Using polarized Raman spectroscopy and the pseudospectral method to characterize molecular structure and function

Andrew L. Weisman

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

ABSTRACT

Using polarized Raman spectroscopy and the pseudospectral method to characterize molecular structure and function

Andrew L. Weisman

Electronic structure calculation is an essential approach for determining the structure and function of molecules and is therefore of critical interest to physics, chemistry, and materials science. Of the various algorithms for calculating electronic structure, the pseudospectral method is among the fastest. However, the trade-off for its speed is more up-front programming and testing, and as a result, applications using the pseudospectral method currently lag behind those using other methods.

In Part I of this dissertation, we first advance the pseudospectral method by optimizing it for an important application, polarized Raman spectroscopy, which is a well-established tool used to characterize molecular properties. This is an application of particular importance because often the easiest and most economical way to obtain the polarized Raman spectrum of a material is to simulate it; thus, utilization of the pseudospectral method for this purpose will accelerate progress in the determination of molecular properties. We demonstrate that our implementation of Raman spectroscopy using the pseudospectral method results in spectra that are just as accurate as those calculated using the traditional analytic method, and in the process, we derive the most comprehensive formulation to date of polarized Raman intensity formulas, applicable to both crystalline and isotropic systems.

Next, we apply our implementation to determine the orientations of crystalline oligothiophenes — a class of materials important in the field of organic electronics — achieving excellent agreement with experiment and demonstrating the general utility of polarized Raman spectroscopy for the determination of crystal orientation. In addition, we derive from first-principles a method for using polarized Raman spectra to establish unambiguously whether a uniform region of a material is crystalline or isotropic. Finally, we introduce free, open-source software that allows a user to determine any of a number of polarized Raman properties of a sample given common output from electronic structure calculations.

In Part II, we apply the pseudospectral method to other areas of scientific importance requiring a deeper understanding of molecular structure and function. First, we use it to accurately determine the frequencies of vibrational tags on biomolecules that can be detected in real-time using stimulated Raman spectroscopy. Next, we

evaluate the performance of the pseudospectral method for calculating excited-state energies and energy gradients of large molecules — another new application of the pseudospectral method — showing that the calculations run much more quickly than those using the analytic method.

Finally, we use the pseudospectral method to simulate the bottleneck process of a solar cell used for water splitting, a promising technology for converting the sun's energy into hydrogen fuel. We apply the speed of the pseudospectral method by modeling the relevant part of the system as a large, explicitly passivated titanium dioxide nanoparticle and simulating it realistically using hybrid density functional theory with an implicit solvent model, yielding insight into the physical nature of the rate-limiting step of water splitting. These results further validate the particularly fast and accurate simulation methodologies used, opening the door to efficient and realistic cluster-based, fully quantum-mechanical simulations of the bottleneck process of a promising technology for clean solar energy conversion.

Taken together, we show how both polarized Raman spectroscopy and the pseudospectral method are effective tools for analyzing the structure and function of important molecular systems.

Contents

List of Figures

List of Tables

Acknowledgments

I am forever grateful to my advisor, Professor Richard Friesner, for six enriching years of guidance, patience, and support. Under his tutelage, I gained more out of my graduate work than I could have ever imagined. I am honored to have been his student. I would also like to thank my wife, Laura Kaplan-Weisman, for her unremitting encouragement, patience, and love. We went through it all together, and I could not have done it without her by my side. In addition, I am thankful for my incredible labmates. On even the most stressful days, they were a source of warmth and comfort, and I am incredibly lucky to have been a part of such an impressive and fun group of people. I regret the times that I worked from home! I would also like to thank all my collaborators I have had the opportunity to work with and learn from. They helped me grow intellectually and made working fun! Finally, I am thankful to my parents, grandparents, family, and friends for their support and love, both inside and outside of my studies, that has kept me motivated and happy throughout my time in graduate school.

To my beloved grandparents

Part I

Polarized Raman Spectroscopy and the Pseudospectral Method

In the first part of this dissertation, the theory and application of polarized Raman spectroscopy is discussed, both in a general context and in the context of its simulation using the pseudospectral method. First, a set of polarized, spontaneous, vibrational Raman intensity formulas that apply to both isotropic and crystalline systems is presented (Chapter 1). This is the most general and comprehensive formulation to date. These formulas provide the basis for a brief discussion of some common theoretical and experimental shortcomings of Raman spectroscopy. Then, an efficient formulation of the energy third derivative required for calculating Raman spectra using the pseudospectral method, an algorithm for quickly calculating the two-electron integrals required in electronic structure calculations, is presented (Chapter 2). It is demonstrated that this methodology produces spectra that are just as accurate as those obtained using the traditional analytic method for calculating the two-electron integrals, thus extending the general speed advantages of the pseudospectral method to calculations of Raman intensities.

Next, this implementation is applied by using density functional theory to characterize the orientations of crystalline oligothiophenes, which approximate the longer polythiophene chains that are used in organic electronics (Chapter 3). Excellent agreement with experimental polarized Raman measurements is shown. In addition, a simulation explicitly demonstrating the utility of polarized Raman signals for determining the precise orientations of crystal unit cells is presented. Then, a general, first-principles method for determining unambiguously whether a uniform region of a material is crystalline or isotropic is introduced (Chapter 4). This utilizes a special polarized Raman signal, in which, under particular experimental conditions, dependence of the signal on all other system and experimental variables drops out. It is further shown how this signal can be used to detect region boundaries in a heterogeneous sample and also to determine whether an arbitrary vibrational mode of a system is degenerate. Finally, open-source software allowing a user to determine any of a number of polarized Raman properties of a sample given common output from electronic structure calculations is introduced (Chapter 5).

Chapter 1

Complete formulation of polarized Raman intensities

1.1 Introduction

In this chapter we first derive spontaneous, non-resonance, vibrational Raman intensity formulas that are at least as general as any we have found in literature. While the derivations resemble those derived in [1] and [2], the following differ by treating the formulas for scattering from oriented and isotropic materials on equal footing, deriving the isotropic formulas only by applying isotropic averaging to the oriented formulas. Further, we apply the most-general description of the polarization properties of the incident and scattered light by using the Stokes vector and Mueller matrix formalism detailed in [3], resulting in Raman (and Rayleigh) intensity formulas that apply for arbitrary values of the three sets of experimental parameters involved: incident light of arbitrary polarization, arbitrary scattering angle, and detected light of arbitrary polarization. The utility of these general formulas will be demonstrated in subsequent chapters. This complete formulation further enables a clear understanding of why common experimental Raman setups and Raman intensity formulas are limited in scope and therefore fail to take advantage of the full breadth of the Raman effect, as explained in Section 1.9.

1.2 Overview of Raman theory

The Raman effect is not a special phenomenon but is rather a general type of inelastic scattering of light by matter that can occur when the wavelength of the incident light is much larger than the scatterer, which for

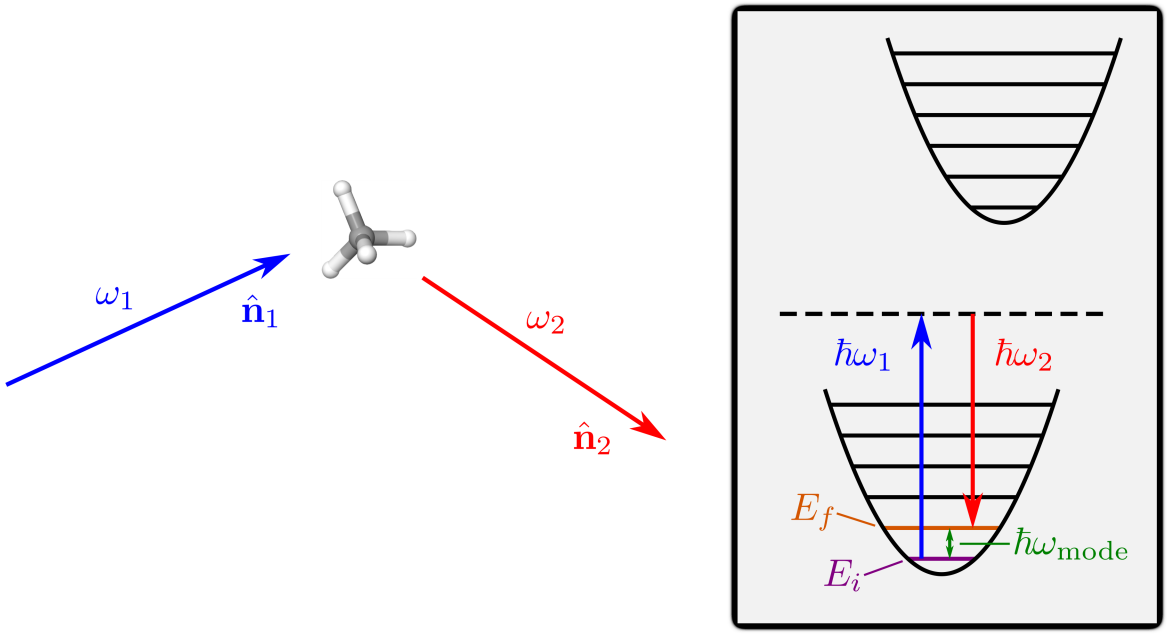

Figure 1.1: Spontaneous Stokes Raman scattering. At left is the incident photon of frequency ω_1 traveling in direction $\hat{\mathbf{n}}_1$, the molecule that scatters the photon, and the scattered photon of frequency ω_2 traveling in direction $\hat{\mathbf{n}}_2$. At right is a schematic diagram of the energy levels of the scatterer. The ground state electronic manifold is depicted by the bottom parabola, and an excited electronic manifold is depicted by the top parabola. Each parabola represents the harmonic approximation to the potential well of the scatterer's vibrational coordinate corresponding to the frequency ω_{mode}. In this approximation, the energy levels are equally spaced by $\hbar\omega_{\text{mode}}$. The diagram depicts the scatterer starting in the vibrational ground state of energy E_i, virtually absorbing the incident photon of energy $\hbar\omega_1$, immediately emitting a photon of energy $\hbar\omega_2$, and ending up in the first excited vibrational state of energy E_f. The dotted line indicates that the intermediate energy level is a virtual energy level, and it is shown well-removed from the top of the ground state electronic manifold and from the bottom of the excited state electronic manifold, as is the case for spontaneous Raman scattering.

an isotropic region of a sample is a molecule and for a crystalline region is a unit cell. The most common type of Raman scattering is spontaneous, non-resonance, vibrational Raman scattering and is characterized by the following quantum mechanical process, illustrated in Fig. 1.1:

1. A photon of energy $\hbar\omega_1$ traveling in direction $\hat{\mathbf{n}}_1$ is incident upon a scatterer that is in an arbitrary vibrational stationary state i of energy E_i.

2. The photon is destroyed by the scatterer and simultaneously the scatterer becomes excited to a virtual state of energy $E_i + \hbar\omega_1$, which is much lower in energy than any excited electronic state (non-resonance case) but is much higher in energy than any vibrational or rotational absorption energy; the photon is not

absorbed.[1]

3. The scatterer decays from the virtual state of energy $E_i + \hbar\omega_1$ to a stationary state f of energy E_f and simultaneously creates a photon of energy

$$\begin{aligned}(E_i + \hbar\omega_1) - E_f &= \hbar\omega_1 - (E_f - E_i) \\ &\equiv \hbar\omega_1 - \Delta E \\ &\equiv \hbar\omega_2.\end{aligned}$$

4. The new photon of energy $\hbar\omega_2$ leaves the scatterer in direction $\hat{\mathbf{n}}_2$.

Throughout the process, we assume that the scatterer remains in its non-degenerate ground electronic state and no magnetic perturbations are present (resulting in real wavefunctions[1]), and the vibrational energy level of the scatterer increases or decreases by one quantum. Such a process is very realistically approximated in experiment[1]. It should be noted that this process encompasses Rayleigh scattering as well, in which case $E_f = E_i$.

This Raman scattering process is important because by studying the properties of the scattered photons such as their frequencies and corresponding intensities — i.e., the Raman spectrum — one can learn about the properties of the scatterers themselves, as we will demonstrate in subsequent chapters. While the frequency of the scattered light provides important information about the quantum energy levels of the scatterers, the intensity of the scattered light — in particular, the intensity of particular polarizations relative to the scattering plane defined by $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$ — provides additional important information about the scattering material. Furthermore, the frequencies of the scattered light cannot be utilized at all if their corresponding intensities are zero or unobservable to begin with. Thus, in order to make definitive conclusions about the scattering system utilizing the Raman effect, one must be able to calculate the expected Raman intensities.

In order to calculate the Raman intensity of the scattered light in step 4, one needs a physical model describing the interaction of the scatterer and the two photons in steps 2 and 3. The model that has been shown over time to accurately describe this interaction is the multipole interaction model[1]. In general, in the multipole interaction model, an incident photon ω_1 interacts with some intrinsic aspect of a scatterer (e.g., its normal vibrations), leaving

[1] In steps 2 and 3, the scatterer does not actually occupy the virtual state; the time spent in the virtual state is so short that during this time there is a large uncertainty in the scatterer's energy, according to the uncertainty principle $\delta E \delta t \geq \hbar/2$. Thus, even though one photon is annihilated and one photon is created by the scatterer, the scatterer's energy does not change by the corresponding opposite amount; energy is not conserved during either of steps 2 or 3. However, energy is conserved *overall*; the scatterer's energy changes by ΔE between the two stationary states and the scattered photon has a difference in energy of $-\Delta E$ relative to the incident photon.

the scatterer with an oscillating multipole moment that produces electromagnetic fields far from the scatterer that are equivalent to those of a monochromatic plane wave, which is associated with a scattered photon ω_2. Typically, the strongest non-trivial multipole moment that is induced is the electric dipole, and in this dissertation we assume the conditions are such that this is the only multipole moment induced by the incident photon.

It has been shown that for the intensities typically associated with Raman scattering, the photons themselves can be treated using classical mechanics (using their associated classical electromagnetic fields), whereas the scatterer must be treated using quantum mechanics[1]. Thus, the starting point for calculating Raman intensities is the formula for the classical intensity I of frequency ω_2 radiated far from a set of N dipole moments oscillating at ω_2. Adapting from [4], this is

$$I = \frac{\omega_2^4}{32\pi^2\epsilon_0 c^3}\left|\sum_{d=1}^{N}\left[(\hat{\mathbf{n}}_2 \times \tilde{\mathbf{p}}_d)\times\hat{\mathbf{n}}_2\right]e^{-i\omega_2\hat{\mathbf{n}}_2\cdot\mathbf{r}_d/c}\right|^2, \tag{1.1}$$

where ϵ_0 is the permittivity of free space, c is the speed of light in vacuum, $\mathbf{r}_d$ is the position vector of dipole moment d, and $\tilde{\mathbf{p}}_d$ is the complex amplitude of the real dipole moment

$$\mathbf{p}_d = \frac{1}{2}\left(\tilde{\mathbf{p}}_d e^{-i\omega_2 t} + \text{c.c.}\right). \tag{1.2}$$

Here and in the rest of this dissertation, "c.c." denotes the complex conjugate of the term preceding it.

Thus, the quantity we need to obtain is the real dipole moment of scatterer d, $\mathbf{p}_d$, from which both the complex amplitude $\tilde{\mathbf{p}}_d$ and the oscillation frequency ω_2 can be obtained. $\mathbf{p}_d$ is the quantity that must be treated quantum mechanically, in which case it becomes the *transition* dipole moment between two vibrational, time-dependent stationary states Ψ_i and Ψ_f of the scatterer:

$$\begin{aligned}\mathbf{p} &\rightarrow \mathbf{p}_{fi}\\ &= \langle\Psi_f|\hat{\mathbf{p}}|\Psi_i\rangle,\end{aligned}$$

where $\hat{\mathbf{p}}$ is the dipole moment operator and we have dropped the subscript d for notational simplicity. Assuming that the energetic effect on the scatterer due to the time-dependent electric field $\mathbf{E}_1$ of the incident photon is small, the wavefunctions Ψ_i and Ψ_f can be calculated using time-dependent perturbation theory, in which the new stationary states are expanded in the perturbation strength as modifications to the unperturbed stationary states $\left|\Psi_i^{(0)}\right\rangle$ and $\left|\Psi_f^{(0)}\right\rangle$:

$$\begin{aligned}|\Psi_i\rangle &= \left|\Psi_i^{(0)}\right\rangle + \left|\Psi_i^{(1)}\right\rangle + \left|\Psi_i^{(2)}\right\rangle + \cdots\\ |\Psi_f\rangle &= \left|\Psi_f^{(0)}\right\rangle + \left|\Psi_f^{(1)}\right\rangle + \left|\Psi_f^{(2)}\right\rangle + \cdots.\end{aligned}$$

The corresponding transition dipole moment can therefore be written to different orders in the perturbation strength:

$$\begin{aligned}
\mathbf{p}_{fi}^{(0)} &= \frac{1}{2}\left\langle \Psi_f^{(0)} \left|\hat{\mathbf{p}}\right| \Psi_i^{(0)} \right\rangle + \text{c.c.} \\
\mathbf{p}_{fi}^{(1)} &= \frac{1}{2}\left(\left\langle \Psi_f^{(0)} \left|\hat{\mathbf{p}}\right| \Psi_i^{(1)} \right\rangle + \left\langle \Psi_f^{(1)} \left|\hat{\mathbf{p}}\right| \Psi_i^{(0)} \right\rangle\right) + \text{c.c.} \\
\mathbf{p}_{fi}^{(2)} &= \frac{1}{2}\left(\left\langle \Psi_f^{(0)} \left|\hat{\mathbf{p}}\right| \Psi_i^{(2)} \right\rangle + \left\langle \Psi_f^{(1)} \left|\hat{\mathbf{p}}\right| \Psi_i^{(1)} \right\rangle + \left\langle \Psi_f^{(2)} \left|\hat{\mathbf{p}}\right| \Psi_i^{(0)} \right\rangle\right) + \text{c.c.} \\
&\vdots
\end{aligned}$$

Note that in order to maintain reality of the transition dipole moments, we have added complex conjugates and multiplied by $\frac{1}{2}$.

In spontaneous Raman scattering, the relevant transition dipole moment is $\mathbf{p}_{fi}^{(1)}$ (as opposed to stimulated Raman scattering, in which the relevant transition dipole moment is $\mathbf{p}_{fi}^{(3)}$). Applying time-dependent perturbation theory using the perturbation $-\hat{\mathbf{p}}\cdot\mathbf{E}_1$ to the Hamiltonian representing the interaction energy between the incident electric field and the dipole moment in the presence of this field, $\mathbf{p}_{fi}^{(1)}$ becomes[2]

$$\mathbf{p}_{fi}^{(1)} = \frac{1}{2} e^{-\Gamma_{fi} t} \sum_{j=1}^{N_f} \alpha_{fi}\left(\omega_j\right) \cdot \tilde{\mathbf{E}}_j e^{-i[(\omega_j - \omega_{fi})t + \phi_{fi}]} + \text{c.c.}, \tag{1.3}$$

where $\Gamma_{fi} \equiv \Gamma_f + \Gamma_i$ is the sum of the half-widths of levels i and f, $\omega_{fi} \equiv \omega_f - \omega_i$ is the frequency difference between stationary states i and f, $\phi_{fi} \equiv \phi_f - \phi_i$ is the phase difference between these stationary states, N_f is the number of applied electric *f*ields, $\tilde{\mathbf{E}}_j$ and ω_j are the complex amplitude and frequency of the jth applied electric field (corresponding to the jth incident photon), and the transition polarizability from state i to state f due to the jth applied field is defined as

$$\alpha_{fi}\left(\omega_j\right) \equiv \frac{1}{\hbar} \sum_{r=1}^{\infty} \left(\frac{\mathbf{p}_{fr}\mathbf{p}_{ri}}{\omega_{ri} - \omega_j - i\Gamma_{ri}} + \frac{\mathbf{p}_{ri}\mathbf{p}_{fr}}{\omega_{rf} + \omega_j + i\Gamma_{rf}} \right), \tag{1.4}$$

where the sum is over all quantum states r that are accessible to the system (electronic, vibrational, and, if resolvable, rotational) and $\mathbf{p}_{st}$ is the matrix element of the dipole moment operator between the unperturbed, *spatial* stationary states ψ_s and ψ_t: $\mathbf{p}_{st} \equiv \langle \psi_s \left|\hat{\mathbf{p}}\right| \psi_t \rangle$.

1.3 Placzek's approximation to the transition polarizability operator

In order to simplify the expression for the required transition dipole moment of eq. 1.3, Placzek[5] made a series of simultaneous, radical approximations to the form of the transition polarizability of eq. 1.4, which were later

[2] See Appendix A for the derivation.

shown to be rigorous[1]. In short, we first write out the transition polarizability as a matrix element between unperturbed spatial stationary states in order to explicitly reveal the transition polarizability *operator* $\hat{\alpha}$:

$$\alpha_{fi}\left(\omega_j\right) = \left\langle \psi_f \left| \hat{\alpha}\left(\omega_j\right) \right| \psi_i \right\rangle .$$

We then explicitly include the dependence on the nuclear coordinates $\{R\}$ that was always implicit in the wavefunctions:

$$\left\langle \psi_f \left| \hat{\alpha}\left(\omega_j\right) \right| \psi_i \right\rangle = \left\langle \psi_f\left(\{R\}\right) \left| \hat{\alpha}\left(\omega_j, \{R\}\right) \right| \psi_i\left(\{R\}\right) \right\rangle .$$

We then utilize the fact that we can always effect a coordinate transformation of the $3N_{\text{nuc}}$ nuclear coordinates $\{R\}$; here we choose the set of $3N_{\text{nuc}} - 6$ mass-weighted normal coordinates $\{Q\}$ obtained by transforming using the eigenvectors of the Hessian:

$$\left\langle \psi_f\left(\{R\}\right) \left| \hat{\alpha}\left(\omega_j, \{R\}\right) \right| \psi_i\left(\{R\}\right) \right\rangle \rightarrow \left\langle \psi_f\left(\{Q\}\right) \left| \hat{\alpha}\left(\omega_j, \{Q\}\right) \right| \psi_i\left(\{Q\}\right) \right\rangle .$$

Next, we assume static polarizabilities, i.e., that the polarizability is independent of the frequency ω_j of the applied electric field, which holds in the limit $\omega_j \rightarrow 0$:

$$\left\langle \psi_f\left(\{Q\}\right) \left| \hat{\alpha}\left(\omega_j, \{Q\}\right) \right| \psi_i\left(\{Q\}\right) \right\rangle \rightarrow \left\langle \psi_f\left(\{Q\}\right) \left| \hat{\alpha}\left(\{Q\}\right) \right| \psi_i\left(\{Q\}\right) \right\rangle .$$

Now we invoke the Born-Oppenheimer approximation, in which the electronic wavefunctions $\{e\}$ can be separated out from the vibrational wavefunctions $\{v\}$ (and, if they were resolvable, from the rotational wavefunctions):

$$\begin{aligned} \left|\psi_s\left(\{Q\}\right)\right\rangle &\rightarrow \left|e_s, v_s\left(\{Q\}\right)\right\rangle \\ &= \left|e_s\right\rangle \left|v_s\left(\{Q\}\right)\right\rangle . \end{aligned}$$

We have excluded any dependence of the electronic wavefunctions on the coordinates $\{Q\}$ since they are *implicit* functions of $\{Q\}$; the normal vibrational wavefunctions explicitly depend on the normal coordinates and we retain this dependence. The transition polarizability then becomes

$$\left\langle \psi_f\left(\{Q\}\right) \left| \hat{\alpha}\left(\{Q\}\right) \right| \psi_i\left(\{Q\}\right) \right\rangle \rightarrow \left\langle e_f, v_f\left(\{Q\}\right) \left| \hat{\alpha}\left(\{Q\}\right) \right| e_i, v_i\left(\{Q\}\right) \right\rangle .$$

Finally, we take the coordinates $\{Q\}$ in the transition polarizability operator $\hat{\alpha}$ to become coordinate *operators* $\left\{\hat{Q}\right\}$:

$$\begin{aligned} \left\langle e_f, v_f\left(\{Q\}\right) \left| \hat{\alpha}\left(\{Q\}\right) \right| e_i, v_i\left(\{Q\}\right) \right\rangle &\rightarrow \left\langle e_f, v_f\left(\{Q\}\right) \left| \hat{\alpha}\left(\left\{\hat{Q}\right\}\right) \right| e_i, v_i\left(\{Q\}\right) \right\rangle \\ &\equiv \alpha_{fi}. \end{aligned}$$

The polarizability operator has been transformed to a simple explicit function of the normal coordinate operators $\left\{\hat{Q}\right\}$, in which we have removed the dependence of the corresponding matrix element on ω_j and on the wavefunctions, frequencies, and decay constants of all accessible quantum states!

1.4 Approximations of electrical and mechanical harmonicity

As an explicit example of the simple dependence of α_{fi} that we will subsequently study, assuming that the normal coordinates $\left\{\hat{Q}\right\}$ are small perturbations from the equilibrium configuration of the scatterer, we can Taylor-expand the transition polarizability operator in what is known as the approximation of electrical harmonicity (the polarizability is linear in the normal coordinates):

$$\hat{\alpha}\left(\left\{\hat{Q}\right\}\right) \approx \left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\{\hat{Q}\}=0} + \sum_{k=1}^{3N_{\text{nuc}}-6}\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\{\hat{Q}\}=0}\hat{Q}_k.$$

Here we explicitly see that the polarizability operator that results in the complicated matrix element of eq. 1.4 has been replaced by the *numbers* $\left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\{\hat{Q}\}=0}$ and $\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\{\hat{Q}\}=0}$ as coefficients to the simpler operators $\left\{\hat{Q}\right\}$ to increasing powers. The corresponding matrix element becomes

$$\begin{aligned}
\alpha_{fi} &= \left\langle e_f, v_f\left(\{Q\}\right)\middle|\hat{\alpha}\left(\left\{\hat{Q}\right\}\right)\middle|e_i, v_i\left(\{Q\}\right)\right\rangle \\
&\approx \left\langle e_f, v_f\left(\{Q\}\right)\middle|\left(\left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\{\hat{Q}\}=0} + \sum_{k=1}^{3N_{\text{nuc}}-6}\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\{\hat{Q}\}=0}\hat{Q}_k\right)\middle|e_i, v_i\left(\{Q\}\right)\right\rangle \\
&= \left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\{\hat{Q}\}=0}\left\langle e_f, v_f\left(\{Q\}\right)|e_i, v_i\left(\{Q\}\right)\right\rangle + \sum_{k=1}^{3N_{\text{nuc}}-6}\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\{\hat{Q}\}=0} \\
&\quad \times\left\langle e_f, v_f\left(\{Q\}\right)\middle|\hat{Q}_k\middle|e_i, v_i\left(\{Q\}\right)\right\rangle.
\end{aligned}$$

Assuming that the scatterer starts off in its ground electronic state ($e_i \to e_g$), then since we assumed from the start the case of non-resonance scattering, the final electronic state must also be the ground electronic state

$(e_f \rightarrow e_g)$. Then the transition polarizability becomes

$$\begin{aligned}
\alpha_{fi} &= \left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\{\hat{Q}\}=0} \langle e_g|e_g\rangle \langle v_f(\{Q\})|v_i(\{Q\})\rangle + \sum_{k=1}^{3N_{\text{nuc}}-6}\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\{\hat{Q}\}=0} \\
&\quad \times \langle e_g|e_g\rangle \left\langle v_f(\{Q\})\left|\hat{Q}_k\right|v_i(\{Q\})\right\rangle \\
&= \left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\{\hat{Q}\}=0} \left\langle n_1^f n_2^f \cdots n_{3N_{\text{nuc}}-6}^f | n_1^i n_2^i \cdots n_{3N_{\text{nuc}}-6}^i\right\rangle \\
&\quad + \sum_{k=1}^{3N_{\text{nuc}}-6}\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\{\hat{Q}\}=0} \left\langle n_1^f n_2^f \cdots n_{3N_{\text{nuc}}-6}^f \left|\hat{Q}_k\right| n_1^i n_2^i \cdots n_{3N_{\text{nuc}}-6}^i\right\rangle,
\end{aligned}$$

where in the latter step we have transformed the vibrational wavefunctions to occupation number notation. Plugging this expression into eq. 1.3 for the first-order transition dipole moment, we have

$$\begin{aligned}
\mathbf{p}_{fi}^{(1)} &= \frac{1}{2}e^{-\Gamma_{fi}t}\sum_{j=1}^{N_f}\left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\{\hat{Q}\}=0}\left\langle n_1^f n_2^f \cdots n_{3N_{\text{nuc}}-6}^f | n_1^i n_2^i \cdots n_{3N_{\text{nuc}}-6}^i\right\rangle \cdot \tilde{\mathbf{E}}_j e^{-i[(\omega_j-\omega_{fi})t+\phi_{fi}]} + \text{c.c.} \\
&\quad + \frac{1}{2}e^{-\Gamma_{fi}t}\sum_{j=1}^{N_f}\sum_{k=1}^{3N_{\text{nuc}}-6}\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\{\hat{Q}\}=0}\left\langle n_1^f n_2^f \cdots n_{3N_{\text{nuc}}-6}^f \left|\hat{Q}_k\right| n_1^i n_2^i \cdots n_{3N_{\text{nuc}}-6}^i\right\rangle \\
&\quad \cdot\tilde{\mathbf{E}}_j e^{-i[(\omega_j-\omega_{fi})t+\phi_{fi}]} + \text{c.c.}
\end{aligned}$$

Now we separate this expression into terms of different frequencies for the case of a single applied electric field ($N_f = 1$), for which $j \rightarrow 1$. The first term is non-zero only if the occupations of all modes remain the same, i.e., if $f = i$. This is the Rayleigh term of the transition dipole moment:

$$\mathbf{p}_{ii}^{\text{Ray}} = \frac{1}{2}e^{-2\Gamma_i t}\left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\{\hat{Q}\}=0}\cdot\tilde{\mathbf{E}}_1 e^{-i\omega_1 t} + \text{c.c.}$$

Assuming mechanical harmonicity (the forces on the nuclei are linear in the normal coordinates), in which $\hat{Q}_k \rightarrow \sqrt{\frac{\hbar}{2\omega_k}}\left(\hat{a}_k^\dagger + \hat{a}_k\right)$,[3] the second term in the expression for $\mathbf{p}_{fi}^{(1)}$ is non-zero for mode k if either $n_k^f = n_k^i + 1$ and the occupations of all other modes remain the same (Stokes Raman term), or $n_k^f = n_k^i - 1$ and again the occupations

[3] Remember that we are using *mass-weighted* normal coordinates defined by $Q_k^{MW} = \sqrt{m_k}Q_k^{NMW}$, where "$NMW$" stands for "non-mass-weighted." So if we had been using non-mass-weighted normal coordinates, the transformation would have been $\hat{Q}_k \rightarrow \sqrt{\frac{\hbar}{2m_k\omega_k}}\left(\hat{a}_k^\dagger + \hat{a}_k\right)$.

of all other modes remain the same (anti-Stokes Raman term), in which cases we have

$$\begin{aligned}
\mathbf{p}_{fi}^{S} &= \frac{1}{2}e^{-\Gamma_{fi}t}\sqrt{\frac{\hbar\left(n_k^i+1\right)}{2\omega_k}}\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\left\{\hat{Q}\right\}=0}\cdot\tilde{\mathbf{E}}_1 e^{-i[(\omega_1-\omega_k)t+\phi_{fi}]}+\text{c.c.} \\
\mathbf{p}_{fi}^{AS} &= \frac{1}{2}e^{-\Gamma_{fi}t}\sqrt{\frac{\hbar n_k^i}{2\omega_k}}\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\left\{\hat{Q}\right\}=0}\cdot\tilde{\mathbf{E}}_1 e^{-i[(\omega_1+\omega_k)t+\phi_{fi}]}+\text{c.c.}
\end{aligned}$$

Comparing these forms of the transition dipole moment to the general form of the real dipole moment of eq. 1.2, we find that for Stokes Raman scattering,

$$\begin{aligned}
\omega_2^S &= \omega_1-\omega_k \\
\tilde{\mathbf{p}}_d^S &= e^{-\Gamma_{fi}t}\sqrt{\frac{\hbar\left(n_k^i+1\right)}{2\omega_k}}\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\left\{\hat{Q}\right\}=0}\cdot\tilde{\mathbf{E}}_1 e^{-i\phi_{fi}},
\end{aligned}$$

for anti-Stokes Raman scattering,

$$\begin{aligned}
\omega_2^{AS} &= \omega_1+\omega_k \\
\tilde{\mathbf{p}}_d^{AS} &= e^{-\Gamma_{fi}t}\sqrt{\frac{\hbar n_k^i}{2\omega_k}}\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\left\{\hat{Q}\right\}=0}\cdot\tilde{\mathbf{E}}_1 e^{-i\phi_{fi}},
\end{aligned}$$

and for Rayleigh scattering,

$$\begin{aligned}
\omega_2^{\text{Ray}} &= \omega_1 \\
\tilde{\mathbf{p}}_d^{\text{Ray}} &= e^{-2\Gamma_i t}\left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\left\{\hat{Q}\right\}=0}\cdot\tilde{\mathbf{E}}_1.
\end{aligned}$$

Making the final (realistic) simplification of infinite lifetimes of the initial and final states ($\Gamma_{fi}\to 0$, $\Gamma_i\to 0$), pulling the position-dependent phase out of $\tilde{\mathbf{E}}_1$, and taking the applied field at the location of scatterer d via

$$\tilde{\mathbf{E}}_1\to\tilde{\mathbf{E}}_1 e^{i(\omega_1\hat{\mathbf{n}}_1\cdot\mathbf{r}_d/c-\phi_1)},$$

the complex amplitudes of the transition dipole moments become

$$\begin{aligned}
\tilde{\mathbf{p}}_d^S &= \sqrt{\frac{\hbar\left(n_k^i+1\right)}{2\omega_k}}\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\left\{\hat{Q}\right\}=0}\cdot\tilde{\mathbf{E}}_1 e^{i(\omega_1\hat{\mathbf{n}}_1\cdot\mathbf{r}_d/c-\phi_1-\phi_{fi})} \\
\tilde{\mathbf{p}}_d^{AS} &= \sqrt{\frac{\hbar n_k^i}{2\omega_k}}\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\left\{\hat{Q}\right\}=0}\cdot\tilde{\mathbf{E}}_1 e^{i(\omega_1\hat{\mathbf{n}}_1\cdot\mathbf{r}_d/c-\phi_1-\phi_{fi})} \\
\tilde{\mathbf{p}}_d^{\text{Ray}} &= \left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\left\{\hat{Q}\right\}=0}\cdot\tilde{\mathbf{E}}_1 e^{i(\omega_1\hat{\mathbf{n}}_1\cdot\mathbf{r}_d/c-\phi_1)}.
\end{aligned}$$

Plugging these expressions into the formula for the electromagnetic intensity far from oscillating dipole sources (eq. 1.1), we have

$$\begin{aligned} I_k^S &= \frac{(\omega_1-\omega_k)^4\hbar}{64\pi^2\epsilon_0 c^3\omega_k}\left(n_k^i+1\right)\left|\left(\hat{\mathbf{n}}_2\times\left(\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\left\{\hat{Q}\right\}=0}\cdot\tilde{\mathbf{E}}_1\right)\right)\times\hat{\mathbf{n}}_2\right|^2 \\ &\quad\times\left|\sum_{d=1}^{N}e^{i(\omega_1\hat{\mathbf{n}}_1\cdot\mathbf{r}_d/c-\phi_1-\phi_{fi})-i(\omega_1-\omega_k)\hat{\mathbf{n}}_2\cdot\mathbf{r}_d/c}\right|^2 \qquad (1.5) \\ I_k^{AS} &= \frac{(\omega_1+\omega_k)^4\hbar}{64\pi^2\epsilon_0 c^3\omega_k}n_k^i\left|\left(\hat{\mathbf{n}}_2\times\left(\left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\left\{\hat{Q}\right\}=0}\cdot\tilde{\mathbf{E}}_1\right)\right)\times\hat{\mathbf{n}}_2\right|^2 \qquad (1.6) \\ &\quad\times\left|\sum_{d=1}^{N}e^{i(\omega_1\hat{\mathbf{n}}_1\cdot\mathbf{r}_d/c-\phi_1-\phi_{fi})-i(\omega_1+\omega_k)\hat{\mathbf{n}}_2\cdot\mathbf{r}_d/c}\right|^2 \\ I^{\mathsf{Ray}} &= \frac{\omega_1^4}{32\pi^2\epsilon_0 c^3}\left|\left(\hat{\mathbf{n}}_2\times\left(\left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\left\{\hat{Q}\right\}=0}\cdot\tilde{\mathbf{E}}_1\right)\right)\times\hat{\mathbf{n}}_2\right|^2\left|\sum_{d=1}^{N}e^{i(\omega_1(\hat{\mathbf{n}}_1-\hat{\mathbf{n}}_2)\cdot\mathbf{r}_d/c-\phi_1)}\right|^2. \qquad (1.7) \end{aligned}$$

1.5 Coherence properties of the detected intensities

Note that since ϕ_i and ϕ_f, which are the phases of the quantum wavefunctions of the initial and final states, vary arbitrarily from scatterer to scatterer within the scattering material, the Raman intensities contain a structure factor of the form

$$\begin{aligned} \left|\sum_{d=1}^{N}e^{i\phi_{\mathsf{rand},d}}\right|^2 &= \sum_{c=1}^{N}e^{-i\phi_{\mathsf{rand},c}}\sum_{d=1}^{N}e^{i\phi_{\mathsf{rand},d}} \\ &= \sum_{c=1}^{N}\sum_{d=1}^{N}e^{i(\phi_{\mathsf{rand},d}-\phi_{\mathsf{rand},c})} \\ &= \sum_{c=1}^{N}e^{i(\phi_{\mathsf{rand},c}-\phi_{\mathsf{rand},c})}+\sum_{c\neq d=1}^{N}e^{i(\phi_{\mathsf{rand},d}-\phi_{\mathsf{rand},c})} \\ &= \sum_{c=1}^{N}1+0 \\ &= N \end{aligned}$$

for large N, showing that Raman scattering as described here is incoherent, meaning that the intensity radiated from N scatterers is N times the intensity radiated from one scatterer. Conversely, the Rayleigh intensity contains

the structure factor

$$\left|\sum_{d=1}^{N} e^{i(\omega_1(\hat{\mathbf{n}}_1-\hat{\mathbf{n}}_2)\cdot\mathbf{r}_d/c-\phi_1)}\right|^2.$$

We see that for forward scattering, in which $\hat{\mathbf{n}}_2 = \hat{\mathbf{n}}_1$, the structure factor is

$$\begin{aligned}\left|\sum_{d=1}^{N} e^{-i\phi_1}\right|^2 &= \sum_{c=1}^{N} e^{i\phi_1}\sum_{d=1}^{N} e^{-i\phi_1} \\ &= \sum_{c=1}^{N}\sum_{d=1}^{N} 1 \\ &= N^2,\end{aligned}$$

showing that forward Rayleigh scattering is coherent. For non-forward scattering, the structure factor has the form

$$\left|\sum_{d=1}^{N} e^{i(\omega_1\mathbf{n}_{12}\cdot\mathbf{r}_d/c-\phi_1)}\right|^2.$$

If the scatterers have a random distribution ($\mathbf{r}_d$ random), then we obtain the incoherent case in which the structure factor equals N for large N. However, for equally-spaced scatterers, i.e., $\mathbf{r}_d = d\mathbf{s}$, as occurs in, e.g., a lattice, the structure factor takes the form

$$\begin{aligned}\left|\sum_{d=1}^{N} e^{i(\omega_1 d\mathbf{n}_{12}\cdot\mathbf{s}/c-\phi_1)}\right|^2 &= \left|\sum_{d=1}^{N} e^{i((\omega_1 n_{12}s/c)d-\phi_1)}\right|^2 \\ &= \left|e^{-i\phi_1}\sum_{d=1}^{N} e^{i(\omega_1 n_{12}s/c)d}\right|^2 \\ &= \sum_{b=1}^{N}\sum_{d=1}^{N} e^{i(\omega_1 n_{12}s/c)(d-b)} \\ &= \sum_{b=1}^{N} e^{i(\omega_1 n_{12}s/c)(b-b)} + \sum_{b\neq d=1}^{N} e^{i(\omega_1 n_{12}s/c)(d-b)} \\ &= N + \sum_{d>b}^{N}\left(e^{i(\omega_1 n_{12}s/c)(d-b)} + e^{-i(\omega_1 n_{12}s/c)(d-b)}\right) \\ &= N + 2\sum_{d>b}^{N}\cos\left[\frac{\omega_1 n_{12}s}{c}(d-b)\right].\end{aligned}$$

In the second term, the quantity $d-b$ takes on positive integer values. If the factor $\frac{\omega_1 n_{12}s}{c}$ is not an integer multiple of 2π, then for large N the entire term will average out to zero (incoherent scattering). However, if this factor *is* an integer multiple of 2π,

$$\frac{\omega_1 n_{12}s}{c} = 2\pi m,$$

then the structure factor becomes

$$\begin{aligned} N + 2\sum_{d>b}^{N} \cos\left[\frac{\omega_1 n_{12} s}{c}(d-b)\right] &= N + 2\sum_{d>b}^{N} \cos\left[2\pi m (d-b)\right] \\ &= N + 2\sum_{d>b}^{N} 1 \\ &= N + 2\left(\frac{N(N-1)}{2}\right) \\ &= N^2, \end{aligned}$$

so that the scattering is coherent. However, it should be noted that for this to occur, i.e., for the scatterers to be regularly spaced by an integer multiple of $s = \frac{2\pi c}{\omega_1 |\hat{\mathbf{n}}_1 - \hat{\mathbf{n}}_2|} = \frac{\lambda_1}{|\hat{\mathbf{n}}_1 - \hat{\mathbf{n}}_2|}$, this spacing must be on the order of the wavelength of the light, which for Rayleigh scattering is orders of magnitude larger than the typical spacing of scatterers in a crystal. Note further that in the opposite extreme, as the material approaches perfect homogeneity ($s \to 0$), Rayleigh scattering is coherent in all directions. Of course, these extremes do not occur in nature, so we conclude simply that Rayleigh scattering is incoherent unless the scattering is forward scattering. Thus, the intensity formulas of eqs. 1.5 to 1.7 become

$$\begin{aligned} I_k^S &= \frac{(\omega_1 - \omega_k)^4 N\hbar}{64\pi^2 \epsilon_0 c^3 \omega_k}\left(n_k^i + 1\right)\left|\left(\hat{\mathbf{n}}_2 \times \left(\left[\frac{\partial \alpha\left(\left\{\hat{Q}\right\}\right)}{\partial \hat{Q}_k}\right]_{\{\hat{Q}\}=0} \cdot \tilde{\mathbf{E}}_1\right)\right) \times \hat{\mathbf{n}}_2\right|^2 \\ I_k^{AS} &= \frac{(\omega_1 + \omega_k)^4 N\hbar}{64\pi^2 \epsilon_0 c^3 \omega_k} n_k^i \left|\left(\hat{\mathbf{n}}_2 \times \left(\left[\frac{\partial \alpha\left(\left\{\hat{Q}\right\}\right)}{\partial \hat{Q}_k}\right]_{\{\hat{Q}\}=0} \cdot \tilde{\mathbf{E}}_1\right)\right) \times \hat{\mathbf{n}}_2\right|^2 \\ I^{\text{Ray}} &= \frac{\omega_1^4}{32\pi^2 \epsilon_0 c^3}\left|\left(\hat{\mathbf{n}}_2 \times \left(\left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\{\hat{Q}\}=0} \cdot \tilde{\mathbf{E}}_1\right)\right) \times \hat{\mathbf{n}}_2\right|^2 \times \begin{cases} N^2 & \text{forward scattering} \\ N & \text{otherwise.} \end{cases} \end{aligned}$$

1.6 Boltzmann averaging of the intensities

Next, we need to address the factors $n_k^i + 1$ and n_k^i in the intensity formulas for Stokes and anti-Stokes scattering. These involve the occupation numbers of vibrational mode k of the scatterers before any external fields are applied. We know that the energy of vibrational mode k having occupation n of a single scatterer is $E_k(n) = \left(n + \frac{1}{2}\right)\hbar\omega_k$. We can therefore determine the Boltzmann-averaged values of $n+1$ and n, since the Boltzmann weighting of a state with energy $E_k(n)$ is $\exp\left(-\frac{E_k(n)}{k_B T}\right)$, where k_B is Boltzmann's constant and T is the temperature of the system in Kelvin. The accessible states here are $n \in \{0, 1, 2, \ldots\}$. Thus, the Boltzmann-averaged values of $n+1$

and n are

$$\begin{aligned}\langle n+1\rangle &= \frac{\sum_{n=0}^{\infty}(n+1)\,e^{-\left(n+\frac{1}{2}\right)\hbar\omega_k/k_BT}}{\sum_{n=0}^{\infty}e^{-\left(n+\frac{1}{2}\right)\hbar\omega_k/k_BT}} \\ &= \frac{1}{1-e^{-\hbar\omega_k/k_BT}}\end{aligned}$$

and

$$\begin{aligned}\langle n\rangle &= \frac{\sum_{n=0}^{\infty}ne^{-\left(n+\frac{1}{2}\right)\hbar\omega_k/k_BT}}{\sum_{n=0}^{\infty}e^{-\left(n+\frac{1}{2}\right)\hbar\omega_k/k_BT}} \\ &= \frac{1}{e^{\hbar\omega_k/k_BT}-1},\end{aligned}$$

so that the Raman and Rayleigh intensity formulas become

$$I_k^S = \frac{(\omega_1-\omega_k)^4 N\hbar}{64\pi^2\epsilon_0 c^3\omega_k\left(1-e^{-\hbar\omega_k/k_BT}\right)}\left|\left(\hat{\mathbf{n}}_2\times\left(\alpha_k^{0,\mathsf{lab}}\cdot\tilde{\mathbf{E}}_1^{\mathsf{lab}}\right)\right)\times\hat{\mathbf{n}}_2\right|^2 \tag{1.8}$$

$$I_k^{AS} = \frac{(\omega_1+\omega_k)^4 N\hbar}{64\pi^2\epsilon_0 c^3\omega_k\left(e^{\hbar\omega_k/k_BT}-1\right)}\left|\left(\hat{\mathbf{n}}_2\times\left(\alpha_k^{0,\mathsf{lab}}\cdot\tilde{\mathbf{E}}_1^{\mathsf{lab}}\right)\right)\times\hat{\mathbf{n}}_2\right|^2 \tag{1.9}$$

$$I^{\mathsf{Ray}} = \frac{\omega_1^4}{32\pi^2\epsilon_0 c^3}\left|\left(\hat{\mathbf{n}}_2\times\left(\alpha^{0,\mathsf{lab}}\cdot\tilde{\mathbf{E}}_1^{\mathsf{lab}}\right)\right)\times\hat{\mathbf{n}}_2\right|^2\times\begin{cases}N^2 & \text{forward scattering}\\ N & \text{otherwise,}\end{cases} \tag{1.10}$$

where we have defined

$$\begin{aligned}\alpha_k^0 &\equiv \left[\frac{\partial\alpha\left(\left\{\hat{Q}\right\}\right)}{\partial\hat{Q}_k}\right]_{\{\hat{Q}\}=0} \\ \alpha^0 &\equiv \left[\alpha\left(\left\{\hat{Q}\right\}\right)\right]_{\{\hat{Q}\}=0}\end{aligned}$$

and stated explicitly that the quantities α_k^0, α^0, and $\tilde{\mathbf{E}}_1$ are in the lab frame (it is already clear from its definition that the remaining directional quantity, $\hat{\mathbf{n}}_2$, is in the lab frame).

1.7 Polarization considerations and "oriented" intensity formulas

We assume the general experimental setup of Fig. 1.2, in which the light radiated from a scatterer located at the origin of the lab frame (XYZ) propagates in the $+Z$ direction: $\hat{\mathbf{n}}_2 = \hat{\mathbf{Z}} = [0,0,1]$. Then the triple product

$\left(\hat{\mathbf{n}}_2 \times \left(\alpha_k^{0,\mathsf{lab}} \cdot \tilde{\mathbf{E}}_1^{\mathsf{lab}}\right)\right) \times \hat{\mathbf{n}}_2$ required in the intensity formulas above has the value

$$\begin{aligned} \mathbf{p}_k &\equiv \left(\hat{\mathbf{n}}_2 \times \left(\alpha_k^{0,\mathsf{lab}} \cdot \tilde{\mathbf{E}}_1^{\mathsf{lab}}\right)\right) \times \hat{\mathbf{n}}_2 \\ &= \begin{bmatrix} \mathbf{X}^T \alpha_k^{0,\mathsf{lab}} \tilde{\mathbf{E}}_1^{\mathsf{lab}} \\ \mathbf{Y}^T \alpha_k^{0,\mathsf{lab}} \tilde{\mathbf{E}}_1^{\mathsf{lab}} \\ 0 \end{bmatrix}, \end{aligned}$$

where we have dropped the dot product since it is redundant, as it represents multiplication of a vector $\tilde{\mathbf{E}}_1^{\mathsf{lab}}$ by a matrix $\alpha_k^{0,\mathsf{lab}}$. Note that we have specified the X and Y components of the vector $\alpha_k^{0,\mathsf{lab}}\tilde{\mathbf{E}}_1^{\mathsf{lab}}$ by defining the vectors

$$\begin{aligned} \mathbf{X} &\equiv \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T \\ \mathbf{Y} &\equiv \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T \end{aligned}$$

and left-multiplying by their transposes.

In order to calculate $\alpha_k^{0,\mathsf{lab}}\tilde{\mathbf{E}}_1^{\mathsf{lab}}$, we need to transform the quantities α_k^0 and $\tilde{\mathbf{E}}_1$ from their "natural" frames (molecule frame and incident polarization frame, respectively) to the lab frame. If the orthogonal matrix R takes the vector $\mathbf{x}$ from the incident polarization frame to the lab frame via $\mathbf{x}^{\mathsf{lab}} = R\mathbf{x}^{\mathsf{inc}}$ and the orthogonal matrix T takes the vector $\mathbf{x}$ from the molecule frame to the lab frame via $\boldsymbol{x}^{\mathsf{lab}} = T\mathbf{x}^{\mathsf{mol}}$ (and therefore takes a rank-two tensor X in the molecule frame to the lab frame via $X^{\mathsf{lab}} = TX^{\mathsf{mol}}T^T$), the quantity $\alpha_k^{0,\mathsf{lab}}\tilde{\mathbf{E}}_1^{\mathsf{lab}}$ equals $T\alpha_k^{0,\mathsf{mol}}T^T R\tilde{\mathbf{E}}_1^{\mathsf{inc}}$. The rotation matrices R and T are[1]

$$\begin{aligned} R &= \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \\ T &= \begin{bmatrix} l_{xX} & l_{yX} & l_{zX} \\ l_{xY} & l_{yY} & l_{zY} \\ l_{xZ} & l_{yZ} & l_{zZ} \end{bmatrix} \\ &= \begin{bmatrix} \cos\beta\cos\alpha\cos\gamma - \sin\alpha\sin\gamma & -\cos\beta\cos\alpha\sin\gamma - \sin\alpha\cos\gamma & \sin\beta\cos\alpha \\ \cos\beta\sin\alpha\cos\gamma + \cos\alpha\sin\gamma & -\cos\beta\sin\alpha\sin\gamma + \cos\alpha\cos\gamma & \sin\beta\sin\alpha \\ -\sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{bmatrix}, \end{aligned}$$

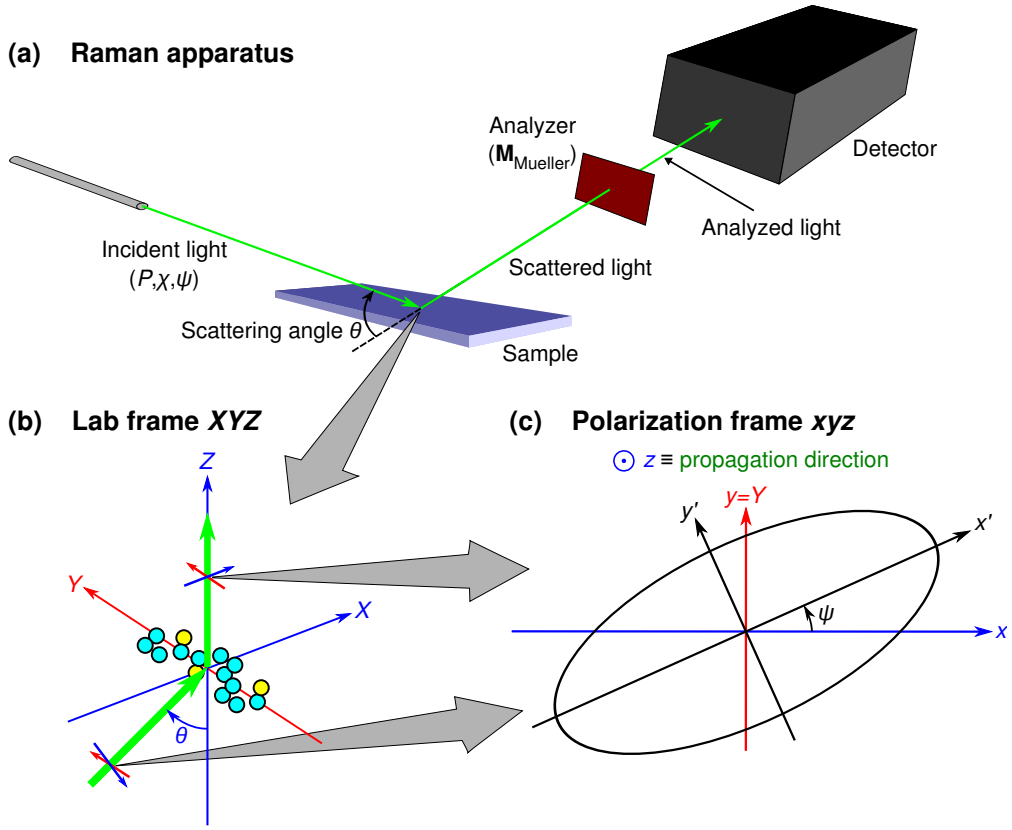

Figure 1.2: Setups and conventions for this dissertation. (a) Schematic of a general Raman apparatus. All controllable, "experimental" variables are shown: $\{P, \chi, \psi, \theta, \mathbf{M}_{\text{Mueller}}\}$. (b) Schematic of the lab frame, defined by the XYZ axes. Note that the optical axes (green arrows) and θ are the same as those in (a). The scattering plane is defined by the XZ plane, which will always be shown in blue. The direction normal to the scattering plane will always be shown in red. The scattered/analyzed light always travels in the $+Z$ direction, and the incident light always travels in the scattering plane. The sample is always located at the origin of the lab frame, and the two small blue/red axes represent the polarization frames of the incident and scattered light. (c) Diagram of a polarization frame, defined by the xyz axes. The xz axes of the polarization frame always lie in the scattering plane XZ and are well-defined by requiring that the light always propagate in the $+z$ direction. The polarization of the light is completely specified by the orientation and shape of the polarization ellipse[3], which is perpendicular to the propagation direction. The orientation of the polarization ellipse is represented by the angle ψ, called the orientation angle, with respect to the scattering plane. Not shown is the ellipticity angle χ, which describes the shape of the polarization ellipse. For example, when $\chi = 0$, the ellipse degenerates to a line along the x' axis; this then represents light that is linearly polarized along the x' axis.

where θ is the angle between the incident and scattered light ($\cos\theta = \hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2$); l_{ij} is the direction cosine specifying the cosine of the angle between the scatterer's i axis and the lab's j axis; and $\{\alpha \in [0, 2\pi], \beta \in [0, \pi], \gamma \in [0, 2\pi]\}$ are Euler angles specifying the orientation of the rotated ("primed", "subscripted") molecule frame relative to the fixed ("unprimed", "non-subscripted") lab frame, consistent with Fig. A5.2 of [1].

We can obtain a complete characterization of the polarization properties of the scattered light by first forming

$$\begin{aligned}
\mathbf{p}_k \mathbf{p}_k^\dagger &= \begin{bmatrix} \mathbf{X}^T T \alpha_k^{0,\mathsf{mol}} T^T R \tilde{\mathbf{E}}_1^{\mathsf{inc}} \\ \mathbf{Y}^T T \alpha_k^{0,\mathsf{mol}} T^T R \tilde{\mathbf{E}}_1^{\mathsf{inc}} \\ 0 \end{bmatrix} \begin{bmatrix} \mathbf{X}^T T \left(\alpha_k^{0,\mathsf{mol}}\right)^* T^T R \left(\tilde{\mathbf{E}}_1^{\mathsf{inc}}\right)^* & \mathbf{Y}^T T \left(\alpha_k^{0,\mathsf{mol}}\right)^* T^T R \left(\tilde{\mathbf{E}}_1^{\mathsf{inc}}\right)^* & 0 \end{bmatrix} \\
&= \frac{E_0^2}{2} \begin{bmatrix} \mathbf{X}^T G_k \mathbf{X} & \mathbf{X}^T G_k \mathbf{Y} & 0 \\ \mathbf{Y}^T G_k \mathbf{X} & \mathbf{Y}^T G_k \mathbf{Y} & 0 \\ 0 & 0 & 0 \end{bmatrix},
\end{aligned}$$

where the 3×3 matrix G_k is defined as

$$G_k = \frac{2}{E_0^2} T \alpha_k^{0,\mathsf{mol}} T^T R \tilde{\mathbf{E}}_1^{\mathsf{inc}} \left(\tilde{\mathbf{E}}_1^{\mathsf{inc}}\right)^\dagger R^T T \left(\alpha_k^{0,\mathsf{mol}}\right)^\dagger T^T$$

and describes the dipole induced in the scatterer by the incident photon. The upper-left 2×2 matrix of $\frac{2}{E_0^2} \mathbf{p}_k \mathbf{p}_k^\dagger$ and G_k is called the coherency matrix J_k^{scat} of the scattered light[3]:

$$\begin{aligned}
J_k^{\mathsf{scat}} &= \frac{2}{E_0^2} A \mathbf{p}_k \mathbf{p}_k^\dagger A^T \\
&= A G_k A^T, \qquad (1.11)
\end{aligned}$$

where we have defined

$$A \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Likewise, the coherency matrix of the incident light J^{inc} is the upper-left 2×2 matrix of what we define as the incident light tensor $\tilde{F}_1^{\mathsf{inc}} \equiv \frac{2}{E_0^2} \tilde{\mathbf{E}}_1^{\mathsf{inc}} \left(\tilde{\mathbf{E}}_1^{\mathsf{inc}}\right)^\dagger$:

$$J^{\mathsf{inc}} = A \tilde{F}_1^{\mathsf{inc}} A^T. \qquad (1.12)$$

Note that the matrix G_k can be represented in terms of the incident light tensor as

$$G_k = T \alpha_k^{0,\mathsf{mol}} T^T R \tilde{F}_1^{\mathsf{inc}} R^T T \left(\alpha_k^{0,\mathsf{mol}}\right)^\dagger T^T. \qquad (1.13)$$

Throughout these calculations, common sense must be used to ensure correct treatment of the units and prefactors.

The coherency matrix J is special because it is equivalent to a four-component vector, called a Stokes vector, that completely specifies the polarization of a photon[3]. The components of the Stokes vector are obtained from J by

$$S_0 = J_{xx} + J_{yy} \tag{1.14}$$

$$S_1 = J_{xx} - J_{yy} \tag{1.15}$$

$$S_2 = J_{xy} + J_{yx} \tag{1.16}$$

$$S_3 = -i\left(J_{xy} - J_{yx}\right). \tag{1.17}$$

The S_0 component represents the total intensity, including that of both the polarized and unpolarized light. The latter three components represent the degree of linear polarization in the 0° and 90° directions, the degree of linear polarization in the 45° and -45° directions, and the degree of circular polarization, respectively. Once a Stokes vector is obtained, the effects of optical elements such as polarizers and retarders is simple because these elements can be represented by 4×4 matrices called Mueller matrices that multiply the Stokes vectors. Common Mueller matrices are listed in Appendix B.

The intensity formulas of eqs. 1.8 to 1.10 equal the S_0 component of the light scattered from a sample located at the origin of the lab frame. As an example of how to measure light intensities at particular polarizations, in order to determine the intensity of the horizontally linearly polarized component of the scattered light, we would first form the scattered Stokes vector using the J_k^{scat} coherency matrix (eq. 1.11 and eqs. 1.14 to 1.17), multiply this Stokes vector by the Mueller matrix representing a horizontal linear polarizer,

$$M = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and *then* take the S_0 component of the resulting Stokes vector $\mathbf{S}_k^{\mathsf{det}}$ as the detected intensity:

$$I_k = f_k \mathbf{B}^T \mathbf{S}_k^{\mathsf{det}},$$

where f_k is the frequency-dependent prefactor to the squared amplitude in eqs. 1.8 to 1.10 and

$$\mathbf{B} \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^T.$$

Similarly, in order to represent *incident* light that is purely *right-handed circularly* polarized, we *could* start with an arbitrary Stokes vector representing light that is circularly polarized to some degree (i.e., having a non-zero value of the S_3 component) and multiply this Stokes vector by the Mueller matrix representing a right-handed circular polarizer,

$$M = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

However, there is a better way to represent the incident light by using the form of the Stokes vectors in terms of the real electric field amplitude E_0, degree of polarization P ($\in [0,1]$), angle of ellipticity χ ($\in \left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$), and orientation angle ψ ($\in [0,\pi]$):

$$S_0 = E_0^2 \tag{1.18}$$

$$S_1 = PE_0^2 \cos 2\chi \cos 2\psi \tag{1.19}$$

$$S_2 = PE_0^2 \cos 2\chi \sin 2\psi \tag{1.20}$$

$$S_3 = PE_0^2 \sin 2\chi. \tag{1.21}$$

This affords a method of obtaining a concrete Stokes vector with which we can start the calculation. For example, we can start with linearly polarized light parallel to (in) the scattering plane using $\{P=1, \chi=0, \psi=0\}$, linearly polarized light perpendicular to the scattering plane using $\left\{P=1, \chi=0, \psi=\frac{\pi}{2}\right\}$, right-handed circularly polarized light using $\left\{P=1, \chi=\frac{\pi}{4}, \psi=\text{arbitrary}\right\}$, left-handed circularly polarized light using $\left\{P=1, \chi=-\frac{\pi}{4}, \psi=\text{arbitrary}\right\}$, unpolarized ("natural") light using $P=0$, or partially polarized light using an intermediate value of P. *Then* we can form the coherency matrix J directly by inverting eqs. 1.14 to 1.17 and using eqs. 1.18 to 1.21,

$$\begin{aligned} J_{xx} &= \frac{E_0^2}{2}\left(1 + P\cos 2\chi \cos 2\psi\right) \\ J_{yy} &= \frac{E_0^2}{2}\left(1 - P\cos 2\chi \cos 2\psi\right) \\ J_{xy} &= \frac{E_0^2}{2}P\left(\cos 2\chi \sin 2\psi + i \sin 2\chi\right) \\ J_{yx} &= \frac{E_0^2}{2}P\left(\cos 2\chi \sin 2\psi - i \sin 2\chi\right), \end{aligned}$$

form the incident light tensor by "inverting" eq. 1.12,

$$\tilde{F}_1^{\mathsf{inc}} = A^T J^{\mathsf{inc}} A \tag{1.22}$$

(i.e., padding the coherency matrix with zeros), and then use eq. 1.13 to form the G_k matrix and proceed with the rest of the calculation.

In summary, we can obtain the Raman and Rayleigh intensities due to an "oriented" scatterer (fixed value of T) by

1. forming the coherency matrix J^{inc} for the incident light defined as

$$\begin{aligned} J^{\mathsf{inc}} &\equiv \frac{2}{E_0^2}\begin{bmatrix} J_{xx} & J_{xy} \\ J_{yx} & J_{yy} \end{bmatrix} \\ &= \begin{bmatrix} 1+P\cos 2\chi\cos 2\psi & P\left(\cos 2\chi\sin 2\psi + i\sin 2\chi\right) \\ P\left(\cos 2\chi\sin 2\psi - i\sin 2\chi\right) & 1-P\cos 2\chi\cos 2\psi \end{bmatrix} \end{aligned}$$

2. forming the G_k matrix for the scattered light by combining eqs. 1.13 and 1.22 to obtain

$$G_k = T\alpha_k^{0,\mathsf{mol}} T^T R A^T J^{\mathsf{inc}} A R^T T \left(\alpha_k^{0,\mathsf{mol}}\right)^\dagger T^T$$

3. forming the Stokes vector $\mathbf{S}_k^{\mathsf{scat}}$ for the scattered light via

$$\mathbf{S}_k^{\mathsf{scat}} = \begin{bmatrix} \mathbf{X}^T G_k \mathbf{X} + \mathbf{Y}^T G_k \mathbf{Y} \\ \mathbf{X}^T G_k \mathbf{X} - \mathbf{Y}^T G_k \mathbf{Y} \\ \mathbf{X}^T G_k \mathbf{Y} + \mathbf{Y}^T G_k \mathbf{X} \\ i\left(\mathbf{Y}^T G_k \mathbf{X} - \mathbf{X}^T G_k \mathbf{Y}\right) \end{bmatrix}$$

4. calculating the detected intensity using

$$I_k^S = \frac{\left(\omega_1-\omega_k\right)^4 N\hbar\mathcal{I}}{64\pi^2\epsilon_0^2 c^4\omega_k\left(1-e^{-\hbar\omega_k/k_B T}\right)}\mathbf{B}^T M^{\mathsf{det}}\mathbf{S}_k^{\mathsf{scat}} \tag{1.23}$$

$$I_k^{AS} = \frac{\left(\omega_1+\omega_k\right)^4 N\hbar\mathcal{I}}{64\pi^2\epsilon_0^2 c^4\omega_k\left(e^{\hbar\omega_k/k_B T}-1\right)}\mathbf{B}^T M^{\mathsf{det}}\mathbf{S}_k^{\mathsf{scat}} \tag{1.24}$$

$$I^{\mathsf{Ray}} = \frac{\omega_1^4\mathcal{I}}{32\pi^2\epsilon_0^2 c^4}\mathbf{B}^T M^{\mathsf{det}}\mathbf{S}^{\mathsf{scat}} \times \begin{cases} N^2 & \text{forward scattering} \\ N & \text{otherwise.} \end{cases} \tag{1.25}$$

Here we have used the relationship between the square of the real field amplitude E_0^2 and the irradiance $\mathcal{I}$,

$$\mathcal{I} = \frac{1}{2}\epsilon_0 c E_0^2,$$

to eliminate the E_0^2 in favor of $\mathcal{I}$. The factors in the Rayleigh intensity formula such as $\mathbf{S}^{\text{scat}}$ follow from calculations similar to those of the corresponding Raman values with k subscripts (such as $\mathbf{S}_k^{\text{scat}}$).

Sample vibrational Raman spectra from "oriented" toluene molecules are shown in Figs. 1.3 and 1.4. The perturbing electric field method (see Chapter 2) of calculating Raman spectra in Jaguar is used, using the B3LYP density functional and the Sadlej basis set. A Lorentzian lineshape having full-width half-maximum of $8\,\text{cm}^{-1}$ is fit to each peak. Note the typical "silent region" from 1700–3000 cm^{-1}.

