# Atomistic Modelling of Aggregate Chlorophyll Systems

By

OLIVER J. H. FEIGHAN


School of Chemistry
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Science.

JANUARY, 2022

Word count: 0

## Abstract

Here goes the abstract

# Dedication and acknowledgements

Here goes the dedication.

## Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE: .....

# Table of Contents

# List of Tables

**Table** **Page**

# List of Figures

**Figure** **Page**

CHAPTER 1

# INTRODUCTION

Naturally occuring light harvesting systems present an interesting scientific challenge. With near perfect efficiency, the energy from a photon will be taken and transfered to a reaction centre, leading to charge transfer processes that culminate in powering biological systems. Making models that can predict and explain these effects are key to making similarly efficient photovoltaic systems.

## 1.1 Quantum Exploits in Light Harvesting Systems

Begins a section.

### 1.1.1 Electronic Energy Transfer

Begins a subsection.

### 1.1.2 Coherence

## 1.2 Light-Matter Response

## 1.3 Electronic Structure for Large Systems

## 1.4 The Aim

CHAPTER 2

# MEAN-FIELD EXCITED STATES

This chapter collates all of the work done on investigating and benchmarking Δ-SCF methods, at both an ab initio DFT level of theory as well as with semi-empirical, tight-binding methods. To start the investigation into what is required for accurate atomistic treatment of chlorophyll systems, we looked at these methods as they have reported as being able to calculate accurate response properties with reduced computational cost. We calculated transition properties for a range of molecules, as well as for a small set of chlorophyll molecules, using variety of different basis sets, density functionals, response methods and electronic structure methods. All of the work was compared to a high-standard EOM-CCSD reference, from which we made conclusions on the accuracy of each method. We also investigated a other issues of the mean-field Δ-SCF method, such as non-orthogonality in the transition properties, and assignment of the symmetries of transition. These investigations, as well as the work on using semi-empirical methods, is all novel work. We found that whilst the high-level methods give reasonable results, the semi-empirical Δ-SCF methods are not as accurate. This results is unexpected in the context of the benchmarking, however has reasonable explanations, and will guide the work of the next chapter on developing a new semi-empirical method. This work sets the groundwork of what could be reasonably achieved against the goals set out in the introductionary chapter.

## 2.1 Theory

### 2.1.1 Δ-SCF and eigenvalue difference

Δ-SCF predicts the excitation energy of a system by comparing the single point energy of the ground state and the excited state. Finding this excited state correctly can be an issue, but is usually assumed to be similar to the ground state. In its simplest form, the Δ-SCF method

calculates the ground state, and then calculates the excited state by rerunning an self-consistent field (SCF) with the excited state occupation numbers. This then gives a full description of both the ground as excited state from the orbital coefficients output from the two SCF procedures.

Initially, the excited state could be calculated by relaxing the orbitals which contain the excited electron and hole in the ground state space, so that the excited state and ground state are orthogonal[4]. However, it was argued that this procedure would exacerbate errors from finding the ground state, and that the excited state was not a proper SCF solution[1]. Alternatively, it was proposed that an SCF like method, where instead of populating orbitals according to the aufbau principle, orbitals which most resemble the previous iteration's orbitals should be occupied. Each iteration in an SCF procedure produces new molecular orbital coefficients by solving the Roothaan-Hall equations[8], generally given as an eigenvalue problem:

$$\mathbf{F}\mathbf{C}^{\text{new}} = \mathbf{S}\mathbf{C}^{\text{new}}\epsilon \tag{2.1}$$

where $\mathbf{C}^{\text{new}}$ are the next orbital coefficient solutions, $\mathbf{S}$ is the overlap, and $\epsilon$ are the orbital energies. The Fock matrix $\mathbf{F}$ is calculated from the previous set of orbital coefficients:

$$\mathbf{F} = f\left(\mathbf{C}^{\text{old}}\right) \tag{2.2}$$

The amount of similarity of orbitals can be estimated from their overlap:

$$\mathbf{O} = \left(\mathbf{C}^{\text{old}}\right)^{\dagger}\mathbf{S}\mathbf{C}^{\text{new}} \tag{2.3}$$

and for a single orbital can be evaluated as a projection:

$$p_j = \sum_i^n O_{ij} = \sum_\nu^N \left[\sum_\mu^N \left(\sum_i^n C_{i\mu}^{\text{old}}\right) S_{\mu\nu}\right] C_{\nu j}^{\text{new}} \tag{2.4}$$

where $\mu, \nu$ are orbital indices. The population can then be given by the set of orbitals with the highest projection $p_j$. This method can be used for any excited state, with the caveat that the orbital solution is in the same region as the ground state solution. For a few low lying states, this is generally true, and so $\Delta$-SCF can be used to calculate a spectrum of excited states[1]. The method of using this orbital overlap is called the maximum overlap method (MOM).

$\Delta$-SCF has been shown to be cheap alternative to TDDFT and other higher level methods, without considerable losses of accuracy in certain cases[9]. Additionally, as the excited state is given as solutions to SCF equations, the gradient of this solution can be given by normal mean-field theory. These gradients would be much cheaper than TDDFT or coupled cluster methods, and so would be advantageous for a dynamic simulation of LHII.

The final descent in response theory would be to eigenvalue difference methods. Here there is assumed to be no response of the orbital energies and shapes when interacting with light. As stated earlier, this would be recovered from the complete Cassida equation if the coupling elements in the **A** and **B** matrices are set to zero. This means that the difference between the excited state energy and the ground state energy is just the difference of the orbital energies between the orbital an electron has been excited to and the orbital has been excited from. Additionally, transition properties can be calculated by calculating transition density matrices from only the ground state. Hence, all the information needed can be given by a single SCF optimization. Generally, eigenvalue difference methods are not seen as accurate response methods, but can offer a quick and easy initial value[2].

### 2.1.2 Transition Density and Dipole Moments

Transition properties, such as the transition dipole moment, can be calculated from the SCF solutions for the ground and excited states. The reduced one-particle transition density matrix can be written as:

$$\mathbf{D}^{21} = |\Psi_1\rangle \langle\Psi_2| \tag{2.5}$$

where $|\Psi_n\rangle$ is the Slater determinant of state $n$, constructed from the set of spin orbitals $\{\phi_j^{(n)}\}$. Expressed in terms of the molecular orbitals coefficients $\mathbf{C}^{(n)}$, the transition density matrix is given by

$$\mathbf{D}^{21} = \mathbf{C}^{(2)} \mathrm{adj}\left(\mathbf{S}^{21}\right) \mathbf{C}^{(1)\dagger} \tag{2.6}$$

. The adjunct of the overlap correspond to Löwdin's normal rules for nonorthogonal determinants. For transition dipole elements, this is:

$$\langle\Psi_2|\hat{\mu}|\Psi_1\rangle = \sum_{jk} \mu_{jk}^{21} \mathrm{adj}\left(\mathbf{S}^{21}\right)_{jk} \tag{2.7}$$

with the determinant of this overlap defined as the inner product of the two states:

$$\left|\mathbf{S}^{21}\right| = \langle\Psi_2|\Psi_1\rangle \tag{2.8}$$

The general definition of the transition dipole

$$\mu^{1\to 2} = \langle\Psi_2|\hat{\mu}|\Psi_1\rangle \tag{2.9}$$

can be expressed with this transition density matrix as:

$$\begin{aligned} \langle \Psi_2 | \hat{\mu} | \Psi_1 \rangle &= \mathrm{tr}\left( \hat{\mu} | \Psi_1 \rangle \langle \Psi_2 | \right) \\ &= \mathrm{tr}\left( \hat{\mu} \mathbf{D}^{21} \right) \end{aligned} \tag{2.10}$$

### 2.1.3 Semi-empirical extensions

We tried to extend the range of DFT methods that could be used for Δ-SCF and eigenvalue difference methods by investigating whether a tight-binding method could predict transition properties. We chose the recently published DFTB method parameterized by the Grimme group for this. This method has been parameterized for geometries, frequencies and non-covalent interactions, and uses an extended version of Hückel theory. The name they present is GFN-xTB, an acronym for "Geometries, Frequencies, Non-Covalent - eXtended Tight Binding". We chose this method for two reasons. The first being that the GFN-xTB method was already implemented in the `QCORE` package. This significantly reduced the amount of effort required for this project. Additionally, there would be other users and developers who would help with implementation of this new method. The second, more scientific reason, is that a similar method has already been published that calculates transition properties. This is the sTDA-xTB method. As development for this method had similar goals as this project, it is useful to look at this in more detail.

#### 2.1.3.1 sTDA-xTB

A similar method to GFN1-xTB has already shown to be accurate at predicting transition properties for large systems, with exceptional speed. This method was published as the sTDA-xTB method[3]. The drop in accuracy for this method is minimal, with the error being around 0.3 - 0.5 eV.

Similar to other xTB methods, the sTDA-xTB method is a tight-binding method that uses empirically fit parameters and a minimal basis set. It was trained on a test set of highly accurate coupled cluster and density functional theory excitation energies, as well as accurate atomic partial charges.

Unlike other xTB methods, the basis set for sTDA-xTB is dependent on the D3 coordination number. This makes this method far more flexible, which would usually be achieved in a fixed basis set by using diffuse or else additional orbitals in the basis set. Additionally, it uses two sets of parameterized basis sets - a smaller valence basis set (VBS) and an extended basis set (XBS).

These two basis sets are used to construct formally similar Fock matrix elements, however in practice these use different global parameters. The core Hamiltonian is similar to other DFTB methods that use a self-consistent charge method, as opposed to an SCF method, to obtain molecular orbital coefficients. It is given by:

$$\langle \psi_\mu | H^{\text{EHT, sTDA-xTB}} | \psi_\mu \rangle = \frac{1}{2}\left(k_\mu^l k_\nu^{l'}\right)\frac{1}{2}\left(h_\mu^l h_\nu^{l'}\right) S_{\mu\nu} - k_T \langle \psi_\mu | \hat{T} | \psi_\nu \rangle \tag{2.11}$$

where, $\mu, \nu, l, l'$ are orbital and shell indices $k_\mu^l$ are shell-wise Hückel parameters, $h$ are effective atomic-orbital energy levels, $S$ is the overlap, $k_T$ is a global constant and $\hat{T}$ is the kinetic energy operator. The charges used in the inter-electronic repulsion function are given by CM5 charges for the XBS Fock matrix. These are calculated using Mulliken charges obtained from diagonalising the Fock matrix with the VBS. The charges for the initial VBS Fock matrix are based on Gasteiger charges, modified by the electronegativities of atoms in the system.

The whole process for determining molecular orbitals can be summarized as:

1. Calculate modified Gasteiger charges for initial guess
2. Diagonalize Fock matrix in the VBS to get the first set of Mulliken charges
3. Compute CM5 charges
4. Diagonalize Fock matrix in the VBS again for final set of Mulliken charges.
5. Recalculate CM5 charges with this final set, and diagonalize the Fock matrix in the XBS. The molecular orbital coefficients from this are then fed to the response theory.

The response theory for this method is based on the previous work in the Grimme group on the simplified Tann-Dancoff Approximation. There are several approximations made between full linear response theory and the sTDA method. First is the Tann-Danncoff approximation, where the B matrix is ignored. The second approximation is to use Mataga-Nishimoto-Ohno-Klopman (MNOK) integrals instead of explicit 2 electron integrals to calculate matrix elements, as well as neglecting the density functional term.

Transition charges are used to calculate these MNOK integrals, where the charges are computed using a Löwdin population analysis. The operator is the MNOK[6][7][5] damped coloub function, with different exponents $y_K$ and $j_J$ for exchange and coulomb integral respectively. The $a_x$ parameter is included to recover the amount of Fock exchange mixing in the original matrix element equation, and is a free parameter.

Third is the truncation of single particle excited space that is used to construct the **A** matrix. This reduces the number of elements that need to be calculated, and so reduces the time taken for diagonalization, whilst also capturing a broad enough spectrum of excitation energies. The sTDA-xTB has many of the same goals as this project, except in one respect, which is the gradient theory. As the sTDA-xTB method still requires constructing and diagonalizing the **A** matrix, albeit with a tight-binding method for molecular orbital coefficients, the gradient of the transition properties would still be difficult to calculate. Hence it wasn't used for this project, but informed us that an xTB like method could be used to get accurate transition properties.

## 2.2 Benchmarking

Having established the hypothesis, that a GFN-xTB based Δ-SCF method (which we name Δ-xTB ) would predict TD-DFT transition properties with decent accuracy, we then tested this on a test set of small molecules, as well as bacterial chlorophylls from the LHII protein. Additionally, we also investigated Δ-SCF results with a DFT method for the ground and excited state solutions, so we could compare the two differences between the Δ-xTB method and our chosen reference method. The first being whether the Δ-SCF method can reproduce response effects, and the second being whether a tight-binding, semi-empirical method could produce decent enough electronic structure.

### 2.2.1 Reference Data and test set

The test set comprised of 109 small molecules. Each system was closed-shell, contained 12 atoms or less, and contained on H, C, N, and F atoms. The size and relative simplicity of this test set was chosen to minimise any other factors that could cause complication in analysing the results. For example choosing a consistently correct basis set and functional for the Δ-SCF results would have been challening if we had to include transition metal complexes, as well as if we had to think about cpu walltime. Additionally, the size of these molecules made symmetries much easier to inspect by hand.

For reference data we calculated the three lowest energy singlet excited states. We used EOM-CCSD with an aug-cc-pVTZ basis set, using Gaussian 16.

### 2.2.2 Small Systems

Transition properties for this test set were calculated using TD-DFT and Δ-SCF , both using a CAM-B3LYP functional and aug-cc-pVTZ basis set. The transitions were assigned to the EOM-CCSD results by comparing transition dipoles, energies and the character of the MOs involved in the transitions. Where the symmetries could be assigned, these were also used, however this was not the case for all systems as many were unsuccessful in labelling symmetry or defaulted to a non-Abelain group. Symmetry labelling was also only available for TD-DFT calculations, as these were performed with Gaussian 16. Δ-SCF calculations were done with the `QCORE` program.

Additionally, the Δ-SCF singlet transition is not a correct representation of a true singlet excitation, as this is a superposition of both spin-same and spin-flipping excitations. The spin-purification formula:

$$\Delta E_S = 2\Delta E^{i,\alpha\rightarrow a,\alpha} - \Delta E^{i,\alpha\rightarrow a,\beta} \tag{2.12}$$

was used to correct for the true singlet excitation energy.

Overall, the excitation energies calculated with Δ-SCF are as accurate at predicting EOM-CCSD energies as TD-DFT. The mean error is 0.35 eV, iwth a standard deviation of 0.25 eV. This is a marginal improvement on the TD-DFT results, which has a mean and standard deviation of 0.41 eV and 0.27 eV respectively. Transition dipoles were similarly accurate to the reference data, although Δ-SCF performs slightly worse in this respect. The mean and standard deviation in the absolute value of transition dipole moment, $|\mu|$, was 0.07 a.u. and 0.08 a.u. respectively. For TDDFT, the mean and standard deviation were 0.03 a.u. and 0.06 a.u. (the atomic unit here being equal to $ea_0$).

In summary, we can see the Δ-SCF can accurately predict transition properties to a EOM-CCSD level of accuracy with as much success as TD-DFT. We could then reasonably expect a tight-binding method, with good electronic structure treatment, to also be appreciably accurate whilst drastically reducing the cost of calculation.

### 2.2.3 Non-orthogonality

There is, however, a caveat with Δ-SCF . The ground and excited states, solutions to two separate SCF cycles, will probably not be orthogonal. In more detail, the Slater determinants $|\Psi_n\rangle$, are constructed from the set of orbitals $\{|\phi_j^{(n)}\rangle\}$. For each state, these spin orbitals will be orthogonal to other orbitals in the same state. However they will not be orthogonal to orbitals of a different state, such that the inner product:

$$S_{jk}^{21} = \langle \phi_j^{(2)} | \phi_k^{(1)} \rangle \tag{2.13}$$

will be non-zero. Similarly, there will be a non-zero transition charge:

$$q^{21} = \langle \Phi^2 | \Phi^2 \rangle \tag{2.14}$$

which breaks the origin-independence property of the transition dipole moment. In this way, any transition dipoles that do not have their centre at the origin will have a systematic error based on this overlap and the distance from the origin. For vertical transitions, this transition charge should be zero, and so all transition dipole moments calculated with non-orthogonal Δ-SCF would always have this error.

