Accounting for Strong Electronic Correlation in Metalloproteins

Edward Baxter Linscott

Corpus Christi College, University of Cambridge

This thesis is submitted for the degree of Doctor of Philosophy

August 2019

Supervised by Dr. D. J. Cole and Prof. M. C. Payne

Department of Physics

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Abstract

Accounting for Strong Electronic Correlation in Metalloproteins

Edward Linscott

Metalloproteins play a crucial role in many key biological processes, from oxygen transport to photosynthesis. In the case of photosynthesis, the oxygen evolving complex (OEC) — a $CaMn_4O_5$ cluster — catalyses water-to-oxygen-gas conversion.

From a computational standpoint, accurately modelling the electronic structure of the OEC and other metalloproteins *ab initio* is difficult, due to two challenges. Firstly, there is that of the strong electronic correlation present due to the partially-filled $3d$-subshells of the transition metal atoms, a classic example of where semi-local density functional theory (DFT) — a go-to method for computational physicists — fails. The second challenge is that of size: as this thesis will demonstrate, we must consider large cluster models that are thousands of atoms in size, which takes us beyond the reach of both plane-wave DFT and quantum chemistry methods.

This thesis explores the capacity of density functional theory-plus-U (DFT + U) and dynamical mean field theory (DMFT) to meet both of these challenges. It will demonstrate how both DFT + U and DMFT can be readily married with linear-scaling DFT, meaning that these theories can be applied to protein systems containing thousands of atoms. In particular, this thesis presents the unification of ONETEP (a linear-scaling DFT code) and TOSCAM (a DMFT solver). It also presents a novel approach for determining Hubbard and Hund's parameters via linear response that is compatible with linear-scaling DFT and resolves inconsistencies between the linear response method and the DFT + U corrective functional.

These techniques are then applied to haem, haemocyanin, and the OEC, providing insight into the role of strong correlation in their electronic structure and function. In so doing, this thesis demonstrates how one can perform large-scale simulations of metalloproteins that account for strong electronic correlation. The results of this thesis are of significant interest due to both the importance of metalloproteins in nature, and the wealth of potential applications that would spring from a thorough understanding of their catalytic and binding properties.

Acknowledgements

First and foremost, I would like to thank Dr. Daniel Cole. He has been a very communicative and helpful supervisor, and I am grateful for his guidance and his (albeit metaphorical) open door. Thanks also to my other supervisor, Prof. Mike Payne, for always being in my corner, and for the financial support during my fourth year.

In addition to Danny and Mike, my thesis has leaned heavily on the expertise of Dr. David O'Regan and Dr. Cédric Weber. David provided guidance on DFT + U, and I'm incredibly grateful for his enthusiasm and his constant appetite to "talk physics". (I regret that I'm not now headed for Dublin.) What David was for DFT + U, Cédric was for DMFT, to whom I'm indebted for his time, support, and insight. These two have been like additional supervisors, and my thesis is much richer for their input.

Thanks to my other collaborators and co-authors, Mohamed Ali, Antoine Georges and Nick Hine. Thanks also to the wider ONETEP developers group for many helpful discussions, and Dr. Michael Rutter for his technical expertise.

I am incredibly privileged to study at Cambridge, and I am grateful to the Royal Society of New Zealand and the Cambridge Commonwealth European and International Trust for funding my studies.

Thanks to my fellow TCMers: Alice, Ben, James, Jamie, Kevin, Matt, Mark, and Will. Without your distraction I might have finished this a little sooner, but I certainly would not of had half the fun doing so.

To my Corpus family — Em (special thanks for the proof-reading), Pat, Kum, Michael, other Michael, Jenny, Josca, Seba, Eleanor, Sarah, and many others — you guys are amazing. The last five years have been a blast thanks to you, and you're going to make leaving Cambridge incredibly hard.

Thanks to Stephen, Esther, and my other NZ friends who have supported me from afar. Hopefully at some stage in the future I will be able to see more of you!

Most of all, enormous thanks to my parents for their constant love and support, and my brothers for all of their very helpful research suggestions.

Great are the works of the LORD, studied by all who delight in them.

Contents

List of Figures

List of Tables

List of Abbreviations

ωB97XD	a range-separated hybrid exchange-correlation functional with dispersion corrections, due to Head-Gordon and co-workers[1]
ωPBEh	hybrid exchange-correlation functional using Henderson *et al.*'s exchange hole[2]
3-21g	split-valence basis set (with three Gaussians representing each core atomic orbital, and two basis functions for each valence orbital: one a linear combination of two Gaussians, the other, one)
6-31g	split-valence basis set (with six Gaussians representing each core atomic orbital, and two basis functions for each valence orbital: one a linear combination of three Gaussians, the other, one)
AE	All-electron (*cf.* pseudopotential approaches)
AIM	Anderson impurity model
aug-cc-pVDZ	correlation-consistent double-zeta valence basis set augmented with diffuse functions
B3LYP	Becke, three-parameter, Lee-Yang-Parr (exchange-correlation functional)
B88	Becke 88 exchange functional
BFGS	Broyden-Fletcher-Goldfarb-Shanno (algorithm)
BLYP	Becke-Lee-Yang-Parr (exchange-correlation functional)
BS	bound singlet
BSSE	basis set superposition error
CASPT2	complete active space perturbation theory
CASSCF	complete active space self-consistent field
CASTEP	Cambridge Serial Total Energy Package
CC	coupled cluster
CG	conjugate gradient
CPT	cluster perturbation theory
cRPA	constrained random phase approximation
cryo-EM	cryogenic electron microscopy
CTQMC	continuous-time quantum Monte Carlo
deoxyHc	deoxygenated haemocyanin
DFT	density functional theory
DFT + U	density functional theory-plus-U

DFT + U + J	density functional theory-plus-U with additional with Hund's coupling terms
DFT + U + V	density functional theory-plus-U with additional inter-atom terms
DMFT	dynamical mean field theory
DMRG	density matrix renormalization group
DMRG-CASPT2	CASPT2 with a DMRG solver
DMRG-CT	DMRG with canonical transformation theory
DNA	deoxyribonucleic acid
DOS	density of states
ED	exact diagonalisation
EDFT	ensemble density functional theory
ENDOR	electron nuclear double resonance
EPR	electron paramagnetic resonance
EXAFS	extended x-ray absorption fine structure
exp	experiment
FCIQMC	full configuration interaction quantum Monte Carlo
FePIm	iron-porphyrin with an axial imidazole ligand
FePImCO	iron-porphyrin with axial imidazole and carbon monoxide
FFT	fast Fourier transform
FTIR	Fourier-transform infrared (spectroscopy)
GGA	generalised gradient approximation
GGA + U	DFT + U in the specific case where the underlying xc functional used is a GGA
GPU	graphics processing unit
HF	Hartree-Fock
HIV	human immunodeficiency virus
HL	Heitler-London
HOMO	highest occupied molecular orbital
HOS	high oxidation state
HPC	high performance computing
KS	Kohn-Sham
LDA	local density approximation (exchange-correlation functional)
LDOS	local density of states
LOS	low oxidation state
LSDA	local spin density approximation
LUMO	lowest unoccupied molecular orbital
LYP	Lee-Yang-Parr (correlation functional)
MINI	minimal Gaussian basis set as presented in Ref. 3
MINI+D	MINI with dispersion corrections
MPI	message passing interface
MRCI	multi-reference configuration interaction
NAO	natural atomic orbital
NBO	natural bonding orbital

NGWF	non-orthogonal generalised Wannier function
NMR	nuclear magnetic resonance
OEC	oxygen evolving complex
ONETEP	Order-N Electronic Total Energy Package
OpenMP	open multi-processing
oxyHc	oxygenated haemocyanin
PAO	pseudo-atomic orbital
PAW	projector-augmented wave
PBE	Perdew-Burke-Ernzerhof (exchange-correlation functional)
PDB	protein data bank (record)
PS	pseudoatom
PSII	photosystem II
PSP	pseudopotential
QM	quantum mechanical
QM/MM	quantum mechanics/molecular mechanics
QMC	quantum Monte Carlo
R-QM/MM	the EXAFS-revised quantum mechanics/molecular mechanics approach of Ref. 4
RAM	random access memory
RASPT2	multiconfigurational second-order perturbation theory restricted active space
RHF	restricted Hartree-Fock
RIXS	resonant inelastic x-ray scattering
RMSD	root-mean-square deviation
RPA	random phase approximation
rr-DMFT	real-space renormalised dynamical mean field theory
SCE	static correlation error
SCF	self-consistent field
SIC	self-interaction correction
SIE	self-interaction error
SQUID	superconducting quantum interference device (spectroscopy)
STO-3G	minimal Slater-type orbital basis set (with each orbital represented by three primitive Gaussian orbitals)
TDDFT	time-dependent density functional theory
TOSCAM	Toolbox for Strongly-Correlated Approaches to Molecules
UV	ultra-violet
vdW	van der Waals (dispersion force)
VWN	Vosko-Wilk-Nusair (local-density approximation to the correlation functional)
XANES	x-ray near edge structure
XAS	x-ray absorption spectroscopy
xc	exchange-correlation
XFEL	x-ray free-electron laser

XRD x-ray diffraction

Preface

Part of Chapter 3 has been published as
E. B. Linscott, D. J. Cole, M. C. Payne, and D. D. O'Regan,
Role of spin in the calculation of Hubbard U and Hund's J parameters from first principles
Phys. Rev. B, **98**, 235157 (2018).

Part of Chapter 4 will be published as
E. B. Linscott, D. J. Cole, N. D. M. Hine, M. C. Payne, and C. Weber,
Introducing TOSCAM: uniting dynamical mean field theory and linear-scaling density functional theory
(in preparation)

Part of Chapter 5 is under review for publication. A preprint can be found as
M. A. Al-Badri, E. B. Linscott, A. Georges, D. J. Cole, and C. Weber,
Super-exchange mechanism and quantum many body excitations in the archetypal hemocyanin/tyrosinase dicopper oxo-bridge
`arXiv:1881.05739` (2018).

That chapter is the outcome of work done in collaboration with my co-authors. Specifically, both the DFT and the DMFT calculations that underpin all of the results of the chapter were run by MB and myself, in consultation with CW. The results of individual sections are generally the product of these calculations or further post-processing:

Section 5.2: I ran the scaling calculations and analysed the results.

Subsection 5.2.1: CW wrote the code to obtain the local axes; MB and CW used this code to determine the optimal local rotation; I compared the results against the NBO analysis.

Subsection 5.3.1: I wrote the code to analyse the projected density matrix; MB and I analysed the results.

Subsection 5.3.2: I wrote the code to extract these quantities from the core calculations; we collectively analysed the results.

Subsection 5.3.3: These results came directly from the core DMFT calculations run by MB and myself.

Subsection 5.3.4: CW extracted these quantities at the suggestion of AG; we collectively analysed the results.

Section 5.4: I performed the NBO analysis and interpreted it with the assistance of DC.

Section 5.5: MB extracted the optics; we collaboratively analysed the results.

Section 5.6: I performed all of the calculations and analysis, with the exception of the TDDFT calculations, which were performed by DC.

The proof in Appendix A.4 was derived collaboratively with Dr. Ben Irwin.

Chapter 1

Introduction

The process of photosynthesis is integral to life on earth as we know it — almost every ecological community is ultimately dependent on a phototroph. Among all of the photosynthetic organisms on earth, oxygenic phototrophs (such as cyanobacteria, algae, and higher plants) are by far the most prolific.

A key step in oxygenic photosynthesis is the oxidation of water:

$$2\mathrm{H_2O} + 4h\nu \rightarrow 4\mathrm{e^-} + \mathrm{O_2} + 4\mathrm{H^+}.$$

This reaction is carried out by the oxygen evolving complex (OEC), a $\mathrm{CaMn_4O_5}$ cluster contained within the photosystem II (PSII) protein complex. The ability of the OEC to oxidise water in very mild physiological conditions is impressive, and remarkably all contemporary phototrophs have an identical reaction centre despite their sheer number and ecological diversity. This makes understanding the structure and function of the OEC (as nature's unique solution to oxygenic light-harvesting) an alluring research problem, and it has been the subject of substantial research already.[7–11] Ultimately, an understanding of the O–O formation mechanism would provide a blueprint for the design of biomimetic catalysts — a field that is garnering an incredible amount of interest.[12–15] However, in spite of all of this attention, the precise structure and mechanism of the OEC is not yet well understood. The next section will elaborate on our present understanding, but for the moment it is worth highlighting that it was only in 2011 that the crystal structure of PSII was determined at atomic resolution[16] — before then, the structure of the OEC was the realm of (educated) guesswork, with some hypothesised models bearing remarkable resemblance to the actual structure.[17]

In parallel with these developments, there have been remarkable advances in the field of computational materials science. When attempting to calculate the electronic structure of atomic systems, the last few decades have seen density functional theory (DFT) establish itself as the go-to method.[18–21] Facilitated by exponentially increasing computing power, modern DFT codes are capable of routinely calculating the electronic structure of systems with hundreds of atoms, opening the door to quantum-mechanical modelling of a vast landscape of systems of considerable scientific interest (such as catalysts, photovoltaics and materials for energy storage). The range of computationally accessible systems has broadened even further with the advent of

linear-scaling DFT codes (that is, codes whose computational cost scales linearly with the number of atoms in the system, rather than the cubic scaling of traditional methods). ONETEP[22] is one such code, notable for its equivalence to plane-wave approaches and its minimal basis size due to the *in situ* optimisation of its basis (a set of local Wannier-like orbitals). Its ability to routinely perform DFT calculations on systems containing thousands of atoms allows more detailed study of nanostructures,[23,24] defects,[25,26] and biological systems.[27–30] Furthermore, extensions to DFT — namely density functional theory-plus-U (DFT + U) and dynamical mean field theory (DMFT) — have facilitated increasingly accurate treatment of strong electronic correlation.

The objective of this thesis, therefore, is to bring these two threads together, developing linear-scaling DFT (and its extensions) to a point where it is capable of performing calculations on metalloproteins such as the OEC that are both sufficiently accurate and computationally tractable. These calculations will allow us to predict important properties of scientific interest, from binding energies to optical spectra.

1.1 The structure and function of the OEC

I will begin by outlining our current understanding of the OEC. Based on the results of x-ray diffraction measurements,[16,31–34] it is known that the OEC contains an inorganic $CaMn_4O_5$ core that catalyses the oxidation process (Figure 1.1). This core lies within PSII, a large cluster of around twenty protein subunits, containing some 120,000 atoms in all (Figure 1.2).

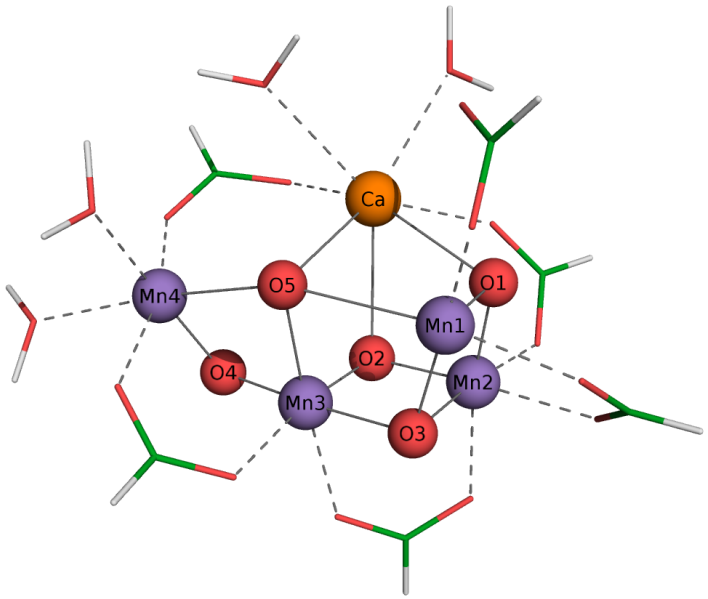


Figure 1.1: The distorted chair structure (solid lines) of the inorganic core of the OEC as found by Suga *et al.* in their seminal XFEL crystal structure.[34] Only the carbonyl groups of the coordinating ligands are shown. A histidine (which coordinates with Mn1) is not shown. Manganese, calcium, oxygen, carbon, and hydrogen atoms are shown in purple, orange, red, green, and white respectively.

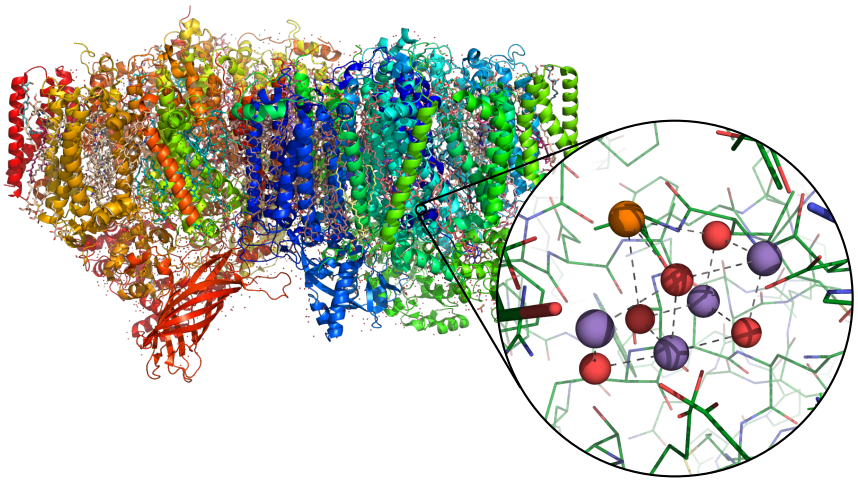


Figure 1.2: The 1.9 Å resolution crystal structure of the entire PSII obtained by x-ray diffraction measurements.[16] The inset reveals the location of an OEC core (there are two such cores within PSII).

In order to oxidise water, the OEC cycles through five states (S_0 to S_4) in what is known as the *Kok cycle* (Figure 1.3).[35,36] The system proceeds to the next intermediate state upon absorption of a photon (at a neighbouring chlorophyll pigment), which prompts the OEC to release an electron and a proton (with the exception of the S_1 to S_2 transition, where no protons are released) and oxidise one of the four manganese atoms. Throughout this process, two water ligands progressively have their hydrogens removed, and bind to one another to form molecular oxygen.

Figure removed due to copyright. The original can be seen in Ref. 37 as Figure 1.

Figure 1.3: The Kok cycle, with proposed (but disputed) structures shown for S_0 to S_3.[36] Manganese, calcium, oxygen, and hydrogen atoms are in purple, yellow, red, and white respectively. The oxidation states for each of the manganese atoms (also disputed) are labelled. Figure taken from Ref. 37.

The atomic structure of the OEC was unknown until 2011, when Umena *et al.* published their 1.9 Å-resolution x-ray diffraction (XRD) structure of the dark-stable S_1 state.[16] However, it came under criticism, as it was suspected that the imaged structure had substantial radiation

damage: both metal-metal and metal-oxygen distances in the reported XRD structure were surprisingly long compared to extended x-ray absorption fine structure (EXAFS) predictions, synthetic complexes, and computational models.[38,39] Because EXAFS is not prone to damaging the OEC via radiation, many authors concluded that radiation damage during the imaging process likely occurred in the case of the XRD structure.[40–42] Other authors disagree, questioning the reliability of EXAFS as a predictive tool.[43]

Then in 2015, Suga *et al.* published a "radiation-damage-free" structure.[34] This work employed a novel imaging technique using an x-ray free-electron laser (XFEL).[†] By using ultrashort, high brilliance x-ray pulses, the structure is imaged (supposedly) before the onset of radiation damage. Since that seminal work, this technique has also enabled measurements of S-states other than the dark-stable S_1 state (which are not amenable to conventional experimental structural determination): in 2017 Suga *et al.* published XFEL data on the S_3 state, and in late 2018 Kern *et al.* published XFEL structures for all the S-states.[46,47]

1.2 Open questions

Building on this increasing body of knowledge, the computational community has proposed a multitude of reaction pathways for O_2 formation. Disagreements between different proposals usually arise around the following points of contention.

1.2.1 Why do XRD, EXAFS, and computational structures disagree?

As discussed already, the 2015 XFEL structure supposedly avoided the radiation damage that dogged its 2011 predecessor. However, the 2015 XFEL structure is also inconsistent with EXAFS spectra,[48] and it is a matter of debate whether or not these structures are genuinely free of radiation damage.[49] The fact that XRD, EXAFS, and computational models remain inconsistent is the source of much consternation.

One particular point of disagreement is the position of O5, the μ_4-oxo-bridge located in the quasi-centre of Mn1 and Mn4 (refer back to Figure 1.1). Specifically, the 2011 and 2015 structures reported unusually equal Mn1–O5 and Mn4–O5 distances, with O5 appearing "stranded" between the two manganese atoms. In comparison, DFT and other methods tend to place O5 as being bound to Mn4, as do recent XFEL structures published by Kern *et al.*.[47] This is not mere nitpicking: given that this oxygen atom is almost certainly one of the two substrate oxygen atoms that ultimately forms O_2,[46] the precise position and chemical bonding of this atom will have far-reaching implications for the entire reaction mechanism. Increasingly accurate simulations and experiments resolving the nature of the bonding of O5 are highly desirable, especially given the potential inaccuracy of existing DFT calculations (which at most use hybrid functionals and/or only consider small cluster models).

[†]Two groups had previously reported XFEL structures for PSII, but these did not achieve atomic resolution.[44,45]

1.2.2 What are the oxidation states of the manganese atoms?

In addition to the protonation state, crystallography cannot unambiguously inform experimentalists as to the oxidation states of the metal ions of the OEC core. Other experimental methods provide some insights: based on multiline electron paramagnetic resonance (EPR) signals of the S_0 and S_2 states, the four Mn ions of the OEC are known to have total spin of $\frac{1}{2}$.[11] Additionally, EPR, electron nuclear double resonance (ENDOR), and x-ray absorption spectroscopy (XAS) analyses agree that S_2 contains a combination of Mn^{3+} and Mn^{4+}. It follows that S_0 must be either $Mn^{2+}(Mn^{3+})_3$, $Mn^{2+}Mn^{3+}(Mn^{4+})_2$, or $(Mn^{3+})_3Mn^{4+}$.[9] This has implications for all the other S-states, as upon each step of the Kok cycle one manganese atom is oxidised. The so-called low oxidation state (LOS) paradigm assumes a S_0 of $Mn^{2+}(Mn^{3+})_3$, and consequently the mean Mn oxidation level is 3.0 in S_1, 3.25 in S_2, and so on. The alternative is the high oxidation state (HOS) paradigm, where the mean Mn oxidation level is 3.5 in S_1, 3.75 in S_2, and so on.

The HOS paradigm is preferred by many authors, largely due to comparisons of the Mn K-edge energy of the OEC with a set of Mn oxy species as measured by XAS.[50] This stance is reaffirmed by Mn Kβ spectroscopy, resonant inelastic x-ray scattering (RIXS) K pre-edge results, and most recently EPR.[51†] Other EPR, ENDOR, and EXAFS data appears to support the LOS paradigm.[53,54] Time-dependent density functional theory (TDDFT) calculations of x-ray near edge structure (XANES) edge positions have also favoured the LOS paradigm,[55] although more recent results have shed doubt on these conclusions.[37] Increasingly accurate calculations could assist the oxidation state assignment.

1.2.3 What is the hydrogenation pattern of the OEC?

Protons are practically invisible in x-ray crystallography and do not appear in crystal structures — and yet, the protonation of a system can have dramatic effects on its electronic structure and the potential mechanisms of O_2 formation. It has been the role of computational studies to deduce the protonation of the OEC in each S-state.[‡,§] On this front, there is little consensus, with competing reaction mechanisms often proposing different protonation configurations.[37,41] Competing protonation schemes are usually delineated on a basis of energetics and the resulting metal-metal and metal-oxygen bond distances. Naturally, the hydrogenation pattern is strongly influenced by the oxidation state paradigm,[56] and is usually investigated using small cluster models for the OEC.

1.2.4 What role do interconvertible states play?

In addition to the multi-line EPR signal indicative of a spin $S = \frac{1}{2}$ ground state of S_2, higher spin signals corresponding to spin $S \geq \frac{5}{2}$ are observed under a variety of preparation conditions. It has been demonstrated that the existence of two interconvertible forms of S_2 rationalise these observations,[57] a proposal that has support from computational investigations.[58–61] The

[†]For a detailed discussion of these results, see Ref. 52.

[‡]Recall that the system loses a hydrogen atom upon each step of the Kok cycle.

[§]Neutron diffraction experiments could also provide insight.

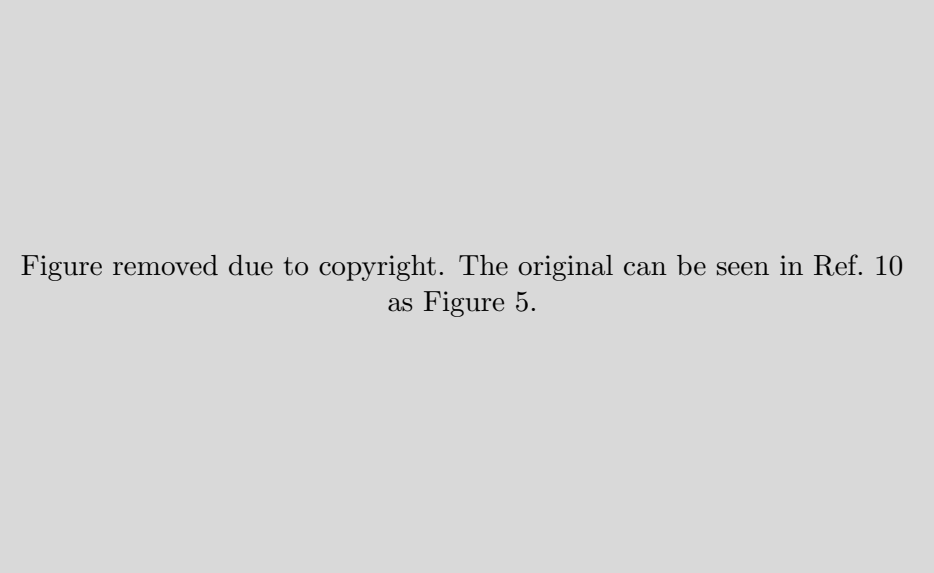


Figure 1.4: Possible pathways for oxygen insertion: (a) W3, via Mn1[62,63,67] (b) W3 (relying on interconvertible S_2 states to move O5 close to Mn1),[57] and (c) the "carousel" or "pivot" movement of water about Mn4.[48,61,68] Figure adapted from Ref. 10.

existence of these interconvertible states has ramifications for potential $S_2 \rightarrow S_3$ mechanisms,[62,63] with some proposed reaction mechanisms involving passing between these two states (such as that depicted in Figure 1.3). The existence and importance of these states is contested; for instance, recent XFEL data failed to detect S_2 interconvertible states.[47]

It is also debated whether or not S_1 has multiple interconvertible structures, or is simply a single rigid structure.[58,64–66] Likewise, interconversion in the S_3 and S_4 states has been proposed.[7]

1.2.5 Which are the substrate water molecules? The S_2 to S_3 transition

There is substantial evidence that the oxo-bridge (known as O5) is one of the two substrate waters.[46,47] The identity of the second is much less clear. Crucially, when the system transitions to the S_3 state, a second oxygen atom ("Ox") appears between Mn1 and Mn4.[46,47] This may be the second substrate oxygen, or alternatively the oxygen that replaces O5 after O_2 release. This leaves several candidates for the second substrate, including W2 (a water molecule bound to Mn4), W3 (bound to Ca), or Ox itself, as illustrated in Figure 1.4.

Increasingly accurate assessments of energetics and reaction barriers will narrow down possible substrate candidates, as will Raman and Fourier-transform infrared (FTIR) spectroscopy, which can help clarify the oxidation state of the manganese atoms and the protonation states of the oxygen atoms.[69]

1.2.6 What is the O_2 bond formation mechanism?

There are two competing proposals for O_2 bond formation. The first (Figure 1.5a) is an oxo-oxyl radical attack mechanism — that is, an unpaired electron on Ox makes it highly reactive, and it forms an O–O bond with O5. This proposal has been championed by Siegbahn and

Figure removed due to copyright. The original can be seen in Ref. 10 as Figure 6.

Figure 1.5: The two competing O_2 bond formation mechanisms: (a) a Mn^{4+}-$O^{\bullet}$ species in the S_4 state could lead to an oxo-oxyl radical attack mechanism, or (b) a Mn^{5+}=O species could lead to a water-nucleophilic attack mechanism (although not necessarily involving the two oxygen atoms pictured here). Figure taken from Ref. 10.

co-workers,[59,63,70,71] among others.[72,73] The alternative (Figure 1.5b) is a nucleophilic attack mechanism, whereby a highly electrophilic Mn^{5+}=O species is attacked by a nucleophilic water (or OH^- molecule). This second mechanism is analogous to many well-characterised inorganic water oxidation catalysts, and is favoured by Cox, Brudvig and others.[7,10,74]

Note that the difference between these two mechanisms lies in the precise location of a hole across the Mn-O bond, so an accurate computational assessment of the spin density of the S_4 state would help to clarify our understanding of this mechanism.

1.3 Thesis outline

Evidently, several of these open questions would benefit from increasingly accurate methods for simulating metalloproteins, where we must be able to simulate large systems while also capturing the complex electronic structure of transition metals. With this in mind, the structure of the rest of this thesis is as follows.

I begin in Chapter 2 by introducing DFT, focusing on the challenges that metalloproteins present to conventional DFT codes. To this end, I explore how DFT — as implemented in ONETEP — achieves linear scaling, and I discuss the plethora of various functionalities that we rely on when simulating biological systems.

The next two chapters explore how to accurately treat the strong electronic correlation present in metalloproteins. Chapter 3 introduces density functional theory-plus-U (DFT + U), a widely used method for electronic structure prediction that provides a pragmatic approach to correcting delocalisation error — a problem which lies at the heart of DFT's underestimation of the band gap. The reliability of any DFT + U calculation hinges on the choice of the parameter U, and in the past decade linear response calculations have become an incredibly popular approach for obtaining these parameters *ab initio*. This chapter also presents a detailed study of the role of spin in linear response calculations. It demonstrates that conventional scalar linear response, which does not distinguish between spin channels, neglects screening by the electrons in the same Hubbard site but of the opposite spin. I present alternative definitions for Hubbard and Hund's parameters that account for this screening, bringing them into line with the contemporary DFT + U functional. These are tested on manganese oxide and hexahydrated transition metals. By establishing a systematic approach for including/excluding screening by the oppo-

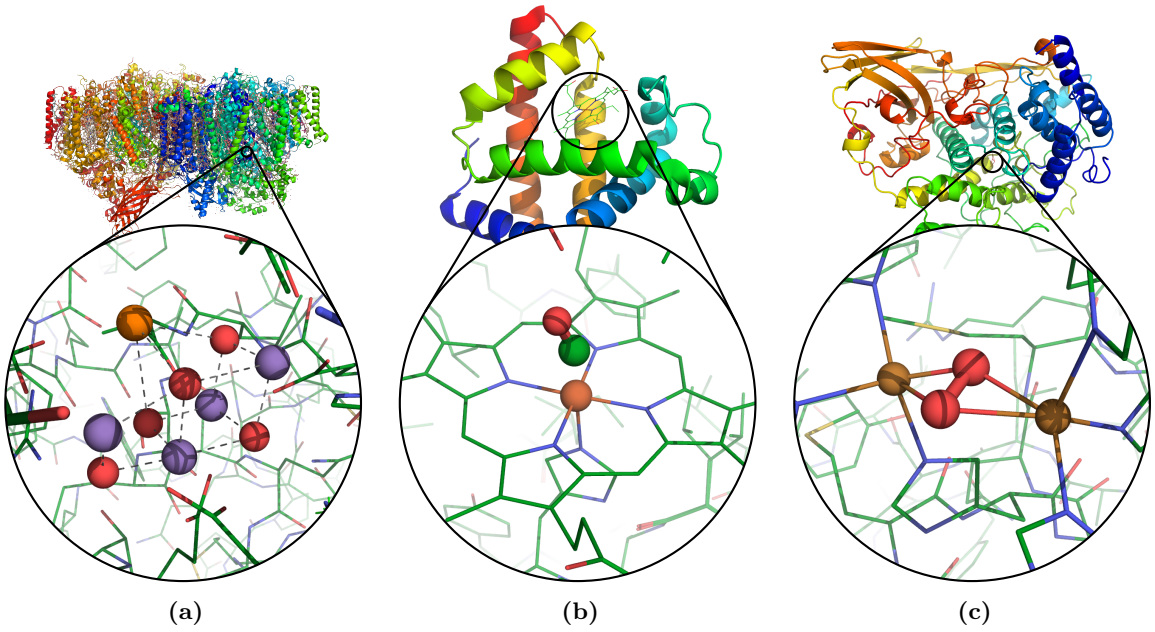


Figure 1.6: Three metalloproteins and their transition metal cores: (a) the OEC core of PSII (PDB record 4UB6),[34] introduced already, (b) the iron porphyrin in carboxymyoglobin (PDB record 1MYZ),[80] and (c) the dicopper core of oxyhaemocyanin (PDB record 1OXY).[81] Green, blue, red, yellow, and purple atoms are carbon, nitrogen, oxygen, sulfur, and manganese; the orange atoms are calcium, iron, and copper in (a-c) respectively.

site spin channel, these developments provide a route forward for performing DFT + $U(+J)$ on spin-polarised systems in a robust and consistent manner. The work of this chapter expands upon the minimum-tracking linear response method, a reformulation of linear response which is applicable in DFT codes that operate via direct minimisation (as opposed to iterating the Kohn-Sham potential and density to self-consistency). Such codes cannot otherwise perform the particular calculations that linear response usually requires. These advances make first-principles DFT + U newly compatible with calculations on large and spatially complex systems such as metalloproteins.

Chapter 4 then introduces an alternative method for treating the transition metal centres: dynamical mean field theory (DMFT). DMFT is a much more sophisticated method that includes quantum dynamical effects and takes into account valence fluctuations, spin fluctuations, and thermal excitations. Although DMFT is routinely used to describe bulk materials, it was recently extended to molecular systems.[75,76] In particular, Refs. 77–79 deployed an implementation of an interface between ONETEP and TOSCAM, a DMFT toolbox. However, this implementation was never distributed – nor, crucially, was it ever properly incorporated into ONETEP. This chapter reports an overhauled implementation of ONETEP+TOSCAM, which (a) restores compatibility with the active version of ONETEP, (b) has no external dependencies that would prevent its distribution, and (c) is straightforward to compile.

At this stage in the thesis, these theoretical and computational developments start to be applied to several metalloproteins, shown in Figure 1.6. The latter half of Chapter 4 demonstrates the capabilities of ONETEP+TOSCAM on an iron porphyrin system (Figure 1.6b). This study provides a logical first stepping-stone for the methodology.

Chapter 5 applies these developments to haemocyanin, a dicopper cluster found in molluscs and arthropods (Figure 1.6c), systematically investigating how the Hubbard Coulomb potential U alters the electronic structure at the Cu_2O_2 site. I present an analysis of the quantum-entangled low energy states and the dominant contributions to the charge and magnetic properties of Cu_2O_2. This enables the identification of a regime of parameters where the singlet is stabilised for the structure *in situ*, in line with experiment. The resulting singlet is in the Heitler-London regime (an entangled quantum superposition of two localised magnetic moments), and is associated with incoherent scattering processes that reduce the lifetime of charge excitations.

Chapter 6 then turns to the inspiration for this thesis: the OEC. Making use of the linear-scaling capacity of ONETEP, I present the geometry optimisation of a 1631-atom cluster model of the OEC (at the level of DFT). Precursory DMFT calculations on a toy 75-atom model of the OEC are then presented. These calculations provide an opportunity to demonstrate several obstacles that arise; I outline in this chapter how they may be overcome.

Finally, Chapter 7 summarises the results of this thesis and discusses several promising topics for future research.

There are also several appendices that provide background and contain several detailed proofs.

Chapter 2

Linear-scaling density functional theory

2.1 The many-body problem and density functional theory

Quantum-mechanical simulations have played a key role in establishing our present understanding of the structure and function of the OEC. A system of many nuclei and electrons such as the OEC is fundamentally described at the quantum-mechanical level by a many-body wavefunction $|\Psi\rangle$ and its Hamiltonian, which is (in atomic units[†])

$$\hat{H} = -\frac{1}{2}\sum_i \nabla_i^2 + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,I}\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} - \sum_I \frac{1}{2M_I}\nabla_I^2 + \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}, \quad (2.1)$$

with the terms being, respectively, the electronic kinetic energy, the electron-electron repulsion, the electron-nuclear attraction, the nuclear kinetic energy, and the nuclear-nuclear repulsion. (Electrons are denoted by lower case indices and nuclei by upper case, with charge Z_I and mass M_I.) The total energy of such a system is given by

$$E = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle}. \quad (2.2)$$

The ground state of the system minimises this energy, and it is this ground state that one usually wants to obtain. However, in full this equation is intractable, so it is necessary to make some simplifying assumptions. It is common to assume that compared to the timescale of nuclear motion, electrons will relax rapidly to their ground-state configuration. This is the *Born-Oppenheimer approximation*. For the purposes of this thesis we will further limit ourselves to electrons in a static array of nuclei, in which case the third term of Equation 2.1 takes the form of an external potential $\sum_i V_{\text{ext}}(\mathbf{r}_i)$, the fourth term can be ignored, and the final term becomes a constant that can be ignored for the purposes of energy minimisation.

However, this system remains vastly complex: the wavefunction $\Psi(\mathbf{r}_1, ..., \mathbf{r}_N)$ is $3N$-dimensional,

[†]In atomic units the electron rest mass m_e, the elementary charge e, and the reduced Planck's constant $\hbar$ are all unity; the resulting unit for length is the Bohr radius $a_0 \approx 5.292 \times 10^{-11}$ m; for energy, the Hartree $E_h \approx 4.359 \times 10^{-18}$ J.

and hence the size of the wavefunction's configuration space scales exponentially with respect to particle number — which is immensely prohibitive. This obstacle led Walter Kohn to — somewhat provocatively — state during his Nobel Lecture that for systems on the order of 100 or more atoms the many-electron wavefunction is "not a legitimate scientific concept", as it cannot be calculated or recorded with sufficient accuracy.[82] Or, in the words of Paul Dirac,[83]

> "The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation."

2.1.1 The Hohenberg-Kohn theorems

Density functional theory (DFT) is one such approximate practical method. Instead of dealing with the $3N$-dimensional many-body wavefunction, DFT recasts the electronic problem in terms of the *three*-dimensional electronic density $n(\mathbf{r})$. This simplification is made possible by means of the Hohenberg-Kohn theorems,[18] which state that there exists a functional of the density such that

$$E[n(r)] \geq E_{GS}, \tag{2.3a}$$

$$E[n_{GS}(r)] = E_{GS}. \tag{2.3b}$$

That is, there exists a functional that always gives values greater than or equal to the ground-state energy E_{GS}, and gives the ground-state energy for the ground-state density.

2.1.2 The Kohn-Sham construction

However, this is only a statement of existence: we do not know the precise form of this functional, as the proofs of the Hohenberg-Kohn theorems only talk about it in very abstract terms.[†] On the basis of what we do know about the underlying Hamiltonian, we can at least decompose the energy functional into several terms

$$E[n] = T_0[n] + \int \mathrm{d}\mathbf{r}\, n(\mathbf{r})[V_{\text{ext}}(\mathbf{r}) + \tfrac{1}{2}\Phi(\mathbf{r})] + E_{xc}[n] \tag{2.4}$$

where T_0 is the kinetic energy, $\Phi(\mathbf{r}) = \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$ is the Hartree term, and $E_{xc}[n]$ is the exchange-correlation (xc) functional, which contains all the terms whose form is unknown and must be approximated (specifically, non-classical electron-electron interactions and the difference in the kinetic energy between the interacting and non-interacting systems.)

[†]In proving Equations 2.3a and 2.3b via the Levy construction, one defines a universal functional $F[n] = \min_{\psi\to n}\langle\psi|\hat{T} + \hat{V}_{ee}|\psi\rangle$ – that is, the lowest expectation value of the kinetic and electron-electron interaction operators of any possible wavefunction that gives rise to the desired density n.[84] ("Universal" because all systems have the same $\hat{T}$ and $\hat{V}_{ee}$ operators, *cf.* V_{ext} which depends on the location and charge of the nuclei.) Written in this way, $F[n]$ — and by extension, $E[n]$ — is a statement of existence and no more.

Kohn and Sham astutely noticed that a system of non-interacting particles subject to the external potential $V_{\text{eff}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + \Phi(\mathbf{r}) + V_{xc}[n](\mathbf{r})$ will share the same ground-state density as this interacting system, where $V_{xc} = \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})}$.[19] (To see this, one need only apply the variational principle to Equation 2.4.) Consequently, the ground-state density of the interacting system can be obtained via the auxiliary problem

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{eff}}[n](\mathbf{r})\right]\varphi_i(\mathbf{r}) = \varepsilon_i\varphi_i(\mathbf{r}), \tag{2.5a}$$

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\varphi_i(\mathbf{r})|^2. \tag{2.5b}$$

where crucially this auxiliary system is non-interacting and will therefore be much easier to solve. These are the Kohn-Sham (KS) equations. Since the effective potential is itself dependent on the system density, the equations must be solved self-consistently: that is, given some initial density $n(\mathbf{r})$ we can generate an effective potential $V_{\text{eff}}[n](\mathbf{r})$, for which we solve Equation 2.5a to obtain the Kohn-Sham orbitals $\{\varphi_i(\mathbf{r})\}$. From these we can obtain an updated density via Equation 2.5b, then update the effective potential, and so on until the density and potential converge. This would yield the exact ground state of the system if the form of the exact exchange-correlation functional was known, but at best it can only be approximated.[†,‡]

2.1.3 Exchange-correlation functionals

Substantial scientific effort has produced a plethora of xc functionals to approximate exchange and correlation. Of these, perhaps the simplest and most naïve is the local density approximation (LDA), which assumes the exchange and correlation are only dependent on the local electron density:

$$E_{xc}^{\text{LDA}}[n] = \int \mathrm{d}\mathbf{r}\, n(\mathbf{r})\varepsilon_{xc}(n(\mathbf{r})) \tag{2.7}$$

where for the exchange part we take the exact result for the homogeneous electron gas

$$E_{x}^{\text{LDA}}[n] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/2}\int \mathrm{d}\mathbf{r}\, n(\mathbf{r})^{4/3} \tag{2.8}$$

and the correlation energy is taken from quantum Monte Carlo (QMC) calculations on the homogeneous electron gas by Ceperley and Alder.[85]

To improve upon this, the next obvious step is to include dependence on the gradient of the

[†]Note that while the real and auxiliary systems have matching densities, their energies differ; to recover the real energy (Equation 2.4) from the eigenenergies of the auxiliary system $\{\varepsilon_i\}$, we make the correction

$$E[n] = \sum_{i}^{\text{occ}} \varepsilon_i + \int \mathrm{d}\mathbf{r}\, n(\mathbf{r})\left(\varepsilon_{xc}[n](\mathbf{r}) - V_{xc}[n](\mathbf{r}) - \frac{1}{2}\Phi(r)\right) \tag{2.6}$$

which removes the double-counting of terms in the Hartree term and the xc functional.

[‡]Also note that the kinetic energy $T_0[n]$ is the kinetic energy of the non-interacting Kohn-Sham system (hence the "0"); the exact xc functional would include the difference between the kinetic energies of the interacting and non-interacting systems.

density:

$$E_{xc}^{\mathrm{GGA}}[n] = \int \mathrm{d}\mathbf{r}\, n(\mathbf{r})\varepsilon_{xc}(n(\mathbf{r}), \nabla n(\mathbf{r})) \tag{2.9}$$

known as generalised gradient approximations (GGAs). By including more and more terms in the xc functional, we can construct a whole hierarchy of functionals varying in accuracy and complexity.

With a suitable xc functional at our disposal, DFT provides a tractable method of obtaining approximate solutions for the ground-state density of many-electron systems. Over the past few decades, DFT has established itself as a go-to method for such calculations.[20,21] This success hinges on the fact that for many systems, it turns out that these approximate xc functionals are able to predict the density, energy, and derived properties of interest with sufficient accuracy.

2.1.4 A standard implementation of DFT: plane waves

When it comes to practically solving Equations 2.5a and 2.5b, it is common to adopt a basis of plane waves in order to take numerical advantage of Bloch's theorem. In this framework, our Kohn-Sham eigenfunction can be written as

$$\varphi_i(\mathbf{r}) = \sum_{\mathbf{q}} c_{i\mathbf{q}} \times \frac{1}{\sqrt{\Omega}} \exp(i\mathbf{q}\cdot\mathbf{r}) = \sum_{\mathbf{q}} c_{i\mathbf{q}} \langle \mathbf{r}|\mathbf{q}\rangle \tag{2.10}$$

where $c_{i\mathbf{q}}$ are the expansion coefficients of the wavefunction now expressed in terms of orthonormal plane waves. In this basis the Schrödinger equation for the Kohn-Sham system (Equation 2.5a) becomes

$$\sum_{\mathbf{q}} \int d\mathbf{r}\, \langle \mathbf{q}'|\mathbf{r}\rangle \left(-\frac{1}{2}\nabla^2 + V_{\mathrm{eff}}[n](\mathbf{r})\right) \langle \mathbf{r}|\mathbf{q}\rangle c_{i\mathbf{q}} = \varepsilon_i c_{i\mathbf{q}'}. \tag{2.11}$$

If the physical system in question is periodic, then the effective potential is also periodic:

$$V_{\mathrm{eff}}[n](\mathbf{r}) = \sum_m V_{\mathrm{eff}}[n]\left(\mathbf{G}_m\right) \exp\left(\mathbf{i}\mathbf{G}_m \cdot \mathbf{r}\right) \tag{2.12}$$

where $\{\mathbf{G}_m\}$ are the reciprocal lattice vectors. Thus

$$\left\langle \mathbf{q}' \left| \hat{V}_{\mathrm{eff}}[n] \right| \mathbf{q} \right\rangle = \sum_m V_{\mathrm{eff}}\left(\mathbf{G}_m\right) \delta_{\mathbf{q}'-\mathbf{q},\mathbf{G}_m} \tag{2.13}$$

– that is, plane waves $|\mathbf{q}\rangle$ and $|\mathbf{q}'\rangle$ are only coupled if they differ by a reciprocal lattice vector. (The kinetic energy, meanwhile, does not couple different plane waves.) So if we define a new wave-vector $\mathbf{k} = \mathbf{q} + \mathbf{G}_m = \mathbf{q}' + \mathbf{G}_{m'}$ then our Schrödinger equation becomes the $\mathbf{k}$-dependent

$$\sum_{m'} H_{mm'}[n](\mathbf{k}) c_{im'}(\mathbf{k}) = \varepsilon_i(\mathbf{k}) c_{im}(\mathbf{k}). \tag{2.14}$$

where

$$H_{mm'}[n](\mathbf{k}) = \frac{1}{2}\left|\mathbf{k} + \mathbf{G}_m\right|^2 \delta_{mm'} + V_{\mathrm{eff}}[n]\left(\mathbf{G}_m - \mathbf{G}_{m'}\right). \tag{2.15}$$

Furthermore, we now have a **k**-dependent eigenfunction

$$\varphi_{i\mathbf{k}}(\mathbf{r}) = \sum_m c_{im}(\mathbf{k}) \times \frac{1}{\sqrt{\Omega}} \exp\left(i\left(\mathbf{k}+\mathbf{G}_m\right)\cdot\mathbf{r}\right) = u_{i\mathbf{k}}(\mathbf{r})\exp(i\mathbf{k}\cdot\mathbf{r}) \tag{2.16}$$

Thus, if we adopt a plane-wave basis, instead of solving Equations 2.5a and 2.5b self-consistently, we now solve

$$\sum_{m'} H_{mm'}[n](\mathbf{k})c_{im'}(\mathbf{k}) = \varepsilon_i(\mathbf{k})c_{im}(\mathbf{k}) \tag{2.17a}$$

$$n(\mathbf{r}) = \frac{1}{N_k}\sum_{i\mathbf{k}} f\left(\varepsilon_{i\mathbf{k}}\right) \underbrace{\frac{1}{\Omega}\sum_{mm'} c^*_{im}(\mathbf{k})c_{im'}(\mathbf{k})\exp\left(i\left(\mathbf{G}_{m'}-\mathbf{G}_m\right)\cdot\mathbf{r}\right)}_{n_{i\mathbf{k}}(\mathbf{r})=|\varphi_{i\mathbf{k}}(\mathbf{r})|^2} \tag{2.17b}$$

self-consistently. (In the calculation of n we have averaged over **k** points, and f is the Fermi-Dirac distribution.)

2.2 The challenge of size

A number of obstacles stand in the way of using plane-wave DFT to accurately simulate systems such as the OEC. Plane waves are well-suited to describing periodic functions, rather than the heterogeneity of proteins, and they struggle especially with large regions of vacuum that inevitably feature in protein systems.[†] But a more insurmountable problem is the issue of scale: photosystem II itself comprises some 120,000 atoms, which is far too large a system to be included in a calculation in its entirety. To perform simulations of the OEC, one must instead consider a system incorporating the OEC core and a selection of the nearby residues, discarding more peripheral atoms to make the system a tractable size.[‡] A good cluster model must be sufficiently large as to accurately capture the behaviour of the system without being prohibitively computationally expensive. This begs the question: how large a cluster does one need?

Intuitively, one might hope that only the chemically active species need be included in any cluster model — but this is not the case. Exclusion of more peripheral residues[§] may induce (or suppress) polarisation or charge transfer across the system, and spurious forces on the periphery of the cluster can propagate to the centre of the cluster.[86] This can ultimately affect the geometry (and hence the electronic structure) of DFT-optimised clusters.

One approach for including more of the biological system at minimal computational cost is quantum mechanics/molecular mechanics (QM/MM). In QM/MM, a central subsystem is treated at the quantum-mechanical level, and a secondary shell of atoms is treated using classical molecular mechanics. In principle, the quantum mechanical (QM)-only and QM/MM methods

[†]Most codes assume periodic boundary conditions, so large areas of vacuum or solvent are required to prevent periodic images from interacting.

[‡]Presently, computational studies of the OEC tend to treat anywhere from 75 to 250 atoms at a quantum-mechanical level of theory.

[§]A *residue* is a single amino acid within a peptide chain.

should give the same result in the large-QM-system limit.

QM/MM studies of biological systems have reported disappointingly slow convergence (with respect to QM region size) of free energies,[86–88] barrier heights,[89] nuclear magnetic resonance (NMR) shieldings,[90,91] solvation effects,[92,93] and excitation energies.[94] In the specific case of the OEC, it has been shown that spectroscopic properties do not adequately converge (with respect to cluster size) for 225-atom systems.[95] Furthermore, QM/MM approaches fail to agree with analogous QM-only calculations in the large-system limit. Studies on other biological systems have reported that properties only begin to converge for systems of the size of 500 atoms.[94,96]

Accurately simulating a system of 500 (or more) atoms is a substantial challenge. In the case of a traditional DFT calculation of N atoms using a plane-wave basis of size M, the requirement to orthogonalise all of the eigenstates (which scales as $\mathcal{O}(N^2) \times \mathcal{O}(M)$) quickly becomes prohibitively expensive.[†]

In order to perform calculations on proteins, therefore, it is necessary to recast the DFT formalism so that it is amenable to linear scaling. Usually this is done by adopting an atom-centred localised basis such as numerical atomic orbitals $\phi_\alpha(\mathbf{r}) = \varphi_{n_\alpha l_\alpha}(r_I) Y_{l_\alpha m_\alpha}(\hat{\mathbf{r}}_I)$ (where $\mathbf{r}_I = \mathbf{r} - \mathbf{R}_I$). If these orbitals $\{\phi_\alpha\}$ are subject to a radial cut-off then the number of non-diagonal elements of the Hamiltonian $H_{\alpha\beta} = \langle\phi_\alpha|\hat{H}_{KS}|\phi_\beta\rangle$ and the overlap matrix $S_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle$ only grows linearly with system size. Nevertheless, finding the KS eigenstates still involves the $\mathcal{O}(N^3)$ diagonalisation of $H_{\alpha\beta}$. Furthermore, by adopting a local basis such as numerical atomic orbitals, one loses the ability to systematically extend a basis set: in the case of a plane-wave basis, the accuracy of the basis is controlled solely by the energy cut-off; in a localised basis one can choose increasingly complex basis sets (for instance, multiple zeta or split valence bases), but these extensions are by no means systematic.

2.3 ONETEP

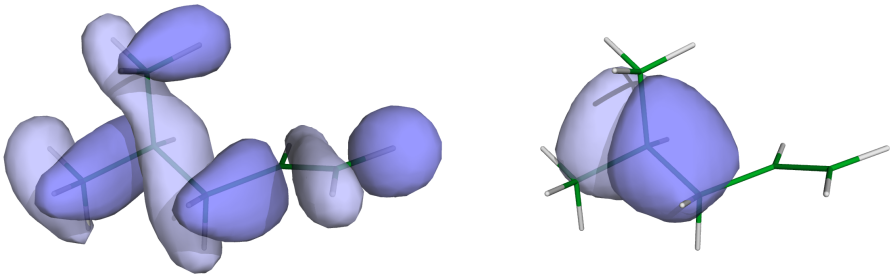
ONETEP (Order-N Electronic Total Energy Package) is a code that implements DFT using a local basis which achieves true linear scaling systematically.[22] The following section explains how this is achieved.

2.3.1 Density kernel and NGWF formalism in ONETEP

The set of KS orbitals $\{\varphi_n(\mathbf{r})\}$ of DFT can be equivalently described by the single-particle density matrix

$$\hat{\rho} = \sum_n f_n |\varphi_n\rangle\langle\varphi_n|; \qquad \rho(\mathbf{r}, \mathbf{r}') = \sum_n f_n \varphi_n^*(\mathbf{r})\varphi_n(\mathbf{r}') \tag{2.18}$$

[†]Naïvely, solving the KS system (Equations 2.5a and 2.5b) would require constructing and diagonalising a M-by-M Hamiltonian, which scales as $\mathcal{O}(M^3)$. With plane waves, M can be of the order of 10^6, which means this matrix would not be able to be stored in RAM, let alone diagonalised. Instead, iterative diagonalisation procedures (such as the Davidson method[97]) can be used to isolate and compute the lowest $N_b \ll M$ energy eigenvectors (where N_b is the number of bands, which is typically on the order of the number of electrons N), resulting in the oft-quoted cubic scaling of plane-wave DFT.



(a) **(b)**

Figure 2.1: A single KS orbital (a) and an optimised NGWF (b) for 4-methyl-1-pentene. The KS orbital extends across the whole system, whereas the NGWF is localised and therefore preferable as a basis function.

where f_n is the occupancy of the n^{th} orbital. The trace of this matrix yields the density of the system ($n(\mathbf{r}) = 2\rho(\mathbf{r}, \mathbf{r})$) and the total energy of the non-interacting auxiliary system is given by

$$E = 2\text{Tr}\left[\hat{\rho}\hat{H}\right], \tag{2.19}$$

which when performing DFT calculations is minimised subject to particle number and idempotency conservation.[†] (Note that this formulation avoids having to perform diagonalisation, which scales cubically.) The energy of the interacting system is then obtained by making the standard double-counting corrections for the Hartree and exchange-correlation terms.

Note that already the system has N occupied states and therefore the size of the density matrix defined in this fashion scales as N^2. To achieve linear scaling, ONETEP transforms from the basis of KS molecular orbitals (which extend over the entire system) to a basis of non-orthogonal generalised Wannier functions (NGWFs),[98] defined as

$$\phi_{\alpha\mathbf{R}}(\mathbf{r}) = \frac{V}{(2\pi)^3}\int_{BZ} e^{-i\mathbf{k}\cdot\mathbf{R}}\left(\sum_n \psi_{n\mathbf{k}}(\mathbf{r})M_{n\alpha}\right)d\mathbf{k}, \tag{2.20}$$

which is a Fourier transform of Bloch states $\psi_{n\mathbf{k}}$ mixed by the matrix $\mathbf{M}$. In the case that $\mathbf{M}$ is the identity, the functions are orthogonal. (ONETEP uses a non-orthogonal form.) The label α allows for the possibility of more than one basis function per unit cell, $\mathbf{R}$ is the lattice vector, and the integral runs over the first Brillouin Zone. NGWFs are spatially localised, which makes them suitable for linear-scaling codes (see Figure 2.1).

To implement this new basis the density matrix is reformulated as

$$\begin{aligned}\rho(\mathbf{r}, \mathbf{r}') &= \sum_n f_n \varphi_n(\mathbf{r})\varphi_n^*(\mathbf{r}') \\ &= \sum_n f_n \frac{V}{(2\pi)^3}\int_{BZ} \psi_{n\mathbf{k}}(\mathbf{r})\psi_{n\mathbf{k}}^*(\mathbf{r}')\text{d}\mathbf{k} \\ &= \sum_{\alpha,\beta,\mathbf{R}} \phi_{\alpha\mathbf{R}}(\mathbf{r})K^{\alpha\beta}\phi_{\beta\mathbf{R}}^*(\mathbf{r}')\end{aligned} \tag{2.21}$$

[†]Note that these expressions apply to systems that are not spin-polarised, hence the various factors of two. It is straightforward to generalise the expressions to account for spin.

where now $K^{\alpha\beta}$ is the density kernel, the representation of the density matrix in the NGWF basis

$$K^{\alpha\beta} = \sum_n N_n^\alpha f_n (N^\dagger)_n^\beta, \tag{2.22}$$

and $\mathbf{N}$ is the inverse of $\mathbf{M}$ from Equation 2.20. In this new framework, the total energy is $E = K^{\alpha\beta}H_{\alpha\beta}$, the idempotency restriction is $K^{\alpha\gamma}S_{\gamma\delta}K^{\delta\beta} = K^{\alpha\beta}$ and the normalisation condition is $K^{\alpha\beta}S_{\beta\alpha} = N$, where $S_{\alpha\beta} = \int \mathrm{d}\mathbf{r}\phi_\alpha(\mathbf{r})\phi_\beta(\mathbf{r})$ is the overlap matrix between NGWFs.

Crucial to the success of linear-scaling DFT is the "near-sightedness" of the density matrix/kernel: it can be shown that the density matrix (and thus also the density kernel) is highly diagonally-dominated, with $\rho(\mathbf{r}, \mathbf{r}')$ decaying very rapidly with $|\mathbf{r} - \mathbf{r}'|$.[99†] Thus the non-zero elements $K^{\alpha\beta}$ of the density-kernel corresponding to NGWFs centred further apart than a cut-off r_k can be discarded. This ensures the density matrix is sufficiently sparse to guarantee linear scaling (although in practice this truncation need not be imposed).

Thus far, this switch to NGWFs is nothing more than a mathematical reformulation. But this is where the approach of ONETEP diverges quite substantially from most DFT codes. When it comes to basis sets, the important considerations are (a) how few basis functions do we need to accurately represent the physical system and (b) if we don't have enough basis functions, how do we generate more? One of the distinct advantages of a plane-wave basis is that its accuracy can be systematically improved: by progressively including higher-k waves in the basis set, we converge on the true solution. (But on the other hand, many plane waves are often required.) At the other end of the spectrum, atomic orbital basis sets require fewer basis functions (because they more closely resemble electronic states) but extending the basis is non-trivial, requiring human expertise and effort.

ONETEP uses an alternative approach: a minimal basis set of NGWFs is used, but the basis functions themselves are optimised during the course of a calculation *in situ*. As they are optimised, they respond and adapt to their surroundings. Although this requires additional computational effort, the resulting basis set is of minimal size and of high accuracy.

The optimisation of the NGWFs is performed in tandem with the optimisation of the density kernel in a self-consistent fashion, as illustrated in Figure 2.2. Below, I will describe how the NGWF and density kernel update steps are achieved.

2.3.2 NGWF optimisation

To facilitate their optimisation, the NGWFs are expanded in terms of frequency-limited delta functions — or "psinc" functions[‡] — shown in Figure 2.3b. A psinc function centred on the point $\mathbf{r}_{klm}$ is given by

$$D_{klm}(\mathbf{r}) = \frac{1}{N_1 N_2 N_3} \sum_{P=-J_1}^{J_1} \sum_{Q=-J_2}^{J_2} \sum_{R=-J_3}^{J_3} e^{i(P\mathbf{b}_1 + Q\mathbf{b}_2 + R\mathbf{b}_3)\cdot(\mathbf{r}-\mathbf{r}_{klm})} \tag{2.23}$$

[†]This decay is exponential for insulators and metals at finite temperatures, but only algebraic for zero-temperature metals.

[‡]Periodic sinc functions, where $\mathrm{sinc}(x) = \sin(x)/x$.

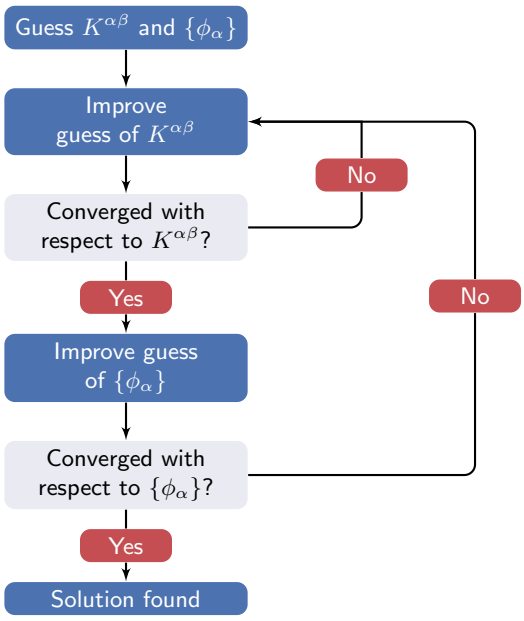


Figure 2.2: Process by which ONETEP finds a self-consistent ground-state solution for $K^{\alpha\beta}$ and $\{\phi_\alpha\}$.

which is zero at all grid points save $\mathbf{r}_{klm}$. Psinc functions are orthogonal by construction. Additionally, the projection of an arbitrary periodic function $f(\mathbf{r})$ onto a psinc function is

$$\int_V \mathrm{d}\mathbf{r}\, D_{klm}(\mathbf{r}) f(\mathbf{r}) = W f_D(\mathbf{r}_{klm}), \tag{2.24}$$

where $f_D(\mathbf{r})$ is the bandwidth-limited $f(\mathbf{r})$ and W the volume per grid point. The NGWFs can be expressed in both the psinc basis

$$\phi_\alpha(\mathbf{r}) = \sum_{k=0}^{N_1-1} \sum_{l=0}^{N_2-1} \sum_{m=0}^{N_3-1} C_{klm,\alpha} D_{klm}(\mathbf{r}) \tag{2.25}$$

and the plane-wave basis

$$\phi_\alpha(\mathbf{r}) = \frac{1}{V} \sum_{P=-J_1}^{J_1} \sum_{Q=-J_2}^{J_2} \sum_{R=-J_3}^{J_3} \tilde{\phi}_\alpha\left(P\mathbf{b}_1 + Q\mathbf{b}_2 + R\mathbf{b}_3\right) e^{i(P\mathbf{b}_1+Q\mathbf{b}_2+R\mathbf{b}_3)\cdot\mathbf{r}}. \tag{2.26}$$

Crucially, the size (and therefore accuracy) of these representations may be related to an energy cut-off (that of the maximum wavevector the grid can support). This means that the basis set of NGWFs expanded in terms of psinc functions can be systematically improved by increasing this cut-off.[100] Compare this to split-valence basis sets, for which there is no systematic approach for expanding the basis.

The amplitudes $\tilde{\phi}_\alpha\left(P\mathbf{b}_1 + Q\mathbf{b}_2 + R\mathbf{b}_3\right)$ are the result of a discrete Fourier transform on the psinc expansion coefficients $C_{klm,\alpha}$, and thus one can still make use of the computational power of Fourier transforms throughout the ONETEP calculation: that is, we are able to calculate energy contributions in real- or reciprocal-space as convenient.[101†]

[†]In order to do all this we must adopt periodic boundary conditions.

