
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

	Page
List of Tables	xi
List of Figures	xiii
1 Introduction	1
1.1 Quantum Exploits in Light Harvesting Systems	1
1.1.1 Electronic Energy Transfer	1
1.1.2 Coherence	1
1.2 Light-Matter Response	1
1.3 Electronic Structure for Large Systems	1
1.4 The Aim	1
2 Mean-Field excited states	3
2.1 Theory	3
2.1.1 Δ -SCF and eigenvalue difference	3
2.1.2 Semi-empirical extensions	4
2.2 Benchmarking	6
2.2.1 Small Systems	7
2.2.2 LHII Chlorophyll	7
2.2.3 GFN methods	7
2.3 Extensibility	10
2.3.1 Embedding	10
2.3.2 Scaling	10
3 Chlorophyll specific methods	11
3.1 The Cassida equation	11
3.1.1 Approximations to Solutions	11
3.1.2 MNOK Integrals	11
3.2 Parameterization	11
3.2.1 Reference Data	11

TABLE OF CONTENTS

3.2.2	Objective Function	11
3.2.3	Minimization Algorithms	11
3.3	Benchmarking	11
3.3.1	Transition properties	11
3.3.2	Potential Energy Surfaces	11
3.3.3	Absorption Spectra	11
4	Exciton Method	29
4.1	Theory	29
4.1.1	Exciton States	29
4.1.2	Embedding	29
4.2	Truncated Chlorophylls	29
4.2.1	Rotation	29
4.2.2	Distance	29
4.3	LHII Pairs	29
4.3.1	Assignment of States	29
4.3.2	Comparison	29
5	Atomistic Modelling of Light Harvesting Complexes	31
5.1	LHII	31
5.1.1	Spectral Density Method	31
5.1.2	Molecular Dynamics Method	31
5.2	Approximating Spectral Densities	31
5.2.1	Hessians	31
5.2.2	Chlorophyll distances	31
5.3	Sites, states and couplings	31
5.3.1	Sites	31
5.3.2	Exciton states	31
5.3.3	Coupling	31
5.4	Excitation Energies	31
6	Discussion	33
6.1	Transition Property Approximations	33
6.2	Further Investigations into LHII	33
6.3	Coherence	33
A	Appendix A	35
A.1	Electronic Structure Codes	35
A.2	Computational Hardware	35

Bibliography	37
---------------------	-----------

LIST OF TABLES

TABLE	Page
-------	------

LIST OF FIGURES

FIGURE	Page
2.1 A caption	8
2.2 Symmetry assigned MOs	10
3.1 correlations of energies	12
3.2 correlations of transition dipole moments	13
3.3 Nelder-mead	14
3.4 SLSQP	15
3.5 optimized parameters from SLSQP procedure.	15
3.6 training scatter	16

INTRODUCTION

Naturally occurring light harvesting systems present an interesting scientific challenge. With near perfect efficiency, the energy from a photon will be taken and transferred 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

MEAN-FIELD EXCITED STATES

Preamble

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 a 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:

$$(2.1) \quad \mathbf{FC}^{\text{new}} = \mathbf{SC}^{\text{new}} \epsilon$$

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

$$(2.2) \quad \mathbf{F} = f(\mathbf{C}^{\text{old}})$$

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

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

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

$$(2.4) \quad p_j = \sum_i^n O_{ij} = \sum_v^N \left[\sum_\mu^N \left(\sum_i^n C_{i\mu}^{\text{old}} \right) S_{\mu v} \right] C_{vj}^{\text{new}}$$

where μ, v 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 Δ -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).

Δ -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 \mathbf{A} and \mathbf{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 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, standing 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 this method has similar goals as this project, it is useful to look at this in more detail.

2.1.2.1 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:

$$(2.5) \quad \langle \psi_\mu | H^{\text{EHT, sTDA-xTB}} | \psi_\mu \rangle = \frac{1}{2} \left(k_\mu^l k_\nu^{l'} \right) \frac{1}{2} \left(h_\mu^l h_\nu^{l'} \right) S_{\mu\nu} - k_T \langle \psi_\mu | \hat{T} | \psi_\nu \rangle$$

where, μ, ν, l, l' are orbital and shell indices k_μ^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 diagonalizing 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 Tamm-Dancoff Approximation. There are several approximations made between full linear response theory and the sTDA method. First is the Tamm-Dancoff 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 coulomb function, with different exponents γ_K and j_J for exchange and coulomb integral respectively. The α_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 a Δ -SCF 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 Small Systems

2.2.1.1 Non-orthogonality

2.2.2 LHII Chlorophyll

2.2.3 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.

Δ -xTB, in this initial form, as well as linear response GFN-xTB, is inaccurate at predicting excitation energies. The range of errors is quite large and there is a significant systematic shift in the errors as well. As the DFT methods, both the linear response and the Δ -SCF methods, are quite accurate at predicting the excitation energy, we conclude that the tight-binding method employed is the leading cause of the error. There could be several factors for this error. For Δ -xTB, the excited state SCF cycle could converge to the wrong state or collapse to the ground state. This would be evident by the symmetry of the excitation, as if the DFT and DFTB methods converged to different states then the symmetry would also be different. We investigated how we might assign symmetries to Δ -xTB transitions, and Δ -SCF in general. This is not a trivial task.