1.8 Isotropic averaging and "isotropic" intensity formulas

We can obtain Raman and Rayleigh intensities due to scattering from isotropically oriented scatterers (i.e., tumbling, rotating scatterers as in a liquid or a gas or randomly oriented scatterers as in a powder or an amorphous solid) by performing isotropic averaging on the "oriented" intensity formulas above. We carry out the calculation below for the Stokes Raman intensity; the calculations for the anti-Stokes Raman and Rayleigh intensities are similar. We express the isotropic average of a quantity using angled brackets $\langle\rangle$. Starting from eq. 1.23:

$$\begin{aligned}
\left\langle I_k^S \right\rangle &= \frac{\left(\omega_1 - \omega_k\right)^4 N\hbar\mathcal{I}}{64\pi^2\epsilon_0^2 c^4 \omega_k \left(1 - e^{-\hbar\omega_k/k_B T}\right)} \mathbf{B}^T M^{\text{det}} \left\langle \mathbf{S}_k^{\text{scat}} \right\rangle \\
&= \frac{\left(\omega_1 - \omega_k\right)^4 N\hbar\mathcal{I}}{64\pi^2\epsilon_0^2 c^4 \omega_k \left(1 - e^{-\hbar\omega_k/k_B T}\right)} \mathbf{B}^T M^{\text{det}} \begin{bmatrix} \left\langle \mathbf{X}^T G_k \mathbf{X} + \mathbf{Y}^T G_k \mathbf{Y} \right\rangle \\ \left\langle \mathbf{X}^T G_k \mathbf{X} - \mathbf{Y}^T G_k \mathbf{Y} \right\rangle \\ \left\langle \mathbf{X}^T G_k \mathbf{Y} + \mathbf{Y}^T G_k \mathbf{X} \right\rangle \\ \left\langle i\left(\mathbf{Y}^T G_k \mathbf{X} - \mathbf{X}^T G_k \mathbf{Y}\right) \right\rangle \end{bmatrix} \\
&= \frac{\left(\omega_1 - \omega_k\right)^4 N\hbar\mathcal{I}}{64\pi^2\epsilon_0^2 c^4 \omega_k \left(1 - e^{-\hbar\omega_k/k_B T}\right)} \mathbf{B}^T M^{\text{det}} \begin{bmatrix} \mathbf{X}^T \left\langle G_k \right\rangle \mathbf{X} + \mathbf{Y}^T \left\langle G_k \right\rangle \mathbf{Y} \\ \mathbf{X}^T \left\langle G_k \right\rangle \mathbf{X} - \mathbf{Y}^T \left\langle G_k \right\rangle \mathbf{Y} \\ \mathbf{X}^T \left\langle G_k \right\rangle \mathbf{Y} + \mathbf{Y}^T \left\langle G_k \right\rangle \mathbf{X} \\ i\left(\mathbf{Y}^T \left\langle G_k \right\rangle \mathbf{X} - \mathbf{X}^T \left\langle G_k \right\rangle \mathbf{Y}\right) \end{bmatrix}.
\end{aligned}$$

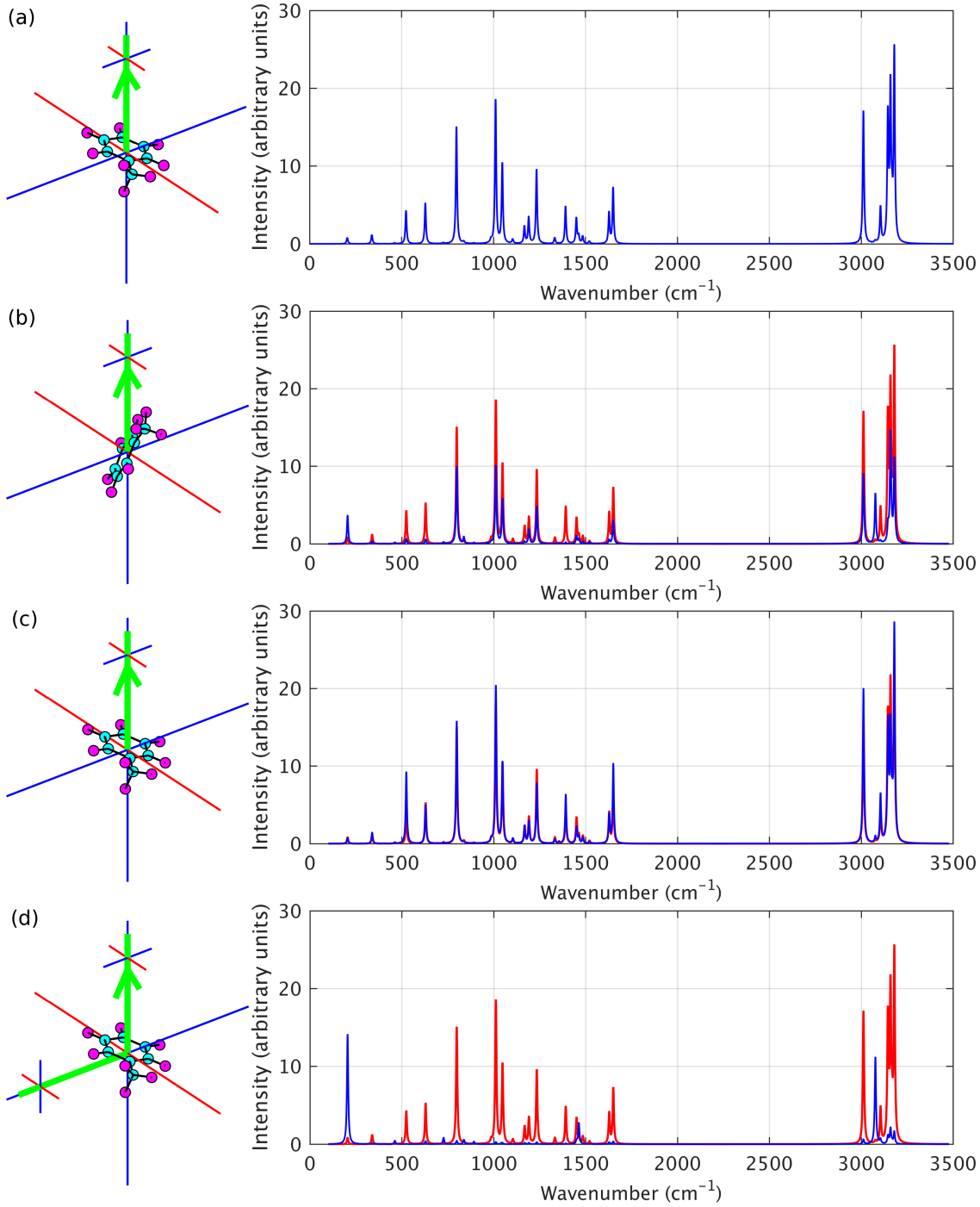

Figure 1.3: "Oriented" Raman spectra of a toluene molecule under various conditions. The rest of the caption is located in the caption of Fig. 1.4.

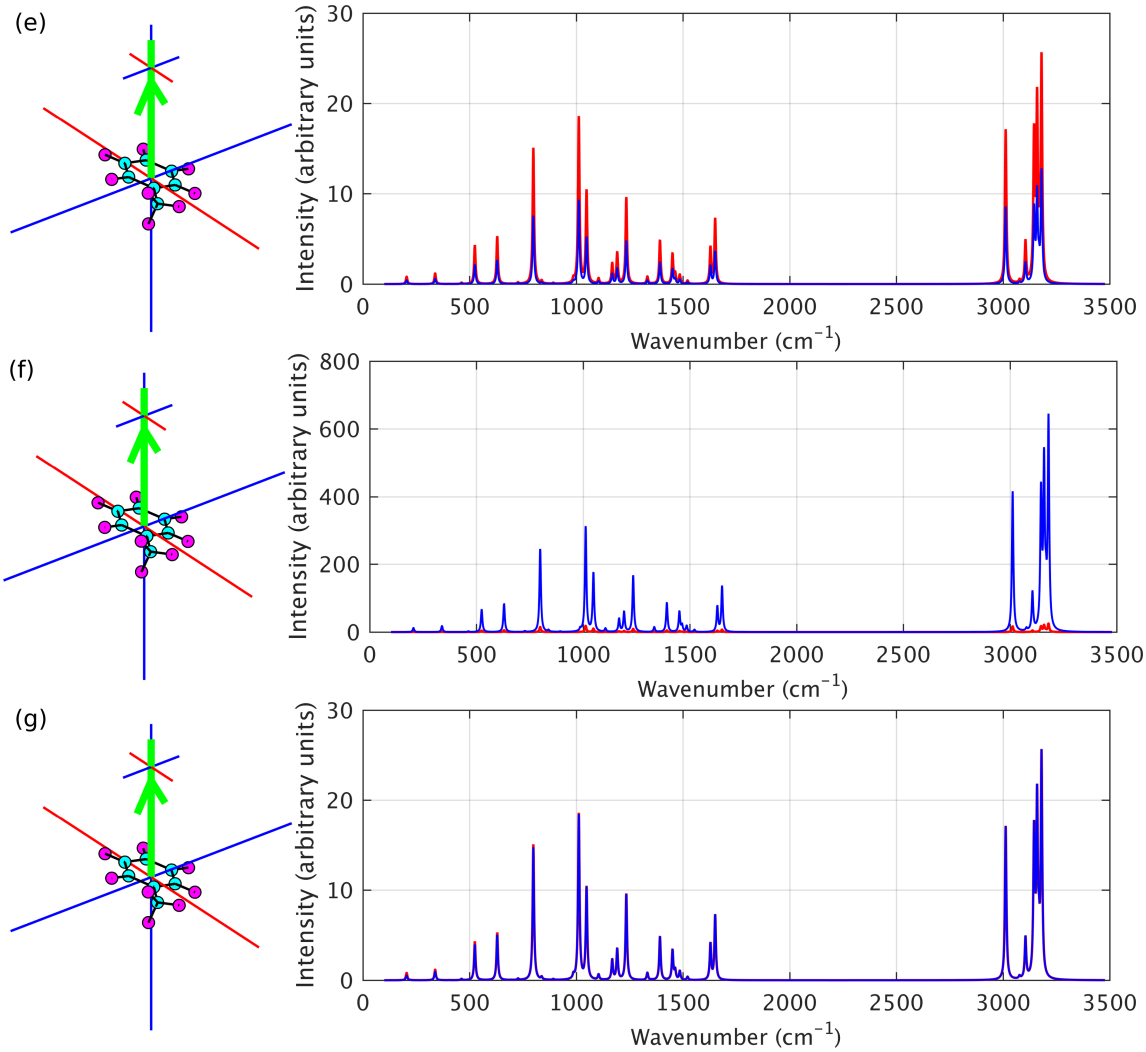

Figure 1.4: "Oriented" Raman spectra of a toluene molecule under various conditions, ctd. This caption encapsulates Fig. 1.3 as well. (a) 740-nm incident light linearly polarized in the scattering plane, backscattering, no analyzer in front of the detector, 300 K. The molecule orientation in the lab frame is always shown on the left, with the incident and scattered light shown in green. The green arrow always points toward the detector, which is always located in the $+Z$-direction in the lab frame. The sample is always located at the origin of the lab frame. The small sets of axes represent the polarization frames of the incident and scattered light. The larger blue axes represent the scattering plane. (b) Same conditions as (a) except for the molecule orientation as shown. The new spectra are always shown in blue and the spectra from Panel (a) are always shown in red. (c) Same conditions as (a) except the incident light is right-handed circularly polarized. (d) Same conditions as (a) except for 90° scattering, as shown on the left. (e) Same conditions as Fig. 1.3a except with a left-handed circular polarization analyzer placed in front of the detector. In this case, the presence of the analyzer halves the spectra. (f) Same conditions as Fig. 1.3a except with incident light of 380 nm. (g) Same conditions as Fig. 1.3a except at a temperature of 1 K. The Raman calculations employed the perturbing field method with a step size of 0.018 a.u. (`refield`=0.018) for the diagonal components and 0.014 a.u. (`refield`=0.014) for the off-diagonal components.

We see that we need to perform isotropic averaging over the G_k matrix, which is easier done in component form:

$$\begin{aligned}
\langle (G_k)_{al} \rangle &= \left\langle T_{ab} \left(\alpha_k^{0,\mathsf{mol}}\right)_{bc} T_{cd}^T R_{de} A_{ef}^T J_{fg}^{\mathsf{inc}} A_{gh} R_{hi}^T T_{ij} \left(\alpha_k^{0,\mathsf{mol}}\right)^\dagger_{jk} T_{kl}^T \right\rangle \\
&= \left(\alpha_k^{0,\mathsf{mol}}\right)_{bc} R_{de} A_{ef}^T J_{fg}^{\mathsf{inc}} A_{gh} R_{hi}^T \left(\alpha_k^{0,\mathsf{mol}}\right)^\dagger_{jk} \left\langle T_{ab} T_{cd}^T T_{ij} T_{kl}^T \right\rangle \\
&= \left(\alpha_k^{0,\mathsf{mol}}\right)_{bc} R_{de} A_{ef}^T J_{fg}^{\mathsf{inc}} A_{gh} R_{hi}^T \left(\alpha_k^{0,\mathsf{mol}}\right)^\dagger_{jk} \left\langle T_{ab} T_{dc} T_{ij} T_{lk} \right\rangle \\
&= \left(\alpha_k^{0,\mathsf{mol}}\right)_{bc} R_{de} A_{ef}^T J_{fg}^{\mathsf{inc}} A_{gh} R_{hi}^T \left(\alpha_k^{0,\mathsf{mol}}\right)^\dagger_{jk} \left\langle l_{ab} l_{dc} l_{ij} l_{lk} \right\rangle \\
&= F_{bcdijk} H_{abcdijkl},
\end{aligned}$$

where

$$\begin{aligned}
F_{bcdijk} &\equiv \left(\alpha_k^{0,\mathsf{mol}}\right)_{bc} R_{de} A_{ef}^T J_{fg}^{\mathsf{inc}} A_{gh} R_{hi}^T \left(\alpha_k^{0,\mathsf{mol}}\right)^\dagger_{jk} \\
&= \left(\alpha_k^{0,\mathsf{mol}}\right)_{bc} \left(\alpha_k^{0,\mathsf{mol}}\right)^\dagger_{jk} \sum_{efgh} R_{de} A_{ef}^T J_{fg}^{\mathsf{inc}} A_{gh} R_{hi}^T
\end{aligned}$$

and

$$\begin{aligned}
H_{abcdijkl} &\equiv \langle l_{ab} l_{dc} l_{ij} l_{lk} \rangle \\
&= \frac{\int l_{ab} l_{dc} l_{ij} l_{lk} \sin\beta d\beta d\alpha d\gamma}{\int \sin\beta d\beta d\alpha d\gamma}.
\end{aligned}$$

The $\{H_{abcdijkl}\}$ are isotropic averages of quadratic products of direction cosines. Note that in these calculations we take the polarizability $\alpha^{0,\mathsf{mol}}$ and its derivatives $\alpha_k^{0,\mathsf{mol}}$ to be real (and symmetric), as is the case for the conditions stated in Section 1.2. Proceeding with the aid of Mathematica, we obtain

$$\begin{aligned}
\left\langle \mathbf{S}_k^{\mathsf{scat}} \right\rangle &= \begin{bmatrix} \mathbf{X}^T \langle G_k \rangle \mathbf{X} + \mathbf{Y}^T \langle G_k \rangle \mathbf{Y} \\ \mathbf{X}^T \langle G_k \rangle \mathbf{X} - \mathbf{Y}^T \langle G_k \rangle \mathbf{Y} \\ \mathbf{X}^T \langle G_k \rangle \mathbf{Y} + \mathbf{Y}^T \langle G_k \rangle \mathbf{X} \\ i\left(\mathbf{Y}^T \langle G_k \rangle \mathbf{X} - \mathbf{X}^T \langle G_k \rangle \mathbf{Y}\right) \end{bmatrix} \\
&= \frac{2}{45} \begin{bmatrix} \frac{1}{4}\left[\left(135a_k^2 + 27\gamma_k^2 + 25\delta_k^2\right)\right] \\ + \left(45a_k^2 + \gamma_k^2 - 5\delta_k^2\right)\left(\cos 2\theta - 2P\cos 2\chi \cos 2\psi \sin^2\theta\right) \\ \\ \frac{1}{4}\left(45a_k^2 + \gamma_k^2 - 5\delta_k^2\right)\left(\left(3 + \cos 2\theta\right) P \cos 2\chi \cos 2\psi - 2\sin^2\theta\right) \\ \left(45a_k^2 + \gamma_k^2 - 5\delta_k^2\right) P \cos\theta \cos 2\chi \sin 2\psi \\ \left(45a_k^2 - 5\gamma_k^2 + 5\delta_k^2\right) P \cos\theta \sin 2\chi \end{bmatrix},
\end{aligned} \tag{1.26}$$

where the isotropic invariants of the polarizability derivative tensor $\alpha_k^{0,\mathsf{mol}}$ have been defined as[1]

- the mean polarizability a_k:

$$a_k \equiv \frac{1}{3}\left(\alpha_{xx,k} + \alpha_{yy,k} + \alpha_{zz,k}\right) \tag{1.27}$$

- the anisotropy γ_k:

$$\begin{aligned}\gamma_k^2 \equiv\ & \frac{1}{2}\left(\left|\alpha_{xx,k} - \alpha_{yy,k}\right|^2 + \left|\alpha_{yy,k} - \alpha_{zz,k}\right|^2 + \left|\alpha_{zz,k} - \alpha_{xx,k}\right|^2\right) \\ & + \frac{3}{4}\left(\left|\alpha_{xy,k} + \alpha_{yx,k}\right|^2 + \left|\alpha_{xz,k} + \alpha_{zx,k}\right|^2 + \left|\alpha_{yz,k} + \alpha_{zy,k}\right|^2\right)\end{aligned} \tag{1.28}$$

- the antisymmetric anisotropy δ_k:

$$\delta_k^2 \equiv \frac{3}{4}\left(\left|\alpha_{xy,k} - \alpha_{yx,k}\right|^2 + \left|\alpha_{yz,k} - \alpha_{zy,k}\right|^2 + \left|\alpha_{zx,k} - \alpha_{xz,k}\right|^2\right). \tag{1.29}$$

In summary, we can obtain the Raman and Rayleigh intensities due to isotropically oriented scatterers by (1) calculating the isotropic invariants a_k, γ_k, and δ_k of eqs. 1.27 to 1.29, (2) forming the scattered Stokes vector $\langle \mathbf{S}_k^{\mathsf{scat}} \rangle$ of eq. 1.26, and (3) calculating

$$\left\langle I_k^S \right\rangle = \frac{\left(\omega_1 - \omega_k\right)^4 N\hbar\mathcal{I}}{64\pi^2\epsilon_0^2 c^4 \omega_k \left(1 - e^{-\hbar\omega_k/k_B T}\right)} \mathbf{B}^T M^{\mathsf{det}} \left\langle \mathbf{S}_k^{\mathsf{scat}} \right\rangle \tag{1.30}$$

$$\left\langle I_k^{AS} \right\rangle = \frac{\left(\omega_1 + \omega_k\right)^4 N\hbar\mathcal{I}}{64\pi^2\epsilon_0^2 c^4 \omega_k \left(e^{\hbar\omega_k/k_B T} - 1\right)} \mathbf{B}^T M^{\mathsf{det}} \left\langle \mathbf{S}_k^{\mathsf{scat}} \right\rangle \tag{1.31}$$

$$\left\langle I^{\mathsf{Ray}} \right\rangle = \frac{\omega_1^4 \mathcal{I}}{32\pi^2\epsilon_0^2 c^4} \mathbf{B}^T M^{\mathsf{det}} \left\langle \mathbf{S}^{\mathsf{scat}} \right\rangle \times \begin{cases} N^2 & \text{forward scattering} \\ N & \text{otherwise.} \end{cases} \tag{1.32}$$

Sample vibrational Raman spectra from "isotropic" toluene molecules are shown in Figs. 1.5 and 1.6. The same calculation and plotting procedures are used here as for the "oriented" spectra above. Finally, we note that while the intensity formulas derived above technically apply only to non-chiral systems[1], they are nevertheless an excellent approximation to chiral systems, as shown in Chapter 3.

1.9 Why modern-day Raman spectroscopy often fails to fully capitalize on the Raman effect

Nowadays the most common Raman scattering apparatus is the Raman microscope, and it appears to have led to an under-appreciation of an important property of Raman scattering: dependence on the scattering angle and the

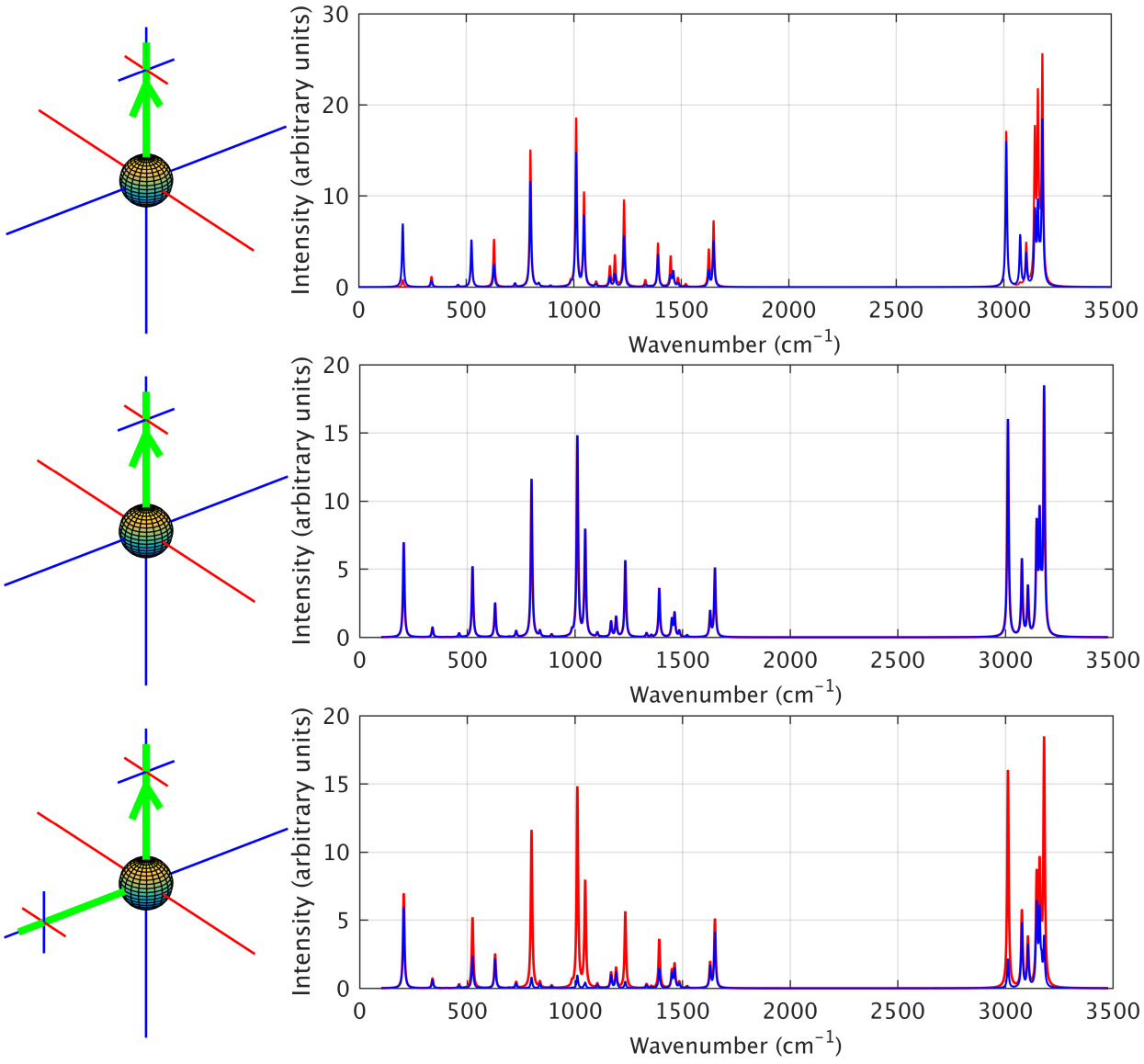

Figure 1.5: "Isotropic" Raman spectra of a toluene molecule under various conditions. (a) 740-nm incident light linearly polarized in the scattering plane (blue axes), backscattering, no analyzer in front of the detector, 300 K. The ellipsoid at left represents the fact that the molecules are isotropically oriented or tumbling isotropically. The corresponding crystalline spectra from Fig. 1.3a are shown in red. (b) Same conditions as (a) except the incident light is right-handed circularly polarized. The new spectra are always shown in blue and the spectra from Panel (a) are always shown in red. In this case, the spectra are the same. (c) Same conditions as (a) except for 90° scattering.

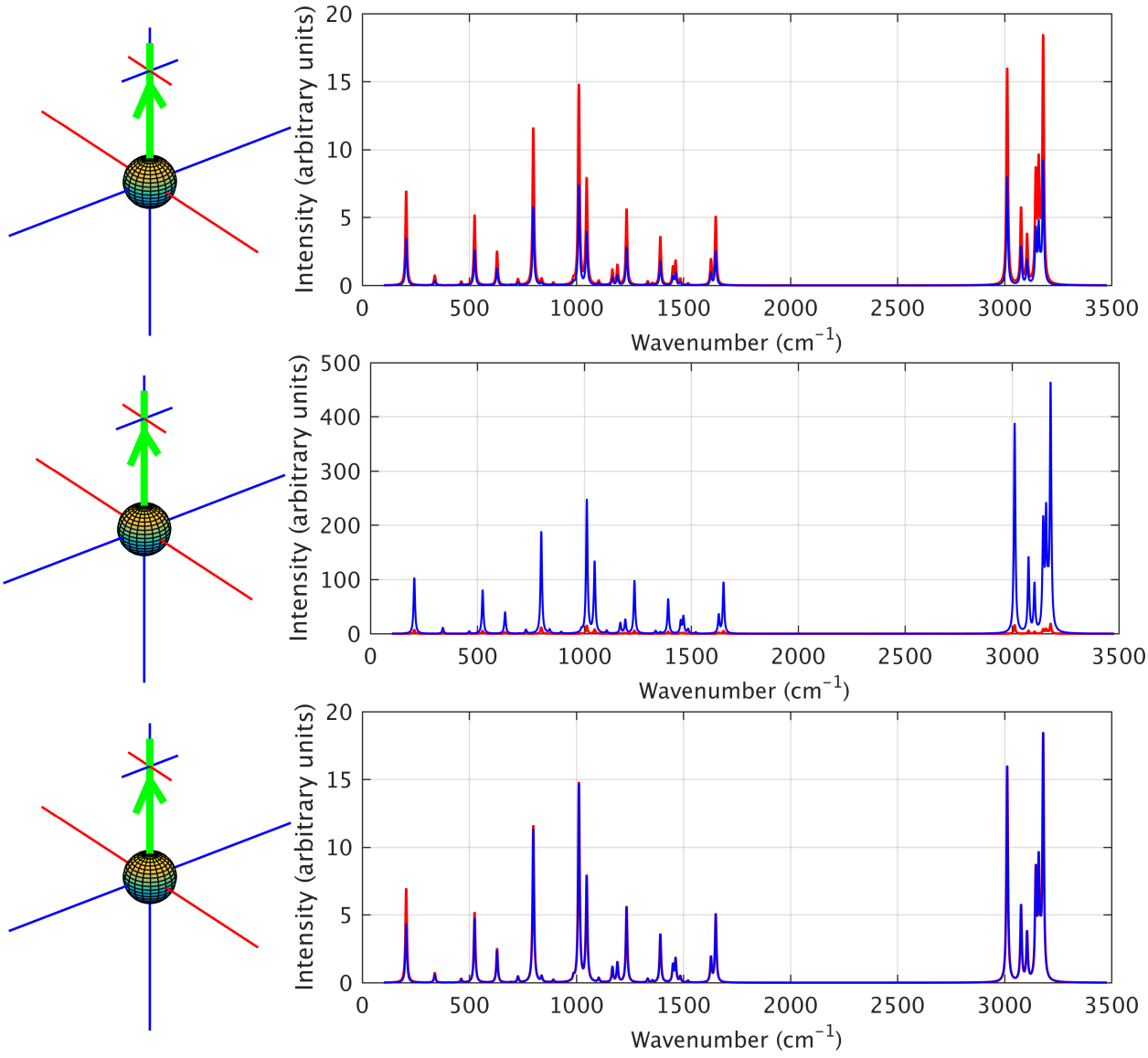

Figure 1.6: "Isotropic" Raman spectra of a toluene molecule under various conditions, ctd. (a) Same conditions as Fig. 1.5a except with a left-handed circular polarization analyzer placed in front of the detector. In this case, the presence of the analyzer halves the spectra. (b) Same conditions as Fig. 1.5a except with incident light of 380 nm. (c) Same conditions as Fig. 1.5a except at a temperature of 1 K.

polarization properties of the incident and scattered light. C.V. Raman himself was excited by the polarization dependence of the effect that became his namesake[6], and it is worth understanding why it is important and what can be done to once again take advantage of these special properties.

An important advantage of the general formulation of the Raman intensity formulas above is that they allow such an understanding to easily be obtained. We see from the final intensity formulas of eqs. 1.23 to 1.25 and 1.30 to 1.32 that, ignoring for clarity the coefficients of the form $\frac{(\omega_1-\omega_k)^4 N\hbar\mathcal{I}}{64\pi^2\epsilon_0^2 c^4 \omega_k \left(1-e^{-\hbar\omega_k/k_B T}\right)}$, all Raman (and Rayleigh) intensities result from the multiplication of

1. a 1×4 vector $\mathbf{B}^T = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$
2. a 4×4 Mueller matrix M^{det} representing the polarization analyzer (or lack thereof) placed in front of the detector
3. a 4×1 Stokes vector (expressed as $\mathbf{S}_k^{\text{scat}}, \mathbf{S}^{\text{scat}}, \langle \mathbf{S}_k^{\text{scat}} \rangle, \langle \mathbf{S}^{\text{scat}} \rangle$ in the intensity equations) representing all the polarization properties of the scattered light. For clarity in what follows, we will represent this Stokes vector simply as $\mathbf{S}^{\text{scat}}$.

Note that the multiplication by $\mathbf{B}^T$ only serves to pick out the first component of the vector $M^{\text{det}}\mathbf{S}^{\text{scat}}$. This means that if no analyzer is placed in front of the detector, in which case M^{det} is the 4×4 unit matrix M^{null}, the detected Raman intensity is directly proportional to the first component of $\mathbf{S}^{\text{scat}}$. For the case of scattering from an isotropic material, we see from eq. 1.26 that this component (traditionally called the zeroth component) is

$$\left(\mathbf{S}^{\text{scat}}\right)_0 = \frac{2}{45}\times\frac{1}{4}\left[\left(135a_k^2+27\gamma_k^2+25\delta_k^2\right)+\left(45a_k^2+\gamma_k^2-5\delta_k^2\right)\left(\cos 2\theta - 2P\cos 2\chi\cos 2\psi\sin^2\theta\right)\right].$$

The important thing to note is that all the polarization dependence of the incident light, i.e., the parameters $\{P,\chi,\psi\}$, are multiplied by $\sin^2\theta$, where θ is the scattering angle. This means that when θ is 0 or π — as occurs, for example, in a Raman microscope, in which the scattering is backscattering — *all polarization dependence of the incident light drops out*, since the factor $2P\cos 2\chi\cos 2\psi$ is multiplied by zero. In this case, the detected Raman intensity is proportional to

$$\begin{aligned}\left(\mathbf{S}^{\text{scat}}\right)_0 &= \frac{2}{45}\times\frac{1}{4}\left[\left(135a_k^2+27\gamma_k^2+25\delta_k^2\right)+\left(45a_k^2+\gamma_k^2-5\delta_k^2\right)\right]\\ &= \frac{2}{45}\times\frac{1}{4}\left(180a_k^2+28\gamma_k^2+20\delta_k^2\right)\\ &= \frac{2}{45}\left(45a_k^2+7\gamma_k^2+5\delta_k^2\right).\end{aligned}$$

Mutiplying this by the constant factor $\frac{45}{2}$, we obtain by far the most common Raman "intensity" formula in literature,

$$I_k = 45a_k^2 + 7\gamma_k^2 + 5\delta_k^2. \tag{1.33}$$

Experimentally, while this intensity formula accurately describes the scattering that takes place in a Raman microscope, its lack of dependence on $\{P, \chi, \psi\}$ makes it useless for backing out information about vibrational mode k the sample itself, which for isotropic samples is equivalent to the values of the isotropic invariants a_k, γ_k, and δ_k. All that can be obtained by an intensity measurement is the linear combination $45a_k^2 + 7\gamma_k^2 + 5\delta_k^2$; additional intensity measurements at different $\{P, \chi, \psi\}$ will obviously add nothing toward determination of the three unknowns $\{a_k, \gamma_k, \delta_k\}$ because the intensity formula has no dependence on $\{P, \chi, \psi\}$. Put another way, ideally we would expect that three intensity measurements $\left\{I_k^{(1)}, I_k^{(2)}, I_k^{(3)}\right\}$ under different experimental conditions (i.e., using three different sets of $\{P, \chi, \psi\}$) would allow determination of the three unknowns $\{a_k, \gamma_k, \delta_k\}$,

$$\begin{aligned} I_k^{(1)} &= D_{11}a_k^2 + D_{12}\gamma_k^2 + D_{13}\delta_k^2 \\ I_k^{(2)} &= D_{21}a_k^2 + D_{22}\gamma_k^2 + D_{23}\delta_k^2 \\ I_k^{(3)} &= D_{31}a_k^2 + D_{32}\gamma_k^2 + D_{33}\delta_k^2, \end{aligned}$$

which in matrix form is

$$D \begin{bmatrix} a_k^2 \\ \gamma_k^2 \\ \delta_k^2 \end{bmatrix} = \begin{bmatrix} I_k^{(1)} \\ I_k^{(2)} \\ I_k^{(3)} \end{bmatrix},$$

but we would find for $\theta = \{0, \pi\}$ that the rank of the data matrix D is 1; all three equations would be linearly dependent on each other. This can be shown, for example, in Mathematica.

By expanding the manipulatable experimental conditions to include the scattering angle θ, i.e., $\{P, \chi, \psi\} \rightarrow \{P, \chi, \psi, \theta\}$, we could obtain two independent equations for the unknowns $\{a_k, \gamma_k, \delta_k\}$. For example, for 90° scattering, in which $\theta = \frac{\pi}{2}$, the intensity formula becomes

$$I_k = \frac{1}{4}\left[\left(135a_k^2 + 27\gamma_k^2 + 25\delta_k^2\right) - \left(45a_k^2 + \gamma_k^2 - 5\delta_k^2\right)(1 + 2P\cos 2\chi \cos 2\psi)\right]. \tag{1.34}$$

Now it is much easier to determine two sets of $\{P, \chi, \psi\}$ whose corresponding intensity measurements lead to two independent equations for the three unknowns $\{a_k, \gamma_k, \delta_k\}$. However, even if we used different $\{\theta\}$ for many different intensity measurements, we would still find that the best we could obtain is a rank-two data matrix D.

Only if we place different polarization analyzers in front of the detector — thus expanding our manipulatable experimental conditions to $\left\{P, \chi, \psi, \theta, M^{\mathsf{det}}\right\}$ — can we obtain all possible information about the sample; i.e., can we make three intensity measurements in order to obtain three *independent* equations for the three unknowns $\{a_k, \gamma_k, \delta_k\}$. It can be shown (in Mathematica, for instance) that this requires that at least one of the measurements involves a circular-polarization analyzer placed in front of the detector. One example set of experimental conditions is

$$\begin{aligned}
&\text{set 1: } \left\{P = 1, \chi = 0, \psi = 0, \theta = 0, M^{\mathsf{det}} = M^{\mathsf{null}}\right\} \\
&\text{set 2: } \left\{P = 1, \chi = 0, \psi = 0, \theta = \frac{\pi}{2}, M^{\mathsf{det}} = M^{\mathsf{null}}\right\} \\
&\text{set 3: } \left\{P = 1, \chi = \frac{\pi}{4}, \psi = 0, \theta = 0, M^{\mathsf{det}} = M^{\mathsf{right}}\right\}.
\end{aligned}$$

(M^{right} represents a right-handed circular polarization analyzer placed in front of the detector; see Appendix B.) In this case the three independent intensity formulas are

$$\begin{aligned}
I_k^{(1)} &= 45a_k^2 + 7\gamma_k^2 + 5\delta_k^2 \\
I_k^{(2)} &= 6\gamma_k^2 + 10\delta_k^2 \\
I_k^{(3)} &= 45a_k^2 + \gamma_k^2 + 5\delta_k^2,
\end{aligned}$$

and the data matrix of the coefficients is highly non-singular and easily rank-three. Inverting this matrix leads to convenient equations for all three unknowns.

Thus, we see that the common intensity formula of eq. 1.33 is applicable to a common yet physically limited set of experimental conditions (for example, a Raman microscope with no analyzer in front of the detector). The misconception that the Raman intensity is a single Raman spectrum — and in particular, one corresponding to this particular set of experimental conditions — appears to be reinforced both experimentally by non-utilization of the polarization and scattering angle dependence of Raman apparatuses (i.e., basic usage of a Raman microscope) and theoretically by implementation in electronic structure codes (including Gaussian, Q-Chem, and Jaguar) of the single intensity formula of eq. 1.33. This is to say nothing of the (less ubiquitous) lack of the coefficients of the form $\frac{(\omega_1 - \omega_k)^4 N \hbar \mathcal{I}}{64\pi^2 \epsilon_0^2 c^4 \omega_k \left(1 - e^{-\hbar\omega_k / k_B T}\right)}$ in the Raman "intensity" formulas implemented in these codes, though Jaguar has recently included this coefficient.

In addition to proposing a reason for the ubiquity of eq. 1.33 as *the* formula for the Raman intensity, the above example of determining the three isotropic invariants using three Raman intensity measurements was meant to

introduce all the experimental parameters that have an effect on the detected Raman intensity. While the example used isotropically-oriented molecules as the scattering material, the dependence of the Raman spectra on all these parameters holds of course for scattering from crystalline systems as well. The transition polarizability on which the Raman effect is based is a tensor in a physical frame, and this leads to beautiful polarization and scattering angle dependence of the spectra. The Raman intensity is essentially equivalent to a scattering cross section; we have shown that, for example, photons of particular polarizations will literally be Raman-scattered more frequently than photons of other polarizations. The experimental parameters leading to eq. 1.33 are just one of an enormous set of combinations of $\left\{P, \chi, \psi, \theta, M^{\mathsf{det}}\right\}$.

One common utility of experimental Raman spectroscopy is obtaining the wavenumbers of the vibrational (or rotational or electronic) modes present in the scattering system, so it is somewhat plausible that using the minimal Raman microscope or the limited Raman "intensity" of eq. 1.33 would suffice. However, there is no guarantee that all modes (or a particular mode of interest) would be Raman active or even simply intense enough under these experimental conditions to be observed in the first place. The only way be sure that the mode(s) of interest would reveal itself, if present, is to ensure that theoretically (e.g., using software), under the current experimental conditions ($\left\{P, \chi, \psi, \theta, M^{\mathsf{det}}\right\}$), the mode should be Raman active and observable. If it is found that the mode of interest is unobservable then that does not mean that Raman spectroscopy cannot detect the mode and instead another method (such as infrared absorption) should be used, but rather simply that the experimental conditions should be adjusted. Only after one finds that no experimental parameters would suffice should he or she move on to another method. However, the phase space of $\left\{P, \chi, \psi, \theta, M^{\mathsf{det}}\right\}$ is so large that it is likely that Raman spectroscopy — and all its associated advantages — would more than suffice.

The beauty of Raman spectroscopy lies in the richness of the spectra that can be obtained upon modifying *experimental* parameters such as $\left\{P, \chi, \psi, \theta, M^{\mathsf{det}}\right\}$. Nothing about the sample needs to be touched; only easily modifiable filters, analyzers, and scattering angles need to be adjusted. These special properties should be understood, appreciated, and applied more frequently in practice, and in this dissertation, we provide two examples of how this can be done. In Chapter 3, we study how Raman spectra can be used to determine the unit cell orientation of a single crystal, even using such a basic apparatus as a Raman microscope (however, modifying the orientation angle ψ of the incident linearly polarized light is key). In Chapter 4, we discuss how Raman spectroscopy can be used to definitively determine whether an arbitrary material is crystalline or isotropic — but only for 90° scattering. This should provide even more motivation for experimentalists and materials scientists to

invoke the richness of Raman spectroscopy by making simple modifications to their experimental setups.

Chapter 2

Implementation of polarized Raman spectra using the pseudospectral method

2.1 Introduction

While the simulated spectra of Chapter 1 (and those of Chapters 3 and 4) show very good agreement with experiment, none of them utilize the pseudospectral method for the required electronic structure calculations. They all employ the Sadlej pVTZ basis set because this basis set, in conjunction with the B3LYP density functional, had previously been shown to provide Raman intensities that agree well with experiment[7, 8]. However, since pseudospectral grids for the Sadlej basis set have not been implemented in Jaguar, the primary electronic structure software to which we have access and that can utilize the pseudospectral method, Jaguar ran all the jobs using the traditional method via an algorithm it calls PRISM.

It thus remains to be seen whether the pseudospectral method — and its associated proven speed-ups for electronic structure calculations[9–18] — can also provide accurate Raman intensities. However, such a study requires that a number of factors be considered and tested. First, as discussed in Section 1.9 and exemplified in the subsequent chapters, one cannot simply "compare Raman intensities" because there are countless intensity formulas corresponding to different experimental conditions (scattering angle and polarizations of the incident and detected light) to consider. Agreement between theory and experiment for one set of experimental conditions and the corresponding intensity formula certainly does not imply agreement in the intensities for a different set of experimental conditions. Instead, from Chapter 1 we saw that the fundamental quantity required for calculating

Raman intensities under *any* set of experimental conditions is the *normal polarizability derivative* $\frac{\partial\alpha}{\partial\mathbf{Q}}$, which is the derivative of the polarizability α with respect to the $3N_{\text{nuc}}-6$ normal coordinates $\{Q\}$.[1] $\frac{\partial\alpha}{\partial\mathbf{Q}}$ itself is obtained by transforming the *Cartesian polarizability derivative* $\frac{\partial\alpha}{\partial\mathbf{R}}$, which is that taken with respect to the $3N_{\text{nuc}}$ Cartesian coordinates $\{R\}$ of the nuclei, using the eigenvectors of the Hessian. Placing these eigenvectors into the columns of a $3N_{\text{nuc}}\times(3N_{\text{nuc}}-6)$ matrix we define as L, the transformation of the ij component of the polarizability $(i,j\in\{1,2,3\})$ is

$$\frac{\partial\alpha_{ij}}{\partial\mathbf{Q}}=L^T\frac{\partial\alpha_{ij}}{\partial\mathbf{R}}.$$

We can therefore further strip down the quantity of focus from $\frac{\partial\alpha}{\partial\mathbf{Q}}$ to $\frac{\partial\alpha}{\partial\mathbf{R}}$ in order to eliminate any propagation of errors in the eigenvectors L to the normal polarizability derivative $\frac{\partial\alpha}{\partial\mathbf{Q}}$.

The next factor to consider is how to calculate $\frac{\partial\alpha}{\partial\mathbf{R}}$, which, since the polarizability is defined as

$$\alpha\equiv\frac{\partial^2E}{\partial\mathbf{F}_1\partial\mathbf{F}_2},$$

where $\mathbf{F}_1$ and $\mathbf{F}_2$ are perturbing electric fields, can further be written as

$$\frac{\partial\alpha}{\partial\mathbf{R}}=\frac{\partial^3E}{\partial\mathbf{R}\partial\mathbf{F}_1\partial\mathbf{F}_2}.$$

From this form we introduce additional useful names for the Cartesian polarizability derivative $\frac{\partial\alpha}{\partial\mathbf{R}}$: the *Raman tensor*, the *Cartesian third derivative*, or simply the *third derivative*. From this form we see that $\frac{\partial\alpha}{\partial\mathbf{R}}$ can be calculated in six unique ways:

1. Analytic calculation of the energy E under $\mathcal{O}\left(3N_{\text{nuc}}\times3\times3\right)=\mathcal{O}\left(27N_{\text{nuc}}\right)$ different conditions (all combinations of $3N_{\text{nuc}}$ perturbed nuclear coordinates and nine pairs of perturbing electric fields).

2. Analytic calculation of the dipole moment

   $$\mu\equiv\frac{\partial E}{\partial\mathbf{F}}$$

   under $\mathcal{O}\left(3N_{\text{nuc}}\times3\right)=\mathcal{O}\left(9N_{\text{nuc}}\right)$ different conditions (all combinations of $3N_{\text{nuc}}$ perturbed nuclear coordinates and three perturbing electric fields).

3. Analytic calculation of the polarizability α under $\mathcal{O}\left(3N_{\text{nuc}}\right)$ different conditions (all $3N_{\text{nuc}}$ perturbed nuclear coordinates).

[1] All derivatives are taken about the equilibrium molecule geometry.

4. Analytic calculation of the (energy) gradient

$$\mathbf{g} \equiv \nabla E \equiv \frac{\partial E}{\partial \mathbf{R}}$$

under $\mathcal{O}(3 \times 3) = \mathcal{O}(9)$ different conditions (nine pairs of perturbing electric fields).

5. Analytic calculation of the atomic polar tensor

$$A \equiv \frac{\partial^2 E}{\partial \mathbf{R} \partial \mathbf{F}}$$

under $\mathcal{O}(3)$ different conditions (three perturbing electric fields).

6. Fully analytic calculation of the third derivative $\frac{\partial^3 E}{\partial \mathbf{R} \partial \mathbf{F}_1 \partial \mathbf{F}_2}$.

Dr. Yixiang Cao of Schrödinger Inc. suggested that the minimal CPU savings, if any, of the fully analytic calculation of the third derivative (Method 6) would not be worth the significant coding efforts required (third derivatives of the density functional theory [DFT] functionals), limiting us to one of the finite difference methods (Methods 1–5). Method 1 does not make too much sense because the quantities required for the other methods (e.g., μ, $\mathbf{g}$, α) have already been shown to be calculated accurately using the pseudospectral method and these methods require fewer such calculations under different conditions than Method 1[10, 18]. Method 5 is not feasible using Jaguar because analytic calculation of the atomic polar tensor A using the pseudospectral method has not been implemented in Jaguar for DFT without first calculating the Hessian $\frac{\partial^2 E}{\partial \mathbf{R}_1 \partial \mathbf{R}_2}$, which is relatively time-consuming.

This leaves us with Methods 2–4 as the plausible methods for calculating the Cartesian polarizability derivative, so we tested all three of these. We define Methods 2 and 3 as the *perturbing coordinate method*, in which Method 2 utilizes numerical polarizability calculations (set by `ipolar=1` in the `&gen` section of the Jaguar input file, which utilizes a finite field perturbation of 0.006 a.u.) and Method 3 utilizes analytic polarizability calculations (set by `ipolar=-1`). A numerical polarizability calculation in Jaguar employs analytic calculation of the dipole moment under $\mathcal{O}(3)$ different conditions (three perturbing electric fields), which is why Method 2 (analytic dipole moments) corresponds to the use of numerical polarizabilities. We define Method 4 as the *perturbing field method*, and we study the perturbing coordinate and perturbing field methods in turn.

First, however, we briefly summarize the pseudospectral method in order to explain why it advantageous to do all this in the first place, i.e., move beyond the traditional, fully spectral, analytic method for electronic structure calculations, using which the calculation of Raman spectra has already been implemented and well-tested.

2.2 Summary of the pseudospectral method for calculation of the two-electron integrals required for electronic structure calculations

In this section we first summarize the traditional method for calculating electronic structure, i.e., determining the ground-state electronic wavefunction. We do this so we can later see where particular quantities come from — starting from first principles — that the pseudospectral method approaches differently. With this our only goal, we leave out many details and in addition focus only on Hartree-Fock theory. Analogous modifications for other theories such as DFT can be inferred and literature can be consulted for more details.

In order to determine the "mechanics" of the electrons in a molecule, one must generally solve the time-independent Schrödinger equation

$$H\Psi = E\Psi$$

for the eigenfunctions $\{\Psi\}$ and eigenvalues $\{E\}$ of the electronic Hamiltonian

$$H\left(\{\mathbf{r}\};\{\mathbf{R}\}\right) = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{j=1}^{M}\frac{Z_j}{|\mathbf{r}_i - \mathbf{R}_j|} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}.$$

Any eigenfunction (wavefunction) can be used to calculate the corresponding eigenvalue (energy) using

$$E\left[\Psi\right] = \left\langle \Psi \left| H \right| \Psi \right\rangle,$$

where here E is expressed as a functional of Ψ.

It is a general theorem in partial differential equations that the smallest eigenvalue of H is that obtained by minimizing the quantity $\langle\Theta|H|\Theta\rangle$, where Θ is any function satisfying the boundary conditions of the system (which in our case means that it must be normalizable):

$$E\left[\Theta\right] = \left\langle \Theta \left| H \right| \Theta \right\rangle \geq E_0.$$

Thus, in the space of eigenfunctions of H,

$$E\left[\Psi\right] = \left\langle \Psi \left| H \right| \Psi \right\rangle \geq E_0.$$

Further, if the eigenfunctions depend on a set of parameters $\{\lambda\}$, then

$$\min_{\{\lambda\}} E\left[\Psi\left(\{\lambda\}\right)\right] \geq E_0,$$

affording a way of best approximating the ground-state wavefunction: simply vary the $\{\lambda\}$ until a minimum value of $E[\Psi]$ is obtained; the Ψ that produces this minimum is our best approximation to the ground-state wavefunction Ψ_0.

In single-determinant Hartree-Fock theory, the wavefunctions $\{\Psi\}$ are approximated by Slater determinants made up of single-particle spatial molecular orbitals $\{\psi(\mathbf{r})\}$ in order to simultaneously account for the indistinguishability of the electrons and the anti-symmetry principle for fermions. The molecular orbitals $\{\psi(\mathbf{r})\}$ then serve as the variational parameters of the wavefunction Ψ, so that

$$\min_{\{\psi(\mathbf{r})\}} E[\Psi(\{\psi(\mathbf{r})\})] \geq E_0.$$

By formally minimizing the energy functional $E[\Psi]$ with respect to the spatial orbitals $\{\psi(\mathbf{r})\}$ while constraining them to be orthonormal using Lagrange's method of undetermined multipliers, one finds that a minimum is obtained when the Hartree-Fock equation is satisfied[19]:

$$f(\mathbf{r})\psi(\mathbf{r}) = \varepsilon\psi(\mathbf{r}).$$

Here, the eigenfunctions of the Fock operator $f(\mathbf{r})$ are the $E[\Psi]$-minimizing spatial orbitals, the eigenvalues $\{\varepsilon\}$ represent the orbital energies, and $f(\mathbf{r})$ is

$$f(\mathbf{r}) = h(\mathbf{r}) + \sum_{b=1}^{N}[J_b(\mathbf{r}) - K_b(\mathbf{r})],$$

where

$$\begin{aligned} h(\mathbf{r}) &\equiv -\frac{1}{2}\nabla^2 - \sum_{j=1}^{M}\frac{Z_j}{|\mathbf{r}-\mathbf{R}_j|} \\ J_b(\mathbf{r}) &\equiv \int d\mathbf{r}'\frac{\psi_b^*(\mathbf{r}')\psi_b(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\ K_b(\mathbf{r})\psi_a(\mathbf{r}) &\equiv \left[\int d\mathbf{r}'\frac{\psi_b^*(\mathbf{r}')\psi_a(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\right]\psi_b(\mathbf{r}). \end{aligned}$$

Finally, in order to make the problem numerically tractable, the orbitals $\{\psi(\mathbf{r})\}$ are typically expanded in a basis $\{\phi(\mathbf{r})\}$,

$$\psi_i(\mathbf{r}) = \sum_{k=1}^{K} C_{ki}\phi_k(\mathbf{r}),$$

so that the matrix elements of the Fock operator in this basis are required so that the orbitals can be obtained by diagonalizing the corresponding Fock matrix. The matrix elements of the terms in the Fock operator are

$$\begin{aligned}
h_{ij} &\equiv \int d\mathbf{r}\phi_i^*(\mathbf{r})\left[h(\mathbf{r})\right]\phi_j(\mathbf{r}) \\
&= -\frac{1}{2}\int d\mathbf{r}\phi_i^*(\mathbf{r})\nabla^2\phi_j(\mathbf{r}) - \sum_{j=1}^{M} Z_j \int d\mathbf{r}\frac{\phi_i^*(\mathbf{r})\phi_j(\mathbf{r})}{|\mathbf{r}-\mathbf{R}_j|} \\
J_{ij} &\equiv \int d\mathbf{r}\phi_i^*(\mathbf{r})\left[\sum_{b=1}^{N} J_b(\mathbf{r})\right]\phi_j(\mathbf{r}) \\
&= \sum_{b=1}^{N}\int d\mathbf{r}d\mathbf{r}'\frac{\phi_i^*(\mathbf{r})\sum_{l=1}^{K} C_{lb}^*\phi_l^*(\mathbf{r}')\sum_{k=1}^{K} C_{kb}\phi_k(\mathbf{r}')\phi_j(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} \\
&= \sum_{kl=1}^{K} P_{kl}\int d\mathbf{r}d\mathbf{r}'\frac{\phi_i^*(\mathbf{r})\phi_l^*(\mathbf{r}')\phi_k(\mathbf{r}')\phi_j(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} \\
K_{ij} &\equiv \int d\mathbf{r}\phi_i^*(\mathbf{r})\left[\sum_{b=1}^{N} K_b(\mathbf{r})\right]\phi_j(\mathbf{r}) \\
&= \sum_{b=1}^{N}\int d\mathbf{r}d\mathbf{r}'\frac{\phi_i^*(\mathbf{r})\sum_{l=1}^{K} C_{lb}^*\phi_l^*(\mathbf{r}')\phi_j(\mathbf{r}')\sum_{k=1}^{K} C_{kb}\phi_k(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} \\
&= \sum_{kl=1}^{K} P_{kl}\int d\mathbf{r}d\mathbf{r}'\frac{\phi_i^*(\mathbf{r})\phi_l^*(\mathbf{r}')\phi_j(\mathbf{r}')\phi_k(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|},
\end{aligned}$$

where we have defined the density matrix

$$P_{kl} \equiv \sum_{b=1}^{N} C_{kb}C_{lb}^*.$$

To summarize, by calculating these matrix elements, forming the Fock matrix using $f_{ij} = h_{ij} + J_{ij} - K_{ij}$, and diagonalizing it to obtain the N spatial orbitals that make up the determinantal wavefunction, we are obtaining our best approximation to the ground-state wavefunction of the electronic Hamiltonian.

The problem with this traditional, analytic approach to determining the electronic structure of a system is the calculation and storage of the four-center, two-electron integrals in the expressions for J_{ij} and K_{ij}. For each of the K^2 elements of each of these matrices, where K is the number of basis functions used and generally scales linearly with the system size, K^2 such integrals must be calculated, so that the assembly of the Fock matrix generally scales with the fourth power of the system size. This is prohibitive for calculations on systems containing hundreds of atoms without the utilization of special techniques for significantly reducing the amount of work required.

One particularly successful such technique is the pseudospectral method, in which more up-front work results in significant speed-ups for calculations on large molecules. The essence of this method can be shown by re-arranging

the expressions above for J_{ij} and K_{ij} as

$$\begin{aligned} J_{ij} &= \int d\mathbf{r}\phi_i^*(\mathbf{r}) \left\{ \phi_j(\mathbf{r}) \sum_{k=1}^{K} \sum_{l=1}^{K} P_{kl} A_{lk}(\mathbf{r}) \right\} \\ K_{ij} &= \int d\mathbf{r}\phi_i^*(\mathbf{r}) \left\{ \sum_{k=1}^{K} \phi_k(\mathbf{r}) \sum_{l=1}^{K} P_{kl} A_{lj}(\mathbf{r}) \right\}, \end{aligned}$$

where we have defined the three-center, one-electron integral

$$A_{lk}(\mathbf{r}) \equiv \int d\mathbf{r}' \frac{\phi_l^*(\mathbf{r}')\phi_k(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}.$$

It has been found[11, 14] that the quantities inside the curly braces, each of which are expansions involving products of two functions of $\mathbf{r}$, can generally be sufficiently expanded in a set of well-chosen functions involving only a single function of $\mathbf{r}$. This new set of basis functions is called the auxiliary basis set $\{\chi(\mathbf{r})\}$ and generally must contain $L > K$ basis functions:

$$\begin{aligned} J_{ij} &\approx \int d\mathbf{r}\phi_i^*(\mathbf{r}) \left\{ \sum_{m=1}^{L} D_{mj}\chi_m(\mathbf{r}) \right\} \\ K_{ij} &\approx \int d\mathbf{r}\phi_i^*(\mathbf{r}) \left\{ \sum_{m=1}^{L} E_{mj}\chi_m(\mathbf{r}) \right\}. \end{aligned}$$

Once the expansion coefficients D_{mj} and E_{mj} are determined, the Coulomb and exchange matrix elements take on simple forms:

$$\begin{aligned} J_{ij} &= \sum_{m=1}^{L} D_{mj} \int d\mathbf{r}\phi_i^*(\mathbf{r})\chi_m(\mathbf{r}) \\ &= \sum_{m=1}^{L} T_{im} D_{mj} \\ &= (TD)_{ij} \\ \rightarrow J &= TD \\ K_{ij} &= \sum_{m=1}^{L} E_{mj} \int d\mathbf{r}\phi_i^*(\mathbf{r})\chi_m(\mathbf{r}) \\ &= \sum_{m=1}^{L} T_{im} E_{mj} \\ &= (TE)_{ij} \\ \rightarrow K &= TE, \end{aligned}$$

where we have defined the $K \times L$ overlap matrix between the auxiliary and original basis set as

$$T_{im} \equiv \int d\mathbf{r}\phi_i^*(\mathbf{r})\chi_m(\mathbf{r}).$$

The pseudospectral method calculates the expansion coefficients D_{mj} and E_{mj} by choosing a grid $\mathbf{r} \rightarrow \{\mathbf{r}_\alpha, \alpha \in \{1, \ldots, G\}\}$ on which the equalities in the curly braces above hold in all significant parts of physical Euclidean space:

$$\begin{aligned} \sum_{m=1}^{L} D_{mj}\chi_m(\mathbf{r}_\alpha) &\approx \phi_j(\mathbf{r}_\alpha) \sum_{k=1}^{K}\sum_{l=1}^{K} P_{kl}A_{lk}(\mathbf{r}_\alpha) \\ \sum_{m=1}^{L} E_{mj}\chi_m(\mathbf{r}_\alpha) &\approx \sum_{k=1}^{K} \phi_k(\mathbf{r}_\alpha) \sum_{l=1}^{K} P_{kl}A_{lj}(\mathbf{r}_\alpha). \end{aligned}$$

Once this grid is chosen, the desired coefficients D_{mj} and E_{mj} can be determined by, for example, a weighted least-squares procedure: defining the matrices

$$\begin{aligned} W_{\alpha j} &\equiv \phi_j(\mathbf{r}_\alpha) \\ R_{\alpha m} &\equiv \chi_m(\mathbf{r}_\alpha) \end{aligned}$$

and a $G \times G$ diagonal matrix w of grid weights, we have for the Coulomb term

$$\begin{aligned} \sum_{m=1}^{L} D_{mj}R_{\alpha m} &= W_{\alpha j}\sum_{k=1}^{K}\sum_{l=1}^{K} P_{kl}A_{lk\alpha} \\ (RD)_{\alpha j} &= W_{\alpha j}\mathsf{tr}\left[PA_\alpha\right] \\ &\equiv W_{\alpha j}\Omega_{\alpha\alpha} \\ &= (\Omega W)_{\alpha j} \\ \rightarrow RD &= \Omega W \\ R^\dagger wRD &= R^\dagger w\Omega W \\ D &= \left(R^\dagger wR\right)^{-1} R^\dagger w\Omega W, \end{aligned}$$

where we have defined Ω as the $G \times G$ diagonal matrix of the traces $\{\mathsf{tr}\left[PA_\alpha\right]\}$:

$$\Omega_{\alpha\beta} \equiv \mathsf{tr}\left[PA_\alpha\right]\delta_{\alpha\beta}.$$

For the exchange term, we have

$$\begin{aligned}
\sum_{m=1}^{L} E_{mj} R_{\alpha m} &= \sum_{k=1}^{K} W_{\alpha k} \sum_{l=1}^{K} P_{kl} A_{lj\alpha} \\
(RE)_{\alpha j} &= \sum_{k=1}^{K} W_{\alpha k} (PA)_{kj\alpha} \\
&= (WPA)_{\alpha j} \\
\rightarrow RE &= WPA \\
R^{\dagger} wRE &= R^{\dagger} wWPA \\
E &= \left(R^{\dagger} wR\right)^{-1} R^{\dagger} wWPA.
\end{aligned}$$

Thus, the expressions for the J and K matrices become

$$\begin{aligned}
J &= T \left(R^{\dagger} wR\right)^{-1} R^{\dagger} w\Omega W \\
&= Q\Omega W \\
K &= T \left(R^{\dagger} wR\right)^{-1} R^{\dagger} wWPA \\
&= QWPA,
\end{aligned}$$

where we have defined

$$Q \equiv T \left(R^{\dagger} wR\right)^{-1} R^{\dagger} w$$

as the $K \times G$ transformation matrix between the physical and spectral bases; J and K in the pseudospectral basis are ΩW and WPA, respectively, and in the purely spectral basis, they are $Q\Omega W$ and $QWPA$, respectively. We note that, importantly, Q is the same for every iteration of the SCF procedure since it does not depend on the density matrix P; this cuts down the total SCF time at the expense of the minimal calculation of Q beforehand.

In summary, whereas in the traditional method for calculating electronic structure, the significant four-center, two-electron integrals out of the total K^4 of them must be re-calculated at each iteration of the SCF procedure (calculating and storing them all beforehand is impossible for large systems), in the pseudospectral method only the significant three-center, one-electron integrals out of the total $K^2 G$ of them must be re-calculated at each SCF iteration, at the expense of up-front calculation of the grids and Q and additional *a priori* programming effort into the determination of the pseudospectral grid parameters and auxiliary basis sets. The end result on large systems, as well-demonstrated in the program Jaguar, has been a significant speed-up for all types of calculations (including energies, gradients, Hessians, dipole moments, polarizabilities, etc.) relative to programs

using traditional methods[9–18]. To date, however, the pseudospectral method has not been applied to the calculation of the quantity $\frac{\partial^3 E}{\partial \mathbf{R} \partial \mathbf{F}_1 \partial \mathbf{F}_2}$ and therefore to the calculation of Raman spectra, as explained in Section 2.1. Thus, below we study just this with the hopes of finding that the pseudospectral method is capable of providing accurate Raman spectra as well, so that the proven speed-ups that the method provides can be further extended to this important application.

2.3 Perturbing coordinate method for calculating the Raman tensor

We began testing the perturbing coordinate method by obtaining Gaussian-calculated data from Dr. Mat Halls, which included Raman activities apparently calculated using the common "intensity" formula of eq. 1.33 for six molecules (H_2O, H_2S, CH_2O, CH_2F_2, CH_2Cl_2, and N_2) using both the 6-31G* and Sadlej pVTZ basis sets and the B3LYP density functional. This is all the Gaussian data we were able to straightforwardly obtain and is what dictated the make-up of our test set of molecules and basis sets and the experimental conditions studied (those leading to eq. 1.33) for the perturbing coordinate method of calculating the Raman tensor. We further calculated the spectra of the same molecules using the same basis sets using Q-Chem, which we confirmed uses the intensity formula of eq. 1.33.[2] We took these Gaussian- and Q-Chem-calculated activities as our benchmark activities; both Gaussian and Q-Chem utilize the traditional spectral electronic structure method and have both been shown to produce accurate Raman spectra.

Then, we calculated the Raman activities using Methods 2 and 3 in Jaguar for the same molecules and basis sets. The Cartesian third derivative for Method 2 is

$$\frac{\partial^3 E}{\partial \mathbf{R} \partial \mathbf{F}_1 \partial \mathbf{F}_2} = \frac{\partial^2 \mu}{\partial \mathbf{R} \partial \mathbf{F}}$$

and that for Method 3 is

$$\frac{\partial^3 E}{\partial \mathbf{R} \partial \mathbf{F}_1 \partial \mathbf{F}_2} = \frac{\partial \alpha}{\partial \mathbf{R}}.$$

As mentioned in the previous section, these two forms of the third derivative are related by $\alpha = \frac{\partial \mu}{\partial \mathbf{F}}$, which corresponds to numerical calculation of the polarizability α (`ipolar=1`); the latter method implies fully analytic calculation of α (`ipolar=-1`). Functionally, we therefore differentiate between Methods 2 and 3 by using the latter form with different settings of the `ipolar` keyword in the `&gen` section. Then, by using a centered finite-difference

[2] Frequencies were not provided in the Gaussian data so for this we used the frequencies calculated by Q-Chem.

scheme, we approximate $\frac{\partial \alpha}{\partial \mathbf{R}}$ numerically in component form by

$$\frac{\partial \alpha_{ij}}{\partial R_k} \approx \frac{\left[\alpha_{ij}\left(R_k + \Delta R\right)\right]_0 - \left[\alpha_{ij}\left(R_k - \Delta R\right)\right]_0}{2\Delta R}, \tag{2.1}$$

where ΔR equals the value corresponding to `pertnd` and the form $\left[\alpha_{ij}\left(R_k + \Delta R\right)\right]_0$ means that the polarizability is calculated for the geometry in which ΔR is added to the kth nuclear coordinate and all other nuclear coordinates remain the same. Thus, for each of the $3N_{\text{nuc}}$ nuclear coordinates, we ran two polarizability jobs — one corresponding to $\left[\alpha_{ij}\left(R_k + \Delta R\right)\right]_0$ and the other corresponding to $\left[\alpha_{ij}\left(R_k - \Delta R\right)\right]_0$ — resulting in a total of $6N_{\text{nuc}}$ polarizability calculations for the perturbing coordinate method. Equation 2.1 was then used to calculate the required third derivative from the results, after which we calculated the Raman activities using eq. 1.33.

We note that all calculations here and throughout this chapter are performed on molecules in the gas phase, but at least based on the excellent agreement between all simulated and experimental spectra of Chapter 3, the simulations of which assumed gas-phase second (the Hessian) and third derivatives (not geometries though) and all substances of which were in either the liquid or solid phases, it appears the gas-phase second and third derivatives sufficiently represent those in the condensed phases as long as the correct molecule geometries are used. This confirms the finding in [20] that solvent effects do not significantly affect Raman spectra.

It should be noted that for the mode(s) around 1100 cm^{-1} for CH_2F_2 for the 6-31G* basis set, Q-Chem found two vibrational modes whereas Jaguar found only a single mode, but by adding activities for modes within 4 cm^{-1} we were then able to compare all calculated activities on a mode-by-mode basis. These raw activities for all six molecules are presented in Section C.2 for perturbing coordinate values of 0.05, 0.10, and 0.25 Bohr using both basis sets, and the corresponding intensities (for all but N_2), obtained by multiplying the activities by the appropriate frequency-dependent prefactors corresponding to incident light of 532 nm and room-temperature conditions, are plotted in Figs. 2.1 (6-31G*, 0.10), 2.2 (Sadlej, 0.10), C.1 (6-31G*, 0.05), C.2 (Sadlej, 0.05), C.3 (6-31G*, 0.25), and C.4 (Sadlej, 0.25).

All six of these figures correspond to backscattering with no polarization analyzer in front of the detector, 532-nm incident light, and room temperature conditions (293 K). Reported and calculated Raman activities were multiplied by the appropriate frequency-dependent prefactors in order to obtain Raman intensities that can be plotted and compared with experiment. Gaussian data are represented by blue lines and a square marker, Q-Chem data are represented by black lines and a diamond marker, Jaguar data calculated using Method 2 (numerical polarizabilities) are represented by green lines and a plus-sign marker, and Jaguar data calculated using Method 3 (analytic polarizabilities) are represented by red lines and a cross marker. We see that the Method 2-calculated

data almost always agree with the Gaussian and Q-Chem data, whereas the Method 3-calculated data sometimes disagree with the Gaussian and Q-Chem data. Note that Gaussian frequencies (not provided) were set equal to the Q-Chem-calculated frequencies, though frequency agreement is inconsequential anyway, as confirmed in [8]. A Lorentzian lineshape having full-width half-maximum of $8\,\mathrm{cm}^{-1}$ is fit to each peak.

The mean absolute deviations (MADs) of the Method 2 and Method 3 data were then independently calculated relative to the Q-Chem and Gaussian values separately. For each perturbing coordinate value, basis set, and molecule, the MAD was calculated over all vibrational modes of the molecule, and then these were averaged over all six molecules, the results of which are presented in Table 2.1. The results in the table confirm what can be expected from the raw activities and the figures: the Raman spectra calculated using Method 2 (numerical polarizabilities) in Jaguar are nearly always in excellent agreement with the benchmark data, whereas those calculated using Method 3 (analytic polarizabilities) in Jaguar sometimes disagree significantly with the benchmark data. For the `pertnd` value of 0.10 Bohr, which led to the most accurate results for both Methods 2 and 3, the MADs for Method 2 were 0.354 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$ relative to Q-Chem (0.369 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$ relative to Gaussian) for the 6-31G* basis set and 0.482 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$ (0.749 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$) for the Sadlej pVTZ basis set, which is clearly excellent given that activities typically vary from 0 to 200 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$. The MADs for Method 3 are not too much worse at 1.969 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$ (1.974 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$) for 6-31G* and 1.186 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$ (1.649 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$) for Sadlej, though this shows that while Method 3 is always faster than Method 2, the degradation of the Raman activities calculated using analytic polarizabilities is not insignificant; analytic polarizabilities should be used with caution. It is particularly interesting to note that whereas using larger coordinate perturbations has little if any effect on the activities calculated using Method 2, increasing the size of the perturbations results in Method 3-calculated activities that are in significant disagreement with the benchmark results: increasing `pertnd` from 0.10 Bohr to 0.25 Bohr increases the MAD for Method 3 by 8.2 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$ for 6-31G* and 8.6 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$ for Sadlej as opposed to 0.16 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$ for 6-31G* and 0.13 $\frac{\mathrm{Ang}^4}{\mathrm{amu}}$ for Sadlej for Method 2.