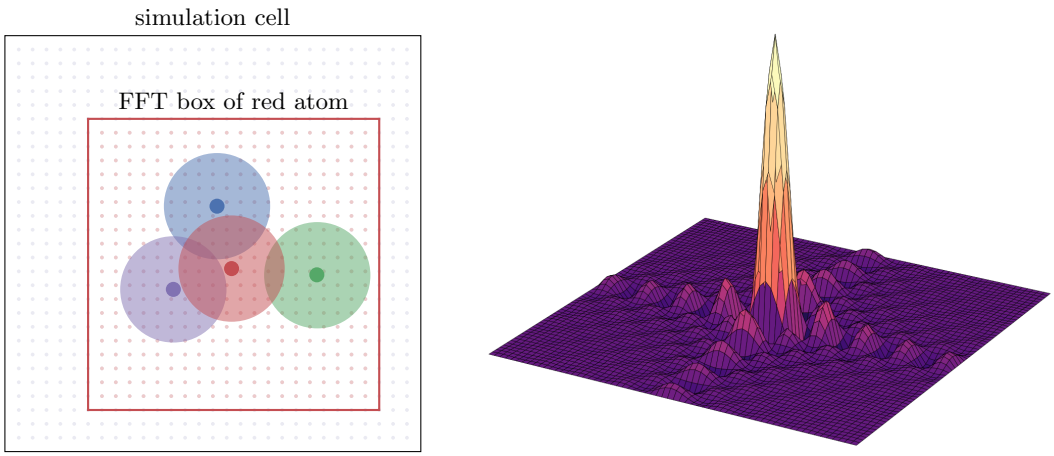


Figure 2.3: (a) The construction of a FFT box in a cartoon system of four atoms, each shown with their nuclei (solid circles) and NGWF radii (translucent). The underlying psinc grid is shown as grey points. The red atom has three neighbouring atoms (blue, green, and purple). Its FFT box encloses its own NGWFs and those of its neighbours. (b) A psinc function. (N.B. The grid shown here serves as a visual aid. It is eight times finer than the psinc grid. Thus, unlike the psinc grid, it has non-zero points in addition to the central peak.)

In the above equations, the indexes k, l, and m spanned the whole simulation cell. But in order to achieve linear scaling, we must apply several restrictions to our representations of the NGWFs. The first is to confine every NGWF to a sphere centred on the atom the NGWF belongs to. The second is to associate a fast Fourier transform (FFT) box with each NGWF. Each FFT box is centred on its corresponding atom, and is sufficiently large to encompass all other NGWFs that overlap with the central NGWF, as illustrated in Figure 2.3a. We then restrict the calculation of all energy terms to operations within FFT boxes (rather the full simulation cell). Roughly, ONETEP will start outperforming plane-wave methods when the FFT box becomes smaller than the simulation cell.

The NGWFs are optimised using the conjugate gradient method.[102] The result is shown in Figure 2.4. By the time the calculation is complete, the now optimised basis set of NGWFs will clearly serve as an excellent basis set for capturing electronic behaviour. This approach is especially well suited for simulating organic molecules: plane waves are good at describing periodic structures such as crystals, but they struggle with heterogeneity and large regions devoid of atoms, both of which are common in biological systems. Because NGWFs are local they do not encounter such difficulties.

2.3.3 Density kernel optimisation

Every time the NGWFs are updated, the density kernel is optimised (refer back to Figure 2.2). The crucial consideration during the density kernel optimisation is that the idempotency of the kernel must be maintained. This is achieved by performing a purification transformation via the method of Li, Nunes, and Vanderbilt.[104] (We also use adaptive purification to monitor extremal eigenvalues, and in some instances we additionally apply a penalty functional method.[105] These will not be discussed here; details can be found in Ref. 106.)

Figure removed due to copyright. The original can be seen in Ref. 103 as Figure 5.

Figure 2.4: A NGWF centred on the carbon of an ethene molecule at the start and end of a calculation. It is initialised as the $2p$ pseudo-atomic orbital (PAO). Figure taken from Ref. 103.

Figure removed due to copyright. The original can be seen in Ref. 106 as Figure 5.

Figure 2.5: An illustration of the purification transform for three different starting orbital occupancies (a) an initial occupancy of $2/5$ converges to 0; (b) an initial occupancy of $6/5$ converges to 1; (c) an initial occupancy of $-2/5$ converges to 1. Figure taken from Ref. 106.

A simple purification algorithm

The purification methods in ONETEP are based on the following simple iterative process by McWeeny[107]

$$\rho_{k+1} = 3\rho_k^2 - 2\rho_k^3. \tag{2.27}$$

In the limit of $k \to \infty$, this procedure (illustrated in Figure 2.5) drives the eigenvalues of ρ_k towards either zero or one, provided the initial eigenvalues lie in the interval $(\frac{1-\sqrt{5}}{2}, \frac{1+\sqrt{5}}{2})$. If the initial eigenvalues lie within the slightly tighter bounds $(-1/2, 3/2)$ then the purified occupancies remain in $[0, 1]$. However, as Figure 2.5c demonstrates, occupancies can flip from "unoccupied" to "occupied" and *vice versa*. This can be avoided if starting occupancies lie within the tighter bounds $(\frac{1-\sqrt{3}}{2}, \frac{1+\sqrt{3}}{2})$.

Li-Nunes-Vanderbilt method

Building upon the simple purification algorithm, the Li-Nunes-Vanderbilt method[104,108] defines the density matrix ρ in terms of an auxiliary matrix σ via

$$\rho = 3\sigma^2 - 2\sigma^3 \tag{2.28}$$

where σ is defined by

$$\sigma(\mathbf{r}, \mathbf{r}') = \sum_{\alpha\beta} \phi_\alpha(\mathbf{r}) L^{\alpha\beta} \phi^*_\beta(\mathbf{r}') \tag{2.29}$$

L is an auxiliary kernel related to the original kernel by $\mathbf{K} = 3\mathbf{LSL} - 2\mathbf{LSLSL}$. Minimising the energy by optimising the matrix elements of the auxiliary kernel in place of the original kernel naturally drives the density matrix to idempotency. (Specifically, if the eigenvalues of L remain within $(-1/2, 3/2)$ then the purified density matrix will be weakly idempotent *i.e.* with eigenvalues in $[0, 1]$.)

In addition to idempotency, we must enforce normalisation. This can be achieved via a simple rescaling of the purification transformation

$$\rho = \frac{N_e}{\mathrm{Tr}\left[3\sigma^2 - 2\sigma^3\right]}(3\sigma^2 - 2\sigma^3) \tag{2.30}$$

and with the revised auxiliary kernel defined via

$$K_{\alpha\beta} = \frac{N_e}{\mathrm{Tr}\left[3LSL - 2LSLSL\right]}\left(3LSL - 2LSLSL\right)_{\alpha\beta}. \tag{2.31}$$

When this is used to generate a search direction, the extra terms in the denominator automatically project out the electron number gradient. During this process the extremal occupancies are monitored, and adaptive purification is applied if necessary.

Real *vs.* imaginary representation and equivalence to *k*-point sampling

ONETEP does not support **k**-point sampling[†] (although work to add this functionality is ongoing). In effect, we only sample $\mathbf{k} = \Gamma$ and consequently our NGWFs and KS eigenfunctions are real. Because ONETEP is designed to tackle large, irregular systems where plane-wave treatments are not well-suited, this is not a critical shortcoming. But if we want to treat periodic systems or benchmark calculations against a plane-wave code, one can achieve the equivalent of **k**-point sampling using super-cells.[‡] While this scales far worse than **k**-point parallelism, it does allow us to sample non-zero **k** if need be.

At this stage in the formalism, ONETEP can perform DFT. But a number of additional functionalities – unnecessary for the most basic of DFT calculations – prove indispensable when it comes to biological systems. These are outlined in the remainder of this section.

[†]That is, performing the sum over **k** in Equations 2.17a and 2.17b.

[‡]Consider a periodic one-dimensional system with n k-points. Assuming time-reversal symmetry, Bloch waves satisfy $\psi_{-k} = \psi_k^*$ and $\varepsilon(-k) = \varepsilon(k)$. This means that instead of working with Bloch waves as our eigenfunctions on $[-a/2, a/2]$ we can equivalently work with the combinations $\tilde{\psi}(r) = \psi_k(r) + \psi_{-k}(r)$ and $i[\psi_k(r) - \psi_{-k}(r)]$ defined over $[-na/2, na/2]$. These new eigenfunctions (a) are real and (b) are Bloch waves with $k = \Gamma$ since $\tilde{\psi}(r) = \exp(i0r)v(r)$ where $v(r) = \psi_k(r) + \psi_{-k}(r)$ is a real periodic function. Furthermore, $v(r)$ must have periodicity na since

$$\begin{aligned}\tilde{\psi}(r + na) &= \psi_k(r + na) + \psi_{-k}(r + na)\\ &= \exp(ik(r + na))u_k(r + na) + \exp(-ik(r + na))u_{-k}(r + na)\\ &= \psi_k(r) + \psi_{-k}(r)\\ &= \tilde{\psi}(r)\end{aligned} \tag{2.32}$$

where the second step follows because $k \in \{-(n-1)\pi/na, ..., \pi/a\}$ and we have assumed n is even. Therefore, instead of the original system with n k-points, we can equivalently work in a system with periodicity na and constrain ourselves to real orbitals.

2.3.4 Implicit solvation

Even with linear scaling it is prohibitively expensive to consider solvent molecules explicitly. To make computation feasible, solvents are usually modelled as continuous polarisable dielectrics. This also avoids the problem of the many possible configurations an explicit solvent can adopt, which would require averaging over inaccessibly long timescales. In effect, a continuous dielectric provides an approximation to a time-average over configurations.

ONETEP includes solvation effects by defining a smooth dielectric cavity around the solvated molecule, whose relative permittivity is defined as

$$\varepsilon(\mathbf{r}) = 1 + \frac{\varepsilon_\infty - 1}{2}\left(1 + \frac{1 - (n(\mathbf{r})/n_0)^{2\beta}}{1 + (n(\mathbf{r})/n_0)^{2\beta}}\right) \tag{2.33}$$

where $n(\mathbf{r})$ is the electronic density, ε_∞ is the bulk relative permittivity, and n_0 and β are free parameters that define the density value where the permittivity drops to $\varepsilon_\infty/2$, and the smoothness of the transition.[27,109] Note that $\lim_{n(\mathbf{r})\to 0}\varepsilon(\mathbf{r}) = \varepsilon_\infty$ and $\lim_{n(\mathbf{r})\to\infty}\varepsilon(\mathbf{r}) = 1$.

The solvation routine solves the non-homogenous Poisson equation $\nabla[\varepsilon(\mathbf{r})\phi(\mathbf{r})] = n_{\text{tot}}(\mathbf{r})$ (where n_{tot} includes nuclear charges) to obtain the potential ϕ due to the molecular density in the dielectric. Because $\varepsilon(\mathbf{r})$ is density-dependent, a self-consistent solution must be found. Once this is achieved, the potential ϕ is then used in place of the Hartree potential in the Kohn-Sham energy functional.

It is expensive to repetitively search for $\varepsilon(\mathbf{r})$ every time $n(\mathbf{r})$ is revised (which happens many times during DFT calculations during the search for a self-consistent ground-state electronic density). To avoid having to perform this inner self-consistency loop, an *in vacuo* calculation for the system is typically performed first, yielding the vacuum density solution $n_{\text{vac}}(\mathbf{r})$. This density is then used to generate the dielectric cavity according to Equation 2.33, and this cavity is then left unchanged during the subsequent solvated calculation. This procedure reduces wall time and computational demands by about an order of magnitude, and yields solvation energies which are within several percent of the full self-consistent solution.[109]

The use of an implicit solvation model is especially important in quantum mechanical calculations of biological systems. Proteins are naturally found in solution, so when modelling small proteins it is standard practice to do so with an implicit solvent. For larger proteins it is important for a different reason: it is a widely held misconception that xc functionals that do not include HF exchange cannot be applied to large systems — a belief that stems from reports of unphysical closing of the gap between the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) in calculations of proteins and water clusters.[110] However, it has been demonstrated that this is due to the emergence of a significant dipole across the entire system (largely due to unterminated hydrogen bonds at the cluster boundary). Placing such systems in an implicit solvent helps to compensate for the anomalous dipole and restores the expected HOMO-LUMO gap.[111]

2.3.5 Dispersion

Dispersion forces — or van der Waals forces — play an important role in determining the structure and behaviour of biological systems. However, semi-local DFT xc functionals are inadequate for describing long-range electronic correlation, as they (a) do not consider instantaneous density fluctuations and (b) are "short-sighted": that is, they consider only local properties in the calculation of exchange and correlation. Consequently, such functionals only give rise to exchange-correlation interactions between two atoms if their electronic densities overlap — and because overlap decays exponentially with distance, they often underestimate dispersion.

There are a number of approaches for incorporating dispersion into DFT calculations. One approach is to construct exchange correlation functionals which explicitly include dispersion via a non-local term of the form $E^{\mathrm{nl}} = \iint dr_1 dr_2\, n(r_1)\varphi(r_1, r_2)n(r_2)$ for some potential φ. This has been done with varying levels of success (see Ref. 112 for details). A more pragmatic approach is to add a pair-wise interatomic R^{-6} correction term to the calculated energy as a post-processing procedure,[†] with parameters empirically fitted to reproduce binding energies of a wide range of

[†]The potential of a classical proton/electron system is given (in atomic units) by

$$V(\mathbf{R}) = \frac{\mathbf{r}_I \cdot \hat{\mathbf{R}}}{R^2}, \tag{2.34}$$

where $\mathbf{r}_I$ is the position of the electron relative to the proton at the origin. If there is a second proton/electron system where the proton is at $\mathbf{R} = R\hat{\mathbf{z}}$ and the electron $\mathbf{r}_J$ relative to the proton, then the energy of the second dipole in the field of the first is given by

$$W_{dd} = -\mathbf{E}\cdot\mathbf{r}_J = \nabla V \cdot \mathbf{r}_J = \frac{1}{R^3}(x_I x_J + y_I y_J - 2z_I z_J). \tag{2.35}$$

The Hamiltonian for the analogous quantum mechanical system is $\hat{H} = \hat{H}_{0I} + \hat{H}_{0J} + \hat{W}_{dd}$, where $\hat{W}_{dd} = \frac{1}{R^3}(\hat{x}_I\hat{x}_J + \hat{y}_I\hat{y}_J - 2\hat{z}_I\hat{z}_J)$. Provided that the atoms are sufficiently distant, there is no interaction between them asides from the dipole term, so in the absence of the dipole interaction

$$(\hat{H}_{0I} + \hat{H}_{0J})|\varphi^I_{nlm};\varphi^J_{n'l'm'}\rangle = (E_n + E_{n'})|\varphi^I_{nlm};\varphi^J_{n'l'm'}\rangle, \tag{2.36}$$

and the ground state is $|\varphi^I_{100};\varphi^J_{100}\rangle$. If the dipole-dipole interaction is treated as a perturbation to the system, the first order correction to the energy is

$$E^{(1)} = \langle\varphi^I_{100};\varphi^J_{100}|\hat{W}_{dd}|\varphi^I_{100};\varphi^J_{1,0,0}\rangle = \frac{1}{R^3}\underbrace{\langle\varphi^I_{100}|\hat{x}_I|\varphi^I_{100}\rangle}_{0}\underbrace{\langle\varphi^J_{100}|\hat{x}_J|\varphi^J_{100}\rangle}_{0} + \text{similar terms for } \hat{y} \text{ and } \hat{z} \;= 0, \tag{2.37}$$

since for any stationary state of the atom, the mean values of the components of the position operator are all zero. The second order energy correction is

$$E^{(2)} = \sum_{n,l,m,n',l',m'} \frac{|\langle\varphi^I_{nlm};\varphi^J_{n'l'm'}|\hat{W}_{dd}|\varphi^I_{100};\varphi^J_{100}\rangle|^2}{2E_1 - E_n - E_{n'}}, \tag{2.38}$$

where $\{n,l,m,n',l',m'\} = \{1,0,0,1,0,0\}$ is excluded from the summation and E_i is the energy as given by $\hat{H}_0|\varphi_{i,l,m}\rangle$. Since $2E_1 - E_n - E_{n'} < 0$ and $\hat{W}_{dd} \sim R^{-3}$ it follows that this term is of the form

$$E^{(2)} = -\frac{C}{R^6}, \tag{2.39}$$

and thus dipole-dipole van der Waals forces are attractive and the energy scales as R^{-6}. Of course, a complete treatment of this system would also have to consider the higher order multipoles (*e.g.* dipole-quadrupole, quadrupole-dipole, quadruple-quadrupole etc.) — and indeed, such treatments exist.[113] It transpires that these terms also have zero contribution at first order, and hence the second-order effect of $\hat{W}_{dd}$ will constitute the largest energy correction. For a more complete discussion, the reader is referred to Ref. 114.

dimers.[115] For example

$$E_{vdW} = -\frac{1}{2}\sum_{I,J} f_6(R_{IJ}, R_I^0, R_J^0) C_{6IJ} R_{IJ}^{-6}, \tag{2.40}$$

where I and J are two atoms in the system, R_{IJ} is the distance between them, R_I^0 and R_J^0 are the atoms' vdW radii, C_{6IJ} is the two-body dispersion coefficient, and the R_{IJ}^{-6} singularity is removed by a short-ranged damping term f_6 (where $\lim_{R_{IJ}\to 0} f_6 = 0$). This damping term is further justified by the fact that any exchange functional should provide an adequate description of short range interactions and therefore a correction term is unwanted at close distances. There are a variety of different damping functions: for example, the damping function of Elstner *et al.* is[115]

$$f_6(R_{IJ}, R_I^0, R_J^0) = \left(1 - e^{-c_{\mathrm{damp}}\left(R_{IJ}/R_{IJ}^0\right)^7}\right)^4; \qquad R_{IJ}^0 = \frac{\left(R_I^0\right)^3 + \left(R_J^0\right)^3}{\left(R_I^0\right)^2 + \left(R_J^0\right)^2} \tag{2.41}$$

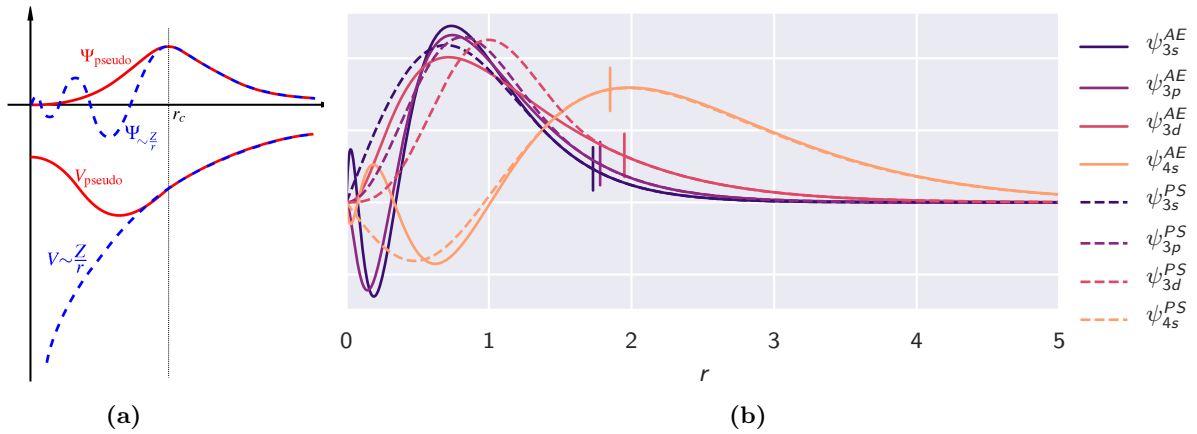
while Grimme proposed[116]

$$f_6(R_{IJ}, R_I^0, R_J^0) = \frac{1}{1 + e^{-c_{\mathrm{damp}}\left(\mathbf{R}_{IJ}/R_{ij}^0 - 1\right)}}; \qquad R_{IJ}^0 = R_I^0 + R_J^0. \tag{2.42}$$

Meanwhile, the C_{6IJ} parameters are typically found by fitting to experimental data or high-accuracy quantum mechanical calculations. The current scheme for calculating vdW corrections in ONETEP assigns to each atom C_{6I} coefficients according to which element it is. These coefficients have been predetermined empirically by fitting dispersion energy calculations to a database of small molecule dimers.[117] From these single-atom parameters the pair parameters C_{6IJ} are constructed (for example, Elstner *et al.* employ the geometric average $C_{6IJ} = \sqrt{C_{6I}C_{6J}}$).[115] Using parameters that are only dependent on the element type of the atom is a substantial approximation, since the C_{6IJ} coefficients ought to be sensitive to the atoms' surroundings. Atom-typing can go some way to addressing this issue, but one class of especially successful approaches involves density partitioning.[118–120] In these schemes, the electronic density is partitioned onto individual atoms, and then the C_6 coefficients are calculated on-the-fly using the resulting atom-in-molecule charges as well as free atom data. Alternatively, the C_6 coefficients can be determined via TDDFT, but this gives average errors of 15-20%, with some values deviating by up to 60%.[118]

2.3.6 Pseudopotentials

The core electrons of an atom are typically closely localised about the nucleus, have several nodes, and are highly insensitive to their surroundings. This means that calculating these deep-lying orbitals will be computationally intensive (requiring a high grid resolution) — disproportionate to their importance in determining the chemical properties. To greatly speed up calculations, the all-electron problem can be recast as a valence-electron problem, involving weaker nuclear potentials called pseudopotentials (PSPs). By absorbing the core electrons into the nuclear potential, our new calculations are less computationally intensive but no less accurate (provided the PSP is well constructed, of course).



(a) **(b)**

Figure 2.6: (a) A cartoon of a norm-conserving PSP, showing the all-electron potential and resulting radial wavefunction (blue, dashed) and a PSP and the resulting pseudo-wavefunction (red, solid). Note that the potentials and wavefunctions match outside of r_c, but within r_c the all-electron wavefunction oscillates much more rapidly. (b) The all-electron (AE) and pseudoatom (PS) valence atomic orbitals for manganese. The cut-off radii are drawn as short vertical lines.

There are a number of methods for constructing pseudopotentials. So-called norm-conserving potentials[121] are constructed such that for the isolated atom

- all-electron and pseudo- valence eigenvalues agree;
- all-electron and pseudo- valence wavefunctions agree beyond some cut-off radius r_c, as do their logarithmic derivatives;
- the integrated charge inside r_c for the all-electron and pseudo- valence wavefunctions match (hence "norm-conserving").

This is illustrated in Figure 2.6. The success of this method hinges on the fact that we can construct PSPs that are *transferable*: that is, a PSP constructed in one environment (typically for the isolated atom) can remain sufficiently accurate in other environments (solids, molecules, *etc.*).

2.3.7 Projector-augmented waves

Norm-conserving pseudopotentials are widely used, but they do have their drawbacks. In particular, PSPs for two-shell elements — that is, those with multiple partially-filled subshells like transition metals — tend to be much less accurate.[122] An alternative to PSPs, the projector-augmented wave (PAW) method[123] gives rise to even smoother wavefunctions near atomic centres.

Motivated by the smoothness of wavefunctions far from atomic centres, and their highly oscillatory nature near them, the PAW method decomposes the wavefunction into parts, so as to formally distinguish between these two regions of differing wavefunction behaviour. To achieve this distinction, we define the following quantities (considering the Hilbert space of all wavefunctions orthogonal to the core states):

$|\psi_n\rangle$ **AE orbitals:** the full solution to the system, treating both nuclear and bonding regions accurately

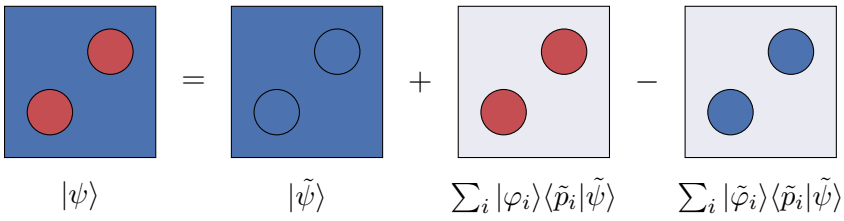


Figure 2.7: A cartoon of the PAW method, showing two augmentation regions (*i.e.* atoms). The full solution $|\psi\rangle$ incorporates accurate, system-independent solutions to the core region (red) and the system-dependent solution of the bonding regions (blue), but the reformulated DFT algorithms need only treat the PS orbitals $|\tilde{\psi}\rangle$.

$|\tilde{\psi}_n\rangle$ **pseudoatom (PS) orbitals:** related to the AE orbitals by a linear transformation $\hat{\mathcal{T}} = 1 + \sum_R \hat{\mathcal{T}}_R$ – that is, identical to the AE result except within atom-centred augmentation regions defined by $\{\hat{\mathcal{T}}_R\}$

$|\varphi_i\rangle$ **AE partial waves:** localised about each nuclei (*i.e.* within the augmentation regions). A natural choice for these are the solutions of the radial Schrödinger equation for the isolated atom

$|\tilde{\varphi}_i\rangle$ **PS partial waves:** a set of smooth functions defined such that $|\varphi_i\rangle = (1+\hat{\mathcal{T}}_R)|\tilde{\varphi}_i\rangle$. These should form a complete set of functions within the augmentation region

The transformation $\hat{\mathcal{T}}$ that satisfies all the properties we desire is given by

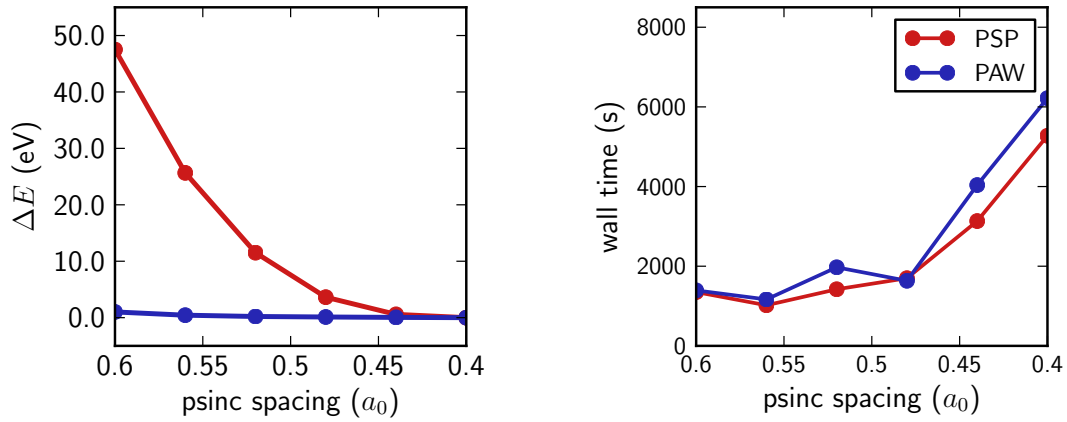
$$\hat{\mathcal{T}} = 1 + \sum_i \left(|\varphi_i\rangle - |\tilde{\varphi}_i\rangle\right)\langle\tilde{p}_i| \tag{2.43}$$

where $|\tilde{p}_i\rangle$ is the dual of $|\tilde{\varphi}_i\rangle$: that is, $\sum_i |\tilde{\varphi}_i\rangle\langle\tilde{p}_i| = 1$. Therefore we have

$$|\psi\rangle = |\tilde{\psi}\rangle + \sum_i \left(|\varphi_i\rangle - |\tilde{\varphi}_i\rangle\right)\langle\tilde{p}_i|\tilde{\psi}\rangle. \tag{2.44}$$

This transformation is loosely depicted in Figure 2.7. Three quantities define this transformation: our choices for the AE and PS partial waves, and the set of duals for the PS partial waves. By recasting operators, the electronic density, the energy, and other such quantities in terms of the PS orbitals, we can work with the much smoother PS representation of the system, but by keeping track of the projection the frozen core can be recovered when required, and the calculations are effectively all-electron. (See Ref. 124 for further details.)

The PAW approach relaxes the norm-conserving constraints on the core region, allowing for a much 'softer' representation within this core region, placing a lower demand on the grid resolution, even for systems that do not contain any transition metals (whose $3d/4s$ orbitals benefit greatly from the PAW treatment). The efficacy of the PAW approach for calculations on hexahydrated manganese is illustrated in Figure 2.8. Clearly, the PAWs outperformed PSPs, since (a) PSPs required a much higher resolution to achieve the same level of energy convergence and (b) PAW calculations came with little to no overhead compared to their PSP counterparts.



(a) **(b)**

Figure 2.8: Comparison of the performance of PSPs and PAWs for the $[Mn(H_2O)_6]^{3+}$ system with respect to psinc spacing. (a) Convergence of total system energy, relative to the energy of the $0.4a_0$ solution; (b) calculation wall time. The PSPs were generated by OPIUM and PAWs with ATOMPAW.[125,126] Figure taken from Ref. 127.

2.3.8 Excited states

Theoretical spectroscopy provides a crucial link between computational methods and experiment. The optical absorption spectrum is given by the imaginary part of the dielectric function

$$\varepsilon(\omega) = \frac{8\pi^2}{\Omega} \sum_{\mathbf{k},v,c} |\langle \psi^c_{\mathbf{k}} | \hat{\mathbf{q}} \cdot \mathbf{r} | \psi^v_{\mathbf{k}} \rangle|^2 \, \delta \left(E^c_{\mathbf{k}} - E^v_{\mathbf{k}} - \omega \right) \tag{2.45}$$

where $|\psi^{v/c}_{\mathbf{k}}\rangle$ is an eigenstate in the valence/conduction band with wavevector $\mathbf{k}$ and energy $E^{v/c}_{\mathbf{k}}$, and $\hat{\mathbf{q}}$ is the direction of polarisation of the photon. Ω is the cell volume; in ONETEP it is assumed that the cell is sufficiently large that the sum over $\mathbf{k}$ points need only include Γ. (This equation is nothing more than Fermi's golden rule.)

A crucial distinction between ONETEP and plane-wave DFT codes is that when ONETEP minimises the energy with respect to the density kernel and NGWFs, the unoccupied conduction bands do not have significant bearing on the total energy, and therefore they can be incorrect. To overcome this, a "conduction" calculation is performed, whereby the lowest-lying conduction states are selectively optimised (see Figure 2.9). For further details, see Ref. 128.

2.4 Applying ONETEP to a protein

I will now demonstrate the capabilities of ONETEP on cyclotide kalata B5.[129] The cyclotides are small disulfide-rich circular proteins found in plants. Their natural role is thought to be as a defence agent, protecting the plant from pests and pathogens. They have proved to be a potent pesticide,[130] and have also demonstrated anti-HIV,[131,132] anti-microbial,[133] and anti-tumour[134] activity. For the purposes of this thesis, cyclotide kalata is a useful test case because it does not contain any transition metal centres, meaning we need not yet worry about strong electronic correlation.

Figure removed due to copyright. The original can be seen in Ref. 128 as Figure 3.

Figure 2.9: Densities of state of a metal-free phthalocyanine, comparing results from ONETEP with and without the conduction state optimisation to the plane-wave code CASTEP. Figure taken from Ref. 128.

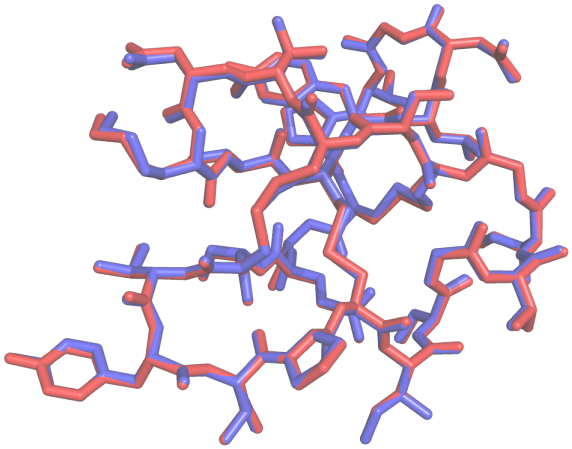


Figure 2.10: The XRD and ONETEP-optimised structures, in blue and red respectively.

In a previous study, Kulik *et al.* investigated the ability of QM-approaches to correctly predict the crystal structure of a number of proteins (starting from experimental structures), including cyclotide kalata B5.[135] However, in order to be able to consider molecules of hundreds of atoms in size, that study only used small, fixed basis sets, which comes at a substantial cost to accuracy and is prone to basis set superposition errors (BSSEs). We are interested in seeing how ONETEP — which is BSSE-immune — compares.

The geometry of the XRD structure of cyclotide kalata B5 was optimised using ONETEP version 4.5 with the PBE xc functional.[136] The calculation was spin-unpolarised, with an energy cut-off of 897 eV. Each nitrogen, carbon, oxygen, and sulfur atom had four NGWFs; hydrogen atoms, one. All NGWFs had a cut-off radius of $8.0\,a_0$. PAW potentials from the JTH dataset were used.[137] The protein was immersed in an implicit solvent parametrised to mimic water ($\varepsilon_\infty = 80$, $n_0 = 0.00035 a_0^{-3}$, and $\beta = 1.3$ following Ref. 138). The geometry was optimised using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, updating the solvent cavity (being a function of the charge density) every five BFGS steps.

Table 2.1: C_α RMSD values with respect to the crystal structure (Å) for restricted Hartree-Fock (RHF), ωPBEh, BLYP (all calculated with TeraChem and using a variety of basis sets[3,139,140]), AMBER (a classical molecular dynamics force field),[141] and ONETEP. All results bar the ONETEP entry are from Ref. 135.

		RHF		
MINI	MINI+D	STO-3G	3-21g	6-31g
0.51	0.41	0. 35	0.38	0.45
		ωPBEh		
MINI	MINI+D	STO-3G	3-21g	6-31g
0.60	0.70	0. 33	0.45	0.44
BLYP 6-31g	AMBER	ONETEP		
0.54	0.55	0. 139		

The resulting optimised structure is compared against the XRD structure in Figure 2.10. In order to quantitatively compare our results to those of Kulik *et al.*, we measured the root-mean-square deviation

$$\mathrm{RMSD}(\{\mathbf{x}_i\},\{\mathbf{y}_i\}) = \sqrt{\frac{1}{N}\sum_{i}^{N}|\mathbf{x}_i-\mathbf{y}_i|^2} \tag{2.46}$$

of the positions of the N alpha carbons of the optimised structure $\{\mathbf{x}_i\}$, relative to the XRD crystal structure $\{\mathbf{y}_i\}$ (Table 2.1). Evidently, ONETEP achieves a vastly superior RMSD to all of the other approaches.

As a brief aside, note that this sort of capability may be of assistance in analysing experimental structures obtained using cryogenic electron microscopy (cryo-EM). Cryo-EM is a x-ray crystallographic technique which (unlike traditional x-ray diffraction) does not depend on the structure being crystalline. This is a substantial advantage, as crystallising proteins is a very challenging process, and many proteins of interest have defied experimentalists' efforts to do so.

Historically, cryo-EM has struggled to achieve sufficiently high resolutions for structural determination. But in recent years, this has been changing, with a number of groups having achieved resolutions as low as 2–3 Å.[142–145] While impressive, these numbers are not quite low enough for atomic resolution. This is where computational approaches have made substantial ground: Monte Carlo sampling methods have been shown to be capable of refining 3–4 Å-resolution XRD structures to within approximately 0.5 Å of high-resolution measurements of the same structures (namely, as implemented in the Rosetta software suite).[146] ONETEP could hypothetically assist in this refinement process. The level of structural refinement performed by the Rosetta suite — involving complete reorientation of side-chains — is not going to be rivalled by ONETEP, whose geometry-optimisation routine will be far slower and will not bring about such substantial conformational changes, but will instead find a local minimum in whichever energy basin the structure happens to start in. That said, ONETEP could be used to polish Rosetta-refined structures. It may be that some of the conformational isomers put forward by Rosetta have enthalpies that differ by orders of k_BT — a problem that would go undetected by Rosetta (which uses a simplistic energetic scoring function), but which ONETEP would reveal immediately.

2.5 Conclusions; the challenge of strong correlation

As this chapter has established, DFT is a powerful tool — especially when reformulated to be linear-scaling. But while linear-scaling DFT does overcome the challenge of size, there remain other challenges to the accurate electronic structure prediction of metalloproteins. For many systems, xc functionals such as the LDA and generalised gradient approximations of Subsection 2.1.3 adequately describe their electronic behaviour. However, for so-called "strongly-correlated" systems, such as the OEC, this is not the case.

Formally, correlation can be defined as the physics/chemistry due to multi-determinantal wavefunctions (that is, beyond-HF).[†] Correlation effects encompass dispersion interactions, particle lifetimes, magnetism, satellites, collective excitations, the Kondo effect, Mott insulators and metal-insulator transitions.[147]

Many of the systems that exhibit these phenomena contain transition element or rare-earth atoms whose $3d$- or $4f$-electron shells are partially filled. Electrons in these shells are in especially close proximity with one another,[‡] and thus their interaction is too pronounced to be adequately described by DFT, which can provide even qualitatively incorrect descriptions of the electronic structure. For example, DFT often yields magnetic moments inconsistent with experiment,[148] predicts some insulators to be metallic,[149,150] and yields equilibrium volumes dramatically different to experiment.[151] DFT also fails to capture important dynamic properties that are enhanced by strong correlation, such as satellite peaks in photoemission spectra,[152,153] magnetic behaviour above the Curie temperature,[153] and the physics of heavy fermion materials.[154]

In the case of the metalloproteins, problems arise from the partially occupied $3d$ orbitals of the transition metals (iron in the case of haem, copper for haemocyanin, and manganese for the OEC). Existing computational studies of the OEC almost exclusively use more advanced xc functionals than the LDA or GGAs. Common approaches are to incorporate Hartree-Fock (HF) exchange via hybrid functionals[58] and DFT + U.[57] These will outperform semi-local functionals, but they are certainly not the most accurate methods available. DFT + U will be explained in the next chapter; hybrid functionals directly include HF exchange into the energy functional

$$E_x^{\mathrm{HF}} = -\frac{1}{2}\sum_{ij\sigma}\int\int d\mathbf{r}\, d\mathbf{r}'\, \frac{\psi_{i\sigma}^*(\mathbf{r})\psi_{j\sigma}(\mathbf{r})\psi_{j\sigma}^*(\mathbf{r}')\psi_{i\sigma}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{2.47}$$

which is mixed with a semi-local xc functional(s). The most prevalent hybrid xc functional is B3LYP:[155,156]

$$E_{xc}^{\mathrm{B3LYP}} = 0.2E_x^{\mathrm{HF}} + 0.08E_x^{\mathrm{LDA}} + 0.72E_x^{\mathrm{B88}} + 0.81E_c^{\mathrm{LYP}} + 0.19E_c^{\mathrm{VWN}} \tag{2.48}$$

where B88 is a GGA,[157] LYP and VWN are correlation functionals,[158,159] and the various coefficients have been derived by fitting to experimental data. B3LYP is widely used and has been

[†]This is different to what a statistician would call 'correlation': the Slater determinant wavefunction enforces the Pauli principle, thereby introducing "correlation" between any two electrons with the same spin, which cannot be in the same single-particle state.

[‡]Because they have no lower $l = 2/l = 3$ orbitals to be orthogonal to, these orbitals are highly localised about the nucleus.

undeniably successful, but there are several issues with its approach. Firstly, the fitting means the calculations are no longer *ab initio* and in principle the functional could fail for systems that are dissimilar to those used during fitting. Secondly, hybrid functionals benefit from cancellation of errors: for example, semi-local DFT underestimates band gaps while HF overestimates band gaps. This makes further development difficult. Finally — and most importantly for our purposes — evaluating Equation 2.47 scales poorly,[†] which prevents hybrid DFT calculations from being performed on sufficiently large cluster models.

The following chapters will explore and develop methods for treating strong electronic correlation. The challenge throughout is that more accurate theories are also more computationally intensive; if the methods and work-flows that we consider are to be of practical use, they must always be integrated within a linear-scaling DFT framework.

[†]It scales as $\mathcal{O}(M^4)$ for a Gaussian basis set with M basis functions; in plane-wave and wavelet representations, the HF exchange can be reformulated to allow its evaluation in $\mathcal{O}(N^3 \log N)$, with N being the number of electrons.

Chapter 3

Density functional theory-plus-U

3.1 Introduction

The first approach this work will explore for improving the treatment of correlation in metalloproteins is the density functional theory-plus-U (DFT + U) method. Inspired by the seminal Hubbard model, this scheme involves adding Hubbard-model-like terms to the DFT framework, in an approach originally designed to capture Mott-Hubbard physics in transition-metal oxides.[160–162] The corrective terms are applied to preselected spatially localised subsystems that are expected to exhibit strong correlation — for instance, the $3d$ orbitals of each Mn atom in the OEC — while the rest of the system is treated with semi-local DFT.

3.1.1 The Hubbard model

According to conventional band structure calculations (where the electron interactions are ignored) a system with a half-filled valence band is metallic. However, the presence of strong Coulombic interactions between electrons can give rise to a phase transition to an insulating phase, known as the Mott metal-insulator transition,[163] which explains (for example) why nickel oxide is not metallic.

The Mott transition is captured by the Hubbard model Hamiltonian:

$$\hat{H} = -t \sum_{\langle mn \rangle} \sum_{\sigma} \hat{c}^{\dagger}_{m\sigma} \hat{c}_{n\sigma} + U \sum_{m} \hat{n}_{m\uparrow} \hat{n}_{m\downarrow}, \tag{3.1}$$

where $\hat{c}_{m\sigma}$ is the annihilation operator for site m and spin σ, and $\hat{n}_{m\sigma} = \hat{c}^{\dagger}_{m\sigma} \hat{c}_{m\sigma}$ is the number operator. U is the on-site repulsion and t is the inter-site hopping parameter (where the sum is over nearest neighbour pairs). Of course, the derivation of this model — contained in full in Appendix A.1 — involves several assumptions. Key among these is the fact that electrostatic interaction is only important for electrons of the same spin on the same site.[†] This leaves us with

[†]While as physicists we are very familiar with the Hubbard model, it is worth drawing attention to how counter-intuitive this assumption is: after all, the unscreened electrostatic interaction is exceptionally long-ranged. In order to realise that an on-site Hubbard parameter is a sensible approximation, we rely on theoretical developments such as Thomas-Fermi screening and the Yukawa effective potential,[164,165] Lindhard screening and Friedel oscillations,[166] and Fermi liquid theory (Landau and others),[167] which explains why the Hubbard model only emerged 34 years after the Heisenberg Hamiltonian.

a comparatively simple Hamiltonian, but one which is notoriously difficult to solve.† Despite this, it has been the subject of intense research because of its ability to exhibit some very non-trivial physics: magnetism, superconductivity, Tomonaga–Luttinger liquids, and, of course, the Mott transition.[169]

The behaviour of the Hubbard model is dictated by competition between U and t. Consider the case of the half-filled system where each site has an average occupancy of one. In the $U \ll t$ limit, the local interactions are only a weak perturbation to the tight-binding model and thus the system will be metallic. In the $U \gg t$ limit, it is unfavourable to doubly occupy any site. This inhibits migration of electrons through the system and makes the system an insulator. In between these two states lies the Mott transition.

3.1.2 The DFT + U correction

In an attempt to have DFT inherit the Hubbard model's ability to describe correlation, one can massage the Hubbard-model formalism so that it can be incorporated into the framework of DFT. By doing so we arrive at the DFT + U correction:

$$E_U[\hat{n}^{I\sigma}] = \sum_{I\sigma} \frac{U^I}{2} \mathrm{Tr}\left[\hat{n}^{I\sigma}(1 - \hat{n}^{I\sigma})\right], \tag{3.2}$$

which is added to the DFT energy functional (here I have used the rotationally-invariant, simplified form).[150,160–162,170,171] The full derivation is contained in Appendix A.2. The density operators $\hat{n}^{I\sigma} = \hat{P}^I \hat{\rho}^\sigma \hat{P}^I$ are projections of the (spin-dependent) KS density operator onto subspaces (indexed I). The projectors $\hat{P}^I = \sum_m |\varphi^I_m\rangle\langle\varphi^I_m|$ are typically constructed from atom-centred, fixed, spin-independent, localised, and orthonormal orbitals φ^I_m (although they may be non-orthogonal[172] and self-consistent[173]). The U^I are externally-defined parameters that determine the strength of the energy corrections for each subspace. The corresponding correction to the KS potential is given by

$$\hat{V}_U = \sum_{I\sigma mn} U^I |\varphi^I_m\rangle \left(\frac{1}{2} - n^{I\sigma}_{mn}\right) \langle\varphi^I_n|. \tag{3.3}$$

This is attractive or repulsive for occupancy matrix eigenvalues greater than or less than one-half, respectively. In the absence of any significant self-consistent response, this will penalise non-integer occupancies of the subspaces, opening an energy gap of order U between any occupied and unoccupied KS orbitals which have a significant overlap with the Hubbard projectors.

In order to correct interactions between unlike spins, DFT + U can be extended to become DFT + U + J.[174–177] This involves a second correction to the total energy,

$$E_J[\hat{n}^\sigma] = \sum_{I\sigma} \frac{J^I}{2} \mathrm{Tr}\left[\hat{n}^{I\sigma}\hat{n}^{I-\sigma}\right], \tag{3.4}$$

where this correction is parametrised by the additional Hund's coupling constants J^I. Addition-

†It has only been solved analytically in one dimension, where there is no Mott transition.[168]

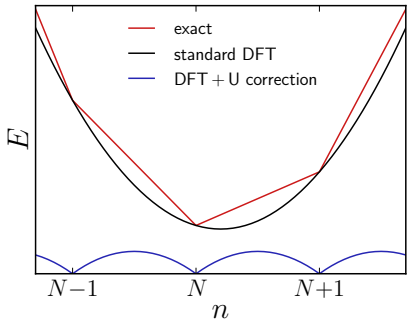


Figure 3.1: Sketch of total energy as a function of number of electrons for a generic atomic system in contact with a reservoir. The Hubbard correction (blue), when added to the result of the LDA (black), can potentially reproduce the piecewise-linear behaviour of the exact energy (red).

ally, the U in Equations 3.2 and 3.3 becomes $U_{\text{eff}} = U - J$. Note that since $\hat{n}^{\sigma}_{mm'} = \hat{c}^{\dagger}_{m\sigma}\hat{c}_{m'\sigma}$, this correction $\text{Tr}\left[\hat{n}^{I\sigma}\hat{n}^{I-\sigma}\right] = n^{I\sigma}_{mm'}n^{I-\sigma}_{m'm}$ can be understood to address "orbital exchange" between electrons of opposite spin (that is, an up electron going from orbital m to m' and a down electron going from m' to m).[176,177]

3.1.3 DFT + U as a correction to self-interaction error

A number of steps in the historical derivation of Equation 3.2 from the Hubbard model are hard to rigorously justify, and consequently this interpretation of DFT + U (that is, as an embedding of the Hubbard model in DFT) has fallen out of favour. Nevertheless, the technique itself remains popular thanks to a reinterpretation of the energy correction term of Equation 3.2 due to Cococcioni, Kulik, Marzari, and co-workers.[178,179]

One of the most prominent failures of many xc functionals is that they do not properly correct for the self-interaction in the Hartree term. Self-interaction error (SIE) — or more generally "delocalisation error"[180,181] — manifests itself as a spurious curvature in total energies with respect to total electron number (see the black curve in Figure 3.1), where instead there should be a derivative discontinuity at integer numbers of electrons and linear behaviour at fractional numbers (red curve).[182] To see why this is, consider an atom in contact with a reservoir of electrons, and let the energy of that atom be given by $E(N)$, where N is the occupancy of the atom. If there is a fractional number $N + \omega$ of atoms in the atomic system, where $0 \leq \omega \leq 1$, the system is a statistical mixture of the N and $N + 1$ states and its energy is given by $(1 - \omega)E(N) + \omega E(N + 1)$ — that is, as a function of N, E has a slope that is piecewise constant.[183] The LDA and GGAs do not reproduce this behaviour: instead, they give rise to an unphysical curvature associated with incorrect treatment of self-interaction for the partially occupied KS orbitals. This failure is closely related to approximate DFT's well-documented underestimation of the band gap.[184,185] But while the origins of the SIE are well understood, it remains a challenge to avoid its introduction when constructing xc functionals, even if exact exchange is incorporated.[180]

Now that we have identified this shortcoming of DFT, let us reconsider the DFT + U formalism. Equation 3.2 can be seen to penalise partial occupancies of the correlated subspaces:

in the basis of localised orbitals $\psi_m^{I\sigma}$ that diagonalise the subspace occupancy matrices such that $\hat{n}^{I\sigma}\psi_m^{I\sigma} = \lambda_m^{I\sigma}\psi_m^{I\sigma}$, the Hubbard correction becomes $\sum_{I\sigma m} U^I \lambda_m^{I\sigma}(1 - \lambda_m^{I\sigma})/2$, which penalises non-integer occupancies of these orbitals $\psi_m^{I\sigma}$ (see the blue curve of Figure 3.1). Hubbard corrections therefore counteract the spurious SIE curvature — and thus, with well-chosen values for $\{U^I\}$, DFT + U calculations may cancel the SIE that is present (although this is not guaranteed).[186] Note that the correction to the total energy vanishes at integer occupancy matrix eigenvalues, where the xc functional is assumed to be correct. In this interpretation, the Hubbard subspaces become localised subspaces in which the SIE is to be selectively addressed.

A substantial advantage of DFT + U (+ J) over other methods that address the SIE (for example, SIC-LSDA,[187–189] Fermi orbital self-interaction correction,[190–192] and Koopman's compliant functionals[193,194]) is its small computational cost: once any Hubbard parameters have been determined, the overhead for incorporating the additional potential and energy terms is insignificant compared to the cost of the DFT calculation itself.[195]

3.1.4 Conventional linear response

In order to apply a Hubbard correction, one must select an appropriate value for the parameters U^I. This can be done pragmatically by picking values on empirical grounds — that is, chosen so that certain system characteristics are reproduced (for example, ionic geometries,[196–199] band gaps,[197–200] and formation enthalpies[199,201,202]). While this approach has seen some success,[203,204] it does not guarantee that the chosen U will correct the SIE energy curvature to the greatest extent achievable, or result in an improved description of other system properties, and this method is not even possible where there is a lack of reliable experimental or higher-level computational data. Furthermore, it opens up DFT + U to criticism that it is not truly an *ab initio* method.

An alternative approach for determining U is the linear response method developed by Cococcioni and de Gironcoli,[178] which built upon the earlier linear response scheme of Pickett and co-workers,[171] and shares many aspects with the constrained LDA approach of Aryasetiawan and co-workers.[205] In this approach, DFT calculations are performed subject to a perturbing potential $\delta\hat{v}_{\text{ext}} = dv_{\text{ext}}^J \hat{P}^J$ confined to the J^{th} Hubbard subspace, for a range of scalar values $\{dv_{\text{ext}}^J\}$. The density operator's response to these perturbations is given by the response operator $\hat{\chi}$:

$$\delta\hat{\rho} = \hat{\chi}\delta\hat{v}_{\text{ext}}. \tag{3.5}$$

The occupancy of the I^{th} Hubbard subspace will change by

$$dn^I = \text{Tr}\left[\hat{P}^I\delta\hat{\rho}\right] = \text{Tr}\left[\hat{P}^I\hat{\chi}\hat{P}^J\right] dv_{\text{ext}}^J \tag{3.6}$$

and thus we can define the projected response matrix[206]

$$\chi_{IJ} \equiv \frac{dn^I}{dv_{\text{ext}}^J} = \text{Tr}\left[\hat{P}^I\hat{\chi}\hat{P}^J\right]. \tag{3.7}$$

A value for U that corresponds to the screened response of the system is given by

$$U^I = \left(\chi_0^{-1} - \chi^{-1}\right)_{II} \tag{3.8}$$

where χ_0 is the response of the non-interacting system, which must be separately measured and removed from the Hubbard correction.[177,207,208] There is also scope here for calculation of off-diagonal terms $V_{IJ} = \left(\chi_0^{-1} - \chi^{-1}\right)_{IJ}$. Introducing corrective inter-site terms is known as DFT + U + V.[209,210]

Satisfyingly, the determination of U via linear response removes any possible arbitrariness of the Hubbard correction: the U parameter is a well-defined property of the system that can be unambiguously measured in theory, rather than a parameter that can be tuned "to taste".[177,179]

Recently the idea of calculating U and J to best emulate subspace-projected KS exact exchange[175] has been further advanced.[211] But because I wish to cancel the systematic errors of approximate DFT[181,182] to the extent possible using functionals of the DFT + U form, we choose to instead develop the linear response formalism that has been shown to successfully achieve this,[171,178,179,212] and that does not incorporate any theory or model (*e.g.* Fock exchange) beyond what is already ordinarily present.

3.1.5 Problems

There are some aspects of the linear response methodology that pose issues. Firstly, delocalisation error is associated with fractional total charge, but the DFT + U functional of Equation 3.2 corrects fractional occupation for each spin channel separately. Conventional linear response, meanwhile, perturbs both spin channels simultaneously. These discrepancies in how we treat spin channels warrant investigation.

Secondly, measuring the non-interacting response χ_0 is not straightforward. The common practice is to follow the example of Ref. 178, and calculate χ_0 via the first iteration of the KS equations during a self-consistent field (SCF) calculation — that is, the response is measured following the initial charge redistribution introduced by the perturbation but before the KS potential is updated. This approach is impractical to implement in codes that use a direct-minimisation procedure of the total energy with respect to the density, KS orbitals, or density matrix. This represents a substantial number of packages, including ONETEP,[22] CONQUEST,[213,214] SIESTA,[215,216] BigDFT,[217] OpenMX,[218] and CP2K[219] (albeit that in some of these the self-consistent field technique is also available). In direct-minimisation, updating the density and potential are not nested separately, so χ_0 cannot be calculated in the manner prescribed above.

Finally, in the case of closed-shell systems, linear response approaches tend to dramatically overestimate U and suffer from numerical instabilities.[220–223] It would be good to overcome, or at least to better understand, these failures.

3.2 Minimum-tracking linear response

Ref. 212 presented an alternative to the established SCF linear response approach for calculating the Hubbard parameters: so-called "minimum-tracking" linear response. This approach is suited for calculating U in direct-minimisation codes, as it makes no reference to the first iteration of the KS loop. In this section I will expand upon this formalism.

The minimum-tracking linear response approach is largely equivalent to SCF linear response, but its derivation centres on the ground-state density for each value of the perturbing potential. As with the SCF approach, a perturbing potential $d\hat{v}_{\text{ext}} = dv^J_{\text{ext}}\hat{P}^J$ is applied to the J^{th} Hubbard subspace. The response of the projected KS potential is given by the chain rule

$$\frac{dv^I_{\text{KS}}}{dv^J_{\text{ext}}} = \frac{dv^I_{\text{ext}}}{dv^J_{\text{ext}}} + \frac{dv^I_{\text{Hxc}}}{dv^J_{\text{ext}}} = \frac{dv^I_{\text{ext}}}{dv^J_{\text{ext}}} + \sum_K \frac{dv^I_{\text{Hxc}}}{dn^K}\frac{dn^K}{dv^J_{\text{ext}}}, \tag{3.9}$$

where the final step follows because while the external potential acting on site J will change the density matrix everywhere, the N-site Hubbard model only sees the N subspace density matrices. Screening due to the residual bath is incorporated within the total derivatives. The projections of one-body operators are given by $O^I = \text{Tr}[\hat{P}^I\hat{O}]/\text{Tr}[\hat{P}^I]$.

Defining the projected Hartree-plus-xc kernel $f_{IJ} \equiv dv^I_{\text{Hxc}}/dn^J$, the inverse dielectric function $(\varepsilon^{-1})_{IJ} \equiv dv^I_{\text{KS}}/dv^J_{\text{ext}}$, and the overlap matrix $\Omega_{IJ} \equiv dv^I_{\text{ext}}/dv^J_{\text{ext}}$, Equation 3.9 becomes

$$\varepsilon^{-1} = \Omega + f\chi \Longrightarrow f = \left(\varepsilon^{-1} - \Omega\right)\chi^{-1}. \tag{3.10}$$

Finally, U can be equated with the projected Hartree-plus-xc kernel, with the residual bath screening in the background.[177] (See Appendix A.3 for a detailed discussion on why we can discard the rest of the system, and Appendix A.4 for a proof of the oft-repeated result that if we include the bath, we can still calculate the difference of the two now non-invertible matrices χ and χ_0.) This yields

$$U^I = \left[\left(\frac{dv_{\text{KS}}}{dv_{\text{ext}}} - 1\right)\left(\frac{dn}{dv_{\text{ext}}}\right)^{-1}\right]_{II}. \tag{3.11}$$

From here on in, I will assume that the overlap matrix $\Omega = \delta_{IJ}$ for simplicity. When Hubbard projectors from different atoms overlap this may become an approximation. I will also reserve f for the matrix measured via linear response, and U for the parameter to be subsequently used in a DFT + U calculation. This distinction will become important.

Equation 3.11 is nothing more than a reformulation of Equation 3.8. We can identify the interacting and non-interacting response matrices

$$\chi_{IJ} = \frac{dn^I}{dv^J_{\text{ext}}}; \tag{3.12a}$$

$$(\chi_0)_{IJ} = \left[\frac{dn}{dv_{\text{ext}}}\left(\frac{dv_{\text{KS}}}{dv_{\text{ext}}}\right)^{-1}\right]_{IJ}. \tag{3.12b}$$

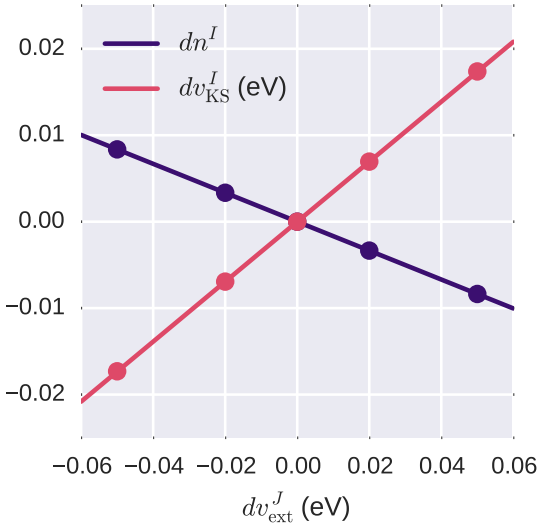


Figure 3.2: A typical linear response plot. Each pair of points represents an individual DFT calculation with a perturbing potential $\delta\hat{v}_{\mathrm{ext}} = dv^J_{\mathrm{ext}}\hat{P}^J$, and the resulting response of the projected density dn^I and KS potential dv^I_{KS} . The slopes of these lines correspond to entries of χ and ε^{-1}. These data have been taken from calculations on $[\mathrm{Cr}(\mathrm{H_2O})_6]^{3+}$, which are covered in detail in Section 3.3.

In this framework, we can see that the removal of the non-interacting response can be rigorously justified as a consequence of the Dyson equation, with U being a measure of the net interaction. These definitions are nothing but a special case of standard linear response theory for DFT (see Appendix A.5). It is crucial that the non-interacting response is calculated as the product of χ and ε, rather than dn^I/dv^J_{KS} directly. dn^I/dv^J_{KS} is both conceptually and numerically arbitrary with respect to the choice of external potential, and so its direct use must be avoided.

Figure 3.2 demonstrates the calculation of elements of χ and ε^{-1} from a typical set of linear response calculations.

Both minimum-tracking and conventional SCF linear response rely on the same external perturbation, and both make use of the Dyson equation. They only differ in their definition of the non-interacting response and the set of densities used in its calculation. In the minimum-tracking procedure, χ_0 is constructed from ground-state densities of the perturbed system, and thus the resulting U is strictly a ground-state property. This is obviously not the case for the SCF approach: there, χ_0 is calculated in reference to an unconverged density and thus the resulting U is not a local property of the ground-state density landscape (but is still a well-defined property of the ground-state KS eigensystem). This distinction is intriguing and worthy of further investigation, but possibly numerically inconsequential in practice.

Already, the minimum-tracking construction reveals an interesting property of the projected χ_0 (and hence f): it is not necessarily symmetric. This is because χ_0 as defined in Equation 3.12b incorporates the total derivative of the potential, which is itself a partial derivative. While the bare χ_0 is certainly symmetric, the response matrices that we deal with here are always screened by the background, and the screening depends on the subspace being perturbed. (In general, χ_0 should not be symmetrised before inversion, even if the resulting U matrix will be.) This observation will also hold for SCF linear response, since it also correctly goes beyond the

symmetric result of first-order perturbation theory.

3.2.1 Accounting for spin

Many transition metal subspaces, such as those found in metalloproteins, exist in a high-spin state, and consequently the response of their two different spin channels will be wildly different. However, all of the above formalism has been blind to this, with both spin channels perturbed during a linear response calculation and their collective response measured. In contrast, the Hubbard energy functional (Equation 3.2) treats different spin channels entirely separately; indeed, spin and sites are treated equivalently, with spin/site indices being totally interchangeable. This raises the question: what happens to the response and interaction parameters if we were to be consistent, and fine-grain linear response down to the level of spin?

In the minimum-tracking formulation it is straightforward to consider spin degrees of freedom. Response matrices become rank-four tensors

$$\chi_{IJ}^{\sigma\sigma'} = \frac{dn^{I\sigma}}{dv^{J\sigma'}}, \tag{3.13}$$

and to measure these elements via linear response, we must perturb spin channels individually. (Practically, this is implemented as a combination of two potentials: a uniform shift applied to both spin-channels and a spin-splitting potential.)

This extension has several consequences. Spin-specific response functions can be visualised by flattening rank-four tensors down to rank-two ones: for example, a two-site system would have response matrices of the form

$$\chi = \begin{pmatrix} \chi_{11}^{\uparrow\uparrow} & \chi_{11}^{\uparrow\downarrow} & \chi_{12}^{\uparrow\uparrow} & \chi_{12}^{\uparrow\downarrow} \\ \chi_{11}^{\downarrow\uparrow} & \chi_{11}^{\downarrow\downarrow} & \chi_{12}^{\downarrow\uparrow} & \chi_{12}^{\uparrow\downarrow} \\ \chi_{21}^{\uparrow\uparrow} & \chi_{21}^{\uparrow\downarrow} & \chi_{22}^{\uparrow\uparrow} & \chi_{22}^{\uparrow\downarrow} \\ \chi_{21}^{\downarrow\uparrow} & \chi_{21}^{\downarrow\downarrow} & \chi_{22}^{\downarrow\uparrow} & \chi_{22}^{\downarrow\downarrow} \end{pmatrix} = \begin{pmatrix} (\chi^{\sigma\sigma'})_{11} & (\chi^{\sigma\sigma'})_{12} \\ (\chi^{\sigma\sigma'})_{21} & (\chi^{\sigma\sigma'})_{22} \end{pmatrix}. \tag{3.14}$$

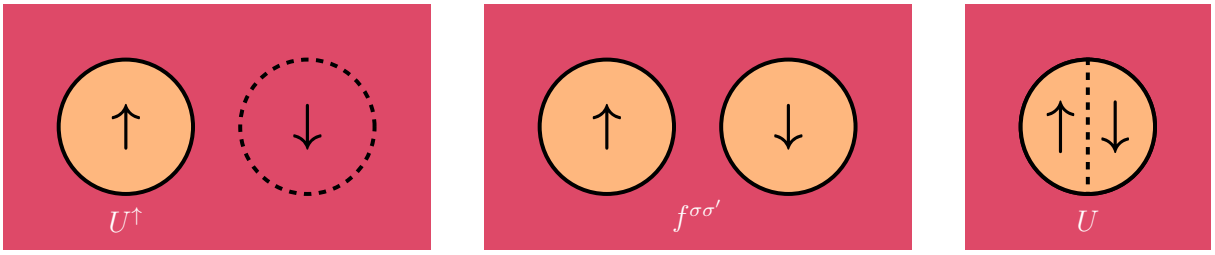
This is not simply aesthetic: it means we are treating spin and atom indices on the same footing, like the DFT + U functional does.

We can construct different models based on how we perform the inversion of this matrix (such as in Equation 3.11): either (1) point-wise inversion, which decouples both sites and spin; (2) atom-wise inversion, with each 2×2 block inverted individually, decoupling sites but not spins; or (3) invert the full matrix, leaving all sites and spins coupled. I will work through each of them in turn.

Point-wise inversion

The Hubbard parameters in this case are screened by the opposite spin on the same site (Figure 3.3a). In this case, Equation 3.11 separates into an independent equation for each atom:

$$f^{\sigma\sigma} = \frac{dv_{\mathrm{KS}}^{\sigma}}{dn^{\sigma}} - \frac{dv_{\mathrm{ext}}^{\sigma}}{dn^{\sigma}} = \frac{dv_{\mathrm{Hxc}}^{\sigma}}{dn^{\sigma}}. \tag{3.15}$$



(a) (b) (c)

Figure 3.3: Schematic diagram illustrating which subspaces screen the Hubbard parameters (pink) and which do not (orange). Individual Hubbard sites are represented by solid circles. Point-wise inversion (a) effectively treats our system as a one-site Hubbard model connected to a bath, where the bath includes the opposite-spin subspace of the same site. Atom-wise inversion (b) is effectively a two-site system connected to a bath. Finally, in conventional linear response (c) both spin channels on a given atom are treated as a single Hubbard site.

