Wannier Functions and Their Role in Improving Density Functional Approximations

by

Aaron Mahler

Department of Physics
Duke University

Date: ______________________

Approved: ______________________

__________________________
Weitao Yang, Supervisor

__________________________
Harold Baranger

__________________________
David Beratan

__________________________
Sara Haravifard

__________________________
Thomas Barthel

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Physics in the Graduate School of Duke University

2021
ABSTRACT

Wannier Functions and Their Role in Improving Density Functional Approximations

by

Aaron Mahler

Department of Physics
Duke University

Date: _________________________

Approved: _____________________

__________________________
Weitao Yang, Supervisor

__________________________
Harold Baranger

__________________________
David Beratan

__________________________
Sara Haravifard

__________________________
Thomas Barthel

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Physics in the Graduate School of Duke University

2021
Abstract

Density functional theory has proven to be an invaluable tool for modeling matter and chemistry. This can be seen from the fact that density functional theory papers are far and away the most cited theory from the physical sciences. While density functional theory excels at predicting total energies and equilibrium geometries, standard approximate functionals can be inadequate for determining some properties such as dissociation energies, reaction barriers, and band gaps. These deficiencies can be traced to delocalization error in density functional approximations. In finite systems, delocalization error can be attributed to the incorrect treatment of fractional electron charge whereby the total energy deviates from the correct behavior of linear interpolation between integer points. For bulk systems the delocalized nature of the orbitals results in a linear total energy at fractional charges, but the slope is incorrect due to delocalization error. Multiple methods have been proposed to fix this deficiency and produce the correct linearity condition such as the Fermi-Löwdin orbital self-interaction correction, Koopmans-compliant functionals, the screened range-separated hybrid functional, the generalized transition state method, and the localized orbital scaling correction. All of these methods rely on spatially localized orbitals for their corrections, highlighting the importance of localized orbitals in modern density functional theory. The traditional method of obtaining localized orbitals minimizes the spatial variance, but here we explore an alternative approach that minimizes the combination of spatial and energetic variance. Minimizing the energetic variance allows for the occupied and unoccupied spaces to be considered together, a feature that is not prescribed in other localization schemes. The localization in energy results in localized orbitals that are more correlated with certain energy ranges, thereby making them more chemically relevant for the states.
that are associated with frontier energies. We show how these localized functions can be used in the localized orbital scaling correction to remedy many of the density functional approximation shortcomings related to delocalization error.
Acknowledgements

I would like to thank Neil Qiang Su for his patience and many hours of tutelage in quantum chemistry theory and programming. His tireless work ethic made the implementation of our theories possible. Additionally his help with mathematical derivations and program debugging in my own work proved invaluable on multiple occasions. His friendship was also one of the highlights of my time at Duke. My colleague Jacob Williams has also been instrumental in implementing certain functionality into this project, which has allowed this work to be at the point where it is today. Particularly, he was the main contributor to deriving and implementing the screened Coulomb interaction. I would like to thank my advisor Weitao Yang for his support during my time in his research group. I would also be remiss without mentioning the support of my girlfriend Leslee Arwood, and her tolerance for my long work hours. This work is also made possible with the funding from MolSSI.
Contents

Abstract iv

Acknowledgements vi

List of Figures x

List of Tables xii

1 Introduction 1
  1.1 Electronic structure theory ............................................. 1
  1.2 Wavefunctions .......................................................... 2
    1.2.1 Basis sets .......................................................... 5
  1.3 Hartree–Fock ............................................................. 7
  1.4 Configuration interaction .............................................. 8
  1.5 Coupled cluster .......................................................... 9
  1.6 Thomas-Fermi theory .................................................. 10
  1.7 Density functional theory ............................................. 11
  1.8 Orbital energies ........................................................ 15
  1.9 Delocalization error ................................................... 17
  1.10 Supplemental material ............................................... 21
    1.10.1 Janak’s theorem ................................................... 21

2 Symmetry in the Localized Orbital Scaling Correction 23
  2.1 Background .............................................................. 24
  2.2 Molecular methods ..................................................... 29
  2.3 Molecular results ....................................................... 34
    2.3.1 Hydrogen .............................................................. 34
| 2.3.2 Annulenes                                   | 35 |
| 2.3.3 Polymers                                    | 36 |
| 2.4 Summary                                       | 38 |
| 2.5 Supplemental material                         | 38 |
| 2.5.1 Relation between symmetry and degeneracy    | 38 |
| 2.5.2 Algorithm for orbitalet localization        | 41 |
| 2.5.3 Relation between symmetry and Bloch theorem | 46 |
| 2.5.4 Some relations between symmetric localized orbitals | 47 |
| 2.5.5 Supplemental molecular test results          | 48 |

3 Dually Localized Wannier Functions

<p>| 3.1 Background                                    | 57 |
| 3.1.1 Cost functions                              | 59 |
| 3.1.2 Occupied and unoccupied spaces              | 61 |
| 3.2 Methods                                       | 62 |
| 3.2.1 Bloch orbital basis                         | 64 |
| 3.2.2 Disentangled orbital basis                  | 65 |
| 3.2.3 Energy cost gradient                        | 66 |
| 3.2.4 Mixing parameter                            | 67 |
| 3.2.5 Computational details                      | 68 |
| 3.3 Results                                       | 69 |
| 3.3.1 Silicon                                     | 70 |
| 3.3.2 Copper                                      | 73 |
| 3.3.3 Ethylene                                    | 74 |
| 3.4 Summary                                       | 74 |
| 3.5 Supplemental material                         | 78 |</p>
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5.1 Appendix of tables</td>
<td>78</td>
</tr>
<tr>
<td>3.5.2 Mixing parameter conversion</td>
<td>80</td>
</tr>
<tr>
<td>3.5.3 Code implementation details</td>
<td>81</td>
</tr>
<tr>
<td>3.5.4 Keyword List</td>
<td>82</td>
</tr>
<tr>
<td>3.5.5 Keyword details</td>
<td>82</td>
</tr>
<tr>
<td>3.5.6 Energy gradient derivation</td>
<td>84</td>
</tr>
<tr>
<td>4 The Localized Orbital Scaling Correction in Periodic Systems</td>
<td>88</td>
</tr>
<tr>
<td>4.1 Periodic systems</td>
<td>88</td>
</tr>
<tr>
<td>4.2 Periodic methods</td>
<td>89</td>
</tr>
<tr>
<td>4.2.1 Energy Correction</td>
<td>90</td>
</tr>
<tr>
<td>4.2.2 Coulomb integrals</td>
<td>93</td>
</tr>
<tr>
<td>4.2.3 Screening</td>
<td>95</td>
</tr>
<tr>
<td>4.3 Periodic results</td>
<td>96</td>
</tr>
<tr>
<td>4.4 Summary</td>
<td>98</td>
</tr>
<tr>
<td>4.5 Supplemental material</td>
<td>101</td>
</tr>
<tr>
<td>4.5.1 Energy correction with disentanglement</td>
<td>101</td>
</tr>
<tr>
<td>4.5.2 Coulomb kernel derivation</td>
<td>103</td>
</tr>
<tr>
<td>4.5.3 Makov-Payne</td>
<td>110</td>
</tr>
<tr>
<td>4.5.4 Choosing the screening parameter</td>
<td>116</td>
</tr>
<tr>
<td>4.5.5 Supplemental sLOSC data</td>
<td>117</td>
</tr>
<tr>
<td>5 Conclusion</td>
<td>121</td>
</tr>
<tr>
<td>5.1 Future Work</td>
<td>122</td>
</tr>
</tbody>
</table>

References 124
List of Figures

1.1 Total energy versus electron number ........................................... 19
1.2 Delocalization error characterized in He clusters ......................... 20
2.1 $\text{H}_2^+$ dissociation curve and densities ................................. 32
2.2 Planar annulene localized orbital comparison ............................... 36
2.3 State degeneracy comparison for polymers .................................. 37
2.4 Extrapolation of trans-polyacetylene oligomers ionization energies ................................. 51
2.5 Orbital isoplots of planar annulenes .......................................... 52
2.6 Orbital isoplots of stretched planar annulenes ............................... 53
2.7 Orbital isoplots of overstretched planar annulenes ....................... 54
2.8 Orbital isoplots mixing $\pi$ bonds of benzene for different geometries ......................................... 55
2.9 Frontier orbital isoplots of $\text{C}_{10}\text{H}_{10}$ for different geometries ......................................... 56
3.1 Degeneracy pattern for the frontier Wannier functions in silicon ........... 71
3.2 Localization cost function versus mixing parameter .......................... 72
3.3 Isoplots of the frontier Wannier functions in silicon ........................ 73
3.4 Isoplots of the frontier Wannier functions in copper ........................ 75
3.5 Isoplots of the frontier Wannier functions in ethylene ...................... 76
4.1 Scatter plot of band gaps versus experiment ................................. 97
4.2 Band structure comparison for silicon ........................................... 98
4.3 Band structure for lithium fluoride ............................................. 99
4.4 Madelung constant calculated for orthorhombic unit cells ................ 114

x
4.5 Coulomb self-energy of Gaussian orbitals in the plane-wave basis . . . . . 115
4.6 Coulomb interaction energy of Gaussian orbitals in the plane-wave basis . 115
4.7 Mean absolute percent error of test sets . . . . . . . . . . . . . . . . . . . . . . 116
4.8 Comparison of experimental fundamental gaps of molecular systems . . . 120
List of Tables

2.1 Character table for $D_{6h}$ point group ........................................ 40
2.2 Character table for $D_{3h}$ point group ........................................ 40
2.3 Character table for $D_3$ point group ........................................ 40
2.4 Correspondence of the irreducible representations among $D_{6h}$, $D_{3h}$ and $D_3$ 41
2.6 Frontier orbital energies of polyacene oligmers ............................ 49
2.5 Mean absolute deviations in G2 test set ...................................... 49
2.7 Ionization energies of trans-polyacetylene oligomers ....................... 50
3.1 Spatial and energy information per DLWF for silicon ....................... 79
3.2 Spatial and energy information per DLWF for copper ...................... 79
3.3 Spatial and energy information per DLWF for ethylene .................... 80
3.4 Parameter file keywords controlling energy localization ................... 82
4.1 Mean absolute percent error of the band gap for test sets ................. 97
4.2 Table of data presented in Fig. 4.1 .............................................. 117
4.3 The molecular sLOSC data .......................................................... 119
Chapter 1

Introduction

1.1 Electronic structure theory

In today’s world we regularly interact with technology that is based on quantum mechanical effects. One example of such a technology is the creation of this document. The contents were composed on a cloud computing service, which means that the physical place that housed the data was hundreds of miles from the location where it was typed. The invention of the solid state transistor and subsequent miniaturization of circuits have allowed for virtually all people to have access to resources around the globe that would be unimaginable to the average person a century ago.