Fortunately, in the grand scheme of things these differences do not matter much because both Methods 2 and 3, i.e., the perturbing coordinate method, scale poorly with system size relative to the perturbing field method (Method 4). This is because the former directly scales with the system size (N_{nuc}) whereas the latter does not: Method 2 requires $\mathcal{O}\left(9N_{\mathrm{nuc}}\right)$ dipole moment calculations and Method 3 requires $\mathcal{O}\left(3N_{\mathrm{nuc}}\right)$ polarizability calculations, whereas Method 4 requires $\mathcal{O}\left(9\right)$ gradient calculations. For this reason, despite studying only the results of a single intensity formula on six small molecules, we end our analysis of the perturbing coordinate method here so that we can focus on the perturbing field method, which is potentially much faster. Nevertheless,

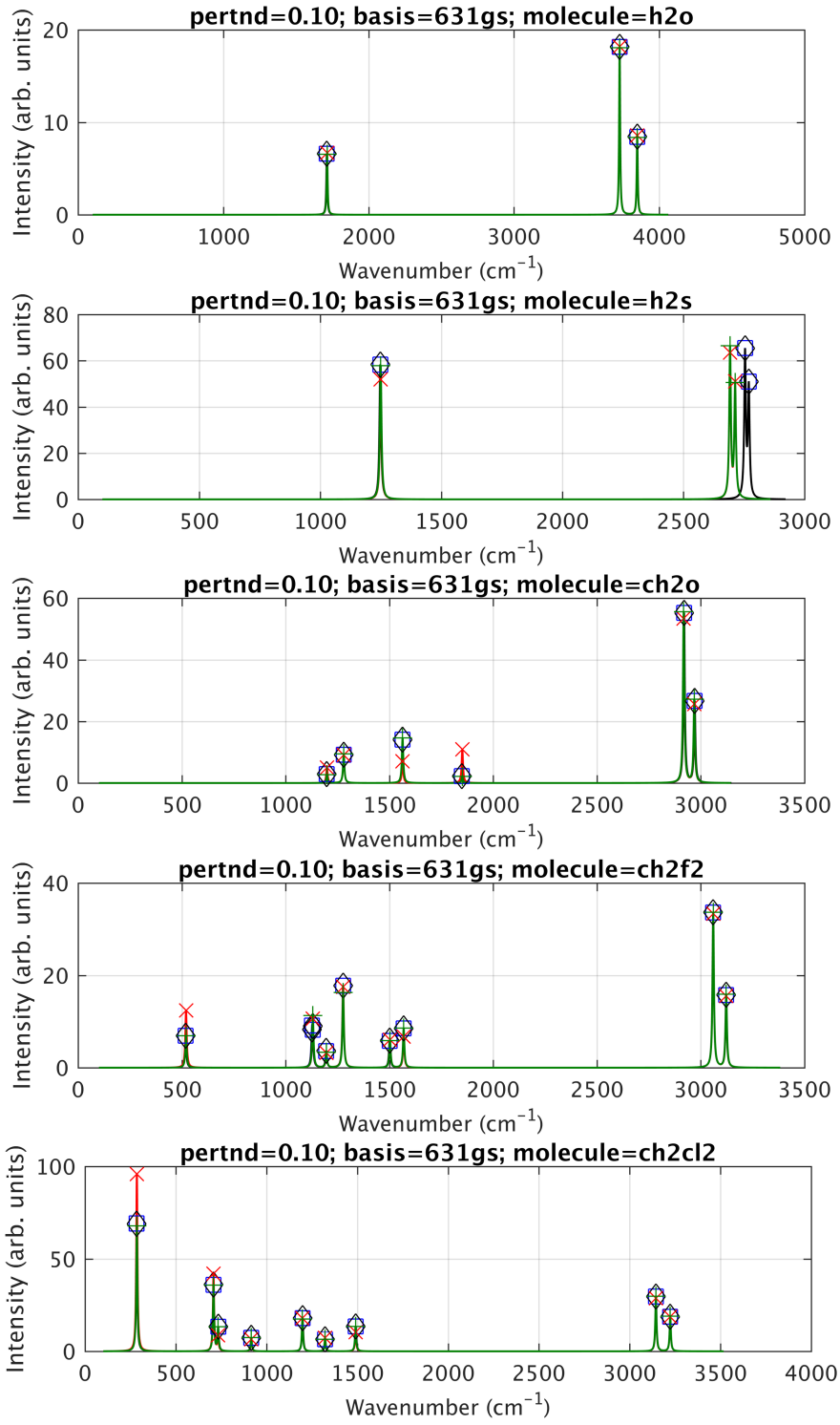

Figure 2.1: Calculated Raman spectra for five molecules using the 6-31G* basis set and the perturbing coordinate method with `pertnd=0.10`. See text for details.

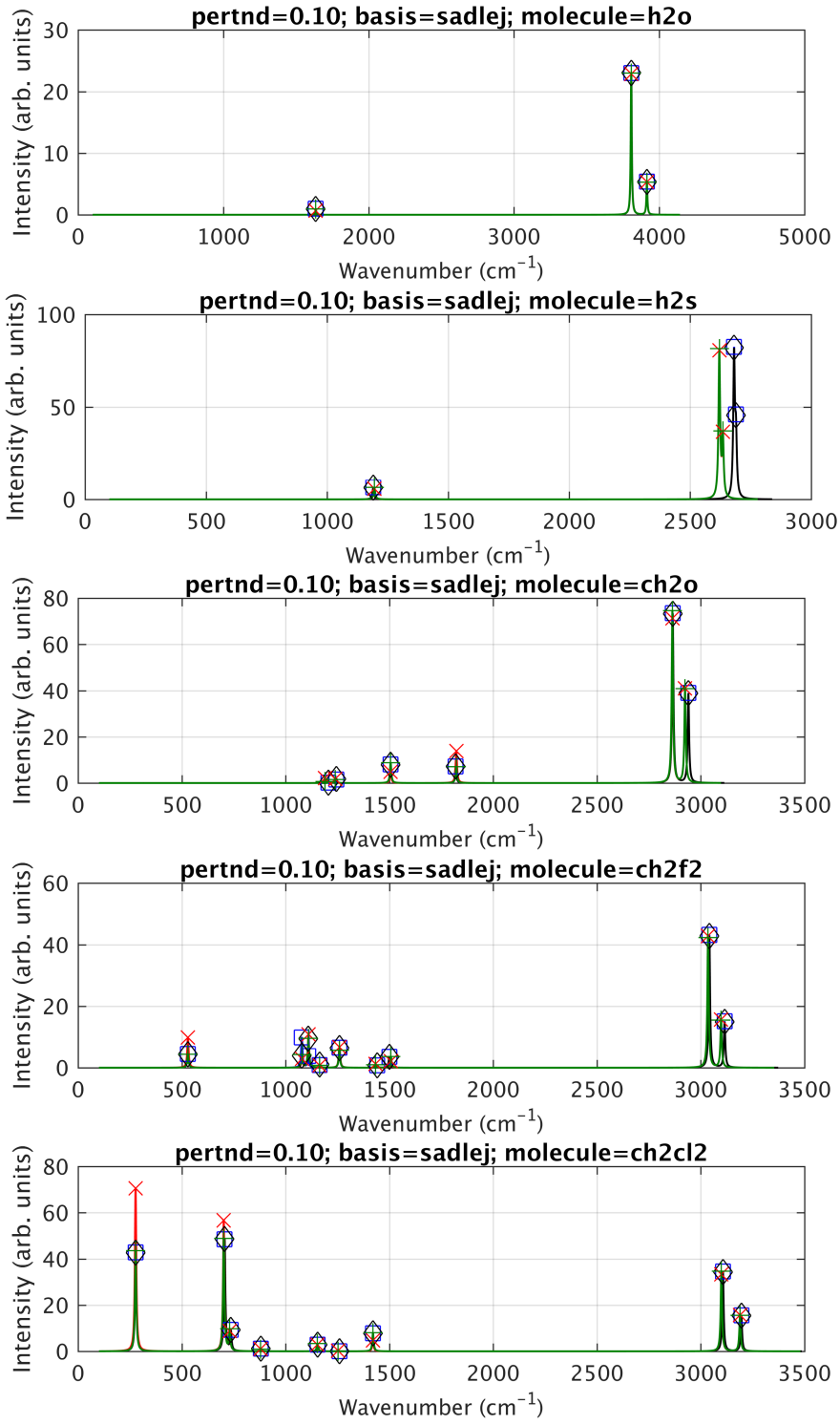

Figure 2.2: Calculated Raman spectra for five molecules using the Sadlej pVTZ basis set and the perturbing coordinate method with `pertnd=0.10`. See text for details.

`pertnd` **(Bohr)**	**MAD relative to Q-Chem** ($\frac{\text{Ang}^4}{\text{amu}}$)	**MAD relative to Gaussian** ($\frac{\text{Ang}^4}{\text{amu}}$)
	*Numerical polarizabilities – 6-31G**	
0.05	0.456	0.473
0.10	0.354	0.369
0.25	0.512	0.519
	*Analytic polarizabilities – 6-31G**	
0.05	1.975	1.980
0.10	1.969	1.974
0.25	10.161	10.171
	Numerical polarizabilities – Sadlej pVTZ	
0.05	0.577	0.873
0.10	0.482	0.749
0.25	0.623	0.859
	Analytic polarizabilities – Sadlej pVTZ	
0.05	1.220	1.678
0.10	1.186	1.649
0.25	9.811	10.190

Table 2.1: Relative errors (MAD) in the Raman activities calculated using the perturbing coordinate method relative to the Q-Chem values (middle column) and Gaussian values (third column). Results are shown for both the 6-31G* (top two sections) and Sadlej pVTZ (bottom two sections) basis sets. As described in the text, Method 2 corresponds to numerically-calculated polarizabilities and Method 3 corresponds to fully analytically-calculated polarizabilities. In each of the four sections the results are shown for various values of the coordinate perturbation, set by the `pertnd` keyword in Jaguar. The errors are calculated by averaging over all modes for each of the six molecules, and then averaging those values over all six molecules. The Raman activities are calculated for a single set of experimental conditions, which correspond for example to backscattering in which no polarization analyzer is placed in front of the detector; the common Raman intensity formula of eq. 1.33 is used for all calculations. For these calculations, activities for modes within $4\,\text{cm}^{-1}$ were added together.

we can already conclude from the results using the 6-31G* basis set and the perturbing coordinate method with numerical polarizabilities that the pseudospectral method certainly *can* provide Raman intensities that agree well with those calculated using the traditional spectral electronic structure method. From the results using the Sadlej pVTZ basis set and the perturbing coordinate method with numerical polarizabilities, we can only conclude that Jaguar's fully spectral method, PRISM, can be used to obtain accurate Raman intensities, though there is no reason to doubt that if pseudospectral grids were implemented for the Sadlej basis set, just-as-accurate Raman spectra could be obtained using the pseudospectral method along with the perturbing coordinate method and numerical polarizabilities. Further, since B3LYP/Sadlej has been shown to provide experimentally accurate Raman spectra, this methodology could be used to predict experimental Raman spectra.

Finally, we note that because we did not perform this testing before utilizing the perturbing coordinate method with *analytic* polarizabilities to produce the Raman spectra of Chapter 3, all spectra in this dissertation said to employ the perturbing coordinate method imply that analytic polarizabilities are used, i.e., Method 3. This was done because, usually, particularly in Jaguar, analytic polarizability calculations are faster than numerical polarizability calculations. However, as shown in Figs. 3.4, 3.5, and 3.7 to 3.9, we see that calculation of Raman intensities using Method 3 still generally produces spectra that agree well with experiment.

2.4 Perturbing field method for calculating the Raman tensor

The form of the Cartesian third derivative for Method 4 is

$$\frac{\partial^3 E}{\partial \mathbf{R} \partial \mathbf{F}_1 \partial \mathbf{F}_2} = \frac{\partial^2 \mathbf{g}}{\partial \mathbf{F}_1 \partial \mathbf{F}_2},$$

which we approximated numerically by applying the centered finite-difference scheme twice:

$$\begin{aligned}
\frac{\partial^2 g_i}{\partial F_j \partial F_k} &= \frac{\partial}{\partial F_j}\frac{\partial g_i}{\partial F_k} \\
&\approx \frac{\partial}{\partial F_j}\left(\frac{g_i\left(\left(\Delta F\right)_k\right) - g_i\left(-\left(\Delta F\right)_k\right)}{2\Delta F}\right) \\
&= \frac{1}{2\Delta F}\left(\frac{\partial g_i\left(\left(\Delta F\right)_k\right)}{\partial F_j} - \frac{\partial g_i\left(-\left(\Delta F\right)_k\right)}{\partial F_j}\right) \\
&\approx \frac{1}{2\Delta F}\left[\frac{g_i\left(\left(\Delta F\right)_j, \left(\Delta F\right)_k\right) - g_i\left(-\left(\Delta F\right)_j, \left(\Delta F\right)_k\right)}{2\Delta F}\right. \\
&\quad \left. -\frac{g_i\left(\left(\Delta F\right)_j, -\left(\Delta F\right)_k\right) - g_i\left(-\left(\Delta F\right)_j, -\left(\Delta F\right)_k\right)}{2\Delta F}\right] \\
&= \frac{1}{4\left(\Delta F\right)^2}\left(g_i\left(\left(\Delta F\right)_j, \left(\Delta F\right)_k\right) - g_i\left(-\left(\Delta F\right)_j, \left(\Delta F\right)_k\right)\right. \\
&\quad \left. -g_i\left(\left(\Delta F\right)_j, -\left(\Delta F\right)_k\right) + g_i\left(-\left(\Delta F\right)_j, -\left(\Delta F\right)_k\right)\right), \qquad (2.2)
\end{aligned}$$

where ΔF corresponds to *half* the value of the `refield` keyword and the notation $g_i\left(\left(\Delta F\right)_j, -\left(\Delta F\right)_k\right)$ means that the gradient is to be calculated in the presence of an electric field of magnitude ΔF in the j direction plus an electric field of magnitude ΔF in the $-k$ direction ($j, k \in \{x, y, z\}$). While it would appear that for all nine pairs of jk four gradient calculations are required, resulting in the need for a total of 36 gradient calculations, in reality only 19 of these gradient calculations are unique. We see this by first writing out the third derivative for the diagonal field components ($j = k$),

$$\begin{aligned}
\frac{\partial^2 g_i}{\partial F_j \partial F_j} &= \frac{1}{4\left(\Delta F\right)^2}\left[g_i\left(\left(\Delta F\right)_j, \left(\Delta F\right)_j\right) - g_i\left(-\left(\Delta F\right)_j, \left(\Delta F\right)_j\right)\right. \\
&\quad \left. -g_i\left(\left(\Delta F\right)_j, -\left(\Delta F\right)_j\right) + g_i\left(-\left(\Delta F\right)_j, -\left(\Delta F\right)_j\right)\right] \\
&= \frac{1}{4\left(\Delta F\right)^2}\left(g_i\left(2\left(\Delta F\right)_j\right) - 2g_i\left(0\right) + g_i\left(-2\left(\Delta F\right)_j\right)\right), \qquad (2.3)
\end{aligned}$$

showing that rather than $3 \times 4 = 12$ gradient calculations for the diagonal components, only 7 are actually needed, and then by recognizing the equality of the mixed derivatives of the off-diagonal components,

$$\frac{\partial^2 g_i}{\partial F_k \partial F_j} = \frac{\partial^2 g_i}{\partial F_j \partial F_k},$$

showing that rather than $6 \times 4 = 24$ gradient calculations for the off-diagonal components, only 12 are actually needed. Thus, a total of 19 gradient calculations are needed: 18 in the presence of various combinations of applied electric fields, and one in the absence of any fields. Once these 19 gradient jobs have completed, eqs. 2.2

and 2.3 can be used to calculate the Cartesian third derivative for the off-diagonal and diagonal components, respectively. It should be noted that based on the Q-Chem output files, this is the same way Q-Chem calculates its third derivative.

Unlike for the perturbing coordinate method in which we were provided with Raman activities for Gaussian rather than the third derivative itself, as long as we focus on comparing our results to only Q-Chem's results (which is somewhat validated from Section 2.3, which showed that the Gaussian and Q-Chem results generally agree), we will now more rigorously test the ability of the pseudospectral method to calculate Raman spectra by comparing the third derivative itself, as explained in Section 2.1. We note that due to a bug in Q-Chem's output in which the xx component of the third derivative is a duplicate of the zz component, we re-calculated Q-Chem's third derivative using the 19-field finite difference method above, checking that we did so correctly by comparing the result to the eight unique components that Q-Chem outputs.

In order to test the pseudospectral method's ability to produce accurate Raman spectra, the basis set used for the jobs must have pseudospectral grids implemented in Jaguar, eliminating the well-tested Sadlej pVTZ basis set as an option. However, recently pseudospectral grids for the Rappoport SVPD basis set (also called def2-SVPD) have been implemented in Jaguar precisely because this basis set leads to results for electric response properties similar to the Sadlej basis set, yet the Rappoport basis set is significantly smaller and therefore results in faster calculations[21]. Thus, unless otherwise specified, for our tests below we use the Rappoport basis set. In addition, since we are no longer limiting ourselves to comparisons to the Gaussian results, we choose a slightly larger ($N = 15$) test set of molecules to study: C_2H_2, C_2H_4, C_2H_6, C_7H_8, CCl_4, CH_2Cl_2, CH_2F_2, CH_2O, CH_4, $CHCl_3$, H_2O, H_2S, H_3N, N_2, and SiO_2.

Preliminary tests showed that calculation of Raman intensities in Jaguar using the finite difference method above worked to various degrees based upon the value of the `refield` keyword; all that needed to be done was to determine the value of `refield` that resulted in a Cartesian third derivative that compared best to that calculated by Q-Chem. In such preliminary tests we found that the ideal `refield` values for optimization of the off-diagonal components of the third derivative were generally different from those that were found to optimize the diagonal components. Thus, we performed separate optimizations of `refield` for the diagonal and off-diagonal components. Lastly, since occasionally the pseudospectral method simply does not work for certain systems and PRISM (the traditional, fully spectral electronic structure method as implemented in Jaguar) needs to be employed by Jaguar in order to make the calculation proceed successfully, we also determined optimum `refield` keywords

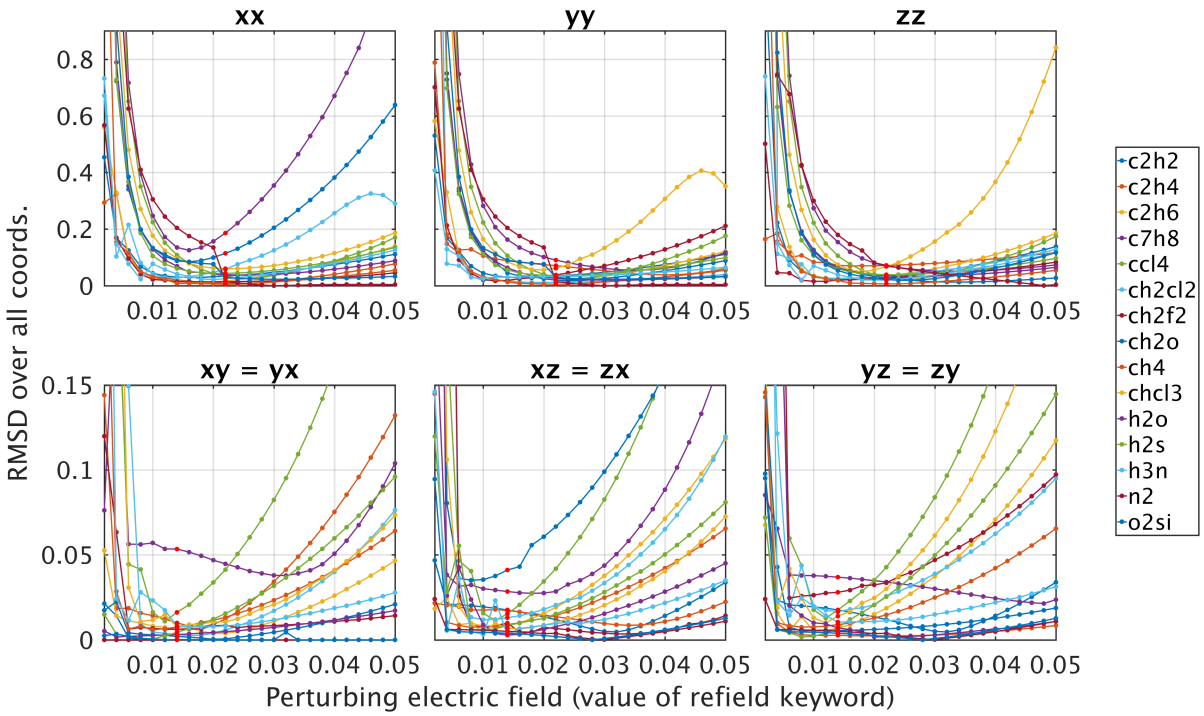

Figure 2.3: RMSD over the $3N_{\text{nuc}}$ Cartesian coordinates of the Raman tensor calculated using the pseudospectral method. Results are shown for all 15 molecules of our test set, all tested values of `refield` (x-axis), and both component types of the Raman tensor: the diagonal components (upper row) and the off-diagonal components (lower row). Datapoints in red denote the optimum (determined later) values of `refield` for both component types. Atomic units are used on both axes.

for the PRISM method in Jaguar.

The results for the pseudospectral method are shown in Fig. 2.3, which shows the RMSD (in a.u.; over all $3N_{\text{nuc}}$ nuclear coordinates) in the third derivative of the Jaguar results relative to Q-Chem for all 25 values of `refield` tested (0.002 a.u. to 0.050 a.u. in steps of 0.002 a.u.) for all 15 molecules and for the diagonal components of the derivative (upper row) and the off-diagonal components of the derivative (lower row). We note that the average ideal `refield` value is higher for the diagonal components than for the off-diagonal components. The ultimately chosen (explained below) ideal `refield` values are shown in red. The corresponding plot for the PRISM method in Jaguar is shown in Fig. 2.4. We note that the off-diagonal agreement at small values of `refield` is generally much better than in the pseudospectral case.

Next we averaged the RMSDs over all 15 molecules, plotting the results for the pseudospectral (blue) and PRISM (red) methods in Fig. 2.5. Now we clearly see that indeed there are optimum values of `refield` in all cases; each of the 12 curves has a clear minimum error value. It is interesting to note that while the pseudospectral

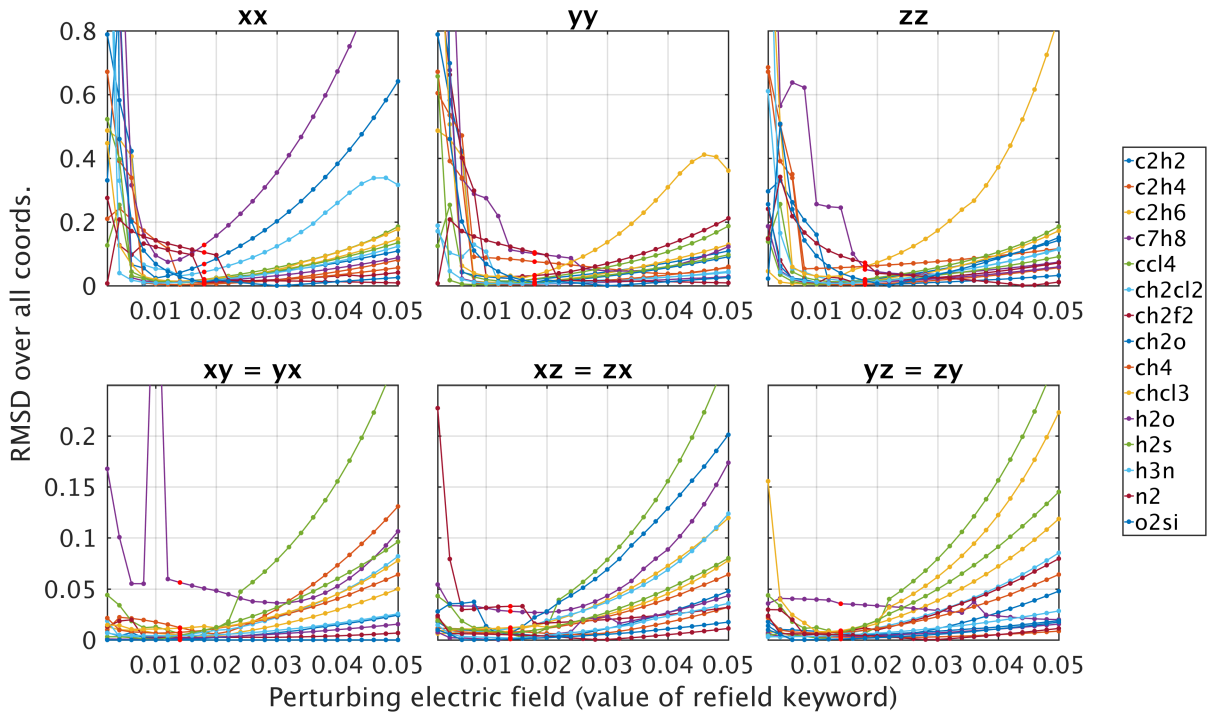

Figure 2.4: RMSD over the $3N_{\text{nuc}}$ Cartesian coordinates of the Raman tensor calculated using the PRISM method. Results are shown for all 15 molecules of our test set, all tested values of `refield` (x-axis), and both component types of the Raman tensor: the diagonal components (upper row) and the off-diagonal components (lower row). Datapoints in red denote the optimum (determined later) values of `refield` for both component types. Atomic units are used on both axes. The aberration of the data for C_7H_8 at 0.010 a.u. for the $xy = yx$ tensor component is the source of the spike in the following figures, but it is ultimately inconsequential.

and PRISM methods differ somewhat significantly for weak perturbing fields, they lead to similar errors in the Raman tensor components for larger perturbing fields. For each individual curve in Fig. 2.5 we also plotted with large circular markers the datapoints whose error is less than three times the minimum error. Only the low-`refield` side of the minima are included in these datapoints because generally the SCFs for both methods have difficulty converging in the presence of larger applied fields. Thus, the larger markers represent the generally *viable* strengths of the perturbing field that yield acceptable errors.

Since the Raman tensor, being a tensor, is very sensitive to the molecule orientation, we finally averaged together the errors for the three diagonal components and for the three unique off-diagonal components, plotting the results in Fig. 2.6 for the diagonal components (blue) and the off-diagonal components (red) for both the pseudospectral and PRISM methods. We again calculate and plot the (low-`refield`) datapoints whose errors are within three times the minimum error for each curve. For the pseudospectral method, these viable perturbing field

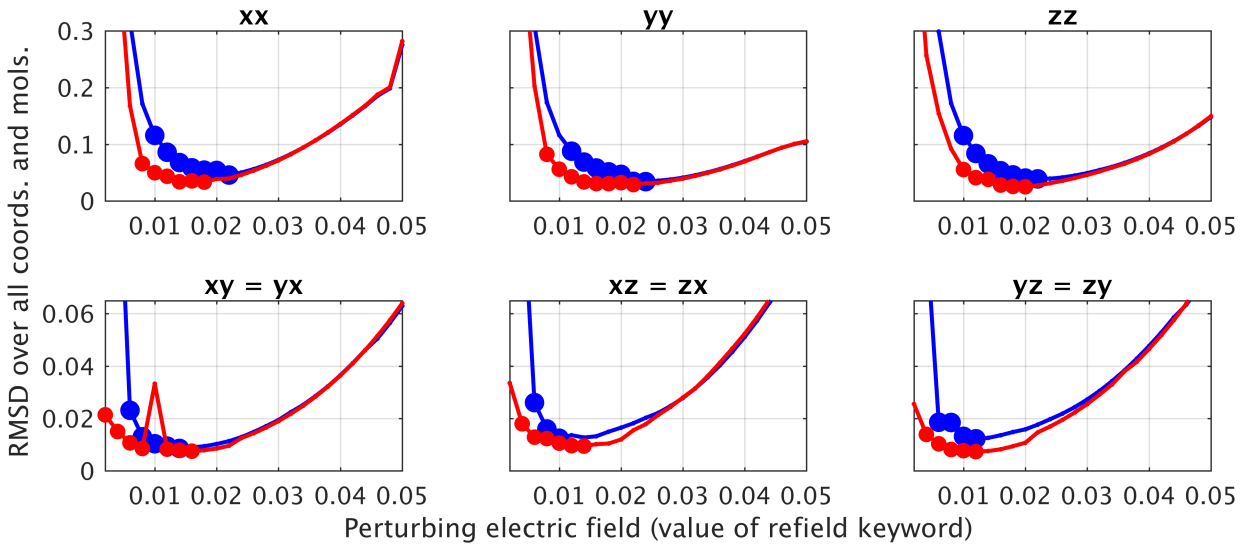

Figure 2.5: RMSD over the $3N_{\text{nuc}}$ Cartesian coordinates of the Raman tensor averaged over all 15 molecules of our test set. The data for the pseudospectral method are shown in blue and those for the PRISM method are shown in red. Results are shown for all tested values of `refield` (x-axis) and both component types of the Raman tensor: the diagonal components (upper row) and the off-diagonal components (lower row). For each curve datapoints corresponding to errors within three times the minimum error are indicated by large circular markers; only the low-`refield` such datapoints are shown because they represent the smaller perturbing field strengths that generally result in SCFs that converge more easily. Atomic units are used on both axes.

values range from 0.010 a.u. to 0.022 a.u. for the diagonal tensor components and from 0.006 a.u. to 0.014 a.u. for the off-diagonal components. For the PRISM method, these viable values of `refield` range from 0.008 a.u. to 0.018 a.u. for the diagonal components and 0.004 a.u. to 0.014 a.u. for the off-diagonal components. In practice, we should start with the upper limits of these ranges (which correspond to minimum error in the tensor components) and, if the jobs have trouble converging, try smaller `refield` values (or PRISM if the pseudospectral method was just tried), stopping at the lower limits of the ranges, below which the error is significantly worse. From Fig. 2.6 we note that PRISM always leads to better Raman tensors than the pseudospectral method, though the disadvantage of course is that the corresponding jobs will be much slower. However, if Raman spectra are obtained that seem to be significantly different than expected, PRISM should be used to check that the correct spectra are being obtained by Jaguar. Finally, we also note that based on the viable `refield` values, usage of different values for the diagonal and off-diagonal tensor components is justified.

In order to get a feel for the Raman activities to which the ideal `refield` values lead, below we list the raw Raman activities calculated using all methods studied so far (using the minimum-error values of `refield` for the perturbing field methods), for the common intensity formula of eq. 1.33 for the same six molecules for

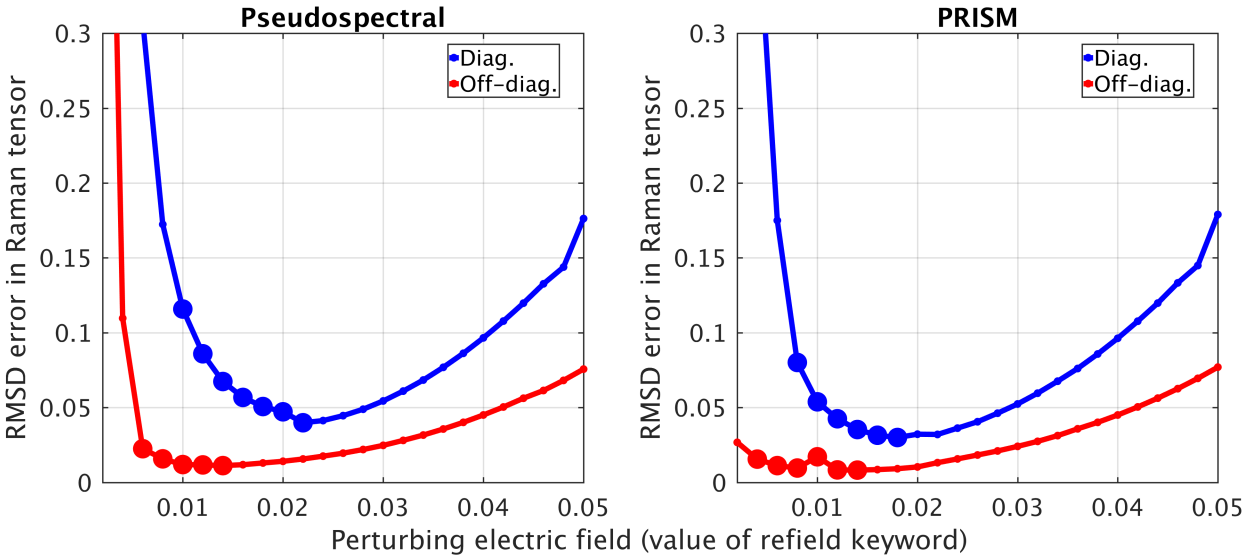

Figure 2.6: RMSD over the $3N_{\text{nuc}}$ Cartesian coordinates of the Raman tensor averaged over all 15 molecules of our test set and over each component type. Results are shown for all tested values of `refield` (x-axis). For each curve datapoints corresponding to errors within three times the minimum error are indicated by large circular markers; only the low-`refield` such datapoints are shown because they represent the smaller perturbing field strengths that generally result in SCFs that converge more easily. Atomic units are used on both axes.

which we have Gaussian data available. Since the Gaussian data do not include results for the Rappoport basis set, we repeated the entire study of determining optimum `refield` values for the 6-31G* basis set, finding error-minimizing `refield` values for the pseudospectral method of 0.036 a.u. for the diagonal components and 0.016 a.u. for the off-diagonal components and for the PRISM method of 0.014 a.u. for the diagonal components and 0.008 a.u. for the off-diagonal components. We therefore list the Raman activities for the perturbing field method using the pseudospectral method (called "`refield, ps`" below) and for the perturbing field method using the PRISM method (called "`refield, prism`" below). For comparison we also list the Raman activities calculated by the perturbing coordinate method using both numerical polarizabilities and analytic polarizabilities, both of which use the "best" value for `pertnd` of 0.10 Bohr. It should be noted that even though it appears that the perturbing field method yields better Raman activities than the perturbing coordinate method, we note that we did not optimize the value of `pertnd` as we did for `refield` since we chose to focus on the more-efficient perturbing field method. It is interesting to note that, at least for the six molecules and single intensity formula studied here, for the perturbing field method, the pseudospectral method appears to provide better Raman activities than the PRISM method; this is confirmed by calculation of the mean absolute deviations presented in

Table 2.2. However, the conclusion is reversed if the data for N_2 are excluded, and based on this limited number of molecules and intensity formula, it is probably unrealistic to declare a superior method. It should be noted that for the full set of 15 molecules for the 6-31G* basis set, whereas for the diagonal component PRISM leads to smaller error (just as seen for the Rappoport data in Fig. 2.6), for the off-diagonal component the pseudospectral method leads to smaller error (unlike for the Rappoport basis set), though the difference is probably insignificant: using PRISM the tensor error is 0.015 a.u. whereas using the pseudospectral method the error is 0.006 a.u; both errors are very small.

```
6-31G*:
  H2O:
    Gaussian:                    7.96   78.84   39.07
    Q-Chem:                      7.97   78.89   39.12
    Jaguar (refield, ps):        7.94   79.16   39.08
    Jaguar (refield, prism):     7.97   79.68   39.11
    Jaguar (numerical, 0.10):    7.92   78.37   38.62
    Jaguar (analytic,  0.10):    8.09   78.87   39.62
  H2S:
    Gaussian:                   46.00  156.18  119.22
    Q-Chem:                     46.01  155.95  119.15
    Jaguar (refield, ps):       45.63  155.40  118.95
    Jaguar (refield, prism):    45.80  156.28  118.93
    Jaguar (numerical, 0.10):   45.77  155.70  117.99
    Jaguar (analytic,  0.10):   41.10  148.37  119.45
  CH2O:
    Gaussian:                    2.12    7.48   14.97    3.01  152.02   75.12
    Q-Chem:                      2.11    7.47   14.96    3.00  151.97   75.05
    Jaguar (refield, ps):        2.11    7.49   14.97    2.99  152.40   75.10
    Jaguar (refield, prism):     2.08    7.48   15.14    3.06  151.24   75.04
    Jaguar (numerical, 0.10):    2.00    7.75   15.60    2.91  152.73   76.63
    Jaguar (analytic,  0.10):    3.75    7.42    7.49   14.76  146.65   71.94
  CH2F2:
    Gaussian:                    1.78    7.72    2.63   14.42    5.83    9.10
100.45   48.58
    Q-Chem:                      1.78    7.71    2.64   14.46    5.86    9.12
100.71   48.66
    Jaguar (refield, ps):        1.78    7.75    2.62   14.49    5.79    9.16
101.02   48.71
    Jaguar (refield, prism):     1.78    7.77    2.64   14.44    5.89    9.14
100.13   48.76
    Jaguar (numerical, 0.10):    1.79    7.84    2.44   13.25    5.87    9.16
100.69   49.05
    Jaguar (analytic,  0.10):    3.20    7.37    2.33   14.25    6.06    7.18
99.80   47.93
  CH2Cl2:
    Gaussian:                    7.53   13.71    5.01    3.84   13.40    5.58
```

Method	MAD relative to Q-Chem ($\frac{Ang^4}{amu}$)	MAD relative to Gaussian ($\frac{Ang^4}{amu}$)
Perturbing field / PS	0.150	0.172
Perturbing field / PRISM	0.186	0.184
Perturbing coord. (numerical, 0.10 Bohr) / PS	0.354	0.369
Perturbing coord. (analytic, 0.10 Bohr) / PS	1.969	1.974

Table 2.2: Average MAD of the Raman activities over all six molecules having Gaussian Raman data, relative to both Q-Chem and Gaussian, using the 6-31G* basis set. Note that the perturbing coordinate methods were not optimized as rigorously as the perturbing field methods; their errors can likely be reduced by optimizing `pertnd`. While it appears for this limited set of molecules and single intensity formula that the pseudospectral (PS) perturbing field method provides superior Raman activities to the PRISM perturbing field method, the data are limited and conclusions cannot be made; for example, if the data for N_2 are excluded, PRISM leads to a smaller error than the pseudospectral method for the perturbing field method. The raw data to which this table pertains are located in the text.

```
13.39   93.14   61.33
    Q-Chem:                     7.52   13.77    5.04    3.84   13.40    5.57
13.41   93.25   61.45
    Jaguar (refield, ps):       7.79   13.66    4.96    3.84   13.43    5.66
13.45   94.64   61.58
    Jaguar (refield, prism):    7.50   13.69    4.99    3.85   13.40    5.62
13.40   93.56   61.47
    Jaguar (numerical, 0.10):   7.46   13.67    4.97    3.96   13.08    5.58
13.47   93.91   62.23
    Jaguar (analytic,  0.10):  10.53   16.10    3.08    3.47   13.37    6.18
10.37   90.33   61.19
  N2:
    Gaussian:                  14.24
    Q-Chem:                    14.24
    Jaguar (refield, ps):      14.23
    Jaguar (refield, prism):   13.97
    Jaguar (numerical, 0.10):  14.09
    Jaguar (analytic,  0.10):  14.32
```

Finally, we briefly study the effect of using the largest-tensor-error `refield` values identified by the left-most large circular markers in Fig. 2.6 instead of using the optimal `refield` values identified by the right-most large circular markers. We do this non-comprehensively by using eq. 1.33 to calculate the Raman activity using the Rappoport basis set for the entire 15-molecule test set by using the two "extreme" sets of `refield` values, for both the pseudospectral and PRISM methods. We find that the average (over all 15 molecules) mean absolute

deviation for the pseudospectral method is 1.12 $\frac{\text{Ang}^4}{\text{amu}}$ (with a maximum deviation of 26.22 $\frac{\text{Ang}^4}{\text{amu}}$) whereas the average MAD for the PRISM method is 0.73 $\frac{\text{Ang}^4}{\text{amu}}$ (with a maximum deviation however of 41.01 $\frac{\text{Ang}^4}{\text{amu}}$).[3] These are both relatively small differences (again, common activities range from 0 to 200 $\frac{\text{Ang}^4}{\text{amu}}$) and it is therefore generally acceptable to use smaller `refield` values if the SCFs have trouble converging. However, due to the occasional large deviations, this should be done with the realization that a couple of the activities may be significantly wrong. In order to obtain a feel for the differences, the raw activities for each set of `refield` values are listed in Appendix D.

2.5 Discussion

The analytic method for electronic structure calculations has long been shown to produce Raman intensities that compare well with experiment, particularly when using the B3LYP density functional and the Sadlej basis set[7, 8]. More recently, the Rappoport basis set has been shown using the analytic method and hybrid density functional theory to produce energy third derivatives that compare well with those calculated using the Sadlej basis set[21]. What we have done in this chapter is show that we obtain Raman intensities calculated by the pseudospectral method that compare well with those calculated by the analytic method, using the B3LYP/Rappoport methodology. Thus, we have indirectly shown that the pseudospectral method, using B3LYP/Rappoport, produces Raman spectra that should compare reasonably well with experiment.

We have not explicitly benchmarked our Raman intensity calculations to show that we obtain Raman spectra using the pseudospectral method more quickly than using the analytic method. However, our methodology for producing Raman spectra, which is the same as that used in Q-Chem, relies on 19 gradient calculations, and it has already been shown in literature[18] and even in this dissertation (Chapter 7) that the pseudospectral method generally calculates gradients faster than the analytic method. Thus, as long as it is assumed the speed advantage of gradient calculations holds under perturbing electric fields, we have indirectly shown that the pseudospectral method calculates Raman spectra faster than the analytic method.

Thus, while a more detailed study of the experimental accuracy and the superiority in calculation speed would confirm our conclusions here, in the meantime we have largely shown that the pseudospectral method calculates Raman spectra that compare well with experiment more quickly than the analytic method. While our test set consisted of 15 small molecules and is enough to convince us of the utility of the pseudospectral method for

[3] The minimum, MAD, and maximum values for each of the 15 molecules are listed in Appendix D.

simulating Raman spectra, this is a subjective conclusion given the small size of the test set used. What is not subjective, however, is the utility of Raman simulations for studying the structure and function of molecules. Thus, in our opinion, we have so far demonstrated an important way in which the pseudospectral method can be used for elucidating molecular structure and function.

Chapter 3

Using polarized Raman spectra to determine unit cell orientation: application to oligothiophene crystals

3.1 Introduction

It is well-known that complex and useful properties of materials such as their conductivity and luminescence depend on their morphologies and in particular on their crystallinity[22, 23] and, within crystalline regions, on their molecular orientation[24, 25]. The *de facto* method for measuring such morphology in a material is X-ray diffraction (XRD), but this has significant drawbacks, such as the need to place the sample in a diffractometer, the possibility of sample degradation due to the high energies required, the time and monetary resources required, and the poor spatial resolution obtained. In contrast, in general Raman spectroscopy can be performed *in situ*, requires only visible-wavelength radiation, is fast and cheap, and can obtain excellent spatial resolution. Thus, it would be beneficial to have Raman spectroscopic methods available to perform the same structure-resolving functionality for crystals and thin films as XRD.

While numerous examples exist of the utilization of Raman spectroscopy to elucidate sample morphology[26–37], none have yet to demonstrate the ability of polarized Raman spectroscopy to determine the unit cell orientations of organic single crystals. Below we do just this for single crystals of oligothiophenes (Fig. 3.1), shorter versions of the class of π-conjugated polymers called polythiophenes, which have been shown to have important

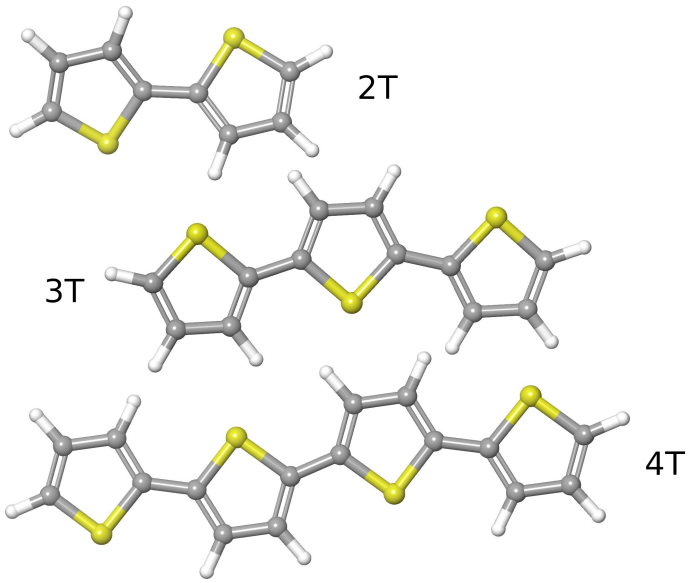

Figure 3.1: Bithiophene (2T), terthiophene (3T), and quaterthiophene (4T) molecules. Longer chains are polythiophenes.

utility in organic electronics[38–42] due fundamentally to their high conductivity when doped. In particular, below we apply the general Raman intensity formulas derived in Chapter 1 to back out the orientation of the unit cell of molecular single crystals by studying the Raman intensities at various polarizations of the light incident on the crystals. At the time of publication[43] this was the most explicit demonstration of such an application of polarized Raman spectroscopy and was certainly the first to demonstrate such an effect on organic single crystals. As an example, we focus on oligothiophene molecular crystals, whose polythiophene analogues find important application in, for example, flexible, silicon-free electronics. We propose that the particular methodologies and results for oligothiophenes can be extrapolated to their more useful form as long polythiophene chains, which are more difficult to control experimentally and to simulate theoretically. Further, while we focus on orientation determination of unit cells of organic molecules, we propose these methods as generally applicable to any crystalline material.

3.2 Motivating polarized Raman spectra as a measure of the physical structure of oligothiophene molecules

We first examine the plausibility of Raman spectroscopy as a tool for determining the physical structure of oligothiophenes by investigating how polarized Raman spectra change as the torsion angle between the two monomers of a bithiophene (2T) molecule changes. The simulated Raman spectra are shown in Fig. 3.2. Panel A shows isotropic spectra whereas Panels B and C show crystalline spectra of two different orientations of the molecule relative to the optical axis (as usual, this is the vertical, z-axis shown in the insets). All calculations in this chapter assume the experimental setup used by the Willets Group (Appendix F): linearly polarized incident light at 532 nm, backscattering (a Raman microscope is used), no polarization analyzer in front of the detector, and room temperature conditions. At the right of each panel we plot the relative energy of each conformation of the molecule. We see that while the isotropic spectra in Panel A resemble the crystalline spectra for the orientation specified in Panel B, they differ from the crystalline spectra for the orientation specified in Panel C, in which the long axis of the molecule is parallel to the optical axis. It should be noted that for all three sets of spectra, as the internal dihedral is rotated, many vibrational modes shift or split by as much as $\sim 81\,\mathrm{cm}^{-1}$ (which corresponds to the splitting of what we call MM1 below). From this example we conclude that the Raman spectra are very sensitive to the molecular conformation and we therefore find it very plausible that the dependence of the spectra on the molecular orientation would be just as sensitive, allowing the spectra to be useful indicators of the orientations of the molecules.

3.3 Ensuring reasonable agreement between experiment and theory

Next it was important that we make sure we could obtain reasonable agreement between the experimental Raman spectra and the simulated spectra. For this purpose we chose to start with liquid ("isotropic") 2T, whose spectra were shown in Fig. 3.2a for a range of dihedral angles. The simulated Raman spectrum at the minimum-energy (for the B3LYP/Sadlej pVTZ level of theory) dihedral of 158° is shown in Fig. 3.3 in green; this is the minimum-energy cross section of Fig. 3.2a. Plotted in blue is the Willets Group's experimental Raman spectrum for liquid 2T, and we note that the simulated intensities do not agree so well. This is because, while the simulated spectrum corresponds to the minimum-energy conformation of the 2T molecule, in reality multiple conformations are accessible at room temperature. In other words, the "components" at particular dihedral angles need to be

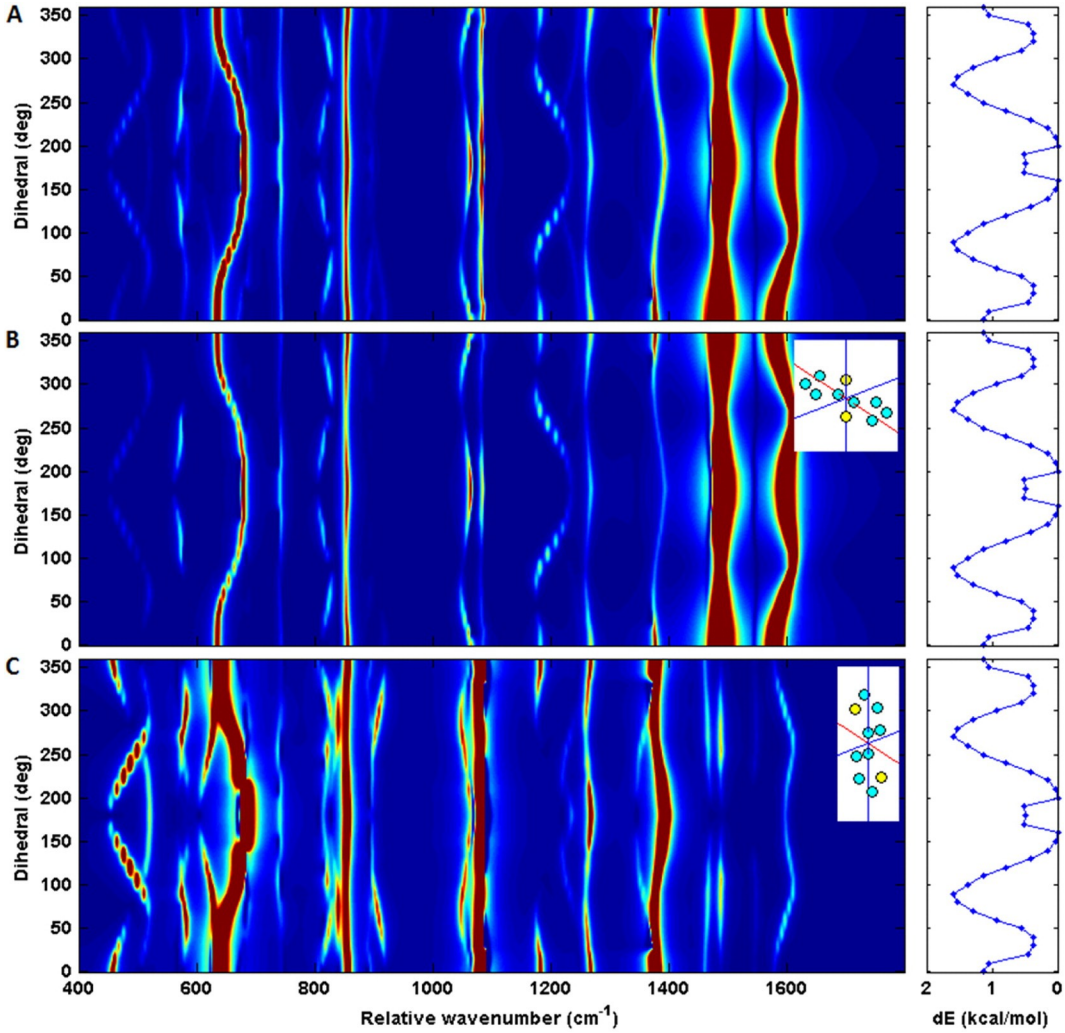

Figure 3.2: Simulated vibrational Raman spectra of a 2T molecule with respect to the torsion angle between the thiophene monomers. (a) Isotropic spectra. Blue indicates a low intensity and dark red indicates a high intensity. To the right is the calculated relative energy of each dihedral conformation, linearly interpolated between data points. These curves denote the energy of the molecule in the gas phase and are the same for all three panels. 0° refers to both monomers in the same plane, with the sulfur atoms on the same side; 180° refers to the planar conformation in which the sulfur atoms alternate sides. (b) Crystalline spectra for the molecule orientation shown in the inset, in which the plane of the molecule is orthogonal to the optical Raman microscope axis, shown as the vertical blue line. In the insets, the hydrogen atoms are not shown, though calculations are made with these atoms present. (c) Crystalline spectra for the molecule oriented with its long axis parallel to the optical axis. Note that "cigar"- or "bubble"-like shapes in the images are artificial features of the interpolation between calculations for different dihedrals. The Raman calculations were done using B3LYP[44]/Sadlej using Jaguar[9] with the perturbing coordinate implementation of Raman spectra with a 0.05-Bohr step size using analytic calculations for the polarizabilities. As for all simulated spectra in this chapter, each mode is fit with an 8-cm^{-1} full-width half-maximum Lorentzian curve. Note that within each panel the intensity scaling is the same, but between panels the intensity scaling is different in order to demonstrate the rich spectral dependence on the state of aggregation (crystalline vs. isotropic) and the molecule dihedral. Graphic taken from Fig. 1 of [43].

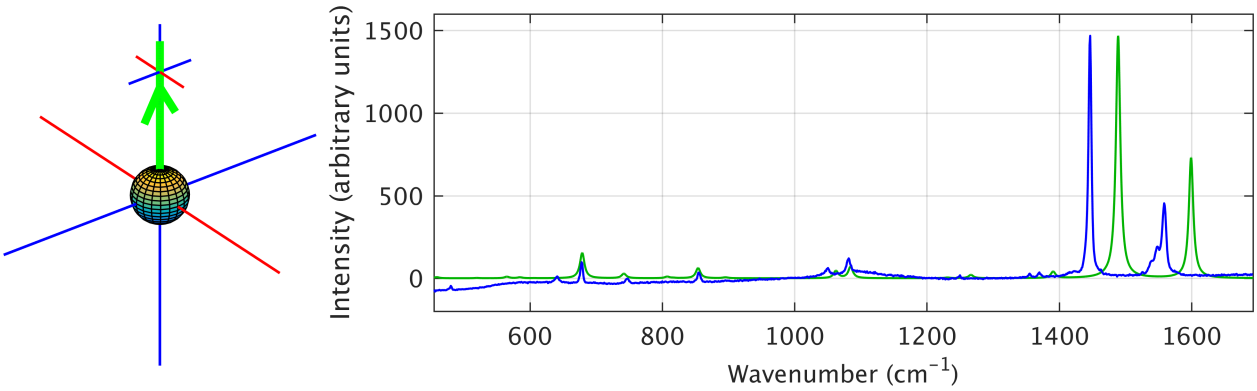

Figure 3.3: Calculated (green) Raman spectrum of liquid 2T for the single minimum-energy dihedral conformation of $158°$ (for the level of theory used, below). This spectrum is *not* expected to agree with the experimental result (blue) because multiple molecule conformations are accessible — not just the $158°$ conformation — and so Boltzmann averaging needs to be applied. The experimental curve is shifted so that the average intensity in the region between 1265 cm^{-1} and 1340 cm^{-1} equals zero, and its maximum is normalized to the maximum of the calculated result. The Raman calculations employed the perturbing field method in Jaguar with a step size of 0.018 a.u. (`refield=0.018`) for the diagonal components and 0.014 a.u. for the off-diagonal components. The B3LYP density functional and the Sadlej pVTZ basis set were used, and the standard analytic calculation of the two-electron integrals was used (as opposed to the pseudospectral method).

averaged together somehow. The correct way to do this is by averaging each "component" spectrum using a weighting equal to the Boltzmann factor e^{-E_i/k_BT}, where E_i is the energy of the conformation corresponding to dihedral i. The level of theory used was B3LYP/6-31+G** for the geometry optimizations and LMP2/cc-pVTZ(-f) for the energy calculations for each fixed dihedral; this level of theory was also used for the energy plots on the right-hand sides of Fig. 3.2. When this averaging is done, the agreement between experiment and theory is excellent; see Fig. 3.4a, which plots the Boltzmann-averaged simulated spectrum in red, the experimental result in blue, and the components (with different intensity scalings) in the background in green for reference. It should be noted that even though the final peak wavenumbers generally disagree, this is to be expected because without applying some sort of frequency-scaling method such as Pulay's SQM method[45–47], DFT-calculated frequencies generally are inaccurate.[1] Conversely, since we are using the reliable[7, 8] Sadlej pVTZ basis set[48–51] for the Raman intensities, the intensities themselves are expected to be accurate, as seen from the excellent agreement in the intensities between theory and experiment in Fig. 3.4a.

[1]It is important to note, however, that the errors in wavenumber do not imply that the peak heights cannot be trusted. The frequencies are the eigenvalues of the Hessian matrix and the intensities are based on how the polarizability changes as the molecule vibrates, which is described by the eigenvectors of the Hessian. Because there are an infinite number of matrices with the same eigenvectors but different eigenvalues, there is not a one-to-one correspondence between the two. It is thus possible to have the correct mode intensities, but incorrect mode wavenumbers.

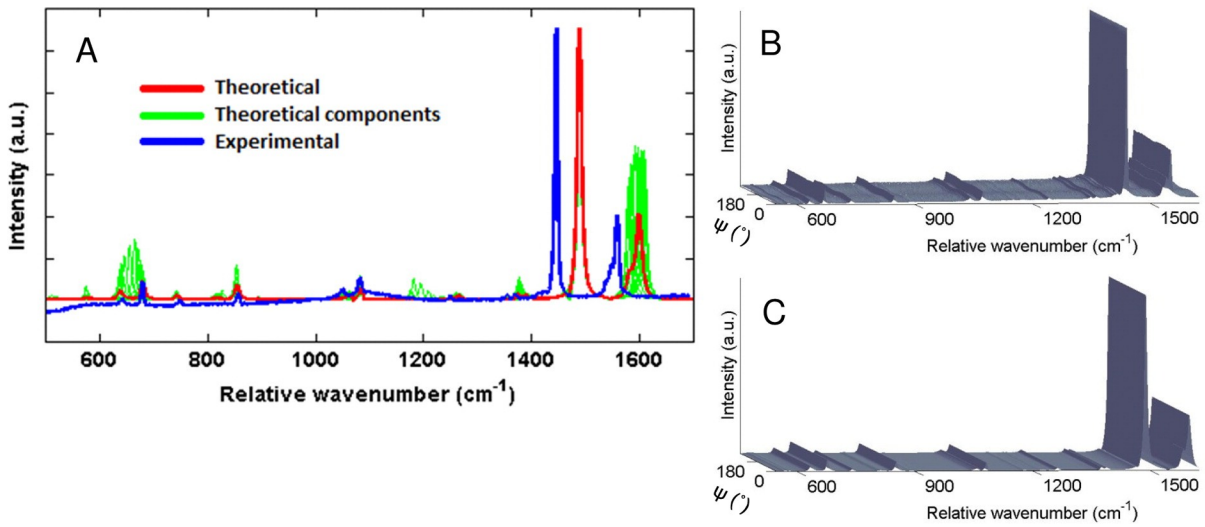

Figure 3.4: Boltzmann-averaged Raman spectra of liquid 2T. The averaging was done over the molecule torsion angle, as explained in the text. (a) Boltzmann-averaged simulated spectrum (red), experimental spectrum (blue), and simulated components (with different intensity scalings; maximum component and experimental peaks are always scaled to the maximum Boltzmann-averaged peak; green). A limited wavenumber range is included because this is the range that was provided by the Willets Group; the lower limit is the lowest wavenumber expected to be accurately measured experimentally, and beyond the upper limit lies the silent region, where no vibrational modes are present. (b) Experimental spectra taken for various values of the orientation angle ψ of the linearly polarized incident light, in which ψ is varied through $360°$. (c) Corresponding Boltzmann-averaged simulated spectra for ψ varied through $360°$. The Raman calculations employed the perturbing coordinate method with a step size of 0.05 Bohr (`pertnd=0.05`). Graphic taken from Fig. 2 of [43].

Finally, in preparation for using what we will shortly define as the "modulation depth" of a set of spectra, the Willets Group observed and the Friesner Group simulated what the Boltzmann-averaged liquid 2T spectra would look like as the orientation angle ψ of the incident linearly polarized light was rotated through $360°$, the results of which are shown in Fig. 3.4b (experiment) and Fig. 3.4c (theory). We see that these spectra are constant as ψ is varied, as expected for isotropic backscattering with no analyzer in front of the detector, as explained in Section 1.9, further bolstering the reliability of the Raman intensity formulas derived in Chapter 1.

While the Boltzmann averaging required for isotropic spectra in which multiple molecule conformations are accessible will not be needed in subsequent sections because there we focus exclusively on crystalline spectra, in which only a single conformation (the $180°$ planar conformation) of the molecules is accessible, the results in the present section have nevertheless demonstrated excellent agreement between experiment and theory on every front considered, so that we can now proceed with further comparisons with confidence.

3.4 Choosing useful polarized Raman signals

Next we identified two relatively intense, vibrational "marker modes" that seemed to be strongly dependent on the orientation of both experimental and simulated oligothiophene crystals. The first marker mode (MM1) was chosen as the one around 670–687 cm^{-1}, corresponding to the in-phase C–S stretch[52], and the second marker mode (MM2) was chosen as that around 1437–1454 cm^{-1}, corresponding to the in-phase C=C stretch[52]. Both these modes can be readily identified in Figs. 3.2 to 3.4.

For the experimental parameters of our experiments of this chapter, unlike for isotropic samples, crystalline samples are sensitive to the orientation angle ψ of the linearly polarized incident light. In general, as can be inferred from the Raman intensity formulas of Chapter 1 (see for example eq. 1.34), the intensity I_k of vibrational mode k varies with ψ as

$$I_k(\psi) = a_k + b_k \cos 2(\psi - \phi_k). \tag{3.1}$$

The intensity $I_k(\psi)$ is always positive, and the parameter a_k describes the offset of the oscillation of I_k with ψ. b_k is the amplitude of this oscillation and ϕ_k refers to the value of ψ at which the maximum intensity $I_k(\psi)$ occurs. The factor of 2 in the cosine argument shows that the period of oscillation is always π. We then replace the parameters a_k and b_k first by defining the modulation depth M_k of mode k as

$$M_k \equiv \frac{2b_k}{a_k + b_k}. \tag{3.2}$$

This quantity describes the relative depth of oscillation of the intensity of mode k as ψ is varied. M_k varies from 0, which indicates that the intensity I_k does not vary at all with ψ, to 1, which indicates that the amplitude of oscillation b_k is equal to the offset a_k itself, i.e., maximal modulation of I_k with ψ. Further, we define

$$A \equiv \frac{\max_\psi [I_2(\psi)]}{\max_\psi [I_1(\psi)]}$$

as the ratio of the largest values of the intensity as ψ is varied for each marker mode. Then, we use as our signals of crystal orientation the five parameters M_1, M_2, ϕ_1, ϕ_2, and A.

Further detailed experimental measurements[43] on precipitated and crystalline 2T showed large variation in the five parameters over three 2T samples, indicating the promise of our five polarized Raman parameters as indicators of unit cell orientation.

3.5 Correlating unit cell orientation with the five marker mode parameters $\{M_1, M_2, \phi_1, \phi_2, A\}$

The Willets Group found that stable 2T single crystals were somewhat difficult to synthesize at room temperature but that stable 3T and 4T single crystals were more stable. Thus, 3T and 4T became the crystals of choice for comparison with simulated data in order to correlate the marker mode parameters with the orientations of the unit cells. The structures of the experimental samples were characterized by XRD and the Willets Group was painstakingly able to report back precise Euler angles describing the orientations of the molecules in the unit cells to within 5°. All crystalline calculations assumed the well-known planar conformations of the molecules in which the sulfur atoms on neighboring monomers alternated sides[52–57] (the 180° conformation), as shown in Fig. 3.1.

Terthiophene

Terthiophene contains eight molecules in a unit cell, four of which have unique orientations, so 12 Euler angles (three are needed for complete specification of a single set of 3D axes) for 3T were reported to the Friesner Group, which then calculated the Raman spectra for each of the four orientations of the 3T molecules, finally averaging the corresponding intensities together (a process corroborated by the oriented gas model[58–61]) in order to simulate the final crystal Raman spectra at the precise sample orientations observed by the Willets Group.

The results are shown in Fig. 3.5 and Table 3.1. Panel A of Fig. 3.5 shows a bright-field image of the experimental sample from which the Raman spectra were measured, with the collection spot indicated by a circle. The physical sample corresponds to the yellow polygon in Panel B, which also displays the XRD-determined orientations (with an error of roughly 15°) of the eight molecules in the unit cell corresponding to the orientation of the sample relative to the optical microscope axis, which is perpendicular to the page. The corresponding measured Raman spectra as a function of the orientation angle ψ are displayed in Panel C, with both marker modes indicated. Panel D shows the corresponding simulated Raman spectra, and the agreement appears to be very good. Panel E shows a polar plot of the relative intensities with respect to ψ on the polar axis, in which the solid lines represent the simulated data and the symbols represent the experimental data. This plot corroborates how accurately the calculations simulate the experimental measurements, with good agreement for the modulation depths M_1 and M_2 (representing the degree to which the plots "pinch in") and the phases ϕ_1 and ϕ_2 (representing the "orientations" of the polar plots, i.e., the directions in which they appear to "point"). The overall scaling in the polar plot for the experimental and theoretical data was done by setting the maximum

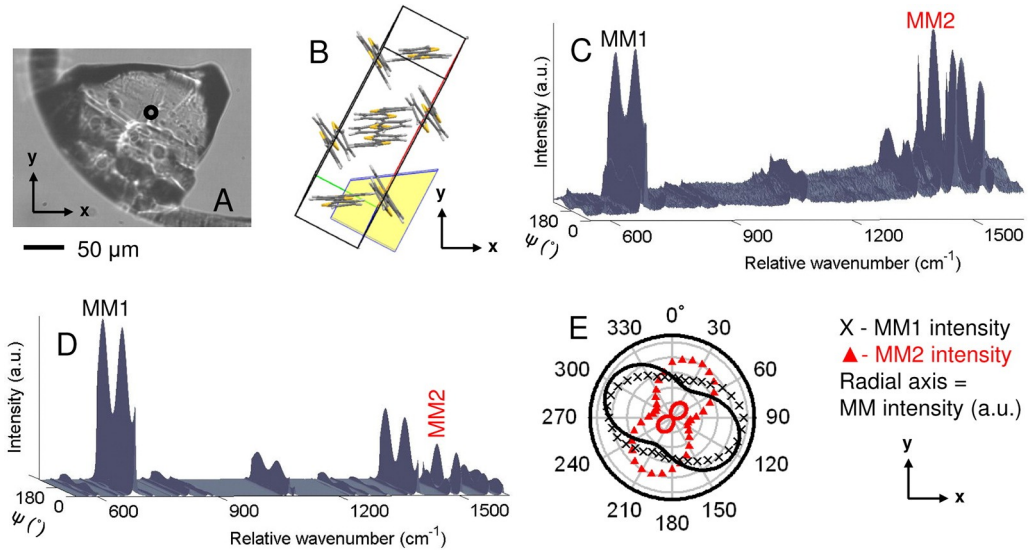

Figure 3.5: Raman spectra of crystalline 3T. (a) Bright-field image of the sample, with the collection spot circled in black. The optical axis is perpendicular to the page. (b) Relationship between the sample in (a), shown as a yellow polygon, to the molecules in the unit cell, as determined by XRD. The optical axis is perpendicular to the page, and note that for this sample orientation the long axes of the molecules are generally parallel to the optical axis. (c) Experimentally measured Raman spectra vs. orientation angle ψ. (d) Corresponding simulated Raman spectra. (e) Polar plots of the two marker mode intensities, in which the solid lines represent the simulated results and the symbols represent the experimental measurements. The Raman calculations employed the perturbing coordinate method with a step size of 0.05 Bohr (`pertnd=0.05`) at the B3LYP/Sadlej pVTZ level of theory. Graphic taken from Fig. 6 of [43].

intensities of the dominant marker mode (MM1) equal.

For all oligothiophene crystals studied experimentally and theoretically, we found a general trend that when the long axes of the molecules are nearly parallel to the optical axis (e.g., see Panel B of Fig. 3.5), the modes in the 1400–$1700\,\mathrm{cm}^{-1}$ region, which includes MM2, become similar in intensity to that of MM1, i.e., A approaches or even dips below 1. The sensitivity of the intensity of these modes to deviations from this orientation helps explain the disagreement between experiment and theory for this region of the spectra in Fig. 3.5, particularly when taking into consideration the registration errors of $\pm 15^\circ$ in determination of the unit cell orientation from the XRD data and of $< 5^\circ$ in the assignment of the Euler angles to each molecule in the unit cell. As the long axis of the molecules is tilted away from this "parallel" orientation, the intensity of the modes in the 1400–$1700\,\mathrm{cm}^{-1}$ region quickly increases, resulting in large values of the parameter A. This trend is illustrated in Fig. 3.6, which shows the intensities of MM1 as solid lines and those of MM2 as dashed lines. The "parallel" orientation is represented by 90° on this plot, and the deviation from this orientation increases as the molecule rotation value

		3T, orientation 1		3T, orientation 2	
		expt.	**theory**	**expt.**	**theory**
MM1	ν_1 (cm^{-1})	687	695 **(8)**	687	695 **(8)**
	M_1	0.43	0.56 **(13%)**	0.73	0.88 **(15%)**
	ϕ_1 (°)	99.7	118.6 **(11%)**	154.3	162.2 **(4.4%)**
MM2	ν_2 (cm^{-1})	1456	1505 **(49)**	1456	1505 **(49)**
	M_2	0.72	0.96 **(24%)**	0.93	1.00 **(7.0%)**
	ϕ_2 (°)	24.8	43.8 **(11%)**	154.7	163.0 **(4.6%)**
other data	A	0.80	0.20 **(-0.60)**	28	32 **(0.06)**
	$\phi_2 - \phi_1$ (°)	-74.9	-74.8 **(0.03%)**	0.4	0.8 **(0.11%)**
		4T, orientation 1		**4T, orientation 2**	
		expt.	**theory**	**expt.**	**theory**
MM1	ν_1 (cm^{-1})	687	698 **(11)**	687	698 **(11)**
	M_1	0.35	0.71 **(36%)**	0.77	0.91 **(14%)**
	ϕ_1 (°)	137.4	119.1 **(-10%)**	160.4	165.7 **(2.9%)**
MM2	ν_2 (cm^{-1})	1454	1492 **(38)**	1454	1492 **(38)**
	M_2	0.80	1.00 **(20%)**	0.81	1.00 **(19%)**
	ϕ_2 (°)	139.1	122.2 **(-9.4%)**	161.5	167.0 **(3.1%)**
other data	A	5.2	22 **(0.63)**	19	39 **(0.31)**
	$\phi_2 - \phi_1$ (°)	1.7	3.1 **(0.39%)**	1.1	1.3 **(0.06%)**

Table 3.1: Experimental and simulated values of the marker mode parameters $\{M_1, M_2, \phi_1, \phi_2, A\}$. The errors in the theoretical values relative to the experimental values are reported in bold in parentheses in the "theory" columns. The formulas for calculation of the errors are located in Section 3.5. Most data were taken from [43].

on the x-axis decreases. From the figure we also note that this variation in A at the two extreme orientations increases as the molecule length increases.

To exemplify this general trend, the Willets Group measured and the Friesner Group simulated the same measurements as in Fig. 3.5 except for the sample having a different orientation in the lab frame in which the molecules in the unit cell are far from the "parallel" orientation just described. The results are shown in Fig. 3.7 and Table 3.1. We note the drastic changes to the five parameters $\{M_1, M_2, \phi_1, \phi_2, A\}$. For this orientation of the crystal, we see that the agreement between experiment and theory for all five parameters, including the relative amplitude A, is very good.

