Local correlation methods for periodic systems

Andrew Zhu

Magdalen College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2025

Declaration

This thesis describes work undertaken in the Department of Chemistry, University of Oxford, between October 2021 and September 2025, under the supervision of Prof. David Tew. The work in Chapters 3-6 contributed to the following publications:

- Zhu, A; Tew, D.P; J. Phys. Chem. A 2024, 128, 39, 8570–8579[1]: *Chapter 3*
- Nejad, A; Zhu, A; Sorathia, K; Tew, D.P; J. Chem. Phys. 2025, 163, 214107[2]: *Chapters 4,5,6*
- Zhu, A; Nejad, A; Komonvasee, P; Sorathia, K; Tew, D.P; J. Chem. Phys. 2025, 163, 214108[3]: *Chapters 4,5,6*

This thesis includes contributions from Dr. Arman Nejad (postdoctoral research fellow- Tew Group) and Poramas Komonvasee (Part II student- Tew Group). Their contributions are outlined in the beginning of each chapter, but also summarised here:

- *Chapter 5:* Arman Nejad developed the formalism and the implementation for the surface-dipole corrected, chargeless local density fitting approach, within the BvK-DLPNO-MP2 scheme, presented in Ref 2.
- *Chapter 6:* Arman Nejad performed the calculations using BvK-DLPNO-MP2, as well as the molecular fragment calculations, presented in Ref 2 and Ref 3
- *Chapter 7:* Poramas Komonvasee performed the calculations presented in this chapter, and participated in the analysis of the results.

All other content within this thesis is my original work.
Andrew Zhu
September 2025

Acknowledgements

First and foremost, I must thank David Tew for his support and supervision over the last four years. Without doubt, David has been one of the most influential figures in the entirety of my education, and I feel very grateful to have been able to develop all the skills that I now possess, thanks to his tutelage. I hope to carry his meticulous and reasoned approach with me in my future work as a researcher.

I would also like to thank Kesha, Arman, and Poramas, who have all been a pleasure to work with and learn from, in our joint task of obtaining periodic DLPNO-MP2. I will always be thankful for Kesha's patience in guiding me through basic solid state physics and quantum chemistry, at the beginning of my DPhil. Arman has been a source of both inspiration and humour, as we concurrently faced the trials and tribulations of implementing tricky code within TURBOMOLE. I am particularly appreciative of his advice on careers as I begun to consider postdoctoral work. And finally, to Mos- thank you for being an awesome Part II student- I wish you a highly successful DPhil in the Tew group.

It has been a fantastic experience working in the Tew group, all of whom I am fond of. I am looking forward to finding out what all of us get up to in the future. Particular thanks must go to Alex, Nick and Jack, who've had to deal with my constant, often non-coherent, ramblings. You have made working in the office very fun, and thesis writing has been unexpectedly relaxed thanks to you all.

As I now reflect on my wider life during the DPhil, I am firstly grateful to all the wonderful people who I have lived with- Anna, Andrew, and Guopeng, Antonio, Sharvi. Your friendship will always be deeply valued, and I'm glad to have learnt so much from you all, over dinners and BBQs.

It would be remiss to not acknowledge the role of everyone in Oxford University Badminton Club who've I've been so lucky to spend time with. It is no exaggeration to state that the badminton club has been my life outside of chemistry, and I'll even ashamedly admit now that OUBaC was a significant factor in wanting to stay at Oxford for my DPhil. I truly did not expect this sport to reveal so many close friends when I started Oxford in 2017. It will always have a place in my heart.

To mum, dad, and my younger brother Leon- thank you for your love and encouragement. No matter where you all are in the world, I am grateful to feel your presence every single day. And finally- to Sophie- thank you. Knowing you has been the highlight of my time here in Oxford.

Contents

Abstract **vii**

List of Figures **viii**

List of Tables **xi**

1 Motivation and Outline **1**

2 Background for Periodic Quantum Chemistry **6**
2.1 The Periodic Hamiltonian 6
2.2 Exploiting Translational Symmetry 8
2.2.1 Bloch's Theorem 9
2.2.2 Periodic Basis Functions 10
2.3 Periodic Hartree–Fock Theory 11

3 Local Wannier Functions **14**
3.1 Introduction 14
3.2 Theory 16
3.2.1 Review of Wannier Functions 16
3.2.2 Generalised Localised Wannier functions 17
3.2.3 Bloch Intrinsic Atomic Orbitals 19
3.2.4 Localisation Procedure 22
3.2.4.1 Diabatic Wannierisation 22
3.2.4.2 Optimisation of PM metric 24
3.3 Results and Discussion 25
3.3.1 Computational Details 25
3.3.2 Overall performance 27
3.3.3 Separate optimisation of valence and core bands 30
3.3.4 Chemical intuition of Bloch IAO generated WFs 32
3.4 Conclusions 34

4 Periodic DLPNO-MP2 Theory **35**
4.1 Correlated wavefunction methods 35
4.1.1 Configuration Interaction 36
4.1.2 Coupled Cluster 37
4.1.3 Møller–Plesset perturbation theory 38
4.2 Local Correlation Methods 39

4.3 Local MP2 for Periodic Systems . . . 43
4.4 DLPNO-MP2 formalism . . . 46
4.4.1 PAOs . . . 46
4.4.2 OSVs . . . 47
4.4.3 PNOs . . . 48
4.4.4 DLPNO-MP2 Residuals and Energy . . . 49

5 Implementing Periodic DLPNO-MP2 51
5.1 Overview of scheme . . . 51
5.2 Challenges with Coulomb lattice summation . . . 53
5.3 Strategy 1: Megacell embedding . . . 55
5.3.1 Translational symmetry . . . 57
5.4 Strategy 2: 'Born–von-Karman' treatment . . . 60
5.4.1 Chargeless Density Fitting . . . 60
5.4.2 Surface-dipole correction . . . 61
5.5 Conclusions . . . 64

6 Periodic DLPNO-MP2 Results 65
6.1 Computational details . . . 65
6.2 Unit cell energies through molecular fragments . . . 66
6.3 Comparison between 'Megacell' and 'BvK' schemes . . . 68
6.3.1 PNO and k-mesh convergence . . . 68
6.3.2 Computational cost comparison . . . 74
6.4 Scaling of Megacell-DLPNO-MP2 . . . 76
6.5 Applications with Megacell-DLPNO-MP2 . . . 78
6.6 Conclusions . . . 80

7 Surface Adsorption using DLPNO-MP2 82
7.1 Introduction . . . 82
7.2 Methods . . . 84
7.2.1 Megacell-DLPNO-MP2 . . . 84
7.2.2 Evaluation of Adsorption Energy . . . 85
7.2.3 Computational Details . . . 86
7.3 Towards to the dilute coverage limit . . . 88
7.3.1 Dilute Regime Results . . . 89
7.4 Probing dense coverages . . . 93
7.4.1 Dense Coverage Results . . . 95
7.5 Conclusions . . . 98

8 Concluding Remarks 101

Appendices

A Additional details on Wannier Localisation **104**

B List of code changes for periodic DLPNO-MP2 **105**

C Unit cell geometries for Chapter 6 **109**

D Basis sets used for Chapter 7 **111**

References **116**

Abstract

Wavefunction-based electronic structure methods, which accurately evaluate the electronic energy of chemical systems, have been increasingly applied to periodic or crystalline materials. Here, the requirement to simulate the in-principle infinite extent of the bulk material significantly exacerbates the computational cost of employing these methods, necessitating the development of reduced-cost algorithms specifically for periodic systems. Domain-based pair natural orbital local correlation (DLPNO) theory has been widely adopted in molecular contexts, and achieves near-linear scaling of computational effort with system size with only modest loss in accuracy by replacing Hamiltonian integrals and excitation amplitudes with low-rank approximations that exploit the inherent locality of electron correlation within non-conducting systems.

In this thesis, we extend DLPNO theory to periodic systems, in order to obtain accurate unit cell electronic energies using Møller–Plesset second order perturbation theory (DLPNO-MP2). Using the existing machinery within the TURBOMOLE quantum chemistry package, we present two complementary implementations for periodic DLPNO-MP2, which arise from different choices to account for the infinite summation of lattice images contained within periodic Coulomb integrals. Proof-of-concept calculations, demonstrating the correct convergence to the infinite bulk limit are shown for both schemes, using a range of one-, two- and three-dimensional systems, and further analysis is conducted comparing the rate of convergence and the computational scaling. The pilot approach of one scheme, in particular, already demonstrates linear and sub-linear scaling with respect to supercell size.

Localised occupied orbitals are an essential starting input to fully leverage the computational savings afforded through DLPNO theory. This thesis also outlines a novel scheme to generate localised Wannier functions, by adapting the molecular intrinsic atomic orbital approach to periodic systems. Finally, in an effort to demonstrate the computational efficiency afforded by periodic DLPNO-MP2, we apply our approach to study surface adsorption interactions, involving calculations with up to 30000 basis functions within the correlation treatment.

List of Figures

2.1 Example crystal lattice structures, with a single unit cell shown on the left, and a $2 \times 2 \times 2$ supercell shown on the right. The three-dimensional unit cell is defined by non-coplanar lattice vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$. 7

3.1 Bloch IAO localised Wannier functions of (4,4) nanotube, showing σ (left) and π (right) bonding character. An isosurface value of 0.05 was used. 27

3.2 Bloch IAO localised Wannier functions of the MgO(001) CO adsorption system. A magnesium 2p-like orbital is shown (left), as well a π bonding orbital on CO (centre). A WF demonstrating oxygen 3p orbital donation, on the MgO surface, to the CO π^* orbital, is also presented (right). Only atoms in the reference cell are shown. An isosurface value of 0.01 was used. 31

5.1 Schematic outlining overall key steps of the DLPNO-MP2 method within the TURBOMOLE software package. A detailed flow chart of the PAO-OSV-PNO cascade to construct PNOs are presented in pink. A summary of adaptations to obtain a periodic scheme are given in green. 52

5.2 Wannier functions for a two-dimensional system defined by a simulation grid consisting of 5x5 unit cells. Dotted lines represent the extent of the WF, beyond which it has sufficiently decayed. The left subfigure shows a pair of WFs that are fully supported within the simulation grid, the middle subfigure shows a pair of WFs that are not. The right subfigure defines the boundaries of the supercell and megacell; all WFs in the supercell are properly represented and form a translationally symmetric orthonormal set that can be used for the correlation treatment. 56

5.3 The PAO domain for pair $i_{\mathbf{0}}j_{\mathbf{l}}$ spans the entire megacell. The deeper shaded orange region shows the range of the PAOs that are fully supported within the megacell. PAOs within the lighter shaded orange regions, such as $\tilde{\mu}_{\mathbf{m}}$, are missing AOs such as $\mu_{\mathbf{n}'}$ to represent them, but the contributions of these AOs to the PAO DF integrals are negligible due to minimal overlap with $i_{\mathbf{0}}$. 59

5.4 Direct lattice summation for $(P_{\mathbf{m}}|Q_{\mathbf{l}})$ with dipole correction. 62

6.1 Molecular cluster fragments that accommodate 1D and 2D BvK supercells (dotted box). The supercells of clusters with sum formula $C_{2l}H_{2l+2}$ and $B_{l^2}N_{l^2}H_{4l}$ are comprised of $(l-2)$ and $(l-2)^2$ unit cells (red box), respectively. 67
6.2 Megacell (circles) and BvK (squares) DLPNO-MP2 correlation energies of C_2H_2 as a function of 1D k-mesh and $\mathcal{T}_{\mathrm{PNO}}$. The horizontal line indicates the canonical thermodynamic limit[2]. 69
6.3 Megacell (circles) and BvK (squares) DLPNO-MP2 correlation energies of C_2HF as a function of 1D k-mesh and $\mathcal{T}_{\mathrm{PNO}}$. The horizontal line indicates the canonical thermodynamic limit[2]. 69
6.4 Megacell (circles) and BvK (squares) DLPNO-MP2 correlation energies of BN as a function of 2D k-mesh and $\mathcal{T}_{\mathrm{PNO}}$. The horizontal line indicates the canonical thermodynamic limit[2]. 70
6.5 Hartree–Fock energy errors compared to the thermodynamic limit as a function of **k**-mesh size, compared with the respective MP2 correlation energy errors for BvK-DLPNO-MP2 (left) and Megacell-DLPNO-MP2 (right). 72
6.6 MP2 correlation energies for MgO, obtained with increasing PNO truncation thresholds and larger supercells. Extrapolations to the complete PNO space limit for each supercell size are given, and an inverse volume extrapolation to the thermodynamic limit is plotted. 74
6.7 Log scaled CPU times for two-dimensional Boron Nitride (left), and three-dimensional Lithium Hydride (right), as a function of log scaled supercell size and PNO truncation threshold. Slope values for the trend lines of the largest supercell sizes as given as m = [slopevalue] 76
6.8 Log scaled CPU times for two key subroutines in the Megacell-DLPNO-MP2 method, as a function of the supercell extent in each dimension, and PNO truncation threshold. The left panel shows the generation of the three index density-fitting integrals, whilst the right panel presents the construction of the approximate external pair density, and the subsequent diagonalization to construct the PNOs. Circle markers denote timings from three-dimensional Lithium Hydride. Square markers denote timings from two-dimensional Boron Nitride. 77
6.9 Supercells of the largest three-dimensional systems explored in this contribution. From left to right: 7x7x7 Magnesium Oxide (rock salt), 5x5x5 Hexagonal Ice, 5x5x5 Silicon (diamond cubic). The number of basis functions spanned within each supercell is 12691, 15000 and 5500, respectively. 79

7.1 Unit cell structures for CO adsorption on an MgO surface corresponding to coverage ratios of $\frac{1}{4}$ and $\frac{1}{9}$. Individual unit cells are depicted on the left, whilst 3^2 supercells are plotted on the right. The reference cell CO molecule is highlighted in white. 85
7.2 The contributions to the adsorption energy of CO on the MgO surface, for any given coverage ratio. The lateral interactions of the gas phase CO molecules must be added to E_{int} to properly distinguish adsorption energies at different coverage ratios. 87
7.3 Hartree–Fock (Top) and MP2 correlation energies (Bottom) for $E_{\text{int,coh}}$, at three dilute coverage ratios. Results using the modified 'DZ' and 'TZ' basis sets are shown, and complete basis set estimate is provided for the MP2 correlation energies. 91
7.4 CO adsorption energies on the MgO surface at varying coverage ratios, up to full monolayer coverage. Hartree–Fock and MP2 correlation energy contributions are plotted as '+' and 'x's, respectively. Results employing the 4×4 surface slab unit cell are also reported for the two most dilute coverage ratios. 95
7.5 2×2 and 4×4 surface slab unit cells used to probe adsorption coverage ratios towards the dense regime. 96

List of Tables

3.1 Insulating and semiconducting systems used to test the Bloch IAO and PM localisation procedure 26
3.2 PM metric values of WFs from the initial SCF calculation, after rotation into the natural gauge, after diabatic Wannierisation and after PM optimisation 27
3.3 Number of iterations to converge PM metric, after initial diabatic preparation, using l-BFGS or steepest ascent. 28
3.4 Number of iterations to converge PM metric with initial diabatic preparation, with core and valence band separation 30
3.5 Bloch IAO partial charges of the non-interacting and equilibrium MgO(001) CO adsorption system 33

5.1 Convergence properties of Coulomb lattice sums between multipole contributions to charge densities $\sum_{\mathbf{L}}(\mathring{\rho}^A|\mathring{\rho}^B_{\mathbf{L}})$. 54

6.1 HF and BvK-DLPNO-MP2 ($\mathcal{T}_{\text{PNO}} = 10^{-7}$) energies per LiH unit (Hartree) for a 3D LiH lattice with two different unit cells. 73
6.2 MP2 correlation energy comparison between Megacell- and BvK-DLPNO-MP2 implementations, varying PNO truncation threshold and supercell size. 73
6.3 Survey of three-dimensional materials probed using Megacell-DLPNO-MP2, given by PNO truncation threshold and largest supercell size employed. Hartree–Fock energies are provided for each supercell. Only valence bands were included in the correlated treatment. All energies given in Hartrees. 80

7.1 Hartree–Fock and MP2 correlation energies for each system involved in evaluating $E_{\text{int,coh}}$, comparing supercell sizes of 3^2 and 5^2. All energies are given in Hartrees, apart from the final row. Calculations employed the 2×2 surface slab unit cell, using the modified 'TZ' basis set. 90
7.2 Comparison of adsorption energies of CO on the MgO (001) surface using Megacell-DLPNO-MP2 and recent results from the literature. 'cl' and 'pbc' refer to cluster and periodic wavefunction schemes, respectively. Experimental results from temperature-programmed desorption (TPD) experiments are also presented. 93

7.3 Comparison of the full cohesive adsorption energy ($E_{\mathrm{int,coh}}$), and the adsorption energy neglecting gas phase CO lateral interactions(E_{int}), at different coverage ratios. 98

1| Motivation and Outline

Computational modelling of the electronic structure of materials is a field of growing importance. Theoretical methods to simulate and solve the Schrödinger equation for extended systems aid in the structure prediction of novel materials, and the evaluation of electronic properties reveals insight into key characteristics of the material, such as semi-conducting[4,5] behaviour, superconductivity[6–9], or magnetic effects[8,10,11]. Amongst the myriad of material properties that are dependent on the electronic structure, a key subset are *ground state properties*. These include structural properties such as lattice constants and equilibrium geometries[12–15]; energetic concepts such as cohesive or formation energies[13–15], phase stabilities[16,17], or even adsorption energies[18–25] for two-dimensional surfaces; and mechanical properties such as bulk modulus[14,15]. This thesis is concerned with methods which broadly aim to probe this subset.

To date, density functional theory[26,27] (DFT) has been the predominant workhorse for electronic structure studies of materials[28,29]. The wealth of research for DFT has largely been aimed towards to the design of increasingly accurate functionals, which map electron densities to energies. In particular, the exchange-correlation (XC) component of the density functional is not known exactly, and has led to the establishment of a rough hierarchy of approximations, featuring a plethora of functionals at each rung of Jacob's ladder[30]. The key attraction of DFT is its favourable cost to accuracy ratio; the original formulation of Kohn-Sham DFT[26,27] has a formal scaling of $\mathcal{O}(n^3)$ with system size, with several reduced cost or linear-scaling approaches[31–33].

Despite this, there are inherent issues to DFT. The development of functionals has historically been guided by the realisation that a particular property was not accurately calculated by then-current functionals, leading to subsequent improvements based on those specific benchmarks. For example, a key issue for DFT has been to capture long-range correlation or dispersion interactions. For

materials, these manifest in the form of non-covalent interactions, such as hydrogen bonding or Van der Waals (VdW) dispersions, which are key to describing systems such as molecular crystals or surface adsorption systems. Significant work has been conducted, involving the development of XC functionals specifically including correction terms for dispersion[23,34–38]. Another long-standing issue is the self interaction error, where use of approximate functionals lead to a unphysical self repulsion of the electrons, which has implications such as underestimated band gaps[39,40]. This issue has been remedied through functionals that incorporate varying degrees of Hartree–Fock exchange[28,41,42]. As a result of the myriad of functionals that now exist, which were developed with differing motivations or with different target benchmarks, a inherent criticism of DFT is that whilst many ground state properties can now be accurately captured by a given functional, there exists no clear framework of functionals which deliver systematic improvement to all chemical properties.

The existing shortcomings of DFT for materials necessitate the development of other, more accurate, computational schemes, where improvements in accuracy can be more systematically defined. These schemes are important in their own right, but also enable robust benchmarks to be formed, and also can be incorporated into the development of novel density functional approaches. Within this field of alternate methods, one key subset are 'wavefunction' schemes, which include Hartree–Fock (HF), Møller-Plesset perturbation (MP) and coupled cluster (CC) theories[43,44]. Other schemes, such as Green's function methods[45–48] or quantum Monte Carlo[49,50], also play significant roles in the simulation of ground and excited state properties of materials. All together, these schemes, alongside DFT, contribute to the richness of computational materials science, competing and complementing each other in the pursuit to improve cost to accuracy ratio.

This thesis is concerned with *correlated wavefunction* or *post-Hartree–Fock* schemes, which build upon a HF reference state to evaluate accurate electronic energies. Correlated wavefunction schemes have had remarkable success within molecular quantum chemistry, and application of these methods to solid-state physics has grown significantly in this century[15,19–22,44,51–77]. In order to generalise

such methods to periodic systems, two main strategies are pursued: embedded or finite cluster approaches, and periodic boundary condition (pbc) methods. The former strategy models periodic systems using finite fragments of the material, which are usually solved to high accuracy with post-HF schemes, often leveraging the same machinery for molecular calculations. The surrounding region of material is often represented at a computationally cheaper level of theory, ranging from point charges to quantum embedding theories[19,20,66,67,70–77]. In contrast, pbc schemes utilise translational symmetry to model unit cells, formally accounting for the infinite collection of cells surrounding the reference unit cell, at the same level of theory, in order to compute an electronic energy per unit cell[15,22,44,51–65]. A key complication with this approach is dealing with the infinite summation over lattice points, which is now present in all the Hamiltonian integrals.

A key concept that is pertinent to solid-state quantum chemistry is the requirement for schemes to converge towards the thermodynamic limit. Cluster approaches and pbc methods both inherently contain a finite-size error, from the limited extent of the cluster or cell size employed, compared to the infinite extent of the material. Therefore, converging to the thermodynamic limit can be achieved through increasing the size of the embedded fragment, or by including more unit cells into the simulation, respectively. Viable schemes for periodic systems should thus have favourable scaling with respect to these parameters.

In this thesis, we focus predominantly post-HF schemes employing periodic boundary conditions. Whilst significant contributions towards the field are now too numerous to outline in detail, we choose to highlight the pioneering work of the CRYSTAL[78] and CRYSCOR[51] codes as early general-purpose packages for Hartree–Fock and post-HF schemes, respectively. A common challenge for all periodic post-HF methods is the expensive computational scaling of the original schemes. Second-order Møller-Plesset perturbation theory (MP2) has a canonical scaling of $\mathcal{O}(n^5)$ with respect to system size, and common variants of CC theory have even more expensive scaling. Compared to DFT, computational bottlenecks would be quickly encountered whenever larger or more complex unit cells are

employed, or when increasing the number of simulated unit cells towards the thermodynamic limit. Domain-based pair natural orbital local correlation[79,80] (DLPNO) theory offers a solution to this problem. DLPNO theory achieves near linear scaling of computational effort with system size with only modest loss in accuracy by replacing Hamiltonian integrals and excitation amplitudes with low-rank approximations that exploit the inherent locality of electron correlation in insulators. For each electron pair, a bespoke set of virtuals are constructed to efficiently capture the pair correlation energies. Whilst DLPNO theory has shown to be highly successful in molecular contexts, to our knowledge, a fully periodic DLPNO scheme has yet to be realised for materials.

This thesis presents work on the development of a local correlation DLPNO code, at MP2 level of theory, for periodic systems. We work directly within the TURBOMOLE[81,82] software package, leveraging the existing periodic Hartree–Fock[83–86] and molecular DLPNO[87–90] codes, to develop two complementary pilot implementations for DLPNO-MP2 with periodic boundary conditions. Convergence results, particularly to the thermodynamic limit, are detailed, and an additional study into surface adsorption applications is also included.

This thesis is structured as follows: Chapter 2 provides a concise overview of the theoretical background for quantum chemistry of periodic systems. In Chapter 3, we outline a novel scheme to obtain localised Wannier functions, which are a necessary prerequisite for periodic DLPNO methods, presenting results on the convergence performance of the localisation scheme. Chapter 4 outlines the theoretical formulation of DLPNO-MP2 theory under periodic boundary conditions. The specific differences in implementation within the TURBOMOLE package, leading to two different pilot schemes of periodic DLPNO-MP2, are then discussed in Chapter 5. Detailed results analysing both schemes are then presented in Chapter 6, where convergence to the thermodynamic limit is discussed, using robust benchmarks, and comparison of the computational efficiencies is conducted. In Chapter 7, we apply our periodic DLPNO-MP2 scheme to study the adsorption energy of carbon monoxide on a magnesium oxide surface, comparing our results to

previous high accuracy correlated wavefunction studies, and extending this model system to more complex scenarios. A summary of the work in this thesis, as well as discussion of future directions, is given in Chapter 8.

2| Background for Periodic Quantum Chemistry

In this chapter, we aim to provide a brief overview of the foundational concepts within periodic quantum chemistry that this thesis builds upon. By no means does this chapter aim to be a complete introduction to all the necessary principles. Instead, we present a concise narrative of the key concepts encountered when solving the periodic Hamiltonian, drawing in parallels from molecular electronic structure theory where necessary, in order to establish the key notation and ideas used frequently in the later chapters. The periodic Hamiltonian is first introduced in Section 2.1, before the key theme of leveraging translational invariance is outlined, where ideas such as k-space and Bloch's theorem are discussed, in Section 2.2. We then finally conclude with a brief discussion of periodic Hartree–Fock theory.

2.1 The Periodic Hamiltonian

Within this thesis, we are interested in solving the Hamiltonian under the Born–Oppenheimer approximation, where the the motion of the nuclei is deemed to be fixed relative to that of the electrons. Under this approximation, the focus is solely on the electronic component of the Hamiltonian,

$$\hat{H} = \hat{T}_{\mathrm{e}} + \hat{V}_{\mathrm{en}} + \hat{V}_{\mathrm{ee}}, \tag{2.1}$$

consisting of the kinetic energy of the electrons, the nuclear-electron Coulombic attraction, and the electron-electron Coulombic repulsion, respectively. We are concerned about the solutions of this electronic Hamiltonian for a periodic, infinitely extending material, which invokes a number of complications compared to modelling molecular systems.

We represent periodic materials as an infinitely repeating array of unit cells, defined by (in three-dimensional cases) primitive lattice vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$, as demonstrated in Figure 2.1. A periodic unit cell can thus be described by the vectors

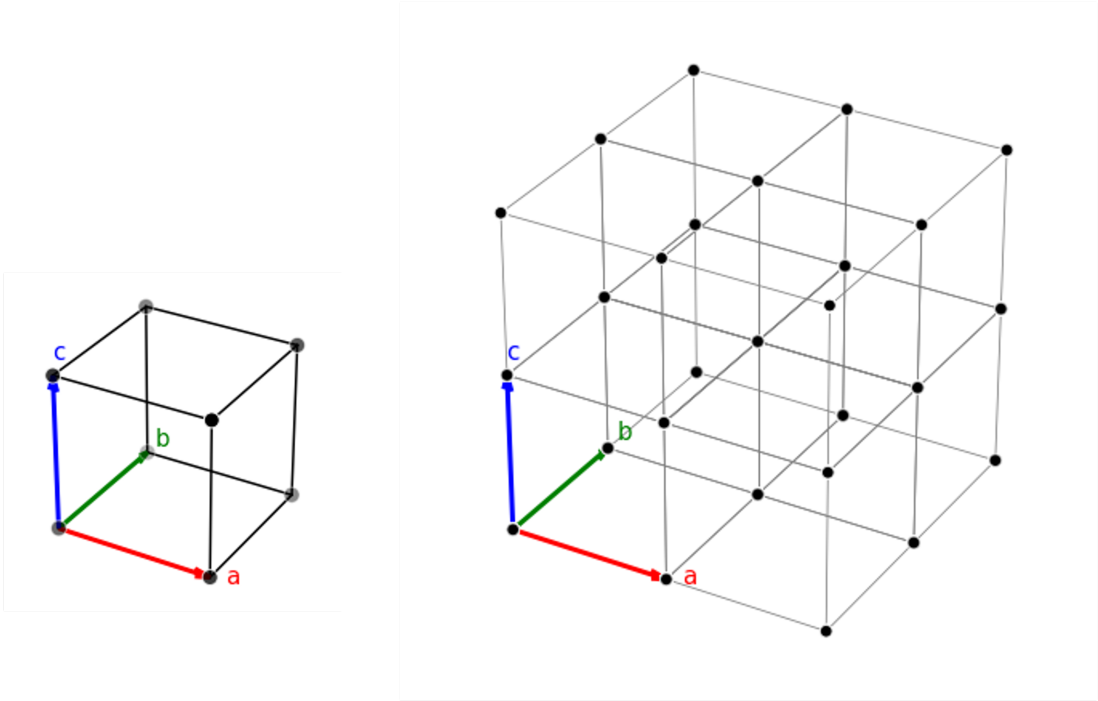

Figure 2.1: Example crystal lattice structures, with a single unit cell shown on the left, and a $2 \times 2 \times 2$ supercell shown on the right. The three-dimensional unit cell is defined by non-coplanar lattice vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$.

which translate from one lattice point to another,

$$\mathbf{l} = n_\mathrm{a}\mathbf{a} + n_\mathrm{b}\mathbf{b} + n_\mathrm{c}\mathbf{c}, \tag{2.2}$$

where $n_\mathrm{a}, n_\mathrm{b}, n_\mathrm{c} \in \mathbb{Z}$. We wish to evaluate the electronic energy for a single unit cell, which is an intensive quantity, but solving the Hamiltonian requires, in principle, accounting for the interactions of the electrons and nuclei situated in the infinite array of cells surrounding the central unit cell in question. Our Coulomb operators in Eq. 2.1, for a unit cell of M nuclei with charges Z, and N electrons, can be represented as,

$$\hat{V}_\mathrm{en} = -\sum_{n}^{N}\sum_{A}^{M}\sum_{\mathbf{l}}^{\infty}\frac{Z_A}{|\mathbf{r}_n - \mathbf{R}_A - \mathbf{l}|} \tag{2.3}$$

$$\hat{V}_\mathrm{ee} = \frac{1}{2}\sum_{n,n'}^{N}\sum_{\mathbf{l}}^{\infty\prime}\frac{1}{|\mathbf{r}_n - \mathbf{r}_{n'} - \mathbf{l}|}. \tag{2.4}$$

The position vectors of the electrons and nuclei are given by $\mathbf{r}_n$ and $\mathbf{R}_A$, respectively, and the primed summation excludes self interactions when $\mathbf{l} = 0$. Atomic units are employed.

Treatment of the infinite lattice summations in $\hat{V}_{\text{en}}$ and $\hat{V}_{\text{ee}}$ are difficult for a number of reasons. The harmonic series here are formally divergent, and treatments that enforce charge neutrality to obtain convergent, intensive values are required, and secondly, the long range interaction of the Coulomb interaction means these sums are slow decaying. The Ewald summation method[91] is a well-known scheme used to solve this issue, in which the Coulomb interaction is divided into short and long range components, which are evaluated as quickly converging absolute sums, in real and reciprocal space, respectively. We do not elaborate on the wealth of literature focused upon this issue any further, apart from a specific case within our own methodology, discussed in Chapter 5. Instead, we now discuss how leveraging translational invariance can provide additional insights in the solutions of this periodic Hamiltonian.

2.2 Exploiting Translational Symmetry

Simulation of an infinitely repeating array of unit cells is difficult, so a common framework is to utilise a *supercell*, consisting of a finite collection of unit cells. For simplicity, we present a one-dimensional system, consisting of N_{a} unit cells, giving a total supercell lattice vector of length $N_{\text{a}}a$. The periodicity of the solid is incorporated through boundary conditions, where any periodic function obeys

$$f(\mathbf{r}) = f(\mathbf{r} + N_{\text{a}}\mathbf{a}). \tag{2.5}$$

These are known as *Born–von-Karman* (BvK) boundary conditions. We introduce a lattice translation operator, which translates a function by a unit cell vector, $\hat{\mathcal{T}}(\mathbf{a})f(\mathbf{r}) = f(\mathbf{r} - \mathbf{a})$. Repeated application of this operator N_{a} times yields

$$\hat{\mathcal{T}}^{N_{\text{a}}}(\mathbf{a})f(\mathbf{r}) = \hat{\mathcal{T}}(N_{\text{a}}\mathbf{a})f(\mathbf{r}) = f(\mathbf{r}). \tag{2.6}$$

This reveals insight into the form of the translation operator, which can be expressed as,

$$1 = e^{iN_{\text{a}}a} = e^{iN_{\text{a}}a\cdot\frac{2\pi n}{N_{\text{a}}a}} = e^{iN_{\text{a}}ka}. \tag{2.7}$$

Here, the first equality reflects the BvK boundary conditions. The cyclic nature of the complex exponential leads to the second equality, where $n \in \mathbb{Z}$, which can be combined in terms of a parameter k, which has dimensions of inverse distance, and is often labelled the crystal momentum. This provides a representation of periodic functions in reciprocal space. The values $k = \frac{2\pi n}{N_\mathrm{a} a}$ are the lengths of reciprocal lattice vectors. Furthermore, there only exists a restricted set of unique k vectors, again as a result of the cyclic periodicity,

$$e^{-ik'a} = e^{-i(k+\frac{2\pi}{a})a} = e^{-ika}, \tag{2.8}$$

meaning that the representation of the periodic system within reciprocal k-space can be restricted to $-\frac{\pi}{a} \leq k < \frac{\pi}{a}$, in a region known as the first Brillouin zone.

Returning to our supercell picture in real space, it can be seen that an increasingly accurate representation of the infinite solid, towards the thermodynamic limit, can be obtained by increasing the supercell size N_a. In doing so, errors associated with the finite size of the supercell are converged out. This directly determines the number of k vectors in the first Brilluoin zone, which must also be N_a. This is an important relationship: the supercell size we choose to employ in our simulations is equivalent to choosing the discretization of the first Brillouin zone.

In the following sections, we explore how the parametrising a periodic system in k-space can aid in characterising the solutions to the Hamiltonian, and how exploiting translational symmetry aids in the evaluation of such eigenstates.

2.2.1 Bloch's Theorem

Since the Hamiltonian has the periodicity of the lattice, it must commute with the translation operator, which implies a shared eigenbasis. The eigenstates of the periodic Hamiltonian can therefore be characterised by considering the action of the translation operator. From Eq. 2.7, we now understand the effect of the lattice translation operator as imparting a phase on our original function,

$$\hat{\mathcal{T}}(\mathbf{a}) f_\mathbf{k}(\mathbf{r}) = f_\mathbf{k}(\mathbf{r} - \mathbf{a}) = e^{-i\mathbf{k}\cdot\mathbf{a}} f_\mathbf{k}(\mathbf{r}). \tag{2.9}$$

A suitable form for $f_{\mathbf{k}}(\mathbf{r})$, which ensures it is an eigenstate of the translation operator, and thus the Hamiltonian, is

$$f_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}). \tag{2.10}$$

This a common form of Bloch's Theorem: the eigenstate consists of a plane wave modulating a function $u_{\mathbf{k}}(\mathbf{r})$, which itself has the periodicity of the lattice. The crystal momentum vector $\mathbf{k}$ parametrises the eigenstate. Within this thesis, we use the term *Bloch functions* or *crystal orbitals* to refer to these eigenstates.

In analogous fashion to molecular quantum chemistry, the many-electron wavefunction solution to the Hamiltonian is constructed through a Slater determinant expansion, which respects Pauli's exclusion principle. The spin-orbitals that form the basis of the Slater determinants are the Bloch functions, in our periodic case.

2.2.2 Periodic Basis Functions

As with molecular quantum chemistry, the single-electron orbitals that construct the wavefunction are projected on a finite set of basis functions. This allows Hamiltonian integrals to be expressed within a computationally tractable, linear algebra framework. The energy obtained from solving the Hamiltonian is thus specific to a given basis set. There exists two main types of periodic basis functions: plane wave (PW) and periodic atom-centred functions.

Whilst an in-depth comparison of the two types of basis function is not presented here, we note that the PW representation has some distinct advantages. The quality and completeness of the basis set is controlled by a single kinetic energy parameter, allowing straightforward convergence to the complete basis set (CBS) limit[92]. Furthermore, the plane wave basis forms an orthonormal set, so Hamiltonian integrals are typically simpler to evaluate. In contrast, atom-centred or atomic orbital (AO) functions, often of the form of Gaussian-type orbitals (GTOs), often encounter linear dependency issues when employing more diffuse basis sets, and continued work is required to develop AO basis sets catered to periodic systems[93–95]. They are, however, a much more natural description of material systems featuring

localised or non-uniform electron densities, and typically much fewer basis functions are required for CBS convergence. AO type basis functions are thus suited for local correlation schemes, and the rest of this thesis works within this framework.

Within a BvK supercell, consisting of N unit cells, the periodic AO in real space is expressed as an infinite sum of non-periodic AOs,

$$|\mu_{\mathbf{l}}\rangle = \sum_{\mathbf{L}}^{\infty} |\mathring{\mu}_{\mathbf{l}+\mathbf{L}}\rangle \,. \tag{2.11}$$

Here, $|\mathring{\mu}_{\mathbf{l}+\mathbf{L}}\rangle$ is a single AO, directly equivalent to its molecular counterpart, where $\mathbf{l}$ is the lattice vector of the unit cell and $\mathbf{L}$ is the supercell lattice vector it sits in. Through discrete Fourier transform, an AO representation in k-space can be constructed,

$$|\mu_{\mathbf{k}}\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{l}}^{N} e^{i\mathbf{k}\cdot\mathbf{l}} |\mu_{\mathbf{l}}\rangle \,. \tag{2.12}$$

These functions fulfil Bloch's theorem, and hence are known as Bloch AOs. In the subsequent section, the Bloch AOs form the basis that enables translational symmetry to significantly reduce the cost of solving the Hamiltonian, which we illustrate through the example of Hartree–Fock theory.

2.3 Periodic Hartree–Fock Theory

The Hartree–Fock (HF) approximation provides a mean-field description of the electronic energy. Through variationally minimising the energy expectation value, the optimal spin orbitals are obtained, which are then used to construct a single Slater determinant as the HF wavefunction.

The central operator in HF theory is the single-particle Fock operator. For initial simplicity, we first present the Fock operator for an N electron system without notion of periodicity,

$$\hat{F} = \sum_{i}^{N} \hat{f}(i) \tag{2.13}$$

$$\hat{f}(i) = \hat{h}(i) + \sum_{j}^{N} (\hat{J}_j(i) - \hat{K}_j(i)), \tag{2.14}$$

where $\hat{h}(1)$ is the one electron operator for the kinetic energy and the electron-nuclear attraction for electron 1,

$$\hat{h}(1) = \frac{1}{2}\nabla_1^2 - \sum_A^M \frac{Z_A}{|\mathbf{r}_1 - \mathbf{R}_A|}. \tag{2.15}$$

$\hat{J}_j$ and $\hat{K}_j$ are the Coulomb and exchange operators,

$$\hat{J}_j(1)\psi_i(\mathbf{x}_1) = \psi_i(\mathbf{x}_1) \int \psi_j^*(\mathbf{x}_2) r_{12}^{-1} \psi_j(\mathbf{x}_2) \mathrm{d}\mathbf{x}_2, \tag{2.16}$$

$$\hat{K}_j(1)\psi_i(\mathbf{x}_1) = \psi_j(\mathbf{x}_1) \int \psi_j^*(\mathbf{x}_2) r_{12}^{-1} \psi_i(\mathbf{x}_2) \mathrm{d}\mathbf{x}_2, \tag{2.17}$$

where i, j index the spin orbitals and $\mathbf{x}_1$ is the combined spin and spatial coordinates for electron 1. Projection of the Fock operator onto a finite basis leads to the Roothan equation, which is a generalised eigenvalue problem, to obtain the orbital eigenvalues ϵ,

$$\mathbf{FC} = \epsilon \mathbf{SC}. \tag{2.18}$$

Here, $S_{ij} = \int \mathrm{d}\mathbf{x}_1 \psi_i^*(\mathbf{x}_1)\psi_j(\mathbf{x}_1)$ is the overlap matrix and $\mathbf{C}$ are the molecular orbital coefficients. The Fock matrix elements, $F_{ij} = \int \mathrm{d}\mathbf{x}_1 \psi_i^*(\mathbf{x}_1)\hat{f}(1)\psi_j(\mathbf{x}_1)$, is obtained through projection of Eq.2.13 onto a basis.