Instead, we applied a transformation to symmetrically orthogonalise the two states involved in the transition, which preserve as much character of the original states as possible. The transformation is given by:

$$|\Psi_{\tilde{\nu}}\rangle = \sum_{\nu} |\Psi_{\nu}\rangle \left[\mathbf{S}^{-\frac{1}{2}}\right]_{\nu\tilde{\nu}} \tag{2.15}$$

It was found that using this method for correcting the non-zero overlap of states, the origin-independence of the transition dipole moment was recovered. It should be noted that whilst this effect is dependent on how large the overlap may be, and it could be argued that with a small overlap this effect may not be large, having any large translation of the molecule (on the order of hundreds of angstroms) can lead to nonphysical transition dipole magnitudes. In a large protein system, where systems can be on this scale, this would obviously present a much larger problem than might be for a vacuum phase small molecule.

### 2.2.4 LHII Chlorophyll

We also tested the accuracy of Δ-SCF on a set of bacterial chlorophyll A (BChla) molecules. These are a much large and more complex system, and so present a tougher challenge for the Δ-SCF method. Additionally, it is a much relevant test for the goal of simulating a whole LHII complex.

As each Bchla is 140 atoms, we chose not to run EOM-CCSD as the reference method but use TDDFT at a PBE0/Def2-SVP level of theory. We ran Δ-SCF with the same density functional and basis set.

We also only looked the $Q_y$ transition with both methods. More detail about this transition and it's importance will be discussed in the next chapter.

We found that Δ-SCF gave transition energies to within the accuracy of TD-DFT, such that we can claim that the fluctuations in transition energy are due to physical intra-molecular reasons rather than random error from the Δ-SCF method. Hence Δ-SCF could be used to accurately predict geometry dependent properties.

However, the error in transition dipole magnitudes was larger than that of the small test set. The error was about 0.42 a.u. larger, but without EOM-CCSD or another high-level method, it's unclear whether this error might be from TD-DFT or Δ-SCF . Additionally, there is a clear correlation between the transition dipole magnitudes from TD-DFT and Δ-SCF , and so whilst quantitative statements cannot be made, we can confidently make qualitative assessments from Δ-SCF .

### 2.2.5 GFN methods

We ran the same benchmarking set for the Δ-xTB method. Here we found it was necessary to extend the number methods that we were using to compare results. Our range of methods were chosen to cover the different approximations for transition properties we were making. Again the reference data was CC2 data, produced by the Grimme group. Our approximations were using Δ-SCF rather than linear response, and semi-empirical rather than DFT. Hence we investigated:

- High level TD-DFT, with a range separated functional and large basis set
- Lower level TD-DFT, with a functional and smaller basis set

- Δ-SCF with range separated functional and large basis set
- Δ-SCF with functional and smaller basis set
- linear response with GFN-xTB
- Δ-SCF with GFN-xTB, Δ-xTB

Concurrent to this work, we also implemented the GFN0-xTB method in QCORE. This method is similar to the GFN-xTB method, but excludes any charge dependent terms in its Fock matrix so is not self-consistent. We also tested whether this would be a possibility for predicting transition properties.

The results are shown in fig-2.2.5.

These results are discussed in more detail in 2.2.5.2, however it's quickly seen that Δ-xTB as well as linear response GFN-xTB is inaccurate at predicting excitation energies. One factor stands out for this error before discussing reasons based on the inherent accuracy of the method. A recurring problem of Δ-SCF methods is that the excited state SCF cycle could converge to the wrong state or collapse to the ground state. This could be seen by the symmetry of the excitation, as the TD-DFT, Δ-SCF and cc2 methods could all be converged to different states. The symmetry of these state could also be different, and so by checking this symmetry we can tell if a comparison between the results given by these different methods is justified. Assigning symmetry to Δ-SCF results however is not a straightforward task.

#### 2.2.5.1 Post-SCF Assignment of Symmetry

Considering symmetry is a common thread in many parts of electronic structure theory. It appears in normal mode analysis, wavefunction analysis and assignment of electronic transitions. For this project, we looked at assigning symmetry to the transitions for Δ-SCF methods, which would require assigning symmetry to the orbitals and overall wavefunction of a molecule. The easiest way to assign the symmetry of transition would be to look at the ground and excited-state MOs by plotting them and inspecting the symmetry. Whilst this method works, it is very time-consuming and could not be automated. Every new method that would be added to the benchmarking would have to have every molecule individually inspected to make sure the symmetry of each transition was correct. Instead, we investigated whether we could assign the symmetry of the resulting orbitals as they are given by the SCF procedure. This was a new feature that had to be implemented in QCORE. Broadly speaking, most electronic structure codes have two choices in assigning symmetry to orbitals - either all of the SCF code will treat symmetry from the outset, or nothing is assigned in the SCF code and assignment will happen post-SCF. Both these approaches have benefits and drawbacks. The first method allows the symmetry to be given at any point in the SCF procedure, and allows the Hamiltonian to be organized into a block diagonal matrix. This can be useful when solving for a large basis set or large system as the matrix diagonalisation

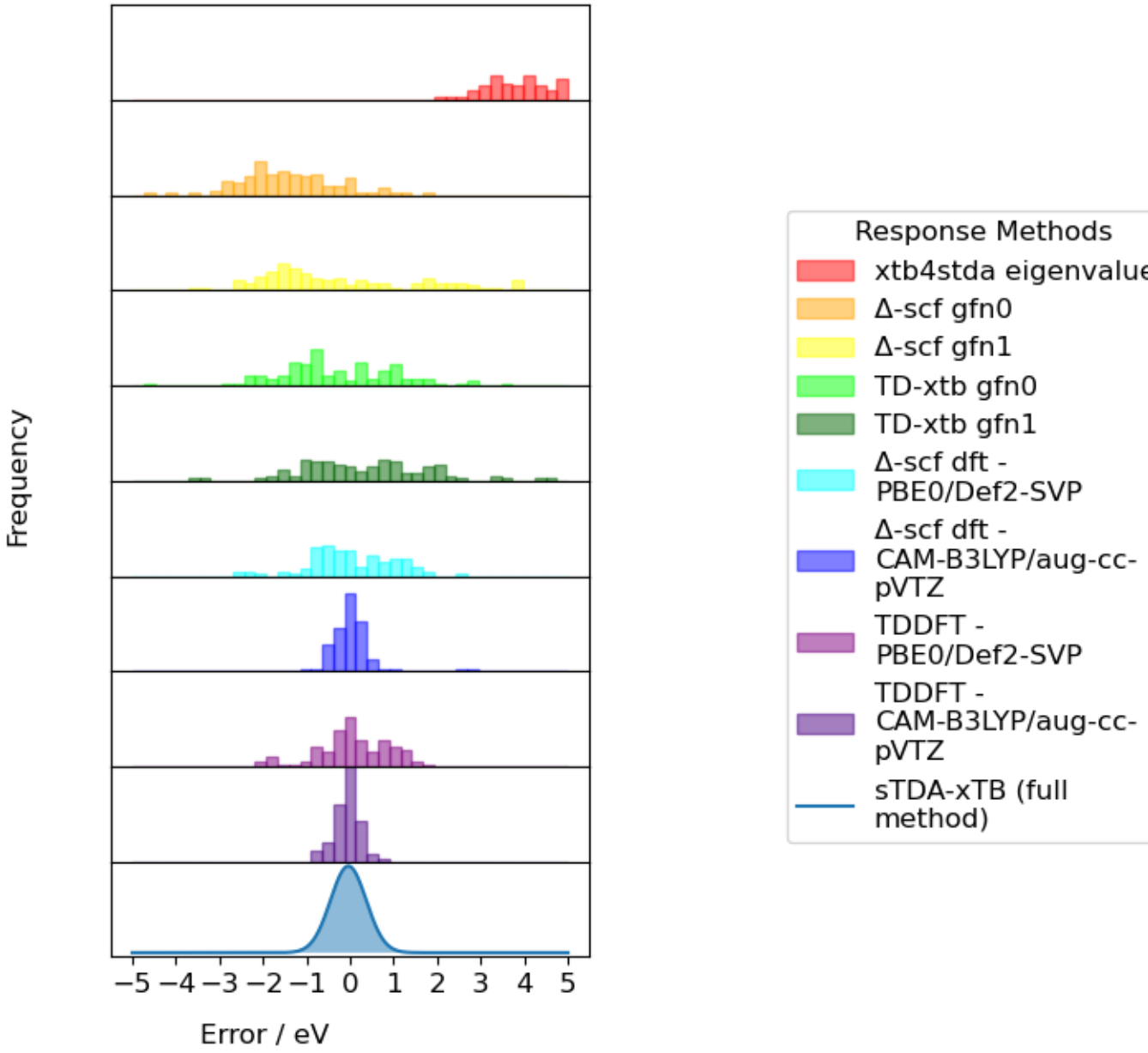



Figure 2.1: The errors of several levels of theory at predicting cc2 transition energies.

can be partitioned and parallelised over several cores or nodes on a cluster computer. However, this works best if the system is highly symmetric, which is often not the case when treating unoptimized systems, such as those from a molecular dynamics simulation, and is definitely not the case when looking at biological systems. The second approach, assigning symmetry after the SCF cycles, doesn't fix these drawbacks, but it does allow for codes which originally didn't have symmetry assignment to be extended without rewriting SCF code. The obvious drawback of doing assignment post-SCF is that symmetry can't be utilized during the SCF procedure. We opted for the second approach, as this was the easiest to implement in `QCORE`. We used the open source library libmsym for point group assignment routines and finding the symmetry adapted linear combination of atomic orbitals. Broadly, the steps for assigning orbital symmetry is as follows:

1. Determine the point group of the molecule, from the atomic positions
2. Setup the atomic orbitals in the libmsym representation.
3. Get the symmetry adapted linear combination (SALC) of atomic orbitals for each subspace. These subspaces are the groups of symmetries that can be found in the point group of the molecule.
4. The SALC can then be used to make the transformation matrix $T$.
5. Assign the one electron molecular orbital (MO) for these subspace characters with the symmetry adapted linear combinations.
6. Multiply the one electron MO symmetries together to find the symmetry of the overall wavefunction.

We implemented this procedure and tested this on methane with a minimal basis set. We found that we could accurately assign the MOs and overall wavefunction of this system with this method for the ground state, which was encouraging.

To assign a label to each molecular orbital, we looked at the character of each orbital in all subspaces. This required transforming the molecular orbital coefficients into the each subspace by using the transformation matrix T:

$$\tilde{\mathbf{C}}_A = \mathbf{T}_A^T \mathbf{S}\mathbf{C} \tag{2.16}$$

and then summing the coefficients to obtain the character:

$$P_A = \sum_{\nu} \left| \tilde{\mathbf{C}}_{A,\mu\nu} \right| \tag{2.17}$$

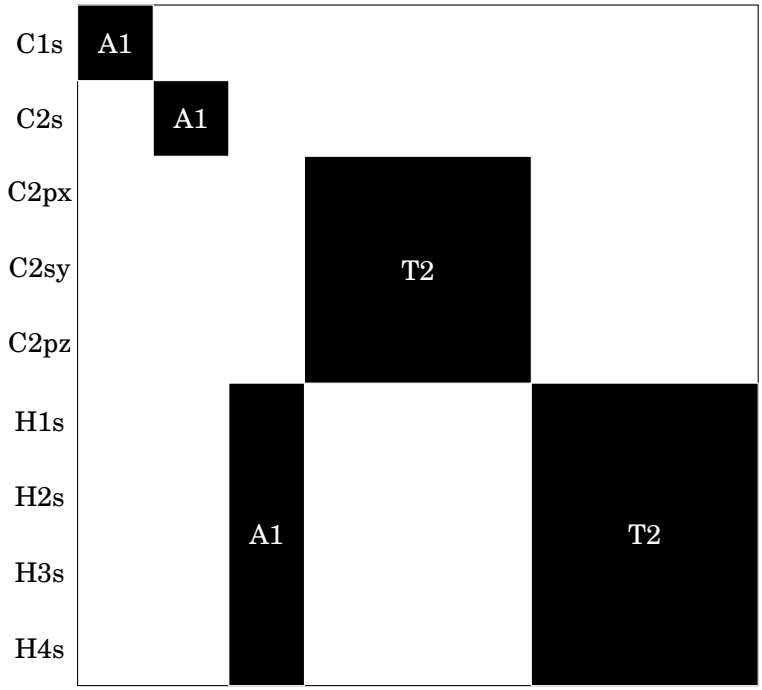



Figure 2.2: Symmetry assigned MOs

. For well defined systems, if the molecular orbital is in the subspace then the character would equal 1. However in practice this was no so clear cut and so the higher subspace character was taken.

We were able to correctly assign the MOs for an optimised methane geometry with an STO-3G basis set - two occupied orbitals and one unoccupied orbital of A1 symmetry, and three occupied and unoccupied orbitals of T2 symmetry. The overall wavefunction symmetry can then be expressed as the product of all MO symmetries, which can be reduced with the reduction formula:

$$n(i) = \frac{1}{h} \sum_{R} \xi_r(R) \xi_i(R) \tag{2.18}$$

. This correctly produced the overall symmetries of ground state systems.

However, we encountered two problems. First was that the assignment of MOs did not word well for excited states, due to the character from the subspace projection being unclear for many MOs. The second was that for non-abelian groups, where there are some degenerate E and T subspaces, this assignment did not work. The reduction of ground state wavefunctions gave non-physical answers.

We discussed some ideas about how to overcome these issues, such as using symmetry decomposition for the non-abelian point groups. We also discussed implementing an analysis based on the antisymmetric component of the direct products. However all of these ideas would

create a more open ended project. Additionally, this may have been a useful feature for testing the benchmarking sets, but chlorophyll molecules would not be symmetric and so type of assignment would not have worked. Hence, we could not confidently assign transitions from Δ-SCF methods with this method. We could still use the non-autonomous method of inspecting transition dipoles and plots of MOs, but as mentioned above this would drastically increase the turnaround time of including more systems in benchmarking, or other investigations. Critically, it would also be impossible to fit a set of parameters to 'learn' transition properties of systems, as will be discussed in the next chapter.

#### 2.2.5.2 Implications of Benchmarking Results

After considering this leading error, we can then make some statements on the data presented.

Overall, the Δ-xTB method was inaccurate - far too inaccurate to be used as a viable method for transition properties of chlorophyll, or any other system. The range of errors is quite large and there is also a significant systematic shift present. However the DFT methods, both the linear response and the Δ-SCF methods, are accurate at predicting excitation energies

We can try to pinpoint where the source of error might be from looking at methods including in this benchmarking. First is that whilst the Δ-xTB method is inaccurate, the sTDA-xTB results shows that a tight-binding approach that doesn't use full linear-response can give very accurate results. Hence we could expect other similar approaches to perform similarly well.

The next comparisons that could be made is whether the difference between linear-response and Δ-SCF methods, or the difference between ab initio or semi-empirical methods, are the leading cause of error. Looking at the results for CAM-B3LYP/aug-cc-pVTZ, we can see that both the linear-response and Δ-SCF results are fairly accurate. There are some outliers in the Δ-SCF results, that we attribute to different transitions to excited states, however overall both have a majority of errors within a 0.5 eV range. Looking at the slightly lower level theory DFT results, PBE0/Def2-SVP, we can see a marked decrease in accuracy. Again, both linear-response and Δ-SCF methods perform similarly, and so we attribute the leading cause of error to be the electronic structure method, and not the response method.

We can continue this trend into the semi-empirical methods. Overall, these methods performed much worse than the ab initio DFT methods, with the exception of the sTDA-xTB method. There is still a relatively small difference between linear-response xtb and Δ-xTB results, however both are fairly inaccurate and so this point is a bit harder to justify.

As the implementation of linear-response and Δ-xTB methods could use both the SCC and non-iterative versions of the gfn-xtb methods (gfn1 and gfn0), we included both in our test sets. We found that both are inaccurate, and there is a marked drop when using gfn0. The systematic shift in the gfn0-Δ-xTB method is especially bad. We conclude from this a response method based on out-of-the-box gfn-xtb would not be viable - this is a conclusion that will be returned to in the following chapter.