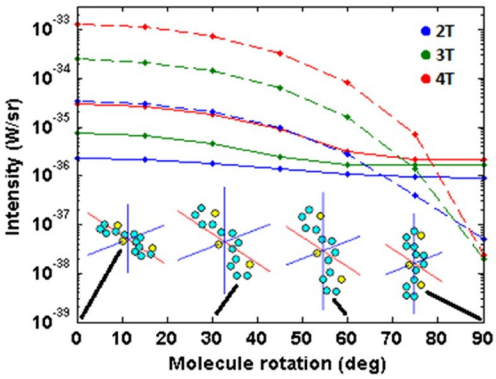

Figure 3.6: Dependence of the marker mode intensities on the orientation of the molecule's long axis relative to the Raman microscope's optical axis. Solid lines show the intensities of MM1, whereas dashed lines show the intensities of MM2. As the molecule is rotated toward an orientation in which its long axis is nearly parallel to the optical axis (represented by the vertical blue axis; 90° molecule rotation), the intensity of MM2 becomes largely suppressed relative to that of MM1 (A near zero). Normally, MM2 is much larger in intensity than MM1 (large A), as seen for smaller degrees of the molecule rotation. Insets of the molecules in the lab frame show 3T, but data are plotted for 2T (blue), 3T (green), and 4T (red). Note the logarithmic scale on the y-axis, which has SI intensity units of Watts per steradian. The Raman calculations employed the perturbing coordinate method with a step size of 0.05 Bohr (`pertnd=0.05`) at the B3LYP/Sadlej pVTZ level of theory. Graphic taken from Fig. 8 of [43].

Quaterthiophene

Analogous calculations were done for 4T, which contains two differently-oriented molecules per unit cell; the Willets Group therefore reported back six Euler angles. The results are shown in Figs. 3.8 and 3.9 and Table 3.1. Again, two sample orientations were used, but unlike for 3T, the orientations for 4T seemed to both be far from the "parallel" orientation described above, in which the long axis of the molecule is roughly parallel to the optical axis (there is no relation between the two orientations of the 3T and 4T samples). This is why Fig. 3.9 is much less different from Fig. 3.8 than Fig. 3.7 is from Fig. 3.5.

To explicitly quantify the degree of agreement between theory and experiment for the five marker mode parameters $\{M_1, M_2, \phi_1, \phi_2, A\}$, in Table 3.1 we have included reasonable calculations of error in bold in parentheses in the "theory" columns. For measurements that have a finite range of values, such as the modulation depth M ($\epsilon\,[0,1]$), phase ϕ ($\epsilon\,[0^\circ, 180^\circ]$), and phase difference $\Delta\phi \equiv \phi_2 - \phi_1$ ($\epsilon\,[-180^\circ, 180^\circ]$), we calculate the percent error in the theoretical calculations relative to the experimental measurements by normalizing the differences between the calculations and measurements by the width of the range of the differences (1, 180°, 360°, respectively). It does not make sense to normalize by the experimental measurements because this would

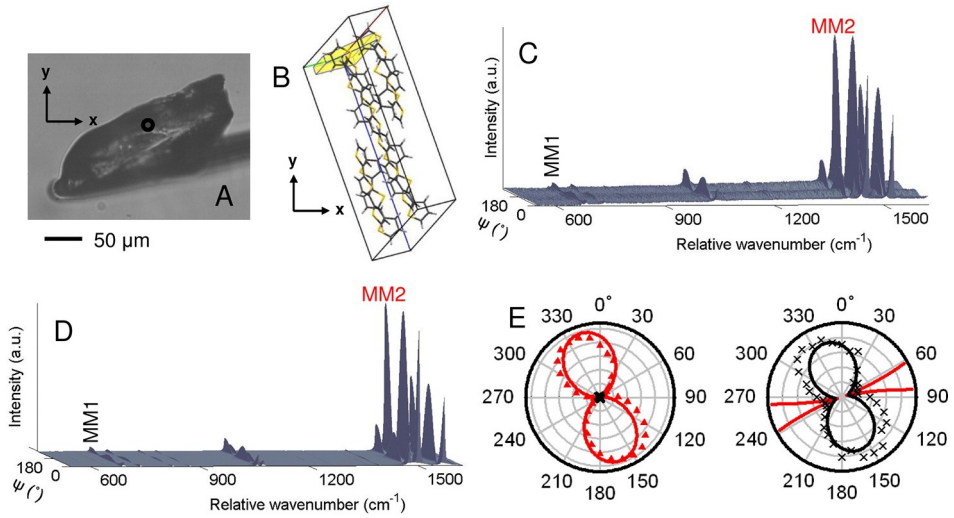

Figure 3.7: Raman spectra of crystalline 3T at a different relative orientation of the sample in the lab frame. (a) Bright-field image of the sample, with the collection spot circled in black. The optical axis is perpendicular to the page. (b) Relationship between the sample in (a), shown as a yellow polygon, to the molecules in the unit cell, as determined by XRD. The optical axis is perpendicular to the page, and note that for this sample orientation the long axes of the molecules are generally perpendicular to the optical axis. (c) Experimentally measured Raman spectra vs. orientation angle ψ. (d) Corresponding simulated Raman spectra. (e) Polar plots of the two marker mode intensities, in which the solid lines represent the simulated results and the symbols represent the experimental measurements. The second polar plot is a zoomed-in version of the first polar plot in order to show the data for MM1 more clearly. Graphic taken from Fig. 7 of [43].

erroneously inflate percent errors for experimental measurements near zero, when in reality any measurement within the possible range is equally likely. On the other hand, for measurements that have an infinite range of values, such as the wavenumber ν ($\epsilon\,[0,\infty]$) and amplitude A ($\epsilon\,[0,\infty]$), we simply take differences between the theoretical and experimental measurements and leave the units as they are specified in the second column (cm^{-1} and unitless, respectively). Further, it is most appropriate to express the error in A as a logarithm. Thus, the formulas we use for the error calculations are

$$\begin{aligned}
\text{err}\,[\nu] &\equiv \nu_{\text{theory}} - \nu_{\text{exp}} \\
\text{err}\,[M] &\equiv (M_{\text{theory}} - M_{\text{exp}}) \times 100 \\
\text{err}\,[\phi] &\equiv \frac{\phi_{\text{theory}} - \phi_{\text{exp}}}{180} \times 100 \\
\text{err}\,[A] &\equiv \log_{10} \frac{A_{\text{theory}}}{A_{\text{exp}}} \\
\text{err}\,[\Delta\phi] &\equiv \frac{\Delta\phi_{\text{theory}} - \Delta\phi_{\text{exp}}}{360} \times 100.
\end{aligned}$$

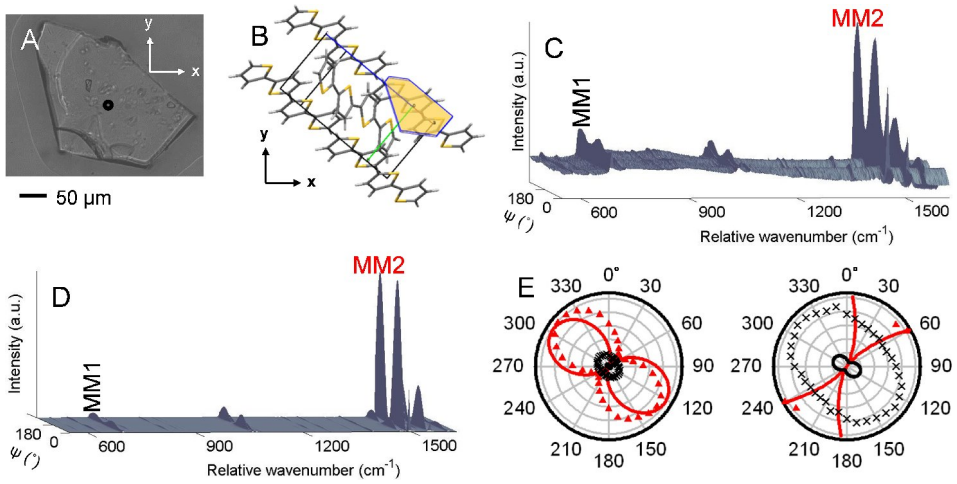

Figure 3.8: Raman spectra of crystalline 4T. (a) Bright-field image of the sample, with the collection spot circled in black. The optical axis is perpendicular to the page. (b) Relationship between the sample in (a), shown as a yellow polygon, to the molecules in the unit cell, as determined by XRD. The optical axis is perpendicular to the page, and note that for this sample orientation the long axes of the molecules are generally perpendicular to the optical axis. (c) Experimentally measured Raman spectra vs. orientation angle ψ. (d) Corresponding simulated Raman spectra. (e) Polar plots of the two marker mode intensities, in which the solid lines represent the simulated results and the symbols represent the experimental measurements. The second polar plot is a zoomed-in version of the first polar plot in order to show the data for MM1 more clearly. Graphic taken from Fig. S-1 of [43].

From the corresponding error values reported in Table 3.1, we see that the average percent error in the eight modulation depth values is about 19% (mean absolute value). Considering the aforementioned large registration errors of $\pm 15^\circ$ for the unit cell and $< 5^\circ$ for the molecules within the cell and the large sensitivity of the spectra to these orientation values, in our opinion 19% error in the modulation depth is very good. The percent error for the eight phase values is even better at 7%. The logarithmic error for the amplitudes, which describes the degree of agreement between theory and experiment in the order of magnitude of the amplitude values, also appears to be very good: all four simulated amplitude values are within a power of 10 of the measured amplitude values (the absolute value of err $[A]$ is always less than 1). In particular, the simulated values of A are always on the same side of 1 as the experimental values of A; the higher-intensity marker mode is always consistent. Finally, it is interesting to note the excellent agreement for the phase *difference* $\Delta\phi$, for which the average percent error is 0.15%! This may further indicate that the simulation methods are accurate and that it is just the orientation registration errors that contribute to the larger errors in the parameter values themselves.

Thus, not only have we shown that the parameters $\{M_1, M_2, \phi_1, \phi_2, A\}$ corresponding to MM1 and MM2 are very strongly correlated with the unit cell orientation, but also that with the theoretical methods presented we

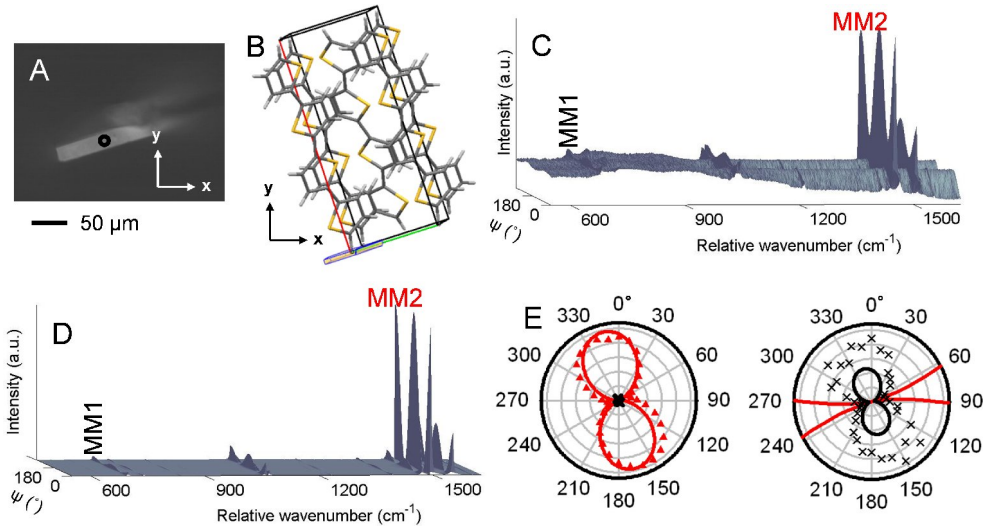

Figure 3.9: Raman spectra of crystalline 4T at a different relative orientation of the sample in the lab frame. (a) Bright-field image of the sample, with the collection spot circled in black. The optical axis is perpendicular to the page. (b) Relationship between the sample in (a), shown as a yellow polygon, to the molecules in the unit cell, as determined by XRD. The optical axis is perpendicular to the page, and note that for this sample orientation the long axes of the molecules are again generally perpendicular to the optical axis. (c) Experimentally measured Raman spectra vs. orientation angle ψ. (d) Corresponding simulated Raman spectra. (e) Polar plots of the two marker mode intensities, in which the solid lines represent the simulated results and the symbols represent the experimental measurements. The second polar plot is a zoomed-in version of the first polar plot in order to show the data for MM1 more clearly. Graphic taken from Fig. S-2 of [43].

can very accurately reproduce such correlations, indicating that, as long as registration error is minimized, many painstaking experimental measurements may be able to be largely replaced by relatively simple simulations for future studies, an example of which we present by the theoretical investigation described next.

3.6 Crystal orientation determination using polarized Raman spectra: theoretical study

Having shown that theory is able to reproduce the experimentally observed polarization-dependent Raman data using experimentally determined molecular orientations, we next asked the question of whether theory could predict molecular orientation given a set of Raman parameters. To this end, we simulated how well we could determine the orientation of a 3T unit cell, including all four uniquely oriented molecules, given its simulated Raman spectra and the orientations of the molecules inside the unit cell, the latter of which can simply be taken from any experimental XRD data.

First, we generated $N_{\text{pool}} = 119,952$ sets of Euler angles to represent a complete, unique, and uniform set of general unit cell orientations. For each unit cell orientation, we simulated the Stokes Raman spectra over a range of orientation angles ($\psi \in \{0, 3, \ldots, 177\}$). All spectra were calculated for the same experimental parameters previously noted in this chapter, i.e., for a backscattering geometry with no polarization analyzer in front of the detector. We thus obtained intensity vs. polarization angle ψ curves for each normal vibrational mode in the spectra for a given unit cell orientation. Unlike for the investigations above, for which we focused on two marker modes, here we expanded the set of marker modes to include all 51 vibrational modes with frequencies above 100 cm^{-1} (including the C–H stretch modes near 3200 cm^{-1}). By fitting each of the 51 polarized Raman intensities to eq. 3.1, we obtained three parameters for a mode k — modulation depth M_k, phase ϕ_k, and maximum amplitude $A_k \equiv \frac{a_k+b_k}{a_1+b_1}$ — for each of the 51 marker modes. As seen from the definition of A_k, the amplitude parameters were calculated relative to the maximum intensity of the lowest-frequency mode, $k = 1$. We repeated this process for each unit cell orientation in N_{pool}, ending up with $3 \times 51 = 153$ parameters for each of the N_{pool} unit cell orientations. We called this set of parameters the "pool" set.

We then generated $N_{\text{test}} = 100$ random unit cell orientations and calculated a set of 153 parameters for each of these orientations, in the same manner as for the pool set. We therefore ended up with 153 parameters for each of N_{test} unit cell orientations, calling this set of parameters the "test" set. Finally, in order to rank the unit cell orientations in the pool set given a set of parameters for any of the unit cell orientations in the test set, we created a very basic, non-fitted scoring function for matching the parameters of the αth "test" orientation to those of the jth "pool" orientation,

$$S_{\alpha j} \equiv \sum_{k=1}^{51} \left\{ \frac{|M_{kj} - M'_{k\alpha}|}{\max_l M_{kl}} + \frac{|A_{kj} - A'_{k\alpha}|}{\max_l A_{kl}} + \frac{1}{2}\sqrt{(\sin\phi_{kj} - \sin\phi'_{k\alpha})^2 + (\cos\phi_{kj} - \cos\phi'_{k\alpha})^2} \right\}.$$

Here, M_{kj} refers to the modulation depth of the kth mode for the jth orientation of the pool set, $M'_{k\alpha}$ refers to the modulation depth of the kth mode for the αth orientation of the test set, and similarly for the other two parameters ϕ and A. Note that each parameter M, ϕ, and A is "normalized" appropriately in order to treat each parameter on equal footing in the scoring function.

The "best" rankings (i.e., 1, 2, 3, etc.) correspond to Euler angles of orientations from the "pool" set being as close as possible to the Euler angles of the current "test" orientation. Because the incident polarization angle is unique up to only 180° due to the symmetry of the electric field vectors, in principle a single set of 153 parameters can correspond to two possible unit cell orientations: one rotated 180° about the optical axis with respect to the other. Therefore, in the following results, we count a pair of such unit cell orientations as a single "match" to a

given set of parameters. For example, say that for a given set of test parameters the following four best-matched unit cell orientations were obtained from the pool set (where the three values represent the associated Euler angles in degrees of the unit cell):

$$\begin{aligned} &\text{match 1:} && [15, 30, 240] \\ &\text{match 2:} && [195, 30, 240] \\ &\text{match 3:} && [280, 160, 10] \\ &\text{match 4:} && [100, 160, 10]\,. \end{aligned}$$

The first pair of unit cell orientations (matches 1 and 2) and the second pair (matches 3 and 4) would be ranked as one and two, respectively.

For the set of 153 parameters corresponding to each unit cell orientation in the test set, the pairs of unit cell orientations in the pool set were ranked using our scoring function. We defined the "true" match in the pool set to be the pair of orientations very similar to the known test orientation, by requiring each of the Euler angles of the true match to be within typically $\pm 5^\circ$ of the actual test Euler angles. In other words, the "true" match corresponded to the first orientation that would allow an experimentalist to make only minimal changes to the crystal in order to obtain the true orientation. No significant modifications were allowed, such as any sort of rotation (however minor) about an axis (aside from rotating by 180° about the optical axis to obtain the other member of the pair, of course). This matching was performed manually, checking each determined match using 3D plots of the orientations (relative to the lab frame) of the four uniquely oriented molecules in the unit cell.

As shown in Table 3.2, for 87 out of our 100 test unit cell orientations, the algorithm found the true pair of matches in the pool set to be ranked number one. For the 13 out of 100 test unit cell orientations that were not determined perfectly, 11 found the true pair of unit cell orientations to be ranked number two, one found the true pair to be ranked number three, and the last found the true pair to be ranked number four. Thus, in our test of determining the unit cell orientation given simulated Raman spectra, we usually determined the roughly correct pair of unit cell orientations as our first guess, and we always determined the correct pair of unit cell orientations by our fourth guess, out of about $\frac{N_{\text{pool}}}{2} = 59,976$ possible guesses.

When the "true" pair of orientations was not found to be ranked first, the orientations that ranked better than the true pair, i.e., false positives, did so most often because the correct contributions to the average spectra came from the wrong molecules within the unit cell. The other false positive orientations ranked better because

Rank of true pair of orientations	**Number of test cases ($N_{\text{test}} = 100$ total cases) obtaining that rank**
1	87
2	11
3	1
4	1
≥ 5	0

Table 3.2: Results of the orientation-determination algorithm. Data taken from [43].

they were rotated about an axis orthogonal to the optical axis yet appeared similar when viewed down the optical axis. Thus, both types of false positive orientations[2] were due to the symmetry of the molecules and unit cell. Likely, for any particular experimental property of interest such as conductivity, such symmetric orientations of the actual crystal would be completely inconsequential. In other words, the "wrong" (symmetric) orientation may very well yield the exact same conductivity as the "right" orientation anyway.

3.7 Discussion

We have shown that we can identify marker modes of oligothiophenes that are dependent upon molecular orientation and vary systematically in the mode intensity as a function of the polarization angle of the excitation light relative to the sample. Theoretical calculations reproduce the experimental marker mode Raman intensity data with very good accuracy; furthermore, at least in the simple idealized test case of Section 3.6, structure can be inferred from the spectra alone with excellent reliability. While there are sources of noise in the experiment, the systematic behavior can be readily perceived above the noise level, and matches the results from theoretical calculations in a robust and consistent fashion. The results shown here are highly encouraging with regard to the possibility of employing polarized Raman spectroscopy to extract useful information concerning unit cell orientation, key torsional conformations, degree of disorder within a crystal (percentage of molecules at a 0° torsion angle as opposed to the lower-energy 180° conformation), local morphology within polymorphic crystals, degree of crystallinity or presence of amorphous molecules in a partially-ordered sample, and distribution of molecules within the unit cell for crystalline and polycrystalline materials composed of conjugated polymers.

[2] See the Supporting Information to [43] for visual examples.

In considering prospects for analyzing more complex systems than those considered here, it is important to begin by summarizing the sources of noise in both the calculations and experimental data. The DFT calculations do not perfectly reproduce experimental vibrational frequencies; there are a number of factors that likely contribute to discrepancies, including (1) the use of single molecule calculations rather than calculations that would include the crystalline environment (e.g., periodic boundary condition DFT methods or the use of large clusters containing neighboring molecules) and (2) intrinsic errors in the DFT functional employed. Both of these issues can be addressed in principle: the former by employing larger models or solid state DFT approaches as noted above and the latter by scaling force constants as is common in the DFT literature[45–47]. Obtaining better frequency agreement would further assist in automating the orientation-determination process. However, as long as one can match the peaks in the DFT spectrum with those in the experimental spectrum, exact matching of calculated and experimental frequencies does not seem to be critical to the determination of structural information, which appears to be more dependent upon comparisons of the computed and experimental Raman intensities[8] as a function of polarization angle. These intensities appear to track quite well in all cases examined herein. Improvements may further be made to the calculated Raman intensities by relaxing assumptions of mechanical and/or electrical harmonicity.

It should be noted that the theoretical approximation of the molecular crystal modes and spectra by the modes/spectra of the constituent non-tumbling planar molecules, ignoring packing effects completely, was quite good; as mentioned, mode frequencies were still useful, and the agreement of the mode intensities with our experimental results was excellent. Such good approximations have been proposed by for example [2] and evidenced by for example [62]. While it is plausible that the use of a plane-wave electronic structure code would yield even better results, one of the largest physical additions it would make would be phonons corresponding to wavelengths larger than the molecule size (i.e., vibrational modes that are not internal to the individual molecules and thus cannot be determined from our method of approximating the crystal modes by the molecular modes). Raman spectra in the lattice phonon region recorded from pentacene[34] and oligothiophene[31, 32] crystals did show peak intensity and position changes at different incident polarizations. However, the degree to which phonon mode intensity fluctuates with incident polarization compared to intramolecular mode intensity fluctuation with incident polarization is unclear at this time.

In this study we noted some issues associated with the experimental data that likely contribute significantly to the discrepancies between the experimental and predicted Raman spectra. One challenge we faced in these studies

was imperfect registration between the XRD laboratory frame and the optical laboratory frame, introducing error into our structural assignments. This is a significant challenge for any correlated optical/structure experiment, yet our Raman data indicate that we may ultimately be able to circumvent this issue by using theoretically calculated Raman data to predict the structure based on matching the polarized response of experimentally measured Raman data. As mentioned previously, assigning the Euler angles of individual molecules within each unit cell based on the XRD data also introduced an additional error source. In addition, we must recognize that XRD provides an ensemble-averaged structure for a crystalline material; if amorphous regions are present in the crystal, these would not necessarily emerge in the XRD data, yet would contribute to the overall measured Raman spectrum. The fact that Raman is sensitive to these sub-domains within an ensemble crystal, as capitalized on in the following chapter, is encouraging for future development and applications of this technique, yet it introduces an additional source of error when trying to match our experimental data to predicted spectra. Finally, the study of crystalline 2T proved difficult due to its instability at room temperature, though such experimental instabilities attest to the value of our simulation methods presented herein.

While using polarized Raman spectra to measure molecular orientation and internal structure remains a work in progress, our initial data demonstrate that we have a very promising technique for using optical measurements to determine the unit cell orientation within organic crystals and crystalline thin films. In the present study, we used the parameters corresponding to all 51 modes above $100\,\mathrm{cm}^{-1}$ for 3T, but it is likely that unique crystal orientations can be determined using far fewer modes, such as limiting the marker modes to one of each type of vibration (e.g., particular bends, stretches, etc.). This is especially important for comparison with experimentally measured Raman data, where low signal-to-noise may suppress observation of all modes. Further, the matching algorithm used in this study is a truly minimal scoring function that normalizes the parameters and determines the total parameter "distance" between a test spectrum and a pool spectrum; it is not fit to any data and can likely be improved greatly. Further avenues for improvements and removal of ambiguities in unit cell orientation include using different scattering angles and using detected radiation at different polarizations, which would likely aid in studies of individual molecule conformation as well.

We have studied crystalline and fluid systems for which the oligomer torsional profiles are well-defined and distinct. This simplifies the problem in that it removes uncertainties that would arise if the torsion angle populations were not known with high certainty, as might happen when studying a less well-characterized conjugated polymer sample. In the latter situation, simulations providing a predicted distribution of torsion angles would be an

essential component of addressing the problem. This situation arises routinely in efforts to compare spectroscopic data with complex systems, for example in the use of two-dimensional IR spectroscopy in elucidating molecular conformational distributions[63]. Both the Willets and Friesner Groups have been developing simulation methods for modeling conjugated polymer distributions[64] and their application in this context is certainly feasible.

Ultimately, it is likely that a combination of spectroscopic methods will be required to provide the maximum possible structural information for conjugated oligo/polymer systems, as the use of multiple probes enables us to overcome noise and ambiguities by requiring that the structural model reproduce all of the data. This is a challenging task, the difficulty of which should not be underestimated. What the present chapter shows is that polarized Raman spectroscopy can make an important, even central, contribution to this effort. Further experiments on increasingly heterogeneous systems should provide indications as to how far the method can be pushed forward.

Chapter 4

A first-principles polarized Raman method for determining whether a uniform region of a sample is crystalline or isotropic

4.1 Introduction

The nano- and micro-scale ordering of a material's components largely determines the macroscopic properties of the material, including charge mobility [65, 66], activation energy for charge carrier transport [66], fluorescence [67–70], photosensitivity [71], infrared [72] and Raman [43, 72] intensity, thermoelectric properties [73], and even appearance [70]. However, the small-scale ordering of some materials, such as that for conducting polymers, is generally quite complex and can be difficult to describe. For example, a largely ordered material can contain small amorphous regions; a material composed of several small crystalline domains can appear amorphous on a longer length scale due to the lack of alignment between the various domains; and even a crystalline material is never perfect. Given this complexity and its influence on functional properties (and how they may spatially vary), having a good way to probe the "crystallinity" at various locations within a given sample is tremendously useful. A less precise — yet more well-defined — characterization than the crystallinity is the "state of internal orientation" $\mathscr{S}$, which can take on either of the two ideal states that lie at the ends of the crystallinity spectrum: "crystalline," in which the molecules are arranged periodically in well-defined unit cells, or "isotropic," in which the molecules are freely rotating (as in a liquid or gas) or randomly oriented (as in an amorphous solid or powder). As the

"degree of crystallinity" of a sample is itself defined by the IUPAC by first approximating the sample as consisting of pockets of uniformly crystalline and uniformly amorphous regions or domains [74], at small enough scales it is indeed reasonable to describe a heterogeneous sample by the states of internal orientation $\{\mathscr{S}_i\}$ of the uniform, homogeneous regions of which it is composed.

Current methods for determining the state of internal orientation $\mathscr{S}$ of a uniform region include X-ray [75–80], electron [81], and neutron [82] diffraction; differential scanning calorimetry [83, 84]; measurement of density [85, 86]; measurement of specific volume, heat, and enthalpy [87, 88]; and infrared [79, 86] and NMR [78, 89] spectroscopy. However, as we will discuss, all of these methods rely on *a priori* knowledge of reference values of properties of purely crystalline regions and purely isotropic regions [87, 88]. Further, these methods can be ambiguous in their assessment of state $\mathscr{S}$ for particular materials and, for a given material, the spectral methods can be ambiguous when using particular peaks — there is no guarantee that a particular peak will be sufficiently sensitive to $\mathscr{S}$ [87, 88].

Raman spectroscopy [78, 80, 90–92] offers advantages over many of these characterization techniques because it provides excellent spatial resolution (limited only by the diffraction limit), is non-destructive to the sample, and can be performed *in situ*. Unfortunately, it too suffers from drawbacks. Previous methods for using Raman spectroscopy to probe the state of internal orientation $\mathscr{S}$ [78, 90, 92–95] involve the analysis of one or more marker modes ($\mathscr{S}$-sensitive modes) that must first be correlated to the states $\mathscr{S}$ of known sample regions and then used to back out the approximate states of unknown regions, all using somewhat involved curve-fitting techniques. In addition, such methods are specific to particular modes of particular systems. Clearly, it would be advantageous to rely on a simpler Raman method for determining the state of internal orientation $\mathscr{S}$ that requires no *a priori* knowledge regarding the behavior of specific peaks and could be applied to any vibrational mode of any system.

The findings presented in Chapter 3 suggest just such an approach. In Fig. 4.1 we show how the Raman intensity scattered from a crystalline terthiophene sample with the unit cell orientation shown in Panel (a) modulates as the polarization of the incident light is changed. Panel (b) shows how the intensity of different vibrational Raman modes changes as the orientation angle ψ of the excitation light is rotated through 360°. From these data, we characterize the magnitude of the oscillation in peak intensity by calculating the "modulation depth" (Panel (c)), which describes the ratio of the difference between the angle-dependent maximum and minimum intensities to the value of the maximum intensity and is defined by eq. 3.2. The modulation depth of a mode for

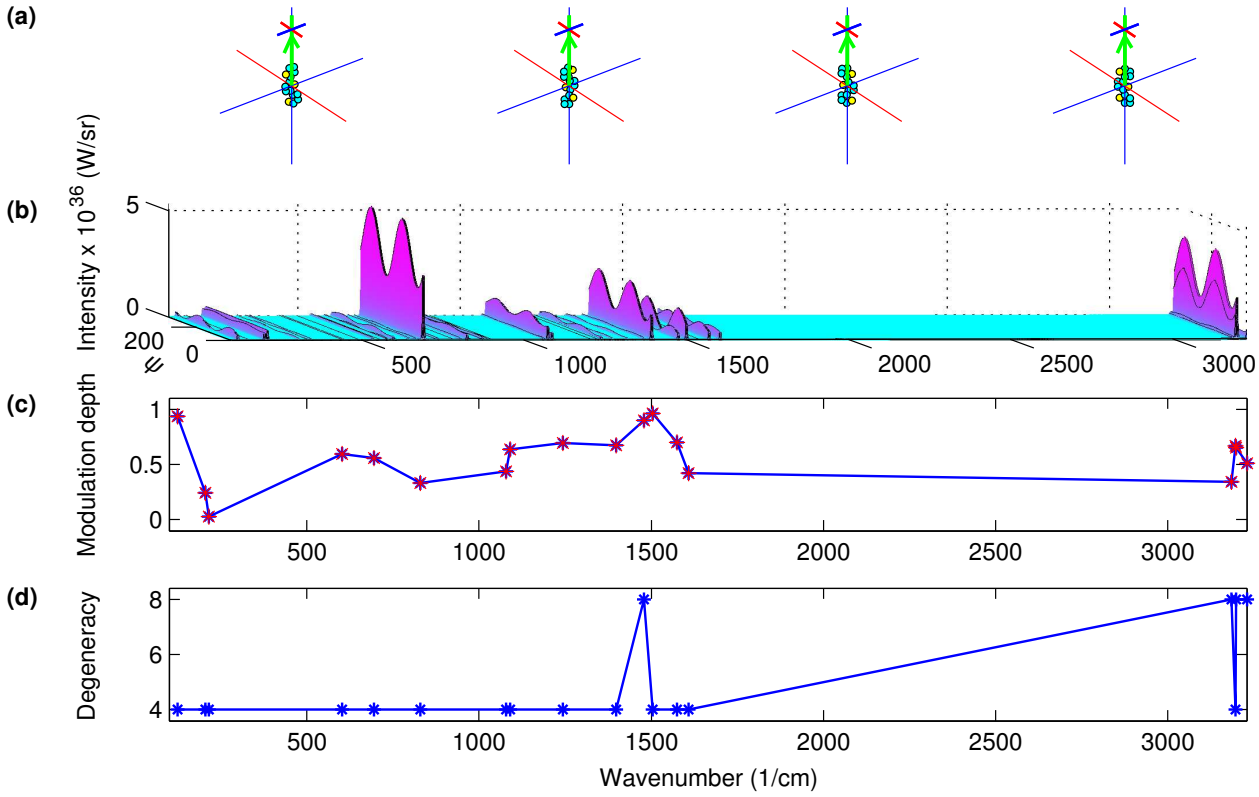

Figure 4.1: Simulated Raman data for crystalline terthiophene under the experimental conditions used in Chapter 3, which showed excellent agreement between these simulations and experiment. (a) Orientation of each of the four molecules in the unit cell relative to the backscattering geometry. (b) Raman spectrum vs. orientation angle ψ ($\in [0°, 360°]$), which refers to the orientation angle of the incident, elliptically polarized light (see Chapters 1 and 3). (c) Modulation depths for the most intense modes in (b), using eq. 3.2. All datapoints are red to indicate that every mode is vibrationally degenerate, which is an approximation used here in the oriented gas model [58, 61]. Note how in general the modulation depths vary largely throughout their domain, $M \in [0, 1]$. (d) Vibrational degeneracies of the most intense modes of the unit cell; the fact that the smallest degeneracy is four is a consequence of the oriented gas model. The elements of this figure are described in more detail in Section 4.2. For all simulations in this chapter, the perturbing coordinate method in Jaguar was used for calculating the Raman spectra at the B3LYP/Sadlej pVTZ level of theory. Lorentzian intensity functions of linewidth (FWHM) $2\,\mathrm{cm}^{-1}$ were fit to each mode.

the experimental setup used in Chapter 3, as for many experimental setups, is highly dependent on the mode's Raman tensor and thus varies substantially from mode to mode. Furthermore, as shown in Chapter 3, the Raman tensor itself depends strongly on the unit cell orientation, so that a full set of modulation depths can be used to back out the orientation of a crystalline sample given the Raman tensors of the modes at a known cell orientation, as demonstrated in Section 3.6.

However, this very dependence of the modulation depth on the mode's Raman tensor and unit cell's orientation

under arbitrary experimental conditions precludes using the modulation depths to determine whether a uniform region of a sample is crystalline or isotropic in the first place, requiring use of alternate techniques (we used X-ray diffraction in the experiments of Chapter 3) to confirm that the region is crystalline before backing out the orientation of the unit cell. Thus, the aim of the present study became to determine whether specific experimental conditions exist that simultaneously (1) remove the dependence of the modulation depth on the Raman tensor and cell orientation, and (2) cause the resulting modulation depth to be different for the crystalline and isotropic states of internal orientation $\mathscr{S}$. Such a set of experimental conditions would allow us to unambiguously assign the states of the uniform regions of a material, after which, if desired, we could use other experimental conditions to back out the relative orientations of the crystalline regions.

In this chapter, we report such a special set of vibrational Raman scattering conditions, allowing us to indeed determine whether a uniform region of a sample is purely crystalline or purely isotropic, using as our signal the modulation depth under this special set of experimental conditions. Our findings are derived, explained, and applied below, based on the mathematical framework rigorously built up from the basic polarized Raman theory of Chapter 1. As such, our conclusions are based on first principles and are general, applicable to any vibrational mode of any material. We discuss what can be inferred from the special modulation depth signal first for vibrationally non-degenerate modes and then for vibrationally degenerate modes. We include simulations on realistic terthiophene systems to provide examples of conclusions that can be drawn from various experiments. Further, we describe how our signal can show whether a mode is vibrationally degenerate, which could be utilized for example in conjunction with mode wavenumber for species identification in a sample of unknown composition. Finally, in Section 4.3 we summarize our findings and discuss the advantages and disadvantages of our method.

4.2 A modulation depth signal to determine region state, sample uniformity, and mode degeneracy

First, we must emphasize that there are two types of non-trivial (i.e., not part of the coefficient $\frac{(\omega_1-\omega_k)^4 N\hbar\mathcal{I}}{64\pi^2\epsilon_0^2 c^4\omega_k\left(1-e^{-\hbar\omega_k/k_B T}\right)}$) variables involved in the intensity formulas derived in Chapter 1. The first type are directly controllable, known, experimental variables; we identify these variables as "experimental" and define a particular set as the "experimental conditions." The second type are uncontrollable, initially unknown, sample-dependent variables; we identify these variables as "sample-dependent." The experimental variables are the incident ($\{P,\chi,\psi\}$) and analyzed (M^{det}) polarizations and the scattering angle θ (Fig. 1.2a). The sample-dependent

variables are the state of internal orientation $\mathscr{S}$ (either crystalline or isotropic), the Raman tensor α_k, and, if the state is crystalline, the unit cell orientation T relative to the lab frame. Note also that while for clarity we use the constant Raman tensor α_k in the unit cell frame and describe the orientation of this frame relative to the lab frame by the matrix T, we could instead absorb T into α_k and define

$$\tilde{\alpha}_k \equiv T^T \alpha_k T, \tag{4.1}$$

producing the variable Raman tensor in the lab frame and disposing of T altogether.

Our goal is to find, using the Raman theory from Chapter 1, a signal of the state of internal orientation $\mathscr{S}$ of a uniform region of a sample, i.e., a signal indicating whether a uniform region is purely crystalline or purely isotropic. (Note that we use the symbol $\mathscr{S}$ to refer to a uniform region that is either purely crystalline or purely isotropic, i.e., $\mathscr{S} \in \{\text{crystalline,isotropic}\}$, and for clarity we will often use "state" to refer to this "state of internal orientation.") We note that there have already been studies that have attempted to determine the other two sample-dependent variables aside from state $\mathscr{S}$: the values of the Raman tensor α_k [62, 96, 97] and, for a crystalline sample, the unit cell orientation T [43, 98, 99]. However, all these studies have applied only to samples of particular compositions. We wanted our method to rather be as transferable and as general as possible; we thus imposed the stipulation that our signal determine the state $\mathscr{S}$ of the sample *independent* of the Raman tensor α_k and unit cell orientation T. Such a signal would thus be extremely general and apply to any sample that could be analyzed using Raman spectroscopy, and it would require no *a priori* knowledge of any sample-dependent correlations.

As mentioned in Section 4.1, we saw potential in the modulation depth $M \in [0, 1]$ as a signal of the state $\mathscr{S}$. We defined M in eq. 3.2 and express it in an equivalent form, for vibrational mode k, as

$$M_k \equiv \frac{\max_\psi I_k - \min_\psi I_k}{\max_\psi I_k}. \tag{4.2}$$

In general the intensity I_k of mode k depends on the variables $\left\{\mathscr{S}, \alpha_k, T; P, \chi, \psi, M^{\text{det}}, \theta\right\}$, so that $M_k = M_k\left(\mathscr{S}, \alpha_k, T; P, \chi, M^{\text{det}}, \theta\right)$. In this notation, the variables before the semicolon are "sample-dependent" and the variables after the semicolon are "experimental," as described above. We thus began by exploring the functional dependence of M_k, and in particular, on the orientation angle ψ. We showed such an investigation in Fig. 4.1, which we described in Section 4.1. Here we note in particular that Panel (b) demonstrates the sinusoidal dependence of the intensity profiles I_k on ψ with a period of $180°$. The modulation depths M_k describe how deep these oscillations are; these values are shown in Panel (c), and we note that they vary greatly for the different

modes, for the experimental conditions used in that experiment. Note that while the spectra in Panel (b) show all vibrational modes of the unit cell, only the observable, most intense modes (which we define as those with intensity profile maxima within a factor of 35 of the most intense maximum) are shown in Panels (c) and (d). This allows us to simulate only what an experimenter might find above the background noise. Further, note that while from the vibrational degeneracies shown in Panel (d) all modes appear to be degenerate, this is in fact an artifact of the use of the oriented gas model, which assumes that there are no interactions between any molecules [58, 61]. In reality far fewer degenerate modes will exist, whose symmetry breaking is caused by the intra- and inter-cellular interactions.

However, we showed in Chapter 3 that for the experimental conditions used, $\left\{P=1, \chi=0, M^{\mathsf{det}}=I, \theta=\pi\right\}$, the dependence of the modulation depth on the Raman tensor and unit cell orientation remained, i.e., $M_k = M_k\left(\mathscr{S}, \alpha_k, T; P=1, \chi=0, M^{\mathsf{det}}=I, \theta=\pi\right)$. Our goal thus became to vary all the controllable, experimental variables $\left\{P, \chi, M^{\mathsf{det}}, \theta\right\}$ until dependence of the modulation depth on the two uncontrollable, sample-dependent variables α_k and T dropped out, i.e., $M_k = M_k\left(\mathscr{S}; P, \chi, M^{\mathsf{det}}, \theta\right)$.

Vibrationally non-degenerate modes

We first focused on modes that were vibrationally non-degenerate. We quickly learned that the general formulas for $M_k\left(\mathscr{S}=\mathsf{crys}, \alpha_k, T; P, \chi, M^{\mathsf{det}}, \theta\right)$ and $M_k\left(\mathscr{S}=\mathsf{iso}, \alpha_k; P, \chi, M^{\mathsf{det}}, \theta\right)$ were too complicated to reasonably write down for arbitrary experimental variables $\left\{P, \chi, M^{\mathsf{det}}, \theta\right\}$ and thus study from the general equations how to vary $\left\{P, \chi, M^{\mathsf{det}}, \theta\right\}$ to free M_k of its dependence on α_k and T. We thus resorted to plugging in common values of the experimental variables $\left\{P, \chi, M^{\mathsf{det}}, \theta\right\}$ and observing with Mathematica whether the resulting formulas for M_k retained any dependence on α_k and T; after all, if we expected our method to be applied experimentally, the experimental variables would certainly have to take on common values.

The values we used for the degree of polarization P were $\left\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\right\}$, and for the angle of ellipticity χ we used the values $\left\{-\frac{\pi}{4}, 0, \frac{\pi}{4}\right\}$. For the analyzer, specified by the matrix M^{det}, we used matrices corresponding to four different orientations of linearly polarized light; a quarter-wave plate, half-wave plate, and five other phase retardations; left- and right-handed light; and no analyzer at all. Finally, for the scattering angle θ we used the values $\left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi\right\}$. For all combinations of these experimental variables $\left\{P, \chi, M^{\mathsf{det}}, \theta\right\}$, we calculated using Mathematica the modulation depths M_k in the crystalline and isotropic cases by analytically varying ψ and observing the dependence of $M_k\left(\mathscr{S}=\mathsf{crys}, \alpha_k, D; P, \chi, M^{\mathsf{det}}, \theta\right)$ and $M_k\left(\mathscr{S}=\mathsf{iso}, \alpha_k; P, \chi, M^{\mathsf{det}}, \theta\right)$ on α_k and

T.

We found many sets of experimental parameters $\left\{P, \chi, M^{\mathsf{det}}, \theta\right\}$ for which $M_k = M_k\left(\mathscr{S} = \mathsf{crys}; P, \chi, M^{\mathsf{det}}, \theta\right)$ and $M_k = M_k\left(\mathscr{S} = \mathsf{iso}; P, \chi, M^{\mathsf{det}}, \theta\right)$, i.e., for which there was simultaneously no dependence of the modulation depth on the Raman tensor and unit cell orientation for both the crystalline and isotropic states. However, in almost all these cases, $M_k\left(\mathscr{S} = \mathsf{crys}; P, \chi, M^{\mathsf{det}}, \theta\right) = M_k\left(\mathscr{S} = \mathsf{iso}; P, \chi, M^{\mathsf{det}}, \theta\right) = 0$, thus providing no way to distinguish between the sample's state of internal orientation $\mathscr{S}$ using M_k under these experimental conditions. However, our exciting finding was that there was indeed one subset of cases, $\left\{P = \text{non-zero}, \chi = 0, M^{\mathsf{det}} = \mathsf{horiz}, \theta = \frac{\pi}{2}\right\}$ ("horiz"[ontal] linear polarization means that only light that is linearly polarized parallel to the scattering plane can pass through — see Fig. 1.2), for which not only the dependence of M_k on α_k and T dropped out for both the crystalline and isotropic states, but also the particular values of M_k were *different* between the crystalline and isotropic states: $M_k\left(\mathscr{S} = \mathsf{crys}; P, \chi = 0, M^{\mathsf{det}} = \mathsf{horiz}, \theta = \frac{\pi}{2}\right) = \frac{2P}{1+P}$ and $M_k\left(\mathscr{S} = \mathsf{iso}; P, \chi = 0, M^{\mathsf{det}} = \mathsf{horiz}, \theta = \frac{\pi}{2}\right) = 0$.

In particular, for completely polarized light ($P = 1$), $M_k\left(\mathscr{S} = \mathsf{crys}; P = 1, \chi = 0, M^{\mathsf{det}} = \mathsf{horiz}, \theta = \frac{\pi}{2}\right) = 1$. To be explicit, upon applying these special experimental conditions to the Stokes intensity equations of Chapter 1 (eq. 1.23 for the crystalline state and eq. 1.30 for the isotropic state), we obtain the intensity profiles

$$\begin{aligned} I_k^{\mathsf{crys}}(\psi) &= \frac{\left(\omega_1 - \omega_k\right)^4 \mathcal{I}\hbar N}{64\pi^2\epsilon_0^2 c^4 \omega_k \left(1 - \exp\left\{-\hbar\omega_k/k_B T\right\}\right)} \left(\tilde{\alpha}_{xy,k}^2 + \tilde{\alpha}_{xz,k}^2\right) \\ &\quad \times \left[1 + \cos\left(2\psi + \tan^{-1}\left(\tilde{\alpha}_{xz,k}^2 - \tilde{\alpha}_{xy,k}^2, -2\tilde{\alpha}_{xy,k}\tilde{\alpha}_{xz,k}\right)\right)\right] \qquad (4.3) \\ I_k^{\mathsf{iso}}(\psi) &= \frac{\left(\omega_1 - \omega_k\right)^4 \mathcal{I}\hbar N}{64\pi^2\epsilon_0^2 c^4 \omega_k \left(1 - \exp\left\{-\hbar\omega_k/k_B T\right\}\right)} \frac{6}{45}\gamma_k^2, \qquad (4.4) \end{aligned}$$

where the quadrant-specifying form of the inverse tangent is $\tan^{-1}(x, y)$, representing $\tan^{-1}\frac{y}{x}$, consistent with Mathematica's definition. Applying eq. 4.2 for the modulation depth, we obtain

$$\begin{aligned} M_k^{\mathsf{crys}} &= 1 \\ M_k^{\mathsf{iso}} &= 0, \end{aligned}$$

which are indeed both fully independent of the Raman tensor α_k and unit cell orientation T. Further, these values of modulation depth M_k for the different states $\mathscr{S}$ are not only different, but they also occupy the extreme values of M_k's domain, $[0, 1]$, making M_k under these special experimental conditions ideal for discriminating between the crystalline and isotropic states of the illuminated region.

It will prove useful later to define the amplitude A_k as the maximum profile intensity over ψ (ignoring the

coefficient $\frac{(\omega_1-\omega_k)^4 N\hbar\mathcal{I}}{64\pi^2\epsilon_0^2 c^4\omega_k\left(1-e^{-\hbar\omega_k/k_B T}\right)}$) and the phase ϕ_k as the value of ψ at this maximum, yielding

$$A_k^{\text{crys}} = 2\left(\tilde{\alpha}_{xy,k}^2+\tilde{\alpha}_{xz,k}^2\right) \quad (4.5)$$

$$A_k^{\text{iso}} = \frac{6}{45}\gamma_k^2$$

$$\phi_k^{\text{crys}} = \frac{1}{2}\tan^{-1}\left(\tilde{\alpha}_{xz,k}^2-\tilde{\alpha}_{xy,k}^2,-2\tilde{\alpha}_{xy,k}\tilde{\alpha}_{xz,k}\right) \quad (4.6)$$

$$\phi_k^{\text{iso}} = \text{undefined.}$$

We see that under the special experimental conditions, for both the crystalline and isotropic states, while the modulation depths M_k are independent of the Raman tensor α_k and unit cell orientation T, the amplitudes and phases are strongly dependent on α_k and T (of course, T is meaningless for an isotropic state).

As an illustration of our modulation depth signal's ability to clearly differentiate between the two states of internal orientation $\mathscr{S}$, see Fig. 4.2 (crystalline case) and 4.3 (isotropic case). In Fig. 4.2, we analyze crystalline terthiophene at the same orientation as in Fig. 4.1, except now we use the special experimental conditions just discussed. Again, due to use of the oriented gas model in our simulations, all vibrational modes are forced to be degenerate, which is why most of the modulation depths are not exactly equal to 1, as explained in the next section. But we see that even in this fully degenerate, "worst-case" scenario, the modulation depths are indeed close to 1. The bottom line is that the modulation depth for any single mode has $M_k \approx 1$ and thus indicates that the illuminated region is crystalline. In contrast, Fig. 4.3 shows clearly that the modulation depth of any single mode, even a degenerate one, has $M_k = 0$ exactly and thus indicates that the illuminated region is isotropic. Importantly, note that because *all* modes must obey this relationship between M_k and state of internal orientation $\mathscr{S}$, observation of more modes will greatly compound the certainty of determination of the region's state.

The modulation depth as defined in eq. (4.2) requires the orientation angle ψ of the incident light to do the varying. For completeness, we applied the same analysis for a modulation depth defined instead with respect to variation of the orientation angle ψ' of a linear analyzer, where $\psi' = 0$ corresponds to a horizontal linear analyzer. We thus obtained the modulation depth by fixing ψ and observing the change in intensity as ψ' was varied:

$$M_k' \equiv \frac{\max_{\psi'} I_k - \min_{\psi'} I_k}{\max_{\psi'} I_k}.$$

We then repeated the same analysis as before, except rather than letting M^{det} take on discrete values, we let the incident orientation angle ψ take on the discrete values $\left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right\}$, where $\psi = 0$ corresponds to, for linearly polarized light, the polarization axis parallel to the scattering plane. The results of the analysis using M_k' were similar to those using M_k: for the experimental conditions $\left\{P = 1, \psi = 0, \chi = 0, \theta = \frac{\pi}{2}\right\}$,

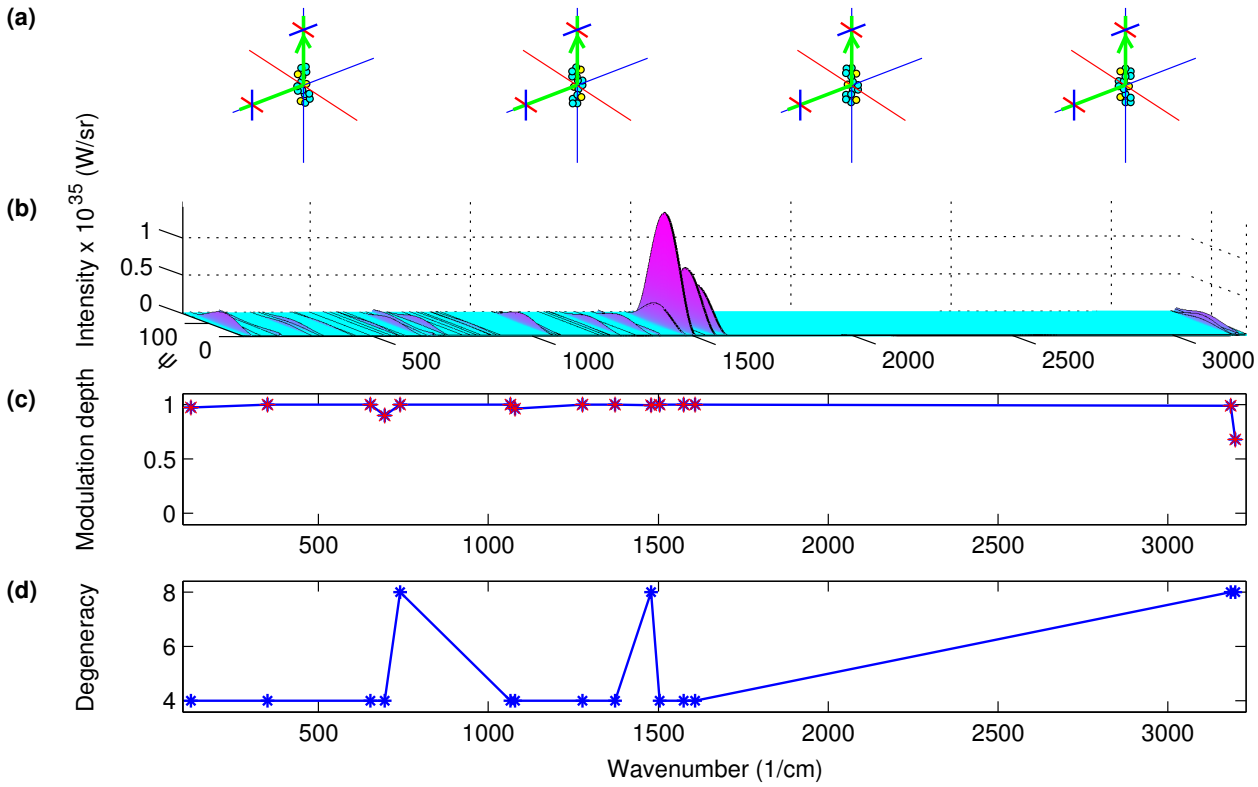

Figure 4.2: Simulated Raman data for crystalline terthiophene under the special experimental conditions that are the subject of this chapter. (a) Orientation of each of the four molecules in the unit cell relative to the $90°$ scattering geometry; these are the same orientations as in Fig. 4.1 for a different set of experimental conditions. (b) Raman spectrum vs. orientation angle ψ ($\in [0°, 180°]$), which again refers to the orientation angle of the incident, linearly polarized light. The analyzer in front of the detector allows through only light that is linearly polarized parallel to the scattering plane, defined by the blue axes in (a). (c) Modulation depths for the most intense modes in (b), using eq. (4.2). Note how in general the modulation depths are largely near 1, as expected for scattering from a crystalline region under these special experimental conditions. (d) Vibrational degeneracies of the most intense modes of the unit cell.

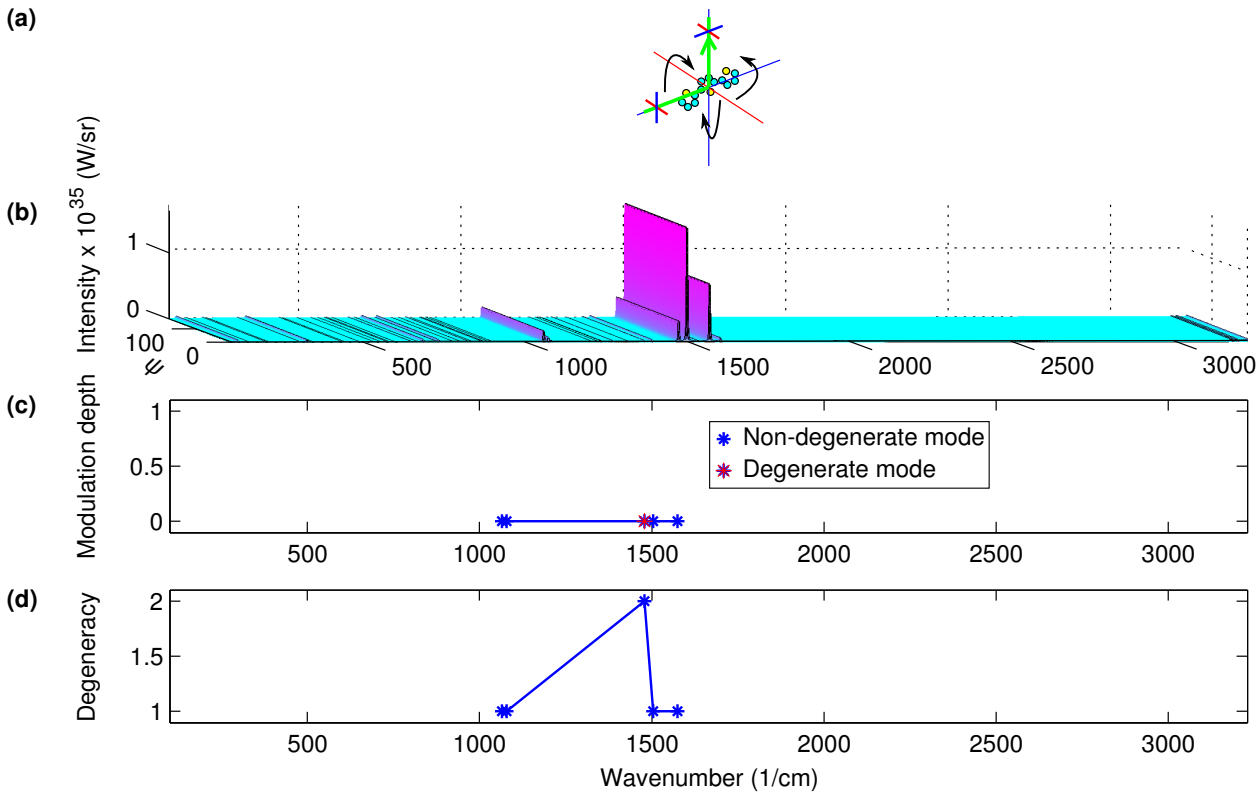

Figure 4.3: Simulated Raman data for isotropically oriented terthiophene molecules under the special experimental conditions. (a) Schematic of the tumbling terthiophene molecule relative to the scattering geometry. (b) Raman spectrum vs. orientation angle ψ ($\in [0°, 180°]$). (c) Modulation depths for the most intense modes in (b). Note that while the experimental conditions are the same as those used in Fig. 4.2 for a crystalline region, now all the modulation depths are exactly 0 — even for the degenerate modes — as is expected for scattering from an isotropic region under these special experimental conditions. (d) Vibrational degeneracies of the most intense modes of the single molecule. Note: The same planar conformation of the terthiophene molecule is used here as in the crystalline case, where the molecule is forced to be planar due to the crystal field. In general the conformation of the molecule in isotropic media is distributed according to the Boltzmann distribution; however, for our purposes we wish to highlight the values of the mode modulation depths, which would remain $M = 0$ even if Boltzmann averaging were applied.

$M'_k\left(\mathscr{S}=\text{crys}; P=1, \psi=0, \chi=0, \theta=\frac{\pi}{2}\right)=1$ and $M'_k\left(\mathscr{S}=\text{iso}; P=1, \psi=0, \chi=0, \theta=\frac{\pi}{2}\right)=0$, independent of the Raman tensor α_k and cell orientation T. Interestingly, while M_k was independent of α_k and T for arbitrary P, we found that M'_k was independent of α_k and T for only $P=1$. Functionally, all formulas for ψ variation hold for ψ' variation if the transformation $\tilde{\alpha}_{xy,k} \rightarrow \tilde{\alpha}_{yz,k}$ is made, and to avoid excess notation, from this point on we will drop any primes on M_k and use ψ to refer to the orientation angle of either the incident or analyzed light. In formulas we will use the Raman tensor component $\tilde{\alpha}_{12,k}$ that refers to rotation of the incident polarization.

To summarize our findings so far, the modulation depth can be defined in two ways: either by rotating the incident polarization and fixing the analyzer, or by fixing the incident polarization and rotating the analyzer. To achieve our special relationship between modulation depth and state of internal orientation $\mathscr{S}$, in the former case, the orientation of the linear analyzer must be fixed parallel to the scattering plane, and in the latter case, the incident light must be linearly polarized parallel to the scattering plane. In both cases, it is required that the scattering angle be 90° and that the incident light be fully $(P=1)$ linearly $(\chi=0)$ polarized (in the former case P can take on any value but only for $P=1$ will $M_k=1$). For a non-degenerate mode, only under either of these two special experimental conditions will the modulation depth unilaterally and unambiguously indicate whether a uniform region is crystalline ($M_k=1$ exactly) or isotropic ($M_k=0$ exactly), fully independent of the Raman scattering tensor α_k and, if crystalline, the unit cell orientation T. Hereafter, when we refer to the "special experimental conditions," we mean either of the two special conditions just described.

It is crucial that for a crystalline region, the intensity profile's phase ϕ_k strongly depends on the unit cell orientation T, as shown in eq. (4.1) and (4.6). This is useful here and in the following section because when multiple intensity profiles (components) contribute to the net intensity profile (resultant) at a particular frequency ω_k, the resultant modulation depth depends on the relative phases and thus on the region orientation. In particular, for our present discussion of non-degenerate modes, two $M=1$ components with similar amplitudes will produce a resultant profile having $M<1$ if the component phases are not equal, which is the case for most orientations of the unit cell.

To illustrate this important concept, consider a polycrystal, which contains differently oriented crystalline domains. For a particular non-degenerate mode k, illumination of only a single domain will yield $M_k=1$ regardless of which domain is being observed, since the modulation depth is independent of cell orientation. However, if multiple differently oriented domains are illuminated simultaneously, the net intensity profile will

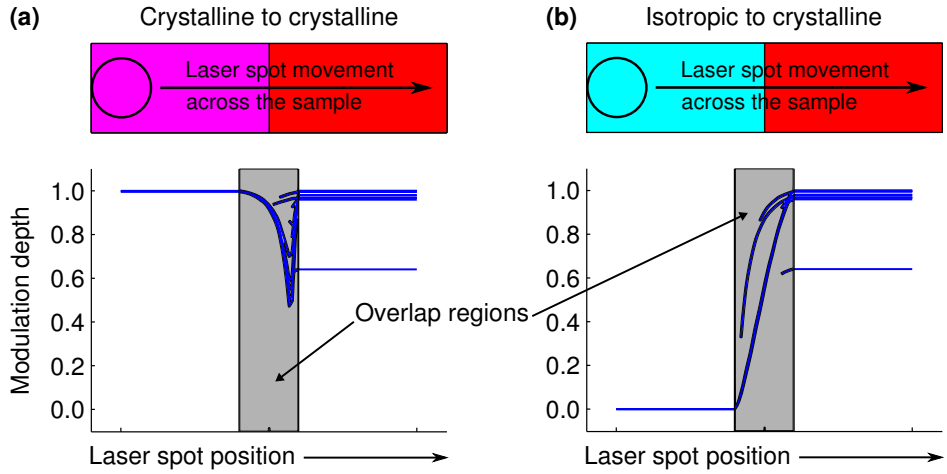

Figure 4.4: Typical simulations of our modulation depth signal upon scanning across a boundary between different regions of a heterogeneous sample. Each blue line corresponds to the signal for one vibrational mode as the sample is scanned, and for each laser spot position, only the most intense, observable modes are plotted. (a) Scanning between two differently oriented crystalline regions. The locations where the laser spot overlaps both regions are depicted in gray. On either region alone, generally $M \approx 1$, but upon illumination of both regions simultaneously, the modulation depths typically dip significantly. (b) Scanning from an isotropic region to a crystalline region. The modulation depths for isotropic regions are always 0 and for crystalline regions they are generally near 1, so that in the overlap region generally a clear transition occurs from $M = 0$ to $M \approx 1$. In both cases, the transition between regions is very clear, particularly when multiple modes are observed; our signal is an excellent detector of a region boundary. Note that all modes are forced to be degenerate due to the use of the oriented gas model; the dips and transitions will generally be even more conspicuous for non-degenerate modes, when $M = 1$ exactly for the crystalline regions.

consist of component, single-domain $M_k = 1$ intensity profiles that do not necessarily add in phase. Thus, in general we will observe $M_k < 1$, so that upon scanning from one domain to the next, a "dip" in modulation depth from 1 will usually clearly be seen, as shown in Fig. 4.4a. While use of the oriented gas model here forces all modes to be degenerate, which as discussed in the next section can lead to $M_k < 1$ on any crystalline region alone, the dips in modulation depth are nevertheless generally significant and clearly observable; our signal clearly indicates that a boundary between domains has been crossed. Further, if multiple domains are being observed, rotation of the sample or the scattering apparatus will generally cause the component intensity profiles to add not-in-phase *differently*, so that rotation will usually force $M_k < 1$ to be clearly observed if it is not already apparent. Thus, certainty that a boundary has been crossed will be increased not only by observing multiple modes but also by observing the sample at multiple orientations. Note that this unique "dipping" feature of our modulation depth signal highlights one of its advantages in boundary detection over using the intensity as a signal, which changes only monotonically upon scanning between regions.

We show in Fig. 4.4b simulations from scanning between an isotropic region and a crystalline region, where observation of $0 < M_k < 1$ in the overlap region is now due not to phase differences but rather simply to addition of the $M = 0$ component of the isotropic region and the $M = 1$ component of the crystalline region. Therefore, scanning from an isotropic region to a crystalline region will produce a smooth "transition" in our modulation depth signal from 0 to 1. Thus, with the general principle of observing $0 < M_k < 1$ for a non-degenerate mode only upon illumination of a non-uniform region, observation of the modulation depth under our special experimental conditions while scanning a sample will generally clearly indicate when a region boundary is being illuminated, allowing straightforward mapping of an entire heterogeneous sample.

The results of this section apply to vibrationally non-degenerate modes, where $M_k = 1$ exactly for illumination of a uniform and crystalline region. It turns out that even though degenerate modes can yield $M_k^{\text{crys}} < 1$ for a uniform region, the conclusions presented here still generally apply. To understand why this is so, we now consider in detail the implications of degenerate modes and the possibility of obtaining $M_k < 1$ on a uniform, crystalline region.

Vibrationally degenerate modes

A vibrationally degenerate mode k means that the scatterer has multiple types of vibrations of frequency ω_k. For illumination of a uniform and crystalline region, each component vibration — which by definition is non-degenerate — contributes an intensity profile at ω_k having $M = 1$ under the special experimental conditions, as in the previous section. However, in general the phases of the component profiles are different (eq. (4.6)), so that the net intensity profile $I_k(\psi)$ at ω_k will generally have $0 \leq M_k^{\text{crys}} \leq 1$. (Note that this same principle of not-in-phase addition of multiple intensity components was seen in the previous section for non-degenerate modes, when multiple crystalline domains were simultaneously illuminated.) On the other hand, the net, degenerate modulation depth for illumination of an isotropic region will remain $M_k^{\text{iso}} = 0$, since the addition of multiple $M = 0$ profiles, which are constant with respect to ψ, will comprise a net profile that is also constant with respect to ψ.

Explicitly, under the special experimental conditions, the net intensity profiles for a degenerate mode k are

$$
\begin{aligned}
I_k^{\mathsf{crys}}(\psi) &= \frac{(\omega_1-\omega_k)^4 \mathcal{I}\hbar N}{64\pi^2\epsilon_0^2 c^4 \omega_k \left(1-\exp\{-\hbar\omega_k/k_B T\}\right)} \sum_{i=1}^{D_k} \left(\tilde{\alpha}_{xy,k,i}^2 + \tilde{\alpha}_{xz,k,i}^2\right) \\
&\quad \times \left[1+\cos\left(2\psi + \tan^{-1}\left(\tilde{\alpha}_{xz,k,i}^2 - \tilde{\alpha}_{xy,k,i}^2, -2\tilde{\alpha}_{xy,k,i}\tilde{\alpha}_{xz,k,i}\right)\right)\right] \\
I_k^{\mathsf{iso}}(\psi) &= \frac{(\omega_1-\omega_k)^4 \mathcal{I}\hbar N}{64\pi^2\epsilon_0^2 c^4 \omega_k \left(1-\exp\{-\hbar\omega_k/k_B T\}\right)} \left(\frac{6}{45}\right) \sum_{i=1}^{D_k} \gamma_{k,i}^2,
\end{aligned}
$$

where D_k is the degeneracy of the frequency ω_k and $\tilde{\alpha}_{mn,k,i}$ is the mn component of the i^{th} of D_k Raman tensors with respect to normal coordinates of frequency ω_k. The corresponding modulation depths at ω_k are, from eq. (4.2),

$$
\begin{aligned}
M_k^{\mathsf{crys}} &= \frac{2\sqrt{\left[\sum_{i=1}^{D_k}\left(\tilde{\alpha}_{xz,k,i}^2 - \tilde{\alpha}_{xy,k,i}^2\right)\right]^2 + \left[2\sum_{i=1}^{D_k}\tilde{\alpha}_{xy,k,i}\tilde{\alpha}_{xz,k,i}\right]^2}}{\sum_{i=1}^{D_k}\left(\tilde{\alpha}_{xy,k,i}^2 + \tilde{\alpha}_{xz,k,i}^2\right) + \sqrt{\left[\sum_{i=1}^{D_k}\left(\tilde{\alpha}_{xz,k,i}^2 - \tilde{\alpha}_{xy,k,i}^2\right)\right]^2 + \left[2\sum_{i=1}^{D_k}\tilde{\alpha}_{xy,k,i}\tilde{\alpha}_{xz,k,i}\right]^2}} \qquad (4.7) \\
M_k^{\mathsf{iso}} &= 0.
\end{aligned}
$$

We see that indeed while $M_k^{\mathsf{iso}} = 0$ regardless of vibrational degeneracy D_k, in general $0 \leq M_k^{\mathsf{crys}} \leq 1$. Note that for $D_k = 1$, M_k^{crys} reduces to 1 exactly, as it should in the non-degenerate case.

We see that for a degenerate mode, while $M_k^{\mathsf{crys}} > 0$ indicates that a uniform region must be crystalline, there is an ambiguity when $M_k \approx 0$ as to whether the region is crystalline or isotropic. Let us define as our "false positive" the case in which we determine the illuminated region to be isotropic when it is actually crystalline. First, it turns out that for any single mode, the false positive of obtaining $M_k = 0$ is rare: this can only occur for a degenerate mode, and this mode must have component vibrations with very particular combinations of phases and amplitudes. For example, if there are two components to a degenerate mode, not only must they be perfectly out of phase, but they must also have equal amplitudes. Second, if we employ the general strategy of observing as many modes as possible, the false positive becomes even more rare, as it would require *all* modes to be degenerate and *all* modes to have particular component vibrations such that their resultant intensities yield $M = 0$. Note that because the only way our signal can yield $M_k > 0$ for any mode in a uniform region is for that region to be crystalline, all that is needed to determine that a region is crystalline is for *any* mode to clearly show $M_k > 0$.

To ensure that our signal would always correctly and unambiguously determine whether a uniform region was crystalline or isotropic, we studied in detail the prevalence and implications of this lone false positive case (obtaining $M = 0$ for all observable modes in a uniform and crystalline region) by simulating our modulation depth

signal for 20 molecules at 176 unique orientations. We found that it is indeed extremely rare to obtain $M \approx 0$ for all observable modes for a crystalline region; the false positive occurs only for highly symmetric molecules in highly symmetric orientations with respect to the scattering apparatus. Next, we studied how stable the false positive was with respect to slight rotation of the sample or apparatus, because as previously discussed, the component phases and amplitudes are very sensitive to unit cell orientation (see eq. (4.1), (4.5), and (4.6)). We figured that upon rotation it would be rare — if not impossible — for the components of all the modes to continue to produce resultant intensity profiles having $M = 0$.

