Computational Investigations into Astrochemical Inorganic Oxides, Ammonia Borane, and Genetic Algorithms

E. Michael Valencia

COMPUTATIONAL INVESTIGATIONS INTO ASTROCHEMICAL INORGANIC
OXIDES, AMMONIA BORANE, AND GENETIC ALGORITHMS

by

E. Michael Valencia

A thesis submitted to the faculty of the University of Mississippi in partial
fulfillment of the requirements of the Sally McDonnell Barksdale Honors College.

Oxford

May 2022

Approved by

_____
Advisor: Dr. Ryan Fortenberry

_____
Reader: Dr. Eden Tanner

_____
Reader: Dr. Pankaj Pandey

Acknowledgements

This work would not have been possible without the relentless support of family throughout my entire life, especially my parents, siblings, and grandparents. I thank them greatly for everything they have enabled me to accomplish and for teaching me important lessons even when I didn't want to hear them. For their financial, emotional and time support, I am greatly indebted. I hope that this work serves as a testament to their sacrifices.

I also gladly acknowledge the support of my friends, Rachel, Austin, Tom and Garrett, who have been with me through some of the most difficult and also the most enjoyable periods in my life. I hope that our friendships will continue long into the future.

I would also like to thank my many teachers throughout the years, whether formally or informally instructing me. To my teachers in my youth, especially Mr. Harden, who instilled in me a love for science and chemistry, I owe very much. Of course, my foremost guide through the forest of chemistry (just as Virgil leads Dante through the woods) has been Dr. Fortenberry. With aptitude, precision and implacable enthusiasm, he has unerringly kept me on the straight path for research and school, and I will forever owe him for accepting me into his group four years ago.

Abstract

The formulations of quantum mechanics in the early 1900s were exciting theoretical discoveries, but were not practical to apply until the advent of computers and the subsequent computational methods in 1951. With the introduction of tractable simplifications, procedures such as Hartree-Fock allowed for determination of properties of non-trivial systems. Presently, huge leads of computational power have allowed for extremely precise, quantitative work that can be applied to the human body, synthesis, or even astrochemical processes. This thesis presents works concerning 1) the history of quantum mechanics; 2) a brief primer on computational chemistry and its methods; 3) inorganic oxides in space that may be important for planetary formation; 4) an elusive molecule computed via a new method; and finally 5) a new method for determining the spectroscopic data of molecules from experimental data utilizing the principles of Darwin.

Contents

Chapter 1

Quantum Mechanics

> Nobody understands quantum
> mechanics.
>
> ―――――――――――――――――――――
>
> *Richard Feynman*

1.1 The Ultraviolet Catastrophe

Before the advent of quantum mechanics, Lord Rayleigh derived an approximation to the energy density of electromagnetic (EM) radiation from a black body in 1900. In its original form, Rayleigh had derived the equation from the equipartition principles of Boltzmann and Maxwell [41].

$$u(\lambda, T) = \frac{c_1 T}{\lambda^4} e^{-\frac{c_2}{\lambda T}} \tag{1.1}$$

In Equation 1.1, c_1 and c_2 are proposed fitting constants which Rayleigh believed could be determined experimentally [41]. The famous Rayleigh-Jeans law would not be proposed for 5 years, in 1905 [25]. Regardless, such a law predicted that as the radiated energy approached the ultraviolet region, the energy emitted would trend toward infinity. Clearly, since the Sun was not outputting infinite energy, and can be

closely approximated to be a black body [112], the proposed law was flawed at higher energy. Rayleigh had not yet considered the revolutionary idea that there was some fundamental proportion that would resolve this problem.

Just after the publication of Rayleigh's work, Max Planck took inspiration from both his teacher, Gustav Kirchoff, as well as Wilheim Wein and Lord Rayleigh, to formulate the now-familiar Planck's Law [25] for the spectral radiance of blackbodies, though Planck did not believe in atoms and was forced to use Boltzmann's atomic theory of entropy for his equation to work [102].

$$B_\lambda(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/(\lambda k_B T)} - 1} \tag{1.2}$$

This resolved the ultraviolet catastrophe. Though five years after Planck's Law, Lord Rayleigh and James Jeans classically determined the Raleigh-Jeans law for the energy density of blackbody radiation.

$$u(\nu, T) = \frac{8\pi k_B T \nu^2}{c^3} \tag{1.3}$$

Boltzmann then, in 1901, defined a new constant of proportionality between the oscillation frequency of his imaginary, proposed oscillator and its energy, since he could not find a physical source of his law without conceding to Boltzmann's atomic theories.

$$\epsilon = h\nu \tag{1.4}$$

Planck, however, did not believe that light was quantized. It was not until the work of Albert Einstein that this concept would propagate [120]. The proportionality constant, h, would also later be termed Planck's constant.

1.2 Energy is Quantized

The work of Albert Einstein in 1905 radically shifted the understanding of physics. His work on the photoelectric effect, which won him the Nobel Prize in 1921, posited that light was organized into discrete packets, called photons. Thus, light can have particle-like behavior. If this is true, it would explain why the ejection of electrons from the surface of a metal varied with the frequency of the light, not the intensity of the light [76]. Einstein asserted that a simple linear equation could model the ejection of electrons through the conservation of energy.

$$h\nu = \Phi + T \tag{1.5}$$

Here, the energy of the photon (recall Equation 1.4) is equivalent to the minimum energy needed to eject the electron (Φ) plus the kinetic energy of the photon (T). Thus, the intensity would only increase the number of ejected electrons but not the energy [72].

Thus, Planck's ideas of discrete energy levels were verified, despite his reluctance. What followed next was a rapid progression into the structure of the atom and quantum mechanics. In 1909, Rutherford determined that the nucleus of the atom was both massive and positively charged [42]. Following the Rutherford model of the nucleus, which would collapse by classical mechanics, Niels Bohr proposed his own model, which attempted to explain the spectral lines explained experimentally by the Rydberg Equation [76]. Unfortunately, this model would only work for the hydrogen atom. His model relied on calculating the energy levels and the difference between them was equivalent to the emitted spectral lines, following Einstein's theory of quanitzed light. Thus,

$$E_n = -\frac{Z^2 R_E}{n^2}, \tag{1.6}$$

where Z is the atomic number of the atom, R_E is the Rydberg constant, and n is the nth energy level. And also,

$$hv = E_{upper} - E_{lower} \qquad (1.7)$$

Equation 1.7 describes the energy change from a higher energy level to a lower energy level, which is equivalent to the photon emitted or absorbed.

Due to the failure of the Bohr model, Louis de Broglie suggested that all particles have the same relationships as those of photons, despite having mass. Thus, he derived the de Broglie wavelength [21].

$$\lambda = \frac{h}{mv} = \frac{h}{p} \qquad (1.8)$$

Therefore, all matter has a de Broglie wavelength. However, ordinary matter is massive. Hence, the de Broglie wavelength is much smaller than the object itself, and classical mechanics prevails. In contrast, for small, rapidly moving objects like electrons, the de Broglie wavelength is larger, and therefore quantum mechanics must be used. In other words, the de Broglie wavelength must be smaller than the spatial extent of the object for classical mechanics to prevail.

The next significant discovery was made by Werner Heisenberg in 1927, famously called the Uncertainty Principle. Since electrons have significant wave-like properties, their diffraction patterns do not adhere to classical mechanics. Heisenberg concluded that determining which band an electron will end up on a diffraction pattern is impossible [53]. Heisenberg, considering a hypothetical experiment where position and momentum of an electron are taken with an optical microscope, came to the conclusion that the uncertainty of the momentum multiplied by the uncertainty in the position could be no less than $h/4\pi$ [76]. Thus,

$$\Delta x \Delta p_x \geq \frac{h}{4\pi} \tag{1.9}$$

1.3 Schrödinger's Equations

Just as the definite future state of a classical system can be predicted with changes in time (kinematics), the *probable* future of a quantum system can be determined with changes in time (recall Equation 1.9). To do so, the wavefunction Ψ, is said to contain all the information about the system. Erwin Schrödinger determined an equation to describe a one-particle, one-dimensional system, which is known as the Time-Dependent Schrödinger Equation.

$$-\frac{\hbar}{i}\frac{\partial \Psi(x,t)}{\partial t} = -\frac{\hbar}{2m}\frac{\partial^2 \Psi(x,t)}{\partial^2} + V(x,t)\Psi(x,t) \tag{1.10}$$

In Equation 1.10, $\hbar$ is the reduced Planck's constant, i is the imaginary unit, Ψ is the wavefunction of the system, m is the mass of the particle, and $V(x,t)$ is the potential energy function of the system [76]. Despite this formulation, Schrödinger struggled with the interpretation of Ψ, believing that it was an intermediate facet of the electric charge density [12]. The interpretation was resolved by Max Born, who posited that the probability of finding the particle at time t was possible.

$$|\Psi(x,t)^2|\ dx \tag{1.11}$$

Another formula, which is much more pliable, can be derived from these conclusions. The Time-Independent Schrödinger equation assumes that the potential energy and the wave function are not dependent on time. The symbol ψ is used for terms that depend only on the coordinate [76], [53].

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x) \tag{1.12}$$

The left hand operator portion of Equation 1.12, is known as the Hamiltonian (energy) operator. Thus, the above equation can be written as an eigenvalue equation.

$$\hat{H}\psi_i = E_i\psi_i, \tag{1.13}$$

Most significantly, Equation 1.13 allows for determination of the energy of a system. Applied with constraints, this equation yields useful approximations such as particle in a box, rigid rotor, and the harmonic oscillator. In braket notation, it can be written as

$$\mathscr{H}|\phi\rangle = \mathscr{E}|\phi\rangle, \tag{1.14}$$

where $\mathscr{H}$ is the Hamiltonian operator for the system. The Hamiltonian is composed of five pieces

$$\mathscr{H} = \hat{T}_n + \hat{T}_e + \hat{U}_{en} + \hat{U}_{ee} + \hat{U}_{nn} \tag{1.15}$$

where the first term corresponds to kinetic energy of each nucleus; the second is the kinetic energy operator for each electron; the third is the potential energy between nuclei and electrons; the fourth is the electron electron repulsion term; and the fifth is the nuclear repulsion energy. Under the ubiquitous Born-Oppenheimer approximation, electrons are moving much more quickly than the nuclei. Thus, the nuclei are essentially fixed in space, and, therefore, the kinetic energy of the nuclei can be neglected. Also, the nuclear-nuclear repulsion is constant. The electronic Hamiltonian is therefore

$$\mathcal{H}_{elec} = \hat{T}_e + \hat{U}_{en} + \hat{U}_{ee} + U_{nn}. \qquad (1.16)$$

This constitutes the main problem of computational chemistry, as the motion of electrons is usually the most important for understanding a system.

Chapter 2

Computational Chemistry

Talk is cheap. Show me the code.

Linus Torvalds

2.1   Introduction

Modern computational chemistry is extremely powerful. Foundationally, it rests upon the quantum mechanical models posed in the early 1900s. The field is a young one, however, as modern computation was not available until 1951, when the first semi-empirical atomic orbital calculations were performed [117]. Now, computational chemistry is capable of modeling complex systems in the body, astrochemical processes in space, and reaction schemes for use in the laboratory. It has also been the source of several Nobel Prizes, notably the 1998 award to Walter Kohn and John Pople [91], as well as the 2013 Nobel Prize to Martin Karplus, Arieh Warshel, and Michael Levitt [92]. Taken altogether, computational chemistry allows the principles of quantum mechanics to be readily applied to everyday life. It, more importantly, has freed chemistry from the confines of conventional laboratory processes. This chapter discusses the implementation of two critical methods in quantum chemistry and its application to spectroscopic data.

2.2   Hartree-Fock (HF) Method

This section was written referencing Szabo and Ostland [123], Levine [76], and lecture notes by Dave Sherill [114] and Ryan Fortenberry [34].

The Hartree-Fock (HF) approximation is extremely critical to computational chemistry as many methods utilize the HF approximation as a starting point for further corrections. It can be summarized as the utilization of a single Slater determinant wave function to perform variational optimization of the energy expectation value under the constraint of spin orthonormality. Essentially, it was developed to solve the electronic Schrödinger equation after the Born-Oppenheimer approximation and allows calculation of useful quantities for a system, like the energy.

Recall that the variational principle states that, for a Hamiltonian H that is time-independent with the lowest-energy eigenvalue of E, if ϕ is a well-behaved, normalized function of the coordinates that satisfy the boundary condition, then the upper bound for the system's ground energy can be calculated [76].

$$\int \phi^* \hat{H} \phi \; d\tau \geq E_1 \tag{2.1}$$

This can be extended for non-normalized ϕ in a useful form, which can be optimized to find the minimum energy of the system.

$$\frac{\int \phi^* \hat{H} \phi d\tau}{\int \phi^* \phi \; d\tau} \geq E_1 \tag{2.2}$$

Where the * notation indicates the complex conjugate of ϕ, and ϕ can be any wavefunction, not necessarily a normalized one. The significance is that the method will always overestimate the energy of the system. Since exactly one electron systems are known, a simple approximation may be to assume that there is no electron repulsion. Beginning with a guess wave function, the spatial motion of more than one electron could be described with the Hartree Product.

$$\psi_{HP}(r_1, r_2, \ldots, r_N) = \phi_1(r_1)\phi_2(r_2) \cdots \phi_N(r_N) \tag{2.3}$$

Unfortunately, this does not satisfy the necessary antisymmetric principle for fermions. However, this can be resolved through the use of Slater determinants, which are a product of spin orbitals (themselves a product of spatial orbitals and either α or β spin). Thus, a new wavefunction, following the antisymmetric constraint can be constructed. For two electrons, it would have the following form.

$$\psi_{x_1, x_2} = -\frac{1}{\sqrt{2}}[\chi_1(x_1)\chi_2(x_2) - \chi_1(x_2)\chi_2(x_1)] \tag{2.4}$$

Where χ_i represents spin orbital i and x_n represents the electron coordinate of electron n. This can further be generalized to any number of electrons electrons. Since a determinant can always be created using the generalized form, a shorthand is often employed using the bra-ket notation.

$$= |\chi_i \chi_j \cdots \chi_k\rangle \tag{2.5}$$

With this in mind, the Hartree-Fock energy is given by the variational principle.

$$E_{HF} = \langle \Psi | \hat{H}_{el} + V_{NN} | \Psi \rangle \tag{2.6}$$

Since V_{NN} does not depend on the electronic coordinate, it can be approximated as a constant value. The $\hat{H}_{el}$ can be divided into a one-electron operator and a two electron operator. The one electron operator consists of its attraction to the nucleus.

$$h(1) = -\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{r_{1A}} \tag{2.7}$$

Furthermore, the two electron operator can be simply defined as:

$$v(i,j) = \frac{1}{r_{ij}} \tag{2.8}$$

Thus, the $\hat{H}_{el}$ operator can be redefined in terms of Equation 2.7 and Equation 2.8.

$$\hat{H}_{el} = \sum_i h(i) + \sum_{i<j} v(i,j) \tag{2.9}$$

Given the single determinant, as in equation 2.5, the energy of Equation 2.6 is a functional of the spin orbitals. Through advanced methods, Equation 2.6 can be rewritten in terms of the integrals of the one and two electron operators.

$$E_{HF} = \sum_i \langle i|h|i \rangle + \frac{1}{2} \sum_{ij} \langle ii|jj \rangle - \langle ij|ji \rangle \tag{2.10}$$

Here,

$$\langle i|h|i \rangle = \int dx_1 \chi_1^*(x_1) h(r_1) \chi_j(x_1), \, and \tag{2.11}$$

$$\langle ij|kl \rangle = \int dx_1 dx_2 \chi_1^*(x_1) \chi_j(x_1) \frac{1}{r_{12}} \chi_k^*(x_2) \chi_l(x_2). \tag{2.12}$$

Applying Lagrange's method of undetermined multipliers to Equation 2.10 allows for an equation that defines the best spin orbitals.

$$h(x_1)\chi_i(x_2) + \sum_{j\neq i} \left[\int dx_2 |\chi_j(x_2)|^2 r_{12}^{-1} \right] \chi_i(x_1) - \sum_{j\neq i} \left[\int dx_2 \chi_j^*(x_2)\chi_i(x_2) r_{12}^{-1} \right] \chi_j(x_1) = \epsilon_i \chi_i(x_1) \tag{2.13}$$

Where ϵ_i is the orbital energy of χ_i.

Thankfully, this equation can be simplified greatly. The first bracketed term, called the Coloumb term, shows that the electron feels the repulsion from the average charge distribution of the other electrons. Thus, the electron density is "smeared"

in the problem space, rather than calculating the exact repulsive energy of all the electrons, which would be problematic due to the N-body problem:

$$\mathscr{J}_j(x_1) = \sum_{j \neq i} \left[\int dx_2 |\chi_j(x_2)|^2 r_{12}^{-1} \right] \chi_i(x_1). \tag{2.14}$$

Similarly, the second bracketed term is named the exchange operator. It has no classical interpretation, but is necessary to preserve the antisymmetric properties:

$$\mathscr{K}_j(x_1)\chi_i(x_1) = \left[\int dx_2 \chi_j^*(x_2)\chi_i(x_2) r_{12}^{-1} \right] \chi_j(x_1). \tag{2.15}$$

A new operator, termed the Fock operator, can also be deduced which will make an eigenvalue equation more clear, given that $j \neq i$.

$$f(x_1) = h(x_1) + \sum_j \mathscr{J}_j(x_1) - \mathscr{K}(x_1) \tag{2.16}$$

Then, Equation 2.13 can be written simply as an eigenvalue equation.

$$f(x_1)\chi_i(x_1) = \epsilon_i \chi_i(x_1) \tag{2.17}$$

However, this equation was not feasible to solve until 1951 ("coincidentally" at the start of computational chemistry) when Roothaan proposed the expansion of the spatial orbitals as linear combinations of the set of one-electron basis functions, $\bar{\chi}_\mu$. This transforms the basis functions into the molecular orbitals. Theoretically, to exactly represent the MOs, an infinite number would be necessary, but practically, a finite number must be used. If enough are used, the error is negligible.

$$\chi_i = \sum_{\mu=1}^{K} C_{\mu i} \bar{\chi}_\mu \tag{2.18}$$

Substitution of Equation 2.18 into Equation 2.17 yields

$$f(x_1) \sum_\nu C_{\nu i} \bar{\chi}_\nu(x_1) = \epsilon \sum_\nu C_{\nu i} \bar{\chi}_\nu(x_1) \tag{2.19}$$

Note that the basis functions have been converted into MOs, denoted by the Greek letters.

Finally, Equation 2.18 can be simplified into a eigenvalue-like equation, where $\mathbf{F}$ represents the Fock Matrix, $\mathbf{S}$ represents the overlap matrix, and $\mathbf{C}$ represents the expansion coefficents.

