capability of GA search techniques to operate without constraints, but the work on high-pressure nitrogen presented in Chapter 4 hinted that sometimes the use of constraints is actually required in order to account for deficiencies in computational

treatments of the configurational space or improper indexing of the powder x-ray diffraction patterns. Thus, in addition, it is further shown in Chapter 4 that the implementation of symmetry constraints into a GA successfully recovers the correct ε-O_2 structure from symmetry characterizations available to the research community since the mid 1990's; the ε-O_2 phase is a high-pressure, solid, molecular phase of oxygen that serves as a precursor to the higher-pressure metallic ζ-O_2 form of oxygen and,[17] in spite of a number of modeling studies geared specifically toward identifying its structure (with GA studies number among them[1]), the correct ε-O_2 structure was not characterized or even proposed in the literature until a single crystal x-ray diffraction study in 2006.[18] In other words, the correct ε-O_2 structure was recovered from GA-inspired methodologies by biasing the GA to seek out structures with the experimentally proposed *C*2/m symmetry. The symmetry-restricted GA procedures were further tested by searching for and successfully recovering the ζ-O_2,[17, 19] post-aragonite $CaCO_3$,[20] and Dy_2B_4C[21] structures directly from experimentally-proposed lattice information derived from powder x-ray diffraction studies. The symmetry-biased searches not only saved substantial computation time over unrestricted searches, but, again, ultimately proved to be necessary in order to recover the experimentally characterized ε-O_2 phase structure.

Due to the absence of any low-energy *P*mma-symmetry ζ-N_2 structures from the aforementioned GA searches on Cell B of molecular nitrogen, the experimentally-derived symmetry for cell B was tested by running the symmetry-restricted GA trials introduced above. These are trials which were biased to search out configurations consistent with either the $P2_12_12_1$/$P2_12_12$/$P2_122$ symmetry of Cell A or the *P*mma symmetry of Cell B. These trials were successful in recovering the same low-enthalpy candidates recovered

from unrestricted searches on cell A,[15] the first set of lattice parameters, but went one step further for cell B by mapping out a number of structures that were overlooked in the unrestricted search. One of these structures was structure C2,[16] which was proposed in Chapter 4 as the structure most consistent with the experimental characterization of ζ-N_2 that was recovered from the GA trials.

Expanding on the practice of introducing constraints on the GA's search space, the potential benefits and limitations of these GA-inspired search procedures in materials design were probed in Chapters 5 and 6 to promote the technology on a broader and more application-specific research scale. It was shown that such a procedure can be used in materials design to search doped metal-oxide materials for structures with either low association energies or optimal electronic and/or ionic properties at the classical and/or DFT level of theory.[22] The advantage of such practices over other search techniques is the ability of the GA to readily explore high-energy structures that show promise toward a specific application, thereby allowing researchers to envision how to go about seeking optimal performances when synthesizing a particular material.

In Chapter 5, a specialized GA-inspired search procedure for doped metal oxides was presented and tested on lanthanide doped ceria (LDC) materials; the procedure allows the study of generic defect association complexes in metal oxide materials at experimentally-relevant dopant concentrations.[22] The key aspect behind these search techniques was the incorporation of mating and mutation routines that preserve the identity of the parent metal oxide;[3] note that the specific mating operations applied herein are readily transferable to any doped metal oxide framework. Although the GA protocols

were geared specifically toward doped metal oxides, note that such materials have a wide appeal in a number of distinct, and diverse, research areas.

To demonstrate the effectiveness of these GA-inspired search procedures on doped metal oxides, it was shown such structure searches of doped ceria reproduce the results of classical, systematic searches for 3.2% LDC ($[L_2O_3]_1[CeO_2]_{30}$); specifically, the trials were run for samarium-, gadolinium-, and lutetium-doped ceria.[23] Most modeling studies of doped ceria build their simulation cells at these 3.2% concentrations because it is possible to systematically consider all 32 possible arrangements of the defects about the 2x2x2 CeO_2 simulation cell;[23-25] it is not possible to do so at higher concentrations. A number of distinct potentials were tested at 3.2% LDC concentrations and the GA protocols successfully recovered the well-documented crossover across the lanthanide series from NNN to NN dopant site-preference to the vacancy at Gd^{3+}.[23, 26] By using tandem classical and density functional theory energy calculations, the GA-inspired search procedures successfully mapped out the higher (and arguably more relevant, see Chapter 5) 6.6% SDC concentration at the DFT level of theory for the first time. The use of a classical optimization scheme in the early stages of the GA greatly accelerated the search process and proved to be an excellent screen for the more rigorous DFT optimizations that followed.