Overall, arguably the absolute worst was the eigenvalue difference method based orbital energies (eigenvalues of the Hamiltonian diagonalisation) from the sTDA-xTB method. To obtain these results, we used the available sTDA-xTB programs, that are in two parts. The first part runs a version of xTB that provide molecular orbital coefficients and energies for the sTDA program to then use to calculate the transition properties. on `xtb4stda` results. For these results we took the difference between linear-response routine in the full sTDA-xTB method. This was the simplest method that we included, and was a test to see whether the optimised electronic structure parameters were inherently good at predicting transition properties. As can be seen, this isn't the case, and so we could expect no real difference in a new method if we started from this xtb method or a version of gfn-xtb. It seems that the response method used can make up a large part of predicting transition properties accurately.

The result that gfn-xtb based methods are not accurate is not unexpected. Tight-binding methods in general are inaccurate methods are electronic interactions, due to not including a large part of the electronic density. Additionally, they are highly parameterised to the certain systems and problems. Whilst the gfn-xtb methods are better than many other methods in this parameterisation, using a few pair-wise parameters as possible to improve extensibility to systems outside the training set, there will always be the problem that these parameters are trained on ground state properties and not excited state properties. Whilst there was the possibility these parameters might have been accurate at both ground and excited state properties, the results show that this just isn't the case.

## 2.3 Conclusions and Further Work

The two main hypotheses tested in this chapter have been whether a $\Delta$-SCF approach to response properties can be as accurate as full linear-response or better, and if a tight-binding based method could also be as accurate.

The answer to both of these are yes, but not at the same time, at least in the forms given in this chapter.

This is an encouraging step towards the overall thesis of this work. All of the $\Delta$-SCF methods tested in this work have required much less cpu work, both in terms of time and in memory, than the full linear-response methods. The difference is greater for the higher levels of theory, as this can amount to hours of cpu time saved. For the tight-binding methods, this is not as stark until we start considering much larger systems and problems.

The remaining issues and further work are outlined below.

### 2.3.1 Symmetry

The main problems as have been discussed have mainly been in assigning the correct transition between different methods. A $\Delta$-SCF has not been as actively investigated as TD-DFT methods,

there are fewer developed tools for this kind of problem, much less other programs or codes that could be used for reference. For example, many TD-DFT codes can provide symmetry in their results, whereas it took alot of work even to attempt to do so with Δ-SCF . This would have been an unfinished problem going forward. However, due to the well defined $Q_y$ transition (discussed in the following chapter), this may not be an issue for chlorophyll. But as further work might include other systems, this would have been a recurring issue to be solved. Additionally, it doesn't give much confidence to the results of the benchmarking.

Further work on this issue would be to return to this investigation. As reported, assigning ground state symmetries was successful, and much of the machinery required was implemented.

### 2.3.2 Embedding

As reported in this chapter, these benchmarking results do not include any investigation into how these methods would behave in an embedded system. This is more of an implementation problem than a theoretical problem for the majority of the methods investigated, with the exception of the xtb based methods. As implemented in `QCORE`, all of the DFT methods as well as the gfn1 based method could have included embedding effects. This would be important to investigate as we would have moved to more biological systems, where embedding effects can be very important to include. Further work on the remaining methods would have involved solving two problems. First would be the formulation of embedding theory for gfn0, as this has not been investigated before. We did perform some work on this area.

Second would be embedding the sTDA-xTB theory, which really would have required a deep investigation into the source code of this method. Given the scope of this project, this was not a realistic consideration, especially for a relatively minor part of the work.

### 2.3.3 Scaling

Whilst the Δ-SCF results are promising in terms of accuracy and computational effort, this method may not be well suited for the issues discussed in the previous chapter. The main issue is scaling up the volume of calculations that could reasonably be done. All of the semi-empirical methods used in this benchmarking all had a cpu walltime in the order of seconds, with very small memory requirements. Additionally, they all could be run on a single core due to the larger overhead of serial calculations as compared to the parallelisable routines - i.e. more time would be spent on running SCF cycles or running functional code, which can only be run in serial, than on constructing the Fock matrices or calculating integral values, which can be done in parallel. The opposite is true for the larger scale DFT calculations. Especially for the large chlorophyll system, the memory requirements were much higher than the semi-empirical methods. For example, using the `QCORE` implementation, all of the chlorophyll calculations required on the order of $10^2$ gb of memory, and would take  90 minutes of cpu walltime. This includes being parallelised over many cores (although not nodes), all of which adds together to make a relatively expensive

calculation. This also means that it is entirely unfeasible to run these calculations on a desktop computer.

This has two drawbacks. First, as said above, is the volume of calculations that can be reasonably be done. This investigation used 27 individual chlorophylls, which would represent about half a days worth of supercomputer time. Expanding this to a larger scale study of LHII, which can require up to thousands to hundreds of thousands of calculations, would take weeks or months of computational time. The first drawback then is that a study based on this method would not be repeatable in the time frame of a single project. To discover a flaw or additional property that is needed, and so rerun all the calculations even twice or three times, would take on the order of a year and so either be unfeasible or drastically reduce the ability to test new ideas, given that nothing should be allowed to fail. The second drawback is the lack of ease of use. As all calculations would have to be run on a supercomputer, no small assays could be done on a local machine or desktop computer. This is not a major problem, as many methods are unable to be run on a desktop computer, especially large scale DFT calculations. The problem here is that the sTDA-xTB results, which were all run on a desktop computer, show that it is possible to run these systems on smaller machines. It would be worth exploring whether there are more, similar methods that could also achieve this, as this would allow more researchers to be able to run larger systems without the previously needed expense of time and energy, as well as allow for a quicker turnaround time on assays. Additionally, it would also just allow more calculations to be run. This topic is revisited in chapter 5.

The reliability, or robustness, of $\Delta$-SCF is another issue in scaling up the volume of calculations. As mentioned at a couple of points in this chapter, when calculating the excited state, $\Delta$-SCF has the occasional tendency to either converge to an unintentional state, collapse back down to the ground state, or not converge at all. It can be very hard to predict when this will happen (for example some Bchla geometries behave like this whilst others do not, which is hard to explain), and requires several tricks to fix. These include Fock matrix damping, changing parts of the SCF convergence algorithm such as the DIIS procedure, or using step-wise initial guesses, such as using an unphysical half-electron excitation as a starting point for a full electron excitation. To work out which trick is needed, and which systems will need them, requires a trial-and-error approach, and it's never guaranteed that something will work. For getting a time series of a transition property from a whole MD simulation, this is obviously a large problem.

CHAPTER 3

# CHLOROPHYLL SPECIFIC METHODS

Preamble this is a preamble.

## 3.1 The Cassida equation

### 3.1.1 Approximations to Solutions

### 3.1.2 MNOK Integrals

## 3.2 Parameterization

### 3.2.1 Reference Data

### 3.2.2 Objective Function

### 3.2.3 Minimization Algorithms

## 3.3 Benchmarking

### 3.3.1 Transition properties

### 3.3.2 Potential Energy Surfaces

### 3.3.3 Absorption Spectra

Excitation Energies correlation

| | PBE0 | dscf | eigdiff | camb3 |
|---|---|---|---|---|
| **PBE0** | 1 | 0.915415 | 0.930766 | 0.9523 |
| **dscf** | 0.915415 | 1 | 0.846607 | 0.8639 |
| **eigdiff** | 0.930766 | 0.846607 | 1 | 0.9250 |
| **camb3lyp** | 0.952334 | 0.863979 | 0.925034 | |
| **BLYP** | 0.730346 | 0.637576 | 0.606743 | 0.6376 |

Figure 3.1: correlations of energies

Transition Dipole Magnitudes correlation

| | PBE0 | dscf | eigdiff | camb |
|---|---|---|---|---|
| PBE0 | 1 | 0.5891 | 0.566434 | 0.76 |
| dscf | 0.5891 | 1 | 0.977046 | 0.80 |
| eigdiff | 0.566434 | 0.977046 | 1 | 0.77 |
| camb3lyp | 0.762861 | 0.800859 | 0.773252 | |
| BLYP | 0.0832043 | -0.0344407 | -0.0098512 | 0.050 |

Figure 3.2: correlations of transition dipole moments

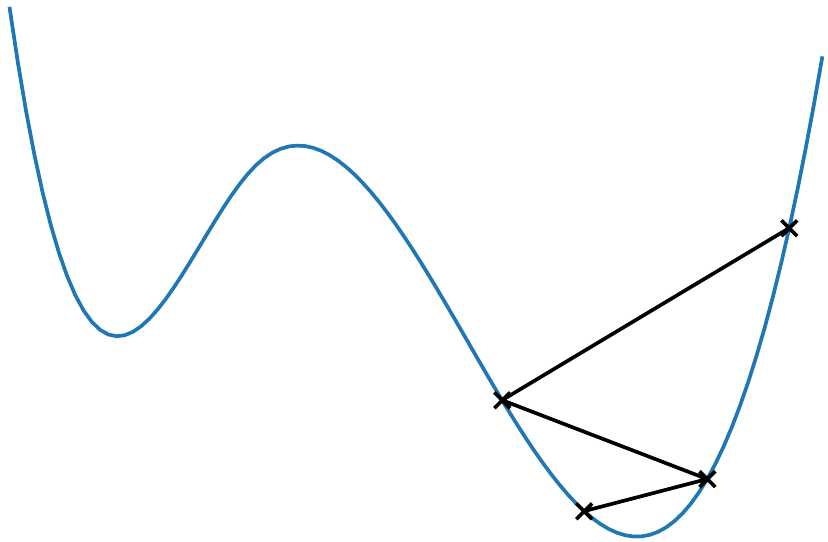


Figure 3.3: Nelder-mead

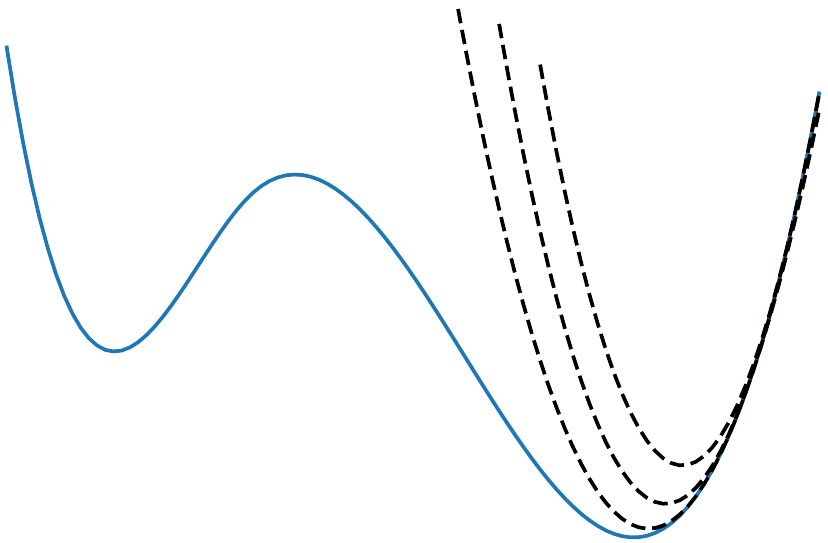
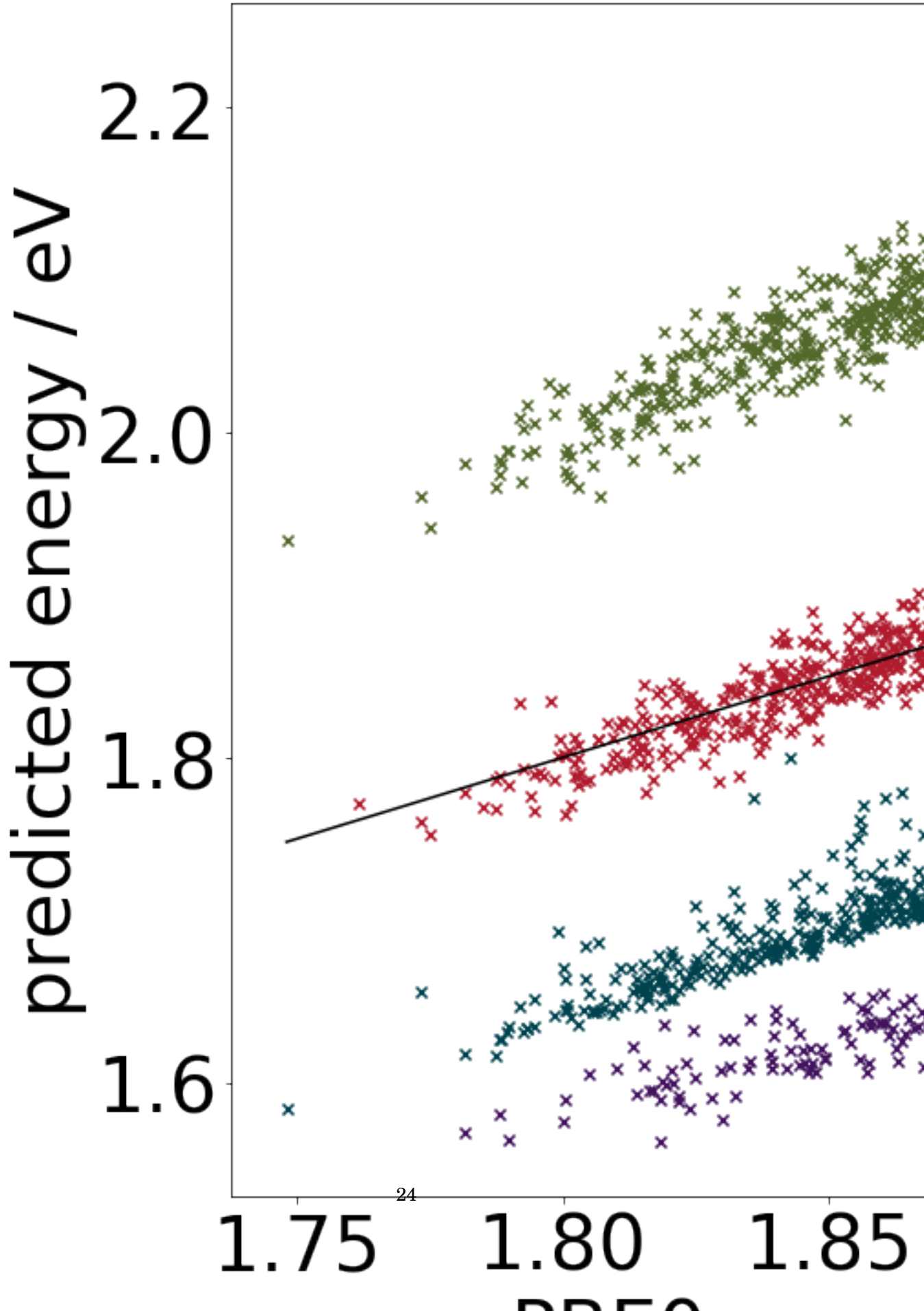
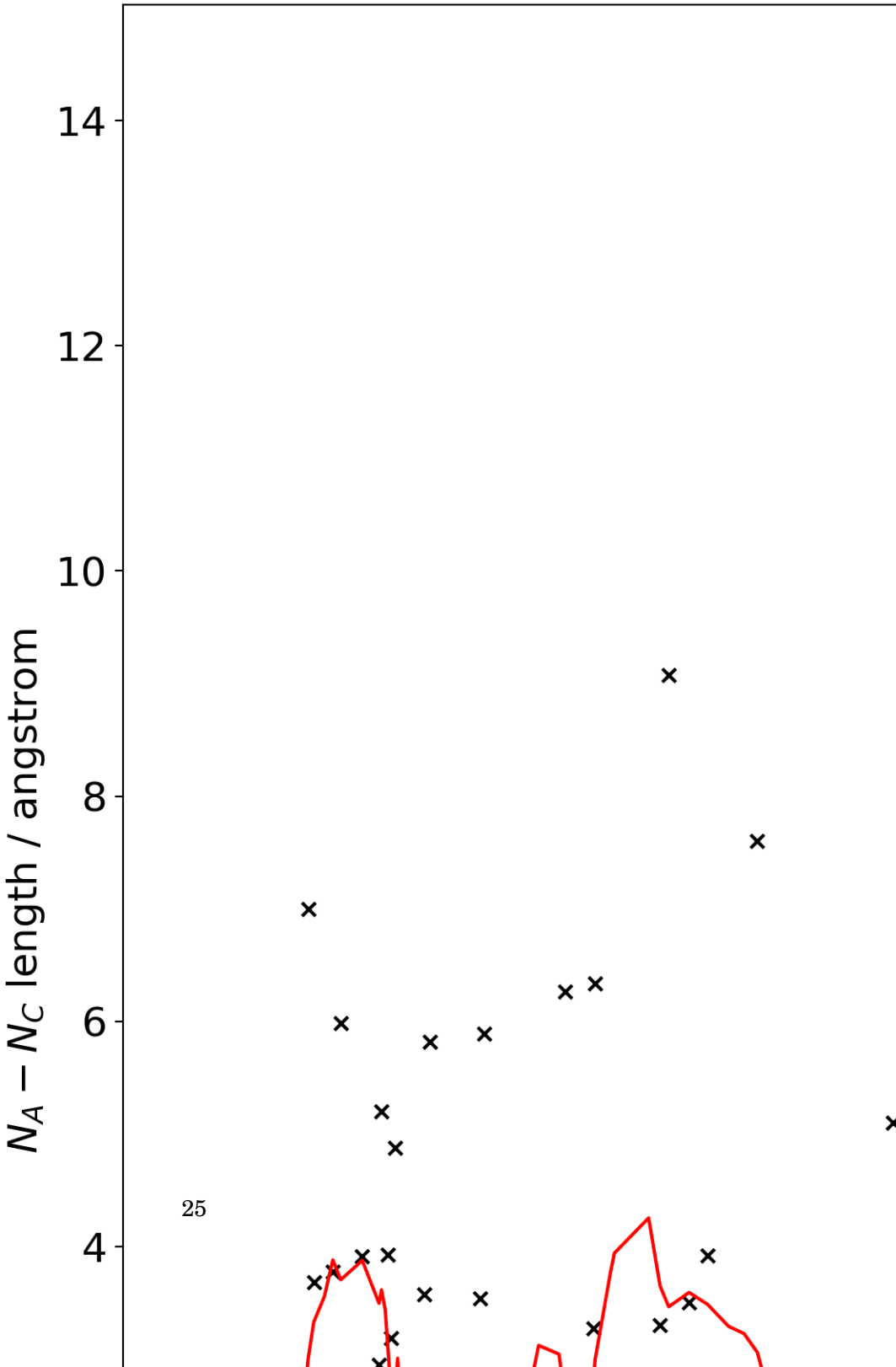