We have dropped the atomic indices for brevity. This simplification affords some numerical cancellation of errors, since inversion is no longer performed. The off-diagonal components of the matrix $f^{\sigma\sigma'}$ are not meaningful in this case. The conventional DFT + U functional requires a spin-independent U; for this we must average the spin-up and spin-down components:

$$U = \frac{1}{2}\left(f^{\uparrow\uparrow} + f^{\downarrow\downarrow}\right). \tag{3.16}$$

This will henceforth be referred to as "averaged 1×1". There is also the option to avoid this approximation and apply a different value of U to each spin channel: $U^{\sigma} = f^{\sigma\sigma}$ ("1×1").

It is interesting to note that Shishkin and Sato[224] have previously advocated removing the off-diagonal components of site-indexed response matrices. This was motivated by the fact that these components were negligible so removing them did not alter the resulting Hubbard parameters. Here, however, the off-diagonal components components correspond to coupling between spin channels on the same atom. These components are sizeable and neglecting them appreciably alters Hubbard and Hund's parameters, as we will see.

Atom-wise inversion

In atom-wise inversion, screening from both the bath and other sites is present in the response matrices, but the resulting $f = \chi_0^{-1} - \chi^{-1}$ is *bare* with respect to inter-spin interactions on the same atom as it is removed by the inversion of the spin-indexed response (Figure 3.3b). Employing this approach amounts to assuming inter-spin interactions will be corrected separately *i.e.* with a $+J$ functional. (This is because in the absence of such a correction, a spin-screened U would be necessary.)

Equation 3.11 reduces to

$$f^{\sigma\sigma'} = \left[\left(\frac{dv_{\mathrm{KS}}}{dv_{\mathrm{ext}}} - 1\right)\left(\frac{dn}{dv_{\mathrm{ext}}}\right)^{-1}\right]^{\sigma\sigma'} \tag{3.17}$$

where each term is a two-by-two matrix indexed by spin channel, and if there are N atoms there are N such equations. For practical use in DFT + U + J, f can be related to the scalar Hubbard

parameter U that, in the minimum-tracking linear response formalism, is defined by

$$U = \frac{1}{2}\frac{dv^{\uparrow}_{\mathrm{Hxc}} + dv^{\downarrow}_{\mathrm{Hxc}}}{d(n^{\uparrow} + n^{\downarrow})} \approx \frac{1}{2}\frac{f^{\uparrow\uparrow}\delta n^{\uparrow} + f^{\uparrow\downarrow}\delta n^{\downarrow} + f^{\downarrow\uparrow}\delta n^{\uparrow} + f^{\downarrow\downarrow}\delta n^{\downarrow}}{\delta(n^{\uparrow} + n^{\downarrow})}. \tag{3.18}$$

In other words, U is given by a weighted average of the elements of $f^{\sigma\sigma'}$, where elements are weighted according to the extent to which the spin-up and -down densities would respond to a perturbation. In the case of spin-unpolarised systems, the two densities would respond equally ($dn^{\uparrow} = dn^{\downarrow}$) and Equation 3.18 simplifies to

$$U = \frac{1}{2}(f^{\uparrow\uparrow} + f^{\uparrow\downarrow}) \tag{3.19}$$

(where we have also taken advantage of the symmetries $f^{\uparrow\uparrow} = f^{\downarrow\downarrow}$ and $f^{\uparrow\downarrow} = f^{\downarrow\uparrow}$). Such a straightforward simplification is not possible for spin-polarised systems. Instead, we must account for the possibility of different spin-up and -down density responses, and we must adopt some approximation if we are to reduce the kernel down to a scalar U.

I propose two alternative approximations. The first, more naïve approach, is simply to approximate this as

$$U = \frac{1}{4}\left(f^{\uparrow\uparrow} + f^{\uparrow\downarrow} + f^{\downarrow\uparrow} + f^{\downarrow\downarrow}\right) \tag{3.20}$$

which I will refer to as "simple 2×2". (This is not as bad as it might seem, for a start, it is exact in the spin-unpolarised limit. But more fundamentally, the assertion that $dn^{\uparrow} = dn^{\downarrow}$ means that we are in effect measuring the curvature of the energy while constraining the magnetic moment to remain constant.)

To derive a more sophisticated approach, consider the ratio

$$\frac{dn^{\uparrow}}{dn^{\downarrow}} = \frac{\sum_{\sigma}\chi^{\uparrow\sigma}dv^{\sigma}_{\mathrm{ext}}}{\sum_{\sigma}\chi^{\downarrow\sigma}dv^{\sigma}_{\mathrm{ext}}}. \tag{3.21}$$

If we focus in particular on a perturbation of the form $dv^{\uparrow}_{\mathrm{ext}} = dv^{\downarrow}_{\mathrm{ext}}$ this simplifies to

$$\frac{\sum_{\sigma}\chi^{\uparrow\sigma}}{\sum_{\sigma}\chi^{\downarrow\sigma}} = \lambda_U. \tag{3.22}$$

Therefore, if we assert that in general $dn^{\uparrow}/dn^{\downarrow}$ can be approximated by λ_U then Equation 3.18 simplifies to

$$U = \frac{1}{2}\frac{\lambda_U(f^{\uparrow\uparrow} + f^{\downarrow\uparrow}) + f^{\uparrow\downarrow} + f^{\downarrow\downarrow}}{\lambda_U + 1}. \tag{3.23}$$

This approximation I will refer to as "scaled 2 × 2". It is a reasonable approximation, but it is certainly not rigorously justified, and is perhaps best interpreted *post hoc* (as the following section will explain).

With atom-wise inversion, Hund's parameters J can be directly calculated in an analogous manner to U: in place of Equation 3.18 we instead define, within the spin-polarised minimum-

tracking linear response formalism,

$$J = -\frac{1}{2}\frac{dv^{\uparrow}_{\mathrm{Hxc}} - dv^{\downarrow}_{\mathrm{Hxc}}}{d(n^{\uparrow} - n^{\downarrow})}. \tag{3.24}$$

Simple 2×2 yields

$$J = -\frac{1}{4}\left(f^{\uparrow\uparrow} - f^{\uparrow\downarrow} - f^{\downarrow\uparrow} + f^{\downarrow\downarrow}\right), \tag{3.25}$$

while for scaled 2×2 one can derive the analogous expression of Equation 3.25 in a very similar manner, except that the scaling factor λ_J is constructed with reference to a perturbation of the form $dv^{\uparrow}_{\mathrm{ext}} = -dv^{\downarrow}_{\mathrm{ext}}$ (that is, one that will most directly affect magnetic moments).[†]

$$J = -\frac{1}{2}\frac{\lambda_J(f^{\uparrow\uparrow} - f^{\downarrow\uparrow}) + f^{\uparrow\downarrow} - f^{\downarrow\downarrow}}{\lambda_J - 1}; \qquad \lambda_J = \frac{\chi^{\uparrow\uparrow} - \chi^{\uparrow\downarrow}}{\chi^{\downarrow\uparrow} - \chi^{\downarrow\downarrow}}. \tag{3.26}$$

Full inversion

Finally, in the case of full matrix inversion, the result is bare with respect to both inter-spin and inter-site interactions by the same logic. This implies that inter-atom interactions require, and are subject to, correction via a $+V$ term. This V term would be doubly spin-dependent, and it may need to be symmetrised with respect to the site indices to retain a Hermitian KS Hamiltonian for each spin. I will not explore this approach in this work.

We emphasise that including each of these successive terms (J and V) should not be viewed as systematic improvements. In the limit that corrective parameters are introduced within and between every single subspace (such that the corresponding screened interactions are removed) the entire system becomes effectively non-interacting. Corrective terms are only appropriate where the corresponding interactions dwarf all others.

3.2.2 Comparisons with the conventional scalar approach

Conventional linear response calculations do not treat spin channels separately (Figure 3.3c); for a single-site system χ, ε^{-1} and f would all be scalars. It is straightforward to relate the spin-indexed response matrices of the previous section to these scalars:

$$dn = dn^{\uparrow} + dn^{\downarrow} \approx \left[\sum_{\sigma\sigma'} \chi^{\sigma\sigma'}\right] dv_{\mathrm{ext}} \Longrightarrow \chi \approx \sum_{\sigma\sigma'} \chi^{\sigma\sigma'}. \tag{3.27}$$

Likewise

$$dv_{\mathrm{KS}} = \frac{1}{2}\left[dv^{\uparrow}_{\mathrm{KS}} + dv^{\downarrow}_{\mathrm{KS}}\right] \approx \frac{1}{2}\left[\sum_{\sigma\sigma'} \left(\varepsilon^{-1}\right)^{\sigma\sigma'}\right] dv_{\mathrm{ext}} \Longrightarrow \varepsilon^{-1} \approx \frac{1}{2}\left[\sum_{\sigma\sigma'} \left(\varepsilon^{-1}\right)^{\sigma\sigma'}\right]. \tag{3.28}$$

These two relations allow us to examine the role of spin-screening in scalar linear response. The Hubbard parameter obtained via spin-indexed, atom-wise inversion (scaled 2×2; Equation 3.23)

[†]Here, one could certainly use λ_U. Neither approximation is obviously superior to the other, and furthermore this choice does not drastically alter the resulting J. After all, Hund's coupling is much less susceptible to screening than Hubbard parameters.

can be rewritten as

$$U = \frac{1}{2}\frac{\sum_{\sigma\sigma'}(f\chi)^{\sigma\sigma'}}{\sum_{\sigma\sigma'}\chi^{\sigma\sigma'}} = \frac{1}{2}\frac{\sum_{\sigma\sigma'}(\varepsilon^{-1}-1)^{\sigma\sigma'}}{\sum_{\sigma\sigma'}\chi^{\sigma\sigma'}} = \frac{\varepsilon^{-1}-1}{\chi}. \tag{3.29}$$

This is nothing less than the scalar expression $U = \chi_0^{-1} - \chi^{-1}$, which is used in scalar linear response (the widely used standard approach). We may conclude that the conventional scalar approach and scaled 2×2 are entirely equivalent.

Therefore, Hubbard parameters obtained by spin-aggregated approaches are not screened by the opposite spin channel on the same site. Since they combine both like and unlike spin interactions (*cf.* Equation 3.23), they do *not* correspond to the like-spin-only interaction $U_{\text{eff}} = U - J$ (as implied elsewhere).[225] We could have anticipated this result: during a scalar linear response calculation there is no shift in the external potential difference between the two spin channels, so there is no external driver for changes in subspace spin polarisation (to first order).

We noted earlier that atom-wise inversion formally necessitates a Hund's correction, but such a correction is not usually included when the conventional linear response approach is employed. Given that these methods are equivalent, I argue that it is more consistent to include a Hund's exchange correction term (*e.g.* calculated using Equation 3.24) when using a Hubbard correction calculated in the conventional manner.

The precise functional form of the $+J$ correction needed is, however, the subject of ongoing research.[226] Recently, for example, Millis and co-workers demonstrated that spin-polarised DFT already possesses some degree of intrinsic exchange splitting, and they have argued convincingly that the contemporary form of the $+J$ correction can overestimate exchange splitting.[227] This finding is corroborated by my own results discussed later in this chapter (*e.g.* Table 3.7).

3.3 Application to a complete series of hexahydrated transition metals and manganese oxide

In the second half of this chapter, I will explore the ramifications of these theoretical developments on two test systems: hexahydrated transition metals, and manganese oxide. In these systems, all of the metal atoms have partially filled $3d$ sub-shells, and thus they are poorly described by local or semi-local xc functionals.[228–231] DFT + U may provide a more accurate description of these systems.[177,186,210,221]

Manganese oxide (MnO) has a rock salt structure. At low temperatures it is antiferromagnetic,[232] and has a band gap of approximately 4 eV that is substantially underestimated by semi-local functionals.[150] Conventional linear response calculations on MnO yield an excessively large Hubbard parameter ($U > 7$ eV).[233] Meanwhile, hexahydrated transition metals comprise of a central first-row transition metal ion surrounded by six water ligands in a tetragonal arrangement (Figure 3.4c). Depending on the electronic structure of the metal, these systems may exhibit Jahn-Teller distortion, resulting in an elongated tetragonal structure with two axial waters being slightly more distant than their four equatorial counterparts (Figure 3.4). While neither of these systems are metalloproteins, they can serve as good proving grounds for our

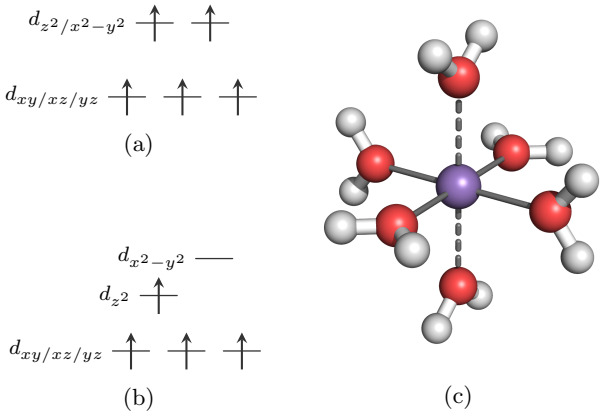


Figure 3.4: The ground state of the $3d$ electrons in (a) $[Mn(H_2O)_6]^{2+}$ and (b) $[Mn(H_2O)_6]^{3+}$. In both systems, the d_{xy}, d_{xz}, and d_{yz} orbitals (the so-called t_{2g} orbitals) have lower energy as they have lobes directed between the ligands (and hence less overlap with the ligand orbitals) compared to the d_{z^2} and $d_{x^2-y^2}$ orbitals (e_g). For the doubly-charged system, the system is symmetric and no Jahn-Teller splitting takes place. In the triply-charged system, the molecule distorts into a D_{2h} symmetry as shown in (c), with the axial bonds (dashed) fractionally longer than the equatorial bonds (solid).

Figure removed due to copyright. The original can be seen in Ref. 195 as Figure 1.

Figure 3.5: Scaling of DFT and DFT + U calculations for NiO nanoclusters of increasing size. Crucially, linear scaling is retained. Figure taken from Ref. 195.

theory. Hexahydrated transition metals bear some resemblance to metalloproteins such as the OEC, and in a loose sense, MnO could be seen as the "bulk limit" of the OEC core.

3.3.1 Computational details

All the following calculations in this chapter were performed using ONETEP[22,88,172,173,195,234,235] (version 4.3) using the PBE xc functional.[136]

One of the major advantages of DFT + U compared to other methods that improve upon the LDA and GGAs (such as hybrid functionals) is its minimal computational cost: once the parameter U has been determined, the penalty for moving from DFT to DFT + U is relatively insignificant, and does not compromise the linear-scaling of ONETEP (see Figure 3.5).

For MnO, a square super-cell containing 512 atoms was simulated under periodic boundary conditions without explicit k-point sampling (but recall in Subsection 2.3.3 I demonstrated how a super-cell can be used in place of k-point sampling). This is a non-diagonal super-cell[236] of the four-atom primitive cell, and gives an equivalent k-point sampling scheme that includes both

Z and Γ.[†] This is crucial for when we measure the band gap, as it is known to be Z-to-Γ. The lattice parameter was set to the experimental value of 4.445 Å.[237] The calculations were spin-polarised, with an energy cut-off of 1030 eV. Each Mn atom had ten NGWFs; O atoms, four. All NGWFs had a cut-off radius of $11.0\,a_0$.

For the hexahydrated metals, all calculations were spin-polarised, with an energy cut-off of 897 eV. Depending on the species, there were nine, ten, or thirteen NGWFs on the transition metal atom, four on each oxygen, and one on each hydrogen. All NGWFs had $14\,a_0$ cut-off radii. An Elstner dispersion correction[115,117] was applied, and electrostatics were treated using a padded cell and a Coulomb cut-off.[238]

For all the calculations, the Hubbard projectors were constructed from solving the neutral atomic problem subject to the pseudopotential of the species in question.[235] Most pseudopotentials were taken from the Rappe group pseudopotential library[239] although those for Co and

[†]We can relate a super-cell with basis vectors $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ to a primitive cell $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ via

$$\begin{pmatrix}\mathbf{A}_1\\ \mathbf{A}_2\\ \mathbf{A}_3\end{pmatrix} = S \begin{pmatrix}\mathbf{a}_1\\ \mathbf{a}_2\\ \mathbf{a}_3\end{pmatrix}. \tag{3.30}$$

An arbitrary k-point can be expressed in terms of either the reciprocal primitive lattice vectors or the reciprocal super-cell lattice vectors. These two alternative representations are related by

$$S\mathbf{q} = \mathbf{Q}. \tag{3.31}$$

As demonstrated by Lloyd-Williams and Monserrat,[236] if $\mathbf{Q}$ is an integer vector then this k-point is commensurate with the super-cell generated by S.

Our antiferromagnetic MnO system has a needle-like primitive cell containing four atoms along the $(1,1,1)$ axis, with primitive cell vectors

$$\mathbf{a}_1 = (2,1,1)l \tag{3.32a}$$
$$\mathbf{a}_2 = (1,2,1)l \tag{3.32b}$$
$$\mathbf{a}_3 = (1,1,2)l \tag{3.32c}$$

where $l = 2.22\,\mathring{A}$. One super-cell that samples both Γ and Z is generated by

$$S = \begin{pmatrix}6 & -2 & -2\\ -2 & 6 & -2\\ -2 & -2 & 6\end{pmatrix} \tag{3.33}$$

since for Γ

$$S \begin{pmatrix}0\\0\\0\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix} \tag{3.34}$$

(which is trivially true of all super-cells), and for Z

$$S \begin{pmatrix}1/2\\1/2\\1/2\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix} \tag{3.35}$$

since in both of these cases $\mathbf{Q}$ is a vector of integers. The super-cell generated by this S is

$$\begin{pmatrix}\mathbf{A}'_1\\ \mathbf{A}'_2\\ \mathbf{A}'_3\end{pmatrix} = \begin{pmatrix}8 & 0 & 0\\ 0 & 8 & 0\\ 0 & 0 & 8\end{pmatrix} l \tag{3.36}$$

which is nothing less than our 512-atom cubic cell.

A more subtle effect at play here is the weighting of k-points, which can change from super-cell to super-cell: the system will change if we change the weighting on different k-points. (Think of an infinite hydrogen chain: if most of the k-point weighting is at Γ then you have a very "bonding-like" density, whereas if most of the weighting is at the Brillouin zone edge, you will have twice as many lobes in the density due to "anti-bonding-like" k-dependent orbitals being over-sampled.) This effect will only be keenly felt if there are too few k-points.

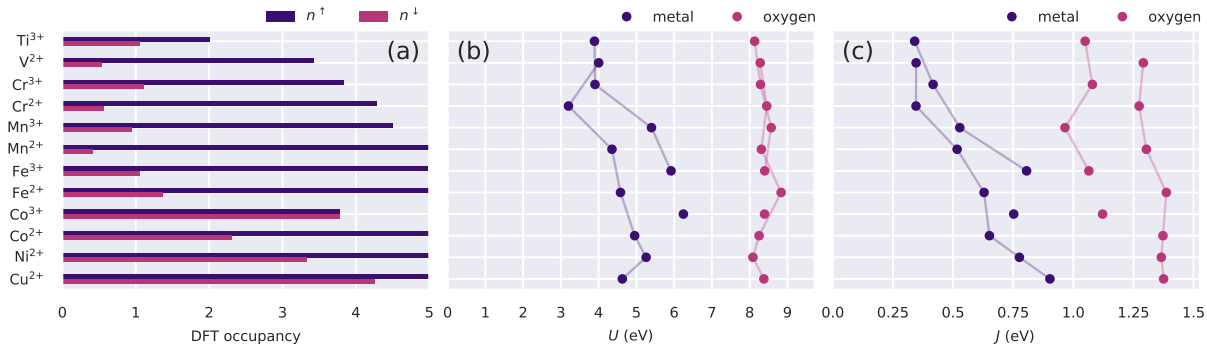


Figure 3.6: (a) The metal ion $3d$ subspace occupancies as given by DFT. The residual spin-down densities for the lighter metals are not formally what one would expect; a Hubbard correction should remedy this. (b) Hubbard parameters and (c) Hund's parameters as calculated via scaled 2×2 (equivalent to the scalar approach). Faint lines link the +2 systems/+3 systems to show the general trends. (Co^{3+}, being the only low-spin system, is not linked.)

Fe were generated in-house using OPIUM.[125,240–245] These were scalar relativistic pseudopotentials[246] with non-linear core corrections.[247] All DFT + U + J calculations used a $+J$ correction to the energy, potential, and ionic forces. I used the energetic correction shown in Equation 3.4 (following the example of Ref. 176 I have omitted the "$n_{\min}$" term that appears in that paper).

Example input and output files can be found at `https://www.repository.cam.ac.uk/handle/1810/288598`.

3.3.2 Calculating Hubbard parameters

Hubbard U and Hund's J parameters were calculated for the set of hexahydrated transition metals and MnO. Prior to the linear response calculations, the geometries of the hexahydrated transition metal complexes were optimised using the PBE xc functional without a Hubbard correction and with the water molecules constrained to their respective planes (refer to Figure 3.4). Various linear response approaches were performed: averaged and non-averaged 1×1, simple and scaled 2×2, as well as the standard scalar approach. While the scalar values reported here will be roughly analogous to conventional linear response reported elsewhere, they were calculated using minimum-tracking linear response, not SCF, which differ in their definitions of χ_0.

Hubbard and Hund's parameters were obtained for two Hubbard subspaces: the $3d$ subspace on the transition metal ion, and the $2p$ subspace on one of the equatorial oxygen atoms, taken as a representative of the six oxygen atoms in the system. The Hubbard parameters that were obtained are listed in Tables 3.1 and 3.2 respectively, and plotted in Figure 3.6. The uncertainties in the Hubbard parameters have also been calculated from the error in the least-square fits of $dv^\sigma_{\mathrm{Hxc}}/dn^{\sigma'}$, $dv^\sigma_{\mathrm{KS}}/dv^{\sigma'}_{\mathrm{ext}}$ and $dn^\sigma/dv^{\sigma'}_{\mathrm{ext}}$ using unbiased Gaussian error propagation. These error estimates prove to be very instructive.

Table 3.1: Values of U and J (eV) for hexahydrated transition metals and a spin-up manganese atom of MnO, calculated using the various linear response schemes introduced in subsection 3.2.1. The linear response calculations for the fully-filled $3d$ subspace in Fe^{3+} were poorly behaved (two different pseudopotentials were tested) and have consequently been excluded.

metal	scalar	averaged 1×1	1×1		simple 2×2		scaled 2×2	
	U	U	$U^{\uparrow}$	$U^{\downarrow}$	U	J	U	J
Ti^{3+}	3.88 ± 0.00	1.66 ± 0.00	1.85 ± 0.01	1.47 ± 0.00	3.90 ± 0.01	0.34 ± 0.00	3.89 ± 0.01	0.34 ± 0.00
V^{2+}	4.00 ± 0.00	2.78 ± 0.00	3.29 ± 0.00	2.28 ± 0.00	4.07 ± 0.01	0.34 ± 0.00	4.00 ± 0.01	0.35 ± 0.00
Cr^{3+}	3.90 ± 0.00	1.78 ± 0.00	1.86 ± 0.00	1.70 ± 0.00	4.04 ± 0.01	0.40 ± 0.00	3.90 ± 0.01	0.42 ± 0.00
Cr^{2+}	3.20 ± 0.00	2.39 ± 0.00	2.75 ± 0.00	2.04 ± 0.00	3.34 ± 0.01	0.33 ± 0.00	3.20 ± 0.01	0.35 ± 0.00
Mn^{3+}	5.40 ± 0.00	2.00 ± 0.00	1.51 ± 0.00	2.50 ± 0.00	5.86 ± 0.01	0.50 ± 0.00	5.40 ± 0.01	0.53 ± 0.00
Mn^{2+}	4.36 ± 0.00	4.05 ± 0.08	4.28 ± 0.15	3.82 ± 0.00	4.90 ± 0.06	0.37 ± 0.06	4.35 ± 0.01	0.52 ± 0.01
Fe^{3+}	5.88 ± 0.01	—	—	5.45 ± 0.02	—	—	5.92 ± 0.02	0.81 ± 0.02
Fe^{2+}	4.58 ± 0.00	5.07 ± 0.09	6.28 ± 0.18	3.86 ± 0.00	6.06 ± 0.09	0.43 ± 0.06	4.58 ± 0.01	0.63 ± 0.01
Co^{3+}	6.25 ± 0.00	1.19 ± 0.00	1.19 ± 0.00	1.19 ± 0.00	6.25 ± 0.00	0.75 ± 0.00	6.25 ± 0.00	0.75 ± 0.00
Co^{2+}	4.95 ± 0.02	6.19 ± 0.02	8.17 ± 0.03	4.22 ± 0.02	7.15 ± 0.02	0.48 ± 0.01	4.96 ± 0.02	0.65 ± 0.01
Ni^{2+}	5.26 ± 0.00	9.84 ± 0.02	15.41 ± 0.05	4.27 ± 0.00	12.35 ± 0.03	0.75 ± 0.02	5.26 ± 0.01	0.78 ± 0.01
Cu^{2+}	4.62 ± 0.00	-2.54 ± 0.03	-9.11 ± 0.05	4.04 ± 0.00	-4.99 ± 0.02	0.85 ± 0.02	4.63 ± 0.01	0.90 ± 0.01
MnO	5.44 ± 0.04	4.63 ± 0.08	5.54 ± 0.15	3.72 ± 0.02	8.38 ± 0.15	0.51 ± 0.05	5.37 ± 0.04	0.49 ± 0.02

Table 3.2: Values of U and J (eV) calculated using the various linear response schemes, for an equatorial oxygen atom within hexahydrated transition metal systems, and for a MnO oxygen atom.

metal	scalar	averaged 1×1	1×1		simple 2×2		scaled 2×2	
	U	U	$U^{\uparrow}$	$U^{\downarrow}$	U	J	U	J
Ti^{3+}	8.16 ± 0.03	5.05 ± 0.01	5.20 ± 0.01	4.89 ± 0.00	8.14 ± 0.02	1.05 ± 0.00	8.13 ± 0.02	1.05 ± 0.00
V^{2+}	8.28 ± 0.00	5.69 ± 0.00	5.70 ± 0.00	5.69 ± 0.00	8.28 ± 0.01	1.29 ± 0.00	8.28 ± 0.01	1.29 ± 0.00
Cr^{3+}	8.29 ± 0.00	5.54 ± 0.00	5.44 ± 0.00	5.65 ± 0.00	8.29 ± 0.02	1.08 ± 0.01	8.29 ± 0.02	1.08 ± 0.01
Cr^{2+}	8.44 ± 0.01	6.28 ± 0.01	6.55 ± 0.01	6.01 ± 0.02	8.45 ± 0.02	1.27 ± 0.01	8.45 ± 0.02	1.27 ± 0.01
Mn^{3+}	8.57 ± 0.00	4.94 ± 0.00	5.53 ± 0.00	4.35 ± 0.00	8.58 ± 0.03	0.97 ± 0.01	8.57 ± 0.03	0.97 ± 0.01
Mn^{2+}	8.30 ± 0.00	6.05 ± 0.01	5.70 ± 0.01	6.39 ± 0.00	8.29 ± 0.01	1.31 ± 0.00	8.31 ± 0.01	1.30 ± 0.01
Fe^{3+}	8.37 ± 0.03	5.55 ± 0.05	4.48 ± 0.07	6.62 ± 0.07	8.59 ± 0.14	1.24 ± 0.06	8.40 ± 0.12	1.06 ± 0.06
Fe^{2+}	8.83 ± 0.01	5.77 ± 0.00	5.43 ± 0.01	6.10 ± 0.00	8.83 ± 0.01	1.40 ± 0.00	8.83 ± 0.01	1.39 ± 0.01
Co^{3+}	8.26 ± 0.00	4.37 ± 0.09	4.27 ± 0.11	4.48 ± 0.15	8.39 ± 0.10	1.12 ± 0.05	8.39 ± 0.10	1.12 ± 0.05
Co^{2+}	8.25 ± 0.06	5.24 ± 0.10	4.89 ± 0.11	5.60 ± 0.15	8.24 ± 0.09	1.38 ± 0.06	8.25 ± 0.09	1.37 ± 0.06
Ni^{2+}	8.09 ± 0.01	4.89 ± 0.00	4.65 ± 0.00	5.14 ± 0.00	8.09 ± 0.01	1.37 ± 0.00	8.09 ± 0.01	1.37 ± 0.00
Cu^{2+}	8.38 ± 0.00	5.08 ± 0.00	4.68 ± 0.00	5.48 ± 0.00	8.36 ± 0.01	1.38 ± 0.00	8.38 ± 0.01	1.38 ± 0.00
MnO	10.88 ± 0.01	5.32 ± 0.04	5.32 ± 0.05	5.32 ± 0.05	10.92 ± 0.12	1.03 ± 0.03	10.92 ± 0.12	1.03 ± 0.03

General trends

Both tables exhibit some general trends: the Hubbard parameters of the metal ions grow slowly as the number of $3d$ electrons increases (Figure 3.6a); oxygen parameters remain relatively stable; the Hund's coupling parameters of the metals appear reasonable. Furthermore, the scalar approach and scaled 2×2 (atom-wise inversion) yield the same result across the board, in keeping with the conclusions of Subsection 3.2.2. The scaled 2×2 approach is marginally less numerically stable, which is reflected by the marginally larger error estimates. Interestingly, however, I found that for the spin channel that matters to strong correlation (the spin-up channel for less-than-half filled sub-shells, and the spin-down channel for more-than-half filled sub-shells), the relevant 1×1 U is very reasonable, and systematically lower in value than the conventional scalar U. This hints at a possible solution for first-principles DFT + U calculations on systems in which the calculated scalar U proves to be unphysically large, and the predominantly empty/full spin channel is already well described by the approximate functional.

One particularly noteworthy result is the substantial spin-screening of the Hubbard parameters of $[Co(H_2O)_6]^{3+}$ observed in averaged and non-averaged 1×1. This is the only complex in a low-spin ground state, so the up and down KS orbitals overlap perfectly and there is very efficient screening between spins. This system also exhibits one of the largest J values. Similarly, the large J values on the oxygen atoms may surprise at first (as Hund's physics is expected to play a very minor role here). This illustrates an important point: the absence of any magnetisation does not imply the absence of magnetisation-related error in the approximate functional. Subsequent calculations demonstrate that applying this J term, large as it is, does not result in the oxygen atoms acquiring magnetic moments.

Some works go one step further and calculate Hubbard parameters in a self-consistent fashion,[179,221,225] with linear response being performed on DFT + U ground states. While it remains to be seen what effect this additional step would have, it will likely be small here because these systems do not undergo qualitative changes in electronic structure upon the application of U:[248] in going from DFT to scalar DFT + U, the root-mean-square and maximum fractional differences in the total $3d$ occupancies are 6% and 15% respectively. For the spin moment $\mu = n^{\uparrow} - n^{\downarrow}$ these are 7% and 14% respectively.

It is important to acknowledge that the authors of Ref. 186 calculated U for this set of molecules (using scalar linear response). In comparison, their values are lower (by 1.4 eV on average) and more species-dependent (a standard deviation of 1.2 eV compared to 0.9 eV for our set of values). In comparison with this work, Ref. 186 (a) used ultra-soft pseudopotentials as opposed to norm-conserving ones; (b) performed all calculations on structures optimised in the 3+ charge state; (c) employed U self-consistency for some calculations; and (d) used SCF linear response. As the following section will demonstrate, details such as (a) and (b) can substantially affect Hubbard parameters.

Comparison of schemes

Table 3.1 illustrates the dangers of averaging across the two spin channels, as performed in averaged 1×1. For systems where both the spin-up and spin-down channels are partially

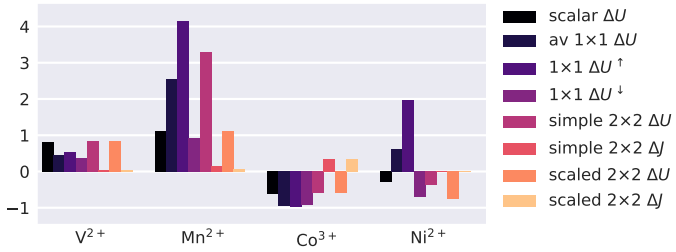


Figure 3.7: The difference in Hubbard parameters for four hexahydrated transition metals, as calculated via the various linear response schemes and using two alternative simulation set-ups (eV).

occupied (see Figure 3.6a) the responses are well-behaved, the Hubbard parameters are both sensible and similar, and averaging is unlikely to have any drastic effects. But for the heavier elements with filled spin-up channels, we are faced with the prospect of averaging two very different values, which in the most extreme cases lead to negative Hubbard parameters. Here, averaging the two values is likely to be a poor approximation.

However, any Hubbard correction will not directly affect a fully-occupied channel, because the Hubbard energy correction term (Equation 3.2) vanishes regardless of the magnitude of U. If it is imperative that the same correction must be applied to both channels, an argument could be made in favour of applying the $U^{\downarrow}$ value in place of an average. Of course, the Hubbard *potential* does not vanish (Equation 3.3) and fictional spin-up KS orbitals that overlap with the Hubbard projectors would be shifted by $U^{\downarrow}$. This inconsistency may have unforeseen effects, and an alternative may be to apply DFT + U to partially-filled spin channels only.

Table 3.1 also demonstrates the shortcomings of simple 2×2, the approximate atom-wise-inversion-based method. In the upper half of the table it yields reasonable values similar to those of scaled 2×2. But in the latter half (where dramatically different response in the spin-up and spin-down channels is expected) the approximation is a very poor one and the resulting parameters are unphysical. Scaled 2×2 encounters no such difficulties, justifying the use of the rescaling factors $\lambda_{U/J}$. I will consider simple 2×2 no further.

Dependence on simulation settings

The results of linear response calculations are sensitive to the precise settings of a calculation. Figure 3.7 shows the difference in Hubbard parameters as obtained using two quite different simulation schemes. Both sets of calculations were performed on the same physical systems, but they differed in (a) the pseudopotentials used (Rappe *vs.* in-house); (b) the electrostatic truncation scheme used (padded cell with a spherical cut-off[238] *vs.* a Martyna-Tuckerman correction[249]); and (c) the resolution of the fine grid used for calculating products of basis functions (a factor of two *vs.* a factor of four finer than the standard grid). The majority of the Hubbard parameters match to within 1 eV, except for those that relate to the response of a nearly-fully occupied subspace, where the response is extremely changeable.

Table 3.3: Values of U and J (eV) for the $3d$ subspace of Zn in hexahydrated zinc, calculated using the various linear response schemes and two alternative sets of Hubbard projectors (as defined by the net charge configuration of the Zn atom in a pseudoatomic solver).

PAO charge		+0	+2
scalar	U	10.05±0.03	34.77±0.01
averaged 1×1	U	11.60±0.04	44.64±0.02
1×1	$U^{\uparrow}$	11.67±0.06	44.65±0.03
	$U^{\downarrow}$	11.53±0.06	44.63±0.02
simple 2×2	U	10.08±0.03	34.79±0.02
	J	1.75±0.05	1.47±0.03
scaled 2×2	U	10.08±0.03	34.79±0.02
	J	1.75±0.05	1.47±0.03

A closed-shell system

Linear response calculations were also performed on $[Zn(H_2O)_6]^{2+}$. Zn^{2+} is not strictly a transition metal, as its $3d$ shell is filled. Linear response calculations on closed-shell systems tend to be troublesome,[220,222] possibly due the response becoming non-linear.[223]

The results of our calculations are listed in Table 3.3. These calculations were performed for two different definitions of the Hubbard projectors. In ONETEP these are defined using PAOs: that is, the DFT solutions of the isolated atom/ion with the pseudopotential.[235,250,251] Table 3.3 lists the Hubbard parameters for when the pseudoatomic problem was solved with a total charge of 0 and +2, keeping the pseudopotential itself fixed. The Hubbard projectors corresponding to the neutral pseudoatom are more diffuse than those for the +2 case.

We find that U is exceptionally large as given by both the scalar and spin-resolved linear response schemes, and with either definition of the Hubbard projectors. The dependence of the result on the Hubbard projectors is very striking, and is the most dramatic case that I have seen. But what is more remarkable is the robustness of these calculations (as shown by the small uncertainties). Crucially, this robustness is not due to the fact that some schemes avoid matrix inversion: the uncertainties are similar for schemes where matrix inversion is necessary (2×2) and those where it is not (1×1), and in no case did I observe evidence of non-linear response.

3.3.3 A comparison with cRPA

For the sake of comparison, it is instructive to study how constrained random phase approximation (cRPA) methods account for the spin-screening of Hubbard parameters.[252,253] In these approaches, the non-interacting response χ_0 is partitioned into components corresponding to response within/between various subspaces. For instance, consider a system consisting of a single site with spin-up and -down channels. The component due to response solely within the spin-up subspace is given by the $(\uparrow,\uparrow)$th entry of χ_0 — that is,

$$(\chi_{0,\uparrow})^{\sigma\sigma'} \equiv \begin{pmatrix} {\chi_0}^{\uparrow\uparrow} & 0 \\ 0 & 0 \end{pmatrix}. \tag{3.37}$$

Table 3.4: Spin-screened Hubbard parameters U (eV) calculated using the cRPA approach. The differences with respect to the corresponding averaged and non-averaged 1×1 results of Table 3.1 are given in parentheses.

metal	average		$U^{\uparrow}$		$U^{\downarrow}$	
Ti^{3+}	0.80	(−0.86)	0.88	(−0.97)	0.71	(−0.76)
V^{2+}	2.22	(−0.56)	2.57	(−0.72)	1.88	(−0.40)
Cr^{3+}	1.16	(−0.62)	1.06	(−0.80)	1.26	(−0.44)
Cr^{2+}	2.07	(−0.32)	2.31	(−0.44)	1.83	(−0.21)
Mn^{3+}	1.17	(−0.83)	0.38	(−1.13)	1.95	(−0.55)
Mn^{2+}	3.47	(−0.58)	3.15	(−1.13)	3.78	(−0.04)
Co^{3+}	1.20	(+0.01)	1.20	(+0.01)	1.20	(+0.01)
Co^{2+}	5.19	(−1.00)	6.23	(−1.94)	4.14	(−0.08)
Ni^{2+}	8.36	(−1.48)	12.39	(−3.02)	4.32	(+0.05)
Cu^{2+}	−3.53	(−0.99)	−11.44	(−2.33)	4.37	(+0.33)

The non-interacting response due to all other contributions is

$$(\tilde{\chi}_{0,\uparrow})^{\sigma\sigma'} \equiv \chi_0 - \chi_{0,\uparrow} = \begin{pmatrix} 0 & \chi_0{}^{\uparrow\downarrow} \\ \chi_0{}^{\downarrow\uparrow} & \chi_0{}^{\downarrow\downarrow} \end{pmatrix}. \tag{3.38}$$

For such a non-interacting response $\tilde{\chi}_{0,\sigma}$ there is a corresponding Dyson equation

$$U^{\sigma}_{RPA} = \left[\left(f^{-1} - \tilde{\chi}_{0,\sigma} \right)^{-1} \right]^{\sigma\sigma} \tag{3.39}$$

where U^{σ}_{RPA} is now screened by everything except for interactions within the spin-σ subspace (as this screening is what $\tilde{\chi}_{0,\sigma}$ pertains to).

Screened interaction parameters U^{σ}_{RPA} for hexahydrated metal systems are tabulated in Table 3.4. In this work, it was shown that point-wise inversion (the averaged and non-averaged 1×1 schemes) yields an interaction screened by both the opposite spin channel on the same site and the remainder of the system, so we expect the results of Table 3.4 to resemble those of Table 3.1. They are correlated, but the match is certainly not exact. This suggests that the random phase approximation (RPA) is not a good approximation for screening between unlike-spins, and that more sophisticated methods (such as that of Ref. 254) are required.

3.4 Properties of MnO

We calculated the band gap (Figure 3.8) and the local magnetic moment of Mn (Figure 3.9) for bulk MnO using Hubbard and Hund's parameters obtained via our novel schemes (and listed in Tables 3.1 and 3.2). Semi-local functionals dramatically underestimate the band gap of MnO; the local/semi-local results presented in Figure 3.8 underestimate it by 2.3 eV on average (with a standard deviation of 1.0 eV). They also underestimate the local magnetic moment (by $0.35 \pm 0.14\,\mu_B$). More sophisticated techniques have been applied with mixed success: hybrid, GW, and other DFT + U studies underestimate the band gap by 1.3 ± 1.0, 1.3 ± 0.7, and 1.1 ± 0.7 eV respectively. Our approaches compare very favourably, with the band gap agreeing with experiment, differing on average by -0.2 ± 0.4 eV. Scaled 2×2 in particular gives both the band gap and magnetic moment in excellent agreement with experiment.

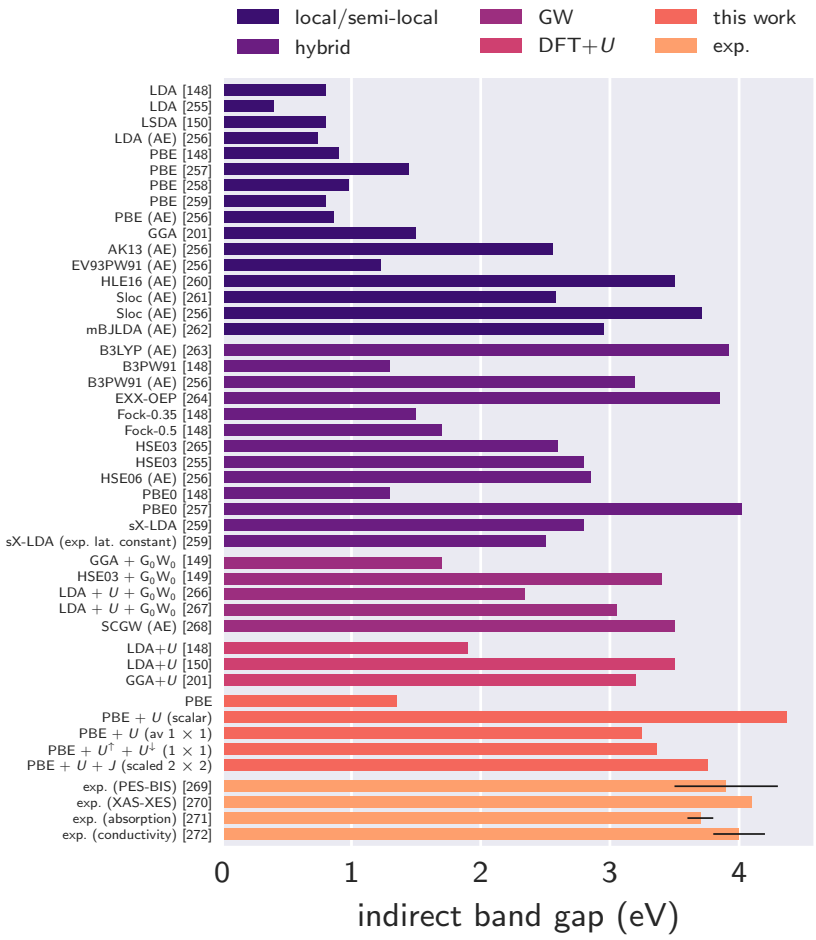


Figure 3.8: The indirect band gap of MnO, as calculated by various computational approaches, as well as experimental results (with error bars). All-electron calculations are denoted "AE".

I found the predicted band gap to be highly sensitive to the choice of pseudopotential, with different pseudopotentials predicting anything from a metal to gaps as large as 2 eV (for PBE). All-electron calculations yield a gap of 0.86 eV.[256] To obtain similar values with a pseudopotential, ensuring accurate descriptions of $4s$ and $4p$ scattering proved to be key.

Transition metal oxides are typically insulating for one of two reasons. Early $3d$ transition metal oxides (such as TiO and VO) are Mott-Hubbard insulators, with the band gap sitting between the lower and upper Hubbard bands. Late $3d$ transition metal oxides (such as CuO and NiO) are charge-transfer insulators, with band gaps formed between the oxygen $2p$ band and the upper metal $3d$ band, separated by the ligand-to-metal charge transfer energy.

MnO sits near the boundary of these two regimes; the valence band edge is neither purely metal $3d$ or oxygen $2p$ in character.[269,275] As Figure 3.10 illustrates, this picture is captured by all schemes, with the valence band edge character sitting between 36 to 59 % Mn. That said, if Hubbard corrections are applied to Mn but not O, the Mn character drops to below 26% in all cases, incorrectly approaching a charge-transfer insulator. This demonstrates the importance of applying corrections to the oxygen orbitals. The valence band in its entirety is plotted in Figure 3.11, and our methods exhibit marked improvement over PBE.

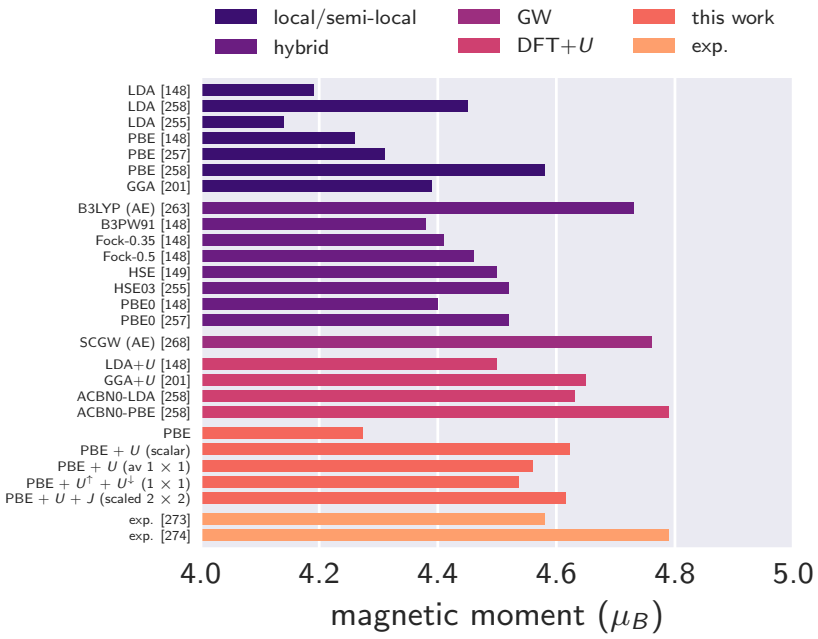


Figure 3.9: The magnetic moment of the manganese atoms in MnO, as calculated by various approaches.

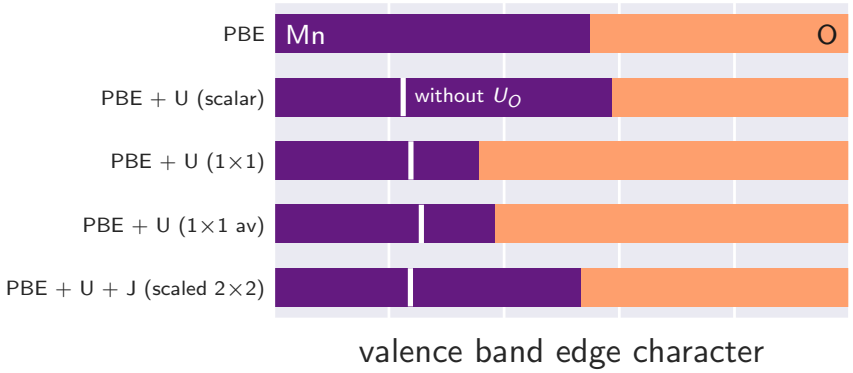


Figure 3.10: The valence band edge character of MnO, showing the fractional contribution of Mn (purple) and O (orange). PBE correctly predicts the valence band edge's mixed character, as do the different corrective schemes. This balance is due largely to the U (and J where relevant) terms applied to the oxygen $2p$ subspaces, which see the Mn fractions increase from unphysically low values (indicated in white).

3.5 Properties of hexahydrated metal complexes

3.5.1 Structural properties

We will now examine how these various Hubbard corrections affect the resulting geometry of the hexahydrated metal systems. Hartree-Fock,[276] hybrid DFT,[277,278] and semi-local xc functionals (such as PBE)[279] already predict bond lengths consistent with experiment,[280] without any need for Hubbard corrections. However, these corrections can dramatically affect structural properties; it would be undesirable for them to do so here.

If only the $3d$ orbitals of the transition metal species are subjected to a Hubbard correction, and the structure optimised, metal-oxygen distances dramatically lengthen (Figure 3.12a and

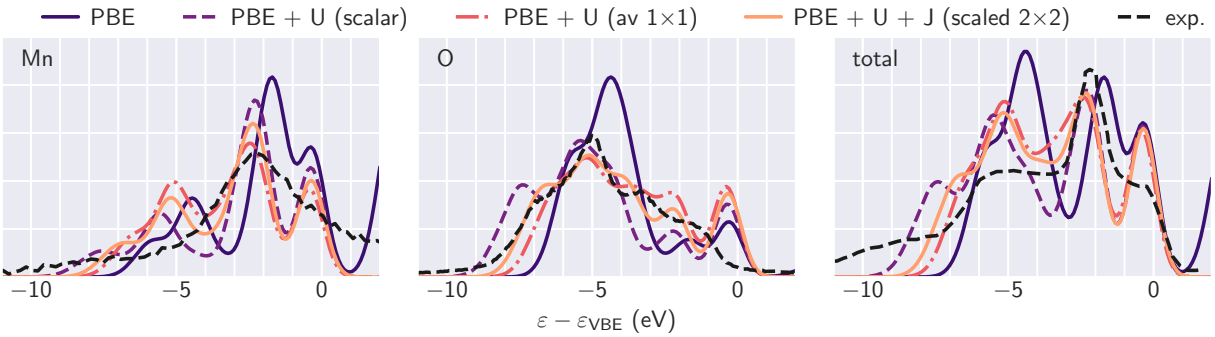


Figure 3.11: The local Mn, O, and total densities of states as obtained by the different schemes. The 1×1 result is similar to the averaged 1×1 result, and so has been excluded for simplicity. The energy scale is shown relative to the valence band edge energy $\varepsilon_{\mathrm{VBE}}$. Experimental results (x-ray emission and photoelectron spectroscopy) from Ref 270 are included for comparison.

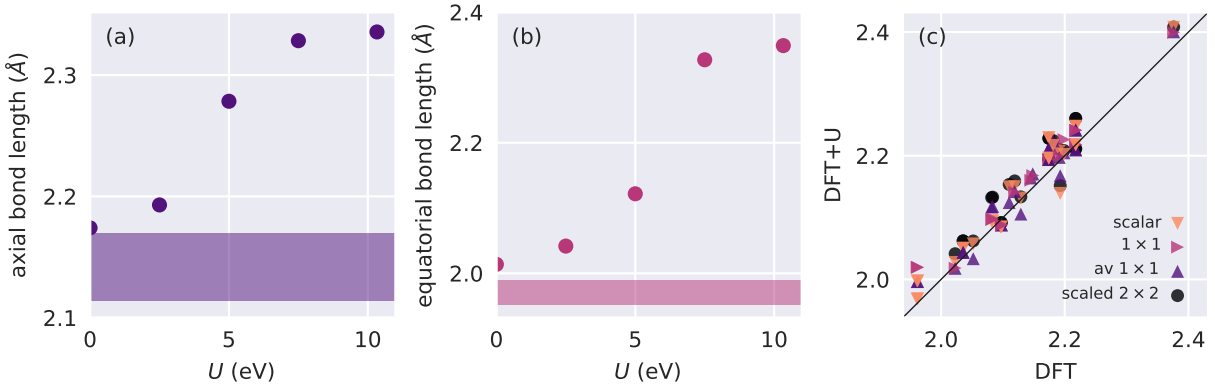
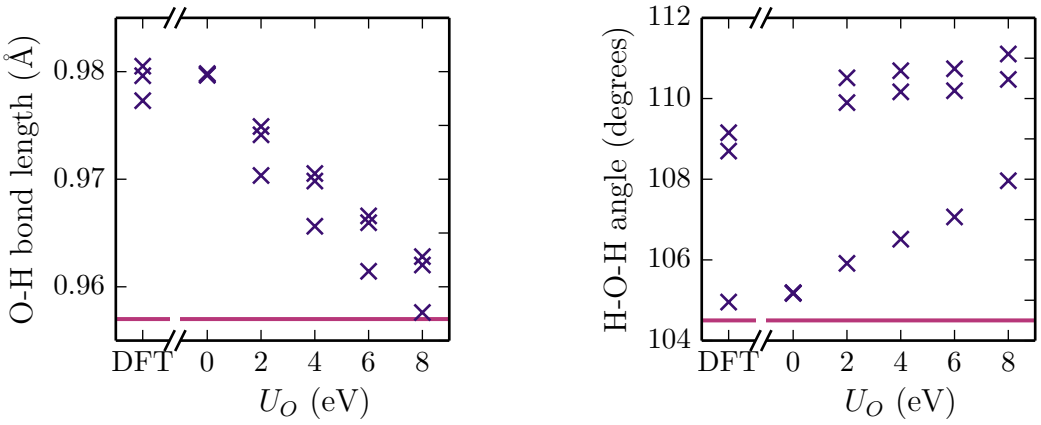


Figure 3.12: The mean (a) axial and (b) equatorial bond lengths of hexahydrated Mn^{3+} when optimised using DFT + U, for various values of U^{Mn} and without adding a Hubbard correction to the oxygen atoms. The shaded regions indicate the range of values reported by other computational studies,[276–279] which are in line with experiment.[280] (c) Metal-oxygen distances as given by DFT + U–optimised structures, now with a first-principles Hubbard U correction to the oxygen $2p$ orbitals, as compared to analogous PBE calculations. Each data-point corresponds to a distinct set of Hubbard parameters from Tables 3.1 and 3.2 (that is, all different transition metal species and schemes for computing Hubbard parameters).

3.12b). This is because any hybridisation that existed between the metal $3d$ orbitals with lone pairs on the water ligands is weakened by the lowering of the energy of any filled $3d$ orbitals. Consequently, the individual species are stabilised and they drift apart. It is clear that this elongation is wholly unphysical, taking bond lengths well outside of the range of experimental values. This failure is not specific to this particular system or any procedure for computing U, but is a well-documented problem.[210,248,281,282]

There are a number of approaches for correcting this issue. One solution is DFT + U + V, whose inter-site interaction correction to the DFT + U energy functional may correctively favour O ($2p$)–metal ($3d$) bonding.[209] Alternatively, adaptive Hubbard projectors can mitigate the problem, as they will be more delocalised and responsive to the bonding environment.[173] But perhaps the most pragmatic approach is to add Hubbard corrections to the $2p$ orbitals of the oxygen atoms.[283,284] This lowers their energies to levels comparable with the $3d$ orbitals, re-establishing the possibility of hybridisation.



(a) **(b)**

Figure 3.13: Water deformation as a result of the addition of Hubbard corrections to the manganese and oxygen orbitals in hexahydrated manganese, showing (a) O–H distances and (b) H–O–H angles (plotted as crosses). Also shown (light purple) is a literature value corresponding to an ab initio calculation of liquid water.[286] The leftmost point corresponds to $U_{Mn,O}$=0 eV; the rest of the plot corresponds to $U_{Mn} = 4.9$ eV and U_0 progressively increasing from 0 to 8 eV. The water molecules for each structure display a range of bond lengths and angles due to the Jahn-Teller effect.

Table 3.5: Spin-flip energies (eV) for various hexahydrated transition metal systems. The quantum chemistry results are from Ref. 278, and the experimental results are from Ref. 287 (and the references therein).

metal	DFT	DFT + U (+J)				CASSCF	CASPT2	MRCI	exp
		scalar	av 1×1	1×1	scaled 2×2				
V^{2+}	1.06	1.10	1.09	1.28	1.10	2.01	1.89	1.98	1.62
Cr^{3+}	1.11	1.04	1.11	1.33	1.04	2.41	2.23	2.35	2.60
Mn^{2+}	2.16	2.41	2.39	2.41	2.42	3.42	2.91	3.25	2.34
Co^{2+}	1.60	1.85	1.85	1.86	1.85	1.96	1.95	1.76	1.98
Ni^{2+}	1.23	1.44	1.48	1.50	1.44	2.30	2.03	2.23	1.91

The success of the latter method is demonstrated in Figure 3.12c, where the addition of these corrections reduces any bond elongation to at most a five percent increase (and in many cases much less). The alignment is particularly remarkable given the range of different U and J values being used.

It is important to note that adding Hubbard terms to the oxygen atoms (a) alters hydrogen-oxygen-hydrogen angles by less than 2%, (b) alters oxygen-hydrogen bond lengths by approximately 1%, and (c) does not result in the oxygen atoms acquiring a magnetic moment (the largest observed was $0.014\,\mu_B$ for DFT and $0.073\,\mu_B$ for DFT + U + J), as shown in Figure 3.13. This level of deformation is acceptable, being within the sorts of errors one would expect from DFT geometries (and also within the range of bond lengths and angles used in molecular models of water[285]).

3.5.2 Spectroscopic properties

Hubbard corrections have significant bearing on spectroscopic properties (given that to first order, they open a gap between the filled and unfilled Hubbard projectors). This section will

Table 3.6: Hubbard parameters calculated via linear response for systems where one electron's spin has been flipped from the ground spin state. The differences to the parameters obtained for the ground state (Table 3.1) are listed in parentheses.

metal	scalar	averaged 1×1	1×1		scaled 2×2	
	U	U	$U^{\uparrow}$	$U^{\downarrow}$	U	J
V^{2+}	3.99 (−0.01)	2.57 (−0.21)	2.72 (−0.57)	2.42 (+0.14)	3.84 (−0.23)	0.35 (+0.01)
Cr^{3+}	4.03 (+0.13)	1.69 (−0.09)	1.71 (−0.15)	1.68 (−0.02)	4.01 (−0.03)	0.41 (+0.01)
Cr^{2+}	3.08 (−0.12)	2.04 (−0.35)	2.12 (−0.63)	1.97 (−0.07)	3.08 (−0.26)	0.31 (−0.02)
Mn^{3+}	5.26 (−0.14)	1.64 (−0.36)	1.59 (+0.08)	1.69 (−0.81)	5.27 (−0.59)	0.50 (+0.00)
Mn^{2+}	4.33 (−0.03)	2.97 (−1.08)	3.19 (−1.09)	2.74 (−1.08)	4.56 (−0.34)	0.38 (+0.01)
Co^{2+}	5.11 (+0.16)	2.85 (−3.34)	2.86 (−5.31)	2.85 (−1.37)	5.12 (−2.03)	0.42 (−0.06)
Ni^{2+}	5.49 (+0.23)	3.32 (−6.52)	3.31 (−12.10)	3.32 (−0.95)	5.48 (−6.87)	0.90 (+0.15)

Table 3.7: Spin-flip energies (eV) for various hexahydrated transition metal systems with U (and J for 2×2) updated following the flip. The quantum chemistry results are from Ref. 278, and the experimental results are from Ref. 287 (and the references therein).

metal	DFT	DFT + U (+J)				CASSCF	CASPT2	MRCI	exp
		scalar	av 1×1	1×1	scaled 2×2				
V^{2+}	1.06	1.11	1.26	1.28	0.80	2.01	1.89	1.98	1.62
Cr^{3+}	1.11	0.94	1.32	1.33	−0.15	2.41	2.23	2.35	2.60
Mn^{2+}	2.16	2.40	2.73	2.74	2.00	3.42	2.91	3.25	2.34
Co^{2+}	1.60	1.83	1.44	1.62	0.72	1.96	1.95	1.76	1.98
Ni^{2+}	1.23	1.81	1.03	1.50	0.41	2.30	2.03	2.23	1.91

focus on *d*-*d* excitation energies, where a single electron transitions between two $3d$ orbitals. While these transitions are formally dipole-dipole forbidden by the Laporte selection rule, they are allowed via vibronic coupling.[288]

The first subset of such transitions are those which involve the flip of the electron's spin. These transitions additionally violate spin selection rules, but vibronic coupling again means that they are observable (albeit weakly). The transition energies are simply calculated as the difference in the total energy between two DFT (+ U) calculations where the total spin differs by $\hbar$. This was done without updating U. As this approach relies only on the accuracy of the total energy, DFT alone (without a Hubbard correction) might give reasonable results. This is indeed what I found (Table 3.5). The results are relatively insensitive to the choice of Hubbard parameters. Surprisingly, the scalar and scaled 2×2 approaches yield near-identical results, despite the fact that the two approaches differ by the value for J and share the same value for U. A Hund's correction ought to have a significant bearing on spin-flip energies, providing further evidence that the precise functional form of the $+J$ functional needs revision.

If instead we update U for the excited state, we get the Hubbard values listed in Table 3.6 and the resulting spin-flip energies listed in Table 3.7. Using updated Hubbard parameters worsened the resulting spin-flip energies, with some cases even predicting the wrong ground state. The scaled 2×2 results are an excellent demonstration of the findings of Millis and co-workers, who showed that the current $+J$ functional wrongly disfavours ferromagnetism.[227]

The other possible *d*-*d* excitations involve the transition of a single electron without changing its spin. These transitions are spin-allowed, and thus will exhibit intensities between those of fully allowed and spin-forbidden transitions. The transition energies are calculated as the difference in energy of the corresponding KS orbitals, and are listed in Table 3.8.

I found that DFT and DFT + U (+ J) have mixed success reproducing these transition en-

ergies. This not surprising. The energy of such transitions is instead directly related to the calculated KS band gap and, as such, DFT (with its well-known underestimation of the band gap) will not give accurate results. Hubbard corrections tend to correctly enlarge KS band gaps, but there is no reason *a priori* why the final gap they produce ought to be accurate.[186] Ongoing efforts are being made to construct generalised DFT + U theories that satisfy Janak's/Koopman's theorem.[289–291] These transition energies will also be highly sensitive to static correlation, a failing of DFT associated with multi-reference ground states. This failing remains unaddressed and may be an important factor in the overestimation of transition energies of Ti^{2+}, Fe^{2+}, and Co^{2+}.[180,181] Adapting DFT + U-like functionals to correct both self-interaction and static correlation error is an area of active research.[292] Furthermore, the excitation energies shown have been computed using a very simplistic approach, neglecting vibronic and solvation effects (among others), which would likely result in significant shifts.[293]

3.6 Conclusions

This chapter has presented the generalisation of the minimum-tracking linear response formalism for calculating U and J to multiple sites and spins.[212] In this formalism, the non-interacting response χ_0 is strictly a ground-state property. Previously, it was not possible to calculate Hubbard parameters via linear response in large, spin-polarised systems such as metalloproteins.[30,294] But because minimum-tracking is compatible with direct minimisation (common to linear-scaling density functional theory packages such as ONETEP), linear response calculations on large and complex systems are now possible.

Crucially, this formalism allowed me to work with spin relatively easily. I demonstrated that the scalar linear response approach, whose use is widespread, yields a Hubbard U that is unscreened by the opposite spin channel of the same site. I presented alternative approaches that account for this screening. Specifically, the opposite spin channel can be included in the bath, which is consistent with the effective decoupling of spins into separate subspaces implied by the standard DFT + U functional (*i.e.* the 1×1 schemes). This lowers the resulting U values. Alternatively (but not equivalently), if inter-spin interactions require correction then a Hund's coupling parameter ought to be used in conjunction with an adjusted Hubbard parameter (scaled 2×2).

Applying these approaches to hexahydrated transition metals revealed significant trends in the Hubbard parameters across the transition metals. The linear response calculations were remarkably stable numerically, offering a possible route forward for closed-shell solids. That said, the best DFT + U like model, and hence the uniquely-defined linear response calculation scheme for that model, seems to be difficult to predict for a given system and underlying exchange-correlation functional.