Specifically, it was found that, at high concentrations of Samarium-doped ceria (SDC) materials (> 3.2% Sm_2O_3), the dopant atoms clearly prefer to occupy the next-nearest neighbor (NNN) site with respect to the vacancies over the NN sites, as expected from the literature. It was interesting to observe this NNN interaction preference in the larger 3x3x3 ceria cells at a 3.8% SDC concentration; the findings were qualitatively

different to those recovered from similar concentrations (3.2%) in 2x2x2 simulation cells, wherein the conclusion drawn from the smaller simulation cells was a preference for NN interactions. In 3.8% SDC, representing the larger 3x3x3 simulation cell, there is a single NN dopant/vacancy association in the six most stable configurations, implying there are multiple low-energy and energetically-competitive configurations of SDC since, by symmetry, there are a number of ways to construct crystal structures with such defect-defect associations. These trends were again seen at the DFT level of theory in 6.6% SDC simulation cells; note that such observations could not be extracted from the 3.2% SDC simulation cells due to the imposed symmetry of the smaller 2x2x2 CeO_2 supercells. It is noted that these energetically-competitive structures would undoubtedly impact the material's oxygen mobility, since their existence provides an abundance of SDC structures with relative association energies near zero and eases the energetic profile of diffusion pathways throughout the lattice. Given the linear dependence of the migration barrier on trivalent lanthanide dopants as one progresses along the lanthanide series, this would then single out SDC, along with GDC, as being promising materials for minimizing the asymmetric thermodynamic profile of migration barriers, as discussed by Wei *et al.*[23] and Nakayama *et al.*[26]

Building on these efforts, the GA-inspired search procedures were shown to function remarkably well even with unprecedented large CeO_2 search spaces. Specifically, they were applied to systems with high Sm concentrations in 3x3x3 simulation cells such that the experimentally-observed optimal concentration of ~11% SDC could be assessed.[27] This involved navigating simulation cells with up to 18 vacancies (and 36 dopant atoms) to reach a peak concentration of ~20.0% SDC.[28] The

observed trends were reminiscent of the NNN-favoring structures recovered from the 6.6% and 3.8% SDC structure searches and, furthermore, the experimentally-expected ordering of the oxygen vacancies in the lowest energy structures was observed beyond ~11% SDC concentrations. In addition, an anisotropy was noted in defect associations at lower SDC concentrations which would help further explain the nature of increased ionic conductivity as the concentration is increased in this regime.

Finally, it was noted that this methodology not only allows general studies of defect association energies, but lays a further foundation to readily be incorporated into proposed "inverse optimization" schemes[4] in order to optimize a particular property of an oxide material. This particular application was discussed at length in Chapter 6, wherein its applications toward optimizing ionic conductivity in doped CeO_2 and electronic mobility in doped ZnO were specifically targeted.

The target of enhancing electron mobility in ZnO, which is used as substrate in gas-sensor technologies, was pursued by applying the GA methodology to optimize dispersive band-structure behaviors of doped metal-oxides.[29] This was demonstrated by using the GA to find optimal dopant atoms and arrangements which increase the dispersion (ie. the curvature) of the native ZnO material's conduction band at the Γ-point at the DFT level of theory; the dispersion of the conduction band has been reported to be proportional to the oxide's electronic mobility.[30] It was observed that 16.7 mol% Hg-doped and 6.25/16.0 mol% Hg-/Cd- doped ZnO materials show the most promise for increasing ZnO's electronic mobility. However, it was also found that 13.9 mol% Hg-doped ZnO would exhibit the most reproducible, optimal electron mobilities since the conductivity was found to be more or less independent of the defect configuration.

Further analysis revealed that Hg and Cd were the only dopant atoms considered which introduced band states with energies near the conduction band minimum, thus explaining their emergence from the algorithm.