Two other definitions are critical for performing the entire HF procedure. The first is the density matrix ($\mathbf{P}$), which describes the probability of finding an electron at point r. In terms of molecular orbitals, it can be shown that

$$\mathbf{P} = 2 \sum_a^{N/2} C_{\mu a} C_{\nu a}^*. \tag{2.20}$$

C remains the same expansion coefficients from Equation 2.18. The second piece is the G matrix. The first term in Equation 2.10 will now be defined as H^{core}, or the core Hamiltonian. Since the core Hamiltonian depends only one one electron elements, and the one electron operator (i.e., Equation 2.7) depends on kinetic energy and the nuclear attraction, then

$$H_{\mu\nu}^{core} = T_{\mu\nu} + V_{\mu\nu}^{nucl} \tag{2.21}$$

Inserting the Roothan expansion into Equation 2.10, the definition of H^{core} and the Fock matrix yields the G matrix, which is the two electron integral piece that also depends on the density matrix.

$$F_{\mu\nu} = H_{\mu\nu}^{core} + G_{\mu\nu} \tag{2.22}$$

Altogether, the HF procedure is as follows.

- A set of nuclear coordinates, atomic numbers, and a basis set must be defined.

- The overlap integrals ($S_{\mu\nu}$), $H_{\mu\nu}^{core}$ and two electron integrals must be calculated.

- $\mathbf{S}$ is diagonalized, yielding the transformation matrix $\mathbf{X}$.

- A guess for $\mathbf{P}$ is calculated.

- $\mathbf{G}$ is calculated from $\mathbf{P}$ and the two electron integrals.

- $\mathbf{G}$ is then added to $\mathbf{H}$ to obtain the $\mathbf{F}$ matrix (Equation 2.22).

- A transformed Fock matrix can be calculated using $\mathbf{X}$ according to matrix diagonalization.

- Finally, the new $\mathbf{F}'$ can be diagonalized to determine $\mathbf{C}'$ and ϵ. The fitting coefficents can then be determined. This continues until the density matrix no longer changes.

A working example programmed by the author in Go can be found here: `https://github.com/mvee18/410hf`.

2.3 Couple-Cluster Methods

This section was written referencing Levine [76].

There are four sources of error in *ab initio* calculations. For non-heavy atoms, the most significant sources are a lack of electron correlation and incomplete basis sets. Since the HF method treats electrons as an average field rather than seeing how electrons interact, a new method is necessary.

The coupled-cluster (CC) method was promulgated around 1958 through Coester and Kümmel further developed by Pople and Bartlett in the 1960s and 1970s, among others, as a way to correct for electron correlation. This correlation is defined as

the difference in energy between the true nonrelativistic energy of a system and the Hartree-Fock nonrelativistic energy.

$$E_{corr} = E_{nonrel} - E_{HF} \tag{2.23}$$

The most important equation for CC is

$$\psi = e^{\hat{T}}\phi_0 \tag{2.24}$$

where ψ is the exact nonrelativistic ground-state wavefunction, ϕ_0 is the normalized ground-state reference wavefunction (often the Hartree-Fock wavefunction is selected), and $e^{\hat{T}}$ is a Taylor series expansion:

$$e^{\hat{T}} \equiv \sum_{k=0}^{\infty} \frac{\hat{T}}{k!}. \tag{2.25}$$

Furthermore, $\hat{T}$ is the cluster operator, which is defined as

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \cdots + \hat{T}_n, \tag{2.26}$$

and this extends for n number of electrons.

Two operators can then be defined for one-electron and two-electron excitations:

$$\hat{T}_1\phi_0 = \sum_{a=n+1}^{\infty} \sum_{i=1}^{n} t_i^a \phi_i^a \tag{2.27}$$

$$\hat{T}_2\phi_0 = \sum_{b=a+1}^{\infty} \sum_{a=n=1}^{\infty} \sum_{j=i+1}^{n} \sum_{i=1}^{n-1} t_{ij}^{ab} \phi_{ij}^{ab}. \tag{2.28}$$

The significance of these operators is that they convert the Slater determinants ϕ_0 into a linear combination of singly or doubly excited states. Furthemore, CC theory utilizes virtual orbitals, which allows the electrons to separate and be excited into

these virtual states. Through this process, electron correlation can be calculated. This is done by finding the cluster coefficents (also called amplitudes), i.e., $t_i^a, t_{ij}^{ab}, \ldots$, with more terms added for greater accuracy at a large computational cost.

To illustrate the process, consider the couple-cluster doubles method, which includes only the $\hat{T}_2$ operator. Thus,

$$\psi_{CCD} = e^{\hat{T}_2} \, \phi_0. \tag{2.29}$$

However, the $e^{\hat{T}_2}$ term includes not only the doubly excited states (which are actually calculated), but also approximations to the quadruple, sextuple, and the further $2n$ substutions, which is shown in Equation 2.28. These approximations are derived from products of the coefficients of the double excitation determinants.

Now, the amplitudes can be solved. Equation 2.24 is substituted into the Schrödinger Equation:

$$\mathscr{H} e^{\hat{T}} \phi_0 = E e^{\hat{T}} \phi_0. \tag{2.30}$$

Multiplication by ϕ_0^* and integrating, while keeping in mind the orthogonality spin orbitals, produces:

$$\left\langle \phi_0 \left| \mathscr{H} \right| e^{\hat{T}} \phi_0 \right\rangle = E. \tag{2.31}$$

Eventually, the $\hat{T}$ value is approximated as $\hat{T}_2$, which yields

$$E_{CCD} = \left\langle \phi_0 \left| \mathscr{H} \right| e^{\hat{T}_2} \phi_0 \right\rangle. \tag{2.32}$$

These can be subsituted into an expansion of Equation 2.30 to yield

$$\left\langle \phi_{ij}^{ab} \left| \mathscr{H} \right| e^{\hat{T}_2} \phi_0 \right\rangle = \left\langle \phi_0 \left| \mathscr{H} \right| e^{\hat{T}_2} \phi_0 \right\rangle \left\langle \phi_{ij}^{ab} | e^{\hat{T}_2} \phi_0 \right\rangle. \tag{2.33}$$

Equation 2.33 can be simplified into

$$\left\langle \phi_{ij}^{ab} \left| \mathscr{H} \right| e^{\hat{T}_2} \phi_0 \right\rangle = E_{HF} + \left\langle \phi_0 \left| \mathscr{H} \right| \hat{T}_2 \phi_0 \right\rangle. \tag{2.34}$$

Through application of Condon-Slater rules and the orthogonality of Slater determinants, the resulting integrals involving Slater determinants can be expressed in terms of integrals over the basis functions, which can be solved as a set of nonlinear simultaneous equations, which are solved iteratively. This is expressed as

$$\sum_{s=1}^{m} a_{rs} x_s + \sum_{t=2}^{m} \sum_{s=1}^{t-1} b_{rst} x_s x_t + c_r = 0. \tag{2.35}$$

where the unknown quantities (t_{ij}^{ab}) are x_m; a, b, c are constants derived from orbital energies and electron-repulsion integrals; m is the number of unknown amplitudes. At last, the amplitudes can be solved iteratively. Once known, the wavefunction has been defined in Equation 2.24, and the energy can be determined from Equation 2.32.

The current "gold standard" for computational chemistry integrates the singles and doubles, as well as perturbative triples, termed CCSD(T). The method is incredibly useful due to its relative quickness and high accuracy for small to medium systems [108].

2.4 Quartic Force Fields

Quartic force fields (QFFs) are fourth-order Taylor series approximations of the potential portion of the Watson internuclear Hamiltonian. Before this can be appreciated, a brief discussion of its necessity is in order. Recall the Schrödinger equation (Equation 1.14) and that the $\mathscr{H}$ of concern is often the electronic $\mathscr{H}$. However, the motion of the nuclei (rovibrational motion) may be necessary to model in some cases. Using a similar approximation to the Born-Oppenheimer, a nuclear Hamiltonian can

be constructed. This construction also can be shown as a potential energy surface for the nuclei. However, the original construction of the nuclear Hamiltonian does not account for the external (rovibrational) motion of the molecule. Uncoupling the external motion and the internal motion was first accomplished by Eckart in 1935 by imposing a frame that rotates with the molecule [31]. Though, it was not until 1968 that Watson derived a rovibrational Hamiltonian using the Eckart space [130] that describes all motions of the nuclei, usually termed the Watson Hamiltonian:

$$\mathscr{H}^{Watson} = -\frac{\hbar^2}{2M_{tot}} \sum_{\alpha=1}^{3} \frac{\partial^2}{\partial X_\alpha^2} + \frac{1}{2} \sum_{\alpha,\beta=1}^{3} \mu_{\alpha\beta}(P_\alpha - \Pi_\alpha)(P_\beta - \Pi_\beta) + U - \frac{\hbar^2}{2} \sum_{s=1}^{3N-6} \frac{\partial^2}{\partial q_2^2} + V.$$

(2.36)

Here, the first term is the center of mass; the second is a rigid rotor rotational term; the third potential term is the Watson term; the fourth is the kinetic energy of the vibrations of the nuclei in terms of nuclear displacement; finally, V is the potential energy. The final term (the internuclear potential of the Watson Hamiltonian) is that which the QFF seeks to approximate. The V term can be approximated with a fourth order Taylor series expansion [47].

$$V = \frac{1}{2} \sum_{ij} F_{ij} \Delta_i \Delta_j + \frac{1}{6} \sum_{ijk} F_{ijk} \Delta_i \Delta_j \Delta_k + \frac{1}{24} \sum_{ijkl} F_{ijkl} \Delta_i \Delta_j \Delta_k \Delta_l \qquad (2.37)$$

where F represent the force constants (changes in energy with respect to distance), and the Δ_i terms represent the displacements in that nuclear coordinate. The exact implementation to determine these power terms varies, and they are further described in chapters 3 and 4.

Chapter 3

Enstatite (MgSiO$_3$) and forsterite (Mg$_2$SiO$_4$) monomers and dimers: highly detectable infrared and radioastronomical molecular building blocks

Abstract

Isolated MgSiO$_3$ and Mg$_2$SiO$_4$ molecules are shown here to exhibit bright IR features that fall close to unattributed astronomical lines observed toward objects known to possess crystalline enstatite and forsterite, minerals of the same respective empirical formulae. These molecules are therefore tantaliz-

ing candidates for explaining the origin of such features. Furthermore the C_{2v} monomer minima of each formula set have dipole moments on the order of 10.0 D or larger making them desirable candidates for radioastronomical observation as enabled through rotational spectroscopic data further provided in this high-level CCSD(T)-F12/cc-pVTZ-F12 quantum chemical study. Astrophysical detection of these molecules could inform the build-up pathways for creating nanocrystals from small molecules in protoplanetary disks or could show the opposite in explaining the destruction of enstatite and forsterite minerals in supernovae events or other high-energy stellar processes. This work also shows that the lowest energy isomers for molecules containing the geologically necessary elements Mg and Si have oxygen bonded between any of the other heavier elements making oxygen the glue for pre-mineralogic chemistry.

3.1   Introduction

Since all the atoms in the Universe heavier than lithium were necessarily formed through stellar nucleosynthesis or some type of post-stellar event (i.e. supernovae or neutron star collision), minerals had to originate first as small molecules before becoming nanocrystals and ultimately arriving at rocky structures.

Small molecules comprised of the same atomic ratios as empirical formulae for crystalline minerals have been shown to exhibit large intensities and longer wavelengths in their IR spectra compared to more commonly examined volatile species [71, 93]. These molecules also have strongly varying infrared peak positions producing very tall, isolated, far-IR spectral features. These properties align quite closely with the requirements for portions of known stellar spectra like that from the Butterfly nebula (NGC 6302) where crystalline enstatite and forsterite have already been observed [90]. Hence, the $MgSiO_3$ and Mg_2SiO_4 molecules, monomers for two of the most common minerals, may be detectable in such astrophysical regions and would

24

provide insights into the formation and destruction of dust particles or even rocky bodies.

The most common mineral in the Earth's mantle is olivine [139] with the magnesium-rich forsterite (Mg_2SiO_4) endmember as the largest contributor [65]. Forsterite and its MgO-cleaved enstatite ($MgSiO_3$) cousin are among the most common minerals in rocky bodies within the Solar System [139]. They are also major energy sinks in the quartz-periclase-corundum-water mineral network which drives much of the chemistry and geology in the upper layers of the Earth [110].

These minerals likely form orginally as interstellar dust grains [65, 43, 69, 118, 66, 80, 141] before coalescing into larger structures and crystals within stellar atmospheres and protoplanetary disks as observations and laboratory/theoretical investigations support [90, 104, 22, 46, 74]. These minerals and small molecules containing Mg, Si, and O have also been seen in the ejecta of supernovae [17] and swirling in the atmospheres of both Mercury and Mars [88, 103]. Hence, small molecules of the same empirical formulae as the larger crystalline minerals may help to show how dust forms in planetary nurseries or whether rocky bodies may have been present before a stellar death.

Recent work supports the idea that enstatite mineral crystals initially form amorphously [80], and the infrared signatures of nanocrystals or amorphous clusters of the same number and type of atoms have notable differences enabling their unique detection between the phases in these pseudo-macroscopic particles [141]. More fundamental than any clusters, however, are monomer and dimer molecules of the simple chemical formulae. Previous density functional theory (DFT) computations on $(MgSiO_3)_N$ and $(Mg_2SiO_4)_N$ have provided structural determination as to the minimum energy structures for various N levels guiding how these molecular clusters nucleate and grow [97, 80, 77]. However, *in situ*, observable evidence for these processes must be obtained in order to verify the validity of such theoretical processes,

and rotational (and some vibrational) spectroscopy has long-been the work horse for molecular detections in various astronomical environments [82, 33, 83].

Laboratory gas-phase synthesis and spectral characterization with high levels of certainty for molecules such as the $MgSiO_3$ and Mg_2SiO_4 monomers can be exceeding difficult. Quantum chemistry, on the other hand, begins with the molecular structure as input for the "experiment" and can be utilized to produce unambiguous molecular spectra. Granted, the problem with the latter method is with regards to accuracy. However, state-of-the-art quantum chemical approaches have been able to produce vibrational frequencies to within 1.0 cm^{-1} and rotational constants to within 15 MHz of gas phase experiment in many cases [60, 61, 63, 36, 64, 142, 37, 38, 70, 33, 13]. Recent work has pushed less expensive but more recently-developed quantum chemical methods to providing vibrational frequencies to within 7.0 cm^{-1} of higher-level theory but for orders of magnitude less computational time [62, 6, 4, 5].

This approach has already produced vibrational frequencies for the monomer, dimer, and trimer of magnesium oxide/periclase/MgO with notable accuracies compared to experiment benchmarked for the monomer [71]. Similar work has also been done for the monomer and dimer of magnesium fluoride/sellaite/MgF_2 [93]. Both of these studies show that these relatively small species may have large dipole moments and IR intensities when allowed by symmetry. For instance, the antisymmetric stretch of the MgF_2 monomer has an intensity 2.5 times that of the "bright" antisymmetric stretch in water [93] while the vibration in the MgO monomer is 15 times that of this same water motion [71]. These large dipoles or induced dipoles are brought about simply due to the extreme electronegativity differences between elements on opposite sides of the periodic table. Additionally, most of the frequencies are towards, if not in, the Far-IR due to the higher masses of the atoms involved. Consequently, this work will employ these proven quantum chemical techniques in order to provide IR and millimeter-wave spectral data for $MgSiO_3$ and Mg_2SiO_4 so that these molecules

Figure 3.1: Visual Depiction of the $MgSiO_3$ Isomer.
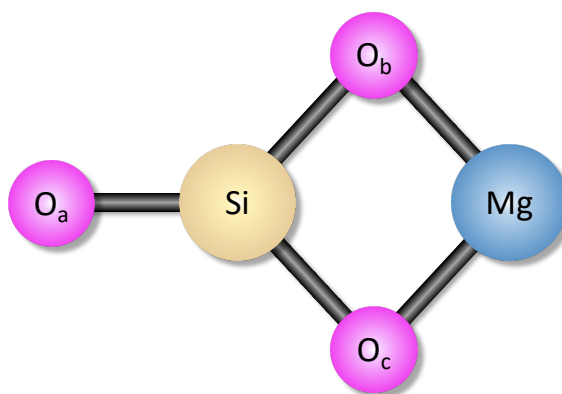
may possibly be observed with the newest generation of observatories, especially the upcoming the *James Webb Space Telescope.*

3.2 Computational Details

All computations make use of the MOLPRO2015.1 quantum chemical program [57, 133, 132]. The only exceptions are the DFT computations and the MP2/6-31+G(d) double-harmonic IR intensities which make use of the Gaussian16 program [89, 55, 40]. The latter intensity computations have been shown to produce semi-quantitative agreement with higher-level theory for heavy reductions in computational time [140, 32].

3.2.1 Anharmonic Spectra

The most accurate fundamental vibrational frequencies and rotational constants computed for the least amount of computational time come from quartic force fields (QFFs), fourth-order Taylor series expansions of the potential portion within the internuclear Watson Hamiltonian [35]. These are computed via geometry optimizations to define the reference geometry and 1613 single-point energies for displacements of 0.005 Å and 0.005 radians of this geometry in terms of the symmetry-internal coor-

dinate system as defined from a lazy-Cartesian algorithm [124]. For $MgSiO_3$, these coordinates are defined from Figure 3.1 as follows:

$$S_1 = \quad r(\text{Si} - \text{Mg}) \tag{3.1}$$

$$S_2 = \quad r(\text{Si} = \text{O}_a) \tag{3.2}$$

$$S_3 = \quad \tfrac{1}{\sqrt{2}}(r(\text{Si} - \text{O}_b) + r(\text{Si} - \text{O}_c)) \tag{3.3}$$

$$S_4 = \quad \tfrac{1}{\sqrt{2}}(\angle(\text{O}_a-\text{Si} - \text{O}_b) + \angle(\text{O}_a-\text{Si} - \text{O}_c)) \tag{3.4}$$

$$S_5 = \quad \tfrac{1}{\sqrt{2}}(r(\text{Si} - \text{O}_b) - r(\text{Si} - \text{O}_c)) \tag{3.5}$$

$$S_6 = \quad \tfrac{1}{\sqrt{2}}(\angle(\text{Mg} - \text{O}_b-\text{Si}) - \angle(-\text{Mg} - \text{O}_c\text{Si})) \tag{3.6}$$

$$S_7 = \quad \tfrac{1}{\sqrt{2}}(\angle(\text{O}_a-\text{Si} - \text{O}_b) - \angle(\text{O}_a-\text{Si} - \text{O}_c)) \tag{3.7}$$

$$S_8 = \quad \text{LIN}(\text{Mg} - \text{Si} - \text{O}_a) \tag{3.8}$$

$$S_9 = \quad \tau(\text{O}_b-\text{Si} - \text{O}_a-\text{O}_c) \tag{3.9}$$

where "LIN" represents a linear bending angle directly out-of-the-plane of the molecule. All QFF computations and the dipole moment determinations make use of coupled cluster theory at the singles, doubles, and perturbative triples level [106] within the F12 explicitly correlated construction [3, 73] along with the F12 triple-zeta basis set [62] referred to as the CCSD(T)-F12/cc-pVTZ-F12 level of theory abbreviated as F12-TZ in this work.