In the case of MnO, a canonical strongly correlated system, these novel approaches gave band gaps, magnetic moments, and valence band edge characters in excellent agreement with experiment, with a satisfyingly small variance compared to hybrid functionals and other methods. In the case of the hexahydrated transition metal complexes all approaches reproduced reasonable bond lengths but none reliably reproduced experimental *d*-*d* excitation energies. The

Table 3.8: KS transition energies (eV) for spin-conserving *d*-*d* excitations. In all cases, corrective terms were applied to both the metal 3*d* and oxygen 2*p* subspaces.

metal	final symmetry	DFT	DFT + U scalar	DFT + U 1a	DFT + U 1b	DFT + U 2b	CASSCF	CASPT2	MRCI	exp
Ti^{3+}	$^1B_{2g}$	0.27	3.16	1.48	1.63	2.79	0.00			
	$^1B_{3g}$	0.28	3.25	1.51	1.66	2.86	0.00			
	^1A_g	1.94	3.98	2.81	2.92	3.76	1.69	1.71	1.76	2.16
	^1A_g	2.38	4.60	3.33	3.44	4.34	1.70	1.72	1.77	2.52
V^{2+}	$^3B_{1g}$	1.97	4.92	3.98	4.35	4.61	1.19	1.26	1.28	1.53
	$^3B_{2g}$	1.97	4.92	3.98	4.35	4.61	1.19	1.26	1.28	1.53
	$^3B_{3g}$	1.97	4.92	3.98	4.35	4.61	1.19	1.26	1.28	1.53
Cr^{3+}	$^3B_{1g}$	2.24	3.98	3.15	3.17	3.56	1.69	1.77	1.79	2.16
	$^3B_{2g}$	2.24	3.98	3.15	3.17	3.56	1.69	1.77	1.79	2.16
	$^3B_{3g}$	2.24	3.98	3.15	3.17	3.56	1.69	1.77	1.79	2.16
Cr^{2+}	^4A_g	0.38	2.06	1.60	1.80	1.83	0.62	0.69	0.64	1.17
	$^4B_{2g}$	1.37	3.28	2.77	2.99	3.04	1.18	1.27	1.19	
	$^4B_{3g}$	1.53	3.44	2.93	3.15	3.22	1.23	1.30	1.23	
	$^4B_{1g}$	1.95	3.88	3.36	3.58	3.61	1.34	1.44	1.36	1.75
Mn^{3+}	^4A_g	0.21	1.28	0.68	0.62	0.97	0.69	0.77	0.72	1.11
	$^4B_{2g}$	0.97	5.08	3.63	1.86	4.87	1.72	1.96	1.78	2.53
	$^4B_{3g}$	2.64	5.28	3.99	3.41	4.95	1.76	1.99	1.82	2.53
	$^4B_{1g}$	3.00	5.72	4.38	3.88	5.42	1.91	2.21	2.00	2.53
Fe^{2+}	$^4B_{2g}$	1.28	5.62	6.03	4.92	4.95	0.00			
	$^4B_{3g}$	1.28	5.63	6.04	4.93	5.04	0.00			
	^4A_g	1.88	5.73	6.07	5.07	5.24	0.75	0.80	0.83	1.29
	^4A_g	3.12	6.99	7.34	6.38	6.33	0.85	0.89	0.91	1.29
Co^{2+}	$^3B_{2g}$	3.02	7.85	8.65	7.03	7.16	0.00			
	$^3B_{3g}$	3.03	7.85	8.67	7.04	7.17	0.00			
	$^3B_{2g}$	3.64	7.91	8.75	7.15	7.23	0.67	0.81	0.65	1.02
	$^3B_{3g}$	3.64	7.91	8.78	7.16	7.24	0.68	0.81	0.65	1.02
	$^3B_{3g}$	5.17	9.41	10.08	8.76	8.69	2.82	2.69	2.62	2.41
	$^3B_{2g}$	5.17	9.42	10.10	8.77	8.70	2.85	2.69	2.62	2.41
Ni^{2+}	$^2B_{3g}$	4.30	8.74	13.39	7.72	7.91	0.75			
	$^2B_{1g}$	4.30	8.74	13.39	7.72	7.91	0.76	0.89	0.85	1.05
	$^2B_{2g}$	4.30	8.74	13.39	7.72	7.91	0.76			
	$^2B_{1g}$	4.30	8.74	13.39	7.72	7.91	1.31	1.48	1.45	1.67
	$^2B_{2g}$	4.31	8.74	13.39	7.72	7.91	1.31			
	$^2B_{3g}$	4.31	8.74	13.39	7.72	7.91	1.31			
Cu^{2+}	^1A_g	1.79	5.06	0.49	4.56	4.47	0.51	0.61	0.52	1.17
	$^1B_{2g}$	2.31	6.19	1.12	5.49	5.33	0.84	1.08	0.85	
	$^1B_{3g}$	2.82	6.68	1.81	5.91	5.82	0.89	1.12	0.89	
	$^1B_{1g}$	3.34	6.85	1.99	5.95	5.88	0.97	1.23	0.99	1.56

1×1 approach gave the best results for spin-flip energies (a well-defined ground-state property), but even these were not in very good agreement with quantum-chemistry results. Here, the electronic structure appears to be too complicated to be accurately described by the standard DFT + U functional, especially while static correlation remains unaddressed. The development of DFT + U methodologies are reliant on ever more accurate quantum chemistry benchmarks (*e.g.* Refs. 295 and 296).

Applying Hubbard corrections to the oxygen 2*p* subspaces proved to be necessary to preserve the correct valence band edge character in MnO and to reproduce bond lengths in hexahydrated transition metals.

By establishing a systematic approach for including/excluding screening by the opposite spin channel, these developments provide a route forward for performing DFT + U (+ J) on spin-polarised systems in a robust and consistent manner.

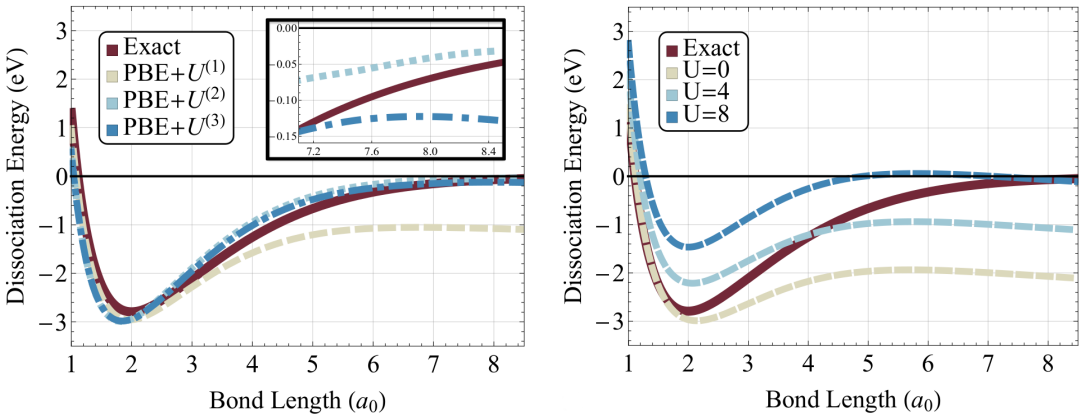


Figure 3.14: (a) Dissociation curves of H_2^+ as given by DFT and various self-consistent DFT + U schemes (differing in their definition of self-consistency). Crucially, U is calculated separately for each different bond length. (b) Dissociation curves generated using fixed values for U. This emphasises the importance of calculating U *ab initio*. Figures taken from Ref. 212.

3.6.1 The dissonance between local and global curvature

A substantial shortcoming of DFT + U as a correction to SIE is the distinction between local and global curvature. SIE gives rise to a curvature in the energy with respect to the total number of atoms in the entire system ("global curvature"), whereas DFT + U subtracts curvature with respect to the occupancy of a local subspace ("local curvature"), which are very different quantities. Ref. 186 demonstrated that for many systems correcting local curvature did address global curvature, but for others this was not the case at all. (This is why I have always stated that DFT + U can *partially* correct the SIE present in a system.)

If we restrict ourselves to a system where global and local curvature are the same, the results are promising. For instance, in calculations on stretched H_2^+ DFT + U can reproduce the dissociation curve (see Figure 3.14).[212] The challenge is to go beyond this point to multi-electron systems. One way of doing so would be to develop functionals with explicit derivative discontinuities — although this would be a dramatic departure from DFT + U-like theories.

3.6.2 Static correlation error

We need not stray too far from H_2^+ to run into issues. DFT exhibits two major inaccuracies when it comes to energies. SIE is one, and the second is static correlation error (SCE).[180] In the classification scheme of Cohen *et al.*, SCE is associated not with fractional charge but fractional spins, and can therefore be observed in systems as simple as He^{x+} ($0 < x < 2$), as shown in Figure 3.15. Furthermore, attempts to correct SIE using DFT + U have been shown to worsen SCE.[292]

Many of the issues discussed in this chapter stemmed from the conflict between (a) the desire to co-opt DFT + U to correct density curvature and (b) the fact that the DFT + U functional corrects spin channels separately. Perhaps a conceptually cleaner approach would be to consider

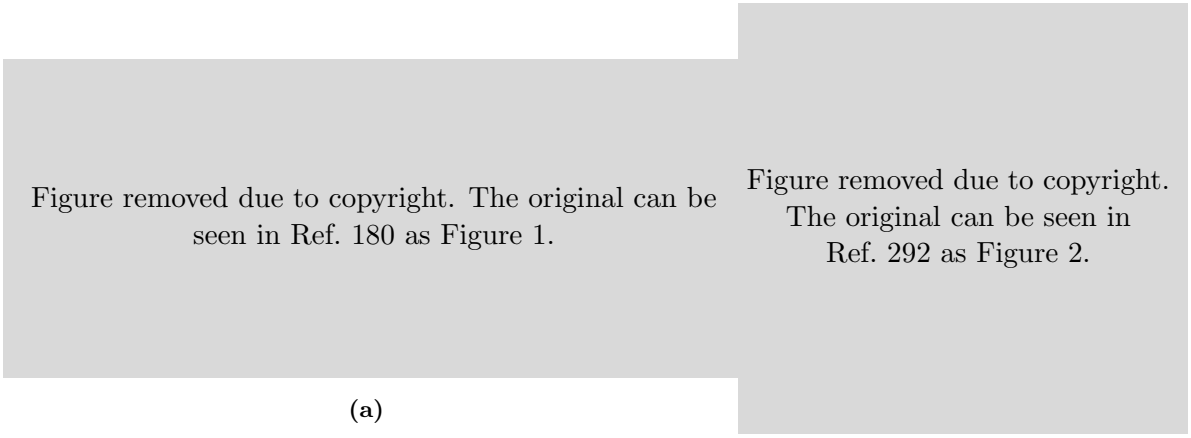


Figure 3.15: (a) The principal systematic errors in practical quantum-mechanical material simulation are self-interaction error (SIE, left) and static correlation error (SCE, right). Figure adapted from Ref. 180. (b) The curvature of the PBE energy of helium as a function of the spin-up (n_α) and spin-down (n_β) occupancies, making clearly visible both the SIE curvature (as $n = n_\alpha + n_\beta$ changes) and the SCE curvature (as $\mu = n_\alpha - n_\beta$ changes). Figure taken from Ref. 292.

alternative energy corrections of the form:

$$E_{\text{correction}} = \frac{1}{2}\frac{\delta^2 E}{\delta n^2}\left(n - n^2\right) - \frac{1}{2}\frac{\delta^2 E}{\delta \mu^2}\left(\mu^2\right) \tag{3.40}$$

which, by design, linearises E as a function of n (that is, removes SIE) and makes it independent of μ (removes SCE). In order to draw some parallels with the DFT + U functional, let us manipulate this a little:

$$\begin{aligned} E_{\text{correction}} =& \frac{1}{2}\frac{\delta^2 E}{\delta n^2}\left(n - n^2\right) + \frac{1}{2}\frac{\delta^2 E}{\delta \mu^2}\left(-\mu^2\right) \\ =& \frac{1}{2}\frac{\delta^2 E}{\delta n^2}\left(\sum_\sigma n^\sigma - \sum_{\sigma\sigma'} n^\sigma n^{\sigma'}\right) + \frac{1}{2}\frac{\delta^2 E}{\delta \mu^2}\left(2n^\uparrow n^\downarrow - \sum_\sigma (n^\sigma)^2\right) \\ =& \frac{1}{2}\left(\frac{\delta^2 E}{\delta n^2} + \frac{\delta^2 E}{\delta \mu^2}\right)\sum_\sigma \left(n^\sigma - (n^\sigma)^2\right) + \frac{1}{2}\left(-\frac{\delta^2 E}{\delta n^2} + \frac{\delta^2 E}{\delta \mu^2}\right)\sum_\sigma n^\sigma n^{-\sigma} - \frac{1}{2}\frac{\delta^2 E}{\delta \mu^2}\sum_\sigma n^\sigma. \end{aligned} \tag{3.41}$$

We can relate the two curvatures to Hubbard parameters:

$$U = \frac{\delta^2 E_{int}}{\delta n^2} = \frac{\delta v_{\text{Hxc}}}{\delta n} = \frac{1}{2}\frac{\delta v^\uparrow_{\text{Hxc}} + \delta v^\downarrow_{\text{Hxc}}}{\delta(n^\uparrow + n^\downarrow)}. \tag{3.42}$$

In the case of a closed-shell system, this simplifies due to the fact that $\delta n^\uparrow = \delta n^\downarrow$, $f^{\uparrow\uparrow} = f^{\downarrow\downarrow}$, and $f^{\uparrow\downarrow} = f^{\downarrow\uparrow}$:

$$U = \frac{1}{2}(f^{\uparrow\uparrow} + f^{\uparrow\downarrow}) \tag{3.43}$$

— that is, the average of the like- and unlike-spin interactions. Likewise for J,

$$J = -\frac{\delta^2 E_{\text{Hxc}}}{\delta\mu^2} = -\frac{1}{2}\frac{\delta v^{\uparrow}_{\text{Hxc}} - \delta v^{\downarrow}_{\text{Hxc}}}{\delta(n^{\uparrow} - n^{\downarrow})} \tag{3.44}$$

which in the closed-shell case becomes

$$J = -\frac{1}{2}(f^{\uparrow\uparrow} - f^{\uparrow\downarrow}) \tag{3.45}$$

that is, a difference between the like- and unlike interactions. Note that the effective U reduces to the like-spin interactions only, as expected ($U_{\text{eff}} = U - J = f^{\uparrow\uparrow}$). In this simplified case our revised energy correction functional becomes

$$\begin{aligned} E_{\text{correction}} = &-\frac{1}{2}f^{\uparrow\uparrow}\sum_{\sigma}(n^{\sigma})^2 - \frac{1}{2}f^{\uparrow\downarrow}\sum_{\sigma} n^{\sigma}n^{-\sigma} \\ &+ \frac{1}{2}(f^{\uparrow\uparrow} + f^{\uparrow\downarrow})\sum_{\sigma} n^{\sigma}. \end{aligned} \tag{3.46}$$

This makes the action of this particular Hubbard correction transparent: the first two terms remove any quadratic interactions within a subspace, and the final term installs linear behaviour in an average-like way. Perhaps a functional such as this may be better suited to counteracting SIE and SCE than DFT + U.

Following similar logic, Kulik and co-workers have recently constructed DFT + U-like functionals for lone atoms and homonuclear diatomic molecules, fitting their parameters (U, J, and several others besides) in order to explicitly recover the flat plane condition.[292,297] This they were able to do, demonstrating that generalised DFT + U-like functionals can simultaneously address both SIE and SCE and recover the flat plane condition. Admittedly in some cases this required exceptionally large corrective terms (*i.e.* $U, |J| > 20\,\text{eV}$ in many instances). It will be interesting to see how these corrective functionals affect system properties.

3.6.3 Koopman's compliance

An entirely different approach is taken by the recently-developed "Koopman's-compliant" functionals.[193,194,298,299] Instead of defining self-interaction in terms of the curvature of the total energy, they consider the individual Kohn-Sham eigenvalues

$$\varepsilon_{i\sigma} = \langle\phi_{i\sigma}|H_{KS}|\phi_{i\sigma}\rangle \tag{3.47}$$

and assert that the system is "self-interaction-free" if these energies are independent of the corresponding occupation of that orbital f_i. The construction of the resulting functional is straightforward:

$$E_{\text{KC}}[n] = E_{\text{DFT}} + \sum_{i\sigma}\alpha_{i\sigma}\Pi_{i\sigma} \tag{3.48}$$

where

$$\Pi_{i\sigma}(f_{i\sigma}) = -\int_0^{f_i}\langle\phi_{i\sigma}|H_{\text{DFT}}|\phi_{i\sigma}\rangle ds + f_{i\sigma}\int_0^1\langle\phi_{i\sigma}|H_{\text{DFT}}|\phi_{i\sigma}\rangle \tag{3.49}$$

which transparently removes the non-linear behaviour of the underlying functional and replaces it with a linear Koopman's term that interpolates between integer occupations. (The $\{\alpha_{i\sigma}\}$ terms are screening coefficients that must be calculated in a similar manner to Hubbard/Hund's parameters in DFT + U.)

One complication of these functionals is that the energy correction is orbital-dependent, taking us beyond the realm of density functional theories. Here, the "variational orbitals" that minimise the functional are different to the "canonical orbitals" that diagonalise the Hamiltonian: the variational orbitals typically become localised to lower the total energy while the canonical orbitals are typically delocalised and interpreted much like KS orbitals.[300]

This class of functionals performs well, yielding accuracy comparable to GW at a fraction of the computational cost[301] — another fine example of the importance of self-interaction and the usefulness of DFT + U-like theory looking forward.

Chapter 4

Dynamical mean field theory

4.1 Introduction

Leaving behind the DFT + U method of the previous chapter, an alternative approach for treating correlated materials is dynamical mean field theory (DMFT). In order to motivate this method, let us revisit some of the ideas underpinning DFT.

Many physical systems can be well described in an independent-particle framework. In this approach, valence electrons are described using a wave-like picture, with Bloch states representing individual wavefunctions. This works well for systems where the electrons are highly itinerant, but it breaks down if electrons instead have a tendency to localise (as in the case of the $U \gg t$ limit of the Hubbard model discussed in Subsection 3.1.1). In these cases, it is better to conceptualise electrons as particles, with wavefunctions localised on individual atomic sites. Strongly-correlated systems often sit at the boundary of these two extremes, with the electrons "hesitating" between itinerant and localised behaviour.[302]

DFT — with its auxiliary non-interacting system — leans heavily on the independent-particle picture. While formally one can prove that the ground-state density uniquely determines the Hamiltonian of a system (and therefore every single system property, both ground- and excited-state), in practice approximate DFT is generally most successful for systems where the electrons are itinerant.

DMFT was developed in an attempt to restore aspects of the "atom-like" picture. Developed by Metzner, Vollhardt, Georges, Kotliar, and others, DMFT[303,304] is a Green's function[305] method that maps the electronic problem onto an impurity Hamiltonian with a self-consistency condition. This model Hamiltonian includes Hubbard and Hund's-like terms, much like the Hubbard model discussed earlier, and local quantum fluctuations are fully taken into account, allowing DMFT to capture complex electronic behaviour such as the intermediate three-peak states of the Mott transition, the transfer of spectral weight, and the finite lifetime of excitations.[306]

Like DFT + U, DMFT can be used in conjunction with DFT to treat localised regions where correlation is important.[150] In the case of DFT + U, each correlated subspace was subjected to an additional potential; in DFT + DMFT, we go substantially further, subjecting these subspaces

to a full Green's function treatment.† The fact that DMFT can be selectively applied is critical, as DMFT alone is prohibitively expensive for studying most realistic systems.

In the past decade, numerous codes have been written to add DMFT functionality to existing DFT packages. These include EDMFTF[307,308] and DFTTools[309] on top of Wien2K,[310] EDMFTF[308] on top of VASP,[311–313] DCore[314] on top of Quantum Espresso[315] and OpenMX,[218,316] TOSCAM[317] on top of CASTEP,[318,319] Amulet[320] on top of Quantum Espresso[315] and Elk,[321] and ComDMFT[322] on top of FlapwMBPT.[323,324] Many of these make use of stand-alone libraries such as TRIQS,[325] ALPS,[326] iQIST,[327] or W2dynamics.[328] This chapter introduces an overhauled implementation of TOSCAM on top of ONETEP. In contrast to the packages mentioned above, this approach uniquely enables us to perform DMFT calculations on large and aperiodic systems such as nanoparticles and metalloproteins.

This ONETEP+TOSCAM code has already seen success: it has been used to explain the insulating M_1 phase of vanadium dioxide,[77] to demonstrate the importance of Hund's coupling in the binding energetics of myoglobin,[78,79] and to reveal the super-exchange mechanism in the dicopper oxo-bridge of haemocyanin and tyrosinase (see Chapter 5).[6] But until now it has not been available to the scientific community at large. As part of my PhD project I have substantially rewritten the DMFT module in ONETEP, which is now included in the latest release (version 5.0). This chapter presents an overview of this methodology, its implementation, and an example of its application to an iron porphyrin system.

4.2 Theory

A DMFT calculation involves the self-consistent calculation of the Green's function $G^{\alpha\beta}(\omega)$ (ω here may be $\omega + i0^+$ or $i\omega_n$ if operating in the finite-temperature Matsubara representation) and the self-energy $\Sigma_{\alpha\beta}$, which are related via

$$G^{\alpha\beta}(\omega) = \left[(\omega+\mu)S - H - \Sigma(\omega)\right]_{\alpha\beta}^{-1} \tag{4.1}$$

where μ is the chemical potential and $S_{\alpha\beta}$ is the NGWF overlap matrix (that is, $S_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle$), which of course is non-diagonal.

Treating most physical systems at the DMFT level would usually be prohibitively expensive (I will explain why later). The DFT + DMFT scheme takes advantage of the fact that strong electronic correlation is often confined to identifiable localised subspaces (for instance, the $3d$ orbitals of a transition metal atom), with the remainder of the system having a delocalised, free-electron character. In such systems, the correlated subspaces can be treated at the DMFT level, while DFT alone should be sufficient everywhere else.

Correlated subspaces are typically defined via a set of local, fixed, atom-centred, spin-independent, and orthogonal orbitals $\{\varphi_m^I\}$. (Here, I is the atom index and m is an orbital index.) In ONETEP, these are defined using the same the Hubbard projectors as in DFT + U (that is, they are defined using PAOs: the Kohn-Sham solutions to the isolated pseudopotential of the correlated atom[235,250,251]).

†For a brief overview of Green's function formalism, see Appendix A.6.

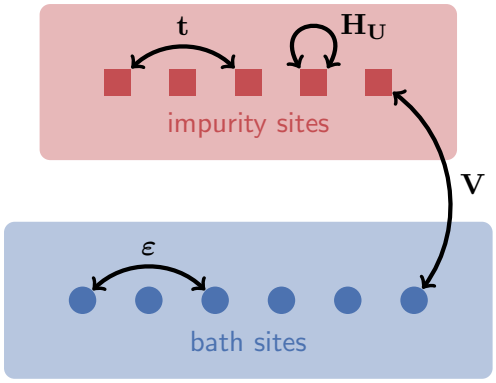


Figure 4.1: Schematic diagram of an Anderson impurity model, showing the impurity sites (red squares), bath sites (blue circles) and the interaction parameters.

4.2.1 The Anderson impurity model

In order to efficiently find a self-consistent solution to Equation 4.1, DMFT relies on mapping correlated subspaces to auxiliary Anderson impurity models (AIMs). An AIM is a simplified Hamiltonian that describes the interaction of a number of sites (known as impurity sites) with a bath of additional electronic levels:

$$\hat{H} = \underbrace{\sum_{ij\sigma}(\varepsilon_{ij} - \mu)\hat{c}^{\dagger}_{i\sigma}\hat{c}_{j\sigma}}_{\hat{H}_{\text{bath}}} + \underbrace{\sum_{im\sigma}\left(V_{mi}\hat{f}^{\dagger}_{m\sigma}\hat{c}_{i\sigma} + h.c.\right)}_{\hat{H}_{\text{mix}}} + \underbrace{\sum_{mm'\sigma}(t_{mm'} - \mu)\hat{f}^{\dagger}_{m\sigma}\hat{f}_{m'\sigma} + \hat{H}_U}_{\hat{H}_{\text{loc}}} \tag{4.2}$$

where $\hat{H}_{\text{bath}}$ describes the non-correlated behaviour of the bath (parametrised by the hopping matrix ε_{ij}), $\hat{H}_{\text{loc}}$ the impurity (parametrised by the impurity hopping $t_{mm'}$ and the interaction Hamiltonian $\hat{H}_U$), and $\hat{H}_{\text{mix}}$ the coupling between the two (parametrised by V_{mi}). The bath and impurity sites have a shared chemical potential μ, and $\hat{c}/\hat{f}$ are the annihilation operators for the bath/impurity. The convention throughout will be that Greek indices correspond to NGWFs, m and m' to Hubbard subspaces and their corresponding impurity sites, and Latin indices to bath sites. σ is the spin index.

The non-interacting Anderson model (*i.e.* $H_U = 0$) has the Green's function

$$G^0_{\text{tot}}(\omega) = \frac{1}{\omega + \mu - T} \tag{4.3}$$

where the full hopping matrix is of the block matrix form

$$T = \begin{pmatrix} t & V \\ V^{\dagger} & \varepsilon \end{pmatrix}. \tag{4.4}$$

It follows that the (non-interacting) impurity Green's function — that is, the top-left-hand block of $G^0_{\text{tot}}(\omega)$ — simplifies to

$$G^0_{\text{imp}}(\omega)^{-1} = \omega + \mu - t - \Delta_{\text{imp}}(\omega), \tag{4.5}$$

where

$$\Delta_{\mathrm{imp}_{mm'}}(\omega) = V_{mi}\left(\frac{1}{\omega+\mu-\varepsilon}\right)_{ij} V^{\dagger}_{jm'}$$

is the so-called impurity *hybridisation function*. This quantity is of particular importance because it encapsulates all of the contributions of the bath sites to the physics of the impurity sites; the AIM impurity Green's function is given by

$$G_{\mathrm{imp}}(\omega)^{-1} = G^{0}_{\mathrm{imp}}(\omega)^{-1} - \Sigma(\omega) = \omega + \mu - t - \Delta_{\mathrm{imp}}(\omega) - \Sigma_{\mathrm{imp}}(\omega). \tag{4.6}$$

4.2.2 A DMFT calculation

This subsection will walk through the steps in a standard DMFT calculation as performed in TOSCAM + ONETEP. It is important to note that DMFT typically invokes a mean field approximation across multiple correlated sites (hence *dynamical "mean field" theory*), an approach that only becomes exact in the limit of infinite coordination (or equivalently, dimensions). This is not the case in our following real-space approach, where instead correlated sites are typically treated via a (possibly multi-site) AIM.

Mapping physical systems to an impurity model

DFT + DMFT utilises an AIM as an auxiliary system: the AIM parameters $\{V_{mi}\}$, $\{\varepsilon_{ij}\}$, and $\{t_{mm'}\}$ are chosen such that the resulting model Hamiltonian reproduces the physics of the real system as closely as possible. This mapping proceeds as follows. Firstly, the Kohn-Sham Hamiltonian, an estimate of the system self-energy (zero is a reasonable starting point), and a total Green's function (obtained via Equation 4.1) are each projected onto the correlated subspaces. For instance, the local Green's function is given by

$$\tilde{G}^{I}_{mm'}(\omega) = W^{I}_{m\alpha} G^{\alpha\beta}(\omega) (W^{I})^{\dagger}_{\beta m'} \tag{4.7}$$

where $W^{I}_{m\alpha} = \langle \varphi^{I}_{m} | \phi_{\alpha} \rangle$ is the overlap of the NGWFs and the Hubbard projectors. In a similar manner one can obtain the projected self energy $\tilde{\Sigma}^{I}(\omega)$ and the projected Kohn-Sham Hamiltonian $\tilde{H}^{I}$.

The impurity hopping parameters $t_{mm'}$ for the auxiliary AIM are set equal to the projected Hamiltonian. Meanwhile, in order to determine $\{V_{mi}\}$ and $\{\varepsilon_{ij}\}$, we define the local hybridisation function for our physical system

$$\tilde{\Delta}^{I}(\omega) = \omega + \mu - (\tilde{G}^{I})^{-1}(\omega) - \tilde{\Sigma}^{I}(\omega) - \tilde{H}^{I} \tag{4.8}$$

which is analogous to the definition of the impurity hybridisation function (Equation 4.6). We choose the impurity model bath parameters such that the AIM hybridisation function matches this local hybridisation function as closely as possible. This is done by minimising the distance function

$$d(\mathbf{V}, \boldsymbol{\varepsilon}) = \sum_{\omega<\omega_c} \frac{1}{\omega^{\gamma}} \left| \Delta_{\mathrm{imp}}(\omega) - \tilde{\Delta}^{I}(\omega) \right|^{2} \tag{4.9}$$

using a conjugate gradient (CG), BFGS, or similar minimisation algorithm. Here, ω_c is a cut-off frequency and γ is a user-specified parameter that can allow for the preferential weighting of agreement at low frequencies.

In order to complete the construction of the auxiliary AIM Hamiltonian we choose H_U to be of the Slater-Kanamori form[329,330]

$$\begin{aligned}\hat{H}_U = U \sum_m \hat{n}_{m\uparrow}\hat{n}_{m\downarrow} + \left(U' - \frac{J}{2}\right) \sum_{m>m'} \hat{n}_m \hat{n}_{m'} \\ -J \sum_{m>m'} (2\hat{\mathbf{S}}_m \hat{\mathbf{S}}_{m'} + \hat{f}^\dagger_{m\uparrow}\hat{f}^\dagger_{m\downarrow}\hat{f}_{m'\uparrow}\hat{f}_{m'\downarrow}).\end{aligned} \tag{4.10}$$

This Hamiltonian is well-suited to capturing multiplet properties of low energy states.[331] Its first term describes intra-orbital Coulomb repulsion. The second describes the inter-orbital repulsion, with $U' = U - 2J$ further renormalised by the Hund's coupling to ensure the rotational invariance of the Hamiltonian. The third and final term captures the Hund's exchange coupling; $\hat{\mathbf{S}}_m$ is the spin of orbital m, given by $(\hat{\mathbf{S}}_m)_i = \frac{1}{2}\sum_{\sigma\sigma'} \hat{f}^\dagger_{m\sigma}(\mathbf{s}_i)_{\sigma\sigma'}\hat{f}_{m\sigma'}$ via the Pauli spin matrices $\{\mathbf{s}_i\}$. The Hubbard parameter U and Hund's coupling J are user-specified parameters that, in principle, could be obtained via linear response[178] but are often chosen empirically or treated as variational parameters.

Now that we have defined ε, V, t, and H_U, the mapping of a real system to an auxiliary AIM is complete. In theory, this mapping can be exact: as long as $\Delta_{\text{imp}}(\omega)$ and $\tilde{\Delta}^I(\omega)$ match exactly, $G_{\text{imp}}(\omega)$ and $\tilde{G}^I(\omega)$ will also. Getting this mapping right is therefore of the utmost importance.

Solving the AIM

Having constructed the AIM Hamiltonian H_{AIM}, the next step is to calculate the Green's function of the AIM (known as the *impurity Green's function*):

$$\begin{aligned}G_{\text{imp}_{mm'}}(\omega) &= \int_{-\infty}^{\infty} e^{i\omega t} G_{\text{imp}_{mm'}}(t)\, dt \\ &= -i \int_0^\infty e^{i\omega t} \langle e^{i\hat{H}t}\hat{c}_m e^{-i\hat{H}t}, \hat{c}^\dagger_{m'}\rangle\, dt \\ &= -i \left(\left\langle \hat{c}_m \int_0^\infty e^{i(\omega - (\hat{H}-E_0))t}\, dt\, \hat{c}^\dagger_{m'} \right\rangle + \left\langle \hat{c}^\dagger_{m'} \int_0^\infty e^{i(\omega + (\hat{H}-E_0))t}\, dt\, \hat{c}_m \right\rangle \right) \\ &= \left\langle \hat{c}_m \frac{1}{\omega - (\hat{H}-E_0)} \hat{c}^\dagger_{m'} \right\rangle + \left\langle \hat{c}^\dagger_{m'} \frac{1}{\omega + (\hat{H}-E_0)} \hat{c}_m \right\rangle\end{aligned} \tag{4.11}$$

where $\langle\ \bullet\ \rangle$ is the thermodynamic average, which at zero temperature becomes $\langle\psi_0|\ \bullet\ |\psi_0\rangle$.

Resolving Equation 4.11 is highly expensive, and becomes one of the most substantial computational barriers in a DMFT calculation. If there are m bath sites and n impurity orbitals, the Hilbert space of this problem scales as 4^{m+n}.[†] This is far larger than any of the other matrix inversions that we need to calculate during the DMFT loop (for instance, $G^{\alpha\beta}$ is only as large as the number of Kohn-Sham orbitals, which in turn will be of the order of the number

[†]For a system containing a single transition metal there will be five impurity orbitals (one for each $3d$ orbital) and then typically six to eight bath sites.

of electrons in the physical system — typically several thousand at most). There are a multitude of approaches for obtaining G_{imp}, such as exact diagonalisation (ED) and continuous time Monte Carlo algorithms. The calculations in this work employ ED via the Lanczos algorithm to evaluate Equation 4.6, a process which is explained in detail n Appendix A.7.

Given a solution G_{imp} (obtained via ED or otherwise), the impurity self-energy can then be obtained via

$$\Sigma(\omega) = [G^0_{\mathrm{imp}}]^{-1}(\omega) - G^{-1}_{\mathrm{imp}}(\omega) \tag{4.12}$$

where the non-interacting impurity Green's function is given by Equation 4.5. Note that this operation is far less expensive than Equation 4.6 because these matrices are only $m \times m$ in size.

Upfolding and double-counting

Having obtained the impurity Green's function Σ^I for each AIM, the final step is to upfold this result to the complete physical system. Since the original DFT solution already contains the influence of the Coulomb interaction to some degree, double-counting becomes an issue. A popular form of the correction is

$$E_{\mathrm{DC}} = \frac{U^{\mathrm{av}}}{2} n\,(n-1) - \frac{J}{2}\sum_{\sigma} n_\sigma(n_\sigma - 1) \tag{4.13}$$

where n is the total occupancy of the subspace, and

$$U^{\mathrm{av}} = \frac{U + 2(N-1)U'}{2N-1} \tag{4.14}$$

with N being the number of orbitals spanning the correlated subspace (and recall that $U' = U - 2J$).[331] This double-counting is derived by attempting to subtract the DFT contributions in an average way; U^{av} is the average of the intra- and inter-orbital Coulomb parameters.

The self-energy is upfolded to the NGWF basis via

$$\Sigma_{\alpha\beta} = \sum_I W^I_{m\alpha}(\Sigma^{I\,mm'} - E_{\mathrm{DC}}\delta^{mm'})W^I_{m'\beta} \tag{4.15}$$

— and with that, we are back where we started, having generated a new estimate of the self-energy $\Sigma_{\alpha\beta}$ for the full system.

The DMFT algorithm in full

To summarise, the scheme is as follows:

1. perform a DFT calculation to construct the system Hamiltonian
2. initialise the self-energy as $\Sigma_{\alpha\beta}(\omega) = 0$
3. obtain the Green's function for the full system (equation 4.1)
4. project the total Green's function and self energy onto the I^{th} Hubbard subspace to obtain the corresponding local quantities (Equation 4.7)

5. calculate the local hybridisation function (Equation 4.8)
6. find the bath parameters ε_{ij} and V_{mi} such that the AIM hybridisation function (equation 4.2.1) matches the local hybridisation function found above
7. explicitly solve the AIM Hamiltonian to obtain the impurity Green's function (equation 4.11)
8. update the impurity self-interaction (equation 4.12)
9. upfold the self-energies from each correlated subspace to obtain the total self-interaction (equation 4.15)

Note that if we only have one correlated site in our system (as is the case for many of the biological systems we will come across in this thesis), this mapping is exact, and the local lattice Green's function at step 9 will already match the impurity Green's function.

This is not the case for bulk systems. There, the mean field approximation that we adopt means that the self-energy of a correlated site is also inherited by the "bath" *i.e.* one would typically solve a single Anderson impurity problem but then in Equation 4.15, the index I would run over all correlated sites. This means that after step 9 we must return to step 3, and repeat this loop until the local lattice and impurity Green's functions match.

Once the calculation is converged, we can extract system properties from the Green's function (such as the density of states and the optical absorption). One can also apply standard ONETEP analysis techniques to the electron density (such as natural bonding orbital analysis). These techniques will be demonstrated in Section 4.3.

4.2.3 Extensions

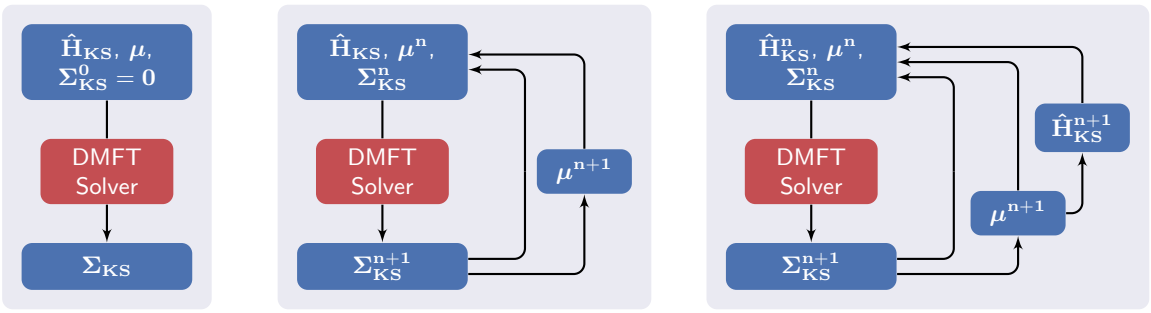
There are several possible extensions to the theory described thus far. These are not essential but often useful.

Enlarged AIM via cluster perturbation theory

If an AIM has too few bath sites at its disposal, it will be insufficiently flexible to fit a given local hybridisation function. The brute-force approach would be to increase the number of bath sites, but in practice the number of bath sites is severely limited due to the exponential growth of Hilbert space with respect to the AIM's total number of sites (bath and impurity). To overcome this barrier, a secondary set of bath levels are coupled to the primary bath levels via cluster perturbation theory (CPT). By indirectly including these sites, the AIM system acquires extra flexibility without expanding the Hilbert space, resulting in a dramatic drop in the distance function. For more details, see Ref. 332.

Self-consistency

For a system with a single correlated site, there is no feedback from the self energy to the hybridisation function, and — provided the AIM is sufficiently representative — the DMFT algorithm



(a) single-shot **(b)** charge-conserving **(c)** self-consistent

Figure 4.2: The three DMFT schemes, in increasing order of complexity.

will converge in a single step. (In this case the algorithm is not a mean-field approximation, but exact.) This scheme is shown in Fig. 4.2a.

However, there are a number of reasons why we may not be content with the resulting solution. For a start, the total number of electrons in the system is related to the total retarded Green's function via

$$N = \int d\omega\, \rho^{\alpha\beta}(\omega) S_{\alpha\beta}; \qquad \rho^{\alpha\beta}(\omega) = \frac{1}{2\pi i}\left(G^{\alpha\beta}(\omega) - G^{\alpha\beta\dagger}(\omega)\right), \tag{4.16}$$

where $\rho^{\alpha\beta}(\omega)$ is the basis-resolved DMFT spectral density matrix.

There is no reason *a priori* why the Green's function, updated via the DMFT loop, should yield the same number of electrons as we started with — in fact, this is almost never the case. For this reason, charge conservation can optionally be enforced by adjusting μ so that $\int_{-\infty}^{\mu} \rho(\omega) = N$. This update is done during each DMFT cycle, which means that our total Green's function (now adjusted by our altered μ) will not necessarily be consistent with the self energy — and consequently more than one DMFT loop will likely be required to iterate to self-consistency (Fig. 4.2b). We will refer to this as "charge-conserving" DMFT; it will be the primary approach deployed in Chapter 5.

Finally, in the DFT formalism, the Hamiltonian is a functional of the density. It could be argued that if we are to be fully self-consistent, whenever the density changes the Hamiltonian should be updated accordingly. In this scheme, one iterates until Σ, H, and μ all converge (Fig. 4.2c). This we will refer to as "self-consistent" DMFT. We use Pulay mixing[333,334] to update the Hamiltonian (via the density kernel) and the self-energy. Performing this double-loop naturally makes the calculations much more expensive, but they remain feasible. This approach was taken in Refs. 335–337, for example.

4.2.4 Practical implementation

In our implementation, ONETEP and TOSCAM are responsible for separate sections of the DMFT loop, as shown in Figure 4.3. As the calculation proceeds, these two programs alternate, with the entire procedure being driven by an overarching script.

This splitting makes our algorithm highly amenable to parallelisation: parallel TOSCAM

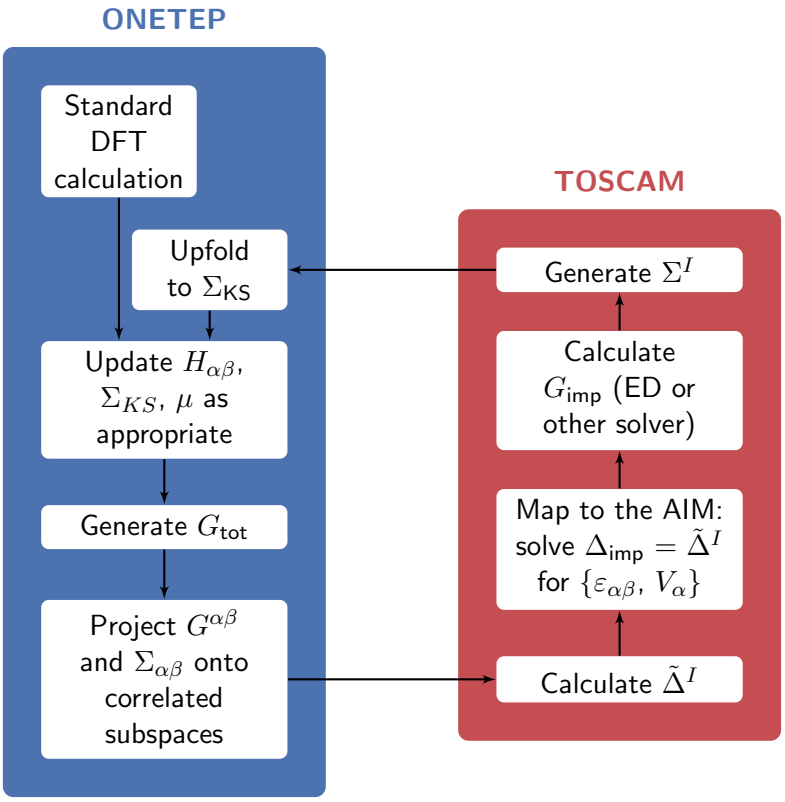


Figure 4.3: A simplified DMFT loop, demonstrating which program (ONETEP or TOSCAM) is responsible for which step.

instances can consider different correlated subspaces in isolation. (That is, a system with many correlated sites is embarrassingly parallel if inter-site correlation can be neglected.) By design, the AIM solver in TOSCAM is as modular as possible. This allows it to be easily interchanged with other solvers that have been independently developed.

A ONETEP+TOSCAM implementation already existed prior to my PhD (and was used to generate the results of Refs. 77–79). However, that version was never integrated into the official ONETEP repository, so subsequent development of ONETEP had made the two codes incompatible. Furthermore, the implementation had a raft of external dependencies[†] that complicated its compilation and distribution. I overhauled this interface, which involved (a) restoring compatibility with the active version of ONETEP, (b) reducing the number of external dependencies as much as possible, to those that we can expect HPC systems to have installed (or are sufficiently lightweight and open-source that we can distribute them directly),[‡] and (c) extensively tidied the code, removing ambiguous syntax, adding documentation, and automating its compilation via a `makefile` procedure.

This overhauled interface will be made freely available for download, along with documentation, tutorials, and example input files. (ONETEP must be obtained separately.)

[†]Specifically, BLAS, LAPACK, FFTW, ARPACK, BLZPACK, DIERICKX, Fields, GSL, PGPLOT, and SLATEC.

[‡]The new interface has external dependencies on BLAS, LAPACK, FFTW, and has an in-house copy of DIERICKX.

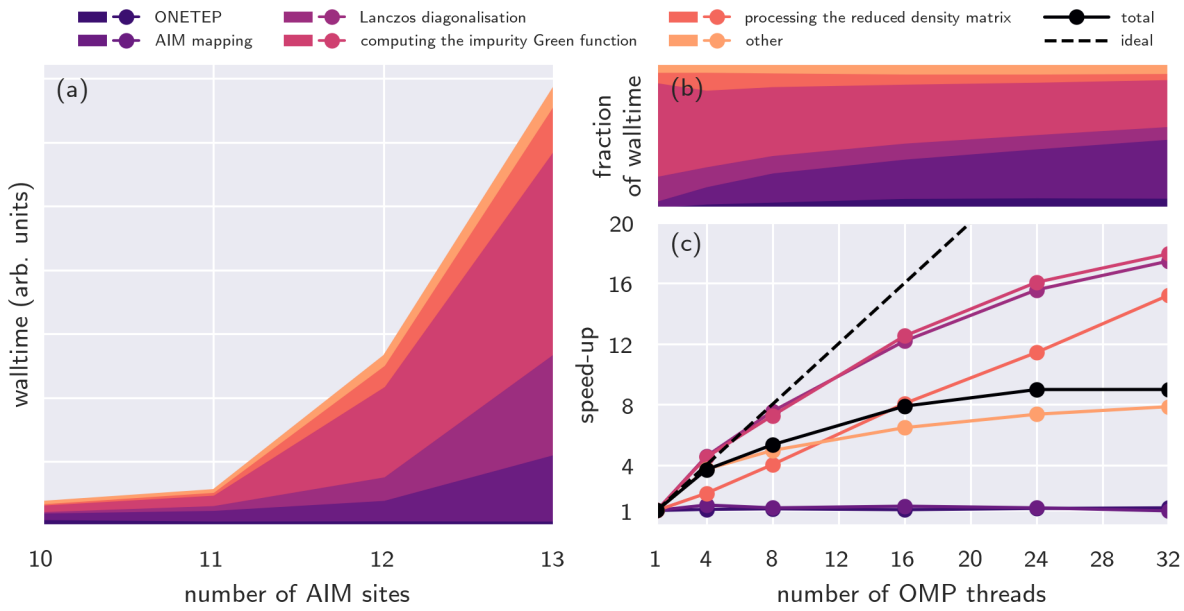


Figure 4.4: The scaling of ONETEP+TOSCAM for calculations on iron porphyrin (see Section 4.3 for details). (a) The scaling with respect to the number of AIM sites; (b) and (c) the fractional wall time and the speed-up with respect to the number of OpenMP threads. "Lanczos diagonalisation" and "computing the impurity Green's function" are two steps involved in solving the AIM; for details refer to Subsection A.7.2.

4.2.5 Scaling

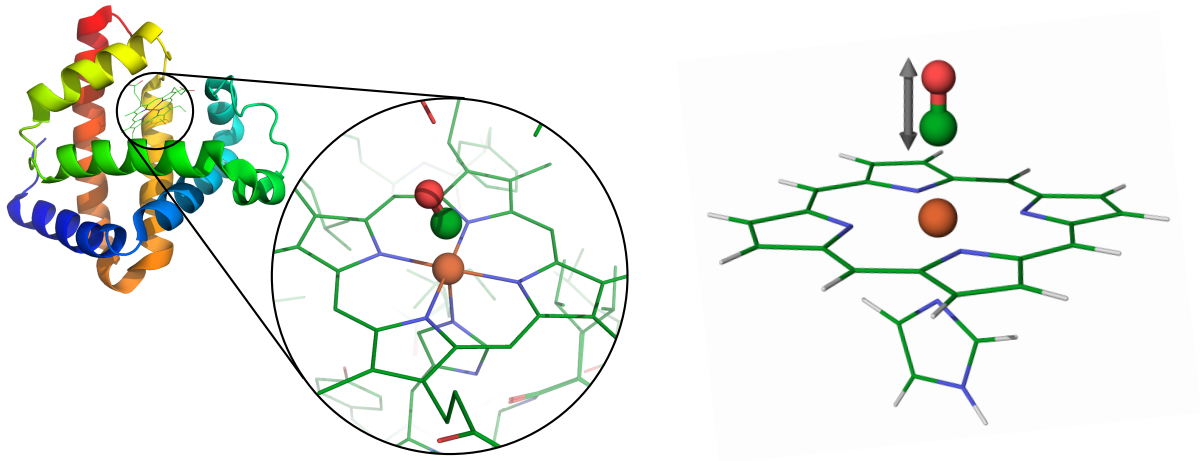
One of our primary considerations is how ONETEP+TOSCAM calculations scale. As discussed already, obtaining the Green's function of the AIM scales very poorly with the number of AIM sites. This is shown in Figure 4.4a. We are not entirely in a position to dictate the number of AIM sites: a $3d$ correlated site is represented as a five-site impurity, and typically we need to include at least six bath sites to give the AIM sufficient flexibility to fit the hybridisation function. Subsequent chapters will explore methods for side-stepping this requirement.

To some extent, poor scaling can be overcome by efficient parallelisation. Both ONETEP and TOSCAM employ hybrid MPI and OpenMP parallelisation schemes. ONETEP's parallelisation is highly optimised. Individual atoms are distributed across MPI threads, with lower-level computationally-intensive operations (including 3D FFT box operations, sparse matrix algebra operations, calculation of integrals, and Ewald summation) being further parallelised with OpenMP.[338]

In the current implementation of TOSCAM, individual MPI tasks are responsible for individual correlated atoms. For systems where we have only one unique correlated atom, MPI becomes redundant. Meanwhile, OpenMP is deployed to speed up lower-level operations (see Figure 4.4b and c).

4.3 Iron porphyrin

In the final section of this chapter, to demonstrate the use of the ONETEP+TOSCAM interface, I present some calculations on an archetypal strongly-correlated system, FePImCO (Figure 4.5).



(a) (b)

Figure 4.5: (a) Carboxymyoglobin, showing the iron binding site.[80] (b) The model complex studied in this chapter: iron porphyrin with axial imidazole and carbon monoxide ligands. Hydrogen, carbon, nitrogen, oxygen and iron atoms are are shown in white, green, blue, red, and orange respectively.

By translating the carbon monoxide molecule perpendicular to the porphyrin plane, we will model the photodissociation of carboxymyoglobin. Myoglobin is one of the most ubiquitous metalloproteins. Previous studies have successfully applied DMFT in order to rationalise its binding energetics,[78,79] so it will serve as a natural starting point for this thesis's exploration of the DMFT method. There are also unresolved questions surrounding the process of carbon monoxide photodissociation (as I will explain below).

4.3.1 Computational details

All DFT calculations were performed using a modified copy of ONETEP.[22,172,173,195,234,235†] All calculations used the PBE xc functional,[136] were spin-unpolarised,[‡] and had an energy cut-off of 908 eV. There were 13 NGWFs on the iron atom, four on each carbon, nitrogen, and oxygen, and one on each hydrogen. All NGWFs had 6.6 Å cut-off radii. Open boundary conditions were achieved using a padded cell and a spherical Coulomb cut-off.[238] Scalar relativistic pseudopotentials were used, generated in-house using OPIUM,[125,240–246] and the Hubbard projectors were constructed from the Kohn-Sham solutions for a lone iron pseudopotential.[235]

The bound structure was taken from Ref. 339, which had been optimised with the B3LYP functional. The other structures were generated by simply translating the carbon monoxide molecule in steps of 0.1Å, without subsequent geometry optimisation of the rest of the system. (An ideal analysis would involve a constrained geometry optimisation, to account for effects such as doming.)

Both charge-conserving and self-consistent calculations were performed, using enlarged AIM Hamiltonians via the CPT extension. Six (or sometimes seven) bath orbitals proved necessary for the AIM to be able to fit the hybridisation function using the BFGS minimisation algorithm,

[†]Version 4.3; those modifications were subsequently included in ONETEP 5.0.
[‡]That is, $n^{\uparrow}(\mathbf{r}) = n^{\downarrow}(\mathbf{r})$.

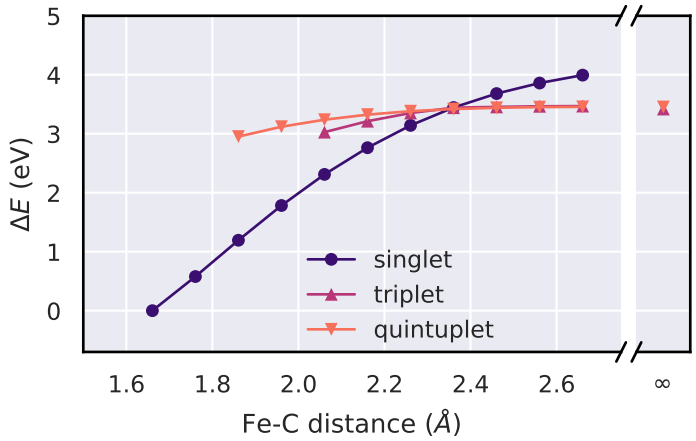


Figure 4.6: Spin state energies as given by DFT. For FePImCO the singlet state is correctly preferred, but for FePIm the triplet is wrongly preferred, albeit only very marginally (by 34 meV).

and the AIM was solved using an ED Lanczos solver. Values of $U = 4.0\,\text{eV}$ and $J = 0.7\,\text{eV}$ were used in the AIM Hamiltonian.

4.3.2 The quantum-mechanical state of the *3d* iron subspace

A lot of effort (largely in the quantum chemistry community) has been made to correctly predict the spin state of Fe(II)P with (and without) a variety of axial ligands. These range from decades-old Hartree-Fock calculations to recent FCIQMC studies.[340–345] FePImCO is one of the simpler cases, with a singlet state universally predicted. Meanwhile, iron-porphyrin with an axial imidazole ligand (FePIm) has proven to be more of a challenge. Experiment characterises FePIm as a quintet. Semi-local DFT wrongly predicts it to be a triplet (as shown in Figure 4.6). DFT + U remedies this,[282] as does HF.[340]

To start, we will examine the charge transfer that takes place during CO dissociation in the DFT + DMFT picture. The Fe atom in FePIm is formally in the 2+ state (d^6). When it binds CO, it moves closer to 1+ (d^7) due to ligand-to-metal charge transfer. This is corroborated by our DFT+DMFT calculations: the occupancy of the $3d$ subspace can be calculated via

$$n_{3d} = \frac{1}{2\pi i} \sum_m \int \mathrm{d}\omega \, {G_{\text{imp}}}_{mm}(\omega) - {G_{\text{imp}}}_{mm}^{\dagger}(\omega). \tag{4.17}$$

This is plotted in Figure 4.7a. The unbinding is plainly visible in a sudden step in the total occupancy, at the same distance that DFT predicted the low-to-high-spin crossover (refer back to Figure 4.6). The effect of DMFT is especially pronounced at large Fe-C distances, where it drives the subspace occupancy towards the expected formal d^6 configuration. (In some sense, DMFT restores the quantised nature of the electrons in the correlated subspace.)

As a means of analysing the spin state of the iron atom during the dissociation process with DMFT, we construct the reduced density matrix

$$\hat{\rho} = \sum_i e^{-\beta E_i} \mathrm{Tr}_B[|i\rangle\langle i|], \tag{4.18}$$

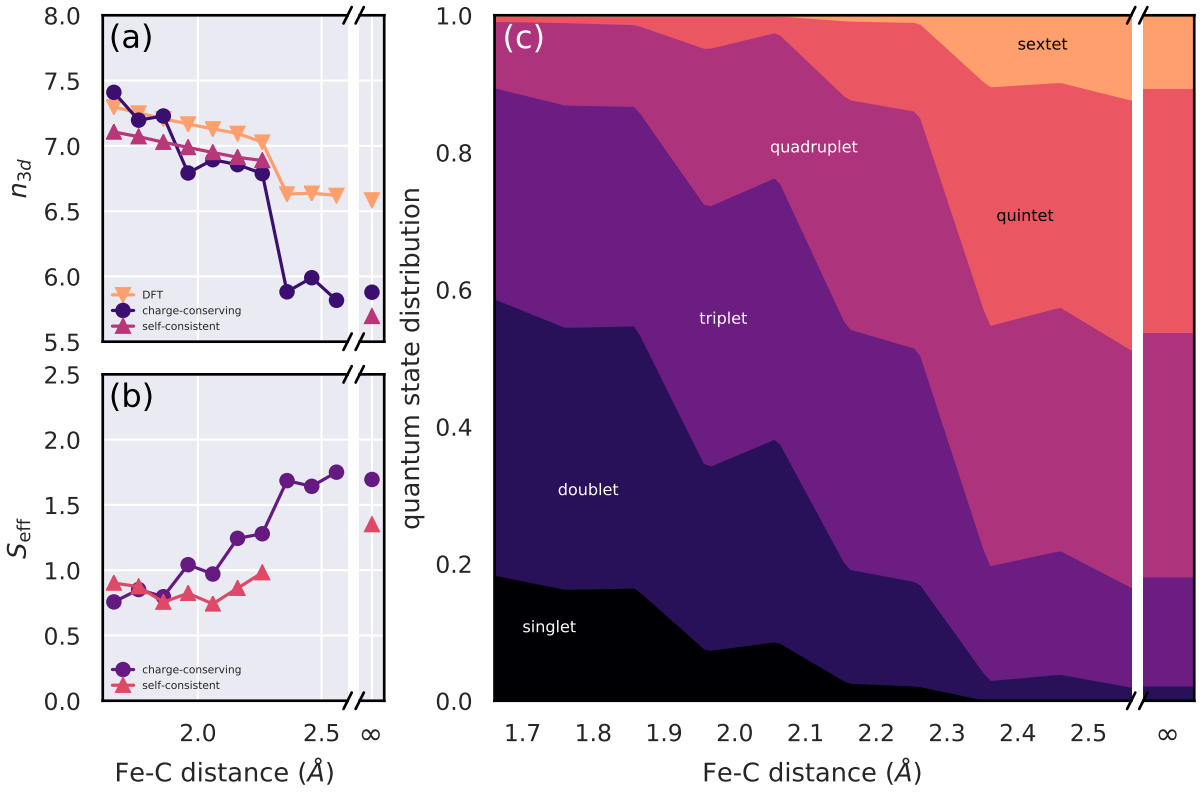


Figure 4.7: The electronic state of iron in FePImCO during CO dissociation. (a) The total occupancy of the Fe-3*d* subspace as given by DFT and two different DMFT schemes. Unfortunately self-consistent DMFT calculations proved very difficult to converge beyond the low-to-high spin transition, so these results have been excluded throughout. Below this transition, the two methods qualitatively agree. (b) The effective spin S_{eff} of the reduced density matrix, defined via $\text{Tr}[\hat{S}^2\hat{\rho}] = \hbar^2 S_{\text{eff}}(S_{\text{eff}}+1)$. (c) The decomposition of the reduced density matrix by spin state. The colours correspond to the respective weights of the different contributions; if a colour occupied all the vertical axis, it would mean that all eigenvectors of the density matrix are in that particular quantum sector.

where we take the partial trace of the low-lying eigenstates of the AIM over the bath degrees of freedom, leaving a mixed density operator for the impurity alone. It is then straightforward to calculate the expectation value of $\hat{S}^2 = \sum_{i,j} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j$ and extract the effective spin S_{eff} (Figure 4.7b). Here we can see that at large distances we approach the quintet $S_{\text{eff}} = 2$. At small distances we are closer to the triplet value $S_{\text{eff}} = 1$. Note that this does not mean that DMFT has failed to predict that FePImCO is a singlet. Rather, this result is compatible with (but does not confirm the existence of) a singlet forming across the Fe-CO bond. By limiting ourselves to the Fe subspace we cannot detect such a singlet.

To inspect the reduced density matrix in more detail, one can construct the spin-projector

$$\hat{P}_S = \sum_{s \in S} |s\rangle\langle s| \tag{4.19}$$

as the sum of the eigenstates $|s\rangle$ of the operator $\hat{S}^2$ with eigenvalue $S(S+1)$. This allows us to evaluate the fraction of the reduced density matrix in singlet, doublet, triplet, and higher states via $\text{Tr}\left[\hat{P}_S\hat{\rho}\hat{P}_S\right]$ for $S = 0,\ \frac{1}{2},\ 1$ etc. Note, however, that this approach is incompatible

with the CPT extension. The CPT extension involves solving an auxiliary AIM Hamiltonian that shares the same impurity Green's function as a larger AIM Hamiltonian, and consequently any quantities derived directly from the Green's function will be unaffected. However, there is no such guarantee for the reduced density matrix, because the hybridisation function of this auxiliary system does not necessarily match that of the physical system. To overcome this, the CPT extension was at first applied in order to obtain an approximate solution, but then removed for the final DMFT step. Typically this final step required the addition of an extra bath site so that the AIM acquired sufficient flexibility to fit the impurity hybridisation function to the local hybridisation function without the assistance of the CPT extension.

The decomposition of the reduced density matrix into spin sectors is displayed in Figure 4.7c. It reveals a large quintet state contribution in the limit of dissociation, but also that, regardless of Fe-C distance, many different spin sectors are important. This would be missed if we only examined S_{eff}. Evidently, a multitude of states play an important role throughout CO-unbinding, and therefore the success of DFT + U and HF in predicting the quintet ground state must be for the wrong reasons, as neither go beyond the single-determinantal picture. (Note that HF is known to overly favour high-spin states.[346])

It should be noted that the precise details of Figure 4.7 are somewhat sensitive to various simulation parameters (most notably the definition of the Hubbard projectors), but qualitatively the results are expected to hold generally.

4.3.3 Photodissociation

The photodissociation mechanism of carboxymyoglobin is already relatively well understood. Irradiation at 570 nm (2.18 eV) causes the excitation of electrons in the porphyrin ring into low lying singlet states with π/π^* character (the so-called Q band).[347] The carbon monoxide ligand then dissociates within 50 fs, as the system adiabatically crosses to a repulsive anti-back-bonding orbital.[348,349] There is a small (but not insignificant) predicted energy barrier of 0.08 eV between these two states, as calculated by B3LYP and TDDFT.[339] The porphyrin then undergoes the "intersystem crossing", a complicated, multi-step process which ultimately takes the dissociated system to its high-spin ground state.

To a large extent, semi-local DFT captures this process. The energies of the lowest unoccupied KS molecular orbitals as predicted via DFT are shown in Figure 4.8. The Q band is present, and the pathway from the Q band to the anti-back-bonding orbital is clearly visible via their crossing at approximately 2.3 Å (the same distance we observe the low-to-high spin crossover in Figure 4.6), with an energy barrier of approximately 0.13 eV. Compared to the TDDFT/B3LYP results of Refs. 348 and 349, PBE calculations place this crossover at a much longer distance (approximately 2.3 Å compared to 2.0 Å), and predict that the energy of the anti-back-bonding orbital drops much more steeply.[†]

To compare the results of DMFT to these KS eigenenergies, the analogous quantity we must extract is the density of states (DOS). The DOS is given by the trace of the many-body density

[†]Head-Gordon and co-workers noted that the very gentle decrease in the energy of the anti-back-bonding orbital as predicted by their TDDFT/B3LYP calculations is at odds with the ~ 50 fs timescale of photodissociation.[349]

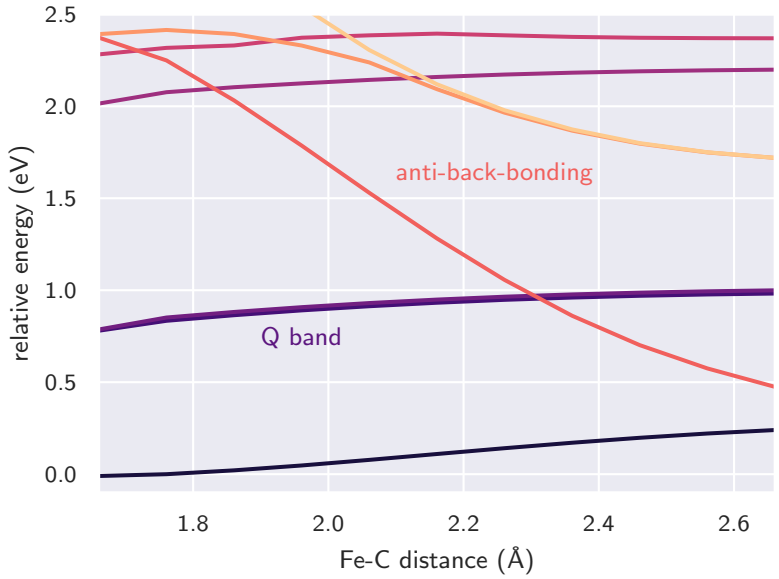


Figure 4.8: Energies of the KS molecular orbitals, measured relative to the highest occupied orbital of the tightly-CO-bound structure.

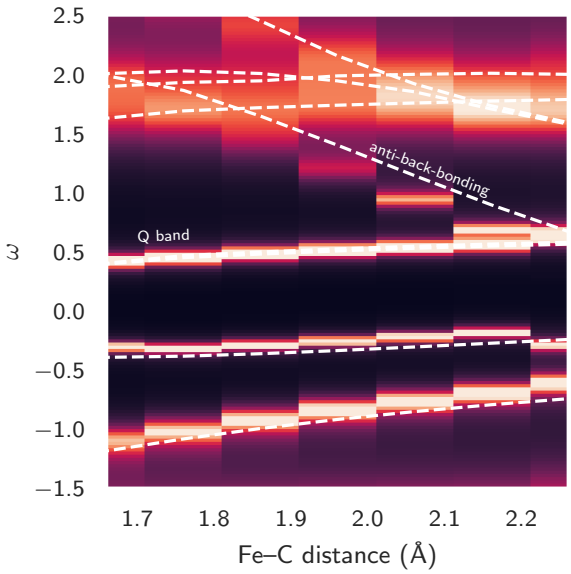


Figure 4.9: The DMFT DOS of FePImCO during dissociation, compared to the KS eigenenergies (white dashed lines), as given by self-consistent DMFT calculations. The DOS and eigenenergies have been aligned to match the Q band, because, being a porphyrin-ring state, it should not be significantly shifted by DMFT.

matrix

$$\rho(\omega) = \sum_{\alpha,\beta} \rho^{\alpha\beta}(\omega) S_{\beta\alpha}. \tag{4.20}$$

The DMFT DOS is compared to the KS eigenenergies in Figure 4.9. Qualitatively, they yield very similar results, although DMFT does provide further details such as the finite lifetime of excitations.