Returning to periodic systems, we now demonstrate why representation of the periodic Hamiltonian in k-space is useful. Any matrix representation of an operator which is invariant to lattice translations, such as the HF Fock operator, can be shown to have a block-diagonal structure when rotated into the Bloch basis. This is the entire motivation for employing this basis, since it allows the overall matrix problem to be decomposed and solved separately, per k-point.

Starting from a Fock matrix element in a basis of Bloch AOs, one can expand the element in terms of real space AOs, following Eq. 2.12,

$$\langle \mu_{\mathbf{k}} | \hat{F} | \nu_{\mathbf{k}'} \rangle = \frac{1}{N} \sum_{\mathbf{l},\mathbf{l}'} e^{-i\mathbf{k}\cdot\mathbf{l}} e^{i\mathbf{k}'\cdot\mathbf{l}'} \langle \mu_{\mathbf{l}} | \hat{F} | \nu_{\mathbf{l}'} \rangle . \tag{2.19}$$

By substituting the summation indices $\mathbf{l}'' = \mathbf{l} - \mathbf{l}'$, and noting that the real-space Fock matrix element depends only on the relative distance between $\mathbf{l}$ and $\mathbf{l}'$, we can

write

$$\langle \mu_{\mathbf{k}} | \hat{F} | \nu_{\mathbf{k}'} \rangle = \frac{1}{N} \sum_{\mathbf{l},\mathbf{l}''} e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{l}} e^{-i\mathbf{k}'\cdot\mathbf{l}''} \langle \mu_{\mathbf{l}''} | \hat{F} | \nu_{\mathbf{0}} \rangle . \tag{2.20}$$

This periodic Fourier identity[96] yields a Kronecker delta, leading to,

$$\langle \mu_{\mathbf{k}} | \hat{F} | \nu_{\mathbf{k}'} \rangle = \delta_{\mathbf{k},\mathbf{k}'} \sum_{\mathbf{l}''} e^{-i\mathbf{k}'\cdot\mathbf{l}''} \langle \mu_{\mathbf{l}''} | \hat{F} | \nu_{\mathbf{0}} \rangle . \tag{2.21}$$

As a result of the conservation of crystal momentum, all matrix elements of one-body operators between different k-points are necessarily zero, proving the block-diagonal structure of translationally invariant operators in this basis. The Roothan equations for periodic systems can thus be solved for each individual reciprocal lattice vector,

$$\mathbf{F}^{\mathbf{k}}\mathbf{C}^{\mathbf{k}} = \epsilon^{\mathbf{k}}\mathbf{S}^{\mathbf{k}}\mathbf{C}^{\mathbf{k}}. \tag{2.22}$$

Solution of these equations yields the band energies $\epsilon^{\mathbf{k}}$, as well as the Bloch functions. The occupied Bloch functions or crystal orbitals are typically expressed in terms of the Bloch AOs,

$$|i_{\mathbf{k}}\rangle = \sum_{\mu} C^{\mathbf{k}}_{\mu,i} |\mu_{\mathbf{k}}\rangle . \tag{2.23}$$

These are analogous to the canonical molecular orbitals, and typically are delocalised across the entire supercell.

Whilst various codes for periodic HF now exist, including the pioneering CRYSTAL[78] package, we choose to only briefly highlight the implementation within the TURBOMOLE[81,82] software package, housed in the `riper`[83–86,97,98] module. Within this scheme, the Coulomb term is computed by dividing the interaction into near-field and far-field components, which are solved quickly using density fitting and continuous fast multipole schemes, respectively[85]. The evaluation of the Fock exchange matrix is done entirely in real space, and utilises a minimum image convention to remove the singularity that occurs with finite k-meshes.[97]

3| Local Wannier Functions

The work in this chapter contributed to the publication Ref.1.

3.1 Introduction

Mean field theories, such as Hartree–Fock (HF) or Kohn-Sham density functional theory (DFT)[26,27], provide a description of the electronic structure of a system through an one-particle orbital model, giving insight to the bonding in molecules and band structure for materials. However, the canonical molecular orbitals (MOs) or Bloch functions, are typically delocalised across the entire system, and thus do not intuitively map to the interpretation of bonding in terms of overlap of atomic orbitals (AOs), which is a local picture. By applying unitary rotations to the occupied orbitals, one can obtain localised objects, commonly known as Wannier functions (WFs)[99] for periodic systems.

Obtaining a localised representation of the occupied space is a key first step in local correlation schemes, which exploit this locality to construct localised domains of virtuals. The work in this chapter thus sets a foundation for the entire periodic DLPNO-MP2 framework of this thesis, by outlining a scheme to generated localised occupied Wannier functions.

Methods to evaluate localised molecular orbitals have focused upon defining a localisation metric or functional; the stationary points of this functional thus correspond to localised orbitals. The two most commonly employed metrics are from Foster and Boys (FB)[100,101], and Pipek and Mezey (PM)[102], both of which have been adapted for periodic systems[103–108]. The FB metric, which minimises the spread of the orbitals, has seen widespread usage, namely through the Wannier90[104] package, which has now established interfaces with various periodic, plane-wave based, codes[109–112]. In contrast, the PM metric, which uses the Mulliken partial atomic charges, is naturally suited to codes employing localised basis sets, under a linear-combination-of-atomic-orbitals (LCAO) framework[81,82,113–116], where AO

coefficients are directly accessible and overlaps are easily computed. In addition, WFs localised with the PM metric produces orbitals with separate σ and π bonding character, giving advantages in chemical interpretation, as opposed to FB. The Mulliken charges, however, are unreliable for non-minimal basis sets. This issue arises from the near-redundancy of the LCAO expansion with large basis sets and is exacerbated in crystals. To avoid this problem, alternate partial charge definitions[106,107] have been utilised to obtain localised WFs in the spirit of the PM method. The intrinsic atomic orbital (IAO) method[117] is one partial charge estimate that has widely been applied for molecules schemes. Its appeal lies in its robustness against basis set choice, ease of computation, and full representation of the occupied space. In this chapter, we present work that generalises the IAO method to periodic systems, generating localised WFs in the spirit of the PM scheme using Bloch IAO partial charge estimates.

In Section 3.2.3 Bloch intrinsic atomic orbitals (Bloch IAOs) are introduced as the natural periodic extension of IAOs, and the overall optimisation scheme to generate localised WFs with Bloch IAOs is presented. By generalising the well-established IAO method to crystals, direct comparison of localised orbitals between periodic and molecular systems becomes feasible. The initial guess is a crucial step in the optimisation and we propose a simple and effective procedure for generating localised orbitals by defining a natural gauge and by constructing diabatic Bloch orbitals and diabatic Wannier functions. We then present and analyse performance and stability of the optimisation, with particular discussion of the solver's performance when separating core and valence bands, in Section 3.3. Finally, in Section 3.3.4, the chemical interpretability of Bloch IAOs are commented upon, using a surface adsorption system as an example.

3.2 Theory

3.2.1 Review of Wannier Functions

The Bloch functions, obtained as solutions of the periodic HF equations, are delocalised across the entire system. By superimposing the Bloch functions of a single band across the first Brillouin zone, a conventional Wannier function[99], centred on a unit cell given by lattice vector $\mathbf{l}$, is given by,

$$|i_{\mathbf{l}}\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}}^{N} e^{-i\mathbf{k}\cdot\mathbf{l}} |i_{\mathbf{k}}\rangle \,. \tag{3.1}$$

The WFs span the same space as their Bloch counterparts, with translational copies found in each unit cell.

Bloch functions are defined up to an arbitrary phase only. However, the spatial distribution of the resultant WFs are highly dependant on the relative phases of the contributing Bloch functions. The WFs are thus gauge variant. To obtain localised conventional WFs, the relative phases of the Bloch functions for each band must be optimised. By rotating the gauge such that the Bloch functions appear smooth in reciprocal space, the resulting WFs in real space are in turn localised, as a property of Fourier transforms,

$$|i_{\mathbf{l}}\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}}^{N} e^{-i\mathbf{k}\cdot\mathbf{l}} e^{i\theta_i^{\mathbf{k}}} |i_{\mathbf{k}}\rangle \,. \tag{3.2}$$

A natural gauge for each Bloch function can be defined by requiring that the scalar product of the coefficients between Bloch functions at $\mathbf{k}$ and the Γ-point $\mathbf{0}$ is real valued. By first computing the phase difference

$$\sum_{\mu} C_{\mu,i}^{*\mathbf{k}} C_{\mu,i}^{\mathbf{0}} = R e^{i\theta_i^{\mathbf{k}}}, \tag{3.3}$$

the Bloch functions can be rotated into their natural gauge $|i_{\mathbf{k}}^{\mathrm{n}}\rangle$, from their original gauge, $|i_{\mathbf{k}}^{\mathrm{o}}\rangle$, straightforwardly,

$$|i_{\mathbf{k}}^{\mathrm{n}}\rangle = |i_{\mathbf{k}}^{\mathrm{o}}\rangle \, e^{-i\theta_i^{\mathbf{k}}}. \tag{3.4}$$

For bands that have small dispersion, or minimal mixing, imposing the natural gauge is often sufficient to produce well-localised Wannier functions.

3.2.2 Generalised Localised Wannier functions

Generalised Wannier functions are defined by allowing Bloch functions from several bands to mix[103],

$$|i_{\mathbf{l}}\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}}^{N} \sum_{j} e^{-i\mathbf{k}\cdot\mathbf{l}} U_{j,i}^{\mathbf{k}} \; |j_{\mathbf{k}}\rangle \, . \tag{3.5}$$

By allowing mixing between bands, WFs can be further localised not only to a unit cell, but also to atomic sites within a unit cell. The resulting generalised WFs are even more strongly 'non-unique' than the conventional definition, and their locality is highly dependant on the choice of $U_{ji}^{\mathbf{k}}$. To ensure real-valued WFs, time-reversal symmetry about the Γ-point must be imposed between the Bloch functions, $|i_{\mathbf{k}}\rangle = (|i_{-\mathbf{k}}\rangle)^*$. The Bloch functions at the Γ-point are real, following convention, and the choice of unitary carries the constraint of $\boldsymbol{U}^{\mathbf{k}} = (\boldsymbol{U}^{-\mathbf{k}})^*$.

Localisation of WFs is achieved by varying $\boldsymbol{U}^{\mathbf{k}}$ to optimise a chosen locality metric. The Foster–Boys[100,101] and Pipek–Mezey[102] metrics are two of the most important examples, which continue to be frequent use. Other examples, such as the Edmiston–Ruedenberg[118] metric, which maximises the Coulomb self-interaction of the orbitals, remain important but less heavily employed due to comparative computational expense.

The FB method localises orbitals by defining a metric which minimises the orbital spread, as given by its variance,

$$\langle O \rangle_{\mathbf{FB}} = \sum_{i} \langle i_{\mathbf{0}} | \, \mathbf{r^2} \, | i_{\mathbf{0}} \rangle - \langle i_{\mathbf{0}} | \, \mathbf{r} \, | i_{\mathbf{0}} \rangle^2 \, . \tag{3.6}$$

Marzari and Vanderbilt[103] generalised the FB approach, originally conceived for molecules, to evaluate localised WFs, creating so called Foster–Boys Wannier functions (FBWFs). The Wannier90 package[104], which employs this method, has been widely used amongst the solid-state community[105], interfacing with both plane wave and Gaussian codes[104].

The original Pipek–Mezey metric was defined as the sum of squares of the Mulliken partial charges[102]. To generate localised WFs using the PM metric (PMWFs), the objective functional is given by,

$$\langle O \rangle_{\mathbf{PM}} = \sum_{\mathbf{l},A,i} |Q_i^{A_\mathbf{l}}|^p = \sum_{\mathbf{l},A,i} \langle i_\mathbf{0}| \hat{P}_{A_\mathbf{l}} |i_\mathbf{0}\rangle^p \,, \tag{3.7}$$

where $Q_i^{A_\mathbf{l}}$ is the Mulliken charge[102] associated with WF i on atom A, situated in unit cell $\mathbf{l}$, evaluated using the WFs located in the reference unit cell. $\hat{P}_{A_\mathbf{l}}$ projects onto a basis of AOs centred on atom $A_\mathbf{l}$, given by,

$$\hat{P}_{A_\mathbf{l}} = \sum_{\mu \in A} \sum_{\nu,\mathbf{l}'} |\mu_\mathbf{l}\rangle \, (S^{-1})^{\nu,\mathbf{l}'}_{\mu,\mathbf{l}} \, \langle \nu_{\mathbf{l}'}| \,, \tag{3.8}$$

where $S^{\nu,\mathbf{l}'}_{\mu,\mathbf{l}} = \langle \mu_\mathbf{l} | \nu_{\mathbf{l}'} \rangle$ is the overlap. Under a BvK LCAO approach, where atomic orbital coefficients and periodic AO overlaps are naturally available, computing the PM metric is extremely straightforward. In addition, the strong interpretability of PMWFs, producing orbitals with σ and π separation[106], as opposed to the 'banana' bonds found within the FBWF scheme, further motivates our choice to utilise this metric.

The value of the penalty exponent, p, is typically 2, or 4. If p is chosen to equal 1, then Eq.3.7 reduces to the normalisation criteria of the bands,

$$\sum_{\mathbf{l},A} \sum_{i}^{\mathrm{n_{occ}}} Q_i^{A_\mathbf{l}} = \mathrm{n_{occ}}, \tag{3.9}$$

where $\mathrm{n_{occ}}$ are the number of occupied bands.

A key issue with the original definition of the PM metric is that the Mulliken partial charges do not possess a complete basis set limit, meaning they are unreliable for non-minimal basis sets. Alternative charge definitions for MOs have been suggested[119–121], that remove this basis set dependence. Lehtola and Jónsson demonstrated that the localised orbitals obtained were largely independent of the chosen partial charge estimate[119], providing significant freedom in choice. In the context of periodic systems, Jónsson et al.[106] first introduced a scheme to generate PMWFs, avoiding the issues surrounding Mulliken charges by using real space partitioning of orbital charge densities. Clement et al. later outline an

alternative charge definition based upon projection onto a pre-determined set of minimal basis functions[107].

The intrinsic atomic orbital method (IAO), as proposed by Knizia[117], is one choice of an alternative partial charge estimate that has been employed successfully for molecules. Using a free-atom minimal basis as a template, contraction coefficients from the original basis to IAOs are defined such that the occupied orbitals are exactly represented, which provides a consistent assignment of charge to atomic centres. Localised MOs using IAOs align well with chemical intuition, and quantitative measures such as partial charges and populations are shown to be resistant to changes in original basis, and are consistent with chemical understanding, leading to the method being implemented in many quantum chemistry packages[81,82,116,122,123]. We thus propose to adapt IAOs to construct a charge metric suitable to localise Wannier functions, $Q_i^{A_1^{\mathrm{IAO}}}$, using Bloch intrinsic atomic orbitals (Bloch IAOs).

3.2.3 Bloch Intrinsic Atomic Orbitals

Given the success of IAOs within molecular schemes, we believe a k-space extension to periodic systems possesses the advantages of being conceptually familiar, whilst also being straightforward to adapt from existing molecular codes. Having the same partial charge estimate for both molecular and periodic systems opens the possibility of making direct comparisons across systems. Schäfer et al.[25] demonstrate the use of IAOs to evaluate localised Wannier functions for a Γ-point-only calculation, following the molecular formulation as described by Janowski[124]. Cui et al.[73] construct crystal IAOs, from which projected AOs are evaluated. We employ similar principles in our generalisation to k-space, but crucially outline the additional augmentations needed to construct localised WFs, optimised using the PM metric, as a full periodic adaption of the IAO method.

We choose to adapt Knizia's method[117], such that a set of intrinsic atomic orbitals are constructed for each k-point within our Monkhorst-Pack quadrature mesh[125]. The Bloch IAOs are able to exactly describe the original occupied Bloch functions, providing a basis independent charge metric for WFs.

The original Bloch functions (Eq.4.14) are expressed in terms of Bloch AOs in the original basis set, labelled B_1. Analogous to Knizia's approach, a minimal basis, B_2, of free-atom AOs is first chosen, from which corresponding Bloch AOs are obtained, $|\rho_{\mathbf{k}}\rangle$ where $\rho \in B_2$, from Eq.2.12.

The following projection operators are defined,

$$\hat{P}_{12}^{\mathbf{k}} = \sum_{\mu,\nu \in \mathrm{B}_1} |\mu_{\mathbf{k}}\rangle \, (S_1^{-1})_{\mu,\nu}^{\mathbf{k}} \, \langle \nu_{\mathbf{k}}| \tag{3.10}$$

$$\hat{P}_{21}^{\mathbf{k}} = \sum_{\rho,\sigma \in \mathrm{B}_2} |\rho_{\mathbf{k}}\rangle \, (S_2^{-1})_{\rho,\sigma}^{\mathbf{k}} \, \langle \sigma_{\mathbf{k}}| \, , \tag{3.11}$$

where $(S_1)_{\mu,\nu}^{\mathbf{k}} = \langle \mu_{\mathbf{k}} | \nu_{\mathbf{k}} \rangle$ and $(S_2)_{\rho,\sigma}^{\mathbf{k}} = \langle \rho_{\mathbf{k}} | \sigma_{\mathbf{k}} \rangle$ are the Bloch-AO overlap matrices in the original and minimal basis sets, respectively. Using these operators, depolarisied occupied Bloch functions are obtained through

$$\{|\check{i}_{\mathbf{k}}\rangle\} = \mathrm{orth}\{\hat{P}_{12}^{\mathbf{k}} \hat{P}_{21}^{\mathbf{k}} \, |i_{\mathbf{k}}\rangle\}, \tag{3.12}$$

or in matrix form:

$$\check{\mathbf{C}}^{\mathbf{k}} = \mathrm{orth}\{\mathbf{P}_{\mathbf{12}}^{\mathbf{k}} \mathbf{P}_{\mathbf{21}}^{\mathbf{k}} \mathbf{C}^{\mathbf{k}}\}. \tag{3.13}$$

Here orth{} denotes symmetric orthogonalisation and the transfer matrices are $\mathbf{P}_{\mathbf{12}}^{\mathbf{k}} = (\mathbf{S}_1^{-1})^{\mathbf{k}} \mathbf{S}_{12}^{\mathbf{k}}$ and $\mathbf{P}_{\mathbf{21}}^{\mathbf{k}} = (\mathbf{S}_2^{-1})^{\mathbf{k}} \mathbf{S}_{21}^{\mathbf{k}}$, where $(S_{12})_{\mu\rho}^{\mathbf{k}} = \langle \mu_{\mathbf{k}} | \rho_{\mathbf{k}} \rangle$ and $\mathbf{S}_{21}^{\mathbf{k}} = \mathbf{S}_{12}^{\mathbf{k}\dagger}$. The projector onto the depolarised occupied Bloch functions is $\check{O}^{\mathbf{k}} = \sum_i |\check{i}_{\mathbf{k}}\rangle \, \langle \check{i}_{\mathbf{k}}|$. The Bloch IAOs are the minimal Bloch AO basis that contains both the depolarised and polarisation contributions and are defined through

$$|\rho_{\mathbf{k}}^{\mathrm{IAO}}\rangle = (O^{\mathbf{k}} \check{O}^{\mathbf{k}} + (1 - O^{\mathbf{k}})(1 - \check{O}^{\mathbf{k}})) \hat{P}_{12}^{\mathbf{k}} \, |\rho_{\mathbf{k}}\rangle \, . \tag{3.14}$$

In matrix notation, Eq.3.14 is given as:

$$\begin{aligned} \mathbf{A}^{\mathbf{k}} =& \mathbf{C}^{\mathbf{k}} \mathbf{C}^{\mathbf{k}\dagger} \mathbf{S}_{\mathbf{1}}^{\mathbf{k}} \check{\mathbf{C}}^{\mathbf{k}} \check{\mathbf{C}}^{\mathbf{k}\dagger} \mathbf{S}_{\mathbf{12}}^{\mathbf{k}} + \\ & (\mathbf{1} - \mathbf{C}^{\mathbf{k}} \mathbf{C}^{\mathbf{k}\dagger} \mathbf{S}_{\mathbf{1}}^{\mathbf{k}})(\mathbf{P}_{\mathbf{12}}^{\mathbf{k}} - \check{\mathbf{C}}^{\mathbf{k}} \check{\mathbf{C}}^{\mathbf{k}\dagger} \mathbf{S}_{\mathbf{12}}^{\mathbf{k}}), \end{aligned} \tag{3.15}$$

where $\mathbf{A}^{\mathbf{k}}$ are the contraction coefficients from Bloch AOs to Bloch IAOs at each k-point, and $\mathbf{1}$ is the identity in the space of B_1. Janowski[124] and Knizia[117] both

outline a simpler definition for the IAOs, which is equivalent under the assumption that B_2 can be directly expressed in B_1. Having implemented both schemes, the output Bloch IAOs are noted to be very similar, with no significant difference in localisation performance. Finally, the coefficients of the occupied Bloch functions in the Bloch IAO basis are given by

$$\mathbf{C}^{\mathbf{k}(\mathrm{IAO})} = (\mathbf{S}^{\mathbf{k}(\mathrm{IAO})})^{-\mathbf{1}}\mathbf{A}^{\mathbf{k}\dagger}\mathbf{S}_{\mathbf{1}}^{\mathbf{k}}\mathbf{C}^{\mathbf{k}}, \tag{3.16}$$

$$\mathbf{S}^{\mathbf{k}(\mathrm{IAO})} = \mathbf{A}^{\mathbf{k}\dagger}\mathbf{S}_{\mathbf{1}}^{\mathbf{k}}\mathbf{A}^{\mathbf{k}}. \tag{3.17}$$

In the original molecular implementation, the output IAO coefficients are symmetrically orthogonalised. However, in the periodic case, the orthogonalisation procedure introduces arbitrary phases to the Bloch functions in the IAO basis, specifically when obtaining the eigenvectors of the IAO coefficient matrix. The relative phase differences between k-points are thus altered compared to the original Bloch functions expressed in B_1, leading to issues when optimising the set of unitary matrices, across the Brillouin zone, in the IAO basis, since they do not correspond to the original Bloch functions. The simplest solution to remove this additional gauge problem is to leave the IAOs un-orthogonalised. The 'depolarised' Bloch functions, given by Eqs.3.12 and 3.13, which are orthogonalised, avoid this issue, because any phase augmentation is cancelled in the projector $\tilde{O}^k$.

In summary, obtaining the Bloch IAOs is numerically straightforward, requiring only a free atom basis, and its corresponding Bloch AO overlaps to perform the matrix multiplication steps. Computation of inverse overlap matrices can be avoided by solving instead with a Cholesky decomposition. The Bloch coefficients in the IAO basis and the IAO overlap matrix can then be used in a Pipek–Mezey style optimisation to obtain optimally local Wannier functions. The PM projector (Eq.3.8), in the Bloch IAO basis, is now defined as

$$\hat{P}_{A_\mathbf{l}}^{\mathrm{IAO}} = \sum_{\rho\in A}\sum_{\sigma,\mathbf{l}'} |\rho_{\mathbf{l}}^{\mathrm{IAO}}\rangle \, (S^{-1})_{\rho,\mathbf{l}}^{\sigma,\mathbf{l}'(\mathrm{IAO})} \, \langle\sigma_{\mathbf{l}'}^{\mathrm{IAO}}| \,, \tag{3.18}$$

where $S_{\rho,\mathbf{l}}^{\sigma,\mathbf{l}'(\mathrm{IAO})} = \langle\rho_{\mathbf{l}}^{\mathrm{IAO}}|\sigma_{\mathbf{l}'}^{\mathrm{IAO}}\rangle$ is the IAO overlap in real space, obtained from Fourier transforming $\mathbf{S}^{\mathbf{k}(\mathrm{IAO})}$.

It should be noted that Clement et al.[107] also employed a minimal basis of free atom AOs to calculate a robust charge estimate. Orbital coefficients in the minimal basis were obtained by computing the psuedoinverse of the overlaps between the Bloch functions in the original basis, and the real space reference cell AOs in the minimal basis. Whilst this charge estimate is also simple to evaluate, and demonstrated to be robust and basis set resistant, Bloch IAOs have the additional advantage of being able to exactly represent the occupied space, as demonstrated first in molecules.

3.2.4 Localisation Procedure

Recent work obtaining PM localised WFs and MOs have involved optimisation algorithms to determine the stationary points of the functional[106–108,119,126]. Schreder and Luber[108] implemented a method that simultaneously applies complex Jacobi rotations to several unitary matrices[127], in order to maximise the the PM functional. Clement et al.[107] recently demonstrated that a solver using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm leads to significantly faster convergence compared to previous steepest ascent (SA) or conjugate gradient implementations. Our localisation procedure uses a BFGS based algorithm, that is employed in conjunction with Bloch IAO charges. To generate an effective initial guess for the optimisation, a novel procedure generates approximately localised WFs, which we call *diabatic* Wannier functions.

3.2.4.1 Diabatic Wannierisation

The initial guess for the WFs is an important step in the localisation procedure in order to avoid encountering local maxima. Methods which project Bloch functions onto a set of trial functions have been outlined[105], whilst other implementations ensure the unitary space is probed fully by running multiple calculations using randomly sampled unitary matrices[106,107]. Clement et al.[107] combine random unitary sampling with a procedure to remove the gauge freedom of the Bloch functions.

As mentioned prior, Bloch functions are defined with an arbitrary gauge (Eq.3.2). By fixing the gauge such that the variations between Bloch functions in k-space are gradual, the Fourier transform produces WFs which are largely localised to a single cell, serving as an excellent starting guess for further optimisation. We have defined the natural gauge to be where the scalar product of the coefficients between Bloch functions within a band at $\mathbf{k}$ and the Γ-point $\mathbf{0}$ is real. For generalised WFs (Eq.3.5), where the gauge uncertainty is increased by mixing bands, the intuition of the natural gauge is extended, to construct diabatic Bloch orbitals and diabatic Wannier functions. Firstly, by an orthogonal transformation $\mathbf{O}$, the Bloch orbitals of the Γ-point are localised,

$$|i_{\mathbf{0}}\rangle = \sum_j |j_{\mathbf{0}}\rangle O_{ji}. \tag{3.19}$$

The Bloch orbitals of the remaining k-points are then chosen to be those with maximal similarity with the Γ-point. The locality of the orbitals of the Γ-point is thus transferred diabatically across the first Brillouin zone. This is obtained by calculating the unitary matrices, outside the Γ-point, which give the minimal least squares difference to the Bloch coefficients of the Γ-point,

$$\min_{\boldsymbol{U}_{\mathbf{k}}} ||C^{\mathbf{0}}_{\mu,p} - \sum_j C^{\mathbf{k}}_{\mu,j} U^{\mathbf{k}}_{j,p}||^2. \tag{3.20}$$

where $||\dots||$ is the Frobenius[128] norm. As this is an example of an Orthogonal Procrustes problem[129], a solution can be easily obtained via the singular value decomposition of the product of Bloch coefficients $\boldsymbol{C}^{\mathbf{k}\dagger}\boldsymbol{C}^{\mathbf{0}}$,

$$|i_{\mathbf{k}}\rangle = \sum_{\mu,j,j'} C^{\mathbf{k}}_{\mu,j} U_{j,j'} V_{j',i} |\mu_{\mathbf{k}}\rangle \tag{3.21}$$

$$\boldsymbol{C}^{\mathbf{k}\dagger}\boldsymbol{C}^{\mathbf{0}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}. \tag{3.22}$$

We choose a convenient approximate localisation procedure for the Bloch orbitals of the Γ-point, where they are replaced simply with the Cholesky vectors of the Γ-point density, ensuring computation of the diabatic WFs is fast.

3.2.4.2 Optimisation of PM metric

A gradient based optimisation method is implemented to obtain the stationary points of the PM functional. Similar to Clement et al.[107] and Lehtola and Jónsson[126], a Riemannian geometry approach is adopted to maintain the unitary constraint, as outlined in Refs. 130, 131. This method has proved successful since the unitary constraint is maintained implicitly, whilst other methods, including Lagrange multipliers[132], may suffer from slow convergence or only obtain a solution which only approximately maintains orthonormality.

Given the extensive discussion of the unitary optimisation algorithm in Refs. 130,131, we only briefly outline our procedure here. Crucially, the PM charge metric and all associated expressions are evaluated in the Bloch IAO basis, using $\mathbf{C}^{\mathbf{k}(\mathrm{IAO})}$ (Eq.3.16). The real space IAO overlaps, $S^{\rho,\mathbf{l}(\mathrm{IAO})}_{\sigma,\mathbf{l}'}$, are also used.

In the following expressions, the IAO labels are omitted for clarity. The Bloch IAO charges are defined as

$$\begin{aligned} Q_i^{A_\mathbf{l}^{\mathrm{IAO}}} &= \langle i_\mathbf{0}| \, \hat{P}^{\mathrm{IAO}}_{A_\mathbf{l}} \, |i_\mathbf{0}\rangle = \\ &\frac{1}{N^2} \sum_{\rho\in A} \Big[\sum_{j,\mathbf{k}} \bar{C}^{*\mathbf{k},\mathbf{l}}_{\rho,j} \, U^{*\mathbf{k}}_{j,i} \Big] \Big[\sum_{j',\mathbf{k}'} C^{\mathbf{k}',\mathbf{l}}_{\rho,j'} \, U^{\mathbf{k}'}_{j',i} \Big], \end{aligned} \tag{3.23}$$

where $C^{\mathbf{k},\mathbf{l}}_{\rho,j} = C^{\mathbf{k}}_{\rho,j} e^{i\mathbf{k}\mathbf{l}}$ and $\bar{C}^{\mathbf{k},\mathbf{l}}_{\rho,j} = \sum_{\sigma,\mathbf{l}'} C^{\mathbf{k},\mathbf{l}'}_{\sigma,j} S^{\rho,\mathbf{l}}_{\sigma,\mathbf{l}'}$. The Euclidean derivative of the PM functional, $\langle O \rangle_{\mathbf{PM}}$, with respect to the unitary at $\mathbf{k}$, is given by:

$$\begin{aligned} \frac{\partial \langle O \rangle_{\mathbf{PM}}}{\partial U^{*\mathbf{k}}_{j,i}} &= \frac{\partial \langle O \rangle_{\mathbf{PM}}}{\partial Q_i^{A_\mathbf{l}^{\mathrm{IAO}}}} \frac{\partial Q_i^{A_\mathbf{l}^{\mathrm{IAO}}}}{\partial U^{*\mathbf{k}}_{j,i}} \\ &= \frac{p}{N^2} \sum_{\mathbf{l},A} |Q_i^{A_\mathbf{l}^{\mathrm{IAO}}}|^{p-1} \sum_{\rho\in A} \bar{C}^{*\mathbf{k},\mathbf{l}}_{\rho,j} \Big[\sum_{j',\mathbf{k}'} C^{\mathbf{k}',\mathbf{l}}_{\rho,j'} U^{\mathbf{k}'}_{j',i} \Big] \\ &+ C^{*\mathbf{k},\mathbf{l}}_{\rho,j} \Big[\sum_{j'',\mathbf{k}''} \bar{C}^{\mathbf{k}'',\mathbf{l}}_{\rho,j''} \, U^{\mathbf{k}''}_{j'',i} \Big]. \end{aligned} \tag{3.24}$$

The Riemannian gradient, $\boldsymbol{G}_\mathbf{k}$, can then be transformed from the Euclidean gradient, $\boldsymbol{\Gamma}_\mathbf{k} = \frac{\partial \langle O \rangle_{\mathbf{PM}}}{\partial \boldsymbol{U}^*_\mathbf{k}}|_{\boldsymbol{U}_\mathbf{k}}$, by

$$\boldsymbol{G}_\mathbf{k} = \boldsymbol{\Gamma}_\mathbf{k}(\boldsymbol{U}_\mathbf{k})^\dagger - \boldsymbol{U}_\mathbf{k}(\boldsymbol{\Gamma}_\mathbf{k})^\dagger. \tag{3.25}$$

The 'two loop recursion' version of the limited-memory BFGS algorithm (l-BFGS)[133] is used, as first implemented for WFs by Clement et al.[107], to obtain a search

direction, $\{\boldsymbol{H}_{\mathbf{k}}\}$. The matrix elements in the upper triangle of the anti-Hermitian matrices for the Riemannian gradient, $\{\boldsymbol{G}_{\mathbf{k}}\}$, form the gradient vector for the l-BFGS algorithm, ensuring the output search direction is located on the unitary manifold. Given the requirement for the output WFs to be real, as mentioned earlier, our gradient vector is comprised of a total $o^2(N-1)/2 + o(o-1)/2$ real numbers, where o is the number of Bloch functions being localised.

To obtain a suitable step size, μ_{opt}, an Armijo[130,134] and Wolfe line search[133] were both implemented. If the line search along the l-BFGS direction fails, the search direction is reset to the steepest ascent vector, and the line search is then repeated.

The unitary matrices at each k-point are updated,

$$\boldsymbol{U}_{\mathbf{k},\text{new}} = e^{\mu_{\text{opt}}\boldsymbol{H}_{\mathbf{k}}}\boldsymbol{U}_{\mathbf{k},\text{old}}, \tag{3.26}$$

until the norm of the Riemannian gradient decreases below a threshold. The output unitary can then be applied directly to the original Bloch functions, to obtain localised WFs in the B_1 basis, using Eq.3.5.

3.3 Results and Discussion

3.3.1 Computational Details

The Bloch IAO procedure and the Pipek–Mezey Wannier function localisation have been implemented in a developmental version of the TURBOMOLE[81,82] package. The initial mean-field Bloch functions were obtained through the periodic Hartree–Fock procedure, within the `riper` module[83–86,97,98]. To generate the IAOs, a minimal free-atom basis was constructed from HF calculations of isolated atoms in the cc-pVTZ[135–137] basis, as already implemented within TURBOMOLE to construct molecular IAOs. The PM functional was evaluated with a penalty exponent of $p = 4$, rather than 2, as shown in Eq.3.7, due to better localisation for π character orbitals, as discussed in prior works[117,119].

An Armijo step size method and a Wolfe line search were tested in the localisation procedure. It is known that a line search fulfilling the Wolfe conditions ensures stability of the BFGS updates, by ensuring the approximate Hessian, within our

Table 3.1: Insulating and semiconducting systems used to test the Bloch IAO and PM localisation procedure

System	Basis B_1	Monkhorst-Pack mesh size
diamond	pob-TZVP[93]	11,11,11
silicon	pob-TZVP	11,11,11
boron nitride	pob-TZVP	15,15
graphene	pob-TZVP	15,15
MgO	def-SVP[138]	11,11,11
SiO2	pob-TZVP	5,5,5
trans-$(C_2H_2)_\infty$	pob-TZVP	101
(4,4) C-nanotube	def2-SVP[138]	11

maximisation problem, is negative definite[133]. However, computation of the line search is costly, since multiple gradient evaluations are required along the trial direction. By contrast, the Armijo line search only requires PM metric values to be evaluated, and we observed its convergence performance to be similar to the Wolfe line search, with shorter total wall times. The Armijo search was employed in all calculations subsequently discussed in this chapter.

The formal computational scaling of the optimisation scheme was obtained by analysing the intermediates involved to evaluate the PM gradient and metric, detailed explicitly in Appendix A. The most expensive intermediate to evaluate has a complexity scaling of $\mathcal{O}(i\mu^2k^2)$, where i, μ and k refer to the number of occupied orbitals per cell, the number of IAOs per cell, and the k-point mesh, respectively. One intermediate, computed only once in the optimisation, has a scaling of $\mathcal{O}(\mu^2k^3)$. Storage of the real space IAO overlaps, $S^{\rho,\mathbf{l}(\mathrm{IAO})}_{\sigma,\mathbf{l}'}$, have the largest memory cost. Although not implemented within the current version of the code, exploiting the sparsity of this overlap matrix appears to be a viable option to reduce the memory and evaluation cost of this tensor, as the information of this overlap tensor will only in practice scale as $\mathcal{O}(\mu k)$ in the long range limit. Within all our testing, these potential memory or computation scaling concerns have not posed any issues, and the wall times for the PM localisation scheme are always smaller or similar to the periodic Hartree–Fock method, and much shorter than the subsequent DLPNO-MP2 calculation.

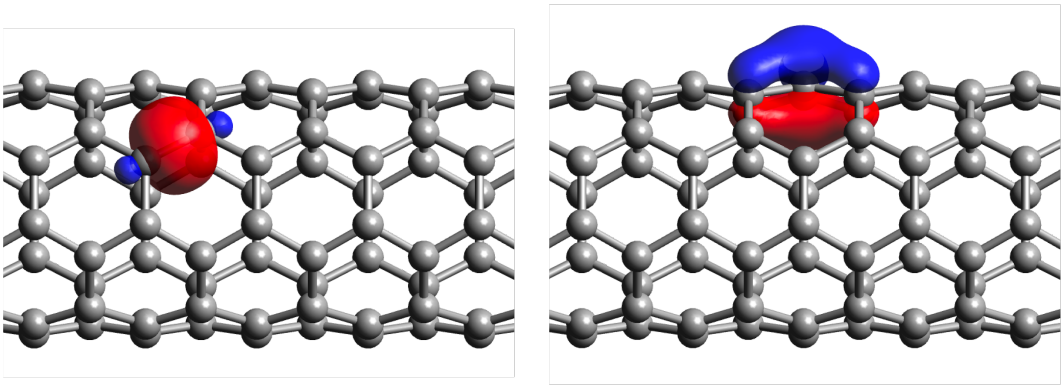

Figure 3.1: Bloch IAO localised Wannier functions of (4,4) nanotube, showing σ (left) and π (right) bonding character. An isosurface value of 0.05 was used.

Table 3.2: PM metric values of WFs from the initial SCF calculation, after rotation into the natural gauge, after diabatic Wannierisation and after PM optimisation

System	$\langle O \rangle_{\mathbf{PM}}$(SCF)	$\langle O \rangle_{\mathbf{PM}}$(Nat.)	$\langle O \rangle_{\mathbf{PM}}$(Dia.)	$\langle O \rangle_{\mathbf{PM}}$ (Opt.)
diamond	1.49×10^{-8}	6.13×10^{-2}	1.92	2.66
silicon	2.40×10^{-2}	6.64×10^{-2}	9.79	10.66
boron nitride	1.20×10^{-5}	2.36	3.19	3.37
graphene	3.37×10^{-6}	8.37×10^{-2}	1.94	2.56
MgO	2.88×10^{-1}	3.87	9.56	9.61
SiO2	5.96×10^{-4}	9.55	30.80	37.29
trans-$(C_2H_2)_\infty$	4.92×10^{-5}	4.25×10^{-2}	2.17	2.76
(4,4) C-nanotube	1.65×10^{-3}	4.34×10^{-3}	33.70	40.60

Table 3.1 details the insulating and semi-conducting systems used to probe the performance of the IAO PM localisation scheme. The original basis set, B_1, used in the mean field calculation, and the Monkhorst-Pack[125] mesh, are shown. For the RI-J approximation, all test systems utilised the universal Coulomb-fitting auxiliary basis sets[138] with the exception of the magnesium oxide and carbon nanotube systems, which employed auxiliary functions optimised for the def-SVP and def2-SVP basis sets, respectively[138]. Unit cell parameters and geometries are provided in the supplementary information. All figures were plotted using Avogadro[139].

3.3.2 Overall performance

Tables 3.2 and 3.3 report the performance of the Bloch IAO localisation scheme. The values of the PM metric are presented for the WFs of the SCF calculation, after rotation into the natural gauge, after diabatic Wannierisation and after PM

Table 3.3: Number of iterations to converge PM metric, after initial diabatic preparation, using l-BFGS or steepest ascent.