After the F12-TZ energies are computed at each point, the QFF function is fit via a least-squares procedure with a sum of squared residuals on the order of 10^{-13} a.u.^{2} A refit of the surface produces the equilibrium geometry and the final force constants (given in the supplemental information, SI). The force constants are transformed into Cartesian coordinates in the INTDER program [8] so that second-order rotational [86] and vibrational perturbation theory (VPT2) [131, 95] can be utilized within the SPECTRO program [44]. Resonances are included in the rovibrational comptutations

in order to increase accuracy of the anharmonically predicted spectral values [79, 78]. $MgSiO_3$ exhibits $2\nu_6 = 2\nu_7 = \nu_3$ and $\nu_6 + \nu_3 = \nu_8 + \nu_2 = \nu_1$ Fermi resonance polyads; a $2\nu_9 = \nu_6$ type-1 Fermi resonance; and $\nu_8 + \nu_6 = \nu_5$ and $\nu_9 + \nu_7 = \nu_4$ type-2 Fermi resonances. Furthermore, there are five Coriolis resonances: ν_9/ν_8, ν_8/ν_7, and ν_7/ν_5 A-type as well as ν_6/ν_5 and ν_4/ν_2 C-type resonances.

The IR features of the Mg_2SiO_4 structures could not be computed in the same manner due to computational costs even with the savings from F12-TZ. All atoms involved are heavy-atoms, and each additional increase grows the number of points required in the QFF geometrically and the cost of each single-point energy exponentially. However, the anharmonic shifts for such heavy-atoms [32], especially the Mg and Si atoms, are relatively small, typically less than 10 cm^{-1}. The anharmonic correction is therefore within the errors for the F12-TZ frequencies anyway indicating that the harmonic frequencies could be good estimates in-and-of-themselves. In order to further refine the Mg_2SiO_4 frequencies, the scaling factor from the $MgSiO_3$ anharmonic corrections is employed for the larger molecules. All modes of $MgSiO_3$ higher than 500 cm^{-1} had anharmonic shifts of 98.45% on the average while those lower than 500 cm^{-1} had average shifts of 99.42%. Hence, these scaling factors will be applied to the Mg_2SiO_4 isomers where the frequencies fall within these ranges.

3.2.2 Relative Energies

In order to point future work towards the processes of building clusters with these empirical formulae, relative energies are computed between found minima for the monomers and dimers of $MgSiO_3$ and Mg_2SiO_4. The correlation consistent composite approach (ccCA) developed by [23] has been shown to produce coupled cluster level accuracy for a significant reduction in computational time. In brief, B3LYP/cc-pVTZ geometry optimizations [11, 138, 75, 99, 67, 30] are utilized to compute MP2 complete basis set (CBS) extrapolations with cc-pVDZ, cc-pVTZ, and cc-pVQZ bases.

Table 3.1: B3LYP/cc-pVTZ and ccCA Relative Energies (in kcal/mol) for the Dimers of $MgSiO_3$ & Mg_2SiO_4 and Monomers of Mg_2SiO_4. Isomers Are Defined in Figure 3.2.

MgSiO$_3$			Mg$_2$SiO$_4$		
Isomer	B3LYP	ccCA	Isomer	B3LYP	ccCA-TZ[a]
Monomers					
			A (D_{2d})	15.36	12.31 (10.31)
			B (C_s)	61.74	61.11 (66.08)
			C (C_{2v})	0.00	0.00 (0.00)
Dimers					
B	3.67	3.35	B	45.43	51.02
D	25.81	36.14	C2	49.62	53.48
E	42.23	45.98	D	33.02	32.86
F	0.00	0.00	E2	131.75	55.13
G	51.69	66.45	F	73.97	65.36
			G	0.00	0.00
			H	124.59	122.15
			I	0.30	31.94

[a]The monomer values in parentheses are for the full ccCA for comparison.

All computations make use of the tight d, cc-pV($X + d$)Z, basis sets for silicon and magnesium.

Then, corrections for higher-order electron correlation [CCSD(T)/cc-pVTZ − MP2/cc-pVTZ], scalar relativity [26] [MP2/cc-VTZ-DK − MP2/cc-pVTZ], core electron correlation [MP2/aug-cc-pCVTZ − MP2/aug-cc-pVTZ], and scaled B3LYP/cc-pVTZ zero-point vibrational energy (ZPVE) are added to the CBS energies. Some of the MP2/aug-cc-pVQZ single point energies for the Mg_2SiO_4 dimers are also too computationally costly so the CBS terms there are replaced simply with MP2/aug-cc-pVTZ giving the ccCA-TZ notation.

3.3 Results and Discussion

3.3.1 Isomeric Stabilities

The most stable isomers for either the enstatite or forsterite families all have one thing in common: a bonding motif where oxygen separates the third-row atoms, and there are no third-row−third-row or oxygen−oxygen bonds. Such behavior could be

Figure 3.2: The MgSiO$_3$ Dimer, Mg$_2$SiO$_4$ Monomer, and Mg$_2$SiO$_4$ Dimer Isomers (Si=silver; Mg=green; & O=red).
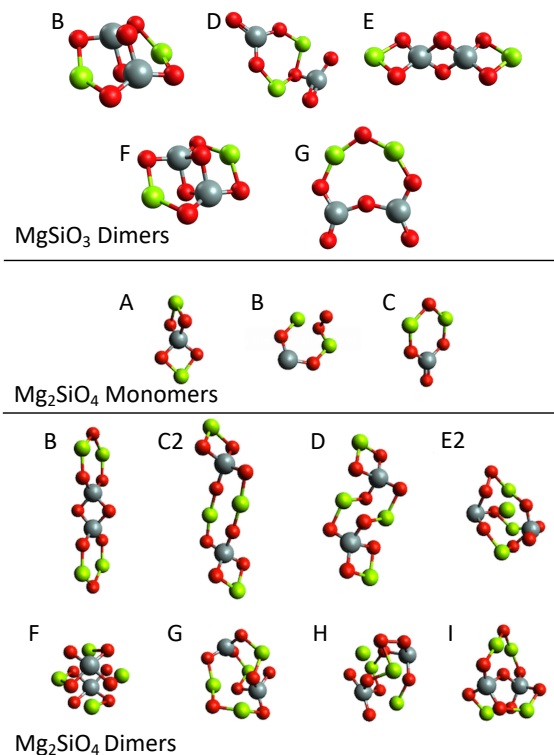
likened to bricks and mortar where the third-row atoms are the former and oxygen is the latter and is likely the result of hard acid-hard base interactions. In other words, oxygen holds the structure together in the most stable configuration. Silicon oxides have been shown to exhibit such behavior [129], but this oxygen bonding motif may be more universal in inorganic and pre-mineralogic chemistry than previously reported.

Previous DFT work on the isomers of the MgSiO$_3$ monomer showed conclusively that the C$_{2v}$ isomer given in Figure 3.1 is the most stable form lying 39.46 kcal/mol below the next-closest isomer which is a ring-opened C$_s$ form [97]. This is corroborated by contemporary DFT studies which agree that this molecule isomer is the most energetically favored for MgSiO$_3$ [77]. Simply attaching a MgO monomer to a quartz (SiO$_2$) monomer produces this isomer in a strongly-favored fashion. Hence, other possible isomers will not be explored presently.

Figure 3.3: Visual Depiction and Atom Labels of the Mg_2SiO_4 lowest-energy C_{2v} Isomer.
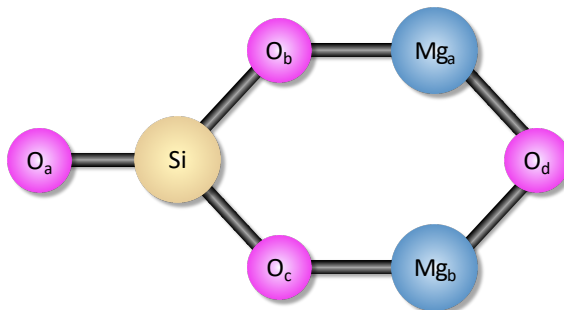
For the dimers of $MgSiO_3$, the minimum energy structure found here for both B3LYP/cc-pVTZ and full ccCA (F from Table 3.1 and Figure 3.2) closely resembles the minimum energy geometry from [77] indicating that both studies are in agreement as to the global minimum. Presently, the dimer is notably more stable than the monomer with 154.22 kcal/mol of energy released upon dimerization. The creation of this dimer can be thought of as two monomers brought together in a slipped-stacked fashion. This work also shows that another low-lying isomer is also present, B from Figure 3.2. This isomer is closely related to the lower-energy F isomer, but the B isomer has the magnesium atoms both pointing in the same direction. In F, they are "trans" or opposite one another. Granted, 3.35 kcal/mol (0.145 eV) from Table 3.1 is an uphill climb, but in regions like supernova remnants or stellar atmospheres where magnesium silicates have already been observed, the conditions of these regions are such that this isomer could be easily accessed in the destruction or formation of larger enstatite clusters or nanocrystals. Additionally, the 4-coordinate silicon in both isomers is like the 4-coordinate silicon for the crystal structure in the larger bulk [54]. The Si=O motif in the monomer is not preferred in the dimer and likely larger clusters. Both results imply a preference for the 4-coordinate silicon over the 3-coordinate. The other three isomers are likely too high in energy to contribute, but the B3LYP/cc-pVTZ and full ccCA results agree semi-quantitatively and produce the same energy ordering of the isomers.

The Mg_2SiO_4 monomer isomers have not been explored previously save for reporting on the lowest energy structure [77]. B3LYP/cc-pVTZ, ccCA, and ccCA-TZ all confirm that the C_{2v} structure (Figures 3.3 and 3.2C) is the energy minimum. However, here a D_{2d} isomer (Figure 3.2A) is shown to lie 10.31 kcal/mol (0.447 eV) above this. While higher in energy, this D_{2d} isomer makes the silicon atom 4-coordinate, again more closely in line with the larger crystal structure and without the exterior Si=O bonding present. Consequently, this isomer is also likely present in regions where the lowest-energy C_{2v} isomer is present, as well. Additionally, the ccCA and ccCA-TZ variants report similar relative energies indicating that when the full CBS computations are not available, cc-pVTZ will suffice for computing semi-quantitative relative energies and isomer energy ordering.

There are no isomers of the Mg_2SiO_4 dimer that energetically compete with the lowest energy isomer (G). While B3LYP indicates that the I isomer is nearly degenerate, ccCA-TZ increases the relative energy notably. The G isomer is a cage structure matching that from [77], again, with all third-row atoms coordinating purely to oxygen atoms. The next-lowest energy isomers, I and D, also follow this pattern and are roughly 32 kcal/mol each above the G. The much-higher E2 and H isomers have O−O bonds present in their structures. In any case, observations of the Mg_2SiO_4 dimer are most likely to be for the G isomer.

3.3.2 Rotational Spectra and Structures of $MgSiO_3$ and Mg_2SiO_4

The $MgSiO_3$ monomer should be easily detectable if it exists in protoplanetary disks or supernova remnants. It has a monstrous 12.78 D dipole moment and is near-prolate. The MP2/6-31+G(d) Mulliken charges have both Mg and Si with more than a single positive charge, and all of the oxygen atoms with nearly a single negative charge each. Consequently, low column-densities of this material should not limit the observation of this molecule. The 100K pure rotational spectrum computed from

Table 3.2: CCSD(T)-F12/cc-pVTZ-F12 Equilibrium and Vibrationally-Averaged Structural Data and Rotational Spectroscopic Constants for $MgSiO_3$ and the three Isotopes of Mg.

	Units	$^{24}MgSiO_3$	$^{25}MgSiO_3$	$^{26}MgSiO_3$
$r_e(Si=O_a)$	Å	1.51964	–	–
$r_e(Si-O_b)$	Å	1.63575	–	–
$r_e(Mg-Si)$	Å	2.42998	–	–
$\angle_e(O_a-Si-O_b)$	°	129.887	–	–
$\angle_e(Mg-O_b-Si)$	°	87.621	–	–
A_e	MHz	10028.4	10028.4	10028.4
B_e	MHz	3005.9	2952.3	2901.8
C_e	MHz	2312.7	2280.8	2250.6
Δ_J	kHz	0.283	0.275	0.268
Δ_K	kHz	8.096	8.154	8.207
Δ_{JK}	kHz	0.793	0.743	0.697
δ_J	kHz	0.074	0.072	0.069
δ_K	kHZ	1.238	1.204	1.173
Φ_J	mHz	0.049	0.047	0.045
Φ_K	mHz	27.024	26.741	26.482
Φ_{JK}	mHz	1.483	1.370	1.269
Φ_{KJ}	mHz	-10.268	-9.829	-9.508
ϕ_j	mHz	0.024	0.023	0.021
ϕ_{jk}	mHz	0.894	0.834	0.780
ϕ_k	mHz	15.289	14.948	14.630
μ	D	12.78	–	–
$r_0(Si=O_a)$	Å	1.52086	1.52084	1.52082
$r_0(Si-O_b)$	Å	1.64025	1.64024	1.64024
$r_0(Mg-Si)$	Å	2.43596	2.43593	2.43591
$\angle_0(O_a-Si-O_b)$	°	129.883	129.882	129.882
$\angle_0(Mg-O_b-Si)$	°	87.588	87.588	87.588
A_0	MHz	9980.4	9980.3	9980.3
B_0	MHz	2996.0	2942.6	2892.3
C_0	MHz	2303.1	2271.5	2241.4

Figure 3.4: The Pure Rotational Spectrum of MgSiO$_3$ at 100K.
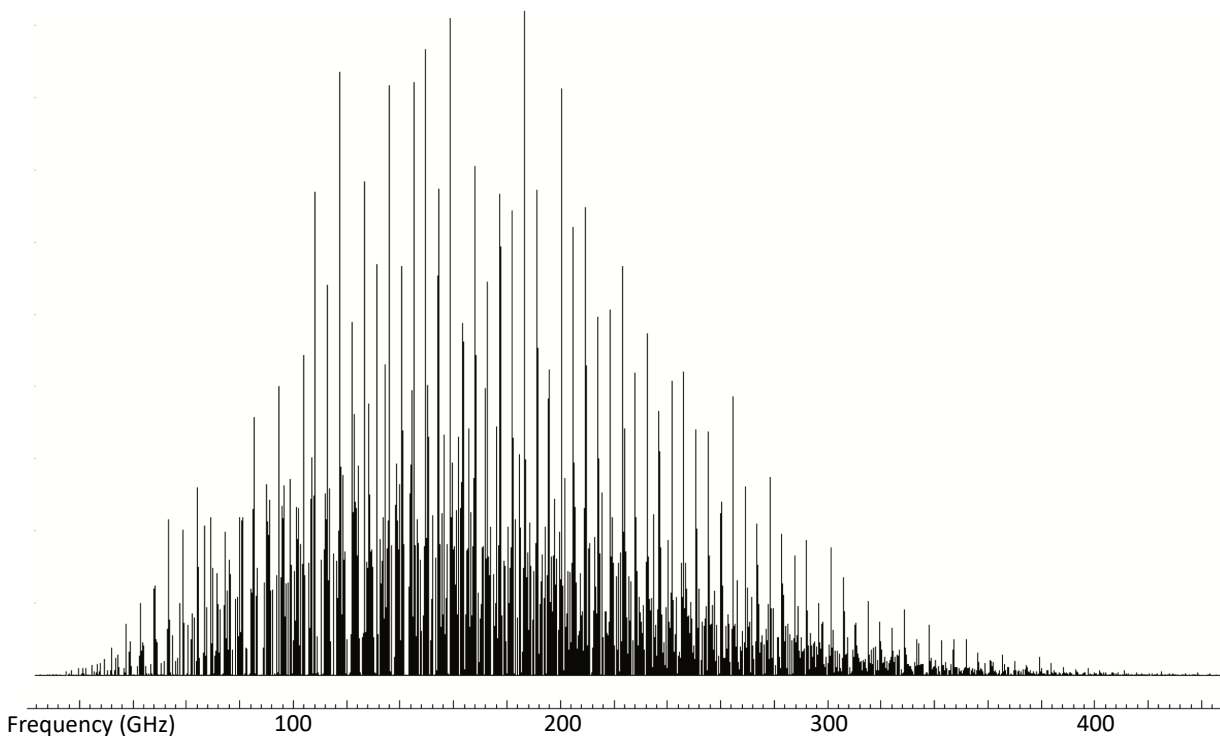
the CCSD(T)-F12/cc-pVTZ-F12 QFF spectroscopic constants reproted in Table 3.3 is given in Figure 3.4 for reference. The anharmonic corrections to the rotational constants are within the accuracy of the method and shift the B and C constants by less than 10 MHz. The Si=O$_a$ bond length at 1.52086 Å in Table 3.3 is more than 0.1 Å longer than the Si−O$_b$ bond length indicating some level of a higher-bond order in the former as a Lewis structure would imply.

The lowest-energy C_{2v} Mg$_2$SiO$_4$ isomer also has an incredibly large CCSD(T)-F12/cc-pVTZ-F12 dipole moment of 9.71 D as reported in Table 3.3. The QFF computations could not be implemented for C_{2v} Mg$_2$SiO$_4$, but the equilibrium rotational constants should be very good estimates for their physical values. The heavier mass of this molecule will push the rotational spectrum to shorter frequencies, but, again, the large dipole moment may also make this molecule observable even if it possesses low column densities. The Si−O bond lengths are comparable between this

35

Table 3.3: CCSD(T)-F12/cc-pVTZ-F12 Equilibrium Structural Data and Rotational Spectroscopic Constants for the Lowest-Energy C_{2v} Isomer of Mg_2SiO_4.

	Units	Mg_2SiO_4
$r_e(Si=O_a)$	Å	1.524
$r_e(Si-O_b)$	Å	1.609
$r_e(Si-O_d)$	Å	3.893
$r_e(Mg-Si)$	Å	3.095
$\angle_e(O_a-Si-O_b)$	°	124.557
$\angle_e(Mg-O_b-Si)$	°	127.487
$\angle_e(Mg-O_d-Mg)$	°	102.325
A_e	MHz	3308.5
B_e	MHz	1309.6
C_e	MHz	938.2
μ	D	9.71

Table 3.4: CCSD(T)-F12/cc-pVTZ-F12 Equilibrium Structural Data and Rotational Spectroscopic Constants for the D_{2d} Isomer of Mg_2SiO_4.