To reveal the contribution of individual atoms (or groups of atoms) towards the DMFT DOS,

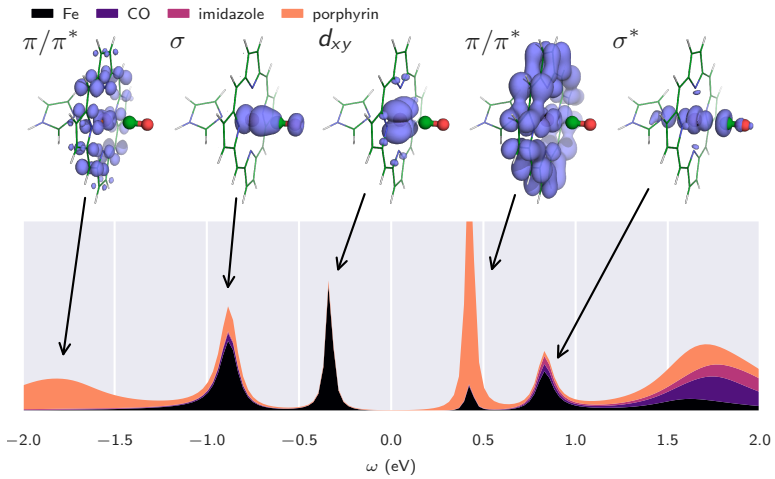


Figure 4.10: Self-consistent DMFT density of state for carboxy-haem with a Fe-C distance of 2.06 Å. The DOS is further decomposed into contributions from the iron atom, CO molecule, imidazole ligand and porphyrin ligand. Above, isosurfaces of $\rho(\mathbf{r}, \omega_{\text{peak}})$ have been plotted for each peak.

it can be decomposed into local densities of state (LDOSs)

$$\rho_I(\omega) = \sum_{\alpha \in I} \sum_{\beta} \rho^{\alpha\beta}(\omega) S_{\beta\alpha}, \tag{4.21}$$

where I denotes a subset of NGWFs typically belonging to atoms that are a particular element or part of a spatially distinct subsystem (*e.g.* all the NGWFs belonging to atoms in the porphyrin ring). One such LDOS is plotted in Figure 4.10, along with isosurfaces of the spectral density at energies corresponding to the various peaks in the DOS. The Q-band π/π^* orbitals and the Fe-CO back- and anti-back-bonding orbitals are all clearly identifiable.

Another important quantity that can be extracted from DMFT calculations is the optical spectrum. The theoretical optical absorption spectrum can be obtained within the linear-response regime (that is, Kubo formalism) as

$$\sigma_{ij}(\omega) = \frac{2\pi}{\Omega} \int d\omega' \frac{f(\omega' - \omega) - f(\omega')}{\omega} \left(\rho^{\alpha\beta}(\omega' - \omega) \mathbf{v}^i_{\beta\gamma} \rho^{\gamma\delta}(\omega') \mathbf{v}^j_{\delta\alpha} \right) \tag{4.22}$$

where Ω the simulation cell volume, $f(\omega)$ is the Fermi-Dirac distribution, ρ is the basis-resolved spectral density, the i and j indices correspond to Cartesian directions, the velocity operator $\mathbf{v}$ is

$$\mathbf{v}^j_{\alpha\beta} = -i\langle\alpha|\nabla_j|\beta\rangle + i\langle\alpha|\left[\hat{V}_{nl}, \mathbf{r}\right]|\beta\rangle \tag{4.23}$$

which includes the effect of non-local pseudopotentials V_{nl} on the velocity operator matrix elements, and adopts the no-vertex-corrections approximation.[350] Optical spectra for haem are typically carried out in liquid or gas phases, and so are described by the isotropic part of the optical conductivity tensor

$$\sigma(\omega) = \frac{1}{3} \sum_i \sigma_{ii}(\omega). \tag{4.24}$$

The optical absorption spectra for carboxy-haem complexes as given by self-consistent DMFT

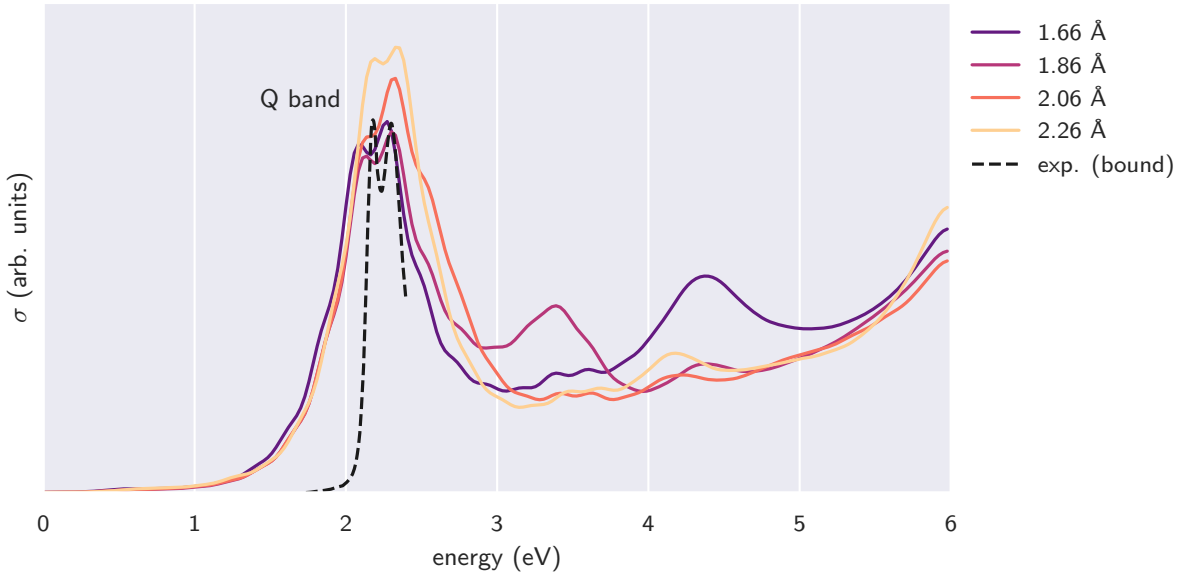


Figure 4.11: Optical spectra of FePImCO calculated using self-consistent DMFT, going from ligated (dark) up to the point of dissociation (light). Also pictured are the Q-band peaks from experimental spectra of carboxymyoglobin.[351]

are plotted in Figure 4.11. These spectra are dominated by a feature at around 2 eV associated with π-π^* transitions on the porphyrin ring — that is, the Q band. The double-peak structure of the Q band is successfully reproduced.† Secondary peaks appear above 3 eV corresponding to direct photoexcitation of the anti-back-bonding orbital.

4.4 Conclusions

This chapter has introduced the DMFT method and described how to interface it with DFT. More specifically, I have detailed its implementation with linear-scaling DFT in the form of ONETEP+TOSCAM. Extensive work was performed to overhaul this code, in order to make it compatible with the active version of ONETEP and prepare it for distribution. Crucially, for the purposes of simulating metalloproteins, this DFT + DMFT implementation does not compromise our ability to model thousands of atoms at the DFT level.

Calculations on the photodissociation of carboxymyoglobin showcased the kind of results one can extract from such a DFT + DMFT calculation on a metalloprotein. The calculations do not present any previously unknown physics – but this was not their purpose. Nevertheless, there is scope here to resolve some unanswered questions surrounding the photodissociation process. In particular, the remarkably fast rate of photodissociation (~ 50 fs) is at odds with the gentle slope of the potential energy surface (discussed above) and the predicted barrier on the order of 0.1 eV (compared to the 0.028 eV zero-point energy of the Fe-C stretching mode).[339] Further study could investigate this apparent contradiction.

†Ref. 79 found that $J > 0$ is necessary to obtain this double-peak feature.

Chapter 5

Haemocyanin

N.B. This chapter is the result of work done jointly by myself and M. A. Al-Badri (of King's College, London). A detailed breakdown of who contributed to the calculations presented in each section is contained in the Preface.

5.1 Introduction

This chapter will apply the methods of the previous chapters to the dicopper core of oxygenated haemocyanin (oxyHc). In addition to the inherent scientific value of studying this system's electronic structure, for the purposes of this thesis oxyHc provides a logical next stepping stone towards simulating the OEC: compared to the haem system of the previous chapter, we now have two correlated sites, but these correlated sites are comparatively simple (with only one hole on each copper).

5.1.1 The structure and function of haemocyanin

Copper-based metalloproteins play a major role in biology as electron or dioxygen (O_2) transporters. Haemocyanin is one of three oxygen transporting proteins found in nature, alongside the iron-based haemrythrin and haemoglobin, and is common to a number of invertebrates, such as molluscs and arthropods. Deoxygenated haemocyanin (deoxyHc) employs two half-spin copper (I) cations, each coordinated with the imidazole rings of three histidine residues, to reversibly bind O_2 as shown in Figure 5.1.

Some type 3 copper-based systems[†] also possess catalytic properties. Haemocyanin can decompose hydrogen peroxide into water and oxygen[353] and synthetic analogues have been shown to reversibly cleave the dioxygen bond[354] — a mechanism that enables tyrosinase and catechol oxidases to oxidise phenols.[355] There is significant interest in the biomimetic application of naturally occurring metal complexes for use in metallodrug design, with Cu(II) complexes recently employed in cancer therapeutics as artificial DNA metallonucleases[356] and tyrosinase mimics.[357] An accurate understanding of the electronic structure (spin and charge) of the Cu_2O_2 core is essential to clarifying the operation of dioxygen transport and would advance the design of synthetic catalysts that employ dioxygen as a terminal oxidant.

[†] *i.e.* systems with a pair of copper atoms, each coordinated with three histidine residues.

5.1.2 Computational challenges

The binding of O_2 to deoxyHc remains a challenging problem, being a spin-forbidden transition. Molecular O_2 is in a spin triplet configuration, and the Cu ions in deoxyHc are known to be in the Cu(I) d^{10} singlet configuration. The combination of triplet O_2 and singlet deoxyHc, to produce the Cu_2O_2 antiferromagnetic singlet in oxyHc, is believed to occur via a simultaneous charge transfer of one electron from each Cu(I) ion to O_2, forming a hybrid Cu(II)-peroxy-Cu(II) configuration. A superexchange pathway is hypothesised to form across the two Cu atoms, stabilising the singlet.[358] This mechanism is supported by SQUID measurements that report a large superexchange coupling between the two Cu centres,[359] and a diamagnetic ground state.[360]

Despite intensive study, theoretical analysis has so far proved to be challenging for many electronic structure methods including *ab initio* quantum chemistry, DFT, and QM/MM methods. In particular, DFT and hybrid-DFT do not predict the correct singlet ground state due to the fact that its multi-reference nature is not accessible in DFT-based approaches.[361–363] (Experiments have alluded to the necessity of characterising the oxyHc ground state as a mixed valence state.[364]) To overcome this limitation, a spin-projection method (also called spin-mixing) is often applied, whereby the different spin-polarised ground states are calculated individually,[346] and the entangled singlet is reconstructed by linear combination of the respective Slater determinants (essentially a combination of the spin-broken symmetry state in the up-down, up-up, down-down configurations to extract an effective singlet state).[365] Although this construction yields insights into the energetics, it does not allow the study of excitations, preventing comparison with experimental data such as the optical absorption.[362] Furthermore, the spin-contamination present in spin-polarised hybrid DFT remains an issue,[358,363,366–368] and typically the broken symmetry state wrongly becomes asymmetric in the Cu_2O_2 core.[362]

Multi-reference wavefunction methods have been extensively applied to the oxyHc core,[367–374] but these approaches are not feasible for systems containing more than several dozen electrons. Flock and Pierloot argue that the inclusion of the imidazole ligands results in steric effects that

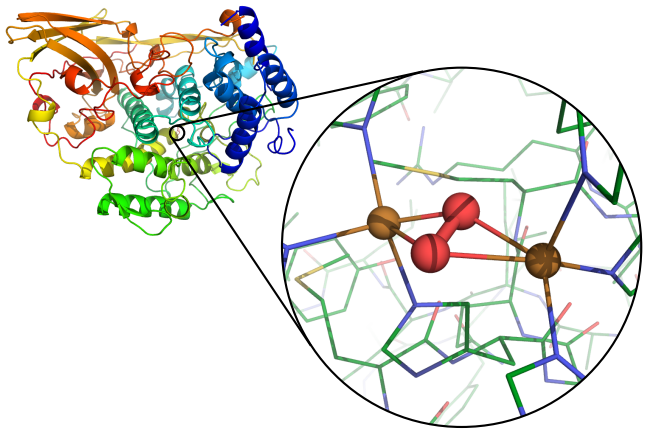


Figure 5.1: The full haemocyanin protein, with an inset showing the binding site for O_2. This protein structure is taken from *Limulus polyphemus* (horseshoe crab), and corresponds to PDB record 1OXY.[352]

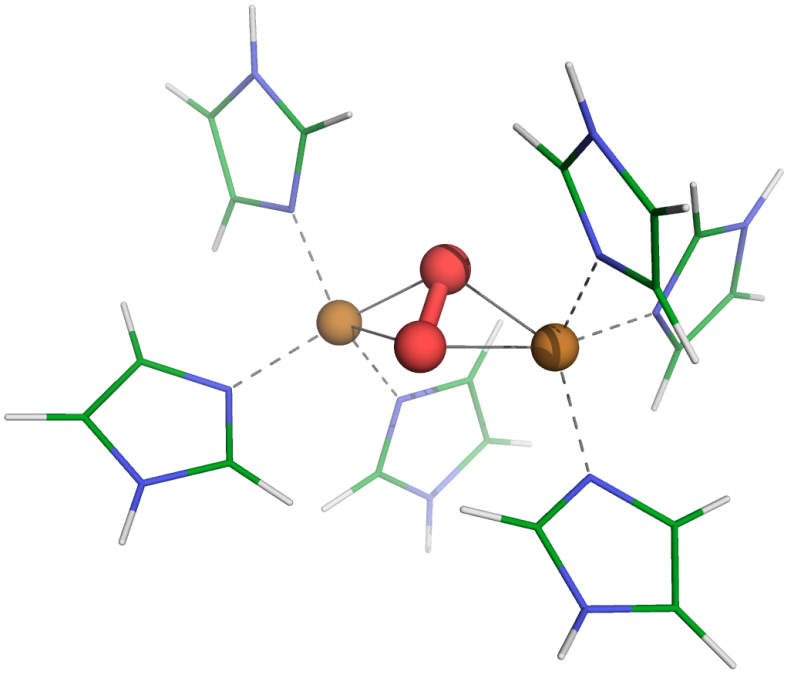


Figure 5.2: The oxyHc model simulated in this chapter, showing the Cu_2O_2 correlated subsystem, which is treated using DMFT, and the surrounding imidazole rings representing the protein environment. (The red, orange, green, blue, and white atoms are oxygen, copper, carbon, nitrogen, and hydrogen.)

are critical for a realistic description of oxygen containing dicopper systems.[370] However, most multi-reference wavefunction studies of this core consider a simplified model with ammonia ligands (including CC,[367,368] CASPT2,[370] MRCI,[371] RASPT2,[372] and DMRG-CASPT2[373]), while others (such as DMRG[374] and DMRG-CT[375]) are limited further to the experimentally inaccessible bare $Cu_2O_2{}^{2+}$ core alone. Furthermore, this system has large active-space requirements given that it likely suffers from triplet instability,[363] and if the number of allowed excitations is too limited, size-extensivity errors arise.[372] Some of these methods also lack dynamical correlation contributions,[376–378] and others strongly over-correct correlation effects.[371]

DMFT accounts for these limitations by treating the many-body effects and the superexchange of the dicopper bridge explicitly (unlike DFT), while limiting this treatment to the correlated subspace of the copper $3d$ electrons, thereby side-stepping the prohibitive scaling of quantum chemistry methods. This chapter will present the first DFT + DMFT simulations of the oxyHc functional complex.[6] These calculations will be performed on a 58-atom model of the core and the coordinating histidine ligands (modelled by imidazole rings), as shown in Figure 5.2. While 58 atoms is within the reach of some less accurate quantum chemistry methods, the overhead of extending DFT + DMFT to include much more of the protein environment using the ONETEP+TOSCAM framework would come at an insignificant computational cost.

As this problem involves direct exchange across two correlated atoms, we used the non-local DMFT implementation (cluster DMFT) needed to capture the superexchange mechanism between the Cu_2 d-orbitals and intermediate p-orbitals, as single-site DMFT can only treat the multiplet structure of each Cu atom separately.

Figure 5.3: The convergence of the system-to-AIM mapping for haemocyanin, as quantified by the distance d (see Equation 4.9), as a function of the total number of sites (impurity and bath).

The Hubbard U correction is crucial for describing many-body effects at play in the oxyHc core. Several competing effects stem from the local Hubbard U physics: charge localisation, exchange of electrons, charge-transfer excitations, and stabilisation of magnetic multiplets. Although typical values for U can be obtained by linear response or cRPA,[282] I will first consider a range of values for U. By artificially manipulating the magnitude of the local many-body effects, we can investigate their influence on the electronic spectral weight and magnetic properties. At the end of the chapter I will return to linear response.

5.2 Methods

The geometry of the 58 atom system was obtained from Ref. 363, which had been optimised using the B3LYP hybrid functional. The resulting structure closely matches the experimentally observed structure.[81,363]

The initial calculations were performed with ONETEP,[22] using an energy cut-off of 897 eV and the PBE exchange-correlation functional.[136] Nine NGWFs were associated with each copper atoms, four with each carbon, nitrogen, and oxygen, and one with each hydrogen. Spin symmetry was imposed. NGWFs were truncated using 7 Å cut-off radii. Open boundary conditions were achieved via a padded cell and a Coulomb cut-off.[238] The Hubbard projectors were constructed from the Kohn-Sham solutions to an isolated copper pseudopotential.[235] The pseudopotentials were generated OPIUM.[125] These pseudopotentials partially account for scalar relativistic effects. (Studies have demonstrated that relativistic effects can play a role in the electronic structure of the Cu_2O_2 core,[379] but relativistic effects cannot be explicitly included in current state-of-the-art DMFT calculations.)

The DFT + DMFT calculations were then performed with TOSCAM, mapping the oxyHc model to an AIM Hamiltonian, and solving it with an extended Lanczos solver[380] to obtain the

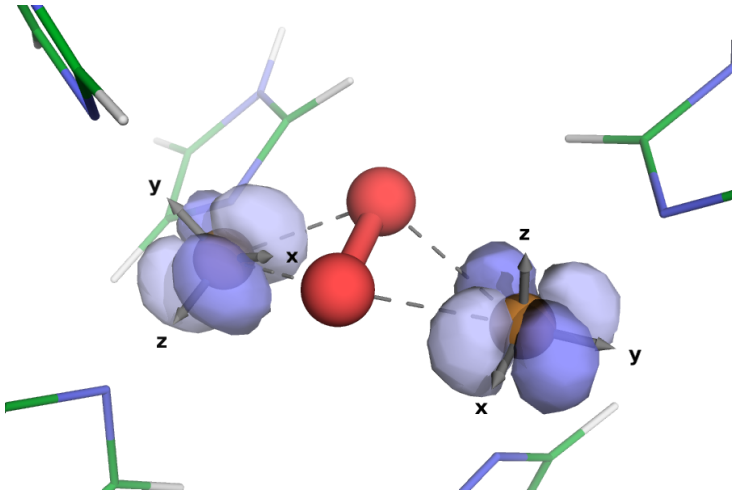


Figure 5.4: The local axes for the Cu $3d$ correlated subspaces, and the two half-filled NBOs for comparison. (Cu_A is on the left; Cu_B is on the right.)

DMFT self energy. The convergence of the mapping is shown in Figure 5.3. If we increase the number sites of our AIM, the AIM Hamiltonian has more parameters, and we stand a better chance of fitting the physical hybridisation function because the AIM hybridisation function is more flexible. As a rule of thumb $d < 10^{-7}$ is generally adequate; if d is much smaller than this it tends to indicate overfitting. Subsequent calculations presented in this chapter used eight sites.

The DMFT calculations were carried out at room temperature ($T = 293\,\mathrm{K}$). The Hubbard U was varied over the range $0-10\,\mathrm{eV}$, with the Hund's coupling fixed at $J = 0.8\,\mathrm{eV}$. We performed self-consistency over the chemical potential but not the Hamiltonian (refer back to Subsection 4.2.3) to minimise computational cost.

5.2.1 Local axes

To identify the best spatial representation of the local $3d$-subspaces in the AIM, we first identified the orthogonal transformation which reduces the off-diagonal elements of the local Green's function for each copper atom. We implemented a minimisation procedure which finds the closest corresponding real space $SO(3)$ rotation of the local Cartesian axis corresponding to the $O(5)$ orthonormal transformation in d-space. The resulting axes for the two Cu subspaces are shown in Figure 5.4. As shown in Table 5.1 these axes localise the holes on single d orbitals; d_{xz} for Cu_A (on the left of the figure) and d_{xy} for Cu_B (on the right). Increasing U decreased the occupancy of these two orbitals.

The two natural bonding orbitals (NBOs) identified as being half-filled are plotted for comparison (see Section 5.4). The NBO method and analysis will be properly explained later in Section 5.4, but for the purposes of this section, these NBOs are orbitals that have been identified as being half-filled via an analysis that is agnostic to the projection procedure used in the cluster DMFT calculations. It is therefore very reassuring that these orbitals align with the axes.

Table 5.1: DMFT $3d$ orbital occupations of Cu in our model of ligated haemocyanin for different Hubbard U values. The $U = 4\,\text{eV}$ values come from a single-site DMFT calculation, where all five orbitals on each Cu atom were explicitly included (but inter-copper correlation had to be neglected due to the exponentially-scaling computational cost of solving the corresponding AIM). The bold font indicates the identified holes. From this calculation the ideal rotation was determined, and the results for other values of U come from cluster-DMFT calculations using this rotation. Note that the orbital labels correspond to the local axes to each Cu atom (as shown in Figure 5.4).

U (eV)	atom	d_{xy}	d_{yz}	$d_{3z^2-r^2}$	d_{xz}	$d_{x^2-y^2}$
0	Cu_A				1.72	
	Cu_B	1.68				
4	Cu_A	2.00	2.00	1.95	**1.50**	1.93
	Cu_B	**1.40**	2.00	1.98	1.99	2.00
8	Cu_A				1.25	
	Cu_B	1.21				
10	Cu_A				1.18	
	Cu_B	1.14				

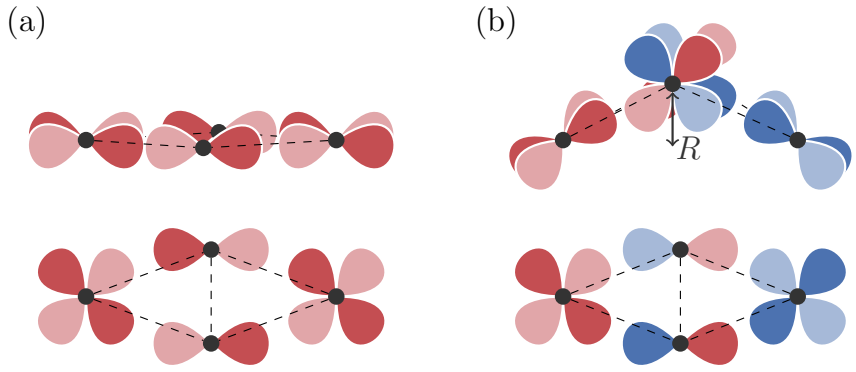


Figure 5.5: The superexchange model of Solomon and co-workers, depicting the Cu_2O_2 core viewed from side-on (top) and above (bottom). (a) In the planar configuration, single ligand orbitals bridge the two copper sites, and superexchange is possible. (b) In a bent configuration, the copper d orbitals overlap with different π^* orbitals. As these two sets of orbitals are orthogonal, hopping between the blue and the red subspaces is not possible and the superexchange mechanism breaks down.

5.3 The ground electronic state of the Cu_2O_2 core

5.3.1 Formation of the singlet

In vivo, the Cu_2O_2 core exists in a low-spin (singlet) state. (This can be identified experimentally via EPR.[381]) In the model of Solomon and co-workers, this low-spin state is stabilised by superexchange via the O_2 ligand orbitals, which relies on the Cu_2O_2 core being planar (Figure 5.5a).[362] As the peroxide molecule unbinds, the core butterflies (*i.e.* the dioxygen moves up out of the plane, leaving the core in a bent configuration). Here, each Cu overlaps with a different oxygen π^* orbital on the peroxide (Figure 5.5b). This removes the superexchange, and the triplet state becomes most favourable. If we measure planarity by $R = |\frac{1}{2}(\mathbf{r}_{\text{CuA}} + \mathbf{r}_{\text{CuB}}) - \frac{1}{2}(\mathbf{r}_{\text{O1}} + \mathbf{r}_{\text{O2}})|$ — that is, the distance between the mean position of the two copper atoms and the mean position of the two oxygen atoms, B3LYP calculations predict that the singlet-to-triplet transition occurs at $R = 0.6\,\text{Å}$.[362]

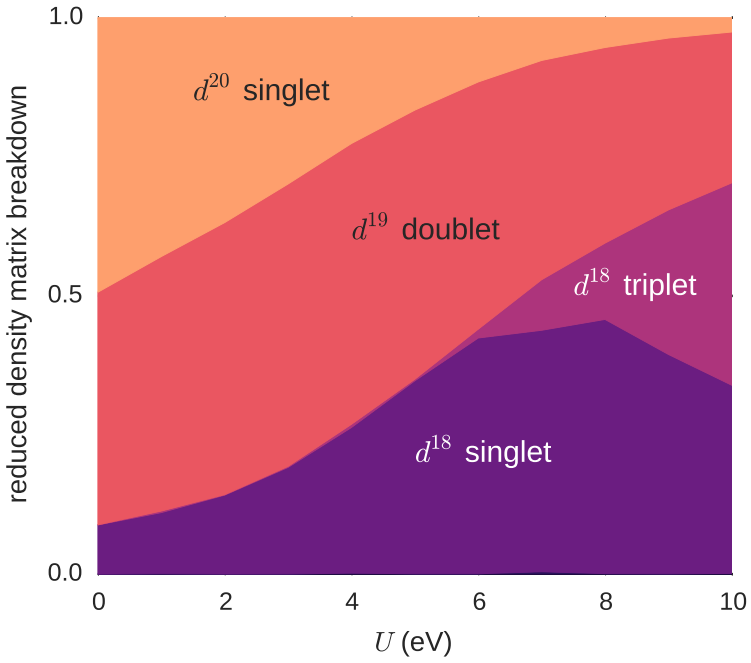


Figure 5.6: Decomposition of the reduced density matrix of the Cu_2 dimer in the different quantum sectors. Note that the d occupation is the sum of both Cu sites (for example, d^{20} means both Cu atoms are in the d^{10} configuration).

However, x-ray structures of the Cu_2O_2 core reveal that the bound singlet state is not planar. In oxyHc $R = 0.47$ Å, and in oxyTy $R = 0.63$ Å— beyond the predicted singlet-to-triplet transition.[81,382] QM/MM studies of the entire oxyHc protein (from which our model complex derives) obtain $R = 0.54$ to 0.71 Å; evidently, the protein scaffolding around the binding site prevents the core from ever reaching the planar structure observed in model complexes.[363]

With this in mind, I examined the reduced finite temperature density matrix of our butterflied model ($R = 0.68$ Å), obtained by tracing out the bath states from the density matrix (a procedure explained more fully in Subsection 4.3.2). The lowest-energy eigenstates of the reduced density matrix provide a detailed picture of the effective electronic structure of the $3d$ subspace of the Cu atoms. Note that in this approach, the ground-state wavefunction is not a pure state with a single allowed value for the spin states (singlet, doublet, triplet, *etc.*). Furthermore, compared to the haem system in Subsection 4.3.2, the quantum state we are probing exists purely within the Cu_2 joint $3d$ subspace, so we will not encounter the issues we had for the FePImCO singlet. The distribution of these states is displayed in Figure 5.6. In the weakly correlated regime ($U < 2$ eV), we find a large contribution from the d^{20} and d^{19} configurations, indicating that the average charge transfer from the Cu to O_2 involves less than one electron per Cu, thus preventing the formation of a singlet (as the Cu $3d$ orbitals are nearly full). As U increases, the total electronic occupation of the Cu dimer decreases (Figure 5.6 and Table 5.1). In the range $U = 6 - 8$ eV, the d^{18} singlet component is maximised, beyond which d^{18} triplet excitations begin to contribute. Since $U \approx 8$ eV in the case of both molecules[383] and solids,[160] it appears that in nature many-body quantum effects stabilise the low-spin singlet in spite of the butterflied structure of the Cu_2O_2 core.

5.3.2 Details of diamagnetism

The existence of this singlet is corroborated by the observed local magnetic moment and spin correlation (Figure 5.7). The spin correlation K reaches half the saturation value for $U = 6-8\,\text{eV}$. Note that the saturation value would only be obtained for a diatomic system in vacuum, which is not hybridised to the rest of the molecule. As the local Cu $3d$ orbital charge and spin are not true quantum numbers in the molecule due to hybridisation, quantum fluctuations reduce the amplitude of the spin correlation to half the full value.

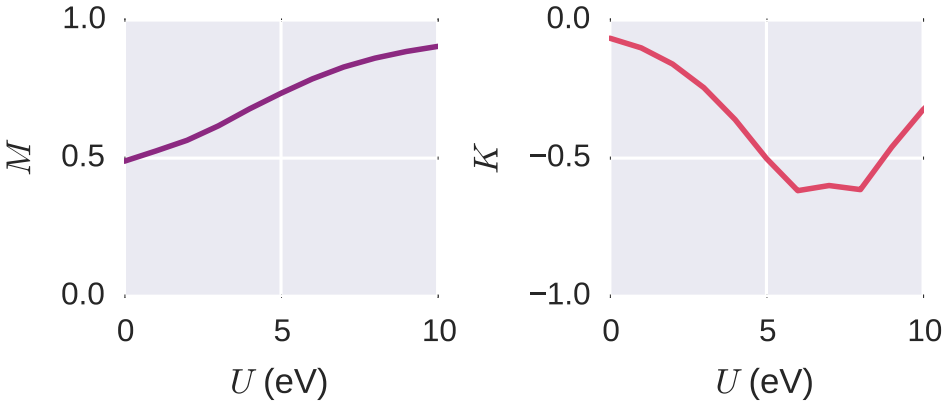


Figure 5.7: The effective magnetic moment $M = \sqrt{\langle \mathbf{S}_1^2 \rangle / 3}$ (normalised by saturation value) and the spin correlation $K = 2\langle \mathbf{S}_1 \cdot \mathbf{S}_2 \rangle$ for varying values of the Hubbard U. For a pure two orbital singlet, $K = -1.5$. In our calculations, as the rest of the molecule hybridises with the Cu orbitals, the spin correlation is renormalised to half its saturation value for $U = 6 - 8\,\text{eV}$.

5.3.3 von Neumann entropy

The importance of multi-determinantal physics can be quantified by the von Neumann entropy. The von Neumann entropy, obtained in the dicopper $3d$ subspace, is given by $\Lambda = \text{Tr}\,[\hat{\rho}_d \log \hat{\rho}_d]$, where $\hat{\rho}_d$ is the dicopper reduced density matrix, traced over the states of the AIM bath environment.

The von Neumann entropy is plotted in Figure 5.8. Interestingly, it grows as U increases, pointing to the importance of many low-lying quantum states. I note the presence of two plateaus, for $U = 4-6\,\text{eV}$ and $U = 7-8\,\text{eV}$, that coincide with the formation of the singlet and triplet configurations in the histogram in Figure 5.6.

5.3.4 The superexchange mechanism

Having identified this singlet in the Cu_2O_2 core at $U \approx 8\,\text{eV}$, let us establish how it forms. Direct hopping between localised Cu d-orbitals is very unlikely due to the large distance by which they are separated, and therefore hopping must proceed via an intermediate oxygen p-orbital (*i.e.* superexchange).[384]

The superexchange process can be visualised in the canonical hydrogen atom dimer system, in which we have a a pair of electrons (one spin-up, one -down) that form a singlet state. In this

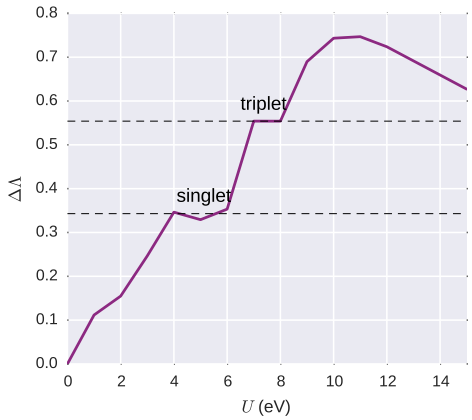


Figure 5.8: The von Neumann entropy Λ of the reduced density matrix and its dependence on the on-site interaction U.

picture, two different limits are possible: (a) at short H-H distances, the H atoms form a bond; the up- and down-electron form a delocalised bound singlet (BS) centred on the bond, with a high degree of double occupancy; (b) in the dissociated case, known as the Heitler-London (HL) limit, the H atoms are far apart and the singlet is a true quantum entangled state of the singly occupied H orbitals. The HL case typically appears in cases of dissociation, with the charge localised around the H atoms. However, it may also occur in systems where the local Hubbard Coulomb repulsion U acts as a Coulomb blockade: many-body effects prevent long-lived charge transfer excitations, and the Coulomb repulsion energy is reduced at the expense of the kinetic energy. A signature of the blockade is typically a large increase in the self energy at the Fermi level, indicating charge localisation and incoherent scattering associated with a short lifetime of charge excitations.

To investigate the nature of the singlet (BS or HL), Figure 5.9a shows the computed self energy of the Cu $3d$ subspace, for various values of U. There is a qualitative difference between $U = 6\,\text{eV}$ and $U = 8\,\text{eV}$: at $U = 8\,\text{eV}$ the self energy develops a pole at $\omega = 0\,\text{eV}$ (Figure 5.9b). The formation of the pole is associated with the regime where excitations are incoherent, which prevents long-lived charge transfer excitations from the Cu $3d$ orbitals to O_2. Here, many-body effects act as a Coulomb blockade and the charge is in turn localised, with weak direct coupling. For $U \approx 6\,\text{eV}$, this pole is absent and therefore the singlet is in the BS limit, where charge excitations allow a direct electron transfer across the oxo-bridge. Note that the observation of the BS-HL crossover is not apparent in averaged quantities, such as in the double occupancies (Figure 5.9c), which evolve smoothly with the Coulomb repulsion.

Turning back to the physical system, superexchange relies on a single ligand orbital bridging the two copper sites. (Superexchange pathways via two p orbitals are possible, but they give rise to ferromagnetic coupling that is significantly weaker than antiferromagnetic coupling from single-ligand orbital pathways.[385]) Examination of the molecular orbitals near the Fermi level (Figure 5.10) reveals that for the HOMO of the bent Cu_2O_2 structure, electronic density of the oxygen ligand is directed into the copper plane, thus providing a pathway for the antiferromagnetic superexchange that we observe. To properly establish this, one could examine the precise

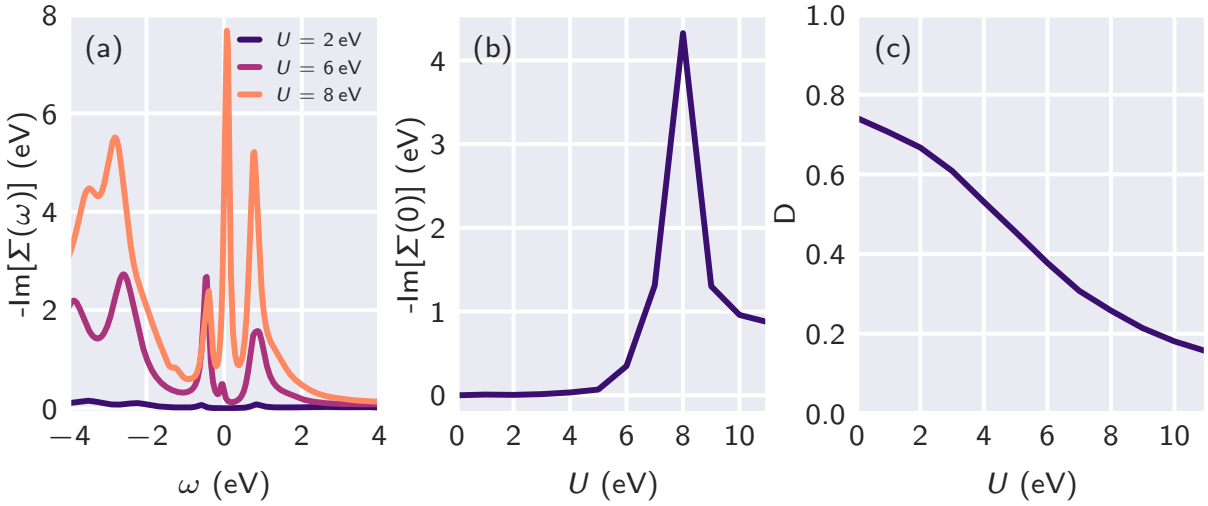


Figure 5.9: (a) The imaginary part of the dynamical mean field local self energy of the Cu-3d empty orbital for Hubbard $U = 2\,\text{eV}$, $6\,\text{eV}$, and $8\,\text{eV}$. At $U = 8\,\text{eV}$, we obtain incoherent excitations at $\omega = 0\,\text{eV}$. (b) The self energy at $\omega = 0$ and (c) the double occupancy D as a function of U. Note that although the double occupancy evolves smoothly with the Coulomb interaction U, $\Sigma(\omega = 0)$ shows a sharp increase near $U = 8$, associated with the stabilisation of a localised singlet. (The lifetimes of quasiparticles are given by $(\text{Re}[\Sigma(\omega)] - 1)/\text{Im}[\Sigma(\omega)]$, so if $-\text{Im}[\Sigma(0)]$ is very large then lifetimes of low-frequency excitations/holes in the joint Cu_2 3d subspace are very small: in other words, charge transfer excitations will have short lifetimes, stabilising the singlet.)

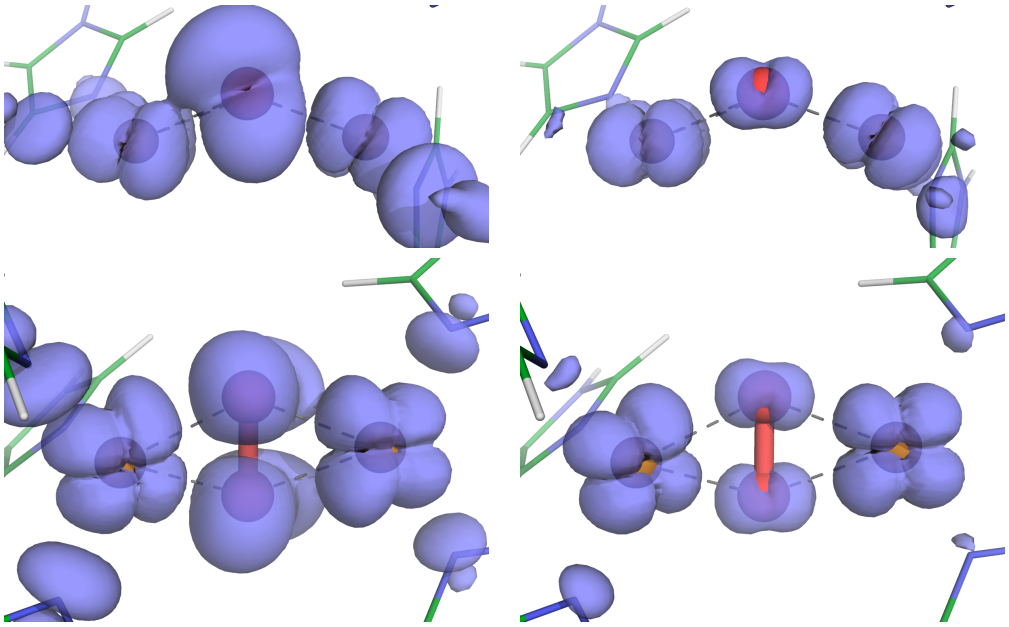


Figure 5.10: Isosurfaces for the HOMO (left) and LUMO (right) densities for $U = 8\,\text{eV}$, as viewed from above (top) and face-on (bottom). Note that because these are extracted from the Green's function via the spectral density, the phase of the orbitals is inaccessible.

parametrisation of AIMs for this system (including the oxygen molecular orbitals explicitly as impurity orbitals), but that is beyond the scope of this work.

Interestingly, we note that these molecular orbitals differ in their energetic ordering compared to those from DFT studies of planar model complexes.[360] In particular, the HOMO involves hybridisation with oxygen π^* rather than σ^* orbitals (with the σ^* oxygen orbital featuring approximately 3 eV above the Fermi level). This reordering will have substantial ramifications for the potential catalytic pathways, especially considering the importance of the σ^* orbital to O_2 bond-breaking.

5.4 Natural bond orbital analysis

This molecular orbital picture is confirmed by NBO analysis, which I performed on the DFT and DMFT densities in order to understand the nature of the bonding in the Cu_2O_2 complex.[386–388] NBO analysis involves a series of diagonalisation and occupancy-weighted orthogonalisation procedures on the single-particle density matrix, transforming it into a set of atom-centred orthogonal natural atomic orbitals (NAOs), then natural hybrid orbitals, and finally the natural bond orbitals $\{|\sigma_i\rangle\}$, which are either one- or two-atom centred. By construction, this procedure decomposes the electronic density into terms resembling Lewis-type chemistry (with bonding and lone pairs of electrons). The NBOs generated from DFT + DMFT densities largely retain the familiar profile of DFT-based NBOs, but their occupancies may be expected to deviate further from integer values due to quantum-mechanical and finite-temperature multi-reference effects captured within DFT + DMFT.

Natural bonding orbital analysis was performed using the *NBO 5* programme.[387] Performing this transformation starting from ONETEP's basis of NGWFs is non-trivial, and is described in Ref. 388.

For haemocyanin, this analysis reveals a hole in one $3d$ orbital for each Cu atom (with $3d$ occupancies of 9.11 and 9.07 for $U = 8$ eV), confirming the expected Cu(II) oxidation state $3d^9 4s^0$ (Table 5.1 and Figure 5.11a) and the expected orientation of the local axes that give rise to this hole (as illustrated in Figure 5.4).

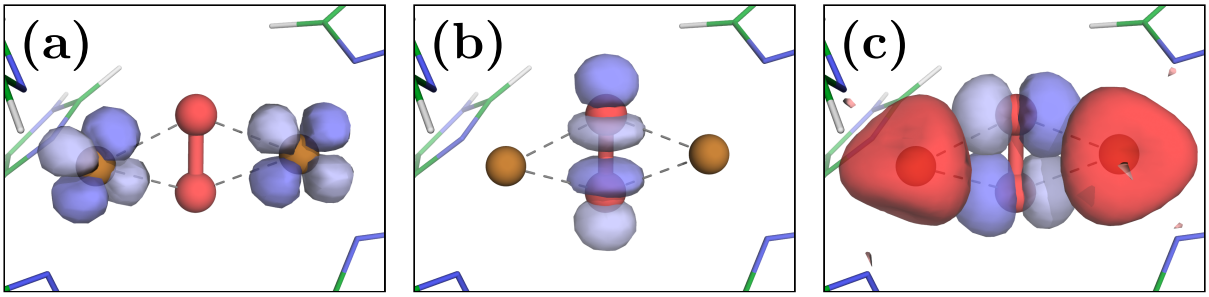


Figure 5.11: Isosurfaces of several natural bonding orbitals for $U = 8$ eV. (a) Two Cu $3d$ orbitals are identified as half-filled by the NBO analysis. (b) The O_2 σ^* anti-bond is empty, and does not hybridise with any Cu orbitals. (c) Instead, O $2p$ (blue) to Cu $4s$ (red) charge transfer is favourable.

A second-order perturbation analysis[†] detects multiple energetically favourable transfers of electronic density from filled to unfilled NBOs, revealing those aspects of the electronic structure that are not well described by Lewis-like chemistry. Early studies of haemocyanin identified back-bonding charge transfer from Cu $3d$ to oxygen σ^* anti-bonding orbitals (Figure 5.11b) as an important factor in explaining the comparatively low $750\,\mathrm{cm}^{-1}$ Raman frequency of the O_2 bond.[389] However, our second-order perturbation analyses find that this back-transfer is not present. For $U = 8\,\mathrm{eV}$ we instead detect favourable charge transfer from O $2p$ orbitals to Cu $4s$ orbitals (Figure 5.11c).

5.5 Optical transitions

As a validation of the DFT + DMFT computational model, and to identify the strength of correlations in oxyHc, we extracted the optical absorption spectrum of ligated haemocyanin (Figure 5.12).

The experimental absorption spectrum of oxyHc[391,392] is dominated by a peak at 4.5 eV corresponding to intense aromatic bands attributed to the wider protein (or, in the case of our smaller cluster model, the imidazole molecules). Several smaller features of the spectrum are associated with the oxygen molecule (making the spectrum qualitatively dependent on the protein's ligation state). A peak at approximately 3.6 eV is attributed to ligand-to-metal charge transfer from the O_2 π anti-bond with lobes oriented towards the copper atoms. This orbital is denoted "π^*_σ"; the π anti-bond with lobes directed perpendicular to the copper atoms is denoted π^*_v and is responsible for a weaker peak at approximately 2.2 eV. An even weaker peak at 1.8 eV supposedly corresponds to metal $d \to d$ transitions.[360,362,390]

The DFT + DMFT spectrum qualitatively reproduces many of these features. The large peak at 4.5 eV peak is present, albeit blue-shifted. As these excitations are unrelated to the Cu_2 correlated subspace, this blue-shift is due to inaccuracy at the DFT level (possibly due to the xc functional, the absence of the wider protein environment, or to the conduction states being insufficiently optimised as discussed in Subsection 2.3.8). This could be corrected by using a more accurate starting point, such as is done in GW + DMFT.[393,394] However, this would violate a requirement of any successful technique for accurately simulating metalloproteins: that it must be able to be scaled up to cluster models containing many hundreds of atoms.

[†]NBO analysis allows electronic delocalisation to be quantified. In the NBO formalism, delocalisation presents itself as deviation from the ideal Lewis description of the system provided by the bonding NBOs. This behaviour is partially captured by the anti-bonding orbitals, but is also considered by searching for energetically favourable delocalisations of electronic density to unfilled orbitals as predicted by second order perturbation theory.

Formally, the Hamiltonian of the system in the NBO representation $\langle\sigma_i|\hat{H}|\sigma_j\rangle$ is decomposed into its diagonal and off-diagonal parts, and then the off-diagonal parts are treated as a perturbation to the diagonal NBO Lewis-like description. The first-order energy correction for this "perturbation" is zero; the second order term is given by

$$\Delta E_i^{(2)} = \sum_{j\neq i} f_i \frac{|\langle\sigma_i|\hat{H}|\sigma_j\rangle|^2}{E_i - E_j} = \sum_{j\neq i} \Delta E^{(2)}_{i\to j} \tag{5.1}$$

and therefore the stabilisation effects of electron transfer from NBO $|\sigma_i\rangle$ to $|\sigma_j\rangle$ can be estimated by $\Delta E^{(2)}_{i\to j}$.

For example, dative bonding appears within this donor-acceptor analysis. This is because dative bonds do not appear amongst the set of bonding NBOs, but are identified by the analysis as lone pairs on the donor atom for which it is highly favourable to donate electrons to unoccupied NBOs on the acceptor.

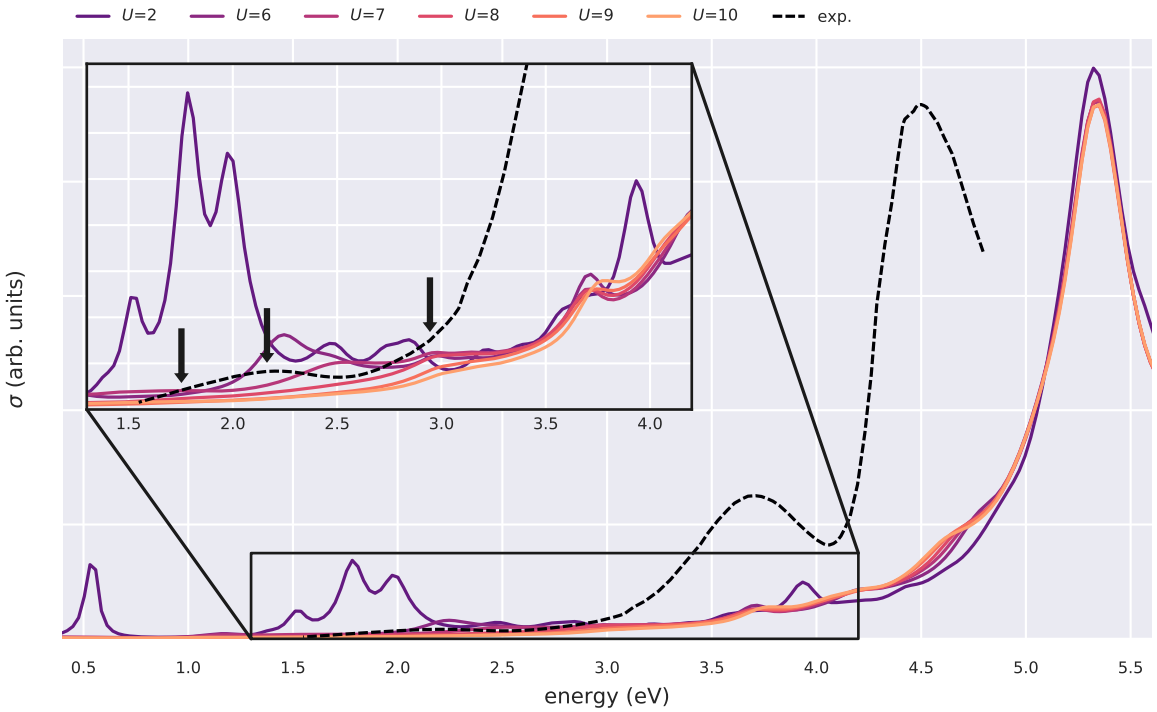


Figure 5.12: Theoretical optical absorption of the Cu_2O_2 core and imidazole rings obtained by DMFT for various values of the Coulomb repulsion U. For comparison, we show the experimental optical absorption[390] in a wide range of wavelengths (infrared to UV). There are several much smaller peaks in the experimental spectra that are not visible at this scale (indicated with arrows).

As for the lower-energy excitations associated with the Cu_2 subspace itself, the most notable feature is that for $U < 6\,\text{eV}$ — that is, prior to the formation of the singlet — there are several erroneous low-energy features which are absent in experiment, although they are also seen in other computational studies.[362] For larger values of U these features are suppressed due to a large increase in incoherent scattering at $\omega = 0\,\text{eV}$ at $U = 8\,\text{eV}$ (Figure 5.9), associated with the localisation of the holes in the Cu $3d$ shell. In comparison, DFT, without extensions, puts a strong emphasis on the near-infrared peak in the optical absorption because the aforementioned scattering processes are absent at this level of theory. This is also evident from the very small HOMO-LUMO gap in the DOS (Figure 5.13) for small values of U.

The second notable difference in the spectra is the unphysical suppression of ligand-to-metal charge transfer (in particular for the peak at 3.6 eV). There are suggestions of some excitations at this particular frequency (see inset of Figure 5.12), but they are weaker than in experiment. This is reminiscent of what we saw in Chapter 3, where corrections to metal centres disrupted their bonding with adjacent oxygen atoms. Further work will investigate if including oxygen-$2p$ orbitals explicitly in the AIM can alleviate this problem.

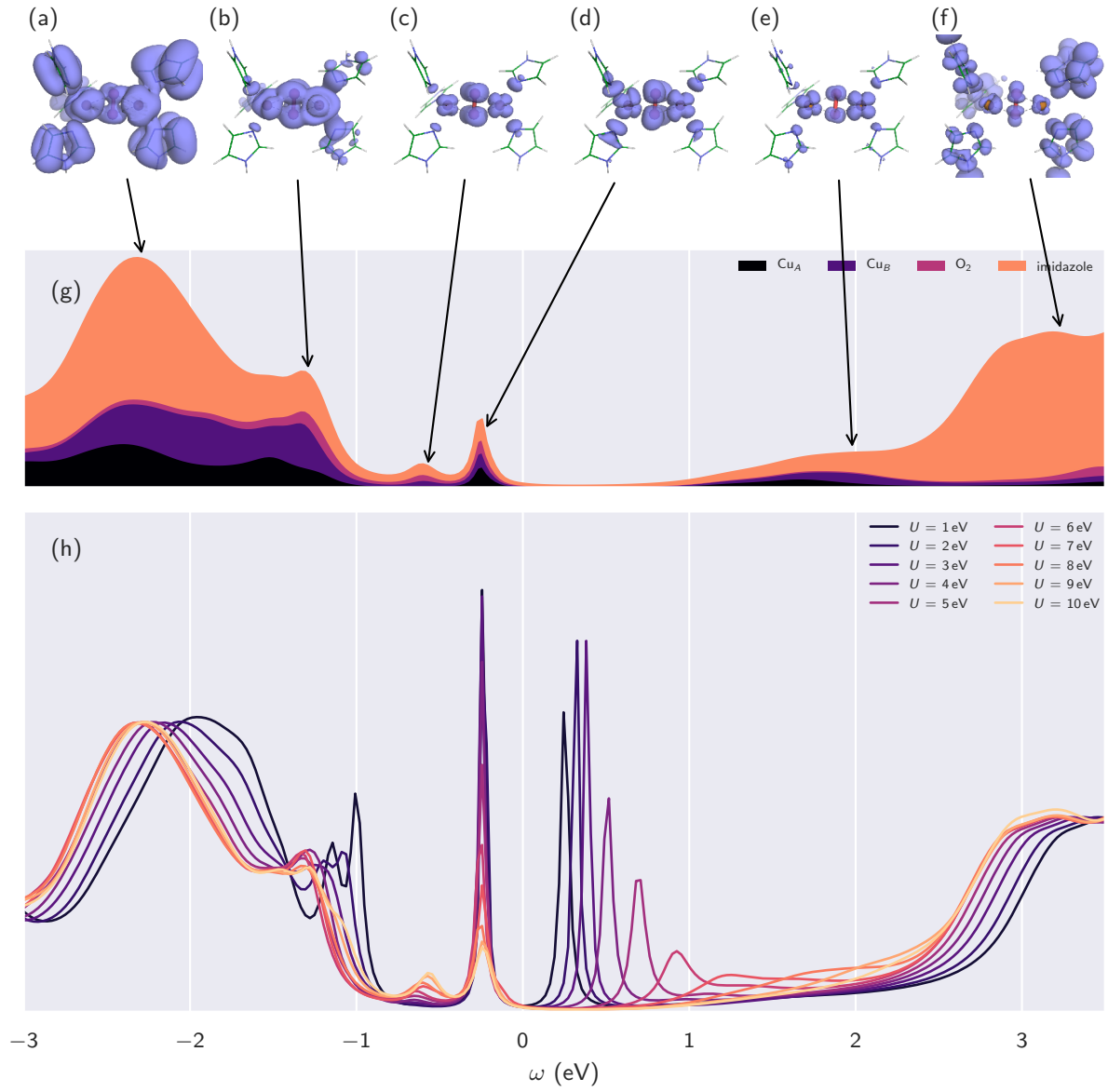


Figure 5.13: (g) The local density of states for $U = 8\,\mathrm{eV}$ and (h) the different total density of states of the system for a range of Hubbard U values from DMFT. Isosurfaces of the spectral density $\rho(\omega)$ are shown for the indicated peaks (a-f).

5.6 What about DFT + U?

Thus far, we have only applied our DMFT framework to the haemocyanin model. It is worthwhile considering if lower-level theories such as DFT + U can recover the same results, but with substantially less effort. We have also been treating U as a variational parameter, when the results of Chapter 3 allow us to determine it *ab initio*, so we will first revisit these results in the context of haemocyanin.

Table 5.2: Values of U and J (eV) for haemocyanin, calculated using the various linear response schemes introduced in Subsection 3.2.1.

atom	scalar	averaged 1×1	1×1		scaled 2×2	
	U	U	$U^{\uparrow}$	$U^{\downarrow}$	U	J
Cu_A	6.83 ± 0.01	2.16 ± 0.01	0.75 ± 0.00	3.56 ± 0.01	6.78 ± 0.05	0.90 ± 0.01
Cu_B	6.87 ± 0.07	2.31 ± 0.01	4.26 ± 0.02	0.36 ± 0.01	6.88 ± 0.07	0.87 ± 0.01
O_1	7.10 ± 0.00	2.18 ± 0.00	2.38 ± 0.00	1.99 ± 0.00	7.10 ± 0.04	0.86 ± 0.00
O_2	7.11 ± 0.00	2.16 ± 0.00	2.11 ± 0.00	2.20 ± 0.00	7.11 ± 0.04	0.86 ± 0.00

5.6.1 Linear response

Our novel linear response methods were applied to the oxyHc system. The resulting values for U and J are contained in Table 5.2. In particular the 2×2 results of $U = 6.8\,\text{eV}$ and $J = 0.9\,\text{eV}$ are very similar to the $U \approx 8\,\text{eV}$ and $J = 0.8\,\text{eV}$ that we chose when thinking of U as a free parameter.

However, there are some subtleties worth highlighting here. The results of Chapter 3 were the product of careful consideration of how linear response relates to the functional form of the DFT + U corrective functional. We should therefore be cautious taking these values and immediately using them in the DFT + DMFT framework, where U and J appear in a different context (namely, the Slater-Kanamori Hamiltonian of Equation 4.10 rather than the DFT + U + J correction of Equations 3.2 and 3.4). More work is required to determine the correct prescription for using the minimal-tracking definition $U = dv_{Hxc}/dN$ in the context of DMFT. (Ref. 253 provides some insights on this front.) That said, it is not entirely unreasonable to take the results of Table 5.2 at face value and use them directly in the context of DFT + DMFT; this would be akin to what others have done using Hubbard parameters determined via cRPA.[395]

5.6.2 Optical spectra

The optical spectrum of our oxyHc system was calculated using DFT, DFT + U, and TDDFT. The resulting spectra are shown in Figure 5.14, alongside the DMFT and experimental spectra for comparison.

The DFT and DFT + U calculations were performed using ONETEP (using the same computational approach as that described in Section 5.2). Two DFT calculations were performed: in the first, spin symmetry was imposed (*i.e.* $n^{\uparrow}(\mathbf{r}) = n^{\downarrow}(\mathbf{r})$); in this scheme, it is impossible for a spin-singlet to form between the copper atoms.[†] The second calculation was allowed to break this symmetry, with the copper $3d$ subspaces artificially polarised such that one is majority spin-up and the other majority spin-down. This achieves the desired spin state to some extent, as the system is antiferromagnetic rather than diamagnetic. (As discussed in Subsection 5.1.2, DFT is unable to capture the ground state's multi-reference character.) Both of these calculations produced spectra with several very low-energy features that are absent in experiment. The DFT + U calculations see some improvement, with these low energy features shifted up by several eV, but not enough to align them with experiment.

[†]This is of course a poor approximation. Nevertheless, we are still interested in this result because this calculation provides the initial density fed to the DMFT calculations.

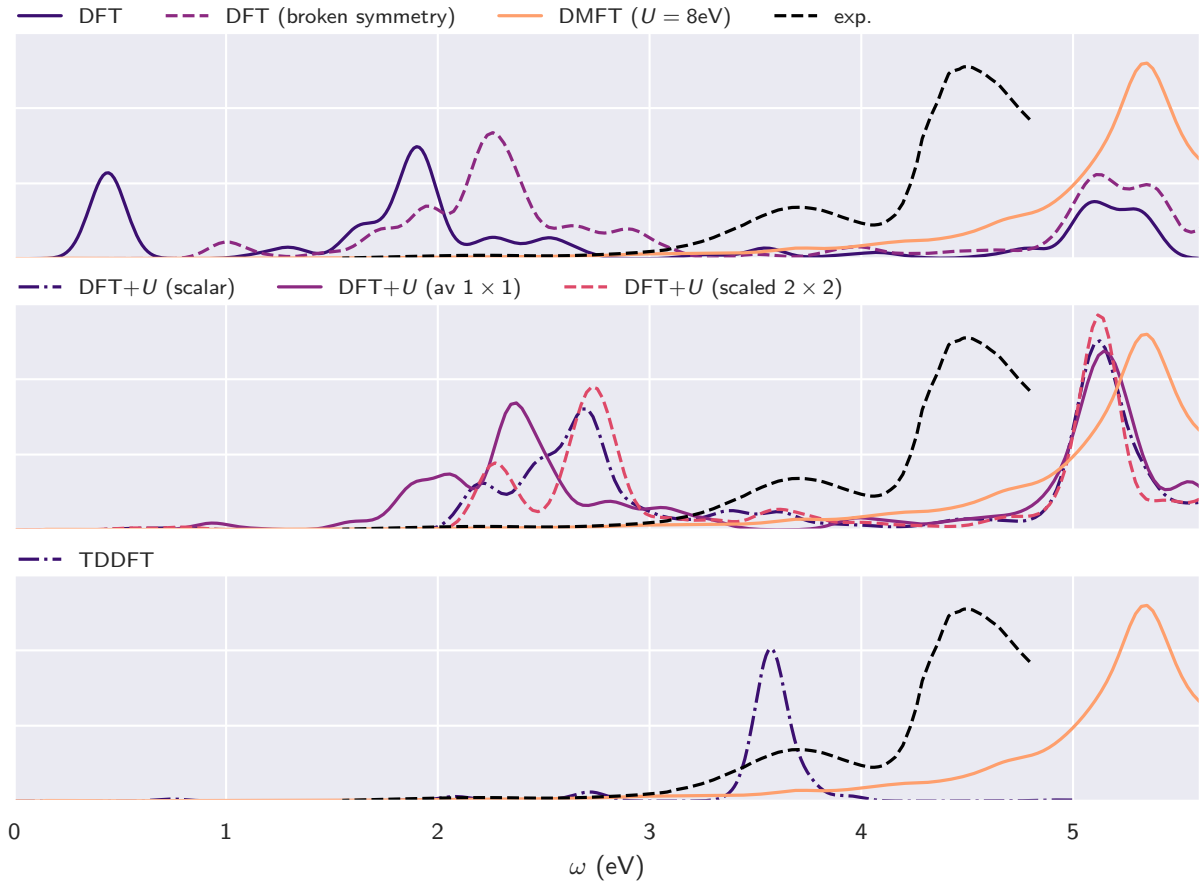


Figure 5.14: The optical spectra of oxyHc given by DFT (PBE), DFT + U, hybrid TDDFT (ωB97XD), DMFT ($U = 8\,\text{eV}$) and experiment.[390]

The spectrum was also calculated with unrestricted TDDFT, using the range-separated hybrid xc functional ωB97XD (which includes Grimme D2 dispersion corrections) and an atom-centred basis set (aug-cc-pVDZ).[1] These calculations were performed with Gaussian09.[396] They predict a strong absorption peak at 3.6 eV, in agreement with experiment, but the feature at 4.5 eV is notably missing. More importantly, the ground-state electronic structure at this level of theory is not the antiferromagnetic singlet state that is observed experimentally. Like DFT, TDDFT cannot capture the multi-reference character of the ground state, so any agreement between TDDFT and experiment is likely to be for the wrong reasons. Thus we see that both (a) adding Hubbard-like physics via a $+U$ correction or (b) adding dynamical effects via TDDFT partially shift the erroneous low-energy excitations upwards, but neither approach individually reproduces the optical absorption spectrum of haemocyanin.

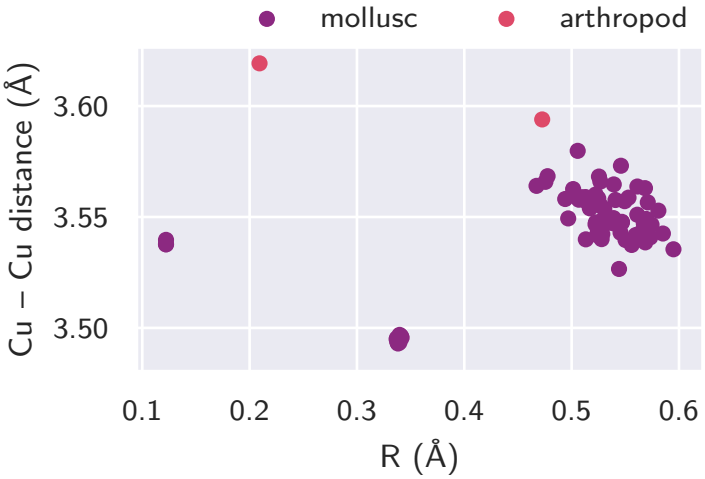


Figure 5.15: The Cu – Cu distance only very weakly anticorrelates with the butterfly effect, as shown by this survey of Cu_2O_2 geometries extracted from experimental XRD structures (Refs. 81, 352, 400–402).