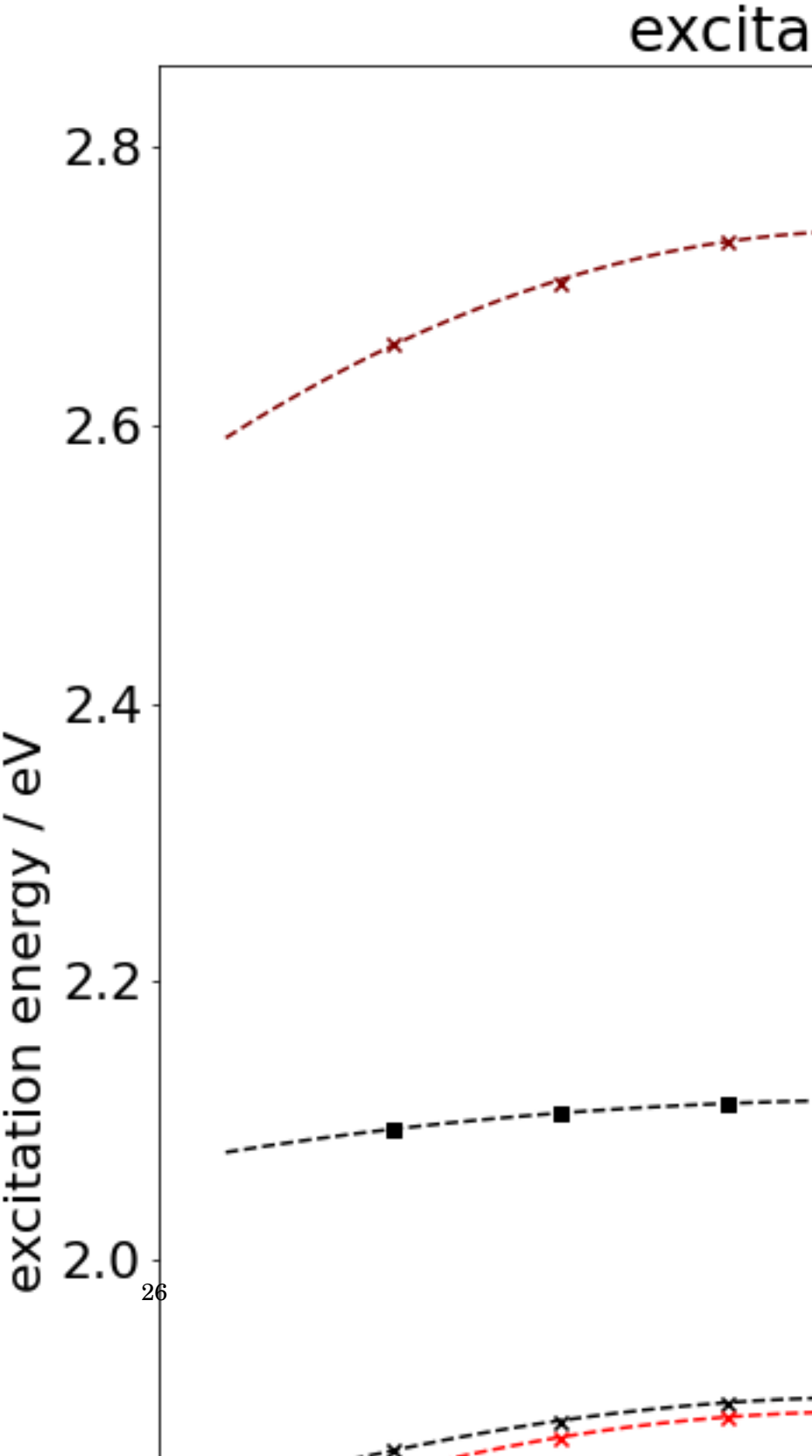
Figure 3.4: SLSQP

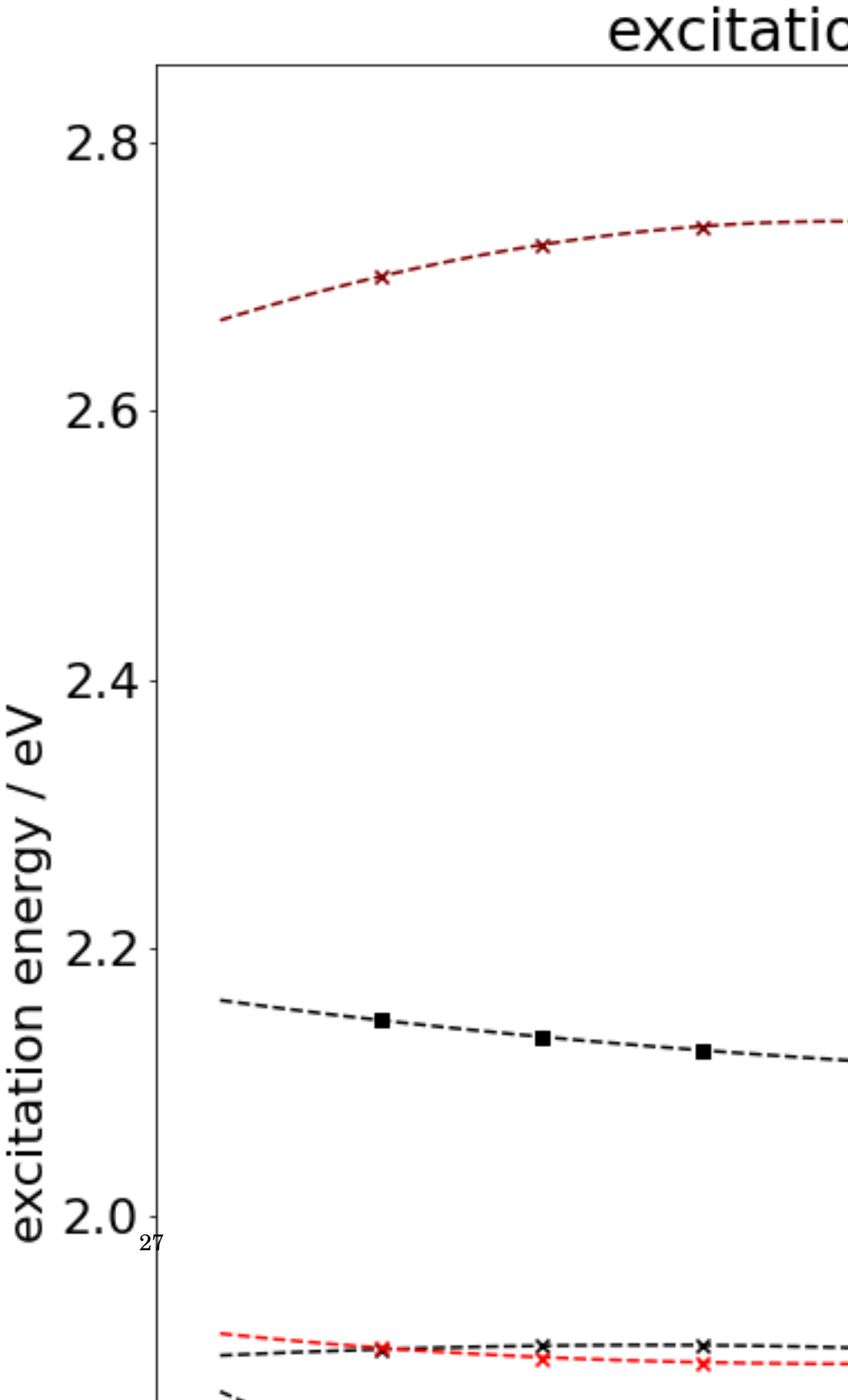
| | |
|---|---|
| $y_K$ | 1.0 |
| $y_J$ | 1.0 |
| $a_x$ | 1.0 |

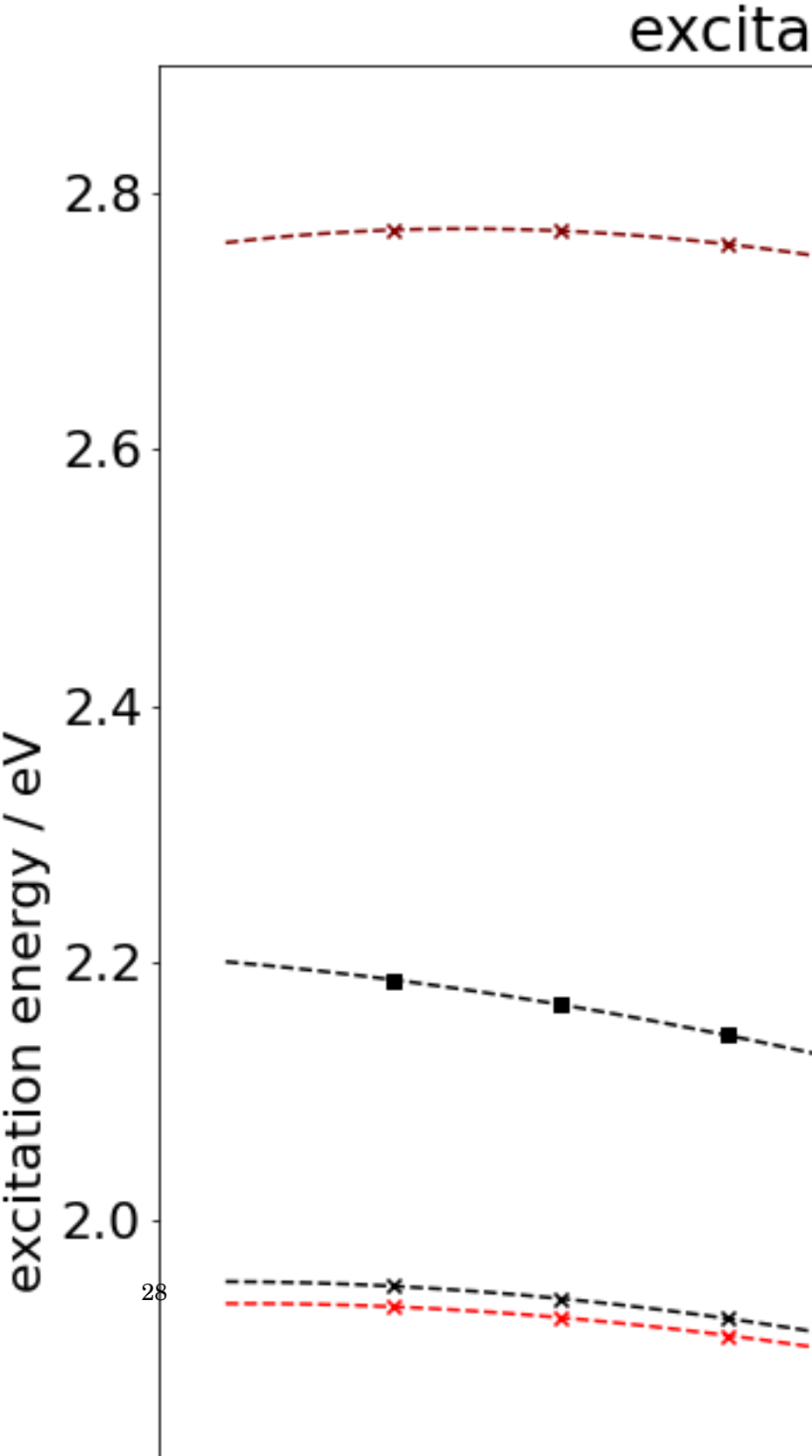
Figure 3.5: optimized parameters from SLSQP procedure.