To further illustrate how these inverse optimization schemes can be incorporated into materials research, they were used in a "proof of principle" concept to optimize ionic conductivity in 10.3% LDC materials. To probe this potential application, the fitness metric was modified to reflect the balance of NN and NNN dopant-vacancy interactions; this built on conclusions drawn from GA searches of SDC concerning its ionic conductivity profile at a range of concentrations (as presented in Chapter 5.4). Since it can be reasoned that, at low temperatures, the mobility of the oxygen ions would be governed by the relative energies of the defect configurations, the relative energy of each structure within each LDC material (eg. LuDC, SDC, GDC, or YDC) was explicitly incorporated into the fitness metric. Using the force field parameters reported by Wei *et. al*,[23] it was found that the procedure successfully singled out SDC and GDC as the most fit candidate materials, as expected from the literature.[24, 26]

7.2 Impact and Outlook

The work presented in this thesis was conceived from two parallel lines of research: 1) from modeling studies of high-pressure polymeric nitrogen, and 2) from GA studies of specific chemical and/or configurational search spaces. As such, the impact of these branches in their particular fields resulted in: a) the identification of previously unreported allotropes of molecular and non-molecular nitrogen at high pressures which related well with experimental studies, b) the proposal of effective search methodologies capable of reliably navigating a number of chemical spaces ranging from

thermodynamically stable covalent crystals to their metastable counterparts, c) the proposal of similar search methodologies capable of navigating low-energy defect associations in crystal frameworks, d) the identification and discussion of the physical natures of dopant/defect interactions in LDC materials that are responsible for their conductivity profile, and e) a procedure to automate inverse optimization schemes of doped metal oxides geared to enhance their roles in their particular technological applications.

The studies on high-pressure polymeric nitrogen have shown that the potential energy surface of polymeric nitrogen is complex, in the sense that there are a number of metastable structures higher in enthalpy than cg-N at high pressures and, in addition, a number of metastable polymeric structures that become lower in enthalpy than cg-N at low pressures. This suggests that cg-N is not only thermodynamically unstable with respect to molecular nitrogen at ambient pressures, but seems particularly unstable with respect to other polymeric forms as well. This casts even more doubt on whether the cg-N structure, the only crystal to be synthesized in the lab at high pressures, is recoverable at ambient pressures and suggests that future work in this area should follow the example set in the molecular nitrogen community and pursue metastable polymeric materials with high nitrogen content. Preliminary work in this area was presented at the end of Chapter 3, when GA searches were run on high-pressure nitrogen-/hydrogen- based materials. This work could easily be expanded upon to search out other materials at varying nitrogen concentrations. In addition, the inverse GA protocols discussed in Chapter 6 could be introduced to study high-pressure polymeric nitrogen in order to ascertain the effect of dopant atoms on target properties, like the band structure or lattice strain, over

relevant pressure ranges in an attempt to identify the most promising "dopant" species to introduce in a diamond anvil cell environment.

The studies on defect associations in doped ceria also lend themselves to be elaborated upon and used as a foundation in future studies. For example, SDC materials were the focus of most of the work presented in Chapter 5. Knowledge of the atomic positions of SDC provided therein could make possible a computational study of the catalysis of this material. In addition to its role as a solid electrolyte, SDC is also prominently used in solid oxide fuel cells as a substrate to catalyze the oxidation of hydrocarbons, yet their mechanism of action remains unknown. These results thereby provide a basis for investigating this mechanism through appropriately-designed surface models using the insight gained from the bulk. In this light, slabs could either be constructed by building the slab directly from the low energy bulk configurations recovered from the conventional GA trials on the bulk oxide, or taken one step further by using the GA itself to optimize the distribution of vacancies near the surface/vacuum interface of the slab. This would be an interesting direction to take if, for example, one is interested in profiling the distribution of vacancies about the interface over a range of different surface cuts.

In addition, it would be interesting to study other LDC materials, like those doped with Gadolinium or Yttrium, in order to gain physical insight into the role of these other dopant atoms on their respective conductivity profiles. In addition, the foundation was laid in Chapter 5 to perform pseudo-exhaustive searches of LDC's potential energy surface at the DFT+U level of theory; since DFT+U is the standard methodology used in the extensive theoretical literature on pure ceria, it would be of interest to the research

community at large to run more extensive GA's on high LDC concentration models at the DFT+U level of theory to further ascertain the effect of additional defect-defect interactions in the simulation cell.

The inverse GA protocols used to reproduce the trends in ionic conductivity of 10.3% LDC seen across the lanthanide series can be used in future work to study co-doped ceria as well, like those of Gd-/Y- doped ceria. In this light, the scope of the calculations presented here can be extended from its current expositive role to assume a more predictive capacity. The general trend seen in doubly-doped LDC materials is an increase in ionic conductivity;[31] however, certain dopant combinations are known to hinder the ionic conductivity performance over that of its singly-doped counterparts at low temperatures.[32] It stands to reason that inverse and conventional GA studies on doubly-doped ceria could shed light on the underlying physical reasons for the enhanced or diminished performances and, in addition, predict optimal doubly-doped materials. Furthermore, the application of inverse GA's to enhance electron mobility in doped metal oxide substrates, as discussed in Chapter 6, could be elaborated upon by considering other doped metal oxide materials and exploring other levels of theory (beyond DFT).