The specific area of quantum mechanics that describes how electrons behave in matter and chemistry is known as electronic structure theory. One method used for the quantum-mechanical characterization of physical systems within the electronic structure theory relies on the time-independent Schrödinger equation using atomic units where \( \hbar = 1 \):

\[
\hat{H} \Psi = \left( -\sum_i \frac{1}{2m_i} \nabla^2 r_i + \hat{V} \right) \Psi = E \Psi. \tag{1.1}
\]

Here \( \hat{H} \) is the Hamiltonian and represents all the terms in parentheses. A deceptively simple equation, it belies the complexity that can be introduced from the potential operator \( \hat{V} \), which can rely on many positions at once. This means that the solution to this equation, the many-body wavefunction \( \Psi \), can have independent degrees of freedom for every particle considered. When considering nuclei and electrons, the many-body wavefunction has three spatial degrees of freedom per nuclei and electron, plus one spin degree of freedom for each electron. The first approximation in electronic structure theory is to fix the positions of the nuclei in space, referred to as the Born-Oppenheimer approximation. This removes the
nuclei degrees of freedom and means the constant nuclear-nuclear repulsions can be ignored while solving. Once again using atomic units so that the electron mass and charge are 1, the electronic Schrödinger equation can then be written as

\[
\left( -\frac{1}{2} \sum_i \nabla_i^2 r_i - \sum_{ij} \frac{Z_i}{|R_i - r_j|} + \sum_{i<j} \frac{1}{|r_i - r_j|} \right) \Psi = E \Psi. \tag{1.2}
\]

The \( Z_i \) refer to the nuclei charges, the \( R_i \) refer to the nuclei positions, and the \( r_i \) refer to the electron positions. This is a linear second-order homogeneous partial differential equation with \( 4N \) degrees of freedom where \( N \) is the number of electrons. This can be written as \( \Psi(x_1, x_2, ..., x_N) \), where \( x_i = \{ r_i, \sigma_i \} \) includes the spatial and spin degrees of freedom respectively. The Coulomb repulsion term between electrons in Eq. (1.2) results in electrons being correlated, which prevents the solution from being separable. The known methods to solve this equation scale exponentially with system size, making them intractable for anything but the smallest systems [1]. The rest of this chapter will give an overview into some of the most well known methods for solving this problem exactly and for solving it approximately.

### 1.2 Wavefunctions

We start by examining the many-body wavefunction for electrons. Since electrons are fermions, a wavefunction for electrons must be antisymmetric under exchange. This means when two electrons are switched the wavefunction with switch signs, \( \Psi(x_1, x_2) = -\Psi(x_2, x_1) \).

Since the expectation of physical quantities depends on the magnitude of the wavefunction squared this has no effect on the expectation of observable quantities. This invariance to particles switching places is the mathematical manifestation of the indistinguishability of elementary particles, a tenant of quantum mechanics.

The antisymmetry of the fermionic wavefunction also means it can written a sum of Slater determinants [2]. A Slater determinant is the normalized determinant of a matrix of single-particle wavefunctions. Single-particle wavefunctions may also be referred to as orbitals.
For simplicity we start by showing the case where we write the many-body wavefunction as a single Slater determinant of orbitals,

$$\Psi(x_1, x_2, \ldots, x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(x_1) & \psi_2(x_1) & \cdots & \psi_N(x_1) \\ \psi_1(x_2) & \psi_1(x_2) & \cdots & \psi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(x_N) & \psi_2(x_N) & \cdots & \psi_N(x_N) \end{vmatrix}.$$  \hspace{1cm} (1.3)

The rows of this matrix are a specific electron’s position, $x_1$, and the columns are specific orbitals, $\psi_1$. The factor of $(N!)^{-1/2}$ is for normalization. Switching two columns of this matrix has the effect of switching the sign of $\Psi$, since that is a property of determinants. This is why using Slater determinants to represent $\Psi$ enforces the antisymmetry of fermions.

The Slater determinant inherently models electron correlation, which can be seen from the joint electron density

$$\rho(x_1, x_2) = \int dx_3 \ldots dx_N |\Psi(x_1, x_2, \ldots, x_N)|^2$$

$$= \frac{1}{N(N-1)} \sum_{ij} \left( |\psi_i(x_1)|^2 |\psi_j(x_2)|^2 - \overline{\psi}_i(x_1)\psi_i(x_2)\overline{\psi}_j(x_2)\psi_j(x_1) \right),$$  \hspace{1cm} (1.4)

where $\overline{\psi}$ is the complex conjugate of $\psi$. This term represents the probability of finding two electrons with spin $\sigma_1$ and $\sigma_2$ at positions $r_1$ and $r_2$ respectively for some small region of space around those points. Now consider the case where the spins are the same, $\sigma_1 = \sigma_2$. Integrating out the spins, and with a bit of abuse in notation we write the spatially dependent part of the orbitals as $\psi_i(r_j)$, we obtain

$$\rho(r_1, r_2) = \frac{1}{2} \left( |\psi_1(r_1)|^2 |\psi_2(r_2)|^2 + |\psi_1(r_2)|^2 |\psi_2(r_1)|^2 - \overline{\psi}_1(r_1)\psi_1(r_2)\overline{\psi}_2(r_2)\psi_2(r_1) - \psi_1(r_1)\overline{\psi}_1(r_2)\psi_2(r_2)\overline{\psi}_2(r_1) \right).$$  \hspace{1cm} (1.5)

The important thing to note here is that $\rho(r_1, r_1)|_{\sigma_1=\sigma_2} = 0$, which says that no matter what spatial orbitals the electrons are occupying, they may not occupy the same point in space when they are equivalent spin. This type of correlation is known as a Fermi hole, and is the Pauli exclusion principle in action.
The single determinant picture also offers a simple way of constructing the single particle density. For an arbitrary many-body wavefunction the single particle density is found by integrating out all of the coordinates except one,

$$\rho(x_1) = \int dx_2...dx_N |\Psi(x_1, x_2, ..., x_N)|^2.$$  \hspace{1cm} (1.6)

The single-particle density is sufficient for finding the expectation of observables that only depend on a single position, such as the polarization. In the single-determinant picture, however, we may find the single-particle density with a simple sum over the occupied orbitals,

$$\rho(x_1) = \sum_i |\psi_i(x_1)|^2.$$ \hspace{1cm} (1.7)

The single-determinant wavefunction does not correlate electrons of opposite spins. We once again use Eq. (1.4) in the simpler case of only two electrons and integrate out the spin for the case of opposite spin, $\sigma_1 \neq \sigma_2$. Writing the spatially dependent part of the orbitals as $\psi_i(r_j)$ this gives

$$\rho(r_1, r_2) = \frac{1}{2} (|\psi_1(r_1)|^2|\psi_2(r_2)|^2 + |\psi_1(r_2)|^2|\psi_2(r_1)|^2).$$ \hspace{1cm} (1.8)

Since this probability is unchanged if we switch $r_1$ and $r_2$ the two electrons are uncorrelated. In the case where the spatial orbitals are the same, $\rho(r_1, r_2) = |\psi_1(r_1)|^2|\psi_1(r_2)|^2$ we may even have two electrons of opposite spin occupy the same point. This same result is also found for many electrons when considering two electrons of opposite spin for single determinants.

The single determinant wavefunction can therefore offer correlation between electrons of the same spin, but a single determinant alone cannot correlate electrons of opposing spin. However, we have seen that the single determinant representation can make calculations more tractable, which we also naively expect to be true when considering the alternative as a sum of determinants. We will use the single determinant as our starting point to introduce the first approximate solution method for the electron Schrödinger equation.
1.2.1 Basis sets

In order to describe the wavefunction, we must first introduce basis sets. Basis sets are sets of functions that are used to represent orbitals as a linear combination of basis functions. Typically for molecules the basis functions correspond to the atomic orbitals for the atoms in the molecule. In order to differentiate these atomic orbitals from the single-particle wavefunctions, we refer to the single-particle wavefunctions as molecular orbitals. A common choice of atomic orbital uses one or more Cartesian Gaussians, which for 1D take the form \((x-x_0)^m\exp[-\alpha(x-x_0)^2]\). A 3D Cartesian Gaussian is then just the product of an \(x\), \(y\), and \(z\) Gaussian. The angular momentum in each dimension is set by the variable \(m\), with an \(s\)-type orbital being \(m = 0\) for all 3 dimensions. Gaussians are useful because they have the amenable property, whereby the product of two Gaussians makes another Gaussian. The spatial derivative of a Gaussian is also comprised of a sum of more Gaussians. This leads to recurrence relations that can be used to solve for their overlaps and kinetic energy analytically. Coupled with the fact that \(s\)-type Gaussians have an analytic solution for Coulomb integrals means that any combination of Gaussians also has an analytic Coulomb integral \([3]–[5]\). Gaussian basis sets also come in spherical varieties, which require fewer Gaussians at higher angular momenta. The downside of using Gaussians is that the basis functions are not orthogonal, therefore operations such as a change of basis require inversion of the overlap matrix for the whole basis set. Additionally, Gaussian basis sets cannot replicate Kato’s cusp condition whereby the wavefunction is discontinuous in its first derivative at nuclei positions \([6]\). This lack of a cusp can produce numerical instabilities for methods such as Quantum Monte Carlo because as the nuclei’s potential energy diverges as \(-Z/r\) the kinetic energy of a wave function using a finite size Gaussian basis set cannot cancel this divergence, and therefore cannot be an eigenfunction near the nuclei \([7]\).