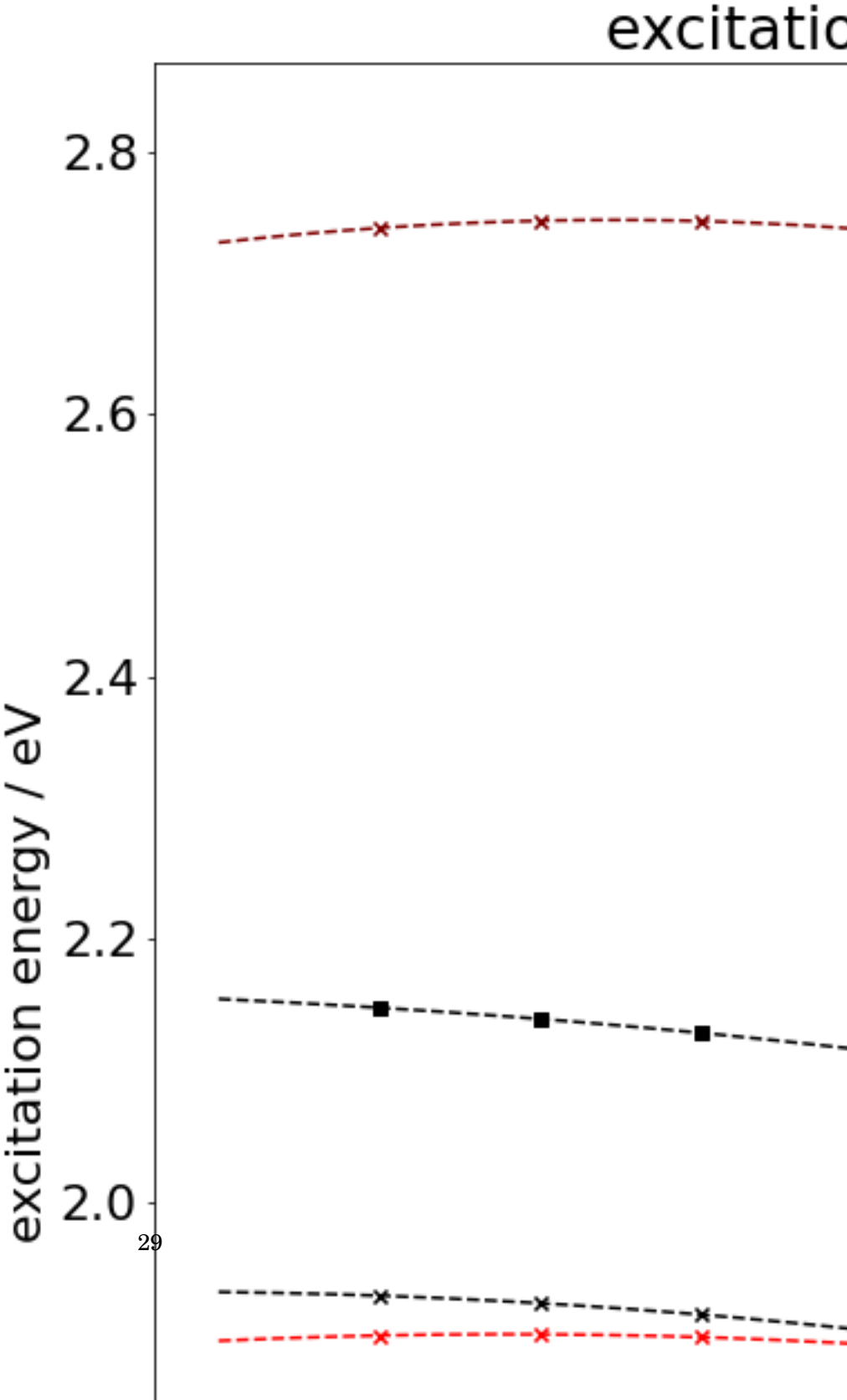

predicted energy / eV
2.2
2.0
1.8
1.6
1.75
1.80
1.85

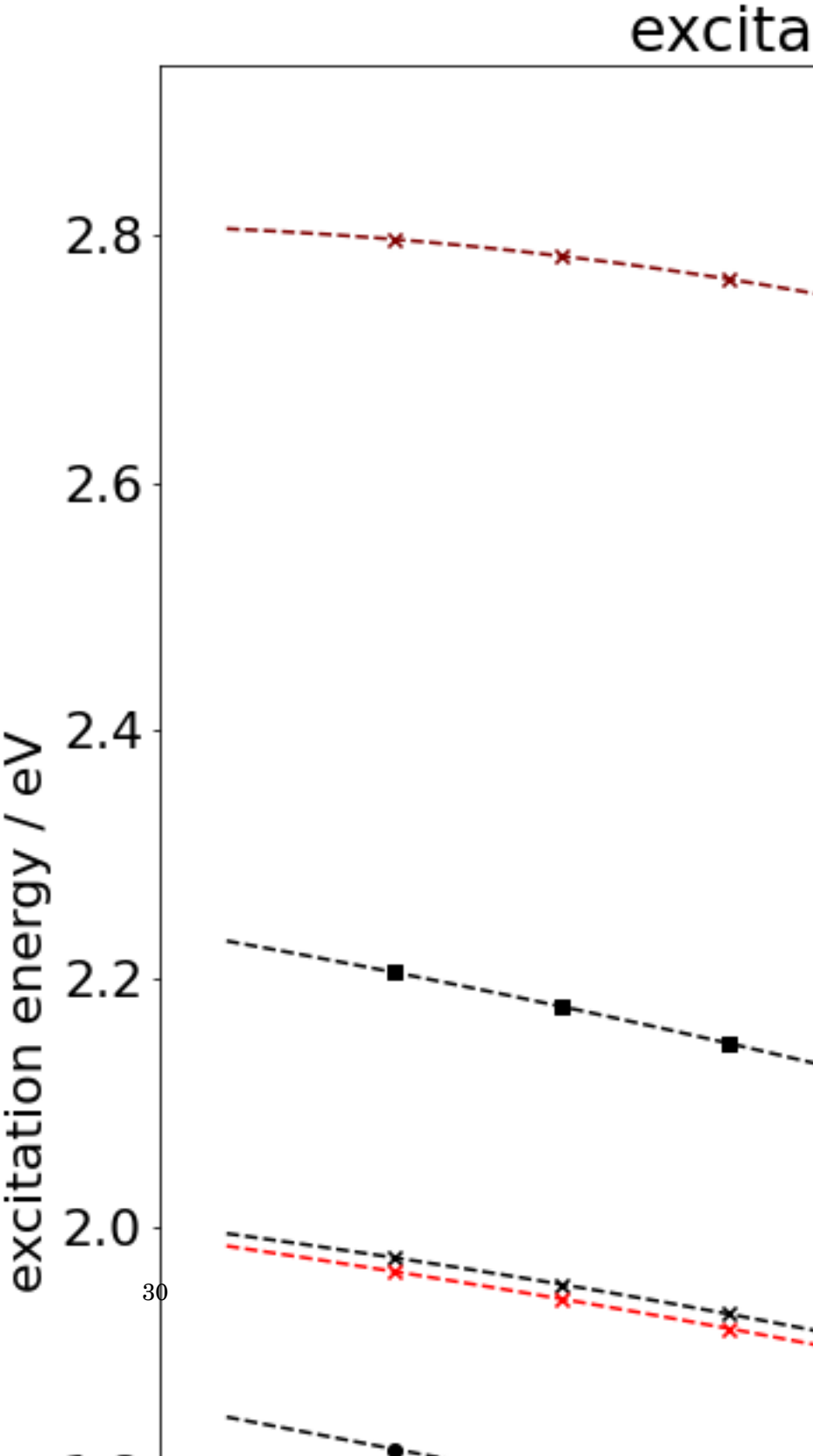

14
12
10
8
6
4
$N_A - N_C$ length / angstrom

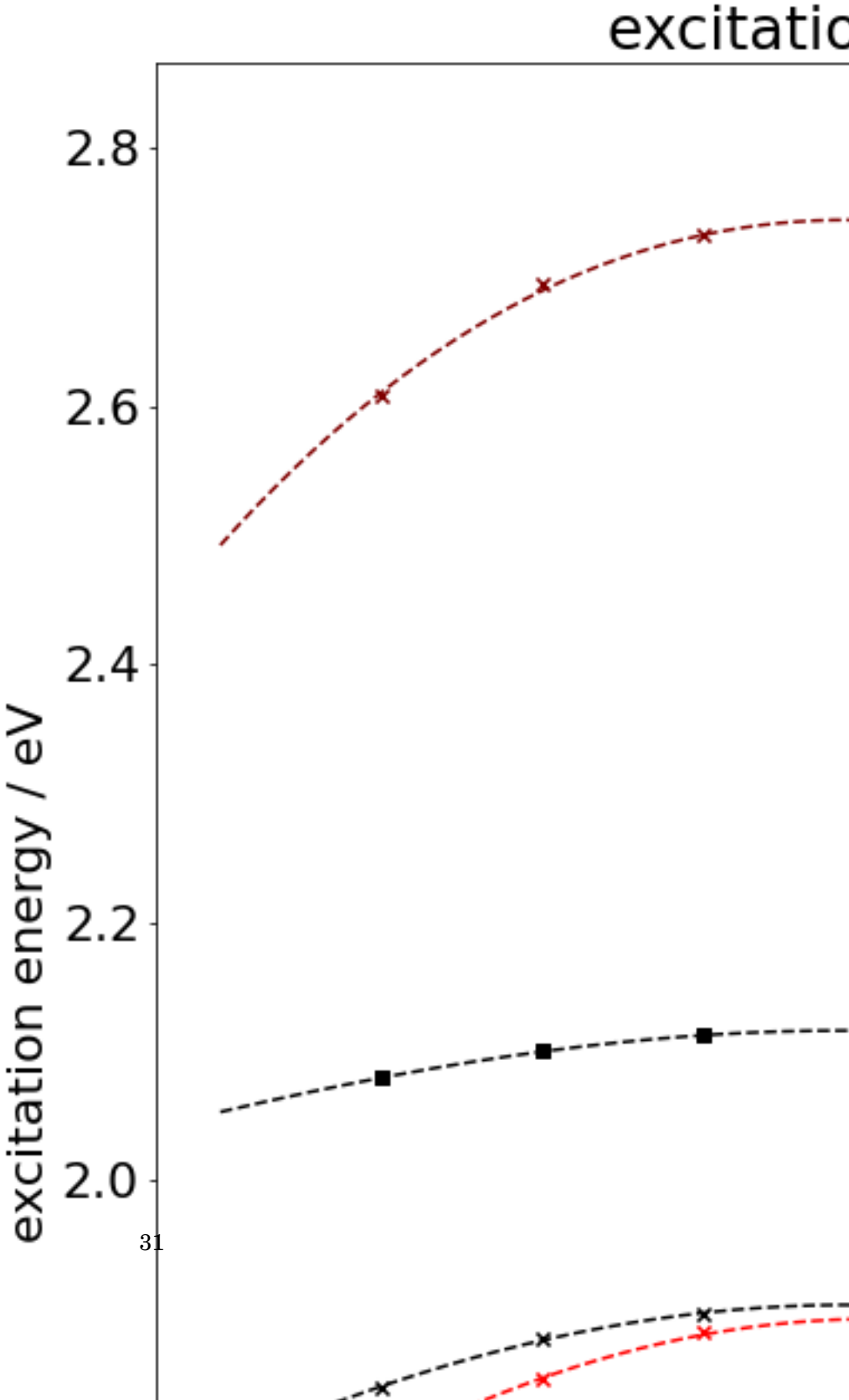

excita
2.8
2.6
2.4
2.2
2.0
excitation energy / eV

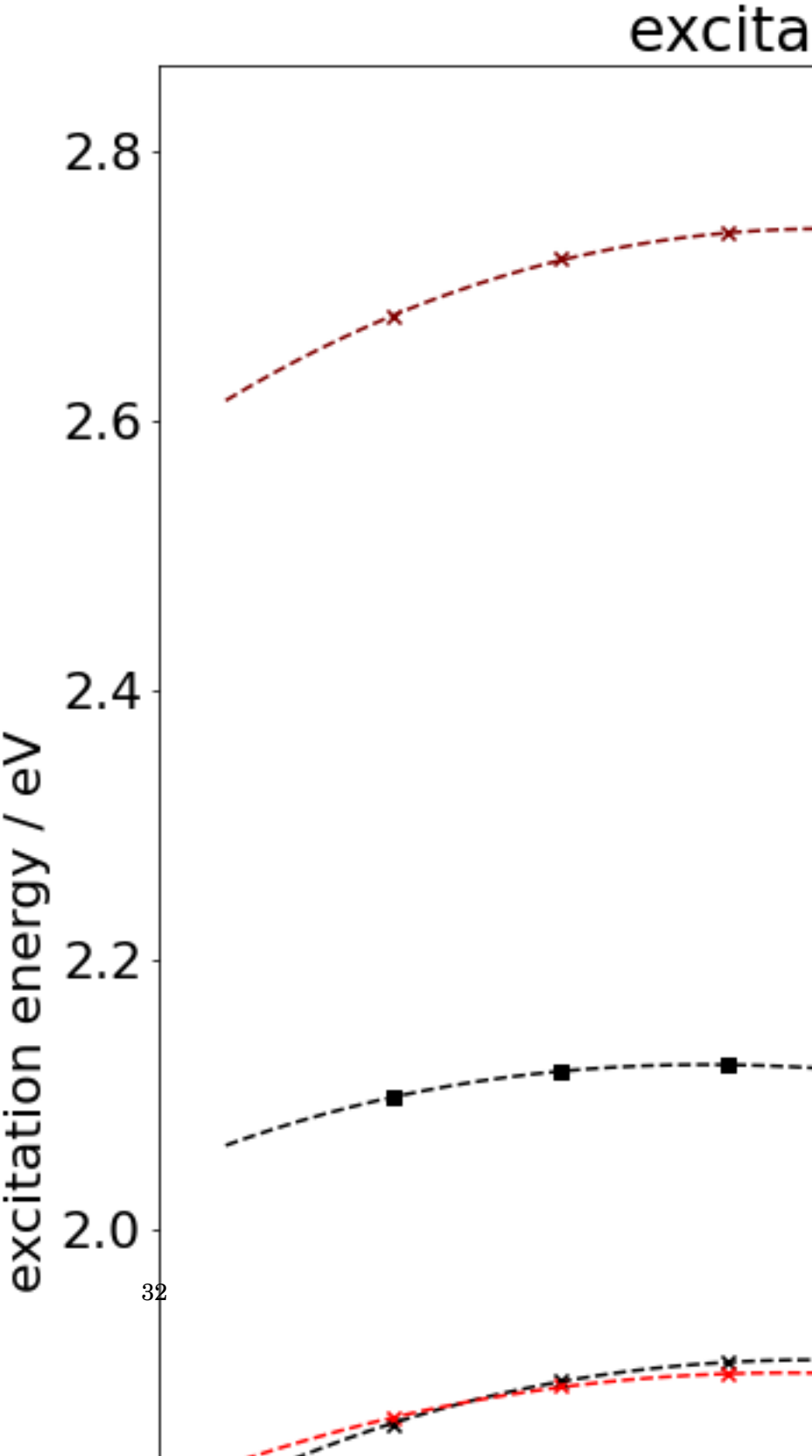

excitatio
2.8
2.6
2.4
2.2
2.0
excitation energy / eV

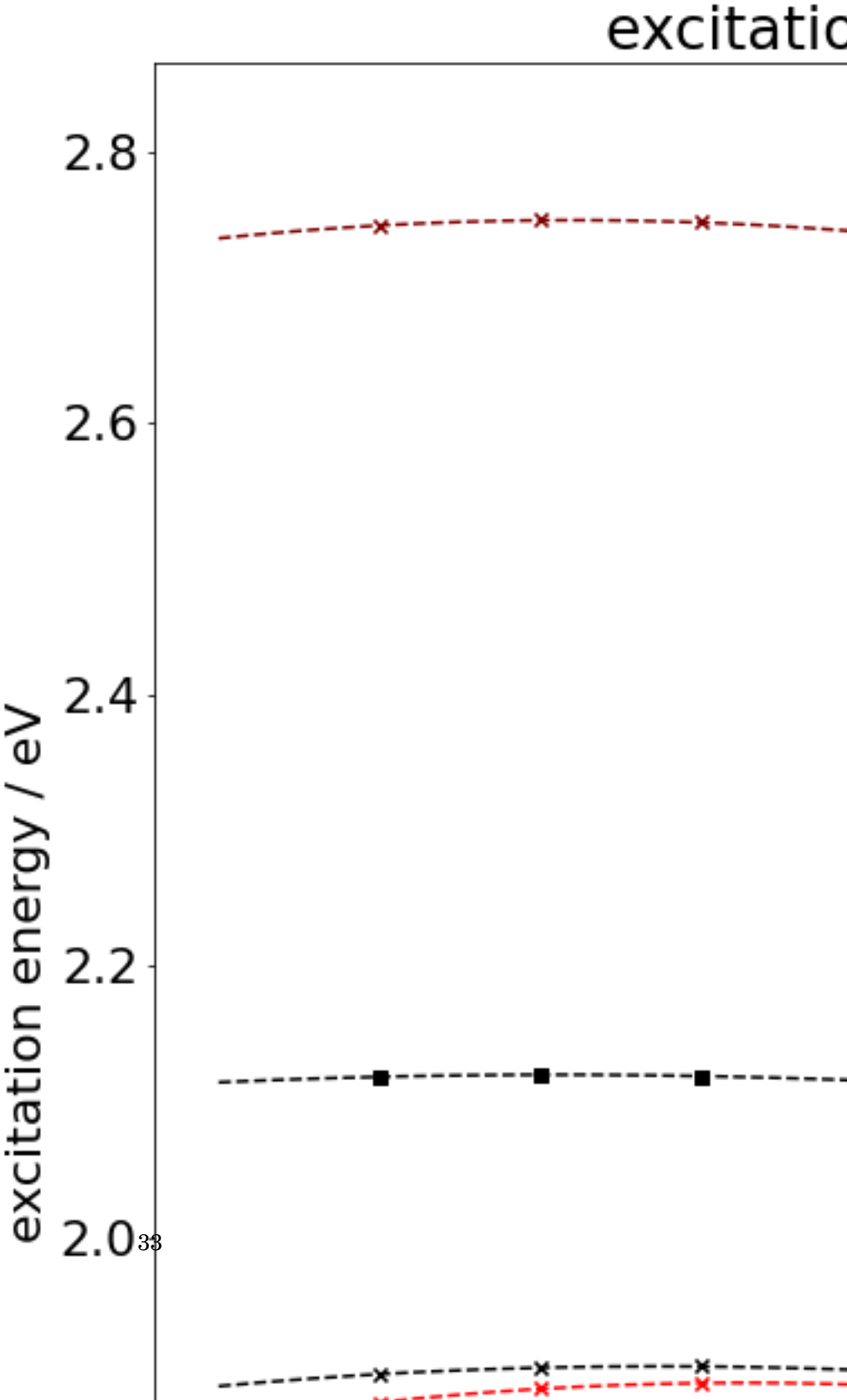

excita
2.8
2.6
2.4
2.2
2.0
excitation energy / eV

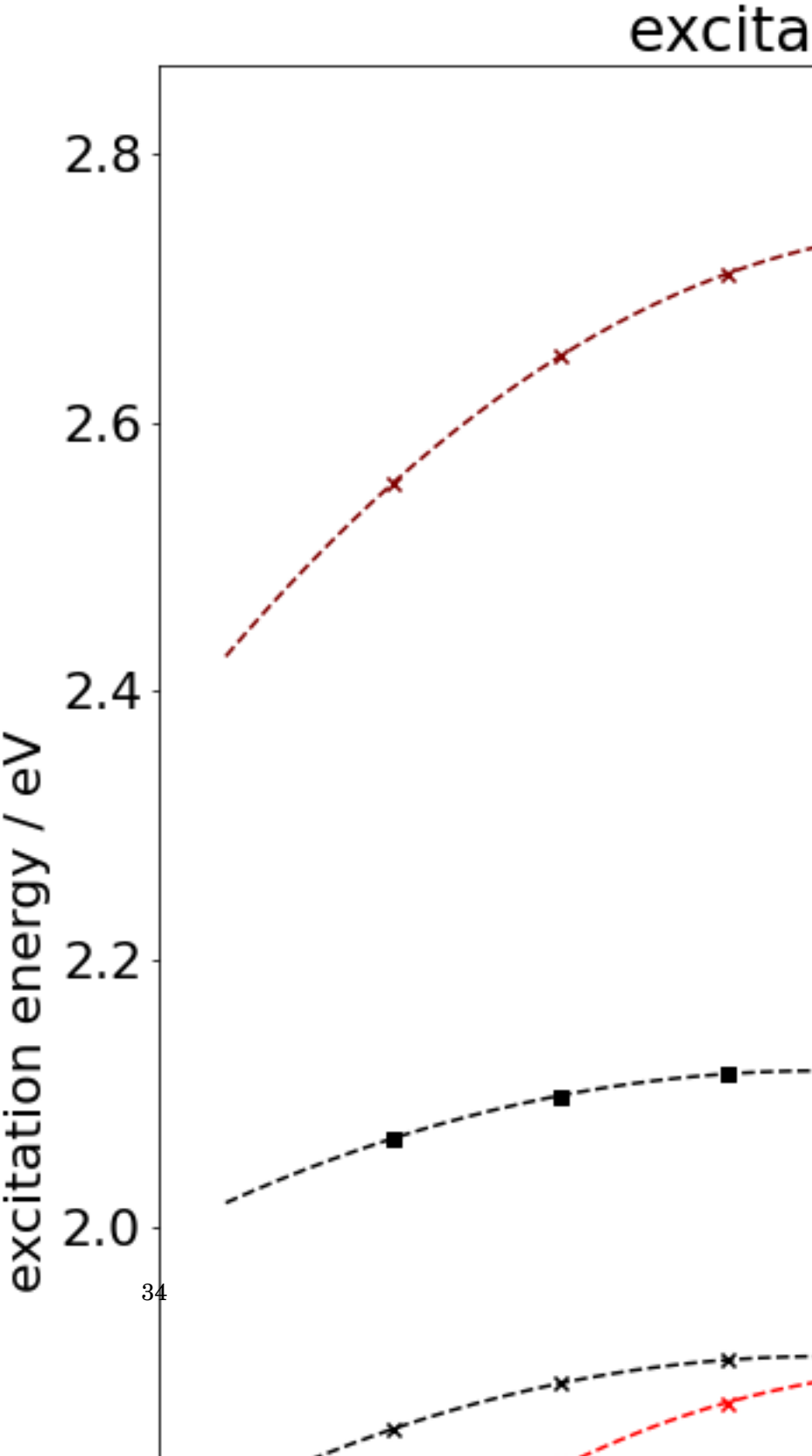

excitatio
2.8
2.6
2.4
2.2
2.0
excitation energy / eV

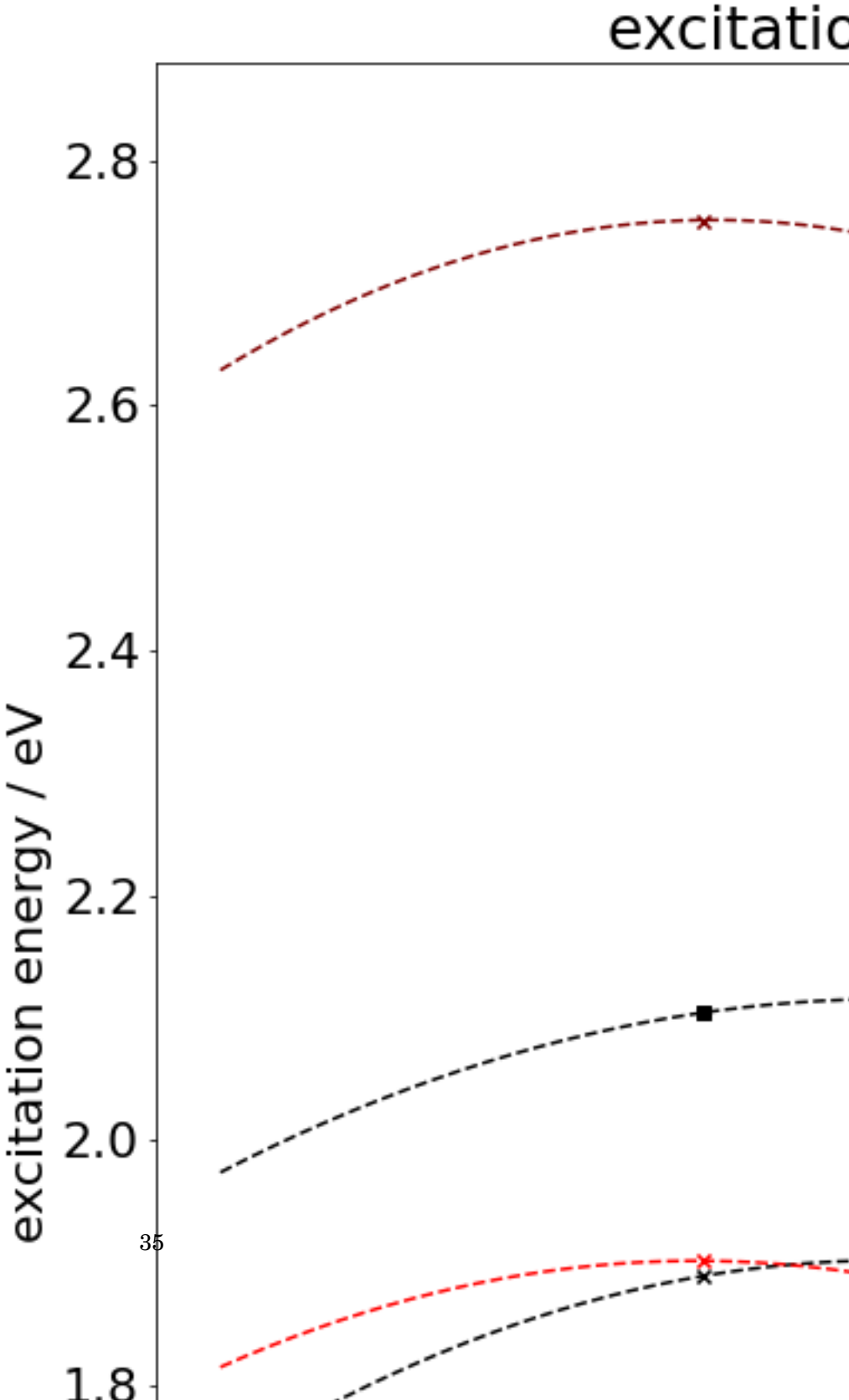

excita
2.8
2.6
2.4