We indeed found that when the false positive did occur, rotation by as little as 1° in any direction usually showed $M > 0$ for some modes, resolving the ambiguity and showing that the region was indeed crystalline. Rotation by 5° always revealed this, and it can thus safely be assumed that if rotation by just 5° in any direction does not change all modes from having $M \approx 0$, then the illuminated region must be isotropic. In general, however, observing $M \approx 0$ for all modes is so rare for a crystalline system that if $M \approx 0$ for all modes is observed, it can safely be assumed with extremely high certainty that the region is isotropic, without having to rotate the sample or scattering apparatus.

A degenerate mode in a crystalline region not only usually has a modulation depth that is clearly greater than 0, but often it is closer to 1, so that in general, for an isotropic region we will find $M_k = 0$ and for a crystalline region, $M_k \approx 1$. In fact, the simulations presented in the previous section on non-degenerate modes (Fig. 4.2 to 4.4) actually exclusively involved modes that, in the crystalline case, were forced to be degenerate due to use of the oriented gas model. So instead of obtaining $M_k^{\text{crys}} = 1$ exactly, we generally obtained $M_k^{\text{crys}} \approx 1$. Thus, the general conclusions presented in the previous section actually apply to all modes, both non-degenerate and degenerate. In particular, upon comparing Panel (c), which contains our modulation depth signal, of Fig. 4.2 and 4.3, there is still no question which figure corresponds to which state of internal orientation, and in Fig. 4.4, there is still no question that the dips (Panel (a)) and transitions (Panel (b)) clearly indicate the boundaries between the distinct regions. Taken together, our signal is an extremely clear and robust detector of region boundaries and indicator of region state, particularly when we employ the general strategies of observing multiple modes and observing the sample at multiple orientations. As none of our conclusions have applied to particular types of materials, our method can be applied unilaterally to any sample of any composition. Further, no information about the sample is needed *a priori*, including the Raman tensor, the unit cell orientation, and even the vibrational degeneracies of the modes.

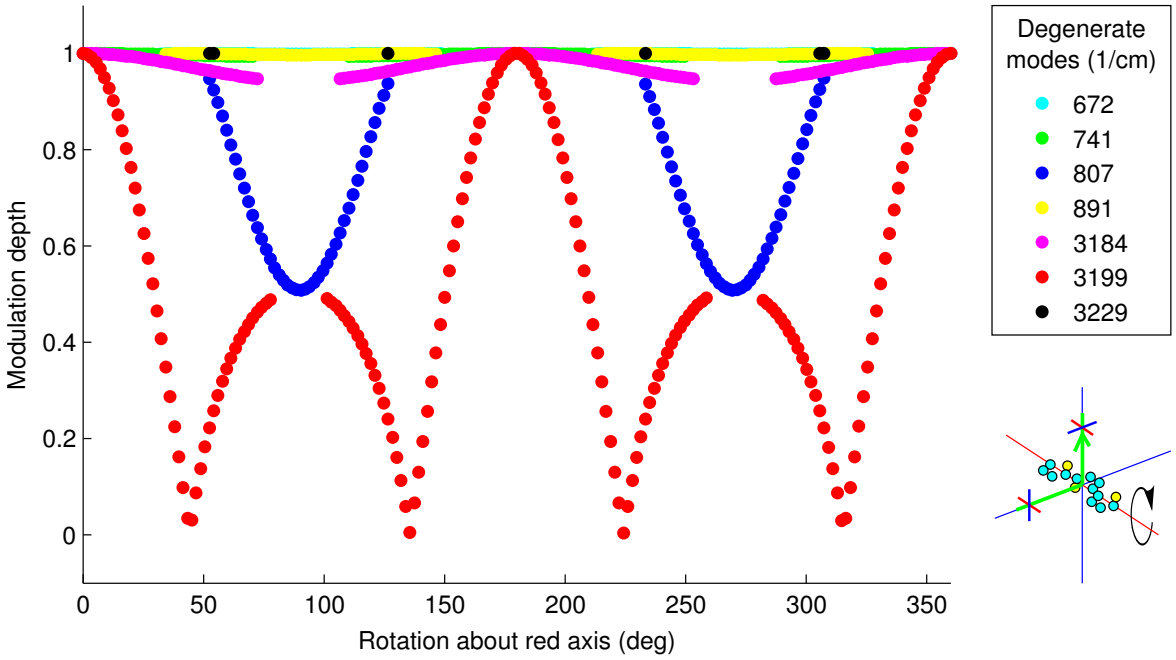

Figure 4.5: Simulated modulation depths of the degenerate modes of the planar terthiophene molecule as it is rotated about its long axis under the special experimental conditions. While for some orientations the modulation depth M of some degenerate modes appears to be approximately 1 — possible for both non-degenerate and degenerate vibrational modes — by rotating the sample or apparatus we can sometimes force M to be noticeably less than 1, indicating that the mode is vibrationally degenerate. Note that for each molecule orientation, only the most intense, observable modes are plotted.

Finally, a particularly interesting application of our modulation depth signal is determination of when a mode is vibrationally degenerate: as seen from eq. (4.7), the only way to obtain $M_k < 1$ for a uniform, crystalline region is for mode k to be degenerate ($D_k > 1$). Note, however, that we could certainly obtain $M_k = 1$ for a degenerate mode for components with particular combinations of the phases and amplitudes. This could occur, for example, when all the components of a degenerate mode are in phase, or when the components are not in phase but the amplitude of one component dominates over the others. Thus, while observation of $M_k < 1$ in a uniform, crystalline region will definitively indicate that mode k is degenerate, observation of $M_k = 1$ does not necessarily mean that the mode is *not* degenerate.

However, here again we can exploit the strong and arbitrary phase/amplitude dependence of the component intensity profiles on unit cell orientation (eq. (4.1), (4.5), and (4.6)): rotation of the sample or apparatus could cause the phases and amplitudes of the component $M = 1$ profiles to add differently so that the net intensity profile clearly yields $M_k < 1$ if it is not already apparent. Of course, this will not always work if, for example, the

amplitude of one component continues to dominate over the amplitudes of the other components, even as the phases change arbitrarily as the unit cell is rotated. These phenomena are demonstrated in Fig. 4.5, which plots the modulation depths of all the degenerate modes of a single terthiophene molecule as it is rotated about its long axis. We see that when the molecule is not rotated at all (0° rotation), all degenerate modes express $M = 1$, yielding no definitive conclusion using our signal as to whether any of the modes are degenerate. However, as the molecule is rotated, we see that the modes with wavenumbers 807, 3184, and 3199 must have components with amplitudes that are the same order of magnitude so that as the component phases arbitrarily change, so do the net modulation depths, leading us to conclude that these modes are indeed degenerate. (More generally, the modulation depth in a uniform, crystalline region can change upon sample rotation only for a degenerate mode, just as the modulation depth of a non-degenerate mode can change upon sample rotation only if a domain boundary is being illuminated.) On the other hand, we see that the modulation depths of the other degenerate modes likely have dominant components at all orientations of the molecule so that even as their component phases arbitrarily change, the net modulation depths remain near 1, rendering these degenerate modes indistinguishable from non-degenerate modes using our signal.

4.3 Discussion

When C.V. Raman discovered the inelastic scattering effect that later became his namesake, he was particularly intrigued by the dependence of the analyzed intensity on the polarization of the light [6]. We feel that we have truly capitalized on this special polarization dependence as he might have hoped: by appropriately modifying the polarizations (and scattering angle) of the incident/scattered light — which are fully controllable, experimental parameters — we have been able to determine a fundamental and important property of the sample itself: whether the scatterers are arranged periodically or isotropically. This effect can be roughly compared to "magic angle spinning" in NMR, in which orientational dependence of the signal falls out under a particular, controllable experimental condition [100–102].

Let us summarize the results we have derived above. We first introduced the modulation depth of a vibrational mode, and then showed how, under a special set of experimental conditions, for a non-degenerate mode the modulation depth M serves as a signal of the state of internal orientation $\mathscr{S}$ of molecules within a uniform region, i.e., whether the region consists of periodically or isotropically oriented molecules. We then showed what this signal would yield from a non-uniform region — either a border between two differently oriented crystalline

Table 4.1: Comparison of methods for determining state $\mathscr{S}$, directly from established formulas for the degree of crystallinity of crystalline polymers. The subscripts c (crystalline) and a (amorphous) refer to reference values for purely crystalline and purely amorphous states, respectively.

Method	**Crystalline if...**	**Amorphous if...**	**Ambiguous if...**
Direct measurement methods:			
Specific enthalpy of fusion Δh (calorimetry)	$\Delta h = \Delta h_c$	$\Delta h = 0$	$\Delta h_c \approx 0$
Specific volume ϕ (or density ρ)	$\phi = \phi_c$ (or $\rho = \rho_c$)	$\phi = \phi_a$ (or $\rho = \rho_a$)	$\phi_c \approx \phi_a$ (or $\rho_c \approx \rho_a$)
Specific heat c	$c = c_c$	$c = c_a$	$c_c \approx c_a$
Specific enthalpy h	$h = h_c$	$h = h_a$	$h_c \approx h_a$
Spectral methods:			
X-ray diffraction	No amorphous contribution to spectrum	No crystalline contribution to spectrum	Crystalline/amorphous contributions are similar
NMR			
Infrared absorption			
Previous Raman methods			
Our method:			
Non-degenerate Raman modulation depth M under special experimental conditions	$M = 1$	$M = 0$	NA

domains or a border between an isotropic region and a crystalline region. This allowed us to characterize the uniformity of a sample, i.e., the locations of the boundaries between the homogeneous regions of a heterogeneous sample. Then, we extended our signal to include vibrationally degenerate modes and described what the signal would then tell us about the sample's states and uniformity, showing that the same conclusions generally hold for degenerate modes as for non-degenerate modes. Finally, we showed that in addition to information about the overall molecular structure of a sample, our signal yields information about an intrinsic property of the scatterers themselves: the degeneracies of their vibrational modes.

Table 4.1 compares previous methods for determining the state $\mathscr{S}$ of a uniform region and highlights the advantages of ours. The formulas in this table are taken directly from established methods for determining the degree of crystallinity of crystalline polymers [87, 88], but their general features hold for all types of materials. All previous methods — direct measurements of specific enthalpy of fusion (i.e., calorimetry), volume, density, heat, and enthalpy, and the spectral methods based on X-ray diffraction, NMR, infrared absorption, and Raman spectroscopy — are heuristic methods that require *a priori* knowledge of reference values for purely crystalline and purely isotropic regions (middle two columns of Table 4.1). For example, determination of state $\mathscr{S}$ using calorimetry requires knowing the specific enthalpy of fusion for a purely crystalline sample, Δh_c. Likewise, determination of $\mathscr{S}$ using measurement of the specific volume ϕ requires *a priori* knowledge of the specific volumes

of the purely crystalline and purely amorphous states, ϕ_c and ϕ_a. And any of the spectral methods — Raman included [78, 90, 92–95] — requires knowledge of the exact shapes and sizes of the $\mathscr{S}$-sensitive peaks for the purely crystalline and purely amorphous states. In contrast, our method is first-principles and does not require knowledge of any such reference information beforehand; our modulation depth signal M need only be compared to the numbers 0 and 1.

In addition, due to this reliance on reference values for purely crystalline and purely isotropic regions, all previous methods are never necessarily deterministic of the state $\mathscr{S}$; the heuristic for $\mathscr{S}$ can be ambiguous for a given material or mode (last column of Table 4.1). For example, if Δh_c for a particular material is very small, calorimetry cannot be used to determine $\mathscr{S}$ since Δh would then no longer differentiate between the crystalline ($\Delta h = \Delta h_c$) and amorphous ($\Delta h = 0$) states $\mathscr{S}$. Likewise, if for a particular material ϕ_c and ϕ_a were roughly equal, measurement of the specific volume ϕ of the sample would not allow us to determine $\mathscr{S}$. And for the spectral methods for determining $\mathscr{S}$, not just any peak will do: the peak shape and size must sufficiently distinguish a purely crystalline region of the sample from a purely isotropic one in the first place. On the other hand, our method, due to its first-principles derivation, is *always* deterministic for *any* observable mode of *any* material: the "reference values" of 0 and 1 are obviously never arbitrarily close to each other. In addition, because our method provides an independent signal from each observable vibrational mode — whereas prior Raman methods tend to utilize the spectral peak properties of only one or two modes to determine $\mathscr{S}$ — the certainty of our assessment of a given region's state of internal orientation is greatly compounded; uncertainty is minimized.

Further advantages of our method include its wide applicability to various kinds of materials, such as thin films (for which X-ray diffraction is generally not used to determine the state of internal orientation $\mathscr{S}$), molecular crystals, and semiconductors. In particular, for obvious reasons this method may be very useful when applied to materials exhibiting amorphous-to-crystalline transitions, such as phase-change memory. For 3D materials, confocal Raman spectroscopy could be utilized in order to map out the full solid in 3D [72]. Further, the required setup and apparatus for our method should be straightforward to implement, and the data collection and post-processing could be fully automated and digitized. Overall, our method allows unambiguous delineation of the uniform regions of the sample, after which other methods [98, 99], such as the mode modulation depths under different experimental conditions [43], could be employed to determine the orientations of the crystalline regions, culminating in a complete characterization of the molecular structure of the sample.

We note that just as experimental observations often do not precisely reflect the theory on which they are

based due to noise or complicating effects, we would not expect experiment to yield $M = 0$ exactly for an isotropic region or $M = 1$ exactly for a crystalline region. What we have discovered here is a basic principle that can be built upon and that we do expect well set-up experimental observations to primarily reflect. Further, while the principles presented here do hold rigorously, due to the linewidth-dependence of the amplitudes, phases, and modulation depths, we recommend, when needed, to focus on non-degenerate modes that are sufficiently isolated in wavenumber from other modes in order to observe the desired effect more clearly. Finally, we note that other experimental conditions may not provide a zero-or-one signal but may still provide a useful signal while being less sensitive to mode degeneracies or bleeding from nearby modes, thus being a useful supplement to the conditions described here. Simulated signals under different experimental conditions can be explored using the open source software mentioned in the following chapter.

A current limitation of our method is that it will not apply to uniform regions that are smaller than the illuminated area or volume. For example, for a crystalline polymer, which by definition is composed of amorphous and crystalline regions, while the amorphous regions can be arbitrarily large, the size of a crystalline domain typically tops out at around 50 nm [74]. In contrast, Raman spot sizes are typically on the order of hundreds of nanometers, even for high numerical apertures, meaning that in such materials the laser spot will rarely illuminate only one crystalline domain. However, as interest in (and the ability to produce) larger crystalline domains continues to grow, we expect this method to prove useful for polymeric materials. For example, polymer single crystals tend to have sizes on the order of microns [74], more than large enough for analysis with our method.

Another consideration to keep in mind is that our method can only determine *whether* a vibrational mode is degenerate, as opposed to *how* degenerate the mode is. In other words, our method can determine either that a single type of vibration occurs at a particular frequency or that multiple types of vibrations occur at a particular frequency, but it cannot say, for example, that two types of vibrations as opposed to three types of vibrations occur at the same frequency. This is due to the fact that while a non-degenerate mode activated in a crystalline region must have $M = 1$, the observation of $M < 1$ could be caused by the not-in-phase addition of any number of modes. Further, there is also a chance that even upon study of a crystalline system at multiple orientations, a vibrationally degenerate mode may appear to have $M \approx 1$ at all these orientations, in which case our signal is inconclusive as to whether the mode is degenerate.

In addition, while our method will determine conclusively whether a uniform region is crystalline or isotropic, it has not yet been investigated for determining intermediate degrees of crystallinity. Other Raman methods have

attempted to do so [78, 90, 92–95], but even for the *sub*task of determining the state of internal orientation $\mathscr{S}$ of a uniform region, these other methods rely on a heuristic procedure, since there is no way of knowing, for example, precisely what the integrated area of a Raman peak should theoretically be for a given crystalline region [2]. In fact, there is no way to tell that the region is 100% uniform in the first place, let alone 100% crystalline. On the other hand, due to the extreme sensitivity of our method to even slight changes in the Raman tensor and cell orientation, all that is required to determine that a region is uniform is to scan the region and ensure that the multitude of modulation depth signals remain constant. And after this is done, for a non-degenerate mode, our signal must unambiguously identify the state $\mathscr{S}$ of the region by signaling either 0 or 1 — there can be no values in between. In any case, the crystallinity of intermediate states is not physically well-defined; indeed, on the molecular level it can be difficult to define whether particular molecules in a heterogeneous region are periodically or isotropically oriented [103], which can be corroborated by the fact that often different methods of measuring intermediate crystallinity yield different values [74]. After all, we could not expect to be able to precisely characterize heterogeneous regions, a generally difficult task [104], without having the ability to fully characterize homogeneous regions in the first place. We therefore conclude that our approach of studying such uniform, well-defined states of internal orientation is a logical step toward rigorously understanding and characterizing more complex sample morphologies.

Finally, we want to emphasize that we have only theoretically demonstrated our method in this work. Although we experimentally confirmed our modulation depth formulas for both the crystalline and isotropic cases in our prior work (under different experimental conditions) [43], the relationship between the modulation depth M and the state of internal orientation $\mathscr{S}$ under the proposed experimental conditions remains to be verified experimentally. Such verification would simply involve an apparatus similar to the one shown in Fig. 1.2a, using a secondary technique — such as one of those listed in Table 4.1 — to confirm sample crystallinity. Thus, we hope we have sufficiently inspired the reader to put this method to the test, after which we hope it will become a primary method in the community for determination of region state, sample uniformity, and mode degeneracy.

Chapter 5

Open-source software: Polarized Raman Explorer

Regarding our method developed in the previous chapter for using the modulation depth M to determine whether a uniform region of a material is crystalline ($M = 1$) or isotropic ($M = 0$), as mentioned in Section 4.3, vibrational modes that are truly degenerate even in the presence of a surrounding crystal field, including the potential effects of other molecules in a single unit cell, can pose a problem: the modulation depth in the crystalline state can potentially not be equal to one. In addition, contamination from nearby vibrational modes due to finite linewidths can add components to the total intensity of interest and cause the modulation depth to be less than one. There are practical complications that must be overcome in any experiment exploiting any physical phenomenon, and these are examples of such complications for the current principle of interest.

If experimentally investigating this phenomenon for a desired system of interest is challenging, one alternative may be to use simulations, which certainly could be a challenge in and of itself. At best, the researcher would already have a general understanding of polarized Raman spectroscopy and an output file from an electronic structure code containing the Raman tensor. From this he or she would need to know the precise orientation of the unit cell (in the crystalline case) and/or plug the Raman tensor components into eq. 1.28 (in the isotropic case) in order to obtain the quantities $\tilde{\alpha}^2_{xy,k}$ and $\tilde{\alpha}^2_{xz,k}$ and/or γ^2_k, respectively, for each vibrational mode k. She would then need to plug these quantities into eqs. 4.3 and/or 4.4 in order to determine the intensities as a function of the orientation angle ψ, fit each intensity to a Lorentzian curve with the appropriate linewidth of the experimental setup, and add the curves as functions of ψ for each mode in order to determine the total overall intensity as a function of both the wavenumber shift and ψ, examples of which were shown in Panel (b) of Figs. 4.1 to 4.3. This process is required to make sure that possible degeneracies and nearby mode contamination do not result in an

overall expected intensity for a particular mode that deviates too much from one in the crystalline case. (Again, it is not possible, even practically, for the overall expected isotropic modulation depth to be anything other than zero; only the crystalline modulation depth needs to be checked.) Such a process is also required to determine the expected amplitudes, in order to determine if the modes of interest are even observable in the first place. If the researcher wants to explore other experimental conditions, she would need to do the same thing except she would need to first work through the appropriate equations in Chapter 1 in order to put together the desired intensity formulas in the first place, as we have not put together intensity formulas for any experimental conditions other than the special ones described in the previous chapter.

Even if the researcher simply wanted a single set of spectra, as opposed to those as a function of ψ, as is generally required for the calculation of the modulation depth, she would need to put together the appropriate equations of Chapter 1 and follow the steps above — the only simplification would be that she would not need to carry through the ψ-dependence. If she were only interested experimental conditions pertaining to a generic Raman microscope for a purely isotropic sample, her job would be relatively easy, requiring her to simply multiply the common intensity formula of eq. 1.33 by the wavenumber-dependent coefficient, though she would still need to fit all modes with the appropriately-wide Lorentzian curves and add them together in order to obtain results she would actually expect to obtain from her experiment. If she were interested in any other experimental conditions for an isotropic sample or *any* conditions — even the Raman microscope conditions — for a crystalline sample, she would first need to put together the general formulas we derived in Chapter 1. Even for a researcher who understood the meaning of these formulas, this task would be annoying and time-consuming at best.

During our studies of polarized Raman spectra we have put together a GUI in MATLAB automating all these tasks, except of course the first-principles electronic structure simulations themselves that produce the Raman tensor in the first place. While this was essentially a requirement for us (all spectra in this dissertation were plotted with some iteration of this software), we think it would be invaluable to anybody who has a remote interest in obtaining Raman spectra that can be compared with experiment, whether an early-year graduate student or a professor or a scientist in industry. There is no better way to appreciate the richness of Raman spectroscopy than to easily and quickly plot the spectra for various samples under various experimental conditions, and further, as far as we have found, there is no real alternative anyway.

We thus present our software here, which we plan to shortly publish as part of a paper, called Polarized Raman Explorer. As there is a constantly-updated, free MATLAB runtime environment available from MATLAB's creator

MathWorks Inc. for any operating system, anyone will be able to use this software free of charge. In addition, we plan to upload the source code to Github so that anybody can see precisely how the formulas work behind the scenes; there is no better way to fully understand a formula than to see how it is actually implemented. Of course, we will also upload this software to MATLAB's own community projects workspace as well so those who already have access to MATLAB can be up and running almost immediately.

Currently, the software requires only a Jaguar output file that has calculated the gradients required for subsequent calculation of the Raman tensor (see Chapter 2). It then extracts the tensor and Hessian eigenvectors and eigenfrequencies using awk and bash (which are included in every Linux distribution) and plots the Raman intensity, activity, modulation depth, phase, amplitude, or 3D intensity-vs.-ψ surface plot for any experimental condition. The user can optionally input a Jaguar output file that has calculated the vibrational frequencies more accurately, as in a calculation employing the Pulay SQM method[45], or even an additional Jaguar output file that can be used for calculating the diagonal and off-diagonal components of the Raman tensor separately in order to obtain more accurate results (see Chapter 2 for why this may be important). The software then employs an eigenvector-matching algorithm in the backend that plots the final Raman intensities at the correct frequencies. Experimental parameters that can be varied include the state of aggregation (crystalline or isotropic), molecule orientation (if crystalline), incident and/or detected polarizations, scattering angle, direction of frequency shift (Stokes vs. anti-Stokes), linewidth, and the parameters affecting the coefficient in the intensity formulas, including the number of scatterers, incident irradiance, excitation wavelength, and temperature. Additional features include plotting of the Raman apparatus relative to the molecule orientation, rotation and other interactivity of either plot, export of the resulting plots at publication quality at any resolution and print size, overplots of which modes may actually be detectable, and, shortly, export of the calculated data in a text file.

Finally, the software will soon not be limited to Jaguar output files; we have already written (just not implemented) a "translator" for Q-Chem output files. Further, as the source code will be freely available, anybody can easily write a script that parses the output file from any electronic structure program; such translators are small and easy to make (the Jaguar translator is an 84-line bash script). We expect that any unfinished parts described above and a manuscript can be completed in just a couple weeks. A screenshot of the GUI is shown in Fig. 5.1.

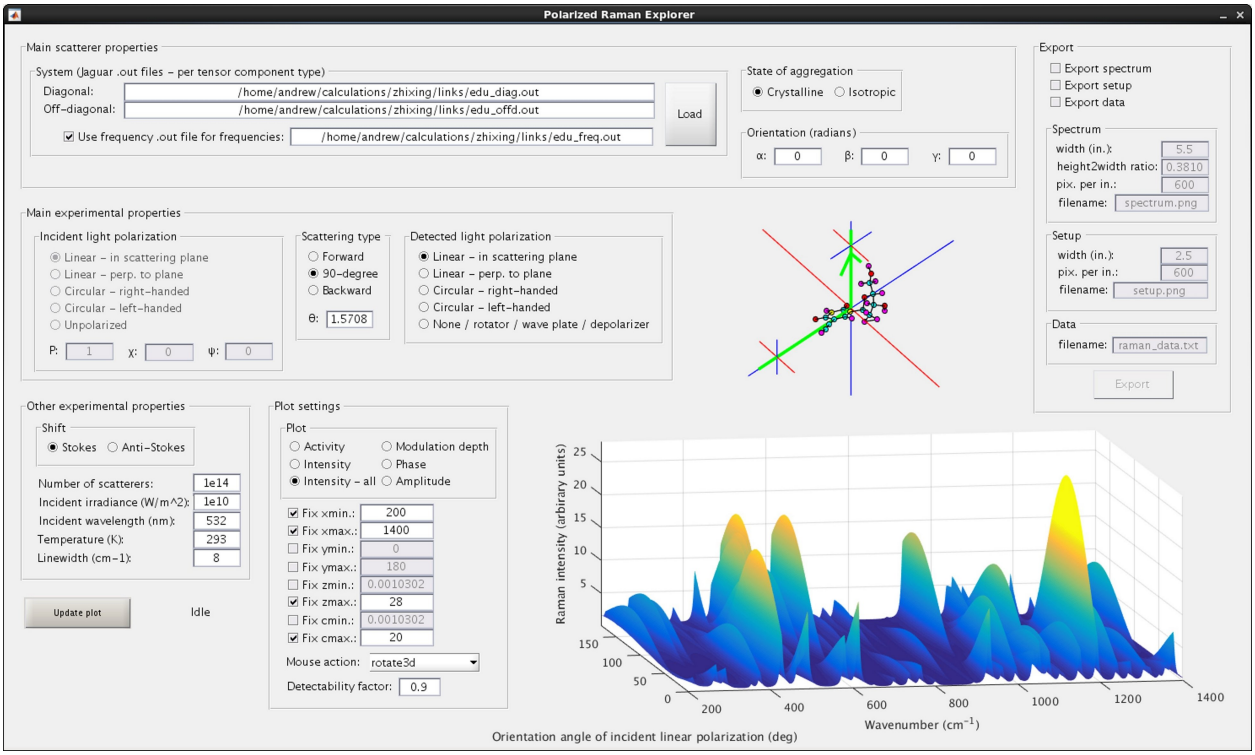

Figure 5.1: Screenshot of the Polarized Raman Explorer software analyzing the EdU molecule.

Part II

Applications of the Pseudospectral Method

In the second part of this dissertation, application of the pseudospectral method is extended to additional areas of scientific importance requiring the understanding of molecular structure and function. First, pseudospectral calculations of vibrational frequencies of solvent-based biomolecules containing alkyne vibrational tags, which allow biomolecule activity to be tracked in real-time using stimulated Raman spectroscopy, are presented, demonstrating the excellent predictive capabilities of the frequency simulation methodology for such experiments (Chapter 6). Next, benchmarking data for the calculation of ground- and excited-state energies and gradients using the pseudospectral method are presented, showing single-processor superiority over the traditional analytic method in all areas except for ground-state energies on molecules using more than 4,800 basis functions (Chapter 7). This demonstrates the advantage of applying the pseudospectral method for calculating the energies and optimized geometries of large systems in either the ground or excited electronic states.

Lastly, a simulation of the first of four proton-coupled electron transfers (PCETs) of the oxidation evolution reaction (OER) at the anode of a water splitting solar cell is introduced, modeling the anode as a 147-atom, explicitly passivated anatase titanium dioxide nanoparticle (Chapter 8). This is done to understand why a water splitting solar cell requires so much more input energy than is ultimately stored in the bonds of the hydrogen and oxygen gases produced, an inefficiency referred to as the "overpotential." A version of Jaguar, electronic structure software based on the pseudospectral method, is optimized for the Stampede supercomputer and is used to realistically simulate the PCET at the nanoparticle catalyst using hybrid DFT (B3LYP) with a Poisson-Boltzmann implicit solvent model, finding a thermodynamic overpotential of 1.15–1.17 V, plus a kinetic contribution of 0.27–0.49 V, resulting in a total anodic overpotential of 1.44–1.64 V. While the kinetic contribution is consistent with other studies, the thermodynamic contribution is too high by about 0.4 V. However, studies are underway on larger, even more realistic nanoparticles, and it is expected that the thermodynamic overpotential will decrease toward the expected value of 0.7–0.8 V. Even now, these results support findings in literature that the cell inefficiency due to a large overpotential, a significant part of which is due to the kinetic activation barrier, is caused by the unstable species present on the surface of the catalyst upon occurrence of the first PCET of the OER. These results further validate the particularly fast and accurate methodologies used, opening the door to efficient and realistic cluster-based, fully quantum-mechanical simulations of the bottleneck process of a promising technology for clean solar energy conversion.

Chapter 6

Live detection of biomolecule activity using stimulated Raman spectroscopy

Use of fluorescence microscopy has long been among the primary ways to detect the location and function of cells in living biological systems[105–112]. However, such techniques for tagging small biomolecules such as nucleic, amino, and fatty acids are known to have significant drawbacks due to the comparatively large size of the dye molecules serving as the tags, which has led to the search for better ways to probe the workings of cells *in vivo*. Professor Wei Min's and Professor Colin Nuckolls' experimental labs at Columbia University have investigated such a method in which two-atom alkyne groups are tagged to biomolecules of interest. The vibrational frequency of the alkyne group's C–C triple bond stretch is special because it lies in the "silent zone" of the vibrational spectrum between about 1700 and 3000 cm^{-1}, in which no vibrations are otherwise present in systems of biological interest. Thus, detection of the alkyne triple bond stretch is completely unobstructed, making its observation using vibrational Raman spectroscopy a useful way for detecting the biomolecules to which they are attached. Further, stimulated Raman spectroscopy (SRS) provides significant enhancement of the generally modest peak found in the spontaneous spectrum (about 4500x in the following experiments), allowing the Raman signal to be detected much more easily. The Min and Nuckolls groups have successfully synthesized molecules containing such alkyne tags and detected them in biological systems using their SRS setup.

By using carbon isotopes in place of the usual carbon atoms that make up the triple bond C–C stretch, the Min and Nuckolls groups further aimed to produce three additional vibrations at frequencies red-shifted to that of the unmodified alkyne groups, thereby producing a four-color vibrational "palette" in which different isotopes

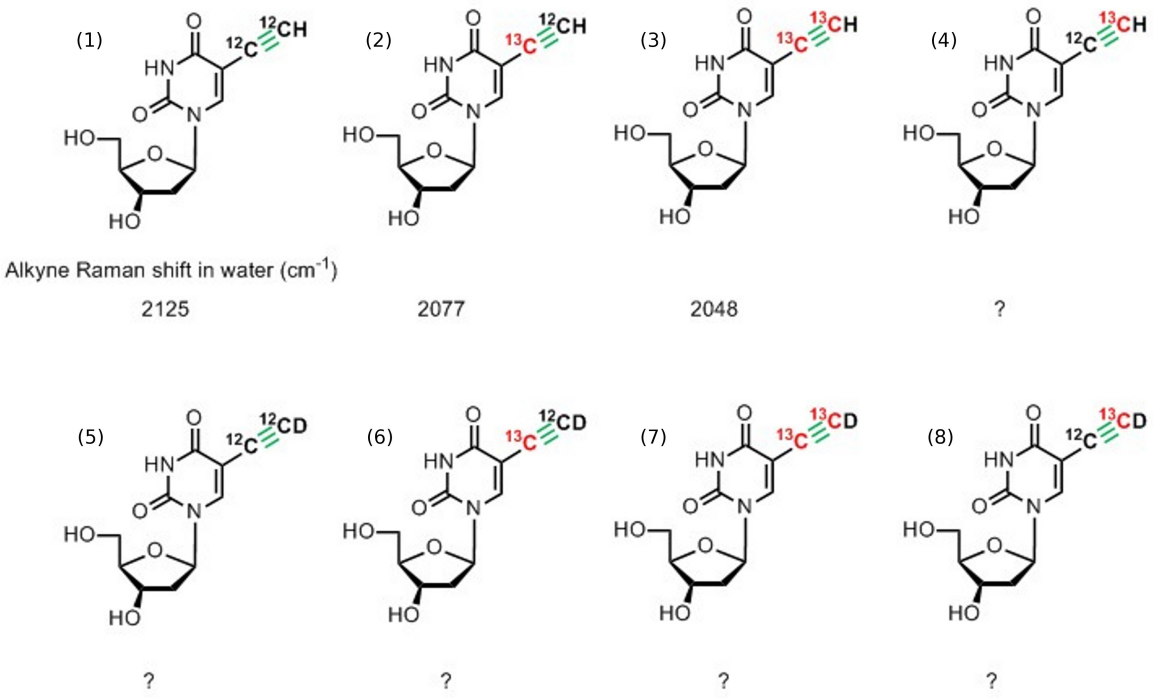

Figure 6.1: "Isotopologue" of various alkyne group isotopes attached to EdU. The latter four molecules contain deuterium atoms in place of the terminal hydrogen atoms. Image courtesy of Dr. Zhixing Chen.

of alkyne groups could be attached to various small biomolecules so that multiple biological phenomena could be observed simultaneously. They further considered how the frequencies might be modified by substituting deuterium for the hydrogen atom terminating the alkyne group, resulting in an eight-molecule "isotopologue" (Fig. 6.1) of alkyne tags for consideration, choosing 5-ethylnyldeoxyuridine ("EdU"; a cell proliferation reporter) as a model compound. However, Dr. Zhixing Chen, a former graduate student in the Min Group and current postdoc at Stanford, found that the synthesis of the different types of alkyne tags on the EdU molecules was sometimes difficult or impossible, so we in the Friesner Group provided simulations and theoretical explanations of the various types of tags. In particular, our contribution was to calculate accurate vibrational frequencies of the eight alkyne tags attached to EdU molecules so that Dr. Chen could ensure their frequencies were sufficiently unique to justify attempts at synthesizing the tags in the first place. These results are discussed below.

Starting from the ChemDraw file of the eight molecules in the isotopologue that Dr. Chen sent us, we converted the coordinates to 3D using OpenBabel, fixed any errors in the conversion manually using Maestro, performed force-field optimizations using Maestro (using the default OPLS 2005 force field), and then began optimizations in

Jaguar using DFT in water using the Poisson-Boltzmann solver. It should be noted that changing the isotopes in Maestro did not have any effect on the force field optimizations; only one of the eight molecules in the isotopologue was needed to obtain molecular mechanics-optimized geometries. As Peter Pulay et. al.'s Hessian-element scaling method[45], which scales individual elements of the Hessian according to the type of vibration and corrects for systematic deficiencies in the wavenumbers calculated at the B3LYP/6-31G* level of theory (implemented in Jaguar using the `isqm=1` setting in the `&gen` section), this is the level of theory we used for all calculations. In addition, for the frequency calculations, we used ultrafine pseudospectral grids (`gdftder2=-13; gdftcphf=-13`), without which the Hessian calculations did not converge. `&atomic` sections were used in the Jaguar input files to identify the isotopes of the relevant atoms for the seven non-native EdU molecules. We note that for the geometry optimizations often Jaguar would not stop upon finding a higher-energy geometry, as we found happens often in solution (typical EdU optimizations in solution took 50–60 iterations). This is due to the fact that Jaguar uses multiple criteria to determine convergence, such as thresholds on the elements of the gradient. In addition, the low-energy twists and turns of the hydroxyls adsorbed to the molecules contributed to the longer geometry optimizations for EdU. Thus, for the frequency calculations we went back through the geometries of the optimizations and used those that were truly the minimum-energy geometries, which we later did automatically by using the `nofail=1` setting.

Results for the frequencies are shown in Table 6.1. The Min group had already determined the vibrational frequencies of the first three molecules in Fig. 6.1, which are included in parentheses in the corresponding rows of Table 6.1. We see that our methodology produces frequencies to excellent accuracy compared to experiment, with an error for these three molecules of $3\,\mathrm{cm}^{-1}$. Furthermore, the Min Group found a typical full-width half-maximum of $14\,\mathrm{cm}^{-1}$ of the alkyne vibrations, allowing us to take into account the shapes of the simulated spectra as well. We estimate the bottom frequencies of the alkyne peaks by adding and subtracting $14\,\mathrm{cm}^{-1}$ from our simulated peak centers, plotting the results in Fig. 6.2, from which we see that molecules 1–4 are generally well-separated throughout the frequency ranges and thus may indeed serve as unique Raman markers. On the other hand, of the deuteriated molecules 5–8, we found that taking into account the peak widths, only two unique Raman markers may be possible; from Fig. 6.2 we see that peak 5 may be too similar to peak 8 and that peak 6 may be too similar to peak 7. This suggests to Min Group that only one in each pair should be synthesized.

It turned out that the deuteriated isotopes (molecules 5–8) that Dr. Chen attempted to synthesize were kinetically unstable, and further he found that molecule 4 was difficult to synthesize. In an effort to save time,

Molecule	Frequency (cm^{-1})
1	2126 (2125)
2	2076 (2077)
3	2051 (2048)
4	2101
5	1981
6	1942
7	1933
8	1974

Table 6.1: DFT-determined frequencies of different alkyne tags attached to EdU molecules. Experimental findings for the first three molecules are in parentheses. The numbering of the molecules corresponds to that in Fig. 6.1.

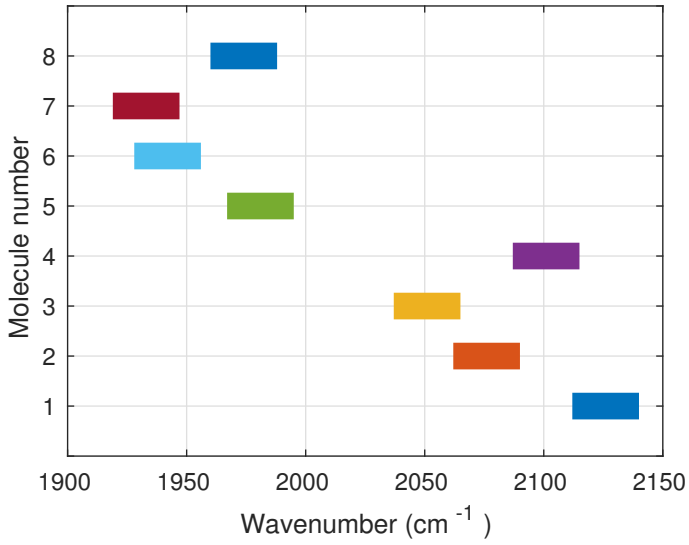

Figure 6.2: Graphical presentation of the DFT-predicted alkyne group frequencies accounting for a linewidth (FWHM) of $14\,\mathrm{cm}^{-1}$. To be safe, we choose the widths of the bars to be $2 \times 14 = 28\,\mathrm{cm}^{-1}$, centered around the calculated frequencies.

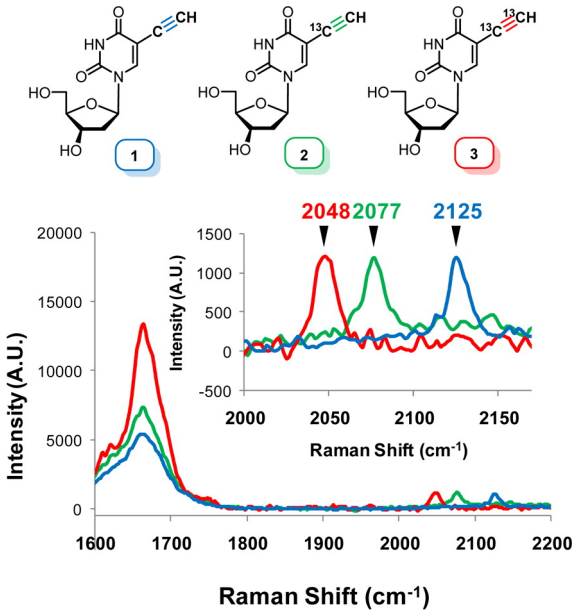

Figure 6.3: Spontaneous Raman spectra of HeLa cells incubated with the first three EdU molecules, showing that experimentally, too, the vibrations are found to be well-separated. The larger plot at the bottom shows the relative wavenumbers and intensities of the peptide bond vibrations ($1655\,\mathrm{cm}^{-1}$) in each of the three EdU samples. Image courtesy of Dr. Zhixing Chen.

the experimentalists started with EdU molecules 1–3 for their study of using the tags for live-cell imaging. They found promise in the discriminatory capabilities of these three types of alkyne tags from experimental spontaneous Raman spectra, shown in Fig. 6.3. We note the larger peaks at around $1655\,\mathrm{cm}^{-1}$, which correspond to the peptide bond vibrations. As a test of live-cell imaging using the alkyne isotope groups as tags, the Min and Nuckolls groups attached tags 1–3 to EdU molecules in HeLa cells that were synthesizing DNA. Three different samples were used, one for each species of tag, and SRS imaging was performed in five channels for each of the three samples: $1655\,\mathrm{cm}^{-1}$ (detecting the high-intensity peptide vibrations), $2000\,\mathrm{cm}^{-1}$ (deliberately off-resonant with any vibration), $2048\,\mathrm{cm}^{-1}$ (detecting the vibration of molecule 3), $2077\,\mathrm{cm}^{-1}$ (detecting the vibration of molecule 2), and $2125\,\mathrm{cm}^{-1}$ (detecting the vibration of EdU with a native alkyne tag; molecule 1). Fig. 6.4 clearly shows that SRS indeed detects the vibration actually present and does not detect any other vibrations, as each of the three rows in the figure shows only Raman signals from a single alkyne channel. The rightmost column shows the composite image, clearly showing where DNA is being synthesized in each of the three samples, since EdU is known to incorporate into newly synthesized DNA.

Finally, later on we ran similar frequency calculations on propyne and phenylacetylene, whose calculations

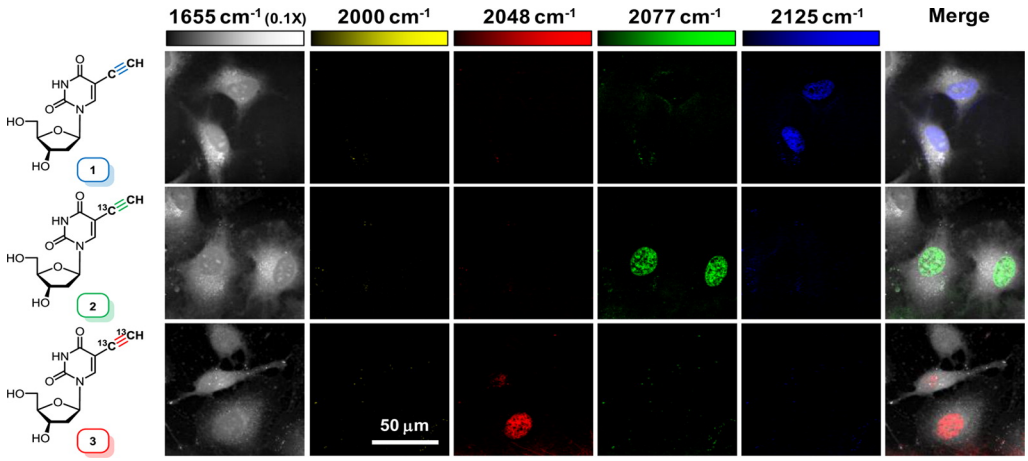

Figure 6.4: SRS imaging of HeLa cells using different EdU molecules in each of the three samples (in rows). The cells are in the process of synthesizing DNA, so it is expected that the EdU molecules be present. Each column represents a different frequency region studied using SRS: the first corresponds to the peptide bond vibrations, the second to an inactive region, the next three corresponding to the vibrational frequencies of the first three EdU molecules, and the last a composite image of all five Raman channels. Each differently-tagged sample shows the EdU molecules present in the HeLa cells, and the absence of the other two channels in any of the samples clearly shows the ability of the alkyne isotopes to differentiate between frequency regions. Image courtesy of Dr. Zhixing Chen.

seemed to have fewer convergence issues (the solution phase optimizations generally completed at energetic minima), for the first three alkyne isotopes. Just as relative to the native EdU molecule the first three alkyne isotope shifts were -50, -75, and $-25\,\mathrm{cm}^{-1}$, we found similar shifts for the propyne and phenylacetylene alkyne isotopes relative to the native molecules: -49, -76, and $-25\,\mathrm{cm}^{-1}$ for propyne and -49, -75, and $-24\,\mathrm{cm}^{-1}$ for phenylacetylene. While the native frequencies were different ($2141\,\mathrm{cm}^{-1}$ and $2121\,\mathrm{cm}^{-1}$ vs. $2126\,\mathrm{cm}^{-1}$ for EdU), the frequency shifts were consistently $25\,\mathrm{cm}^{-1}$ apart, showing the potential for the alkyne isotopes as *general* vibrational tags of biomolecules. Computationally, these results show that careful electronic structure calculations using the pseudospectral method are capable of predicting vibrational frequencies that agree well with experiment, and it is clear that laborious chemical synthesis can be obviated when the theoretical predictions show no or marginal gain of producing vibrational tags of similar frequencies.

For the sake of completeness, a year or so after completing this project, Dr. Chen reached out asking for theoretical insight regarding his and Prof. Min's continued investigation into live cell analysis using multi-color Raman probes along a slightly different path. This lead to another successful collaboration in which we were again able to provide insight into their observations, this time investigating intensities in addition to the frequencies. However, currently they have a paper under review on the experimental imaging methods involved, and in an

effort to keep this private in the meantime they have requested we hold off on any discussion of this new work. However, a paper including the results of our collaboration is in preparation[113].

Chapter 7

Benchmarking the pseudospectral method for obtaining excited-state energies and gradients of large molecules

As mentioned in Chapter 2, the pseudospectral method for calculating the two-electron integrals required for electronic structure calculations[1] has been shown to provide properties of atoms and molecules to chemical or near-chemical accuracy at least as quickly as any other electronic structure algorithm[9–18]. In particular, Jaguar, the primary electronic structure software to which we have access and that can utilize the pseudospectral method, has been shown to be one of the fastest electronic structure programs out there for non-periodic systems. However, the cost of implementing such a complex algorithm as the pseudospectral method is a slower development pace, as each new application utilizing the algorithm requires more time to implement and test.

In particular, while the traditional spectral method for electronic structure calculations has been shown to produce accurate excited-state energies and excited-state energy gradients at reasonable speeds, the pseudospectral method has until recently not even implemented these excited-state applications. Recently, however, Schrödinger Inc., the current developer of Jaguar, has completed implementing accurate calculation of excited-state energies and gradients, and so it remains to be seen how the pseudospectral method compares with fully spectral methods in terms of calculation speed for these properties. Such a study is important because without a significant speed-up using the pseudospectral method — while maintaining accurate excited-state energy and gradient values — there

[1] See Section 2.2 on page 37 for a brief explanation of the pseudospectral implementation of these integrals.

Vendor	AMD
CPU	Opteron 2376
Number of cores	4
Number of threads	8
Base speed	2300 MHz
CPU	800 MHz
System name	foct04
Total memory	32173 Mb

Table 7.1: Properties of the node used for the benchmarking.

would be little reason to utilize the pseudospectral method in the first place, for either research or commercial applications.

Thus, our aim was to carry out this comprehensive benchmarking of the pseudospectral method for calculating excited-state energies and their gradients, which we did by comparing these calculations in Jaguar to those in Q-Chem, an excellent atomic orbital-based electronic structure program that utilizes the traditional spectral method and that is known for producing accurate results at fast speeds. These results are presented below. First, however, we carry out benchmarking for the calculation of ground-state energies and their gradients in an effort to confirm the already-known speed advantages of the pseudospectral method compared to other methods and to get a feel for the degree of speed-up, if any, for the subsequent performance of the corresponding excited-state calculations.

All calculations were done on a single logical processor on a single eight logical-processor, four-core node, the properties of which are described in Table 7.1. We chose to perform the benchmarking on a single processor in order to eliminate the complex consideration and testing on multiple cores. We note that this does not take into consideration the different parallelization implementations between the two methods, and rather focuses on the algorithms themselves. However, as all codes nowadays are designed for multiple CPUs, this benchmarking could be more realistic. In particular, it is certainly possible that one of the two methods studied is better-suited for parallelization. Nevertheless, numerous factors go into the testing of even the serial jobs alone, and for the time being we focus on timing comparisons on a single processor.

In general, four jobs were run at once on the node, so that the maximum CPU usage of the node was 50%. We took care to make sure appropriate memory settings were chosen in the codes so that the node's memory would not be saturated at any time. For timing comparisons we compared the total wall time using our own external

timestamps; however, since these were serial jobs, this is technically equivalent to the total user+sys CPU time. We chose as default options as possible for both Jaguar and Q-Chem, since each program has already put much work into determining default values of cutoff/convergence criteria that result in accurate results obtained as quickly as possible. The one non-default option we chose for both programs was turning off the use of symmetry, since the different molecules studied have varying degrees of symmetry, but it would be useful to us to be able to rely on only the number of basis functions as our primary independent variable.

Our test set consisted of 46 molecules falling into three different categories: organic molecules ($N = 18$), nanotubes ($N = 8$), and C_n fullerenes of different sizes n ($N = 20$). These molecules ranged in sizes of 30–228 atoms and 450–3180 basis functions, respectively, calculated at the B3LYP/6-31G** level of theory. Four distinct jobs (ground-state energy, ground-state gradient, one singlet excited-state energy, and excited-state gradient) were run for each molecule, for each program, corresponding to 368 jobs total. As the Tamm-Dancoff approximation (TDA) to time-dependent DFT was well-implemented in each program (at the time, full linear response excited-state gradient calculations crashed Q-Chem), the excited-state calculations were all run using this approximation. The same molecule geometries were used for both programs, though we first perturbed the coordinates of all molecules slightly using a random Gaussian distribution in order to obtain non-zero ground-state gradients. We used the latest available versions of each program at the time of benchmarking (v2014-3 of Jaguar and v4.2 of Q-Chem).

First, we made sure we were able to obtain results in Jaguar that compared well in accuracy to those calculated using Q-Chem. For the ground- and excited-state energy comparisons, we calculated the mean absolute deviation for all molecules, obtaining values of 0.0072 Ha (0.0003%) for the ground-state energies and 0.016 eV (0.8%) for the excited-state energies. For the ground- and excited-state gradient comparisons, we used the RMSD of the elements of the gradients of Jaguar relative to Q-Chem, resulting in an average RMSD disagreement of 0.00065 Ha/Bohr (5%) for the ground-state gradients and 0.00052 Ha/Bohr (3%) for the first excited-state singlet gradients. It is interesting that we achieved excited-state gradient agreement that compared better with Q-Chem than the ground-state gradients. It should be noted however that for the 6x6x4 nanotube, Jaguar calculated a negative excitation energy (-0.14 eV), which is clearly wrong. We nevertheless include the corresponding timing data, which is justified by the fact that, as seen in the top-right plot of Fig. 7.3, the job time still fits perfectly in line with that of the other nanotubes. The excitation energy for tetraphenylporphyrin was also an outlier, differing from Q-Chem's calculation by 0.20 eV. Again, the corresponding timing results do not appear to be affected, and

we include them in the aggregate analysis below. Overall, however, it is clear that Jaguar's excited-state energies and gradients compare well with Q-Chem's.

In Tables 7.2 to 7.7 for each molecule in our test set we report the number of atoms in the molecule, the number of basis functions b used, the time t in hours the job took for each program, and the speed-up of the Jaguar job relative to the Q-Chem job, calculated by dividing the Q-Chem runtime by the Jaguar runtime. Within each molecule class (organics, nanotubes, fullerenes), we sort the tables by the number of basis functions, and at the bottom of each table we report the average, minimum, and maximum speed-ups, as well as the prefactors C and exponents N for the function $t = Cb^N$ fitted to each set of data. As expected, there is less speed-up for the ground-state calculations than the excited-state calculations. The average speed-ups were 1.2, 1.4, and 1.9 for the ground-state energy jobs for the organics, nanotubes, and fullerenes, respectively, and the ground-state gradient speed-ups were 2.7, 4.1, and 5.4. Taking the averages and noting the specifications above, we therefore say that on average, the pseudospectral algorithm is 1.5 times faster than the fully spectral algorithm for energy calculations and 4.1 times faster for gradient calculations, for jobs run on a single CPU. There is much more of an improvement for the excited-state jobs (using the TDA approximation): speed-ups of 5.6, 6.8, and 10.9 for energy calculations and 8.0, 10.5, and 14.7 for the gradient calculations, so that we can say that on average, the pseudospectral algorithm is 7.8 times faster then the fully spectral algorithm for TDA excited-state energy calculations and 11.0 times faster for excited-state gradient calculations. We therefore expect that the extra up-front work put into implementing the pseudospectral algorithm is well worth it.

Figs. 7.1 to 7.4 show plots of the data from these tables, indicating a polynomial time-dependence, and Fig. 7.5 shows the calculated prefactors and exponents based on a least-squares fit. The prefactors indicate how fast the jobs will run on small molecules (the smaller the prefactor, the faster the job), whereas the exponents give us a better idea of the scaling on larger molecules (the smaller the exponent, the faster the job). Thus, from the results of this figure we see that for organic molecules, Jaguar is faster for smaller molecules for all calculation types (ground-state energy, ground-state gradient, excited-state energy, excited-state gradient), but is slower for larger molecules for the first three calculation types and slightly overtakes Q-Chem for excited-state gradient calculations. The results for the nanotubes show that for smaller molecules Jaguar is faster only for ground-state energy and excited-state gradient calculations, and for larger molecules Jaguar is slightly slower for ground-state energy calculations but faster for all other calculation types. For the fullerenes, for smaller molecules Jaguar is faster for all calculation types except for excited-state gradients, and for larger molecules, Jaguar is slower for

			Ground-state energy			Ground-state gradient		
molname	**natom**	**nbfunc**	**Q-Chem**	**Jaguar**	**Speed-up**	**Q-Chem**	**Jaguar**	**Speed-up**
phenylporphyrin	48	540	0.11	0.08	**1.3**	0.35	0.14	**2.6**
ltn870	61	685	0.17	0.11	**1.6**	0.54	0.19	**2.8**
cycloartenol	81	715	0.24	0.18	**1.3**	0.75	0.28	**2.7**
brassinolide	82	750	0.26	0.21	**1.2**	0.83	0.29	**2.9**
photochromic-crownether1	72	760	0.29	0.20	**1.5**	1.07	0.31	**3.5**
rubrene	70	770	0.29	0.19	**1.5**	0.99	0.29	**3.4**
ltn828	70	810	0.18	0.15	**1.3**	0.59	0.24	**2.4**
ltn864	76	860	0.20	0.17	**1.2**	0.68	0.29	**2.4**
tetraphenyl-porphyrin	78	870	0.28	0.21	**1.3**	0.95	0.34	**2.8**
bp4mpy	98	1090	0.33	0.27	**1.2**	1.12	0.43	**2.6**
vinblastine	117	1175	0.73	0.56	**1.3**	2.57	0.86	**3.0**
vincristine	116	1180	0.75	0.69	**1.1**	2.66	0.96	**2.8**
paclitaxel	113	1185	0.75	0.67	**1.1**	2.64	0.93	**2.8**
twisted_acene	120	1340	0.93	0.60	**1.5**	3.38	1.01	**3.3**
demoxytocin	133	1353	0.79	0.91	**0.9**	2.77	1.18	**2.3**
borromeanring	122	1370	0.47	0.43	**1.1**	1.65	0.65	**2.5**
oxytocin	135	1373	0.75	1.07	**0.7**	2.57	1.38	**1.9**
carbetocin	138	1384	0.72	0.97	**0.7**	2.43	1.26	**1.9**
				Average:	**1.2**		*Average:*	**2.7**
				Min:	0.7		*Min:*	1.9
				Max:	1.6		*Max:*	3.5
		Prefactor:	3.2E-07	8.9E-09		5.2E-07	5.1E-08	
		Exponent:	2.03	2.53		2.14	2.34	

Table 7.2: Total calculation time in hours for the ground-state energy and gradient calculations for the 18 organic molecules. "Speed-up" is that of the pseudospectral algorithm (Jaguar) relative to the traditional, fully analytic algorithm (Q-Chem). These times are "scaled" in order to roughly take into account the machine load, since generally four jobs were run simultaneously. The reported times can thus be taken to be those one would observe if only that job were run on the testing machine.

both energy calculations, faster for excited-state gradient calculations, and equally as fast for ground-state energy calculations.

Averaging the three sets of polynomial parameters together, we find the overall parameter fits shown in Table 7.8. For the cases in which the polynomial curves cross, we report the basis set size overtake in the rightmost column to indicate how large of a molecule is needed for the program in the preceding column to apply. For ground-state energies, Jaguar appears to be faster for systems of up to about $4,800$ basis functions in size, at which point Q-Chem becomes faster. For ground-state gradients, Jaguar is faster for systems of all sizes. For excited-state energies, while Q-Chem has a slightly lower exponent, Jaguar has a significantly lower prefactor so

			TDA energy			TDA gradient		
molname	**natom**	**nbfunc**	**Q-Chem**	**Jaguar**	**Speed-up**	**Q-Chem**	**Jaguar**	**Speed-up**
phenylporphyrin	48	540	0.80	0.13	**6.4**	1.87	0.28	**6.7**
ltn870	61	685	0.79	0.14	**5.5**	2.53	0.39	**6.5**
cycloartenol	81	715	1.53	0.26	**5.8**	4.85	0.60	**8.0**
brassinolide	82	750	1.42	0.26	**5.4**	4.94	0.58	**8.5**
photochromic-crownether1	72	760	1.62	0.26	**6.3**	5.83	0.61	**9.6**
rubrene	70	770	1.43	0.24	**6.0**	5.20	0.58	**9.0**
ltn828	70	810	1.37	0.24	**5.7**	3.23	0.52	**6.2**
ltn864	76	860	1.08	0.22	**4.8**	3.39	0.57	**6.0**
tetraphenyl-porphyrin	78	870	1.80	0.34	**5.3**	5.35	0.68	**7.9**
bp4mpy	98	1090	2.16	0.40	**5.4**	5.98	0.90	**6.6**
vinblastine	117	1175	7.34	0.99	**7.4**	19.55	1.86	**10.5**
vincristine	116	1180	3.81	0.80	**4.8**	15.81	1.77	**8.9**
paclitaxel	113	1185	7.01	0.96	**7.3**	19.77	1.89	**10.4**
twisted_acene	120	1340	4.91	0.75	**6.5**	19.46	1.93	**10.1**
demoxytocin	133	1353	8.46	1.39	**6.1**	21.70	2.45	**8.9**
borromeanring	122	1370	2.33	0.59	**3.9**	7.99	1.46	**5.5**
oxytocin	135	1373	7.40	1.51	**4.9**	19.25	2.67	**7.2**
carbetocin	138	1384	5.46	1.34	**4.1**	16.15	2.46	**6.6**
				Average:	**5.6**		*Average:*	**8.0**
				Min:	3.9		*Min:*	5.5
				Max:	7.4		*Max:*	10.5
		Prefactor:	2.0E-07	1.2E-08		4.4E-07	1.1E-07	
		Exponent:	2.38	2.53		2.42	2.33	

Table 7.3: Total calculation time in hours for the excited-state energy and gradient calculations for the 18 organic molecules.

			Ground-state energy			Ground-state gradient		
molname	**natom**	**nbfunc**	**Q-Chem**	**Jaguar**	**Speed-up**	**Q-Chem**	**Jaguar**	**Speed-up**
ntube_6x6x1	60	660	0.15	0.13	**1.2**	0.54	0.19	**2.8**
ntube_6x6x2	84	1020	0.49	0.32	**1.6**	1.94	0.54	**3.6**
ntube_6x6x3	108	1380	1.03	0.62	**1.7**	4.12	0.96	**4.3**
ntube_6x6x4	132	1740	1.74	1.25	**1.4**	7.47	1.97	**3.8**
ntube_6x6x5	156	2100	2.70	2.11	**1.3**	11.61	2.73	**4.2**
ntube_6x6x6	180	2460	3.86	2.96	**1.3**	16.70	3.73	**4.5**
ntube_6x6x7	204	2820	5.26	3.55	**1.5**	25.29	5.01	**5.0**
ntube_6x6x8	228	3180	6.83	5.85	**1.2**	33.28	7.33	**4.5**
				Average:	**1.4**		*Average:*	**4.1**
				Min:	1.2		*Min:*	2.8
				Max:	1.7		*Max:*	5.0
		Prefactor:	2.8E-08	1.5E-08		3.1E-08	7.5E-08	
		Exponent:	2.40	2.44		2.58	2.28	

Table 7.4: Total calculation time in hours for the ground-state energy and gradient calculations for the eight nanotubes.

			TDA energy			TDA gradient		
molname	**natom**	**nbfunc**	**Q-Chem**	**Jaguar**	**Speed-up**	**Q-Chem**	**Jaguar**	**Speed-up**
ntube_6x6x1	60	660	0.64	0.15	**4.3**	2.52	0.37	**6.8**
ntube_6x6x2	84	1020	2.69	0.39	**6.9**	9.93	1.03	**9.6**
ntube_6x6x3	108	1380	5.35	0.74	**7.3**	22.66	1.90	**12.0**
ntube_6x6x4	132	1740	9.84	1.44	**6.8**	42.37	3.97	**10.7**
ntube_6x6x5	156	2100	15.78	2.40	**6.6**	67.50	5.76	**11.7**
ntube_6x6x6	180	2460	24.36	3.34	**7.3**	97.39	8.69	**11.2**
ntube_6x6x7	204	2820	31.99	4.09	**7.8**	133.68	11.73	**11.4**
ntube_6x6x8	228	3180	48.25	6.32	**7.6**	180.05	17.43	**10.3**
				Average:	**6.8**		*Average:*	**10.5**
				Min:	4.3		*Min:*	6.8
				Max:	7.8		*Max:*	12.0
		Prefactor:	2.3E-08	2.7E-08		7.8E-08	4.8E-08	
		Exponent:	2.66	2.38		2.68	2.44	

Table 7.5: Total calculation time in hours for the excited-state energy and gradient calculations for the eight nanotubes.

			Ground-state energy			Ground-state gradient		
molname	**natom**	**nbfunc**	**Q-Chem**	**Jaguar**	**Speed-up**	**Q-Chem**	**Jaguar**	**Speed-up**
c30-c2v-2	30	450	0.20	0.10	**2.0**	0.72	0.16	**4.5**
c34-c2-1	34	510	0.29	0.13	**2.2**	1.06	0.21	**5.1**
c36-c1-3	36	540	0.37	0.12	**3.1**	1.48	0.23	**6.4**
c44-d2-2	44	660	0.61	0.23	**2.7**	2.25	0.39	**5.7**
c46-c1-4	46	690	0.62	0.24	**2.6**	2.33	0.39	**5.9**
c48-c2-1	48	720	0.62	0.27	**2.3**	2.45	0.41	**6.0**
c60-ih	60	900	0.68	0.37	**1.8**	2.79	0.53	**5.3**
c70-d5h	70	1050	0.94	0.81	**1.2**	4.38	1.09	**4.0**
c72-d6d	72	1080	1.07	0.63	**1.7**	4.25	0.86	**4.9**
c76-d2	76	1140	1.19	0.77	**1.6**	5.32	1.04	**5.1**
c78-d3-1	78	1170	1.19	0.74	**1.6**	5.43	1.01	**5.4**
c80-d5d-1	80	1200	1.37	0.75	**1.8**	6.07	1.06	**5.7**
c84-c1-12	84	1260	1.49	0.93	**1.6**	6.94	1.17	**5.9**
c86-c1-1	86	1290	1.73	0.88	**2.0**	7.46	1.22	**6.1**
c90-c1-3	90	1350	1.71	1.03	**1.7**	7.58	1.40	**5.4**
c92-c1-2	92	1380	1.93	1.31	**1.5**	8.70	1.68	**5.2**
c94-c1-2	94	1410	1.89	1.18	**1.6**	8.60	1.68	**5.1**
c96-c1-7	96	1440	2.30	1.19	**1.9**	9.61	1.61	**6.0**
c98-c1-2	98	1470	2.24	1.23	**1.8**	9.81	1.80	**5.4**
c100-c1-9	100	1500	2.12	1.28	**1.6**	10.42	2.11	**4.9**
				Average:	**1.9**		*Average:*	**5.4**
				Min:	1.2		*Min:*	4.0
				Max:	3.1		*Max:*	6.4
		Prefactor:	4.1E-06	1.4E-07		4.2E-06	7.6E-07	
		Exponent:	1.80	2.19		2.01	2.01	

Table 7.6: Total calculation time in hours for the ground-state energy and gradient calculations for the 20 fullerenes.

			TDA energy			TDA gradient		
molname	**natom**	**nbfunc**	**Q-Chem**	**Jaguar**	**Speed-up**	**Q-Chem**	**Jaguar**	**Speed-up**
c30-c2v-2	30	450	1.47	0.13	**11.4**	3.10	0.32	**9.8**
c34-c2-1	34	510	1.66	0.16	**10.4**	4.04	0.40	**10.1**
c36-c1-3	36	540	2.11	0.16	**13.6**	5.39	0.48	**11.1**
c44-d2-2	44	660	3.05	0.27	**11.4**	8.49	0.73	**11.7**
c46-c1-4	46	690	4.28	0.31	**13.6**	10.52	0.79	**13.4**
c48-c2-1	48	720	3.96	0.32	**12.2**	10.73	0.82	**13.2**
c60-ih	60	900	6.37	0.49	**13.0**	17.81	1.10	**16.2**
c70-d5h	70	1050	13.14	1.02	**12.8**	30.18	1.88	**16.0**
c72-d6d	72	1080	6.64	0.71	**9.3**	24.38	1.58	**15.4**
c76-d2	76	1140	8.74	0.89	**9.8**	29.35	1.85	**15.9**
c78-d3-1	78	1170	7.57	0.85	**8.9**	28.11	1.86	**15.1**
c80-d5d-1	80	1200	8.31	0.85	**9.8**	30.54	1.92	**15.9**
c84-c1-12	84	1260	12.95	1.14	**11.3**	38.55	2.27	**17.0**
c86-c1-1	86	1290	11.10	1.05	**10.6**	38.29	2.28	**16.8**
c90-c1-3	90	1350	11.57	1.21	**9.6**	40.63	2.51	**16.2**
c92-c1-2	92	1380	12.95	1.46	**8.9**	45.17	2.88	**15.7**
c94-c1-2	94	1410	13.06	1.34	**9.7**	45.16	2.87	**15.7**
c96-c1-7	96	1440	14.26	1.37	**10.4**	50.65	3.05	**16.6**
c98-c1-2	98	1470	14.70	1.45	**10.2**	51.11	3.20	**15.9**
c100-c1-9	100	1500	15.55	1.50	**10.4**	53.21	3.48	**15.3**
				Average:	**10.9**		*Average:*	**14.7**
				Min:	8.9		*Min:*	9.8
				Max:	13.6		*Max:*	17.0
		Prefactor:	1.3E-05	3.0E-07		2.7E-06	3.2E-06	
		Exponent:	1.91	2.11		2.30	1.89	

Table 7.7: Total calculation time in hours for the excited-state energy and gradient calculations for the 20 fullerenes.

that effectively Jaguar is faster for systems of all sizes (up to systems having about 1.5×10^8 basis functions, which is effectively infinite). Finally, for excited-state gradients, the prefactors for Q-Chem and Jaguar are so close that even though Q-Chem's is barely lower, Jaguar's significantly lower exponent implies that effectively, Jaguar is always the faster program. Thus, under our assumption that Jaguar is representative of a cutting-edge implementation of the pseudospectral algorithm for the two-electron integrals and that Q-Chem is representative of a cutting-edge fully spectral algorithm, we conclude that implementation of the pseudospectral algorithm is well-worth the extra effort for excited-state energy and gradient calculations, as independently also concluded from the raw speed-up values bolded in Tables 7.2 to 7.7. The finding here that Q-Chem is faster for ground-state energies of molecules larger than $4,800$ basis functions is not obtained from these tables because all molecules in our test set had fewer than $4,800$ basis functions.

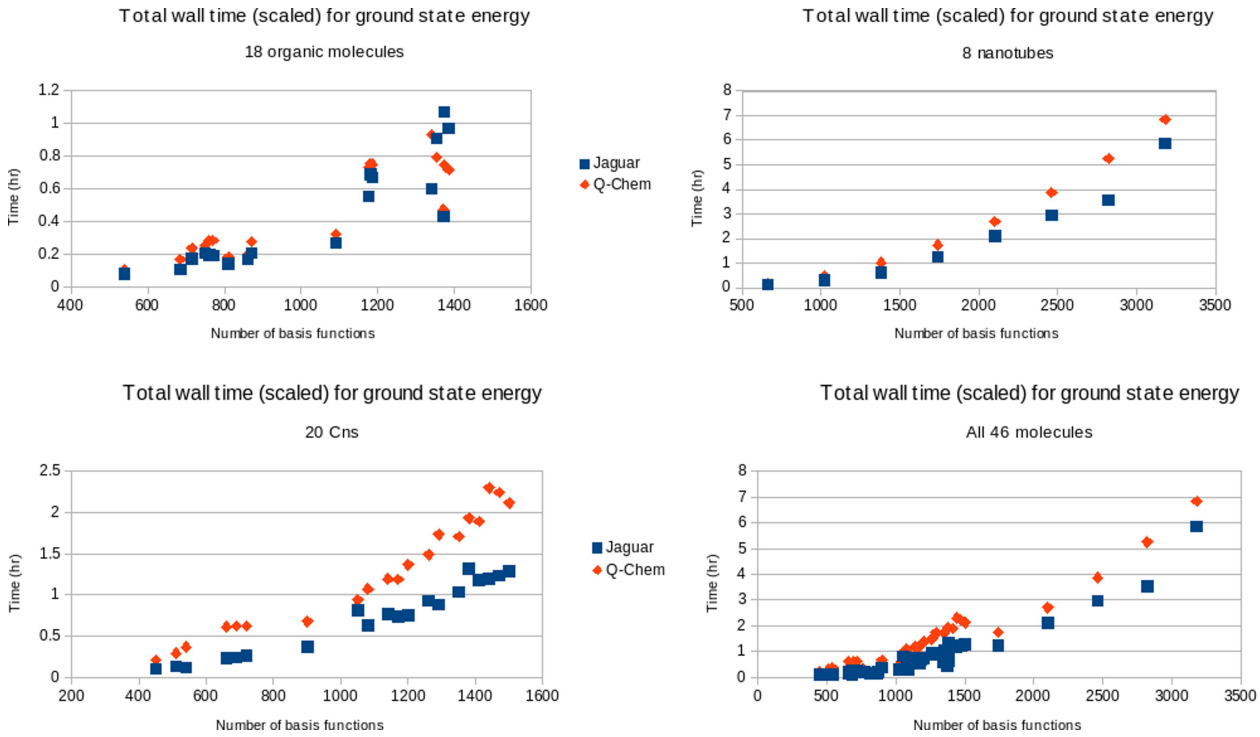

Figure 7.1: Ground-state energy calculation times. While these data are "scaled" in an attempt to take into account machine load, the corresponding "unscaled" data, and in particular, their trends, are very similar, as shown in Fig. 7.5.