5.7 Conclusions

This chapter has presented the application of our DFT + DMFT approach to oxyHc, a molecule of important biological function containing multiple correlated centres. The reduced density matrix of the $3d$ subspace of the two Cu atoms revealed the presence of fluctuating spin states, in which a Cu_2 d^{18} singlet component is maximised at $U = 8\,\text{eV}$ in spite of the butterfly distortion of the Cu_2O_2 core. The Hubbard U is necessary to capture the multi-reference character of the ground state, placing oxyHc in the limit of a true quantum entangled singlet in the limit of the Heitler-London model, with the highest occupied molecular orbital likely providing a pathway for antiferromagnetic superexchange. However, the DFT + DMFT approach had mixed success reproducing the experimentally-observed peaks in the absorption spectrum at around 2.2 eV, 3.6 eV, and 4.5 eV.

It has been previously suggested[397] that the catalytic properties of haemocyanin are related to reduced butterfly distortion, as mollusc haemocyanins are 30 times more active than their arthropod counterparts,[353] and EXAFS measurements[398,399] indicate that the Cu – Cu distance is greater (and ergo one would infer that the butterfly distortion is likely weaker) in molluscs. However, XRD structures[81,352,400–402] show shorter Cu – Cu distances in molluscs, and only weak anticorrelation between Cu – Cu distance and the butterfly effect, as shown in Figure 5.15. More work is required to explain the remarkable activity of mollusc haemocyanins.

This chapter provides a starting point for studying biological activity of oxyHc and related type 3 Cu-based enzymes by (a) establishing that the singlet can survive the butterfly distortion, thereby resolving a prior inconsistency between structural data, spectroscopy, and first-principles calculations, and (b) by providing a framework for subsequent studies to account for the effect of the protein "scaffolding" in which the active site sits, without compromising the accurate treatment of strong electronic correlation.

Chapter 6

The oxygen evolving complex

We come now to the system that motivated this thesis: the OEC of PSII (Figure 6.1).

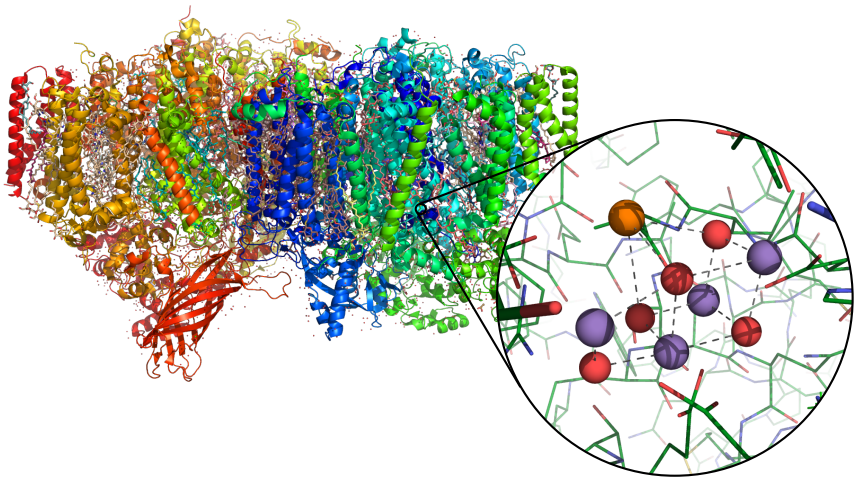


Figure 6.1: Photosystem II and the oxygen evolving complex (inset), as revealed via XFEL spectroscopy in PDB 4UB6.[34]

6.1 A large cluster model of the OEC

As Section 2.2 argued, the accurate simulation of metalloproteins requires many hundreds of atoms to be simulated at the quantum-mechanical level. There, I enumerated various failures of small cluster models. In the specific case of the OEC, studies have shown that spectroscopic properties do not adequately converge (with respect to cluster size) for 225-atom systems,[95] and QM/MM approaches fail to agree with analogous QM-only calculations in the large-system limit. Motivated by these observations, I generated a cluster model of unprecedented size of the S_1 state of the OEC, starting from the 2015 "radiation-damage-free" XFEL crystal structure,[34] and optimised its geometry at the semi-local DFT level. This cluster model will serve as a starting point for future DFT + U and DMFT calculations.

6.1.1 Cluster preparation

An approximately spherical cluster of radius 13 Å around the OEC was cut out from monomer A of the "radiation-damage-free" XFEL crystal structure (PDB record 4UB6), corresponding to the dark-stable S_1 state of the OEC. The cluster includes the core $CaMn_4O_5$ cluster itself, 91 amino acid residues, the two chloride ions, and 48 water molecules. In all, the cluster contains a total of 1631 atoms. Wherever a peptide chain extended out well beyond 13 Å, it was truncated and capped with a neutral methyl group.

Because XRD does not detect the presence of hydrogen atoms, they had to be added separately to the cluster. For the purposes of this work, all of the oxygen atoms within the cluster were assumed to be unprotonated, and all exterior oxygen atoms were assumed to be water molecules, in line with Ref. 403 (among many others). The hydrogens added to the water molecules needed to form an optimal hydrogen-bonding network. Establishing and verifying the hydrogen-bonding network proved to be a time-consuming process. Candidate networks were generated using the molecular modelling interface MAESTRO, which performs automated hydrogen-bond optimisation.[404] The optimisation procedure used the OPLS 2.1 force field[405,406] to find the lowest energy configuration via a Monte-Carlo algorithm, allowing for the reorientation of water molecules and 180° flips of asparagine, glutamine and histidine. The hydrogen-bond optimisation was performed with protonation states of residues predicted for a pH of 7.0, as determined by a PROPKA pK_a prediction.[407,408]

The structure was then screened for poorly oriented hydrogen atoms. By flagging pairs of atoms that ought to form hydrogen bonds but did not, and also checking for unrealistically short H–H distances, candidate hydrogen-bonding networks were evaluated and improved upon. (*N.B.* the above process was carried out as part of my Master's.[127])

6.1.2 Geometry optimisation

The geometry of this cluster model was then optimised using ONETEP versions 4.5 to 5.1 with the PBE xc functional.[136] The calculation was spin-polarised, with an energy cut-off of 897 eV. Each manganese atom had thirteen NGWFs, the calcium had ten; nitrogen, carbon, oxygen, and chlorine atoms four; hydrogen atoms, one. All NGWFs had a cut-off radius of 13.0 a_0. PAW potentials from the JTH dataset were used.[137] The protein was immersed in an implicit solvent parametrised to mimic water ($\varepsilon_\infty = 80$, $n_0 = 0.00035 a_0^{-3}$, and $\beta = 1.3$ following Ref. 138). The overall cluster had a charge of -2, which was calculated assuming the HOS paradigm. Ensemble density functional theory (EDFT) was used in order to stabilise the possibly non-insulating vacuum state for generating the solvent cavity. (I later modified the code to use the initial guess for the density to immediately generate a solvent cavity, which would then be used instead of a vacuum to calculate the final cavity. While this cavity would not be especially accurate, at the very least it ought to avoid any pathologies associated with the vacuum calculations.)

The geometry was optimised using the BFGS algorithm, updating the solvent cavity (being a function of the charge density) every five BFGS steps. The positions of the innermost 200 atoms were optimised; the rest were fixed in place (see Figure 6.2).[28]

Some development of ONETEP was required to perform these calculations, including over-

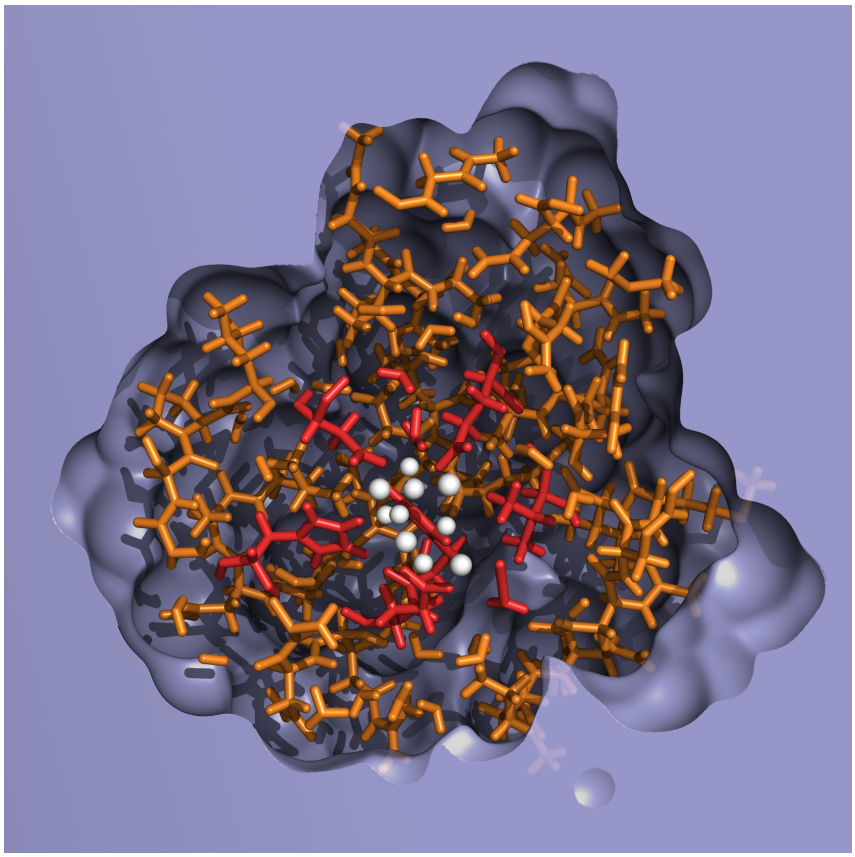


Figure 6.2: A cross-section of the 1631-atom, 13 Å radius cluster model of the OEC, showing the oxygen evolving complex (white), with surrounding protein (with those atoms that were optimised in red, and those that were constrained in orange) and the $\varepsilon = \varepsilon_\infty/2$ isosurface of the implicit solvent (blue).

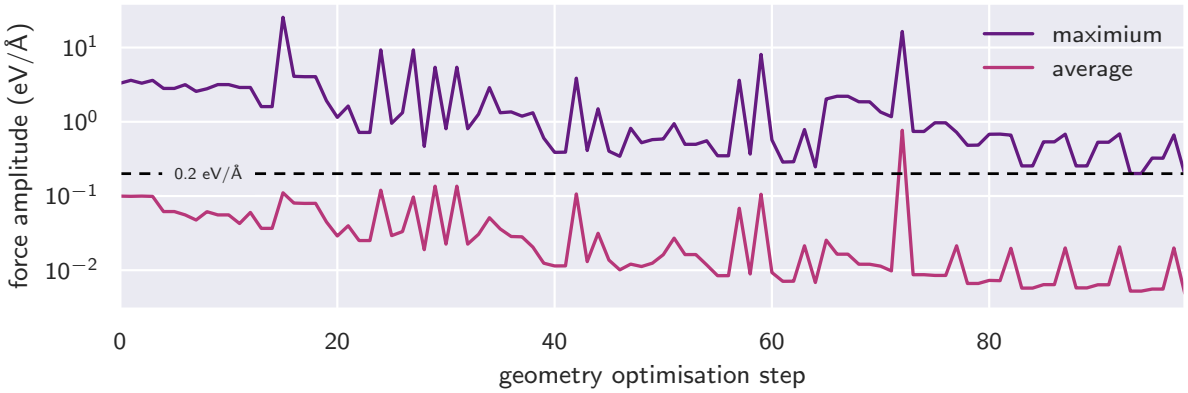


Figure 6.3: The maximum and average force on the atoms in the OEC cluster model as it was optimised.

hauling logic associated with when to regenerate the implicit solvent during a geometry optimisation, and developing smooth implicit solvent exclusion regions in order to prevent pockets of implicit solvent forming within the cluster model. (This is undesired: all the water molecules within the cluster were explicitly included, so the implicit solvent should be restricted to outside the cluster.) This is discussed further in Appendix A.9.

The convergence of the atomic forces during the optimisation of the OEC are shown in Figure 6.3. The final structure obtained has all atomic forces less than 0.2 eV/Å, with the average atomic force much lower than that.

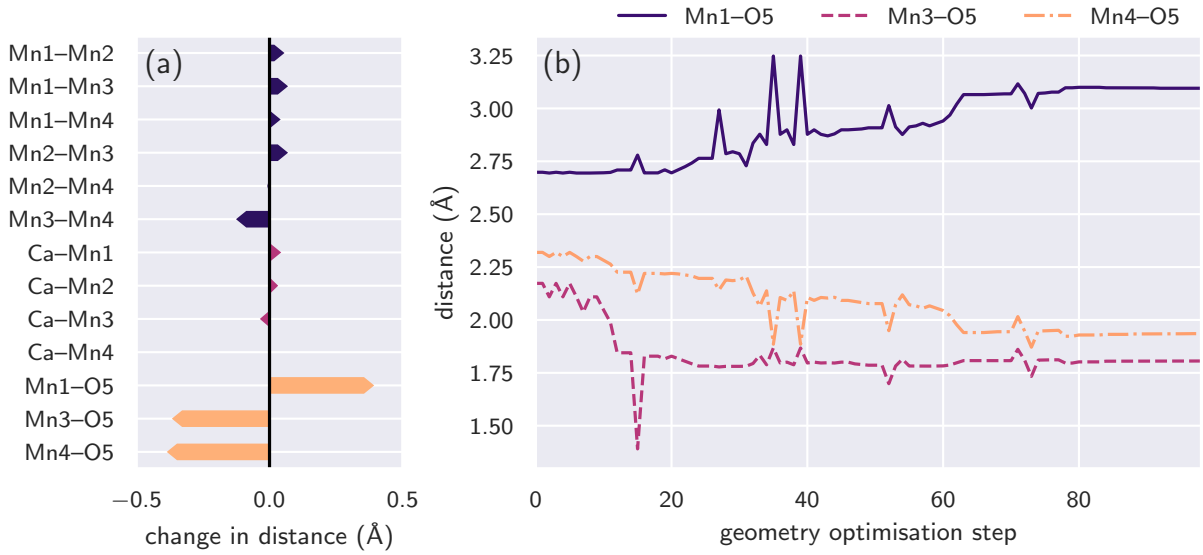


Figure 6.4: (a) The change in selected metal-metal and metal-oxygen distances after optimisation of the 1631-atom cluster model of the OEC, and (b) the change in Mn–O5 distances during the optimisation process. Refer to Figure 6.5 for the atom labels.

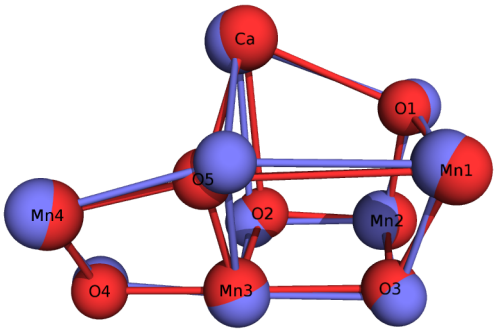


Figure 6.5: The geometry of the OEC core of the 1631-atom cluster model before (blue) and after (red) optimisation.

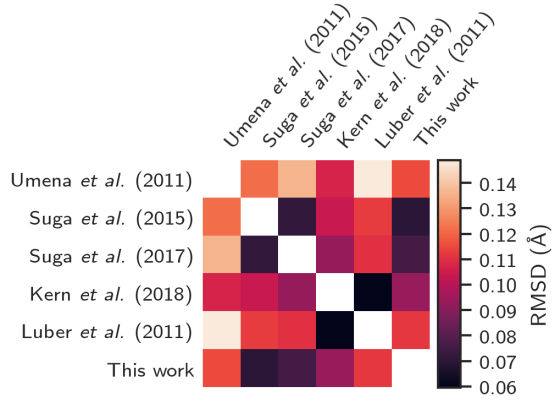


Figure 6.6: RMSDs of the metal-metal distances in the OEC between several experimental and computational studies.

6.1.3 Analysis of the optimised geometry

The changes in selected metal-metal and metal-oxygen distances during optimisation are shown in Figure 6.4. The characteristic shift of O5 away from Mn1 and towards Mn4 is clear, with the Mn3– and Mn4–O distances finishing at the more typical ~ 1.9 Å for Mn–O bonds.[38,39] Given that these are the first calculations on the OEC of this size, we can already rule out ambient protein effects as the source of the inconsistencies in the Mn–O5 distances in the experimental structures. Metal-metal distances are compared in Table 6.1 and Figure 6.6. The optimised distances of my cluster model compare very well to XRD structures; there are different experimental structures that differ more with one another than with my optimised structure.

Table 6.1: Metal-metal distances in the OEC, as given by various experiments, as well as the two computationally-optimised structures used in this chapter. Refer to Figure 6.5 for the atom labels. All of these results agree qualitatively; the subtle differences are captured in Figure 6.6.

Authors	Method	MnX–MnY						Ca–MnX			
		1–2	1–3	1–4	2–3	2–4	3–4	1	2	3	4
Yano *et al.*	EXAFS	2.7 to 2.8 × 3; 3.3 × 1						3.4 × 2			
Dau *et al.*	EXAFS	≤ 2.9 × 2; 3.3 × 1 or 2; 3.7 × 1						∼ 3.3 × 2			
Umena *et al.*	XRD	2.82	3.28	4.95	2.90	5.41	2.93	3.52	3.34	3.42	3.80
Suga *et al.*	XFEL	2.64	3.21	4.96	2.69	5.17	2.85	3.48	3.32	3.43	3.81
Suga *et al.*	XFEL	2.64	3.16	4.90	2.74	5.17	2.76	3.60	3.44	3.51	3.85
Kern *et al.*	XFEL	2.81	3.26	4.86	2.84	5.24	2.74	3.42	3.41	3.52	3.90
Luber *et al.*	R-QM/MM	2.72	3.32	4.79	2.81	5.15	2.72	3.35	3.37	3.58	3.88
This work	1631-atom QM	2.72	3.31	4.99	2.77	5.16	2.74	3.56	3.38	3.43	3.76

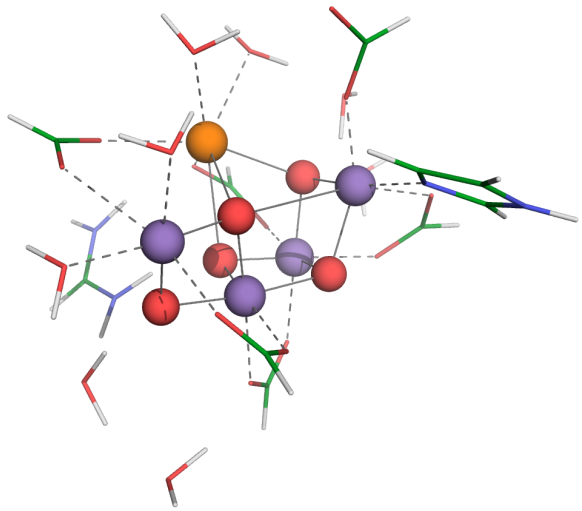


Figure 6.7: The 75-atom model of the OEC (the QM region of the QM/MM calculations of Ref. 40).

6.1.4 A smaller cluster model

The large optimised cluster model of the OEC will be invaluable for future DFT + DMFT calculations. But for the purposes of initial exploratory calculations into the electronic structure of the OEC, I used a much smaller 75-atom cluster model taken from Ref. 40. This structure has been optimised using the so-called revised-QM/MM (R-QM/MM) approach: the atomic positions of the 2011 XRD structure[16] were first optimised using QM/MM, with the QM using the B3LYP xc functional and the classical region being treated with the AMBER force field.[141] The structure was then optimised a second time, but now it was minimised with respect to the sum of the squared deviations between calculated and experimental EXAFS spectra, plus a quadratic penalty function for displacing atoms from their reference QM/MM configuration.[4] The resulting structure is in good agreement with EXAFS measurements, while also exhibiting features common to DFT calculations of the OEC (for example, the position of O5 is tightly bound to Mn4). The 75 atoms used in this thesis correspond to the QM region of the QM/MM calculations, which is pictured in Figure 6.7.

Table 6.2: Values of U and J (eV) for OEC atoms, calculated using the various linear response schemes introduced in Subsection 3.2.1.

atom	scalar	averaged 1×1	1×1		scaled 2×2	
	U	U	$U^{\uparrow}$	$U^{\downarrow}$	U	J
Mn1	4.75 ± 0.00	1.95 ± 0.00	2.26 ± 0.00	1.65 ± 0.00	4.75 ± 0.02	0.52 ± 0.00
Mn2	5.04 ± 0.00	0.95 ± 0.00	1.11 ± 0.00	0.80 ± 0.01	5.04 ± 0.03	0.54 ± 0.00
Mn3	5.02 ± 0.00	0.82 ± 0.00	0.65 ± 0.00	0.99 ± 0.01	5.02 ± 0.05	0.54 ± 0.00
Mn4	4.84 ± 0.00	1.68 ± 0.02	1.49 ± 0.02	1.87 ± 0.02	4.84 ± 0.05	0.53 ± 0.01
O1	7.83 ± 0.00	1.87 ± 0.00	1.48 ± 0.00	2.25 ± 0.00	7.83 ± 0.06	0.81 ± 0.01
O2	8.07 ± 0.00	2.19 ± 0.00	2.40 ± 0.00	1.98 ± 0.00	8.06 ± 0.06	0.82 ± 0.01
O3	7.88 ± 0.00	1.99 ± 0.00	1.78 ± 0.00	2.20 ± 0.00	7.88 ± 0.06	0.81 ± 0.01
O4	7.98 ± 0.00	2.27 ± 0.00	2.38 ± 0.00	2.16 ± 0.00	7.98 ± 0.05	0.83 ± 0.01
O5	7.97 ± 0.00	2.03 ± 0.00	2.06 ± 0.00	2.00 ± 0.00	7.97 ± 0.06	0.82 ± 0.01

6.1.5 Linear response

In preparation for subsequent DFT + U and DMFT calculations, the linear response schemes of Chapter 3 were applied to the 75-atom cluster model. These calculations used the same computational set-up as described in Subsection 6.1.2, with the exception of the pseudopotentials used, which were norm-conserving PSPs generated in-house with OPIUM[125] as opposed to PAWs. (The linear response schemes are not yet compatible with PAW; this will be discussed further in Chapter 7.) The results of the linear response calculations are listed in Table 6.2. As was the case for the hexahydrated transition metals and MnO, the Hubbard and Hund's corrections on the oxygen $2p$ subspaces are significant. The different chemical environments of the individual atoms have a small but noticeable effect on the corrective parameters. In the future the geometry of the 1631-atom cluster model will be optimised using these parameters.

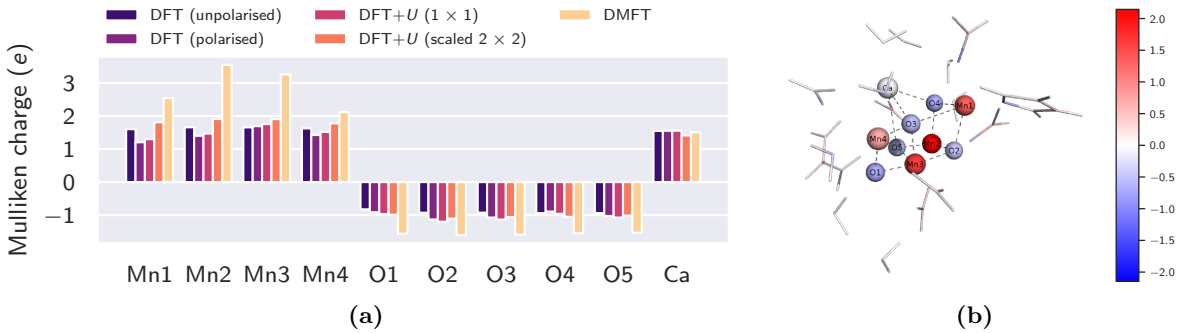
6.2 Single-site DMFT

The OEC presents some new challenges when constructing DMFT calculations. Firstly, it is impossible to perform cluster DMFT on all four manganese atoms at once — the twenty $3d$ orbitals alone (plus the several other bath orbitals one would require) takes us well beyond the reach of the Lanczos algorithm, so we must somehow reduce the size of our Hilbert space.

The first approach is to map each manganese atom to a separate AIM. This means that we are solving four AIMs with five impurity orbitals each, which will be computationally equivalent to the FePImCO calculations of Chapter 4. One of the downsides of this approach is that direct inter-manganese correlation is excluded, as the other manganese atoms are coupled via the bath. This is certainly an approximation (although how poor an approximation this is remains to be seen).

6.2.1 Computational details

Charge-conserving DMFT calculations were performed on the 75-atom Luber model, using enlarged AIM Hamiltonians with six bath sites and a further three sites coupled to each bath site via the CPT extension. The hybridisation function was fitted using the BFGS minimisation algorithm, and the AIM was solved using an ED Lanczos solver. The U and J values for each



(a) (b)

Figure 6.8: (a) The charge of the atoms in the OEC core, as given by a Mulliken analysis.[410] (b) The change in the DMFT Mulliken charges compared to those given by spin-polarised DFT.

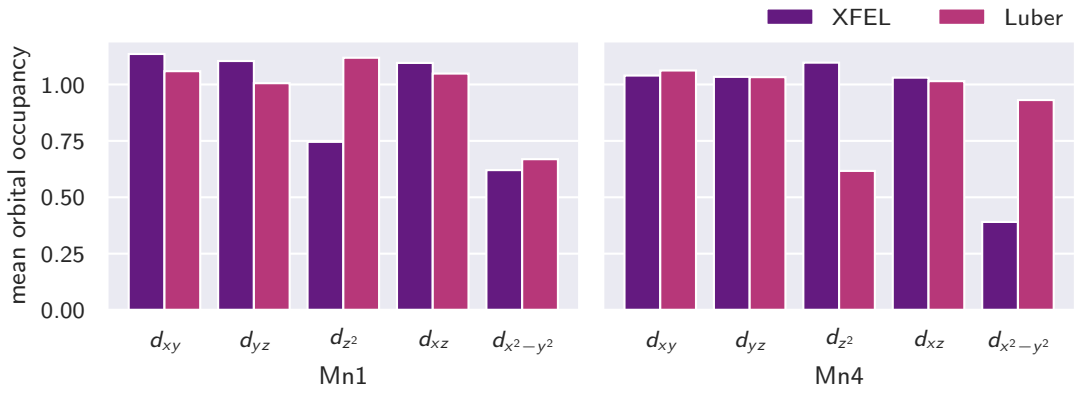
manganese atom determined via scaled 2×2 linear response in Subsection 6.1.5 were used in the AIMs (with the caveats of Subsection 5.6.1 applying here, too). In order to increase numerical stability, nominal valence double-counting was imposed — that is, the double-counting was set to correspond to the nominal occupancy of the $3d$ subspaces assuming the HOS paradigm. (Recall that in the S_1-state the formal charges of the four manganese atoms assuming the HOS are 3+, 4+, 4+, and 3+ respectively.)

6.2.2 The electronic state of the OEC

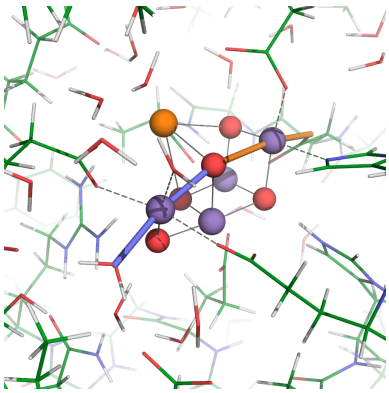
The charges of the individual atoms in the OEC as given by DFT, DFT + $U(+J)$, and DMFT are listed in Figure 6.8a. All methods report smaller values than the formal charges due to hybridisation, but DMFT yielded substantially larger local charges. These shifts in charge are localised to the OEC core (see Figure 6.8b). It would be misleading to say that strong electronic correlation has driven the charge to localise. Instead, this shift is largely a consequence of the choice of double-counting scheme, which acts as a penalty function, driving the Mn $3d$ occupancies towards the values used in the double-counting term. Ultimately it would be desirable to allow the double counting to update, but currently this prevents the DMFT loop from converging.

DMFT calculations were also performed on a 75-atom subset of the 2015 XFEL structure (in other words, the unoptimised core of the 1631-atom model), taking the same 75 atoms as the Luber structure. This provides us with an opportunity to examine the Jahn-Teller physics at play in the OEC. Recall from Section 3.3 that Jahn-Teller distortion in tetrahedrally-coordinated Mn d^4 complexes lowers the energy of the system by elongation along the z-axis, lowering the energy of the d_{z^2} orbital at the expense of the $d_{x^2-y^2}$. Alternatively, tetragonal compression can achieve the reverse.

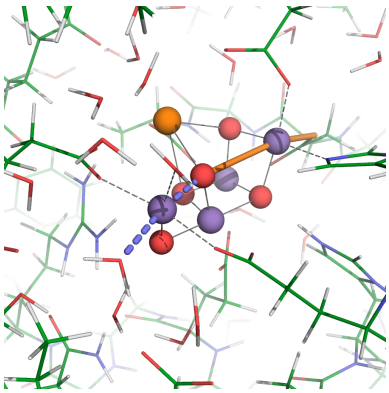
We can clearly observe the Jahn-Teller effect in the occupancies of the $3d$ orbitals of Mn1 and 4 (Figure 6.9). In the case of the XFEL structure, there is tetragonal elongation along the Mn4–O5 axis (Figure 6.9b), which is reflected in the vastly larger d_{z^2} occupancy. This is reversed in the Luber structure (Figure 6.9c), where we instead have tetragonal compression. The picture is slightly less clear with Mn1: in both structures there is unambiguously tetragonal elongation along Mn1–O5, and in the Luber structure this is clearly reflected in the d-orbital occupancies, but this is not true of the XFEL structure. Perhaps this departure from the conventional



(a)



(b) before optimisation (XFEL)



(c) after optimisation (Luber-like)

Figure 6.9: (a) The orbital occupancies of Mn1 and Mn4 in the Luber and XFEL models as given by the DMFT reduced density matrix. Note that these orbital labels correspond to the local axes defined in Figure 6.11, which differ marginally between models. (b) and (c) The Jahn-Teller axes of Mn1 (orange) and Mn4 (blue) in the 1631-atom cluster before and after geometry optimisation, showing elongation (solid) and compression (dashed).

Jahn-Teller picture is due to the substantially electronegative carbonyl ligand coordinating with Mn1. Note that since Mn2 and 3 have formal occupancies of d^3, we do not expect them to be Jahn-Teller-active.

Turning now to the spin state, it is known that the S_1 state of the OEC cluster has no net spin, as identified by the absence of an EPR signal.[51,411] The manganese atoms will individually be in high spin states, with Mn1 and Mn4 being spin quintets; Mn2 and 3, quadruplets (as naïvely determined by their formal charge in the HOS paradigm). The effective spin S_{eff} of the manganese atoms as predicted by single-site, paramagnetic DMFT are 1.93, 1.62, 1.66, and 2.00 respectively.[†] A more detailed breakdown is provided by Figure 6.10.

In order to achieve zero net spin, the spins of Mn1 and 3 are opposite to that of Mn2 and 4. This is reproduced by spin-polarised DFT and DFT + U: for example, DFT + U + J calculations on the Luber structure using the Hubbard and Hund's parameters listed in Table 6.2 give $n^{\uparrow}_{3d} - n^{\downarrow}_{3d} = 3.8$, -3.3, 3.4, and -4.0 for Mn1 through 4.

The single-site DMFT approach has a substantial flaw: by removing all inter-manganese correlation, it does not — or more accurately, cannot — exhibit any spin ordering. This severely

[†]Recall that we defined S_{eff} via $\text{Tr}[\hat{S}^2\hat{\rho}] = \hbar^2 S_{\text{eff}}(S_{\text{eff}}+1)$.

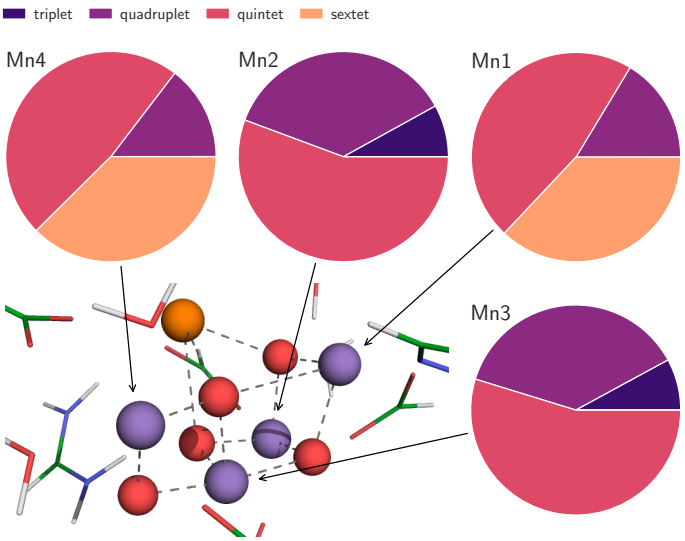


Figure 6.10: The spin states of the Mn 3*d* correlated subspaces in the OEC, calculated from the DMFT reduced density matrices as described in Subsection 4.3.2.

hampers our ability to study the OEC: for example, theories of interconvertibility in the S_2 state hinge on the existence of a low- and a high-spin state (refer back to Subsection 1.2.4). The single-site DMFT calculations as they stand would be incapable of investigating this.

6.3 Going beyond single-site paramagnetic DMFT

In the future, there are several approaches we could take to overcome these issues.

6.3.1 Ferromagnetic single-site DMFT

In order to re-introduce spin ordering, one approach is to perform non-paramagnetic DMFT. In this method, we start with the spin-unpolarised DFT density and Hamiltonian as before,[†] but now the Green's function, self energy, Hamiltonian, and electronic density are permitted to break spin-symmetry; that is, they gain a spin index. In order to break the spin-symmetry of the initial DFT density, the self-energy is initialised to a small, spin-dependent and static value.[‡] Ferromagnetic DMFT calculations on the OEC are underway and ongoing.

6.3.2 Cluster DMFT

Projection

If we ever want to include inter-atomic correlation, multiple manganese atoms would need to be included in the same AIM. We have done this already in the case of haemocyanin, so let us return to consider the strategy of projection that we used there. Projection requires local axes to be constructed for the four Mn subspaces; these are shown in Figure 6.11.

[†]We do not start from a spin-polarised DFT calculation because the DMFT calculation would inherit problems such as spin contamination.

[‡]Compare this to the standard approach, where the self-energy is initialised to be zero.

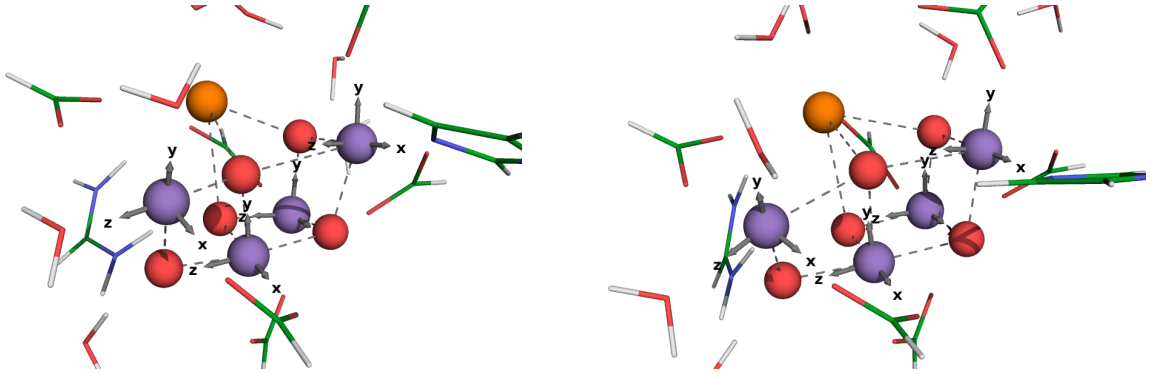


Figure 6.11: The local axes for the Mn $3d$ correlated subspaces in the OEC for (a) the Luber model and (b) the XFEL model.

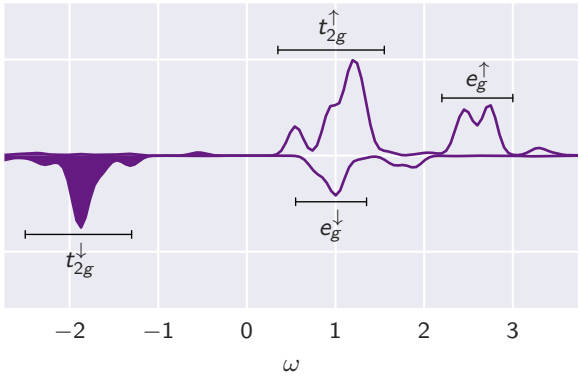


Figure 6.12: The LDOS of Mn2, projected onto the $3d$ Hubbard subspace and split by spin channel, as given by spin-polarised DFT (PBE). The t_{2g} and e_g orbitals are labelled accordingly.

However, there are some distinct differences between the cases of haemocyanin and the OEC. Haemocyanin was remarkably well-suited to projection: there was only a single hole on each copper atom and there were several low-lying $3d$ orbitals that were guaranteed not to play a part in the dynamics, so they could be projected out without concern. The case of the OEC is much more complicated. In the S_1 state, the manganese atoms will have formal occupancies of d^3 or d^4. It is possible that the t_{2g}–e_g splitting[†] of the $3d$ orbitals will mean that some half-filled orbitals will play a lesser role, but that is by no means guaranteed.

If we inspect the LDOS of the manganese atoms (as given by spin-polarised DFT), it becomes evident that projection would be ill-advised. For the projection to be successful, we need the $3d$ correlated subspace to be divided into distinct energy windows: one, containing those orbitals near the Fermi surface which will be retained, and the other, containing orbitals far away from that energy window that will have a negligible contribution to the Green's function. However, this is not the case, as Figure 6.12 demonstrates: the t_{2g} and e_g orbitals are equally close to the chemical potential, so freezing one of these subsets out would be unreasonable as it would drastically affect the physics of the AIM.

These local axes are still useful: they can still be used for all DMFT calculations, just not

[†]Refer back to Figure 3.4.

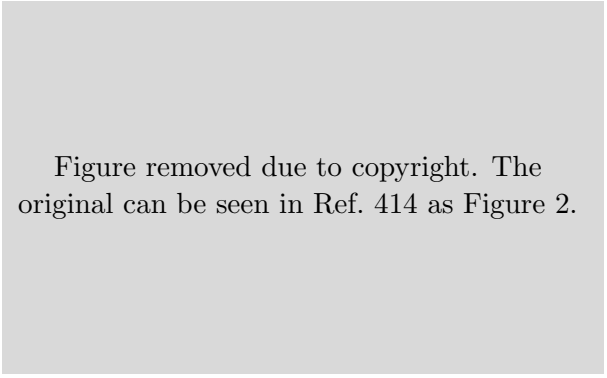


Figure 6.13: A cartoon of the rr-DMFT approach for four correlated sites, showing the construction of multiple dimer DMFT problems from an original four-site problem. $C_{1\text{-}4}$ are smaller clusters that collectively make up the full cluster problem; in our case each would correspond to a manganese $3d$ subspace. Figure adapted from Ref. 414.

for projection.[†] Having a nearly-diagonal Green's function is numerically convenient as it will lead to simpler AIMs.

Dimer DMFT

Having ruled out projection, the next step could be to perform "dimer" DMFT. This involves mapping the OEC to two AIMs, each representing two manganese atoms, thereby reintroducing explicit correlation within the manganese dimers. A framework for this functionality exists within ONETEP+TOSCAM, but it is currently untested and disabled. Further code development would enable these calculations. These dimer AIMs would also be larger than any we have solved thus far, and would likely require either (a) further optimisation of the ED solver or (b) a different solver that scales less poorly with AIM size, such as CTQMC.[412] This dimer approximation may be a good one: DMRG calculations on the OEC core showed that the Mn1 and 2 are relatively disentangled from Mn3 and 4 in the S_0 and S_1 states.[42‡] Additionally, Ref. 53 proposed that a coupled dimer model can best explain the hyperfine ^{55}Mn structure of EPR structure of the S_2 state (although these calculations assumed the LOS).[413]

rr-DMFT

A further extension to this approach that could be explored is the so-called real-space renormalised dynamical mean field theory (rr-DMFT) approach of Imada and co-workers.[414] This method involves decomposing n-site clusters into $\binom{n}{2}$ dimer DMFT calculations, as shown in Figure 6.13. The Green's functions of these individual dimer calculations are then amalgamated to form an approximate solution to the original n-site problem.

[†]Indeed, these local axes were used in the definition of the individual d orbitals whose occupancies were listed in Subsection 6.2.2.

[‡]I would caution against treating these calculations as gospel truth because they neglect the ambient protein environment.

6.4 Conclusions

This chapter has laid the groundwork for large-scale DFT + DMFT calculations on the OEC. The geometry optimisation of a 1631-atom cluster model provides the perfect starting point. Preliminary DMFT calculations have demonstrated that DMFT has the potential to yield unique insights into the electronic structure of the OEC, with the structure of the OEC clearly being highly sensitive to the electronic structure of the manganese atoms. These calculations have also demonstrated several of the challenges that remain: primarily, how to properly account for inter-manganese correlation while keeping the calculations tractable. This will be discussed further in the final chapter. As always, comparison with experiment remains a rigorous test of computation. There is scope for comparison: for example, magnetic circular dichroism spectra — which we can calculate — have been recently proposed to be an excellent probe of the OEC.[415]

If we can perform DMFT calculations on the OEC, it will be a major breakthrough. Studies of biological systems considering thousands of atoms at the DMFT level of theory have already borne fruit: simulations of myoglobin have shown that DMFT corrects the fact that DFT gives an unphysical imbalance between the binding of myoglobin with oxygen and carbon monoxide.[79] In this scenario, the explicit inclusion of many-body effects via DMFT prompted a rearrangement of the Fe 3*d* orbital occupancies and enhanced electron transfer to the O_2 molecule, resulting in a crucial shift in the Fe–O_2 binding energy. Similar shifts in Mn 3*d* orbital occupancies and altered electron transfer between the manganese and oxygen ions for the case of the toy 75-atom model may prompt altered bonding affinities between the ions of the OEC and/or redirection of Jahn-Teller axes, which could hypothetically account for the unresolved disagreement between experimental and computational models of the S_1 state (Subsection 1.2.1). Similar changes could also favour (or rule out) a particular oxidation state paradigm (Subsection 1.2.2), and a detailed picture of the electronic structure of the S_3 state could help clarify the bond formation mechanism (Subsection 1.2.6).

Chapter 7

Conclusions

This thesis has identified the challenges to accurate quantum-mechanical simulation of metalloproteins, and has developed techniques for overcoming them.

In Chapter 2 I demonstrated how linear-scaling DFT is a powerful tool for performing quantum-mechanical simulations on systems comprising of thousands of atoms, which I argued was necessary for the accurate simulation of proteins (transition-metal-containing or not). The capabilities of the linear-scaling DFT code ONETEP were demonstrated on the cyclotide kalata B5 protein. I also discussed the shortcomings of semi-local DFT for accurately treating electronic correlation.

To overcome these shortcomings, I explored, extended, and implemented two methodologies: DFT + U and DMFT. In the case of DFT + U (Chapter 3), a case was made for determining the Hubbard parameters via linear response, and a novel approach for accounting for the non-interacting contribution — so-called "minimum-tracking" linear response — was developed and expanded upon. I also noted that discrepancies exist between conventional linear response and the contemporary DFT + U functional. To resolve these, I paid particular attention to spin and associated screening, proposing revised definitions for Hubbard and Hund's parameters (Subsection 3.2.1). By comparing scalar linear response to my spin-specific theory, I demonstrated that the treatment of inter-spin screening in conventional linear response is somewhat inconsistent with the DFT + U functional as it is most commonly employed (Subsection 3.2.2). While I do not claim to have arrived at an ultimate solution to this inconsistency, I do provide a simple technique by which inter-spin screening of the Hubbard U may be suppressed. This results in spin-dependent U parameters that are generally lower in value than the canonical U for the partially-filled spin channel of a localised subspace (the spin channel that usually harbours the strong correlation effects) and that, in principle, could be applied to that spin channel alone. This hints at a possible solution to the widespread finding that first-principles U parameters can be rather too large in practice, leading to over-correction by DFT + U.

In the latter half of that chapter I applied our theoretical developments to a complete set of hexahydrated transition metal complexes from Ti to Zn. I calculated Hubbard and Hund's parameters using conventional and novel approaches, and then performed DFT + U calculations using these parameters to predict structural and spectroscopic properties. The numerical stability of the minimum-tracking formalism (in which Hubbard parameters are a strictly ground-state

property) allowed me to investigate closed-shell cases with confidence. The Hubbard corrections to oxygen $2p$ subspaces proved to be far from negligible, and helped to obtain sensible structural predictions. Spectroscopic simulations of coordination complexes using DFT + U saw mixed success, whereas indirect band gap results for the long-standing challenge material MnO are very promising when compared against a wide range of more computationally demanding approximations.

The development of the DFT + U method is ongoing: issues remain relating to local *vs.* global curvature, static correlation error, and the precise formulation of $+J$ corrections (among many others). One issue with DFT + U that I have not yet directly addressed is the matter of energy comparability. With Hubbard parameters obtained via linear response being dependent on the chemical environment, it is unclear if DFT + U energies of different structures (now with different corrective parameters) are comparable.[416] Ultimately, DFT + U is a correction to an incorrect energy curvature, so, in principle, comparison of total energetics should improve with the addition of a Hubbard correction. In practice, there is little consensus on this front. Indeed, many authors[417–420] keep U fixed across a wide range of chemical environments — which can become very problematic when the electronic structure of the system changes appreciably. Other authors[421] update U and obtain reasonable results, and there are even some who mix GGA and GGA + U energies.[422,423] Resolving this conflict would dramatically improve DFT + U-based *ab initio* random structure searching.[424]

One self-contained avenue for immediate work is the extension of our minimum-tracking methodology to the PAW formalism (Subsection 2.3.7). This should be a straightforward extension to account for PAW augmentation density when measuring projected potentials. As Subsection 2.3.7 illustrated, the PAW method is a powerful and accurate method of treating core electrons, and being able to make use of it would be advantageous. (Since the Hubbard and Hund's parameters are sensitive to the atomic potential used, one cannot apply parameters determined using norm-conserving PSPs to a system that utilises PAWs.)

The theoretical and computational advances of Chapter 3 are by no means specific to metalloproteins. That said, they do address several key concerns when performing DFT + U on metalloproteins, not least by providing an approach for performing linear response within a linear-scaling framework.

Chapter 4 introduced the DFT + DMFT method. This approach is much more computationally intensive than DFT + U, but — in principle — should more accurately describe strong electronic correlation. The overhauled implementation of ONETEP+TOSCAM was presented. This code can account for strongly correlated electronic behaviour while simultaneously including the effects of protein environments, making it ideally suited for studying biological activity in metalloproteins. Developing and improving this code accounted for a substantial portion of my PhD: the ONETEP DMFT module was updated and incorporated into the active development version, meaning the most recent (and all future versions) of ONETEP will be compatible with TOSCAM. The TOSCAM code was significantly restructured to remove unnecessary external dependencies and to simplify the compilation procedure, with an eye on distributing the code in the near future. This ONETEP+TOSCAM interface will continue to be developed and tested. In particular, a GPU implementation of the ED solver will be incorporated, as well as a CTQMC

solver (that should in principle allow us to solve much larger AIMs).

The capabilities of the ONETEP+TOSCAM code were first demonstrated on the toy example of FePImCO, where I presented results on the details of its electronic state and the CO dissociation mechanism. Several of these results — the spin decomposition of the reduced density matrix, for example — are not obtainable with DFT or DFT + U. The ONETEP+TOSCAM approach was then applied to oxyHc, a dicopper complex that binds molecular oxygen. These calculations allowed us to establish a mechanism for explaining the spin state of the butterflied dicopper core (as opposed to the planar core observed in model complexes). This is a pertinent illustration of the importance of both (a) the ambient protein environment, which imposes the butterflied structure and (b) the accurate simulation of the copper $3d$ electrons.

However, there are some issues that remain: metal-ligand charge transfer excitations were underestimated in the optical absorption spectrum. This could be investigated by incorporating oxygen $2p$ Hubbard subspaces — or, indeed, the dioxygen π^* orbitals — as explicit impurity sites in the DMFT AIM.[†] This approach goes slightly against the spirit of DMFT — in that all of the physics of the oxygen sites ought to be captured by the hybridisation function — but it would allow us to be more definitive about the superexchange pathway. More generally, this approach would allow us to robustly consider radical ligands which cannot be modelled in DFT — for example, in the case of the binding of NO to cyctochrome c′,[425,426] and the oxo-oxyl radical attack mechanism for oxygen formation in the OEC (refer to Subsection 1.2.6). That said, it is worthwhile keeping in mind relative strengths and weaknesses of DMFT compared to competing methods. Namely: our DFT + DMFT approach does not compromise our ability to deploy linear-scaling DFT to the rest of the protein. This provides a useful middle ground between (a) density functional theory based simulations of metalloproteins, which depend heavily on the choice of functional and incorrectly describe multi-reference effects, and (b) multi-reference quantum chemistry approaches, whose unfavourable scaling prevents studies on models of realistic size. After all, if we were to include more and more orbitals explicitly in the AIM, at some stage we are effectively performing quantum chemistry. There already exist several quantum chemistry techniques for selecting active spaces and efficiently solving large active spaces (for example, the FCIQMC-CASSCF technique of Alavi and co-workers, which has already been applied to iron porphyrin[345]).

DMFT can also provide some insights that are not accessible via DFT or quantum chemistry methods. One such quantity is the dynamical spin susceptibility $\chi^{zz}(\mathbf{q}, i\omega)$. This can be calculated, for example, via TDDFT.[427,428] However, the dynamical spin susceptibility is itself an averaged quantity

$$\chi(\mathbf{q}, i\omega) = \sum_{\nu,\nu'} \tilde{\chi}_{\mathbf{q}}(i\nu, i\nu', i\omega) \tag{7.1}$$

where the individual elements $\tilde{\chi}_{\mathbf{q}}(i\nu, i\nu', i\omega)$ relate to the contributions of particular two-particle excitations that collectively make up the dynamical susceptibility. In the scope of DMFT, these

[†]Work has been carried out in the ONETEP community for defining two-centred Hubbard projectors.

individual elements can be related to the local susceptibilities[306]

$$\chi^{\sigma\sigma'}_{\mathrm{imp}}\left(i\nu, i\nu', i\omega\right) = \int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \int_0^\beta d\tau_3 e^{i(\nu\tau_1+\nu'\tau_2+\omega\tau_3)} \left\langle \mathcal{T}_\tau c_\sigma\left(\tau_1\right) c^\dagger_\sigma\left(\tau_2\right) c_{\sigma'}\left(\tau_3\right) c^\dagger_{\sigma'}(0)\right\rangle \quad (7.2)$$

which can be extracted from DMFT via a Lanczos procedure (or alternatively via a QMC approach).[429,430] By inspecting the individual elements — rather than the averaged quantity $\chi(\mathbf{q}, i\omega)$ — we can gain unique insight into the nature of spin transitions (*e.g.* the low- to high-spin transition of FePImCO during photodissociation, or the singlet-triplet transition in haemocyanin).

Finally, Chapter 6 presented calculations on the OEC of PSII. A 1631-atom cluster model, optimised at the level of semi-local DFT, represents the largest QM calculation on the OEC to date, and will serve as an excellent starting point for future DFT + U and DMFT calculations. Preliminary DMFT calculations on a much smaller cluster model demonstrated that these DMFT calculations remain especially challenging, in spite of the advances contained in this thesis, stemming largely from the sheer computational size of the problem if it was to be simply brute-forced. Chapter 6 discussed several strategies for overcoming this computational hurdle, both theoretical (including single-site ferromagnetic, dimer DMFT, and rr-DMFT) and computational (using AIM solvers with superior scaling to ED and implementing a GPU-accelerated version of the code). Meanwhile, I demonstrated that projecting out some of the manganese $3d$ orbitals would severely compromise the accuracy of the calculations.

More generally, there is a need to improve the numerical stability of these DMFT calculations. Currently, a number of issues prevent these calculations from being "routine". This is especially true in the case of the self-consistent calculations, where achieving self-consistency is a challenging process. Further work could consider the optimal strategies for the mixing of the self-energy, the mixing of the density, and the form of the double-counting term.[431,432] Tackling these challenges will serve to drive the development of ONETEP+TOSCAM forward.

There is no shortage of interesting systems to study while honing these techniques. For example, binuclear hydrogenases (NiMn, FeNi, and FeFe)[433] could be studied while perfecting the dimer and rr-DMFT methods. Copper metallothioneins — which contain $\mathrm{Cu}_x\mathrm{S}_y$ clusters where x ranges from one to over ten — could provide an excellent proving ground given that (a) the number of impurity sites can be easily varied and (b) the Hilbert space will be comparatively small compared to the OEC (given that projection could be applied to the nearly-filled copper $3d$ subspaces as in the case of haemocyanin).[434] Cluster DMFT approaches in periodic systems exhibit some non-intuitive behaviour (especially with respect to which atoms are bundled into the same AIM[435]); perhaps the same will be true in biological systems.

It would also be short-sighted to dismiss DFT + U — being less sophisticated than DMFT — as being unable to provide novel insights into the function of the OEC. After all, DFT + U calculations of the OEC provided breakthroughs explaining the results of EPR experiments.[57,436] This thesis has presented the first DFT + U calculations (to my knowledge) on the OEC using U determined via linear response,[†] and will allow re-optimisation of the OEC cluster at the

[†]Refs. 57 and 436 used a value determined via extrapolation of exchange coupling constants to match B3LYP results.

DFT + U + J level. It may be that the sensitivity of U to the environment of each manganese atom (which is captured by linear response) is incredibly important.

To summarise, this thesis has explored how strong electronic correlation can be accounted for within the scope of linear-scaling DFT for the purposes of accurately modelling metalloproteins. Firstly, it has advanced our understanding of linear response calculations of Hubbard and Hund's parameters for open-shell systems (and made it compatible with linear-scaling codes). Secondly, it has presented an overhauled implementation of linear-scaling DFT + DMFT in the form of ONETEP+TOSCAM, yielding insights into the importance of strong electronic correlation in the haemocyanin dicopper complex. Finally, it has established a strategy and framework for DFT + U and DFT + DMFT calculations on the OEC. It is my hope that we will soon be able to perform these calculations routinely, and to then investigate the electronic structure and reaction mechanism of the OEC in great detail.

References

1. J.-D. Chai & M. Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **10,** 6615 (2008).
2. T. M. Henderson, B. G. Janesko & G. E. Scuseria. Generalized gradient approximation model exchange holes for range-separated hybrids. *J. Chem. Phys.* **128,** 194105 (2008).
3. S. Huzinaga *et al. Gaussian basis sets for molecular calculations* (Elsevier, Amsterdam, 1984).
4. E. M. Sproviero *et al.* Quantum mechanics/molecular mechanics study of the catalytic cycle of water splitting in photosystem II. *J. Am. Chem. Soc.* **130,** 3428 (2008).
5. E. B. Linscott, D. J. Cole, M. C. Payne & D. D. O'Regan. Role of spin in the calculation of Hubbard U and Hund's J parameters from first principles. *Phys. Rev. B* **98,** 235157 (2018).
6. M. A. Al-Badri *et al.* Super-exchange mechanism and quantum many body excitations in the archetypal hemocyanin/tyrosinase di-Cu oxo-bridge. arXiv: `1811.05739` (2018).
7. N. Cox & J. Messinger. Reflections on substrate water and dioxygen formation. *BBA - Bioenerg.* **1827,** 1020 (2013).
8. A. Grundmeier & H. Dau. Structural models of the manganese complex of photosystem II and mechanistic implications. *BBA - Bioenerg.* **1817,** 88 (2012).
9. D. J. Vinyard, G. M. Ananyev & G. Charles Dismukes. Photosystem II: the reaction center of oxygenic photosynthesis. *Annu. Rev. Biochem.* **82,** 577 (2013).
10. D. J. Vinyard & G. W. Brudvig. Progress toward a molecular mechanism of water oxidation in photosystem II. *Annu. Rev. Phys. Chem.* **68,** 101 (2017).
11. W. Lubitz, M. Chrysina & N. Cox. Water oxidation in photosystem II. *Photosynth. Res.* 1 (2019).
12. C. Zhang *et al.* A synthetic Mn_4Ca-cluster mimicking the oxygen-evolving center of photosynthesis. *Science* **348,** 690 (2015).
13. D. W. Wakerley & E. Reisner. Oxygen-tolerant proton reduction catalysis: much O_2 about nothing? *Energy Environ. Sci.* **8,** 2283 (2015).
14. C. Bozal-Ginesta & J. R. Durrant. Artificial photosynthesis – concluding remarks. *Faraday Discuss.* **215,** 439 (2019).

15. R. Matheu *et al.* The development of molecular water oxidation catalysts. *Nat. Rev. Chem.* **3,** 331 (2019).

16. Y. Umena, K. Keisuke, J.-R. Shen & N. Kamiya. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9Å. *Nature* **473,** 55 (2011).

17. H. Dau, A. Grundmeier, P. Loja & M. Haumann. On the structure of the manganese complex of photosystem II: extended-range EXAFS data and specific atomic-resolution models for four *S*-states. *Phil. Trans. R. Soc. B* **363,** 1237 (2008).

18. P. Hohenberg & W. Kohn. Inhomogeneous electron gas. *Phys. Rev.* **136,** B864 (1964).

19. W. Kohn & L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140,** A1133 (1965).

20. R. O. Jones. Density functional theory: its origins, rise to prominence, and future. *Rev. Mod. Phys.* **87,** 897 (2015).

21. A. Jain, Y. Shin & K. A. Persson. Computational predictions of energy materials using density functional theory. *Nat. Rev. Mater.* **1,** 15004 (2016).

22. C.-K. Skylaris, P. D. Haynes, A. A. Mostofi & M. C. Payne. Introducing ONETEP: linear-scaling density functional simulations on parallel computers. *J. Chem. Phys.* **122,** 084119 (2005).

23. M. Heiss *et al.* Self-assembled quantum dots in a nanowire system for quantum photonics. *Nat. Mater.* **12,** 439 (2013).

24. N. Todorova *et al.* Dimensionality of carbon nanomaterials determines the binding and dynamics of amyloidogenic peptides: multiscale theoretical simulations. *PLoS Comput. Biol.* **9,** e1003360 (2013).

25. N. D. M. Hine, P. D. Haynes, A. A. Mostofi & M. C. Payne. Linear-scaling density-functional simulations of charged point defects in Al_2O_3 using hierarchical sparse matrix algebra. *J. Chem. Phys.* **133,** 114111 (2010).

26. F. Corsetti & A. A. Mostofi. System-size convergence of point defect properties: the case of the silicon vacancy. *Phys. Rev. B* **84,** 035209 (2011).

27. J. Dziedzic *et al.* Large-scale DFT calculations in implicit solvent – a case study on the T4 lysozyme L99A/M102Q protein. *Int. J. Quantum Chem.* **113,** 771 (2013).

28. G. Lever *et al.* Large-scale density functional theory transition state searching in enzymes. *J. Phys. Chem. Lett.* **5,** 3614 (2014).

29. A. S. Fokas *et al.* Evidence of correlated static disorder in the Fenna-Matthews-Olson complex. *J. Phys. Chem. Lett.* **8,** 2350 (2017).

30. D. J. Cole & N. D. M. Hine. Applications of large-scale density functional theory in biology. *J. Phys. Condens. Matter* **28,** 393001 (2016).

31. K. N. Ferreira *et al.* Architecture of the photosynthetic oxygen-evolving center. *Science* **303,** 1831 (2004).

32. B. Loll *et al.* Towards complete cofactor arrangement in the 3.0Å resolution structure of photosystem II. *Nature* **438,** 1040 (2005).

33. A. Guskov *et al.* Cyanobacterial photosystem II at 2.9-Å resolution and the role of quinones, lipids, channels and chloride. *Nat. Struct. Mol. Biol.* **16,** 334 (2009).

34. M. Suga *et al.* Native structure of photosystem II at 1.95Å resolution viewed by femtosecond x-ray pulses. *Nature* **517,** 99 (2015).

35. P. Joliot, G. Barbieri & R. Chabaud. Un nouveau modele des centres photochimiques du systeme II. *Photochem. Photobiol.* **10,** 309 (1969).

36. B. Kok, B. Forbush & M. McGloin. Cooperation of charges in photosynthetic O_2 evolution – I. A linear four step mechanism. *Photochem. Photobiol.* **11,** 457 (1970).

37. V. Krewald *et al.* Metal oxidation states in biological water splitting. *Chem. Sci.* **6,** 1676 (2015).

38. W. Ames *et al.* Theoretical evaluation of structural models of the S_2 state in the oxygen evolving complex of photosystem II: protonation states and magnetic interactions. *J. Am. Chem. Soc.* **133,** 19743 (2011).

39. A. Robertazzi, A. Galstyan & E. W. Knapp. Can oxidation states and the protonation pattern of oxomanganese complexes be recongised from their structures? *Cryst. Eng. Comm.* **13,** 6369 (2011).

40. S. Luber *et al.* S_1-state model of the O_2-evolving complex of photosystem II. *Biochem.* **50,** 6308 (2011).

41. A. Galstyan, A. Robertazzi & E. W. Knapp. Oxygen-evolving Mn cluster in photosystem II: the protonation pattern and oxidation state in the high-resolution crystal structure. *J. Am. Chem. Soc.* **134,** 7442 (2012).

42. Y. Kurashige, G. K.-L. Chan & T. Yanai. Entangled quantum electronic wavefunctions of the Mn_4CaO_5 cluster in photosystem II. *Nat. Chem.* **5,** 660 (2013).

43. M. A. Beckwith *et al.* How accurately can extended x-ray absorption spectra be predicted from first principles? Implications for modeling the oxygen-evolving complex in photosystem II. *J. Am. Chem. Soc.* **137,** 12815 (2015).

44. J. Kern *et al.* Taking snapshots of photosynthetic water oxidation using femtosecond x-ray diffraction and spectroscopy. *Nat. Commun.* **5,** 4371 (2014).

45. C. Kupitz *et al.* Serial time-resolved crystallography of photosystem II using a femtosecond x-ray laser. *Nature* **513,** 261 (2014).

46. M. Suga *et al.* Light-induced structural changes and the site of O=O bond formation in PSII caught by XFEL. *Nature* **543,** 131 (2017).

47. J. Kern *et al.* Structures of the intermediates of Kok's photosynthetic water oxidation clock. *Nature* **563,** 421 (2018).

48. M. Askerka *et al.* Analysis of the radiation-damage-free x-ray structure of photosystem II in light of EXAFS and QM/MM data. *Biochem.* **54,** 1713 (2015).

49. M. Amin *et al.* X-ray free electron laser radiation damage through the S-state cycle of the oxygen evolving complex of photosystem II. *J. Phys. Chem. B* **121,** 9382 (2017).

50. H. Dau, P. Liebisch & M. Haumann. X-ray absorption spectroscopy to analyze nuclear geometry and electronic structure of biological metal centers—potential and questions examined with special focus on the tetra-nuclear manganese complex of oxygenic photosynthesis. *Anal. Bioanal. Chem.* **376,** 562 (2003).

51. N. Cox *et al.* Electronic structure of the oxygen-evolving complex in photosystem II prior to O-O bond formation. *Science* **345,** 804 (2014).

52. R. J. Pace, L. Jin & R. Stranger. What spectroscopy reveals concerning the Mn oxidation levels in the oxygen evolving complex of photosystem II: x-ray to near infra-red. *Dalt. Trans.* **41,** 11145 (2012).

53. R. Terrett, S. Petrie, R. Stranger & R. J. Pace. What computational chemistry and magnetic resonance reveal concerning the oxygen evolving centre in photosystem II. *J. Inorg. Biochem.* **162,** 178 (2016).

54. H. Chen, D. A. Case & G. C. Dismukes. Reconciling structural and spectroscopic fingerprints of the oxygen-evolving complex of photosystem II: a computational study of the S_2 state. *J. Phys. Chem. B* **122** (2018).