System	l-BFGS	SA
diamond	35	100
silicon	42	114
boron nitride	55	731
graphene	42	60
MgO	8	744
SiO2	338	1794
trans-$(C_2H_2)_\infty$	43	1517
(4,4) C-nanotube	62	261

optimisation. The number of iterations required to localise the PM objective function with l-BFGS, compared to steepest ascent (SA), using diabatic WFs as the initial guess, are also given. The convergence threshold for the PM gradient norm was set to 10^{-5}, with the exception of the threshold for the boron nitride (BN) system, which was set to 10^{-6}. All occupied orbitals are included in the optimisation.

In all cases, the final values of the PM metric from the l-BFGS and steepest ascent optimisations were equal, to a precision of 10^{-5} (or 10^{-6} for BN), confirming that the output WFs were equally localised. A tighter threshold was required for BN in order for the output WFs from l-BFGS and SA to agree to target precision. As demonstrated by Clement et al.[107], we confirm that utilising l-BFGS, compared to SA, markedly improves convergence performance. In some systems, a tenfold reduction in iterations to converge is observed, if not greater. The PM optimiser performs robustly across the range of insulating and semiconducting materials explored, successfully localising every test system. With the exception of silicon dioxide, all test systems converge within 100 iterations with l-BFGS. A larger number of iterations was required for silicon dioxide. We observed that the step size was very small, which indicates that the Hessian description of the landscape in this case may be poor. Despite this, l-BFGS still converges five times faster compared to SA, for this example, showing the robustness of the scheme.

We stress that using diabatic WFs as the initial preparation has an important role in the robustness and quality of the final localised WFs. As seen in Table 3.2, the values of the PM metric after diabatic Wannierisation are remarkably

close to the final optimised values, showing that a significant degree of locality has been captured through the diabatisation. Our experiments using randomly generated unitary matrices as the initial guess led to final WFs with metric values that consistently were smaller than that obtained from the diabatic preparation, and never greater. The choice of objective functional, and the parameterisation employed for the gradient, gives an optimisation landscape with many local maxima and use of an appropriate initial guess, such as diabatic WFs, is required to ensure higher valued maxima are located, compared to random unitary sampling. Direct comparison of the number of iterations required to converge, using a random guess and the diabatic preparation, is often not possible since the final WFs are usually inequivalent. Although not verified in this thesis, we predict that localising the Bloch functions of the Γ-point with an IAO procedure instead of via Cholesky decomposition would further increase $\langle O \rangle_{\mathbf{PM}}$(Dia.) and would reduce the number of iterations required for full optimisation.

The initial values of the PM metric from direct Wannierisation of the SCF Bloch functions are very small. It should be noted that due to the different gauges of the Bloch functions in separate calculations, the values for $\langle O \rangle_{\mathbf{PM}}$(SCF) can vary arbitrarily. The values reported in Table 3.2 are in fact the largest value of $\langle O \rangle_{\mathbf{PM}}$(SCF) taken from 10 separate calculations. Throughout our testing, we have never observed any example where $\langle O \rangle_{\mathbf{PM}}$(SCF) has been greater or similar in magnitude to $\langle O \rangle_{\mathbf{PM}}$(Dia.). Applying the natural gauge to the Bloch functions increases the values of the PM metric in all cases, confirming that the natural gauge smooths the Bloch functions in reciprocal space. For crystals with well-separated bands formed from weakly interacting AOs, simply applying the natural gauge results in well-localised WFs. In all cases, $\langle O \rangle_{\mathbf{PM}}$(Nat.)) sit in between the SCF and diabatic values.

Figure 3.1 presents example localised WFs of the (4,4) nanotube system, described in Table 3.1. The left subfigure clearly shows a carbon-carbon σ bond, whilst the right subfigure illustrates a π character WF. This π and σ separation is

Table 3.4: Number of iterations to converge PM metric with initial diabatic preparation, with core and valence band separation

System	Initial core PM gradient	core	valence
diamond	1.08×10^{-4}	2	20
silicon	8.73×10^{-4}	3	24
boron nitride	9.70×10^{-4}	3	27
graphene	1.91×10^{-4}	2	22
MgO	5.17×10^{-5}	2	4
SiO2	6.55×10^{-3}	2	113
trans-$(C_2H_2)_\infty$	1.67×10^{-4}	3	22
(4,4) C-nanotube	6.81×10^{-2}	3	29

observed in the other test systems, demonstrating that the Bloch IAO localisation procedure retains the advantages of the original PM metric.

3.3.3 Separate optimisation of valence and core bands

It is often desirable to localise WFs for the core and valence bands separately. For example, the accurate calculation of correlation energies within local correlation methods requires localised occupied orbitals without contributions from uncorrelated core orbitals[140]. In view of this, the Bloch IAO localisation scheme performs exceptionally well in localising core and valence bands independently. This is primarily due to the quality of the diabatic WFs, which almost immediately generates localised core orbitals.

Table 3.4 shows the performance of the Bloch IAO localisation, with core and valence separation, for all test systems. The gradient norm of the PM metric for the core bands after the diabatic initial preparation is presented. Subsequent l-BFGS iterations required to localise using the PM metric for both the core and valence bands are also given. The PM gradient norms of our initial core WFs are all already close to the convergence threshold (10^{-5}, or 10^{-6} for BN), and localise within 3 l-BFGS iterations, demonstrating that the diabatic preparation yields nearly optimally local core WFs. The number of iterations required to localise the core increases typically only by 1 or 2 iterations when the convergence threshold is increased to 10^{-8}. Across all test systems, the difference in value of the PM metric between the final localised core orbitals and n_{core} were within 3×10^{-2}, where n_{core}

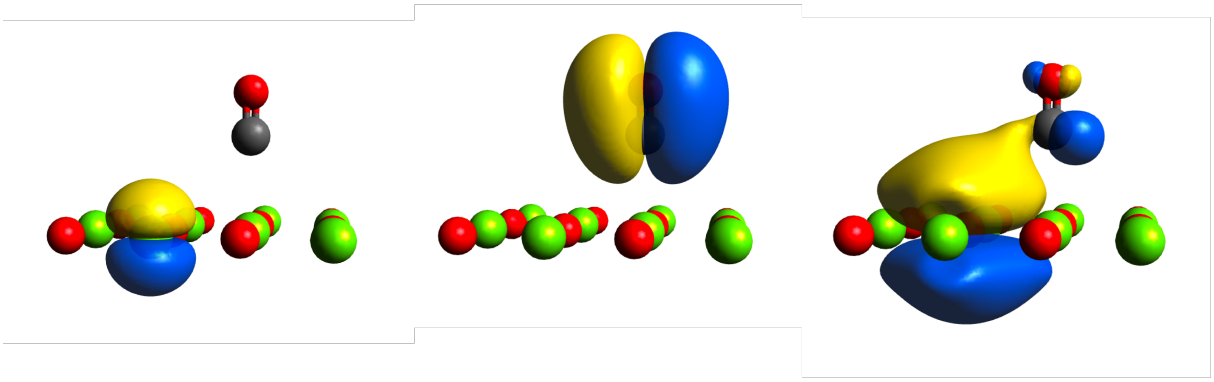

Figure 3.2: Bloch IAO localised Wannier functions of the MgO(001) CO adsorption system. A magnesium 2p-like orbital is shown (left), as well a π bonding orbital on CO (centre). A WF demonstrating oxygen 3p orbital donation, on the MgO surface, to the CO π^* orbital, is also presented (right). Only atoms in the reference cell are shown. An isosurface value of 0.01 was used.

is given by Eq.3.9, summing only over core bands. This indicates that the core orbitals are localised to the global maxima, given that n_{core} is the upper bound for the PM metric for the core. Localisation of the valence bands also occurs rapidly, although less markedly than in the core case, and the total number of iterations required to localise the core and valence bands separately is less than that required to localise the full occupied space (Table 3.2), across all the test systems. This is to be expected, given that the dimensionality of the optimisation problem is reduced by separating the core and valence bands.

Diabatic preparation is particularly effective in localising WFs from bands composed of weakly interacting AOs. The core bands and valence bands with strong ionic character, for example in MgO, require only a few optimisation steps for optimisation. Since the localised valence WFs are usually bonding in character, they are typically centred between atoms, and inherently less local than their core counterparts and require more steps for optimisaiton. Even for the core bands, using diabatic WFs as an initial preparation is vital for the robustness and accuracy of the Bloch IAO scheme. Using a random initial unitary for core bands frequently leads to output WFs with PM metric values significantly smaller than those obtained with the diabatic guess, showing that encountering local maxima is a common problem without correct preparation of the WFs.

3.3.4 Chemical intuition of Bloch IAO generated WFs

One of the key strengths of the original IAO scheme is the clear interpretation of these molecular orbitals, and the direct connection to chemical intuition and concepts. Knizia[117] demonstrated that IAOs allow robust, basis set independent, partial charges and orbital populations to be computed. This enables quantitative measures for electronegativities and oxidation states for molecules, which align with empirical understanding, to be evaluated. In analogous fashion, we demonstrate that Bloch IAO localised WFs provide chemical understanding in periodic systems. Bloch IAO partial charges can be computed in similar fashion to molecular IAOs, by summing the atomic contributions across all the cells within the supercell,

$$q_A = Z_A - \sum_{\mathbf{l},i} Q_i^{A\mathbf{l}^{\mathrm{IAO}}}, \tag{3.27}$$

where Z_A is the atom's nuclear charge.

IAO partial charges can be computed for any periodic system, such as those listed in Table 3.1. However, it is also worth stressing that IAOs can be used for both periodic and molecular systems, on equal footing, meaning that they are a robust and consistent partial charge estimate for probing both material and molecular systems. This opens the possibility to investigate interesting chemical scenarios, such as systems containing interactions between material and molecules.

The adsorption of CO onto the MgO(001) surface has been heralded as the 'hydrogen molecule of surface science[141], and an important case study for the theoretical understanding of heterogenous catalysis. This adsorption example is an ideal case to demonstrate the ability that Bloch IAO localised WFs have to provide insight into the underlying chemistry of the system. Although higher level quantum chemical methods have been used elsewhere to attempt to accurately model the weak van der Waals interactions dominating the adsorption, and to evaluate accurate adsorption energies, we stress that our current motivation is to use this system to exhibit the use of IAO WFs for chemical intuition. Since orbitals provide a zeroth order description to the motion of the electrons, and are a result from mean-field, effective one electron theories, using solely DFT to

Table 3.5: Bloch IAO partial charges of the non-interacting and equilibrium MgO(001) CO adsorption system

System	C (CO)	O (CO)	Mg (MgO)	O (MgO)
non-interacting	0.42	-0.42	1.68	-1.68
equilibrium	0.32	-0.41	1.70[1]	-1.65[2]

model this picture serves our purposes. In Chapter 7, we use periodic DLPNO-MP2 to evaluate accurate adsorption energies.

To model the system, a unit cell consisting of a $4 \times 4 \times 1$ slab of MgO was constructed. CO, orientated perpendicular to the surface, was positioned with a C-Mg equilibrium distance of 2.479Å[142]. To obtain mean-field Bloch functions, a periodic DFT (PBE[143]) calculation was conducted in the `riper` module, using the pob-TZVP[93] basis set, on a (3,3) Monkhorst-Pack mesh to sample the Brillouin zone. Bloch IAO localised WFs were then obtained.

Figure 3.2 presents three Bloch IAO localised WFs of the system, all three of which align with chemical understanding. The left and centre subfigures show a localised 2p orbital, centred on Mg, and a π bonding orbital, centred on CO, respectively. Similar WFs are observed when WFs for the surface and adsorbate are computed separately. A more interesting WF is presented on the right, where back-bonding from the nearest neighbour oxygen atom, on the MgO surface, to the π^* orbital of carbon monoxide, is shown. Although the role of back-bonding within metal oxide adsorption has been scrutinised[24,144] , we wish to make the point that IAO localised WFs provide direct and intuitive chemical understanding, consistent with the level of theory used to generate the original Bloch functions.

The Bloch IAO partial charges of the MgO(001) CO adsorption system, in the equilibrium geometry, were also calculated. In Table 3.5, we compare these partial charge values to charges obtained from separate periodic calculations of the MgO surface slab, and the CO molecule, representing a non-interacting scenario. Most significantly, a reduction in positive partial charge on the carbon of CO is observed moving from the non-interacting to equilibrium geometry, as well as a decrease in negative charge on the oxygen nearest neighbours of MgO. This can be

rationalised from the back-bonding process presented in Figure 3.2, and once again shows that quantitative measures, such as Bloch IAO partial charges, are consistent with chemical intuition, at the level of theory employed in the mean field picture.

3.4 Conclusions

In this chapter, we have generalised the intrinsic atomic orbital method to periodic solids. Bloch IAOs form a minimal basis which exactly represent the occupied bands, and thus remove the well known issue of using Mulliken charges for non-minimal basis sets. They thus enable localised Wannier functions, optimised using the Pipek–Mezey metric, to be robustly evaluated, as first introduced by Jónsson et al.[106] As observed with other PM localisation schemes, separate σ and π character orbitals are generated, in comparison to the Foster–Boys metric. We outline a localisation scheme, which prepares the initial Bloch functions by diabatically transferring locality imposed at the gamma point through the Brillouin zone, before localising according to the PM metric. Clement et al.[107] demonstrated the improved performance using l-BFGS compared to other gradient based solvers, and we confirm this. Our WF localisation scheme is fast and robust, producing localised orbitals similar to other PM derived schemes, whilst using a periodic generalisation of the widely employed IAOs, which we demonstrate can be straightforwardly adapted.

This scheme works efficiently across a range of semiconducting and insulating solids, and in particular we highlight the ability of the diabatic WFs to localise atom-centred WFs almost immediately. Using the example of CO adsorption onto MgO(001), we demonstrate that Bloch IAO partial charges can provide chemical insight into systems, through visualisation of the localised WFs, and through computing measures such as partial charges. We expect that Bloch IAOs will provide a bridge for understanding chemical phenomena within periodic systems. Bloch IAOs are not solely restricted to LCAO methods, but can also be applied with plane wave basis sets by computing the overlap between plane waves and the minimal Gaussian AO basis[25]. In subsequent chapters, we utilise the novel Wannier localisation scheme for the starting point of our periodic DLPNO-MP2 method.

4| Periodic DLPNO-MP2 Theory

The work in this chapter contributed to the publications Ref.2 and Ref.3

This chapter provides an overview of the theoretical concepts underpinning periodic DLPNO-MP2. Building on the ideas of Hartree–Fock (HF) theory in Chapter 2, we first briefly discuss correlated wavefunction schemes in Section 4.1, and give a short account of three key families of post-HF methods, to facilitate the overall context of DLPNO-MP2. Reduced scaling schemes known as local correlation methods are introduced in Section 4.2, and an overview of the use of these methods within the literature is given, for both molecular and, more recently, periodic contexts. We then directly present the periodic local MP2 and the DLPNO-MP2 equations that underpin this thesis, highlighting the PAO-OSV-PNO cascade, in Sections 4.3 and 4.4.

4.1 Correlated wavefunction methods

HF theory provides a mean-field solution to the electronic structure problem. Although generally a good approximation, the HF energy is not equal to the exact electronic energy. In many cases, the neglected portion of the true energy is often critical to the quantitatively accurate prediction of chemical properties, and in fewer cases, even qualitative trends can be incorrectly evaluated. Whilst HF does account for the correlation of electrons due to Pauli antisymmetry, it fails to fully account for the pairwise Coulombic repulsion of the electrons relative to each other, since it only describes the interaction of an electron within an average potential generated by all other electrons. The aim of post-Hartree–Fock methods is to accurately evaluate the missing energy contribution, which is commonly termed the *correlation energy*,

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}. \tag{4.1}$$

Since the HF energy provides a variational upper bound to the exact energy, E_{corr} is defined to be negative, since the correlated motion of the electrons reduces the overall

electron-electron repulsion relative to the mean field picture. Discussion of the nature of correlation energy has led to the use of two broad classifying terms within the quantum chemistry community. *Dynamic correlation* describes small, predominantly pairwise, deviations from the mean field description due to the instantaneous electron-electron repulsions within a system. Here, the single HF Slater determinant is a good approximation to the full many body wavefunction, and strategies to capture dynamic correlation can be considered as perturbative treatments on the HF reference state. On the other hand, *static correlation* describes chemical scenarios where the single HF Slater determinant is a poor reference for the exact wavefunction. Here, additional configurations have significant contributions, and strategies to treat static correlation consider multiple Slater determinants on equal footing.

The remaining part of this section provides a brief account of three post-HF formalisms, Configuration Interaction (CI), Coupled Cluster (CC) and Møller–Plesset perturbation theory (MPPT). We by no means provide a comprehensive account of each family, but our aim is communicate the relative accuracy and degree of approximation in each formalism; the associated computational cost of its commonly used implementations; and the degree to which static and dynamic correlation are accounted for. In turn, we hope to motivate and provide context for a periodic DLPNO-MP2 scheme.

4.1.1 Configuration Interaction

The exact electronic correlation energy, for a given basis set, can in principle be obtained through a full configuration interaction (FCI) expansion. Here, the wavefunction is constructed as a linear combination of the full set of possible determinants,

$$|\Psi_{\mathrm{FCI}}\rangle = |\Psi_0\rangle + \sum_{ia} c_i^a \, |\Psi_i^a\rangle + \sum_{\substack{i<j \\ a<b}} c_{ij}^{ab} \, |\Psi_{ij}^{ab}\rangle + \sum_{\substack{i<j<k \\ a<b<c}} c_{ijk}^{abc} \, |\Psi_{ijk}^{abc}\rangle + \cdots , \tag{4.2}$$

where $|\Psi_0\rangle$ is a reference Slater determinant, often taken to be the HF state, and $|\Psi_i^a\rangle, |\Psi_{ij}^{ab}\rangle, |\Psi_{ijk}^{abc}\rangle$ are respectively singly, doubly and triply excited determinants relative to the reference state. Coefficients $c_i^a, c_{ij}^{ab}, c_{ijk}^{abc}$ are solved to obtain the exact many-body wavefunction. Whilst formally exact, FCI calculations quickly

become impractical because the number of possible determinants grows rapidly with system size. For a system of m spin-orbitals, consisting of n electrons, the number of determinants is,

$$n_{\text{det}} = \binom{m}{n}, \tag{4.3}$$

which scales exponentially in the large n limit. Calculations with FCI are thus restricted to only the smallest molecular systems, where the FCI correlation energy is used to benchmark other wavefunction methods. More practical CI methods involve truncating the number of determinants used in the expansion, such as selected CI[145]. These truncated schemes however omit disconnected excitations and hence do not correctly describe the energy of two non-interacting systems, when calculated together and separately. They are thus not size-extensive.

4.1.2 Coupled Cluster

Coupled Cluster refers to a family of methods which employ a wavefunction ansatz where excitations to different determinants are presented in an exponential form,

$$|\Psi_{\text{CC}}\rangle = \exp(\hat{T})\,|\Psi_0\rangle\,, \tag{4.4}$$

where the reference state is usually taken to be the HF ground state. The cluster operator, $\hat{T}$, consists of a summation of excitation operators,

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots \tag{4.5}$$

$$= \sum_{ia} t_i^a \hat{\tau}_i^a + \sum_{\substack{i<j \\ a<b}} t_{ij}^{ab} \hat{\tau}_{ij}^{ab} + \sum_{\substack{i<j<k \\ a<b<c}} t_{ijk}^{abc} \hat{\tau}_{ijk}^{abc} + \cdots\,, \tag{4.6}$$

where $\hat{\tau}$ are one, two, three, $\cdots$, n-body excitation operators, and the amplitudes $t_{ijk\cdots}^{abc\cdots}$ parametrise the CC wavefunction.

For an n-electron system, if all excitation operators up to $\hat{T}_n$ are included in the cluster operator, then the CC wavefunction is equivalent to FCI, and the exact correlation energy, in that given basis, is obtained. However, the same concerns of practical cost for FCI are also present in this variant of CC. The distinguishing advantage of the CC wavefunction ansatz lies in the properties of the cluster operator

when truncated to a certain excitation level. For example, in the Coupled-Cluster singles-and-doubles (CCSD) approach, where the cluster operator is truncated to $\hat{T} = \hat{T}_1 + \hat{T}_2$, determinants of every excitation level, through disconnected cluster amplitudes, are still included through the exponential form,

$$|\Psi_{\text{CCSD}}\rangle = \left(1 + (\hat{T}_1 + \hat{T}_2) + \frac{1}{2}(\hat{T}_1 + \hat{T}_2)^2 + \cdots\right)|\Psi_0\rangle\,. \tag{4.7}$$

Employing the linked-cluster theorem, this choice of parametrisation of the wavefunction ensures size-extensivity of the energy, in direct contrast to truncated CI.

Commonly employed variants of CC theory include the aforementioned CCSD, which accounts for the important pairwise electron interactions, and the orbital relaxations induced from such pair excitations. CCSD has a canonical scaling of n^6. To obtain higher accuracy, directly accounting for all triples excitations is required, that is, $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3$, in the CCSDT model, but a more common approach is to solve the triples in a perturbative fashion, in the CCSD(T) approach, to reduce the computational scaling from n^8 to n^7. Since these models do not fully account for the higher order excitations, there is an inherent bias to the reference state $|\Psi_0\rangle$ that the excitations are built from, although multireference CC theories have been proposed to better capture static correlation effects[146,147].

4.1.3 Møller–Plesset perturbation theory

For chemical systems where the HF single determinant is a good approximation, a perturbative approach is suitable to approximate the correlation energy. Møller–Plesset perturbation theory, which is based within the Rayleigh-Schödinger framework, provides a correction, by defining the zeroth-order Hamiltonian as the HF Fock operator,

$$\hat{H}_0 = \hat{F} = \sum_i^N \hat{f}(i) \;\text{ and }\; \hat{V} = \hat{H} - \hat{F}. \tag{4.8}$$

$\hat{V}$ is the perturbation operator and $\hat{H}$ and $\hat{F}$ are the full electronic Hamiltonian and Fock operators, as defined earlier in Sections 2.1 and 2.3, respectively.

Since the sum of the zeroth order and first order energies defines the HF energy, $E_{\mathrm{HF}} = E^{(0)} + E^{(1)}$, the first correction value for the correlation energy is given by the second-order energy (MP2),

$$E^{(2)} = -\frac{1}{4}\sum_{ijab}\frac{|g_{ab}^{ij} - g_{ba}^{ij}|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}, \tag{4.9}$$

which features the HF orbital eigenvalues for occupied spin-orbitals i, j and virtuals a, b in the denominator, and the antisymmetrized electron repulsion integrals in the numerator, where $g_{ab}^{ij} = \langle ij|r_{12}^{-1}|ab\rangle$, within a spin-orbital formalism.

Integrating out the spin coordinates, we arrive at the MP2 energy expression in spatial orbitals,

$$E^{(2)} = -\sum_{ijab}\frac{(2g_{ab}^{ij} - g_{ba}^{ij})g_{ab}^{ij}}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}. \tag{4.10}$$

Canonical MP2 scales as n^5 with respect to the number of orbitals, arising from the transformation of the ERIs from the AO to MO basis. It is thus less computationally expensive than FCI, CCSD(T) and CCSD, but also inherently is the most approximate treatment for the correlation energy, only treating pairwise excitations. Higher orders of MPPT incur a greater computational cost, and are less commonly employed.

4.2 Local Correlation Methods

Whilst post-HF methods are shown to provide a systematically improvable framework for treating many body correlation, the steep polynomial scaling of even the cheapest canonical method, MP2 (n^5), still leads to memory and CPU bottlenecks when targeting larger systems. Particularly in the context of electronic structure studies of materials, the poor scalability of these canonical methods lead to practical issues when more complex unit cells are employed, or when increasing the number of simulated unit cells to approach the thermodynamic limit.

Within molecular electronic structure theory, significant work has taken place to reduce the computational cost of post-HF schemes, whilst maintaining the accuracy. One category of methods which has emerged are known as *local correlation schemes*.

These methods are motivated by the idea that, in insulating systems, electron correlation decays with a leading inverse sixth power, and hence is predominantly short ranged. In this view, the steep scaling of canonical wavefunction methods is physically unfounded, originating from the delocalised nature of the canonical orbitals, and instead the locality of electron correlation can be exploited. Unitary transformation of the occupied orbitals to a localised orbital basis allows the correlation treatment to be restricted to occupied orbitals which are physically close to each other.

Local correlation methods follow two main strategies. The first separates the system into smaller fragments, in which correlated wavefunction techniques are used to solve for the energy of each fragment, before the total correlation energy is approximated through the sum of contributions. Examples of fragmentation based implementations include the divide-expand-consolidate method[148–151], the cluster-in-molecule (CIM) method[69,152–156], and the local natural orbital (LNO) schemes[157–159]. The alternative strategy instead solves for the entire system, but uses local approximations to reduce computational cost. The work in this thesis directly follows this second approach, and we now clearly outline the details of this method.

By decomposing the total correlation energy into a sum of energy contributions from occupied pairs, employing two key local approximations presents a route towards a linear scaling wavefunction scheme. In the pair approximation, all occupied pairs, in this local basis, are screened by distance or by approximate pair correlation energy. Pairs falling outside of a given threshold are neglected in the final correlation treatment. In turn, a local virtual basis can be constructed from functions spatially close to the occupied orbitals, such that only a restricted domain of virtual functions are used to describe the excitations of a given orbital. Within the domain approximation, the union of the respective domains for the occupied pair allow pair correlation energies to be evaluated within a truncated virtual space, which reduces computational cost compared to the canonical treatment. At the asymptotic limit, the number of significant pairs scales linearly with n, and the size of the pair domains becomes independent of system size, leading in principle to

a near-linear or linear scaling algorithm. By relaxing the thresholds for the pair screening and expanding the pair domain sizes, the canonical correlation energy can be fully recovered, although this also reclaims the canonical computational cost.

The work of Pulay and Saebø, starting from the 1980s, is considered pioneering in the context of local correlation methods. Projected atomic orbitals (PAOs), constructed by projecting out the occupied space from AOs, were employed to represent the virtual space[160–162]. Initial formulations predetermined the size of the PAO domains[162]. Later, in 2009, Neese and co-workers[79,80] revived Meyer's pair natural orbitals[163,164] (PNOs), where a bespoke set of virtuals are constructed for each electron pair to efficiently capture the pair correlation energies, known as domain-based pair natural local correlation (DLPNO) theory. Later, Yang et al. introduced orbital specific virtuals[165] (OSVs) as another choice, in which a bespoke set of virtuals are constructed for each occupied orbital. PNOs are the most compact representation of the virtual space, followed by OSVs and then PAOs. PNOs offer an advantage over OSVs of being pair specific, and notably a significantly smaller PNO pair domain is required for distant pairs compared to OSVs, where the diminishing interaction does not influence the OSV domain size.

In the context of molecular electronic structure theory, the DLPNO approach has established itself as the standard local correlation method, and has been successfully applied to perturbative[87,166–169] and coupled cluster formalisms[89,140,170–174], to explicitly correlated theory[88,169,175,176], to multireference methods[177–179] and to excited states[180–183], greatly extending the range of applicability of these approaches for computing energies and properties of molecular systems.

In the same period where the DLPNO approach has gained traction and widespread applicability, a number of works have considered hydrid schemes, combining different types of local virtuals, in order to further increase computational efficiency. In particular, we highlight the current molecular approach adopted in the TURBOMOLE software package, which employs a hybrid PAO-OSV-PNO method[87–90]. Here, OSVs are used as an intermediate, where the truncated set of OSVs are first used to estimate pair correlation energies to screen out insignificant

pairs. Since the OSV domain sizes are smaller than the PAO domains, they are also used to efficiently transform to the PNO basis. This approach within TURBOMOLE has been successfully applied to MP2, CCSD, CCSD(T), as well as to explicitly correlated formulations.

In contrast to molecules, wavefunction-based local correlation approaches for periodic systems are significantly less extensive, although a few methods have previously been proposed. In particular, the pioneering work within the CRYSCOR[43,51–53,184–186] package leverages projected atomic orbitals[162], and orbital specific virutals[165] to obtain a low-scaling MP2 implementation for non-conducting systems, under BvK periodic boundary conditions. Fragment based approaches for periodic systems also exist. The cluster-in-molecule approach has been extended to periodic systems by Li and coworkers[68,69,156], and Kallay's LNOs[157–159,187] have also been applied to solids, by Ye and Berkelbach[60]. Both frameworks have implementations of up to CCSD(T) levels of theory. Among these works, the contribution from Wang *et al.*[156] is particularly notable: DLPNO theory is used to solve the correlation space for each WF within their periodic CIM scheme. In contrast to our DLPNO formalism, which we outline in Section 4.3, the DLPNO solver used here is equivalent to molecular implementations, and hence no periodicity is enforced within the PNOs. Overall, these methods have been used successfully to obtain properties such as lattice constants, cohesive energies[53,60,188] and adsorption energies[19,21,189,190] for surface interactions.

Considering the success of DLPNO methods in molecular quantum chemistry, afforded by the advantageous compression in the correlation space compared to PAOs and OSVs, a fully periodic DLPNO scheme appears highly desirable. Hence, in this thesis, we present a real-space, local MP2 method based on DLPNO theory, employing periodicity. We use local approximations to evaluate unit cell correlation energies, without fragmenting the material. Through this scheme, we hope to probe ground state properties such as lattice constants and cohesive energies, for insulating and semi-conducting systems, and to probe surface adsorption interactions. We now proceed with outlining the formalism our periodic DLPNO-MP2 scheme.

4.3 Local MP2 for Periodic Systems

We return to the periodic Hartree–Fock equations, first outlined in Section 2.3. The overlap and Fock matrices are block diagonal in the Bloch AO basis, due to translational symmetry, and the HF orbitals and eigenvalues satisfy

$$\mathbf{F}^{\mathbf{k}}\mathbf{C}^{\mathbf{k}} = \epsilon^{\mathbf{k}}\mathbf{S}^{\mathbf{k}}\mathbf{C}^{\mathbf{k}} \tag{4.11}$$

where

$$\langle \mu_{\mathbf{k}} | \nu_{\mathbf{k}'} \rangle = \delta_{\mathbf{k}\mathbf{k}'} S^{\nu_{\mathbf{k}'}}_{\mu_{\mathbf{k}}} \tag{4.12}$$

$$\langle \mu_{\mathbf{k}} | \hat{F} | \nu_{\mathbf{k}'} \rangle = \delta_{\mathbf{k}\mathbf{k}'} F^{\nu_{\mathbf{k}'}}_{\mu_{\mathbf{k}}}. \tag{4.13}$$

Solution of the HF equations yields the canonical crystal orbitals (COs) $|p_{\mathbf{k}}\rangle$, also referred to as Bloch functions, and orbital eigenvalues $\epsilon_{p_{\mathbf{k}}}$,

$$|p_{\mathbf{k}}\rangle = \sum_{\mu} |\mu_{\mathbf{k}}\rangle \, C^{\mathbf{k}}_{\mu,p}. \tag{4.14}$$

Under a spin-free MP2 formalism, the amplitude and energy equations in the canonical basis are

$$0 = g^{i_{\mathbf{k}_1} j_{\mathbf{k}_2}}_{a_{\mathbf{k}_3} b_{\mathbf{k}_4}} + (\epsilon_{a_{\mathbf{k}_3}} + \epsilon_{b_{\mathbf{k}_4}} - \epsilon_{i_{\mathbf{k}_1}} - \epsilon_{j_{\mathbf{k}_2}}) t^{i_{\mathbf{k}_1} j_{\mathbf{k}_2}}_{a_{\mathbf{k}_3} b_{\mathbf{k}_4}}, \tag{4.15}$$

$$E_{\text{corr}} = \frac{1}{N} \sum_{i_{\mathbf{k}_1} j_{\mathbf{k}_2} a_{\mathbf{k}_3} b_{\mathbf{k}_4}} (2 g^{i_{\mathbf{k}_1} j_{\mathbf{k}_2}}_{a_{\mathbf{k}_3} b_{\mathbf{k}_4}} - g^{i_{\mathbf{k}_1} j_{\mathbf{k}_2}}_{b_{\mathbf{k}_4} a_{\mathbf{k}_3}}) t^{i_{\mathbf{k}_1} j_{\mathbf{k}_2}}_{a_{\mathbf{k}_3} b_{\mathbf{k}_4}}, \tag{4.16}$$

where $g^{i_{\mathbf{k}_1} j_{\mathbf{k}_2}}_{a_{\mathbf{k}_3} b_{\mathbf{k}_4}} = \langle i_{\mathbf{k}_1} j_{\mathbf{k}_2} | r_{12}^{-1} | a_{\mathbf{k}_3} b_{\mathbf{k}_4} \rangle$ are the Bloch electron repulsion integrals (ERIs) and E_{corr} is the second-order correlation energy per unit cell. Here, we have introduced the doubles amplitudes $t^{i_{\mathbf{k}_1} j_{\mathbf{k}_2}}_{a_{\mathbf{k}_3} b_{\mathbf{k}_4}}$. Since the Fock matrix is diagonal, the amplitudes can be substituted to obtain a closed form expression for the correlation energy, exactly like Eq.4.10.

The canonical COs are delocalized throughout the entire crystal and are not suitable for local correlation approximations. By applying an inverse FT, one can generate Wannier Functions (WFs)[99] $|p_{\mathbf{l}}\rangle$, which are a real space representation of the orbitals, each centered on a unit cell given by lattice vector $\mathbf{l}$,

$$|p_{\mathbf{l}}\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}}^{N} e^{-i\mathbf{k}\cdot\mathbf{l}} \, |p_{\mathbf{k}}\rangle \, . \tag{4.17}$$

Wannier functions obey translational symmetry, whereby each function $|p_{\mathbf{l}}\rangle$ is a translational copy of the corresponding function in the reference cell $|p_{\mathbf{0}}\rangle$, noting that the translation may wrap around the periodic supercell boundary. Wannier functions can be further localized by rotating the Bloch functions $|p_{\mathbf{k}}\rangle$ at each k-point among themselves prior to Fourier transformation. We employ our Wannier localisation scheme[1], presented in Chapter 3, to obtain well-localised Wannier functions by optimizing a fourth-order Pipek–Mezey[102] metric using atomic charges from Bloch intrinsic atomic orbitals (IAOs)[117].

In the Wannier basis, the MP2 amplitude and energy equations become

$$\begin{aligned} 0 =& g^{i_{\mathbf{l}_1} j_{\mathbf{l}_2}}_{a_{\mathbf{l}_3} b_{\mathbf{l}_4}} + \sum_{c_{\mathbf{l}_5}} (f^{c_{\mathbf{l}_5}}_{a_{\mathbf{l}_3}} t^{i_{\mathbf{l}_1} j_{\mathbf{l}_2}}_{c_{\mathbf{l}_5} b_{\mathbf{l}_4}} + f^{c_{\mathbf{l}_5}}_{b_{\mathbf{l}_2}} t^{i_{\mathbf{l}_1} j_{\mathbf{l}_2}}_{a_{\mathbf{l}_3} c_{\mathbf{l}_5}}) \\ & - \sum_{k_{\mathbf{l}_5}} (t^{k_{\mathbf{l}_5} j_{\mathbf{l}_2}}_{a_{\mathbf{l}_3} b_{\mathbf{l}_4}} f^{i_{\mathbf{l}_1}}_{k_{\mathbf{l}_5}} + t^{i_{\mathbf{l}_1} k_{\mathbf{l}_5}}_{a_{\mathbf{l}_3} b_{\mathbf{l}_4}} f^{j_{\mathbf{l}_2}}_{k_{\mathbf{l}_5}}), \end{aligned} \tag{4.18}$$

$$E_{\mathrm{corr}} = \frac{2}{1 + \delta_{i_{\mathbf{0}} j_{\mathbf{l}}}} \sum_{i_{\mathbf{0}} \leq j_{\mathbf{l}_2}} \sum_{a_{\mathbf{l}_3} b_{\mathbf{l}_4}} (2 g^{i_{\mathbf{0}} j_{\mathbf{l}_2}}_{a_{\mathbf{l}_3} b_{\mathbf{l}_4}} - g^{i_{\mathbf{0}} j_{\mathbf{l}_2}}_{b_{\mathbf{l}_4} a_{\mathbf{l}_3}}) t^{i_{\mathbf{0}} j_{\mathbf{l}_2}}_{a_{\mathbf{l}_3} b_{\mathbf{l}_4}}, \tag{4.19}$$

where the electron repulsion integrals (ERIs) in the BvK AO basis are evaluated as

$$g^{\kappa_{\mathbf{l}_3} \lambda_{\mathbf{l}_4}}_{\mu_{\mathbf{l}_1} \nu_{\mathbf{l}_2}} = \sum_{\mathbf{L}_2 \mathbf{L}_3 \mathbf{L}_4}^{\infty} \langle \mathring{\mu}_{\mathbf{l}_1} \mathring{\nu}_{\mathbf{l}_2 + \mathbf{L}_2} | r_{12}^{-1} | \mathring{\kappa}_{\mathbf{l}_3 + \mathbf{L}_3} \mathring{\lambda}_{\mathbf{l}_4 + \mathbf{L}_4} \rangle. \tag{4.20}$$

The Fock matrix elements are

$$f^{q_{\mathbf{m}}}_{p_{\mathbf{l}}} = \frac{1}{N} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{l}} f^{q_{\mathbf{k}}}_{p_{\mathbf{k}}} e^{i\mathbf{k}\cdot\mathbf{m}} \tag{4.21}$$

$$\mathbf{f}^{\mathbf{k}} = \mathbf{U}^{\mathbf{k}\dagger} \mathbf{F}^{\mathbf{k}} \mathbf{U}^{\mathbf{k}}, \tag{4.22}$$

where $\mathbf{U}^{\mathbf{k}}$ are the unitary matrices obtained from the Bloch IAO localization procedure. The restricted summation in Eq.4.19 counts each pair interaction once and only includes pairs where at least one orbital is in the reference unit cell, which is at lattice displacement $\mathbf{0}$. We adopt the convention that the reference cell is at the centre of the supercell and use an odd number of k-points in each dimension. The amplitude equations are solved iteratively, since the Fock matrix elements couple together.

Having transformed to the Wannier basis, local approximations can be applied to reduce the computational costs. Electrons in distant orbitals $j_{\mathbf{l}}$ have negligible correlation with electrons in orbital $i_{\mathbf{0}}$ of the reference unit cell and the number of pairs $i_{\mathbf{0}}j_{\mathbf{l}}$ with pair energy greater than ϵ tends to a constant as the size of the supercell is increased. In the PNO approach to local correlation, a model pair density is used to determine an $\mathcal{O}(1)$ set of pair-specific localized virtual orbitals $\{\tilde{a}_{i_{\mathbf{0}}j_{\mathbf{l}}}\}$ adapted to describe the correlation of each occupied pair $i_{\mathbf{0}}j_{\mathbf{l}}$. The error incurred due to discarding virtuals is proportional to $\sqrt{\mathcal{T}_{\mathrm{PNO}}}$, where $\mathcal{T}_{\mathrm{PNO}}$ is the PNO occupation number threshold that defines the PNO subspace[191,192]. In the DLPNO approach, the PNOs are expanded in a pair-specific domain of PAOs $\{\tilde{\mu}_{\mathbf{m}}\}$

$$|\tilde{a}_{i_{\mathbf{0}}j_{\mathbf{l}}}\rangle = \sum_{\tilde{\mu}_{\mathbf{m}} \in \mathcal{D}_{i_{\mathbf{0}}j_{\mathbf{l}}}} |\tilde{\mu}_{\mathbf{m}}\rangle \tilde{C}^{\tilde{a}_{i_{\mathbf{0}}j_{\mathbf{l}}}}_{\tilde{\mu}_{\mathbf{m}}}. \tag{4.23}$$

The energy and amplitude equations are solved iteratively in the PNO basis, analogous to the molecular case.