For the case of molecules, the Dirichlet boundary conditions are \(\lim_{r\to\infty} \psi(\mathbf{r}) = 0\), which is also referred to as open boundary conditions. Another common boundary condition is the one used for solid state calculations called periodic boundary conditions, which for 1D
appear as $\psi(x + L) = \psi(x)$ for a system with periodicity length $L$. In that case the basis functions are chosen to have the same periodicity as the system in question, thereby ensuring that a linear combination of the basis functions also obeys the system periodicity. A typical basis set in the periodic case is plane waves, $\exp(\textbf{i}\textbf{G} \cdot \textbf{r})$, where $\textbf{G}$ is vector in the reciprocal space of the periodic unit cell. This leads to orbitals being represented as a discrete Fourier transform. As a complete basis set, Fourier series are capable of representing any function. The expansion approaches the true shape exactly on the periodicity interval as more basis functions are included. In practical terms this means only a finite number of plane waves are needed for a desired precision level [8]. The starting shape of a plane wave is much farther away from the desired solution than atomic orbitals though, so they require much larger basis sets. This poses a problem for diagonalizing the Hamiltonian using this basis, since dense matrix diagonalization scales with the cube of the matrix size. This problem was addressed by Davidson, who showed that if we only desire the lowest eigenvalues then we may use an iterative procedure that solves for the desired number of eigenvalues in a smaller subspace [9]. Since the basis functions are orthogonal this also offers advantages such as the Coulomb integral of two Fourier expansions being linear in the basis set size, see Sec. 4.5.2 for more details. The kinetic energy is also diagonal in the plane wave basis set. Due to the fact that the molecular orbital solutions are orthogonal, the way that higher energy orbitals remain orthogonal to the lower energy core states is to become highly oscillatory in the core region. This leads to even more plane waves being required to represent the molecular orbitals. To mitigate this, many solid state calculations employ pseudopotentials. Pseudopotentials mimic the effect of core states, but do so as a smoothly varying potential, thereby significantly reducing the number of plane wave basis states required. See [10] for a detailed description of modern pseudopotential use and generation.
1.3 Hartree–Fock

One of the oldest methods for finding a single determinant wavefunction is to consider each orbital individually. The coupling introduced by the Coulomb repulsion between electrons can be substituted for an average over all the electrons, thereby creating a set of $N$ coupled equations for $N$ electrons. This approach is named Hartree-Fock (HF), after the authors who developed it during the 1920s and ’30s. Despite its initial obscurity due to the methods computational demands and the technological limitations of the time, it is now a commonly used method in quantum mechanics. [11], [12]. In the HF scheme we wish to find the eigenfunctions of the Fock operator [13],

$$\hat{f}(\mathbf{x})\psi_j(\mathbf{x}) = \left( \hat{h}(\mathbf{x}) + \sum_i \left( \hat{J}_i(\mathbf{x}) - \hat{K}_i(\mathbf{x}) \right) \right) \psi_j(\mathbf{x}) = \epsilon_j \psi_j(\mathbf{x}) \quad (1.9)$$

$\hat{h}$ is the one body Hamiltonian containing the kinetic energy of electron $\mathbf{x}$ and nuclear repulsion. $\hat{J}_i(\mathbf{x})$ is the Coulomb operator, which is

$$\hat{J}_i(\mathbf{x}) = \int d\mathbf{x}' \frac{\psi_i(\mathbf{x}') \overline{\psi}_i(\mathbf{x}')} {||\mathbf{r} - \mathbf{r}'||}, \quad (1.10)$$

where $\psi_i(\mathbf{x})$ is a the molecular orbital. Summing over the molecular orbitals in the $\hat{J}_i$ of Eq. (1.9) leads to the single particle density as show in Eq. (1.7). In this manner the electron-electron repulsion integrals are treated with a mean field. $\hat{K}_i(\mathbf{x})$ is the exchange term, which is defined by its action upon an orbital:

$$\hat{K}_i(\mathbf{x})\psi_j(\mathbf{x}) = \left[ \int d\mathbf{x}' \frac{\overline{\psi}_i(\mathbf{x}') \psi_j(\mathbf{x}')} {||\mathbf{r} - \mathbf{r}'||} \right] \psi_i(\mathbf{x}). \quad (1.11)$$

The exchange is a non-local operator in that it depends on orbital $\psi_j$’s value over more than one point in space. It’s crucial to combine the exchange operator with the mean field Coulomb term since the combination cancels when acting on the molecular orbital of the same index, i.e. $(\hat{J}_i - \hat{K}_i)\psi_i = 0$. This avoids the self-interaction term that would otherwise be included when using the Coulomb interaction of the total density. The exchange operator
arises from representing the wavefunction as a Slater determinant in order to satisfy the anti-symmetric property of fermions.

The fact that the Fock operator of Eq. (1.9) depends all the molecular orbitals means the individual differential equations are coupled to each other. Starting from an initial guess, the system is then evolved to find the minimum total energy.

$$E^{HF} = \sum_i \langle i|\hat{h}|i\rangle + \frac{1}{2} \sum_{i,j} \langle ij||ij \rangle$$  \hspace{1cm} (1.12)

\(\hat{h}\) is the same as before and the notation \(\langle ij||ij \rangle\) signifies a four-center electron Coulomb integral minus the corresponding exchange integral. Since the HF equations are nonlinear they are solved using a self consistent field (SCF) method to obtain the coefficients that comprise the orbitals. This means that after a trial solution is proposed, the density is found for that trial. The orbitals are then found which diagonalize the effective Hamiltonian for that trial density. Then the density is updated and the cycle is repeated until the density does not change. This allows the calculation to scale as \(O(N^3)\) with electron count and \(O(K^4)\) with a non-orthogonal basis set size due to two-electron integrals [14]. This makes HF fairly quick compared to higher order methods. The mean field approach to electron correlation means it is also one of the least accurate.

### 1.4 Configuration interaction

Many methods use the HF orbitals as a starting point and are known as post-HF methods. The limited accuracy of Hartree-Fock can mainly be attributed to its mean field treatment of electron repulsion and because it only uses a single determinant to represent the wavefunction. By allowing the wavefunction to also be represented by an expansion using excited state determinants, the whole space spanned by a basis set can be reached. Given a basis set of size \(K\), there are \((\binom{2K}{N})\) determinants in this expansion, which means it grows exponentially with basis set size. Diagonalizing the full Hamiltonian with respect to these wavefunctions will give the coefficients for this expansion. The smallest eigenvalue of this
diagonalization will be the exact ground state energy in the Born-Oppenheimer approximation for a given basis set. This method is known as configuration interaction (CI) since the wavefunction is represented as a configuration of states. The origin of the method can be traced to Slater in the same paper where he introduced Slater determinants [2], and the rules governing the interactions between configuration state functions was shown in [15].

Using all possible $\binom{2K}{N}$ choices is known as full CI, but for a large basis set and even modest amount of atoms this becomes computationally infeasible. To mitigate this, the expansion can be cut off at some higher excited state determinant; this is known as truncated CI. This method, however, suffers from size-consistency issues whereby the energy of two molecules far apart does not equal twice the energy of a lone molecule [13].

1.5 Coupled cluster

Another method came about as an extension to truncated CI. It was found that systems that were size inconsistent in truncated CI at the double excitation level could be made size consistent with different coefficients. This solution was found by first approximating the quadruple exciton coefficients as the double exciton coefficients squared. Just as one can gain accuracy from treating pairs of electrons with a double excitation, more accuracy can be obtained from treating pairs of pairs. Adapting this method from quantum field theory that had originally been used to deal with nuclear physics, it was brought to electronic state calculations by Čížek [16].

This method is known as coupled cluster (CC), which treats pairs of pairs specifically with the cluster operator $e^{\mathcal{J}_1}$, where $\mathcal{J}_1 = \sum_{r,a} c_a^r \hat{a}_r^\dagger \hat{a}_a$. The $\hat{a}_r^\dagger$ and $\hat{a}_a$ are the creation and annihilation operators used in second quantization to create an excited state from the ground state. Successive applications of the $\hat{a}_r^\dagger \hat{a}_a$ operator for different $a$ and $r$ can create any possible excited state. Since the exponentiation of an operator can be represented as a sum of all positive powers, this will give more than just singly excited states. In practice due to a finite number of occupied orbitals this expansion will naturally truncate itself.
The CC wavefunction is then $|\psi_{CC}\rangle = e^{i\alpha}|\psi_0\rangle$, and the goal of the CC method is to find the coefficients that satisfy the Schrödinger equation for this ansatz. In this example we only used the single excitation operator but the true cluster operator would exponentiate all possible excitations, $e^{i\alpha_1+i\alpha_2+i\alpha_3+...}$. In reality the level of excitation is truncated for computational feasibility. If it is truncated after the double excitation, $I_2$, it is known as CCSD; truncating after $I_3$ is CCSDT. CC has the advantage of being size consistent and much more accurate than HF. Even at the CCSD level, which scales as $O(K^6)$ with basis set size, it is unfeasible for systems with more than tens of electrons [13].

1.6 Thomas-Fermi theory

All the aforementioned schemes have attempted to approximate solutions the Schrödinger equation, namely the wavefunction $\psi$. The goal of solving the Schrödinger equation is to obtain the expectation value of observables, most commonly the total energy. A tenant of quantum mechanics is that observables can also be obtained from the density, $\langle \psi | H | \psi \rangle = \text{tr}(\hat{\rho} \hat{H})$. Densities, however, have an inherent limitation not present in wavefunction methods. All the schemes presented up to now rely on the variational principle. What this principle says is that for any possible ansatz, $\psi_a$, the energy for that ansatz is always greater than or equal to the true ground state energy:

$$\langle \psi_a | \hat{H} | \psi_a \rangle \geq \langle \psi_0 | \hat{H} | \psi_0 \rangle \quad (1.13)$$

This means we can always know our approximation is greater than the true ground state energy. To solve for the ground state then we can vary our $\psi_a$ until we find the smallest possible value. The normal constraints on a density, such as being Hermitian and having a trace equal to the number of particles, is not enough to guarantee an arbitrary density will obey the variational principle. This drawback plagued much of the early work that attempted to use densities since they are not bound by the true ground state energy.

One of the first methods to use the density was the Thomas-Fermi method, which
attempted to solve the functional [17], [18]

\[ E^{TF}[\rho] = T[\rho] + V_{ne}[\rho] + J[\rho] \quad (1.14) \]

In this approach they used a kinetic energy formulated from the HF calculation of a non-interacting uniform electron gas to obtain the functional

\[ T[\rho] = \frac{3}{10} (3\pi)^{2/3} \int [\rho(r)]^{5/3} dr \quad (1.15) \]

The energy due to the nuclei is

\[ V_{ne} = \int \rho(r)v(r)dr; \quad v(r) = \sum_a \frac{Z_a}{|r-r_a|} \quad (1.16) \]

and the Coulomb repulsion energy took the form

\[ J[\rho] = \frac{1}{2} \int \int \frac{\rho(r)\rho(r')}{|r-r'|} dr dr' \quad (1.17) \]

The nuclear and Coulomb terms are obtained from their classical interpretations. Later Dirac added an exchange energy obtained from a non-interacting electron gas [19],

\[ E_x[\rho] = -\frac{3}{4} \left( \frac{3}{\pi} \right)^{1/3} \int [\rho(r)]^{4/3} dr. \quad (1.18) \]

This then became the Thomas-Fermi-Dirac (TFD) model. The TFD model suffers from quite a few shortcomings that preclude it from being used in chemistry, including not reproducing atomic shell structures and not binding molecules together. Attempts to improve upon this model, such as the gradient correction to the kinetic energy proposed by von-Weizacker in 1935, did not remedy these problems [20].