55. A. R. Jaszewski, S. Petrie, R. J. Pace & R. Stranger. Toward the assignment of the manganese oxidation pattern in the water-oxidizing complex of photosystem II: a time-dependent DFT study of XANES energies. *Chem. Eur. J.* **17,** 5699 (2011).

56. H. Chen, G. C. Dismukes & D. A. Case. Resolving ambiguous protonation and oxidation states in the oxygen evolving complex of photosystem II. *J. Phys. Chem. B* **122,** 8654 (2018).

57. D. Bovi, D. Narzi & L. Guidoni. The S_2 state of the oxygen-evolving complex of photosystem II explored by QM/MM dynamics: spin surfaces and metastable states suggest a reaction path towards the S_3 state. *Angew. Chem. Int. Ed.* **52,** 11744 (2013).

58. H. Isobe *et al.* Theoretical illumination of water-inserted structures of the $CaMn_4O_5$ cluster in the S_2 and S_3 states of oxygen-evolving complex of photosystem II: full geometry optimizations by B3LYP hybrid density functional. *Dalt. Trans.* **41,** 13727 (2012).

59. P. E. M. Siegbahn. Water oxidation mechanism in photosystem II, including oxidations, proton release pathways, O–O bond formation and O_2 release. *BBA - Bioenerg.* **1827,** 1003 (2013).

60. D. A. Pantazis *et al.* Two interconvertible structures that explain the spectroscopic properties of the oxygen-evolving complex of photosystem II in the S_2 state. *Angew. Chem. Int. Ed.* **51,** 9935 (2012).

61. M. Retegan *et al.* A five-coordinate Mn(IV) intermediate in biological water oxidation: spectroscopic signature and a pivot mechanism for water binding. *Chem. Sci.* **7,** 72 (2016).

62. C. J. Kim & R. J. Debus. Evidence from FTIR difference spectroscopy that a substrate H_2O molecule for O_2 formation in photosystem II Is provided by the Ca ion of the catalytic Mn_4CaO_5 cluster. *Biochem.* **56,** 2558 (2017).

63. P. E. M. Siegbahn. The S_2 to S_3 transition for water oxidation in PSII (photosystem II), revisited. *Phys. Chem. Chem. Phys.* **20,** 22926 (2018).

64. M. Kusunoki. S_1-state Mn_4Ca complex of photosystem II exists in equilibrium between the two most-stable isomeric substates: XRD and EXAFS evidence. *J. Photochem. Photobiol. B* **104,** 100 (2011).

65. S. Petrie, P. Gatt, R. Stranger & R. J. Pace. The interaction of His337 with the Mn_4Ca cluster of photosystem II. *Phys. Chem. Chem. Phys.* **14,** 4651 (2012).

66. K. M. Davis & Y. N. Pushkar. Structure of the oxygen evolving complex of photosystem II at room temperature. *J. Phys. Chem. B* **119,** 3492 (2015).

67. I. Ugur, A. W. Rutherford & V. R. Kaila. Redox-coupled substrate water reorganization in the active site of photosystem II – the role of calcium in substrate water delivery. *Biochim. Biophys. Acta - Bioenerg.* **1857,** 740 (2016).

68. D. Narzi, M. Capone, D. Bovi & L. Guidoni. Evolution from S_3 to S_4 states of the oxygen-evolving complex in photosystem II monitored by quantum mechanics/molecular mechanics (QM/MM) dynamics. *Chem. - A Eur. J.* **24,** 10820 (2018).

69. M. Shoji *et al.* Elucidation of the entire Kok cycle for photosynthetic water oxidation by the large-scale quantum mechanics/molecular mechanics calculations: comparison with the experimental results by the recent serial femtosecond crystallography. *Chem. Phys. Lett.* **730,** 416 (2019).

70. P. E. M. Siegbahn. O–O bond formation in the S_4 state of the oxygen-evolving complex in photosystem II. *Chem. Eur. J.* **12,** 9217 (2006).

71. P. E. M. Siegbahn. Nucleophilic water attack is not a possible mechanism for O-O bond formation in photosystem II. *Proc. Natl. Acad. Sci. U.S.A.* **114,** 4966 (2017).

72. K. Kawashima *et al.* O2 evolution and recovery of the water-oxidizing enzyme. *Nat. Commun.* **9,** 1247 (2018).

73. M. Shoji *et al.* Nonadiabatic one-electron transfer mechanism for the O–O bond formation in the oxygen-evolving complex of photosystem II. *Chem. Phys. Lett.* **698,** 138 (2018).

74. M. Shoji *et al.* Large-scale QM/MM calculations of the CaMn4O5 cluster in the S3 state of the oxygen evolving complex of photosystem II. Comparison between water-inserted and no water-inserted structures. *Faraday Discuss.* **198,** 83 (2017).

75. N. Lin, C. A. Marianetti, A. J. Millis & D. R. Reichman. Dynamical mean-field theory for quantum chemistry. *Phys. Rev. Lett.* **106,** 96402 (2011).

76. D. Jacob, K. Haule & G. Kotliar. Dynamical mean-field theory for molecular electronics: electronic structure and transport properties. *Phys. Rev. B* **82,** 195115 (2010).

77. C. Weber *et al.* Vanadium dioxide: a Peierls-Mott insulator stable against disorder. *Phys. Rev. Lett.* **108,** 256402 (2012).

78. C. Weber *et al.* Importance of many-body effects in the kernel of hemoglobin for ligand binding. *Phys. Rev. Lett.* **110,** 106402 (2013).

79. C. Weber, D. J. Cole, D. D. O'Regan & M. C. Payne. Renormalization of myoglobin–ligand binding energetics by quantum many-body effects. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 5790 (2014).

80. D. Bourgeois *et al.* Complex landscape of protein structural dynamics unveiled by nanosecond Laue crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 8704 (2003).

81. K. A. Magnus *et al.* Crystallographic analysis of oxygenated and deoxygenated states of arthropod hemocyanin shows unusual differences. *Proteins Struct. Funct. Genet.* **19,** 302 (1994).

82. W. Kohn. Nobel lecture: electronic structure of matter-wave functions and density functionals. *Rev. Mod. Phys.* **71,** 1253 (1999).

83. P. A. M. Dirac. Quantum mechanics of many-electron systems. *Proc. R. Soc. A* **123,** 714 (1929).

84. M. Levy. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proc. Natl. Acad. Sci. U.S.A.* **76,** 6062 (1979).

85. D. M. Ceperley & B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.* **45,** 566 (1980).

86. I. Solt *et al.* Evaluating boundary dependent errors in QM/MM simulations. *J. Phys. Chem. B* **113,** 5728 (2009).

87. L. Bondesson, E. Rudberg, Y. Luo & P. Sałek. A linear scaling study of solvent-solute interaction energy of drug molecules in aqua solution. *J. Phys. Chem. B* **111,** 10320 (2007).

88. S. J. Fox *et al.* Electrostatic embedding in large-scale first principles quantum mechanical calculations on biomolecules. *J. Chem. Phys.* **135,** 224107 (2011).

89. K. Sadeghian *et al.* Ribose-protonated DNA base excision repair: a combined theoretical and experimental study. *Angew. Chem. Int. Ed.* **53,** 10044 (2014).

90. D. Flaig, M. Beer & C. Ochsenfeld. Convergence of electronic structure with the size of the QM region: example of QM/MM NMR shieldings. *J. Chem. Theory Comput.* **8,** 2260 (2012).

91. E. R. Johnson & G. A. DiLabio. Convergence of calculated nuclear magnetic resonance chemical shifts in a protein with respect to quantum mechanical model size. *J. Mol. Struct. THEOCHEM* **898,** 56 (2009).

92. C. M. Isborn *et al.* Electronic absorption spectra from MM and ab initio QM/MM molecular dynamics: environmental effects on the absorption spectrum of photoactive yellow protein. *J. Chem. Theory Comput.* **8,** 5092 (2012).

93. T. J. Zuehlsdorff *et al.* Solvent effects on electronic excitations of an organic chromophore. *J. Chem. Theory Comput.* **12,** 1853 (2016).

94. D. J. Cole *et al.* Toward ab initio optical spectroscopy of the Fenna–Matthews–Olson complex. *J. Phys. Chem. Lett.* **4,** 4206 (2013).

95. M. Retegan, F. Neese & D. A. Pantazis. Convergence of QM/MM and cluster models for the spectroscopic properties of the oxygen-evolving complex in photosystem II. *J. Chem. Theory Comput.* **9,** 3832 (2013).

96. H. J. Kulik, J. Zhang, J. P. Klinman & T. J. Martínez. How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol O -Methyltransferase. *J. Phys. Chem. B* **120,** 11381 (2016).

97. E. R. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *J. Comput. Phys.* **17,** 87 (1975).

98. C.-K. Skylaris *et al.* Nonorthogonal generalized Wannier function pseudopotential plane-wave method. *Phys. Rev. B* **66,** 035119 (2002).

99. C. Brouder *et al.* Exponential localization of Wannier functions in insulators. *Phys. Rev. Lett.* **98,** 046402 (2007).

100. A. A. Mostofi, P. D. Haynes, C.-K. Skylaris & M. C. Payne. Preconditioned iterative minimization for linear-scaling electronic structure calculations. *J. Chem. Phys.* **119,** 8842 (2003).

101. D. Baye & P.-H. Heenen. Generalised meshes for quantum mechanical problems. *J. Phys. A. Math. Gen.* **19,** 2041 (1986).

102. M. R. Hestenes & E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand. (U.S.)* **49,** 409 (1952).

103. P. D. Haynes, C.-K. Skylaris, A. A. Mostofi & M. C. Payne. ONETEP: linear-scaling density-functional theory with plane waves. *Psi-k Newsl.* **72,** 78 (2005).

104. X.-P. Li, R. W. Nunes & D. Vanderbilt. Density-matrix electronic-structure method with linear system-size scaling. *Phys. Rev. B* **47,** 10891 (1993).

105. P. D. Haynes & M. C. Payne. Corrected penalty-functional method for linear-scaling calculations within density-functional theory. *Phys. Rev. B* **59,** 12173 (1999).

106. P. D. Haynes, C.-K. Skylaris, A. A. Mostofi & M. C. Payne. Density kernel optimization in the ONETEP code. *J. Phys. Condens. Matter* **20,** 294207 (2008).

107. R. McWeeny. Some recent advances in density matrix theory. *Rev. Mod. Phys.* **32,** 335 (1960).

108. R. W. Nunes & D. Vanderbilt. Generalization of the density-matrix method to a nonorthogonal basis. *Phys. Rev. B* **50,** 17611 (1994).

109. J. Dziedzic *et al.* Minimal parameter implicit solvent model for ab initio electronic-structure calculations. *Eur. Lett.* **95,** 43001 (2011).

110. E. Rudberg. Difficulties in applying pure Kohn–Sham density functional theory electronic structure methods to protein molecules. *J. Phys. Condens. Matter* **24,** 072202 (2012).

111. G. Lever *et al.* Electrostatic considerations affecting the calculated HOMO–LUMO gap in protein molecules. *J. Phys. Condens. Matter* **25,** 152101 (2013).

112. J. Klimeš & A. Michaelides. Perspective: advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* **137,** 120901 (2012).

113. A. Tkatchenko, R. A. DiStasio, R. Car & M. Scheffler. Accurate and efficient method for many-body van der Waals interactions. *Phys. Rev. Lett.* **108,** 236402 (2012).

114. C. Cohen-Tannoudji, B. Diu & F. Laloë. *Quantum mechanics: volume two* (John Wiley & Sons, 1977).

115. M. Elstner *et al.* Hydrogen bonding and stacking interactions of nucleic acid base pairs: a density-functional-theory based treatment. *J. Chem. Phys.* **114,** 5149 (2001).

116. S. Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27,** 1787 (2006).

117. Q. Hill & C.-K. Skylaris. Including dispersion interactions in the ONETEP program for linear-scaling density functional theory calculations. *Proc. R. Soc. A* **465,** 669 (2009).

118. A. Tkatchenko & M. Scheffler. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.* **102,** 73005 (2009).

119. T. Bučko, S. Lebègue, J. Hafner & J. G. Ángyán. Improved density dependent correction for the description of London dispersion forces. *J. Chem. Theory Comput.* **9,** 4293 (2013).

120. T. C. Lillestolen & R. J. Wheatley. Atomic charge densities generated using an iterative stockholder procedure. *J. Chem. Phys.* **131,** 144101 (2009).

121. D. R. Hamann, M. Schlüter & C. Chiang. Norm-conserving pseudopotentials. *Phys. Rev. Lett.* **43,** 1494 (1979).

122. G. B. Bachelet, D. R. Hamann & M. Schlüter. Pseudopotentials that work: from H to Pu. *Phys. Rev. B* **26,** 4199 (1982).

123. P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B* **50,** 17953 (1994).

124. N. D. M. Hine. Linear-scaling density functional theory using the projector augmented wave method. *J. Phys. Condens. Matter* **29,** 024001 (2017).

125. *Opium - pseudopotential generation project.* `http://opium.sourceforge.net`.

126. N. A. W. Holzwarth, A. R. Tackett & G. E. Matthews. A projector augmented wave (PAW) code for electronic structure calculations, Part I: Atompaw for generating atom-centered functions. *Comput. Phys. Commun.* **135,** 329 (2001).

127. E. Linscott. *Strong correlation effects in the electronic structure of the photosystem II complex.* Master's thesis (University of Cambridge, 2015).

128. L. E. Ratcliff, N. D. M. Hine & P. D. Haynes. Calculating optical absorption spectra for large systems using linear-scaling density functional theory. *Phys. Rev. B* **84,** 165131 (2011).

129. M. R. Plan *et al.* Structural and biochemical characteristics of the cyclotide kalata B5 from Oldenlandia affinis. *Pept. Sci.* **94,** 647 (2010).

130. M. R. R. Plan, I. Saska, A. G. Cagauan & D. J. Craik. Backbone cyclised peptides from plants show molluscicidal activity against the rice pest Pomacea canaliculata (golden apple snail). *J. Agric. Food Chem.* **56,** 5237 (2008).

131. K. R. Gustafson *et al.* Circulins A and B. Novel human immunodeficiency virus (HIV)-inhibitory macrocyclic peptides from the tropical tree Chassalia parvifolia. *J. Am. Chem. Soc.* **116,** 9337 (1994).

132. N. L. Daly, K. R. Gustafson & D. J. Craik. The role of the cyclic peptide backbone in the anti-HIV activity of the cyclotide kalata B1. *FEBS Lett.* **574,** 69 (2004).

133. J. P. Tam, Y. A. Lu, J. L. Yang & K. W. Chiu. An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 8913 (1999).

134. E. Svangård *et al.* Cytotoxic cyclotides from *Viola tricolor*. *J. Nat. Prod.* **67,** 144 (2004).

135. H. J. Kulik, N. Luehr, I. S. Ufimtsev & T. J. Martinez. Ab initio quantum chemistry for protein structures. *J. Phys. Chem. B* **116,** 12501 (2012).

136. J. P. Perdew, K. Burke & M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77,** 3865 (1996).

137. F. Jollet, M. Torrent & N. Holzwarth. Generation of projector augmented-wave atomic data: a 71 element validated table in the XML format. *Comput. Phys. Commun.* **185,** 1246 (2014).

138. J.-L. Fattebert & F. Gygi. Density functional theory for efficient ab initio molecular dynamics simulations in solution. *J. Comp. Chem.* **23,** 662 (2002).

139. W. J. Hehre, R. F. Stewart & J. A. Pople. Self-consistent molecular-orbital methods. I. Use of Gaussian expansions of Slater-type atomic orbitals. *J. Chem. Phys.* **51,** 2657 (1969).

140. R. Ditchfield, W. J. Hehre & J. A. Pople. Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **54,** 724 (1971).

141. W. D. Cornell *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117,** 5179 (1995).

142. A.-C. Milazzo *et al.* Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy. *J. Struct. Biol.* **176,** 404 (2011).

143. X. Li *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Meth.* **10,** 584 (2013).

144. A. Amunts *et al.* The structure of the human mitochondrial ribosome. *Science* **348,** 95 (2015).

145. A. Bartesaghi *et al.* 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. *Science* **348,** 1147 (2015).

146. F. DiMaio *et al.* Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Meth.* **12,** 361 (2015).

147. R. M. Martin, L. Reining & D. M. Ceperley. *Interacting electrons* (Cambridge University Press, Cambridge, 2016).

148. F. Tran, P. Blaha, K. Schwarz & P. Novák. Hybrid exchange-correlation energy functionals for strongly correlated electrons: applications to transition-metal monoxides. *Phys. Rev. B* **74,** 155108 (2006).

149. C. Rödl, F. Fuchs, J. Furthmüller & F. Bechstedt. Quasiparticle band structures of the antiferromagnetic transition-metal oxides MnO, FeO, CoO, and NiO. *Phys. Rev. B* **79,** 235114 (2009).

150. V. I. Anisimov, F. Aryasetiawan & A. I. Lichtenstein. First-principles calculations of the electronic structure and spectra of strongly correlated systems: the LDA + U method. *J. Phys. Condens. Matter* **9,** 767 (1997).

151. J. D. Becker, J. M. Wills, L. Cox & B. R. Cooper. Electronic structure of Pu compounds with group-IIIB metals: two regimes of behavior. *Phys. Rev. B* **54,** R17265 (1996).

152. A. K. McMahan, K. Held & R. T. Scalettar. Thermodynamic and spectral properties of compressed Ce calculated using a combined local-density approximation and dynamical mean-field theory. *Phys. Rev. B* **67,** 075108 (2003).

153. A. I. Lichtenstein, M. I. Katsnelson & G. Kotliar. Finite-temperature magnetism of transition metals: an ab initio dynamical mean-field theory. *Phys. Rev. Lett.* **87,** 067205 (2001).

154. F. Lechermann, S. Biermann & A. Georges. Interorbital charge transfers and Fermi-surface deformations in strongly correlated metals: models, $BaVS_3$ and Na_xCoO_2. *Prog. Theor. Phys. Suppl.* **160,** 233 (2005).

155. A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98,** 5648 (1993).

156. P. J. Stephens, F. J. Devlin, C. F. Chabalowski & M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **98,** 11623 (1994).

157. A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38,** 3098 (1988).

158. C. Lee, W. Yang & R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37,** 785 (1988).

159. S. H. Vosko, L. Wilk & M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **58,** 1200 (1980).

160. V. I. Anisimov, J. Zaanen & O. K. Andersen. Band theory and Mott insulators: Hubbard U instead of Stoner I. *Phys. Rev. B* **44,** 943 (1991).

161. V. I. Anisimov *et al.* Density-functional theory and NiO photoemission spectra. *Phys. Rev. B* **48,** 16929 (1993).

162. S. L. Dudarev *et al.* Electron-energy-loss spectra and the structural stability of nickel oxide: an LSDA+U study. *Phys. Rev. B* **57,** 1505 (1998).

163. N. F. Mott. The basis of the electron theory of metals, with special reference to the transition metals. *Proc. Phys. Soc. Sect. A* **62,** 416 (1949).

164. N. W. Ashcroft & N. D. Mermin. *Solid State Physics* 33rd ed. (Holt, Rinehart and Winston, New York, 1976).

165. H. Yukawa. On the interaction of elementary particles. I. *Proc. Phys.-Math. Soc. Japan.* **17,** 48 (1935).

166. J. Lindhard. On the properties of a gas of charged particles. *Kgl. Danske Vidensk. Selsk. Mat.-fys. Medd.* **28,** 1 (1954).

167. E. M. Lifshitz & L. P. Pitaevskii. *Statistical physics: part 2* (Pergamon, Oxford, 1980).

168. E. H. Lieb & F. Y. Wu. Absence of Mott transition in an exact solution of the short-range, one-band model in one dimension. *Phys. Rev. Lett.* **20,** 1445 (1968).

169. H. Tasaki. The Hubbard model - an introduction and selected rigorous results. *J. Phys. Condens. Matter* **10,** 4353 (1998).

170. V. I. Anisimov & O. Gunnarsson. Density-functional calculation of effective Coulomb interactions in metals. *Phys. Rev. B* **43,** 7570 (1991).

171. W. E. Pickett, S. C. Erwin & E. C. Ethridge. Reformulation of the LDA + U method for a local-orbital basis. *Phys. Rev. B* **58,** 1201 (1998).

172. D. D. O'Regan, M. C. Payne & A. A. Mostofi. Subspace representations in ab initio methods for strongly correlated systems. *Phys. Rev. B* **83,** 245124 (2011).

173. D. D. O'Regan, N. D. M. Hine, M. C. Payne & A. A. Mostofi. Projector self-consistent DFT+U using nonorthogonal generalized Wannier functions. *Phys. Rev. B* **82,** 081102 (2010).

174. I. V. Solovyev, P. H. Dederichs & V. I. Anisimov. Corrected atomic limit in the local-density approximation and the electronic structure of d impurities in Rb. *Phys. Rev. B* **50,** 16861 (1994).

175. N. J. Mosey & E. A. Carter. Ab initio evaluation of Coulomb and exchange parameters for DFT + U calculations. *Phys. Rev. B* **76,** 155123 (2007).

176. B. Himmetoglu, R. M. Wentzcovitch & M. Cococcioni. First-principles study of electronic and structural properties of CuO. *Phys. Rev. B* **84,** 115108 (2011).

177. B. Himmetoglu, A. Floris, S. de Gironcoli & M. Cococcioni. Hubbard-corrected DFT energy functionals: the LDA+U description of correlated systems. *Int. J. Quant. Chem.* **114,** 14 (2014).

178. M. Cococcioni & S. de Gironcoli. Linear response approach to the calculation of the effective interaction parameters in the LDA + U method. *Phys. Rev. B* **71,** 035105 (2005).

179. H. J. Kulik, M. Cococcioni, D. A. Scherlis & N. Marzari. Density functional theory in transition-metal chemistry: a self-consistent Hubbard U approach. *Phys. Rev. Lett.* **97,** 103001 (2006).

180. A. J. Cohen, P. Mori-Sánchez & W. Yang. Insights into current limitations of density functional theory. *Science* **321,** 792 (2008).

181. A. J. Cohen, P. Mori-Sánchez & W. Yang. Challenges for density functional theory. *Chem. Rev.* **112,** 289 (2012).

182. J. P. Perdew, R. G. Parr, M. Levy & J. L. Balduz. Density-functional theory for fractional particle number: derivative discontinuities of the energy. *Phys. Rev. Lett.* **49,** 1691 (1982).

183. J. P. Perdew *et al.* Exchange and correlation in open systems of fluctuating electron number. *Phys. Rev. A* **76,** 40501 (2007).

184. P. Mori-Sánchez, A. J. Cohen & W. Yang. Localization and delocalization errors in density functional theory and implications for band-gap prediction. *Phys. Rev. Lett.* **100,** 146401 (2008).

185. B. Kaduk, T. Kowalczyk & T. Van Voorhis. Constrained density functional theory. *Chem. Rev.* **112,** 321 (2012).

186. Q. Zhao, E. I. Ioannidis & H. J. Kulik. Global and local curvature in density functional theory. *J. Chem. Phys.* **145,** 054109 (2016).

187. M. Lüders *et al.* Self-interaction correction in multiple scattering theory. *Phys. Rev. B* **71,** 205109 (2005).

188. M. Däne *et al.* Self-interaction correction in multiple scattering theory: application to transition metal oxides. *J. Phys. Condens. Matter* **21,** 045604 (2009).

189. I. D. Hughes *et al.* Lanthanide contraction and magnetism in the heavy rare earth elements. *Nature* **446,** 650 (2007).

190. M. R. Pederson, A. Ruzsinszky & J. P. Perdew. Communication: self-interaction correction with unitary invariance in density functional theory. *J. Chem. Phys.* **140,** 121103 (2014).

191. M. R. Pederson. Fermi orbital derivatives in self-interaction corrected density functional theory: applications to closed shell atoms. *J. Chem. Phys.* **142,** 064112 (2015).

192. T. Hahn, S. Liebing, J. Kortus & M. R. Pederson. Fermi orbital self-interaction corrected electronic structure of molecules beyond local density approximation. *J. Chem. Phys.* **143,** 224104 (2015).

193. I. Dabo *et al.* Koopmans' condition for density-functional theory. *Phys. Rev. B* **82,** 115121 (2010).

194. G. Borghi *et al.* Koopmans-compliant functionals and their performance against reference molecular data. *Phys. Rev. B* **90,** 075135 (2014).

195. D. D. O'Regan, N. D. M. Hine, M. C. Payne & A. A. Mostofi. Linear-scaling DFT+U with full local orbital optimization. *Phys. Rev. B* **85,** 085107 (2012).

196. C. Loschen, J. Carrasco, K. M. Neyman & F. Illas. First-principles LDA + U and GGA + U study of cerium oxides: dependence on the effective U parameter. *Phys. Rev. B* **75,** 035115 (2007).

197. C. W. M. Castleton, J. Kullgren & K. Hermansson. Tuning LDA+U for electron localization and structure at oxygen vacancies in ceria. *J. Chem. Phys.* **127,** 244704 (2007).

198. B. J. Morgan & G. W. Watson. A DFT+U description of oxygen vacancies at the TiO_2 rutile (110) surface. *Surf. Sci.* **601,** 5034 (2007).

199. C. Franchini *et al.* Ground-state properties of multivalent manganese oxides: density functional and hybrid density functional calculations. *Phys. Rev. B* **75,** 195128 (2007).

200. A. Rohrbach, J. Hafner & G. Kresse. Ab initio study of the (0001) surfaces of hematite and chromia: influence of strong electronic correlations. *Phys. Rev. B* **70,** 125426 (2004).

201. L. Wang, T. Maxisch & G. Ceder. Oxidation energies of transition metal oxides within the GGA + U framework. *Phys. Rev. B* **73,** 195107 (2006).

202. S. P. Ong *et al.* Voltage, stability and diffusion barrier differences between sodium-ion and lithium-ion intercalation materials. *Energy Environ. Sci.* **4,** 3680 (2011).

203. C. J. Fennie & K. M. Rabe. Magnetic and electric phase control in epitaxial $EuTiO_3$ from first principles. *Phys. Rev. Lett.* **97,** 267602 (2006).

204. E. Finazzi, C. Di Valentin, G. Pacchioni & A. Selloni. Excess electron states in reduced bulk anatase TiO_2: comparison of standard GGA, GGA+U, and hybrid DFT calculations. *J. Chem. Phys.* **129,** 154113 (2008).

205. F. Aryasetiawan, K. Karlsson, O. Jepsen & U. Schönberger. Calculations of Hubbard U from first-principles. *Phys. Rev. B* **74,** 125106 (2006).

206. D. D. O'Regan & G. Teobaldi. Optimization of constrained density functional theory. *Phys. Rev. B* **94,** 035159 (2016).

207. M. Cococcioni. *A LDA+U study of selected iron compounds.* PhD thesis (Scuola Internazionale Superiore di Studi Avanzati, 2002).

208. P. H. Dederichs, S. Blügel, R. Zeller & H. Akai. Ground states of constrained systems: application to cerium impurities. *Phys. Rev. Lett.* **53,** 2512 (1984).

209. V. Leiria Campo Jr & M. Cococcioni. Extended DFT + U + V method with on-site and inter-site electronic interactions. *J. Phys. Condens. Matter* **22,** 055602 (2010).

210. H. J. Kulik & N. Marzari. Transition-metal dioxides: a case for the intersite term in Hubbard-model functionals. *J. Chem. Phys.* **134** (2011).

211. L. A. Agapito, S. Curtarolo & M. Buongiorno Nardelli. Reformulation of DFT + U as a pseudohybrid Hubbard density functional for accelerated materials discovery. *Phys. Rev. X* **5,** 011006 (2015).

212. G. Moynihan, G. Teobaldi & D. D. O'Regan. A self-consistent ground-state formulation of the first-principles Hubbard U parameter validated on one-electron self-interaction error. arXiv: 1704.08076 (2017).

213. D. R. Bowler, T. Miyazaki & M. J. Gillan. Recent progress in linear scaling ab initio electronic structure techniques. *J. Phys. Condens. Matter* **14,** 2781 (2002).

214. M. J. Gillan, D. R. Bowler, A. S. Torralba & T. Miyazaki. Order-N first-principles calculations with the CONQUEST code. *Comput. Phys. Commun.* **177,** 14 (2007).

215. J. M. Soler *et al.* The SIESTA method for ab initio order-N materials simulation. *J. Phys. Condens. Matter* **14,** 2745 (2002).

216. E. Artacho *et al.* The SIESTA method; developments and applicability. *J. Phys. Condens. Matter* **20,** 064208 (2008).

217. L. Genovese *et al.* Daubechies wavelets for high performance electronic structure calculations: the BigDFT project. *Comptes Rendus Mécanique* **339,** 149 (2011).

218. T. Ozaki & H. Kino. Efficient projector expansion for the ab initio LCAO method. *Phys. Rev. B* **72,** 045121 (2005).

219. V. Weber, J. VandeVondele, J. Hutter & A. M. N. Niklasson. Direct energy functional minimization under orthogonality constraints. *J. Chem. Phys.* **128,** 084113 (2008).

220. S. J. Hu, S. S. Yan, M. W. Zhao & L. M. Mei. First-principles LDA + U calculations of the Co-doped ZnO magnetic semiconductor. *Phys. Rev. B* **73,** 245205 (2006).

221. H. J. Kulik & N. Marzari. Systematic study of first-row transition-metal diatomic molecules: a self-consistent DFT+U approach. *J. Chem. Phys.* **133,** 114103 (2010).

222. K. Yu & E. A. Carter. Communication: comparing ab initio methods of obtaining effective U parameters for closed-shell materials. *J. Chem. Phys.* **140,** 121105 (2014).

223. B. Huang. The screened pseudo-charge repulsive potential in perturbed orbitals for band calculations by DFT+U. *Phys. Chem. Chem. Phys.* **19,** 8008 (2017).

224. M. Shishkin & H. Sato. Self-consistent parametrization of DFT + U framework using linear response approach: application to evaluation of redox potentials of battery cathodes. *Phys. Rev. B* **93,** 085135 (2016).

225. H. J. Kulik & N. Marzari. A self-consistent Hubbard U density-functional theory approach to the addition-elimination reactions of hydrocarbons on bare FeO^+. *J. Chem. Phys.* **129,** 134314 (2008).

226. M. Shishkin & H. Sato. DFT+ U in Dudarev's formulation with corrected interactions between the electrons with opposite spins: the form of Hamiltonian, calculation of forces, and bandgap adjustments. *J. Chem. Phys.* **151,** 024102 (2019).

227. H. Chen & A. J. Millis. Spin-density functional theories and their + U and + J extensions: a comparative study of transition metals and transition metal oxides. *Phys. Rev. B* **93,** 045133 (2016).

228. J. N. Harvey. On the accuracy of density functional theory in transition metal chemistry. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.,* **102,** 203 (2006).

229. C. J. Cramer & D. G. Truhlar. Density functional theory for transition metals and transition metal chemistry. *Phys. Chem. Chem. Phys.* **11,** 10757 (2009).

230. H. S. Yu, S. L. Li & D. G. Truhlar. Perspective: Kohn-Sham density functional theory descending a staircase. *J. Chem. Phys.* **145,** 130901 (2016).

231. E. R. Johnson & A. D. Becke. Communication: DFT treatment of strong correlation in 3d transition-metal diatomics. *J. Chem. Phys.* **146,** 211105 (2017).

232. C. G. Shull, W. A. Strauser & E. O. Wollan. Neutron diffraction by paramagnetic and antiferromagnetic substances. *Phys. Rev.* **83,** 333 (1951).

233. J. S. Lim, D. Saldana-Greco & A. M. Rappe. Improved pseudopotential transferability for magnetic and electronic properties of binary manganese oxides from DFT + U + J calculations. *Phys. Rev. B* **94,** 165151 (2016).

234. N. D. M. Hine *et al.* Accurate ionic forces and geometry optimization in linear-scaling density-functional theory with local orbitals. *Phys. Rev. B* **83,** 195102 (2011).

235. Á. Ruiz-Serrano, N. D. M. Hine & C.-K. Skylaris. Pulay forces from localized orbitals optimized in situ using a psinc basis set. *J. Chem. Phys.* **136,** 234101 (2012).

236. J. H. Lloyd-Williams & B. Monserrat. Lattice dynamics and electron-phonon coupling calculations using nondiagonal supercells. *Phys. Rev. B* **92,** 184301 (2015).

237. R. W. G. Wyckoff. *Crystal Structures* 2nd ed. (Interscience Publishers, New York, 1963).

238. N. D. M. Hine, J. Dziedzic, P. D. Haynes & C.-K. Skylaris. Electrostatic interactions in finite systems treated with periodic boundary conditions: application to linear-scaling density functional theory. *J. Chem. Phys.* **135,** 204103 (2011).

239. *Perdew-Burke-Ernzerhof GGA pseudopotentials.* `http://www.sas.upenn.edu/rappegroup/research/pseudo-potential-gga.html`.

240. G. P. Kerker. Non-singular atomic pseudopotentials for solid state applications. *J. Phys. C Solid State Phys.* **13,** L189 (1980).

241. L. Kleinman & D. M. Bylander. Efficacious form for model pseudopotentials. *Phys. Rev. Lett.* **48,** 1425 (1982).

242. D. R. Hamann. Generalized norm-conserving pseudopotentials. *Phys. Rev. B* **40,** 2980 (1989).

243. A. M. Rappe, K. M. Rabe, E. Kaxiras & J. D. Joannopoulos. Optimized pseudopotentials. *Phys. Rev. B* **41,** 1227 (1990).

244. X. Gonze, R. Stumpf & M. Scheffler. Analysis of separable potentials. *Phys. Rev. B* **44,** 8503 (1991).

245. N. J. Ramer & A. M. Rappe. Designed nonlocal pseudopotentials for enhanced transferability. *Phys. Rev. B* **59,** 12471 (1999).

246. I. Grinberg, N. J. Ramer & A. M. Rappe. Transferable relativistic Dirac-Slater pseudopotentials. *Phys. Rev. B* **62,** 2311 (2000).

247. S. G. Louie, S. Froyen & M. L. Cohen. Nonlinear ionic pseudopotentials in spin-density-functional calculations. *Phys. Rev. B* **26,** 1738 (1982).

248. H. J. Kulik. Perspective: treating electron over-delocalization with the DFT+U method. *J. Chem. Phys.* **142,** 240901 (2015).

249. G. J. Martyna & M. E. Tuckerman. A reciprocal space based method for treating long range interactions in ab initio and force-field-based calculations in clusters. *J. Chem. Phys.* **110,** 2810 (1999).

250. O. F. Sankey & D. J. Niklewski. Ab initio multicenter tight-binding model for molecular-dynamics simulations and other applications in covalent systems. *Phys. Rev. B* **40,** 3979 (1989).

251. E. Artacho *et al.* Linear-scaling ab-initio calculations for large and complex systems. *Phys. Status Solidi* **215,** 809 (1999).

252. F. Aryasetiawan *et al.* Frequency-dependent local interactions and low-energy effective models from electronic structure calculations. *Phys. Rev. B* **70,** 195104 (2004).

253. H. Sakakibara *et al.* Model-mapped RPA for determining the effective Coulomb interaction. *J. Phys. Soc. Japan* **86,** 044714 (2017).

254. E. Şaşıoğlu *et al.* Wannier-function approach to spin excitations in solids. *Phys. Rev. B* **81,** 054434 (2010).

255. M. Marsman, J. Paier, A. Stroppa & G. Kresse. Hybrid functionals applied to extended systems. *J. Phys. Condens. Matter* **20,** 064201 (2008).

256. F. Tran & P. Blaha. Importance of the kinetic energy density for band gap calculations in solids with density functional theory. *J. Phys. Chem. A* **121,** 3318 (2017).

257. C. Franchini *et al.* Density functional theory study of MnO by a hybrid functional approach. *Phys. Rev. B* **72,** 045132 (2005).

258. P. Gopal *et al.* Improved electronic structure and magnetic exchange interactions in transition metal oxides. *J. Phys. Condens. Matter* **29,** 444003 (2017).

259. R. Gillen & J. Robertson. Accurate screened exchange band structures for the transition metal monoxides MnO, FeO, CoO and NiO. *J. Phys. Condens. Matter* **25,** 165502 (2013).

260. P. Verma & D. G. Truhlar. HLE16: a local Kohn–Sham gradient approximation with good performance for semiconductor band gaps and molecular excitation energies. *J. Phys. Chem. Lett.* **8,** 380 (2017).

261. K. Finzel & A. I. Baranov. A simple model for the Slater exchange potential and its performance for solids. *Int. J. Quantum Chem.* **117,** 40 (2017).

262. F. Tran & P. Blaha. Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. *Phys. Rev. Lett.* **102,** 226401 (2009).

263. X. Feng. Electronic structure of MnO and CoO from the B3LYP hybrid density functional method. *Phys. Rev. B* **69,** 155107 (2004).

264. E. Engel & R. N. Schmid. Insulating ground states of transition-metal monoxides from exact exchange. *Phys. Rev. Lett.* **103,** 036404 (2009).

265. A. Schrön, C. Rödl & F. Bechstedt. Energetic stability and magnetic properties of MnO in the rocksalt, wurtzite, and zinc-blende structures: influence of exchange and correlation. *Phys. Rev. B* **82,** 165109 (2010).

266. H. Jiang, R. I. Gomez-Abal, P. Rinke & M. Scheffler. First-principles modeling of localized d states with the GW@LDA+U approach. *Phys. Rev. B* **82,** 045108 (2010).

267. S. Kobayashi, Y. Nohara, S. Yamamoto & T. Fujiwara. GW approximation with LSDA + U method and applications to NiO, MnO, and V_2O_3. *Phys. Rev. B* **78,** 155112 (2008).

268. S. V. Faleev, M. van Schilfgaarde & T. Kotani. All-electron self-consistent GW approximation: application to Si, MnO, and NiO. *Phys. Rev. Lett.* **93,** 126406 (2004).

269. J. van Elp *et al.* Electronic structure of MnO. *Phys. Rev. B* **44,** 1530 (1991).

270. E. Z. Kurmaev *et al.* Oxygen x-ray emission and absorption spectra as a probe of the electronic structure of strongly correlated oxides. *Phys. Rev. B* **77,** 165127 (2008).

271. R. N. Iskenderov, I. A. Drabkin, L. T. Emel'yanova & Y. M. Ksendzov. Absorption spectrum of MnO single crystals. *Sov. Phys. – Solid State* **10,** 2031 (1969).

272. I. A. Drabkin, L. T. Emel'yanova, R. N. Iskenderov & Y. M. Ksendzov. Photoconductivity of single crystals of MnO. *Sov. Phys. – Solid State* **10,** 2428 (1969).

273. A. K. Cheetham & D. A. O. Hope. Magnetic ordering and exchange effects in the antiferromagnetic solid solutions $Mn_xNi_{1-x}O$. *Phys. Rev. B* **27,** 6964 (1983).

274. B. E. F. Fender, A. J. Jacobson & F. A. Wedgwood. Covalency parameters in MnO, α-MnS, and NiO. *J. Chem. Phys.* **48,** 990 (1968).

275. A. Fujimori *et al.* Electronic structure of MnO. *Phys. Rev. B* **42,** 7580 (1990).

276. R. Åkesson, L. G. M. Pettersson, M. Sandstroem & U. Wahlgren. Theoretical calculations of the Jahn-Teller effect in the hexahydrated Cu(II), Cr(II), and Mn(III) ions, hexaaquacopper(2+), hexaaquachromium(2+) and hexaaquamanganese(3+), and comparisons with the hexahydrated Cu(I), Cr(III), and Mn(II) clusters. *J. Phys. Chem.* **96,** 150 (1992).

277. B. Kallies & R. Meier. Electronic structure of 3*d* $[M(H_2O)_6]^{3+}$ ions from Sc^{III} to Fe^{III}: a quantum mechanical study based on DFT computations and natural bond orbital analyses. *Inorg. Chem.* **40,** 3101 (2001).

278. Y. Yang, M. A. Ratner & G. C. Schatz. Multireference ab initio study of ligand field *d*–*d* transitions in octahedral transition-metal oxide clusters. *J. Phys. Chem. C* **118,** 29196 (2014).

279. J. Li *et al.* Calculation of redox potentials and pK_a values of hydrated transition metal cations by a combined density functional and continuum dielectric theory. *Inorg. Chem.* **35,** 4694 (1996).

280. J. K. Beattie, S. P. Best, B. W. Skelton & A. H. White. Structural studies on the caesium alums, $CsM[SO_4]_2 \cdot 12H_2O$. *J. Chem. Soc. Dalt. Trans.* 2105 (1981).

281. H. J. Kulik & N. Marzari. Accurate potential energy surfaces with a DFT+U(R) approach. *J. Chem. Phys.* **135** (2011).

282. D. A. Scherlis, M. Cococcioni, P. Sit & N. Marzari. Simulation of heme using DFT + U: a step toward accurate spin-state energetics. *J. Phys. Chem. B* **111,** 7384 (2007).

283. I. A. Nekrasov, M. A. Korotin & V. I. Anisimov. Coulomb interaction in oxygen *p*-shell in LDA+U method and its influence on calculated spectral and magnetic properties of transition metal oxides. arXiv: `0009107 [cond-mat]` (2000).

284. C. Cao, S. Hill & H.-P. Cheng. Strongly correlated electrons in the $[Ni(hmp)(ROH)X]_4$ single molecule magnet: a DFT + U study. *Phys. Rev. Lett.* **100,** 167206 (2008).

285. B. Guillot. A reappraisal of what we have learnt during three decades of computer simulations on water. *J. Mol. Liq.* **101,** 219 (2002).

286. P. L. Silvestrelli & M. Parrinello. Structural, electronic, and bonding properties of liquid water from first principles. *J. Chem. Phys.* **111,** 3572 (1999).

287. C. K. Jørgensen. *Absorption spectra and chemical bonding in complexes.* (Pergamon Press, Oxford, 1962).

288. D. C. Harris & M. D. Bertolucci. *Symmetry and spectroscopy,* 336 (Oxford University Press, New York, 1978).

289. J. F. Janak. Proof that $\partial E/\partial n_i = \varepsilon$ in density-functional theory. *Phys. Rev. B* **18,** 7165 (1978).

290. T. Koopmans. Über die zuordnung von wellenfunktionen und eigenwerten zu den einzelnen elektronen eines atoms. *Physica* **1,** 104 (1934).

291. G. Moynihan, G. Teobaldi & D. D. O'Regan. Inapplicability of exact constraints and a minimal two-parameter generalization to the DFT+ U based correction of self-interaction error. *Phys. Rev. B* **94,** 220104 (2016).

292. A. Bajaj, J. P. Janet & H. J. Kulik. Communication: recovering the flat-plane condition in electronic structure theory at semi-local DFT cost. *J. Chem. Phys.* **147,** 191101 (2017).

293. M. Radoń, K. Gassowska, J. Szklarzewicz & E. Broclawik. Spin-state energetics of Fe(III) and Ru(III) aqua complexes: accurate ab initio calculations and evidence for huge solvation effects. *J. Chem. Theory Comput.* **12,** 1592 (2016).

294. D. J. Cole, D. D. O'Regan & M. C. Payne. Ligand discrimination in myoglobin from linear-scaling DFT+U. *J. Phys. Chem. Lett.* **3,** 1448 (2012).

295. S. Song *et al.* Benchmarks and reliable DFT results for spin gaps of small ligand Fe(II) complexes. *J. Chem. Theory Comput.* **14,** 2304 (2018).

296. Q. M. Phung, M. Feldt, J. N. Harvey & K. Pierloot. Toward highly accurate spin state energetics in first-row transition metal complexes: a combined CASPT2/CC approach. *J. Chem. Theory Comput.* **14,** 2446 (2018).

297. A. Bajaj, F. Liu & H. J. Kulik. Non-empirical, low-cost recovery of exact conditions with model-Hamiltonian inspired expressions in jmDFT. *J. Chem. Phys.* **150,** 154115 (2019).

298. A. Ferretti, I. Dabo, M. Cococcioni & N. Marzari. Bridging density-functional and many-body perturbation theory: orbital-density dependence in electronic-structure functionals. *Phys. Rev. B* **89,** 195134 (2014).

299. N. L. Nguyen, N. Colonna, A. Ferretti & N. Marzari. Koopmans-compliant spectral functionals for extended systems. *Phys. Rev. X* **8,** 021051 (2018).

300. M. R. Pederson, R. A. Heaton & C. C. Lin. Local-density Hartree–Fock theory of electronic states of molecules with self-interaction correction. *J. Chem. Phys.* **80,** 1972 (1984).

301. N. Colonna, N. L. Nguyen, A. Ferretti & N. Marzari. Koopmans-compliant functionals and potentials and their application to the GW100 test set. *J. Chem. Theory Comput.* **15,** 1905 (2019).

302. A. Georges. *Strongly correlated electron materials: dynamical mean-field theory and electronic structure. AIP Conf. Proc.* **715** (AIP, 2004).

303. W. Metzner & D. Vollhardt. Correlated lattice fermions in d = ∞ dimensions. *Phys. Rev. Lett.* **62,** 324 (1989).

304. A. Georges & G. Kotliar. Hubbard model in infinite dimensions. *Phys. Rev. B* **45,** 6479 (1992).

305. M. C. M. Wright. Green function or Green's function? *Nat. Phys.* **2,** 646 (2006).

306. A. Georges, G. Kotliar, W. Krauth & M. J. Rozenberg. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.* **68,** 13 (1996).

307. K. Haule, C.-H. Yee & K. Kim. Dynamical mean-field theory within the full-potential methods: electronic structure of $CeIrIn_5$, $CeCoIn_5$, and $CeRhIn_5$. *Phys. Rev. B* **81,** 195107 (2010).

308. *DFT + Embedded DMFT Functional.* `http://hauleweb.rutgers.edu/tutorials/index.html`.

309. M. Aichhorn *et al.* TRIQS/DFTTools: a TRIQS application for ab initio calculations of correlated materials. *Comput. Phys. Commun.* **204,** 200 (2016).

310. *WIEN 2k.* `http://susi.theochem.tuwien.ac.at/`.

311. G. Kresse & J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6,** 15 (1996).

312. G. Kresse & J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54,** 11169 (1996).

313. *The VASP site.* `https://www.vasp.at/`.

314. *DCore — DCore 1.0.0 documentation.* https://issp-center-dev.github.io/DCore/index.html.

315. P. Giannozzi *et al.* QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21,** 395502 (2009).

316. *OpenMX website.* http://www.openmx-square.org/.

317. E. Plekhanov *et al.* Many-body renormalization of forces in f-electron materials. *Phys. Rev. B* **98,** 075129 (2018).

318. S. J. Clark *et al.* First principles methods using CASTEP. *Z. Kristallogr. Cryst. Mater.* **220,** 567 (2005).

319. *CASTEP website.* http://www.castep.org/.

320. *Amulet website.* http://amulet-code.org/.

321. *The Elk Code.* http://elk.sourceforge.net/.

322. S. Choi *et al.* ComDMFT: a massively parallel computer package for the electronic structure of correlated-electron systems. arXiv: 1810.01679 (2018).

323. *FlapwMBPT.* https://www.bnl.gov/cmpmsd/flapwmbpt/.

324. A. L. Kutepov. Electronic structure of Na, K, Si, and LiF from self-consistent solution of Hedin's equations including vertex corrections. *Phys. Rev. B* **94,** 155101 (2016).

325. O. Parcollet *et al.* TRIQS: a toolbox for research on interacting quantum systems. *Comput. Phys. Commun.* **196,** 398 (2015).

326. A. Gaenko *et al.* Updated core libraries of the ALPS project. *Comput. Phys. Commun.* **213,** 235 (2017).

327. L. Huang *et al.* iQIST: an open source continuous-time quantum Monte Carlo impurity solver toolkit. *Comput. Phys. Commun.* **195,** 140 (2015).

328. M. Wallerberger *et al.* w2dynamics: local one- and two-particle quantities from dynamical mean field theory. *Comput. Phys. Commun.* **235,** 388 (2019).

329. J. C. Slater. The ferromagnetism of nickel. *Phys. Rev.* **49,** 537 (1936).

330. J. Kanamori. Superexchange interaction and symmetry properties of electron orbitals. *J. Phys. Chem. Solids* **10,** 87 (1959).

331. M. Imada, A. Fujimori & Y. Tokura. Metal-insulator transitions. *Rev. Mod. Phys.* **70,** 1039 (1998).

332. C. Weber, A. Amaricci, M. Capone & P. B. Littlewood. Augmented hybrid exact diagonalization solver for dynamical mean field theory. *Phys. Rev. B* **86,** 115136 (2012).

333. P. Pulay. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett.* **73,** 393 (1980).

334. P. Pulay. Improved SCF convergence acceleration. *J. Comput. Chem.* **3,** 556 (1982).

335. L. V. Pourovskii, B. Amadon, S. Biermann & A. Georges. Self-consistency over the charge density in dynamical mean-field theory: a linear muffin-tin implementation and some physical implications. *Phys. Rev. B* **76,** 235101 (2007).

336. H. Park, A. J. Millis & C. A. Marianetti. Computing total energies in complex materials using charge self-consistent DFT + DMFT. *Phys. Rev. B* **90,** 235103 (2014).

337. S. Bhandary, E. Assmann, M. Aichhorn & K. Held. Charge self-consistency in density functional theory combined with dynamical mean field theory: k-space reoccupation and orbital order. *Phys. Rev. B* **94,** 155131 (2016).

338. K. A. Wilkinson, N. D. M. Hine & C.-K. Skylaris. Hybrid MPI-OpenMP parallelism in the ONETEP linear-scaling electronic structure code: application to the delamination of cellulose nanofibrils. *J. Chem. Theory Comput.* **10,** 4782 (2014).

339. K. Michał & A. Dijkstra. Personal communication (2016).

340. S. Obara & H. Kashiwagi. Ab initio MO studies of electronic states and Mössbauer spectra of high-, intermediate-, and low-spin Fe(II)-porphyrin complexes. *J. Chem. Phys.* **77,** 3155 (1982).

341. Y.-K. Choe, T. Hashimoto, H. Nakano & K. Hirao. Theoretical study of the electronic ground state of iron(II) porphine. *Chem. Phys. Lett.* **295,** 380 (1998).

342. Y.-K. Choe, T. Nakajima, K. Hirao & R. Lindh. Theoretical study of the electronic ground state of iron(II) porphine. II. *J. Chem. Phys.* **111,** 3837 (1999).

343. K. Pierloot. The CASPT2 method in inorganic electronic spectroscopy: from ionic transition metal to covalent actinide complexes. *Mol. Phys.* **101,** 2083 (2003).

344. A. R. Groenhof, M. Swart, A. W. Ehlers & K. Lammertsma. Electronic ground states of iron porphyrin and of the first species in the catalytic reaction cycle of cytochrome P450s. *J. Phys. Chem. A* **109,** 3411 (2005).

345. G. Li Manni, S. D. Smart & A. Alavi. Combining the complete active space self-consistent field method and the full configuration interaction quantum Monte Carlo within a super-CI Framework, with application to challenging metal-porphyrins. *J. Chem. Theory Comput.* **12,** 1245 (2016).

346. F. Neese. Prediction of molecular properties and molecular spectroscopy with density functional theory: from fundamental theory to exchange-coupling. *Coord. Chem. Rev.* **253,** 526 (2009).

347. S. Franzen, L. Kiger, C. Poyart & J.-L. Martin. Heme photolysis occurs by ultrafast excited state metal-to-ring charge transfer. *Biophys. J.* **80,** 2372 (2001).

348. A. Dreuw, B. D. Dunietz & M. Head-Gordon. Characterization of the relevant excited states in the photodissociation of CO-ligated hemoglobin and myoglobin. *J. Am. Chem. Soc.* **124,** 12070 (2002).

349. B. D. Dunietz, A. Dreuw & M. Head-Gordon. Initial steps of the photodissociation of the CO ligated heme group. *J. Phys. Chem. B* **107,** 5623 (2003).

350. A. J. Millis. *Strong interactions in low dimensions* (eds D. Baeriswyl & L. Degiorgi) 195 (Springer Netherlands, Dordrecht, 2004).

351. W. A. Eaton *et al.* Optical spectra of oxy- and deoxyhemoglobin. *J. Am. Chem. Soc.* **100,** 4991 (1978).

352. B. Hazes *et al.* Crystal structure of deoxygenated limulus polyphemus subunit II hemocyanin at 2.18 Å resolution: clues for a mechanism for allosteric regulation. *Protein Sci.* **2,** 597 (2008).

353. F. Ghiretti. The decomposition of hydrogen peroxide by hemocyanin and by its dissociation products. *Arch. Biochem. Biophys.* **63,** 165 (1956).

354. J. A. Halfen *et al.* Reversible cleavage and formation of the dioxygen O-O bond within a dicopper complex. *Science* **271,** 1397 (1996).

355. H. W. Duckworth & J. E. Coleman. Physicochemical and kinetic properties of mushroom tyrosinase. *J. Biol. Chem.* **245,** 1613 (1970).

356. T. J. P. McGivern, S. Afsharpour & C. J. Marmion. Copper complexes as artificial DNA metallonucleases: from Sigman's reagent to next generation anti-cancer agent? *Inorganica Chim. Acta* **472,** 12 (2018).

357. C. J. Nunes *et al.* Reactivity of dinuclear copper(II) complexes towards melanoma cells: correlation with its stability, tyrosinase mimicking and nuclease activity. *J. Inorg. Biochem.* **149,** 49 (2015).

358. B. F. Gherman & C. J. Cramer. Quantum chemical studies of molecules incorporating a $Cu_2O_2^{2+}$ core. *Coord. Chem. Rev.* **253,** 723 (2009).

359. D. M. Dooley *et al.* Magnetic susceptibility studies of laccase and oxyhemocyanin (copper proteins/variable temperature measurements/antiferromagnetism). *Proc. Natl. Acad. Sci. U.S.A.* **75,** 3019 (1978).

360. E. I. Solomon *et al.* Copper dioxygen (bio)inorganic chemistry. *Faraday Discuss.* **148,** 11 (2011).

361. Y. Takano *et al.* Theoretical studies on the magnetic interaction and reversible dioxygen binding of the active site in hemocyanin. *Chem. Phys. Lett.* **335,** 395 (2001).

362. M. Metz & E. I. Solomon. Dioxygen binding to deoxyhemocyanin: electronic structure and mechanism of the spin-forbidden two-electron reduction of O_2. *J. Am. Chem. Soc.* **123,** 4938 (2001).

363. T. Saito & W. Thiel. Quantum mechanics/molecular mechanics study of oxygen binding in hemocyanin. *J. Phys. Chem. B* **118,** 5034 (2014).

364. B. Loeb, I. Crivelli & C. Andrade. Oxy-hemocyanin: a peroxo copper (II) complex? A mixed-valence alternative view. *Comments Inorg. Chem.* **20,** 1 (1998).

365. J. Li, L. Noodleman & D. A. Case. *Inorganic electronic structure and spectroscopy, volume 1* 6610724 (John Wiley & Sons, New York, 1999).

366. P. E. M. Siegbahn. The performance of hybrid DFT for mechanisms involving transition metal complexes in enzymes. *J. Biol. Inorg. Chem.* **11,** 695 (2006).

367. C. J. Cramer *et al.* Theoretical characterization of end-on and side-on peroxide coordination in ligated Cu_2O_2 models. *J. Phys. Chem. A* **110,** 11557 (2006).

368. C. J. Cramer *et al.* Theoretical models on the Cu_2O_2 torture track: mechanistic implications for oxytyrosinase and small-molecule analogues. *J. Phys. Chem. A* **110,** 1991 (2006).

369. M. Roemelt, S. Guo & G. K-L Chan. A projected approximation to strongly contracted N-electron valence perturbation theory for DMRG wavefunctions. *J. Chem. Phys.* **144,** 204113 (2016).

370. M. Flock & K. Pierloot. Theoretical study of the interconversion of O2-binding dicopper complexes. *J. Phys. Chem. A* **103,** 95 (1999).

371. M. F. Rode & H.-J. Werner. Ab initio study of the O_2 binding in dicopper complexes. *Theor. Chem. Acc.* **114,** 309 (2005).

372. Å. Malmqvist *et al.* Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **128,** 5648 (2008).

373. Y. Kurashige, J. Chalupský, T. N. Lan & T. Yanai. Complete active space second-order perturbation theory with cumulant approximation for extended active-space wavefunction from density matrix renormalization group. *J. Chem. Phys.* **141,** 174111 (2014).

374. Y. Kurashige & T. Yanai. High-performance ab initio density matrix renormalization group method: applicability to large-scale multireference problems for metal compounds. *J. Chem. Phys.* **130,** 234114 (2009).

375. T. Yanai, Y. Kurashige, E. Neuscamman & G. K.-L. Chan. Multireference quantum chemistry through a joint density matrix renormalization group and canonical transformation theory. *J. Chem. Phys.* **132,** 24105 (2010).

376. P.-Å. Malmqvist & B. O. Roos. The CASSCF state interaction method. *Chem. Phys. Lett.* **155,** 189 (1989).

377. B. O. Roos. *Advances in chemical physics: ab initio methods in quantum chemistry part 2, volume 69* (ed K. Lawley) 399 (John Wiley & Sons, Ltd, Chichester, 1987).

378. B. O. Roos, P. R. Taylor & P. E. Siegbahn. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chem. Phys.* **48,** 157 (1980).

379. D. G. Liakos & F. Neese. Interplay of correlation and relativistic effects in correlated calculations on transition-metal complexes: the $(Cu_2O_2)^{2+}$ core revisited. *J. Chem. Theory Comput.* **7,** 1511 (2011).

380. M. Aichhorn, M. Daghofer, H. G. Evertz & W. von der Linden. Low-temperature Lanczos method for strongly correlated systems. *Phys. Rev. B* **67,** 161103 (2003).

381. T. Nakamura & H. Mason. An electron spin resonance study of copper valence in oxyhemocyanin. *Biochem. Biophys. Res. Commun.* **3,** 297 (1960).

382. Y. Matoba *et al.* Crystallographic evidence that the dinuclear copper center of tyrosinase is flexible during catalysis. *J. Biol. Chem.* **281,** 8981 (2006).

383. S. V. Didziulis, S. L. Cohen, A. A. Gewirth & E. I. Solomon. Variable photon energy photoelectron spectroscopic studies of copper chlorides: an experimental probe of metal-ligand bonding and changes in electronic structure on ionization. *J. Am. Chem. Soc.* **110,** 250 (1988).

384. A. Lichtenstein *et al. Correlated electrons: from models to materials* (eds E. Pavarini, E. Koch, F. Anders & M. Jarrel) (Forschungszentrum Jülich GmbH, Jülich, 2012).

385. P. W. Anderson. Theory of magnetic exchange interactions: exchange in insulators and semiconductors. *Solid State Phys.* **14,** 99 (1963).

386. A. E. Reed, L. A. Curtiss & F. Weinhold. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **88,** 899 (1988).

387. E. D. Glendening *et al. NBO 5.9 and the NBO 5.9 Manual.* `http://www.chem.wisc.edu/~nbo5`. Madison, 2011.

388. L. P. Lee, D. J. Cole, M. C. Payne & C.-K. Skylaris. Natural bond orbital analysis in the ONETEP code: applications to large protein systems. *J. Comp. Chem.* **34,** 429 (2013).

389. M. J. Baldwin *et al.* Spectroscopic studies of side-on peroxide-bridged binuclear copper(II) model complexes of relevance to oxyhemocyanin and oxytyrosinase. *J. Am. Chem. Soc.* **114,** 10421 (1992).

390. N. H. Andersen *et al.* Raman, UV–vis, and CD spectroscopic studies of dodecameric oxyhemocyanin from Carcinus aestuarii. *Chem. Lett.* **40,** 1360 (2011).

391. R. S. Himmelwright, N. C. Eickman, C. D. LuBien & E. I. Solomon. Chemical and spectroscopic comparison of the binuclear copper active site of mollusc and arthropod hemocyanins. *J. Am. Chem. Soc.* **102,** 5378 (1980).

392. K. Heirwegh, H. Borginon & R. Lontie. Separation and absorption spectra of α-and β-haemocyanin of Helix pomatia. *Biochim. Biophys. Acta* **48,** 517 (1961).

393. P. Sun & G. Kotliar. Extended dynamical mean-field theory and GW method. *Phys. Rev. B* **66,** 085120 (2002).

394. S. Biermann, F. Aryasetiawan & A. Georges. First-principles approach to the electronic structure of strongly correlated systems: combining the GW approximation and dynamical mean-field theory. *Phys. Rev. Lett.* **90,** 086402 (2003).

395. P. Werner, R. Sakuma, F. Nilsson & F. Aryasetiawan. Dynamical screening in La_2CuO_4. *Phys. Rev. B* **91,** 125142 (2015).

396. M. J. Frisch *et al. Gaussian 09*. Wallingford, 2013.

397. E. I. Solomon *et al.* Copper active sites in biology. *Chem. Rev.* **114,** 3659 (2014).

398. J. M. Brown *et al.* Structural studies of the hemocyanin active site. 1. Extended x-ray absorption fine structure (EXAFS) analysis. *J. Am. Chem. Soc.* **102,** 4210 (1980).

399. G. L. Woolery *et al.* EXAFS studies of binuclear copper site of oxy-, deoxy-, and metaquo-, metfluoro-, and metazidohemocyanin from arthropods and molluscs. *J. Am. Chem. Soc.* **106,** 86 (1984).

400. M. E. Cuff, K. I. Miller, K. E. van Holde & W. A. Hendrickson. Crystal structure of a functional unit from Octopus hemocyanin. *J. Mol. Biol.* **278,** 855 (1998).

401. C. Gatsogiannis & J. Markl. Keyhole limpet hemocyanin: 9-Å cryoEM structure and molecular model of the KLH1 didecamer reveal the interfaces and intricate topology of the 160 functional units. *J. Mol. Biol.* **385,** 963 (2009).

402. Z. Gai *et al.* Crystal structure of the 3.8-MDa respiratory supermolecule hemocyanin at 3.0 Å resolution. *Structure* **23,** 2204 (2015).

403. R. Pal *et al.* S_0-state model of the oxygen-evolving complex of photosystem II. *Biochem.* **52,** 7703 (2013).

404. *Maestro version 10.2*. Schrödinger, LLC, New York, 2015.

405. W. L. Jorgensen & J. Tirado-Rives. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110,** 1657 (1988).

406. W. L. Jorgensen, D. S. Maxwell & J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118,** 11225 (1996).

407. C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski & J. H. Jensen. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK_a values. *J. Chem. Theory Comput.* **7,** 2284 (2011).

408. M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski & J. H. Jensen. PROPKA3: consistent treatment of internal and surface Residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7,** 525 (2011).

409. J. Yano *et al.* Where water is oxidized to dioxygen: Structure of the photosynthetic Mn_4Ca cluster. *Science* **314,** 821 (2006).

410. R. S. Mulliken. Electronic population analysis on LCAO–MO molecular wave functions. I. *J. Chem. Phys.* **23,** 1833 (1955).

411. T. Yamauchi *et al.* Parallel polarization electron paramagnetic resonance studies of the S1-state manganese cluster in the photosynthetic oxygen-evolving system. *Biochem.* **36,** 7520 (1997).

412. E. Gull *et al.* Continuous-time Monte Carlo methods for quantum impurity models. *Rev. Mod. Phys.* **83,** 349 (2011).

413. L. Jin *et al.* Electronic structure of the oxygen evolving complex in photosystem II, as revealed by ^{55}Mn Davies ENDOR studies at 2.5 K. *Phys. Chem. Chem. Phys.* **16,** 7799 (2014).

414. D. Kubota, S. Sakai & M. Imada. Real-space renormalized dynamical mean field theory. *Phys. Rev. B* **93,** 205119 (2016).

415. J. Morton *et al.* Structured near-infrared magnetic circular dichroism spectra of the Mn4CaO5 cluster of PSII in T. vulcanus are dominated by Mn(IV) d-d 'spin-flip' transitions. *Biochim. Biophys. Acta - Bioenerg.* **1859,** 88 (2018).

416. J. A. Santana, J. Kim, P. R. C. Kent & F. A. Reboredo. Successes and failures of Hubbard-corrected density functional theory: the case of Mg doped $LiCoO_2$. *J. Chem. Phys.* **141,** 164706 (2014).

417. L. Wang, T. Maxisch & G. Ceder. A first-principles approach to studying the thermal stability of oxide cathode materials. *Chem. Mater.* **19,** 543 (2007).

418. D. Kramer & G. Ceder. Tailoring the morphology of $LiCoO_2$: A first principles study. *Chem. Mater.* **21,** 3799 (2009).

419. Y. Koyama *et al.* Defect chemistry in layered $LiMO_2$ (M = Co, Ni, Mn, and $Li_{1/3}Mn_{2/3}$) by first-principles calculations. *Chem. Mater.* **24,** 3886 (2012).