A linear scaling algorithm for the correlation energy can be constructed in the same way as for the molecular case by forming the PNOs through a PAO to OSV to PNO subspace compression cascade and using local density (DF) fitting for the ERIs. In the local DF approximation, the ERIs are computed using the robust formula

$$g^{i_{\mathbf{0}}j_{\mathbf{l}}}_{\tilde{a}\tilde{b}} = \sum_{Q_{\mathbf{m}}, P_{\mathbf{n}}} (i_{\mathbf{0}}\tilde{a}|Q_{\mathbf{m}})(Q_{\mathbf{m}}|P_{\mathbf{n}})^{-1}(P_{\mathbf{n}}|j_{\mathbf{l}}\tilde{b}), \tag{4.24}$$

where $Q_{\mathbf{m}}, P_{\mathbf{n}}$ are the subset of auxiliary basis functions local to pair $i_{\mathbf{0}}j_{\mathbf{l}}$. The local fitting domain is the set of functions with overlaps $(i_{\mathbf{0}}i_{\mathbf{0}}Q_{\mathbf{m}}Q_{\mathbf{m}})$ or $(j_{\mathbf{l}}j_{\mathbf{l}}P_{\mathbf{n}}P_{\mathbf{n}})$ above a DF threshold, which is linked to the PNO threshold. The largest PAO domains at the start of the PAO-OSV-PNO cascade are also selected on the basis of integral estimates, which ultimately determines the pair-specific PAO domains for PNOs. As supercell size increases, the auxiliary basis and PAO domains for a given pair grow to a constant size.

In contrast to the molecular case, however, the BvK boundary conditions require that all functions are periodic over the supercell. Consequently, overlap, Fock and

ERI integrals involve summing each function's contribution within all periodic images of the supercell, up to the thermodynamic limit, as shown in Eq 2.11 and Eq 4.20. This lattice summation presents a significant complication, in particular to the ERIs, as mentioned in Section 2.1. In the following section, the general details of the PAO-OSV-PNO cascade are outlined, but the specific strategy for the evaluation of the periodic ERIs are not discussed. The focus of the Chapter 5 concerns two different treatments for this problem, leading consequently to two significantly differing implementations of periodic DLPNO-MP2.

4.4 DLPNO-MP2 formalism

4.4.1 PAOs

The BvK PAOs are the projection of the BvK AO onto the space spanned by the virtual Wannier orbitals $a_{\mathbf{l}}$

$$|\tilde{\mu}_{\mathbf{0}}\rangle = \sum_{a_{\mathbf{l}}} |a_{\mathbf{l}}\rangle\langle a_{\mathbf{l}}|\mu_{\mathbf{0}}\rangle = \sum_{\nu_{\mathbf{m}}} |\nu_{\mathbf{m}}\rangle \hat{C}^{\mu_{\mathbf{0}}}_{\nu_{\mathbf{m}}}, \tag{4.25}$$

where the coefficient matrix can be expressed as

$$\hat{C}^{\mu_{\mathbf{0}}}_{\nu_{\mathbf{m}}} = \sum_{a_{\mathbf{l}}\kappa_{\mathbf{n}}} C^{a_{\mathbf{l}}}_{\nu_{\mathbf{m}}} C^{a_{\mathbf{l}}}_{\kappa_{\mathbf{n}}} S^{\kappa_{\mathbf{n}}}_{\mu_{\mathbf{0}}}. \tag{4.26}$$

The overlap of the BvK AO functions within the BvK cell is evaluated as a sum over regular overlap integrals for the infinite periodic images of the individual AOs

$$S^{\kappa_{\mathbf{n}}}_{\mu_{\mathbf{0}}} = \langle \mu_{\mathbf{0}}|\kappa_{\mathbf{n}}\rangle = \sum_{\mathbf{N}}^{\infty} \langle \mathring{\mu}_{\mathbf{0}}|\mathring{\kappa}_{\mathbf{n}+\mathbf{N}}\rangle. \tag{4.27}$$

Since the Wannier functions and BvK AOs are all translational copies of the corresponding functions in the reference cell $\mathbf{0}$, the PAOs also inherit translational symmetry $\hat{C}^{\mu_{\mathbf{l}}}_{\nu_{\mathbf{m}}} = \hat{C}^{\mu_{\mathbf{0}}}_{\nu_{\mathbf{m}-\mathbf{l}}}$. Since the PAOs form a redundant set, there is no unique expansion of the virtuals in terms of PAOs. We choose to define a set of contravariant PAOs, whereby the transformation matrix from the virtuals to the contravariant PAOs is obtained from the right-hand Moore–Penrose pseudo inverse of $\hat{C}^{\mu_{\mathbf{l}}}_{\nu_{\mathbf{m}}}$ [90].

Our approach to forming OSVs employs a numerical Laplace transformation and we require Laplace transformed BvK PAOs. These are defined as

$$|\tilde{\mu}^z_{\mathbf{0}}\rangle = \sum_{a_{\mathbf{k}}} |a_{\mathbf{k}}\rangle e^{-(\epsilon_{a_{\mathbf{k}}}-\epsilon_F)t_z}\langle a_{\mathbf{k}}|\tilde{\mu}_{\mathbf{0}}\rangle \tag{4.28}$$

$$= \sum_{\nu_{\mathbf{m}}} |\nu_{\mathbf{m}}\rangle \hat{C}^{\mu_{\mathbf{0}},z}_{\nu_{\mathbf{m}}} \tag{4.29}$$

where ϵ_F is the Fermi level and t_z is a Laplace integration grid point. The expansion coefficients after the second equality are expressed as

$$\hat{C}^{\mu_{\mathbf{0}},z}_{\nu_{\mathbf{m}}} = \sum_{a\mathbf{n}\mathbf{n}'\kappa_{\mathbf{m}}} C^{a_{\mathbf{n}'}}_{\nu_{\mathbf{m}}} L^{a,z}_{\mathbf{n}'\mathbf{n}} C^{a_{\mathbf{n}}}_{\kappa_{\mathbf{l}}} S^{\kappa_{\mathbf{l}}}_{\mu_{\mathbf{0}}} \tag{4.30}$$

$$L^{a,z}_{\mathbf{n}'\mathbf{n}} = \frac{1}{N}\sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{n}'} e^{-(\epsilon_{a_{\mathbf{k}}}-\epsilon_F)t_z} e^{i\mathbf{k}\cdot\mathbf{n}}. \tag{4.31}$$

PAOs $|\tilde{\mu}^z_{\mathbf{l}}\rangle$ are also all translational copies of $|\tilde{\mu}^z_{\mathbf{0}}\rangle$.

4.4.2 OSVs

Orbital specific virtuals are PNOs for diagonal pairs $|i_{\mathbf{l}}i_{\mathbf{l}}\rangle$. Our approach is to closely follow the molecular PNO-MP2 implementation and to obtain OSVs as an intermediary for the purpose of accelerating the determination of PNOs. We obtain OSVs for each orbital $|i_{\mathbf{l}}\rangle$ as eigenvectors of an approximate external density matrix for pair $|i_{\mathbf{l}}i_{\mathbf{l}}\rangle$, which we construct in a prinicpal domain of PAOs $\tilde{\mu} \in \mathcal{D}_{i_{\mathbf{l}}}$ selected using the greedy algorithm described in Ref.90.

The external density is computed from first-order amplitudes for diagonal pairs approximated using a numerical Laplace transformation.[87,193,194]. The integration points z and weights w_z are determined from the orbital eigenvalues of the supercell in the same way as for molecular calculations,

$$t^{i_{\mathbf{0}}i_{\mathbf{0}}}_{\tilde{\nu}_{\mathbf{l}}\hat{\kappa}_{\mathbf{n}}} = \sum_z w_z \langle \tilde{\nu}^z_{\mathbf{l}} \hat{\kappa}^z_{\mathbf{n}} | i^z_{\mathbf{0}} i^z_{\mathbf{0}}\rangle \tag{4.32}$$

$$D^{\tilde{\nu}_{\mathbf{m}}}_{\tilde{\mu}_{\mathbf{l}}} = \sum_{\kappa_{\mathbf{n}}} 4 t^{i_{\mathbf{0}}i_{\mathbf{0}}}_{\tilde{\mu}_{\mathbf{l}}\hat{\kappa}_{\mathbf{n}}} t^{i_{\mathbf{0}}i_{\mathbf{0}}}_{\tilde{\nu}_{\mathbf{m}}\hat{\kappa}_{\mathbf{n}}}. \tag{4.33}$$

Specifically, the OSV coefficients $C^{\tilde{a}}_{\tilde{\nu}_{\mathbf{m}}}$ and occupation numbers $n_{\tilde{a}}$ for orbital $i_{\mathbf{0}}$ are obtained via

$$\sum_{\nu_{\mathbf{m}}} D^{\tilde{\nu}_{\mathbf{m}}}_{\tilde{\mu}_{\mathbf{l}}} C^{\tilde{a}}_{\tilde{\nu}_{\mathbf{m}}} = \sum_{\nu_{\mathbf{m}}} S^{\tilde{\nu}_{\mathbf{m}}}_{\tilde{\mu}_{\mathbf{l}}} C^{\tilde{a}}_{\tilde{\nu}_{\mathbf{m}}} n_{\tilde{a}}. \tag{4.34}$$

Here $S_{\tilde{\mu}}^{\tilde{\nu}}$ is the overlap between the selected PAOs $S_{\tilde{\mu}_\mathbf{l}}^{\tilde{\nu}_\mathbf{m}} = \langle \tilde{\mu}_\mathbf{l} | \tilde{\nu}_\mathbf{m} \rangle$ and $C_{\tilde{\nu}_\mathbf{m}}^{\tilde{a}}$ are the transformation coefficients from the PAOs to the OSVs $|\tilde{a}_{i_\mathbf{0}}\rangle$, which are particular to each orbital $|i_\mathbf{0}\rangle$. Only those OSVs $\{\tilde{a}_{i_\mathbf{0}}\}$ with occupation numbers $n_{\tilde{a}}$ greater than a user defined threshold are retained. The OSVs for $|i_\mathbf{l}\rangle$ are simply $C_{\tilde{\nu}_{\mathbf{m}+\mathbf{l}}}^{\tilde{a}}$, obtained from those of $|i_\mathbf{0}\rangle$ by translation.

Once the OSVs are determined, all $i_\mathbf{0} j_\mathbf{l}$ pairs are screened to select only pairs with significant correlation energies. The OSV-SOS-MP2 (spin-opposite-scaled) energy is used to compute approximate pair energies for the purpose of discarding insignificant pairs and providing an estimate of their contribution to the total energy,

$$E_{\mathrm{SOS}}^{i_\mathbf{0} j_\mathbf{l}} = -\sum_{\tilde{a}\in[i_\mathbf{0}],\tilde{b}\in[j_\mathbf{l}]} \frac{g_{\tilde{a}\tilde{b}}^{i_\mathbf{0} j_\mathbf{l}} g_{\tilde{a}\tilde{b}}^{i_\mathbf{0} j_\mathbf{l}}}{\varepsilon_{\tilde{a}} + \varepsilon_{\tilde{b}} - f_{i_\mathbf{0}}^{i_\mathbf{0}} - f_{j_\mathbf{l}}^{j_\mathbf{l}}} \,. \tag{4.35}$$

The OSV-SOS-MP2 energy does not contain exchange contributions, but this is well-motivated since the MP2 correlation energy is dominated by the Coulomb part, particularly for distant pairs. It can be evaluated efficiently using asymmetric density fitting.

4.4.3 PNOs

Having obtained OSVs for every orbital $|i_\mathbf{l}\rangle$, we construct initial pair specific virtuals $\{\bar{a}_{i_\mathbf{0} j_\mathbf{l}}\}$ for each pair $|i_\mathbf{0} j_\mathbf{l}\rangle$ by merging the OSVs for $i_\mathbf{0}$ and $j_\mathbf{l}$. In addition, we construct initial PAO and DF pair domains from the union of the respective domains of $i_\mathbf{0}$ and $j_\mathbf{l}$. For each pair, the PNOs are formed as eigenvectors of an approximate external density matrix constructed within this pair specific subspace. First-order amplitudes are constructed using the semi-canonical approximation, where the pair specific orbitals are canonicalised to diagonalise the virtual block of the Fock matrix for each pair $f_{\bar{a}}^{\bar{b}} = \varepsilon_{\bar{a}} \delta_{\bar{a}\bar{b}}$ and off-diagonal occupied Fock matrix elements are neglected,

$$t_{\bar{a}\bar{b}}^{i_\mathbf{0} j_\mathbf{l}} = -(\varepsilon_{\bar{a}} - \varepsilon_{\bar{b}} - f_{i_\mathbf{0}}^{i_\mathbf{0}} - f_{j_\mathbf{l}}^{j_\mathbf{l}})^{-1} g_{\bar{a}\bar{b}}^{i_\mathbf{0} j_\mathbf{l}}. \tag{4.36}$$

The approximate external density is then computed by

$$D_{\bar{a}}^{\bar{b}} = 2\sum_{\bar{c}}(t_{\bar{a}\bar{c}}^{i_{\mathbf{0}}j_{\mathbf{l}}} u_{\bar{b}\bar{c}}^{i_{\mathbf{0}}j_{\mathbf{l}}} + t_{\bar{c}\bar{a}}^{i_{\mathbf{0}}j_{\mathbf{l}}} u_{\bar{c}\bar{b}}^{i_{\mathbf{0}}j_{\mathbf{l}}}) \tag{4.37}$$

$$u_{\bar{a}\bar{b}}^{i_{\mathbf{0}}j_{\mathbf{l}}} = 2t_{\bar{a}\bar{b}}^{i_{\mathbf{0}}j_{\mathbf{l}}} - t_{\bar{b}\bar{a}}^{i_{\mathbf{0}}j_{\mathbf{l}}}. \tag{4.38}$$

Finally, diagonalisation of the of the density yields the PNOs as eigenvectors,

$$\sum_{\bar{b}} D_{\bar{a}}^{\bar{b}} C_{\bar{b}}^{\tilde{a}} = n_{\tilde{a}} C_{\bar{a}}^{\tilde{a}}. \tag{4.39}$$

Optionally, a further truncation of the PAO domains to principle pair domains can be performed prior to finding the eigenvectors of the external pair density matrix. The final space of retained PNOs are those with occupation numbers greater than the user defined PNO threshold $n_{\tilde{a}} > \mathcal{T}_{\mathrm{PNO}}$ and the fully coupled first-order amplitude equations are solved in this subspace. Since the PNOs and corresponding integrals and amplitudes for pairs $|i_{\mathbf{l}}j_{\mathbf{m}}\rangle$ are translational copies of $|i_{\mathbf{0}}j_{\mathbf{l}}\rangle$, it is only necessary to compute and store those for $|i_{\mathbf{0}}j_{\mathbf{l}}\rangle$.

4.4.4 DLPNO-MP2 Residuals and Energy

The MP2 energy under the semi-canonical approximation (DLPNO-SC-MP2) can be computed as a byproduct of forming PNOs,

$$E_{\mathrm{SC}}^{i_{\mathbf{0}}j_{\mathbf{l}}} = \sum_{\bar{a}\bar{b}\in[i_{\mathbf{0}}j_{\mathbf{l}}]} u_{\bar{a}\bar{b}}^{i_{\mathbf{0}}j_{\mathbf{l}}} g_{\bar{a}\bar{b}}^{i_{\mathbf{0}}j_{\mathbf{l}}}. \tag{4.40}$$

The difference between the DLPNO-SC-MP2 energy before and after truncation of the PNO space to $\{\tilde{a}_{i_{\mathbf{0}}j_{\mathbf{l}}}\}$ provides an estimate for the contribution of the discarded PNOs to the total energy. The SC-MP2 energy neglects the coupling due to the off-diagonal Fock elements. The final amplitude equations for periodic DLPNO-MP2 are analogous to those of non-periodic DLPNO-MP2, and are solved in the space of retained PNOs,

$$\begin{aligned} 0 =& g_{\tilde{a}\tilde{b}}^{i_{\mathbf{0}}j_{\mathbf{l}}} + (\varepsilon_{\tilde{a}} + \varepsilon_{\tilde{b}}) t_{\tilde{a}\tilde{b}}^{i_{\mathbf{0}}j_{\mathbf{l}}} \\ &- \sum_{k_{\mathbf{m}}} \sum_{\tilde{c}\tilde{d}\in[k_{\mathbf{m}}j_{\mathbf{l}}]} S_{\tilde{a},i_{\mathbf{0}}j_{\mathbf{l}}}^{\tilde{c},k_{\mathbf{m}}j_{\mathbf{l}}} S_{\tilde{b},i_{\mathbf{0}}j_{\mathbf{l}}}^{\tilde{d},k_{\mathbf{m}}j_{\mathbf{l}}} t_{\tilde{c}\tilde{d}}^{k_{\mathbf{m}}j_{\mathbf{l}}} f_{k_{\mathbf{m}}}^{i_{\mathbf{0}}} \\ &- \sum_{k_{\mathbf{m}}} \sum_{\tilde{c}\tilde{d}\in[i_{\mathbf{0}}k_{\mathbf{m}}]} S_{\tilde{a},i_{\mathbf{0}}j_{\mathbf{l}}}^{\tilde{c},i_{\mathbf{0}}k_{\mathbf{m}}} S_{\tilde{b},i_{\mathbf{0}}j_{\mathbf{l}}}^{\tilde{d},i_{\mathbf{0}}k_{\mathbf{m}}} t_{\tilde{c}\tilde{d}}^{i_{\mathbf{0}}k_{\mathbf{m}}} f_{k_{\mathbf{m}}}^{j_{\mathbf{l}}} \end{aligned} \tag{4.41}$$

where $\tilde{a}\tilde{b} \in [i_{\mathbf{0}}j_{\mathbf{l}}]$ and $S^{\tilde{c},k_{\mathbf{m}}j_{\mathbf{l}}}_{\tilde{a},i_{\mathbf{0}}j_{\mathbf{l}}}$ is the overlap between PNOs for pair $i_{\mathbf{0}}j_{\mathbf{l}}$ and pair $k_{\mathbf{m}}j_{\mathbf{l}}$. The amplitude equations are solved iteratively for every pair $|i_{\mathbf{0}}j_{\mathbf{l}}\rangle$ and the correlation energy per unit cell is then evaluated as

$$E_{\mathrm{corr}} = \frac{2}{1+\delta_{i_{\mathbf{0}}j_{\mathbf{l}}}} \sum_{i_{\mathbf{0}} \leq j_{\mathbf{l}}} \sum_{\tilde{a}\tilde{b} \in [i_{\mathbf{0}}j_{\mathbf{l}}]} (2g^{i_{\mathbf{0}}j_{\mathbf{l}}}_{\tilde{a}\tilde{b}} - g^{i_{\mathbf{0}}j_{\mathbf{l}}}_{\tilde{b}\tilde{a}}) t^{i_{\mathbf{0}}j_{\mathbf{l}}}_{\tilde{a}\tilde{b}} + \Delta \tag{4.42}$$

where Δ is the correction term, formed by summing the differences between the DLPNO-SC-MP2 energies before and after truncation, as well as the OSV-SOS-MP2 energy estimates of the discarded pairs.

Since the WFs and PNO domains are rigorously translationally symmetric, the amplitudes, overlaps and Fock matrix elements spanning the entire supercell can be generated from the translationally unique set of objects where at least one index resides in the reference cell. For example, $f^{j_{\mathbf{l}}}_{k_{\mathbf{m}}} = f^{j_{\mathbf{0}}}_{k_{\mathbf{m-l}}}$ and $t^{k_{\mathbf{m}}j_{\mathbf{l}}}_{\tilde{c}\tilde{d}} = t^{k_{\mathbf{0}}j_{\mathbf{l-m}}}_{\tilde{c}\tilde{d}}$, where it is understood that modulo arithmetic is applied to the lattice vectors. The periodic DLPNO-MP2 formalism is completely analogous to the molecular case, except that orbital coefficients, integrals and amplitudes obey translational invariance, and that the integrals contain an in-principle infinite summation over periodic supercell images. The molecular DLPNO implementation[87–90] within TURBOMOLE can thus be repurposed to treat periodic systems, which also enables both molecular and periodic settings to be treated on an equal footing.

5| Implementing Periodic DLPNO-MP2

The work in this chapter contributed to the publications Ref.2 and Ref.3. Arman Nejad developed the formalism and the implementation for the surface-dipole corrected, chargeless local density fitting approach, within the BvK-DLPNO-MP2 scheme, presented in Ref.2.

In this chapter, we present the work involved in implementing periodic DLPNO-MP2. After reviewing the overall scheme, within the TURBOMOLE[81,82] software package, we directly address the challenges of periodic Coulomb sums, previously ignored in Chapter 4, in Section 5.2. The challenges in evaluating the electron repulsion integrals (ERIs) lead us to present two different implementations of periodic DLPNO-MP2. We present the 'Megacell' approach in Section 5.3, where a supercell correlation treatment is embedded within a megacell, and translational symmetry is rigorously imposed. In Section 5.4, a differing approach, enforcing Born–von-Karman (BvK) boundary conditions, is outlined, and we detail the chargeless and surface-dipole corrected density fitting scheme required to obtain convergent ERIs.

5.1 Overview of scheme

In Chapter 4, we outlined the underlying theory for a DLPNO-MP2 method under periodic boundary conditions, outlining the specific PAO-OSV-PNO[87–90] cascade that has already been implemented for molecular systems, inside the TURBOMOLE software package.

In this chapter, we shift to focus on the key challenges in implementing such a method, within the existing functionalities of TURBOMOLE. In Figure 5.1, the key steps within the DLPNO-MP2 method are presented in the central flowchart, where the steps for the calculations of the PNOs, including the PAO-OSV-PNO cascade, are presented in further detail, on the flowchart on the right, in pink. The molecular DLPNO-MP2 code is housed in the `pnoccsd` module.

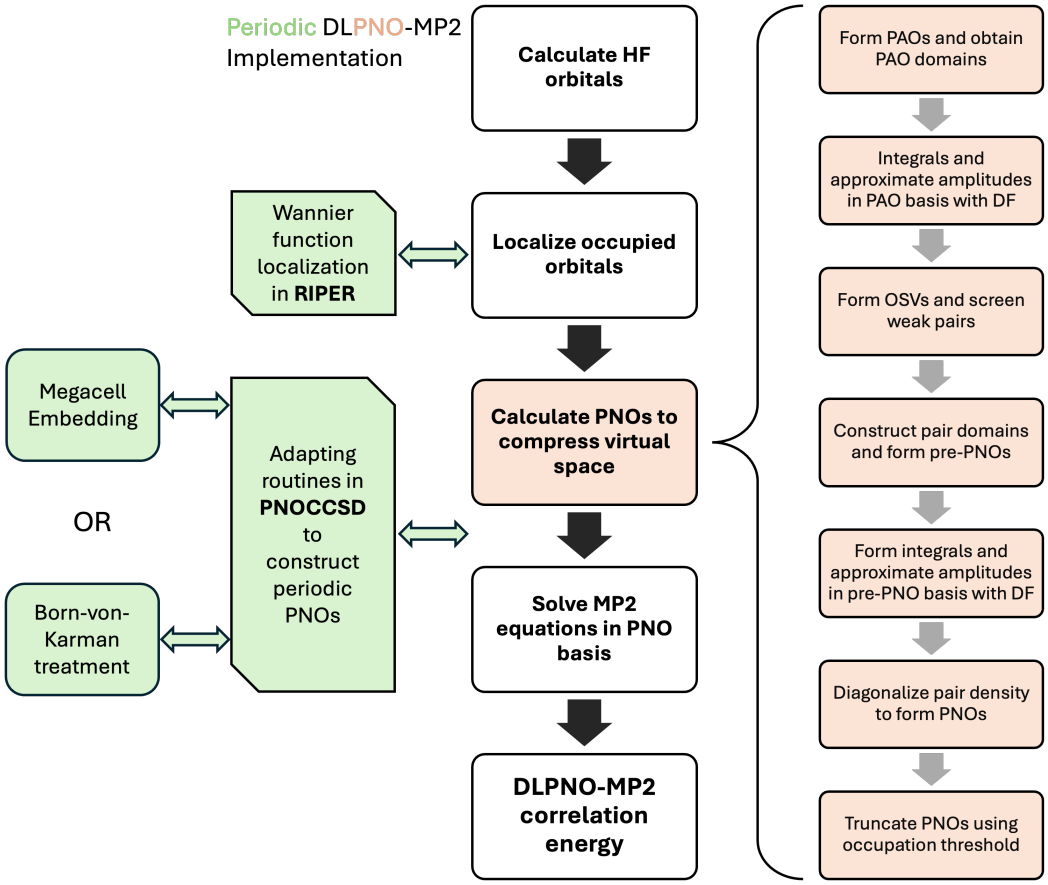

Figure 5.1: Schematic outlining overall key steps of the DLPNO-MP2 method within the TURBOMOLE software package. A detailed flow chart of the PAO-OSV-PNO cascade to construct PNOs are presented in pink. A summary of adaptations to obtain a periodic scheme are given in green.

The key implementation changes, to adapt the routines within TURBOMOLE for a periodic DLPNO-MP2 method, are given in green, in Figure 5.1. The Wannier localisation scheme, the focus of Chapter 3, has been implemented in the `riper` module, which also houses the developmental periodic HF scheme, from Sierka and coworkers[83–86,97,98]. The Wannier localisation scheme has been implemented as a post-processing function after the self-consistent orbitals have been generated. The translationally unique set of Wannier function coefficients and Wannier gauge Fock matrix elements are provided as output of the `riper` program, and used subsequently as input in the `pnoccsd` program.

Several adaptations to the routines within the `pnoccsd` module are required to create a scheme suitable for periodic systems. Although the overall theory

for periodic DLPNO-MP2 is presented in Chapter 4, we omitted any discussion surrounding the treatment of periodic electron repulsion integrals (ERIs). The correct treatment of these Coulomb sums is a vital question, arguably the most important within this thesis. In the remaining sections of this chapter, we explain the challenges of correcting computing Coulomb lattice sums, and then describe two separate strategies for periodic DLPNO-MP2, which we denote in shorthand as the 'Megacell' and 'BvK' (Born–von-Karman) approaches, respectively.

5.2 Challenges with Coulomb lattice summation

We revisit the ERI, in the BvK AO basis,

$$g_{\mu_{\mathbf{l}_1}\nu_{\mathbf{l}_2}}^{\kappa_{\mathbf{l}_3}\lambda_{\mathbf{l}_4}} = \sum_{\mathbf{L}_2\mathbf{L}_3\mathbf{L}_4}^{\infty} \langle \mathring{\mu}_{\mathbf{l}_1} \mathring{\nu}_{\mathbf{l}_2+\mathbf{L}_2} | r_{12}^{-1} | \mathring{\kappa}_{\mathbf{l}_3+\mathbf{L}_3} \mathring{\lambda}_{\mathbf{l}_4+\mathbf{L}_4} \rangle, \tag{5.1}$$

which features a infinite summation of all the periodic supercell images of the AOs $\mathring{\nu}, \mathring{\kappa}, \mathring{\lambda}$. PNO methods require ERIs in a different virtual basis for each occupied pair. Direct integral transformation of these integrals would incur prohibitively high costs, and so the density fitting (DF) approximation is employed. The electron density is approximated as a linear combination of auxiliary basis functions $\{ P_{\mathbf{n}} \}$,

$$|i_{\mathbf{l}}\tilde{a}) \approx \sum_{P_{\mathbf{n}}} C_{i_{\mathbf{l}}\tilde{a}}^{P_{\mathbf{n}}} |P_{\mathbf{n}}), \tag{5.2}$$

where the fitting coefficients of the auxiliary basis functions are solved by minimising the least-squares error of the true density with the Coulomb metric. This allows for the decomposition of the ERI into a product of two and three-index intermediates,

$$g_{\tilde{a}\tilde{b}}^{i_{\mathbf{0}}j_{\mathbf{l}}} = \sum_{Q_{\mathbf{m}},P_{\mathbf{n}}} (i_{\mathbf{0}}\tilde{a}|Q_{\mathbf{m}})(Q_{\mathbf{m}}|P_{\mathbf{n}})^{-1}(P_{\mathbf{n}}|j_{\mathbf{l}}\tilde{b}), \tag{5.3}$$

which accelerates integral transformation. The two and three-index intermediates implicitly involve the same infinite lattice summation of supercell images present in Eq.5.1, and consideration must be taken to the convergence character of these integrals.

Due to the inherently long-range nature of electrostatic interactions, the Coulomb energy between a charge density $\mathring{\rho}^A(\mathbf{r}_1)$ and an infinite lattice of charge densities

Table 5.1: Convergence properties of Coulomb lattice sums between multipole contributions to charge densities $\sum_{\mathbf{L}}(\mathring{\rho}^A|\mathring{\rho}^B_{\mathbf{L}})$.

$(\rho^A\|\rho^B)$	Decay	1D	2D	3D
$(s\|s)$	$(r_{12})^{-1}$	divergent	divergent	divergent
$(s\|p)$	$(r_{12})^{-2}$	absolute	conditional	divergent
$(s\|d)$	$(r_{12})^{-3}$	absolute	absolute	conditional
$(p\|p)$	$(r_{12})^{-3}$	absolute	absolute	conditional
$(s\|f)$	$(r_{12})^{-4}$	absolute	absolute	absolute
$(p\|d)$	$(r_{12})^{-4}$	absolute	absolute	absolute

$\mathring{\rho}^B_{\mathbf{L}}(\mathbf{r}_2)$ is not necessarily absolutely convergent, and for non-neutral charge densities, even divergent. Table 5.1 lists the convergence properties of lattice summations of the multipolar contributions to a charge distribution constituting of charges, dipoles, quadrupoles etc. which are denoted by s, p, d, etc. Since the relevant densities $\rho = i_{\mathbf{l}}\tilde{a}$ in MP2 are chargeless, the longest length-scale decay originates from dipole-dipole interactions. The conditional convergence of the dipole-dipole terms in 3D arises from a residual surface energy, whose value depends on the shape of the surface of the lattice sum as it expands to infinity. Obtaining the correct integrals requires that this surface energy is removed. An instructive example that illustrates the geometric nature of this effect is the Madelung constant of an NaCl cell. The cubic unit cell arrangement has zero cell dipole moment, resulting in a fast absolutely convergent lattice sum, whereas the rhombohedral unit cell only yields the correct value if the surface effect, due to the cell dipole moment, is properly considered.[195,196].

After the pioneering works by Born and von Kármán, Madelung, Ewald and others,[91,197–199] the treatment of periodic lattice sums for efficient electronic structure theory remains an active field of interest. An inherent complication is that this property manifests differently in direct or reciprocal space treatments[200–202]. Contributions which are particularly relevant to this thesis include the many works in the context of LCAO-based Hartree–Fock,[84,196,203,204] and MP2,[43,52,59,205] periodic density fitting[84,206,207] and the multipole method.[84,85,202,208–211]

We now proceed to outline the two strategies we adopt to correctly account for the correct convergence of the Coulomb lattice sums.

5.3 Strategy 1: Megacell embedding

One possible strategy to a valid representation of the ERIs is to avoid the explicit enforcement of BvK boundary conditions, thus removing the infinite summation of supercell images altogether and circumventing the issues of conditional convergence. The complications associated with lattice summation can be removed under the condition that the functions are sufficiently close to the thermodynamic limit, since the summation imposed by the BvK boundary conditions disappears in the limit of an infinite sized supercell. To retain translational invariance, where orbitals at the edge of the simulated supercell are treated identically to those at the centre, it is necessary to account for a larger portion of the material surrounding and including the supercell, which we refer to as the megacell. This approach is thus called Megacell-DLPNO-MP2. The motivation of this scheme stems from significantly reduced costs of evaluating the ERIs, which are now essentially molecular, and the fact that the r^{-6} decay of the pair correlation energies is formally quicker than the r^{-3} of the MP2 ERIs.

In the BvK AO basis, the WFs are

$$|p_{\mathbf{l}}\rangle = \sum_{\mu_{\mathbf{m}}} |\mu_{\mathbf{m}+\mathbf{l}}\rangle \, C^{p\mathbf{0}}_{\mu_{\mathbf{m}}} \tag{5.4}$$

$$= \sum_{\mu_{\mathbf{m}}} \sum_{\mathbf{M}}^{\infty} |\mathring{\mu}_{\mathbf{m}+\mathbf{l}+\mathbf{M}}\rangle \, C^{p\mathbf{0}}_{\mu_{\mathbf{m}}}, \tag{5.5}$$

where the summation over $\mu_{\mathbf{m}}$ includes all functions in the supercell and we have used the translational symmetry of the WFs $C^{p\mathbf{l}}_{\mu_{\mathbf{m}}} = C^{p\mathbf{0}}_{\mu_{\mathbf{m}-\mathbf{l}}}$, where it is understood that $\mathbf{m}-\mathbf{l}$ is subject to modulo arithmetic within the supercell lattice. If the WF $|p_{\mathbf{l}}\rangle$ decays to zero within half a supercell extent in all periodic directions, then one can define WFs for the infinite crystal without periodic boundary conditions as

$$|p_{\mathbf{l}+\mathbf{L}}\rangle = \sum_{\mu_{\mathbf{m}}} |\mathring{\mu}_{\mathbf{m}+\mathbf{l}+\mathbf{L}}\rangle \, C^{p\mathbf{0}}_{\mu_{\mathbf{m}}}, \tag{5.6}$$

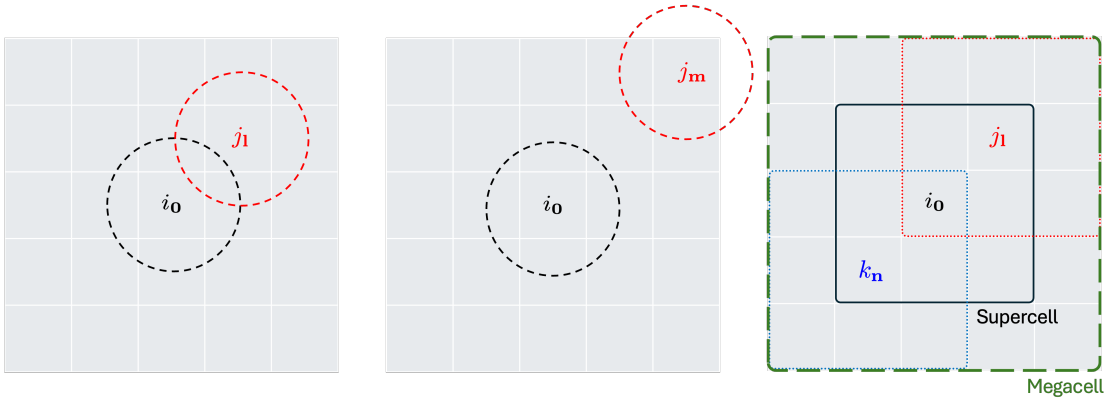

Figure 5.2: Wannier functions for a two-dimensional system defined by a simulation grid consisting of 5x5 unit cells. Dotted lines represent the extent of the WF, beyond which it has sufficiently decayed. The left subfigure shows a pair of WFs that are fully supported within the simulation grid, the middle subfigure shows a pair of WFs that are not. The right subfigure defines the boundaries of the supercell and megacell; all WFs in the supercell are properly represented and form a translationally symmetric orthonormal set that can be used for the correlation treatment.

where the summation over $\mu_\mathbf{m}$ only extends over one supercell. Provided that the WFs have decayed to zero at the supercell boundary, these WFs form an orthonormal set

$$\langle p_\mathbf{l} | q_\mathbf{m} \rangle = \delta_{pq} \delta_\mathbf{lm}, \tag{5.7}$$

where the overlap refers to a simple direct-space integral, without periodic boundary conditions.

To make use of this simplification in a practical calculation, every orbital used for the correlation treatment must be fully supported within the simulation, so that artifacts due to edge effects are avoided, thereby ensuring full translational symmetry. We therefore embed the *supercell* in a larger *megacell* such that all WFs centered in the supercell are sufficiently decayed at the boundary of the megacell. Only the WFs of the supercell are included in the correlation treatment. This is exemplified in Figure 5.2, which depicts, in a two-dimensional example, Wannier functions centered in different cells, with their radius of decay.

The megacell boundary adds on half of the extent of the supercell in each lattice vector. Specifically, the megacell lattice extent is given as $k_\text{mega} = 2k_\text{super} - 1$, where k_super is the supercell size. We find that well localised occupied WFs often

decay rapidly and the orthonormality error is less than one part in a million in our calculations. The condition that virtual WFs are sufficiently decayed to zero at the supercell boundary is rarely satisfied in practice, and is highly basis set dependent. The DLPNO approach, however, makes use of PAOs instead of virtual WFs, which do decay rapidly to zero, since their extent is directly linked to the occupied orbitals.

The MP2 residual and energy equations in the Wannier basis (Eqs.4.41 and 4.42) can now be employed to correlate all orbitals within the supercell without reference to BvK boundary conditions. Translational symmetry of the integrals and amplitudes remains, but translation extends into the megacell rather than wrapping around the periodic boundary. Coulomb and overlap integral evaluation therefore become identical to that of molecular calculations, without lattice summation. The thermodynamic limit is approached by increasing the size of the supercell to account for all significant pairwise correlations, which requires a commensurate increase in the size of the megacell.

5.3.1 Translational symmetry

The Megacell-DLPNO-MP2 method proceeds by first performing a periodic HF calculation in the megacell. Wannier occupied and virtual orbitals are obtained from the Bloch functions and the Fock matrix is transformed to the Wannier basis. The orbitals and Fock elements are then fed into the molecular DLPNO-MP2 code, which applies the PAO-OSV-PNO cascade and local density fitting to compute the correlation energy of the supercell. Periodic boundary conditions are not enforced and so all overlap metrics are molecular in the sense that no lattice summation is included. In contrast to the molecular case, each quantity has the translational symmetry of the crystal lattice. Therefore it is only necessary to compute and store the minimal set of unique quantities. In the following, we discuss how translational symmetry can be used to obtain sub-linear scaling in computational costs. A full account of all the subroutines changed and added to the `pnoccsd` module is presented in Appendix B. Only key steps and their subsequent efficiency improvements are presented here.

Computation of the ERIs in the PAO basis, using the local DF approximation, is the first computationally expensive task in the DLPNO procedure. Translational symmetry is exploited here to accelerate this step. Firstly, to preserve translational invariance, we restrict the domain of auxiliary basis functions to within the local supercell surrounding the occupied WF, so the combined domain for $i_{\mathbf{0}}j_{\mathbf{l}}$ does not extend beyond the megacell. The PAO domains are also similarly restricted to span the supercell surrounding the occupied WF. The two-index and three-index integrals exhibit translational symmetry

$$(Q_{\mathbf{m}}|P_{\mathbf{n}}) = (Q_{\mathbf{0}}|P_{\mathbf{n-m}}), \tag{5.8}$$

$$(i_{\mathbf{l}}\tilde{\mu}_{\mathbf{n}}|Q_{\mathbf{m}}) = (i_{\mathbf{0}}\tilde{\mu}_{\mathbf{n-l}}|Q_{\mathbf{m-l}}). \tag{5.9}$$

It is only necessary to compute and store the unique integrals where one index is in the reference cell. The computational cost of evaluating $(i_{\mathbf{0}}\tilde{\mu}_{\mathbf{n}}|Q_{\mathbf{m}})$ is asymptotically independent of supercell size once AO integral screening is applied, since the DF and PAO domains are of $\mathcal{O}(1)$ once the supercell is large enough for these domains to saturate.