### 1.7 Density functional theory

Due to the above referenced limitations, density methods were largely neglected until 1964, which is the year Hohenberg and Kohn laid down the theory that justified using the density
to solve for the ground state of the Schrödinger equation [21]. Their first theorem stated that there is a one-to-one mapping between the ground state density and an external potential up to a constant. The second theorem is that if a density can be made from a ground state wavefunction of any external potential, known as \( \nu \)-representability, the energy of that density using the true potential is greater than the true ground state energy. Mathematically this can be represented as

\[
E_{v_0}[\rho_0(r)] < E_{v_0}[\rho_1(r)]; \forall \rho_1 \in \mathcal{V}_N; \rho_0 \neq \rho_1.
\]  

(1.19)

In this equation the \( v_0 \) subscript on \( E \) signifies the true external potential, \( \rho_0 \) is the true ground state, and \( \mathcal{V}_N \) is the \( \nu \)-representability set. Essentially, the theorem allows us to recover the variational principle using densities. If we can find a potential that is more easily solvable, any density arising from the wavefunctions of that solution will give us an energy greater than the true ground state. At first glance this result may seem remarkable, but it does not simplify solving for solutions to the Schrödinger equation in practice. Following this, Kohn and Sham went a step further by proving this also holds true for densities corresponding to a non-interacting reference system [22]. The domain of this formulation, \( \mathcal{V}_N^0 \), differs from the \( \mathcal{V}_N \) in the Hohenberg and Kohn theorems in that it comes from a non-interacting system. It assumes that a density satisfying such a variational principle would lie in the intersection of \( \mathcal{V}_N^0 \) and \( \mathcal{V}_N \). This can be represented mathematically as

\[
E_{\nu}[\rho_0(r)] \leq E_{\nu}[\rho_1(r)]; \forall \rho_1 \in \mathcal{V}_N \cap \mathcal{V}_N^0
\]  

(1.20)

In practice this minimization transforms the problem of finding the ground state one of one N-body problem into N one-body problems. We can state this as

\[
E_0 = \min_{\rho \in \mathcal{V}_N^0} E_{\nu}[\rho] = \min_{\\{\phi_i\}} E_{\nu}[\rho]
\]  

(1.21)

We reformulate the minimization over the density to one over a set of orbitals because the kinetic energy formula is a function of the orbitals. The density is explicitly dependent on

\footnote{These energy bounded density theorems rely on a non-degenerate ground state assumption that can be circumvented by alternative formulations such as [23].}
the orbitals,

\[ \rho(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2; \quad \langle \phi_i | \phi_j \rangle = \delta_{ij} \]  

(1.22)

This makes the minimization implicitly over the density. We can now employ a Lagrangian to write out our system of equations, using the orthonormality of our orbitals to obtain Lagrangian multipliers.

\[ \mathcal{L}_\nu = E_v[\rho] + \sum_{i,j=1}^{N} \lambda_{ij} [\langle \phi_i | \phi_j \rangle - \delta_{i,j}] . \]  

(1.23)

Employing the standard Lagrangian technique that the derivative of a stationary state is zero with functional derivatives we arrive at

\[ h_s \phi_i \equiv \frac{\delta E[\rho]}{\delta \phi_i^*} = \left[ -\frac{1}{2} \nabla^2 + v_s \right] \phi_i = \sum_j \lambda_{ij} \phi_j \]  

(1.24)

for our individual differential equations. Here \( h_s \) is the Kohn Sham (KS) Hamiltonian, where \( v_s \) is the KS effective potential.

\[ v_s(\mathbf{r}) = v_{ne}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{xc}(\mathbf{r}); \quad v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} \]  

(1.25)

\( v_{xc} \) is called the local exchange-correlation potential. Since \( \rho \) is Hermitian this implies that \( v_s \) is also Hermitian, which in turn implies \( h_s \) is Hermitian. Finally, this also results in \( \overline{\lambda_{ij}} = \overline{\lambda_{ji}} \), where the bar above \( \lambda_{ij} \) denotes the complex conjugate. The fact that \( h_s \) only depends explicitly on \( \rho \) tells us it is invariant under unitary transform. A transform must then exist that will diagonalize the \( \lambda_{ij} \)'s, leading to the celebrated KS equation,

\[ h_s \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}). \]  

(1.26)

We can now see the trick of how the KS formulation decouples the Schrödinger Equation. Since the \( h_s \) is implicitly dependent on the \( \phi_i \)'s through the density these equations are nonlinear and are solved using an SCF procedure. To find the ground state then one would
find the set of orbitals that minimized the overall energy given by

\[ E_v[\rho] = T_s[\rho] + J[\rho] + \int \rho(r)v_{ne}(r)dr + E_{xc}[\rho] \]  

(1.27)

We have written \( T_s \) as a functional of \( \rho \) for consistency, but it is evaluated using the eigenstates of \( h_s \), which are implicit functionals of \( \rho \). At this point the KS scheme is theoretically exact in the Born-Oppenheimer picture. Unfortunately, there is a glaring problem in that the \( E_{xc}[\rho] \) is unknown.

The first form for the \( E_{xc} \) functional was proposed in the seminal paper that also laid down the KS equations. It was suggested that for systems with slowly varying density the \( E_{xc} \) for a uniform electron gas could be used, which is called the local density approximation (LDA). The exchange part of this energy was already known from Dirac, as mentioned above. Originally the Wigner approximation was used for the correlation, which limited the accuracy of LDA to below what was useful for chemistry. In 1980 Ceperly and Alder published their procedure for simulating the electron gas using a Stochastic Monte Carlo process that allowed calculating the correlation of any density to desired precision [24]. Later this was fit to an analytic function for interpolating the correlation at any density known as VWN5. This put LDA in the accuracy range of around 70 kcal/mol.

For systems that have highly non-uniform densities the next step is to use the gradient term in a series expansion. These type of functionals are known as generalized gradient approximations (GGAs). A naive Taylor series expansion of the \( E_{xc} \) however, can lead to functionals that perform worse than LDA because they do not satisfy the electron hole sum rules [25]. Arguably the two most well known GGAs are BLYP, a combination of the Becke88 exchange with the Lee-Yang-Parr correlation functional, and the Perdew-Burke-Ernzerhof (PBE) functionals. These are only slightly more computationally intensive than LDA and can bring accuracies down to around 10 kcal/mol. After noticing that inclusion of a portion of exact exchange (i.e. HF exchange) improved upon GGAs, hybrid functionals were introduced. They mix portions of GGA, LDA, and exact exchange. The mixing pro-
portions are found by fitting to thermochemical data. These can fall into three categories: global, local, and range separated. The global hybrid uses a fixed ratio; these are the most common, with B3LYP being one of the most popular DFT methods. Local hybrids are similar except the mixing ratio can depend on space. Range separated functionals attempt to mimic a screened Coulomb operator. They use one set of mixing ratios for close-range interaction, and another set for long range. A popular choice for molecules is CAM-B3LYP, which uses exact exchange at long range. This is problematic in periodic systems though and is reversed in the HSE functional, which is more commonly used in bulk calculations [26]. The inclusion of different functional types once again slightly increases the computational complexity of hybrids in relation to GGAs. In the case of B3LYP it can lower the accuracy down to the neighborhood of 4 kcal/mol.

1.8 Orbital energies

One of the most relevant quantities after calculating the total energy is the spectral function, which tells us the energy levels of the system. This can be probed experimentally with photoemission spectroscopy, which shines a specific frequency of light on a sample and measures the energy of any emitted electrons. The peaks measured in the photoemission spectrum correspond to quantized energy states. From a theoretical calculation this can be compared with a density of states that gives the likelihood of an electron in the system having a specific energy. In the theory of single-particle pictures such as HF and DFT, this is correlated with the eigenvalues of the molecular orbitals.

Another related quantity is the chemical potential of a system that refers to the energy change of the system upon release or absorption of an electron, \( \mu \equiv \partial E/\partial N \). The chemical potential, therefore, differs whether it is approached from above or below,

\[
\mu(N) = \begin{cases} 
-I(N) = E(N) - E(N - 1); & dN < 0 \\
-A(N) = E(N + 1) - E(N); & dN > 0.
\end{cases}
\]

I(N) is the ionization potential (IP) and A(N) is the electron affinity (EA). The difference
of these two quantities is referred to as the integer gap:

\[ E_{\text{gap}}^{\text{integer}} = I - A \]

\[ = [E(N-1) - E(N)] - [E(N) - E(N+1)] \]  \hspace{1cm} (1.29)

Also known as the fundamental gap, this quantity is of key importance because it signifies the smallest amount of energy required to excite the system from the ground state. For a conducting material the fundamental gap is zero. In semiconductors and insulators, this quantity is greater than zero and predicting the fundamental gap accurately is key to allowing for the discovery of new materials. Another related quantity is the derivative gap,

\[ E_{\text{gap}}^{\text{deriv}} = \frac{\partial E}{\partial N} \bigg|_+ - \frac{\partial E}{\partial N} \bigg|_- . \]  \hspace{1cm} (1.30)

Comparing Eqs. 1.29 and 1.30, we see if the lines connecting \( E(N-1) \) to \( E(N) \) and from \( E(N) \) to \( E(N+1) \) were linear, then these two quantities would be the same. The behavior for systems at non-integer number of electrons is actually defined within the context of grand canonical ensemble theory. Using this connection it was proven that for the ground state, the behavior of the energy curve between integer points should be a linear interpolation. This was shown in 1982 by the authors Perdew, Parr, Levy, and Balduz and we therefore refer to this linearity condition as the PPLB condition [27]. More current literature also exists that shows the linearity condition using pure states [28]. The result of this remarkable connection though is that for an exact theory the derivative gap and integer gaps should be equivalent.

The relevance of the derivative gap can be found starting from Janak’s theorem, which says that derivative of the total energy with respect to an individual orbital occupancy is that orbital’s eigenvalue:

\[ \frac{\partial E}{\partial n_i} = \epsilon_i . \]  \hspace{1cm} (1.31)

This theorem states that if we take the derivative of total energy with the highest occupied molecular orbital (HOMO) occupation we should get the HOMO energy. The same argu-
ment applies to the lowest unoccupied molecular orbital (LUMO). Since the HOMO energy marks the end of the occupied orbital energies and the LUMO energy marks the beginning of the unoccupied orbital energies, they are referred to as frontier orbitals.