	Q-Chem	Jaguar	Faster for smaller molecules	Faster for larger molecules	Basis size overtake
Ground-state energy	$t = \left(1.5 \times 10^{-6}\right) b^{2.08}$	$t = \left(5.6 \times 10^{-8}\right) b^{2.39}$	Jaguar	Q-Chem	4,800
Ground-state gradient	$t = \left(1.6 \times 10^{-6}\right) b^{2.24}$	$t = \left(3.0 \times 10^{-7}\right) b^{2.21}$	Jaguar	Jaguar	NA
Excited-state energy	$t = \left(4.4 \times 10^{-6}\right) b^{2.32}$	$t = \left(1.1 \times 10^{-7}\right) b^{2.34}$	Jaguar	Jaguar	NA
Excited-state gradient	$t = \left(1.1 \times 10^{-6}\right) b^{2.47}$	$t = \left(1.1 \times 10^{-6}\right) b^{2.22}$	Jaguar	Jaguar	NA

Table 7.8: Overall scalings for the different job types for each program, with determinations of the faster program for different molecule sizes, and, if applicable, the basis set size at which one program's speed overtakes that of the other.

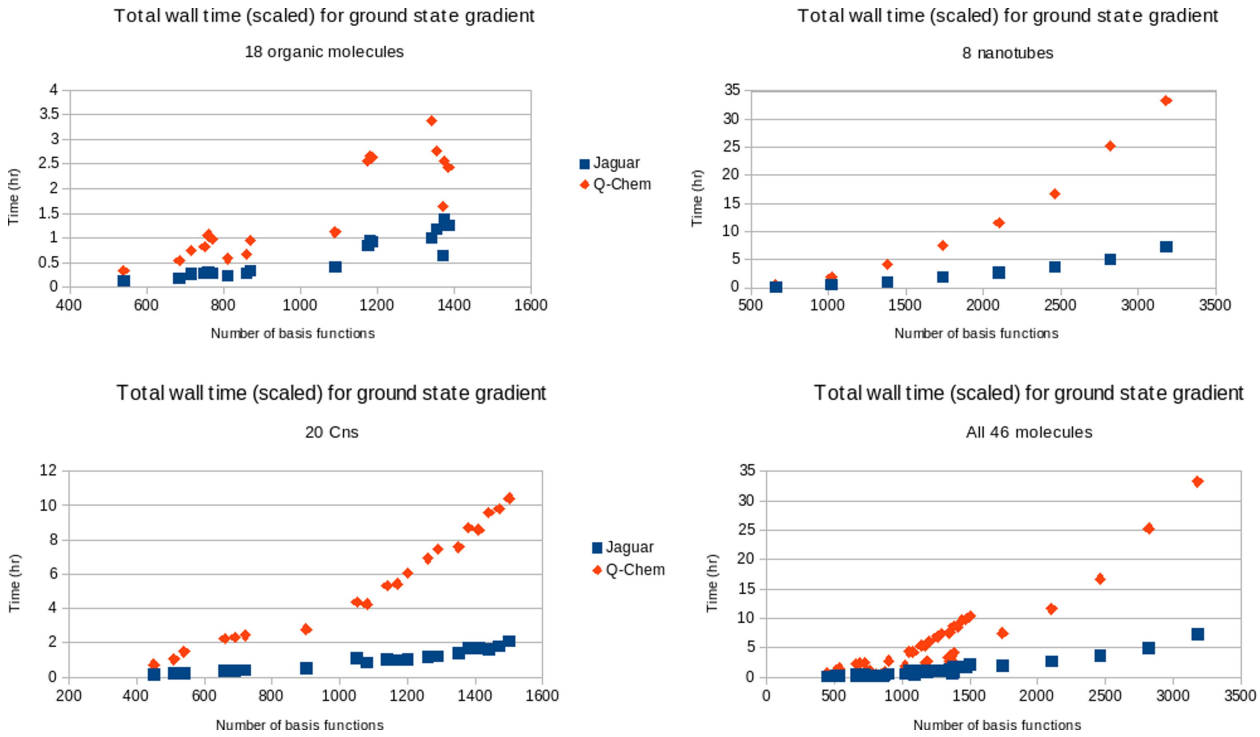

Figure 7.2: Ground-state gradient calculation times.

Finally, we note that for particular calculations (such as transition state calculations) one sometimes desires a basis set including more diffuse basis functions. We ran some less-rigorous excited-state energy tests on a subset of the molecules (the 20 molecules we included in [114] and that ultimately converged) using the 6-31G**++ basis set, and as often happens when adding diffuse basis functions, we found both programs to have a few convergence issues. We found that for the organic molecules and nanotubes, the speed-ups are reduced by about 10%, whereas for the fullerenes, they are reduced by about 50%. For Jaguar, this is likely due to the larger pseudospectral grids required for diffuse basis functions. However, we found agreement in accuracy for all 20 molecules to be just as good as for the 6-31G** basis set, and further Jaguar is still generally faster.

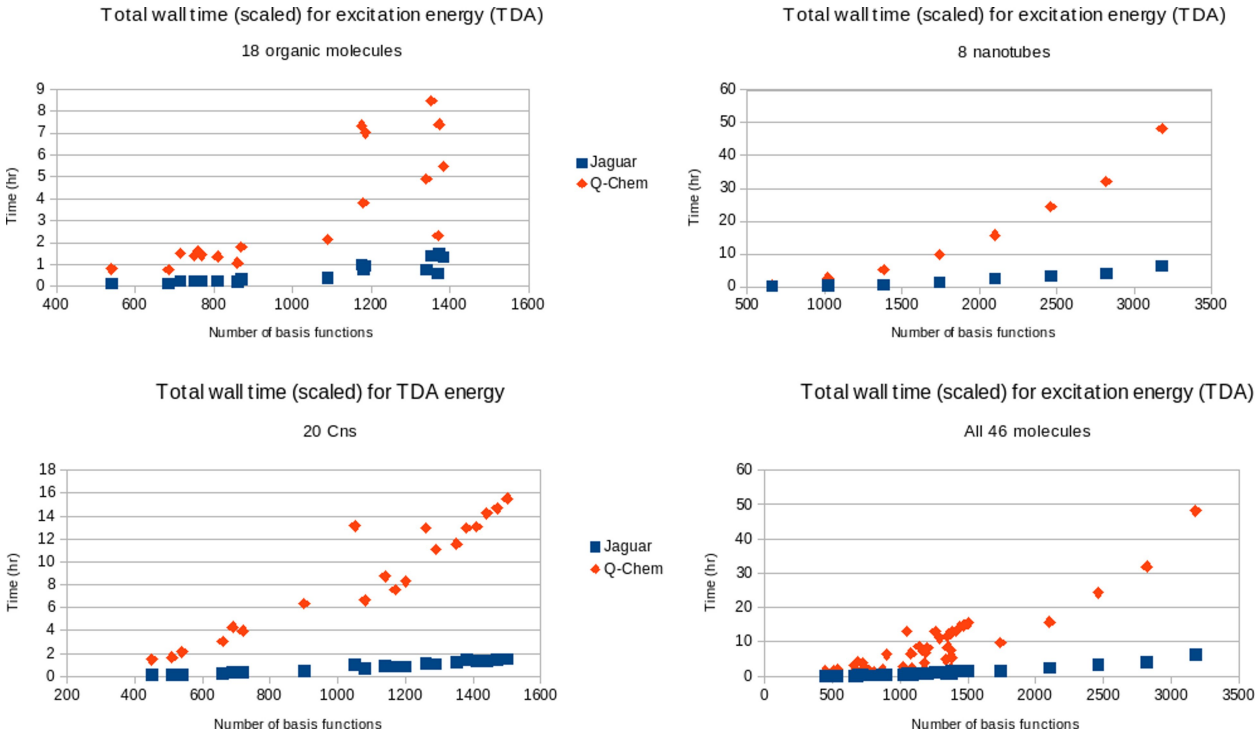

Figure 7.3: TDA energy calculation times.

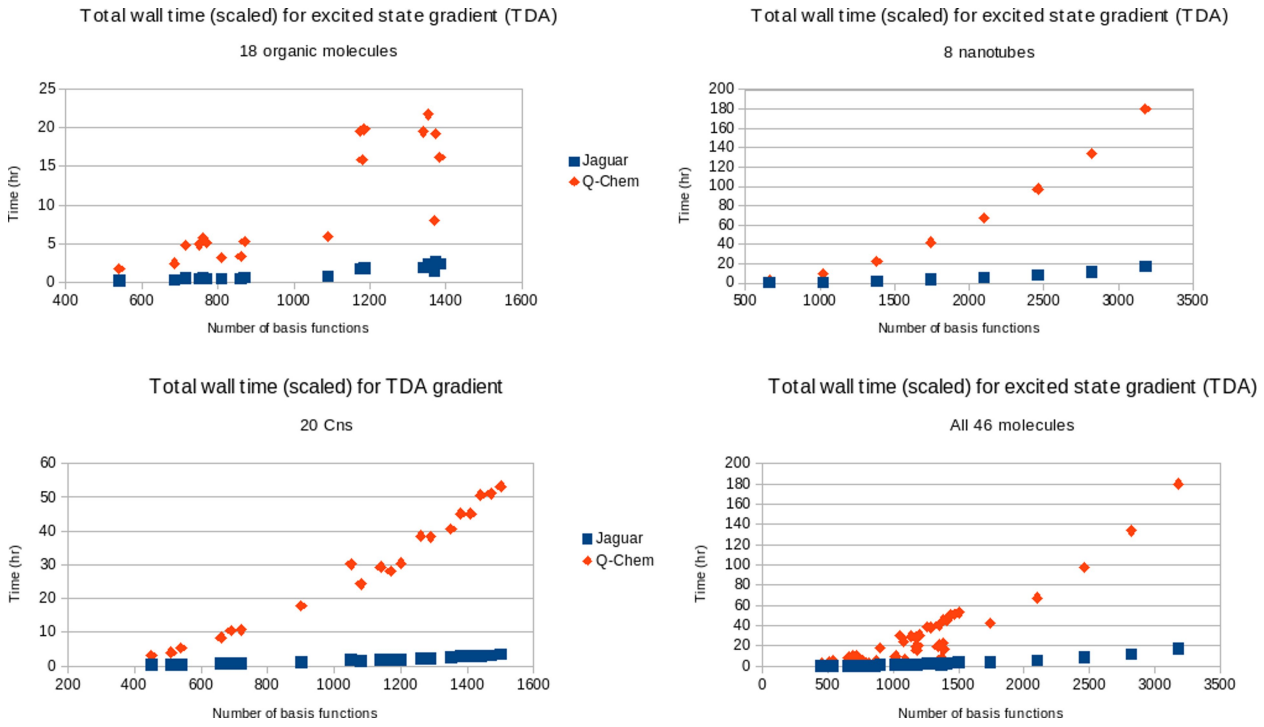

Figure 7.4: TDA gradient calculation times.

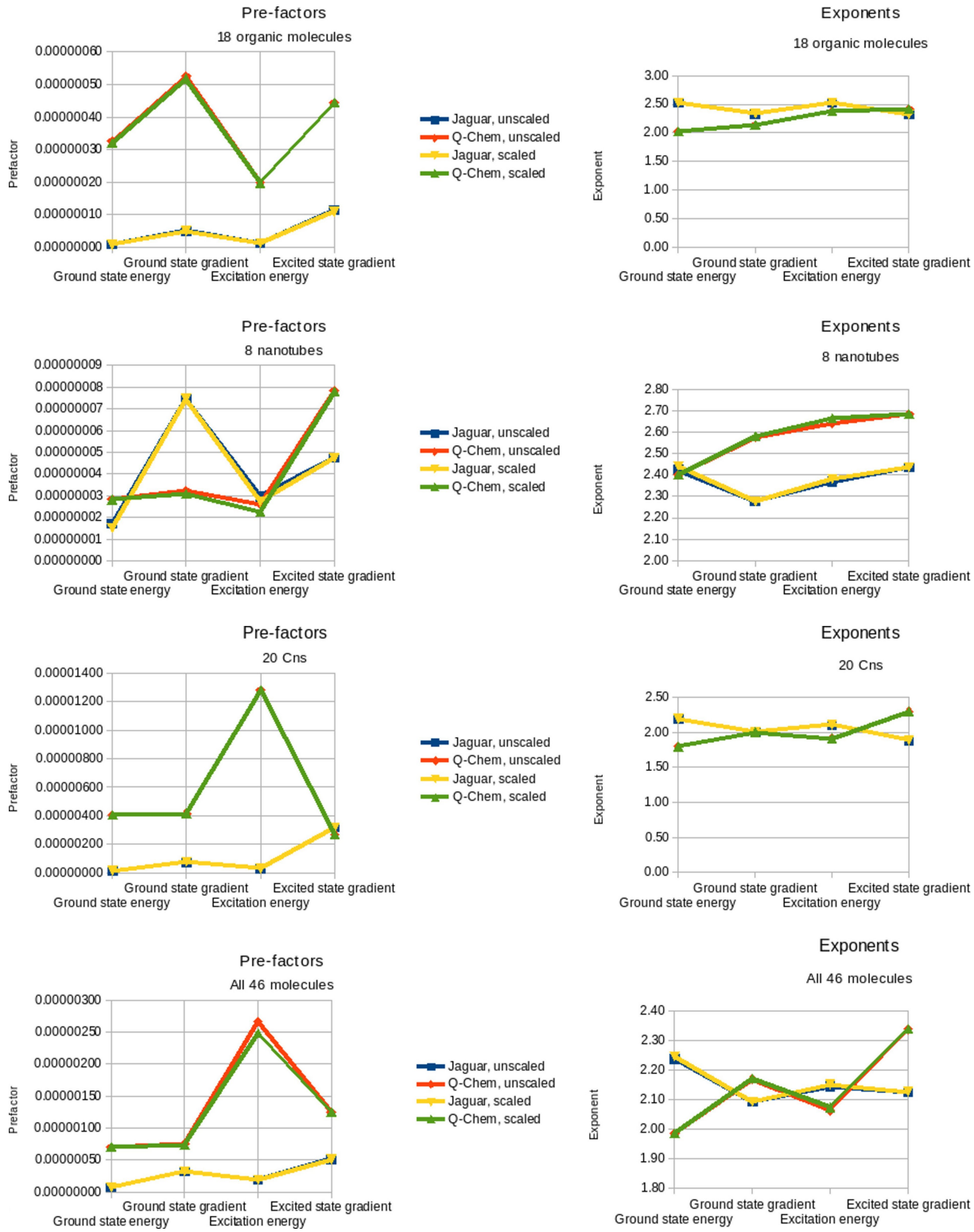

Figure 7.5: Scaling prefactors and exponents.

Chapter 8

Density functional theory study of the rate-limiting step of water splitting on large titanium dioxide nanoparticles

8.1 Introduction

Water splitting is the chemical reaction

$$2H_2O + \text{energy} \longrightarrow O_2 + 2H_2,$$

and it is intriguing because the hydrogen gas (H_2) produced can be used as a clean energy source. This is because the reverse reaction,

$$O_2 + 2H_2 \longrightarrow 2H_2O + \text{energy},$$

not only produces useful energy, but in addition the only by-product is water, which of course has additional utility as, for example, a water source in areas also experiencing drought. Of course, the oxygen gas (O_2) required to combust with the H_2 to produce water is plentiful since it makes up 21% of the Earth's atmosphere[115]. Finally, the idea is made even more enticing if the forward water splitting reaction is driven using only sunlight, an essentially infinite source of energy.

However, researchers excited by this idea have discovered that it requires a lot more energy than is stored in the hydrogen and oxygen bases produced[116–120], making the process unexpectedly inefficient and difficult. This

difficulty is compounded because the reason for the extra energy is unclear; if the reason were better understood, this "overpotential" problem could more easily be overcome. Researchers have honed in on the source, but a precise understanding has proved elusive.

In an electrochemical cell, the water splitting reaction is broken up into two half-reactions, called the oxidation evolution reaction (OER), which occurs at the anode, and the hydrogen evolution reaction (HER), which occurs at the cathode:

$$\begin{aligned} \text{OER:} \quad & 2H_2O + 4h^+ \longrightarrow 4H^+ + O_2 \\ \text{HER:} \quad & 4H^+ + 4e^- \longrightarrow 2H_2, \end{aligned}$$

where h^+ is a hole, H^+ is a proton, and e^- is an electron. While the HER is generally an efficient process — and in a recent case even found to occur with 100% efficiency[121] — the OER is not[122], and this inefficiency is thought to be the source of the overpotential. The OER typically proceeds via four distinct proton-coupled electron transfers (PCETs)[122]:

$$\begin{aligned} \text{PCET } \#1: \quad & {}^*H_2O + h^+ \longrightarrow {}^*OH + H^+ \\ \text{PCET } \#2: \quad & {}^*OH + h^+ \longrightarrow {}^*O + H^+ \\ \text{PCET } \#3: \quad & {}^*O + H_2O + h^+ \longrightarrow {}^*OOH + H^+ \\ \text{PCET } \#4: \quad & {}^*OOH + h^+ \longrightarrow {}^*O_2 + H^+. \end{aligned}$$

Here, an asterisk signifies a catalyst for the reaction, which is typically a semiconducting metal-oxide nanoparticle, and $^*X$ signifies that X is adsorbed to the surface of the catalyst. It is widely thought that the inefficiency of the OER lies in PCET #1 taking a significant amount of energy[119, 122], effectively resulting in the reduction potential of the OER increasing significantly from its thermodynamic value of 1.23 V[123, 124]. Thus, by studying possible contributions to the overpotential of PCET #1, we can identify the experimental modifications to be made that will minimize the overpotential and therefore allow water splitting to be achieved more efficiently. For example, it was suggested in [122, 125] that modification of the pH of the water in which the electrodes are submerged would improve system efficiency.

It is generally thought that the overpotential present in PCET #1 is due to multiple phenomena; here, we focus primarily on the thermodynamic and kinetic contributions arising from the instability of the intermediate species, $^*OH$, adsorbed on the surface of the catalyst. For example, it was found in [119] that the energy difference between the reactants and products of PCET #1 was 0.7 eV, indicating an overpotential of 0.7 V. [122] found that accounting for the kinetics of the reaction, i.e., the *way* the reactants transform into the products, contributed an additional 0.2–0.5 eV to the energy required, increasing the overpotential at the anode to 0.9–1.2 V. Our

goal is to study both the thermodynamics and kinetics of PCET #1 using our extremely fast methodology, the pseudospectral method, as implemented in the electronic structure software Jaguar.

Jaguar's speed affords us the ability to employ methodologies for particularly realistic simulations. In particular, our group has previously developed an accurate methodology for the modeling of TiO_2 clusters[126] in which all aspects of the problem are treated at least as realistically as in previous studies, and we think that applying this methodology to the study of the overpotential problem will contribute a deeper understanding to electrochemical research. The underlying theory we use is completely quantum-mechanical: density functional theory (DFT) using the (hybrid) B3LYP density functional and the atomic orbital-based LACVP basis set. We use as the nanoparticle TiO_2 in the anatase polymorph, widely thought to be the most photocatalytically active phase of TiO_2[127, 128], which is a prototypical material used in photocatalytic research. We have simulated cuts of such nanoparticles from their bulk structure (next section), including two that are nearly life-size anatase TiO_2 nanoparticles, each of which primarily exposes one of the two most photocatalytically active (yet stable) surfaces: the (001) and the (101). There is debate in literature regarding which surface is more photocatalytically active[129, 130] and we further hope to help settle this debate. In addition, for each TiO_2 nanoparticle, we explicitly passivate all surfaces with the water ligands that are likely to be adsorbed when the nanoparticles are soaked in neutral-pH water (as would exist in the prototypical water splitting cell); such passivation of all surfaces of the nanoparticle is impossible to do using plane wave-based basis sets and is a great advantage of our method. Finally, we run all our calculations in continuum solvent using the Poisson-Boltzmann model of implicit solvation in water.

8.2 Cutting anatase TiO_2 nanoparticles from their bulk structure

In order to proceed with simulations using anatase TiO_2 nanoparticles, we needed the Cartesian coordinates of the atoms in the nanoparticles. We started by obtaining the coordinates of the atoms in the unit cell of bulk anatase TiO_2 from a high-quality submission in the ICSD Web database[131] (collection code 63711), and by copying these unit cells in all three spatial directions in MATLAB, we were able to simulate the coordinates of the *bulk* of an anatase TiO_2 nanoparticle.

While anatase has been shown to be the most reactive, stable polymorph of TiO_2, it is debated which *surface* of such a nanoparticle is the most reactive[129]. The two primary surfaces of these nanoparticles are the (101) and (001) surfaces, as shown using the Wulff construction and verified experimentally. In recent years there has been a surge of experimental efforts[128, 132–138] in producing anatase TiO_2 nanoparticles exposing primarily (001)

facets due to earlier studies suggesting that the (001) surface is the most reactive, but recently this assumption has been called into question[129, 130, 137, 139]. We therefore aimed to produce nanoparticles exposing various amounts of both surfaces so that we could voice our own opinion as to which surface is the most reactive, in particular for the oxidation evolution reaction.

Thus, we needed to figure out how to "cut" our simulated bulk anatase TiO_2 structures along the (101) and (001) surfaces. In the $\langle 101 \rangle$ and $\langle 001 \rangle$ directions there are six and three different atomic planes, respectively, between which we could make our cuts. We first operated under the principle of having autocompensated, non-polar surfaces[140], i.e., making the cuts in which an equal number of Ti–O and O–Ti bonds would be broken in order to obtain surfaces of neutral charge. This defined well the cut to be made in the $\langle 001 \rangle$ direction but only narrowed down to two the number of cuts we could make in the $\langle 101 \rangle$ direction, as also mentioned in [141]. We then operated under the principle of cutting the fewer bonds per unit area, which would generally be the lower-energy cut of the bulk; this then defined well the particular cut to be made in the $\langle 101 \rangle$ direction. We verified our chosen cuts by viewing structures of anatase TiO_2 nanoparticles in literature (for example, [129]). Continuing to use MATLAB, we cut out from the bulk nanoparticles of various sizes and shapes in which cuts along these two surfaces and their symmetric equivalents were made, generally resulting in the expected truncated bipyramid structure, as shown in Fig. 8.1.

8.3 Explicit passivation of the surfaces of the nanoparticles with water ligands

The next step was figuring out how to appropriately passivate the (101) and (001) surfaces of the anatase TiO_2 nanoparticles with water ligands in a realistic water splitting environment (explicit passivation). We searched literature but were unable to find a justified consensus in the passivation schemes, so we came up with our own passivation scheme using basic atomic orbital theory. We assumed that the ideal water splitting environment for TiO_2 nanoparticles would simply be neutral water, in which case the molecules available to passivate the incorrectly-valenced and incorrectly-coordinated — i.e., "unhappy" — atoms at the just-cut surfaces would be H_2O (a "happy" molecule), *neutral* OH (missing one electron; "unhappy"), and *neutral* H (missing one electron; "unhappy"). The "unhappy" species of atoms at the appropriately-cut (101) surface were Ti atoms that were five-coordinated and in a $+3$ oxidation state ("Ti(1,-1)"; happy Ti atoms in bulk are six-coordinated and have oxidation states of $+4$) and O atoms that were two-coordinated and in a -1 oxidation state ("O(1,1)"; happy O atoms in bulk are three-coordinated and have oxidation states of -2). The unhappy species of atoms at the

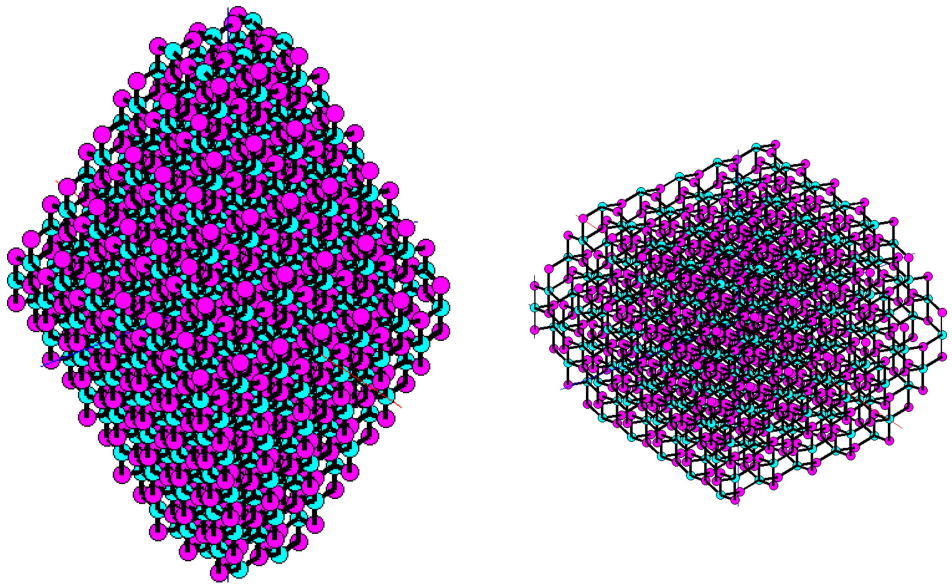

Figure 8.1: Unpassivated anatase TiO_2 nanoparticles exposing different degrees of the (101) and (001) surfaces, exhibiting the truncated bipyramid morphology. (Left) Primary exposure of the eight (101) surfaces. (Right) Primary exposure of the two (001) surfaces (top and bottom surfaces). Note that this latter structure is obtained by further truncating the bipyramid that is more clearly seen in the former. Ti atoms are in cyan and O atoms are in magenta. Interatomic distances are actually drawn to scale between the two nanoparticles; it is the atoms that are rendered at different sizes.

(001) surface were Ti atoms that were five-coordinated and already in the $+4$ oxidation state ("Ti(1,0)") and O atoms that were two-coordinated and already in the -2 oxidation state ("O(1,0)"). The unhappy Ti(1,-1) surface species wants to be satisfied by a single adsorbate since it is missing one coordination (hence the 1 in "(1,-1)") and wants to give away one electron to this adsorbate in order to obtain its correct oxidation state (hence the "-1"). In bulk this is normally satisfied by one of the four covalent Ti–O bonds extending from a Ti atom; of the available adsorbates in the neutral water environment, this could be satisfied by either the O atom of a neutral OH or a neutral H atom. We chose the OH and therefore passivated all Ti(1,-1) species with neutral hydroxyls. Similarly, because the unhappy Ti(1,0) surface species wants to be satisfied by a single adsorbate that needs no change in oxidation state, which in bulk is normally satisfied by one of the two donor-acceptor Ti–O bonds extending from a Ti atom, we passivated all these surface species by the O atoms of neutral water molecules. We likewise passivated the O(1,1) surface atoms with neutral H atoms, though we chose not to passivate the O(1,0) species of atoms, as discussed in Section 8.6. Note that other unhappy species of surface Ti atoms, Ti(2,-2) and Ti(2,-1), which occurred at corners and edges between surfaces, were passivated using similar reasoning. Two

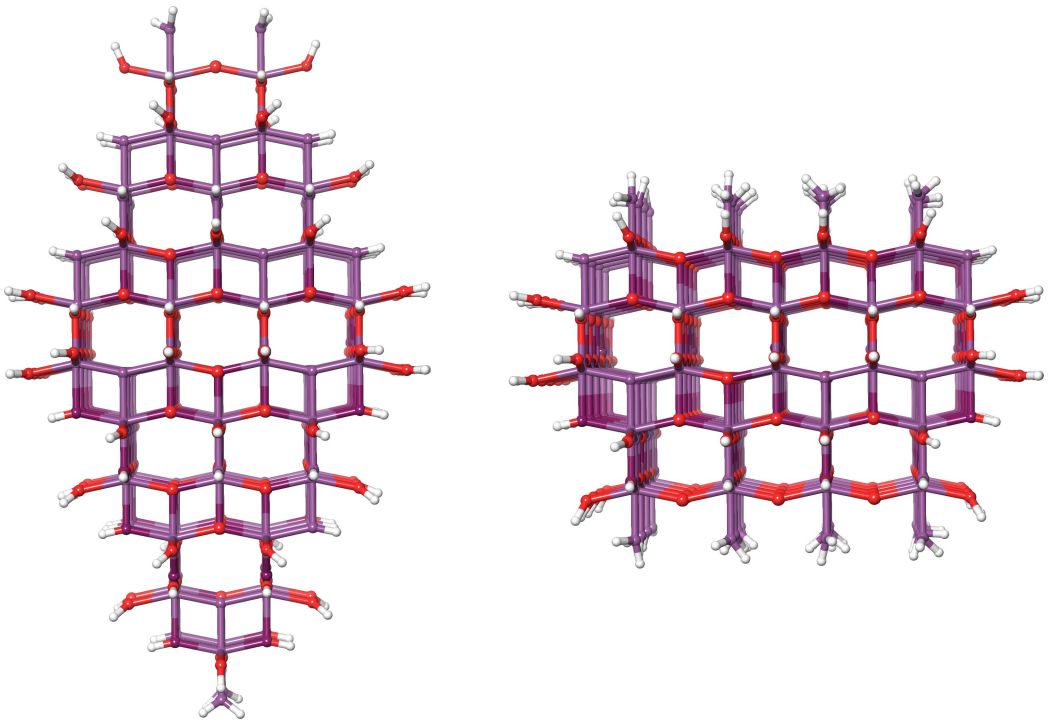

Figure 8.2: Sample passivated nanoparticles, a bit smaller than those of Fig. 8.1. Note that this view enables observation of the bulk structure of anatase within the nanoparticles. Deep purple atoms are artificial atoms that help the bonds display correctly in the program Maestro.

sample passivated nanoparticles are shown in Fig. 8.2.

8.4 Optimizing the nanoparticles using DFT

In order to use DFT to optimize the geometries of the nanoparticles we created in MATLAB, we first needed good initial guesses of the wavefunctions. However, for any of the nanoparticles larger than 58 atoms in size, none of the initial guesses built into Jaguar specifically to handle systems containing transition metals worked, so we needed to use the Fragmented Initial Guess (FIG) code developed by Dr. Jing Zhang and Prof. Friesner. However, we could not use the code outright because the physical partitioning scheme it used worked only for rutile polymorphs. We therefore developed our own, general partitioning scheme that should work for an arbitrary nanoparticle.

Physical partitioning scheme of arbitrary nanoparticles for obtaining working initial quantum wavefunctions for the self-consistent field procedure

Molecular systems containing transition metal atoms are well known to have complicated electronic structures due to the occupation of the *d* orbitals of these atoms. Likewise, calculation of the electronic structure of such systems has proven difficult, and numerous methods for such systems have emerged for initializing the quantum mechanical wavefunction needed for the self-consistent field (SCF) procedure that is the nexus of any electronic structure calculation. However, all such algorithms have failed for titanium dioxide nanoparticle clusters of both the rutile and anatase polymorphs. Dr. Jing Zhang, an ex-graduate student in the Friesner Group, produced a working method for initializing the wavefunction of rutile TiO_2 nanoparticles that works by iteratively converging the wavefunctions of the fragments of a larger cluster, calculating the resulting charge fields of the fragments at their neighbors, and re-converging the wavefunctions of the fragments in the presence of these charge fields using a Poisson-Boltzmann solver. Such an approach, called the Fragmented Initial Guess (FIG) method, has proven successful in generating initial guesses of the quantum wavefunction for rutile TiO_2 nanoparticles that can further be refined using the SCF procedure without failing.

However, such an approach for successfully initializing the quantum wavefunction requires careful partitioning of the cluster into individual fragments in the first place. The specific scheme used by Dr. Zhang for rutile TiO_2 could not work for the anatase polymorph because it partitioned the rutile cluster into sets of planes in a way that is not well-defined for the anatase polymorph. Motivated by the fact that anatase is the more catalytically active polymorph of TiO_2 for use as an anode of an electrochemical cell and our desire to realistically simulate the oxidation evolution reaction of water splitting on TiO_2, we set out to develop a fragmentation scheme that would not work only for an anatase TiO_2 cluster but also for a general nanoparticle. Therefore, the method needed to be based on general physical principles that could be applied to an arbitrary system.

Our idea was to break the nanoparticle up into fragments that are individually as stable as possible so that the nanoparticle resembles a system of realistic — i.e., energetically favorable — smaller particles interacting with each other electrostatically. The logic is that the more realistic these interacting fragments are, the better our DFT methods will describe the system, as these methods have been tested on realistic molecular systems for decades. We decided to form these low-energy fragments from basic units that themselves have the same crystal structure as the larger nanoparticle of interest. These basic units must have relatively low surface energies because they are themselves very small nanoparticles of a naturally forming crystal structure. For example, for

anatase TiO_2 we decided to use as our basic unit the smallest constructible bipyramidal chunk of atoms taken from anatase bulk. It turns out that such a chunk is the linear O–Ti–O group of atoms. Since the combination of two low-surface-energy units must itself yield a low-surface-energy unit (all exposed surfaces are still low-energy), we decided to use as our fragments in the FIG code clumps of the basic three-atom TiO_2 units, and we wrote a simple algorithm to break the non-passivated nanoparticle up into such fragments. We stipulated that: (1) each fragment consist of 34 to 99 atoms, (2) the last-formed fragment be at least 70% of the average size (in number of atoms) of all the other fragments, and (3) each fragment have a (# of O atoms) / (# of Ti atoms) ratio of at least 1.75. Finally, we set the last fragment to be the much-larger layer of passivating ligands.

In all of our runs of the FIG code on various-sized passivated anatase nanoparticles, including one containing 483 atoms (97 titanium atoms; a nearly realistically occurring size) and 4,796 basis functions, fragmenting the nanoparticle using this algorithm always produced at least one initial guess wavefunction whose SCF calculation converged. This occurred even when all other initial guess methods in the latest versions of Jaguar, including those developed specifically for transition metal systems, failed. Thus, this fragmentation scheme together with the FIG procedure has the potential to be extremely robust, and we expect that it can be applied to any polymorph of any crystal for which standard initial guess approaches experience difficulty.

Next, in order to optimize the nanoparticles in water (implicit passivation), we generally needed to first optimize them in vacuum for five iterations or so. After the solvent optimizations subsequently completed, our general approach was to then knock off an electron (with the assumption that the electron transfer occurs before the proton transfer during the first PCET of the OER) and then optimize the resulting cationic clusters in water. Keywords used for the optimizations and subsequent properties calculations are shown in Appendix E.

8.5 Optimizing Jaguar for the Stampede supercomputer

Previous studies on rutile TiO_2 for the study of the electron trapping and transport properties in dye-sensitized solar cells by Dr. Jing Zhang required that he compile Jaguar to work on the supercomputer Stampede in order to obtain manageable calculation runtimes. However, the version of Jaguar Dr. Zhang compiled was very old and unlicensed, and at the time it was a good idea to attempt to get the latest, licensed version of Jaguar set up on Stampede so that new developments in the code could be taken advantage of for use on our even-larger anatase nanoparticles. However, as the Jaguar codebase had grown significantly in size, making the code difficult to compile on even a single node not located at Schrödinger Inc., and since Schrödinger Inc. had begun phasing

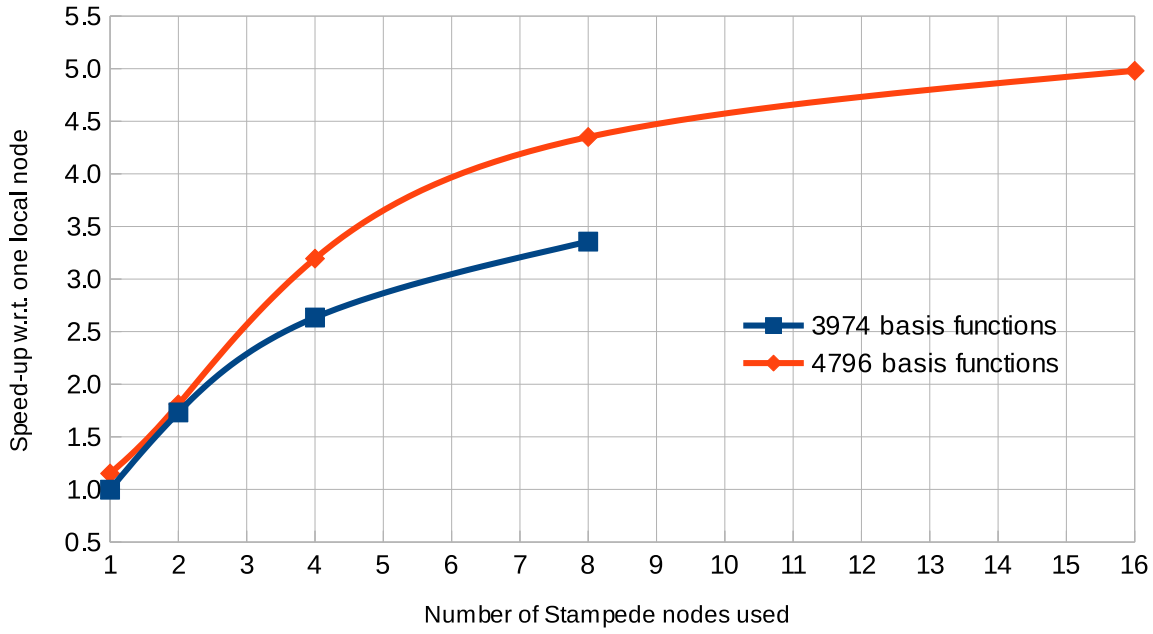

Figure 8.3: Speed-up (with respect to one local node) vs. number of Stampede nodes for our build of Jaguar optimized for Stampede. The speed was defined as the number of SCF calculations the job completed for however long the job was run, which was limited to 48 hours for the Stampede jobs and about 60 hours for the two local jobs. Since this is a rough estimation, only one job was run per data point. The two local jobs were run on very similar 24-core AMD machines, and with threading, the jobs used all 48 threads per node. The Stampede nodes consist of 16 cores and threading is turned off, so these jobs used 16 threads per node. When multiple nodes were used, one MPI process was spawned per node. Benchmarking was run on two systems, one exposing primarily the (101) surface of anatase TiO_2 (3974 basis functions; left image of Fig. 8.2) and the other exposing primarily the (001) surface (4,796 basis functions; right image of Fig. 8.2). The results show that Jaguar scales very well up to about four nodes (64 cores) and reasonably well up to about eight nodes (128 cores).

out internode communications (MPI; currently Jaguar is designed to run only on a single node), this turned out to be a difficult and time-consuming endeavor. However, with the help of Dr. Jerome Vienne at the Texas Advanced Computing Center at UT Austin, which hosts the Stampede supercomputer, we eventually got the full functionality of Jaguar working on Stampede. We further were able to optimize the code for Stampede's hardware using Intel's optimization tools, resulting in particularly excellent performance, as shown in Fig. 8.3. We leave out further details except to note that, after attempting to use the better-supported Open MPI and MPICH2 parallel libraries, we ultimately got internode communications implemented using Intel MPI, which is designed for the Intel Xeon nodes that make up Stampede anyway.

8.6 Reevaluation of the explicit passivation of the nanoparticle surfaces

Our two largest-optimized anatase TiO_2 nanoparticles primarily exposing different surfaces were shown in Fig. 8.2. However, initially we did not passivate the oxygen atoms exposed on the (101) surface ("O(1,1)", in the terminology of Section 8.3). Despite multiple meetings reviewing the optimization and orbital results of similarly passivated, increasingly larger nanoparticles, none of us caught something we eventually noticed in all the optimized neutral nanoparticles as well as their subsequent cationic variations: the formation of O–O bonds. The difference in passivation of the MATLAB geometry is shown in Panels (a) ("incorrect" passivation) and (b) ("correct" passivation) of Fig. 8.4. The difference in the subsequent solvent-optimized structures is shown in Panels (c) and (d), respectively. The results of Fig. 8.4 reflect the effects of the different passivations we saw for every nanoparticle observed so far: with the incorrect passivation (no passivation of any oxygen atoms), the bonds one or two layers from the surface deform significantly, and in particular some types of bonds, such as the O–O bonds, appear in these regions that were not initially present. The bulk of the nanoparticle somewhat deceivingly does stay intact, as we demonstrate in Fig. 8.5 with larger nanoparticles; the bulk of the nanoparticles in Panels (a) and (c) do indeed resemble the original structure. We may therefore generally expect that properties of the nanoparticles more or less described by the "inside" of the clusters, such as the band gap, should remain what they would be had the correct passivation been used, as shown in the densities of states plots of Fig. 8.6. However, due to significant deformation of the surface and subsurface layers, the nanoparticles can overall appear significantly altered from their bulk state. In particular, as the oxidation evolution reaction primarily takes place at the surface of nanoparticle catalysts, we do not expect the results of our DFT study of the OER to be realistic if incorrectly passivated nanoparticles are used, and in subsequent sections we use correctly passivated nanoparticles.

While in hindsight we wished we had caught this passivation issue earlier, the very fact that passivation of the O(1,1) surface atoms makes such a significant difference in the optimized nanoparticle structures is a useful finding in its own right. Not only does this suggest how these nanoparticles need to be passivated in order to model these systems correctly, but also it suggests to experimentalists that care must be taken to ensure their TiO_2 catalysts are correctly passivated as described above. If they find that their samples are deforming or measure surface-related properties that have unexpected values, one factor they might consider and address is that the surfaces of their samples have incorrect passivations.

In addition to the stability of the bulk of the nanoparticle, our findings show that appropriate passivation of the oxygen atoms directly impacts how tightly the (001) surface holds on to the adsorbed waters; we see in Panel

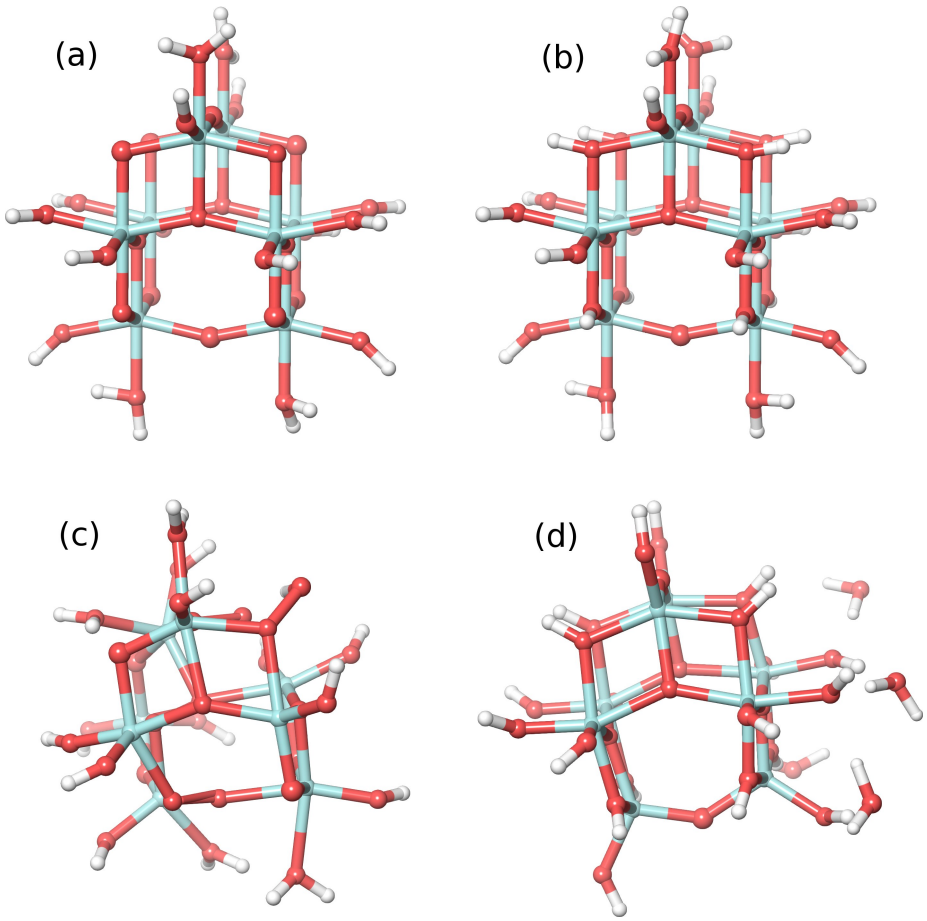

Figure 8.4: A small anatase nanoparticle with different passivations of the eight oxygen atoms having an incorrect oxidation state. (a) MATLAB geometry (i.e., cut from bulk) with the eight oxygen atoms unpassivated. (b) MATLAB geometry with the eight oxygen atoms passivated by neutral hydrogen atoms. (c) Solvent-optimized structure from (a). Note the formation of O–O–H (top right) and O=O (bottom left) bonds. (d) Solvent-optimized structure from (b). Now the overall structure of the nanoparticle stays intact.

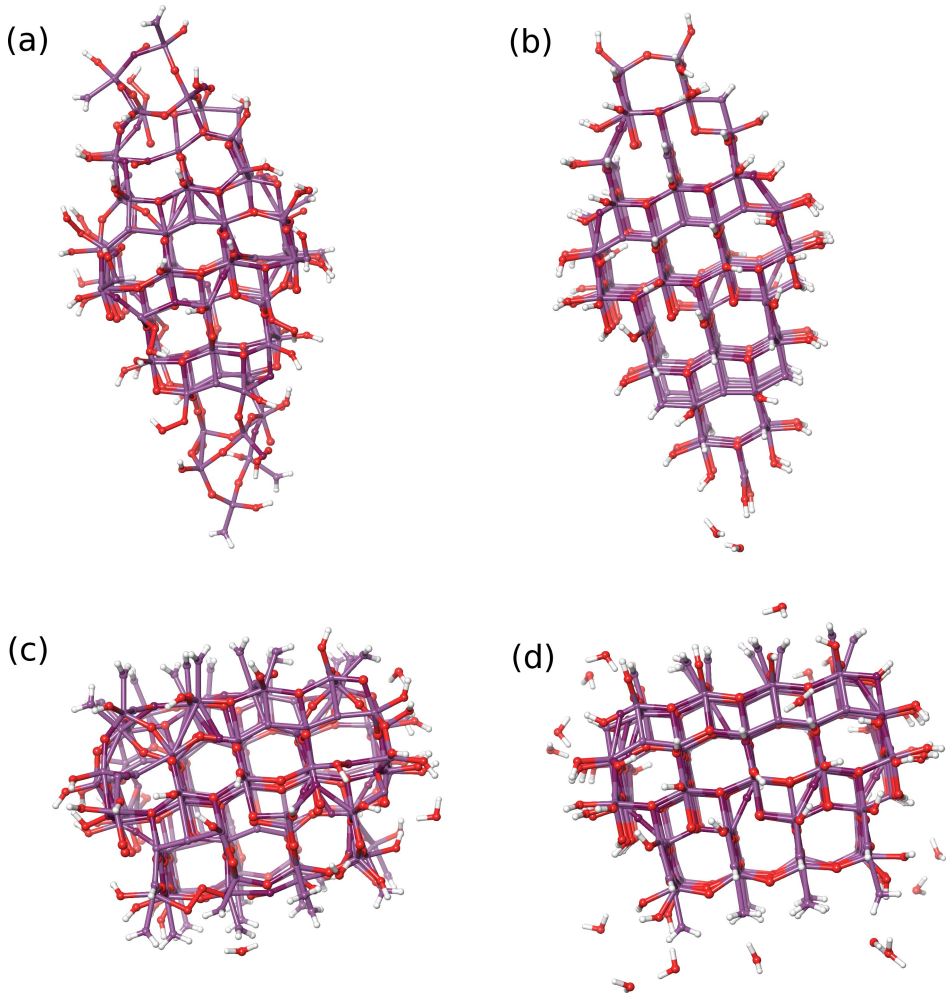

Figure 8.5: Large, neutral anatase TiO_2 nanoparticles optimized in water. (a) Optimized version of the nanoparticle in the left panel of Fig. 8.2 with the incorrect passivation. (b) Optimized version of the same nanoparticle with the correct passivation. (c) Optimized version of the nanoparticle in the right panel of Fig. 8.2 with the incorrect passivation. (d) Optimized version of the same nanoparticle with the correct passivation. The nanoparticles in Panels (b) and (d) have not yet completed optimizing, though at this point any further modifications to their geometries would be with regard to the external waters and adsorbed water ligands; the bulk of the nanoparticle is likely finished optimizing. Note that in the Maestro software used here, some bonds are rendered that do not exist and some bonds that do exist are not rendered.

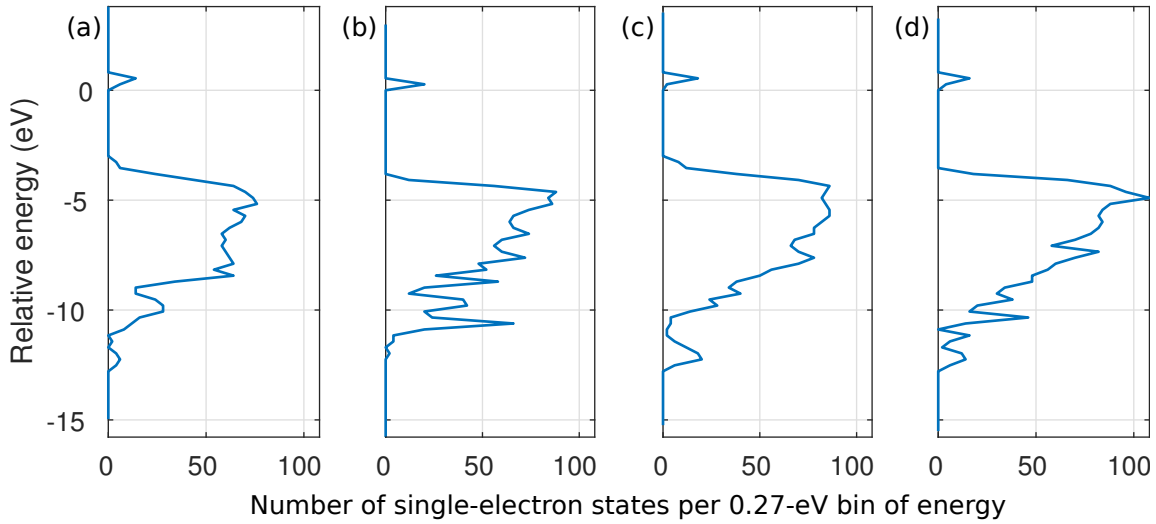

Figure 8.6: Densities of states for the nanoparticles shown in Fig. 8.5. The entire valence band is shown, though the conduction bands only display 10 spatial orbitals (20 single-electron states) because this was the number of virtual orbitals we requested be output in the Jaguar calculations. As previously discussed, the nanoparticles with the correct passivations, corresponding to Panels (b) and (d), have not yet completed optimizing in solution, though the bulk of the nanoparticle has likely completed optimizing. We do expect the orbitals to shift around slightly in energy as the nanoparticles complete their optimizations, so Panels (b) and (d) should be expected to change slightly. Note we have shifted these densities of states in energy in order to align the top of the band gap with 0 eV. The HOMO-LUMO gaps for the structures shown here are 3.61, 4.30, 3.68, and 4.05 eV, respectively; it remains to be seen as the optimizations finish and as we use better-converged wavefunctions whether the true gaps differ from 4.30 and 4.05 eV for the correctly-passivated nanoparticles. Note that the number of available electronic states increases abruptly as the orbital energy decreases from the band gap for the better-passivated nanoparticles of Panels (b) and (d) than for the incorrectly-passivated nanoparticles of (a) and (c), where we see a slightly more gradual increase in the densities of states with respect to decreasing orbital energy. This appears to be part of the reason the band gaps are larger in Panels (b) and (d) than in (a) and (c). We note that we do not expect these gaps to compare particularly well with the corresponding experimental values, as it has been shown that hybrid functionals tend to overestimate the band gap of anatase TiO_2 of 3.2 eV[129].

(c) of Fig. 8.5, which incorrectly passivates the nanoparticle, that waters generally stay adsorbed to the (001) surface, whereas Panel (d), which correctly passivates the nanoparticle, shows that many of the waters pop off upon optimizing the structure. We note that the adsorption of the water ligands of the majority (101)-exposing nanoparticle of Panels (a) and (b) is generally unaffected by the correct passivation of the oxygen atoms. The type and degree of water passivation on the (001) surface of anatase TiO_2 is debated in literature, in particular whether adsorption occurs molecularly or dissociatively; we have shown that using our methodology, which we expect would be an accurate representation of experimental findings, water molecules do not adsorb to the (001) surface in high densities, and when they do adsorb, they do so molecularly. The adsorption of water molecules on

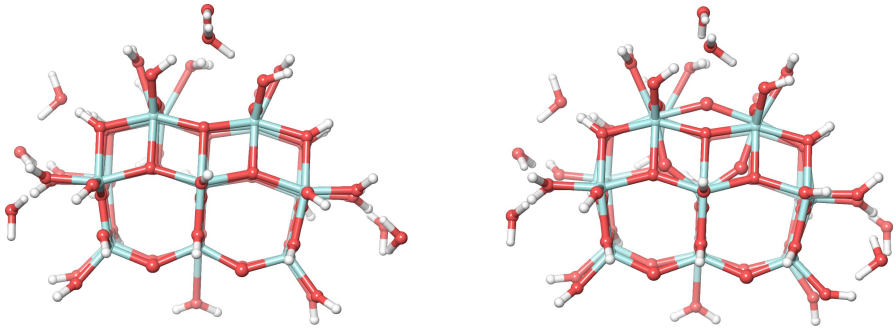

Figure 8.7: Optimized-in-water neutral (left) and cationic (right) clusters that are the focus of our PCET studies. The latter is obtained by increasing the charge of the former by one atomic unit and then optimizing the system again. We see that the subsequent optimization of the cationic cluster creates a surface reconstruction at the top-center of the nanoparticle. It remains to be determined when and why surface reconstruction occurs, and in particular whether it occurs on larger clusters.

the (101) surface of anatase TiO_2 is also debated in literature[119, 129, 140, 142, 143] and here we show that adsorption of dissociated waters, i.e., hydroxyls on the exposed Ti atoms and hydrogen atoms on the exposed O atoms, may be a viable passivation of the (101) surfaces.

8.7 Surface reconstruction on our prototype nanoparticle

Due to our setback involving the incorrect passivation of the nanoparticles mentioned in the previous section, we had to start anew, re-optimizing the larger, correctly-passivated nanoparticles while simultaneously studying the first PCET of the OER on smaller, correctly-passivated nanoparticles. The rest of this chapter reports the results we have so far obtained using a 147-atom, 21-Ti atom nanoparticle optimized in water (1,217 basis functions using the LACVP basis set), shown in Fig. 8.7 and whose properties are shown in Fig. 8.8. The left panel of Fig. 8.7 shows the optimized neutral nanoparticle and the right panel shows the optimized cationic nanoparticle, and from this view we see that surface reconstruction has taken place, in which the three-coordinated oxygen atom in the middle of the top layer of atoms has protruded directly upward, breaking the bond it had with the central Ti atom and pulling two other O atoms bonded to the Ti atom upward and inward, creating a six-sided ring of bonds in the center of the top two layers of atoms.

While this is so far the only significantly-sized, correctly-passivated, cationic nanoparticle for which we have seen surface reconstruction, at this time it is unclear how common such reconstruction events are — in particular, whether surfaces of larger cationic nanoparticles reconstruct as well. At the time of writing we are optimizing

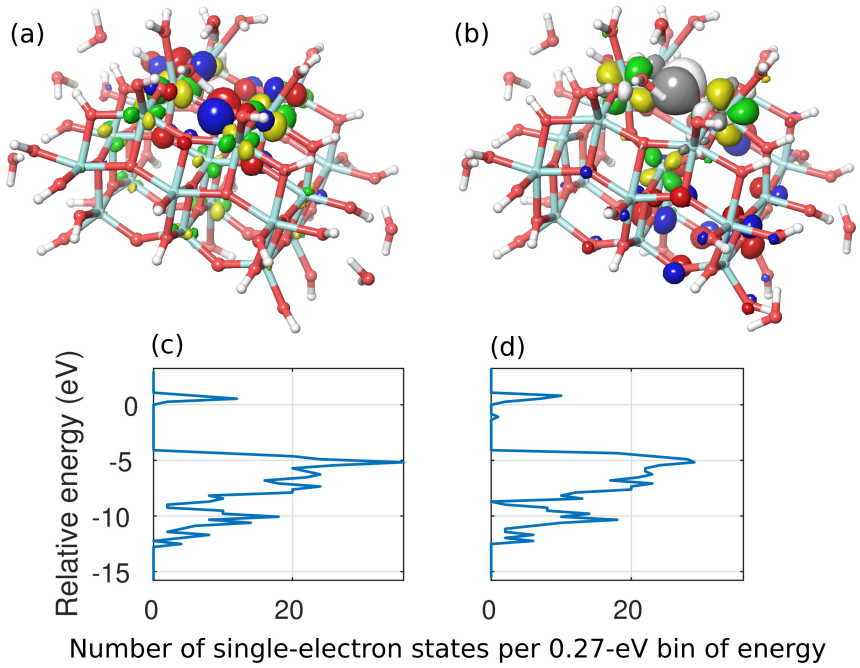

Figure 8.8: Properties of the clusters shown in Fig. 8.7. The neutral, unreconstructed nanoparticle corresponds to Panels (a) and (c), and the cationic, reconstructed nanoparticle corresponds to Panels (b) and (d). In (a) and (b) we show the surfaces of the HOMO (blue/red; O2p orbitals) and LUMO (green/yellow; Ti3d orbitals), and, for the odd-electron cationic cluster of Panel (b), the "hole" (gray/white; primarily an O2p orbital), which corresponds to the true LUMO. In (c) and (d) we show the densities of states of the two clusters, showing the full valence band, the band gap, and the 20 lowest-energy virtual spin orbitals that we outputted in our DFT calculations. In the valence band of Panel (c) we note there are 420 single-electron states and in that of Panel (d) there are 419 single-electron states, as expected because the corresponding latter cluster is forced to have one fewer electron. The energy of the "surface state" or "hole" is shown as the one single-electron state present in the band gap of Panel (d). We further note the plausibility of the location of the hole on the top-center O atom of Panel (b) because the HOMO of Panel (a) contains two electrons, and since we removed one of them, presumably from this orbital, we might expect the hole of Panel (b) to resemble the HOMO of Panel (a), as it does. The computed gap of the neutral nanoparticle is 4.62 eV and that of the cationic nanoparticle is 4.58 eV, excluding the "hole" state inside the gap, which is 3.12 eV above HOMO and 1.46 eV below LUMO.

the neutral and cationic versions of such larger nanoparticles on the Stampede supercomputer. However, surface reconstruction on TiO_2(001) nanoparticles has certainly been observed in literature[144–146], many studies of which disagree on the nature of the reconstruction, and it is interesting in this case that the reconstruction was catalyzed simply by kicking an electron off the corresponding neutral nanoparticle and optimizing the result.

What we do know, however, from the eight cationic optimizations of our prototype nanoparticle starting from slightly different geometries and using slightly different settings for the optimizations (different step sizes and different values of `cut20` and `maxit`), is that the (001) surface always reconstructs ($N = 6$) unless there are two waters adsorbed to the Ti atoms surrounding the reconstructing O3c atom and these waters are both hydrogen-bonded to nearby, unattached waters ($N = 2$), an example of which is shown in Fig. 8.9d. When even one of the nearby H-bonded waters is not present ($N = 1$), the surface resumes reconstructing. We further note that we can make no general conclusion regarding the energies of the reconstructed vs. unreconstructed nanoparticles. Relative to the average solution phase energy of the eight cationic optimizations of our prototype nanoparticle, the energies of the *N*=6 reconstructed results ranged from -0.66 eV to 0.69 eV, whereas the energies of the *N*=2 unreconstructed results were -0.05 eV and 0.04 eV. The one difference we can take away from the optimizations, however, is the location of the hole for the eight cationic optimizations. While the plot of the true LUMO in all cases shows the hole localized on the reconstructing O3c atom (whether or not reconstruction occurs), for the two unreconstructing cases, the charge on this atom (calculated using the electrostatic potential) is similar to that of the other five O3c atoms on the (001) surfaces, ranging from about -1 a.u to -1.5 a.u. In the six reconstructing cases, the charge on the reconstructing O3c atom stabilizes to around -0.6 a.u to -0.7 a.u. In other words, the nearby H-bonded water molecules in the unreconstructing cases seem to prevent localization of the $+1$ nanoparticle charge on the reconstructing O3c, though at this time it is unclear whether this is what causes the reconstruction, is caused *by* the reconstruction, or is simply correlated with the reconstruction.

To see whether the surface reconstruction was dependent on the charge on the nanoparticle, we added back an electron to our primary prototype system and let the resulting neutral system optimize once more (calling the result the "reverse-optimized" structure), finding that the surface reconstruction remained: it appeared to be irreversible, despite the fact that the reverse-optimized (still reconstructed) structure was 0.46 eV lower in energy than the original neutral unreconstructed structure. To determine the energy required to force the neutral nanoparticle to unreconstruct, we carried out a transition state job in Jaguar (Quadratic Synchronous Transit [QST] with the guess midway between the reactant and product), in which the reactant was very nearly the neutral version of

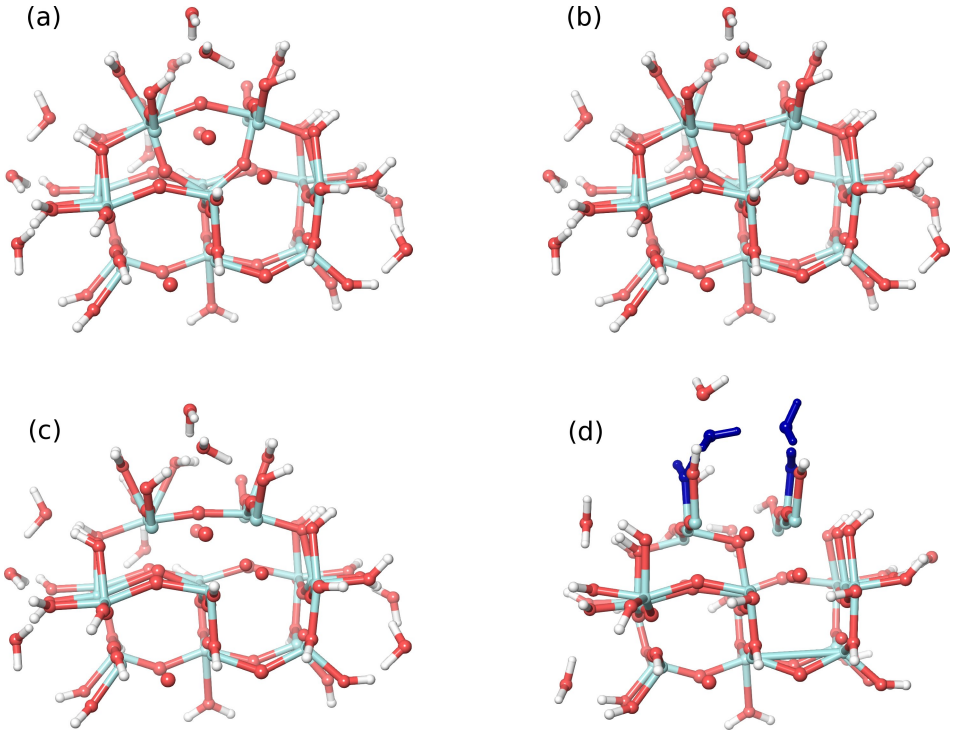

Figure 8.9: Transition state calculation (QST) attempting to force the reconstructed surface of the neutral nanoparticle to unreconstruct. An electron is initially added back to what is nearly the reconstructed cationic cluster and makes up the reactant, shown in Panel (a). The product, shown in Panel (b), is the same as the reactant except with the reconstructing oxygen atom pushed downward. The guess transition state (not shown) was automatically forced to be the average of the two geometries using the `qstinit=0.5` setting in Jaguar. The calculated transition state is shown in Panel (c), and in Panel (d) we show an example of an optimized cationic nanoparticle that does not reconstruct due to nearby water molecules hydrogen-bonding to the two adsorbed water molecules surrounding the reconstructing oxygen atom; all four of these water molecules are shown in dark blue. Note that some bonds that do exist in the system are not rendered in the images, and vice versa.

the cationic optimized cluster and the product was the unreconstructed version of this. The reactant, product, and transition state are shown in Fig. 8.9a–c. The calculated transition state is 0.20 eV higher in the energy than the original unreconstructed neutral nanoparticle and 0.67 eV higher in energy than the reverse-optimized structure. According to the room-temperature Maxwell-Boltzmann distribution, this makes the transition state about 10^3–10^4 times less likely than the original structure (and 10^{11} times less likely than the reverse-optimized structure), showing indeed that the neutral nanoparticle will never spontaneously reconstruct, even though the neutral reconstructed structure is much lower in energy. The only way for the surface to reconstruct is when the nanoparticle is $+1$-charged, when it happens spontaneously unless two nearby waters are hydrogen-bonded as described above. Finally, we noticed one more occurrence of surface reconstruction during one of the pulling

experiments described below; see Section 8.8.

In the PCET-related jobs that follow we start from the reconstructed cationic nanoparticle, but we note that an important follow-up study, in addition to using larger nanoparticles, would be to perform the same simulated experiments from the unreconstructed cationic nanoparticle, which appears to require two nearby hydrogen-bonded waters as described above. Whether this is a viable configuration can only practically be determined using molecular dynamics simulations in which the positions of many external water molecules can be optimized. Our attempt to simulate such a water environment is with the use of our implicit solvent model, though clearly this is incapable of sufficiently simulating additional explicit water molecules, which clearly can have a significant effect on both the system geometry and atomic charges. An alternative would be to perform the following experiments from the bottom of the nanoparticle, which remains unreconstructed, though this may be undesirable because the hole is located on top of the nanoparticle. Regardless, from the information at hand we see no reason that the reconstructed surface used for the following studies is invalid or even unlikely in any way.

8.8 Simulating the first PCET of the OER

Setup

In order to study the thermodynamic and kinetic contributions to the overpotential of the first PCET, we set out to construct as best as possible the potential energy curve for the entirety of the proton transfer, including both transfer of the proton from an adsorbate to a nearby water and from the resulting hydronium molecule close to the nanoparticle to the hydronium molecule "pulled" far away from the nanoparticle. We perform this study on a $+1$-charged ("cationic") nanoparticle to simulate what may actually occur in an experiment (i.e., we assume the photoexcited electron leaves the nanoparticle relatively quickly), and for reference we perform the study on a neutral nanoparticle as well. Further, we choose two adsorbed waters from which we pull off the proton (by pulling off a hydrogen atom; the charge on the pulled-off hydrogen atom was always $+1$): one adsorbed to an "edge" titanium site located at the intersection of (101) and (001) surfaces and one adsorbed to a titanium site located on the top (001) surface. We note that this latter site is directly bonded to the top reconstructing O atom mentioned in Section 8.7 and is thus closer to the primary location of the hole by two bonds. Thus, we have four distinct cluster-site combinations from which we perform our pull-off-the-H experiment, as shown in Fig. 8.10: from both the cationic and neutral nanoparticles and from both the "edge" and "surface" adsorbed waters.

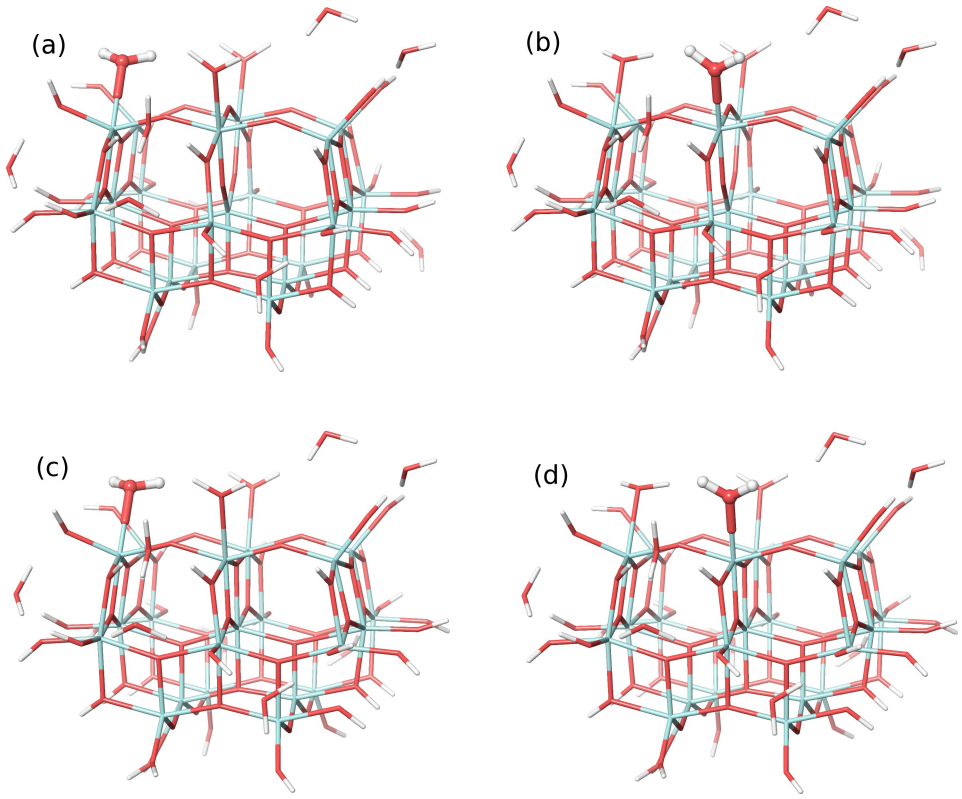

Figure 8.10: Initial nanoparticle setup for each of the four pull-off-the-H cases. For clarity, the nearby waters are not shown, but they are shown in subsequent figures. (a) Cationic nanoparticle with the "edge" adsorbed water highlighted, (b) cationic nanoparticle with the "surface" adsorbed water highlighted, (c) neutral nanoparticle with the "edge" adsorbed water highlighted, and (d) neutral nanoparticle with the "surface" adsorbed water highlighted. The H atom that transferred between the adsorbed and nearby waters was the rightmost H atom of the water depicted for the "edge" cases (cases (a) and (c)) and the leftmost H atom of the water depicted for the "surface" cases ((b) and (d)). In each case these transferring H atoms were deleted and a hydronium molecule was placed at a reasonable distance near the adsorbate, with one of the H atoms of the hydronium pointing toward the O atom of the resulting adsorbed hydroxyl. Then a geometry optimization was performed, freezing the coordinates of five Ti atoms and of the O atom of the hydronium. Then, the hydronium was displaced farther away from the nanoparticle, the six sets of coordinates were again frozen, and the system was again optimized. This was repeated for various distances from the nanoparticle to the hydronium (i.e., different values of the reaction coordinate) in order to produce a smooth potential energy curve. The direction of "pulling" of the hydronium was $\langle 001 \rangle$.