2.2
2.0
excitation energy / eV

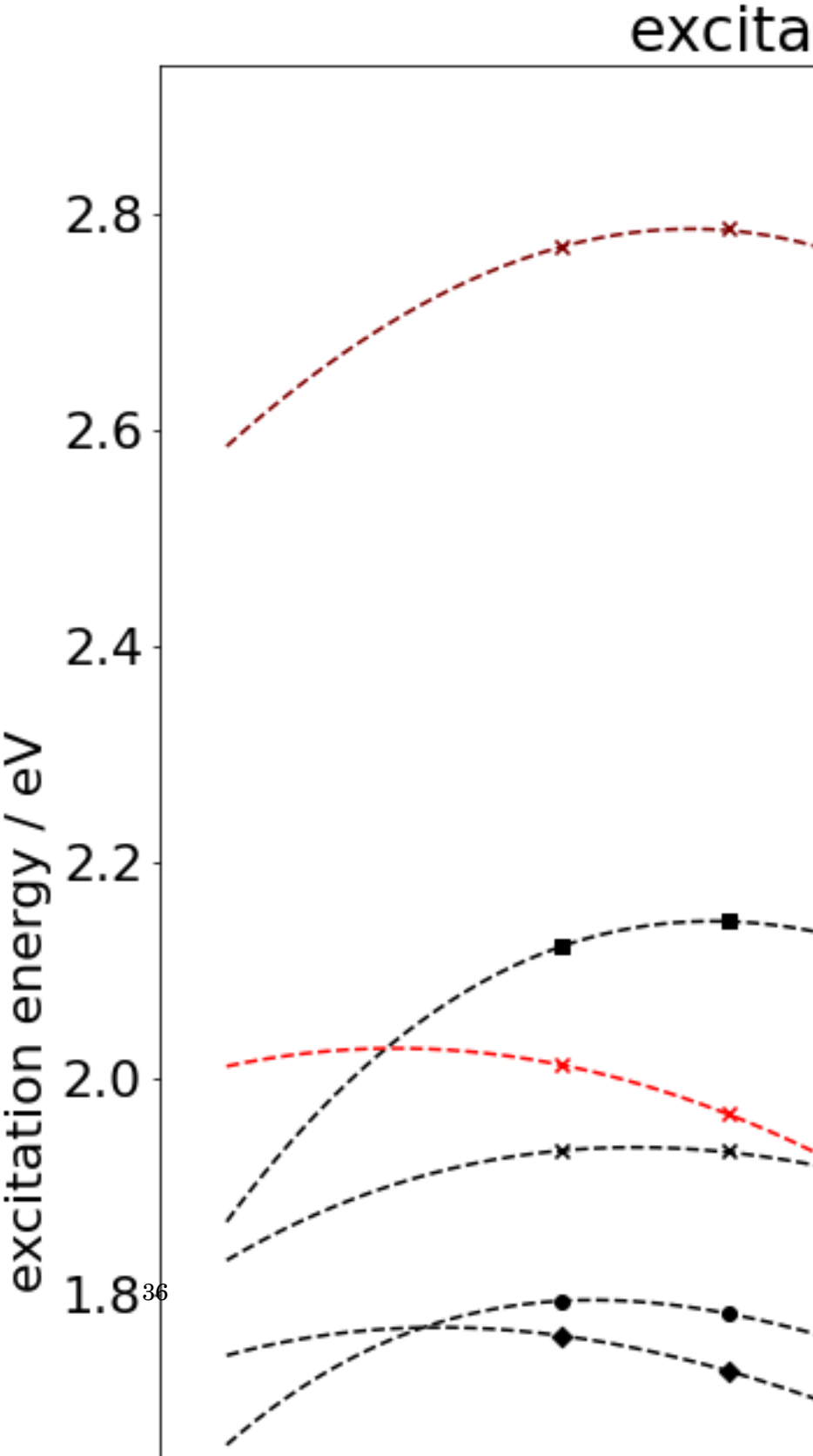

excitatio
2.8
2.6
2.4
2.2
2.0
excitation energy / eV


excita
2.8
2.6
2.4
2.2
2.0
excitation energy / eV


excitatio
2.8
2.6
2.4
2.2
2.0
excitation energy / eV


excita
2.8
2.6
2.4
2.2
2.0
excitation energy / eV


excitatio
2.8
2.6
2.4
2.2
2.0
1.8
excitation energy / eV


excita
2.8
2.6
2.4
2.2
2.0
1.8
excitation energy / eV

CHAPTER 4

# EXCITON METHOD

Preamble

## 4.1 Theory

### 4.1.1 Exciton States

### 4.1.2 Embedding

## 4.2 Truncated Chlorophylls

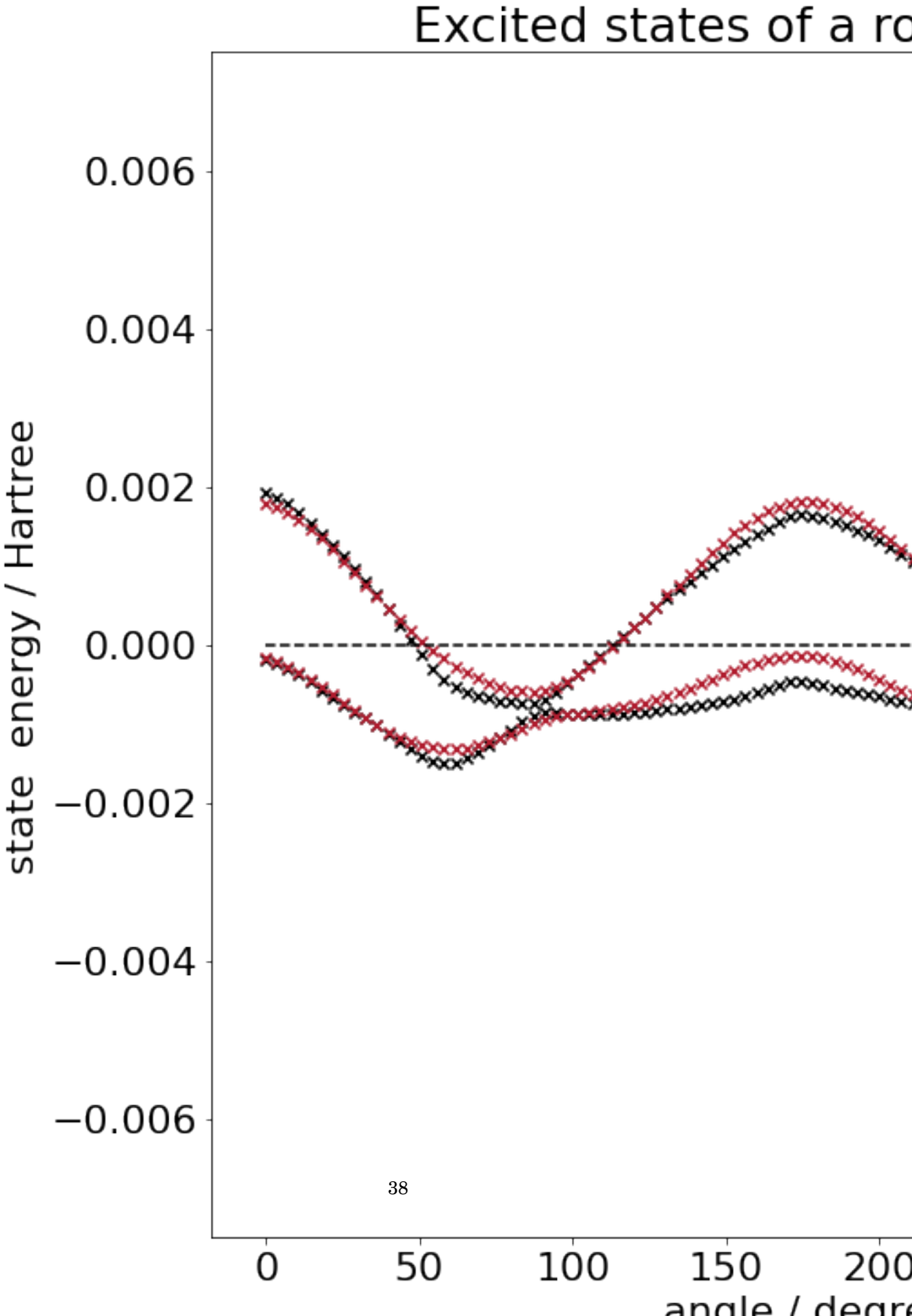
### 4.2.1 Rotation

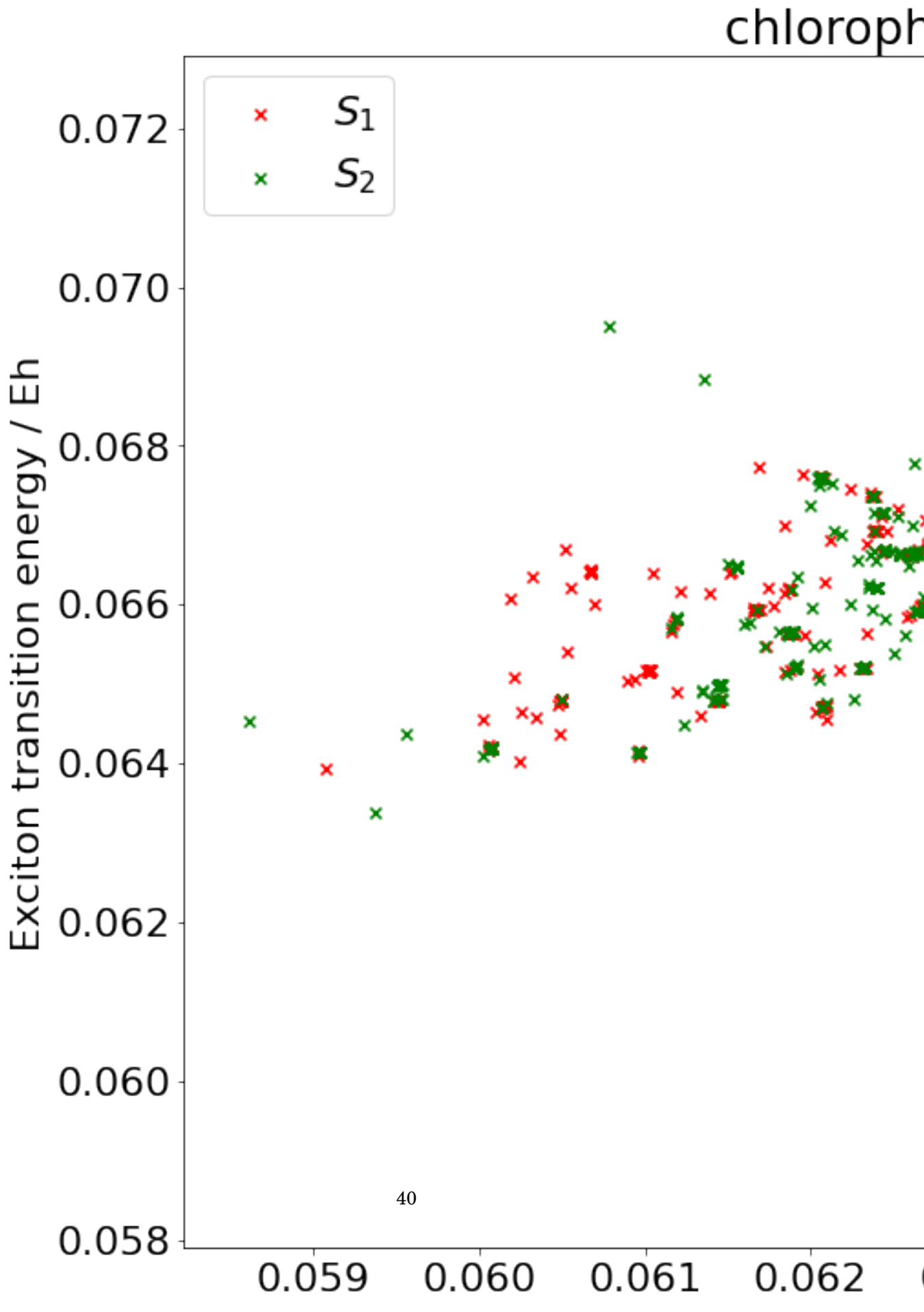