In the case of exchange-correlation functionals that explicitly depend on the single-particle density such as LDA and GGAs it has been shown that taking the derivative of the total energy with respect to particle number from below will yield the HOMO energy [29]. Taking the same derivative from above will give the LUMO energy, i.e.

\[
\frac{\partial E_v}{\partial N} = \epsilon_f.
\] (1.32)

Here \( E_v \) is used to show this particular result only applies to explicitly density-dependent functionals and \( \epsilon_f \) stands for frontier orbital energy. For results pertaining to other methods in the generalized KS framework see [29]. This means we can find the derivative gap by simply taking the difference of the frontier orbital energies. Coupled with the PPLB condition this implies that for an exact functional the fundamental gap can be found using just the frontier orbital energies from a DFT calculation. In practice, though, it is well known that the derivative gap of LDA and GGAs greatly underestimates the integer gap. However, a neutral \( E(N) \) system can be calculated and an ionic system at \( E(N \pm 1) \) can also be calculated, which will be accurate since DFT gives good total energies for integer electron calculations of small to medium sized molecules. The IP and EA can then be found from the difference of the total energies, leading to this type of calculation being called \( \Delta \text{SCF} \). \( \Delta \text{SCF} \) calculations show much better agreement with experiment than the derivative gap found from the neutral species calculation. The source of this shortcoming is explored in the next section.

### 1.9 Delocalization error

Despite the success of DFT, it is known to not make accurate predictions for certain properties. These include underestimating band gaps and reaction barriers [30], undervaluing
dissociation curves \[31\], overestimating conductance and polarizability \[32\], incorrect energy level alignment and charge transfer across interfaces \[33\], \[34\], and a lack of dispersion forces \[35\]. Many of these inaccuracies can be traced back to the exchange-correlation energy. Correlation errors are typically classified as static (strong) and dynamic (weak). The weak correlation error has to do with DFT incorrectly modeling long-range interactions. In molecules this partly stems from using Gaussian basis sets, which decay faster than Slater type orbitals. The static correlation error affects the electronic structure and arises in cases with closely spaced energy levels. This can be critical for properties that directly use orbital energies such as band gaps.

Historically, many of DFT’s shortcomings have been attributed to self-interaction error (SIE). In the SIE an electron repulsed itself in the classical Coulomb energy calculation since it uses the total density. We can see this is treated correctly in HF calculations for one-electron systems because the Coulomb self-energy exactly cancels with the exchange. In real systems an electron would induce a Fermi hole of reduced density around itself to account for the Pauli exclusion principle. In traditional DFAs, however, the $E_x$ does not exactly cancel the Coulomb term for a single electron, resulting in the SIE. Attempts to construct an SIE free functional were developed in the MCY2 \[36\] and B05 \[37\] functionals, but they were found to still suffer from problems such as incorrect dissociation limits.

More recent literature has traced the errors associated with orbital self interaction as stemming from delocalization error. Delocalization error manifests as the violation of the PPLB condition for fractional electrons in finite systems. Standard DFAs undervalue dissociation curves due to an underestimation of energy at a non-integer number of electrons, in violation of the PPLB condition. In large systems and in the periodic limit delocalization error gives an energy curve that is linear but with incorrect slope, leading to incorrect total energies for ionized systems \[38\], \[39\].

In the simplest case this can be seen in systems with one electron, as in $H^+_2$. In the dissociation limit each hydrogen atom has half an electron. Since DFAs undervalue non-
integer energies this system has less energy than it should as an H and H\(^+\). Delocalization error gives DFAs a tendency to unphysically broaden the electron density and underestimate non-integer electron energies. This leads to a convex behavior of DFAs in non-integer cases. Recent work has also shown the DFA error is not purely convex and can be concave at points approaching integer values \[40\]. HF tends to over localize electrons, which leads to a concave behavior of overestimation between integers known as localization error, see Fig. 1.1. The mirrored nature of these errors gives an intuitive reasoning for mixing exact exchange and GGAs in hybrid functionals.

![Figure 1.1](image)

**Figure 1.1**: Total energy curve versus number of electrons for hypothetical functionals. The true behavior corresponding to the PPLB condition is the black line labeled as exact. The delocalization of DFAs leads to a concave underestimation at fractional electrons as shown in blue. The over localization of HF leads to a convex overestimation at fractional electrons as shown in green. The inherently delocalized nature of the orbitals in the bulk limit leads to the incorrect linear behavior as shown for the dashed purple line.

Many of the errors in DFT may actually be attributed to delocalization error. In the case of band gaps we noted previously that DFT gave derivative gaps that were too small. Looking at Fig. 1.1 we can see that this derivative discontinuity will be severely underestimated using DFAs. Now we can also understand why this occurs through the breaking of the PPLB condition and the relation of the frontier orbitals to the chemical
Figure 1.2: (a) is the difference for $\epsilon_{HOMO}$ and $I_{ve}$ on top and the difference in $I_{ve}$ and experiment on bottom. (b) is the difference in the fractional energy $E(He^{\delta+})$ and the linear interpolation at that point. Reproduced from [41].

Potential. The incorrect prediction of a band gaps leads to poor energy level alignment at interfaces. Since charge transfer depends on the difference in band gap levels this type of calculation is beyond current standard KS DFT. DFAs also suffer from underestimating reaction barriers due to underestimating transition state barriers. These can be improved by increasing the amount of HF exchange, indicating they rely on fractional charges [35]. Polarizability is also overestimated due to the delocalization of fractional charge at the edge of a molecule [42].

In this work we will explore corrections that attempt to address delocalization error both in finite and bulk calculations. The next chapter deals with finding localized orbitals, specifically those used in the periodic case. This localized transformation of the effective Hamiltonian states is key to representing the fractional charges that are incorrectly treated in DFAs. In the last chapter we then explore how the amount of fractional charge in the localized picture can be used to restore the linearity condition and correct delocalization error.
1.10 Supplemental material

1.10.1 Janak’s theorem

We can define the density for arbitrary occupancy:

$$\rho(r) = \sum_i n_i |\psi_i|^2,$$

where \(n_i\) is the occupancy of orbital \(\psi_i\). Then we rewrite the usual KS kinetic energy to include occupancy

$$T_s[\rho] = \sum_i t_i \implies \tilde{T} = \sum_i n_i t_i.$$

This gives a total energy of

$$\tilde{E} \equiv \tilde{T} + U[\rho] + E_{xc}[\rho].$$

Taking the derivative of the energy with respect to occupancy:

$$\frac{\partial \tilde{E}}{\partial n_i} = \epsilon_i + \sum_j n_j \frac{\partial t_j}{\partial n_i} + \int v_{\text{eff}}(|\psi_i|^2 + \sum_j n_j \frac{\partial |\psi_j|^2}{\partial n_i}) dr.$$ 

For the individual kinetic energies we have

$$t_i = \int \overline{\psi(r)_i}(-\nabla^2)\psi_i(r) dr = \epsilon_i - \int \overline{\psi_i(r)}v_{\text{eff}}(r)\psi_i(r) dr$$

Plugging that form of orbital kinetic energy back into the total energy derivative we obtain

$$\frac{\partial \tilde{E}}{\partial n_i} = \epsilon_i + \sum_j n_j \left( \frac{\partial t_j}{\partial n_i} + \int dr v_{\text{eff}}(r) \frac{\partial |\psi_j(r)|^2}{\partial n_i} \right).$$

We then find the derivative of the orbital kinetic energy with respect to arbitrary occupancy and find

$$\frac{\partial t_j}{\partial n_i} = \int \frac{\overline{\psi_j(r)}(-\nabla^2)\psi_j(r) dr}{\partial n_i} + \text{c.c.},$$
where c.c stands for complex conjugate of the preceding term. This can then be plugged back into the total energy derivative

\[ \frac{\partial \tilde{E}}{\partial n_i} = \epsilon_i + \sum_j n_j \left( \int d\mathbf{r} \frac{\overline{\psi}_j(\mathbf{r})}{\partial n_i} \left( -\nabla^2 + v_{\text{eff}}(\mathbf{r}) \right) \psi_j(\mathbf{r}) + \text{c.c.} \right) \]

From the orthogonality of the orbitals we know the relation

\[ \sum_j n_j \epsilon_j \frac{\partial}{\partial n_i} \int |\psi_j|^2 d\mathbf{r} = 0 \]

This means for the total energy derivative the only surviving term is the orbital energy eigenvalue, which is the result of Janak’s theorem \[43\]:

\[ \frac{\partial E}{\partial n_i} = \epsilon_i \]

There has been some publication disputing the validity of Janak’s proof in the context of density functional theory since it takes derivatives of discontinuous quantities, namely the occupations. This is a valid point since the occupations are taken to be integers. However, there are alternative proofs that rely on a Lagrange multiplier that is analogous to the chemical potential and are well defined for arbitrary occupations, see Sec. III. of \[44\]. The result of this and other alternative proofs, however, concludes the same result as Janak. It is not surprising considering the fact that within the KS framework there is nothing forbidding non-integer occupancy. Indeed calculations may be carried out with explicitly fractional occupations in order to observe the behavior of the \( E(N) \) curve for fractional number of electrons.
Chapter 2

Symmetry in the Localized Orbital Scaling Correction

Density functional theory offers reasonable accuracy at acceptable computational cost, but as noted previously it suffers from delocalization error. This results in inaccurate predictions of quantities related to energy differences, such as band gaps. The localized orbital scaling correction (LOSC) was developed to correct delocalization error by using orbitals localized in space and energy [41]. These localized orbitals span the occupied and unoccupied spaces in order to address frontier orbital degeneracy and offer corrections to the energy eigenvalues. In its original incarnation LOSC was shown to be an effective method to reduce the delocalization error associated with commonly used density functional approximations; it significantly improves the descriptions of dissociation of cationic species, band gaps, and photoemission spectra. However, the original method suffered from problems such as broken symmetry and non-smooth dissociation energy curves. In this chapter we explore modifications to the original LOSC method in molecules that reduces the number of fitted parameters and results in a more robust treatment of system symmetry and degeneracy while retaining the same level of accuracy as the original method [45]. The principle change in the method that allows for a more robust treatment of system symmetries involves introducing a new localization scheme that places spatial and energetic localization on equal footing. Additionally, the curvature matrix is redefined to smooth dissociation curves. However, before introducing the LOSC method and its modifications, we first review the motivation of how orbital-dependent density functionals are able to correct delocalization error\(^1\).

\[^1\]Parts of this chapter has been adapted from the following previously published article: Neil Qiang Su, Aaron Mahler, and Weitao Yang, *Preserving Symmetry and Degeneracy in the Localized Orbital Scaling Correction Approach*, J. Phys. Chem. Lett. 2020, 11, 4, 1528–1535 [45].
2.1 Background