Finally, the GA methodology can readily be used to treat an entire surface slab as the chemical system. In this role, one could envision a fitness metric that somehow measures the influence of a guest molecule on some inherent property when the guest is adsorbed onto the metal oxide surface. Since metal oxide substrates play fundamental roles in heterogeneous catalysis, such studies could readily complement mechanistic studies of surface reactions with the aim of finding an optimal doping arrangement of the slab that promotes its activity. When coupled with more "standard" inverse optimization

schemes on the parent bulk metal-oxide, such a practice could deliver a complete theoretical profile, both explanatory and predictive, of an entire catalytic process.

7.3 Thesis-Related Publications

Accepted or Submitted Publications

Zahariev, F.; Hu, A.; **Hooper, J.;** Zhang, F.; Woo, T.K. "New layered single-bonded non-molecular phase of nitrogen from first principles simulation" *Physical Review B*, **2005**, *72*, 214108 (1-6).

Zahariev, F.; Hu, A.; **Hooper, J.**; Zhang, F.; Woo, T.K. *Ab-initio* based simulations of high-pressure phases of nitrogen" AIP Conference Proceedings (2006), 845(Pt. 1, Shock Compression of Condensed Matter--2005, Part 1), 373-376

Zahariev, F.; Dudiy, S.V.; **Hooper, J.**; Zhang, F.; Woo, T.K. "Systematic Method to New Phases of Polymeric Nitrogen Under High-pressure" *Physical Review Letters*, **2006**, *97*, 155503/1-4

Zahariev, F.; **Hooper, J.**; Alavi, S.; Zhang, F.; Woo, T.K. "New allotrope of single-bonded nitrogen by first-principle computational experiments on one-dimensional helical chains" proceedings of the 2006 13th International Detonation Symposium, (Norfolk, VA, July 23-28)

Zahariev, F.; **Hooper, J.**; Alavi, S.; Zhang, F.; Woo, T.K. "New low-pressure metastable phase of single-bonded polymeric nitrogen from a helical structure motif and first-principles calculations" Physical Review B, **2007**, *75*, 140101. (rapid communication)

Hooper, J.; Hu, A., Zhang, F.; Woo, T.K. "A Genetic Algorithm and First Principles DFT Study of the High Pressure Molecular Nitrogen" *Physical Review B*, **2009**, *80*, 104117/1-8.

Hooper, J.; Ismail, A.; Giorgi, J. B.; Woo, T.K. "A Genetic Algorithm Based Approach to Investigate Doped Metal-Oxide Materials: Application to Lanthanide-Doped Ceria" *Physical Review B*, **2010**, *81*, 224104 (1-11).

Hooper, J.; Ismail, A.; Giorgi, J. B.; Woo, T.K. "Computational Insights into the Nature of Increased Ionic Conductivity in Concentrated Samarium-Doped Ceria: A Genetic Algorithm Study" *Physical Chemistry Chemical Physics*, accepted.

Hooper, J.; Woo, T.K. "On the Merits of Introducing Symmetry Constraints to a Genetic Algorithm Based Structure Search: Interweaving Theory and Experiment", submitted.

Hooper, J.; Hu, A.; Campana, C.; Woo, T.K. "A Genetic Algorithm Based Exploration of Low Pressure Polymeric Nitrogen Structures", submitted.

Ismail, A.; **Hooper, J.;** Giorgi, J.B.; Woo, T.K. "A Genetic Algorithm and First-Principles DFT+U Study of Samarium-Doped Ceria", submitted.

Downie, L. E..; Smith, J. S.; Desgrenniers, S.; Savard, D.; Murugesu, M.; **Hooper, J.;** Woo, T.K.; Hu, A.; Zhang, F. "A High Pressure Study of Ammonium Azide up to 65 GPa", submitted.

Manuscripts in Preparation

Hooper, J.; Woo, T.K. "Probing the Effects of Local Chemical Environments of Singly- and Doubly- Doped Ceria on Ionic Conductivity: A Genetic Algorithm Study", in preparation.

Hooper, J.; Campana, C.; Woo, T.K. "A Genetic-Algorithm Based Inverse Optimization Scheme to Tune Electronic Properties of Doped ZnO Materials", in preparation.