Using the PAO ERIs, only the OSVs for the reference cell are computed, since those for the other cells can be obtained through translational symmetry. In the limit of supercell sizes where the number of DF functions and PAOs above the screening thresholds saturate, the cost of forming the OSVs becomes $\mathcal{O}(1)$ for each $i_{\mathbf{0}}$ and therefore $\mathcal{O}(1)$ overall. In the subsequent pair screening stage, only OSV-SOS-MP2 pair estimates, given by Eq.4.35, contributing to the energy of the reference cell need to be evaluated. The computational cost of this step is $\mathcal{O}(N)$ if all pairs $i_{\mathbf{0}}j_{\mathbf{l}}$ are considered. In the limit of large supercell sizes, simple prescreening approaches such as distance based cutoffs or dipole approximations can be used to truncate the pair list to obtain $\mathcal{O}(1)$ scaling.

Only PNOs of pairs with at least one index in the reference cell, $i_{\mathbf{0}}j_{\mathbf{l}}$, are evaluated and stored. Computing the Coulomb and exchange integrals in the PNO basis, for each these pairs, is also one of the most costly steps within the scheme. Translational invariance is once again exploited to compute the unique integrals belonging to the

reference cell, but now the PAO and DF domains span the pair union space. This step scales linearly with the size of the supercell up to the point where the number of significant pairs $i_{\mathbf{0}}j_{\mathbf{l}}$ within the supercell plateaus. The domain of PAOs for a pair $i_{\mathbf{0}}j_{\mathbf{l}}$ touches the boundary of the megacell and the AOs required to form a PAO at the megacell boundary extend beyond the megacell. This is illustrated in Figure 5.3 using a two-dimensional example. However, the contribution to $(i_{\mathbf{0}}\tilde{\mu}_{\mathbf{m}}|Q_{\mathbf{l}})$ from an AO outisde of the megacell $(i_{\mathbf{0}}\mu_{\mathbf{n}'}|Q_{\mathbf{l}})$ is completely negligible since the overlap of those AOs with $i_{\mathbf{0}}$ vanishes and the contribution would have been screened out. Since the corresponding integrals for $j_{\mathbf{l}}$ are translational copies from the reference cell, the robustness in the integrals is preserved for all occupied orbitals in the supercell.

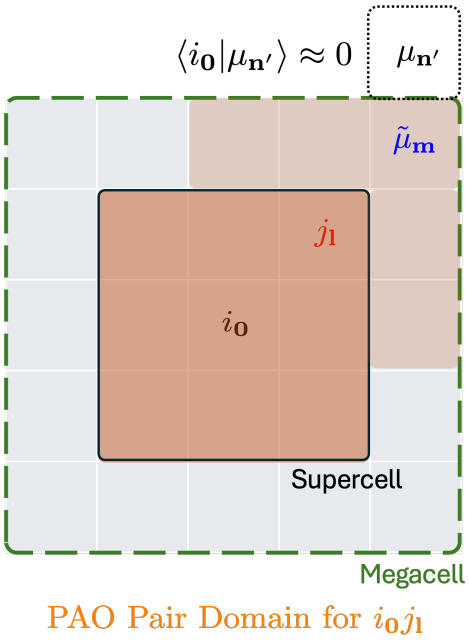

Figure 5.3: The PAO domain for pair $i_{\mathbf{0}}j_{\mathbf{l}}$ spans the entire megacell. The deeper shaded orange region shows the range of the PAOs that are fully supported within the megacell. PAOs within the lighter shaded orange regions, such as $\tilde{\mu}_{\mathbf{m}}$, are missing AOs such as $\mu_{\mathbf{n}'}$ to represent them, but the contributions of these AOs to the PAO DF integrals are negligible due to minimal overlap with $i_{\mathbf{0}}$.

The final DLPNO-MP2 amplitude equations (Eq. 4.41) are solved iteratively in the space of retained PNOs. The coupling introduces amplitudes and overlaps for pairs $\{k_{\mathbf{m}}j_{\mathbf{l}}\}$ where neither index is in the reference cell, but these are translational copies of a pair with one index in the reference cell, and only overlaps spanning

$\{ i_{\mathbf{0}} j_{\mathbf{l}} \}$ pair list are thus required. The residual equations become

$$\begin{aligned} 0 =& g^{i_{\mathbf{0}} j_{\mathbf{l}}}_{\tilde{a}\tilde{b}} + (\varepsilon_{\tilde{a}} + \varepsilon_{\tilde{b}}) t^{i_{\mathbf{0}} j_{\mathbf{l}}}_{\tilde{a}\tilde{b}} \\ & - \sum_{k_{\mathbf{m-1}}} \sum_{\tilde{c}\tilde{d} \in [k_{\mathbf{m-1}} j_{\mathbf{0}}]} S^{\tilde{c}, k_{\mathbf{m-1}} j_{\mathbf{0}}}_{\tilde{a}, i_{-1} j_{\mathbf{0}}} S^{\tilde{d}, k_{\mathbf{m-1}} j_{\mathbf{0}}}_{\tilde{b}, i_{-1} j_{\mathbf{0}}} t^{k_{\mathbf{m-1}} j_{\mathbf{0}}}_{\tilde{c}\tilde{d}} f^{i_{\mathbf{0}}}_{k_{\mathbf{m}}} \\ & - \sum_{k_{\mathbf{m}}} \sum_{\tilde{c}\tilde{d} \in [i_{\mathbf{0}} k_{\mathbf{m}}]} S^{\tilde{c}, i_{\mathbf{0}} k_{\mathbf{m}}}_{\tilde{a}, i_{\mathbf{0}} j_{\mathbf{l}}} S^{\tilde{d}, i_{\mathbf{0}} k_{\mathbf{m}}}_{\tilde{b}, i_{\mathbf{0}} j_{\mathbf{l}}} t^{i_{\mathbf{0}} k_{\mathbf{m}}}_{\tilde{c}\tilde{d}} f^{j_{\mathbf{0}}}_{k_{\mathbf{m-1}}}. \end{aligned} \tag{5.10}$$

The residual equations can be solved entirely self-consistently within the PNOs spanned by $\{ i_{\mathbf{0}} j_{\mathbf{l}} \}$.

5.4 Strategy 2: 'Born–von-Karman' treatment

As discussed earlier, correct evaluation of the ERIs, under Born–von-Karman periodic boundary conditions, involves accounting for the convergence behaviour of the charge and higher multipole interactions. The BvK-DLPNO-MP2 scheme directly addresses this, providing convergent lattice summations with no shape-dependant surface energy contribution in 3D. This is achieved through a chargeless local density fitting approach to correctly represent the MP2 densities, and a surface-dipole correction scheme to account for the long-range dipole-dipole interactions. Wannier functions and a Wannier gauge Fock matrix are obtained from a periodic HF calculation, spanning the supercell, and then fed into an adapted version of the DLPNO-MP2 code, which correctly addresses the periodic ERIs.

5.4.1 Chargeless Density Fitting

Since the MP2 densities $\rho = i_{\mathbf{l}} \tilde{a}$ are chargeless, the fitted densities, which are expanded with an auxiliary basis set, must also be charge-neutral, to avoid divergences under lattice summation. We adapt the procedure outlaid by Burow *et al.*[84,212]. The linear combination of DF functions forming the approximate density can be resolved into a charged ($\|$) and chargeless ($\perp$) component

$$\sum_{Q_{\mathbf{l}}} |\mathring{Q}_{\mathbf{l}}) C_{Q_{\mathbf{l}}} = \sum_{Q_{\mathbf{l}}} |\mathring{Q}_{\mathbf{l}}) C^{\|}_{Q_{\mathbf{l}}} + \sum_{Q_{\mathbf{l}}} |\mathring{Q}_{\mathbf{l}}) C^{\perp}_{Q_{\mathbf{l}}} \tag{5.11}$$

where

$$\mathbf{C}^{\perp} = \left(\mathbf{1} - \frac{\mathbf{n}\mathbf{n}^{\intercal}}{|\mathbf{n}|^2}\right)\mathbf{C} = \mathbf{P}^{\perp}\mathbf{C} \tag{5.12}$$

$$n_{Q_{\mathbf{l}}} = \int_{-\infty}^{\infty} \mathring{Q}_{\mathbf{l}}(\mathbf{r})\,\mathrm{d}^3\mathbf{r} \tag{5.13}$$

The vector $\mathbf{n}$ is the linear combination of $\mathring{Q}_{\mathbf{l}}$ that results in a charge and is specific to the local density fitting domain $\mathcal{D}$. To fit a chargeless density, the charged component of the fitting basis is projected out before solving for the fitting coefficients, which results in

$$(\rho^A|\rho^B) = \boldsymbol{\gamma}_A^{\perp\dagger}(\mathbf{V}^{\perp})^{-1}\boldsymbol{\gamma}_B^{\perp} \tag{5.14}$$

$$\mathbf{V}^{\perp} = \mathbf{P}_A^{\perp\dagger}\mathbf{V}\mathbf{P}_B^{\perp} \tag{5.15}$$

$$\boldsymbol{\gamma}_A^{\perp} = \boldsymbol{\gamma}_A\mathbf{P}_A^{\perp}, \tag{5.16}$$

where $\mathbf{V}$ and $\boldsymbol{\gamma}$ represent the two and three-index density fitting intermediates in Eq.5.3,

$$V_{P_{\mathbf{m}}Q_{\mathbf{l}}} = (P_{\mathbf{m}}|Q_{\mathbf{l}}) = \sum_{\mathbf{L}}(\mathring{P}_{\mathbf{m}}|\mathring{Q}_{\mathbf{l}+\mathbf{L}}) \tag{5.17}$$

$$\gamma_{Q_{\mathbf{l}}}^{A} = (\rho^A|Q_{\mathbf{l}}) = \sum_{\mathbf{L}}(\mathring{\rho}^A|\mathring{Q}_{\mathbf{l}+\mathbf{L}}). \tag{5.18}$$

5.4.2 Surface-dipole correction

The dipole of a charge density can be represented as a set of six point charges placed at the centers of the faces of the BvK supercell. Periodically repeating these dipoles within a finite volume $\mathcal{M} = \mathcal{M}_a\mathcal{M}_b\mathcal{M}_c$ results in a cancellation of all point charges within the shape, but leaves uncancelled positive and negative charges on opposite faces. The electrostatic interaction of these charges with the dipole of the charge density in the reference supercell gives rise to a surface energy that vanishes in 1D and 2D, but converges to a shape-dependent value as $\mathcal{M}$ increases in 3D. The correct integrals for BvK boundary conditions are obtained from lattice summation protocols that impose 'tin foil'[213] conditions at the surface, where there is no artificial surface contribution to the energy.

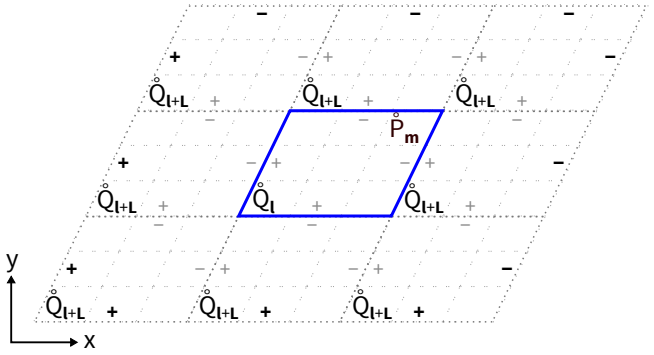

Figure 5.4: Direct lattice summation for $(P_\mathbf{m}|Q_\mathbf{l})$ with dipole correction.

To eliminate the surface contribution, we follow the method by Stolarczyk and Piela,[203] which has been adapted to the periodic LCAO Hartree–Fock method in `riper`[84], described in detail in Ref.212. To each auxiliary basis function in the lattice summation, we add a set of six point charges located at the faces of the BvK supercell within which the function resides, that exactly cancel the dipole of the function.

$$(P_\mathbf{m}|Q_\mathbf{l}) = \sum_\mathbf{L} (\mathring{P}_\mathbf{m}|\mathring{Q}_{\mathbf{l}+\mathbf{L}} + \mathring{d}^{Q_\mathbf{l}}_\mathbf{L}) \tag{5.19}$$

$$(\rho^A|Q_\mathbf{l}) = \sum_\mathbf{L} (\mathring{\rho}^A|\mathring{Q}_{\mathbf{l}+\mathbf{L}} + \mathring{d}^{Q_\mathbf{l}}_\mathbf{L}) \tag{5.20}$$

$$\mathring{d}^{Q_\mathbf{l}}_\mathbf{L}(\mathbf{r}) = \sum_\mathbf{I} q^{Q_\mathbf{l}}_\mathbf{I} \delta(\mathbf{r} - \tfrac{1}{2}\mathbf{R}_\mathbf{I} - \mathbf{R}_\mathbf{L}) \tag{5.21}$$

Here, $\mathbf{I}$ are the unit BvK supercell displacements,

$$\pm\mathbf{A} = (\pm 1, 0, 0),\ \pm\mathbf{B} = (0, \pm 1, 0),\ \pm\mathbf{C} = (0, 0, \pm 1). \tag{5.22}$$

In one- and two-dimensional calculations, $\mathbf{I}$ only includes the periodic lattice vectors, and two and four point charges are used, respectively, to cancel the dipole in the periodic degrees of freedom. A 2D case is illustrated in Fig. 5.4.

The Cartesian dipole components of a function $\mathring{Q}_\mathbf{l}$ are given by

$$p^{Q_\mathbf{l}}_\alpha = -\int_{-\infty}^{\infty} \alpha\, \mathring{Q}_\mathbf{l}(\mathbf{r})\, \mathrm{d}^3\mathbf{r} \qquad \alpha = x, y, z \tag{5.23}$$

where the minus is due to the electronic charge. The values of the point charges $q^{Q_\mathbf{l}}_\mathbf{I}$ that cancel the dipole are obtained by transforming $\mathbf{p}^{Q_\mathbf{l}}$ to the BvK lattice to obtain $p^{Q_\mathbf{l}}_\mathbf{a}, p^{Q_\mathbf{l}}_\mathbf{b}, p^{Q_\mathbf{l}}_\mathbf{c}$. The point charges that cancel the dipole in unit cell lattice vector

$\mathbf{a}$, placed at supercell displacements of $(\pm 0.5, 0, 0)$, are $\mp p_a^{Q_\mathbf{l}}/|\mathbf{A}|$, and similarly for the other lattice vectors.

In our pilot implementation, we evaluate the integrals through direct space lattice summation without applying any fast multipole acceleration techniques.[85,202,209] The lattice summation is truncated at a sufficiently large volume $\mathcal{M}$, which we monitor through the convergence of the integrals to a given precision. The surface correction term only applies to s-type and p-type DF functions, as all other functions have zero dipole. The net dipole arising from the s-type functions in the density fitting integrals is well defined because the net charge is projected out. For 1D and 2D calculations, the surface correction is only applied in the periodic directions.

Obtaining computational savings through translational symmetry is significantly complicated by the charge projection and surface correction terms. While the corrected integrals are translationally symmetric after charge projection, the individual contributions are not. In particular, in order to ensure that the surface term is properly eliminated, the lattice summation range and correction terms for s-type DF functions must be evaluated with reference to a common origin. The full set of AO integrals and correction terms in the supercell are thus computed, using the centre of the supercell as the origin. Integral evaluation through direct lattice summation is the main CPU bottleneck. A further subtlety is that the correction term breaks the symmetry of the Coulomb metric $V_{P_\mathbf{m}Q_\mathbf{l}}$. Symmetry is restored in the limit of large $\mathcal{M}$. We therefore symmetrise $V_{P_\mathbf{m}Q_\mathbf{l}}$, which we find accelerates convergence with lattice extent.

Work is currently underway to implement all compatible translational symmetry savings within the Megacell-DLPNO-MP2 scheme (reported fully in Appendix B) to the BvK-DLPNO-MP2 approach. All results and validation in Chapter 6 used a version of the BvK-DLPNO-MP2 approach which leveraged translational invariance to evaluate the PNOs and solve the non-redundant set of residual equations, as reported earlier in Section 5.3.1.

5.5 Conclusions

In this chapter, two implementations, Megacell-DLPNO-MP2 and BvK-DLPNO-MP2, have been introduced and outlined. The origin of the differing approaches stems from the difficulties in obtaining absolute convergence for periodic integrals with the Coulomb metric. Both approaches utilise periodicity to evaluate unit cell correlation energies. In the BvK approach, the surface-corrected, charge-projected, density fitting scheme produces absolutely convergent MP2 integrals, which inherently accounts for periodic boundary conditions. In contrast, the Megacell scheme requires translational symmetry to be explicitly enforced, since the DLPNO-MP2 equations are approximated to be at the thermodynamic limit, and so lattice summation is removed from the integral evaluation.

In Chapter 1, we highlight two main strategies for electronic structure calculations of periodic systems: embedded cluster or periodic boundary condition approaches. BvK-DLPNO-MP2 can be straightforwardly understood as a complete pbc approach, since lattice summation is included in the evaluation of all functions. Megacell-DLPNO-MP2, whilst still working within a periodic formalism, is perhaps slightly more difficult to define. The use of an encompassing megacell shares some conceptual similarities with embedding schemes, although it is motivated by the need to support the periodicity of objects centred in the supercell, and not by the need to use a higher accuracy theory. In addition, one could draw parallels between the Megacell approach and the PBC-CIM-DLPNO-CCSD(T) scheme from Wang *et al.*[156], if a single individual cluster in the latter scheme could be expanded to span the entire supercell, and the local fragment were solved in the domain space spanning the megacell. We would expect this particular strategy, however, to be prohibitively expensive, without exploitation of translational symmetry.

6| Periodic DLPNO-MP2 Results

The work in this chapter contributed to the publications Ref.2 and Ref.3. Arman Nejad performed the calculations using BvK-DLPNO-MP2, as well as the molecular fragment calculations.

In this chapter, the results of the two periodic DLPNO-MP2 implementations are presented and analysed, for a range of one-, two- and three-dimensional insulating systems. We first present an approach to obtain unit cell energies from molecular, canonical MP2 calculations, to form a benchmark for both periodic DLPNO schemes, in Section 6.2. The accuracy of BvK- and Megacell-DLPNO-MP2 are then analysed in Section 6.3.1, considering the PNO and k-mesh convergence. After discussing the relative computational costs of both approaches, the remainder of the chapter focuses on the performance of Megacell-DLPNO-MP2, where plots of computational scaling are shown, and an overview of application systems are presented, demonstrating the current capabilities of the approach.

6.1 Computational details

Periodic LCAO-based Hartree–Fock calculations using k-point sampling are utilised in the `riper` module[83–86,97,98], the output of which provides the HF Bloch functions and band energies required for MP2. All HF calculations utilized the universal Coulomb-fitting auxiliary basis sets[138] for the RI-J approximation. In this work we adopt the Monkhorst-Pack grid for our k-point grid[125]. The Wannier function localization procedure, as described in Chapter 3, has also been implemented in a developmental version of the `riper` module[1]. For all the systems explored within this thesis, the periodic HF scheme was highly efficient and never presented itself as a computational bottleneck, when compared to the subsequent DLPNO-MP2 calculations.

Born–von-Karman and Megacell domain-based local PNO-MP2 (BvK-DLPNO-MP2 and Megacell-DLPNO-MP2) have been implemented in a developmental

version of the TURBOMOLE[81] program, within the `pnoccsd` module[87,90]. All periodic systems explored in this chapter employed the pob-TZVP[93] orbital basis sets, which have specifically been optimized for LCAO simulations in the solid state.[93,94]. The density fitting approximation used in the MP2 calculations employed the def2-TZVP auxiliary basis sets[214]. Unless otherwise indicated, the frozen-core approximation is applied. For all systems, primitive unit cells were chosen. DLPNO-MP2 calculations are performed for a series of increasingly accurate PNO thresholds ($\mathcal{T}_{\mathrm{PNO}} = 10^{-X}$, $X = 6, 7, 8, 9$). Motivated by the observation that the largest discarded amplitude is proportional to the square root of the PNO threshold, the complete PNO space (CPS) limit is estimated through a square root extrapolation using the two most accurate PNO energies.[191,192] All calculations were run on a single node (Intel(R) Xeon(R) Gold 6248R CPU) with a maximum RAM limit of 386 GB and 1.8 Tb disk, with OMP parallelisation. In all cases, we utilise primitive unit cells. Lattice constants, unit cell geometries, and computed energies for all calculations are collected in Appendix C.

6.2 Unit cell energies through molecular fragments

To obtain reliable reference values for unit cell energies against which to test the performance of the two periodic DLPNO-MP2 implementations, we utilise the molecular RI-MP2[215] infrastructure in TURBOMOLE[81,82]. Through this molecular fragment scheme, canonical unit cell energies for an effective supercell size can be obtained, which we subsequently converge to obtain a thermodynamic limit estimate. Such comparison is useful to verify that the periodic DLPNO-MP2 methods are internally consistent within TURBOMOLE, through benchmarking against the robustly tested molecular code. This also enables us to assess the numerical stability of the periodic DLPNO-MP2 schemes, to examine the rate of convergence to the thermodynamic limit, and to analyse the accuracy of the PNO approximation, which should converge towards the canonical RI-MP2 benchmark with increasingly tight thresholds. In Ref.2, we also apply the same fragment approach using the molecular

DLPNO-MP2 program, allowing for a comparison of the PNO approximations between molecular and periodic schemes, but we omit discussion of this within this chapter for the sake of conciseness and clarity.

In the molecular fragment approach, the energy of a unit cell is obtained by modelling the energy of a molecular fragment as a core of unit cells, plus contributions from faces, edges and corners. Starting from the periodic supercell structures, the fragments are generated by saturating the dangling bonds with hydrogens, as illustrated in Fig. 6.1. The energies of a series of increasing molecular fragments can then be used to subtract the surface contributions to obtain the correlation energy per unit cell, which converges to the thermodynamic limit as the size of the molecular fragments increases.

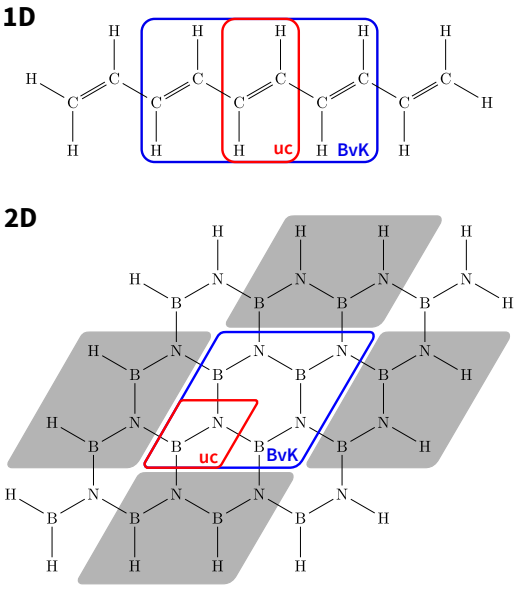

Figure 6.1: Molecular cluster fragments that accommodate 1D and 2D BvK supercells (dotted box). The supercells of clusters with sum formula $C_{2l}H_{2l+2}$ and $B_{l^2}N_{l^2}H_{4l}$ are comprised of $(l-2)$ and $(l-2)^2$ unit cells (red box), respectively.

The molecular fragment method was used to obtain benchmark unit cell energies of 1D and 2D periodic systems. 3D examples were not possible due to the high computational expense of this approach. For the 1D systems, the molecular energies of $C_{2l}H_{2l+2}$ and $C_{2l}H_{l+2}F_l$ are modelled as

$$E_{l,\mathrm{tot}} = 2E_{\mathrm{corner}} + (l-2)E_{\mathrm{uc}}. \tag{6.1}$$

Using two fragment calculations, the corner energy can be subtracted to obtain an estimate of the unit cell energy

$$E_{\mathrm{uc}}^{\mathrm{1D}} = \frac{E_{l,\mathrm{tot}} - E_{l-2,\mathrm{tot}}}{2}. \tag{6.2}$$

The largest supercell size accommodated in the largest fragment corresponds to $k_{\mathit{eff}} = (l-2)$. For the 2D system, the total energy may be decomposed as

$$E_{l,\mathrm{tot}} = 4E_{\mathrm{corner}} + 4(l-2)E_{\mathrm{edge}} + (l-2)^2 E_{\mathrm{uc}}. \tag{6.3}$$

To extract the unit cell energy for an effective mesh $k_{\mathit{eff}}^2 = (l-2)^2$, we now require three molecular clusters

$$E_{\mathrm{uc}}^{\mathrm{2D}} = \frac{E_{l,\mathrm{tot}} - E_{l-2,\mathrm{tot}} + E_{l-4,\mathrm{tot}}}{8}. \tag{6.4}$$

In the following results, molecular fragment calculations were performed with increasing k_{eff} until converged to a precision of 1×10^{-5} Ha, to provide a robust estimate of the thermodynamic limit unit cell energy.

6.3 Comparison between 'Megacell' and 'BvK' schemes

In this section, we perform a detailed comparison between Megacell-DLPNO-MP2 and BvK-DLPNO-MP2, first investigating the accuracy of both schemes, across a range of non-conducting materials, using the molecular fragment benchmark described above. Further analysis on the convergence behaviours to the thermodynamic and canonical limits are also given. We then compare the current computational cost of both implementations and comment on the practical merits of both schemes.

6.3.1 PNO and k-mesh convergence

Firstly, three relatively simple systems, one-dimensional linear alkene polymers C_2H_2 and C_2HF, as well as two-dimensional hexagonal boron nitride (BN) sheets, were chosen to validate the performance of BvK- and Megacell-DLPNO-MP2. C_2HF was selected in order to include long-range dipole-dipole interactions from the polar

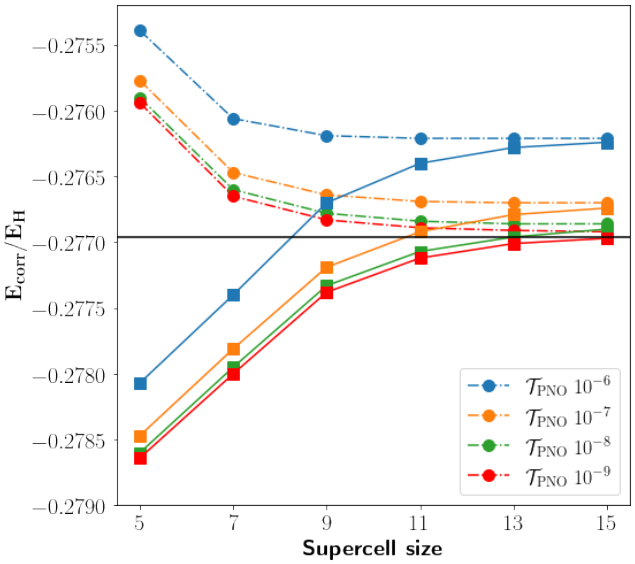

Figure 6.2: Megacell (circles) and BvK (squares) DLPNO-MP2 correlation energies of C_2H_2 as a function of 1D k-mesh and $\mathcal{T}_{PNO}$. The horizontal line indicates the canonical thermodynamic limit[2].

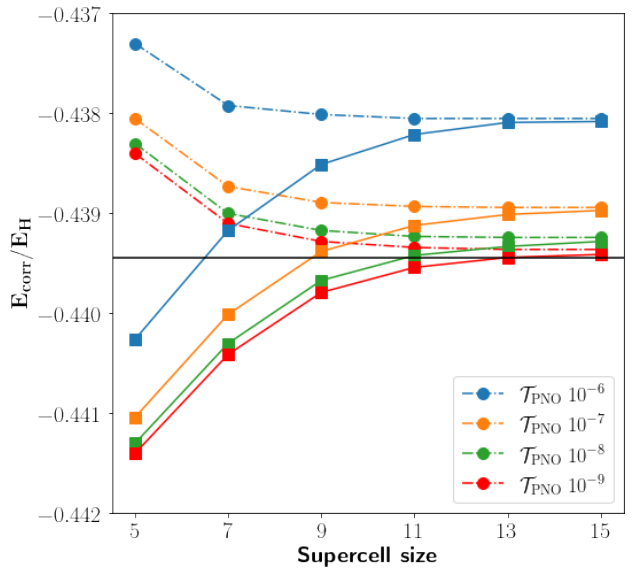

Figure 6.3: Megacell (circles) and BvK (squares) DLPNO-MP2 correlation energies of C_2HF as a function of 1D k-mesh and $\mathcal{T}_{PNO}$. The horizontal line indicates the canonical thermodynamic limit[2].

C−F bonds, to investigate any difference in convergence behaviour compared to C_2H_2. Hexagonal BN is isoelectronic and structurally analogous to graphene, but avoids Fermi-level degeneracies, behaving as an insulator with a much higher band gap ($\sim 6\,eV$).[216], ensuring stable periodic MP2 energies. We present MP2 correlation

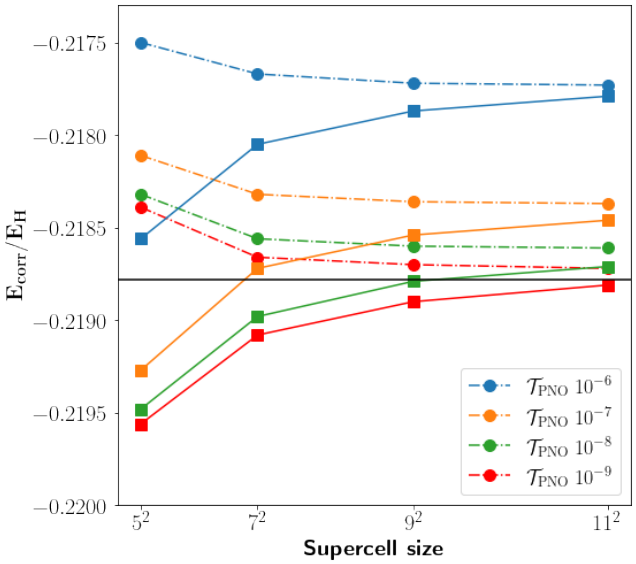

Figure 6.4: Megacell (circles) and BvK (squares) DLPNO-MP2 correlation energies of BN as a function of 2D k-mesh and $\mathcal{T}_{\mathrm{PNO}}$. The horizontal line indicates the canonical thermodynamic limit[2].

energies, obtained through BvK- and Megacell-DLPNO-MP2, for C_2H_2, C_2HF and BN, in Figures 6.2, 6.3 and 6.4, respectively, at varying $\mathcal{T}_{\mathrm{PNO}}$ thresholds and with increasing supercell size. In each figure, the black line is the thermodynamic and canonical limit, obtained via the RI-MP2 molecular fragment scheme.

In all three examples, the BvK and Megacell schemes converge to the same energy at the thermodynamic limit for each PNO threshold. For the one-dimensional systems, agreement between BvK and Megacell at the largest supercell size is on the order of 1×10^{-5} Ha, and on the order of 1×10^{-4} Ha for boron nitride. This remarkable agreement between the three different schemes at the canonical and thermodynamic limit underlines the consistency in the PNO local correlation approximation and the stability of the PAO-OSV-PNO cascade. It provides strong validation for both periodic schemes, demonstrating internal consistency with the existing molecular TURBOMOLE routines. Comparing the systems, we note no appreciable difference in the general rate of convergence with increasing supercell size between C_2H_2 and C_2HF, suggesting that the performance of neither method is impacted by the presence of a net dipole in the unit cell. We expect 1×10^{-5} Ha

agreement for 2D BN to also be possible, but our current computational limitations prevent 13×13 or larger supercell simulations.

Comparing the convergence of the BvK and Megacell approaches to the thermodynamic limit, we find that BvK correlation energies converge from below, whilst the Megacell values converge from above. In all three examples, the rate of convergence for the Megacell approach is faster than that of BvK. We attribute the main source of this difference to be due the different HF orbitals and band energies used for BvK and Megacell, for a MP2 calculation of the same supercell size. As discussed in Section 5.3, the Megacell schemes utilises WFs obtained from HF calculation employing a denser k-mesh, $k_{\text{mega}} = 2k_{\text{super}} - 1$, which are therefore closer to the thermodynamic limit than those of the BvK calculations. The spatial locality of the WFs as well as the orbital eigenvalues influences the subsequent MP2 calculation, which can impact convergence of the MP2 correlation energy.

In order to further comment on the impact of the HF convergence error for BvK and Megacell, Figure 6.5 plots the HF error with increasing k-mesh size, compared to the thermodynamic limit. On the left plot, the relative BvK-DLPNO-MP2 energy errors, at two $\mathcal{T}_{\text{PNO}}$ thresholds, using the HF orbitals at that given k-mesh, are presented. The right graph plots the Megacell-DLPNO-MP2 energy errors. In the case of BvK, the HF error is of greater or similar magnitude to the MP2 errors, indicating that part of the slower convergence can be attributed to significantly changing WFs and band energies with growing supercell size. Whilst for Megacell, the HF error is significantly smaller than the MP2 error, suggesting that the Megacell scheme largely utilises fully converged WFs and band energies, especially from $\mathbf{k} = 13$ onwards. Finite-size effects in virtual HF orbitals and eigenvalues are not reflected in the HF energy, but also have a direct impact on the MP2 correlation energy, and we expect these to follow a similar convergence pattern as Figure 6.5.

As discussed in Section 5.2, the formal conditional convergence of dipole-dipole interactions in three-dimensional systems introduces an additional complexity to obtaining the correct unit cell energies. In addition, benchmark unit cell energies from the molecular fragment approach were unavailable, due to the computational

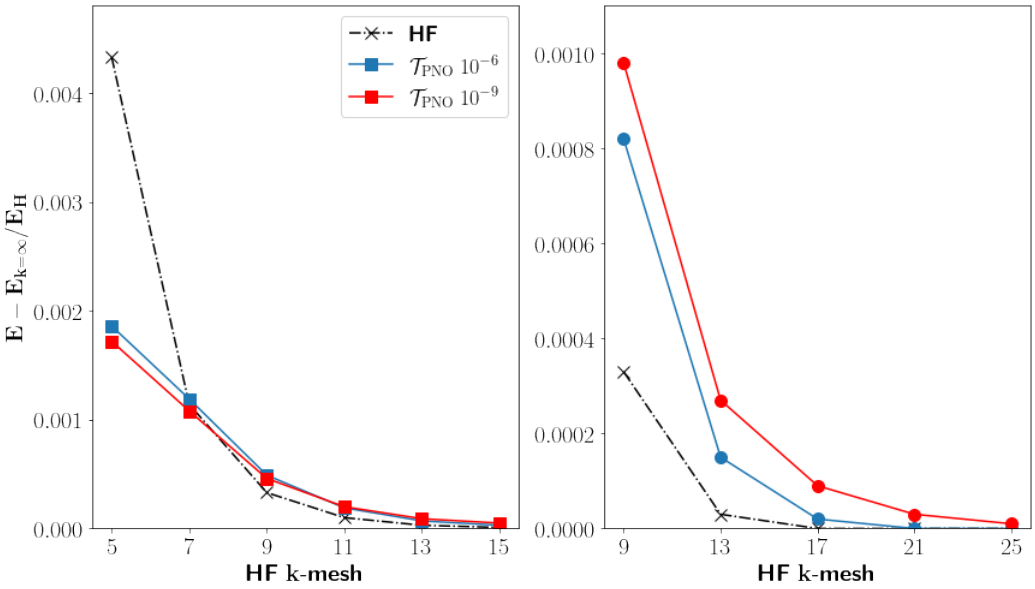

Figure 6.5: Hartree–Fock energy errors compared to the thermodynamic limit as a function of **k**-mesh size, compared with the respective MP2 correlation energy errors for BvK-DLPNO-MP2 (left) and Megacell-DLPNO-MP2 (right).

cost of simulating large 3D clusters using the molecular RI-MP2 code. First, for BvK-DLPNO-MP2, we verify that the dipole correction scheme correctly removes the surface energy contributions in 3D systems, by computing the correlation energy of the same lattice through two different unit cells. At the thermodynamic limit, these calculations should yield the same HF and correlation energies. The results for cubic and rhombohedral LiH unit cells are reported in Table 6.1. The values are converged with respect to lattice image summation to the degree of precision that they are presented. Since the cubic cell contains four LiH units and the rhombohedral only one, the former energies are divided by four. The HF energies per LiH unit for the 3×3×3 cubic and 5×5×5 rhombohedral calculations agree to within 0.25 mHa, reducing to 10^{-5} Ha for 5×5×5 and 7×7×7. The DLPNO-MP2 correlation energies per LiH unit for the 3×3×3 cubic and 5×5×5 rhombohedral calculations agree to within 0.15 mHa, if the dipole correction is applied. If it is not, the energies deviate by 621 mHa. This data indicates that the dipole-corrected density fitting implementation is correct.

We now directly compare the performance of the two schemes for 3D systems. Table 6.2 compares the Megacell- and BvK-DLPNO-MP2 correlation energies for

Table 6.1: HF and BvK-DLPNO-MP2 ($\mathcal{T}_{\mathrm{PNO}} = 10^{-7}$) energies per LiH unit (Hartree) for a 3D LiH lattice with two different unit cells.

	Cubic		Rhombohedral	
k-mesh	3×3×3	5×5×5	5×5×5	7×7×7
n_{atom}	216	1000	250	686
E_{HF}	-8.06035	-8.06058	-8.06060	-8.06059
E_{corr}				
... dip-cor	-0.03060		-0.03075	-0.03092
... no dip-cor	-0.077		-0.698	

Table 6.2: MP2 correlation energy comparison between Megacell- and BvK-DLPNO-MP2 implementations, varying PNO truncation threshold and supercell size.

System	$\mathcal{T}_{\mathrm{PNO}}$	mega	BvK	mega	BvK
k_{super}		5^3		7^3	
LiH	6	-0.03063	-0.03066	-0.03083	-0.03083
	7	-0.03070	-0.03075	-0.03091	-0.03093
	8	-0.03073	-0.03078	-0.03096	-0.03096
k_{super}		3^3		5^3	
MgO	6	-0.19747	-0.19631	-0.19868	-0.19853
	7	-0.19765	-0.19651	-0.19897	-0.19882
	8	-0.19774	-0.19660	-0.19908	-0.19902

Lithium Hydride (LiH) and Magnesium Oxide (MgO) using $\mathcal{T}_{\mathrm{PNO}} = 10^{-6}, 10^{-7}, 10^{-8}$ at a series of k-mesh supercell sizes. For LiH, agreement to better than 0.1 millihartree is already achieved for a 5^3 supercell, at each $\mathcal{T}_{\mathrm{PNO}}$ value. For MgO, the rapid convergence of the two schemes supercells is also evident, with differences between Megacell and BvK decreasing by an order of magnitude from 3^3 to 5^3. In contrast to the earlier one- and two-dimensional examples, both the Megacell and BvK schemes appear to converge to the thermodynamic limit from above. Further comparison of the asymptotic behaviour of both approaches will be the subject of future work. Overall, this data pleasingly validates both schemes for probing 3D systems.