	Units	Mg_2SiO_4
$r_e(Si-O)$	Å	1.664
$r_e(Mg-Si)$	Å	3.095
$\angle_e(O-Si-O)$	°	99.636
$\angle_e(O-Si-O)$	°	114.602
$\angle_e(Mg-O-Si)$	°	86.696
A_e	MHz	4888.9
B_e	MHz	1248.3
C_e	MHz	1248.3

forsterite monomer and the previously discussed $MgSiO_3$ monomer with the symmetric $Si-O_b$ lengths shortening slightly in the heavier form from 1.636 Å to 1.609 Å. The D_{2d} Mg_2SiO_4 isomer has no permanent dipole moment by symmetry, is clearly an oblate-top by the A>B=C rotational constant relationship given in Table 3.4, and the four equivalent $Si-O$ bond lengths all lengthen to 1.664 Å which is, again, closer to crystalline behavior (1.649 Å) at 77 K and 1 atm [54].

3.3.3 Vibrational Spectral Considerations

All three molecules analyzed in this work exhibit exceptionally bright infrared transitions. The brightest is the D_{2d} Mg_2SiO_4 isomer's ν_1 (b_2) antisymmetric $Si-Mg$ stretch where the silicon atom is shuttling between the two magnesium atoms. The benchmark for the antisymmetric stretch of water is right at 70 km/mol, and this

Table 3.5: CCSD(T)-F12/cc-pVTZ-F12 Harmonic and Anharmonic (QFF and scaled[a]) Vibrational Frequencies (in cm^{-1}) for Monomers of ^{24}MgSiO$_3$ and ^{24}Mg$_2$SiO$_4$ (C_{2v} & D_{2h}) with Intensities in Parentheses (in km/mol).

| | | MgSiO$_3$ | | Mg$_2$SiO$_4$ | | |
				C_{2v}		D_{2d}
ω_1	a_1	1278.6 (211)	a_1	1258.5 (315)	b_2	940.5 (639)
ω_2	b_2	872.7 (297)	b_2	1017.6 (320)	e	803.7 (298)
ω_3	a_1	824.7 (84)	a_1	943.9 (231)	a_1	765.1
ω_4	a_1	710.1 (142)	b_2	782.7 (174)	b_2	729.5 (170)
ω_5	b_2	618.1 (32)	a_1	687.8 (117)	a_1	723.8
ω_6	a_1	483.2 (22)	b_2	621.1 (1)	e	642.3 (13)
ω_7	b_1	431.6 (99)	a_1	550.8 (18)	b_2	566.5 (181)
ω_8	b_2	310.0 (5)	b_1	421.2 (100)	e	401.1 (15)
ω_9	b_1	165.4 (14)	a_1	399.9 (7)	a_1	355.8
ω_{10}			b_2	298.0 (52)	b_1	241.4
ω_{11}			a_1	272.0 (28)	e	118.6 (85)
ω_{12}			b_1	230.9 (117)		
ω_{13}			a_2	139.2		
ω_{14}			b_2	112.8 (57)		
ω_{15}			b_1	69.3 (5)		
ZPVE		2837.3		3844.0		4064.8
ν_1	a_1	1261.8	a_1	1239.0	b_2	925.9
ν_2	b_2	858.8	b_2	1001.8	e	791.2
ν_3	a_1	812.0	a_1	929.2	a_1	753.3
ν_4	a_1	700.3	b_2	770.6	b_2	718.2
ν_5	b_2	608.2	a_1	677.1	a_1	712.5
ν_6	a_1	479.9	b_2	611.4	e	632.3
ν_7	b_1	428.7	a_1	542.2	b_2	563.2
ν_8	b_2	308.2	b_1	414.7	e	398.8
ν_9	b_1	164.7	a_1	397.5	a_1	353.7
ν_{10}			b_2	296.3	b_1	240.0
ν_{11}			a_1	270.4	e	117.9
ν_{12}			b_1	229.6		
ν_{13}			a_2	138.4		
ν_{14}			b_2	112.1		
ν_{15}			b_1	68.9		

[a] The MgSiO$_3$ values are fully computed QFFs. The Mg$_2$SiO$_4$ are scaled based on the MgSiO$_3$ average corrections of 98.55% and 99.42%.

fundamental's motion is reported here to be nearly ten times that at 639 km/mol as given in Table 3.5. The anharmonic frequency of this motion at 925.9 cm^{-1} (10.80 μm) falls in a region that is little-populated by likely PAH frequencies [90, 84].

Table 3.5 also contains all of the harmonic and anharmonic vibrational data computed here. Most notably, MgSiO$_3$'s ν_2 and ν_1 at 1261.8 cm^{-1} (7.93 μm) and 858.8 cm^{-1} (11.64 μm) also have intensities above 200 km/mol. The Mg$_2$SiO$_4$ C_{2v} isomer's ν_1 through ν_3 and D_{2d} ν_2 fundamentals all have intensities above 200 km/mol. These are present in the 12.60 μm to 7.93 μm range. Nearly all of the fundamentals for these three molecules that have symmetry-allowed intensities are in what most chemists consider the "bright" range. The intensities are lower for lower frequencies and longer wavelengths on the whole, but Mg$_2$SiO$_4$ isomers each have far-IR frequencies with intensities above 85 km/mol. For instance, the lower-energy D_{2d} isomer has a 100 km/mol ν_9 (b_1) fundamental vibrational frequency at 414.7 cm^{-1} (24.11 μm) approaching the THz range.

The largest anharmonicity present in MgSiO$_3$ is in the ν_1 (a_1) Si=O$_a$ stretch as would be expected for the highest frequency mode. The total average anharmonicity reduces the harmonic frequencies by 98.49 %. However, the modes above 500 cm^{-1} differ from those below 500 cm^{-1}. Those above are 98.45 % while those below are 99.42 %. Furthermore, 0.58% reduction below 500 cm^{-1} is less than 3.0 cm^{-1} within the expected accuracy for the F12-TZ method utilized here. Hence, the lower frequency fundamentals are nearly unaffected by anharmonic corrections which take significantly longer to compute than scaled harmonics. Additionally, the included Fermi resonances refine the anharmonic frequencies in MgSiO$_3$ by less than 1.0 cm^{-1}. As a result, high-level, costly computations can be utilized for computing the harmonic frequencies of these molecules. The given, scaled data for the larger Mg$_2$SiO$_4$ isomers are likely in error compared to the full QFF results by less than 0.01 %. Furthermore, the largest correction excluded here in the QFF results are the core-electron correlation effects,

but these typically only influence stretches with bond orders above one. Granted, the Si=O_a stretches may be in error, but, again, this should be less than an average of 7.0 cm^{-1} [5]. The vibrational frequencies for ^{25}MgSiO$_3$ and ^{26}MgSiO$_3$ are given in Table S1 of the SI.

3.4 Astrophysical Implications and Conclusions

The "uncertain area of the spectrum" for NGC 6302 where crystalline enstatite and forsterite have been observed [90] is ripe for the consideration of these small, magnesium silicates. Their brightest features fall in the 8.0 μm to 18.0 μm range. The computed IR features for the four highest-frequency/shortest-wavelength fundamentals of MgSiO$_3$ lie in ranges where there are unattributed peaks in astronomical IR spectra. While C−rich particles and O−rich crystalline solids have been observed in these areas, the shapes of these other narrow features more closely match the behavior of small molecules rather than networked solids like polycyclic aromatic hydrocarbons, dust, or ice. The spectral behavior is further continued with D_{2d} and C_{2v} Mg$_2$SiO$_4$. The brightest peak of those computed herein is for D_{2d} Mg$_2$SiO$_4$ at 10.80 μm close to notable peaks reported for NGC 6302 [90]. Additionally, the 929.9 cm^{-1} (10.76 μm) and 414.7 cm^{-1} (24.11 μm) features of C_{2v} Mg$_2$SiO$_4$ are enticingly close to other observed features giving at least a striking coincidence for astronomical correlation.

Additionally, the C_{2v} MgSiO$_3$ and Mg$_2$SiO$_4$ monomers have enormous dipole moments making them attractive targets for radioastronomical observation from the ground. Furthermore, radio detection for any of these monomers would then imply that their spectral features should appear in the IR, as well, giving better estimates for column densities and chemical reaction pathways. Neither the rotational constants nor the vibrational frequencies provided herein are grossly affected by anharmonic

corrections, and the CCSD(T)-F12/cc-pVTZ-F12 level of theory has already been shown to provide exceptionally accurate quantum chemical descriptions of molecular behavior. Consequently, the presently computed vibrational and rotational spectral data should be reliable for laboratory comparison if not astronomical observation.

Finally, the most stable structures of the monomers and dimers show that oxygen is crucial for molecular buildup of inorganic species. Oxygen is the central atom that holds most geochemically-relevant molecules together. Dozens of isomers not reported in the presented tables were attempted in this work, and if they contained $O-O$ or third-row$-$third-row bonds, such species either would dissociate or isomerize into a third-row$-O$ bonding motif. Occasionally, some would optimize, but these are so high in energy that they did not warrant further exploration.

3.5   Acknowledgements

Chapter 4

Aharmonic Vibrational Frequencies of Ammonia Borane

Abstract

The fundamental vibrational frequency of the B-N stretch in BH$_3$NH$_3$ has eluded gas-phase experimental observation for decades, despite the interest in the molecule's utility for hydrogen storage. This work offers a theoretical anharmonic prediction of this mode to be 644 cm^{-1}, using a Cartesian quartic force field at the CCSD(T)-F12/cc-pVTZ-F12 level of theory. The other fundamental frequencies reported herein have a mean absolute error of only 5 cm^{-1} from the seven available gas-phase experimental frequencies, making the anharmonic vibrational frequencies and rotational constants the most accurate computational data available for BH$_3$NH$_3$ to date. Ammonia borane also represents one of the largest molecules ever studied by quartic force fields, making this work an important step in extending the breadth of application for these theoretical rovibrational techniques.
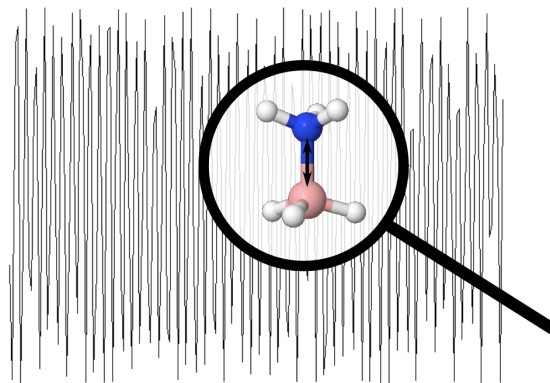
Figure 4.1: Table of Contents Graphic

4.1 Introduction

Ammonia borane is a promising hydrogen storage medium for use in fuel cells due to its high hydrogen density[113, 121, 98, 119]. More recently, it has gained additional prominence as a potential precursor to hexagonal boron nitride, an interesting substrate for subsequent formation of graphene and other 2D nanomaterials[10]. For such a potentially useful molecule, surprisingly little is known about its most informative vibrational mode, the B-N stretch. Spectroscopic investigation into ammonia borane stretches back to the early 1970s when Smith, Seshadri, and White reported its infrared spectrum and that of two of its deuterated isotopologues as captured in an argon matrix[116]. However, the only gas-phase investigations would not take place for nearly a decade when the work of Suenram and Thorne revealed the microwave spectrum of $^{11}BH_3NH_3$ and $^{10}BH_3NH_3$ along with some derived geometrical parameters[122]. Shortly thereafter, Thorne, Suenram, and Lovas updated this list to include a total of nine isotopologues of BH_3NH_3, including substitutions of ^2H, ^{10}B, and ^{15}N, but relatively little further spectroscopic data has been generated since then[126].

The relative dearth of infrared data forced many investigations on the applications of ammonia borane to rely on the matrix data of Smith, *et al.*[116], with some assignments augmented by the SCF/6-31G(d) and MP2/6-31G(d)

computations of Dillen and Verhoeven[24]. Finally in 2012, the work of Sams *et al.* provided a combined gas-phase experimental and computational investigation of the rovibrational characteristics of BH_3NH_3[113]. Unfortunately, the B-N stretch was not detected in the experiment due to its low intensity, but the CCSD(T)/aug-cc-pVTZ harmonic frequencies combined with MP2/aug-cc-pVTZ second-order vibrational perturbation theory (VPT2) anharmonicities provide the best estimate of that stretching mode currently available at 630 cm^{-1}. Despite being the best theoretical data available, the computed values offered by Sams *et al.* are not without their limitations. Although the harmonic portions of the fundamental frequencies were computed at the CCSD(T)/aug-cc-pVTZ level, the addition of the MP2 anharmonicities cannot offer as high accuracy as a uniform CCSD(T) treatment. This type of composite construction of the anharmonic frequencies does not allow for a robust handling of Fermi resonances at the CCSD(T) level, potentially limiting its accuracy[79, 78].

As interest in ammonia borane continues due to its importance in alternative fuels and nanomaterials, a more accurate picture of its B-N stretch becomes ever more necessary. In particular, the B-N stretching frequency has been suggested as an important identifier for the molecule's environment[119, 24, 56], making its accurate gas-phase determination necessary for understanding how changes to its frequency are indicative of shifts in the surrounding chemical environment in applied studies. The same exquisite sensitivity that could make the B-N stretch useful also complicates experimental observation. Previous computational work[27] on the BH_3NH_3 dimer has shown that the harmonic frequency of the B-N stretch likely shifts by up to 40 cm^{-1} due to intermolecular dihydrogen bonding. Similarly, IR and Raman experiments[56, 94] on solid BH_3NH_3 show even larger shifts of nearly 150 cm^{-1} from the frequency calculated by Sams *et al.* Therefore, only pure gas-phase spectra can produce the necessary vibrational frequency for the B-N stretch. As described in Ref. 113, a veritable "armamentarium" of experimental techniques have been used to in-

vestigate this molecule, but the addition of gas-phase infrared spectroscopy will provide specific information for this mode. Such is only possible if the baseline fundamental vibrational frequency can be clearly determined. Consequently, the present work seeks to provide a comprehensive anharmonic analysis at a consistent level of theory, along with accuracy-boosting resonance corrections.

The state-of-the-art quantum chemical approach to calculating rovibrational spectral data for the lowest computational cost is the quartic force field (QFF). QFFs are fourth-order Taylor series expansions of the internuclear potential term in the Watson Hamiltonian[35]. When joined with coupled cluster theory at the singles, doubles, and perturbative triples level[107] within the F12 explicitly correlated construction (CCSD(T)-F12b)[3, 73] and a triple-ζ basis set, QFFs can typically achieve accuracy of 5 to 7 cm^{-1} compared to experiment[62, 6, 4, 5, 134]. A limitation of the QFF methodology as previously used by our group is its reliance on symmetry internal coordinates (SICs) for the generation of the displaced geometries needed to produce the QFF. While SICs take advantage of molecular symmetry to reduce significantly the number of points that must be computed, their determination is often more art than science and can be particularly difficult for larger and more symmetric molecules, like those with C_{3v} symmetry. As such, another purpose of the present study is to demonstrate the use of Cartesian coordinates in generating the QFF displacements. Applications include molecules that would otherwise be too large or complicated to investigate via SIC-based QFFs and those where the SICs and normal coordinates are not well-aligned, requiring a broader sampling of the potential energy surface. The improvement of Cartesian QFF methodologies will allow for the faster generation of gas-phase spectra for such application-rich but size-prohibitive molecules as ammonia borane.

4.2 Computational Details

All of the CCSD(T)-F12b computations performed in this study, including geometry optimizations, harmonic frequencies, and single-point energies, utilize the Molpro 2015.1 software package[133] with the cc-pVTZ-F12 basis set[62, 101, 58] under the frozen core approximation and with very tight convergence criteria for all of the integrals and for the gradients in the geometry optimization. This CCSD(T)-F12b/cc-pVTZ-F12 scheme is abbreviated as F12-TZ in the following. The anharmonic vibrational frequencies and rotational constants are calculated using a Cartesian QFF methodology at this F12-TZ level. In order to map out the fourth-order Taylor series approximation of the internuclear potential energy surface, i.e. the QFF, displacements of 0.005 Å are taken along each Cartesian coordinate. These are simply vectors in the x, y, and z directions for each atom. At each displaced geometry, a single-point energy is computed and this energy is used to calculate the corresponding force constant by central finite differences. The Cartesian QFF requires a nominal total of 271780 single-point calculations (compared to the 39605 required for an SIC QFF), assuming every component energy of each finite difference is computed. However, saving the reference energy and an array of second-derivative energies for use in the fourth-derivative finite difference calculations reduces the number of these points that actually have to be computed roughly by half.

The resulting force constants are then used by the second-order rotational perturbation theory[86] and VPT2[131, 95] in the SPECTRO[45] program. Type 1 and 2 Fermi resonances and polyads, Coriolis resonances, and Darling-Dennison resonances are taken into account to further increase the accuracy of the anharmonic values[79, 78]. The resonances are shown in the Supporting Information (SI) in Table S1. The anharmonic analysis was repeated for the $^{10}BH_3NH_3$ and the $BH_3{}^{15}NH_3$ isotopologues. These results are also shown in the SI in Tables S2 and S3, while the rest of the rotational and geometrical data

is shown in Tables S4 and S5. Additional MP2[89], CCSD[105], and canonical CCSD(T)[107] computations were carried out to systematically examine the effects of dynamical electron correlation, basis set size, diffuse functions and the frozen core approximation on the optimized structures and harmonic vibrational frequencies of BH_3NH_3. The associated computational details and corresponding results (Tables S6–S16) are shown in the SI. The abbreviations XZ, aXZ, and CXZ are hereafter used to describe the cc-pVXZ[29, 100, 136], aug-cc-pVXZ[29, 67], and cc-pCVXZ[29, 137] basis sets used therein. Within the CXZ computations, FC denotes that the frozen core approximation was used, while AE indicates that all electrons were correlated.

4.3 Results and Discussion

Table 4.1: Harmonic vibrational frequencies (in cm^{-1}) of BH_3NH_3.

Mode	Description	MP2 5Z	CCSD 5Z	CCSD(T) aTZ[a]	CCSD(T) 5Z	CCSD(T)-F12 F12-TZ QFF
$\omega_1(a_1)$	symm. N-H str.	3498	3507	3466	3476	3474
$\omega_2(a_1)$	symm. B-H str	2486	2457	2430	2447	2445
$\omega_3(a_1)$	symm. NH_3 def.	1329	1355	1353	1339	1340
$\omega_4(a_1)$	symm. BH_3 def.	1213	1212	1196	1200	1199
$\omega_5(a_1)$	N-B str.	687	679	673	678	679
$\omega_6(a_2)$	torsion	264	259	274	259	259
$\omega_7(e)$	antisymm. N-H str.	3628	3620	3580	3592	3590
$\omega_8(e)$	antisymm. B-H str.	2553	2512	2493	2505	2503
$\omega_9(e)$	antisymm. NH_3 def.	1675	1692	1679	1677	1678
$\omega_{10}(e)$	antisymm. BH_3 def.	1219	1212	1181	1202	1201
$\omega_{11}(e)$	antisymm. BH_3 rock	1081	1085	1061	1075	1076
$\omega_{12}(e)$	antisymm. NH_3 rock	648	654	656	646	647

[a]From Ref. 113

The substantial differences in many of the harmonic frequencies of BH_3NH_3, as shown in Table 4.1, in going from MP2 to CCSD to CCSD(T) demonstrate the effects of the method on the frequency computations. As shown in Tables S6–S8 of the SI, moving from triple-ζ to even 6-ζ quality within a given basis set series is not as impactful. In the case of the MP2 harmonic frequencies in Table S6, the maximum difference between the 5Z and 6Z values is only

1 cm^{-1}, suggesting that 5-ζ quality is sufficient in the other computations as well. Hence, the agreement of the CCSD(T)/5Z results with the Cartesian F12-TZ QFF values demonstrates the fortuitous performance of the substantially cheaper F12-TZ combination. The largest difference between these two data sets is 2 cm^{-1}, but most of the deviations are within 1 cm^{-1}. For ω_5 and ω_6, even CCSD/5Z produces good agreement with the CCSD(T)/5Z and F12-TZ values, and all three are within 1 cm^{-1} of each other. Such agreement indicates at least some degree of method independence for these modes and bodes well for the reliability of the F12-TZ methodology in the anharmonic computations.