2.2.3.1 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

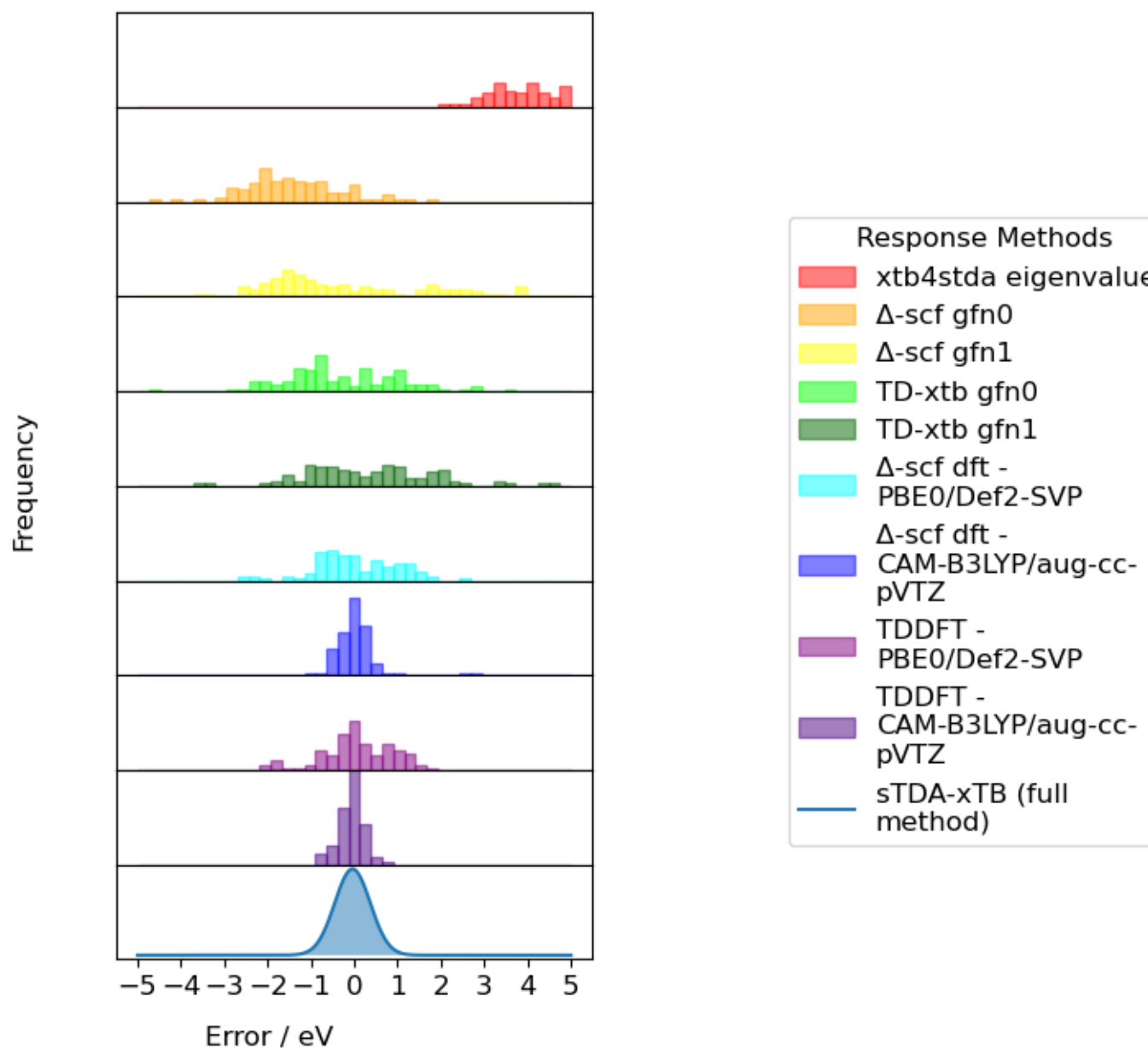


Figure 2.1: A caption

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. 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, so allows the Hamiltonian to be organized into a block diagonal matrix, which can be useful when solving for a large basis set or large system. However, this is only true if the system is highly symmetric, which is often not the case when treating systems from a molecular dynamics simulation, and definitely not the case when looking at biological systems. The second approach 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. We used the open source library libmsym for point group assignment routines and finding the symmetry adapted linear combination of atomic orbitals. Broadly, the outline of 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. However, we encountered two problems. First was that the assignment of MOs broke down for excited states. The second was that for non-abelian groups, where there are some degenerate subspaces such as E and T, this assignment also didn't work. We discussed some ideas about using symmetry decomposition for the non-abelian point groups, but as this was a short study we decided that this might not be the most prudent path. Hence, we could not confidently assign transitions from Δ -SCF methods with this method. The importance of this is that symmetry couldn't be assigned automatically. We still

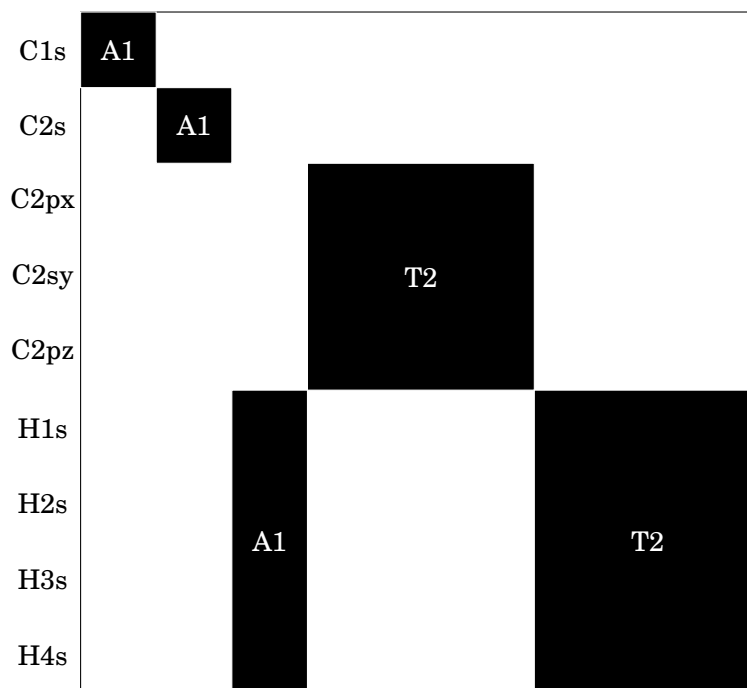


Figure 2.2: Symmetry assigned MOs

had the method of assigning transitions by looking at transition dipoles and plots of molecular orbitals, but these methods require a person to look at the problem at hand. This makes any benchmarking a lengthy process, which isn't repeatable as we extend the methods included in the benchmarking.

2.2.3.2 Implications of Benchmarking Results

There may also be inaccurate treatment of inter-electronic repulsion and correlation which is found in other tight binding methods. Additionally, the parameters in GFN1 and GFN0 are not optimized for excited states.

We also investigated an eigenvalue difference method, using eigenvalues from the `xtb4stda` program. This is the program that provides the molecular orbital coefficients for the `stda` program. This was done to initially investigate whether an eigenvalue difference method for an excited state parameterized method would be accurate, but this proves to not be the case.

2.3 Extensibility

2.3.1 Embedding

2.3.2 Scaling

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	camb3lyp
PBE0	1	0.5891	0.566434	0.762861
dscf	0.5891	1	0.977046	0.800859
eigdiff	0.566434	0.977046	1	0.773252
camb3lyp	0.762861	0.800859	0.773252	1
BLYP	0.0832043	-0.0344407	-0.0098512	0.0500000