A quantity of particular interest in DFT is the meaning of the orbital energies. In the Hartree–Fock approximation, the ionization energy and electron affinity of a system are equal to the negative of the highest occupied and lowest unoccupied orbital eigenvalues respectively, assuming frozen orbitals \[46\]. A similar connection can be shown for the chemical potential \( \mu = \frac{\partial E}{\partial N} \), which shows how the energy changes with respect to particle number. Perdew, Parr, Levy, and Balduz showed in [27] that \( E(N) \) under the exact functional is piecewise linear between points of integer occupancy,

\[
E(N + \delta) = \begin{cases} 
(1 + \delta)E(N) + \delta E(N - 1) & \delta < 0, \\
(1 - \delta)E(N) + \delta E(N + 1) & \delta > 0,
\end{cases}
\]

which is referred to as the PPLB condition. A theory that obeys the PPLB condition will thus have a chemical potential given by

\[
\mu(N) = \begin{cases} 
-I(N) = E(N) - E(N - 1) & dN < 0, \\
-A(N) = E(N + 1) - E(N) & dN > 0.
\end{cases}
\]

Here, \( I(N) \) is the ionization potential (IP) of the \( N \)-electron system, and \( A(N) \) is its electron affinity (EA). Justifying the use of chemical potentials as theoretical prediction of \( -\text{IP} \) and \( -\text{EA} \), Cohen et al. showed in [29] that for any local or nonlocal exchange-correlation functional continuous in the electron density the chemical potential \( \frac{\partial E}{\partial N} \) is given by the derivative of the energy with respect to the frontier occupation numbers, \( n_{\text{HOMO}} \) and \( n_{\text{LUMO}} \), which are equal to the respective frontier orbital energy eigenvalues:

\[
\frac{\partial E}{\partial N} = \begin{cases} 
\frac{\partial E}{\partial n_{\text{HOMO}}} = \epsilon_{\text{HOMO}} & \partial N < 0, \\
\frac{\partial E}{\partial n_{\text{LUMO}}} = \epsilon_{\text{LUMO}} & \partial N > 0.
\end{cases}
\]

Here, \( \epsilon_{\text{HOMO}} \) is the highest occupied molecular orbital (HOMO) energy and \( \epsilon_{\text{LUMO}} \) is the lowest unoccupied molecular orbital (LUMO) energy. The second equality in Eq. (2.3) is the result of Janak’s theorem [43]. Taken together, these facts mean that the negative IP
(EA) is equivalent to the HOMO (LUMO) orbital energy for a theory that obeys the PPLB condition. In other words, the HOMO (LUMO) orbital energy can be used as theoretical prediction of negative IP (EA) and EA from the DFA used, without carrying out the finite difference calculation in Eq. (2.2). The difference between the IP and EA is the integer gap of the system,

\[ E_{\text{integer gap}} = I - A = E(N-1) - 2E(N) + E(N+1). \]  

\[ (2.4) \]

\( E_{\text{integer gap}} \), also called the fundamental gap, signifies the difference between positively and negatively ionizing the system and is a crucial part of the accurate modeling of semiconductors’ electronic structure. Note that fundamental gap can also be computed as the chemical potential difference. This is the derivative gap, defined as

\[ E_{\text{deriv gap}} = \left. \frac{\partial E}{\partial N} \right|_+ - \left. \frac{\partial E}{\partial N} \right|_- = \epsilon_{\text{LUMO}} - \epsilon_{\text{HOMO}}. \]

\[ (2.5) \]

For a theory that obeys the PPLB condition, the derivative gap will be equal to the integer gap [29]. For molecular systems, the chemical potential and fundamental gap can then be determined from the orbital energies of the single \( N \)-electron calculation; alternatively, the \( (N \pm 1) \)-electron energies can be obtained to compute the integer gap directly (the \( \Delta \text{SCF} \) method). In bulk systems, calculations of charged systems are much more difficult, so the derivative gap offers the easiest way to determine the fundamental gap of the system.

The delocalization error has a dramatic size-dependent manifestation. In finite systems, standard DFAs fail to obey the PPLB linearity condition, so the derivative gap is not equal to the integer gap. This is due to the error in the approximate exchange-correlation functional, which yields a convex \( E(N) \) curve, underestimating the linear interpolation prescribed by the PPLB condition. This convex deviation has been identified as the cause for an unphysical smearing of the electron density in space, as well as underestimation of the total energy in a delocalized electron density; thus, we may identify it with delocalization.
error. In bulk systems, the delocalized nature of the orbitals produces a total energy linear with respect to fractional charge, appearing to yield no deviation from the PPLB condition; however, delocalization error manifests as an incorrect slope of the $E(N)$ line [38].

The effects of delocalization error include the underestimation of band gaps and reaction barriers [30], undervaluation of dissociation curves [31], [47], [48], overestimation of conductance and polarizability [32], and incorrect energy level alignment and charge transfer across interfaces [33]–[35]. To capture the full derivative discontinuity, and hence band gap, it has been shown that the exact functional, either local or nonlocal, cannot be a differentiable functional of the electron density, or the Kohn-Sham density matrix [49], [50]. To reduce the systematic delocalization error, many approaches have been developed, including range-separated functionals [51]–[57], the screened range-separated hybrid functional [58], self-interaction error corrected functionals [48], [59]–[65], Koopmans-compliant functionals [66], [67], and generalized transition state methods [68] along with the related development using localized Wannier functions [69].

The shape of the non-linear delocalization error has been characterized as quadratic in shape [70]. This naturally leads to a linearity restoring correction that subtracts a square term and adds back a linear term proportional to the fractional occupation,

$$
\Delta E_{\text{linear}} \propto n_f - n_f^2,
$$

(2.6)

where $n_f$ is the occupation of the frontier state. We first note that this term offers no correction to the energy at integer occupancies, which is good since DFAs are known to give accurate total energies for systems with an integer number of electrons. This means we may extend the form of the correction to all orbitals and only the fractionally occupied orbitals will contribute to the energy correction,

$$
\Delta E_{\text{linear}} \propto \sum_i \alpha_i \left(n_i - n_i^2\right).
$$

(2.7)

The second thing to note is the form of orbital energy correction this offers. We can deduce
the form of this correction by applying Janak’s theorem for a given orbital occupation,

$$\Delta \epsilon_{j}^{\text{linear}} \propto \frac{\partial \Delta E_{\text{linear}}}{\partial n_{j}} \propto \alpha_{j} \left(\frac{1}{2} - n_{j}\right)$$  \hspace{1cm} (2.8)

This shows that for occupied states this type of correction will lower the energy since $1/2 - 1 < 0$. For unoccupied states the opposite is true and the correction will raise the energy. We have assumed that the $\alpha_{i} > 0$, but this condition must be true for underestimated energy curves as is the case for DFAs. The lowering of occupied energy states and raising of unoccupied energy states has a history within DFT, and has been empirically implemented in the past to match DFA data to experiment. It is referred to as the scissor method [71], and is still used today, see [72], although as an empirical method it has mostly been replaced by *ab initio* methods.

This form of correction for the eigenvalues also has zero correction for half-occupied states. Looking at Fig 1.1, which shows the deviation from linearity for DFAs, a quadratic deviation is parallel to the linear curve halfway between the integer points. Since the slope of the total energy curve is equivalent to the frontier orbital energy, it makes sense that there should be no correction to the frontier eigenvalue at the point where the slopes are equivalent. This is also in agreement with Slater transition theory, which approximates excitation energy using the ”transisition state”. Slater takes the transition state to be the point where the excited process is halfway between the initial and final states [2].

There are numerous DFT correction methods that employ an orbital-dependent correction that can be cast into a form similar to Eqs. (2.7) and (2.8). The method proposed in [69] uses MLWFs for the orbitals in a LDA+U scheme, which attempts to emulate nearest-neighbor interaction terms from the Hubbard model. Their orbital energy corrections take the form

$$\Delta \epsilon_{i}^{\text{LDA+U}} = U \left(\frac{1}{2} - n_{i}\right),$$  \hspace{1cm} (2.9)

where $U$ is analogous to the nearest-neighbor interaction strength from the Hubbard model, which they determine from a constrained DFT calculation. The generalized transition state
(GTS) method put forth in [68], uses Wannier functions to obtain orbital energy corrections using Slater transition theory. The GTS orbital energy correction can be written as

\[ \Delta \epsilon_{GTS}^j(n_j) = \left( \frac{1}{2} - n_j \right) \frac{\partial \epsilon_j(q_j)}{\partial q_j}. \]  

(2.10)

Another group of corrections that attempts to restore the total energy linearity are the Koopman’s compliant functionals. In that theory they do not specifically model the deviation as linear, but instead subtract all non-linear contributions and add back a linear term. There are multiple flavors of Koopman’s compliant functionals, some of which rely on Slater transition theory to approximate the average slope of the non-linear deviation. Even without employing this Slater transition theory, the Koopman’s integral method offers orbital energy corrections similar to Eq. (2.8), see [67]. Most Koopman’s compliant functionals cannot be cast strictly into the form of Eq. (2.8) though because they have an additional term stemming from explicit treatment of the Perdew-Zunger self-interaction correction [59]. The Perdew-Zunger self interaction correction has also inspired another class of corrections that show promise for correcting the DFT deviation from linearity known as the Fermi-Löwdin self-interaction correction (FLO-SIC) [73]. The self-interaction corrections attempt to correct for the mean-field approximation to the Coulomb repulsion in the single-particle Hamiltonian, but do not directly address delocalization error.

As a side note, some literature suggests that any orbital-dependent density functional method can be referred to as DFT+U [8]. We find this convention to be a bit cumbersome since it implies that all orbital-dependent methods derive their corrections from a Hubbard style interaction term. While this may have been a historical precedent for certain orbital-dependent methods, the amount of methods that do not derive their theory based on a connection with the Hubbard model makes this a somewhat spurious naming convention. Some orbital-dependent methods that make no appeal to the Hubbard model include the localized orbital scaling correction [41], Koopman’s compliant functionals [66], and the Perdew-Zunger self-interaction correction [59], the last of which predates DFT+U [74].
2.2 Molecular methods

In order to impose the PPLB condition on global and local regions in a dynamical way, orbitalets \[41\], a set of localized orbitals (LOs) that are localized in both physical and energy spaces, are utilized. For small systems at equilibrium with large orbital energy difference, orbitalets remain canonical orbitals (COs), hence the correction is imposed in global region; while for stretched or big systems with small band gaps, orbitalets become LOs on each fragments, thus the correction is imposed in local regions. This allows LOSC to correct the total energy and orbital energies in a size-consistent manner.