Finally, to assess the degree to which Megacell-DLPNO-MP2 is able to accurately estimate the canonical and thermodynamic limit, in Figure 6.6 we plot the MP2 correlation energies of MgO as a function of inverse supercell size (volume), at varying PNO truncation thresholds. For each supercell size, the complete PNO space (CPS) limit is obtained through a square root extrapolation[191,192] of the PNO truncation

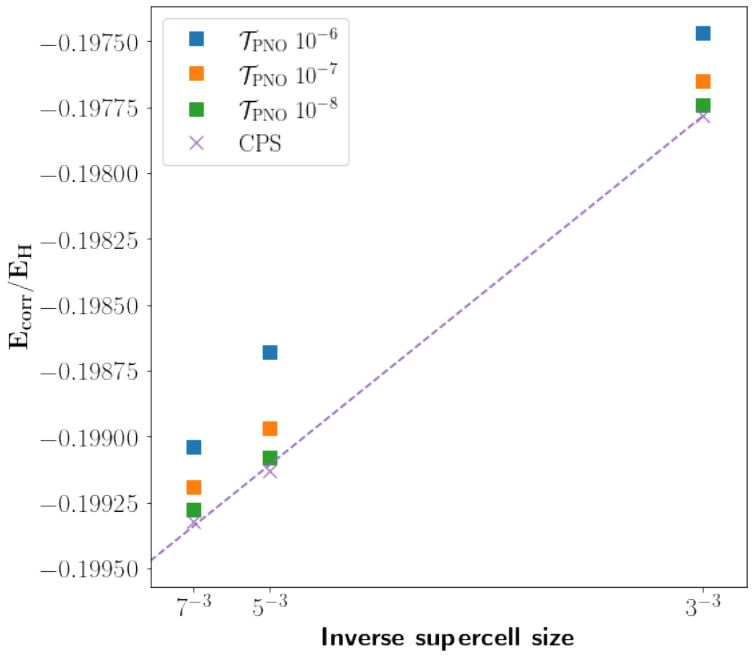

Figure 6.6: MP2 correlation energies for MgO, obtained with increasing PNO truncation thresholds and larger supercells. Extrapolations to the complete PNO space limit for each supercell size are given, and an inverse volume extrapolation to the thermodynamic limit is plotted.

threshold, motivated by the observation that the largest discarded amplitude is proportional to the square of the threshold. An inverse volume extrapolation[62,217] is then performed using the CPS values to obtain a canonical and thermodynamic limit value for the MP2 correlation energy per cell. To a very good approximation, the Megacell-DLPNO-MP2 values lie on the expected line following the inverse volume convergence. The consistent behaviour of the Megacell-DLPNO-MP2 scheme thus makes it possible to extrapolate to both the CPS and thermodynamic limits.

6.3.2 Computational cost comparison

Having established that both BvK- and Megacell-DLPNO-MP2 converge to the correct limit value with k-mesh and PNO truncation threshold, a key point of discussion is the relative computational cost between both approaches. Overall, comparing the current states of implementation, the Megacell scheme is significantly more computationally efficient than BvK, largely due to the differing CPU costs of computing the ERIs.

The computational bottleneck of computing the ERIs within the BvK scheme originates from the large volume extent $\mathcal{M}$ required to obtain sufficiently converged integrals in Eqs.5.19 and 5.20. As described in Section 5.4, implementing translational symmetry to compute only translationally unique integrals is currently incompatible with the charge projection and surface correction, which require a consistent common origin for the entire supercell. Hence, the scaling of this step is asymptotically $\mathcal{O}(k_{\text{super}})$ when the supercell is large enough for screening to be effective, but with a large prefactor from the naive lattice summation of multiple supercell images up to a volume $\mathcal{M}$. In contrast, no lattice summation is required in the ERI evaluation for the Megacell scheme, and translational symmetry is exploited to compute only the unique set of integrals, which scale asymptotically as $\mathcal{O}(1)$. As a result, the present implementation of Megacell-DLPNO-MP2 has significantly further scope to simulate larger systems, particularly in 3D, and tentative estimates of the wall times required by the BvK scheme for the same systems were deemed to be impractically long for Ref.2 and this thesis.

Formal comparison of RAM and disk memory requirements of both approaches is difficult, due to ongoing code changes in both pilot schemes. For Megacell-DLPNO-MP2, exploitation of translational symmetry removes storage costs scaling with the size of the megacell, such as in the case of the density fitting ERIs, which would otherwise rapidly render the approach to be impractical. However, full implementation of translational symmetry to every possible routine in the `pnoccsd` module is difficult, and so work is still ongoing to address memory bottlenecks which subsequently appear. At present, the current memory bottleneck is the storage of the full Fock matrix in the PAO basis, which scales as $\mathcal{O}(k^2_{\text{mega}})$, but work is ongoing to store only the translationally unique Fock matrix elements, which scale as $\mathcal{O}(k_{\text{mega}})$. Since the BvK approach employs just a supercell, formal memory scaling costs are inherently less severe, and it is currently the CPU expense that prevents BvK-DLPNO-MP2 from simulating larger systems.

Having concluded that Megacell-DLPNO-MP2 is currently computationally more efficient than BvK-DLPNO-MP2, we choose to focus on this scheme for

the remainder of the chapter, in order to elaborate on its scaling performance and applicability in further detail.

6.4 Scaling of Megacell-DLPNO-MP2

For molecules, the combined effect of pair screening, integral screening, local density fitting and PNO compression leads to an asymptotically linear-scaling DLPNO-MP2 method with respect to system size.[82,87,169] In essence the complexity for correlating each orbital i is asymptotically $\mathcal{O}(1)$ since correlation is short-ranged and linear scaling results from there being $\mathcal{O}(N)$ orbitals in a molecule. For periodic systems, when increasing the number of unit cells within the supercell, this equates to an asymptotic scaling that is linear in the number of orbitals in a unit cell, which is $\mathcal{O}(1)$. This formal scaling of computational effort is only observed in practice if appropriate local approximations are applied to every step in the program workflow, and the supercell size is sufficiently large for the locality savings to take effect. In the following, we report the observed scaling of our current implementation of Megacell-DLPNO-MP2 for 2D Boron Nitride and 3D Lithium Hydride systems in realistic calculations. All CPU timings were evaluated on a single OMP thread.

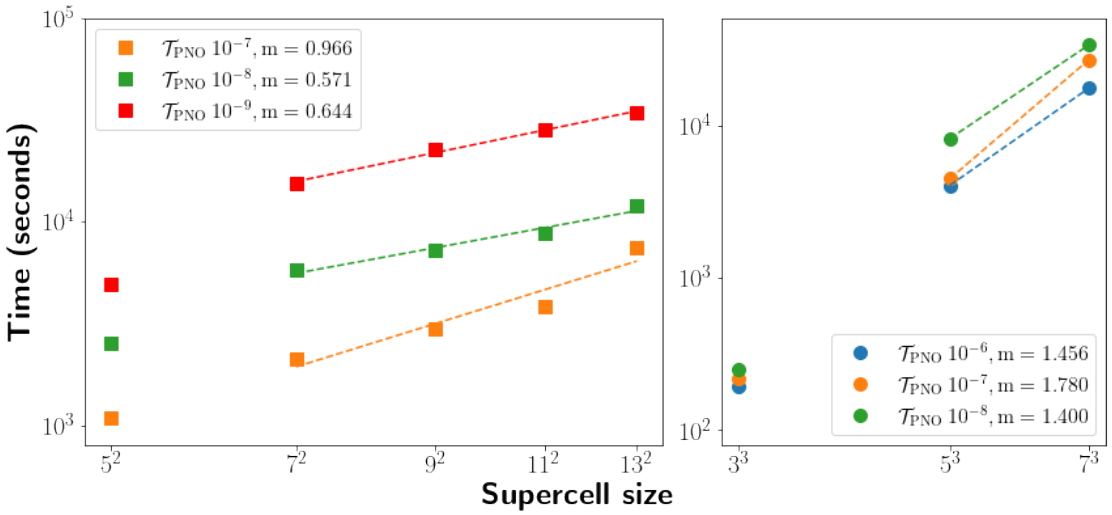

Figure 6.7: Log scaled CPU times for two-dimensional Boron Nitride (left), and three-dimensional Lithium Hydride (right), as a function of log scaled supercell size and PNO truncation threshold. Slope values for the trend lines of the largest supercell sizes as given as m = [slopevalue]

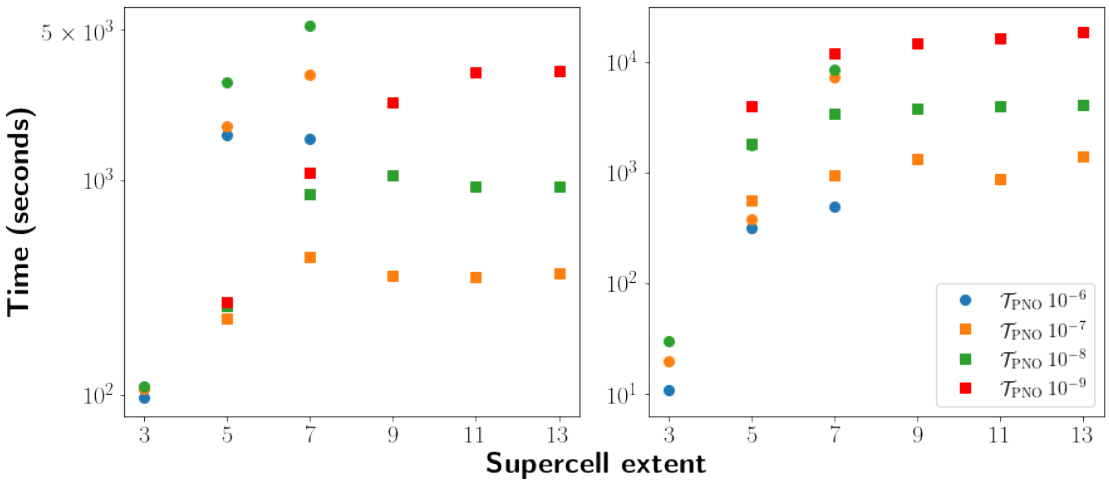

Figure 6.8: Log scaled CPU times for two key subroutines in the Megacell-DLPNO-MP2 method, as a function of the supercell extent in each dimension, and PNO truncation threshold. The left panel shows the generation of the three index density-fitting integrals, whilst the right panel presents the construction of the approximate external pair density, and the subsequent diagonalization to construct the PNOs. Circle markers denote timings from three-dimensional Lithium Hydride. Square markers denote timings from two-dimensional Boron Nitride.

In Figure 6.7, the total CPU times for Boron Nitride (left) and Lithium Hydride (right) are plotted, at increasing supercell sizes. Lines of best fit of the log-scaled graphs, discarding the smallest supercell size, are also presented, and the gradients are shown. For BN, the fitted lines demonstrate overall linear and sub-linear CPU scaling in this size regime. This is a direct consequence of the combination of the local domains, pair screening and translational invariance savings. Whilst the linear and sub-linear scaling from our BN results are encouraging, the theoretical $\mathcal{O}(1)$ scaling is not attained, even though the interaction pair list $i_{\mathbf{0}}j_{\mathbf{l}}$ has saturated. We attribute this to several subroutines that have not yet been fully adapted from the molecular scheme to leverage translational invariance. Particularly for $\mathcal{T}_{\mathrm{PNO}} = 10^{-7}$, these routines, which were not considered to have expensive scaling or prefactors earlier in the code development, now have a measurable impact of the overall cost.

The CPU scaling for LiH currently shows only sub-quadratic cost at the largest two supercell sizes, for all $\mathcal{T}_{\mathrm{PNO}}$ values. This is partly due to the aforementioned unadapted routines, which have a greater impact for larger three-dimensional systems. More significantly, however, the largest supercells employed for LiH

have not yet entered the regime in which the interaction sphere of $i_{\mathbf{0}}j_{\mathbf{l}}$ pairs becomes fully saturated. A transition to sub-linear scaling at larger supercell sizes is expected, but we have not been able to demonstrate this due to the limitations of our available hardware.

The two most expensive steps in the Megacell-DLPNO-MP2 method are the evaluation of the three-index integrals $(\mu_{\mathbf{m}} i_{\mathbf{0}}|Q_{\mathbf{l}})$ and the formation of the PNOs, which involves the construction of the ERIs in the pre-PNO basis $(\bar{a}i_{\mathbf{0}}|\bar{b}j_{\mathbf{l}})$. The individual CPU timings for these key steps in Megacell-DLPNO-MP2 are presented in Figure 6.8, for both BN and LiH. The log-scaled times are plotted against the supercell extent in each dimension, to enable comparison of the field of interaction between two-dimensional BN and three-dimensional LiH.

The left panel displays the CPU times for the evaluation of the three-index integrals. For the 2D BN example, the observed CPU cost tends to a constant, reflecting the correct $\mathcal{O}(1)$ scaling due to the saturation of the density fitting domains and efficacy of integral screening. The DF threshold is linked to the PNO threshold and saturation of the DF domain requires larger supercells for tighter PNO thresholds. The 3D LiH example appears to follow a similar trend, but the supercell diameter of 7 unit cells is not sufficiently large for the DF domain to saturate for the asymptotic scaling to manifest. The right panel displays the CPU times for the subroutine where the PNOs are formed. For BN, the observed timings again tend towards $\mathcal{O}(1)$ scaling after a diameter of 9 unit cells. Although the LiH data follows a similar trend for $\mathcal{T}_{\mathrm{PNO}} = 10^{-6}$ and 10^{-8}, the timings for 10^{-7} scales linearly in this regime. This is due to sub-optimal I/O batching that has not yet been adjusted from the molecular code to account for reduced pair list. Although the code can already be used to evaluate large supercells at tight PNO thresholds, improved performance can be found through further optimisation.

6.5 Applications with Megacell-DLPNO-MP2

To demonstrate the current scope and performance of the Megacell-DLPNO-MP2 approach we apply our method to simulate large supercells of three-dimensional

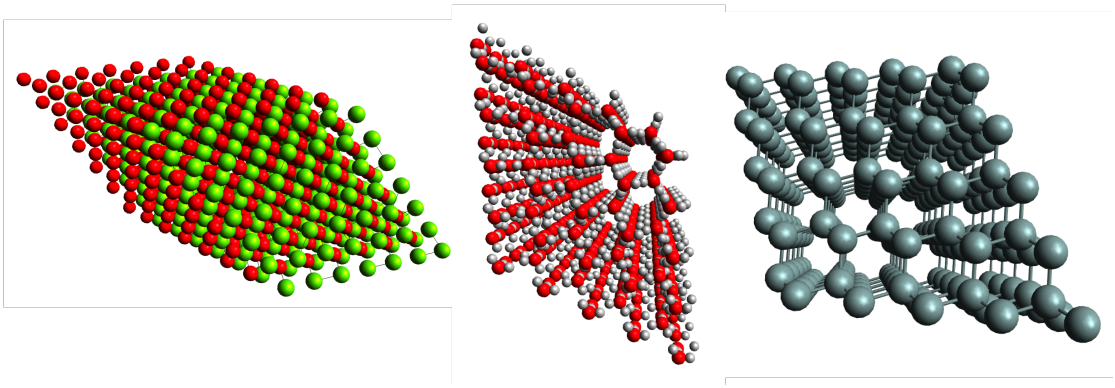

Figure 6.9: Supercells of the largest three-dimensional systems explored in this contribution. From left to right: 7x7x7 Magnesium Oxide (rock salt), 5x5x5 Hexagonal Ice, 5x5x5 Silicon (diamond cubic). The number of basis functions spanned within each supercell is 12691, 15000 and 5500, respectively.

materials. Two rocksalt structures (LiH, MgO), one diamond cubic structures (Si), as well a hexagonal polymorph of ice, are probed. Table 6.3 presents MP2 correlation energies for the valence bands for each of these systems, at varying PNO truncation thresholds. Canonical pob-TZVP energies are estimated via extrapolation to the complete PNO space, and inverse volume extrapolations are applied to obtain the canonical and thermodynamic limit estimates, $E_{\mathrm{corr,TDL}}$.

The supercells for our three largest calculations are displayed in Figure 6.9. The largest, a 5^3 supercell of hexagonal ice, contains 15000 basis functions. With the exception of Si, the MP2 correlation energies obtained from the largest supercell and tightest $\mathcal{T}_{\mathrm{PNO}}$ threshold are all within 0.2 millihartree of the estimated $E_{\mathrm{corr,TDL}}$ value, demonstrating the extent to which the Megacell-DLPNO-MP2 can sample the thermodynamic limit. At these system sizes, the preceding periodic HF and Wannier localisation calculations still remain much faster than Megacell-DLPNO-MP2, and have an insignificant impact on overall computational demands. Combining the HF and localisation steps, no systems possessed wall times exceeding two hours for these large three-dimensional systems, with most finishing in under half an hour. All MP2 calculations were completed within two days on our machines, using OMP parallelisation over 48 threads. We are currently prevented from performing larger calculations due to some memory and disk bottlenecks of our pilot implementation and the hardware we have access to. In particular, storage of

the OSV overlaps, which currently span the entire supercell, exceeds our current disk size. Nevertheless, the accuracy and efficiency of our approach is encouraging and we anticipate that Megacell-DLPNO-MP2 method can be readily applied to obtain ground state properties of insulating systems.

Table 6.3: Survey of three-dimensional materials probed using Megacell-DLPNO-MP2, given by PNO truncation threshold and largest supercell size employed. Hartree–Fock energies are provided for each supercell. Only valence bands were included in the correlated treatment. All energies given in Hartrees.

k_{super}		3^3		5^3		7^3		
System	$\mathcal{T}_{\text{PNO}}$	E_{corr}	E_{HF}	E_{corr}	E_{HF}	E_{corr}	E_{HF}	$E_{\text{corr,TDL}}$
LiH	6	-0.02934	-8.06060	-0.03063	-8.06059	-0.03083	-8.06059	
	7	-0.02938		-0.03070		-0.03091		
	8	-0.02940		-0.03073		-0.03096		
	CPS	-0.02941		-0.03074		-0.03098		-0.03112
MgO	6	-0.19747	-274.68452	-0.19868	-274.68455	-0.19904	-274.68455	
	7	-0.19765		-0.19897		-0.19919		
	8	-0.19774		-0.19908		-0.19928		
	CPS	-0.19778		-0.19913		-0.19932		-0.19947
Si	6	-0.16401	-577.84603	-0.17394	-577.84720			
	7	-0.16507		-0.17637				
	8	-0.16542		-0.17703				
	CPS	-0.16558		-0.17734				-0.18057
ice I_h	6	-0.74541	-304.26198	-0.74620	-304.26198			
	7	-0.74677		-0.74738				
	8	-0.74727		-0.74781				
	CPS	-0.74750		-0.74801				-0.74815

6.6 Conclusions

In this chapter, proof-of-concept results are presented for BvK- and Megacell-DLPNO-MP2. We compare correlation energies computed using the BvK and Megacell schemes for 1D, 2D and 3D materials, finding that the correlation energies all converge to the same thermodynamic limit for each PNO threshold, showing that the level of accuracy of the PNO truncation scheme is entirely consistent for all systems explored. Using canonical benchmark values obtained through a molecular fragment approach, we demonstrate the efficacy of the PNO approximation, and the smooth convergence towards the canonical limit, validating both periodic approaches. Whilst both schemes converge to the same limit values, we note that the Megacell scheme appears to have quicker convergence with respect to supercell size, and we

attribute a large source of this to the input HF band energies and WFs, which are closer to the thermodynamic limit, spanning the commensurate megacell size. Ongoing work in the Tew group is currently focused on analysing the source of the finite-size error of both approaches in greater detail, in order to discern if the convergence rate can be further accelerated.

The current computational bottlenecks for BvK-DLPNO-MP2, arising from the lattice summation in the ERIs, means that the pilot version of the Megacell scheme is computationally more practical. Using Megacell-DLPNO-MP2, scaling benchmarks with respect to supercell size confirm the theoretical $\mathcal{O}(1)$ scaling using 2D BN, whilst 3D LiH results indicate that the asymptotic regime is close to being achieved with our computational capabilities. Simulations of large 3D supercells, featuring up to 15000 basis functions, is readily achievable with the pilot scheme.

We believe continued optimisation of both schemes will allow more complex and larger supercells to be simulated in the future. In particular, implementation of fast multipole acceleration techniques would significantly reduce the prefactor for evaluating the periodic ERIs in the BvK scheme, leading it to be more competitive with the Megacell approach. The corresponding memory and storage scaling of both methods could then be analysed more systematically.

Overall, the results in this chapter demonstrate the resounding success of the initial pilot implementations of BvK- and Megacell-DLPNO-MP2, and demonstrate that ground state chemical properties such as lattice constants, cohesive and adsorption energies for non-conducting systems are already within reach using these schemes. In the next chapter, we present a case study demonstrating the efficacy of the periodic DLPNO-MP2 approach for one of these applications.

7| Surface Adsorption using DLPNO-MP2

Poramas Komonvasee performed the calculations presented in this chapter, and participated in the analysis of the results.

7.1 Introduction

The interaction of a molecule as it adsorbs onto a surface is a key chemical phenomena, in which accurate modelling of such as system provides benefits to a number of important applications. In particular, computational evaluation of the adsorption energy (E_{ads}) gives mechanistic insight to reactions for heterogenous catalysis, gas storage and surface lubrication, among others[218–225]. Due to its computational efficiency, studies of adsorption systems predominantly have employed DFT, but accurate modelling of the dispersion interactions involved is difficult unless corrections are incorporated into the density functional treatment[23,34–38]. Even with these dispersion-corrected functionals, there is a need for higher accuracy quantum chemistry methods, which inherently capture dispersive effects, to create a trusted benchmark.

The adsorption of a single carbon monoxide (CO) molecule onto a pristine magnesium oxide (MgO-001) surface has become the consensus toy model system to probe surface adsorption interactions, within the electronic structure community. Within the last ten years, there have been a number of studies, employing high accuracy correlated wavefunction methods, which have aimed to evaluate the adsorption energy of CO on MgO at the dilute coverage limit[18–21]. These schemes can be roughly divided into finite-cluster methods, or approaches employing periodic boundary conditions. Through carefully converging out errors associated with parameters affected by basis set and finite-size effects, excellent agreement now exists between a number of computational schemes, as well as with experiment.

There are, however, still unanswered questions for simulating adsorption processes. In reality, adsorption reactions rarely feature a single adsorbate molecule,

and consideration for the optimal ratio of surface site availability is a key concern for heterogenous catalysis. For example, experimental literature for CO adsorption onto an MgO surface reports varying reactivities associated with dilute to dense coverage regimes[226–230]. Whilst an agreed benchmark for the adsorption energy of a single CO molecule on a pristine MgO surface has now been established, simulations incorporating multiple CO adsorption sites, approaching denser monolayer coverages (ML), are still scarce. These systems are inherently more difficult to simulate, given the expanded configuration space to now optimise over, and the need to incorporate lateral interactions between the CO monomers. The computational expense required to simulate larger unit cells or fragments also poses a challenge for current correlated wavefunction methods. In this thesis, we have outlined the theory and implementation of periodic DLPNO-MP2, and verified that the pilot scheme of Megacell-DLPNO-MP2 is particularly computationally efficient, able to simulate large supercells of materials and showing asymptotic constant scaling with respect to supercell size. Our aim in this chapter is to use the Megacell scheme to reveal further insight into interactions of the MgO + CO adsorption system.

In this chapter, we first outline our computational methodology employing Megacell-DLPNO-MP2 to model surface adsorption systems. Then, Section 7.3, we verify the validity of Megacell-DLPNO-MP2 for surface interactions by evaluating the adsorption energy towards the infinitely dilute regime, which has now been benchmarked rigorously by several high accuracy computational and experimental schemes. Having established agreement with these previous studies, we then use Megacell-DLPNO-MP2 to probe CO coverage ratios approaching the fully filled monolayer (1ML) limit, in Section 7.4. This is achieved through leveraging the novel capability of periodic DLPNO-MP2 in simulating multiple adsorption sites within a supercell, whilst ensuring the thermodynamic limit of the surface is consistently simulated.

7.2 Methods

7.2.1 Megacell-DLPNO-MP2

Earlier in this thesis, we introduce Megacell-DLPNO-MP2 as a local correlation scheme to obtain MP2 energies for periodic systems. The use of PNOs offers a significant compression of the virtual space, and we demonstrate sub-linear scaling with respect to supercell size in our earlier work. The degree of compression is controlled through a single variable, the PNO occupation number threshold, $\mathcal{T}_{\mathrm{PNO}}$, and the error incurred due to discarding virtuals is proportional to $\sqrt{\mathcal{T}_{\mathrm{PNO}}}$.

Previous computational studies of the surface adsorption energy of CO on MgO have incorporated higher accuracy correlated wavefunction methods, notably CCSD(T). Whilst our periodic DLPNO procedure only explores these adsorption systems at MP2 accuracy, we note that these previous studies highlight the comparable accuracy that MP2 achieves within their computational studies compared to their highest accuracy values[21,23,231], suggesting significant value in a periodic MP2 method that can efficiently simulate the thermodynamic bulk of the surface layer in an adsorption reaction.

The efficacy of the PNO approximation, combined with the translational symmetry savings, enables multiple unit cells, each featuring a CO adsorption site, to be simulated. Figure 7.1 presents unit cells corresponding to coverage ratios of $\frac{1}{4}$ and $\frac{1}{9}$, respectively, as well as 3^2 supercells of each system, where the CO molecule in the reference (central) unit cell are highlighted in white. Evaluating the adsorption energy for the reference cell CO using Megacell-DLPNO-MP2 crucially enables a separation between probing the correct thermodynamic limit of the bulk surface, which is achieved through including more unit cells within the supercell, and investigating different coverage ratios, which can be controlled by the size of the surface slab in the unit cell. In this contribution we predominantly employ 3^2 supercells, using 5^2 supercells occasionally to verify the thermodynamic limit convergence of our calculations. We believe this work is the first instance where a periodic post-HF scheme simulates multiple adsorption

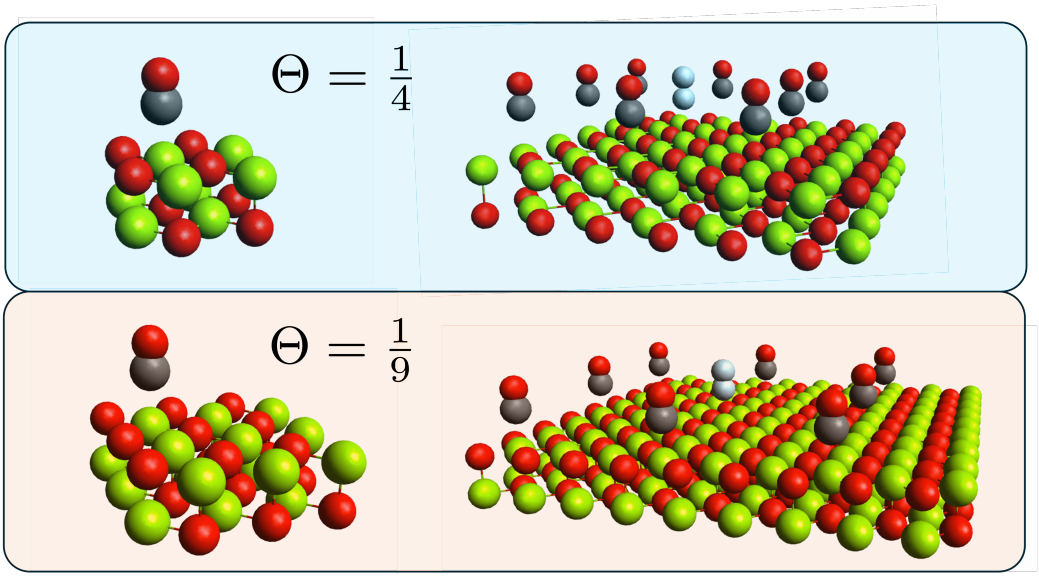

Figure 7.1: Unit cell structures for CO adsorption on an MgO surface corresponding to coverage ratios of $\frac{1}{4}$ and $\frac{1}{9}$. Individual unit cells are depicted on the left, whilst 3^2 supercells are plotted on the right. The reference cell CO molecule is highlighted in white.

sites in a supercell calculation, in contrast to previous studies which employ only single point (Γ) calculations.

7.2.2 Evaluation of Adsorption Energy

The adsorption energy is defined as the energy difference between the adsorbed CO molecule on the MgO surface, and the energies of the individual non-interacting components,

$$E_{\text{ads}} = E_{[\text{MgOCO}]} - E_{[\text{MgO}]} - E_{[\text{CO}]}, \tag{7.1}$$

where [MgO] is the pristine MgO surface and [CO] is the isolated gas-phase CO molecule. In line with previous works[19–21], the actual computation of E_{ads} is performed subtly differently, where the geometry of the non-interacting MgO surface and CO molecule are frozen to that of the interacting CO + MgO system, instead of the pristine or isolated geometries. A closely related quantity, E_{int}, is given by,

$$E_{\text{int}} = E_{[\text{MgOCO}]} - E_{[\text{MgO},\bar{\text{CO}}]} - E_{[\text{CO},\bar{\text{MgO}}]}. \tag{7.2}$$

As described by Shi and coauthors[20], computing E_{int} ensures basis set superposition error (BSSE) is accounted for. In our chosen notation, a system $[X, \bar{Y}]$ contains the

X subsystem with ghost functions representing $\bar{Y}$, in order to remove BSSE. To obtain the correct adsorption energy from E_{int}, an additional geometry relaxation term, Δ_{geom}, is also required to account for the energy change from the non-interacting, pristine geometries to the adsorbed one. Ye and Berkelbach[21] and Shi et al.[20] both compute Δ_{geom} using informed choices of DFT functionals, incorporating dispersion corrections, reaching a similar agreement of approximately 1.0 kJ mol^{-1}.

As highlighted earlier, the Megacell-DLPNO-MP2 scheme simulates multiple unit cells within the supercell, meaning the overall simulation cell features multiple CO molecules adsorbed onto the surface. With this approach, an additional lateral energy contribution must be considered, describing the cohesive energy when an isolated CO molecule is brought into a lattice of COs, without the interactions with the surface,

$$E_{\text{int,coh}} = E_{\text{int}} + E_{[\text{CO}]} - E_{[\text{CO},\bar{\text{C}}\bar{\text{O}}]}. \tag{7.3}$$

$E_{[\text{CO}]}$ is the energy of the gas-phase CO molecule within the lattice of COs. $E_{[\text{CO},\bar{\text{C}}\bar{\text{O}}]}$ is the energy of a single isolated CO molecule, but with ghost functions spanning all the neighbouring COs in the $E_{[\text{CO}]}$ system, in order to account for BSSE. Figure 7.2 is a schematic visualizing the contributions within $E_{\text{int,coh}}$. The final adsorption energy in this chapter is thus given by,

$$E_{\text{ads}} = E_{\text{int,coh}} + \Delta_{\text{geom}}. \tag{7.4}$$

7.2.3 Computational Details

Within our study, Megacell-DLPNO-MP2 is employed to evaluate all quantities contributing to $E_{\text{int,coh}}$, apart from $E_{[\text{CO},\bar{\text{C}}\bar{\text{O}}]}$. This calculation is aperiodic, featuring only a single CO molecule surrounded by ghost functions, which we compute using the existing molecular DLPNO-MP2 implementation[87,90] in the `pnoccsd` module, within the TURBOMOLE program. The Hartree–Fock orbitals for this system are obtained from the `ridft` module.

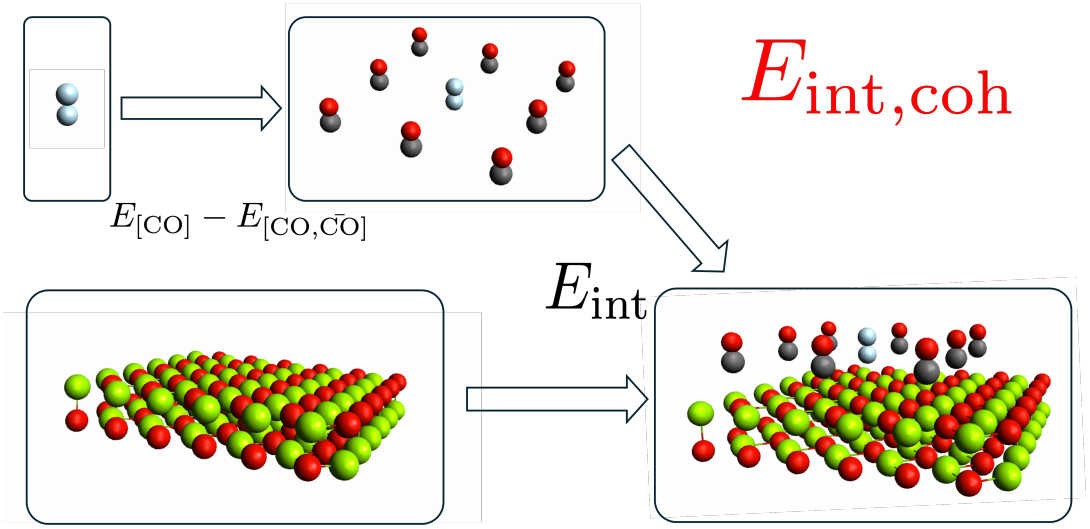

Figure 7.2: The contributions to the adsorption energy of CO on the MgO surface, for any given coverage ratio. The lateral interactions of the gas phase CO molecules must be added to E_{int} to properly distinguish adsorption energies at different coverage ratios.

The importance of using sufficiently expansive basis sets to capture the adsorption interactions has been highlighted extensively in previous works. In the initial phase of testing, we encountered issues converging the periodic HF calculations using all-electron cc-pVTZ[135,136] orbital basis sets, which we attributed to linear dependency issues arising from the diffuse basis functions centered on Mg, leading to divergent exchange contributions. HF convergence using the all electron pob-TZVP[93] and pob-TZVP-rev2[94,232] basis sets presented no issues, but since diffuse and additional polarization functions are vital for accurately capturing the dispersion interactions in adsorption systems, the lack of these functions within these basis sets also presented an issue. Our solution was to employ the cc-pVXZ orbital basis sets for C and O, and to use modified pob-XZVP-rev2 basis sets for Mg (X=D,T). These modified basis sets contained additional valence polarization and diffuse functions from the cc-pVXZ basis sets, as well as core polarisation functions from the cc-pwCVXZ basis sets. We thus performed two sets of adsorption calculations, at approximate 'DZ' and 'TZ' quality basis sets, enabling an estimate of the complete basis set limit values. The basis sets employed are included in Appendix D for clarity.

For the RI-J approximation within the periodic HF calculations, the cc-pVTZ auxiliary basis sets[138] were employed for carbon and oxygen atoms, whilst the

def2-TZVP auxiliary basis sets[138] were used for magnesium, for all calculations. The same auxiliary basis sets were also employed for the density-fitting treatment within the MP2 calculations, apart from for the ghost atoms, which did not feature any auxiliary basis functions.

Previous work from Shi *et al.*[20] highlights the underestimation of E_{int} if correlation from core electrons, particularly on magnesium, is ignored. We thus proceed with just freezing the 1s core shells for C, O and Mg, in our DLPNO-MP2 treatment. All calculations were run on a single node (Intel(R) Xeon(R) Gold 6248R CPU) with a maximum RAM limit of 386GB and 1.8Tb disk, with OMP parallelisation of up to 48 threads. No single calculation in this contribution exceeded two days of wall time, highlighting the efficiency of the Megacell-DLPNO-MP2 approach. Our largest calculations featured supercells containing just under 30000 orbital basis functions, representing the largest calculations undertaken by Megacell-DLPNO-MP2 thus far.

7.3 Towards to the dilute coverage limit

In this section, we study the adsorption of a single CO molecule on the pristine MgO surface, modeling the surface at the thermodynamic limit, and probing the adsorption energy towards the infinitely dilute limit. As mentioned earlier, a number of works, incorporating accurate correlated wavefunction methods, have already studied the adsorption energy of this system at the dilute regime. Alessio *et al.*[19] and Boese and Sauer[18] both employ hybrid MP2: DFT-D embedded cluster calculations, incorporating single point CCSD(T) calculations to obtain the final adsorption energy estimate, with values of -21.2 ± 0.5 kJ mol^{-1} and -21.0 ± 1.0 kJ mol^{-1}, respectively. Alessio *et al.* also employ periodic local MP2 to demonstrate agreement with their embedded cluster MP2 estimates. Shi *et al.*[20] used CCSD(T) calculations within their embedded cluster SKZCAM approach, obtaining an adsorption energy of -19.2 ± 1.0 kJ mol^{-1}, which they show agreement with canonical periodic CCSD(T) and diffusion Monte Carlo. Finally, Ye and Berkelbach[21] calculate adsorption energies using periodic local natural orbital (LNO) schemes at MP2 and CCSD(T) levels, calculating a value of -20.0 ± 0.5 kJ mol^{-1}. Considering that the

touted standard of 'chemical accuracy' is within 4.2 kJ mol^{-1}, the overall consensus of these methods is remarkable, considering the differences in the approaches taken and the need to converge out errors in all the method specific wavefunction and basis related parameters.

Experimental enthalpies for the adsorption of have been determined using temperature-programmed desorption (TPD) techniques, of which we highlight the work from Dohnálek *et al.*[228], and Wichtendahl *et al.*[229]. Both studies produce TPD spectra at varying CO coverages, reporting desorption values at minimal coverages of around 0.3ML, which provide the closest estimate to the adsorption energy at the dilute limit. As noted by Boese and Sauer, as well as Shi *et al.*, conversion of these enthalpy values to a comparable adsorption energy requires removal of thermal and zero-point energy contributions, as well as accounting for the PV term. From experiments from Dohnálek *et al.*, Boese and Sauer[18] obtained a TPD-derived energy of -20.6 ± 2.4 kJ mol^{-1}, whilst Shi *et al.*[20] obtain an experimental E_{ads} estimate of -19.2 ± 1.0 kJ mol^{-1}, where further optimization of the pre-exponential factor in the Redhead equation was considered, based off averaging the two experimental values[228,229]. With both experiment-derived and theoretical studies from a number of different works agreeing to well within chemical accuracy, the adsorption energy of CO on MgO at the dilute regime appears to have reached a robust consensus, providing a rigorous benchmark for our study using periodic DLPNO-MP2.

7.3.1 Dilute Regime Results

In our study to probe the dilute limit adsorption energy, we follow similar protocols to Ye and Berkelbach[21] and Shi *et al.*[20]. We model the MgO (001) surface using a two-layer slab, which both studies previously determine is sufficient to converge the adsorption energy with respect to the layer depth. The CO molecule is adsorbed in a perpendicular fashion to the surface, with the C end pointing towards the five-fold coordinated Mg site, with a Mg-C interaction distance set to 2.460$\mathring{A}$, to enable direct comparison with the aforementioned two works.

We employ three types of unit cell to probe the dilute regime: a 2×2, 3×3 and 4×4 surface slab of 2-layer MgO (2×2 and 3×3 unit cells are presented in Figure 7.1). We directly use the optimized equilibrium geometries for the MgO CO system reported by Ye and Berkelbach, which they obtain through optimization with DFT-D functionals. Our geometry relexation energy, Δ_{geom}, is thus also 1.1 kJ mol^{-1}. The crucial difference with our study is that our supercell periodic approach can indepedently examine different coverage ratios (in this case $\frac{1}{4}$, $\frac{1}{9}$ and $\frac{1}{16}$ ML), whilst also converging the thermodynamic limit of the surface through increasing the supercell size.

Table 7.1: Hartree–Fock and MP2 correlation energies for each system involved in evaluating $E_{\text{int,coh}}$, comparing supercell sizes of 3^2 and 5^2. All energies are given in Hartrees, apart from the final row. Calculations employed the 2×2 surface slab unit cell, using the modified 'TZ' basis set.

System	HF		$E_{\text{corr}}(\mathcal{T}_{\text{PNO}} = 10^{-7})$		$E_{\text{corr}}(\mathcal{T}_{\text{PNO}} = 10^{-8})$		CPS	
n_{super}	3^2	5^2	3^2	5^2	3^2	5^2	3^2	5^2
[MgOCO]	-2310.10796	-2310.10796	-4.14200	-4.14224	-4.14370	-4.14412	-4.14449	-4.14499
[MgO, $\bar{\text{C}}$O]	-2197.33009	-2197.33009	-3.77718	-3.77739	-3.77861	-3.77898	-3.77927	-3.77972
[CO, $\bar{\text{M}}$gO]	-112.77868	-112.77868	-0.35807	-0.35807	-0.35814	-0.35814	-0.35817	-0.35817
[CO]	-112.77842	-112.77842	-0.35712	-0.35712	-0.35714	-0.35714	-0.35715	-0.35715
[CO, $\bar{\text{C}}$O]	-112.77855	-112.77855	-0.35685	-0.35685	-0.35687	-0.35687	-0.35693	-0.35693
$E_{\text{int,coh}}$	0.00094	0.00094	-0.00702	-0.00705	-0.00723	-0.00727	-0.00732	-0.00738
$E_{\text{int,coh}}$(kJ mol^{-1})	2.47	2.47	-18.42	-18.51	-18.97	-19.11	-19.23	-19.39

We first verify the thermodynamic limit convergence of the adsorption calculations. Table 7.1 presents the HF and MP2 correlation energies (E_{corr}) for each system involved in the evaluation of $E_{\text{int,coh}}$, for the 2×2 surface slab unit cell at 'TZ' quality, for supercell sizes of 3^2 and 5^2. Computing energies for the 3×3 and 4×4 surface slab unit cells at 5^2 supercell sizes is currently prohibited due to the current memory bottlenecks within our computational setup of Megacell-DLPNO-MP2, but given that the 2×2 unit cell represents the densest coverage ratio, we expect the thermodynamic limit convergence for the larger unit cells to be faster than this. The MP2 correlation energies are computed at two $\mathcal{T}_{\text{PNO}}$ thresholds, which are then used to estimate the complete PNO space (CPS) limit using a square root extrapolation.