In contrast, the MP2 frequencies are typically farthest from both the CCSD(T)/5Z and F12-TZ results. The addition of the diffuse functions in the aTZ computations of Tables S6–S8 also decreases all the frequencies compared to the TZ values, introducing larger deviations than any within the XZ series. Moving to the larger aQZ basis set restores agreement with the XZ results, but neither canonical CCSD(T)/aTZ nor MP2 seems very promising on its own in light of these data. On the other hand, the effects of the frozen-core approximation seem relatively minor. The largest deviation across the three methods between the CXZ FC and AE results is 6 cm^{-1}, with many of the differences well inside that margin. As such, maintaining the frozen-core approximation in the F12-TZ QFF is well worth the commensurate computational savings.

Only the F12-TZ QFF approach is carried forward to the anharmonic analysis because the harmonic results suggest it strikes a good balance between accuracy and efficiency. As shown in Table 4.2, the anharmonic fundamental frequency of the B-N stretch, ν_5, is computed here to be 644 cm^{-1} at the F12-TZ level after the inclusion of the Fermi, Coriolis, and Darling-Dennison resonances. Whereas the value before the resonance correction of 631 cm^{-1} agrees nearly perfectly with the previous computational value[113] of 630 cm^{-1}, the inclusion of the type 1 Fermi resonance between $2\nu_6$ and ν_5 increases ν_5 by 13 cm^{-1} compared to the unperturbed value. This seemingly small but truly sub-

Table 4.2: Anharmonic vibrational frequencies (in cm^{-1}) of BH$_3$NH$_3$

Mode	Matrix[a]	Gas[b]	Previous Theory[c]	F12-TZ QFF[d] Pert.	F12-TZ QFF[d] Unpert.
ν_1	3337	-	3321	3333	3307
ν_2	2340	2298.861(8)	2449	2275	2338
ν_3	1301	1288.6384(3)	1297	1287	1287
ν_4	1052	1177.56(3)	1208	1174	1177
ν_5	603	-	630	644	631
ν_6	-	-	252	300	300
ν_7	3386	3417.81(6)	3410	3415	3415
ν_8	2415	2405.58(2)	2392	2401	2392
ν_9	1608	1610.62(2)	1608	1610	1622
ν_{10}	1186	-	1165	1169	1171
ν_{11}	968	1042.316(2)	1027	1047	1044
ν_{12}	-	-	648	649	649

[a]Ar-matrix infrared frequencies from Ref. 116 with reassignment of peaks from Ref. 24 by SCF and MP2/6-31G(d) computations
[b]Gas-phase infrared frequencies from Ref. 113
[c]CCSD(T)/aug-cc-pVTZ harmonics plus VPT2 MP2/aug-cc-pVTZ anharmonicities from Ref. 113
[d]Modes 1, 2, 4, 5, 8, 9, 10, 11 are affected by Fermi resonances. Unperturbed values are prior to the resonance correction.

stantial difference indicates the importance of accounting for Fermi resonances and reflects the strength of methods that support their incorporation throughout the analysis, such as the QFF methodology conjoined with the SPECTRO program used herein.

As shown in Table 4.2, resonance corrections do not guarantee moving closer to individual experimental results. In the case of ν_4 and ν_{11}, the resonance-corrected F12-TZ values are actually slightly farther from the gas-phase experimental values than the uncorrected ones. However, in aggregate, the performance of the resonance-corrected F12-TZ data is substantially better than that of the other computational set. Whereas the F12-TZ frequencies have a mean absolute error (MAE) of only 5 cm^{-1} compared to the seven available gas-phase values[113], the MAE of the CCSD(T)/aTZ harmonics augmented with MP2/aTZ anharmonicities from Ref. 113 is 24 cm^{-1}. Such performance

is to be expected of F12-TZ QFFs, which typically offer accuracy within 5 to 7 cm^{-1} of gas-phase experiment[62]. In fact, the deviations from experiment in all but ν_2 are less than the lower bound of 5 cm^{-1}, but the substantially larger difference of 24 cm^{-1} in ν_2 drives up the average. Regardless, the F12-TZ data represents the most accurate theoretical data available for ammonia borane.

Potential further evidence for the identification of ν_5 at 644 cm^{-1} lies in Fig. 9 of Ref. 113. In this infrared absorption spectrum of BH_3NH_3 between 550 and 810 cm^{-1}, there is a clear increase in the baseline from 670 to 630 cm^{-1}. The authors identify an actual peak above the noise potentially belonging to BH_2NH_2 in this region, but the presence of other peaks and the baseline elevation suggests that more may be happening in this portion of the spectrum. Most enticingly, there appear to be several unattributable peaks extending from the baseline in this region. Unfortunately, the absorbances of these fundamental frequencies are even less than that of the BH_2NH_2 peak classified as a "bump" by the authors at less than 0.004 absorbance units[113]. The low computed CCSD(T)/5Z intensity of only 12 km mol^{-1} for the B-N stretch could allow it to remain hidden in this area of the spectrum and would require a higher-resolution experiment to fully elucidate.

Also shown in Table S2 of the SI is a comparison of the F12-TZ computed fundamental frequencies with the four experimental $^{10}BH_3NH_3$ frequencies obtained by Sams *et al.* The MAE for these four modes is only 3.1 cm^{-1}, suggesting that the F12-TZ QFF captures the frequency changes due to isotopic substitution very effectively. While experimental frequencies for the $BH_3{}^{15}NH_3$ isotopologue are unavailable, the F12-TZ results for that molecule are presented in Table S2 as well. Table S4 shows that both the F12-TZ and Sams *et al.* computational data sets additionally agree well with the experimental geometrical parameters, except for the F12-TZ value of the H-B-N bond angle, which appears to overshoot the experimental measurement by nearly a whole degree. Additional rotational constants, including the first vibrationally excited prin-

ciple rotational constants for the $^{10}BH_3NH_3$ and $BH_3{}^{15}NH_3$ isotopologues are available in the SI Tables S3 and S5.

4.4   Conclusions

In conclusion, the anharmonic vibrational frequencies and rotational constants of BH_3NH_3 reported herein agree well with available gas-phase experimental data. While previous computational results using a combined CCSD(T) and MP2 approach offered anharmonic frequencies with an average deviation from experiment of 24 cm^{-1}, the F12-TZ QFF values in the present study achieve agreement to within 5 cm^{-1}. Consequently, these represent the most accurate theoretical values currently available for BH_3NH_3, and in the case of the elusive B-N stretch the only gas-phase-comparable data. While the low intensity of the B-N stretch has allowed it to elude experimental observation, it is identified here computationally to be 644 cm^{-1}, offering new insight into the chemical environment of an application-rich molecule. The impressive agreement of the other vibrational frequencies with available experiment lends additional credence to the fidelity of this assignment from the purely computational value for the dark B-N stretch, as well. More generally, the success of a new Cartesian QFF methodology in describing the anharmonic rovibrational characteristics of a molecule quite complicated to examine with user-defined internal coordinates is promising for the future of QFF analysis. Additionally, Cartesian coordinates provide a more complete description of atomic movement, unconfined to specific internal coordinate schemes. This allows for the treatment of floppy systems for which SICs become ill-defined. Further improvements to the current software, such as using symmetry in the Cartesian treatment to further reduce redundant computations, will continue to extend the domain of vibrational problems that can be examined with QFFs.

4.5   Acknowledgement

Chapter 5

Genetic Algorithms

> We will now discuss in a little more
> detail the Struggle for Existence.

Charles Darwin

5.1 Darwin's Principles of Evolution

When the principles of evolution were first promulgated by Darwin, he could not have imagined that it would be applied to "thinking machines" of the future. A brief discussion of evolution is in order since its concepts are immediately applicable to astrochemical genetic algorithms.

Evolution is the way that organisms change over time. According to Mayr, the theory has the following critical components: fertility, the uniqueness of each individual, artifical selection, and natural selection [81]. For ferility, Darwin believed the ability for exponential increase of each species was possible, but was limited by outside factors such as limited resources [20]. Darwin also noticed how animal breeders were able to select favorable traits in their stock through choice breedings, since each individual animal was not unique, just as no two humans are the same [20, 81]. The most important piece of the evolutionary puzzle, however, was natural selection, or, "the struggle for existence."

Natural selection can be defined as "a non-random difference in reproductive output among replicating entities . . . leading to an increase in the proportion of beneficial, heritable characteristics" [50]. Due to the limited resources, organisms must compete. The individuals with greater fitness in the environment are more likely to pass on their genes to the next generation, therefore, hopefully improving the overall fitness of the population. If the environment were to have unlimited resources the population would grow exponentially. For the process of evolution to work, however, there must be both genetic inheritance and variation to introduce new genes [50]. Due to advances in genetic understanding, these are now understood to be the workings of DNA and crossing-over [18]. Recall also that human genes are sorted into chromosomes.

Taken together, the principles of evolution can follow this simplified process. First, a population of organisms grows exponentially until the resources limit its growth. Then, the organisms compete for limited resources and those that are successful are able to pass on their genes. In doing so (for sexual reproduction proccesses), the offspring inherit genes from both parents, as well as new genes due to genetic mutations. If repeated over a long period of time, or if the environmental demands change, the species will evolve and express different traits [50]. It is this process which can be translated into evolutionary computational algorithms.

5.2 Introduction to Genetic Algorithms

The above principles can be directly translated into computer algorithms, termed genetic algorithms (GAs). At their core, genetic algorithms are employed to solve optimization problems. This is, in essence, the goal of minimizing things. To do so, the weights of various inputs are changed so that the output's difference is minimized. The inputs are variables, the evaluation of the inputs is known as the fitness or loss function, and the output is the fitness

[52].

The first genetic algorithms were developed by John Holland in the 1960s as he sought to abstract the process of biological evolution to computational procedures [87]. In Holland's original formulation, the organisms had chromosomes that were encoded by the fundamental units of computation, 0 and 1, rather than the biological purines and pyrimidines [87, 59]. The main draw of genetic optimization is that it can leverage the processes of Darwin's principles to evaluate an extraordinarily large problem space that would be computationally impossible to traverse exhaustively as the genetic solution adapts to the problem space. Furthermore, GAs are capable of solving problems that are poorly modeled due to factors such as lack of continuity, derivatives, linearity, or other difficult features [15]. In the present work, a GA is employed since the fitness function is unable to be evaluated directly.

5.2.1 The Genetic Algorithm Process

A GA utilizes the principles of Darwinism evolution. While the implementation of each piece may be unique to specific GAs, the following basic components are present in all GAs.

- A population of organisms with chromosomes (DNA).

- A fitness function to evaluate the fitness of organisms in the population.

- A method of natural selection to determine which organisms will survive and pass on genes.

- A crossover method to determine how genes will be inherited.

- A mutation method to introduce variation in the population.

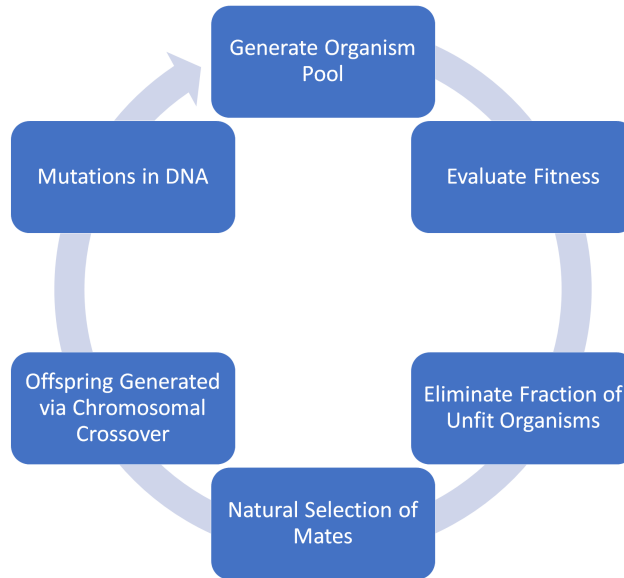The above process is illustrated by Figure 5.1:

Figure 5.1: The GA optimization process.

First, a population of organisms is generated through a random process or via an initial guess near the global minimum if additional information is known about the problem space. These organisms and their DNA are simply lists of values, whether discrete or continuous, real-valued genes [52, 59]. These are usually encoded as a vector element for each dimension of the problem. Therefore, a problem with n dimensions would have organisms encoded as the following chromosome with n genes [49]:

$$\text{chromosome} = [x_1, x_2, \ldots, x_n] \tag{5.1}$$

The organisms must now be judged on their fitness in their environment. In Darwin's principles, fitness is intrinsic to the traits bestowed upon each organism by their genetic code. Similarly, the genetic organisms' fitness comes from their genes. Some common fitness functions include the Bent Cigar function, the Rotated Discus function, and the Rosenbrock function. The Rosenbrock function in 2D is shown below and the organisms have a chromosome with two elements [109].
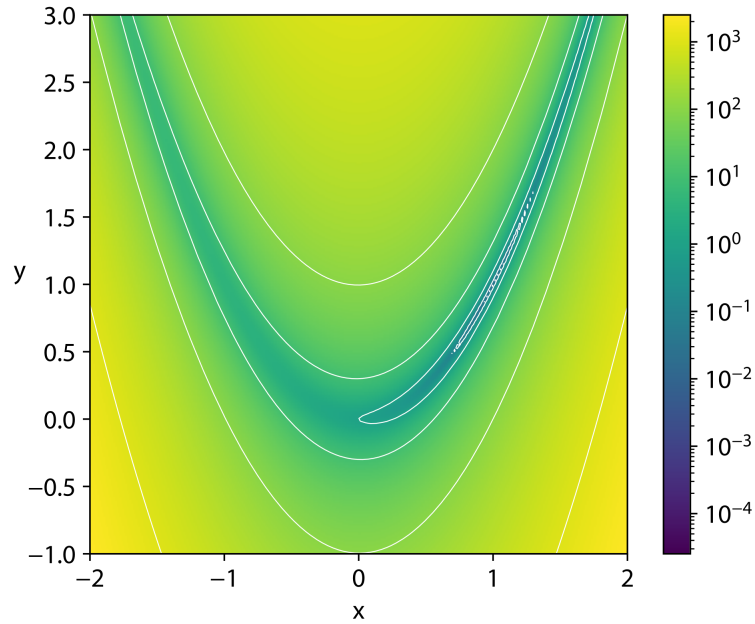
Figure 5.2: The Rosenbrock function in 2D. Its function is $f(x, y) = (a - x)^2 + b(y - x^2)^2$, where a = 1 and b = 100. The absolute minimum is at (1, 1).

Thus, the Rosenbrock fitness function tests each organism to determine its distance from the global minimum. The further the organism from the global minimum, the less fit. However, the fitness function is a generic operator and can be defined in any way that models the problem space. A GA is most useful compared to other methods when the number of dimensions is great (approximately 10 or more) [49].

Now, some organisms must die. Just as the struggle for existence favors the most fit organisms in a biological environment, the natural selection operator favors the most fit organisms in a GA. Usually, the lower 50% is immediately removed from the mating pool. From the remaining pool, there are many methods to determine which organisms pass on their genes, but the two most common are the roulette wheel selection, as described by Haupt & Haupt [52], and the tournament selection method, as described by Miller & Goldberg [85]. In the roulette wheel method, organisms are given a weighted chance to be selected based on their fitness versus the overall fitness, usually referred to as

rank weighting. Here, P_n is the probability of organism n being chosen, f_n is the fitness of organism n, and the sum corresponds to the total fitness of the population.

$$P_n = \frac{f_n}{\sum_{i=1}^{n}} \tag{5.2}$$

For an example population with 5 organisms, the probabilities are tabulated below.

Table 5.1: An example of roulette wheel selection. Notice that the probability is greatest in the organism with the highest fitness.

Organism	Fitness	P_n	$\sum_{i=1}^{n}$
1	5.00	0.33	0.33
2	4.00	0.27	0.60
3	3.00	0.20	0.80
4	2.00	0.13	0.93
5	1.00	0.07	1.00

On the other hand, tournament selection utilizes an "arena" where random organisms are selected to compete. From this arena, the most fit organism emerges and is allowed to mate while the others perish and are removed from the mating pool. This selection method is useful since it is not necessary to sort the organisms by their fitness [52, 85].
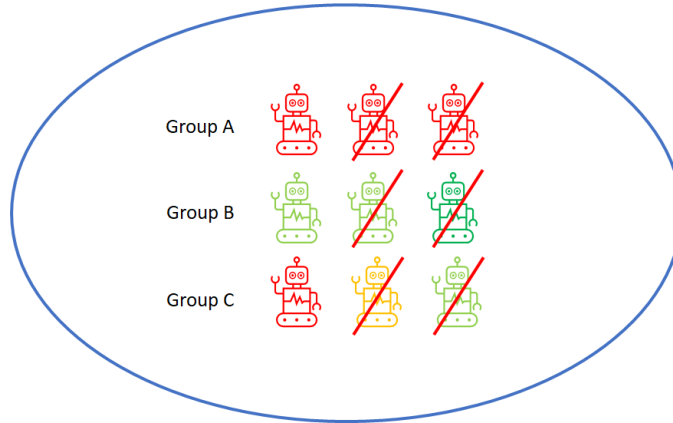
Tournament Selection

Figure 5.3: The above is an example of tournament selection. The most fit organisms are red, and the least fit are green. The most fit organism survives, even if it is possibly less fit than the other winners.