Figure 3.2: correlations of transition dipole moments

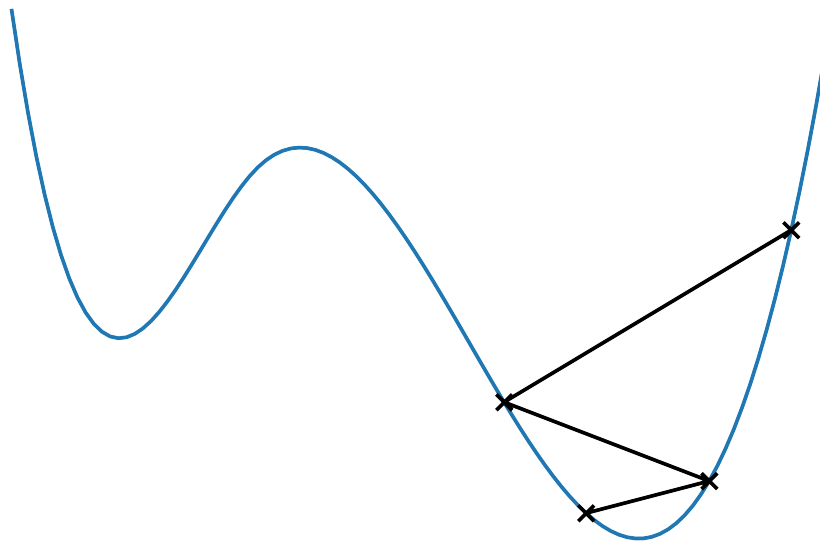


Figure 3.3: Nelder-mead

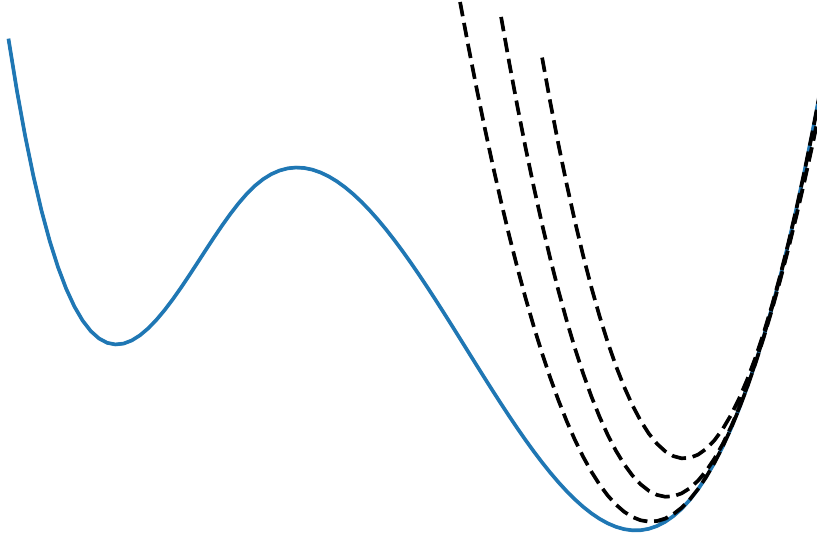
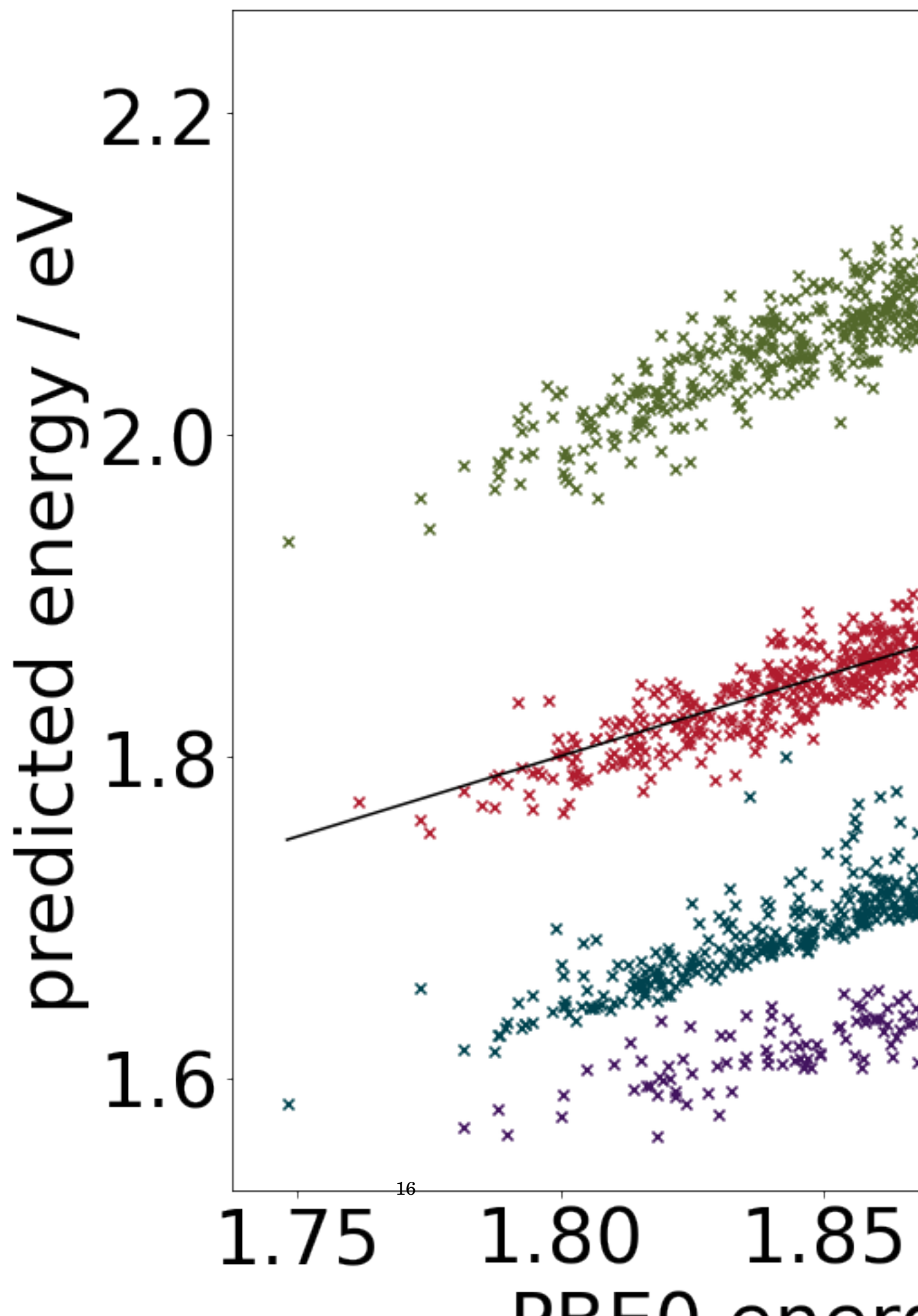
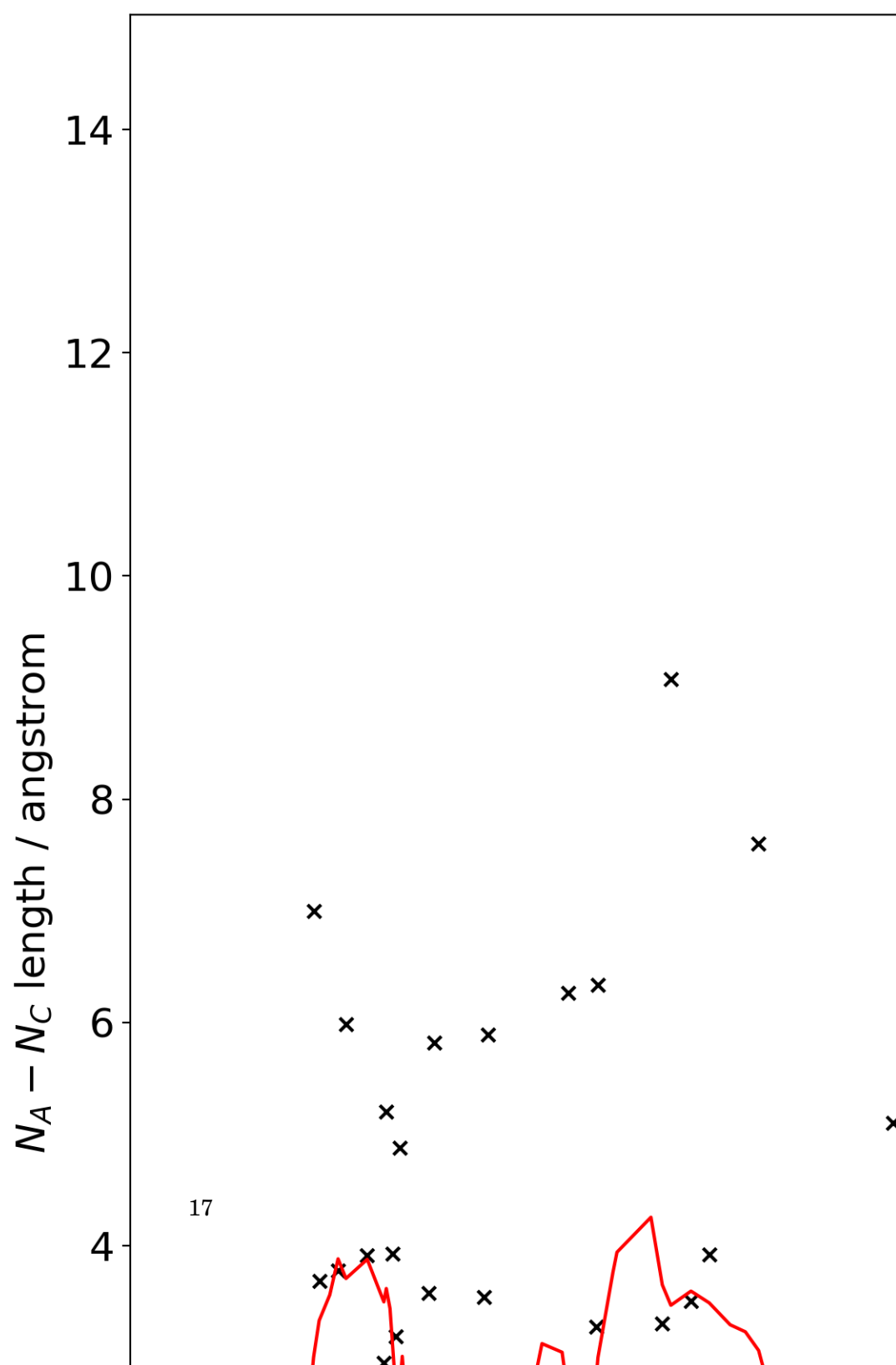