First examining the HF energies, we note exact agreement between 3^2 and 5^2 supercells, for each system, to a precision of 10^{-5} Hartrees, indicating thermodynamic limit convergence. For the MP2 correlation energies, only the systems

simulating the MgO surface layer, [MgOCO] and [MgO, $\bar{\text{C}}$O], show small deviations between the the two supercell sizes, on the order of 10^{-4} Hartrees, for both $\mathcal{T}_{\text{PNO}}$ thresholds. However, these deviations largely cancel in evaluating $E_{\text{int,coh}}$, leading to energies with differences on the order of 0.1 kJ mol^{-1}. The difference between the extrapolated CPS limit and $\mathcal{T}_{\text{PNO}} = 10^{-8}$ energies are also of a similar magnitude. The close agreement of these energies, especially when compared to the magnitude of uncertainty when converging to the basis set limit, which we subsequently discuss, allows us to conclude that both the thermodynamic limit of the surface slab, and the canonical limit have been sufficiently converged.

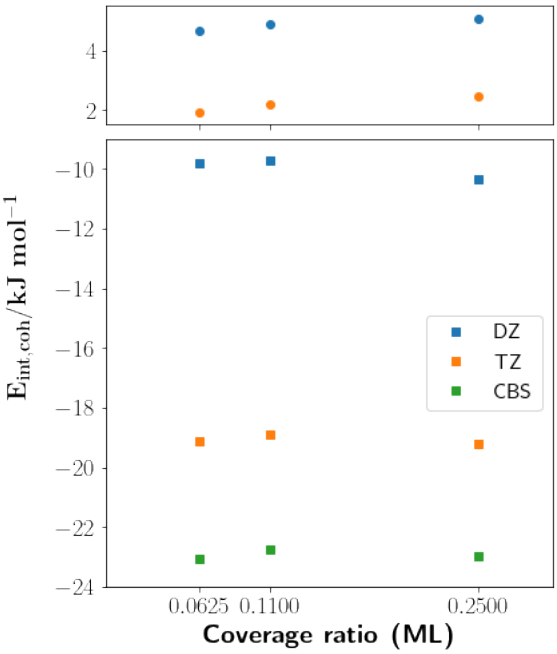

Figure 7.3: Hartree–Fock (Top) and MP2 correlation energies (Bottom) for $E_{\text{int,coh}}$, at three dilute coverage ratios. Results using the modified 'DZ' and 'TZ' basis sets are shown, and complete basis set estimate is provided for the MP2 correlation energies.

Figure 7.3 presents the $E_{\text{int,coh}}$ energies at increasing coverage ratios, obtained from the 4×4, 3×3 and 2×2 surface slab unit cells, respectively. The top panel presents the HF energies, whilst the bottom plots the MP2 correlation energies. For both cases, the cohesive energies for different coverage ratios are evaluated using the 'DZ' and 'TZ' basis sets. A 3^2 supercell was employed in all cases. For the MP2 correlation energies, an inverse cube extrapolation is performed to estimate the basis set limit value, using Helgaker's two-point extrapolation formula.

We first highlight that the cohesive energies are largely similar in magnitude across the different coverage ratios, indicating that all three coverage ratios capture the dilute coverage regime. Whilst a slight increase in MP2 correlation energy is observed from 0.25 to 0.11 coverage ratios, this deviation is on a similar order to the uncertainties incurred from the CPS and thermodynamic limit extrapolations, and thus we do not deem it to be significant. The HF cohesive energies show a very slight monotonic decrease towards further dilute coverages, but the differences are again so small that any extrapolation to the infinitely dilute regime would not be insightful. We stress that these constant coverage ratio values are only possible due to the large 3^2 supercell size, afforded by the Megacell-DLPNO-MP2 method, which is able to accurately sample the thermodynamic limit of the MgO surface, for the adsorbed CO molecule in the reference unit cell. In contrast, Ye and Berkelbach[21] show a significant decrease in correlation energy from the 2×2 to 4×4 unit cells, with their periodic MP2 scheme. We attribute this to the single point calculations used, which do not capture the full extent of the surface, and hence the larger surface sizes lead to substantial increases in the magnitude of the interaction energy.

Figure 7.3 shows that extrapolation of the 'DZ' and 'TZ' quality basis sets to a basis set limit estimate gives a significant decrease in MP2 correlation cohesive energies compared to the 'TZ' values, on the order of 4 kJ mol^{-1}, across all coverage ratios. This source of error is the largest component of uncertainty within our scheme, and is admittedly significantly larger than previous computational schemes, which we record in Table 7.2. We point to our current limitations in converging periodic HF calculations using cc-pVTZ or higher quality basis sets as the source of this issue, preventing of us conducting calculations at QZ quality basis sets, and subsequent TZ, QZ extrapolations to obtain a narrower error bar, which Ye and Berkelbach[21] conduct.

The final adsorption energies obtained via Megacell-DLPNO-MP2 are presented in Table 7.2, compared to previous periodic or cluster based schemes, as well as experimental TPD results. Our values are obtained by summing the HF energy for the 'TZ' basis with the CBS extrapolated MP2 correlation energy. The three

Table 7.2: Comparison of adsorption energies of CO on the MgO (001) surface using Megacell-DLPNO-MP2 and recent results from the literature. 'cl' and 'pbc' refer to cluster and periodic wavefunction schemes, respectively. Experimental results from temperature-programmed desorption (TPD) experiments are also presented.

Coverage ratio Θ	E_{ads} (kJ mol^{-1})	Methodology
0.2500	-19.4 ± 2.1	Megacell-DLPNO-MP2
0.1111	-19.5 ± 2.1	Megacell-DLPNO-MP2
0.0625	-20.0 ± 2.2	Megacell-DLPNO-MP2
dilute limit	-21.2 ± 0.5	cl MP2:DFT-D+ΔCC[19]
dilute limit	-21.6 ± 0.3	pbc MP2[19]
0.125	-21.0 ± 1.0	cl MP2:DFT-D+ΔCC[18]
dilute limit	-18.8 ± 0.3	pbc MP2[21]
dilute limit	-20.0 ± 0.5	pbc CCSD(T)[21]
dilute limit	-18.5 ± 0.5	cl MP2[20]
dilute limit	-19.2 ± 0.6	cl CCSD(T)[20]
0.3	-20.6 ± 2.4	TPD[18,228]
0.3	-19.2 ± 1.0	TPD[20,228,229]

coverage ratios probed all fall within the dilute coverage regime, and we note the general excellent agreement of our values, within 2 kJ mol^{-1} of all other values presented on the Table. This is perhaps unexpected, given previous evidence of the good agreement of periodic and cluster MP2 calculations with higher level coupled cluster results. Clearly, our larger error bars, obtained from halving the difference between the 'TZ' and CBS extrapolated correlation energies, are a point of concern, stemming from our limitations in basis set. Despite this, the overall agreement of the scheme gives promise for Megacell-DLPNO-MP2 as a robust scheme to probe different adsorption coverages.

7.4 Probing dense coverages

Whilst a number of wavefunction-based studies probing the adsorption energy of CO on MgO at the dilute coverage limit now achieve excellent agreement with experimental values, considerably less work has been done to simulate the adsorption energy dependence at higher coverage ratios. From the perspective of simulations, the inherent difficulty is now the vastly increased degrees of freedom that now need to be analyzed. Due to the lateral interactions between CO molecules, which are

now significant at higher coverage densities, differing (i.e. non-perpendicular) CO orientations, configurations or clusters now need to be considered, in order to obtain the lowest energy state for a given coverage ratio. Whilst some experiments give details of highly ordered c(4x2) phases[233,234] at particularly high coverage ratios (0.9-1 ML), where lateral repulsions between COs dominate and a significantly less exothermic adsorption energy is reported, relatively little consensus has been established for favored phases between 0.25 to 0.8 ML coverages, providing no benchmark systems for computational approaches. Dohnálek *et al.*[228] do provide an experimentally-derived adsorption energy plot as a function of relative CO coverage, which confirms the effect of increasing CO lateral repulsions on the adsorption energy.

Cluster calculations are inherently disadvantaged, since modeling multiple CO adsorption sites to capture the lateral interactions between a central CO in principle involves building a much larger fragment. Recent work by Shi *et al.*[235] provide a solution by evaluating the cohesive energy of the lattice of COs in the absence of the surface at CCSD(T) accuracy, and incorporating a correction for the effect of the MgO surfaces, using DFT. By decomposing the adsorption energy in this way, the subsequent evaluation of E_{int} just computes the interaction of the individual monomers on the MgO surface, although the lateral interactions under the effect of the surface are never directly explored to correlated wavefunction accuracy.

Wavefunction-based periodic approaches are more naturally suited to modelling denser coverage ratios, since the adsorbed layer at higher ratios increasingly resemble periodic models. However, large supercells, beyond single point calculations, are required to correctly capture the full extent of the periodic lateral interactions. Furthermore, larger unit cells are needed if differing CO orientations or configurations are to be probed. Computational studies are scarce, but we highlight the work by Minot *et al.*[236], using periodic Hartree–Fock , which reports a combination of perpendicular and bent CO geometries at $\Theta = 0.75$, presenting two different arrangements which are energetically comparable. To our knowledge, no periodic correlated wavefunction studies, targeting denser coverage ratios, exist, and studies with DFT are also scarce.

We expect Megacell-DLPNO-MP2 to be a suitable method to accurately model denser coverages at the thermodynamic limit, since its excellent cost to accuracy ratio enables multiple unit cells to be simulated, as verified in our earlier study of the dilute coverage regime. In this section, we choose not to explore the full parameter space involving non-perpendicular CO geometries or differing configurations, which would require a substantially more involved systematic study. Instead, we use Megacell-DLPNO-MP2 to simulate unit cells featuring perpendicular COs at increasingly dense coverage ratios, approaching $\Theta = 1$. Although not directly comparable to experimental TPD studies, we still expect to capture the effect of the lateral repulsions between COs with our simplified scheme, and observe the trend of decreasing magnitude of E_{ads}, going towards $\Theta = 1$.

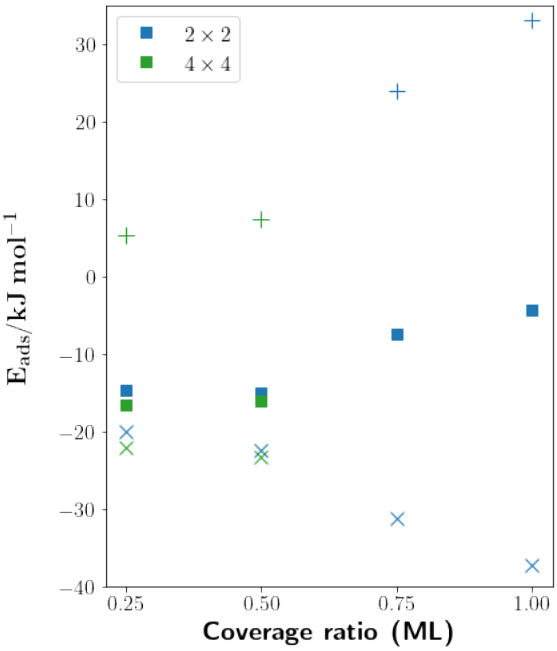

Figure 7.4: CO adsorption energies on the MgO surface at varying coverage ratios, up to full monolayer coverage. Hartree–Fock and MP2 correlation energy contributions are plotted as '+' and 'x's, respectively. Results employing the 4×4 surface slab unit cell are also reported for the two most dilute coverage ratios.

7.4.1 Dense Coverage Results

In our study to probe larger coverage ratios, we use a pristine version of the 2×2 surface slab unit cell employed earlier, using the same Mg-C interaction distance.

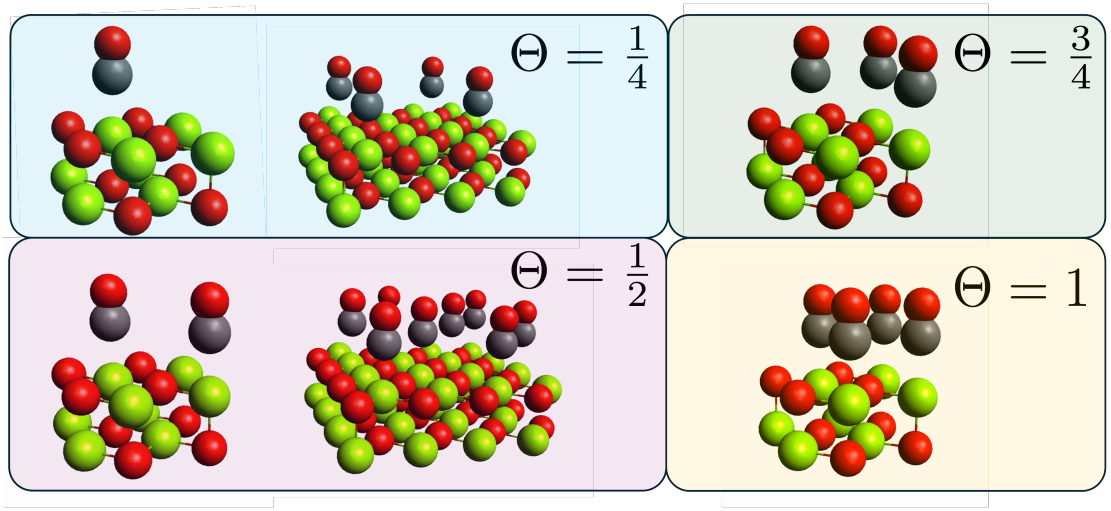

Figure 7.5: 2×2 and 4×4 surface slab unit cells used to probe adsorption coverage ratios towards the dense regime.

Since no geometry optimization has been performed, Δ_{geom} is zero. We populate the unit cells with one, two (separated diagonally), three and four CO adsorption sites, in order to capture coverage ratios of 0.25, 0.5, 0.75 and 1, respectively. We also utilize a pristine 4×4 surface slab unit cell, with 4 and 8 CO adsorption sites, to again probe 0.25 and 0.5 surface coverages, to verify the consistency of the E_{ads} calculations using Megacell-DLPNO-MP2. We present all unit cells employed in Figure 7.5. A supercell size of 3^2 is used throughout. Due to a bottleneck in disk memory size within the current implementation, evaluation of multiple CO sites in a given unit cell is currently limited to PNO thresholds of 10^{-6}. We use the same modified 'TZ' basis set as before.

With multiple CO adsorbates now in a given unit cell, the expression to evaluate the adsorption energy now needs to be modified,

$$E_{\text{ads}} = E_{\text{int,coh}} = -E_{[\text{CO},\bar{\text{C}}\bar{\text{O}}]} + \frac{E_{[\text{MgOCO}]} - E_{[\text{MgO},\bar{\text{C}}\bar{\text{O}}]} - E_{[\text{CO},\bar{\text{M}}\bar{\text{g}}\bar{\text{O}}]} + E_{[\text{CO}]}}{n_{\text{CO}}} \tag{7.5}$$

where the first equality is due to the lack of geometry relaxation energy in this specific scheme. n_{CO} is the number of COs per unit cell, and $E_{[\text{CO},\bar{\text{C}}\bar{\text{O}}]}$ is the energy of a single CO molecule, with ghost functions representing all remaining COs spanning the reference unit cell and neighboring cells.

Figure 7.4 presents the adsorption energy of CO on the MgO surface, at varying coverage ratios, up to full monolayer coverage ($\Theta = 1$). The HF and MP2 correlation energy contributions are plotted as '+' and 'x', respectively. Pleasingly, the simulated adsorption energy values show an overall decrease in magnitude from $\Theta = 0.25$ to $\Theta = 1$, with the least exothermic adsorption energy of -4.3 kJ mol^{-1} obtained at full monolayer coverage, reflecting the increased lateral repulsion exhibited at higher ratios. One notes the comparative larger increase of the HF repulsion, compared to the attractive dispersive MP2 correlation contribution, as the source of this overall behavior. In particular, we can attribute the marked increase in adsorption energy from $\Theta = 0.5$ to $\Theta = 0.75$ directly to the inclusion of nearest neighbor CO repulsions, which are present in the $\Theta = 0.75, 1$ unit cells (see Figure 7.5). In comparison, we can reason that the diagonal neighbor CO lateral repulsions are much less significant, given the similar adsorption energies between $\Theta = 0.25$ to $\Theta = 0.5$.

Again, we reiterate caution in directly comparing to the TPD-derived adsorption energy curves presented by Dohnálek *et al.*[228], since no account for variable CO orientation, or phases of CO occupation, are considered in our calculations. Furthermore, we also do not provide complete basis set or PNO space estimates for these calculations, due to computational limitations in our current scheme. For example, examining Figure 4 in Ref.228, it is clear that there is disagreement in the adsorption energy at full monolayer coverage, but we have not simulated the c(4x2) phase that the authors report to be present. Nonetheless, the qualitative behavior of the lateral repulsions, captured entirely within a consistent MP2 level of theory, indicates promise in simulating unit cells featuring multiple sites. In addition, we can posit that no nearest neighbor adsorption configurations can be responsible for the experimentally-derived adsorption energies between 0.3 to 0.8 ML, since our calculations report significantly smaller adsorption energies. With further improvements in the Megacell-DLPNO-MP2 framework, realistic simulation of even more sophisticated geometries appears feasible.

Comparison between the 2×2 and 4×4 unit cells at equivalent coverages point to slight disagreement, on the order of around 1 to 2 kJ mol^{-1}, originating in the MP2

correlation energy. This difference remains even when validating the thermodyanmic limit convergence using 5^2 supercells. We believe this is due to the loose PNO threshold (10^{-6}) used, which may inconsistently capture the pair correlation energies between the two different unit cells. Again, further implementation improvements to the scheme will reveal whether these differences are converged out at the correct complete PNO space limit.

Finally, to further analyze the origin of the lateral interactions, we also compute the interaction energy without the cohesive contribution, E_{int}, per CO monomer. In Table 7.3, these values are shown to be relatively consistent across the various coverage ratios, in contrast to complete cohesive adsorption energies. We thus suggest that the difference in lateral repulsion between the CO monomers in the gas phase, compared to with the influence of the MgO surface, is negligible. In turn, this also validates the approach taken in Ref.235, where the gas-phase CO interactions are used to approximate the cohesive contribution to the adsorption energy, as mentioned earlier. Again, calculations at stricter PNO thresholds are required in order to solidify this statement.

Table 7.3: Comparison of the full cohesive adsorption energy ($E_{\text{int,coh}}$), and the adsorption energy neglecting gas phase CO lateral interactions(E_{int}), at different coverage ratios.

Coverage ratio Θ	$E_{\text{int,coh}}$ (kJ mol^{-1})	E_{int} (kJ mol^{-1})
0.25	-14.6	-14.4
0.50	-15.1	-13.5
0.75	-7.5	-13.9
1.00	-4.3	-14.4

7.5 Conclusions

Correlated wavefunction approaches have been increasingly applied to study adsorption problems, of which CO adsorption on the MgO(001) surface is one of the most fundamental and commonly used case studies. Prior to this work, cluster and periodic wavefunction methods employing MP2 and Coupled Cluster theory have established excellent agreement with experimentally-derived values for the adsorption energy of a single CO molecule onto a pristine MgO surface.

In this study, we demonstrate the capabilities of Megacell-DLPNO-MP2 in probing the adsorption of CO on MgO. Crucially, a key focus has been isolating the coverage ratio from the convergence of the surface towards the full thermodynamic limit, which is enabled through our supercell approach. We successfully demonstrate similar estimates of the dilute regime adsorption energy (around -20 kJ mol^{-1}), at three different coverage ratios, compared to previous computational and experimental approaches, albeit with larger uncertainties due to our current basis set restrictions.

Tackling the adsorption energy of denser coverages is an inherently trickier problem, due to the vastly larger configuration space that needs to be considered, as well as differing non-perpendicular geometries. Here, we employ a simplified model consisting of just perpendicularly adsorbed COs, evaluating adsorption energies as a function of monolayer coverage ratio, factoring in the effect of the lateral repulsion between COs. We achieve a qualitative agreement with experiment of decreased exothermic behavior at higher coverage densities, although clearly our values are hindered by our current PNO truncation and basis set choices, and a deeper study is required to explore the larger configuration space.

Whilst the agreement of the CO adsorption energy on the pristine MgO surface between modern quantum chemistry approaches and experimental schemes is already impressive, further work is still required to model surface interactions that may be present in realistic heterogeneous catalysis applications. Surfaces are not completely pristine in reality, and the presence of terraces, dopants, defects and kinks on the surface are now understood to have significant influences on the binding with adsorbates[237–244]. Dohnálek *et al.*[228], for example, report significantly more exothermic adsorption energies associated with defect-related sites, which become the dominant mechanism at extremely low coverages. Periodic correlated wavefunction methods need to continue to improve to model increasingly large unit cells if these effects are to be captured, and whilst embedded cluster approaches with DFT have been reported, to our knowledge there is significantly fewer studies employing correlated wavefunction methods. In this regard, with further improvements, we

believe periodic DLPNO-MP2 to be a viable wavefunction based candidate to explore sophisticated adsorption scenarios, treating periodic monolayer, defect, kink or terraced surfaces all on equal footing, thanks to its computational efficiency for simulating supercells.

8| Concluding Remarks

The use of correlated wavefunction schemes to obtain ground state properties for periodic systems has grown considerably in the last two decades. This is highly beneficial, since it provides a practical alternate option to DFT, for the study of materials, and establishes robust, systematically improvable benchmarks to aid in the development of future methods or density functionals. In order for post-HF methods to be viable choices for simulating large, complex unit cells, or to converge to the thermodynamic limit, consideration must be taken to avoid the expensive canonical scaling. Adapting local correlation schemes to periodic systems provides a solution to this problem. DLPNO theory has established itself as a highly successful local correlation scheme for molecular systems, offering the tightest compression of the virtual space compared to other choices of virtual orbitals.

We present two novel DLPNO-MP2 schemes which evaluate unit cell correlation energies for crystalline systems, solving under periodic boundary conditions. The BvK-DLPNO-MP2 scheme explicitly accounts for the infinite lattice summation within the integral evaluation, employing a charge projected and surface dipole corrected local density fitting scheme to obtain convergent sums. In contrast, Megacell-DLPNO-MP2 uses the DLPNO-MP2 equations at the thermodynamic limit, taking the lattice summation outside of the integral evaluation, which thereby necessitates the use of an encompassing megacell around the supercell correlation treatment, to preserve periodicity. In both schemes, we prepare localised Wannier functions using an optimisation scheme employing charges from Bloch IAOs, which is detailed and analysed in Chapter 3.

Results from Chapter 6 show that the PNO approximation is consistent for both Megacell and BvK schemes, obtaining the same value at thermodynamic limit for a given PNO truncation threshold. Molecular fragment benchmarks demonstrate that both schemes converge to the correct canonical limit when approaching the CPS limit, validating the accuracy of both methods. Although proof-of-concept results

of one-, two- and three-dimensional systems are presented for both implementations, we note that the slow convergence of the lattice summation for the ERIs within the pilot version of BvK-DLPNO-MP2 forms the main CPU bottleneck, hindering the capacity to simulate larger unit cells or larger supercells, compared to the Megacell scheme. The inclusion of fast multiple acceleration schemes will vastly reduce this prefactor, in the future. In contrast, our current version of Megacell-DLPNO-MP2 encounters bottlenecks from our RAM and disk limits, and full implementation of translational symmetry savings is needed to improve this.

In spite of this, Megacell-DLPNO-MP2 can already compute correlation energies for some very large systems, for example when applied to study surface adsorption, in Chapter 7, featuring supercells of close to 30000 basis functions. Here, we show that Megacell-DLPNO-MP2 reproduces the well-benchmarked adsorption energy of a single CO on the MgO surface, and we extend this model to consider dense coverage regimes of up to eight adsorption sites within a unit cell. To reduce our reported errors, we require higher quality basis sets that are suitable for periodic systems. All in all, this thesis has outlined implementations that demonstrate the validity of periodic DLPNO-MP2 theory, and shown that ground state properties, such as adsorption energies, are now accessible using these methods.

A periodic DLPNO-MP2 method lays the foundation for a DLPNO-CCSD(T) treatment for crystalline systems. CCSD(T) is one of the most accurate schemes for incorporating dynamic correlation within molecular quantum chemistry. Adaption of this method, within a DLPNO framework, under periodic boundary conditions promises highly accurate, large-scale calculations for ground state properties of crystalline systems. In the longer term future, we again note the widespread applications of DLPNO theory, including for multireference systems[177–179] or for excited states[180–183], within molecular electronic structure theory. With a periodic DLPNO-MP2 implementation now established, one can envisage generalising these theories in analogous fashion to develop schemes which can probe excited state phenomena in periodic systems, or simulate strongly correlated materials, contributing to the richness of modern day methods in computational materials science.

Appendices

A| Additional details on Wannier Localisation

The IAO charge and PM metric gradient expression, as given by Eqs. 3.23 and 3.24, are again presented here,

$$\begin{aligned} Q_i^{A_{\boldsymbol{R}}^{\mathrm{IAO}}} &= \langle \phi_{i,\mathbf{0}} | \, \hat{P}^{\mathrm{IAO}}_{A_{\boldsymbol{R}}} \, | \phi_{i,\mathbf{0}} \rangle = \\ &\frac{1}{N^2} \sum_{\mu \in A} \Big[\sum_{j,\mathbf{k}} \bar{C}^{*\mathbf{k},\boldsymbol{R}}_{\mu,j} \, U^{*\mathbf{k}}_{j,i} \Big] \Big[\sum_{j',\mathbf{k}'} C^{\mathbf{k}',\boldsymbol{R}}_{\mu,j'} \, U^{\mathbf{k}'}_{j',i} \Big], \end{aligned} \tag{A.1}$$

$$\begin{aligned} \frac{\partial \langle O \rangle_{\mathbf{PM}}}{\partial U^{*\mathbf{k}}_{j,i}} &= \frac{\partial \langle O \rangle_{\mathbf{PM}}}{\partial Q_i^{A_{\boldsymbol{R}}^{\mathrm{IAO}}}} \frac{\partial Q_i^{A_{\boldsymbol{R}}^{\mathrm{IAO}}}}{\partial U^{*\mathbf{k}}_{j,i}} \\ &= \frac{p}{N^2} \sum_{\boldsymbol{R},A} |Q_i^{A_{\boldsymbol{R}}^{\mathrm{IAO}}}|^{p-1} \sum_{\mu \in A} \bar{C}^{*\mathbf{k},\boldsymbol{R}}_{\mu,j} \Big[\sum_{j',\mathbf{k}'} C^{\mathbf{k}',\boldsymbol{R}}_{\mu,j'} U^{\mathbf{k}'}_{j',i} \Big] \\ &+ C^{*\mathbf{k},\boldsymbol{R}}_{\mu,j} \Big[\sum_{j'',\mathbf{k}''} \bar{C}^{\mathbf{k}'',\boldsymbol{R}}_{\mu,j''} \, U^{\mathbf{k}''}_{j'',i} \Big]. \end{aligned} \tag{A.2}$$

where $C^{\mathbf{k},\boldsymbol{R}}_{\mu,j} = C^{\mathbf{k}}_{\mu,j} e^{i\mathbf{k}\boldsymbol{R}}$ and $\bar{C}^{\mathbf{k},\boldsymbol{R}}_{\mu,j} = \sum_{\nu,\boldsymbol{R}'} C^{\mathbf{k},\boldsymbol{R}'}_{\nu,j} S^{\mu,\boldsymbol{R}}_{\nu,\boldsymbol{R}'}$. To reduce complexity cost, the following intermediates are defined,

$$X_{\mu,l,\mathbf{R}} = \sum_{\mathbf{k}} V_{\mu,l,\mathbf{k}} e^{i\mathbf{k}\mathbf{R}} = \sum_{j,\mathbf{k}} C^{\mathbf{k}}_{\mu,j} U^{\mathbf{k}}_{j,l} e^{i\mathbf{k}\mathbf{R}} \tag{A.3}$$

$$Z_{l,\mu,\mathbf{R}} = \sum_{\nu,\mathbf{R}'} X_{l,\nu,\mathbf{R}'} S^{\mu,\mathbf{R}}_{\nu,\mathbf{R}'} = \sum_{\nu,\mathbf{R}',j,\mathbf{k}} C^{\mathbf{k}}_{\nu,j} U^{\mathbf{k}}_{j,l} e^{i\mathbf{k}\mathbf{R}'} S^{\mu,\mathbf{R}}_{\nu,\mathbf{R}'} \tag{A.4}$$

$$Y_{j,\mu,\mathbf{R},\mathbf{k}} = \sum_{\nu} C^{\mathbf{k}}_{\nu,j} W_{\nu,\mu,\mathbf{R},\mathbf{k}} = \sum_{\nu,\mathbf{R}'} C^{\mathbf{k}}_{\nu,j} e^{i\mathbf{k}\mathbf{R}'} S^{\mu,\mathbf{R}}_{\nu,\mathbf{R}'} \tag{A.5}$$

Using these tensors, the IAO charge simplifies to

$$Q_l^{A_{\boldsymbol{R}}^{\mathrm{IAO}}} = \frac{1}{N^2} \sum_{\mu \in A} Z^*_{l,\mu,\mathbf{R}} \, X_{\mu,l,\mathbf{R}}. \tag{A.6}$$

The PM metric gradient, Eq. 3.24, is now given by

$$\frac{\partial \langle O \rangle_{\mathbf{PM}}}{\partial U^{*\mathbf{k}}_{j,i}} = \frac{p}{N^2} \sum_{\mathbf{R},A} |Q_i^{A_{\mathbf{R}}}| \sum_{\mu \in A} (Y^*_{j,\mu,\mathbf{R},\mathbf{k}} \, X_{\mu,i,\mathbf{R}} + C^{*\mathbf{k}}_{\mu,j} e^{-i\mathbf{k}\mathbf{R}} Z_{i,\mu,\mathbf{R}}). \tag{A.7}$$

The tensors $Z_{l,\mu,\mathbf{R}}$ and $X_{\mu,l,\mathbf{R}}$ are evaluated at each iteration step, since they have a dependence on the current unitary, $U^{\mathbf{k}}_{j,l}$. $Y_{j,\mu,\mathbf{R},\mathbf{k}}$ is evaluated once at the beginning of the optimisation. In terms of complexity, evaluation of $Z_{l,\mu,\mathbf{R}}$ has the highest scaling cost for each iteration, with an order $l\mu^2\mathbf{k}^2$ cost. Constructing $Y_{j,\mu,\mathbf{R},\mathbf{k}}$ scales with $\mu^2\mathbf{k}^3$, which has the largest cost with respect to $\mathbf{k}$-grid size, but is only evaluated once. In terms of memory cost, the largest tensor stored is the real space overlap $S^{\mu,\boldsymbol{R}}_{\nu,\boldsymbol{R}'}$.

B| List of code changes for periodic DLPNO-MP2

The following is a list of all the modified or newly generated subroutines within the `pnoccsd` module in the developmental TURBOMOLE package that this thesis is based on. Most of these code changes were motivated by the need to exploit translational symmetry. It thus provides a summary of all the subroutine changes for Megacell-DLPNO-MP2, and some of the translational symmetry accelerations for BvK-DLPNO-MP2. It does not include the code changes specifically for the surface-corrected, charge-projected density fitting scheme, for BvK-DLPNO-MP2, which were implemented by Arman Nejad. It also does not include the routines for the novel Wannier localisation scheme, outlined in Chapter 3, which are housed in the `riper` module. Code changes specific to BvK-DLPNO-MP2 are denoted by †, whilst Megacell-DLPNO-MP2 $*$. Newly created routines are highlighted by •. The routines are ordered in approximate running sequence.

`pnoccsd`, `pno_getpnos`

- Logical variable control of `lbvk` and `ltranssym`†$*$

`num3idxovlnew`•

- A replacement to `num3idxovl`, with reduction to stack cost through reordering allocations outside OMP loop†$*$
- CPU saving through screening of basis functions only in reference cell, rather than megacell$*$

`type_cells`

- Added data structures for Megacell implementation, including the arrays providing information on the supercell boundary for each cell within the reference supercell$*$

`pno_locdom`

- Restrict auxiliary basis function domains for each LMO to the reference supercell- CPU saving for `pno_mkbqocc`∗

`pno_paoselect_den`

- Restrict PAO domains for each LMO to the supercell around the LMO∗

`pno_osvs`

- Megacell to supercell RAM saving introduced for `cmoz` and `dmoz` arrays∗

`pno_muizq`

- Saving only translationally invariant stripe of `dmoz` and `cmoz`- associated RAM savings through reorganising scratch arrays- inserted an additional memory check∗

`pno_mkbqocc`

- 1st call- use `riusemega` to restrict auxilary basis function summation for CPU and disk savings, only save i_0 integrals for disk savings∗
- 2nd call- only save for i_0 integrals for disk savings∗
- Both calls- AO screening just for LMOs in reference cell∗
- Restrict evaluation of raw $(\mu\nu|P)$ integrals to only auxiliary basis functions in the reference cell†
- Generates LMO integrals $(\mu i|P)$ integrals through copying/reordering/reindexing of $(\mu i|P_0)$†

`pno_iterosvgen`

- Only compute reference cell OSVs for compute savings∗
- Regenerating full `cmoz/dmoz` in a temporary scratch array∗

`locbas_setprs`

- New `iopt=4` which ensure `%nstore` spans the full megacell- to work with `cmoz dmoz` disk savings∗

`pno_intdiag`

- Skipping evaluation if not in reference cell for CPU savings∗

`pno_intdiag`, `pno_bqocctraf`, `pno_bqosv`, `pno_rdosvints`

- Teething with `riusemega` for CPU savings∗
- `ioffmega` offset implemented for disk savings- storing only translationally unique 3-index integrals∗

`pno_pbcosvs`•

- Enforces translational invariance of osvs by copying reference cell osvs to all other cells∗†
- Repoints associated PAO information to enforce translational invariance of osvs∗†

`pno_bqosv`,`pno_bqosv_megacell`•

- Evaluates B(Q,ai) intermediates in the OSV basis only for the reference cell∗†
- Regenerates non-reference cell intermediates through translational copying from reference cell quantities∗†

`pno_pairscreen`

- Restrict to only reference cell contributing pairs in the `peridist` coarse screening∗
- Restrict pairs to $i_0 j_l$ or $i_l j_0$ where l is within the supercell in the final pair selection∗†

`pno_prepnos`

- Inserted a thread reduction memory check∗

`pno_pdens`

- Restrict pairs to $i_0 j_l$∗†
- Skipping batch if i_0 or $i_0 j_l$ pairs aren't within that batch- saves on I/O with the integrals∗

`pno_3idxget_pbc_mega`•

- Extracts 3 index integrals only belonging from the reference cell, and constructs integrals in other cells through reordering reference cell quantities∗

`pno_pnotranssym`•

- Repopulates PNO information of the full supercell using translational invariance from the reference cell; Copies and reorganises PNO integrals (`xijab`), Fock matrix elements (`fdiag`); pre-PNO -> PNO coefficients are copied to correct positions in the full list∗†
- Added option to regenerate to $i_0 j_l$ or $i_l j_0$ list- changed list lengths and allocations∗

`pno_osv2pno`

- Restricts evaluation of OSV->PNO coefficients to only those beginning in the reference cell∗†
- Copies and reorders OSV->PNO transformation matrices of reference cell to all other cells∗†

`pno_sett2res`

- Changed repointing structure such that the translation is evaluated immediately, and the reference cell `nkmo`, `novlij`, `iadrovlp` quantities are read in- this ensures the number of pairs are kept consistent∗†

`pno_calcener`

- Adding $i_0 j_l$ pair energies together to get a reference cell correlation energy∗†

C| Unit cell geometries for Chapter 6

The geometries and cell parameters are reported in the Turbomole format

```
$cell
  a b c  α   β   γ
```

where all lengths are in Bohr and angles in degree. Primitive unit cells were chosen in all cases.