The LOSC correction to total energy \[75\] is formulated as

\[
\Delta E = \frac{1}{2} \sum_{pq} \lambda_{pq} (\delta_{pq} - \lambda_{pq}) \kappa[\rho_p, \rho_q],
\]

(2.11)

where \( \lambda \) and \( \kappa \) are occupation and curvature matrices respectively and \( \tau \) denotes the complex conjugate of \( z \). Since the off-diagonal elements of the \( \lambda \) matrix may be complex values in periodic system calculations using plane-wave basis sets, the LOSC correction to total energy has been slightly modified in Eq. (2.11) to rely on both \( \lambda_{pq} \) and it’s complex conjugate \( \lambda_{pq}^{\ast} \), to guarantee that the correction is a real number. Note this modification of using \( \lambda_{pq} \) makes no difference in the molecular case when only real-valued wavefunctions are considered. The occupation matrix is calculated via \( \lambda_{pq} = \langle \phi_p | \rho_s | \phi_q \rangle \), thus \( \lambda_{pq} = \lambda_{qp} \).

The curvature, \( \kappa[\rho_p, \rho_q] \), takes the form of

\[
\kappa[\rho_p, \rho_q] = (1 - d_{HF}^{\alpha}) \left\{ \int \int dr \, dr' \frac{\rho_p(r) \rho_q(r')}{|r - r'|} - \frac{2\tau C_X}{3} \int dr |\rho_p(r)\rho_q(r)|^{2/3} dr \right\}
\]

(2.12)

where \( \tau = 1.2378 \), \( C_X = \frac{3}{4} (\frac{6}{\pi})^{1/3} \), and \( d_{HF}^{\alpha} \) is the amount of HF exchange energy in the parent functional. Here, a set of LOs and their densities, \( \rho_p(r) = |\phi_p(r)|^2 \), are needed to define the \( \lambda \) and \( \kappa \) matrices.

The desired LOs in LOSC are obtained by unitary transformation upon both occupied and virtual COs, i.e. \( \phi_p(r) = \sum_q U_{pq} \varphi_q(r) \), which are different from LOs of Foster-Boys
localization [76], which only considers mixing occupied or virtual COs amongst themselves. The transformation matrix $U$ is obtained by minimizing the following objective function [75],

$$F = \sum_p \left( \langle \phi_p | r^2 | \phi_p \rangle - |\langle \phi_p | r | \phi_p \rangle|^2 \right) + \sum_{pq} 2\omega_{pq} |\langle \phi_p | \phi_q \rangle|^2,$$

(2.13)

where the first term is for the physical space localization, while the second term for the energy space localization. It should be noted that the penalty function, $\omega_{pq}$ in the energetic cost part, is a monotonically increasing function that depends on CO energy difference, $|\varepsilon_p - \varepsilon_q|$; it thus restricts the mixing of COs with large energy difference. This strategy of combining space and energy costs produces dynamic localization in the orbitalets. The orbitalets thus allow the appropriate amount of correction to the delocalization error in common DFAs at different geometries, which cannot be achieved with a fixed set of atomic orbitals as in LDA+U [77], [78]. In contrast, the generalized transition state (GTS) method [79] and the work by Ma and Wang [80] make use of LOs from mixing of only occupied or virtual COs, thus they cannot change the total energies for physical systems with integer number of electrons; hence these energy functionals are not size consistent and can only correct orbital energies (with failure, e.g. for the $H_2^+$ dissociation).

The LOSC correction to the one-electron Hamiltonian can be obtained by $\delta \Delta E / \delta \rho_s$. For simplicity, we use the frozen-orbital approximation where the contribution from the orbital relaxation is not taken into account. The original work introducing LOSC showed how this actually works quite well [75]. Here, the correction to the Hamiltonian is formulated in a form similar to ref. [81], i.e.

$$\Delta \hat{v} = \sum_{pq} |\phi_q\rangle \Lambda_{pq} \langle \phi_p|,$$

(2.14)

where $\Lambda_{pq} = (\frac{1}{2} \delta_{pq} - \lambda_{pq}) \kappa[\rho_p, \rho_q]$, which depends not only on the density matrix but also on orbitalets. Ideally, the LOSC corrected Hamiltonian should obey the system symmetry and thus

$$[\Delta \hat{v}, \hat{R}_n] = 0$$

(2.15)
for the symmetry operators $\hat{R}_n$ from the symmetry group of the system. Hence, the symmetry of the LOSC Hamiltonian is decided by orbitalets, and the satisfaction of Eq. (2.15) requires that the orbitalets obey

$$\hat{R}_n \{ \phi_p \} \equiv \{ \phi_p \}, \quad (2.16)$$

which means that for each LO with a symmetry operator acting on it, the resulting LO should be able to be reproduced by some LO from the original set of LOs up to a multiplicative constant; see the analysis in Sec. 2.5.4. Eq. (2.16) poses a strict requirement on orbitalets, but as shown in Sec. 2.5.1, a Hamiltonian obeying a subset of the system symmetry can also maintain state degeneracy, thus lowering the symmetry requirements of the orbitalets. Note that normally COs cannot satisfy this requirement.

The corrections to orbital energies are calculated via $\Delta \varepsilon_p = \langle \varphi_p | \Delta \hat{v} | \varphi_p \rangle$ [75]. The LOs obtained by minimizing Eq. (2.13) do not normally obey Eq. (2.16), which thus leads to the destruction of Hamiltonian symmetry and state degeneracy. In perturbation theory, energy level splitting occurs when a perturbation of broken symmetry acts on the system under study, thus the destruction of degeneracy is easy to understand if the LOSC correction, Eq. (2.14), is viewed as a perturbation to the Hamiltonian of the parent DFA.

The aforementioned benefits from using orbitalets and their affect on the Hamiltonian therefore suggest that it is necessary to consider the influence on symmetry when orbitalets are used in orbital-dependent DFT. The symmetry of the Hamiltonian decides the state degeneracy (see Sec. 2.5.1) and is closely related to the Bloch theorem applied in the calculation of periodic systems (see Sec. 2.5.3). Take for example the orbitalets of $\text{H}_2^+$ at different H-H distances using the BLYP functional as shown in Fig. 2.1(a). The orbitalets are the same as the COs at equilibrium bond length, but upon dissociation the orbitalets transform through a series of unsymmetric LOs before turning into two symmetric LOs at the dissociation limit. Thus the original LOSC orbital localization can incorrectly break molecular symmetries and associated state degeneracies.

To address this symmetry issue, we introduced an objective function taking the form
Figure 2.1: LO densities plotted along the bonding axis of H$_2^+$ at the internuclear distances of $R_{HH} = 1.0$, 2.1 and 5.0 Å, respectively for (a) LOSC-BLYP and (b) LOSC2-BLYP. The two H atoms locate at $x = 0.0$ and $x = R_{HH}$. (c) Potential energy curves for H$_2^+$ dissociation. The energy of a doublet H atom is set to zero. Basis set cc-pVTZ [82] is used. All densities and energies are in a.u.

The objective function treats the space and the energy costs on equal footing, where the parameter $\gamma$ allows for a continuous change between spatial localization and energetic localization. When $\gamma = 0$, the resulting orbitalets are maximally localized in physical space but delocalized in physical space; when $\gamma = 1$, then the orbitalets remain COs which are localized in energy space while delocalized in physical space. A similar idea was suggested by Gygi et al. [83] to include the Hamiltonian in the objective function so as to obtain orbitals that are simultaneously localized in space, while still being associated with a well defined region of the eigenvalue spectrum. The desired orbitalets required for LOSC can be obtained by using a
specific value of $\gamma$. As compared to the spatial cost, the energetic cost is much smaller in value and is therefore multiplied by the factor of $C = 10^3$. This factor does not change any results, but it avoids the difficulty of parameter optimization when the desired value of $\gamma$ is too close to the extreme value of 1. We found a value of $\gamma = 0.707$ was optimal to obtain a balanced behavior on IPs and EAs of small sized molecules and polymers, see Sec. 2.5.5 for details. The algorithm of minimizing Eq. (2.17) through a series of $2 \times 2$ rotations is provided in Sec. 2.5.2, which can be efficiently computed through an easily coded parallel implementation.

The form of cost function shown in Eq. (2.17) is able to much more effectively retain system symmetries as shown in the next section. However, it can discontinuously jump between localizations during molecular dissociations. To address this fact, a modified curvature is introduced in LOSC2. In the energy correction of Eq. (2.11), the diagonal terms are positive, which restore the PPLB condition on each local region; while the off-diagonal terms can be negative, which are necessary for retrieving the correct asymptotic behavior when two fragments are far apart[75]. However, the off-diagonal elements of the $\kappa$ matrix that are associated with two LOs on different fragments decay too fast as the fragments move away from each other, thereby the negative terms in Eq. (2.11) decay rapidly, resulting in a sharp upward shift in the dissociation curve; see Fig. 2.1(c). To fix this, the LOSC2 curvature matrix is redefined as

$$
\tilde{\kappa}[\rho_p, \rho_q] = \text{erf}(\tau S_{pq}) \sqrt{\kappa[\rho_p, \rho_p]\kappa[\rho_q, \rho_q]} + \text{erfc}(\tau S_{pq})\kappa[\rho_p, \rho_q],
$$

(2.18)

where $\kappa[\rho_p, \rho_q]$ is from Eq. (2.12), and $S_{pq}$ is calculated via $\int \rho_p(r)\rho_q(r)dr$. This new definition changes the off-diagonal elements while keeping the diagonal elements unchanged, i.e. $\tilde{\kappa}[\rho_p, \rho_p] = \kappa[\rho_p, \rho_p]$, which is essential to guarantee adherence to the PPLB condition on local regions. For an off-diagonal element associated with two LOs on different fragments, $\sqrt{\kappa[\rho_p, \rho_p]\kappa[\rho_q, \rho_q]}$ reduces the decay rate of $\tilde{\kappa}[\rho_p, \rho_q]$. Moreover, due to the error functions in Eq. (2.18), $\tilde{\kappa}[\rho_p, \rho_q]$ would reduce to $\kappa[\rho_p, \rho_q]$ at the dissociation limit, which ensures the correct asymptotic behavior. The parameter $\tau$, which is set at 8.0, allows for adjusting the
rate at which the error functions change, it is optimized for a smooth dissociation curve for 
$H_2^+$. Fig. 2.1(c) shows how the new definition of curvature smooths the $H_2^+$ dissociation 
curve.

We have now constructed the new version of LOSC for molecules, denoted as LOSC2 in this chapter. It mainly has three improvements. First, the correction to total energy has been slightly modified to guarantee that the correction of Eq. (2.11) is a real number and the corresponding correction to the Hamiltonian is Hermitian for complex orbitals. Second, the new objective function in Eq. (2.17) treats spatial and energetic localizations in the same manner so that the orbitalets acquire more symmetry, which is very important for the calculation of the electronic structure for both molecular and bulk systems. Third, the curvature matrix is redefined in Eq. (2.18) to smooth dissociation curves.