The winners must replace the organisms that were lost to natural selection, which corresponds to "Offspring Generated via Chromosomal Crossover" in Figure 5.1. There are several methods for doing so, but the most useful for real-valued genes is known as the quadratic crossover method [2]. This method utilizes three parents to make an offspring organism, rather than the traditional two in the biological world. In doing so, the three points can be fitted to a parabola and the maximal fitness can be determined. The following equations demonstrate the process:

$$h_j(\zeta) = a_j\zeta^2 + b_j\zeta + c_j, \qquad (5.3)$$

$$a_j = \frac{1}{v_{3j} - v_{2j}}\left[\frac{f_3 - f_1}{v_{3j} - v_{1j}} - \frac{f_2 - f_1}{v_{2j} - v_{1j}},\right] \qquad (5.4)$$

$$b_j = \frac{f_2 - f_1}{v_{2j} - v_{1j}} - a_j(v_{2j} + v_{1j}), and \qquad (5.5)$$

$$c_j = f_1 - a_j v_{1j}^2 - b_j v_{1j}. \qquad (5.6)$$

In the above equations, f_n describes the fitness of the nth parent and v_{nj} is the gene value of the parent at the given index. Therefore, when the coefficents in Equations 5.4, 5.5, and 5.6 are solved and inserted into Equation 5.3, a parabolic description of the three parents is given with the fitness on the y axis and the gene value on the x axis. Simple differentiation of Equation 5.3, yields the following result:

$$\frac{d}{d\zeta}h_j(\zeta) = b_j + 2a_j\zeta. \tag{5.7}$$

When setting Equation 5.7 equal to 0, the critical point is $-\frac{b_j}{2a_j}$. If the second derivative is negative, then the critical point is a maximum, and the fitness is optimized at the corresponding x value [2]. The fitting is illustrated in Figure 5.4
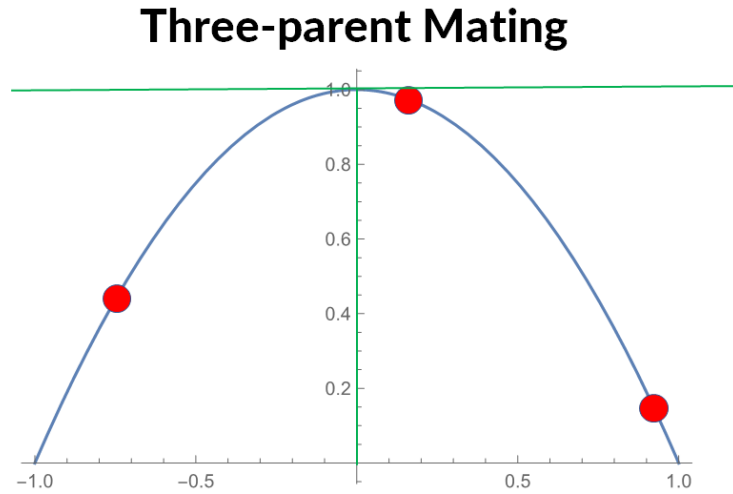

Figure 5.4: An idealized parabola fitting with 3 parents. The green lines represent the absolute maximum y and the corresponding x value. The best gene value is 0.

However, there are cases where there is no maximum or the interpolation fails. In such a case, a linear interpolation is performed between the greatest and least fit of the 3 parents:

$$d_k = r(m_{1k} - m_{2k}) + m_{1k}. \tag{5.8}$$

In the above equation, the d_k term is the gene value for the kth gene in the chromosome. The r term is a random value between $[0, 1]$. Then, the m_{1k} and m_{2k} terms correspond to the gene value of the most and least fit parent at the kth gene index, respectively. However, this can also fail by generating a term outside of the domain interval for the genes. To fix this, the value of r is halved several times, checking each time to see if d_k is acceptable. If it is unable to generate an acceptable value, one of the parents' gene values is assigned to the child [2].

The final step of the GA process is to introduce genetic variance, which emulates the mutations that would be introduced due to crossover in biological systems. This is the simplest process where each gene has a chance to be mutated. If a gene is selected to be mutated, then the original gene replaced with another random value in the domain [52]. An alternative method is to add or subtract a normally distributed value to the gene value, thereby displacing the value rather than completely replacing it. The correct rates of mutation, however, should be implemented as a dyanmically changing value with respect to the convergence rate of the population, rather than a static value [51].

$$p_n' = r, r \in (L_b, U_b) \tag{5.9}$$

$$p_n' = p_n + \sigma N_n(0, 1) \tag{5.10}$$

Equation 5.8 shows the random replacement of a gene by selecting a random value within the lower bound, L_b, and the upper bound U_b. In contrast, Equation 5.10 shows the additive version of the genetic mutation with a normally distributed value. One cycle of this entire process is termed a "generation" and it continues until a fitness threshold has been reached.

5.3   Applications to Astrochemistry

It may be that the QFFs discussed previously are too expensive or complex to calculate in some cases, especially for molecules with more than 7 atoms present. The QFFs approximate the force constants very closely, but the accuracy diminishes with each Taylor series term. It may be useful, therefore, to reverse the QFF process in order to obtain some useful data without constructing all of the energy displacements. For example, one may wish to extract the equilibrium rotational force constants of molecules. Instead of beginning with a structure, one may instead begin with the experimental frequencies and derive the force constants from them. Figure 5.5 demonstrates the simplified QFF process.

Figure 5.5: The simplified QFF process and the proposed reverse transformation.
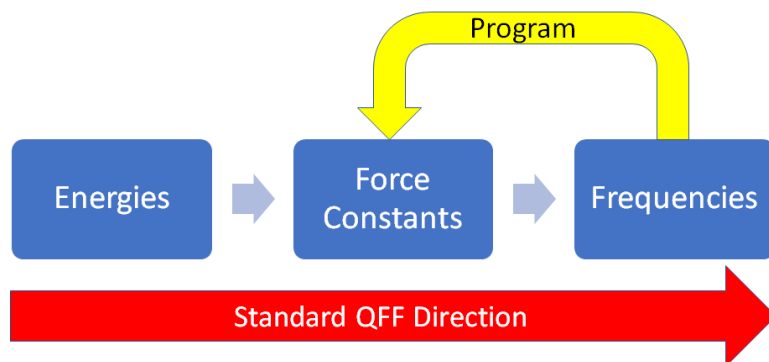
But how can this be accomplished? Similar to cryptographic analysis, the P vs NP problem [125] continues to plague computer algorithms. In essence, the forward step is "easy" while reversing the process is "hard". Of course, one could always try every exhaustive linear combination of force constants that combine to produce the experimental frequencies, but such a process would take a painstaking amount of time.

In this particular problem space, utilizing neural networks is infeasible. QFFs are computationally expensive and have only been completed for a small subset of molecules. Without a large sample (thousands) of many, many per-

mutations of atoms, the model fitting is likely to be insufficient. Furthermore, gradient descent methods, while extremely useful for other optimization problems, are precluded by a Jacobian matrix that would span thousands of dimensions and would furthermore be crippled by the thousands of iterations in each force constant. The main issue stems from the fact that the second order perturbation methods (VPT2) [39] used are too slow to incrementally change each force constant by some amount and evaluating the change in the loss function, especially since the domain is real-valued numbers (see Equation 5.1. Implementation of a "gradient descent-like" method, similar to particle swarm optimization [68], as suggested by Gozali [49], shows at least two orders of magnitude worse convergence than traditional and psuedo genetic algorithms. This is worsened since the force constants can only be evaluated indirectly, with VPT2 acting as a "black box".

Genetic algorithms, on the other hand, can be applied for several reasons. First, the change in force constants is motivated by the genetic algorithm, rather than constantly evaluating the loss function. Instead of evaluating thousands (maybe millions) of times for each organism, each organism is only evaluated once for fitness. Second, the genetic algorithm needs no knowledge of the problem space to effectively find the minimum, which is useful since we evaluate the loss function indirectly.

To begin, the implementation for solving this problem using Figure 5.1 is detailed. First, a pool of organisms is generated. A unique feature of the organism's DNA is that, like human genes, they are organized into chromosomes. This is a two dimensional matrix. The first dimension is the collection of chromosomes, with the first index corresponding to the second order force constants; the second to the third order force constants; and the third to the fourth order force constants. Then, inside each chromosome is the list of force constants corresponding to each derivative order. Therefore, each organism's DNA has the following form:

$$\begin{bmatrix} g_{1,1} & g_{1,2} & g_{1,3} & \cdots & g_{1,i} \\ g_{2,1} & g_{2,2} & g_{2,3} & \cdots & g_{2,j} \\ g_{3,1} & g_{3,2} & g_{3,3} & \cdots & g_{3,k} \end{bmatrix} \qquad (5.11)$$

where $g_{l,m}$ is the $(l+1)$-th order force constant at index j, which are also referred to as genes. Also, the following relationship holds for the lengths of each chromosome: $i < j < k$. To illustrate the notation, $g_{1,1}$ is the first force constant in the list of the second order force constants. This separation is necessary to keep the genes separate, since they will segregate independently and therefore follow Mendel's laws of genetics. A random population can be generated given domain bounds for each order of force constant, or an initial guess at a quicker level of theory (e.g., B3LYP) can be input.

Next, the organisms are evaluated for fitness after evaluating their force constants and converting them into frequencies via VPT2. Equation 5.12 shows the fitness function used:

$$F = \sum_{i=1}^{n} (x_i - \tilde{x}_i)^2. \qquad (5.12)$$

Where x_i represents the organism's outputted frequency in index i, and $\tilde{x}_i$ represent the experimental frequency at index i. Thus, the organisms that are closest to the wanted values are the most fit. While it may seem contradictory to use a smaller fitness value as "more fit", the fitness definition is arbitrary. However, the quadratic fitting crossover must instead find the minimum, rather than the maximum.

The rest of the process follows from the introduction section. The organisms are sorted and those below the median in fitness are killed off and do not mate. Then, the remaining organisms undergo tournament selection to create a pool of mating organisms. Next, three organisms are selected to mate and the offspring are generated via the quadratic method (as in Equation 5.3).

Finally, each organism is mutated slightly to preserve genetic diversity. This process is then repeated until the threshold criteria for fitness is met. The other useful feature of the genetic algorithm is that, for the most part, the only parameters that significantly change between problem spaces are the definition of the organism and the fitness function. It could, as a result, optimize a wide variety of problems.

5.4 Multi-Island Method

The issue with a single island model is that genetic diversity is not well-preserved, which leads to premature convergence [9] [111]. Therefore, separating the species populations could be an excellent place to begin, since the populations will evolve separately. Additionally,, a new evolution method could also be chosen to ensure that the species will evolve significantly different. The methods describe by Gozali [49] [48] describe a method using three islands with three different evolution strategies. Figure 5.6 shows the general idea of one master island and several worker nodes.
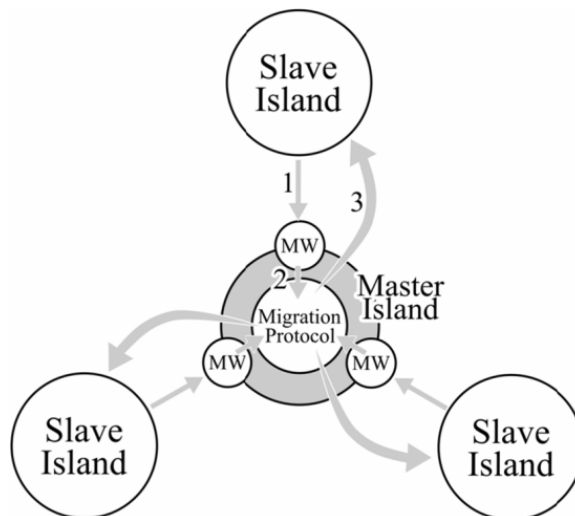

Figure 5.6: A diagram of migration protocols, adapted from Gozali 2017 [48].

The first worker island is the traditional genetic algorithm (TGA). The TGA

implements all of the steps described in the previous sections and in Figure 5.1.

The second worker island is a pseudo genetic algorithm (PGA). The current implementation is the one described by Chen, Zhong, and Zhang [16]. The significant difference between a TGA and a PGA is that the PGA does not implement natural selection of mating parents, nor is there quadratic crossover for the generation of new offspring. Thus, the only steps are the initialization of the population, evaluation of the fitness and a very simple mating method. All of the organisms are created in pairs, and the genes are exactly complementary. For real-valued genes, the complement of $g_{l,m}$ is $-g_{l,m}$. Furthermore, the mating process is simply a random chance of swapping the gene value at each chromosome. Equation 5.13 illustrates the process for a simple one chromosomal organism, starting with both parents as exact complements.

$$\begin{bmatrix} g_{1,1} & g_{1,2} & g_{1,3} \end{bmatrix} \mathcal{M}_{pga} \begin{bmatrix} -g_{1,1} & -g_{1,2} & -g_{1,3} \end{bmatrix} \rightarrow \begin{bmatrix} -g_{1,1} & -g_{1,2} & g_{1,3} \end{bmatrix}, \ C_2 \tag{5.13}$$

This can be represented as

$$P_1 \ \mathcal{M}_{pga} \ P_2 \rightarrow C_1, \ C_2 \tag{5.14}$$

where the first term organism is Parent 1, the second is Parent 2. Mating them with the PGA mating operator ($\mathcal{M}_{pga}$) yields the first child organism, as well as the complement C_2, which is the same as $C_2 = -1 * C_1$.

The reasoning behind this mating operator is due to a concept called the Hamming distance, which is essentially a measure of the difference of two organisms [14]. The PGA mating operation ensures that the Hamming Distance between two organisms is maximized, since the maximum distance between two organisms is one with opposite gene values. [16]. This is beneficial since more genetic diversity ensure the problem space will be traversed effectively, as well as avoiding premature convergence. It is defined, for real value genes, as

$$\delta(P', P'') = \sum_{l=1}^{L} \left| \left[\frac{P'_l + UB}{2UB} - \frac{P''_l + UB}{2UB} \right] \right| \tag{5.15}$$

where δ is the Hamming distance between the two organisms, P' and P''. The result of this process is a much faster genetic algorithm, though the pool of genes never changes, only the mixture within the population. Therefore, no genes are lost, which preserves genetic diversity. At the end of each generation, the most fit organism is sent to the master island for migration.

The final genetic algorithm is called an informed genetic algorithm (IGA), loosely based on the works of Kennedy [68] and Shi [115]. The IGA possesses all of the features of the TGA, but it does have a modified organism structure and mutation procedure. Essentially, the IGA implements an internal gradient descent method, where $\frac{\partial F}{\partial g_{l,m}}$ is evaluated. This is the change in fitness with respect to a change in gene value. The exact implementation of this method, as described by Gozali [49], reduces the typical velocity and acceleration components of a particle swarm to a simple Boolean. Therefore, each gene value $g_{l,m}$ has an exact mapping to $\chi_{l,m}$, where χ is either UP ($\uparrow$) or DOWN ($\downarrow$), which are implemented as positive or negative values. These values give a direction to the mutation, which was previous in a random direction, and convert it from a scalar to a vector. Thus, the new organism structure is similar to Equation 5.11, with an extra underlying array.

$$\left[\begin{bmatrix} g_{1,1} & g_{1,2} & g_{1,3} & \cdots & g_{1,i} \\ g_{2,1} & g_{2,2} & g_{2,3} & \cdots & g_{2,j} \\ g_{3,1} & g_{3,2} & g_{3,3} & \cdots & g_{3,k} \end{bmatrix} \begin{bmatrix} \chi_{1,1} & \chi_{1,2} & \chi_{1,3} & \cdots & \chi_{1,i} \\ \chi_{2,1} & \chi_{2,2} & \chi_{2,3} & \cdots & \chi_{2,j} \\ \chi_{3,1} & \chi_{3,2} & \chi_{3,3} & \cdots & \chi_{3,k} \end{bmatrix} \right] \tag{5.16}$$

The benefit of this approach is that the mutation now has a direction, which will reduce the noise introduced by mutation, since it should trend toward the one which optimizes fitness. The direction of $\chi_{l,m}$ is evaluated by taking the

initial fitness, F_0 and displacing $g_{l,m}$ with a mutation δ:

$$\begin{cases} \chi_{l,m} & \frac{\partial F}{\partial g'_{l,m}} \geq 1 \\ -\chi_{l,m} & \frac{\partial F}{\partial g'_{l,m}} < 1. \end{cases} \tag{5.17}$$

Here, $g'_{l,m}$ is the new gene value after mutation. This equation shows that if the mutation reduces the fitness (i.e, the change is negative), the direction of the mutation is reversed. If the current direction improves the fitness, the direction remains. The other steps of the GA remain the same. Due to the extra evaluation of direction, however, the IGA converges much more slowly than the TGA or PGA. At the end of each generation, the most fit organism is moved to the master island for migration.

5.4.1 Migration Protocol

Unfortunately, migrating organisms from one population to another while ensuring genetic diversity is maximized is non-trivial, especially since the islands will generate organisms at different rates. Following the protocol established by Gozali [49] is shown in Figure 5.7, adapted from the same work.
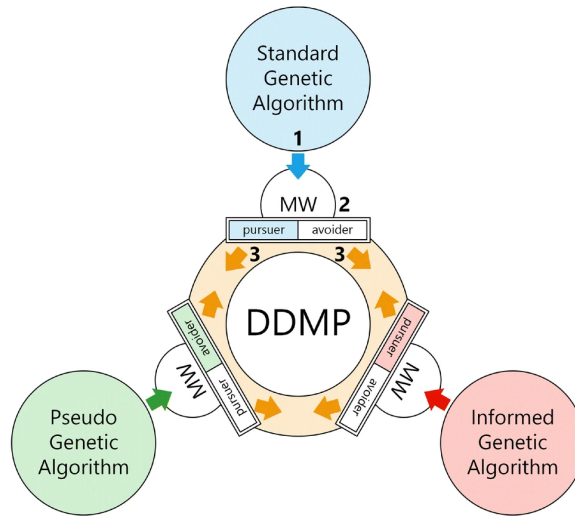

Figure 5.7: The Dual Migration Localized Island Model Genetic Algorithm.