As preliminary calculations showed that the cluster/hydronium state was higher-energy than the cluster/water state, our idea was to perform, for each of the four pull-off-the-H experiments, relaxed geometry optimizations starting from the cluster/hydronium state, at a variety of distances between the cluster and the hydronium. These distances were maintained throughout each optimization by fixing the coordinates of five titanium atoms at the bottom and center of the nanoparticle (whose positions we would not expect to change much since the nanoparticle had already been optimized) and the oxygen atom of the nearby hydronium. We pulled the hydronium exactly in the $\langle 001 \rangle$ direction in each of the four cases in order to simulate the transfer of positive charge in a photoelectrochemical cell from the anode to the electrolyte. At the completion of each optimization, properties calculations were done with a tighter SCF energy convergence criterion, and the six highest-energy occupied orbitals and six lowest-energy unoccupied orbitals were generated, in addition to other surfaces such as the spin, electrostatic potential, and electron density.

Results

Snapshots of the pulling experiments in each of the four cases are shown in Figs. 8.11 to 8.14. In Fig. 8.15 we plot the optimized energy at each point along the reaction coordinate (distance between the nanoparticle and the hydronium ion) for each of the four cases. We have put each of the four curves on equal footing by defining the "distance" the same way for each case, as the length of the vector between the fixed, central Ti atom and the O atom of the hydronium projected onto the $\langle 001 \rangle$ direction. We see that generally more energy is required to pull a proton off the neutral nanoparticle than the cationic nanoparticle, and more energy is required to pull a proton off the edge water than the surface water. The properties of the potential energy surfaces are tabulated in Table 8.1. The actual proton transfer in each curve in Fig. 8.15 is identified by the steepest change in energy with reaction coordinate (i.e., the abscissa), where the nearby molecule at the smaller reaction coordinate is a water and at the larger reaction coordinate is a hydronium ion.

In all cases, just before the proton transfer, the adsorbed water "reaches" up toward the nearby water, pulling on nearby nanoparticle bonds and changing the nanoparticle geometry significantly. Once the proton transfers from the adsorbate to the nearby water and the resulting hydronium gets pulled away from the nanoparticle, one of the two O–Ti–O angles whose Ti atom the adsorbate is attached to becomes flatter (closer to $180°$) by $17.2°$, $7.3°$, $18.5°$, and $14.7°$ for each respective case. In addition, the HOMO on the nanoparticle eventually localizes on O2*p* sites near the adsorbate; the pulling-away of the hydronium seems to have a lasting effect on the location

System / adsorbate	Thermodynamic energy (eV)	Kinetic energy (eV)	Barrier height (eV)
Cationic / edge	1.15	1.64	0.49
Cationic / surface	1.17	1.44	0.27
Neutral / edge	1.17	1.85	0.68
Neutral / surface	1.20	1.69	0.49

Table 8.1: Properties of the potential energy curves shown in Fig. 8.15. Thermodynamic energy refers to the difference between the two extremes of the reaction coordinate. Kinetic energy refers to the difference between the minimum and maximum energies. The barrier height is the difference between the kinetic and thermodynamic energies, i.e., the difference between the maximum energy and the final energy along the reaction coordinate.

of the HOMOs. In general, the energy of the composite system initially (after the proton jumps to the nearby water) increases as the hydronium is pulled away: the hydronium generally still has an effect on the nanoparticle until it is pulled farther away; it is not as if the system simply equilibrates once the proton transfers to the nearby water. This shows that it is important to remember to study the second part of the PCET in which the hydronium is pulled from nearby to far away, not just the first part in which the proton transfers from the adsorbate to the nearby water. Finally, it is interesting to note that the thermodynamic energy of all four configurations at the maximum distances is the same: 1.15–1.20 eV.

In the two cationic cases, the location of the hole remains the same throughout the entire process, in the $2p$ orbital of the top-center, originally three-coordinated oxygen atom (as also found in [122], though they were studying the (101) surface). In addition, from Fig. 8.15 we see that the distance at which the proton transfers to the nearby water molecule is larger for the cationic nanoparticles than for the neutral nanoparticles by about 0.3 Ang. Further, in the cationic cases we see that less energy is required overall for the proton transfer by about 0.2 eV.

In general, the Ti–O bond at an adsorbed hydroxyl is smaller than that at an adsorbed water, so after the proton is transferred as the reaction coordinate increases, we would expect the Ti–O bond to shorten. This is true for the surface adsorbate cases, but we note that for the edge adsorbate cases, the Ti–O bond starts shortening even before the proton transfers to the nearby water. Further, while for the surface adsorbate after the proton transfer the Ti site of the adsorbed hydroxyl smoothly settles back to near its initial position and the Ti–O bond shortens gradually, for the edge adsorbate the Ti site snaps back simultaneously with the proton transfer and the Ti–O bond shortens immediately. In addition, for the edge adsorbate cases, the LUMOs remain in their initial positions after the proton transfer, whereas in the surface adsorbate cases, the LUMOs change locations after the

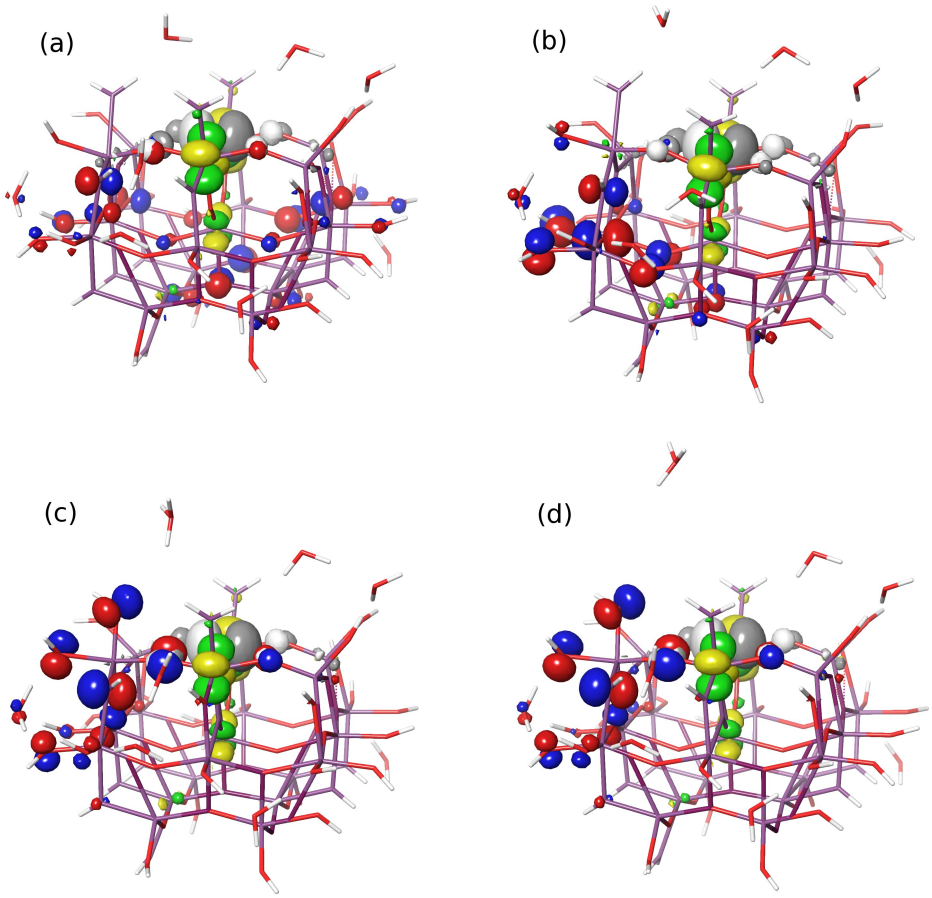

Figure 8.11: Four selected snapshots (out of 16) of the pulling experiment for the cationic nanoparticle + edge adsorbate case. The nearby water/hydronium is the top external molecule shown in each snapshot. In snapshots (a) and (b), we see that the initial hydronium placement is so close to the adsorbed hydroxyl that one of the hydronium's H atoms gets transferred to the adsorbate upon optimizing the geometry of the system. In snapshots (c) and (d), we see that the initial hydronium placement is far enough away from the nanoparticle that upon optimization the hydronium molecule retains all three of its H atoms. Snapshots (b) and (c) bookend the proton transfer, and so we see from snapshot (b) that right before (in terms of the reaction coordinate) the proton transfer, the H atom pulls the entire adsorbed water, nearby bonds, and single-particle HOMO (blue/red surfaces) towards the nearby water; the geometry changes significantly. We see from snapshot (c) that right after the proton transfer, the adsorbate snaps back toward the nanoparticle, and from (d) we see that after starting out (as in (a)) more delocalized, the HOMO remains around the O atom of the adsorbed hydroxyl as the hydronium molecule is pulled away from the nanoparticle. Throughout the process, the single-particle "hole" (gray/white surfaces; this is the true LUMO) remains localized on a $2p$ orbital of the primary reconstructing O atom, and the single-particle LUMO (green/yellow surfaces; this is the true LUMO+1 orbital) remains primarily on the nearby Ti3*d* orbitals.

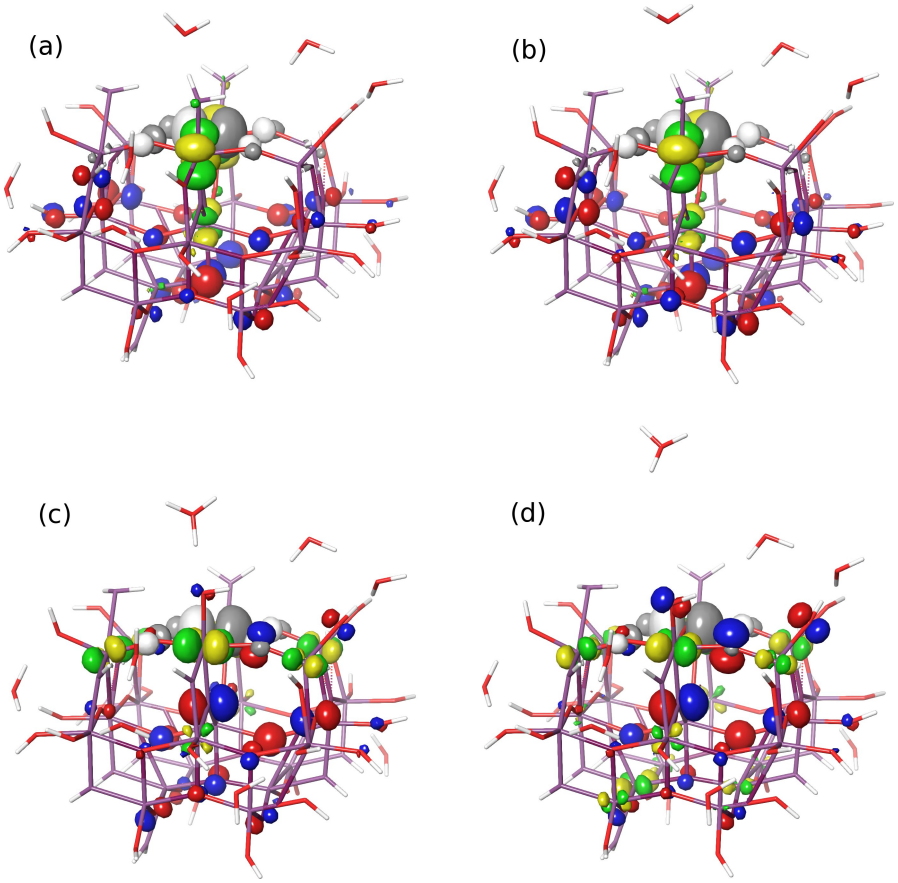

Figure 8.12: Four selected snapshots (out of 17) of the pulling experiment for the cationic nanoparticle + surface adsorbate case. Again, snapshots (b) and (c) bookend the proton transfer. Now the LUMO changes once the proton is transferred and the adsorbate does not "snap" back until the hydronium is pulled farther away; it is more of a gradual settling-back.

proton transfer.

As found in [122], the order of the proton and electron transfer is thought to have an effect on the activation barrier, so we briefly studied this, taking the electron transfer to be indicated by the charge on the oxygen atom of the adsorbate from which the proton is transferred. We show in Fig. 8.16 that for pulling the protons from the surface adsorbates, the proton transfers well before the charge on the adsorbate O atom increases, indicating perhaps that the proton transfer long precedes the electron transfer. In contrast, for the edge cases, which have higher activation barriers than the surface cases, we note that the electron and proton transfers occur simultaneously. Thus, although often concertedness of the electron and proton transfers of a PCET results in a smaller activation barrier, here we find that the opposite is true.

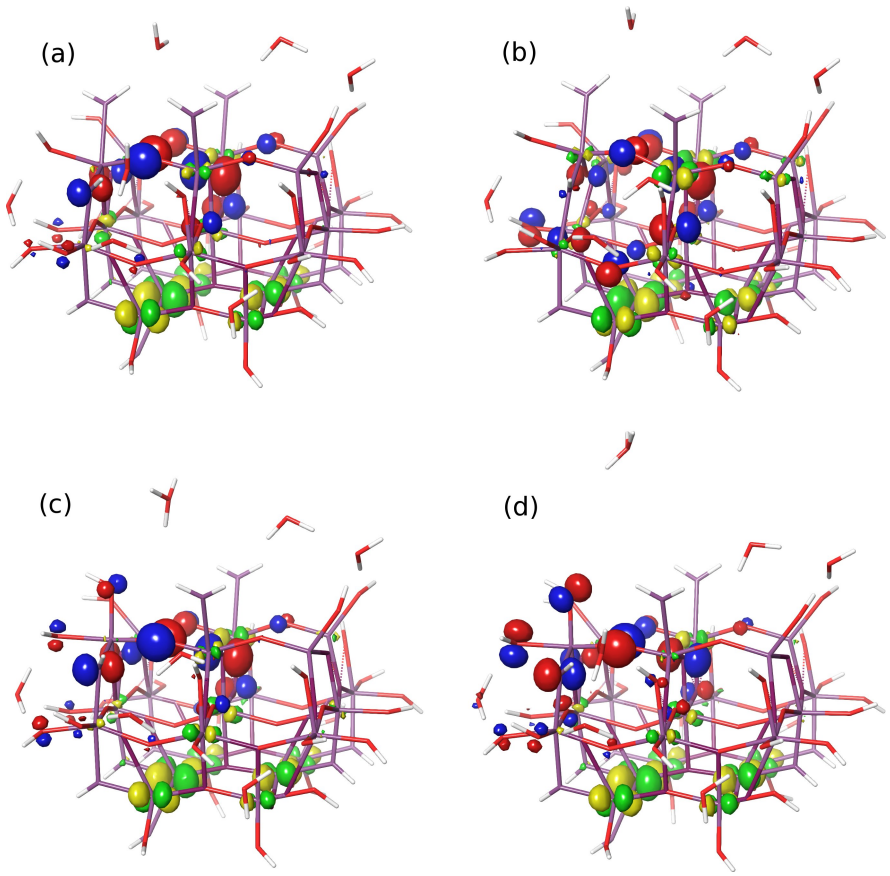

Figure 8.13: Four selected snapshots (out of 15) of the pulling experiment for the neutral nanoparticle + edge adsorbate case. Again, snapshots (b) and (c) bookend the proton transfer. Everything is the same as for the cationic case, except that now the LUMO is located at the bottom of the nanoparticle.

We note that as the hydronium is pulled away in both surface adsorbate cases, the edge adsorbed water often pops off and on again, during which the hydronium rotates around, its original O–H bond not always pointing toward the surface adsorbate. As soon as the edge water adsorbs back on the nanoparticle, the hydronium's bonds realign. This is to be contrasted with how tightly the Ti atom seems to hold onto the edge water when the pulling is done from this water: the presence of the proton on the surface water seems to have an effect three bonds away, a somewhat delocalized effect. We saw this effect too in preliminary optimizations and transition state calculations. Since quite a bit of energy is required to pull the proton off the edge water when the surface water stays intact, we posit that if the surface water became a hydroxyl, then pulling a proton from the edge water may well instead dislocate the entire edge water rather than just the proton, since the edge Ti–O bond seems to be weaker in this case. This may have an effect on the OER as a whole, suggesting perhaps that in order

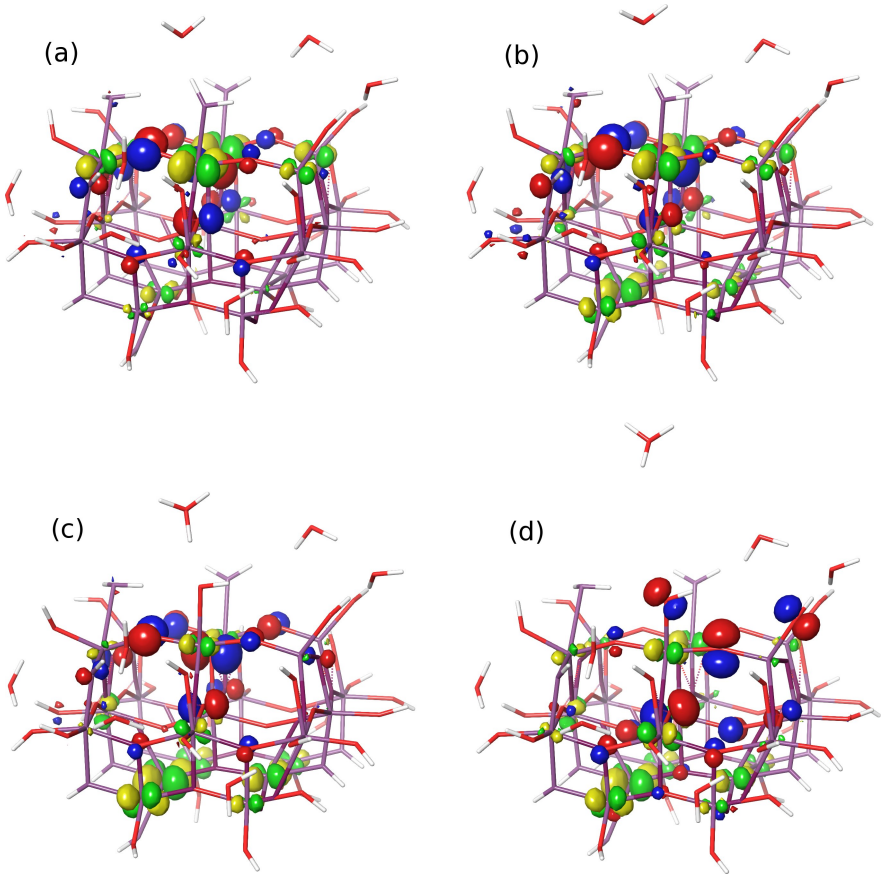

Figure 8.14: Four selected snapshots (out of 18) of the pulling experiment for the neutral nanoparticle + surface adsorbate case. Again, snapshots (b) and (c) bookend the proton transfer. Everything is the same as for the cationic case, except that now the LUMO stays more delocalized after the proton is transferred.

for the first PCET to occur at all, nearby adsorbates must remain as waters rather than hydroxyls. One of the four covalent bonds of the Ti holding on to the edge water is attached to a passivating hydroxyl, whereas the Ti atom holding onto the surface water has all four of its covalent bonds attached to bulk oxygens, so perhaps this somewhat delocalized effect only occurs for this sort of relative geometry. The effect of the water desorbing and adsorbing again can be seen in the curves of Fig. 8.15, in which the two surface adsorbate cases show jaggedness around the peaks of the curves, corresponding to the edge water popping off (black lines) and on again (green lines). The curves may be best thought of by ignoring the interior regions between each pair of black/green lines. However, as this does not affect the absolute maximum value of the plot, the properties of the curves, listed in Table 8.1, are unaffected.

Interestingly, we note that for pulling from the surface adsorbate of the neutral nanoparticle, at one of the

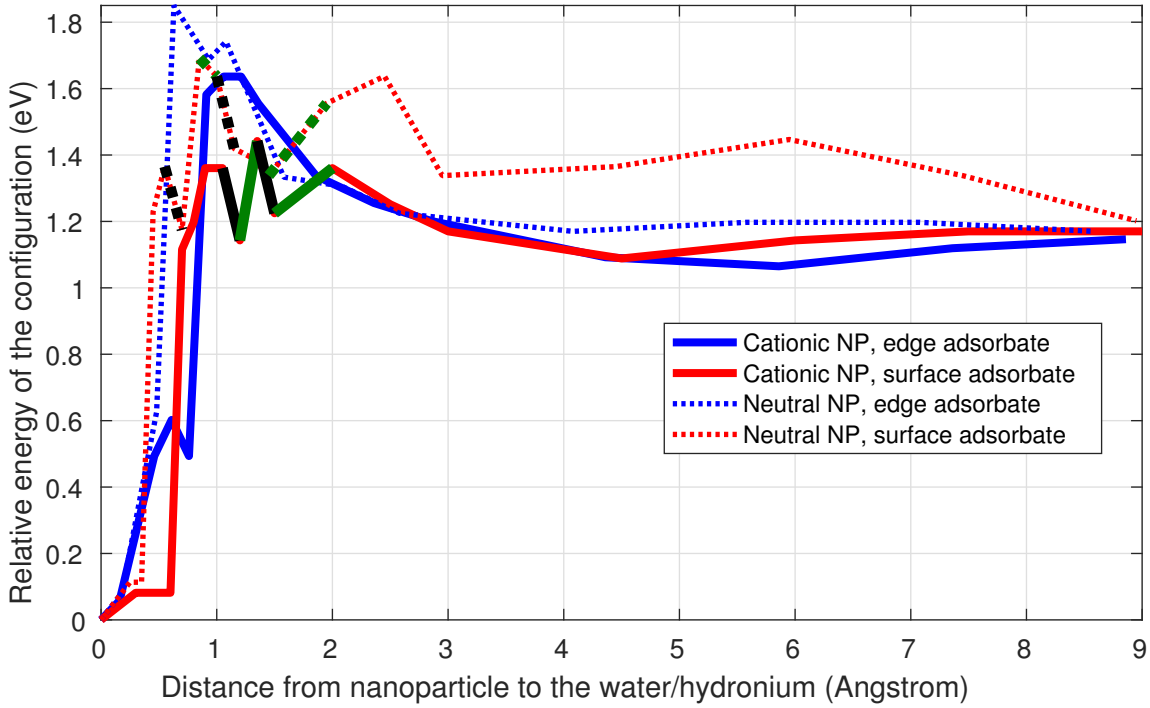

Figure 8.15: Solution phase energy as a function of the reaction coordinate, which is the distance between the water/hydronium oxygen atom and the lower bulk of the nanoparticle. One can tell from the curves at which distances the nearby molecule is a water or a hydronium: on the left side of the steepest increase between datapoints the molecule is a water; on the right side it is a hydronium. Black pathways indicate that a water pops off between the datapoints and green pathways indicate that the water re-adsorbs; this wavering water is the "edge" water.

distances soon after the proton transfer, when the edge water pops off it becomes H-bonded to the opposite bridging O2c (in addition to a nearby adsorbed hydroxyl, to which it was always H-bonded). This appears to be the cause of some surface reconstruction, though the optimization completes before the surface finishes reconstructing. We noted such effects the H bonds at the surface have on the surface reconstruction in Section 8.7. An effect of this can be seen in the bottom-right plot of Fig. 8.16, in which the charge on the reconstructing oxygen atom temporarily increases to the same as on the adsorbate's O atom. This optimization was run on our local cluster, and we further note that upon running the exact same input file on Stampede with our optimized version of Jaguar, we did not find such surface reconstruction, as the edge water popped off but never H-bonded to the O2c. Tangentially, we ran two other sub-jobs for the pulling experiments on both Stampede and our local cluster (which use different versions of Jaguar), and while one achieved very similar results, another (pulling from the edge adsorbate of the cationic nanoparticle) determined the proton to be on the nanoparticle instead of the

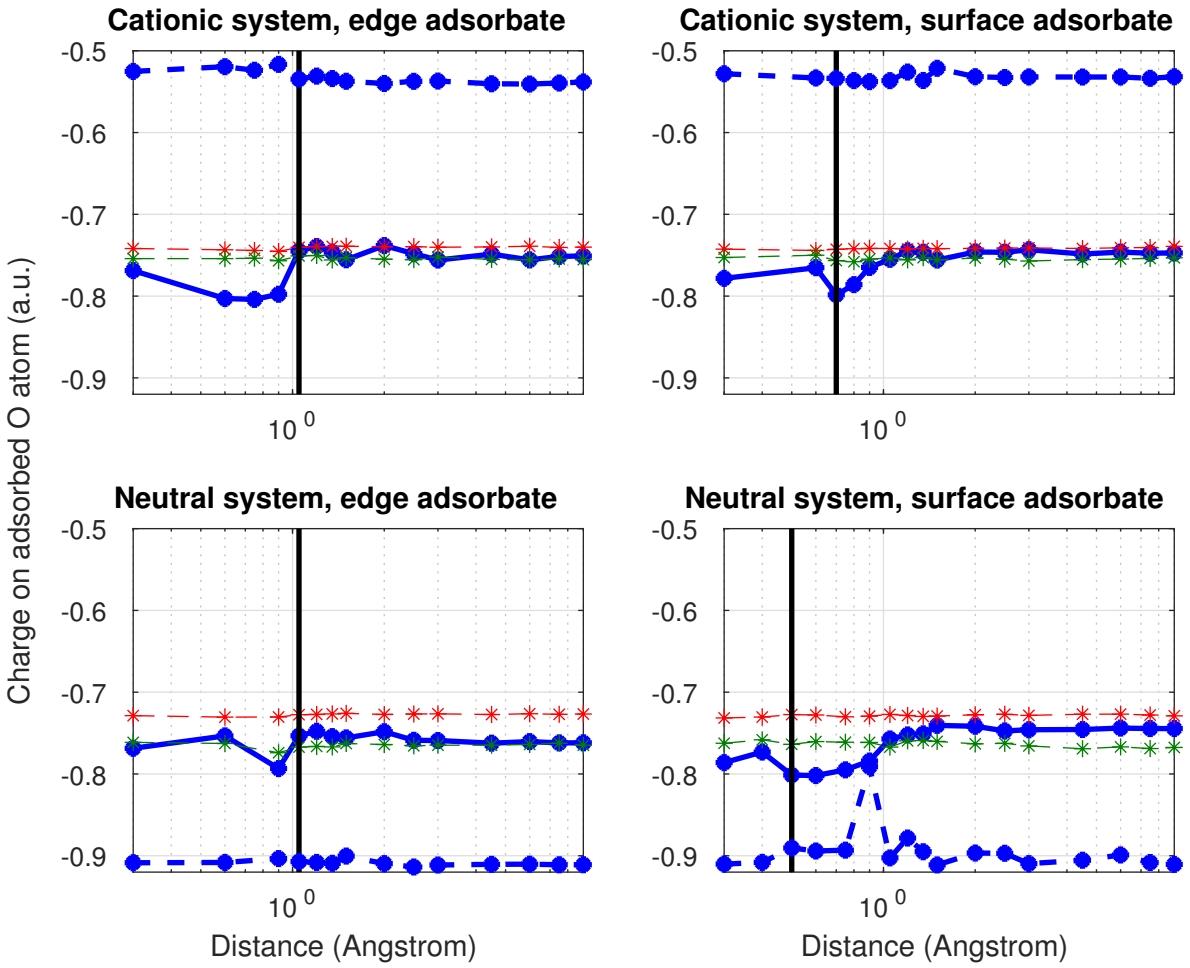

Figure 8.16: Concertedness of the electron and proton transfers in each of the four pulling experiments. The pulling distances are on the x-axes, and the y-axes show charge in atomic units as calculated by a Mulliken population analysis. Blue datapoints (solid line) indicate the charge on the oxygen atom of the adsorbate from which the proton transfers. As reference values, red datapoints indicate the charge on the O atom of an adsorbed water on the bottom of the nanoparticle, and green datapoints indicate the charge on the O atom of an adsorbed hydroxyl on the bottom of the nanoparticle. Dashed blue datapoints indicate the charge on the reconstructing oxygen atom. Vertical black lines indicate the distance at which the proton transfers to the nearby water molecule.

nearby water, changing which sub-jobs bookended the proton transfer. Thus, together with the reconstruction results of Section 8.7, we can say our results are sometimes quite sensitive to the precise system geometries and calculation methods used.

Discussion

As discussed in the Introduction, the total energy required for each pulling experiment represents the anodic overpotential in our model photoelectrochemical (PEC) cell due to the instability of the intermediate species, *OH, adsorbed on the surface of the nanoparticle catalyst. Thus, from the cationic results above, we see that the thermodynamic overpotentials are about 1.16 eV and that the total overpotentials, including the kinetic components, are about 1.44–1.64 eV, corresponding to kinetic contributions (i.e., activation energies) of 0.27–0.49 eV. In literature[119, 120, 122, 147, 148], typical thermodynamic overpotentials are found to be 0.7–0.8 eV, total overpotentials 0.3–1.1 eV, and kinetic contributions 0.2–0.5 eV. (We note that two of these sources[147, 148] report typical total experimental overpotentials, which include additional important contributions to the overpotential such as electron-hole recombination and hole transfer[149], whereas the other sources[119, 120, 122] report simulated overpotentials due primarily to the instability of the intermediate species.) Using these data as comparisons, we see we have obtained kinetic overpotentials in agreement with literature, whereas our thermodynamic overpotentials are too high. However, given that at least one similar study[124] on one-titanium clusters yields a thermodynamic overpotential of 3.43 eV, we expect that as we study larger clusters, the thermodynamic overpotential may decrease. We thus conclude that so far, our model reasonably predicts both thermodynamic and kinetic contributions to the anodic overpotential. This potentially validates our particularly fast and accurate methodologies consisting of a combination of hybrid DFT, explicit passivation and an implicit solvation model, and the pseudospectral method for realistically simulating a crucial piece of an economically important PEC system. As this study is still underway, particularly on larger clusters, we hope to obtain more reasonable thermodynamic overpotentials in the near future to confirm this and allow for additional simulations to help push the field forward.

We hypothesize two potential mechanisms for the kinetic contribution to the overpotential, which is clearly significant. The first is that this activation barrier exists because as the proton transfers from the adsorbate to the nearby water, the electrons that remain near the reaction site on the nanoparticle electrostatically resist this transfer by attracting the proton. This is exemplified by what physically appears to be the nearby water pulling so hard on the proton that the nanoparticle geometry distorts significantly, as seen particularly in the edge cases

shown in Figs. 8.11 and 8.13. As the proton is pulled from the edge adsorbate of the cationic nanoparticle, as shown in Fig. 8.11, the HOMO quickly migrates toward the reaction site after initially being delocalized. However, as the proton is pulled from the surface adsorbate of the cationic nanoparticle, as shown in Fig. 8.12, the negative charge migrates toward the reaction site more slowly, accumulating less negative charge near the active site, and therefore pulling less strongly on the proton being transferred. We may attribute the relatively diminished amount of negative charge near the active site to the presence of the hole on the primary reconstructing oxygen atom, which is closer to the active site in the surface case. The hole represents an O2p site that is too high in energy (3.12 eV above the HOMO) for the electrons to occupy, despite the fact that we are already putting in energy (but only $\sim$1.44 eV as shown in Table 8.1) to make the proton transfer in the first place. This hole site is one of the closest O2p sites for the negative charge to occupy in the surface case, and since it is too high-energy to be occupied, overall there is less negative charge near the active site and therefore less of an electrostatic attraction to the transferring proton from the surface adsorbate. Conversely, for transfer from the edge adsorbate, there is no hole nearby and the highest-energy electrons can occupy all of the O2p sites near the active site, causing a larger buildup of negative charge around the transferring proton and thus requiring more energy for the proton transfer to occur.

The fact that we see the same sort of potential energy curves, as shown in Fig. 8.15, for pulling from the neutral nanoparticle as for the cationic nanoparticle, at first seems to weaken the argument just made regarding the buildup of charge near the active site; in the neutral case the O2p site is occupied so we do not expect particularly less negative charge buildup around the active site in the surface case as for the edge case. However, we note that in the neutral case the surface is unreconstructed and therefore the now-occupied O2p site is *farther away* from the proton being transferred. This still results in there being less negative charge near the transferring proton in the surface case than in the edge case, in which once again the HOMO migrates to all O2p sites near the proton transfer, perhaps implying a *closer* buildup of negative charge in the edge case than in the surface case.

Of course, there could just be more negative charge on all the oxygen atoms near the edge adsorbate than near the surface adsorbate. We would need to do a more detailed analysis of the charge, for example using Mulliken population analysis, in order to quantify this. Regardless of the particular details, it makes sense that the activation barrier could be caused by a buildup of negative charge around the reaction site that opposes the proton transfer, and that the role of the hole may be as a forbidden site for the most mobile electrons, causing

relatively less charge buildup around any nearby active sites than if the hole were not present.

A second possible mechanism for the kinetic contribution to the overpotential could be related to the apparent geometric correlation in the nanoparticle due to its electronic structure. We have so far encountered two particular examples of changes in the geometry of the nanoparticle having an effect on the geometry relatively far (at least a couple bond lengths) away: (1) the fact that passivation of the exposed oxygen atoms by hydrogen atoms causes both more overall stability in the nanoparticle and the (001) surfaces of the nanoparticle to hold on much less strongly to any adsorbed waters (compare for example Panels (c) and (d) of Fig. 8.5), and (2) the fact that once the proton is transferred from the surface adsorbate to a nearby water, the edge water appears to be very loosely bound to the nanoparticle and frequently detaches, seen in the pull-off-the-H experiments when pulling was done from the surface water, for both the cationic and neutral cases. We thus feel it is plausible that, as long as the nanoparticle is passivated correctly and is therefore electronically and geometrically "correlated," the fact that there is a larger activation barrier for the edge adsorbate than for the surface adsorbate is somehow due to these non-local effects.

Our next steps for this investigation start with performing the same "pulling" experiments on larger nanoparticles, which are currently underway. Along the way we will learn whether reconstruction occurs on larger clusters or is perhaps limited to medium-sized clusters such as the one presented here. If we find surface reconstruction to persist, we may investigate further how the results change when performing the experiments at and away from reconstructed parts of the surface. Similarly, as H bonds are known to affect surface phenomena, as we have found in Section 8.7 and possibly here, we could study how such bonds affect our results; for example, we could perform a pulling experiment using an H atom that is H bonded to a nearby water to this very nearby water. Further, since we have found here that pulling from a surface water requires less energy than pulling from an edge water, studying larger clusters will allow us to further isolate any particular edge effects by pulling from the center of a surface, which is farther away from the edges[150]. In addition, some[119] have found that the local surface structure of TiO_2 nanoparticles has little bearing on the overpotential, instead depending more on the position of the valence band maximum, and since we are working on larger clusters exposing the (101) surface, we will be able to gauge how much difference the particular facet makes on the overpotential. Finally, we may attempt to calculate reaction rates, which is possible here because, as discussed in the Introduction, the first PCET is generally thought to be the rate-limiting step of the OER. Along the way, if we continue to study neutral clusters as reference points, we may run unrestricted calculations to confirm whether this makes a difference in the results,

since essentially unrestricted calculations allow more electronic degrees of freedom to be sampled. Our work may be further aided by understanding how the nuclear charges and bond lengths change as the H^+/H_3O^+ is pulled away from the surface.

Assuming we are able to continue validating our methodologies by comparing our results to experiment and other simulations, there are many important aspects to the problem that are ripe with potential for investigation using our methodologies. For example, pH affects the implicit and explicit solvations and the PCET steps themselves, and we will be able to modify all these variables within our model accordingly. While we have implicitly studied the degree of surface coverage of water ligands on our cluster by realistically optimizing the system starting from a completely passivated surface, we did not allow for different species to passivate each surface. For example, although we determined that H_2O passivates the (001) surface and that OH passivates the (101) surface on physical grounds and therefore started with surface passivations by those species, we did not allow for OH to passivate the (001) or H_2O to passivate the (101) surfaces, as has been found in some studies to be the case[151, 152]. Subsequent study of the PCET on those clusters would help scientists understand how to best prepare the anode of a PEC cell. We could further study the other three PCETs of the OER to confirm that indeed the first PCET is the rate-limiting step. In addition, we could perform the same types of investigations on rutile, as there has been at least one instance[129] in which the catalytic superiority of anatase for the use in PEC cells has been called into question. Further, the true oxidation power of the holes in a PEC cell depends on the quasi-Fermi level (electrochemical potential) for the holes[153] (and in general on the hole kinetics[149]), which changes with illumination intensity, and studying this could be a way to understand a water splitting device at an even more realistic level.

Finally, another more realistic study could be that in which we have not assumed the photoexcited electrons leave the nanoparticle relatively quickly, perhaps instead stabilizing in trap states at the surface. It is known that the true functioning of a PEC cell consists of a non-equilibrium steady state of electrons and holes in the anode[154], and we could pursue this more computationally challenging, but again more realistic, scenario, determining the resulting overpotential in order to make contact with measurable experimental quantities. On the other hand, an important and somewhat more accessible, yet still challenging, factor could be the effects that surface defects, such as oxygen vacancies and steps, have on the overpotential. In addition, we could study the edge effects more realistically (on larger nanoparticles) and in more detail than we have here, which could be particularly illuminating because often catalytic activity is found to increase at such defects[127, 140], contrary to

the results of our basic study of edge effects above. Regardless, in the end, any of these phenomena significantly affects the functioning of a PEC cell, and its understanding is therefore crucial in order to make a promising type of solar cell physically and economically feasible. Thus, in addition to our insight into the overpotential of the OER, our work here has shown, and will continue to show, how such phenomena can be efficiently and robustly simulated using first principles techniques.

Bibliography

[1] D. A. Long, *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules* (John Wiley and Sons, 2002).

[2] M. M. Sushchinskii, *Raman Spectra of Molecules and Crystals* (Israel Program for Scientific Translations, 1972).

[3] D. H. Goldstein, *Polarized Light*, 3rd ed. (CRC Press, 2010).

[4] J. D. Jackson, *Classical electrodynamics* (Wiley, 1999).

[5] G. Placzek (Akademische Verlagsgesellschaft, Leipzig, 1934) pp. 205–374.

[6] D. A. Long, Int. Rev. Phys. Chem. **7**, 317 (1988).

[7] M. Halls and H. Schlegel, The Journal of chemical physics **111**, 8819 (1999).

[8] E. E. Zvereva, A. R. Shagidullin, and S. A. Katsyuba, The Journal of Physical Chemistry A **115**, 63 (2011).

[9] A. D. Bochevarov, E. Harder, T. F. Hughes, J. R. Greenwood, D. A. Braden, D. M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang, and R. A. Friesner, Int. J. Quantum Chem. **113**, 2110 (2013).

[10] Y. Cao and R. A. Friesner, The Journal of Chemical Physics **122**, 104102 (2005).

[11] R. A. Friesner, The Journal of Chemical Physics **86**, 3522 (1987).

[12] R. A. Friesner, J. Phys. Chem. **92**, 3091 (1988).

[13] R. A. Friesner, Chemical Physics Letters **116**, 39 (1985).

[14] R. A. Friesner, The Journal of Chemical Physics **85**, 1462 (1986).

[15] J.-M. Langlois, R. P. Muller, T. R. Coley, W. A. Goddard, M. N. Ringnalda, Y. Won, and R. A. Friesner, The Journal of Chemical Physics **92**, 7488 (1990).

[16] M. N. Ringnalda, Y. Won, and R. A. Friesner, The Journal of Chemical Physics **92**, 1163 (1990).

[17] M. N. Ringnalda, M. Belhadj, and R. A. Friesner, The Journal of Chemical Physics **93**, 3397 (1990).

[18] Y. Won, J.-G. Lee, M. N. Ringnalda, and R. A. Friesner, The Journal of Chemical Physics **94**, 8152 (1991).

[19] A. Szabó and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory* (Courier Dover Publications, 1996).

[20] L. Jensen and G. C. Schatz, The Journal of Physical Chemistry A **110**, 5973 (2006).

[21] D. Rappoport and F. Furche, Journal of Chemical Physics **133**, 134105 (2010).

[22] J. H. Park, J.-I. Park, D. H. Kim, J.-H. Kim, J. S. Kim, J. H. Lee, M. Sim, S. Y. Lee, and K. Cho, J. Mater. Chem. **20**, 5860 (2010).

[23] J. H. Schön, C. Kloc, and B. Batlogg, Synthetic Metals **115**, 75 (2000).

[24] R. Joseph Kline, M. D. McGehee, and M. F. Toney, Nat Mater **5**, 222 (2006).

[25] H. Sirringhaus, P. J. Brown, R. H. Friend, M. M. Nielsen, K. Bechgaard, B. M. W. Langeveld-Voss, A. J. H. Spiering, R. a. J. Janssen, E. W. Meijer, P. Herwig, and D. M. de Leeuw, Nature **401**, 685 (1999).

[26] F. Rubio-Marcos, M. A. Bañares, J. J. Romero, and J. F. Fernandez, J. Raman Spectrosc. **42**, 639 (2011).

[27] M. Castriota, A. Fasanella, E. Cazzanelli, L. De Sio, R. Caputo, and C. Umeton, Optics Express **19**, 10494 (2011).

[28] E. Venuti, R. G. Della Valle, L. Farina, A. Brillante, M. Masino, and A. Girlando, Phys. Rev. B **70**, 104106 (2004).

[29] M. Huang, H. Yan, C. Chen, D. Song, T. F. Heinz, and J. Hone, PNAS **106**, 7304 (2009).

[30] Z. Q. Lu, T. Quinn, and H. S. Reehal, Journal of Applied Physics **97**, 033512 (2005).

[31] P. Ranzieri, A. Girlando, S. Tavazzi, M. Campione, L. Raimondo, I. Bilotti, A. Brillante, D. Valle, R. G, and E. Venuti, ChemPhysChem **10**, 657 (2009).

[32] A. Brillante, I. Bilotti, F. Biscarini, R. G. Della Valle, and E. Venuti, Chemical Physics **328**, 125 (2006).

[33] J. M. Winfield, C. L. Donley, R. H. Friend, and J.-S. Kim, Journal of Applied Physics **107**, 024902 (2010).

[34] A. Brillante, R. G. Della Valle, L. Farina, A. Girlando, M. Masino, and E. Venuti, Chemical Physics Letters **357**, 32 (2002).

[35] T. V. Basova and B. A. Kolesov, Thin Solid Films **325**, 140 (1998).

[36] A. Crisci, F. Baillet, M. Mermoux, G. Bogdan, M. Nesládek, and K. Haenen, Phys. Status Solidi A **208**, 2038 (2011).

[37] D. T. James, B. K. C. Kjellander, W. T. T. Smaal, G. H. Gelinck, C. Combe, I. McCulloch, R. Wilson, J. H. Burroughes, D. D. C. Bradley, and J.-S. Kim, ACS Nano **5**, 9824 (2011).

[38] L. Sicot, C. Fiorini, A. Lorin, P. Raimond, C. Sentein, and J.-M. Nunzi, Solar Energy Materials and Solar Cells **63**, 49 (2000).

[39] M. Berggren, O. Inganäs, G. Gustafsson, M. R. Andersson, T. Hjertberg, and O. Wennerström, Synthetic Metals **71**, 2185 (1995).

[40] D. Braun, G. Gustafsson, D. McBranch, and A. J. Heeger, Journal of Applied Physics **72**, 564 (1992).

[41] A. Tsumura, H. Koezuka, and T. Ando, Applied Physics Letters **49**, 1210 (1986).

[42] S. Glenis, G. Horowitz, G. Tourillon, and F. Garnier, Thin Solid Films **111**, 93 (1984).

[43] J. C. Heckel, A. L. Weisman, S. T. Schneebeli, M. L. Hall, L. J. Sherry, S. M. Stranahan, K. H. DuBay, R. A. Friesner, and K. A. Willets, J. Phys. Chem. A **116**, 6804 (2012).

[44] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).

[45] G. Rauhut and P. Pulay, J. Phys. Chem. **99**, 3093 (1995).

[46] T. Sundius, Vibrational Spectroscopy **29**, 89 (2002).

[47] K. Szczepaniak, M. M. Szczesniak, and W. B. Person, J. Phys. Chem. A **104**, 3852 (2000).

[48] D. Feller, Journal of Computational Chemistry **17**, 1571 (1996).

[49] A. J. Sadlej, Theor. Chem. Acc. **79**, 123 (1991).

[50] A. J. Sadlej, ChemPlusChem **53**, 1995 (1988).

[51] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T. L. Windus, J. Chem. Inf. Model. **47**, 1045 (2007).

[52] G. Louarn, J. P. Buisson, S. Lefrant, and D. Fichou, J. Phys. Chem. **99**, 11399 (1995).

[53] E. Agosti, M. Rivola, V. Hernandez, M. Del Zoppo, and G. Zerbi, Synthetic Metals **100**, 101 (1999).

[54] S. Millefiori, A. Alparone, and A. Millefiori, Journal of Heterocyclic Chemistry, Journal of Heterocyclic Chemistry **37, 37**, 847, 847 (2000).

[55] G. J. Visser, G. J. Heeres, J. Wolters, and A. Vos, Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry **24**, 467 (1968).

[56] T. Siegrist, R. Fleming, R. Haddon, R. Laudise, A. Lovinger, H. Katz, P. Bridenbaugh, and D. Davis, Journal of Materials Research **10**, 2170 (1995).

[57] D. Fichou, B. Bachet, F. Demanze, I. Billy, G. Horowitz, and F. Garnier, Adv. Mater. **8**, 500 (1996).

[58] E. J. Ambrose, A. Elliott, and R. B. Temple, Proc. R. Soc. Lond. A **206**, 192 (1951).

[59] T. R. Gilson and P. Hendra, *Laser Raman Spectroscopy* (John Wiley & Sons Ltd, 1970).

[60] G. Turrell, *Infrared and Raman Spectra of Crystals* (Academic Pr, 1972).

[61] G. C. Pimentel and A. L. McClellan, J. Chem. Phys. **20**, 270 (1952).

[62] P. Hermet, N. Izard, A. Rahmani, and P. Ghosez, J. Phys. Chem. B **110**, 24869 (2006).

[63] M. T. Zanni, S. Gnanakaran, J. Stenger, and R. M. Hochstrasser, J. Phys. Chem. B **105**, 6520 (2001).

[64] G. Bounos, S. Ghosh, A. K. Lee, K. N. Plunkett, K. H. DuBay, J. C. Bolinger, R. Zhang, R. A. Friesner, C. Nuckolls, D. R. Reichman, and P. F. Barbara, J. Am. Chem. Soc. **133**, 10155 (2011).

[65] Y. Shirota and H. Kageyama, Chem. Rev. **107**, 953 (2007).

[66] N. Vukmirovic, Phys. Chem. Chem. Phys. **15**, 3543 (2013).

[67] S. Jayanty and T. P. Radhakrishnan, Chem. Eur. J. **10**, 791 (2004).

[68] A. Patra, N. Hebalkar, B. Sreedhar, M. Sarkar, A. Samanta, and T. P. Radhakrishnan, Small **2**, 650 (2006).

[69] Y. Dong, J. W. Y. Lam, A. Qin, Z. Li, J. Sun, H. H.-Y. Sung, I. D. Williams, and B. Z. Tang, Chem. Commun. , 40 (2007).

[70] C. G. Chandaluri and T. P. Radhakrishnan, Angew. Chem. Int. Ed. **51**, 11849 (2012).

[71] B. Z. Tang, H. Z. Chen, R. S. Xu, J. W. Y. Lam, K. K. L. Cheuk, H. N. C. Wong, and M. Wang, Chem. Mater. **12**, 213 (2000).

[72] L. Sawyer and D. Grubb, *Polymer Microscopy* (Chapman and Hall, London; New York, 1996).

[73] A. I. Abutaha, S. R. S. Kumar, and H. N. Alshareef, Appl. Phys. Lett. **102**, 053507 (2013).

[74] S. V. Meille, G. Allegra, P. H. Geil, J. He, M. Hess, J.-I. Jin, P. Kratochvil, W. Mormann, and R. Stepto, Pure Appl. Chem. **83**, 1831 (2011).

[75] L. Segal, J. J. Creely, A. E. Martin, and C. M. Conrad, Text. Res. J. **29**, 786 (1959).

[76] Z. Mo, K. B. Lee, Y. B. Moon, M. Kobayashi, A. J. Heeger, and F. Wudl, Macromolecules **18**, 1972 (1985).

[77] M. Putkonen and L. Niinisto, Thin Solid Films **514**, 145 (2006).

[78] C. Mutungi, L. Passauer, C. Onyango, D. Jaros, and H. Rohm, Carbohydr. Polym. **87**, 598 (2012).

[79] Y. Liu, D. Thibodeaux, G. Gamble, P. Bauer, and D. VanDerveer, Appl. Spectrosc. **66**, 983 (2012).

[80] U. P. Agarwal, R. R. Reiner, and S. A. Ralph, J. Agric. Food Chem. **61**, 103 (2013).

[81] K. M. Paralikar and S. M. Betrabet, J. Appl. Polym. Sci. **21**, 899 (1977).

[82] E. Girardin, P. Millet, and A. Lodini, J. Biomed. Mater. Res. **49**, 211 (2000).

[83] X. Pan, T. Julian, and L. Augsburger, AAPS PharmSciTech **7**, E72 (2006).

[84] I. Kushida, J. Pharm. Pharmacol. **64**, 366 (2012).

[85] O. P. Obande and M. Gilbert, J. Appl. Polym. Sci. **37**, 1713 (1989).

[86] N. Vasanthan, J. Chem. Educ. **89**, 387 (2012).

[87] S. Kavesh and J. M. Schultz, Polym. Eng. Sci. **9**, 452 (1969).

[88] Z. Mo and H. Zhang, J. Macromol. Sci., Polym. Rev. **35**, 555 (1995).

[89] L. Zhang, E. W. Hansen, I. Helland, E. Hinrichsen, A. Larsen, and J. Roots, Macromolecules **42**, 5189 (2009).

[90] R. P. Paradkar, S. S. Sakhalkar, X. He, and M. S. Ellison, J. Appl. Polym. Sci. **88**, 545 (2003).

[91] S. S. Cherukupalli and A. A. Ogale, Polym. Eng. Sci. **44**, 1484 (2004).

[92] W. Lin, M. Cossar, V. Dang, and J. Teh, J. Polymer Testing **26**, 814 (2007).

[93] M. J. Pelletier, *Analytical Applications of Raman Spectroscopy* (Blackwell Science, Osney Mead, Oxford; Malden, MA, 1999).

[94] G. Choong, E. Vallat-Sauvain, X. Multone, L. Fesquet, U. Kroll, and J. Meier, J. Phys. D: Appl. Phys. **46**, 235105 (2013).

[95] W.-E. Hong and J.-S. Ro, J. Appl. Phys. **114**, 073511 (2013).

[96] I. R. Beattie, M. J. Gall, and G. A. Ozin, J. Chem. Soc. A , 1001 (1969).

[97] A. M. Mintairov, N. A. Sadchikov, T. Sauncy, M. Holtz, G. A. Seryogin, S. A. Nikishin, and H. Temkin, Phys. Rev. B **59**, 15197 (1999).

[98] G. Pezzotti, A. Matsutani, and W. Zhu, J. Am. Chem. Soc. **93**, 256 (2010).

[99] G. Xiao, Y. Guo, Y. Lin, X. Ma, Z. Su, and Q. Wang, Phys. Chem. Chem. Phys. **14**, 16286 (2012).

[100] E. R. Andrew, A. Bradbury, and R. G. Eades, Nature **182**, 1659 (1958).

[101] I. J. Lowe, Phys. Rev. Lett. **2**, 285 (1959).

[102] J. W. Hennel and J. Klinowski, in *New Techniques in Solid-State NMR*, Topics in Current Chemistry No. 246, edited by J. Klinowski (Springer Berlin Heidelberg, 2005) pp. 1–14.

[103] J. Perez, *Physics and Mechanics of Amorphous Polymers* (A.A. Balkema, Rotterdam; Brookfield, VT, 1998).

[104] T. Blythe and D. Bloor, *Electrical Properties of Polymers* (Cambridge University Press, 2008).

[105] N. C. Shaner, P. A. Steinbach, and R. Y. Tsien, Nature Methods **2**, 905 (2005).

[106] L. D. Lavis and R. T. Raines, ACS Chemical Biology **3**, 142 (2008).

[107] X. Michalet, F. F. Pinaud, L. A. Bentolila, J. M. Tsay, S. Doose, J. J. Li, G. Sundaresan, A. M. Wu, S. S. Gambhir, and S. Weiss, Science **307**, 538 (2005).

[108] P. Yan, C. D. Acker, W.-L. Zhou, P. Lee, C. Bollensdorff, A. Negrean, J. Lotti, L. Sacconi, S. D. Antic, P. Kohl, H. D. Mansvelder, F. S. Pavone, and L. M. Loew, Proceedings of the National Academy of Sciences **109**, 20443 (2012).

[109] Y. Zhao, S. Araki, J. Wu, T. Teramoto, Y.-F. Chang, M. Nakano, A. S. Abdelfattah, M. Fujiwara, T. Ishihara, T. Nagai, and R. E. Campbell, Science **333**, 1888 (2011).

[110] M. Bates, B. Huang, G. T. Dempsey, and X. Zhuang, Science **317**, 1749 (2007).

[111] G. Donnert, J. Keller, C. A. Wurm, S. O. Rizzoli, V. Westphal, A. Schönle, R. Jahn, S. Jakobs, C. Eggeling, and S. W. Hell, Biophysical Journal **92**, L67 (2007).

[112] A. Miyawaki, Annual Review of Biochemistry **80**, 357 (2011).

[113] Z. Chen, A. L. Weisman, L. Wei, R. A. Friesner, and W. Min, In preparation (2017).

[114] Y. Cao, T. Hughes, D. Giesen, M. D. Halls, A. Goldberg, T. R. Vadicherla, M. Sastry, B. Patel, W. Sherman, A. L. Weisman, and R. A. Friesner, Journal of Computational Chemistry **37**, 1425 (2016).

[115] T. W. Lyons, C. T. Reinhard, and N. J. Planavsky, Nature **506**, 307 (2014).

[116] C. C. L. McCrory, S. Jung, I. M. Ferrer, S. M. Chatman, J. C. Peters, and T. F. Jaramillo, J. Am. Chem. Soc. **137**, 4347 (2015).

[117] J. H. Lee and A. Selloni, Phys. Rev. Lett. **112**, 196102 (2014).

[118] I. C. Man, H.-Y. Su, F. Calle-Vallejo, H. A. Hansen, J. I. Martínez, N. G. Inoglu, J. Kitchin, T. F. Jaramillo, J. K. Nørskov, and J. Rossmeisl, ChemCatChem **3**, 1159 (2011).

[119] Y.-F. Li, Z.-P. Liu, L. Liu, and W. Gao, J. Am. Chem. Soc. **132**, 13008 (2010).

[120] Á. Valdés, Z.-W. Qu, G.-J. Kroes, J. Rossmeisl, and J. K. Nørskov, The Journal of Physical Chemistry C **112**, 9872 (2008).

[121] P. Kalisman, Y. Nakibli, and L. Amirav, Nano Lett. **16**, 1776 (2016).

[122] J. Chen, Y.-F. Li, P. Sit, and A. Selloni, J. Am. Chem. Soc. **135**, 18774 (2013).

[123] G. Wang, Y. Ling, H. Wang, L. Xihong, and Y. Li, Journal of Photochemistry and Photobiology C: Photochemistry Reviews **19**, 35 (2014).

[124] F. Rodríguez-Hernández, D. C. Tranca, B. M. Szyja, R. A. van Santen, A. Martínez-Mesa, L. Uranga-Piña, and G. Seifert, The Journal of Physical Chemistry C **120**, 437 (2016).

[125] M. Huynh, D. K. Bediako, and D. G. Nocera, J. Am. Chem. Soc. **136**, 6002 (2014).

[126] J. Zhang, T. F. Hughes, M. Steigerwald, L. Brus, and R. A. Friesner, J. Am. Chem. Soc. **134**, 12028 (2012).

[127] X.-Q. Gong, A. Selloni, M. Batzill, and U. Diebold, Nature Materials **5**, 665 (2006).

[128] W.-J. Ong, L.-L. Tan, S.-P. Chai, S.-T. Yong, and A. R. Mohamed, Nanoscale **6**, 1946 (2014).

[129] F. De Angelis, C. Di Valentin, S. Fantacci, A. Vittadini, and A. Selloni, Chemical Reviews (2014), 10.1021/cr500055q.

[130] T. R. Gordon, M. Cargnello, T. Paik, F. Mangolini, R. T. Weber, P. Fornasiero, and C. B. Murray, Journal of the American Chemical Society **134**, 6751 (2012).

[131] G. Bergerhoff and I.D. Brown, in *Crystallographic Databases* (International Union of Crystallography, Chester, 1987).

[132] W. Yang, J. Li, Y. Wang, F. Zhu, W. Shi, F. Wan, and D. Xu, Chem. Commun. **47**, 1809 (2011).

[133] M.-H. Jung, M.-J. Chu, and M. G. Kang, Chem. Commun. **48**, 5016 (2012).

[134] J. Fan, W. Cai, and J. Yu, Chemistry – An Asian Journal **6**, 2481 (2011).

[135] J. Yu, J. Fan, and K. Lv, Chem. Commun. **2**, 2144 (2010).

[136] J. Miao and B. Liu, RSC Adv. **3**, 1222 (2012).

[137] W. Q. Fang, X.-Q. Gong, and H. G. Yang, The Journal of Physical Chemistry Letters **2**, 725 (2011).

[138] B. Li, Z. Zhao, F. Gao, X. Wang, and J. Qiu, Applied Catalysis B: Environmental **147**, 958 (2014).

[139] J. Pan, G. Liu, G. Q. M. Lu, and H.-M. Cheng, Angewandte Chemie International Edition **50**, 2133 (2011).

[140] W. Hebenstreit, N. Ruzycki, G. S. Herman, Y. Gao, and U. Diebold, Phys. Rev. B **62**, R16334 (2000).

[141] F. F. Sanches, G. Mallia, L. Liborio, U. Diebold, and N. M. Harrison, Physical Review B **89**, 245309 (2014).

[142] U. Diebold, *Growth, surface characterization, and reactivity of TiO2 anatase films - EPSCOR*, Tech. Rep. 835107 (Office of Scientific and Technical Information, 2004).

[143] A. S. Barnard, P. Zapol, and L. A. Curtiss, Journal of Chemical Theory and Computation **1**, 107 (2005).

[144] G. S. Herman, M. R. Sievers, and Y. Gao, Physical Review Letters **84**, 3354 (2000).

[145] R. Hengerer, B. Bolliger, M. Erbudak, and M. Grätzel, Surface Science **460**, 162 (2000).

[146] M. Lazzeri and A. Selloni, Physical Review Letters **87**, 266105 (2001).

[147] L. M. Peter, Electroanalysis **27**, 864 (2015).

[148] A. J. Nozik, Nature **257**, 383 (1975).

[149] A. J. Cowan, C. J. Barnett, S. R. Pendlebury, M. Barroso, K. Sivula, M. Grätzel, J. R. Durrant, and D. R. Klug, Journal of the American Chemical Society **133**, 10134 (2011).

[150] Q. Kuang, X. Wang, Z. Jiang, Z. Xie, and L. Zheng, Accounts of Chemical Research **47**, 308 (2014).

[151] M. A. Henderson, Surface Science Reports **46**, 1 (2002).

[152] A. Selloni, Nature Materials **7**, 613 (2008).

[153] L. M. Peter and K. G. Upul Wijayantha, ChemPhysChem **15**, 1983 (2014).

[154] L. M. Peter, Journal of Solid State Electrochemistry **17**, 315 (2013).

[155] A. L. Weisman, K. H. DuBay, K. A. Willets, and R. A. Friesner, The Journal of Chemical Physics **141**, 224702 (2014).

[156] Z. Chen, D. W. Paley, L. Wei, A. L. Weisman, R. A. Friesner, C. Nuckolls, and W. Min, Journal of the American Chemical Society **136**, 8027 (2014).

[157] J. Zhang, A. L. Weisman, P. Saitta, and R. A. Friesner, International Journal of Quantum Chemistry **116**, 357 (2016).

Part III

Appendices

Appendix A

Transition dipole moment

A.1 General solution to the time-dependent Schrödinger equation

The time-dependent Schrödinger equation for the N electrons of a scatterer is

$$H\left(\{\mathbf{r}\}\right)\Psi\left(\{\mathbf{r}\},t\right)=i\hbar\frac{\partial\Psi\left(\{\mathbf{r}\},t\right)}{\partial t},$$

where $\{\mathbf{r}\}$ represents the set of $3N$ spatial coordinates of all N electrons. As long as the Hamiltonian H is time-independent, we can solve this partial differential equation using separation of variables with $\Psi\left(\{\mathbf{r}\},t\right)=u\left(\{\mathbf{r}\}\right)T\left(t\right)$:

$$\begin{aligned}H\left(\{\mathbf{r}\}\right)u\left(\{\mathbf{r}\}\right)T\left(t\right) &= i\hbar\frac{\partial u\left(\{\mathbf{r}\}\right)T\left(t\right)}{\partial t}\\ T\left(t\right)H\left(\{\mathbf{r}\}\right)u\left(\{\mathbf{r}\}\right) &= i\hbar u\left(\{\mathbf{r}\}\right)\frac{\partial T\left(t\right)}{\partial t}\\ \frac{H\left(\{\mathbf{r}\}\right)u\left(\{\mathbf{r}\}\right)}{u\left(\{\mathbf{r}\}\right)} &= i\hbar\frac{1}{T\left(t\right)}\frac{\partial T\left(t\right)}{\partial t}.\end{aligned}$$

Since the sides of this equation are functionally independent, they must both equal the same constant, which we call E:

$$\begin{aligned}H\left(\{\mathbf{r}\}\right)u\left(\{\mathbf{r}\}\right) &= Eu\left(\{\mathbf{r}\}\right)\\ \frac{\partial T\left(t\right)}{\partial t} &= -i\omega T\left(t\right),\end{aligned}$$

where here and elsewhere we use $\omega\equiv\frac{E}{\hbar}$ and $\omega_{ab}\equiv\frac{E_a-E_b}{\hbar}$. If we have solved the former equation, the time-independent Schrödinger equation, completely for the eigenfunctions $\{u\left(\{\mathbf{r}\}\right)\}$ and eigenvalues $\{E\}$ of the

Hamiltonian $H(\{\mathbf{r}\})$, then the solution to the latter equation is

$$T(t) = T(0)\, e^{-i\omega t}.$$

The full solution is a linear combination of all the eigensolutions:

$$\begin{aligned}\Psi(\{\mathbf{r}\},t) &= \sum_{i=1}^{\infty} \Psi_i(\{\mathbf{r}\},t) \\ &= \sum_{i=1}^{\infty} a_i u_i(\{\mathbf{r}\})\, e^{-i\omega_i t}.\end{aligned}$$

Since the Hamiltonian is Hermitian, its set of spatial eigenfunctions $\{u(\{\mathbf{r}\})\}$ is orthonormal,

$$\int u_i^*(\{\mathbf{r}\})\, u_j(\{\mathbf{r}\})\, d\{\mathbf{r}\} = \delta_{ij}, \tag{A.1}$$

and its eigenvalues $\{E\}$ are real. However, as a crude way of incorporating finite lifetimes of the eigenstates, we transform $\omega_i \to \omega_i - i\Gamma_i$, where Γ_i is the is the half-width of level i and both ω_i and Γ_i are real. Thus, the general unperturbed solution to the Schrödinger equation is

$$\Psi(\{\mathbf{r}\},t) = \sum_{i=1}^{\infty} a_i u_i(\{\mathbf{r}\})\, e^{-i(\omega_i - i\Gamma_i)t}. \tag{A.2}$$

We note that throughout this entire analysis we make no restriction on the spatial wavefunctions $\{u(\{\mathbf{r}\})\}$, although they are usually real except in cases involving magnetic effects or degenerate states[1]. We also note that for notational simplicity we leave out the overall phase present in the time-dependent factors of the stationary states, though in the main text when this is important we address it explicitly. The modification to eq. A.2 would be

$$\Psi(\{\mathbf{r}\},t) \to \sum_{i=1}^{\infty} a_i u_i(\{\mathbf{r}\})\, e^{-i(\omega_i - i\Gamma_i)t + i\phi_i},$$

and a trick for keeping track of the phase in what follows is to assume that whatever happens to the frequency ω_i in phase factors like $e^{-i(\omega_i - i\Gamma_i)t + i\phi_i}$ will also happen to ϕ_i.