420. G. W. Mann *et al.* First-principles Hubbard U approach for small molecule binding in metal-organic frameworks. *J. Chem. Phys.* **144,** 174104 (2016).

421. M. Cococcioni & N. Marzari. Energetics and cathode voltages of $LiMPO_4$ olivines (M = Fe, Mn) from extended Hubbard functionals. *Phys. Rev. Mater.* **3,** 033801 (2019).

422. A. Jain *et al.* Formation enthalpies by mixing GGA and GGA+U calculations. *Phys. Rev. B* **84,** 045115 (2011).

423. M. Aykol & C. Wolverton. Local environment dependent GGA + U method for accurate thermochemistry of transition metal compounds. *Phys. Rev. B* **90,** 115105 (2014).

424. C. J. Pickard & R. J. Needs. Ab initio random structure searching. *J. Phys. Condens. Matter* **23,** 053201 (2011).

425. M. A. Hough & C. R. Andrew. Cytochromes c′: structure, reactivity and relevance to haem-based gas sensing. *Adv. Microb. Physiol.* **67,** 1 (2015).

426. A. Manole *et al.* Conformational control of the binding of diatomic gases to cytochrome c′. *J. Biol. Inorg. Chem.* **20,** 675 (2015).

427. E. Runge & E. K. U. Gross. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* **52,** 997 (1984).

428. K. L. Liu & S. H. Vosko. A time-dependent spin density functional theory for the dynamical spin susceptibility. *Can. J. Phys.* **67,** 1015 (1989).

429. A. Dolfen. *Strong electronic correlations in low-dimensional systems.* PhD thesis (RWTH Aachen University, 2010).

430. M. Jarrell. Hubbard model in infinite dimensions: a quantum Monte Carlo study. *Phys. Rev. Lett.* **69,** 168 (1992).

431. K. Haule. Exact double counting in combining the dynamical mean field theory and the density functional theory. *Phys. Rev. Lett.* **115,** 196403 (2015).

432. O. Kristanovski, A. B. Shick, F. Lechermann & A. I. Lichtenstein. Role of nonspherical double counting in DFT+DMFT: total energy and structural optimization of pnictide superconductors. *Phys. Rev. B* **97,** 201116 (2018).

433. V. Artero *et al.* From enzyme maturation to synthetic chemistry: the case of hydrogenases. *Acc. Chem. Res.* **48,** 2380 (2015).

434. G. Henkel & B. Krebs. Metallothioneins: zinc, cadmium, mercury, and copper thiolates and selenolates mimicking protein active site features – structural aspects and biological implications. *Chem. Rev.* **104,** 801 (2004).

435. M. Potthoff. *DMFT: from infinite dimensions to real materials* (eds E. Pavarini, E. Koch, A. Lichtenstein & D. Vollhardt) chap. 5 (Forschungszentrum Jülich GmbH, Jülich, 2018).

436. M. Capone, D. Bovi, D. Narzi & L. Guidoni. Reorganization of substrate waters between the closed and open cubane conformers during the S_2 to S_3 transition in the oxygen evolving complex. *Biochem.* **54,** 6439 (2015).

437. A. Altland & B. Simons. *Condensed matter field theory* 2nd ed. (Cambridge University Press, Cambridge, 2010).

438. R. G. Parr & W. Yang. *Density-functional theory of atoms and molecules* (Oxford University Press, Oxford, 1989).

439. A. G. Petukhov, I. I. Mazin, L. Chioncel & A. I. Lichtenstein. Correlated metals and the LDA+U method. *Phys. Rev. B* **67,** 153106 (2003).

440. E. K. U. Gross & W. Kohn. Local density-functional theory of frequency-dependent linear response. *Phys. Rev. Lett.* **55,** 2850 (1985).

441. S. Baroni, S. de Gironcoli, A. Dal Corso & P. Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **73,** 515 (2001).

442. M. Marques & E. Gross. Time-dependent density functional theory. *Annu. Rev. Phys. Chem.* **55,** 427 (2004).

443. E. Koch. *DMFT: from infinite dimensions to real materials* (eds E. Pavarini, E. Koch, A. Lichtenstein & D. Vollhardt) chap. 11 (Forschungszentrum Jülich GmbH, Jülich, 2018).

444. V. M. Galitskii & A. B. Migdal. Application of quantum field theory methods to the many body problem. *J. Exptl. Theor. Phys.* **34,** 139 (1958).

445. W. H. Press, S. A. Teukolsky, W. T. Vetterling & B. P. Flannery. Numerical recipes in C (1992).

446. A. Dolfen. *Massively parallel exact diagonalization of strongly correlated systems.* Master's thesis (RWTH Aachen University, 2006).

Appendices

A.1 The Hubbard model

This appendix presents the derivation of the Hubbard model.

To develop a Hamiltonian suitable for explaining the Mott transition, consider a one-dimensional lattice of identical atoms. The Hamiltonian of an electron gas is given by

$$\hat{H} = \hat{H}^0 + \frac{1}{2}\int \mathrm{d}x \int \mathrm{d}x' \sum_{\sigma\sigma'} \hat{c}^\dagger_\sigma(x)\hat{c}^\dagger_{\sigma'}(x')\frac{1}{|x-x'|}\hat{c}_{\sigma'}(x')\hat{c}_\sigma(x), \tag{A.1}$$

where $\hat{H}^0$ is the single-particle Hamiltonian

$$\hat{H}^0 = \int \mathrm{d}x \sum_\sigma \hat{c}^\dagger_\sigma(x)\left(-\frac{\nabla^2}{2} + V_{ext}(x)\right)\hat{c}_\sigma(x), \tag{A.2}$$

and $\hat{c}_\sigma(x)$ is the field operator corresponding to the removal of an electron of spin σ located at x.

Given a periodic external potential, the eigenstates of this system will be Bloch states $\varphi_{ks}(x)$ that are extended across the system (k being the state's momentum and $s = 0, 1, ...$ the band or orbital index). For our purposes it is more instructive to consider the system via a basis of localised functions

$$|\psi_{ns}\rangle = \frac{1}{\sqrt{N}} \sum_{k\in\left[-\frac{\pi}{a},\frac{\pi}{a}\right]} e^{ikna}|\varphi_{ks}\rangle, \tag{A.3}$$

where the sum runs over those k in the first Brillouin zone. $|\psi_{ns}\rangle$ are known as *Wannier functions*. If the lattice is widely spaced, these Wannier functions will vary little from the s^{th} bound state of the isolated atom for the low-energy orbitals.

Restricting ourselves to the $s = 0$ band, the operator $\hat{c}^\dagger_{n\sigma} = \int_0^L \mathrm{d}x\ \langle x|\psi_n\rangle\hat{c}^\dagger_\sigma(x)$ can be shown to be the creation operator for an electron at site $|\psi_n\rangle$ with spin σ. Using these new field operators, the Hamiltonian of Equation A.1 can be reformulated as

$$\hat{H} = -\sum_{mn}\sum_\sigma t_{mn}\hat{c}^\dagger_{m\sigma}\hat{c}_{n\sigma} + \sum_{mnm'n'}\sum_{\sigma\sigma'} U_{mnm'n'}\hat{c}^\dagger_{m\sigma}\hat{c}^\dagger_{n\sigma'}\hat{c}_{m'\sigma'}\hat{c}_{n'\sigma}, \tag{A.4}$$

where t_{mn} are the single-particle matrix elements

$$t_{mn} = -\langle\psi_m|\hat{H}^0|\psi_n\rangle = \frac{1}{N}\sum_k e^{i(n-m)ka}\varepsilon_k \tag{A.5}$$

and the interaction matrix elements are given by

$$U_{mnm'n'} = \frac{1}{2}\int_0^L \mathrm{d}x \int_0^L \mathrm{d}x' \psi_m^*(x)\psi_n^*(x')\frac{1}{|x-x'|}\psi_{m'}(x')\psi_{n'}(x). \tag{A.6}$$

At this point, a number of approximations are made which are suitable for when the atoms are sufficiently separated and the overlap between neighbouring orbitals is weak:

(a) all but density-density fluctuation interactions between electrons on the same site are ignored;

(b) interactions between sites that are not nearest neighbours are ignored;

(c) the strength of the terms are assumed to be the same from site to site.

This reduces Equation A.4 to the *Hubbard Hamiltonian*

$$\hat{H} = -t\sum_{\langle mn\rangle}\sum_{\sigma}\hat{c}_{m\sigma}^{\dagger}\hat{c}_{n\sigma} + U\sum_{m}\hat{n}_{m\uparrow}\hat{n}_{m\downarrow}, \tag{A.7}$$

where $\hat{n}_{m\sigma}$ is the number operator $\hat{c}_{m\sigma}^{\dagger}\hat{c}_{m\sigma}$, and t and U are constants. Intuitively, the two terms of the Hubbard Hamiltonian describe the tunnelling between neighbouring lattice sites, and the local Coulomb interaction between atoms on the same site. The behaviour of our system will ultimately depend on the competition between these two terms as discussed in Subsection 3.1.1. For a more complete discussion on the Hubbard model, the reader is referred to Ref. 437.

A.2 DFT + U as derived from the Hubbard model

This appendix follows the historical derivation of the Hubbard correction used in DFT + U.

Let a correlated subspace be defined by a set of basis orbitals (known as *Hubbard projectors*). Within this subspace, the operator associated with electron-electron interactions is

$$\hat{U} = \sum_{mnm'n'}\sum_{\sigma\sigma'} U_{mnm'n'}\hat{c}_{m\sigma}^{\dagger}\hat{c}_{n\sigma'}^{\dagger}\hat{c}_{m'\sigma'}\hat{c}_{n'\sigma}, \tag{A.8}$$

where $\{m, n, m', n'\}$ are Hubbard projector labels and $\{\sigma\}$ are spin indices, and $\hat{c}_{m\sigma}^{\dagger}$ are the associated creation operators. One can show that

$$\begin{aligned} E_{\mathrm{Hub}} = \langle\hat{U}\rangle = \frac{1}{2}\sum_{\substack{mnm'n'\sigma \\ m\neq n, m'\neq n'}} & (U_{mnm'n'} - U_{mnn'm'})\langle n', \sigma; m', \sigma|\hat{\rho}_2|n, \sigma; m, \sigma\rangle \\ + \frac{1}{2}\sum_{mnm'n'\sigma} & U_{mnm'n'}\langle n', \sigma; m', -\sigma|\hat{\rho}_2|n, -\sigma; m, \sigma\rangle) \\ & - U_{mnn'm'}\langle n', -\sigma; m', \sigma|\hat{\rho}_2|n, -\sigma; m, \sigma\rangle. \end{aligned} \tag{A.9}$$

where $\hat{\rho}_2$ is the two-body density matrix. Adopting the ansatz that the many-body wavefunction is a Slater determinant of single-particle states, the two-body density matrices $\hat{\rho}_2$ can be

decomposed as determinants of single-body density.[438] In this case

$$E_{\text{Hub}} = \frac{1}{2} \sum_{\substack{mnm'n'\sigma \\ m\neq n, m'\neq n'}} (U_{mnn'm'} - U_{mnm'n'}) n^{\sigma}_{mm'} n^{\sigma}_{nn'} + \frac{1}{2} \sum_{mnm'n'\sigma} U_{mnm'n'} n^{\sigma}_{mn'} n^{-\sigma}_{nm'}, \tag{A.10}$$

where $n^{\sigma}_{mm'} = \langle m|\hat{\rho}^{\sigma}|m'\rangle$. At this stage the only approximation that has been introduced is the assertion that the state corresponds to a Slater determinant. If $U_{mnm'n'}$ is obtained using the unscreened Coulomb potential, then Equation A.10 is equivalent to a Hartree-Fock treatment of the system.

Now, all but two-site terms are ignored. Due to the symmetries of $U_{mnm'n'}$, this leaves only two types of terms: U_{mnnm} and U_{mnmn}. These are then averaged over the Hubbard projectors to yield two scalars:

$$U = \frac{1}{(2l+1)^2} \sum_{mn} U_{mnnm}; \qquad J = \frac{1}{(2l+1)^2} \sum_{mn} U_{mnmn}. \tag{A.11}$$

Using these average values in place of the tensorial terms simplifies A.10 to

$$\begin{aligned} E_{\text{Hub}} =& \frac{1}{2} \sum_{mn\sigma} U (n^{\sigma}_{mm} n^{\sigma}_{nn} - n^{\sigma}_{mn} n^{\sigma}_{nm} + n^{\sigma}_{mm} n^{-\sigma}_{nn}) \\ &+ \frac{1}{2} \sum_{mn\sigma} J (n^{\sigma}_{mn} n^{\sigma}_{nm} - n^{\sigma}_{mm} n^{\sigma}_{nn} + n^{\sigma}_{mn} n^{-\sigma}_{nm}) \\ =& \sum_{\sigma} \frac{U}{2} \left((n^{\sigma})^2 + n^{\sigma} n^{-\sigma} - \text{Tr}\left[\mathbf{n}^{\sigma}\mathbf{n}^{\sigma}\right] \right) + \frac{J}{2} \left(\text{Tr}\left[\mathbf{n}^{\sigma}\mathbf{n}^{\sigma} + \mathbf{n}^{\sigma}\mathbf{n}^{-\sigma}\right] - (n^{\sigma})^2 \right) \end{aligned} \tag{A.12}$$

where $n^{\sigma} = \text{Tr}\left[\mathbf{n}^{\sigma}\right]$. If at this stage Equation A.12 was to be incorporated directly into the DFT formalism, interactions associated with the subsystems that are already being handled by the conventional exchange-correlation functional would be double-counted. To avoid this, the fully localised limit[439] is considered, where all correlated subspaces have integer occupancy. In this approximation

$$\text{Tr}\left[\mathbf{n}^{\sigma}\mathbf{n}^{\sigma}\right] \to n^{\sigma}; \qquad \text{Tr}\left[\mathbf{n}^{\sigma}\mathbf{n}^{-\sigma}\right] \to n^{\sigma_{\min}}, \tag{A.13}$$

where $\sigma_{\min}$ denotes the minority spin. Thus in the fully localised limit, the double counting term becomes

$$E_{DC} = \frac{U}{2} n(n-1) - \frac{J}{2} \sum_{\sigma} n^{\sigma}(n^{\sigma} - 1) + J n^{\sigma_{\min}}, \tag{A.14}$$

where $n = \sum_{\sigma} n^{\sigma}$ and hence

$$E_{\text{Hub}} - E_{DC} = \sum_{I\sigma} \frac{U^I - J^I}{2} \text{Tr}\left[\mathbf{n}^{I\sigma}(1 - \mathbf{n}^{I\sigma})\right] + \sum_{I\sigma} \frac{J^I}{2} \left(\text{Tr}\left[\mathbf{n}^{I\sigma}\mathbf{n}^{I-\sigma}\right] - 2\delta_{\sigma\sigma_{\min}} n^{I\sigma} \right). \tag{A.15}$$

Note that the entire expression has now been generalised to allow for the possibility of multiple sites (labelled with the index I), to each of which a correction term is applied. As a final approximation, terms arising from interaction between opposite spin (those contained in the

second sum) are neglected. This leaves

$$E_U = E_{\text{Hub}} - E_{DC} = \sum_{I\sigma} \frac{U^I_{\text{eff}}}{2} \text{Tr}\left[\mathbf{n}^{I\sigma}(1 - \mathbf{n}^{I\sigma})\right], \tag{A.16}$$

where the on-site Coulomb repulsion parameter U^I has been effectively reduced by J^I to U^I_{eff}. (This is Equation 3.2 in the main text.) The DFT + U correction to the KS potential is given by

$$\hat{V}_U = \sum_{I\sigma mn} U^I |m\rangle \left(\frac{1}{2} - n^{I\sigma}_{mn}\right) \langle n| \tag{A.17}$$

(Equation 3.3 in the main text). With this, the derivation is complete: the Hubbard-model formalism is in a form which can be incorporated into the framework of DFT.

A.3 Ignoring the system outside of Hubbard subspaces

Suppose for a N-site Hubbard model we want to consider the existence of the bath explicitly as an additional row/column in the response matrices. Due to the fact that shifting one level by dv_{ext} is equivalent to shifting all other sites by $-dv_{\text{ext}}$ the rows/columns of the $(N+1)\times(N+1)$ response matrices must sum to zero *i.e.* the matrices will be of the form:

$$X = \begin{pmatrix} \chi & -\chi\mathbf{1} \\ -\mathbf{1}^T\chi & \mathbf{1}^T\chi\mathbf{1} \end{pmatrix} \tag{A.18}$$

where χ is the original $N \times N$ response matrix excluding the bath, and $\mathbf{1}$ is a $N \times 1$ array of ones. Since

$$\begin{aligned} X_0 \begin{pmatrix} \chi_0^{-1} - \chi^{-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix} X &= \begin{pmatrix} \chi_0 & -\chi_0\mathbf{1} \\ -\mathbf{1}^T\chi_0 & \mathbf{1}^T\chi_0\mathbf{1} \end{pmatrix} \begin{pmatrix} \chi_0^{-1} - \chi^{-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix} \begin{pmatrix} \chi & -\chi\mathbf{1} \\ -\mathbf{1}^T\chi & \mathbf{1}^T\chi\mathbf{1} \end{pmatrix} \\ &= \begin{pmatrix} I - \chi_0\chi^{-1} & \mathbf{0} \\ -\mathbf{1}^T(I - \chi_0\chi^{-1}) & 0 \end{pmatrix} \begin{pmatrix} \chi & -\chi\mathbf{1} \\ -\mathbf{1}^T\chi & \mathbf{1}^T\chi\mathbf{1} \end{pmatrix} \\ &= \begin{pmatrix} \chi - \chi_0 & -(\chi - \chi_0)\mathbf{1} \\ -\mathbf{1}^T(\chi - \chi_0) & \mathbf{1}(\chi - \chi_0)\mathbf{1} \end{pmatrix} \\ &= X - X_0 \end{aligned} \tag{A.19}$$

it follows that if f is a solution to the $N \times N$ Dyson equation then

$$F \equiv \begin{pmatrix} f & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix} = \begin{pmatrix} \chi_0^{-1} - \chi^{-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix} \tag{A.20}$$

is a solution to the $(N+1) \times (N+1)$ Dyson equation $X = X_0 + X_0FX$. This new kernel F clearly corresponds to the original N-site Hubbard model coupled to a non-interacting bath.

A.4 The difference of the inverses of two non-invertible matrices

The $(N+1)\times(N+1)$ response matrices of Appendix A.3 are non-invertible, so some authors rightly point out that $X_0^{-1}-X^{-1}$ is poorly defined. However, one can still calculate the *difference* $X_0^{-1}-X^{-1}$ via a trick first proposed (to our knowledge) by Cococcioni,[207] who stated (without proof) that the difference can instead be calculated via

$$(\mathbf{X_0}+\lambda\mathbf{1})^{-1}-(\mathbf{X}+\lambda\mathbf{1})^{-1}, \tag{A.21}$$

where $\mathbf{1}$ is now a $N+1$ by $N+1$ matrix with all entries equal to 1, and λ is some non-zero constant. In this appendix, we prove this. Let $\mathbf{X}$ be a generic response matrix that satisfies the aforementioned properties

$$\sum_j (\mathbf{X})_{ij}=0 \qquad \forall j, \tag{A.22a}$$

$$\sum_i (\mathbf{X})_{ij}=0 \qquad \forall i, \tag{A.22b}$$

and define

$$\mathbf{Q}=\mathbf{I}-\frac{1}{N+1}\mathbf{1}. \tag{A.23}$$

Theorem 1

$$\mathbf{X}=(\mathbf{X}+\lambda\mathbf{1})\mathbf{Q} \tag{A.24}$$

Proof: we start by making two observations. Firstly, it is clear that

$$\mathbf{1}^2=(N+1)\mathbf{1}. \tag{A.25}$$

Secondly, $(\mathbf{M1})_{ij}=\sum_k M_{ik}$ for any matrix $\mathbf{M}$. In the specific case of $\mathbf{X}$, we have the property that all row or column sums are 0, and thus

$$\mathbf{X1}=\mathbf{0}. \tag{A.26}$$

It follows that

$$(\mathbf{X}+\lambda\mathbf{1})\mathbf{Q}=(\mathbf{X}+\lambda\mathbf{1})\left(\mathbf{I}-\frac{1}{N+1}\mathbf{1}\right)=\mathbf{X}+\lambda\mathbf{1}-\frac{1}{N+1}((N+1)\lambda\mathbf{1})=\mathbf{X}. \tag{A.27}$$

Theorem 2: $\mathbf{X}+\lambda\mathbf{1}$ has an inverse.

This is not true of all non-invertible matrices X: it is not true of any matrix whose null space is a subspace of the null space of $\mathbf{1}$.[†] This is actually quite restrictive: there must exist some mutually exclusive sets of indices S_1 and S_2 of identical length such that $\sum_{j\in S_1} X_{ij}=\sum_{j\in S_2} X_{ij}$.

[†] If $\mathbf{v}$ is in the null space of $\mathbf{X}+\lambda\mathbf{1}$ then $(\mathbf{X}+\lambda\mathbf{1})\,\mathbf{v}=\mathbf{0}\Rightarrow\mathbf{1}\,(\mathbf{X}+\lambda\mathbf{1})\,\mathbf{v}=\mathbf{0}\Rightarrow\lambda(N+1)\mathbf{1v}=\mathbf{0}\Rightarrow\mathbf{1v}=\mathbf{0}$.

If this is not the case, then there exists some λ which will break the linear dependence of the matrix's rows/columns. We will proceed assuming this is the case.

Theorem 3:

$$\mathbf{XQ} = \mathbf{X} \tag{A.28}$$

This immediately follows from the definition of $\mathbf{Q}$ and Equation A.26.

Thus, for any two response matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ it follows that

$$\mathbf{X}_1(\mathbf{X}_2 + \lambda\mathbf{1})^{-1}\mathbf{X}_2 = \mathbf{X}_1\mathbf{Q} = \mathbf{X}_1, \tag{A.29}$$

where the first equality follows from Theorems 1 and 2, and the second equality follows from Theorem 3. It follows that

$$\mathbf{X}_1\left[(\mathbf{X}_2 + \lambda\mathbf{1})^{-1} - (\mathbf{X}_1 + \lambda\mathbf{1})^{-1}\right]\mathbf{X}_2 = \mathbf{X}_1 - \mathbf{X}_2. \tag{A.30}$$

Compare this to the behaviour of invertible matrices in a Dyson-like context: if $\mathbf{A}$ and $\mathbf{B}$ are invertible then

$$\mathbf{A}(\mathbf{B}^{-1} - \mathbf{A}^{-1})\mathbf{B} = \mathbf{A} - \mathbf{B}. \tag{A.31}$$

Thus the proof is complete: by comparison, for non-invertible response matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ we can calculate the difference of their inverses indirectly, via the difference of two invertible matrices of the form $\mathbf{X}_i + \lambda\mathbf{1}$.

A.5 Details of linear response theory

In this appendix I outline the standard formalism for linear response DFT, following Refs. 440–442 and many others.

Suppose for a given system we perturb the external potential by some small $\delta v_{\text{ext}}(\mathbf{r})$. The resulting change in the density is given by

$$\delta n(\mathbf{r}) = \int d\mathbf{r}'\chi(\mathbf{r},\mathbf{r}')\delta v_{\text{ext}}(\mathbf{r}') \tag{A.32}$$

where $\chi(\mathbf{r},\mathbf{r}')$ is the response function to this perturbation. For the same perturbation, we can choose to define a second response function $\chi_0(\mathbf{r},\mathbf{r}')$ as

$$\delta n(\mathbf{r}) = \int d\mathbf{r}'\chi_0(\mathbf{r},\mathbf{r}')\delta v_{\text{KS}}(\mathbf{r}'). \tag{A.33}$$

The KS potential is given as $v_{\text{KS}}(\mathbf{r}) = v_{\text{Hxc}}[n](\mathbf{r}) + v_{\text{ext}}(\mathbf{r})$ — that is, the sum of the Hartree and exchange-correlation potential, and the external potential (which includes the atomic potentials as well as the perturbing potential). It follows that $\delta v_{\text{KS}}(\mathbf{r}) = \delta v_{\text{Hxc}}[n](\mathbf{r}) + \delta v_{\text{ext}}(\mathbf{r})$. By the KS

construction, the change in the Hubbard-plus-xc-potential can be recast as

$$\delta v_{\mathrm{Hxc}}[n](\mathbf{r}') = \int d\mathbf{r}'' f[n_{GS}](\mathbf{r}', \mathbf{r}'')\delta n(\mathbf{r}''), \tag{A.34}$$

where we have defined the Hartree plus exchange-correlation kernel as

$$f[n_{GS}](\mathbf{r}', \mathbf{r}'') = \left.\frac{\delta v_{\mathrm{Hxc}}(\mathbf{r}')}{\delta n(\mathbf{r}'')}\right|_{n=n_{GS}}. \tag{A.35}$$

Combining Equations A.32–A.35 we can see that χ, χ_0, and f are related via a Dyson-like equation for the Hartree plus exchange-correlation kernel:

$$\chi(\mathbf{r}, \mathbf{r}') = \chi_0(\mathbf{r}, \mathbf{r}') + \int d\mathbf{r}'' \int d\mathbf{r}''' \chi_0(\mathbf{r}, \mathbf{r}''') f[n_{GS}](\mathbf{r}''', \mathbf{r}'')\chi(\mathbf{r}'', \mathbf{r}') \tag{A.36}$$

and we can identify $\chi_0(\mathbf{r}, \mathbf{r}') = \delta n(\mathbf{r})/\delta v_{\mathrm{KS}}(\mathbf{r}')$ as the non-interacting response. For subspaces defined by projection operators $\hat{P}^J$, Equation 3.12b defines the projected non-interacting response, which is used in the minimum-tracking formalism for U.

A.6 A brief introduction to Green's functions

This appendix provides a brief overview of Green's functions.

The idea of a *correlation function* is of immense interest and importance in both experimental and theoretical physics; they are the measure of the response of a system to a perturbation. That is, we are interested in the quantity $X(t) = \langle \hat{X}^F(t) \rangle$: the response of a system in the presence of a perturbation $\hat{H} = \hat{H}_0 + F(t)\hat{Y}$. One can demonstrate that

$$X(t) = \int dt' C_{XY}(t - t')F(t') \tag{A.37}$$

where the retarded response function $C^{XY}(t)$ is given by

$$C^{XY}(t) = -i\Theta(t)\langle\{\hat{X}(t), \hat{Y}(0)\}\rangle. \tag{A.38}$$

(see Ref. 437 for the derivation). If we now focus on correlation functions for fermionic creation and annihilation operators, we get what we call the (fermionic) retarded Green's function

$$G^+_{\alpha\beta}(t) = -i\Theta(t)\langle\{c_\alpha(t), c^\dagger_\beta(0)\}\rangle \tag{A.39}$$

and one can similarly define the advanced Green's function

$$G^-_{\alpha\beta}(t) = i\Theta(-t)\langle\{c_\alpha(t), c^\dagger_\beta(0)\}\rangle. \tag{A.40}$$

It is also useful (for reasons that will be explained later) to define the imaginary-time Green's

function

$$G_{\alpha\beta}(\tau) = -\left\langle T\left[c_\alpha(\tau)c_\beta^\dagger(0)\right]\right\rangle = -\begin{cases}\langle c_\alpha(\tau)c_\beta^\dagger(0)\rangle, & \tau > 0\\ -\langle c_\beta^\dagger(0)c_\alpha(\tau)\rangle, & \tau \le 0\end{cases} \tag{A.41}$$

where the time-evolution of the operators is defined through the imaginary-time Heisenberg representation

$$\hat{X}(\tau) = e^{\tau(\hat{H}-\mu\hat{N})}\hat{X}e^{-\tau(\hat{H}-\mu\hat{N})}, \tag{A.42}$$

and $\langle\hat{X}\rangle = \mathcal{Z}^{-1}\mathrm{Tr}\left[\hat{X}\exp -\beta(\hat{H}-\mu\hat{N})\right]$.[†] All of these Green's functions are measures of the probability amplitude for the propagation of a particle (or hole) excitation in an equilibrium state.

By analogy with the imaginary-time Green's function we can also define the real-time Green's function,

$$G_{\alpha\beta}^T(t) = -i\langle T_t c_\alpha(t)c_\beta^\dagger(0)\rangle \tag{A.43}$$

but while this substitution may seem like the most natural analogue to the imaginary-time Green's function, the retarded/advance Green's functions turn out to be of much more physical significance. A Fourier transform of the imaginary-time Green's function yields the Matsubara Green's function

$$G_{\alpha\beta}(i\omega_n) = \int_0^\beta d\tau\, G_{\alpha\beta}(\tau)e^{i\omega_n\tau} \tag{A.44}$$

with Matsubara frequencies $i\omega_n = \frac{2\pi}{\beta}(n+\frac{1}{2})$.

It is useful to adopt the Lehmann representation of the Green's function:

$$G_{\alpha\beta}(i\omega_n) = \int_{-\infty}^{\infty} d\omega\, \frac{A_{\alpha\beta}(\omega)}{i\omega_n - \omega} \tag{A.45}$$

where

$$A_{\alpha\beta}(\omega) = \frac{1}{\mathcal{Z}}\sum_{m,n}\langle n|c_\beta^\dagger|m\rangle\langle m|c_\alpha|n\rangle(e^{-\beta E_m} - e^{-\beta E_n})\delta(\omega - (E_n - E_m)) \tag{A.46}$$

and $\{E_i\}$ are the eigenvalues of $\hat{H}-\mu\hat{N}$. Note that in practice the spectral function is not often explicitly considered, but it does give us insight into the nature of Green's functions, as explored below.

Before proceeding, it is worth noting that the retarded Green's function $G_{\alpha\beta}^+(\omega)$ can be obtained by analytic extension of the Matsubara Green's function $G_{\alpha\beta}(i\omega_n)$: that is, we can expand the definition of G across the entire complex plane:

$$G(z) = \int_{-\infty}^{\infty} d\omega \frac{A(\omega)}{z-\omega} \tag{A.47}$$

in which case we have

$$G^+(\omega) = G(\omega + i0^+); \tag{A.48a}$$

$$G^-(\omega) = G(\omega - i0^+). \tag{A.48b}$$

[†]Also note that $G_{\alpha\beta}(\tau) = -G_{\alpha\beta}(\tau+\beta)$ for $-\beta < \tau < 0$ due to the cyclic properties of the trace.

This vindicates the earlier choice to consider retarded and advanced Green's functions on the real time axis. At zero temperature, we get the important identity $A_{\alpha\alpha} = -\frac{1}{\pi}\mathrm{Im}[G^+_{\alpha\alpha}(\omega)]$.

A.6.1 Green's function of the non-interacting system

Consider a system of free (fermionic) particles with Hamiltonian

$$\hat{H} - \mu\hat{N} = \sum_\alpha (\varepsilon_\alpha - \mu) c^\dagger_\alpha c_\alpha = \sum_\alpha \xi_\alpha c^\dagger_\alpha c_\alpha. \tag{A.49}$$

The eigenstates $|n\rangle$ are antisymmetrical combinations of single-particle eigenstates $\{|\alpha\rangle\}$, and their energies are simply $E_n = \sum_\alpha^{\mathrm{occ}} \xi_\alpha$. The only non-vanishing terms in the spectral function $A_{\alpha\beta}(\omega)$ are those where $\beta = \alpha$, and $|n\rangle$ and $|m\rangle$ only differ in the occupation of the single-particle eigenstate $|\alpha\rangle$. Thus we have

$$\begin{aligned}
A_{\alpha\beta}(\omega) &= \frac{1}{\mathcal{Z}}\delta_{\alpha\beta} \sum_{m,n} \langle n|c^\dagger_\alpha|m\rangle\langle m|c_\alpha|n\rangle \left(e^{-\beta E_m} - e^{-\beta(E_m+\xi_\alpha)}\right) \delta(\omega - \xi_\alpha) \\
&= \frac{1}{\mathcal{Z}}\delta_{\alpha\beta}\delta(\omega-\xi_\alpha)\left(1 - e^{-\beta\xi_\alpha}\right) \sum_m \langle m|c_\alpha c^\dagger_\alpha|m\rangle e^{-\beta E_m} \\
&= \frac{1}{\mathcal{Z}}\delta_{\alpha\beta}\delta(\omega-\xi_\alpha)\left(1 - e^{-\beta\xi_\alpha}\right) \sum_m \langle m|(1 - c^\dagger_\alpha c_\alpha)|m\rangle e^{-\beta E_m} \\
&= \frac{1}{\mathcal{Z}}\delta_{\alpha\beta}\delta(\omega-\xi_\alpha)\left(1 - e^{-\beta\xi_\alpha}\right) \mathcal{Z}\left(1 - n_F(\xi_\alpha)\right) \\
&= \delta_{\alpha\beta}\delta(\omega - \xi_\alpha)
\end{aligned} \tag{A.50}$$

and hence

$$G^0_{\alpha\beta}(\omega) = \frac{\delta_{\alpha\beta}}{\omega - \xi_\alpha}. \tag{A.51}$$

This makes physical sense. Harking back to our initial definition of the retarded Green's function, consider adding a particle to a non-interacting system in its ground state, where the particle is added into $|\alpha\rangle$ at time $t = 0$. We know that it will evolve as $|\psi(t)\rangle = |\alpha\rangle e^{-i\xi_\alpha t}$, and consequently the probability amplitude that it will be in state $|\beta\rangle$ at some later time is

$$\langle\beta|\psi(t)\rangle = \delta_{\alpha\beta}\Theta(t)e^{-i\xi_\alpha t}. \tag{A.52}$$

(Note the Heaviside function enforcing t is a *later* time.) This is of course nothing less than the retarded Green's function $iG^+_{\alpha\beta}(t)$, the Fourier transform of which is simply

$$G^+_{\alpha\beta}(\omega) = -i\int_{-\infty}^{\infty} dt\, \delta_{\alpha\beta}\Theta(t)e^{i(\omega-\xi_\alpha)t} = \delta_{\alpha\beta}\left.\frac{e^{i(\omega-\xi_\alpha)t}}{\omega-\xi_\alpha}\right|_0^\infty = \frac{\delta_{\alpha\beta}}{\omega + i0^+ - \xi_\alpha} \tag{A.53}$$

where we resolve the oscillating exponential at infinity with the standard trick of shifting the singularity off the real axis slightly. This matches the result we already found earlier (Equation A.51).

A.6.2 Extending to interacting systems

Building on the results for the non-interacting case, it is customary to introduce the Green's function operator as

$$\hat{G}^0(\omega) = \frac{1}{\omega + \mu - \hat{h}} \tag{A.54}$$

where $\hat{h}$ is the single-particle Hamiltonian. We can extend this to the interacting case by introducing the *self energy* Σ:

$$\hat{G}(\omega) = \frac{1}{\omega + \mu - \hat{h} - \Sigma(\omega)}. \tag{A.55}$$

Because it is defined in this way, the self-energy is a measure of the difference between the interacting and non-interacting Green's functions:

$$G_{\alpha\beta}(\omega)^{-1} = G^0_{\alpha\beta}(\omega)^{-1} - \Sigma_{\alpha\beta}(\omega) \tag{A.56}$$

and the Green's function obeys a Dyson equation $G = G^0 + G^0 \Sigma G$.

A.6.3 Why imaginary time?

The choice to work in imaginary time may seem like a strange decision, but there is a good reason for it, and that is to do with finite temperature.

For $T \neq 0$, the expectation value of a time-independent Hamiltonian in the grand-canonical ensemble is

$$\langle \hat{O} \rangle = \frac{1}{Z} \mathrm{Tr}\left[e^{-\beta \hat{K}} \hat{O} \right] \tag{A.57}$$

where $Z = \mathrm{Tr}\left[e^{-\beta \hat{K}} \right]$ and $\hat{K} = \hat{H} - \mu \hat{N}$. In the Heisenberg picture, time-evolution is offloaded onto operators, and

$$\hat{\psi}(x,t) = e^{i\hat{K}t} \hat{\psi}(x) e^{-i\hat{K}t}. \tag{A.58}$$

The time-ordered Green's function for $t > t'$ becomes

$$G^T_{\alpha\beta}(t,t') = -\frac{i}{Z} \mathrm{Tr}\left[e^{-\beta \hat{K}} c_\alpha(t) c^\dagger_\beta(t') \right] = -\frac{i}{Z} \mathrm{Tr}\left[e^{-\beta \hat{K}} \left(e^{i\hat{K}t} \hat{c}_\alpha e^{-i\hat{K}t} \right) \left(e^{-i\hat{K}t'} \hat{c}^\dagger_\beta e^{i\hat{K}t'} \right) \right]. \tag{A.59}$$

The presence of these various thermal factors means that G^T cannot be cast in the form of a spectral function *à la* Equation A.47. However, this problem is resolved if we move to imaginary frequencies. It is easy to demonstrate that

$$\begin{aligned} G^T_{\alpha\beta}(t + i\beta, t') =& -\frac{i}{Z} \mathrm{Tr}\left[e^{-\beta \hat{K}} \left(e^{i\hat{K}(t+i\beta)} \hat{c}_\alpha e^{-i\hat{K}(t+i\beta)} \right) \left(e^{-i\hat{K}t'} \hat{c}^\dagger_\beta e^{i\hat{K}t'} \right) \right] \\ =& \frac{i}{Z} \mathrm{Tr}\left[e^{-\beta \hat{K}} \left(e^{i\hat{K}t} \hat{c}_\alpha e^{-i\hat{K}t} \right) \left(e^{-i\hat{K}t'} \hat{c}^\dagger_\beta e^{i\hat{K}t'} \right) \right] \\ =& -G^T_{\alpha\beta}(t,t') \end{aligned} \tag{A.60}$$

— that is, it is anti-periodic in imaginary time with period β. Likewise, for $t < t'$, $G^T_{\alpha\beta}(t-i\beta, t') = -G^T_{\alpha\beta}(t, t')$. We can exploit this by working in imaginary time, in which case $G(\tau)$ can be expressed as a Fourier series

$$G(\tau) = \frac{1}{\beta} \sum_{n=-\infty}^{\infty} e^{-i\omega_n \tau} G(i\omega_n) \tag{A.61}$$

where $\omega_n = (2n+1)\pi/\beta$. The reverse transform is

$$G(i\omega_n) = \frac{1}{2} \int_{-\beta}^{\beta} d\tau \, e^{i\omega_n \tau} G(\tau). \tag{A.62}$$

This is the end result: for non-zero temperatures, the Green's function is determined solely by its values at the Matsubara frequencies, and thus it makes sense to work in this imaginary-time framework to take advantage of this fact. That said, the process of analytic continuation (reconstructing real-time Green's functions from their imaginary-time counterparts) is a difficult and sensitive process (for example, see Ref. 443).

A.6.4 Extracting system properties from Green's functions

Knowing the one-particle Green's function allows one access to a raft of system properties.

A generic operator

The expectation value of an operator $\hat{O} = \sum_{\alpha\beta} O_{\alpha\beta} c^\dagger_\alpha c_\beta$ is given by

$$\langle \hat{O} \rangle = \sum_{\alpha\beta} O_{\alpha\beta} \langle c^\dagger_\alpha c_\beta \rangle = -i \sum_{\alpha\beta} O_{\alpha\beta} \lim_{t' \to t^+} G^T_{\alpha\beta}(t, t') = \frac{1}{2i\pi} \sum_{\alpha\beta} O_{\alpha\beta} \lim_{\eta \to 0^+} \int d\omega e^{i\eta\omega} G^T_{\alpha\beta}(\omega). \tag{A.63}$$

This can be rewritten in terms of the spectral function

$$\langle \hat{O} \rangle = \sum_{\alpha\beta} O_{\alpha\beta} \int d\omega \, \frac{A_{\alpha\beta}(\omega)}{1 + e^{\beta(\omega-\mu)}} \tag{A.64}$$

or in terms of the Matsubara frequencies

$$\langle \hat{O} \rangle = \frac{1}{\beta} \sum_{\alpha\beta n} O_{\alpha\beta} G_{\alpha\beta}(i\omega_n). \tag{A.65}$$

Density and the density matrix

For the density matrix $\hat{\rho}$

$$\rho_{\alpha\beta} = \int d\omega \, \frac{A_{\alpha\beta}(\omega)}{1 + e^{\beta(\omega-\mu)}}. \tag{A.66}$$

At zero temperature we have

$$\rho_{\alpha\alpha} = \int_{-\infty}^{\mu} d\omega\, A_{\alpha\alpha}(\omega) = \frac{1}{\pi}\int_{-\infty}^{\mu} d\omega\, \mathrm{Im}[G^{+}_{\alpha\alpha}(\omega)] \tag{A.67}$$

and for finite temperature

$$\rho_{\alpha\alpha} = G_{\alpha\alpha}(\tau = 0^{-}) = \frac{1}{\beta}\sum_{-\infty}^{\infty} e^{-i\omega 0^{-}} G_{\alpha\alpha}(i\omega_n). \tag{A.68}$$

Total energy

The total energy of a system of interacting electrons is related to its Green's function via the Galitskii-Migdal formula[444]:

$$E = \frac{1}{2}\int d\mathbf{r} \lim_{t'\to t^{+}} \lim_{\mathbf{r}'\to\mathbf{r}} \left(\frac{\partial}{\partial t} - i h_0(\mathbf{r})\right) G(\mathbf{r}, t; \mathbf{r}', t') \tag{A.69}$$

where $h_0(\mathbf{r})$ is the independent-particle Hamiltonian. For further details on Green's functions, the reader is referred to Refs. 147 and 437.

A.7 Solving an AIM via exact diagonalisation

One of the most computationally-intensive steps in a DMFT calculation is obtaining AIM impurity Green's function from the AIM Hamiltonian. These are related via Equation 4.11, which was

$$G_{\mathrm{imp}_{mm'}}(\omega) = \left\langle \hat{c}_m \frac{1}{\omega - (\hat{H} - E_0)} \hat{c}^{\dagger}_{m'} \right\rangle + \left\langle \hat{c}^{\dagger}_{m'} \frac{1}{\omega + (\hat{H} - E_0)} \hat{c}_m \right\rangle \tag{A.70}$$

where $\langle\ \bullet\ \rangle$ is the thermodynamic average, which at zero temperature becomes $\langle\psi_0|\ \bullet\ |\psi_0\rangle$.

This appendix describes how we can resolve this equation using exact diagonalisation (ED) via the Lanczos algorithm.

A.7.1 The Lanczos algorithm

The Lanczos algorithm is an approach for obtaining the eigenvectors and eigenvalues of a Hermitian matrix A, without ever having to perform a full diagonalisation.

Starting with some arbitrary normalised vector $|0\rangle$, we compute $\varepsilon_0 = \langle 0|A|0\rangle$. Then we construct $|\tilde{1}\rangle = \hat{A}|0\rangle - \varepsilon_0|0\rangle$, and normalise to obtain $|1\rangle$. Importantly, the resulting vector $|1\rangle$ is orthogonal to $|0\rangle$.

We can now generate a third vector $|\tilde{2}\rangle = A|1\rangle - \varepsilon_1|1\rangle - k_1|0\rangle$, where $k_1 = \langle 0|A|1\rangle$, and normalise to obtain $|2\rangle$. Again, $|0\rangle$, $|1\rangle$, and $|2\rangle$ are orthogonal by construction.

Now suppose we were to continue to generate orthogonal vectors according to this pattern

$$|i+1\rangle = \frac{1}{\sqrt{\langle i|(A-\varepsilon_i)^2|i\rangle + {k_i}^2}} A|i\rangle - \varepsilon_i|i\rangle - k_i|i-1\rangle \tag{A.71}$$

to obtain a basis of Lanczos vectors $\{|i\rangle\}$. In this basis, the matrix A is tridiagonal:[†]

$$A_{ij} = \begin{pmatrix} \varepsilon_0 & k_1 & 0 & \cdots & \cdots \\ k_1 & \varepsilon_1 & k_2 & 0 & \cdots \\ 0 & k_2 & \varepsilon_2 & k_3 & 0 \\ \vdots & 0 & k_3 & \varepsilon_3 & \ddots \\ \vdots & \vdots & 0 & \ddots & \ddots \end{pmatrix}_{ij} . \tag{A.72}$$

From here, it is straightforward to calculate the eigenvectors and eigenvalues of A.

As an approximate scheme, one need only consider the first $L+1$ Lanczos vectors. In this case, $\tilde{A}_{ij} = \sum_{kl}^{L} \langle i|k\rangle\langle k|A|l\rangle\langle l|j\rangle$ is an $(L+1)$-by-$(L+1)$ tridiagonal matrix, the eigenvalue problem $\tilde{A}c^\nu = E_\nu c^\nu$ is straightforward to solve, and the eigenvectors of $\tilde{A}$ are approximated by $|\nu\rangle = \sum_i^L c_i^\nu |i\rangle$. By progressively increasing L and periodically recalculating $\{E_0, ..., E_L\}$ one can converge to the eigenvectors and energies of A without ever doing the full diagonalisation.

Note that this algorithm is very cheap; multiplication by $\tilde{A}$ is the most expensive step, and scales as $\mathcal{O}(L^2)$. It also is worthwhile noting that because the Lanczos basis is generated via repeated action of A on the previous Lanczos vector, the Lanczos algorithm rapidly finds the vectors $|i\rangle$ for which $A|i\rangle$ is large — another advantage of the method.

A.7.2 Applying the Lanczos method to the AIM

Let us return now to the problem at hand: we would like to calculate

$$G_{\text{imp}}^{\alpha\beta}(\omega) = \left\langle \psi_0 \left| \hat{c}_\alpha \frac{1}{\omega^+ - (\hat{H} - E_0)} \hat{c}_\beta^\dagger \right| \psi_0 \right\rangle + \left\langle \psi_0 \left| \hat{c}_\beta^\dagger \frac{1}{\omega^+ + (\hat{H} - E_0)} \hat{c}_\alpha \right| \psi_0 \right\rangle .$$

Obtaining $|\psi_0\rangle$ is straightforward: we can obtain it by performing the Lanczos algorithm on $\hat{H}$, as described in the previous section. Given $|\psi_0\rangle$, some additional tricks are necessary to arrive at the Green's function. Let us first focus on the diagonal components $G_{\text{imp}}^{\alpha\alpha}[\omega]$, in which case we are interested in quantities of the form

$$\left\langle \psi_0 \left| \mathcal{O}^\dagger \frac{1}{z - H} \mathcal{O} \right| \psi_0 \right\rangle . \tag{A.73}$$

for some generic operator $\mathcal{O}$. To calculate this, we perform the Lanczos algorithm on H — but now, instead of starting with a random vector, we choose

$$|0\rangle = \frac{\mathcal{O}|\psi_0\rangle}{\sqrt{\langle\psi_0|\mathcal{O}^\dagger\mathcal{O}|\psi_0\rangle}}. \tag{A.74}$$

[†]This is straightforward to show. For example, $\langle j|A|i\rangle = \langle j| \left(|i \tilde{+} 1\rangle + \varepsilon_i|i\rangle + k_i|i-1\rangle \right) = 0$ if $i \leq j-2$. The other entries can be obtained via similar logic.

In the Lanczos basis generated using this vector, we have

$$(z-H)_{ij} = \begin{pmatrix} z-\varepsilon_0 & -k_1 & 0 & \cdots & \cdots \\ -k_1 & z-\varepsilon_1 & -k_2 & 0 & \cdots \\ 0 & -k_2 & z-\varepsilon_2 & -k_3 & 0 \\ \vdots & 0 & -k_3 & z-\varepsilon_3 & \ddots \\ \vdots & \vdots & 0 & \ddots & \ddots \end{pmatrix}_{ij} \tag{A.75}$$

Crucially, the quantity we ultimately want to obtain (Equation A.73) is $(z-H)^{-1}_{00}$, which is given[†] by the continued fraction

$$\cfrac{1}{z-\varepsilon_0-\cfrac{|k_1|^2}{z-\varepsilon_1-\cfrac{|k_2|^2}{z-\varepsilon_2-\cdots}}} \tag{A.81}$$

which can be numerically evaluated (via, for example, the modified Lentz method[445]). Thus we can calculate the diagonal terms $G^{\alpha\alpha}_{\text{imp}}[\omega]$ by setting $\mathcal{O} = \hat{c}_\alpha$. The off-diagonal terms, meanwhile, require some clever trickery: it can be shown[446] that

$$G^{\alpha\beta}_{imp} = \mathcal{G}^{\alpha\beta} - \frac{1}{2}\left(G^{\alpha\alpha}_{imp} + G^{\beta\beta}_{imp}\right) \tag{A.82}$$

where $\mathcal{G}^{\alpha\beta}$ is the result of repeating the above process for the diagonal elements, but now using the initial Lanczos matrix $\mathcal{O} = \frac{1}{\sqrt{2}}(\hat{c}_\alpha + c_\beta)$. (This avoids a vanishing denominator $\langle\psi_0|c^\dagger_\alpha c_\beta|\psi_0\rangle$ if we were to blindly proceed with the same procedure as for the diagonal elements.)

[†]The ij-element of the inverse of A is given by

$$(A^{-1})_{ij} = (-1)^{i+j}\frac{\det\Delta_{ij}}{\det A} \tag{A.76}$$

where Δ_{ij} is the sub-matrix of A obtained by eliminating from A the i-th row and j-th column. In the case of a tridiagonal matrix,

$$\det A = \det\begin{pmatrix} A_{00} & A_{01} & 0 & & \\ A_{10} & A_{11} & A_{12} & 0 & \\ 0 & A_{21} & A_{22} & A_{23} & 0 \\ & 0 & A_{32} & A_{33} & A_{34} \\ & & 0 & A_{43} & A_{44} \end{pmatrix} = A_{00}\det\begin{pmatrix} A_{11} & A_{12} & 0 & \\ A_{21} & A_{22} & A_{23} & 0 \\ 0 & A_{32} & A_{33} & A_{34} \\ & 0 & A_{43} & A_{44} \end{pmatrix} - A_{01}A_{10}\det\begin{pmatrix} A_{22} & A_{23} & 0 \\ A_{32} & A_{33} & A_{34} \\ 0 & A_{43} & A_{44} \end{pmatrix}. \tag{A.77}$$

If D_i is determinant of the matrix A having removed the first i rows and columns, it follows that

$$\frac{D_0}{D_1} = \frac{A_{00}D_1 - |A_{01}|^2 D_2}{D_1} = A_{00} - \frac{A_{01}A_{10}}{D_1/D_2}. \tag{A.78}$$

This reasoning can be extended to

$$\frac{D_l}{D_{l+1}} = A_{ll} - \frac{|A_{ll+1}|^2}{D_{l+1}/D_{l+2}} \tag{A.79}$$

and thus the first element of the inverse of A is given by the continued fraction

$$(A^{-1})_{00} = \frac{1}{D_0/D_1} = \cfrac{1}{A_{00}-\cfrac{|A_{01}|^2}{A_{11}-\cfrac{|A_{12}|^2}{A_{22}-\cdots}}} \tag{A.80}$$

as claimed.

A.8 Is superexchange via empty orbitals viable?

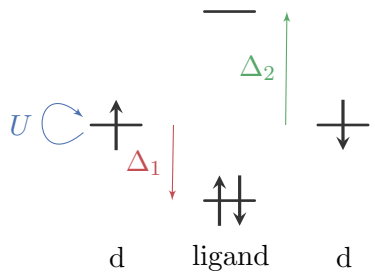


Figure A.1: A sketch of two d orbitals bridged by a ligand orbital.

Consider the system displayed in Figure A.1. Conventional superexchange goes via the filled ligand orbital (and hence coupling is dependent on Δ_1). But what about pathways via the empty ligand orbital? Do such pathways give rise to any spin coupling? And, if so, is this coupling ferromagnetic or antiferromagnetic?

For this subsystem, the Hamiltonian is

$$H = -t\sum_{i\sigma}\left(c_{i\sigma}^{\dagger}c_{p\sigma} + c_{p\sigma}^{\dagger}c_{i\sigma}\right) + \Delta_2\sum_{\sigma} n_{p\sigma} + U\sum_{i} n_{i\uparrow}n_{i\downarrow} \tag{A.83}$$

where t is the metal-to-ligand hopping, i is the metal index, p signifies the ligand site, σ is the spin index, Δ_2 is the difference in the energies of the empty ligand orbital and the metal sites $\varepsilon_p - \varepsilon_d$, and U is the on-site Hubbard interaction on the metal sites.

Let us first explore the Hilbert space of two spin-up electrons, which has the basis $\{c_{1\uparrow}^{\dagger}c_{2\uparrow}^{\dagger}|0\rangle$, $c_{1\uparrow}^{\dagger}c_{p\uparrow}^{\dagger}|0\rangle$, $c_{2\uparrow}^{\dagger}c_{p\uparrow}^{\dagger}|0\rangle\}$. In this Hilbert space, the Hamiltonian is

$$H = \left(\begin{array}{c|cc} 0 & -t & t \\ \hline -t & \Delta_2 & 0 \\ t & 0 & \Delta_2 \end{array}\right) = \begin{pmatrix} \mathbf{A} & \mathbf{T_{01}} \\ \mathbf{T_{10}} & \mathbf{B} \end{pmatrix} \tag{A.84}$$

where lines (and corresponding block matrices) delineate the subspaces where there is or isn't an electron on the ligand site. Downfolding the ligand states gives the energy-dependent effective Hamiltonian

$$H_{\text{eff}}(\omega) = \mathbf{A} + \mathbf{T_{01}}(\omega - \mathbf{B})^{-1}\mathbf{T_{10}} \approx \mathbf{A} + \mathbf{T_{01}}\mathbf{B}^{-1}\mathbf{T_{10}} = -\frac{2t^2}{\Delta_2} \tag{A.85}$$

where in the second step we take the static limit ($\omega \to 0$).

Now let us analogously explore two-electron states with opposite spin. Now our basis is much larger: $\{c_{1\uparrow}^{\dagger}c_{2\downarrow}^{\dagger}|0\rangle$, $c_{2\uparrow}^{\dagger}c_{1\downarrow}^{\dagger}|0\rangle$, $c_{1\uparrow}^{\dagger}c_{p\downarrow}^{\dagger}|0\rangle$, $c_{2\uparrow}^{\dagger}c_{p\downarrow}^{\dagger}|0\rangle$, $c_{p\uparrow}^{\dagger}c_{1\downarrow}^{\dagger}|0\rangle$, $c_{p\uparrow}^{\dagger}c_{2\downarrow}^{\dagger}|0\rangle$, $c_{1\uparrow}^{\dagger}c_{1\downarrow}^{\dagger}|0\rangle$, $c_{2\uparrow}^{\dagger}c_{2\downarrow}^{\dagger}|0\rangle$,

$c^\dagger_{p\uparrow}c^\dagger_{p\downarrow}|0\rangle\}$. Our Hamiltonian is

$$H = \left(\begin{array}{cc|ccccc|ccc} 0 & 0 & -t & 0 & 0 & -t & 0 & 0 & 0 \\ 0 & 0 & 0 & -t & -t & 0 & 0 & 0 & 0 \\ \hline -t & 0 & \Delta_2 & 0 & 0 & 0 & -t & 0 & -t \\ 0 & -t & 0 & \Delta_2 & 0 & 0 & 0 & -t & -t \\ 0 & -t & 0 & 0 & \Delta_2 & 0 & -t & 0 & -t \\ -t & 0 & 0 & 0 & 0 & \Delta_2 & 0 & -t & -t \\ \hline 0 & 0 & -t & 0 & -t & 0 & U & 0 & 0 \\ 0 & 0 & 0 & -t & 0 & -t & 0 & U & 0 \\ 0 & 0 & -t & -t & -t & -t & 0 & 0 & 2\Delta_2 \end{array}\right) = \left(\begin{array}{ccc} \mathbf{A} & \mathbf{T_{01}} & \\ \mathbf{T_{10}} & \mathbf{B} & \mathbf{T_{12}} \\ & \mathbf{T_{21}} & \mathbf{C} \end{array}\right) \tag{A.86}$$

where now we have three subspaces: that with the two electrons on the two metal sites, those states which are coupled to the first subspace, and those states which are not coupled to the first subspace. Downfolding as before gives

$$\begin{aligned} H_{\text{eff}}(\omega) =& \mathbf{A} + \mathbf{T_{01}}(\omega - \mathbf{B} - \mathbf{T_{12}}(\omega - \mathbf{C})^{-1}\mathbf{T_{21}})^{-1}\mathbf{T_{10}} \\ \approx& \mathbf{A} + \mathbf{T_{01}}\mathbf{B}^{-1}\mathbf{T_{12}}\mathbf{C}^{-1}\mathbf{T_{21}}\mathbf{B}^{-1}\mathbf{T_{10}} \\ =& -\frac{2t^2}{\Delta_2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \frac{4t^4}{\Delta_2^2}\left(\frac{1}{\Delta_2} + \frac{1}{U}\right)\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} \end{aligned} \tag{A.87}$$

where in the second step we take the static limit ($\omega \to 0$). Defining

$$\alpha = -\frac{2t^2}{\Delta_2}; \qquad \beta = -\frac{4t^4}{\Delta_2^2}\left(\frac{1}{\Delta_2} + \frac{1}{U}\right) \tag{A.88}$$

then the Hamiltonian

$$\begin{pmatrix}\alpha + \beta & \beta \\ \beta & \alpha + \beta\end{pmatrix} \tag{A.89}$$

obviously has eigenvectors and eigenvalues

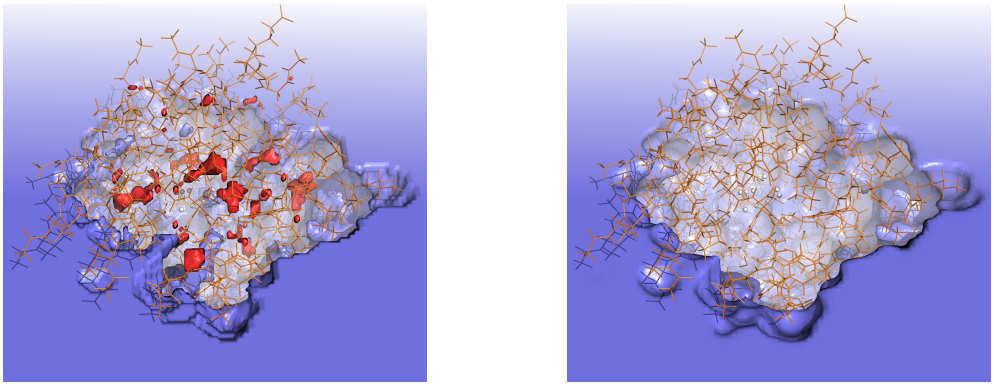
$$\varepsilon_1 = \alpha; \qquad |1\rangle = \frac{1}{\sqrt{2}}\left(c^\dagger_{1\uparrow}c^\dagger_{2\downarrow} - c^\dagger_{2\uparrow}c^\dagger_{1\downarrow}\right)|0\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle\right) \tag{A.90}$$

$$\varepsilon_2 = \alpha + 2\beta; \qquad |2\rangle = \frac{1}{\sqrt{2}}\left(c^\dagger_{1\uparrow}c^\dagger_{2\downarrow} + c^\dagger_{2\uparrow}c^\dagger_{1\downarrow}\right)|0\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow,\downarrow\rangle - |\downarrow,\uparrow\rangle\right) \tag{A.91}$$

— the triplet and singlet, respectively. (Note that getting the correct ordering of the creation operators is crucial!) The triplet state has the same energy α as the aligned-spin case (Equation A.85). Meanwhile, the singlet state is lower in energy; *i.e.* this superexchange mechanism gives rise to an *antiferromagnetic* spin coupling

$$J = -2\beta = \frac{8t^4}{\Delta_2^2}\left(\frac{1}{\Delta_2} + \frac{1}{U}\right). \tag{A.92}$$

(Note that the divergence as $\Delta_2 \to 0$ is not physical: in this limit the downfolding is no longer valid.)



(a) without exclusion region **(b)** with exclusion region

Figure A.2: The cluster model of the OEC (white spheres) and ambient protein (orange), showing the implicit solvent isosurfaces $\varepsilon = \varepsilon_\infty/2$. The ambient solvent (blue) is removed above the $x = 0$ plane to reveal the enclosed protein cluster: (a) the default implicit solvent cavity. Pockets of solvent that are disconnected from the ambient solvent are highlighted in red. Because the cluster already contains explicit water atoms, the presence of implicit solvent within the cluster is undesirable; (b) upon application of a smooth spherical exclusion region, the implicit solvent was expelled from the cluster.

This result is very similar to conventional superexchange via the filled ligand orbital, which gives the effective Hamiltonians

$$H_{\text{eff}} = -\frac{2t^2}{U - \Delta_1} \tag{A.93}$$

for parallel spins and

$$H_{\text{eff}} = -\frac{2t^2}{U - \Delta_1}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{2t^4}{(U - \Delta_1)^2}\left(\frac{1}{U} + \frac{1}{U - \Delta_1}\right)\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{A.94}$$

for anti-parallel spins, where again Δ_1 is the difference of the orbital energies $\varepsilon_p - \varepsilon_d$, but now for the filled p orbital (typically, as in Figure A.1, $\Delta_1 < 0$). The resulting spin-coupling is

$$J = \frac{4t^4}{(U - \Delta_1)^2}\left(\frac{1}{U} + \frac{1}{U - \Delta_1}\right). \tag{A.95}$$

In hindsight, the result of this appendix is obvious: superexchange via empty orbitals is nothing less than superexchange of holes.

A.9 Smooth implicit solvent exclusion regions

During the optimisation of the OEC (Section 6.1) I quickly noticed some anomalous solvation energies, caused by locations within the cluster where the density dropped so low that pockets of implicit solvent started to form (see the red surfaces in Figure A.2a). This is undesired: all the water molecules within the cluster were explicitly included, so the implicit solvent should be restricted to outside the cluster.

In order to overcome this, I implemented smooth implicit solvent exclusion regions in ONETEP. Implicit solvent exclusion regions were already implemented, but they had hard-walled boundaries which had to coincide with areas of high electronic density, since the Poisson equation

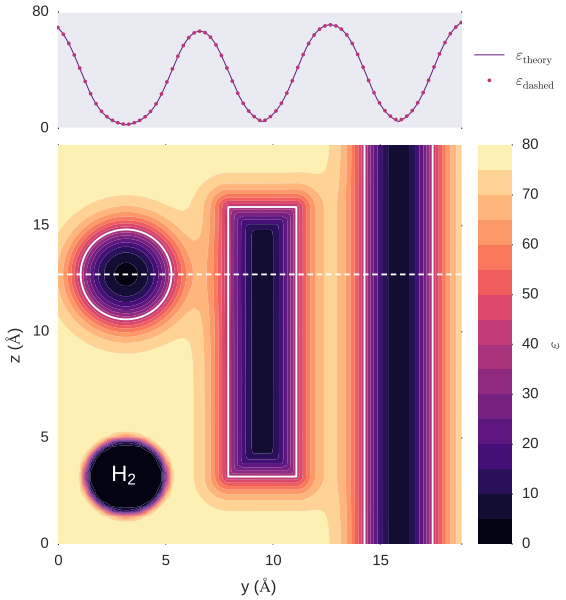


Figure A.3: A cross-section of the dielectric permittivity ε for a simulation cell containing a spherical, a rectangular, and a cylindrical smoothed dielectric exclusion region, as well as a H_2 molecule. The edges of the exclusion regions are shown with solid white lines. Above the main figure is the profile of the permittivity along the dashed line as given by ONETEP ($\varepsilon_{\text{dashed}}$) as well as the theoretical result for a smearing length of $r_0 = 1\,a_0$.

solver fails when it encounters a discontinuity in the permittivity. Hard-walled exclusion regions are perfectly adequate for, say, a carbon nanotube, where it would be straightforward to use a hard-walled cylinder to exclude implicit solvent from within the tube. However, this is not possible for our irregular cluster model of the OEC without being very conservative with the exclusion region. Instead, I added smoothing to the boundary of any exclusion region *i.e.* allowing for multiple exclusion regions we have

$$\varepsilon = \varepsilon_\infty - (\varepsilon_\infty - 1)\min\left\{1.0, \sum_S f_S(\mathbf{r})\right\} \tag{A.96}$$

where the smoothing of an exclusion region S is given by a Fermi-Dirac distribution centred on the exclusion region surface and with some characteristic smoothing width r_0 *i.e.*

$$f_S(r) = \frac{1}{1 + \exp(r/r_0)} \tag{A.97}$$

where r is the distance to the nearest exclusion region surface, is positive if outside the region, and negative inside. The result is demonstrated in Figure A.3, where a series of different solvent exclusion regions were tested in a simulation cell containing a lone H_2 molecule. The smearing is clearly visible, and the Poisson module could successfully solve this system. (If these regions were hard-walled, the solver would crash.) This functionality was implemented in ONETEP v4.5.15.19. The result of applying a spherical, smoothed exclusion region to the OEC cluster model is shown in Figure A.2b: it expelled the pockets of implicit solvent within the cluster.