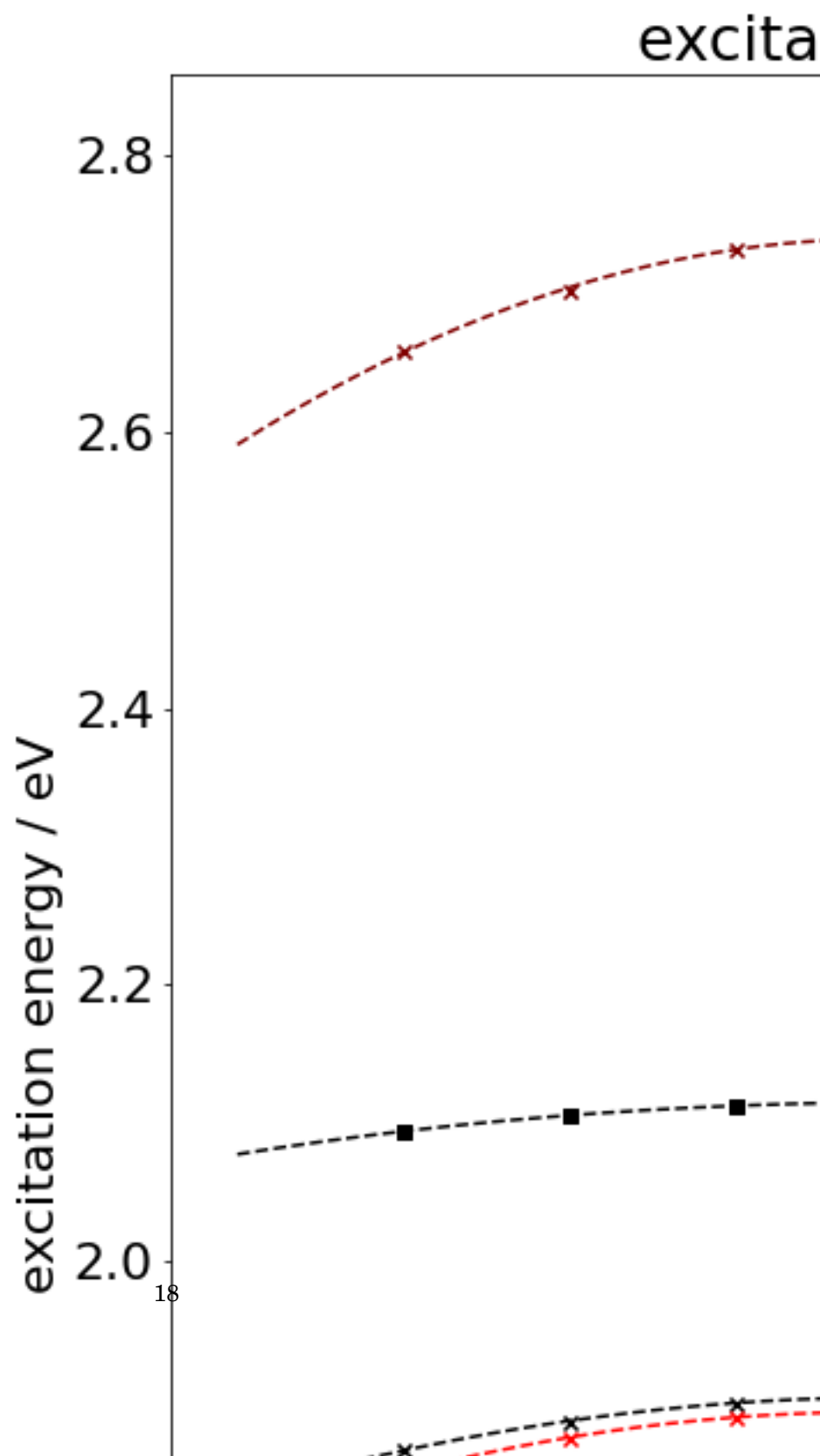
Figure 3.4: SLSQP

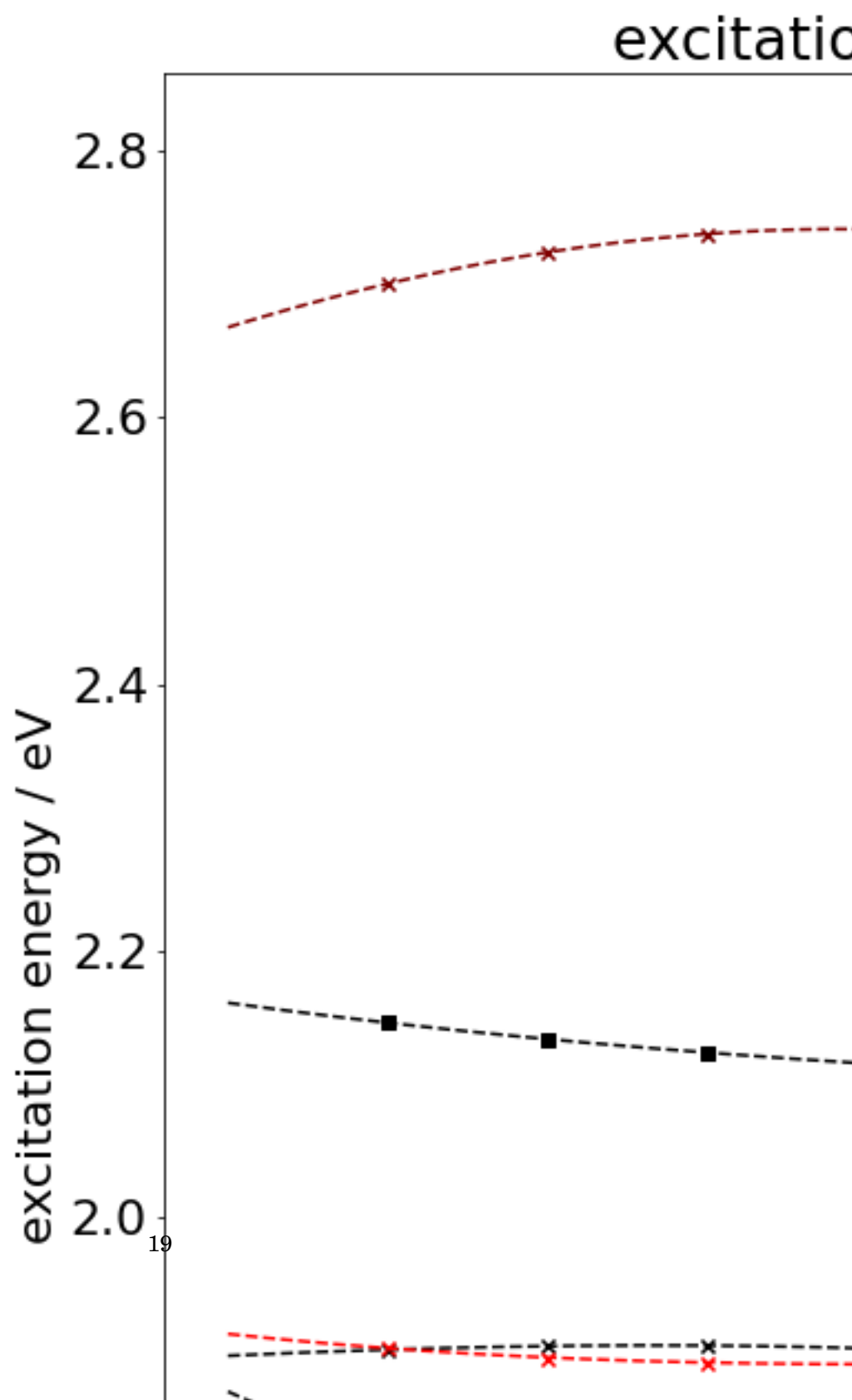
y_K	1.0
y_J	1.0
a_x	1.0

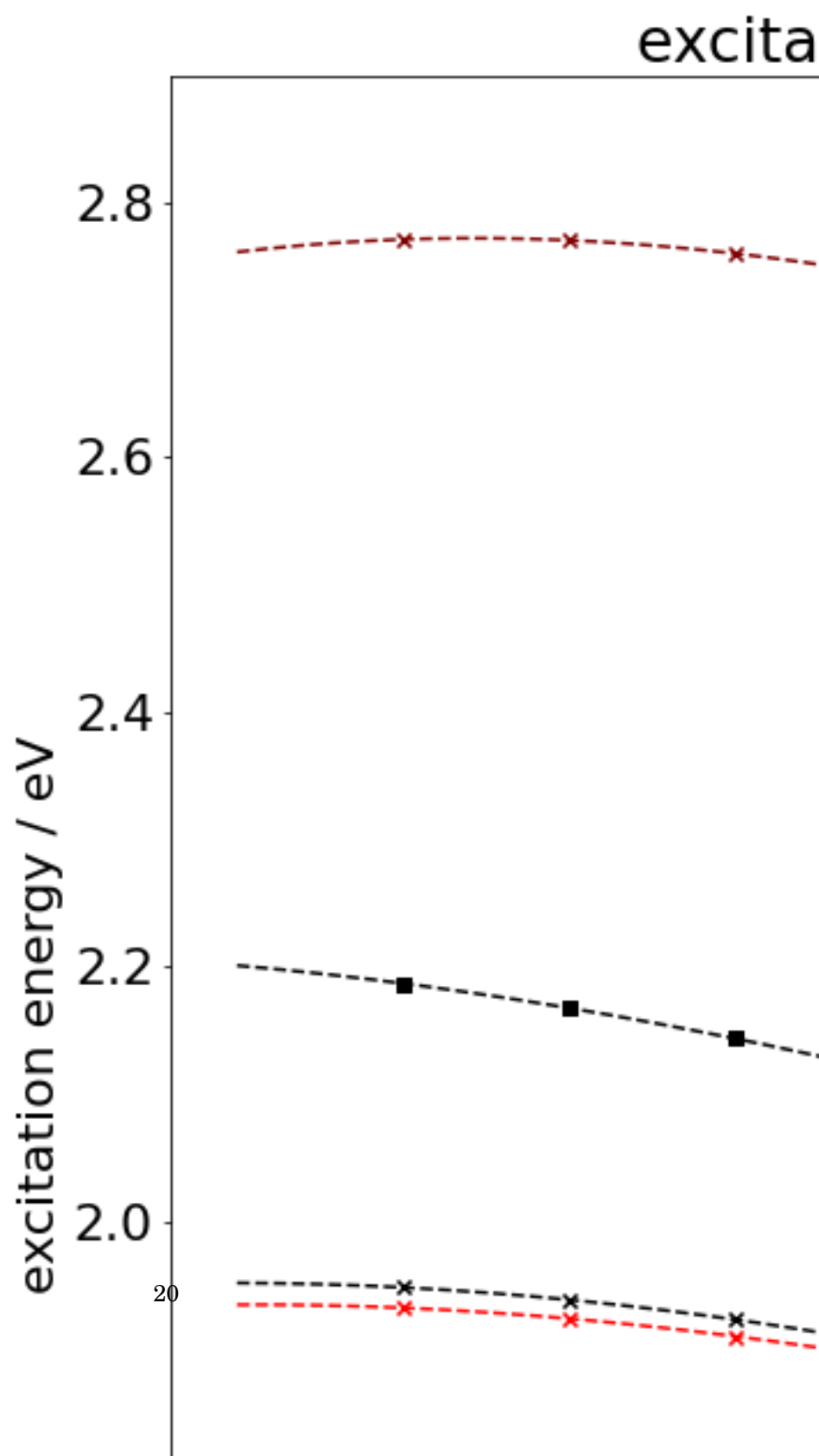
Figure 3.5: optimized parameters from SLSQP procedure.