A.2 Time-dependent perturbation theory

If a time-dependent perturbation $V(\{\mathbf{r}\},t)$ to the Hamiltonian is introduced, so that the new Hamiltonian is

$$H(\{\mathbf{r}\},t) = H^{(0)}(\{\mathbf{r}\}) + \lambda V(\{\mathbf{r}\},t),$$

where λ is a real number between 0 and 1 that keeps track of the order of the perturbation to the original Hamiltonian $H^{(0)}(\{\mathbf{r}\})$, we predict that the new wavefunction $\Psi(\{\mathbf{r}\},t)$ can be written as a power series in λ as a perturbation to the original solution $\Psi^{(0)}(\{\mathbf{r}\},t)$,

$$\begin{aligned}\Psi(\{\mathbf{r}\},t) &= \Psi^{(0)}(\{\mathbf{r}\},t)+\lambda\Psi^{(1)}(\{\mathbf{r}\},t)+\lambda^2\Psi^{(2)}(\{\mathbf{r}\},t)+\cdots \\ &= \sum_{j=0}^{\infty}\lambda^j\Psi^{(j)}(\{\mathbf{r}\},t),\end{aligned}$$

where each correction $\Psi^{(j)}(\{\mathbf{r}\},t)$ to the original wavefunction is written as a linear combination of the original eigensolutions with time-dependent coefficients:

$$\Psi^{(j)}(\{\mathbf{r}\},t)=\sum_{i=1}^{\infty}a_i^{(j)}(t)\,u_i(\{\mathbf{r}\})\,e^{-i(\omega_i-i\Gamma_i)t}. \tag{A.3}$$

We assume that the perturbation $V(\{\mathbf{r}\},t)$ contains no time derivatives and is generally a multiplicative function of time. The left-hand side of the new time-dependent Schrödinger equation is therefore

$$\begin{aligned}\text{LHS} &= H(\{\mathbf{r}\},t)\,\Psi(\{\mathbf{r}\},t) \\ &= \Big(H^{(0)}(\{\mathbf{r}\})+\lambda V(\{\mathbf{r}\},t)\Big)\sum_{j=0}^{\infty}\lambda^j\Psi^{(j)}(\{\mathbf{r}\},t) \\ &= H^{(0)}(\{\mathbf{r}\})\sum_{j=0}^{\infty}\lambda^j\sum_{i=1}^{\infty}a_i^{(j)}(t)\,u_i(\{\mathbf{r}\})\,e^{-i(\omega_i-i\Gamma_i)t}+\lambda V(\{\mathbf{r}\},t)\sum_{j=0}^{\infty}\lambda^j\sum_{i=1}^{\infty}a_i^{(j)}(t)\,u_i(\{\mathbf{r}\})\,e^{-i(\omega_i-i\Gamma_i)t} \\ &= \sum_{j=0}^{\infty}\lambda^j\sum_{i=1}^{\infty}a_i^{(j)}(t)\,H^{(0)}(\{\mathbf{r}\})\,u_i(\{\mathbf{r}\})\,e^{-i(\omega_i-i\Gamma_i)t}+\lambda\sum_{j=0}^{\infty}\lambda^j\sum_{i=1}^{\infty}a_i^{(j)}(t)\,V(\{\mathbf{r}\},t)\,u_i(\{\mathbf{r}\})\,e^{-i(\omega_i-i\Gamma_i)t} \\ &= \sum_{j=0}^{\infty}\lambda^j\sum_{i=1}^{\infty}a_i^{(j)}(t)\,E_iu_i(\{\mathbf{r}\})\,e^{-i(\omega_i-i\Gamma_i)t}+\sum_{j=0}^{\infty}\lambda^{j+1}\sum_{i=1}^{\infty}a_i^{(j)}(t)\,V(\{\mathbf{r}\},t)\,u_i(\{\mathbf{r}\})\,e^{-i(\omega_i-i\Gamma_i)t},\end{aligned}$$

and the right-hand side is

$$\begin{aligned}\text{RHS} &= i\hbar\frac{\partial\Psi(\{\mathbf{r}\},t)}{\partial t} \\ &= i\hbar\frac{\partial}{\partial t}\sum_{j=0}^{\infty}\lambda^j\Psi^{(j)}(\{\mathbf{r}\},t) \\ &= i\hbar\frac{\partial}{\partial t}\sum_{j=0}^{\infty}\lambda^j\sum_{i=1}^{\infty}a_i^{(j)}(t)\,u_i(\{\mathbf{r}\})\,e^{-i(\omega_i-i\Gamma_i)t} \\ &= i\hbar\sum_{j=0}^{\infty}\lambda^j\sum_{i=1}^{\infty}u_i(\{\mathbf{r}\})\,\frac{\partial}{\partial t}a_i^{(j)}(t)\,e^{-i(\omega_i-i\Gamma_i)t} \\ &= i\hbar\sum_{j=0}^{\infty}\lambda^j\sum_{i=1}^{\infty}u_i(\{\mathbf{r}\})\,e^{-i(\omega_i-i\Gamma_i)t}\dot{a}_i^{(j)}(t)+\sum_{j=0}^{\infty}\lambda^j\sum_{i=1}^{\infty}E_iu_i(\{\mathbf{r}\})\,a_i^{(j)}(t)\,e^{-i(\omega_i-i\Gamma_i)t},\end{aligned}$$

where we have applied the zeroth-order time-independent Schrödinger equation for the ith eigenfunction, $H^{(0)}(\{\mathbf{r}\})u_i(\{\mathbf{r}\}) = E_i u_i(\{\mathbf{r}\})$, on the left-hand side and expanded the time derivative on the right-hand side. Subtracting the quantity

$$\sum_{j=0}^{\infty}\lambda^j\sum_{i=1}^{\infty}a_i^{(j)}(t)\,E_i u_i(\{\mathbf{r}\})\,e^{-i(\omega_i-i\Gamma_i)t}$$

from both sides, the Schrödinger equation becomes

$$\sum_{j=0}^{\infty}\lambda^{j+1}\sum_{i=1}^{\infty}a_i^{(j)}(t)\,V(\{\mathbf{r}\},t)\,u_i(\{\mathbf{r}\})\,e^{-i(\omega_i-i\Gamma_i)t} = i\hbar\sum_{j=0}^{\infty}\lambda^j\sum_{i=1}^{\infty}u_i(\{\mathbf{r}\})\,e^{-i(\omega_i-i\Gamma_i)t}\dot{a}_i^{(j)}(t).$$

Left-multiplying both sides by $u_k^*(\{\mathbf{r}\})$, and integrating over all spatial coordinates, we have

$$\begin{aligned}\sum_{j=0}^{\infty}\lambda^{j+1}\sum_{i=1}^{\infty}a_i^{(j)}(t)\,V_{ki}(t)\,e^{-i(\omega_i-i\Gamma_i)t} &= i\hbar\sum_{j=0}^{\infty}\lambda^j\sum_{i=1}^{\infty}\delta_{ik}e^{-i(\omega_i-i\Gamma_i)t}\dot{a}_i^{(j)}(t)\\ &= i\hbar\sum_{j=0}^{\infty}\lambda^j e^{-i(\omega_k-i\Gamma_k)t}\dot{a}_k^{(j)}(t),\end{aligned}$$

where we have used the orthonormality of the spatial eigenfunctions (eq. A.1) and defined

$$V_{ki}(t) \equiv \int u_k^*(\{\mathbf{r}\})\,V(\{\mathbf{r}\},t)\,u_i(\{\mathbf{r}\})\,d\{\mathbf{r}\}. \tag{A.4}$$

Equating terms of similar degrees of smallness, i.e., of the same powers of λ, we have for the $\lambda^0 = 1$ terms,

$$\begin{aligned}i\hbar e^{-i(\omega_k-i\Gamma_k)t}\dot{a}_k^{(0)}(t) &= 0\\ \int_{-\infty}^{t'}\dot{a}_k^{(0)}(t)\,dt &= 0\\ a_k^{(0)}(t) &= a_k^{(0)}(-\infty)\\ &\equiv c_k.\end{aligned}$$

For the general λ^M terms ($M \geq 1$) we have

$$\begin{aligned}i\hbar e^{-i(\omega_k-i\Gamma_k)t}\dot{a}_k^{(M)}(t) &= \sum_{i=1}^{\infty}a_i^{(M-1)}(t)\,V_{ki}(t)\,e^{-i(\omega_i-i\Gamma_i)t}\\ \dot{a}_k^{(M)}(t) &= -\frac{i}{\hbar}\sum_{i=1}^{\infty}a_i^{(M-1)}(t)\,V_{ki}(t)\,e^{i(\omega_{ki}-i\Gamma_{ki})t}\\ \int_{-\infty}^{t'}\dot{a}_k^{(M)}(t)\,dt &= -\frac{i}{\hbar}\sum_{i=1}^{\infty}\int_{-\infty}^{t'}a_i^{(M-1)}(t)\,V_{ki}(t)\,e^{i(\omega_{ki}-i\Gamma_{ki})t}dt\\ a_k^{(M)}(t) &= -\frac{i}{\hbar}\sum_{i=1}^{\infty}\int_{-\infty}^{t}a_i^{(M-1)}(t')\,V_{ki}(t')\,e^{i(\omega_{ki}-i\Gamma_{ki})t'}dt' + a_k^{(M)}(-\infty).\end{aligned} \tag{A.5}$$

Note that whereas $\omega_{ki} = \omega_k - \omega_i$, $\Gamma_{ki} = \Gamma_k + \Gamma_i$.

A.3 Transition dipole moment due to perturbing electric fields

We assume that the perturbation $V(\{\mathbf{r}\},t)$ to the Hamiltonian is due to the total electric field $\mathbf{E}(t)$ interacting with the dipole moment $\mathbf{p}(\{\mathbf{r}\})$ it induces in the scatterer with interaction energy

$$V(\{\mathbf{r}\},t) = -\mathbf{p}(\{\mathbf{r}\}) \cdot \mathbf{E}(t).$$

We have left out the spatial dependence of $\mathbf{E}$ because we assume that at time t the field is constant over the entire scatterer, as is the case in Raman and Rayleigh scattering. If N_f monochromatic plane waves contribute to the total electric field, then

$$\mathbf{E}(t) = \sum_{i=1}^{N_f} \left(\tilde{\mathbf{E}}_i e^{-i\omega_i t} + \text{c.c.} \right).$$

Then, from eq. A.4, the matrix element of the energy perturbation in the basis of the spatial eigenfunctions is

$$\begin{aligned} V_{ki}(t) &= -\int u_k^*(\{\mathbf{r}\})\, \mathbf{p}(\{\mathbf{r}\}) \cdot \mathbf{E}(t)\, u_i(\{\mathbf{r}\})\, d\{\mathbf{r}\} \\ &= -\left(\int u_k^*(\{\mathbf{r}\})\, \mathbf{p}(\{\mathbf{r}\})\, u_i(\{\mathbf{r}\})\, d\{\mathbf{r}\} \right) \cdot \mathbf{E}(t) \\ &= -\mathbf{p}_{ki} \cdot \sum_{j=1}^{N_f} \left(\tilde{\mathbf{E}}_j e^{-i\omega_j t} + \tilde{\mathbf{E}}_j^* e^{i\omega_j t} \right), \end{aligned} \tag{A.6}$$

where we have defined the transition dipole moment $\mathbf{p}_{ki}$ as the appropriate matrix element of the dipole operator in the basis of the unperturbed spatial eigenfunctions:

$$\mathbf{p}_{ki} \equiv \int u_k^*(\{\mathbf{r}\})\, \mathbf{p}(\{\mathbf{r}\})\, u_i(\{\mathbf{r}\})\, d\{\mathbf{r}\}.$$

In Dirac notation, this is written as

$$\mathbf{p}_{ki} = \langle u_k(\{\mathbf{r}\}) | \mathbf{p} | u_i(\{\mathbf{r}\}) \rangle.$$

The new wavefunction resulting from the perturbation then follows from the time-dependent perturbation theory developed above. We assume that at $t = -\infty$ there is no perturbing electric field, so that $a_k^{(M)}(-\infty) = 0$ for $M > 0$. Further, we assume that in the absence of perturbing electric fields (i.e., at $t = -\infty$), the scatterer is in state s so that $a_k^{(0)}(-\infty) = c_k = \delta_{ks}$. Since $a_k^{(0)}(t) = a_k^{(0)}(-\infty)$, we therefore know all the zeroth-order coefficients at time t, and all that remains is to determine the higher-order coefficients at t. The first-order

coefficients for our system follow from eq. A.5:

$$\begin{aligned} a_k^{(1)}(t) &= -\frac{i}{\hbar}\sum_{i=1}^{\infty} c_i \int_{-\infty}^{t} V_{ki}(t')\, e^{i(\omega_{ki}-i\Gamma_{ki})t'} dt' \\ &= -\frac{i}{\hbar}\sum_{i=1}^{\infty} \delta_{is} \int_{-\infty}^{t} V_{ki}(t')\, e^{i(\omega_{ki}-i\Gamma_{ki})t'} dt' \\ &= -\frac{i}{\hbar}\int_{-\infty}^{t} V_{ks}(t')\, e^{i(\omega_{ki}-i\Gamma_{ki})t'} dt'. \end{aligned}$$

Plugging in the expression for $V_{ks}(t')$ from eq. A.6, we have

$$\begin{aligned} a_k^{(1)}(t) &= -\frac{i}{\hbar}\int_{-\infty}^{t} \left(-\mathbf{p}_{ks}\cdot\sum_{j=1}^{N_f}\left(\tilde{\mathbf{E}}_j e^{-i\omega_j t'} + \tilde{\mathbf{E}}_j^* e^{i\omega_j t'}\right)\right) e^{i(\omega_{ks}-i\Gamma_{ks})t'} dt' \\ &= \frac{i}{\hbar}\mathbf{p}_{ks}\cdot\sum_{j=1}^{N_f}\left(\tilde{\mathbf{E}}_j \int_{-\infty}^{t} e^{-i(\omega_j-(\omega_{ks}-i\Gamma_{ks}))t'} dt' + \tilde{\mathbf{E}}_j^* \int_{-\infty}^{t} e^{i(\omega_j+(\omega_{ks}-i\Gamma_{ks}))t'} dt'\right) \\ &= \frac{i}{\hbar}\mathbf{p}_{ks}\cdot\sum_{j=1}^{N_f}\left(\tilde{\mathbf{E}}_j \left[\frac{e^{-i(\omega_j-(\omega_{ks}-i\Gamma_{ks}))t'}}{-i\left(\omega_j-(\omega_{ks}-i\Gamma_{ks})\right)}\right]_{-\infty}^{t} + \tilde{\mathbf{E}}_j^* \left[\frac{e^{i(\omega_j+(\omega_{ks}-i\Gamma_{ks}))t'}}{i\left(\omega_j+(\omega_{ks}-i\Gamma_{ks})\right)}\right]_{-\infty}^{t}\right) \\ &= \frac{1}{\hbar}\mathbf{p}_{ks}\cdot\sum_{j=1}^{N_f}\left(-\tilde{\mathbf{E}}_j \frac{e^{-i(\omega_j-(\omega_{ks}-i\Gamma_{ks}))t}}{\omega_j-(\omega_{ks}-i\Gamma_{ks})} + \tilde{\mathbf{E}}_j^* \frac{e^{i(\omega_j+(\omega_{ks}-i\Gamma_{ks}))t}}{\omega_j+(\omega_{ks}-i\Gamma_{ks})}\right), \end{aligned} \tag{A.7}$$

where we have assumed that at $t'=-\infty$ the terms $e^{-i(\omega_j-(\omega_{ks}-i\Gamma_{ks}))t'}$ and $e^{i(\omega_j+(\omega_{ks}-i\Gamma_{ks}))t'}$ equal zero on the physical grounds that the applied fields were not present way back then.

Knowing the coefficients $a_k^{(1)}(t)$ as in eq. A.7 allows us to write down what the new total wavefunction is in the presence of the perturbation, at least to first order in perturbation strength. This is useful for calculating expectation values or transition matrix elements in the presence of the perturbation. For example, if we wanted to know the expectation value of the dipole moment of a scatterer in unperturbed state s in the presence of perturbing electric fields — or more precisely, of a scatterer starting in state s, being perturbed by applied electric fields, and then ending up in state s — we would need

$$\langle\widetilde{\mathbf{p}}\rangle_s = \langle\Psi_s(\{\mathbf{r}\},t)\,|\mathbf{p}|\,\Psi_s(\{\mathbf{r}\},t)\rangle.$$

Here we have drawn a *wide* tilde $\widetilde{\ }$ above the expectation value to indicate that it is a complex value (not to be confused with a regular tilde $\tilde{\ }$ which indicates that the quantity is a complex *amplitude*). Below, we will take the real part in order to obtain the physical transition dipole moments, which are real quantities.

If instead we wanted to know the *transition* dipole moment corresponding to a scatterer starting in state s, being perturbed by applied electric fields, and then ending up in state t, we would need

$$\langle \widetilde{\mathbf{p}} \rangle_{ts} = \langle \Psi_t \left(\{\mathbf{r}\}, t \right) |\mathbf{p}| \, \Psi_s \left(\{\mathbf{r}\}, t \right) \rangle .$$

In both cases, we need the *full* wavefunction of the form

$$|\Psi_s \left(\{\mathbf{r}\}, t \right)\rangle = \left| \Psi_s^{(0)} \left(\{\mathbf{r}\}, t \right) \right\rangle + \lambda \left| \Psi_s^{(1)} \left(\{\mathbf{r}\}, t \right) \right\rangle + \lambda^2 \left| \Psi_s^{(2)} \left(\{\mathbf{r}\}, t \right) \right\rangle + \cdots ,$$

where, from eq. A.3,

$$\begin{aligned}
\Psi_s^{(0)} \left(\{\mathbf{r}\}, t \right) &= u_s \left(\{\mathbf{r}\} \right) e^{-i(\omega_s - i\Gamma_s)t} \\
\Psi_s^{(1)} \left(\{\mathbf{r}\}, t \right) &= \frac{1}{\hbar} \sum_{i=1}^{\infty} \left(\mathbf{p}_{is} \cdot \sum_{j=1}^{N_f} \left(-\tilde{\mathbf{E}}_j \frac{e^{-i(\omega_j + (\omega_s - i\Gamma_s))t}}{\omega_j - (\omega_{is} - i\Gamma_{is})} + \tilde{\mathbf{E}}_j^* \frac{e^{i(\omega_j - (\omega_s - i\Gamma_s))t}}{\omega_j + (\omega_{is} - i\Gamma_{is})} \right) \right) u_i \left(\{\mathbf{r}\} \right) \\
&\vdots
\end{aligned}$$

Note that we have used eq. A.7 and the fact that $a_i^{(0)}(t) = \delta_{is}$.

Thus, collecting terms of similar order in the perturbation (using the perturbation strength parameter λ), we find that the transition dipole moments to increasing orders of perturbation strength are

$$\begin{aligned}
\langle \mathbf{p} \rangle_{ts}^{(0)} &= \frac{1}{2} \left\langle \Psi_t^{(0)} \left(\{\mathbf{r}\}, t \right) |\mathbf{p}| \, \Psi_s^{(0)} \left(\{\mathbf{r}\}, t \right) \right\rangle + \text{c.c.} \\
\langle \mathbf{p} \rangle_{ts}^{(1)} &= \frac{1}{2} \left(\left\langle \Psi_t^{(0)} \left(\{\mathbf{r}\}, t \right) |\mathbf{p}| \, \Psi_s^{(1)} \left(\{\mathbf{r}\}, t \right) \right\rangle + \left\langle \Psi_t^{(1)} \left(\{\mathbf{r}\}, t \right) |\mathbf{p}| \, \Psi_s^{(0)} \left(\{\mathbf{r}\}, t \right) \right\rangle \right) + \text{c.c.} \\
\langle \mathbf{p} \rangle_{ts}^{(2)} &= \frac{1}{2} \Big(\left\langle \Psi_t^{(0)} \left(\{\mathbf{r}\}, t \right) |\mathbf{p}| \, \Psi_s^{(2)} \left(\{\mathbf{r}\}, t \right) \right\rangle + \left\langle \Psi_t^{(1)} \left(\{\mathbf{r}\}, t \right) |\mathbf{p}| \, \Psi_s^{(1)} \left(\{\mathbf{r}\}, t \right) \right\rangle \\
&\quad + \left\langle \Psi_t^{(2)} \left(\{\mathbf{r}\}, t \right) |\mathbf{p}| \, \Psi_s^{(0)} \left(\{\mathbf{r}\}, t \right) \right\rangle \Big) + \text{c.c.} \\
&\vdots
\end{aligned}$$

We prepare for calculating these transition dipole matrix elements by writing out the required eigenstates in the Hilbert space of unperturbed spatial eigenfunctions $\{u(\{\mathbf{r}\})\}$ up to first order in the perturbation:

$$\begin{aligned}
\left| \Psi_s^{(0)} \left(\{\mathbf{r}\}, t \right) \right\rangle &= |u_s \left(\{\mathbf{r}\} \right)\rangle e^{-i(\omega_s - i\Gamma_s)t} \\
\left\langle \Psi_t^{(0)} \left(\{\mathbf{r}\}, t \right) \right| &= \langle u_t \left(\{\mathbf{r}\} \right)| e^{i(\omega_t + i\Gamma_t)t} \\
\left| \Psi_s^{(1)} \left(\{\mathbf{r}\}, t \right) \right\rangle &= \frac{1}{\hbar} \sum_{i=1}^{\infty} \left(\mathbf{p}_{is} \cdot \sum_{j=1}^{N_f} \left(-\tilde{\mathbf{E}}_j \frac{e^{-i(\omega_j + (\omega_s - i\Gamma_s))t}}{\omega_j - \omega_{is} + i\Gamma_{is}} + \tilde{\mathbf{E}}_j^* \frac{e^{i(\omega_j - (\omega_s - i\Gamma_s))t}}{\omega_j + \omega_{is} - i\Gamma_{is}} \right) \right) |u_i \left(\{\mathbf{r}\} \right)\rangle \\
\left\langle \Psi_t^{(1)} \left(\{\mathbf{r}\}, t \right) \right| &= \frac{1}{\hbar} \sum_{i=1}^{\infty} \left(\mathbf{p}_{it}^* \cdot \sum_{j=1}^{N_f} \left(-\tilde{\mathbf{E}}_j^* \frac{e^{i(\omega_j + (\omega_t + i\Gamma_t))t}}{\omega_j - \omega_{it} - i\Gamma_{it}} + \tilde{\mathbf{E}}_j \frac{e^{-i(\omega_j - (\omega_t + i\Gamma_t))t}}{\omega_j + \omega_{it} + i\Gamma_{it}} \right) \right) \langle u_i \left(\{\mathbf{r}\} \right)| .
\end{aligned}$$

Then,

$$
\begin{aligned}
\langle \mathbf{p} \rangle_{ts}^{(0)} &= \frac{1}{2} \left\langle u_t \left(\{\mathbf{r}\} \right) |\mathbf{p}| \, u_s \left(\{\mathbf{r}\} \right) \right\rangle e^{i(\omega_{ts} + i\Gamma_{ts})t} + \text{c.c.} \\
&= \frac{1}{2} \mathbf{p}_{ts} e^{i(\omega_{ts} + i\Gamma_{ts})t} + \text{c.c.} \\
\langle \mathbf{p} \rangle_{ts}^{(1)} &= \frac{1}{2} \Bigg(\left\langle u_t \left(\{\mathbf{r}\} \right) \right| e^{i(\omega_t + i\Gamma_t)t} \mathbf{p} \frac{1}{\hbar} \sum_{i=1}^{\infty} \left(\mathbf{p}_{is} \cdot \sum_{j=1}^{N_f} \left(-\tilde{\mathbf{E}}_j \frac{e^{-i(\omega_j + (\omega_s - i\Gamma_s))t}}{\omega_j - \omega_{is} + i\Gamma_{is}} + \tilde{\mathbf{E}}_j^* \frac{e^{i(\omega_j - (\omega_s - i\Gamma_s))t}}{\omega_j + \omega_{is} - i\Gamma_{is}} \right) \right) \left| u_i \left(\{\mathbf{r}\} \right) \right\rangle \\
&\quad + \frac{1}{\hbar} \sum_{i=1}^{\infty} \left(\mathbf{p}_{it}^* \cdot \sum_{j=1}^{N_f} \left(-\tilde{\mathbf{E}}_j^* \frac{e^{i(\omega_j + (\omega_t + i\Gamma_t))t}}{\omega_j - \omega_{it} - i\Gamma_{it}} + \tilde{\mathbf{E}}_j \frac{e^{-i(\omega_j - (\omega_t + i\Gamma_t))t}}{\omega_j + \omega_{it} + i\Gamma_{it}} \right) \right) \left\langle u_i \left(\{\mathbf{r}\} \right) |\mathbf{p}| \, u_s \left(\{\mathbf{r}\} \right) \right\rangle e^{-i(\omega_s - i\Gamma_s)t} \Bigg) \\
&\quad + \text{c.c.} \\
&= e^{-\Gamma_{ts} t} \sum_{j=1}^{N_f} \frac{1}{2\hbar} \sum_{i=1}^{\infty} \left(\frac{\mathbf{p}_{ti} \mathbf{p}_{is}}{\omega_{is} - \omega_j - i\Gamma_{is}} \cdot \tilde{\mathbf{E}}_j e^{-i(\omega_j - \omega_{ts})t} + \frac{\mathbf{p}_{ti} \mathbf{p}_{is}}{\omega_{is} + \omega_j - i\Gamma_{is}} \cdot \tilde{\mathbf{E}}_j^* e^{i(\omega_j + \omega_{ts})t} \right) + \text{c.c.} \\
&\quad + e^{-\Gamma_{ts} t} \sum_{j=1}^{N_f} \frac{1}{2\hbar} \sum_{i=1}^{\infty} \left(\frac{\mathbf{p}_{is} \mathbf{p}_{ti}}{\omega_{it} - \omega_j + i\Gamma_{it}} \cdot \tilde{\mathbf{E}}_j^* e^{i(\omega_j + \omega_{ts})t} + \frac{\mathbf{p}_{is} \mathbf{p}_{ti}}{\omega_{it} + \omega_j + i\Gamma_{it}} \cdot \tilde{\mathbf{E}}_j e^{-i(\omega_j - \omega_{ts})t} \right) + \text{c.c.}
\end{aligned}
$$

The second and third terms of the expression for $\langle \mathbf{p} \rangle_{ts}^{(1)}$ do not pertain to Rayleigh/Raman scattering[1]. Thus,

$$
\langle \mathbf{p} \rangle_{ts}^{(1)} = e^{-\Gamma_{ts} t} \frac{1}{2} \sum_{j=1}^{N_f} \frac{1}{\hbar} \sum_{i=1}^{\infty} \left(\frac{\mathbf{p}_{ti} \mathbf{p}_{is}}{\omega_{is} - \omega_j - i\Gamma_{is}} + \frac{\mathbf{p}_{is} \mathbf{p}_{ti}}{\omega_{it} + \omega_j + i\Gamma_{it}} \right) \cdot \tilde{\mathbf{E}}_j e^{-i(\omega_j - \omega_{ts})t} + \text{c.c.}
$$

If we define the transition polarizability as

$$
\left\langle \alpha \left(\omega_j \right) \right\rangle_{ts} \equiv \frac{1}{\hbar} \sum_{i=1}^{\infty} \left(\frac{\mathbf{p}_{ti} \mathbf{p}_{is}}{\omega_{is} - \omega_j - i\Gamma_{is}} + \frac{\mathbf{p}_{is} \mathbf{p}_{ti}}{\omega_{it} + \omega_j + i\Gamma_{it}} \right),
$$

then the first-order transition dipole moment becomes

$$
\langle \mathbf{p} \rangle_{ts}^{(1)} = \frac{1}{2} e^{-\Gamma_{ts} t} \sum_{j=1}^{N_f} \left\langle \alpha \left(\omega_j \right) \right\rangle_{ts} \cdot \tilde{\mathbf{E}}_j e^{-i(\omega_j - \omega_{ts})t} + \text{c.c.}
$$

Appendix B

Mueller matrices representing common optical components

Mueller matrices are four-by-four matrices that represent optical components and can operate on Stokes vectors in order to describe how the optical components affect the light described by the Stokes vectors. This complements the usage of Stokes vectors for describing the polarization properties of the light, which can be polarized, unpolarized, or partially polarized. Common Mueller matrices representing optical components include[3]:

- horizontal linear polarizer (in-[scattering-]plane polarized; $\theta = 0$):

$$M = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- vertical linear polarizer (perpendicular-to-plane polarized; $\theta = \frac{\pi}{2}$):

$$M = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- linear polarizer at $\frac{\pi}{4}$ ($\theta = \frac{\pi}{4}$) in the counterclockwise direction relative to the scattering plane:

$$M = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- linear polarizer at angle θ (the idea is to rotate the Stokes vector by $-\theta$, act on it with a horizontal linear polarizer, and then rotate by θ):

$$M = \frac{1}{2}\begin{bmatrix} 1 & \cos 2\theta & \sin 2\theta & 0 \\ \cos 2\theta & \cos^2 2\theta & \cos 2\theta \sin 2\theta & 0 \\ \sin 2\theta & \cos 2\theta \sin 2\theta & \sin^2 2\theta & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- rotator (of the polarization ellipse by θ in the counterclockwise direction):

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\theta & -\sin 2\theta & 0 \\ 0 & \sin 2\theta & \cos 2\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- quarter-wave plate:

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}$$

- half-wave plate:

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

- right circular polarizer:

$$M = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

- left circular polarizer:

$$M = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$$

- ideal depolarizer:

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Of course, the Mueller matrix that does nothing to the Stokes vector on which it operates is the identity matrix:

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Appendix C

Further analysis of the perturbing coordinate method for calculating Raman spectra

C.1 Raman spectra using 0.05-Bohr and 0.25-Bohr coordinate perturbations

See figures. A Lorentzian lineshape having full-width half-maximum of $8\,\mathrm{cm}^{-1}$ is fit to each peak.

C.2 Numerical values of the Raman activities

Note that in order to obtain these data, activities for modes within $4\,\mathrm{cm}^{-1}$ were added together. This only affected the mode(s) around $1100\,\mathrm{cm}^{-1}$ for CH_2F_2 for the 6-31G* basis set, for which Q-Chem determined there were two modes whereas Jaguar found only a single mode.

```
6-31G*:
  H2O:
    Gaussian:                    7.96    78.84    39.07
    Q-Chem:                      7.97    78.89    39.12
    Jaguar (numerical, 0.05):    7.91    78.34    38.69
    Jaguar (numerical, 0.10):    7.92    78.37    38.62
    Jaguar (numerical, 0.25):    8.14    77.92    38.23
    Jaguar (analytic,  0.05):    8.04    78.94    39.76
    Jaguar (analytic,  0.10):    8.09    78.87    39.62
    Jaguar (analytic,  0.25):    7.58    79.19    10.53
  H2S:
    Gaussian:                   46.00   156.18   119.22
    Q-Chem:                     46.01   155.95   119.15
    Jaguar (numerical, 0.05):   45.76   155.78   117.76
```

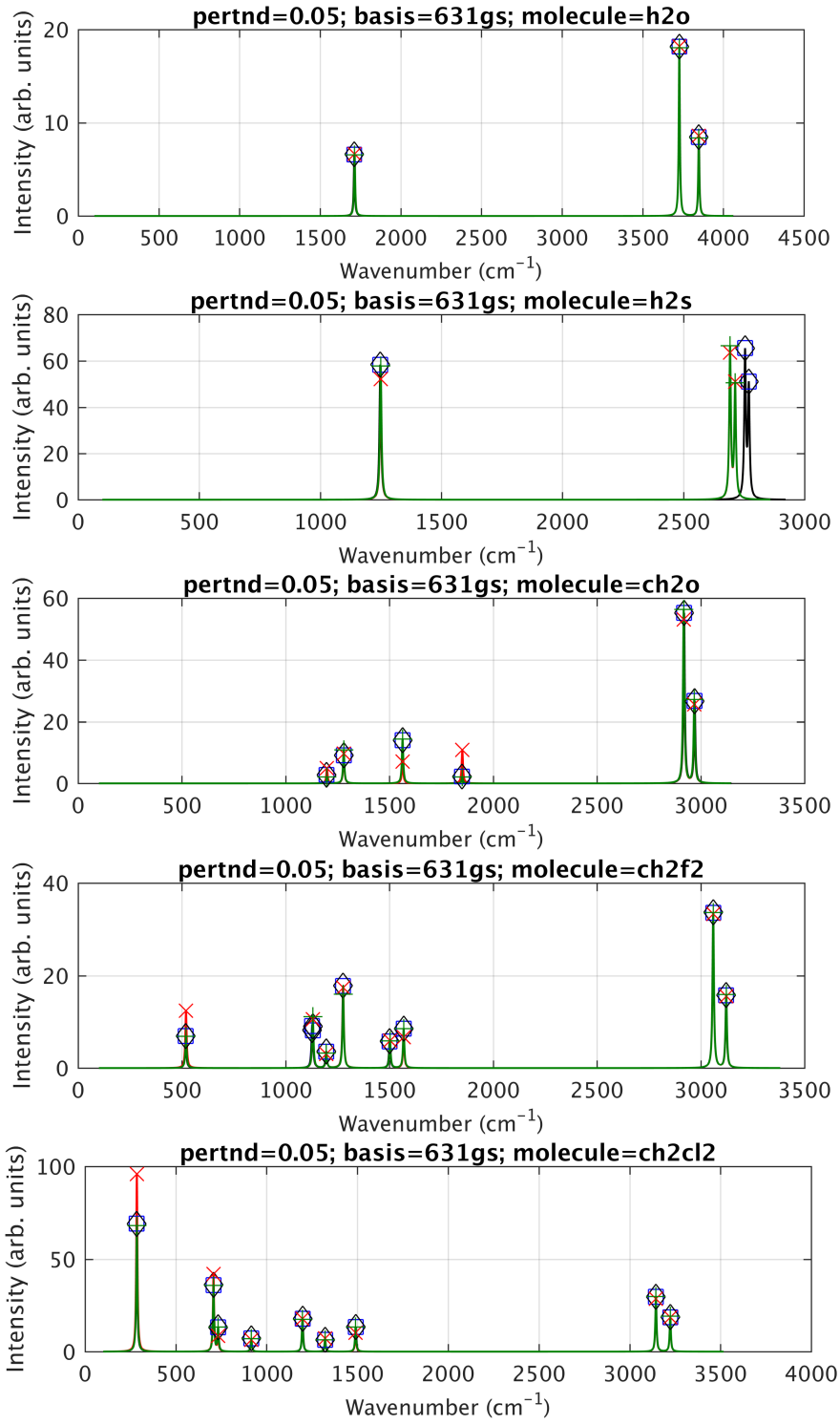

Figure C.1: Calculated Raman spectra for five molecules using the 6-31G* basis set and the perturbing coordinate method with `pertnd=0.05`. See text in Section 2.3 for details.

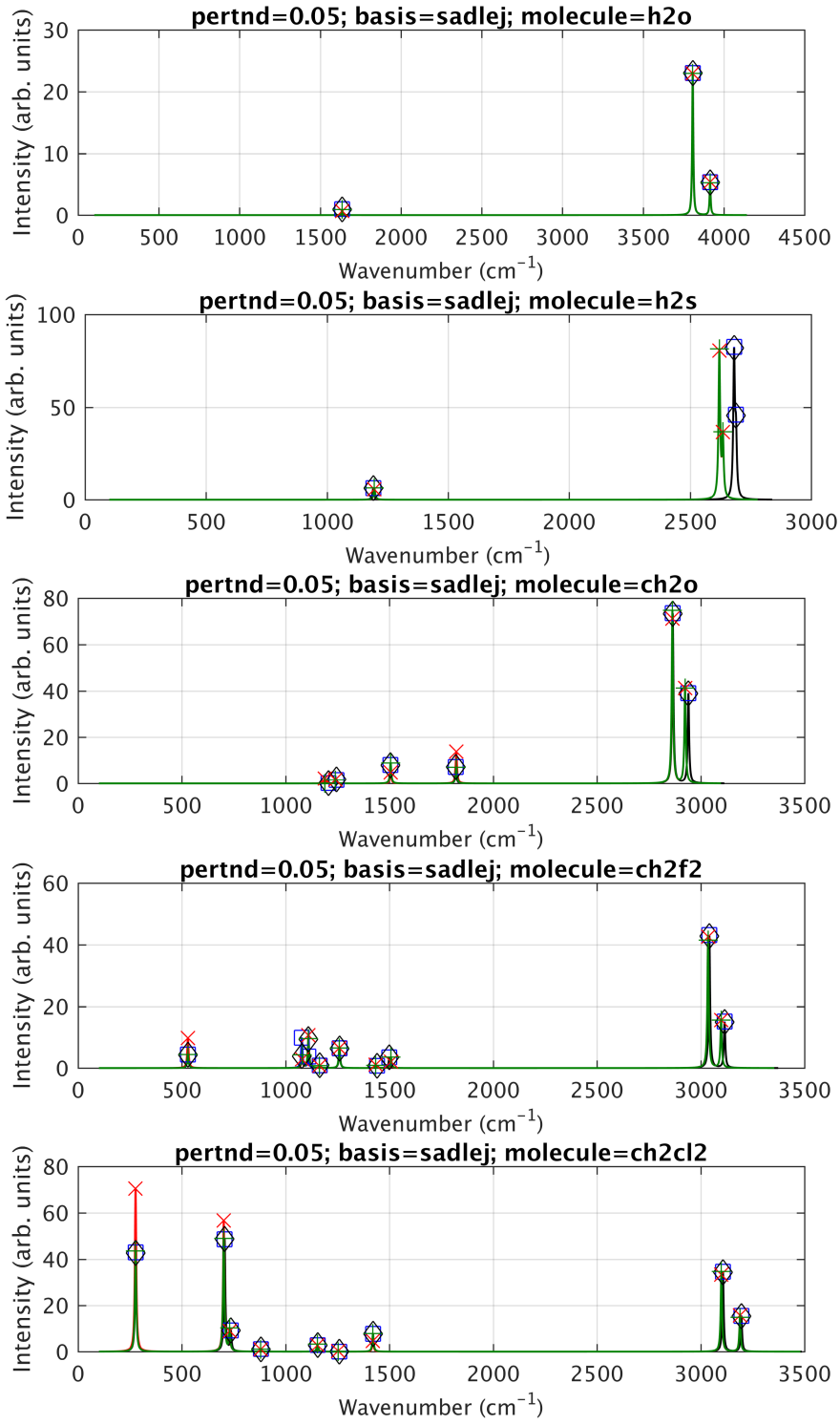

Figure C.2: Calculated Raman spectra for five molecules using the Sadlej pVTZ basis set and the perturbing coordinate method with `pertnd=0.05`. See text in Section 2.3 for details.

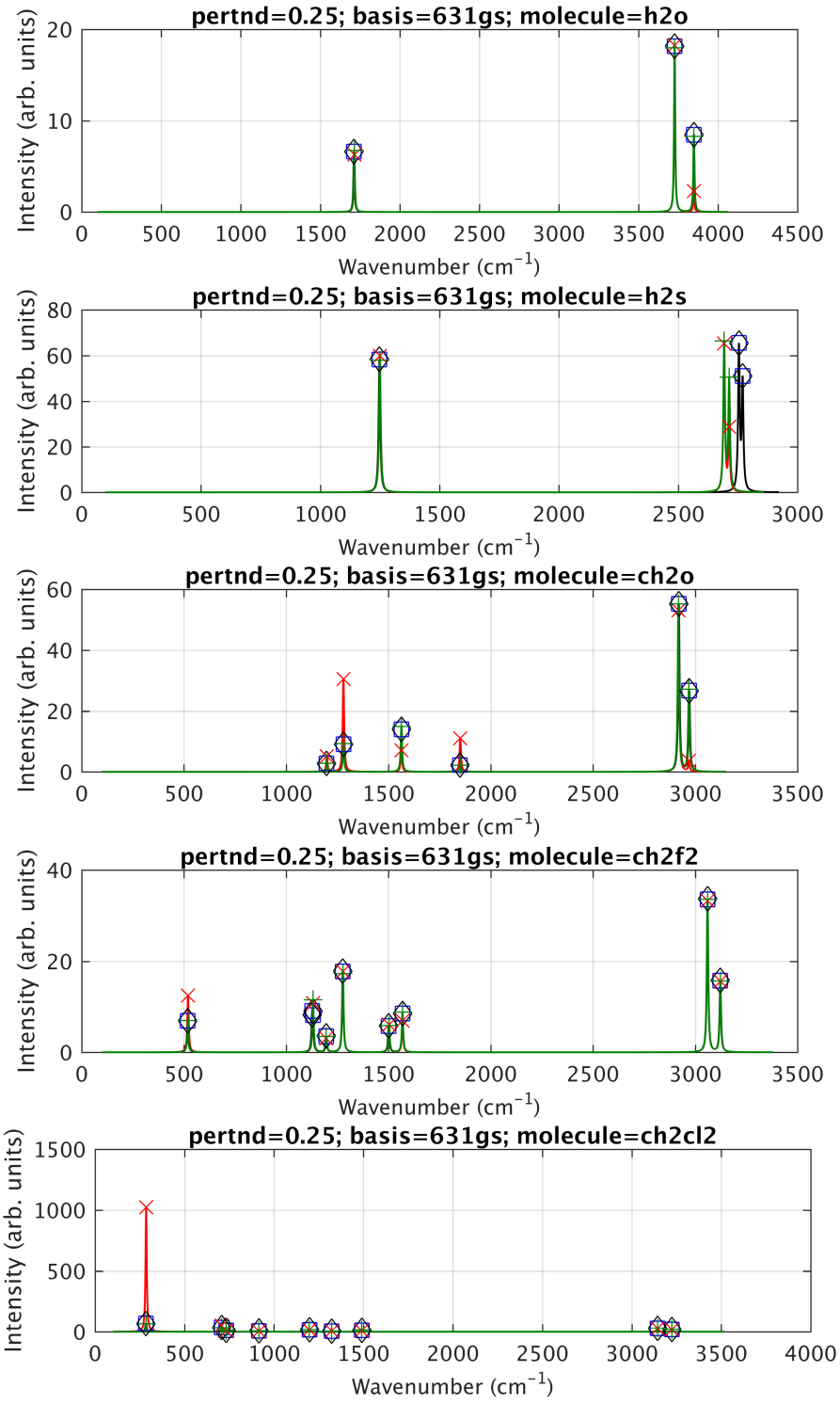

Figure C.3: Calculated Raman spectra for five molecules using the 6-31G* basis set and the perturbing coordinate method with `pertnd=0.25`. See text in Section 2.3 for details.

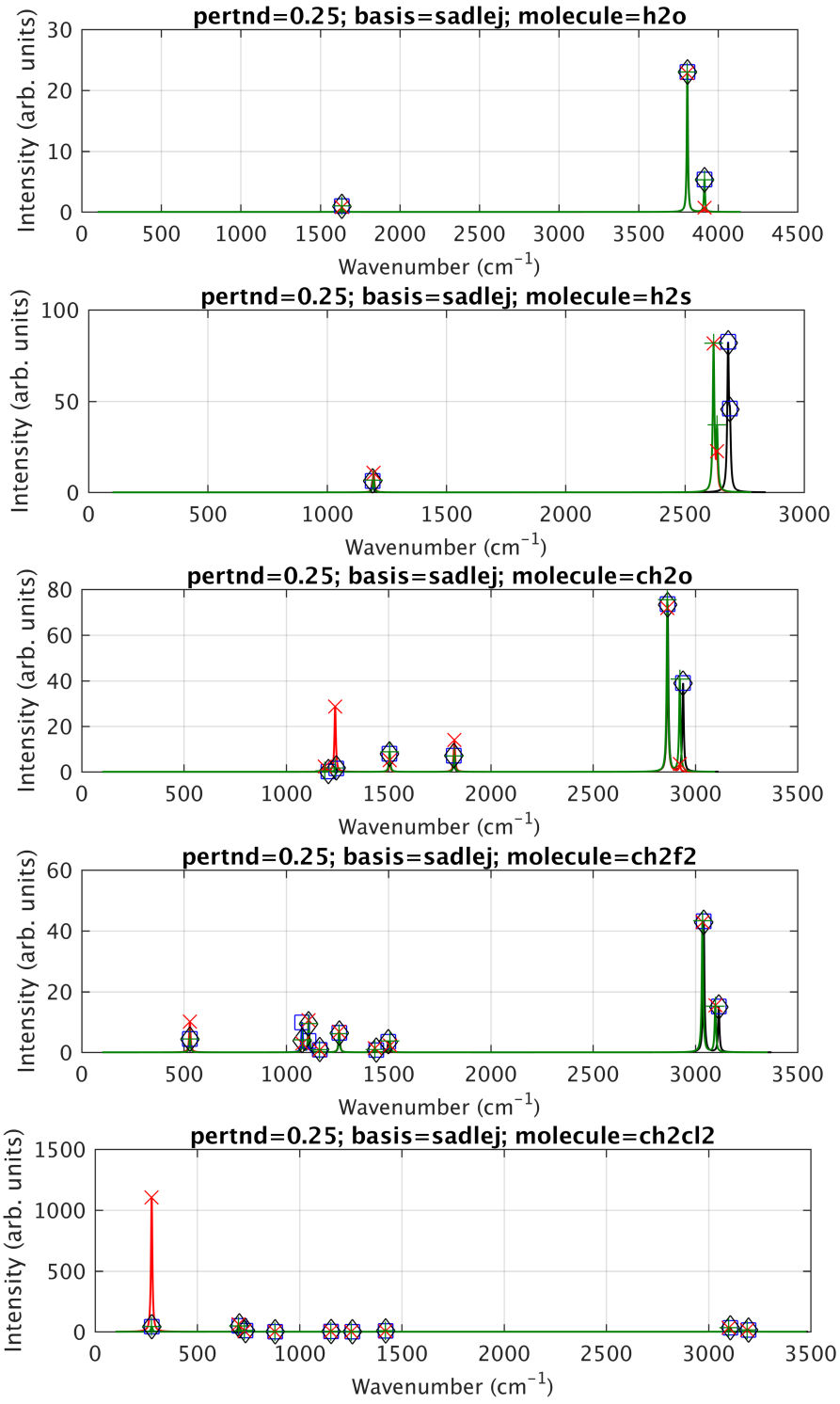

Figure C.4: Calculated Raman spectra for five molecules using the Sadlej pVTZ basis set and the perturbing coordinate method with `pertnd=0.25`. See text in Section 2.3 for details.

```
    Jaguar (numerical, 0.10):   45.77  155.70  117.99
    Jaguar (numerical, 0.25):   45.85  155.34  117.87
    Jaguar (analytic,  0.05):   41.24  148.55  119.22
    Jaguar (analytic,  0.10):   41.10  148.37  119.45
    Jaguar (analytic,  0.25):   47.36  154.61   64.86
  CH2O:
    Gaussian:                   2.12    7.48   14.97    3.01  152.02   75.12
    Q-Chem:                     2.11    7.47   14.96    3.00  151.97   75.05
    Jaguar (numerical, 0.05):   1.60    8.87   15.39    2.87  154.87   76.33
    Jaguar (numerical, 0.10):   2.00    7.75   15.60    2.91  152.73   76.63
    Jaguar (numerical, 0.25):   2.12    7.59   15.91    3.07  151.78   76.42
    Jaguar (analytic,  0.05):   3.74    7.92    7.55   14.70  146.11   72.07
    Jaguar (analytic,  0.10):   3.75    7.42    7.49   14.76  146.65   71.94
    Jaguar (analytic,  0.25):   3.84   24.91    7.56   15.01  146.33    9.61
  CH2F2:
    Gaussian:                   1.78    7.72    2.63   14.42    5.83    9.10
100.45   48.58
    Q-Chem:                     1.78    7.71    2.64   14.46    5.86    9.12
100.71   48.66
    Jaguar (numerical, 0.05):   1.78    7.74    2.38   12.97    5.93    9.12
100.73   49.09
    Jaguar (numerical, 0.10):   1.79    7.84    2.44   13.25    5.87    9.16
100.69   49.05
    Jaguar (numerical, 0.25):   1.81    8.02    2.56   14.05    5.96    9.45
100.43   48.43
    Jaguar (analytic,  0.05):   3.20    7.33    2.27   14.16    6.02    7.14
99.81   47.88
    Jaguar (analytic,  0.10):   3.20    7.37    2.33   14.25    6.06    7.18
99.80   47.93
    Jaguar (analytic,  0.25):   3.22    7.50    2.27   14.44    6.13    7.44
99.56   48.12
  CH2Cl2:
    Gaussian:                   7.53   13.71    5.01    3.84   13.40    5.58
13.39   93.14   61.33
    Q-Chem:                     7.52   13.77    5.04    3.84   13.40    5.57
13.41   93.25   61.45
    Jaguar (numerical, 0.05):   7.48   13.67    5.06    3.80   13.07    5.54
13.42   93.90   63.32
    Jaguar (numerical, 0.10):   7.46   13.67    4.97    3.96   13.08    5.58
13.47   93.91   62.23
    Jaguar (numerical, 0.25):   7.45   13.82    4.90    3.87   13.13    5.63
14.03   94.69   62.44
    Jaguar (analytic,  0.05):  10.53   16.07    3.09    3.47   13.37    6.14
10.30   90.12   61.26
    Jaguar (analytic,  0.10):  10.53   16.10    3.08    3.47   13.37    6.18
10.37   90.33   61.19
    Jaguar (analytic,  0.25): 112.48   19.58    3.05    3.41   13.42    6.22
11.08   91.31   61.39
  N2:
```

```
    Gaussian:                   14.24
    Q-Chem:                     14.24
    Jaguar (numerical, 0.05):   14.20
    Jaguar (numerical, 0.10):   14.09
    Jaguar (numerical, 0.25):   13.60
    Jaguar (analytic,  0.05):   14.34
    Jaguar (analytic,  0.10):   14.32
    Jaguar (analytic,  0.25):   14.15
Sadlej pVTZ:
  H2O:
    Gaussian:                   1.10  104.08   25.57
    Q-Chem:                     1.07  104.21   25.41
    Jaguar (numerical, 0.05):   1.08  104.19   25.22
    Jaguar (numerical, 0.10):   1.09  104.26   25.19
    Jaguar (numerical, 0.25):   1.19  104.42   25.31
    Jaguar (analytic,  0.05):   0.76  104.16   25.71
    Jaguar (analytic,  0.10):   0.77  103.65   25.47
    Jaguar (analytic,  0.25):   0.94  103.43    3.36
  H2S:
    Gaussian:                   4.79  182.09   73.99
    Q-Chem:                     4.73  181.11   73.12
    Jaguar (numerical, 0.05):   4.86  182.31   72.78
    Jaguar (numerical, 0.10):   4.85  182.55   73.31
    Jaguar (numerical, 0.25):   5.06  182.98   73.32
    Jaguar (analytic,  0.05):   4.28  180.43   72.32
    Jaguar (analytic,  0.10):   4.27  180.62   72.39
    Jaguar (analytic,  0.25):   8.06  185.07   39.62
  CH2O:
    Gaussian:                   0.14    1.19    8.10    9.38  196.14  108.37
    Q-Chem:                     0.13    1.38    8.07    9.38  195.44  108.00
    Jaguar (numerical, 0.05):   0.03    1.22    9.02    9.31  198.90  113.24
    Jaguar (numerical, 0.10):   0.43    1.21    8.85    9.38  198.70  112.39
    Jaguar (numerical, 0.25):   0.31    1.34    8.95    9.34  200.87  111.93
    Jaguar (analytic,  0.05):   1.59    1.31    4.90   18.21  189.49  113.29
    Jaguar (analytic,  0.10):   1.60    1.32    4.89   18.27  189.79  112.60
    Jaguar (analytic,  0.25):   1.68   22.45    5.27   18.46  191.34    9.00
  CH2F2:
    Gaussian:                   1.18    6.42    2.49    0.61    5.16    0.75
3.57  127.90   46.35
    Q-Chem:                     1.15    2.46    6.37    0.68    5.10    0.75
3.57  126.77   45.88
    Jaguar (numerical, 0.05):   1.17    2.74    6.47    0.49    5.22    0.83
3.71  122.56   47.43
    Jaguar (numerical, 0.10):   1.17    2.69    6.42    0.52    4.54    0.86
3.67  125.02   46.90
    Jaguar (numerical, 0.25):   1.19    2.64    6.44    0.55    4.91    0.91
3.84  128.03   46.21
    Jaguar (analytic,  0.05):   2.59    1.79    7.27    0.50    5.23    1.06
2.16  125.68   47.44
```

```
    Jaguar (analytic,  0.10):   2.60    1.79    7.31    0.52    5.32    1.17
2.20  126.36   47.45
    Jaguar (analytic,  0.25):   2.68    1.84    7.21    0.52    5.43    1.31
2.37  126.58   47.06
  CH2Cl2:
    Gaussian:                   4.53   18.56    3.40    0.55    2.08    0.04
7.30  106.33   50.34
    Q-Chem:                     4.50   18.63    3.39    0.53    2.07    0.04
7.31  105.59   49.91
    Jaguar (numerical, 0.05):   4.57   18.55    3.76    0.52    2.39    0.03
7.47  106.31   48.10
    Jaguar (numerical, 0.10):   4.58   18.56    3.53    0.47    2.39    0.04
7.51  106.43   50.21
    Jaguar (numerical, 0.25):   4.65   18.67    3.31    0.48    2.36    0.05
7.81  107.56   51.18
    Jaguar (analytic,  0.05):   7.40   21.49    3.17    0.48    2.03    0.21
4.42  102.17   50.57
    Jaguar (analytic,  0.10):   7.39   21.51    3.09    0.49    2.06    0.22
4.44  102.44   50.61
    Jaguar (analytic,  0.25): 115.76   24.42    2.91    0.52    2.17    0.25
4.67  103.87   50.92
  N2:
    Gaussian:                  26.67
    Q-Chem:                    25.55
    Jaguar (numerical, 0.05):  25.51
    Jaguar (numerical, 0.10):  25.48
    Jaguar (numerical, 0.25):  25.28
    Jaguar (analytic,  0.05):  25.55
    Jaguar (analytic,  0.10):  25.52
    Jaguar (analytic,  0.25):  25.29
```

Appendix D

Comparing Raman activities calculated using non-optimal `refield` values vs. optimal `refield` values

Raman activities are calculated using the Rappoport basis set using the single intensity formula of eq. 1.33 for our test set of 15 small molecules. Results are presented for the pseudospectral ("ps") and PRISM ("prism") methods. "Best" refers to use of the error-minimizing `refield` values shown in Fig. 2.6, which for the pseudospectral method are 0.022 a.u. for the diagonal tensor components and 0.014 a.u. for the off-diagonal components, and for the PRISM method are 0.018 a.u. for the diagonal components and 0.014 a.u. for the off-diagonal components. "Worst" refers to use of the smallest `refield` values that produce tensor errors of at most three times the errors of the optimal `refield` values: for the pseudospectral method, these are 0.010 a.u. for the diagonal tensor components and 0.006 a.u. for the off-diagonal components, and for the PRISM method, these are 0.008 a.u. for the diagonal components and 0.004 a.u. for the off-diagonal components.

```
c2h2:
  ps:
    Best:      2.67     2.67     0.00     0.00    99.05     0.00    57.10
    Worst:     2.65     2.65     0.00     0.00    98.31     0.02    56.17
  prism:
    Best:      2.74     2.74     0.00     0.00    98.60     0.00    56.76
    Worst:     2.71     2.71     0.00     0.00    92.96     0.00    53.98
c2h4:
  ps:
```

```
    Best:      0.00     0.00     3.29     0.00     0.23    36.19     0.00    26.39
0.00   221.01   121.45     0.00
    Worst:     0.00     0.00     3.27     0.00     0.24    36.39     0.00    26.43
0.00   217.23   120.74     0.00
  prism:
    Best:      0.00     0.00     3.40     0.00     0.23    35.92     0.00    26.39
0.00   220.50   121.02     0.00
    Worst:     0.00     0.00     3.37     0.00     0.23    36.14     0.00    24.26
0.00   210.22   120.28     0.00
c2h6:
  ps:
    Best:      0.00     0.00     0.00    11.57     0.05     0.04     0.00     0.08
9.98    10.03     0.00     0.00   342.19     0.00   122.88   123.72     0.00     0.00
    Worst:     0.00     0.00     0.00    11.46     0.04     0.04     0.00     0.09
10.01    10.03     0.00     0.00   333.45     0.00   122.14   122.32     0.00     0.00
  prism:
    Best:      0.00     0.00     0.00    11.29     0.04     0.05     0.00     0.08
9.91     9.93     0.00     0.00   342.94     0.00   122.78   123.44     0.00     0.00
    Worst:     0.00     0.00     0.00    11.19     0.05     0.05     0.00     0.09
9.81     9.81     0.00     0.00   342.54     0.00   122.46   122.74     0.00     0.00
c7h8:
  ps:
    Best:      0.43     1.59     0.46     0.00     0.16     5.83     3.66     0.02
0.29    18.21     0.18     0.17     0.46     0.00     0.01    30.02     0.56    17.80
0.62     3.74     4.92    14.60     1.16     0.61    13.39     6.50     6.52     0.68
0.27     8.00    25.34   244.31    79.16    55.96    56.00   125.92    35.37    59.74
299.02
    Worst:     0.43     1.60     0.49     0.00     0.17     6.13     3.66     0.01
0.30    19.69     0.17     0.16     0.37     0.00     0.02    32.10     0.71    18.85
0.63     3.79     5.16    15.98     1.10     0.55     9.08     6.36     6.37     0.63
0.29     8.00    24.22   218.08    78.51    55.40    55.55   122.84    35.01    59.07
293.79
  prism:
    Best:      0.43     1.59     0.44     0.00     0.16     5.83     3.67     0.02
0.28    17.53     0.18     0.17     0.53     0.00     0.01    29.11     0.51    17.66
0.72     3.83     4.90    13.42     1.12     0.58    13.26     6.33     6.44     0.72
0.19     7.98    24.78   249.96    78.97    56.80    56.37   124.28    33.29    60.36
297.30
    Worst:     0.41     1.59     0.47     0.00     0.18     6.22     3.68     0.01
0.27    17.50     0.17     0.17     0.58     0.00     0.01    28.02     0.50    13.17
0.41     3.49     4.82    12.65     1.52     0.52    10.54     6.52     6.40     0.57
0.88     8.05    23.03   237.37    78.67    54.45    54.31   123.98    40.39    57.81
256.29
ccl4:
  ps:
    Best:      2.31     2.31     3.08     3.08     3.08    22.60     5.31     5.31
5.33
    Worst:     2.28     2.29     3.07     3.07     3.07    20.37     5.48     5.49
5.58
```

```
  prism:
    Best:      2.30     2.30     3.09     3.09     3.09    23.07     5.19     5.20
5.20
    Worst:     2.30     2.30     3.10     3.10     3.10    22.61     5.02     5.02
5.02
ch2cl2:
  ps:
    Best:      4.29    18.65     3.78     0.53     2.02     0.07     7.43   100.82
48.27
    Worst:     4.30    18.36     3.63     0.52     2.03     0.08     7.28   101.08
48.12
  prism:
    Best:      4.33    18.56     3.75     0.54     2.02     0.07     7.36    99.96
48.18
    Worst:     4.41    18.82     3.77     0.55     2.03     0.06     7.26    95.15
47.96
ch2f2:
  ps:
    Best:      1.15     2.37     6.16     0.70     4.64     0.81     3.48   120.71
42.96
    Worst:     1.15     2.34     6.26     0.70     4.64     0.84     3.48   118.97
42.75
  prism:
    Best:      1.16     2.27     6.19     0.64     4.52     0.78     3.50   121.14
42.87
    Worst:     1.19     2.11     6.36     0.57     4.19     0.48     3.50   105.04
43.32
ch2o:
  ps:
    Best:      0.28     1.40     8.32     9.54   176.38   100.81
    Worst:     0.28     1.39     8.31     8.74   164.18   100.40
  prism:
    Best:      0.29     1.36     8.30     9.74   177.51    99.71
    Worst:     0.26     1.31     8.46     9.57   172.64   100.73
ch4:
  ps:
    Best:      0.00     0.00     0.00     3.97     3.99   208.98    44.80    44.78
44.78
    Worst:     0.00     0.00     0.00     4.06     4.08   202.97    44.71    44.75
44.74
  prism:
    Best:      0.00     0.00     0.00     4.00     4.00   210.32    44.62    44.61
44.61
    Worst:     0.00     0.00     0.00     4.06     4.06   211.67    44.09    44.08
44.09
chcl3:
  ps:
    Best:      2.78     2.76     8.41    14.73     4.62     4.36     2.41     2.34
62.17
```

```
    Worst:    2.77    2.75    8.24   13.04    4.34    4.52    2.37    2.33
60.55
  prism:
    Best:     2.76    2.76    8.48   15.03    4.52    4.33    2.37    2.30
61.57
    Worst:    2.76    2.77    8.53   14.99    4.34    4.16    2.31    2.25
59.76
h2o:
  ps:
    Best:     0.91  102.03   25.54
    Worst:    1.04  110.71   24.32
  prism:
    Best:     0.90   99.70   25.59
    Worst:    0.92  100.08   25.51
h2s:
  ps:
    Best:     4.68  179.00   72.21
    Worst:    4.99  186.79   69.49
  prism:
    Best:     4.64  175.88   72.06
    Worst:    4.77  175.45   71.33
h3n:
  ps:
    Best:     0.15    1.81    1.82  151.03   35.86   36.20
    Worst:    0.04    1.82    1.80  155.13   35.72   36.04
  prism:
    Best:     0.15    1.81    1.81  149.86   35.59   35.84
    Worst:    0.17    1.81    1.81  148.65   35.42   35.51
n2:
  ps:
    Best:    18.34
    Worst:           16.06
  prism:
    Best:    17.96
    Worst:           17.29
o2si:
  ps:
    Best:     0.00    0.00   46.29    0.00
    Worst:    0.00    0.00   47.27    0.01
  prism:
    Best:     0.00    0.00   45.85    0.00
    Worst:    0.00    0.00   43.93    0.00
```

Here are the minimum, MAD, and maximum activities for each molecule, followed by the aggregate minimum, average MAD, and aggregate maximum values over all 15 molecules:

PS:

Mol	Min	MAD	Max
c2h2	0.00	0.24	0.93
c2h4	0.00	0.39	3.78
c2h6	0.00	0.61	8.75
c7h8	0.00	1.28	26.22
ccl4	0.01	0.32	2.23
ch2cl2	0.00	0.11	0.29
ch2f2	0.00	0.24	1.74
ch2o	0.00	2.24	12.20
ch4	0.00	0.71	6.02
chcl3	0.01	0.44	1.69
h2o	0.12	3.34	8.68
h2s	0.31	3.61	7.79
h3n	0.01	0.76	4.10
n2	2.28	2.28	2.28
o2si	0.00	0.25	0.98
Aggregate:	0.00	1.12	26.22

PRISM:

Mol	Min	MAD	Max
c2h2	0.00	1.21	5.64
c2h4	0.00	1.12	10.27
c2h6	0.00	0.10	0.70
c7h8	0.00	2.10	41.01
ccl4	0.00	0.11	0.46
ch2cl2	0.00	0.61	4.80
ch2f2	0.01	1.96	16.10
ch2o	0.02	1.05	4.87
ch4	0.00	0.34	1.35
chcl3	0.00	0.26	1.81
h2o	0.02	0.16	0.38
h2s	0.13	0.43	0.73
h3n	0.00	0.29	1.21
n2	0.68	0.68	0.68
o2si	0.00	0.48	1.92
Aggregate:	0.00	0.73	41.01

Appendix E

Electronic structure software keywords

The following keywords are most likely the ones used for most jobs, though there may be some variations not noted below.

E.1 Keywords used for Raman calculations

Jaguar	Q-Chem	
basis=rappoport-svpd	basis	gen
dftname=b3lyp	method	b3lyp
ifreq=1	jobtype	freq
iraman=1	doraman	true
refield=xxxx		

For example, to achieve the optimal settings for Raman calculations in Jaguar, as discussed in Chapter 2, we set `refield=0.022` for the diagonal components of the Raman tensor and `refield=0.014` for the off-diagonal components. Note that the "gen" is set for the `basis` keyword in Q-Chem because we had to enter a custom basis in order to use the Rappoport SVPD basis, which is not implemented in Q-Chem.

E.2 Jaguar keywords used for "Live detection of biomolecule activity using stimulated Raman spectroscopy" (Chapter 6)

Geometry optimizations and accurate frequency calculations

```
basis=6-31g*
dftname=b3lyp
gdftcphf=-13
gdftder2=-13
ifreq=1
igeopt=1
isolv=2
isqm=1
maxitg=10
nofail=1
nogas=2
```

In addition, here is a sample `&atomic` section for specifying isotopes:

```
atom    isotope
c5      13
```

E.3 Keywords used for "Benchmarking the pseudospectral method for obtaining excited-state energies and gradients of large molecules" (Chapter 7)

Jaguar benchmarking

Base keywords (ground state energies):

```
basis=6-31g**
dftname=b3lyp
ip351=2
ip6=-3
isymm=0
```

Modifications for ground state gradients:

```
igeopt=-1
```

Modifications (of base keywords) for excited state energies:

```
itda=1
itddft=1
rsinglet=1
rtriplet=0
```

Modifications for excited state gradients:

```
igeopt=-1
itda=1
itddft=1
rsinglet=1
rtriplet=0
```

Modification for excited state energies with diffuse basis functions:

```
basis=6-31g****
```

Note for the diffuse basis function jobs, some jobs needed additional keywords to prevent the SCF procedure from dying due to large DIIS errors:

```
iacc=1
iguess=11
nofail=1
```

Q-Chem benchmarking

Base keywords (ground state energies):

```
basis                    6-31g**
cistr_print              true
exchange                 b3lyp
geom_opt_symflag         0
scf_final_print          1
scf_print                1
sym_ignore               true
symmetry_decomposition   0
symmetry                 false
```

Modifications for ground state gradients:

```
jobtype force
```

Modifications (of base keywords) for excited state energies:

```
cis_n_roots     1
cis_singlets    true
cis_triplets    false
```

Modifications for excited state gradients:

```
cis_n_roots        1
cis_singlets       true
cis_state_deriv    1
cis_triplets       false
jobtype            force
```

Modification for excited state energies with diffuse basis functions:

```
basis     6-31g**++
```

Q-Chem benchmarking using resolution-of-the-identity gradients

```
ari                          true
aux_basis                    rimp2-cc-pvtz
basis                        6-31g**
cistr_print                  true
exchange                     b3lyp
geom_opt_symflag             0
jobtype                      force
scf_final_print              1
scf_print                    1
sym_ignore                   true
symmetry_decomposition       0
symmetry                     false
```

E.4 Jaguar keywords used for "Density functional theory study of the rate-limiting step of water splitting on large titanium dioxide nanoparticles" (Chapter 8)

Note that for very large clusters, a lower `cut20` may be needed in order to obtain the correct LUMO (as seen in orbital surface plots). In one case (an older 05x05 cluster), 1.2e-3 was needed. If there are SCF convergence issues, `cut20` may need to be raised, though note that SCF convergence issues may also arise if `cut20` is too large. For the 03x03 cluster, 1.6e-3 seemed to work well. In addition, sometimes we set `maxit=8` to speed up

the geometry optimizations; this should be fine to do at least early on in the optimizations. Finally, for the bare-minimum-sized nanoparticles (such as those containing a single Ti atom), `iacscf=3` was needed.

Fragmented Initial Guess (FIG) code

```
allow_scf_failure=1
basis=lacvp
cut20=1.50e-003
dconv=3.00e-006
dftname=b3lyp
econv=2.00e-005
iaccg=3
iacscf=4
iconv=1
igeopt=0
iguess=11
iorb1a=0
iorb2a=0
isymm=0
itradj=1
ldips=1
maxit=3
maxitg=300
molchg=0
mulken=0
multip=1
nogas=2
nops=0
solvent=water
```

The call of the FIG code generally requires older versions of Jaguar that employ MPI, and a sample call to the FIG code on a 48-core machine is:

```
export SCHRODINGER=/home/friesner/alw2165/software/schrodinger-2015-1
nohup python ~/fig/fig-orig.py -V -p 48 -m 15 -k -j /scratch -i \
  333_with_unhappy_o_atoms.in 1> output.txt 2> errors.txt &
```

A simpler but less rigorous way to fragment the nanoparticle (instead of using the more complex MATLAB code), using only a `&gen` section as the input, is:

```
./fragment_coords.sh 333_with_happy_o_atoms.in 68
```

Geometry optimizations

```
allow_scf_failure=1
basis=lacvp
cut20=1.60e-003
dconv=3.00e-006
dftname=b3lyp
econv=2.00e-005
iaccg=3
iacscf=4
igeopt=1
iguess=1
iorb1a=0
iorb1b=0
iorb2a=0
iorb2b=0
ipvirt=10
isolv=2
isymm=0
itradj=1
ldips=1
maxitg=300
molchg=xxxx
mulken=0
multip=xxxx
nogas=2
nops=0
solvent=water
tradmn=xxxx
tradmx=xxxx
trust=xxxx
```

Step sizes

	Small	Medium 1	Medium 2	Large
tradmn	0.06	0.06	0.06	0.06
tradmx	0.20	0.50	0.65	0.85
trust	0.15	0.45	0.60	0.80

Properties calculations

```
allow_scf_failure=0
basis=lacvp
cut20=1.60e-003
```

```
dconv=3.00e-006
dftname=b3lyp
econv=1.00e-005
iaccg=3
iacscf=4
igeopt=0
iguess=1
iorb1a=homo-5
iorb1b=homo-5
iorb2a=lumo+5
iorb2b=lumo+5
ipvirt=10
isolv=2
isymm=0
itradj=1
ldips=5
maxit=250
maxitg=1
molchg=xxxx
mulken=1
multip=xxxx
nogas=2
nops=0
solvent=water
```

If excited state energies are desired:

```
itddft=1
nroot=5
rsinglet=xxxx
rtriplet=xxxx
```

Transition state and intrinsic reaction coordinate (IRC) calculations

Transition state calculation	IRC calculation
allow_scf_failure=0 basis=lacvp cut20=1.60e-003 dconv=3.00e-006 dftname=b3lyp econv=1.00e-005 iaccg=3 iacscf=4 igeopt=2 iguess=1 iorb1a=homo-1 iorb1b=homo-1 iorb2a=lumo+1 iorb2b=lumo+1 ipvirt=10 iqst=1 isolv=2 isymm=0 itradj=1 itrvec=-5 ldips=1 maxit=250 maxitg=10 molchg=xxxx mulken=1 multip=xxxx nogas=2 nops=0 solvent=water tradmn=0.06 tradmx=0.2 trust=0.15	allow_scf_failure=0 basis=lacvp cut20=1.60e-003 dconv=3.00e-006 dftname=b3lyp econv=1.00e-005 iaccg=3 iacscf=4 iguess=1 inhess=4 ipvirt=10 irc=2 ircmax=7 ircmxcyc=8 ircstep=0.3 isolv=2 isymm=0 itradj=1 ldips=1 maxit=250 maxitg=20 molchg=1 mulken=1 multip=2 nogas=2 nops=0 solvent=water tradmn=0.06 tradmx=0.65 trust=0.60

Appendix F

Papers and collaborators

All collaborators were at Columbia at the time of study unless otherwise specified. Prof. Richard Friesner is of course an integral part to all projects.

- General polarized Raman intensity formulas (Chapter 1), Raman misconceptions (Section 1.9), polarized Raman software (Chapter 5)
  - *First-author paper in preparation*
- Implementation of polarized Raman spectroscopy using the pseudospectral method (Chapter 2)
  - Collaborators: Yixiang Cao (Schrödinger Inc.)
  - *First-author paper in preparation*
- Using polarized Raman spectra to determine unit cell orientation: application to oligothiophene crystals (Chapter 3)
  - Collaborators:
    - UT Austin (experiment): John Heckel, Leif Sherry, Sarah Stranahan, Katherine Willets
    - Columbia (theory): Severin Schneebeli, Michelle Hall, Kateri DuBay
  - *First-author paper (co-authored with John Heckel) published in 2012[43]*
- A first-principles polarized Raman method for determining whether a uniform region of a sample is crystalline or isotropic (Chapter 4)

  - Collaborators: Kateri DuBay, Katherine Willets (UT Austin)
  - *First-author paper published in 2014[155]*

- Live detection of biomolecule activity using stimulated Raman spectroscopy (Chapter 6)
  - Collaborators: Zhixing Chen, Daniel Paley, Lu Wei, Colin Nuckolls, Wei Min
  - *Regular-author paper published in 2014 on alkyne vibrational palette[156]*
  - *Regular-author paper in preparation on design and synthesis of next generation Raman probes for multi-color live cell analysis*

- Benchmarking the pseudospectral method for obtaining excited-state energies and gradients (Chapter 7)
  - Primary collaborators: Yixiang Cao (Schrödinger Inc.)
  - *Regular-author paper published in 2016 on pseudospectral implementation of excited-state energies[114]*
  - *Regular-author paper in preparation on pseudospectral implementation of excited-state energy gradients*

- Study of the first proton-coupled electron transfer of the oxidation evolution reaction on anatase titanium dioxide nanoparticles (Chapter 8)
  - Collaborators: Jing Zhang, Patrick Saitta, Jerome Vienne (Texas Advanced Computing Center, UT Austin), Michael Steigerwald, Louis Brus
  - *Regular-author paper published in 2015 including our general fragmentation scheme for input into the FIG code[157]*
  - *First-author paper in preparation on the first PCET of the OER*