7.4 References

1 A. R. Oganov and C. W. Glass, J. Chem. Phys. **124**, 244704 (2006).

2 D. M. Deaven and K. M. Ho, Phys. Rev. Lett **75**, 288 (1995).

3 G. H. Johannesson, T. Bligaard, A. V. Ruban, H. L. Skriver, K. W. Jacobsen, and J. K. Norskov, Phys. Rev. Lett. **88**, 255506 (2002).

4 K. Kim, P. A. Graf, and W. B. Jones, Journal of Computational Physics **208**, 735 (2005).

5 C. Mailhiot, L. H. Yang, and A. K. McMahan, Phys. Rev. B **46**, 14419 (1992).

6 J. Kotakoski and K. Albe, Phys. Rev. B **77**, 144109 (2008).

7 Y. Ma, A. R. Oganov, Z. Li, Y. Xie, and J. Kotakoski, Phys. Rev. Lett **102**, 065501 (2009).

8 E. Gregoryanz, A. F. Goncharov, C. Sanloup, M. Somayazulu, H.-K. Mao, and R. J. Hemley, J. Chem. Phys. **126**, 184505 (2007).

9 F. Zahariev, A. Hu, J. Hooper, F. Zhang, and T. K. Woo, Phys. Rev. B **72**, 214108 (2005).

10 F. Zahariev, J. Hooper, S. Alavi, F. Zhang, and T. K. Woo, Phys. Rev. B **75**, 140101 (2007).

11 R. Caracas and R. J. Hemley, Chem. Phys. Lett **442**, 65 (2007).

12 F. Zahariev, S. Dudiy, J. Hooper, F. Zhang, and T. K. Woo, Phys. Rev. Lett. **97**, 155503 (2006).

13 A. R. Oganov, C. W. Glass, and S. Ono, Earth Planet.Sci.Lett. **241**, 95 (2006).

14 Y. Yao, J. S. Tse, and K. Tanaka, Phys. Rev. B **77**, 052103 (2008).

15 M. I. Eremets, A. G. Gavriliuk, N. R. Serebryanaya, I. A. Trojan, D. A. Dzivenko, R. Boehler, H. K. Mao, and R. J. Hemley, J. Chem. Phys. **121**, 11296 (2004).

16 J. Hooper, A. Hu, F. Zhang, and T. K. Woo, Phys. Rev. B **80**, 104117 (2009).

17 G. Weck, S. Desgrenniers, P. Loubeyre, and M. Mezouar, Phys. Rev. Lett **102**, 255503 (2009).

18 L. F. Lundegaard, G. Weck, M. I. McMahan, S. Desgrenniers, and P. Loubeyre, Nature **443**, 201 (2006).

19 Y. Ma, A. R. Oganov, and C. W. Glass, Phys. Rev. B **76**, 064101 (2007).

20 S. Ono, T. Kikegawa, Y. Ohishi, and J. Tsuchiya, American Mineralogist **90**, 667 (2005).

21 V. Babizhetskyy, C. Zheng, H. Mattausch, and A. Simon, Journal of Solid State Chemistry **180**, 3515 (2007).

22 J. Hooper, A. Ismail, J. B. Giorgi, and T. K. Woo, Phys. Rev. B **81**, 224104 (2010).

23 X. Wei, W. Pan, L. Cheng, and B. Li, Solid State Ion. **180**, 13 (2009).

[24] D. A. Andersson, S. I. Simak, N. V. Skorodumova, I. A. Abrikosov, and B. Johansson, PNAS **103**, 3518 (2006).
[25] P. P. Dholabhai, J. B. Adams, P. Crozier, and R. Sharma, PCCP **12**, 7904 (2010).
[26] M. Nakayama and M. Martin, PCCP **11**, 3241 (2009).
[27] F.-Y. Wang, B.-Z. Wan, and S. Cheng, J. Solid State Electrochem **9**, 168 (2005).
[28] J. Hooper, A. Ismail, J. B. Giorgi, and T. K. Woo, Phys. Chem. Chem. Phys., in revision (2010).
[29] J. E. Medvedev, Europhysics Letters **78**, 57004 (2007).
[30] J. E. Medvedev and A. J. Freeman, Europhysics Letters **69**, 583 (2005).
[31] X. Guan, H. Zhou, Z. Liu, Y. Wang, and J. Zhang, Materials Research Bulletin **43**, 1046 (2008).
[32] X. Sha, Z. Lu, X. Huang, J. Miao, Z. Ding, X. Xin, and W. Su, Journal of Alloys and Compounds **428**, 59 (2007).