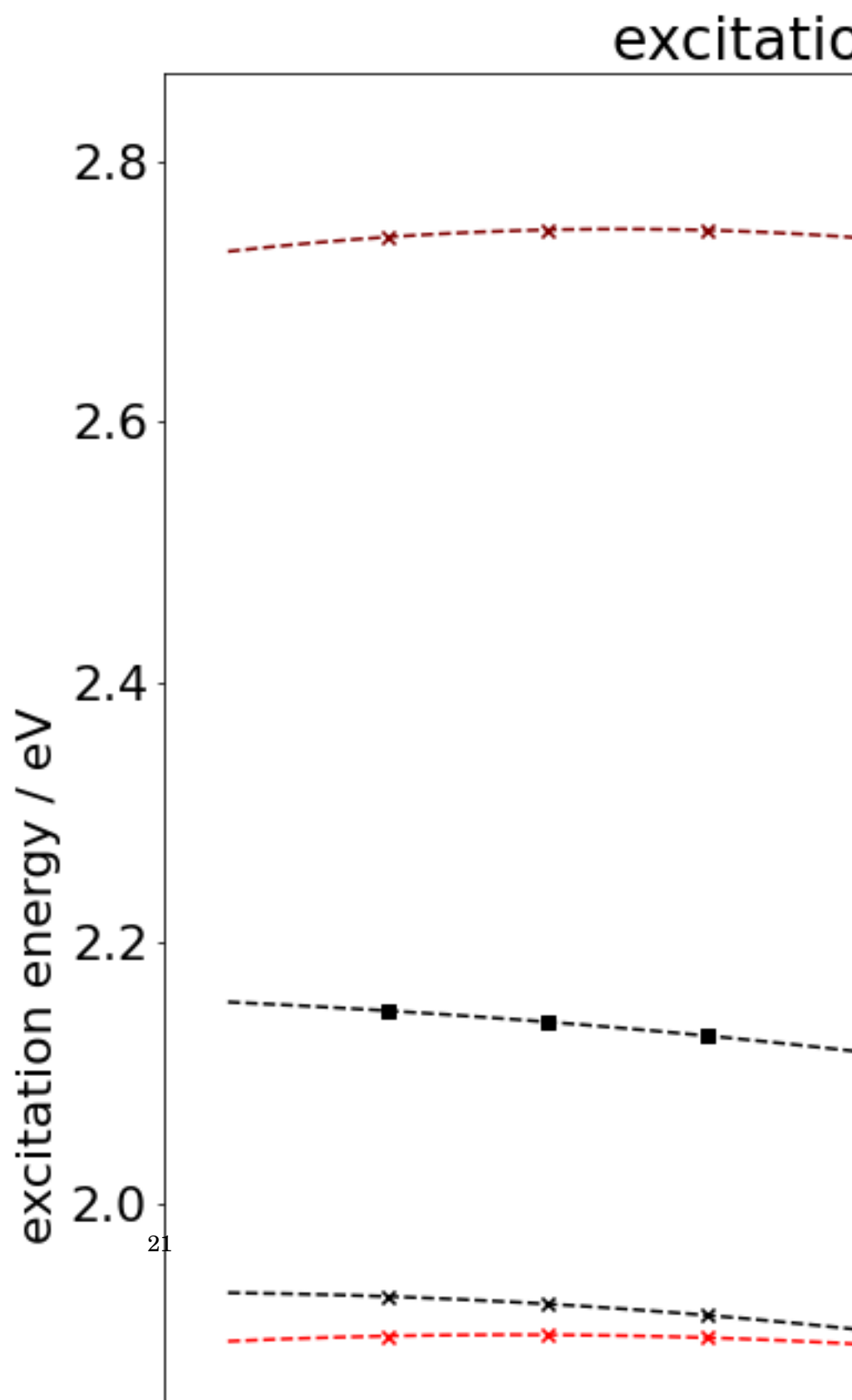


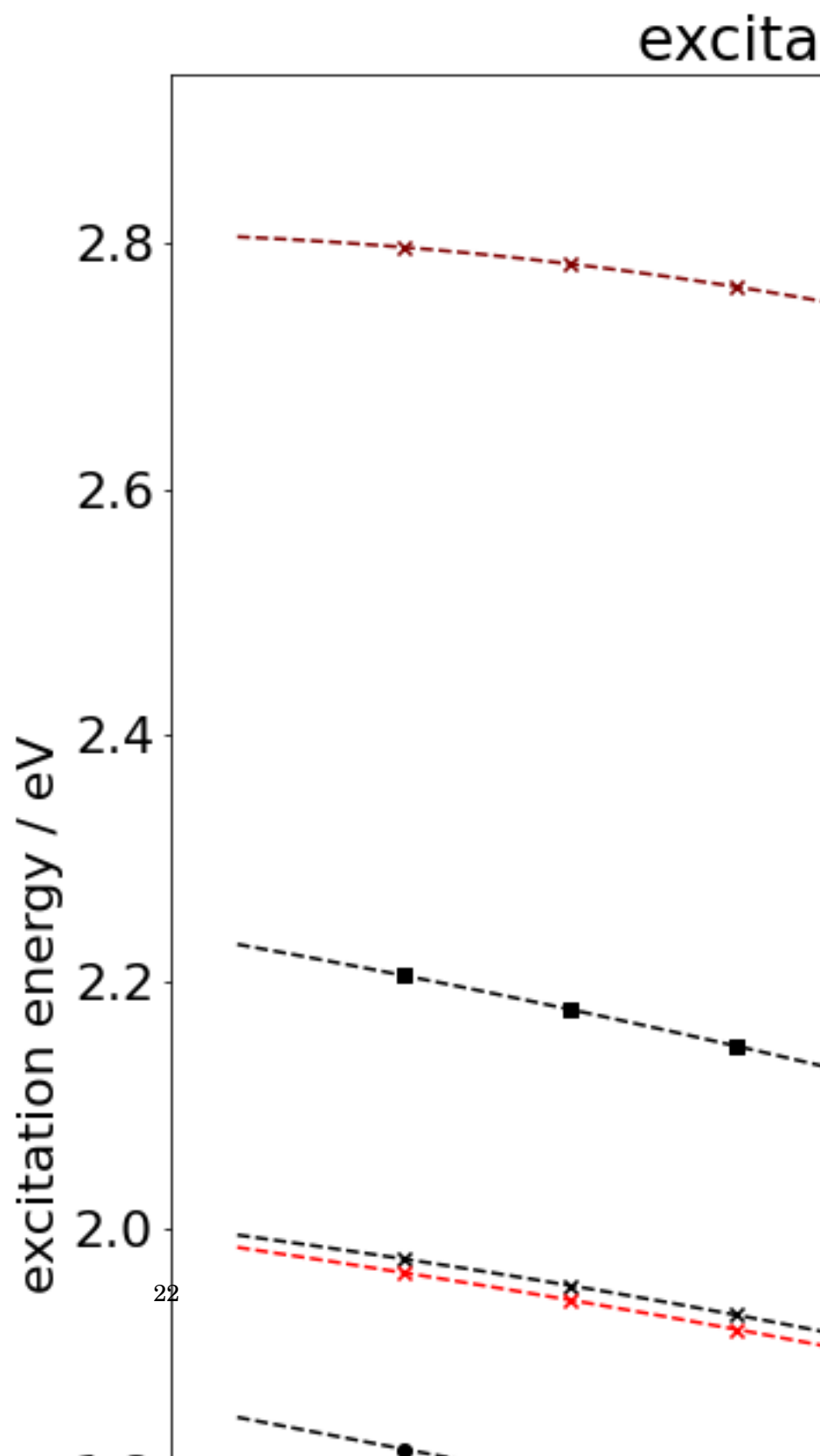


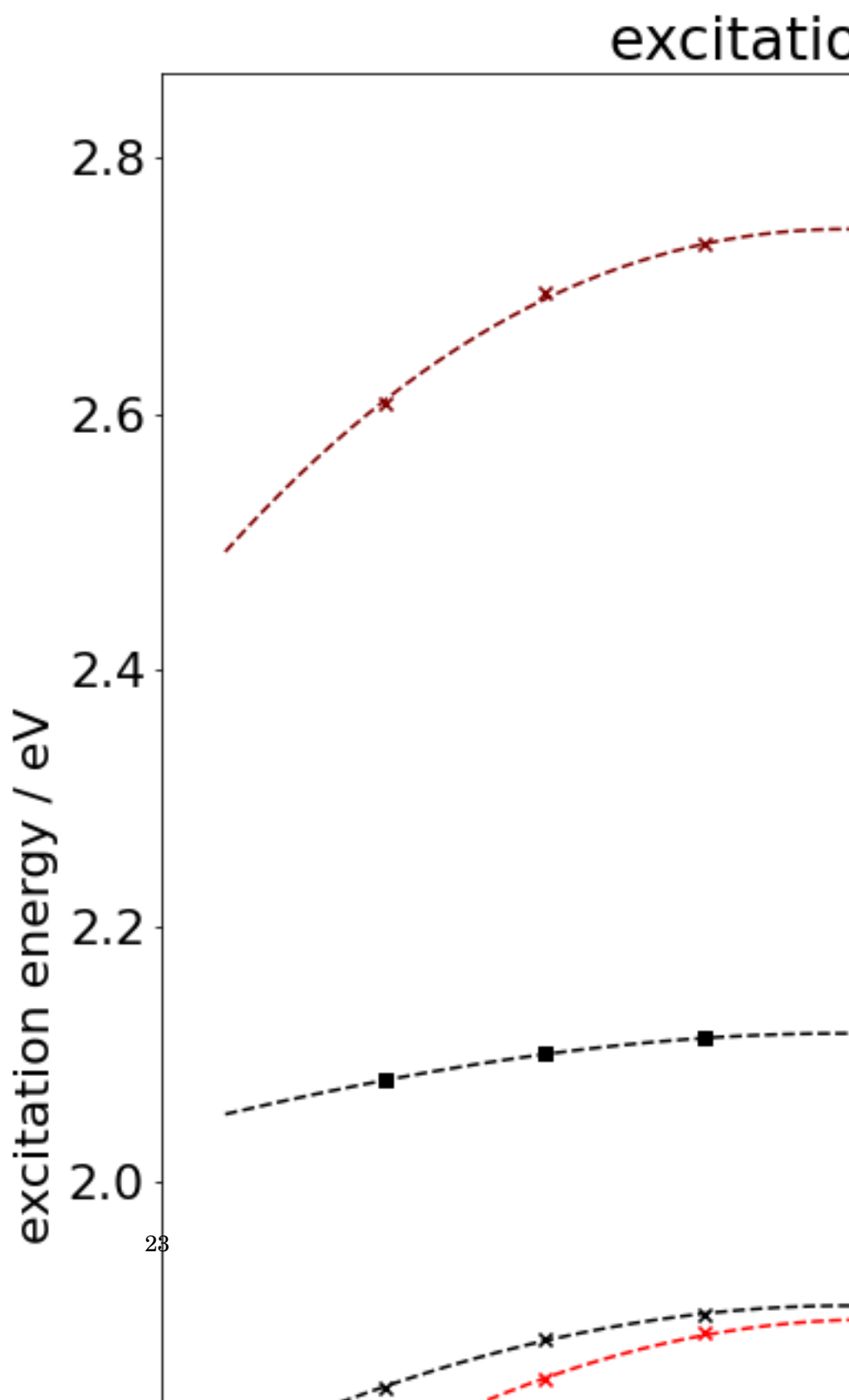


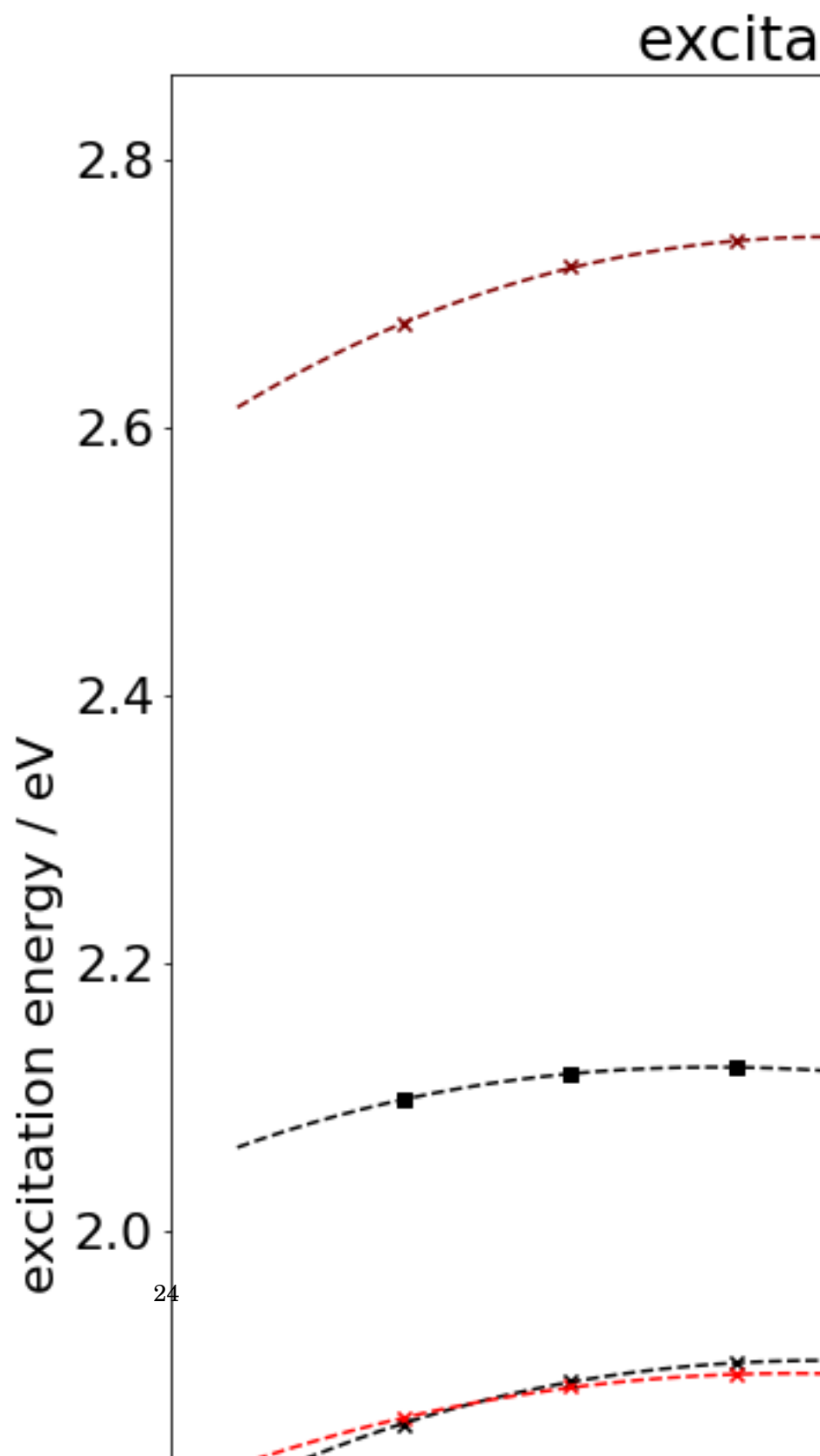


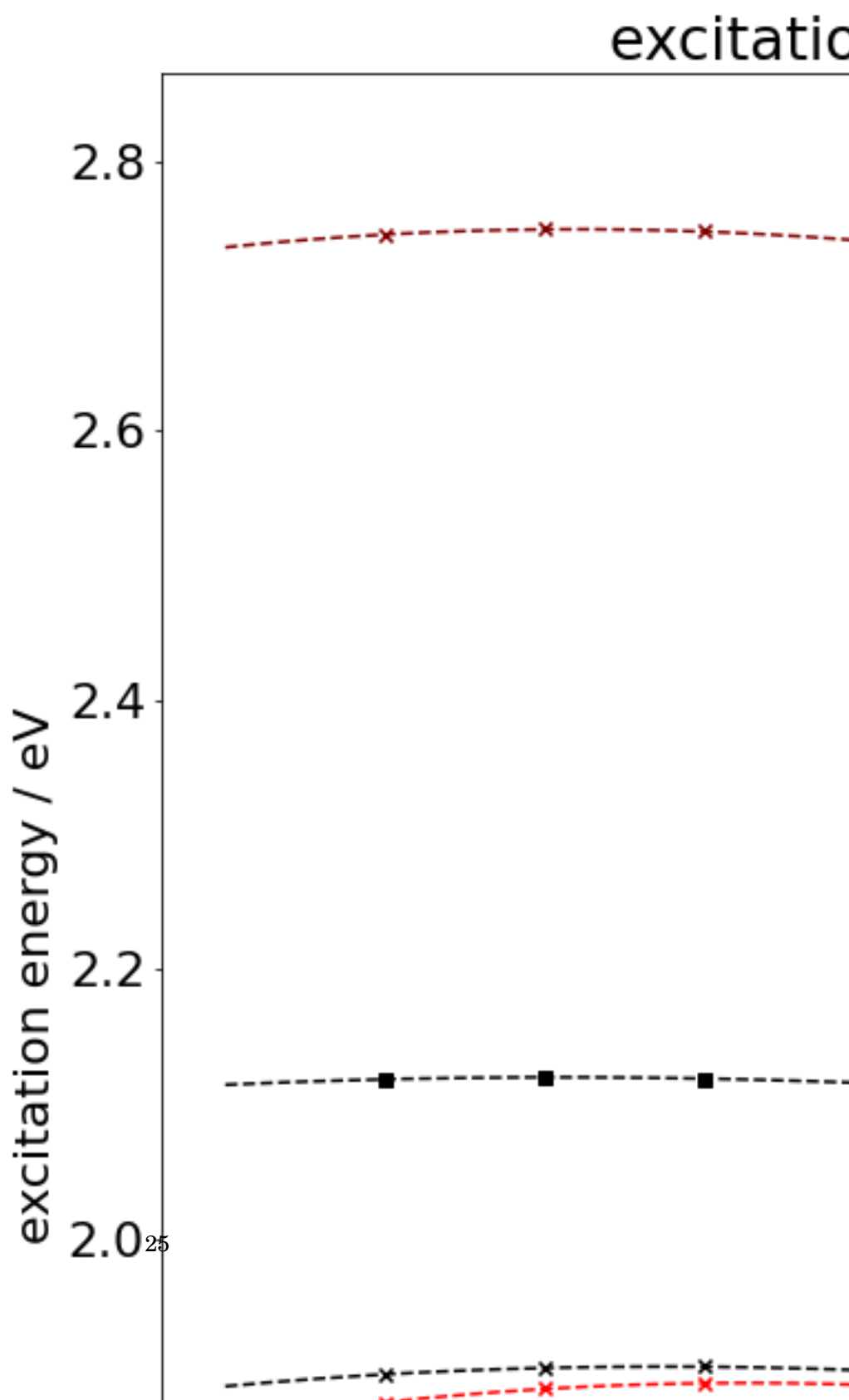


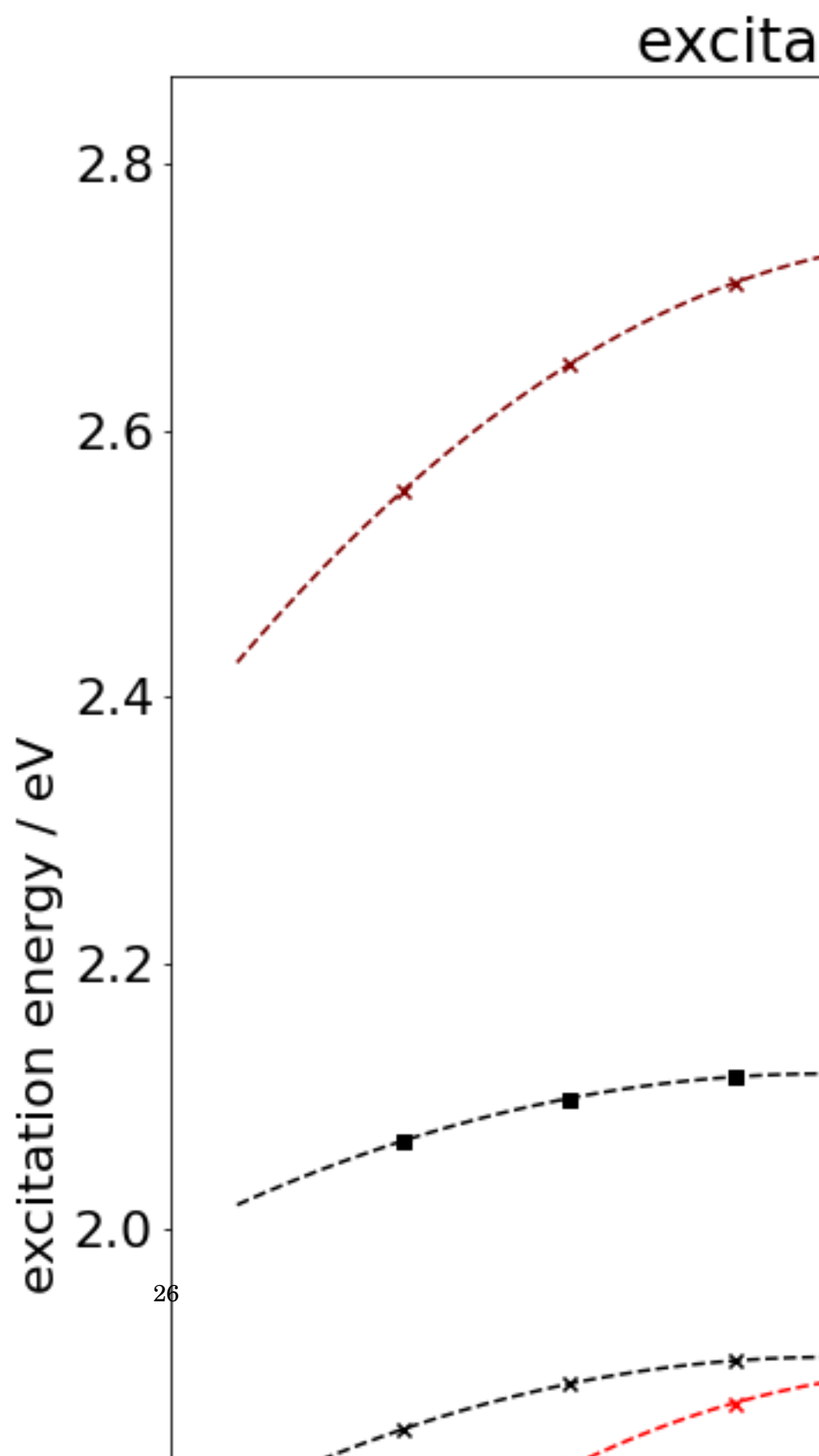