The most significant difference from Figure 5.6 is the introduction of a "pursuer" and "avoider" distinction on the incoming migrants. Jan Paredis introduced such a distinction, where the pursuers seek to minimize the Hamming distance between itself and other organisms, while evaders seek to maximize the distance [96]. To determine which type a migrant is, two terms must be determined. The first is the bias value, inherited from forking genetic algorithms [127], which is a measure of the overall genetic diversity of a population. It is defined as

$$B_i^t = \frac{1}{NL} \sum_{l=1}^{N} \left(\left| \sum_{n=1}^{N} g_{i,n,l}^t - \frac{N}{2} \right| + \frac{N}{2} \right). \tag{5.18}$$

which is the bias of population i at generation t. N represents the number of organisms, L is the length of the chromosome, and $g_{i,n,l}$ is the gene value of the population i, organism n at index l. Above a threshold, θ, the migrating organisms will be designated an avoider. Below the threshold, it will be considered an evader. In such a case, the attractiveness of each island is instead calculated [28].

$$\alpha_i = \alpha_i^{prev} + (\eta_i^{pop} + \eta_i^{mig}) \; i = 1, 2, \ldots, I \tag{5.19}$$

where I is each island in the model, α_i^{prev} is the islands accumulated attractiveness from previous generations, and $\eta_i^{pop} + \eta_i^{mig}$ describe the attractiveness coefficients of the population and the migrant, respectively. The attractiveness is calculated for each migrant and island separately, and the

Thus, the steps can be as follows for migration. First, island i sends the best individual (m). Then, if there is more than one migrant in the pool, the bias value is calculated for the island. If the bias value is greater than θ, the avoider phase occurs. Here, all the migrants in the pool are compared to m. The one with the greatest Hamming distance is sent to island i. If instead the bias value is below θ, then the pursuer phase occurs. In the pursuer phase,

the attractiveness of all the islands and migrants are computed. The migrant which maximizes the attractiveness is sent to island i instead. Figure 5.8 shows the process.
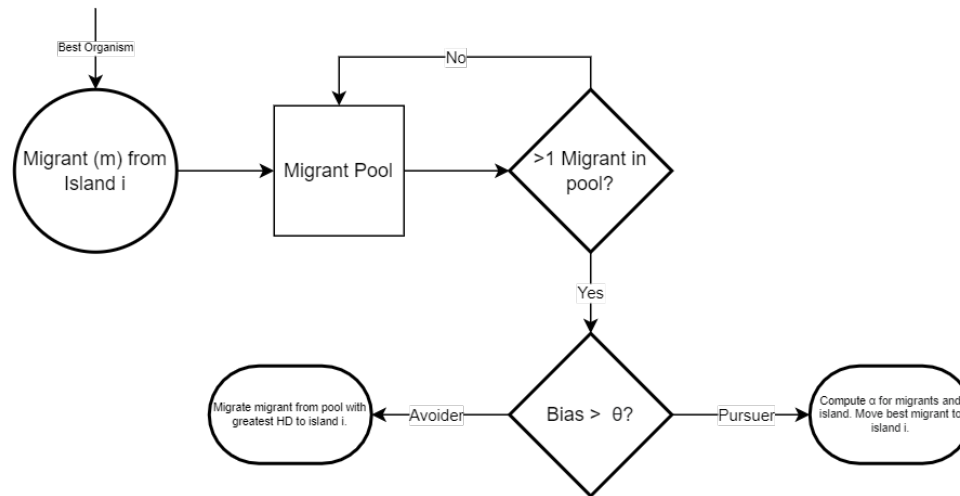

Figure 5.8: The DM-LIMGA migration protocol.

5.5 Results

The described genetic algorithm performs well, considering the extreme dimensionality of the problem. The program was benchmarked against known CCSSD(T)-F12/aug-cc-pVTZ QFF data.
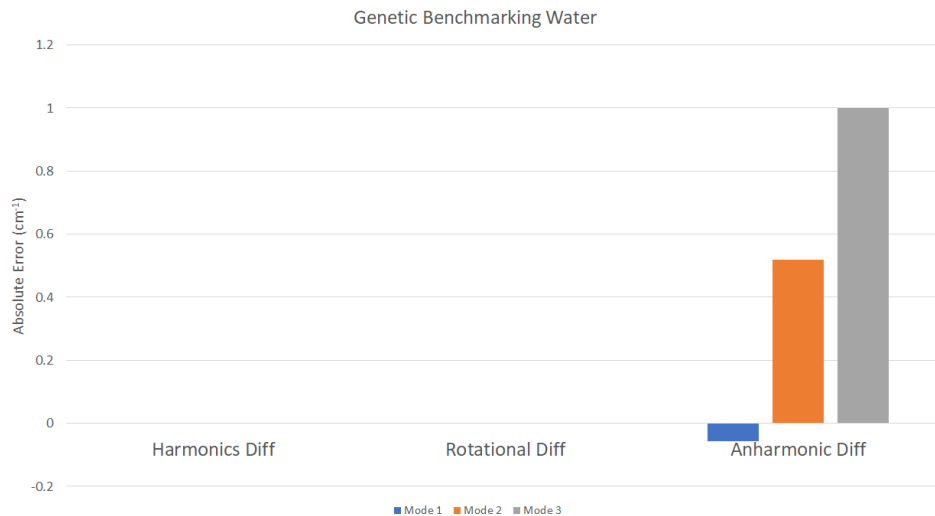
Figure 5.9: Benchmark against water QFFs.

Here, the harmonic frequencies and rotational constants are exactly recreated. The anharmonics have a maximum error of 1 cm^{-1} in the third mode, which is excellent agreement in 24 hours and 900 generations.

Difficulty is posed for larger systems, as the number of dimensions scales geometrically. Proposed solutions include principal component analysis, which can greatly reduce the dimensionality of the problem in exchange for slightly worse numerical solving [1] or the use of a combined Bayesian-GA method, which may be able to parameterize the GA before solving [19]. Greater parallelism is also a possible solution, such as utilizing GPU cores [7]. Nonetheless, a large amount of computational power will be necessary to solve larger systems, as this process is essentially the same as reversing cryptographic hashing.

The source code for this project can be found at `https://github.com/mvee18/ForceGenetics`.

Chapter 6

Conclusions

> And we walked out once more beneath the Stars.
>
> *Dante, Canto XXXIV*

Humanity is well-positioned to make vast advances in understanding questions that have long plagued us. What is our place among the stars? Can we understand the Universe? How can humanity spread beyond Earth and the solar system? In the last hundred years, humanity has made incredible advances in computational power that have allowed for groundbreaking discoveries. Yet, with each piece of unearthed knowledge, more lurks beneath. Once seemingly irrefutable precepts, such as classical mechanics, failed to explain the world at the microscopic level. It may come that tomorrow all of the theories unravel. Undaunted, however, humanity continues to make sense of the world and Universe surrounding us.

Science is a testament to the human spirit. The discoveries of today in the field of astrochemistry may never be applied for generations, but it is done (directly or indirectly) for some greater purpose. The confluence of experiment and theory due to the improvement of theory is a harbinger to the discoveries which will be made in the future. Science is not perfect, however, just as

humans have flaws, but this in no way diminishes its importance.

The most significant lesson I have learned, therefore, is the deeper meaning of research and science. It is a process which can question an individual's intuition and resilience greatly, though the reward is well-worth it. At the same time I was investigating molecules in space, I was also learning about myself through these projects.

Humanity is well-positioned to make great advances with the advent of the information age. It is my sincerest hope that we will take advantage of it.

Bibliography

[1] H. Abdi and L. J. Williams. Principal component analysis. *WIREs Computational Statistics*, 2(4):433–459, 2010. doi: https://doi.org/10.1002/wics.101. URL `https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wics.101`.

[2] A. Adewuya. New methods in genetic search with real-valued chromosomes. 08 2005.

[3] T. B. Adler, G. Knizia, and H.-J. Werner. A simple and efficient CCSD(T)-F12 approximation. *J. Chem. Phys.*, 127:221106, 2007.

[4] D. Agbaglo and R. C. Fortenberry. The performance of ccsd(t)-f12/aug-cc-pvtz for the computation of anharmonic fundamental vibrational frequencies. *Int. J. Quantum Chem.*, 119:e25899, 2019.

[5] D. Agbaglo and R. C. Fortenberry. The performance of explicitly correlated wavefunctions [ccsd(t)-f12b] in the computation of anharmonic vibrational frequencies. *Chem. Phys. Lett.*, 734:136720, 2019.

[6] D. Agbaglo, T. J. Lee, R. Thackston, and R. C. Fortenberry. A small molecule with pah vibrational properties and a detectable rotational spectrum: c-(c)c$_3$h$_2$, cyclopropenylidenyl carbene. *Astrophys. J.*, 871:236, 2019.

[7] E. Alba and M. Tomassini. Parallelism and evolutionary algorithms.

Evolutionary Computation, IEEE Transactions on, 6:443 – 462, 11 2002. doi: 10.1109/TEVC.2002.800880.

[8] W. D. Allen and coworkers, 2005. $INTDER$ 2005 is a General Program Written by W. D. Allen and Coworkers, which Performs Vibrational Analysis and Higher-Order Non-Linear Transformations.

[9] J. Andre, P. Siarry, and T. Dognon. An improvement of the standard genetic algorithm fighting premature convergence in continuous optimization. *Advances in Engineering Software*, 32(1):49–60, 2001. ISSN 0965-9978. doi: https://doi.org/10.1016/S0965-9978(00)00070-3. URL `https://www.sciencedirect.com/science/article/pii/S0965997800000703`.

[10] V. Babenko, G. Lane, A. A. Koos, A. T. Murdock, K. So, J. Britton, S. S. Meysami, J. Moffat, and N. Grobert. Time dependent decomposition of ammonia borane for the controlled production of 2d hexagonal boron nitride. *Sci. Rep.*, 7:14297, 2017.

[11] A. D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.

[12] M. Bitbol. *The Controversy between Schrödinger and the Göttingen-Copenhagen Physicists in the 1950's*, pages 1–33. Springer Netherlands, Dordrecht, 1996. ISBN 978-94-009-1772-9. doi: 10.1007/978-94-009-1772-9_1. URL `https://doi.org/10.1007/978-94-009-1772-9_1`.

[13] L. Bizzocchi, V. Lattanzi, J. Laas, S. Spezzano, B. M. Giuliano, D. Prudenzano, C. Endres, O. Sipilä, and P. Caselli. Accurate sub-millimetre rest frequencies for $HOCO^+$ and $DOCO^+$ ions. *Astron. Astrophys.*, 602: A34, 2017.

[14] A. Bookstein, V. A. Kulyukin, and T. Raita. Generalized hamming distance. *Information Retrieval*, 5(4):353–375, Oct 2002. ISSN 1573-7659. doi: 10.1023/A:1020499411651. URL https://doi.org/10.1023/A:1020499411651.

[15] J. A. Carr. An introduction to genetic algorithms. 2014.

[16] Q. Chen, Y. Zhong, and X. Zhang. A pseudo genetic algorithm. *Neural Computing and Applications*, 19(1):77–83, Feb 2010. ISSN 1433-3058. doi: 10.1007/s00521-009-0237-3. URL https://doi.org/10.1007/s00521-009-0237-3.

[17] I. Cherchneff and E. Dwek. The chemistry of population iii supernova ejecta. ii. the nucleation of molecular clusters as a diagnostic for dust in the early universe. *Astrophys. J.*, 713:1–24, 2010.

[18] E. Coe and L. B. Kass. Proof of physical exchange of genes on the chromosomes. *Proceedings of the National Academy of Sciences*, 102(19):6641–6646, 2005. doi: 10.1073/pnas.0407340102. URL https://www.pnas.org/doi/abs/10.1073/pnas.0407340102.

[19] C. Contaldi, F. Vafaee, and P. C. Nelson. Bayesian network hybrid learning using an elite-guided genetic algorithm. *Artificial Intelligence Review*, 52(1):245–272, Jun 2019. ISSN 1573-7462. doi: 10.1007/s10462-018-9615-5. URL https://doi.org/10.1007/s10462-018-9615-5.

[20] C. Darwin. *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. London : John Murray, 1859, 1859. URL https://search.library.wisc.edu/catalog/9934839413602122.

[21] De Broglie, Louis. Recherches sur la théorie des quanta. *Ann. Phys.*,

10(3):22–128, 1925. doi: 10.1051/anphys/192510030022. URL `https://doi.org/10.1051/anphys/192510030022`.

[22] B. L. de Vries, B. Acke, J. A. D. L. Blommaert, C. Waelkens, L. B. F. M. Waters, B. Vandenbussche, M. Min, G. Olofsson, C. Dominik, L. Decin, M. J. Barlow, A. Brandeker, J. Di Francesco, A. M. Glauser, J. Greaves, P. M. Harvey, W. S. Holland, R. J. Ivison, R. Liseau, E. E. Pantin, G. L. Pilbratt, P. Royer, and B. Sibthorpe. Comet-like mineralogy of olivine crystals in an extrasolar proto-kuiper belt. *Nature*, 490:74–76, 2012.

[23] N. J. DeYonker, T. R. Cundari, and A. K. Wilson. The correlation consistent composite approach (ccca): An alternative to the gaussian *n*-methods. *J. Chem. Phys.*, 124:114104, 2006.

[24] J. Dillen and P. Verhoeven. The end of 30-year-old controversy? a computational study of the b-n stretch frequency of bh_3-nh_3 in the solid state. *J. Phys. Chem. A*, 107:2570–2577, 2003.

[25] R. C. Dougal. The presentation of the planck radiation formula (tutorial). *Physics Education*, 11(6):438–443, sep 1976. doi: 10.1088/0031-9120/11/6/008. URL `https://doi.org/10.1088/0031-9120/11/6/008`.

[26] M. Douglas and N. Kroll. Quantum electrodynamical corrections to the fine structure of helium. *Ann. Phys.*, 82:89–155, 1974.

[27] K. M. Dreux, L. E. McNamara, J. T. Kelly, A. M. Wright, N. I. Hammer, and G. S. Tschumper. Probing dative and dihydrogen bonding in ammonia borane with electronic structure computations and raman under nitrogen spectroscopy. *J. Phys. Chem. A*, 121:5884–5893, 2017.

[28] G. Duarte, A. Lemonge, and L. Goliatt. A dynamic migration policy to the island model. In *2017 IEEE Congress on Evolutionary Computation (CEC)*, pages 1135–1142, 2017. doi: 10.1109/CEC.2017.7969434.

[29] T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90:1007–1023, 1989.

[30] T. H. Dunning, K. A. Peterson, and A. K. Wilson. Gaussian basis sets for use in correlated molecular calculations. x. the atoms aluminum through argon revisited. *J. Chem. Phys.*, 114:9244–9253, 2001.

[31] C. Eckart. Some studies concerning rotating axes and polyatomic molecules. *Phys. Rev.*, 47:552–558, Apr 1935. doi: 10.1103/PhysRev. 47.552. URL https://link.aps.org/doi/10.1103/PhysRev.47.552.

[32] B. Finney, R. C. Fortenberry, J. S. Francisco, and K. A. Peterson. A spectroscopic case for spsi detection: The third-row in a single molecule. *J. Chem. Phys.*, 145:124311, 2016.

[33] R. C. Fortenberry. Quantum astrochemical spectroscopy. *Int. J. Quant. Chem.*, 117:81–91, 2017.

[34] R. C. Fortenberry. Scf/hartree-fock theory, October 2021.

[35] R. C. Fortenberry and T. J. Lee. Computational vibrational spectroscopy for the detection of molecules in space. *Ann. Rep. Comput. Chem.*, 15: 173–202, 2019.

[36] R. C. Fortenberry, X. Huang, J. S. Francisco, T. D. Crawford, and T. J. Lee. Quartic force field predictions of the fundamental vibrational frequencies and spectroscopic constants of the cations hoco$^+$ and doco$^+$. *J. Chem. Phys.*, 136:234309, 2012.

[37] R. C. Fortenberry, X. Huang, T. D. Crawford, and T. J. Lee. Quartic force field rovibrational analysis of protonated acetylene, $C_2H_3{}^+$, and its isotopologues. *J. Phys. Chem. A*, 118:7034–7043, 2014.

[38] R. C. Fortenberry, T. J. Lee, and H. S. P. Müller. Excited vibrational level rotational constants for SiC_2: A sensitive molecular diagnostic for astrophysical conditions. *Molec. Astrophys.*, 1:13–19, 2015.

[39] P. R. Franke, J. F. Stanton, and G. E. Douberly. How to vpt2: Accurate and intuitive simulations of ch stretching infrared spectra using vpt2+k with large effective hamiltonian resonance treatments. *J. Phys. Chem. A*, 125:1301–1324, 2021.

[40] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian 16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.

[41] L. R. F.R.S. Liii. remarks upon the law of complete radiation. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 49(301):539–540, 1900. doi: 10.1080/14786440009463878. URL https://doi.org/10.1080/14786440009463878.

[42] P. E. R. F.R.S. Lxxix. the scattering of α and β particles by matter

and the structure of the atom. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 21(125):669–688, 1911. doi: 10.1080/14786440508637080. URL `https://doi.org/10.1080/14786440508637080`.

[43] H.-P. Gail and E. Sedlmayr. Mineral formation in stellar winds i. condensation sequence of silicate and iron grains in stationary oxygen rich outflows. *Astron. Astrophys.*, 347:594–616, 1999.

[44] J. F. Gaw, A. Willets, W. H. Green, and N. C. Handy. SPECTRO: A program for the derivation of spectrscopic constants from provided quartic force fields and cubic dipole fields. In J. M. Bowman and M. A. Ratner, editors, *Advances in Molecular Vibrations and Collision Dynamics*, pages 170–185. JAI Press, Inc., Greenwich, Connecticut, 1991.

[45] J. F. Gaw, A. Willets, W. H. Green, and N. C. Handy, 1996. *SPECTRO program*, version 3.0.

[46] D. Gobrecht, I. Cherchneff, A. Sarangi, J. M. C. Plane, and S. T. Bromley. Dust formation in the oxygen-rich agb star ik tauri. *Astron. Astrophys.*, 585:A6, 2016.

[47] J. Z. Gong, D. A. Matthews, P. B. Changala, and J. F. Stanton. Fourth-order vibrational perturbation theory with the watson hamiltonian: Report of working equations and preliminary results. *The Journal of Chemical Physics*, 149(11):114102, 2018. doi: 10.1063/1.5040360. URL `https://doi.org/10.1063/1.5040360`.

[48] A. A. Gozali and S. Fujimura. Performance analysis of localization strategy for island model genetic algorithm. In *2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pages 425–430, 2017. doi: 10.1109/SNPD.2017.8022757.

[49] A. A. Gozali and S. Fujimura. Dm-limga: Dual migration localized island model genetic algorithm—a better diversity preserver island model. *Evolutionary Intelligence*, 12(4):527–539, Dec 2019. ISSN 1864-5917. doi: 10.1007/s12065-019-00253-2. URL `https://doi.org/10.1007/s12065-019-00253-2`.

[50] T. R. Gregory. Understanding natural selection: Essential concepts and common misconceptions. *Evolution: Education and Outreach*, 2(2):156–175, Jun 2009. ISSN 1936-6434. doi: 10.1007/s12052-009-0128-1. URL `https://doi.org/10.1007/s12052-009-0128-1`.

[51] A. Hassanat, K. Almohammadi, E. Alkafaween, E. Abunawas, A. Hammouri, and V. B. S. Prasath. Choosing mutation and crossover ratios for genetic algorithms—a review with a new dynamic approach. *Information*, 10(12), 2019. ISSN 2078-2489. doi: 10.3390/info10120390. URL `https://www.mdpi.com/2078-2489/10/12/390`.

[52] R. L. Haupt and S. E. Haupt. *Practical genetic algorithms*. Wiley, 2004.

[53] D. O. Hayward. *Quantum Mechanics for Chemists*. Royal Society of Chemistry, 2002.

[54] R. M. Hazen. Effects of temperature and pressure on the crystal structure of forsterite. *Am. Mineral.*, 61:1280–1293, 1976.

[55] W. J. Hehre, R. Ditchfeld, and J. A. Pople. Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.*, 56: 2257, 1972.