1D C_2H_2

```
$cell
  4.638536401 0 0 0 0 0
$coord
  0.0000000000          0.0000000000          0.0000000000      H
  0.0000000000          2.0541321496          0.0000000000      C
  2.2449855889          3.3051307540          0.0000000000      C
  2.2449855889          5.3592629040          0.0000000000      H
```

1D C_2HF

```
$cell
  4.638536401
$coord
  0.0000000000           0.0000000000        0.0000000000       F
  0.0000000000           2.5511300846        0.0000000000       C
  2.2449855889           3.8021286890        0.0000000000       C
  2.2449855889           5.8562608390        0.0000000000       H
```

2D BN

```
$cell
  4.7259 4.7259 60.
$coord
  0.000000000000       0.000000000000       0.000000000000       B
  2.363000000000       1.364000000000       0.000000000000       N
```

3D LiH

```
$cell
  5.4572 5.4572 5.4572 60. 60. 60.
$coord frac
  0.000000000000       0.000000000000       0.000000000000       Li
 -0.500000000000       0.500000000000       0.500000000000       H
```

3D MgO

```
$cell
  5.798338236 5.798338236 5.798338236 60. 60. 60.
$coord frac
  0.000000000000      0.000000000000      0.000000000000      Mg
  0.500000000000      0.500000000000      0.500000000000      O
```

3D Si

```
$cell
 7.25670856068 7.25670856068 7.25670856068 60.0 60.0 60.0
$coord
    0.000000000000     0.000000000000     0.000000000000  Si
    3.628354280339     2.094831320469     1.481269432146  Si
```

3D Ice I_h

```
$cell
  8.2014107904 8.2014107904 13.356028993 90 90 60
$coord
  0.0000000000      9.4430366559      3.9974358616    H
  4.0880486072     11.8901120652     10.6754508388    H
  0.0000000000      7.6732529979      6.4581971589    H
  4.0880486072     13.6598957231     13.1362121361    H
  5.6005868902     10.9208120851      6.9062369367    H
  2.5755103242     10.9208120851      6.9062369367    H
  6.6635589314     10.4123366360      0.2282219595    H
  1.5125382830     10.4123366360      0.2282219595    H
  0.0000000000      9.4624928981      5.8791642495    O
  4.0880486072     11.8706558230     12.5571792267    O
  4.0880486072     11.8396325860      7.5434853160    O
  0.0000000000      9.4935161350      0.8654703388    O
```

D| Basis sets used for Chapter 7

'TZ' Basis Set

```
c
*
8  s
8236.0000000       0.53100000000E-03
1235.0000000       0.41080000000E-02
280.80000000       0.21087000000E-01
79.270000000       0.81853000000E-01
25.590000000       0.23481700000
8.9970000000       0.43440100000
3.3190000000       0.34612900000
0.36430000000      -0.89830000000E-02
8  s
8236.0000000      -0.11300000000E-03
1235.0000000      -0.87800000000E-03
280.80000000      -0.45400000000E-02
79.270000000      -0.18133000000E-01
25.590000000      -0.55760000000E-01
8.9970000000      -0.12689500000
3.3190000000      -0.17035200000
0.36430000000       0.59868400000
1  s
0.90590000000        1.0000000000
1  s
0.12850000000        1.0000000000
3  p
18.710000000       0.14031000000E-01
4.1330000000       0.86866000000E-01
1.2000000000       0.29021600000
1  p
0.38270000000        1.0000000000
1  p
0.12090000000        1.0000000000
1  d
1.0970000000        1.0000000000
1  d
0.31800000000        1.0000000000
1  f
0.76100000000        1.0000000000
*

o
*
8  s
15330.000000       0.50800000000E-03
2299.0000000       0.39290000000E-02
522.40000000       0.20243000000E-01
147.30000000       0.79181000000E-01
```

```
47.550000000       0.23068700000
16.760000000       0.43311800000
6.2070000000       0.35026000000
0.68820000000     -0.81540000000E-02
8   s
15330.000000     -0.11500000000E-03
2299.0000000     -0.89500000000E-03
522.40000000     -0.46360000000E-02
147.30000000     -0.18724000000E-01
47.550000000     -0.58463000000E-01
16.760000000     -0.13646300000
6.2070000000     -0.17574000000
0.68820000000      0.60341800000
1   s
1.7520000000       1.0000000000
1   s
0.23840000000      1.0000000000
3   p
34.460000000       0.15928000000E-01
7.7490000000       0.99740000000E-01
2.2800000000       0.31049200000
1   p
0.71560000000      1.0000000000
1   p
0.21400000000      1.0000000000
1   d
2.3140000000       1.0000000000
1   d
0.64500000000      1.0000000000
1   f
1.4280000000       1.0000000000
*

mg
*
7   s
31438.3495550              0.00060912311326
4715.51533540             0.00470661964650
1073.16292470             0.02413582065700
303.572387680            0.09362895983400
98.6262510420           0.26646742093000
34.9438084170           0.47890929917000
12.8597851990           0.33698490286000
3   s
64.8769130040           0.01918088930700
19.7255207770           0.09091370439200
2.89518043390          -0.39563756125000
2   s
1.19604547100           1.68276033730000
0.54329451156           0.52141091954000
1   s
0.83471188300           1.00000000000000
1   s
0.14506887000           1.00000000000000
1   s
```

```
2.587700E+01     1.000000E+00
1  s
3.040200E+00     1.000000E+00
5   p
179.871896120            0.00537995490180
42.1200693760           0.03931801409800
13.1205030320           0.15740129476000
4.62575036090          0.35919094128000
1.66952110160          0.45533379310000
1   p
0.56631001000          1.00000000000000
1   p
0.18813966000          1.00000000000000
1   p
1.705300E+01     1.000000E+00
1   p
3.954100E+00     1.000000E+00
1  d
0.126000000          1.00000000
1  d
0.294000000          1.00000000
1  d
4.550700E+00     1.000000E+00
1  d
1.105100E+00     1.000000E+00
1  f
0.252000000          1.00000000
1  f
1.298500E+00     1.000000E+00
*
```

‘DZ’ Basis Set

```
c
*
8  s
6665.0000000      0.69200000000E-03
1000.0000000      0.53290000000E-02
228.00000000      0.27077000000E-01
64.710000000      0.10171800000
21.060000000      0.27474000000
7.4950000000      0.44856400000
2.7970000000      0.28507400000
0.52150000000      0.15204000000E-01
8  s
6665.0000000     -0.14600000000E-03
1000.0000000     -0.11540000000E-02
228.00000000     -0.57250000000E-02
64.710000000     -0.23312000000E-01
21.060000000     -0.63955000000E-01
7.4950000000     -0.14998100000
2.7970000000     -0.12726200000
0.52150000000      0.54452900000
```

```
1  s
0.15960000000        1.0000000000
3  p
9.4390000000       0.38109000000E-01
2.0020000000       0.20948000000
0.54560000000       0.50855700000
1  p
0.15170000000        1.0000000000
1  d
0.55000000000        1.0000000000
*

o
*
8  s
11720.000000       0.71000000000E-03
1759.0000000       0.54700000000E-02
400.80000000       0.27837000000E-01
113.70000000       0.10480000000
37.030000000       0.28306200000
13.270000000       0.44871900000
5.0250000000       0.27095200000
1.0130000000       0.15458000000E-01
8  s
11720.000000      -0.16000000000E-03
1759.0000000      -0.12630000000E-02
400.80000000      -0.62670000000E-02
113.70000000      -0.25716000000E-01
37.030000000      -0.70924000000E-01
13.270000000      -0.16541100000
5.0250000000      -0.11695500000
1.0130000000       0.55736800000
1  s
0.30230000000        1.0000000000
3  p
17.700000000       0.43018000000E-01
3.8540000000       0.22891300000
1.0460000000       0.50872800000
1  p
0.27530000000        1.0000000000
1  d
1.1850000000        1.0000000000
*

mg
*
5  s
4953.8339196      -0.57778967498E-02
745.18044154      -0.43124761082E-01
169.21604972      -0.19268216987
47.300672019      -0.48641439116
14.461336973      -0.42550894077
3  s
24.768174789       0.87956969984E-01
2.4940945349      -0.55165058128
```

```
0.87807584530      -0.53443294833
1   s
0.34506887000        1.0000000000
1   s
0.15005399000        1.0000000000
1   s
2.911600E+00      1.000000E+00
5   p
98.053010494      -0.14480564601E-01
22.586932277      -0.95495750787E-01
6.8391509842      -0.30787672651
2.2332843818      -0.49936292886
0.71606599390      -0.31503476213
1   p
0.24692323000        1.0000000000
1   p
6.031000E+00      1.000000E+00
1    d
1.932000D-01            1.0000000
1   d
1.370200E+00      1.000000E+00
*
```

References

(1) Zhu, A.; Tew, D. P. *The Journal of Physical Chemistry A* **2024**, *128*, 8570–8579.

(2) Nejad, A.; Zhu, A.; Sorathia, K.; Tew, D. P. *The Journal of Chemical Physics* **2025**, *163*, 214107.

(3) Zhu, A.; Nejad, A.; Komonvasee, P.; Sorathia, K.; Tew, D. P. *The Journal of Chemical Physics* **2025**, *163*, 214108.

(4) Majewski, J. A.; Birner, S.; Trellakis, A.; Sabathil, M.; Vogl, P. *phys. stat. sol. (c)* **2004**, *1*, 2003–2027.

(5) Efros, A. L.; Rosen, M. *Annu. Rev. Mater. Sci.* **2000**, *30*, 475–521.

(6) Pickett, W. E. *Rev. Mod. Phys.* **1989**, *61*, 433–512.

(7) Jain, A.; Shin, Y.; Persson, K. A. *Nat Rev Mater* **2016**, *1*, 15004.

(8) Kordyuk, A. A. *Low Temperature Physics* **2012**, *38*, 888–899.

(9) Liu, X.; Zhao, L.; He, S.; He, J.; Liu, D.; Mou, D.; Shen, B.; Hu, Y.; Huang, J.; Zhou, X. J. *J. Phys.: Condens. Matter* **2015**, *27*, 183201.

(10) Hautier, G.; Jain, A.; Ong, S. P. *J Mater Sci* **2012**, *47*, 7317–7340.

(11) Zhang, H. *Electron. Struct.* **2021**, *3*, 033001.

(12) Haas, P.; Tran, F.; Blaha, P. *Phys. Rev. B* **2009**, *79*, 085104.

(13) Schimka, L.; Gaudoin, R.; Klimeš, J.; Marsman, M.; Kresse, G. *Phys. Rev. B* **2013**, *87*, 214102.

(14) Tran, F.; Stelzl, J.; Blaha, P. *The Journal of Chemical Physics* **2016**, *144*, 204120.

(15) Goldzak, T.; Wang, X.; Ye, H.-Z.; Berkelbach, T. C. *J. Chem. Phys.* **2022**, *157*, 174112.

(16) Milman, V.; Winkler, B.; White, J. A.; Pickard, C. J.; Payne, M. C.; Akhmatskaya, E. V.; Nobes, R. H. *Int. J. Quant. Chem.* **2000**, *77*, 895–910.

(17) Bartel, C. J. *J Mater Sci* **2022**, *57*, 10475–10498.

(18) Boese, A. D.; Sauer, J. *Phys. Chem. Chem. Phys.* **2013**, *15*, Publisher: The Royal Society of Chemistry, 16481–16493.

(19) Alessio, M.; Usvyat, D.; Sauer, J. *J. Chem. Theory Comput.* **2019**, *15*, 1329–1344.

(20) Shi, B. X.; Zen, A.; Kapil, V.; Nagy, P. R.; Grüneis, A.; Michaelides, A. *J. Am. Chem. Soc.* **2023**, *145*, 25372–25381.

(21) Ye, H.-Z.; Berkelbach, T. C. *Faraday Discuss.* **2024**, *254*, Publisher: The Royal Society of Chemistry, 628–640.

(22) Tsatsoulis, T.; Hummel, F.; Usvyat, D.; Schütz, M.; Booth, G. H.; Binnie, S. S.; Gillan, M. J.; Alfè, D.; Michaelides, A.; Grüneis, A. *J. Chem. Phys.* **2017**, *146*, 204108.

(23) Ugliengo, P.; Damin, A. *Chemical Physics Letters* **2002**, *366*, 683–690.

(24) Pacchionia, G.; Neyman, K. M.; Rosch, N. *Journal of Electron Spectroscopy and Related Phenomena* **1994**, *69*, 13–21.

(25) Schäfer, Tobias and Gallo, Alejandro and Irmler, Andreas and Hummel, Felix and Grüneis, Andreas *The Journal of Chemical Physics* **2021**, *155*, 244103.

(26) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, Publisher: American Physical Society, B864–B871.

(27) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, Publisher: American Physical Society, A1133–A1138.

(28) Maurer, R. J.; Freysoldt, C.; Reilly, A. M.; Brandenburg, J. G.; Hofmann, O. T.; Björkman, T.; Lebègue, S.; Tkatchenko, A. *Annu. Rev. Mater. Res.* **2019**, *49*, 1–30.

(29) Hafner, J.; Wolverton, C.; Ceder, G. *MRS Bull.* **2006**, *31*, 659–668.

(30) Perdew, J. P. In *AIP Conference Proceedings*; ISSN: 0094243X, AIP: Antwerp (Belgium), 2001; Vol. 577, pp 1–20.

(31) Fonseca Guerra, C.; Snijders, J. G.; Te Velde, G.; Baerends, E. J. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **1998**, *99*, 391–403.

(32) Mohr, S.; Ratcliff, L. E.; Genovese, L.; Caliste, D.; Boulanger, P.; Goedecker, S.; Deutsch, T. *Phys. Chem. Chem. Phys.* **2015**, *17*, 31360–31370.

(33) Nakata, A.; Baker, J. S.; Mujahed, S. Y.; Poulton, J. T. L.; Arapan, S.; Lin, J.; Raza, Z.; Yadav, S.; Truflandier, L.; Miyazaki, T.; Bowler, D. R. *The Journal of Chemical Physics* **2020**, *152*, 164112.

(34) Grimme, S. *J Comput Chem* **2006**, *27*, 1787–1799.

(35) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(36) Grimme, S. *WIREs Comput Mol Sci* **2011**, *1*, 211–228.

(37) Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 073005.

(38) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246401.

(39) Perdew, J. P. *Int. J. Quantum Chem.* **2009**, *28*, 497–523.

(40) Heaton, R. A.; Harrison, J. G.; Lin, C. C. *Phys. Rev. B* **1983**, *28*, 5992–6007.

(41) Muscat, J.; Wander, A.; Harrison, N. *Chemical Physics Letters* **2001**, *342*, 397–401.

(42) Garza, A. J.; Scuseria, G. E. *J. Phys. Chem. Lett.* **2016**, *7*, 4165–4170.

(43) Pisani, C.; Maschio, L.; Casassa, S.; Halo, M.; Schütz, M.; Usvyat, D. *J. Comput. Chem.* **2008**, *29*, 2113–2124.

(44) Zhang, I. Y.; Grüneis, A. *Front. Mater.* **2019**, *6*, 123.

(45) Onida, G.; Reining, L.; Rubio, A. *Rev. Mod. Phys.* **2002**, *74*, 601–659.

(46) Reining, L. *WIREs Comput Mol Sci* **2018**, *8*, e1344.

(47) Georges, A. In *AIP Conference Proceedings*; ISSN: 0094243X, AIP: Salerno (Italy), 2004; Vol. 715, pp 3–74.

(48) Paul, A.; Birol, T. *Annu. Rev. Mater. Res.* **2019**, *49*, 31–52.

(49) Foulkes, W. M. C.; Mitas, L.; Needs, R. J.; Rajagopal, G. *Rev. Mod. Phys.* **2001**, *73*, 33–83.

(50) Shi, H.; Zhang, S. *The Journal of Chemical Physics* **2021**, *154*, 024107.

(51) Pisani, C.; Schütz, M.; Casassa, S.; Usvyat, D.; Maschio, L.; Lorenz, M.; Erba, A. *Phys. Chem. Chem. Phys.* **2012**, *14*, Publisher: The Royal Society of Chemistry, 7615–7628.

(52) Maschio, L.; Usvyat, D.; Manby, F. R.; Casassa, S.; Pisani, C.; Schütz, M. *Phys. Rev. B* **2007**, *76*, Publisher: American Physical Society, 075101.

(53) Usvyat, D.; Maschio, L.; Manby, F. R.; Casassa, S.; Schütz, M.; Pisani, C. *Phys. Rev. B* **2007**, *76*, Publisher: American Physical Society, 075102.

(54) McClain, J.; Sun, Q.; Chan, G. K.-L.; Berkelbach, T. C. *J. Chem. Theory Comput.* **2017**, *13*, 1209–1218.

(55) Gruber, T.; Liao, K.; Tsatsoulis, T.; Hummel, F.; Grüneis, A. *Phys. Rev. X* **2018**, *8*, Publisher: American Physical Society, 021043.

(56) Hirata, S.; Shimazaki, T. *Phys. Rev. B* **2009**, *80*, Publisher: American Physical Society, 085118.

(57) Hirata, S.; Podeszwa, R.; Tobita, M.; Bartlett, R. J. *J. Chem. Phys.* **2004**, *120*, 2581–2592.

(58) Haritan, I.; Wang, X.; Goldzak, T. An Efficient scaled opposite-spin MP2 method for periodic systems, 2025.

(59) Ayala, P. Y.; Kudin, K. N.; Scuseria, G. E. *J. Chem. Phys.* **2001**, *115*, 9698–9707.

(60) Ye, H.-Z.; Berkelbach, T. C. *J. Chem. Theory Comput.* **2024**, *20*, 8948–8959.

(61) Katouda, M.; Nagase, S. *J. Chem. Phys.* **2010**, *133*, 184103.

(62) Booth, G. H.; Grüneis, A.; Kresse, G.; Alavi, A. *Nature* **2013**, *493*, ISBN: 1476-4687, 365–370.

(63) Schäfer, T.; Ramberger, B.; Kresse, G. *J. Chem. Phys.* **2017**, *146*, 104101.

(64) Grüneis, A.; Booth, G. H.; Marsman, M.; Spencer, J.; Alavi, A.; Kresse, G. *J. Chem. Theory Comput.* **2011**, *7*, 2780–2785.

(65) Grüneis, A. *J. Chem. Phys.* **2015**, *143*, 102817.

(66) Stoll, H. *The Journal of Chemical Physics* **1992**, *97*, 8449–8454.

(67) Müller, C.; Usvyat, D.; Stoll, H. *Phys. Rev. B* **2011**, *83*, 245136.

(68) Li, W.; Wang, Y.; Ni, Z.; Li, S. *Acc. Chem. Res.* **2023**, *56*, 3462–3474.

(69) Wang, Y.; Ni, Z.; Li, W.; Li, S. *J. Chem. Theory Comput.* **2019**, *15*, 2933–2943.

(70) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46–53.

(71) Stoll, H. *Chemical Physics Letters* **1992**, *191*, 548–552.

(72) Manby, F. R.; Alf?, D.; Gillan, M. J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5178.

(73) Cui, Z.-H.; Zhu, T.; Chan, G. K.-L. *Journal of Chemical Theory and Computation* **2020**, *16*, 119–129.

(74) Nusspickel, M.; Booth, G. H. *Phys. Rev. X* **2022**, *12*, 011046.

(75) Tuma, C.; Sauer, J. *Chemical Physics Letters* **2004**, *387*, 388–394.

(76) Tuma, C.; Sauer, J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3955–3965.

(77) Matsui, M.; Sakaki, S. *J. Phys. Chem. C* **2017**, *121*, 20242–20253.

(78) Dovesi, R. et al. *The Journal of Chemical Physics* **2020**, *152*, 204111.

(79) Neese, F.; Hansen, A.; Liakos, D. G. *J. Chem. Phys.* **2009**, *131*, Publisher: American Institute of Physics, 064103.

(80) Neese, F.; Wennmohs, F.; Hansen, A. *The Journal of Chemical Physics* **2009**, *130*, 114108.

(81) Balasubramani, S. G. et al. *J. Chem. Phys.* **2020**, *152*, 184107.

(82) Franzke, Y. J. et al. *J. Chem. Theory Comput.* **2023**, *19*, 6859–6890.

(83) Burow, A. M.; Sierka, M. *J. Chem. Theory Comput.* **2011**, *7*, ISBN: 1549-9618 Publisher: American Chemical Society, 3097–3104.

(84) Burow, A. M.; Sierka, M.; Mohamed, F. *J. Chem. Phys.* **2009**, *131*, 214101.

(85) Łazarski, R.; Burow, A. M.; Sierka, M. *J. Chem. Theory Comput.* **2015**, *11*, ISBN: 1549-9618 Publisher: American Chemical Society Type: doi: 10.1021/acs.jctc.5b00252, 3029–3041.

(86) Łazarski, R.; Burow, A. M.; Grajciar, L.; Sierka, M. *J. Comput. Chem.* **2016**, *37*, 2518–2526.

(87) Schmitz, G.; Helmich, B.; and, C. H. *Molecular Physics* **2013**, *111*, Publisher: Taylor & Francis, 2463–2476.

(88) Schmitz, G.; Hättig, C.; Tew, D. P. *Phys. Chem. Chem. Phys.* **2014**, *16*, Publisher: The Royal Society of Chemistry, 22167–22178.

(89) Schmitz, G.; Hättig, C. *J. Chem. Phys.* **2016**, *145*, 234107.

(90) Tew, D. P. *J. Chem. Theory Comput.* **2019**, *15*, 6597–6606.

(91) Ewald, P. P. *Annalen der Physik* **1921**, *369*, 253–287.

(92) Shepherd, J. J.; Grüneis, A.; Booth, G. H.; Kresse, G.; Alavi, A. *Phys. Rev. B* **2012**, *86*, 035111.

(93) Peintinger, M. F.; Oliveira, D. V.; Bredow, T. *J. Comput. Chem.* **2013**, *34*, 451–459.

(94) Laun, J.; Bredow, T. *J. Comput. Chem.* **2022**, *43*, 839–846.

(95) Ye, H.-Z.; Berkelbach, T. C. *J. Chem. Theory Comput.* **2022**, *18*, 1595–1606.

(96) Ashcroft, N. W.; Mermin, N. D., *Solid State Physics*; Harcourt College Publishers: Toronto, 1976.

(97) Irmler, A.; Burow, A. M.; Pauly, F. *J. Chem. Theory Comput.* **2018**, *14*, 4567–4580.

(98) Müller, C.; Sharma, M.; Sierka, M. *J. Comput. Chem.* **2020**, *41*, 2573–2582.

(99) Wannier, G. H. *Phys. Rev.* **1937**, *52*, Publisher: American Physical Society, 191–197.

(100) Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, Publisher: American Physical Society, 296–299.

(101) Foster, J. M.; Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, Publisher: American Physical Society, 300–302.

(102) Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*, 4916–4926.

(103) Marzari, N.; Vanderbilt, D. *Phys. Rev. B* **1997**, *56*, Publisher: American Physical Society, 12847–12865.

(104) Pizzi, G. et al. *J. Phys.: Condens. Matter* **2020**, *32*, Publisher: IOP Publishing, 165902.

(105) Marzari, N.; Mostofi, A. A.; Yates, J. R.; Souza, I.; Vanderbilt, D. *Rev. Mod. Phys.* **2012**, *84*, Publisher: American Physical Society, 1419–1475.

(106) Jónsson, E. Ö.; Lehtola, S.; Puska, M.; Jónsson, H. *J. Chem. Theory Comput.* **2017**, *13*, 460–474.

(107) Clement, M. C.; Wang, X.; Valeev, E. F. *J. Chem. Theory Comput.* **2021**, *17*, ISBN: 1549-9618 Publisher: American Chemical Society Type: doi: 10.1021/acs.jctc.1c00238, 7406–7415.

(108) Schreder, L.; Luber, S. *J. Chem. Phys.* **2024**, *160*, 214117.

(109) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47*, Publisher: American Physical Society, 558–561.

(110) Giannozzi, P. et al. *Journal of Physics: Condensed Matter* **2017**, *29*, 465901.

(111) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. I. J.; Refson, K.; Payne, M. C. *Zeitschrift für Kristallographie - Crystalline Materials* **2005**, *220*, 567–570.

(112) Gonze, X. et al. *Computer Physics Communications* **2016**, *205*, 106–131.

(113) Erba, A.; Desmarais, J. K.; Casassa, S.; Civalleri, B.; Donà, L.; Bush, I. J.; Searle, B.; Maschio, L.; Edith-Daga, L.; Cossard, A.; Ribaldone, C.; Ascrizzi, E.; Marana, N. L.; Flament, J.-P.; Kirtman, B. *Journal of Chemical Theory and Computation* **2023**, *19*, 6891–6932.

(114) Pisani, C.; Dovesi, R. *International Journal of Quantum Chemistry* **1980**, *17*, 501–516.

(115) Dovesi, R.; Pisani, C.; Roetti, C. *International Journal of Quantum Chemistry* **1980**, *17*, 517–529.

(116) Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N. S.; Bogdanov, N. A.; Booth, G. H.; Chen, J.; Cui, Z.-H., et al. *The Journal of chemical physics* **2020**, *153*, Publisher: AIP Publishing.

(117) Knizia, G. *J. Chem. Theory Comput.* **2013**, *9*, Publisher: ACS Publications, 4834–4843.

(118) Edmiston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, Publisher: American Physical Society, 457–464.

(119) Lehtola, S.; Jónsson, H. *Journal of Chemical Theory and Computation* **2014**, *10*, 642–649.

(120) Cioslowski, J. *Journal of Mathematical Chemistry* **1991**, *8*, ISBN: 1572-8897, 169–178.

(121) Alcoba, D. R.; Lain, L.; Torre, A.; Bochicchio, R. C. *Journal of Computational Chemistry* **2006**, *27*, 596–608.

(122) Werner, H.-J.; Knowles, P. J.; Manby, F. R.; Black, J. A.; Doll, K.; Heßelmann, A.; Kats, D.; Köhn, A.; Korona, T.; Kreplin, D. A., et al. *The Journal of chemical physics* **2020**, *152*, Publisher: AIP Publishing.

(123) Saue, T.; Bast, R.; Gomes, A. S. P.; Jensen, H. J. A.; Visscher, L.; Aucar, I. A.; Di Remigio, R.; Dyall, K. G.; Eliav, E.; Fasshauer, E., et al. *The Journal of chemical physics* **2020**, *152*, Publisher: AIP Publishing.

(124) Janowski, T. *Journal of Chemical Theory and Computation* **2014**, *10*, Publisher: ACS Publications, 3085–3091.

(125) Monkhorst, H. J.; Pack, J. D. *Phys. Rev. B* **1976**, *13*, Publisher: American Physical Society, 5188–5192.

(126) Lehtola, S.; Jónsson, H. *Journal of Chemical Theory and Computation* **2013**, *9*, 5365–5372.

(127) Cardoso, J.-F.; Souloumiac, A. *SIAM Journal on Matrix Analysis and Applications* **1996**, *17*, 161–164.

(128) Horn, R. A.; Johnson, C. R., *Matrix analysis*; Cambridge university press: 2012.

(129) Gower, J. C.; Dijksterhuis, G. B., *Procrustes problems*; OUP Oxford: 2004; Vol. 30.

(130) Abrudan, T. E.; Eriksson, J.; Koivunen, V. *IEEE Transactions on Signal Processing* **2008**, *56*, 1134–1147.

(131) Abrudan, T.; Eriksson, J.; Koivunen, V. *Signal Processing* **2009**, *89*, 1704–1714.

(132) Hyvärinen, A.; Karhunen, J.; Oja, E., *Independent Component Analysis*; Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control; Wiley: 2001.

(133) Nocedal, J.; Wright, S. J., *Numerical Optimization*, 2e; Springer: New York, NY, USA, 2006.

(134) Polak, E., *Optimization: Algorithms and Consistent Approximations*; Springer-Verlag: Berlin, Heidelberg, 1997.

(135) Dunning, T. *J. Chem. Phys.* **1989**, *90*, 1007.

(136) Woon, D. E.; Dunning Thom H., J. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(137) Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning, T. H.; Wilson, A. K. *Theor. Chem. Acc.* **2011**, *128*, ISBN: 1432–2234, 69–82.

(138) Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, Publisher: The Royal Society of Chemistry, 1057–1065.

(139) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *Journal of Cheminformatics* **2012**, *4*, ISBN: 1758-2946, 17.

(140) Ma, Q.; Werner, H.-J. *WIREs Computational Molecular Science* **2018**, *8*, e1371.

(141) Sauer, J. *Accounts of Chemical Research* **2019**, *52*, 3502–3510.

(142) Mitra, A.; Hermes, M. R.; Cho, M.; Agarawal, V.; Gagliardi, L. *The Journal of Physical Chemistry Letters* **2022**, *13*, Publisher: American Chemical Society Type: doi: 10.1021/acs.jpclett.2c01915, 7483–7489.

(143) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, Publisher: American Physical Society, 3865–3868.

(144) Neyman, K. M.; Rösch, N. *Chemical Physics* **1992**, *168*, 267–280.

(145) Huron, B.; Malrieu, J. P.; Rancurel, P. *The Journal of Chemical Physics* **1973**, *58*, 5745–5759.

(146) Evangelista, F. A. *The Journal of Chemical Physics* **2018**, *149*, 030901.

(147) Köhn, A.; Hanauer, M.; Mück, L. A.; Jagau, T.-C.; Gauss, J. *WIREs Comput Mol Sci* **2013**, *3*, 176–197.

(148) Ziółkowski, M.; Jansík, B.; Kjærgaard, T.; Jørgensen, P. *The Journal of Chemical Physics* **2010**, *133*, 014107.

(149) Kristensen, K.; Høyvik, I.-M.; Jansik, B.; Jørgensen, P.; Kjærgaard, T.; Reine, S.; Jakowski, J. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15706.

(150) Kjærgaard, T. *The Journal of Chemical Physics* **2017**, *146*, 044103.

(151) Rebolini, E.; Baardsen, G.; Hansen, A. S.; Leikanger, K. R.; Pedersen, T. B. *J. Chem. Theory Comput.* **2018**, *14*, 2427–2438.

(152) Li, S.; Ma, J.; Jiang, Y. *J Comput Chem* **2002**, *23*, 237–244.

(153) Li, W.; Li, S. *The Journal of Chemical Physics* **2004**, *121*, 6649–6657.

(154) Li, S.; Shen, J.; Li, W.; Jiang, Y. *The Journal of Chemical Physics* **2006**, *125*, 074109.

(155) Li, W.; Piecuch, P.; Gour, J. R.; Li, S. *The Journal of Chemical Physics* **2009**, *131*, 114109.

(156) Wang, Y.; Ni, Z.; Neese, F.; Li, W.; Guo, Y.; Li, S. *J. Chem. Theory Comput.* **2022**, *18*, 6510–6521.

(157) Rolik, Z.; Kállay, M. *J. Chem. Phys.* **2011**, *135*, 104111.

(158) Rolik, Z.; Szegedy, L.; Ladjánszki, I.; Ladóczki, B.; Kállay, M. *J. Chem. Phys.* **2013**, *139*, 094105.

(159) Nagy, P. R.; Kállay, M. *J. Chem. Phys.* **2017**, *146*, 214106.

(160) Pulay, P. *Chemical Physics Letters* **1983**, *100*, 151–154.

(161) Sæbø, S.; Pulay, P. *Chemical Physics Letters* **1985**, *113*, 13–18.

(162) Saebo, S.; Pulay, P. *Annual Review of Physical Chemistry* **1993**, *44*, 213–236.

(163) Meyer, W. *Int. J. Quantum Chem.* **2009**, *5*, 341–348.

(164) Meyer, W. *The Journal of Chemical Physics* **1973**, *58*, 1017–1035.

(165) Yang, J.; Kurashige, Y.; Manby, F. R.; Chan, G. K. L. *J. Chem. Phys.* **2011**, *134*, 044123.

(166) Schütz, M.; Hetzer, G.; Werner, H.-J. *J. Chem. Phys.* **1999**, *111*, 5691–5705.

(167) Pinski, P.; Riplinger, C.; Valeev, E. F.; Neese, F. *J. Chem. Phys.* **2015**, *143*, 034108.

(168) Hättig, C.; Tew, D. P.; Helmich, B. *J. Chem. Phys.* **2012**, *136*, 204105.

(169) Tew, D. P.; Hättig, C. *International Journal of Quantum Chemistry* **2013**, *113*, 224–229.

(170) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. *J. Chem. Phys.* **2016**, *144*, 024109.

(171) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. *J. Chem. Phys.* **2018**, *148*, 011101.

(172) Schwilk, M.; Ma, Q.; Köppl, C.; Werner, H.-J. *J. Chem. Theory Comput.* **2017**, *13*, 3650–3675.

(173) Schmitz, G.; Hättig, C. *J. Chem. Theory Comput.* **2017**, *13*, 2623–2633.

(174) Liakos, D. G.; Guo, Y.; Neese, F. *The Journal of Physical Chemistry A* **2020**, *124*, 90–100.

(175) Ma, Q.; Werner, H.-J. *J. Chem. Theory Comput.* **2015**, *11*, 5291–5304.

(176) Tew, D. P. In *New Electron Correlation Methods and their Applications, and Use of Atomic Orbitals with Exponential Asymptotes*, Musial, M., Hoggan, P. E., Eds.; Advances in Quantum Chemistry, Vol. 83; Academic Press: 2021, pp 83–106.

(177) Kats, D.; Werner, H.-J. *J. Chem. Phys.* **2019**, *150*, 214107.

(178) Saitow, M.; Yanai, T. *J. Chem. Phys.* **2020**, *152*, 114111.

(179) Guo, Y.; Sivalingam, K.; Valeev, E. F.; Neese, F. *J. Chem. Phys.* **2016**, *144*, 094111.

(180) Helmich, B.; Hättig, C. *J. Chem. Phys.* **2011**, *135*, 214106.

(181) Frank, M. S.; Hättig, C. *J. Chem. Phys.* **2018**, *148*, 134102.

(182) Dutta, A. K.; Neese, F.; Izsák, R. *J. Chem. Phys.* **2016**, *145*, 034102.

(183) Peng, C.; Clement, M. C.; Valeev, E. F. *J. Chem. Theory Comput.* **2018**, *14*, 5597–5607.

(184) Usvyat, D.; Maschio, L.; Schütz, M. *J. Chem. Phys.* **2015**, *143*, 102805.

(185) Usvyat, D. *J. Chem. Phys.* **2013**, *139*, 194101.

(186) Usvyat, D.; Civalleri, B.; Maschio, L.; Dovesi, R.; Pisani, C.; Schütz, M. *J. Chem. Phys.* **2011**, *134*, 214105.

(187) Nagy, P. R.; Samu, G.; Kállay, M. *J. Chem. Theory Comput.* **2018**, *14*, 4193–4215.

(188) Müller, C.; Usvyat, D. *J. Chem. Theory Comput.* **2013**, *9*, 5590–5598.

(189) Usvyat, D.; Sadeghian, K.; Maschio, L.; Schütz, M. *Phys. Rev. B* **2012**, *86*, Publisher: American Physical Society, 045412.

(190) Mullan, T.; Maschio, L.; Saalfrank, P.; Usvyat, D. *J. Chem. Phys.* **2022**, *156*, 074109.

(191) Sorathia, K.; Tew, D. P. *J. Chem. Phys.* **2020**, *153*, 174112.

(192) Sorathia, K.; Frantzov, D.; Tew, D. P. *J. Chem. Theory Comput.* **2024**, *20*, 2740–2750.

(193) Almlöf, J. *Chemical Physics Letters* **1991**, *181*, 319–320.

(194) Häser, M.; Almlöf, J. *J. Chem. Phys.* **1992**, *96*, 489–494.

(195) Stuart, S. N. *J. Comput. Phys.* **1978**, *29*, 127–132.

(196) Harris, F. E. In *Theoretical Chemistry: Advances and Perspectives*, Eyring, H., Henderson, D., Eds.; Theoretical Chemistry, Vol. 1; Academic Press: New York, 1975, pp 147–218.

(197) Born, M.; von Kármán, T. *Phys. Z.* **1912**, *13*, 297–309.

(198) Born, M.; von Kármán, T. *Phys. Z.* **1913**, *14*, 15–19.

(199) Madelung, E. *Phys. Z.* **1918**, *19*, 524–533.

(200) Makov, G.; Payne, M. C. *Phys. Rev. B* **1995**, *51*, 4014–4022.

(201) Nijboer, B. R. A.; Wette, F. W. D. *Physica* **1958**, *24*, 422–431.

(202) Challacombe, M.; White, C.; Head-Gordon, M. *J. Chem. Phys.* **1997**, *107*, 10131–10140.

(203) Stolarczyk, L. Z.; Piela, L. *International Journal of Quantum Chemistry* **1982**, *22*, 911–927.

(204) Saunders, V. R.; Freyria-Fava, C.; Dovesi, R.; Salasco, L.; Roetti, C. *Mol. Phys.* **1992**, *77*, 629–665.

(205) Bintrim, S. J.; Berkelbach, T. C.; Ye, H.-Z. *J. Chem. Theory Comput.* **2022**, *18*, 5374–5381.

(206) Schütz, M.; Usvyat, D.; Lorenz, M.; Pisani, C.; Maschio, L.; Casassa, S.; Halo, M. In *Accurate Condensed-Phase Quantum Chemistry*, Manby, F. R., Ed.; CRC Press: Boca Raton, 2010, pp 29–55.

(207) Ye, H.-Z.; Berkelbach, T. C. *J. Chem. Phys.* **2021**, *154*, 131104.

(208) Wette, F. W. D.; Nijboer, B. R. A. *Physica* **1958**, *24*, 1105–1118.

(209) Kudin, K. N.; Scuseria, G. E. *Chem. Phys. Lett.* **1998**, *283*, 61–68.

(210) Kudin, K. N.; Scuseria, G. E. *Chem. Phys. Lett.* **1998**, *289*, 611–616.

(211) Kudin, K. N.; Scuseria, G. E. *J. Chem. Phys.* **2004**, *121*, 2886–2890.

(212) Burow, A. M., *Methoden zur Beschreibung von chemischen Strukturen beliebiger Dimensionalität mit der Dichtefunktionaltheorie unter periodischen Randbedingungen*; Universitätsbibliothek der Humboldt-Universität zu Berlin: Berlin, 2011.

(213) Herce, H. D.; Garcia, A. E.; Darden, T. *J. Chem. Phys.* **2007**, *126*, 124106.

(214) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143–152.

(215) Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331–340.

(216) Cassabois, G.; Valvin, P.; Gil, B. *Nat. Photonics* **2016**, *10*, 262–266.

(217) Liao, K.; Grüneis, A. *J. Chem. Phys.* **2016**, *145*, 141102.

(218) Christensen, C. H.; Nørskov, J. K. *The Journal of Chemical Physics* **2008**, *128*, 182503.

(219) Somorjai, G. In *Advances in Catalysis*; Elsevier: 1977; Vol. 26, pp 1–68.

(220) Chen, B. W. J.; Xu, L.; Mavrikakis, M. *Chem. Rev.* **2021**, *121*, 1007–1048.

(221) Ertl, G. *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 1219–1227.

(222) Schauermann, S.; Nilius, N.; Shaikhutdinov, S.; Freund, H.-J. *Acc. Chem. Res.* **2013**, *46*, 1673–1681.

(223) Vogt, C.; Weckhuysen, B. M. *Nat Rev Chem* **2022**, *6*, 89–111.

(224) Morris, R. E.; Wheatley, P. S. *Angew Chem Int Ed* **2008**, *47*, 4966–4981.

(225) Jahanmir, S.; Beltzer, M. *A S L E Transactions* **1986**, *29*, 423–430.

(226) Shigeishi, R.; King, D. A. *Surface Science* **1976**, *58*, 379–396.

(227) Tait, S. L.; Dohnálek, Z.; Campbell, C. T.; Kay, B. D. *The Journal of Chemical Physics* **2006**, *125*, 234308.

(228) Dohnálek, Z.; Kimmel, G. A.; Joyce, S. A.; Ayotte, P.; Smith, R. S.; Kay, B. D. *The Journal of Physical Chemistry B* **2001**, *105*, _eprint: https://doi.org/10.1021/jp003174b, 3747–3751.

(229) Wichtendahl, R.; Rodriguez-Rodrigo, M.; Härtel, U.; Kuhlenbeck, H.; Freund, H.-J. *physica status solidi (a)* **1999**, *173*, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291521-396X%28199905%29173%3A1%3C93%3A%3AAID-PSSA93%3E3.0.CO%3B2-4, 93–100.

(230) Schmid, M.; Parkinson, G. S.; Diebold, U. *ACS Phys. Chem Au* **2023**, *3*, 44–62.

(231) Herschend, B.; Baudin, M.; Hermansson, K. *J. Phys. Chem. B* **2006**, *110*, 5473–5479.

(232) Vilela Oliveira, D.; Laun, J.; Peintinger, M. F.; Bredow, T. *J Comput Chem* **2019**, *40*, 2364–2376.

(233) Audibert, P.; Sidoumou, M.; Suzanne, J. *Surface Science* **1992**, *273*, L467–L471.

(234) Panella, V.; Suzanne, J.; Hoang, P. N. M.; Girardet, C. *J. Phys. I France* **1994**, *4*, 905–920.

(235) Shi, B. X.; Rosen, A. S.; Schäfer, T.; Grüneis, A.; Kapil, V.; Zen, A.; Michaelides, A. An accurate and efficient framework for modelling the surface chemistry of ionic materials, _eprint: 2412.17204, 2025.

(236) Minot, C.; Hove, M. A. V.; Biberian, J.-P. *Surface Science* **1996**, *346*, 283–293.

(237) Guo, Y.-n.; Lu, X.; Zhang, H.-p.; Weng, J.; Watari, F.; Leng, Y. *J. Phys. Chem. C* **2011**, *115*, 18572–18581.

(238) Nolan, M. *J. Phys. Chem. C* **2009**, *113*, 2425–2432.

(239) Lousada, C. M.; Korzhavyi, P. A. *Phys. Chem. Chem. Phys.* **2015**, *17*, 1667–1679.

(240) Camarillo-Cisneros, J.; Liu, W.; Tkatchenko, A. *Phys. Rev. Lett.* **2015**, *115*, 086101.

(241) Starr, D. E.; Campbell, C. T. *J. Am. Chem. Soc.* **2008**, *130*, 7321–7327.

(242) Árnadóttir, L.; Stuve, E. M.; Jónsson, H. *Surface Science* **2010**, *604*, 1978–1986.

(243) Wandelt, K. *Surface Science* **1991**, *251-252*, 387–395.

(244) Ovcharenko, R.; Voloshina, E.; Sauer, J. *Phys. Chem. Chem. Phys.* **2016**, *18*, 25560–25568.