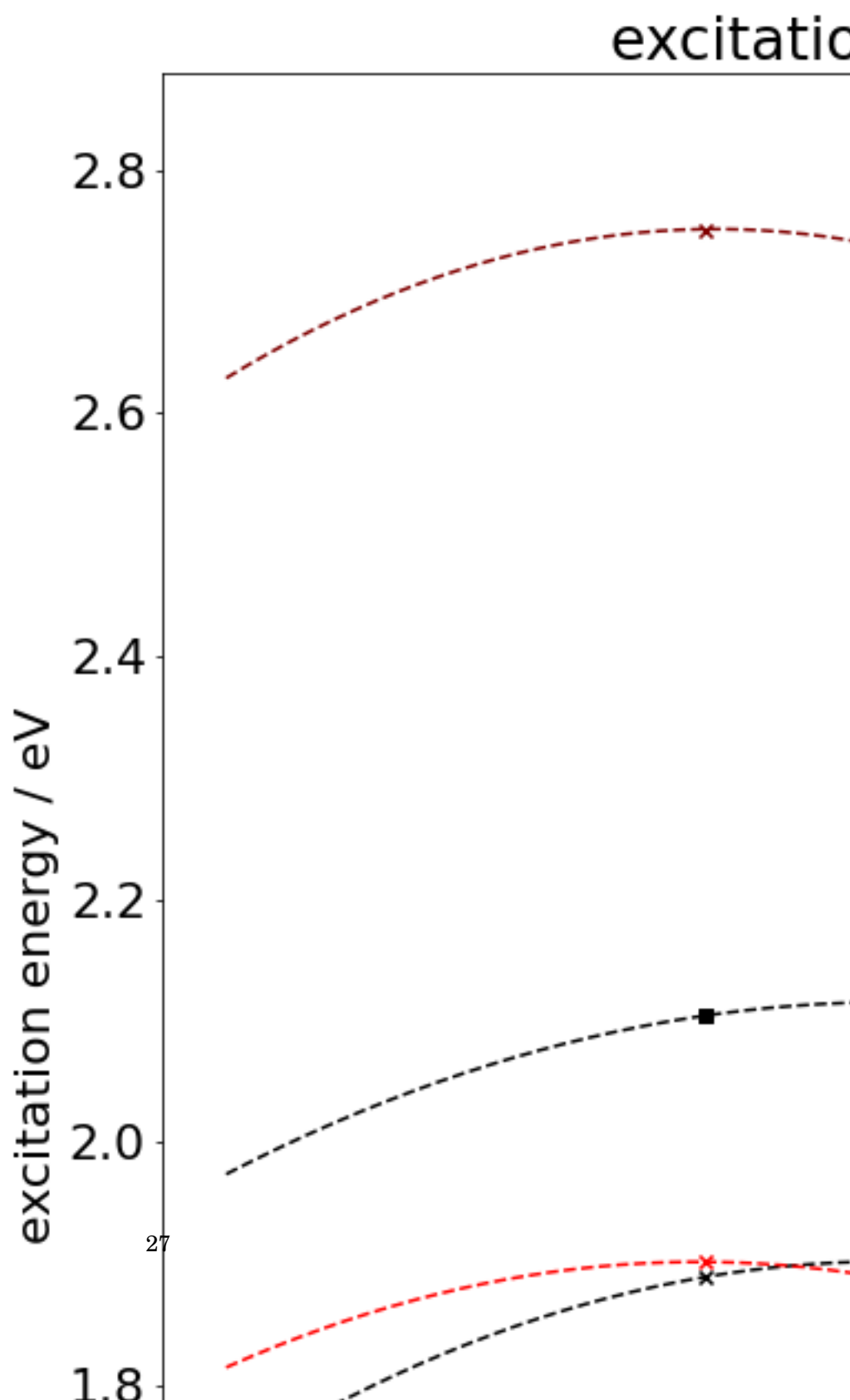


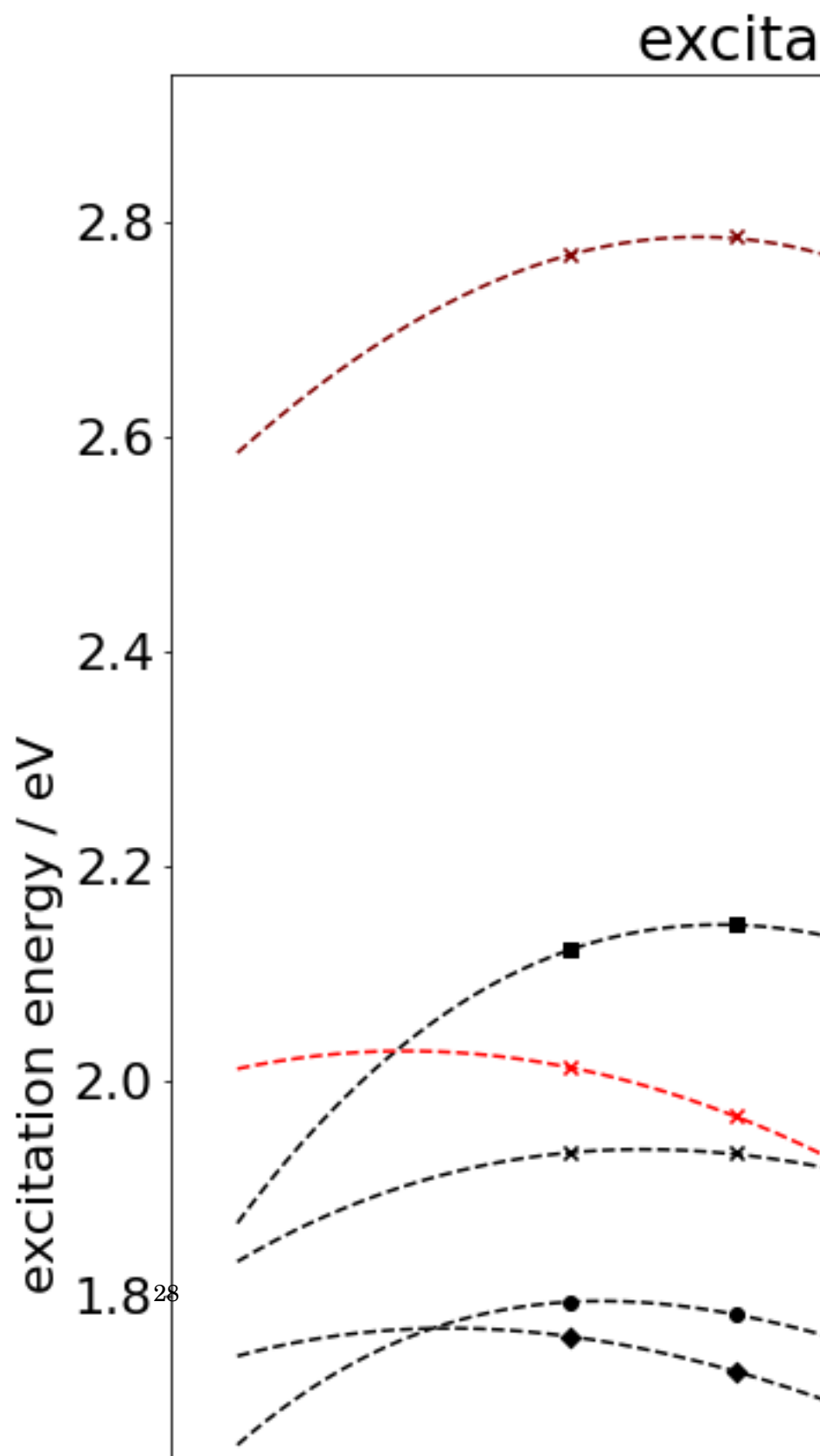












EXCITON METHOD

P^{reamble}

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

ATOMISTIC MODELLING OF LIGHT HARVESTING COMPLEXES

P^{reamble} **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 Chlorophyll distances

5.3 Sites, states and couplings

5.3.1 Sites

5.3.2 Exciton states

5.3.3 Coupling

5.4 Excitation Energies

P^{reamble}

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 Excitation 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*, Zeitschrift für 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. ROOHTAAN, *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.