[56] N. J. Hess, M. E. Bowden, V. M. Parvanov, C. Mundy, S. M. Kathmann, G. K. Schenter, and T. Autrey. Spectroscopic studies of the phase transition in ammonia borane: Raman spectroscopy of single crystal nh_3bh_3 as

a function of temperature from 88 to 330k. *J. Chem. Phys.*, 128:034508, 2008.

[57] J. G. Hill and K. A. Peterson. Correlation consistent basis sets for explicitly correlated wavefunctions: Valence and core-valence basis sets for li, be, na, and mg. *Phys. Chem. Chem. Phys.*, 12:10460–10468, 2010.

[58] J. G. Hill, S. Mazumder, and K. A. Peterson. Correlation consistent basis sets for molecular core-valence effects with explicitly correlated wave functions: The atoms b-ne and al-ar. *J. Chem. Phys.*, 132:054108, 2010.

[59] J. H. Holland. Genetic algorithms. *Scientific American*, 267(1):66–73, 1992. ISSN 00368733, 19467087. URL `http://www.jstor.org/stable/24939139`.

[60] X. Huang and T. J. Lee. A procedure for computing accurate *ab initio* quartic force fields: Application to HO_2^+ and H_2O. *J. Chem. Phys.*, 129: 044312, 2008.

[61] X. Huang and T. J. Lee. Accurate *ab initio* quartic force fields for NH_2^- and CCH^- and rovibrational spectroscopic constants for their isotopologs. *J. Chem. Phys.*, 131:104301, 2009.

[62] X. Huang, E. F. Valeev, and T. J. Lee. Comparison of one-particle basis set extrapolation to explicitly correlated methods for the calculation of accurate quartic force fields, vibrational frequencies, and spectroscopic constants: Application to h_2o, n_2h^+, no_2^+, and c_2h_2. *J. Chem. Phys.*, 133:244108, 2010.

[63] X. Huang, P. R. Taylor, and T. J. Lee. Highly accurate quartic force field, vibrational frequencies, and spectroscopic constants for cyclic and linear $C_3H_3^+$. *J. Phys. Chem. A*, 115:5005–5016, 2011.

[64] X. Huang, R. C. Fortenberry, and T. J. Lee. Protonated nitrous oxide, NNOH$^+$: Fundamental vibrational frequencies and spectroscopic constants from quartic force fields. *J. Chem. Phys.*, 139(8):084313, 2013.

[65] C. Jäger, F. Molster, J. Dorschner, T. Henning, H. Mutschke, and L. Waters. Toward interstellar silicate mineralogy iv. the crystalline revolution. *Astron. Astrophys.*, 339:904–916, 1998.

[66] B. B. Karki. First-principles computation of mantle materials in crystalline and amorphous phases. *Phys. Earth Planet. Interiors*, 240:43–69, 2015.

[67] R. A. Kendall, T. H. Dunning, and R. J. Harrison. Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions. *J. Chem. Phys.*, 96:6796–6806, 1992.

[68] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948 vol.4, 1995. doi: 10.1109/ICNN.1995.488968.

[69] Y. Kimura and J. A. Nuth III. A seed of solar forsterite and possible new evolutional scenario of cosmic silicates. *Astrophys. J.*, 697:L10–L13, 2009.

[70] M. J. R. Kitchens and R. C. Fortenberry. The rovibrational nature of closed-shell third-row triatomics: HOX and HXO, X = Si$^+$, P, S$^+$, and Cl. *Chem. Phys.*, 472:119–127, 2016.

[71] K. A. Kloska and R. C. Fortenberry. Gas-phase spectra of mgo molecules: A possible connection from gas-phase molecules to planet formation. *Mon. Not. Royal Astron. Soc.*, 474:2055–2063, 2018.

[72] R. D. Knight. *Physics for scientists and engineers: A strategic approach.* Pearson/Addison Wesley, 2004.

[73] G. Knizia, T. B. Adler, and H.-J. Werner. Simplified CCSD(T)-F12 Methods: Theory and benchmarks. *J. Chem. Phys.*, 130:054104, 2009.

[74] M. Komatsu, T. J. Fagan, A. N. Krot, K. Nagashima, M. I. Petaev, M. Kimura, and A. Yamaguchi. First evidence for silica condensation within the solar protoplanetary disk. *Proc. Nat. Acad. Sci. USA*, 115: 7497–7502, 2018.

[75] C. Lee, W. T. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B.*, 37:785–789, 1988.

[76] I. N. Levine. *Quantum Chemistry.* Pearson, 2014.

[77] A. Macià Escatllar, T. Lazaukas, S. M. Woodley, and S. T. Bromley. Structure and properties of nanosilicates with olivine $(mg_2sio_4)_n$ and pyroxene $(mgsio_3)_n$ compositions. *ACS Earth Space Chem.*, 2019.

[78] J. M. L. Martin and P. R. Taylor. Accurate *ab initio* quartic force field for *trans*-hnnh and treatment of resonance polyads. *Spectrochim. Acta A*, 53:1039–1050, 1997.

[79] J. M. L. Martin, T. J. Lee, P. R. Taylor, and J.-P. François. The anharmonic force field of ethylene, c_2h_4, by means of accurate *ab initio* calculations. *J. Chem. Phys.*, 103(7):2589–2602, 1995.

[80] C. M. Mauney and D. Lazzati. The formation of astrophysical mg-rich silicate dust. *Mol. Astrophys.*, 12:1–9, 2018.

[81] E. B. Mayr. *The growth of biological thought: Diversity, evolution, and inheritance.* Belknap Press, 2003.

[82] M. C. McCarthy and P. Thaddeus. Microwave and laser spectroscopy of carbon chains and rings. *Chem. Soc. Rev.*, 30:177–185, 2001.

[83] B. A. McGuire. 2018 census of interstellar, circumstellar, extragalactic, protoplanetary disk, and exoplanetary molecules. *Astrophys. J. Suppl. Ser.*, 239:17, 2018.

[84] B. A. McGuire, A. M. Burkhardt, S. Kalenskii, C. N. Shingledecker, A. J. Remijan, E. Herbst, and M. C. McCarthy. Detection of the aromatic molecule benzonitrile (c-c_6h_5cn) in the interstellar medium. *Science*, 359: 202–205, 2018.

[85] B. L. Miller and D. E. Goldberg. Genetic algorithms, tournament selection, and the effects of noise. *Complex Syst.*, 9, 1995.

[86] I. M. Mills. Vibration-rotation structure in asymmetric- and symmetric-top molecules. In K. N. Rao and C. W. Mathews, editors, *Molecular Spectroscopy - Modern Research*, pages 115–140. Academic Press, New York, 1972.

[87] M. Mitchell. Genetic algorithms: An overview. *Complexity*, 1(1):31–39, 1995. doi: https://doi-org.umiss.idm.oclc.org/10.1002/cplx.6130010108. URL `https://onlinelibrary-wiley-com.umiss.idm.oclc.org/doi/abs/10.1002/cplx.6130010108`.

[88] R. M.Killen, A. E. Potter, R. J. Vervack Jr., E. T. Bradley, W. E. Mc-Clintock, C. M.Anderson, and M. H. Burgera. Observations of metallic species in mercury's exosphere. *Icarus*, 209:75–87, 2010.

[89] C. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46:618–622, 1934.

[90] F. J. Molster, T. L. Lim, R. J. Sylvester, L. B. F. M. Waters, M. J. Barlow, D. A. Beintema, M. Cohen, P. Cox, and B. Schmitt. The complete iso spectrum of ngc 6302. *Astron. Astrophys.*, 372:165–172, 2001.

[91] NobelPrize.org. The nobel prize in chemistry 1998, . URL `https://www.nobelprize.org/prizes/chemistry/1998/summary/`.

[92] NobelPrize.org. The nobel prize in chemistry 2013, . URL `https://www.nobelprize.org/prizes/chemistry/2013/summary/`.

[93] C. Z. Palmer and R. C. Fortenberry. Rovibrational considerations for the monomers and dimers of magnesium hydride (MgH_2) and magnesium fluoride (MgF_2). *J. Phys. Chem. A*, 122:7079–7088, 2018.

[94] A. Paolone, F. Teocoli, S. Sanna, O. Palumbo, and T. Autrey. Temperature dependence of the infrared spectrum of ammonia borane: Librations, rotations, and molecular vibrations. *J. Phys. Chem. C*, 117: 729–734, 2013.

[95] D. Papousek and M. R. Aliev. *Molecular Vibration-Rotation Spectra*. Elsevier, Amsterdam, 1982.

[96] J. Paredis. Exploring the evolution of genotype phenotype mappings. In *2017 IEEE Congress on Evolutionary Computation (CEC)*, pages 279–285, 2017. doi: 10.1109/CEC.2017.7969324.

[97] A. B. C. Patzer, C. Chang, M. John, U. Bolick, and D. Sülzle. Theoretical study of stationary points of the mgsio$_3$ molecule. *Chem. Phys. Lett.*, 363: 145–151, 2002.

[98] B. Peng and J. Chen. Ammonia borane as an efficient and lightweight hydrogen storage medium. *Energ. Environ. Sci.*, 4:479–483, 2008.

[99] K. A. Peterson and T. H. Dunning. Benchmark calculations with correlated molecular wave functions. vii. binding energy and structure of the hf dimer. *J. Chem. Phys.*, 102:2032–2041, 1995.

[100] K. A. Peterson, D. E. Woon, and T. H. Dunning, Jr. Benchmark calculations with correlated molecular wave functions. iv. the classical barrier height of the h+h$_2$→h$_2$+h reaction. *J. Chem. Phys.*, 100:7410–7415, 1994.

[101] K. A. Peterson, T. B. Adler, and H.-J. Werner. Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms H, He, B-Ne, and Al-Ar. *J. Chem. Phys.*, 128:084102, 2008.

[102] M. Planck, M. v. Laue, and F. Gaynor. *Scientific autobiography, and other papers.* Citadel Press Incorporated, 1949.

[103] J. M. C. Plane, J. D. Carrillo-Sanchez, T. P. Mangan, M. M. J. Crismani, N. M. Schneider, and A. Määttänen. Meteoric metal chemistry in the martian atmosphere. *J. Geophys. Res.: Planets*, 123:695–707, 2018.

[104] C. A. Poteet, S. T. Megeath, D. M. Watson, N. Calvet, I. S. Remming, M. K. McClure, B. A. Sargent, W. J. Fischer, E. Furlan, L. E. Allen, J. E. Bjorkman, L. Hartmann, J. Muzerolle, J. J. Tobin, and B. Ali. A *Spitzer* spectrograph detection of crystalline silicates in a protostellar envelope. *Astrophys. J. Lett.*, 733:32, 2011.

[105] G. D. Purvis and R. J. Bartlett. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.*, 76:1910–1918, 1982.

[106] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.*, 157:479–483, 1989.

[107] K. Raghavachari, G. W. Trucks, J. A. Pople, and E. Replogle. Highly correlated systems: Structure, binding energy and harmonic vibrational frequencies of ozone. *Chem. Phys. Lett.*, 158:207–212, 1989.

[108] R. O. Ramabhadran and K. Raghavachari. Extrapolation to the gold-standard in quantum chemistry: Computationally efficient and accurate ccsd(t) energies for large molecules using an automated thermochemical hierarchy. *Journal of Chemical Theory and Computation*, 9(9):3986–3994, 2013. doi: 10.1021/ct400465q. URL `https://doi.org/10.1021/ct400465q`. PMID: 26592394.

[109] H. H. Rosenbrock. An Automatic Method for Finding the Greatest or Least Value of a Function. *The Computer Journal*, 3(3):175–184, 01 1960. ISSN 0010-4620. doi: 10.1093/comjnl/3.3.175. URL `https://doi.org/10.1093/comjnl/3.3.175`.

[110] D. M. Roy and R. Roy. Synthesis and stability of the minerals in the system mgo-al$_2$o$_3$-sio$_2$-h$_2$o. *Am. Mineral.*, 40:147–178, 1955.

[111] G. Rudolph. Convergence analysis of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, 5(1):96–101, 1994. doi: 10.1109/72.265964.

[112] J. R. Rumble, editor. *CRC Handbook of Chemistry and Physics*. CRC Press, Boca Raton, 99 edition, 2018.

[113] R. L. Sams, S. S. Xantheas, and T. A. Blake. Vapor phase infrared spectroscopy and ab initio fundamental anharmonic frequencies of ammonia borane. *J. Phys. Chem. A*, 116:3124–3136, 2012.

[114] D. Sherill. An introduction to hartree-fock molecular orbital-theory. `http://vergil.chemistry.gatech.edu/notes/hf-intro/hf-intro.html`, June 2000.

[115] Y. Shi and R. Eberhart. A modified particle swarm optimizer. In *1998 IEEE International Conference on Evolutionary Computation Pro-*

ceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360), pages 69–73, 1998. doi: 10.1109/ICEC.1998.699146.

[116] I. W. M. Smith and R. Zellner. Rate measurements fo reactions of oh by resonance absorption. part 2: Reactions of oh with co with c_2h_4 and c_2h_2. *J. Chem. Soc. Faraday Trans.*, 69:1617–1627, 1973.

[117] S. J. Smith and B. T. Sutcliffe. *The Development of Computational Chemistry in the United Kingdom*, pages 271–316. John Wiley & Sons, Ltd, 1996. ISBN 9780470125878. doi: https://doi.org/10.1002/ 9780470125878.ch5. URL `https://onlinelibrary.wiley.com/doi/ abs/10.1002/9780470125878.ch5`.

[118] A. K. Speck, A. G. Whittington, and A. M. Hofmeister. Disordered silicates in space: A study of laboratory spectra of "amorphous" silicates. *Astrophys. J.*, 740:93, 2011.

[119] A. Staubitz, A. P. M. Robertson, and I. Manners. Ammonia-borane and related compounds as dihydrogen sources. *Chem. Rev.*, 110:4079–4124, 2010.

[120] P. Stehle. *Order, Chaos, Order: The Transition from Classical to Quantum Physics.* Oxford University Press, 1949.

[121] F. H. Stephens, V. Pons, and R. T. Baker. Ammonia-borane: The hydrogen source *par excellence. Dalton Trans.*, 25:2613–2626, 2007.

[122] R. D. Suenram and L. R. Thorne. Microwave spectrum and dipole moment of bh_3nh_3. *Chem. Phys. Lett.*, 78:157–160, 1981.

[123] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory.* Dover, Mineola, NY, 1996.

[124] R. Thackston and R. C. Fortenberry. An efficient algorithm for the determination of force constants and displacements in numerical definitions

of a large, general order taylor series expansion. *J. Math. Chem.*, 56: 103–119, 2018.

[125] C. Thomas. *Introduction to algorithms.* MIT Press, 2003.

[126] L. R. Thorne, R. D. Suenram, and F. J. Lovas. Microwave spectrum, torsional barrier, and structure of bh$_3$nh$_3$. *J. Chem. Phys.*, 78:167–171, 1983.

[127] S. Tsutsui, Y. Fujimoto, and A. Ghosh. Forking Genetic Algorithms: GAs with Search Space Division Schemes. *Evolutionary Computation*, 5 (1):61–80, 03 1997. ISSN 1063-6560. doi: 10.1162/evco.1997.5.1.61. URL https://doi.org/10.1162/evco.1997.5.1.61.

[128] E. M. Valencia, C. J. Worth, and R. C. Fortenberry. Enstatite (mgsio$_3$) and forsterite (mg$_2$sio$_4$) monomers and dimers: Highly-detectable infrared and radioastronomical molecular building blocks. *Mon. Not. Royal Astron. Soc.*, 492:276–282, 2019.

[129] L.-S. Wang, S. R. Desai, H. Wu, and J. B. Nicholas. Small silicon oxide clusters: Chains and rings. *Z. Phys. D*, 40:36–39, 1997.

[130] J. K. Watson. Simplification of the molecular vibration-rotation hamiltonian. *Molecular Physics*, 15(5):479–490, 1968. doi: 10.1080/00268976800101381. URL https://doi.org/10.1080/00268976800101381.

[131] J. K. G. Watson. Aspects of quartic and sextic centrifugal effects on rotational energy levels. In J. R. During, editor, *Vibrational Spectra and Structure*, pages 1–89. Elsevier, Amsterdam, 1977.

[132] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz. Molpro: A General-Purpose Quantum Chemistry Program Package. *WIREs Comput. Mol. Sci.*, 2:242–253, 2012.

[133] H.-J. Werner, P. J. Knowles, F. R. Manby, J. A. Black, K. Doll, A. Heßel-mann, D. Kats, A. Köhn, T. Korona, D. A. Kreplin, Q. Ma, T. F. Miller, A. Mitrushchenkov, K. A. Peterson, I. Polyak, G. Rauhut, and M. Sibaev. Molpro, version 2020.1, a package of *ab initio* programs. *WIREs Comput. Mol. Sci.*, 2020. see http://www.molpro.net.

[134] B. R. Westbrook and R. C. Fortenberry. Anharmonic frequencies of $(MO)_2$ & related hydrides for M = Mg, Al, Si, P, S, Ca, & Ti and heuristics for predicting anharmonic corrections of inorganic oxides. *J. Phys. Chem. A*, 124:3191–3204, 2020.

[135] B. R. Westbrook, E. M. Valencia, S. C. Rushing, G. S. Tschumper, and R. C. Fortenberry. Anharmonic vibrational frequencies of ammonia borane (bh_3nh_3). *J. Chem. Phys*, 154(4):041104, 2021.

[136] A. K. Wilson, T. van Mourik, and T. H. Dunning. Gaussian basis sets for use in correlated moleculear calculations. vi. sextuple zeta correlation consistent basis sets for boron through neon. *J. Mol. Struct.*, 388:339–349, 1996.

[137] D. A. Woon and T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. v. core-valence basis sets for boron through neon. *J. Chem. Phys.*, 103:4572–4585, 1995.

[138] W. T. Yang, R. G. Parr, and C. T. Lee. Various functionals for the kinetic energy density of an atom or molecule. *Phys. Rev. A*, 34:4586–4590, 1986.

[139] A. A. Yaroshevsky and A. G. Bulakh. The mineral composition of the earth's crust, mantle, meteorites, moon, and planets. In A. S. Marfunin, editor, *Advanced Mineralogy*, pages 27–36. Springer, Berlin, 1994.

[140] Q. Yu, J. M. Bowman, R. C. Fortenberry, J. S. Mancini, T. J. Lee, T. D. Crawford, W. Klemperer, and J. S. Francisco. The structure, anharmonic

vibrational frequencies, and intensities of nnhnn$^+$. *J. Phys. Chem. A*, 119: 11623–11631, 2015.

[141] L. Zamirri, A. Macià Escatllar, J. Mariñoso Guiu, P. Ugliengo, and S. T. Bromley. What can infrared spectra tell us about the crystallinity of nanosized interstellar silicate dust grains? *ACS Earth Space Chem.*, 2019.

[142] D. Zhao, K. D. Doney, and H. Linnartz. Laboratory gas-phase detection of the cyclopropenyl cation (c-C$_3$H$_3{}^+$). *Astrophys. J. Lett.*, 791:L28, 2014.