### 4.2.2 Distance

## 4.3 LHII Pairs

### 4.3.1 Assignment of States

### 4.3.2 Comparison


Excited states of a ro
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
state energy / Hartree
0
50
100
150
200
angle / degre

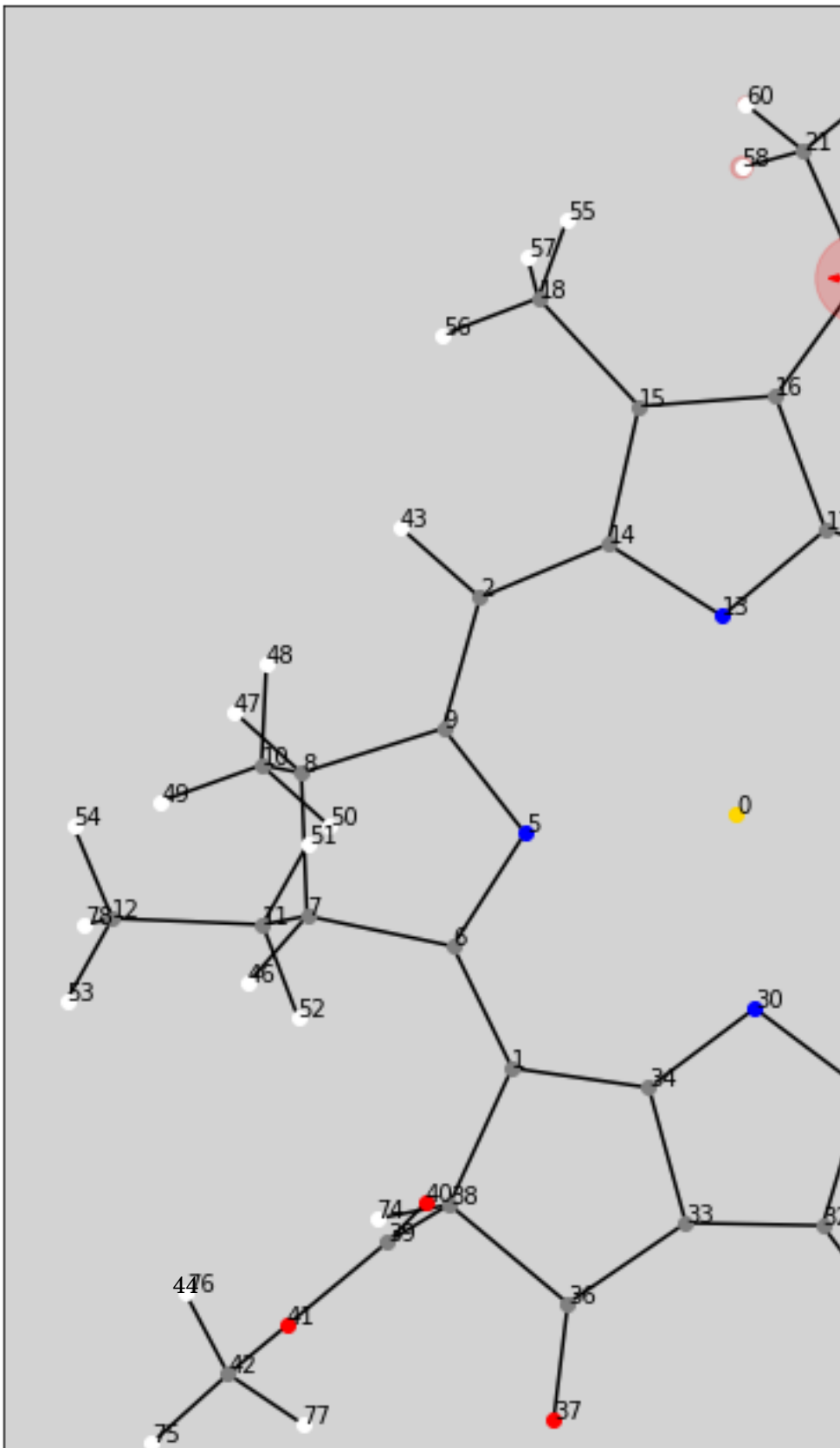

chloroph
$S_1$
$S_2$
Exciton transition energy / Eh
0.072
0.070
0.068
0.066
0.064
0.062
0.060
0.058
0.059
0.060
0.061
0.062

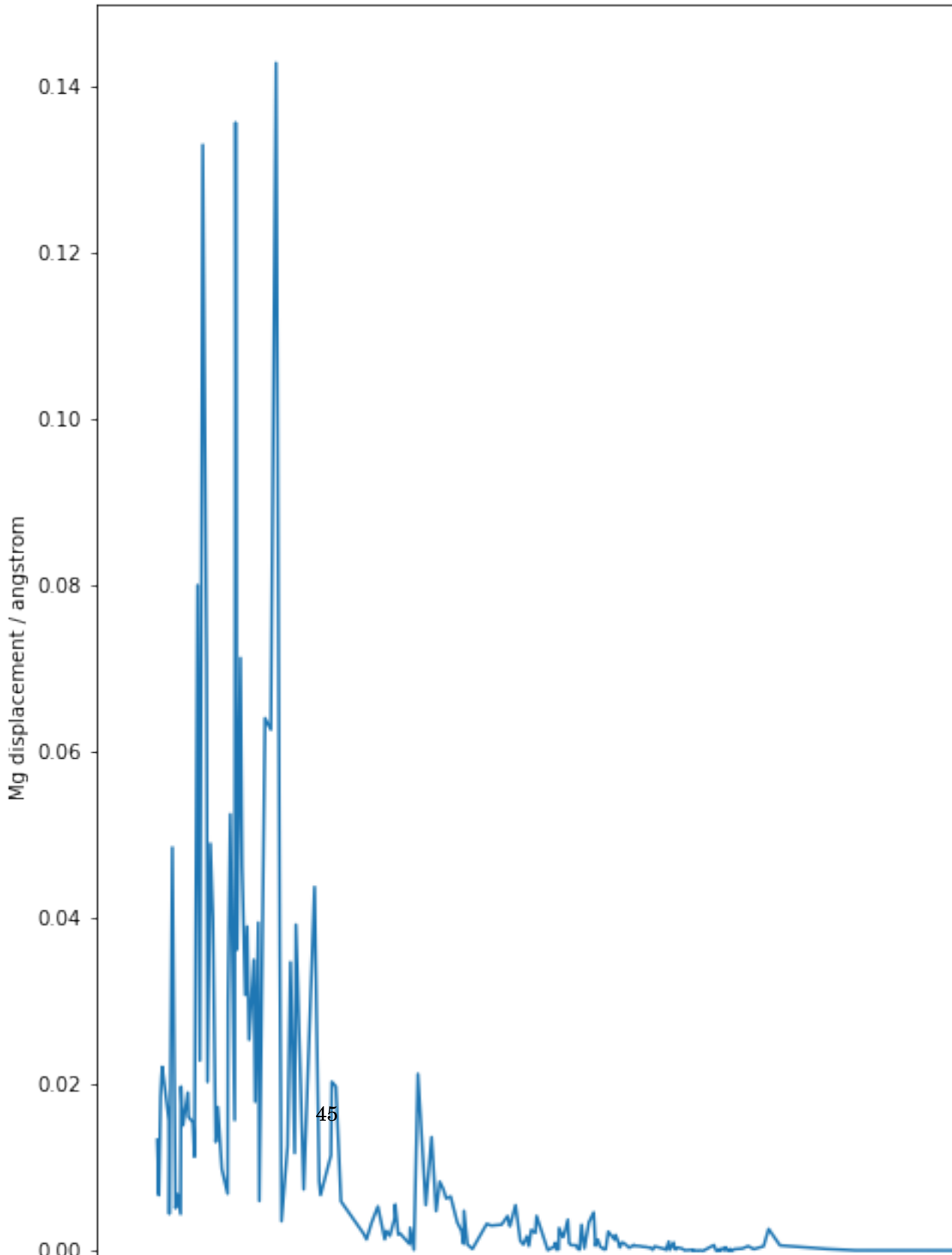
energy ordered
centre ordered

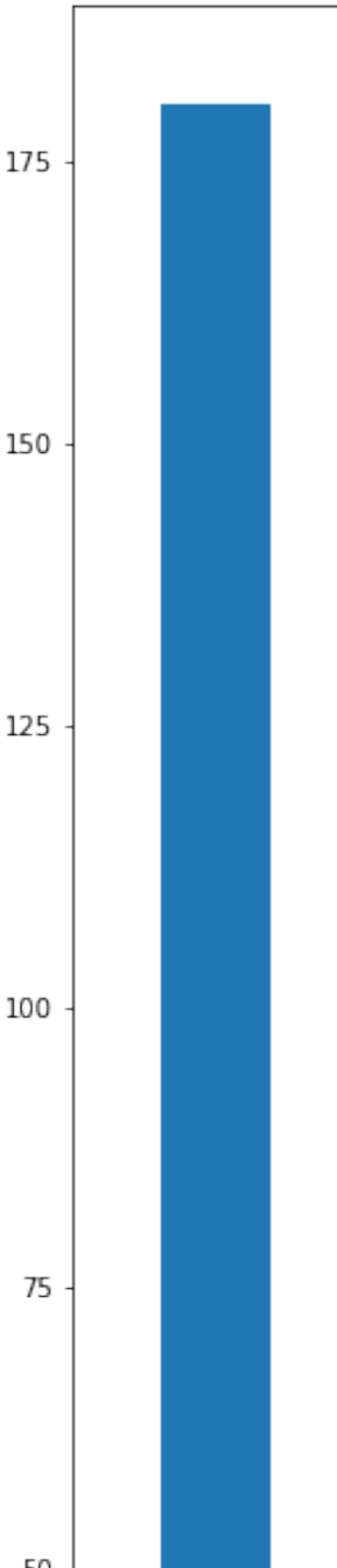
CHAPTER 5

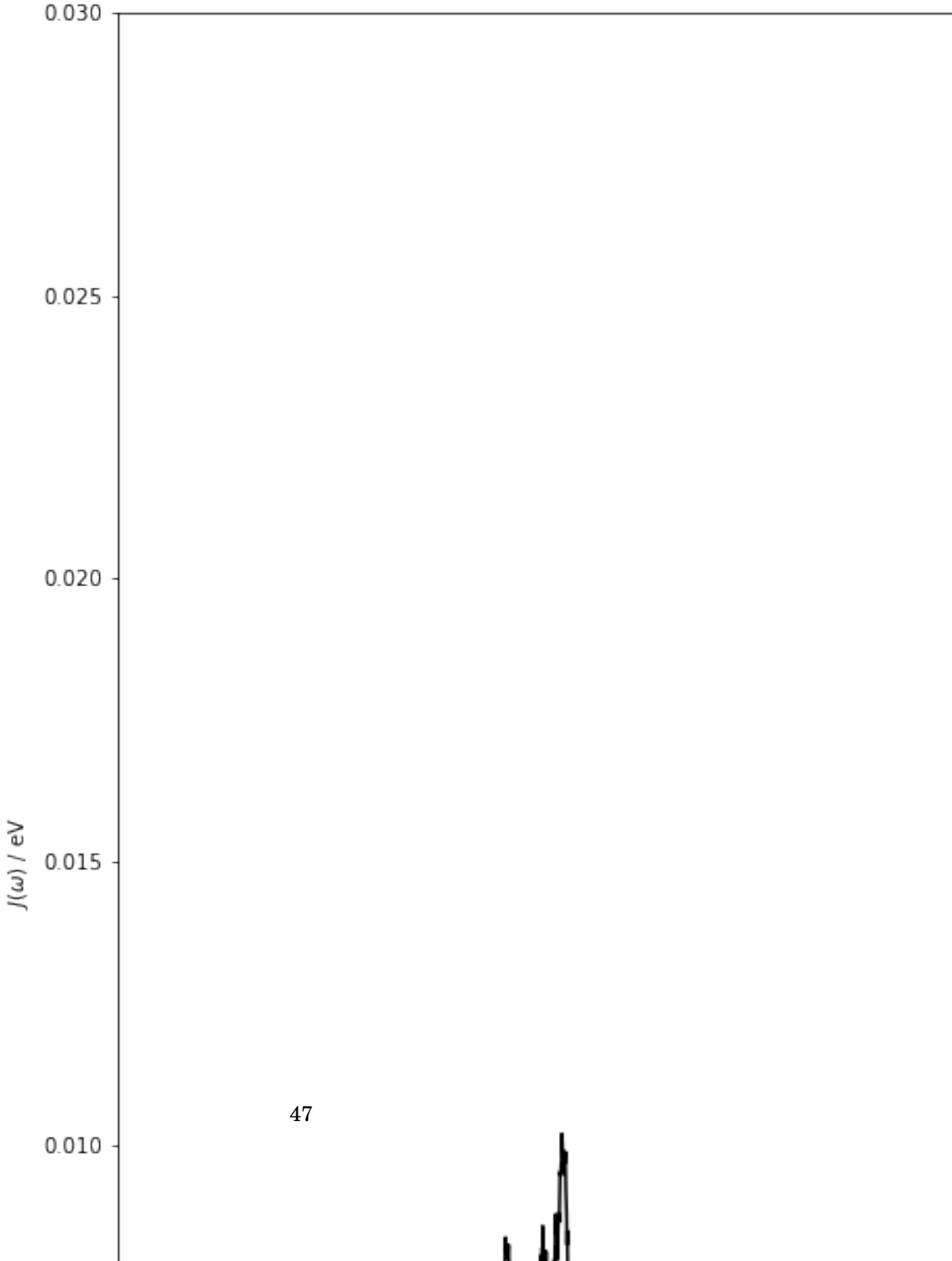
# Atomistic Modelling of Light Harvesting Complexes

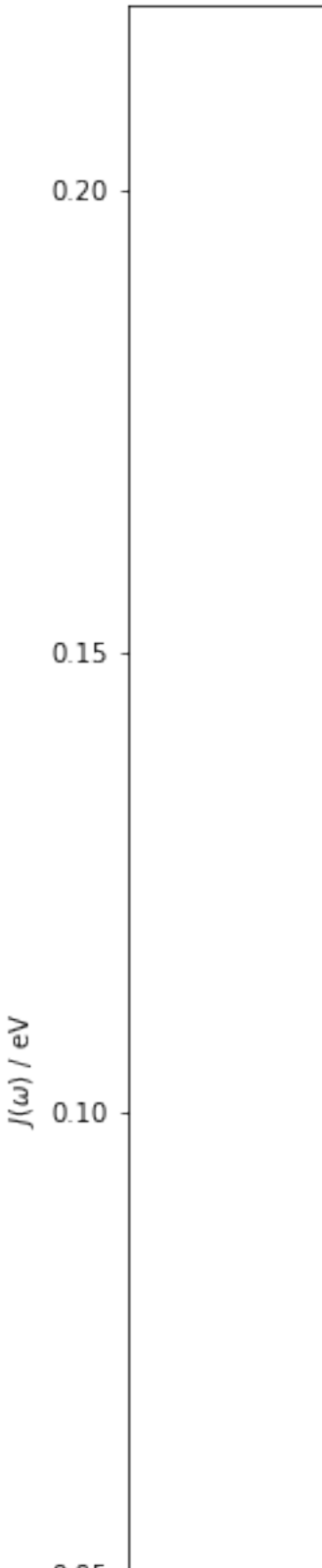
Preamble

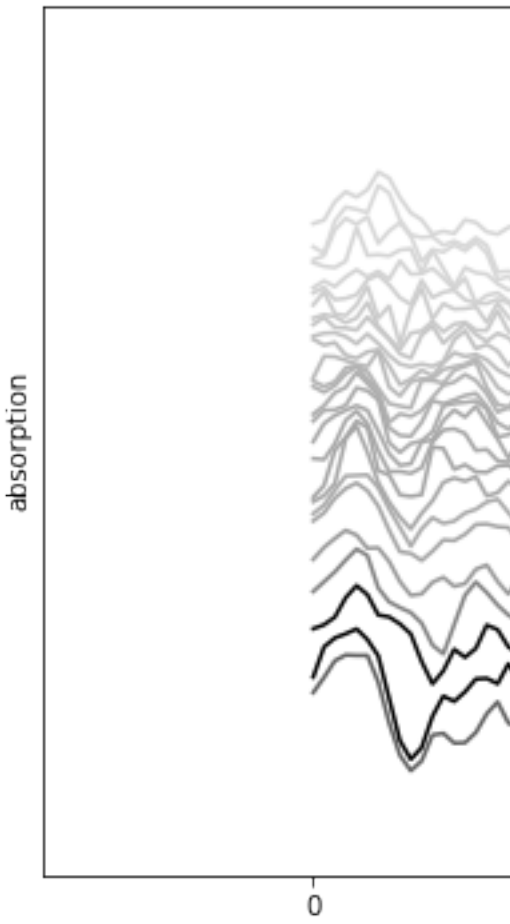

60
21
58
55
57
18
56
15
16
43
14
2
48
47
9
10
8
49
50
54
51
5
0
12
11
7
6
46
53
52
30
1
34
40
38
74
33
39
44
76
41
36
42
77
37
75

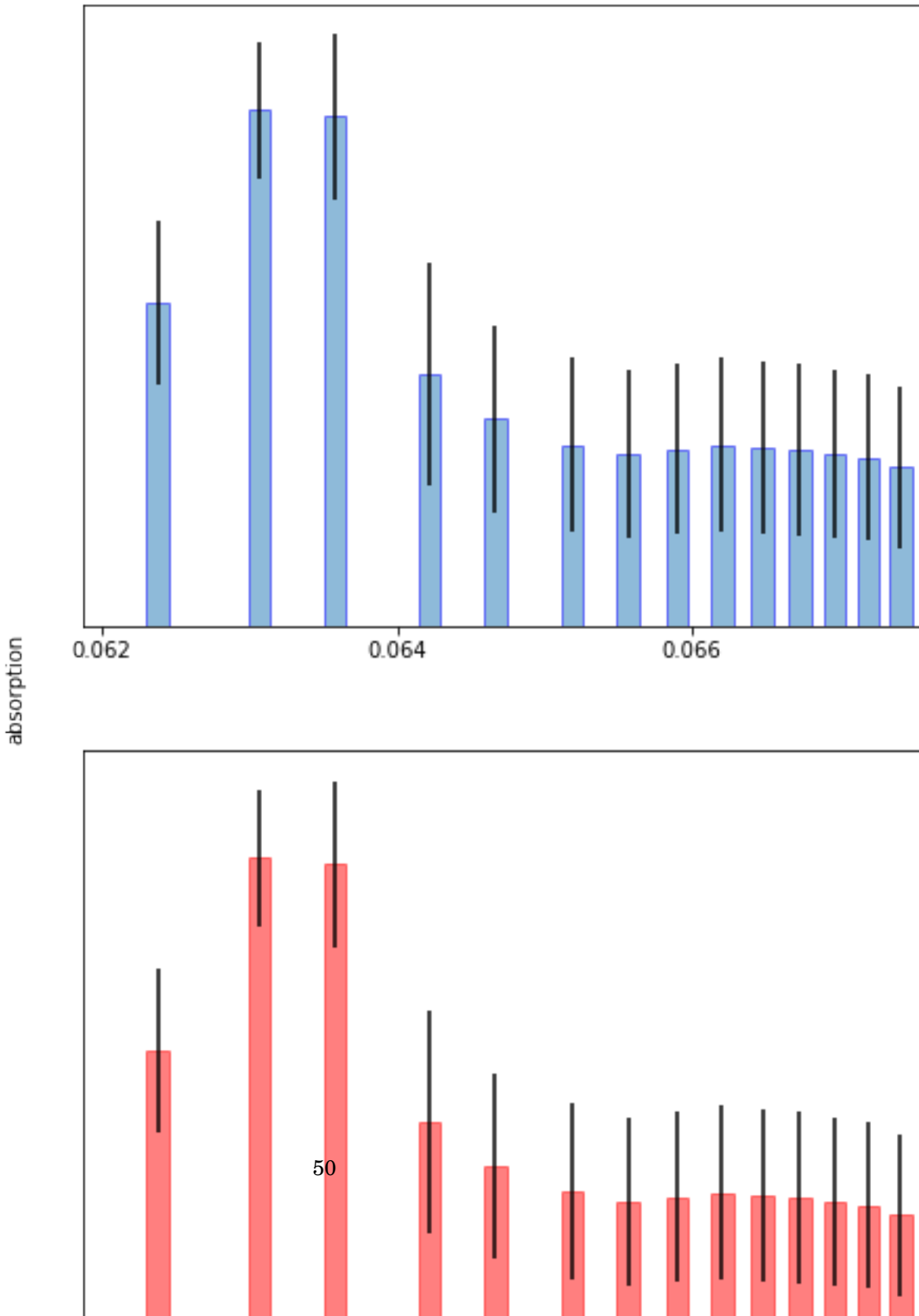

0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Mg displacement / angstrom


175
150
125
100
75


0.030
0.025
0.020
0.015
0.010
J(ω) / eV


0.20
0.15
0.10
J(ω) / eV


absorption
0


absorption
0.062
0.064
0.066

## 5.1 LHII

### 5.1.1 Spectral Density Method

### 5.1.2 Molecular Dynamics Method

## 5.2 Approximating Spectral Densities

### 5.2.1 Hessians

### 5.2.2 Huang Rhys Factors

### 5.2.3 Chlorophyll distances

## 5.3 Environmental Effects

### 5.3.1 Screening

### 5.3.2 Embedding

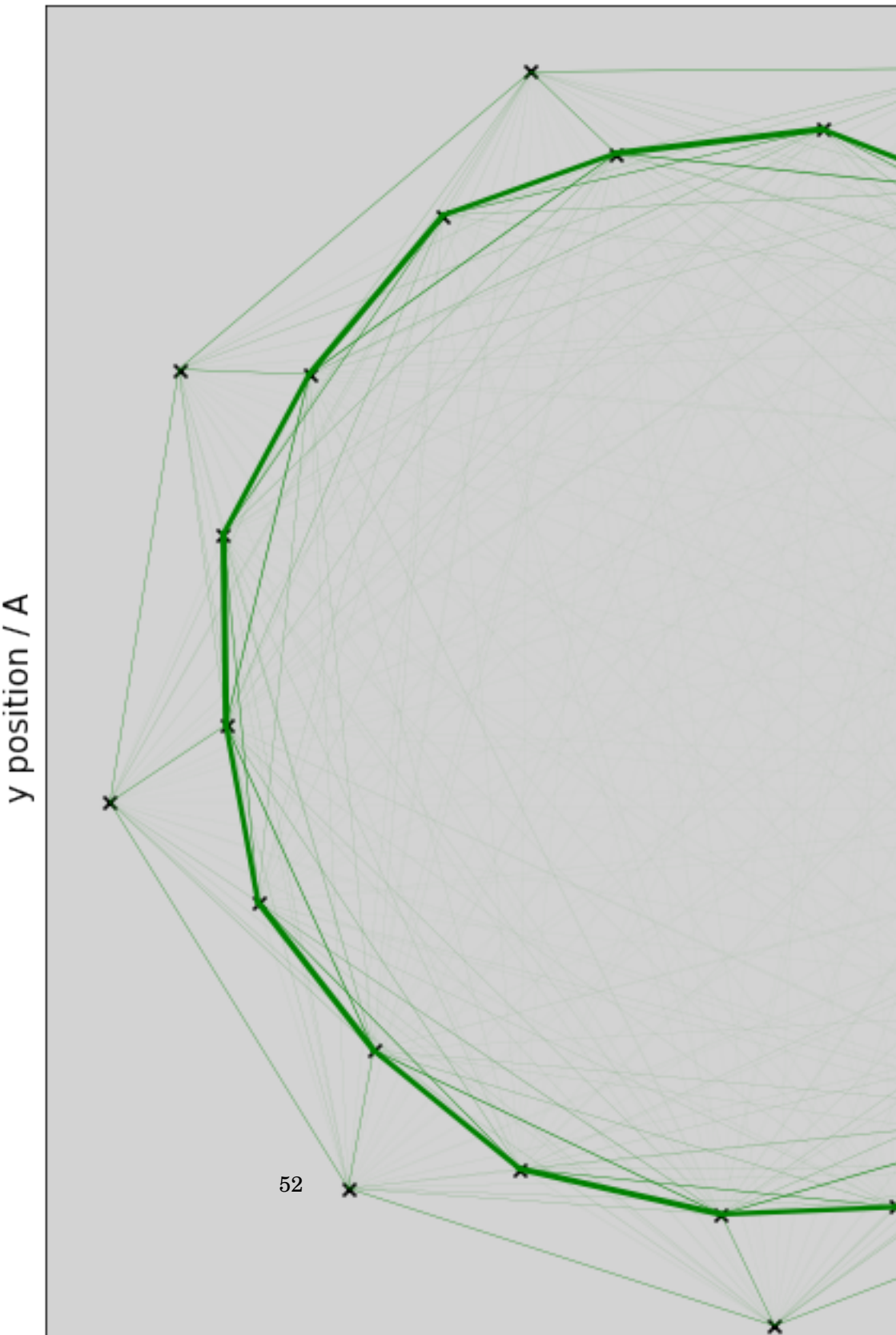
## 5.4 Sites, states and couplings

### 5.4.1 Sites

### 5.4.2 Exciton states

### 5.4.3 Coupling

## 5.5 Excitation Energies


y position / A

CHAPTER 6

# DISCUSSION

Preamble

## 6.1 Transition Property Approximations

## 6.2 Further Investigations into LHII

## 6.3 Coherence


# APPENDIX A

This appendix covers the common computational details of this work. Included are the software packages, hardware used. These are not exhaustive list, and additional details are provided in the main chapters. However, wherever implementations or methodology details are missing, the information will be found here.

## A.1 Electronic Structure Codes

This project has primarily used the QCORE software that is found as part of the ENTOS project. This is a software package for DFT and DFTB electronic structure calculations that has been written as a joint venture between the Miller group in California Institute of Technology and the Manby group in the University of Bristol. It is now being hosted by Entos Inc. It is a novel C++ implementation, with a focus on modularity, functional code and modern development practices to enable easier, cleaner and more reuseable code. All novel methods discussed in the chapters have been implemented in the QCORE package.

## A.2 Computational Hardware

# BIBLIOGRAPHY


[1] A. T. B. GILBERT, N. A. BESLEY, AND P. M. W. GILL, *Self-Consistent Field Calculations of Excited States Using the Maximum Overlap Method (MOM) †*, J. Phys. Chem., 112 (2008), p. 13164.

[2] T. GIMON, A. IPATOV, A. HESSELMANN, AND A. GÖRLING, *Qualitatively Correct Charge-Transfer Exictation Energies in HeH+ by Time-Dependent Density-Functional Theory Due to Exact Exchange Kohn-Sham Eigenvalue Differences*, J. Chem. Theory Comput, 5 (2009), p. 27.

[3] S. GRIMME AND C. BANNWARTH, *Ultra-fast computation of electronic spectra for large systems by tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB)*, J. Chem. Phys., 145 (2016), p. 54103.

[4] S. W. J. HUNT AND W. A. GODDARD, *Excited States of H2O using improved virtual orbitals*, Chem. Phys. Lett., 3 (1969), pp. 414–418.

[5] G. KLOPMAN, V. 86, J. LINEVSKY, K. S. SESHADRI, AND D. WHITE, *A Semiempirical Treatment of Molecular Structures. II. Molecular Terms and Application to Diatomic Molecules*, J. Am. Chem. Soc., 86 (1964), pp. 4550–4557.

[6] K. NISHIMOTO AND N. MATAGA, *Electronic Structure and Spectra of Some Nitrogen Heterocycles*, Zeitshrift fur Phys. Chemie Neue Folge, 12 (1957), pp. 335–338.

[7] K. OHNO, *Some Remarks on the Pariser-Parr-Pople Method*, Theor. chim. Acta, 2 (1964), pp. 219–227.

[8] C. C. J. ROOTHAAN, *New Developments in Molecular Orbital Theory*, Rev. Mod. Phys., 23 (1951), p. 69.

[9] S. B. WORSTER, O. FEIGHAN, AND F. R. MANBY, *Reliable transition properties from excited-state mean-field calculations*, J. Chem. Phys, 154 (2021), p. 124106.