2.3 Molecular results

In this section we show how using a combined spatial and energetic objective function allows localized orbitals to acquire more symmetry, which is important in calculating the electronic structure of both molecules and periodic systems. Although the symmetry of the orbitalets obtained from minimizing Eq. (2.17) cannot be strictly derived in general, in the following tests, it will be shown that the new orbitalets are able to maintain more symmetry than their original counterparts.

2.3.1 Hydrogen

The orbitalets from the new localization for $H_2^+$ dissociation can be found in Fig 2.1(b). At both equilibrium bond length and the dissociation limit, the orbitalets are the same as those from the localization of Eq (2.13). The main difference is that orbitalets are now symmetric in the middle range of bond lengths, which indicates that the new objective function of Eq. (2.17) is able to maintain a system symmetry for orbitalets.
2.3.2 Annulenes

Next the LO symmetry and state degeneracy of planar annulenes (C\(_{4n+2}\)H\(_{4n+2}\)) that satisfy the Hückel’s 4\(n\) + 2 rule for aromaticity are checked; see Fig. 2.2. The benzene uses C-C bond distance of \(R_{CC} = 2.0\ \text{Å}\) and planar [10]annulene \(C_{10}H_{10}\) uses \(R_{CC} = 1.5\ \text{Å}\) and both systems are tested with a minimal basis set. For \(C_6H_6\) \((D_{6h}\) symmetry\) at around equilibrium geometry, both LOSC-BLYP and LOSC2-BLYP rarely mix the six \(\pi\)-bond COs (three occupied and three virtual) with each other due to their small spatial variances and large differences in energy. As the C-C bonds stretch, the \(\pi\)-bond COs become larger and their energy difference becomes smaller so that they will mix with each other into fractionally occupied LOs. Fig. 2.2(a) shows the three LOs with larger occupations at \(R_{CC} = 2.0\ \text{Å}\), LOSC-BLYP cannot generate symmetric LOs, thus breaking the degeneracy of highest occupied molecular orbitals (HOMOs), i.e. \(e_{1g}\) states, by around 0.2 eV. The occupation of each CO or LO, i.e. the expectation \(\langle \rho_s \rangle\) is also shown for each orbital, which is proved to be the same for symmetric LOs in Sec. 2.5.4. The details of additional orbitals can be found in Sec. 2.5.5. For LOSC2-BLYP, the LOs possess the symmetry of \(D_{3h}\), i.e. LOs obey Eq. (2.16) for any symmetry operator from \(D_{3h}\), hence the LOSC2 Hamiltonian is of \(D_{3h}\) symmetry too. As shown in Sec. 2.5.1, this lower symmetry of Hamiltonian does not lift the degeneracy of states in \(C_6H_6\), so that LOSC2-BLYP is able to hold the state degeneracy. Different from \(C_6H_6\), the \(\pi\)-bond COs of \(C_{10}H_{10}\) at around equilibrium geometry already have a large spatial variance; thus the COs tend to mix with each other into LOs. Similarly, LOSC2-BLYP can provide symmetric LOs and maintain state degeneracy for \(C_{10}H_{10}\), while LOSC-BLYP cannot; see Fig. 2.2(b). The examples of different sized annulenes shown here are useful for a preliminary understanding of the calculation of bulk systems. In bulk systems, COs become more delocalized over physical space, implying they will inevitably mix with each other into LOs. Hence, LOSC2 is particularly useful because it is able to provide symmetric LOs. For comparison, LOs of Foster-Boys are also plotted, the LOs also show lower symmetries, which are of \(C_{3h}\).
symmetry for C\textsubscript{6}H\textsubscript{6} and of \( C_{5h} \) symmetry for C\textsubscript{10}H\textsubscript{10}. LOSC-BLYP breaks the degeneracy of \( e_{1g} \) (C\textsubscript{6}H\textsubscript{6}), \( e_{1g} \) and \( e_{2u} \) (C\textsubscript{10}H\textsubscript{10}) by around 0.2 eV, while LOSC2-BLYP guarantees the degeneracy of all the states. When cc-pVTZ is used, the energy of C\textsubscript{10}H\textsubscript{10} HOMOs (\( e_{2u} \)) by BLYP is -5.4 eV, which deviates greatly from -8.1 eV by CCSD(T) [84]–[86]. LOSC2-BLYP improves the energy to -8.5 eV.

\[ \text{Figure 2.2: Comparison of LO symmetry and state degeneracy between LOSC-BLYP and LOSC2-BLYP. (a) benzene (C\textsubscript{6}H\textsubscript{6}) and (b) planar [10]annulene (C\textsubscript{10}H\textsubscript{10}). For each system the (1) \( \pi \)-bond of delocalized COs of BLYP and LOs of (2) Foster-Boys, (3) LOSC-BLYP and (4) LOSC2-BLYP are plotted. Each number in parentheses is occupation of each orbital. (5) The degeneracy of a pair of states (\( e_{1g} \)) for C\textsubscript{6}H\textsubscript{6}, and two pairs of states (\( e_{1g} \) and \( e_{2u} \)) for C\textsubscript{10}H\textsubscript{10}; the y-axis represents orbital energy in eV. The dashed lines are just guides for the eye.} \]

\[ \text{2.3.3 Polymers} \]

It has been found that odd numbered allenes possess doubly degenerate HOMOs which are comprised of the right and left-handed helices orientated 90° from each other, hence they are
Misprediction of the degeneracy of the two HOMOs would lead to misjudgment of the molecular chirality. As shown in Fig. 2.3(a), BLYP has no state degeneracy issue, but it overestimates HOMO energies. LOSC-BLYP improves the orbital energies, however, it splits the degenerate HOMOs by around 0.5 eV. In contrast, LOSC2-BLYP improves the orbital energies while guaranteeing the state degeneracy. Another example is vertical IPs of triphenylene; see Fig. 2.3(b). $1e''$ and $2e''$ are two pairs of doubly degenerate states. Both LOSC-BLYP and LOSC2-BLYP improve the prediction of IP values. However, LOSC-BLYP breaks the degeneracy of both $1e''$ and $2e''$, while LOSC2-BLYP is able to maintain the degeneracy. For comparison, Fig. 2.3 provides also the results of LOSC2-B3LYP, which further improves the orbital energies of LOSC2-BLYP. All geometries are optimized by B3LYP.

**Figure 2.3:** Comparison of the energy level and state degeneracy among BLYP, LOSC-BLYP, LOSC2-BLYP and LOSC2-B3LYP. (a) HOMOs of odd numbered alkenes. The y-axis represents orbital energy in eV. The CCSD(T) results are used as reference. (b) Vertical ionization potentials of triphenylene ($D_{3h}$ symmetry) are tested, where $1e''$ and $2e''$ are two pairs of doubly degenerate states. The x-axis represents IPs in eV. Experimental results are from ref. [88].

Additionally, LOSC2 has been tested on a variety of properties. The tests include
reaction barrier heights (on HTBH38/08 and NHTBH38/08 [89]), thermochemistry (on G2-97 [90]), IPs and EAs of small sized molecules (on G2-97 [90]) and polymers (polyacenes and polyacetylenes); a summary of these results can be found in Sec. 2.5.5.

2.4 Summary

In summary, the excellent performance of the LOSC approach reflects the capability of orbitalets for constructing functionals, making it clear that orbitalet symmetry must be considered due to its effect on state degeneracy. To address the issue of broken symmetry and degeneracy in the original version of LOSC, the new localization utilized here allows orbitalets to preserve more symmetry and maintain the desired degeneracy in LOSC2, which is important for electronic structure calculations. Furthermore, LOSC2 outperforms LOSC1 in generating smooth dissociation curves, while maintaining the similar accuracy as LOSC1 over many properties. All the tests here demonstrate the advantage of LOSC2 in the calculation of molecular systems and its potential in the application of bulk systems.

2.5 Supplemental material

2.5.1 Relation between symmetry and degeneracy

All the symmetry operations associated with the geometry of a particular system, \{\hat{S}_k\}, have the property of permuting atoms in identical environment in the system and form a symmetry group (point group or space group). For commonly used density functional approximations (DFAs), such as local density approximation (LDA), generalized gradient approximation (GGA) etc., the one-electron Hamiltonian, \hat{h}, which depends on the density and the geometry (due to nuclear-attraction operator), commutes with symmetry operators from \{\hat{S}_k\}, i.e.

\[ [\hat{S}_k, \hat{h}] = 0 \Rightarrow \hat{S}_k \hat{h} = \hat{h} \hat{S}_k. \]  

(2.19)
Let \( \{ \varepsilon_p \} \) and \( \{ \varphi_p(r) \} \) be the eigenvalues and eigenstates of \( \hat{h} \) respectively. Now applying a symmetry operator, \( \hat{S}_k \), on both side of the eigenequation,

\[
\hat{h}\varphi_p(r) = \varepsilon_p\varphi_p(r),
\]

we obtain

\[
\hat{S}_k\hat{h}\varphi_p(r) = \hat{h}\hat{S}_k\varphi_p(r) = \varepsilon_p\hat{S}_k\varphi_p(r).
\]

Thus, \( \hat{S}_k\varphi_p(r) \) is also an eigenstate of \( \hat{h} \) associated with eigenvalue \( \varepsilon_p \). If \( \varphi_p(r) \) is not a degenerate state, then it can differ from \( \hat{S}_k\varphi_p(r) \) only by a constant factor, i.e. \( \hat{S}_k\varphi_p(r) = C\varphi_p(r) \). If \( \varphi_p(r) \) and \( \varphi_q(r) \) form a pair of doubly degenerate states, then \( \hat{S}_k\varphi_p(r) \) can be obtained by linear combination of \( \varphi_p(r) \) and \( \varphi_q(r) \), i.e.

\[
\hat{S}_k\varphi_p(r) = \varphi_p(r)U^k_1 + \varphi_q(r)U^k_2 \Rightarrow \hat{S}_k[\varphi_p(r), \varphi_q(r)] = [\varphi_p(r), \varphi_q(r)]U^k.
\]

Thus, \( \varphi_p(r) \) and \( \varphi_q(r) \) carry a two-dimensional irreducible representation of \( \{ \hat{S}_k \} \), and \( U^k \) is the matrix representation of the symmetry operator \( \hat{S}_k \) on \( \varphi_p(r) \) and \( \varphi_q(r) \). Similarly, if \( \varphi_p(r) \), \( \varphi_q(r) \), and etc. are \( n \)-fold degenerate states (excluding occasionally degenerate), then they carry a \( n \)-dimensional irreducible representation of \( \{ \hat{S}_k \} \). Therefore, the symmetry of the system decides the degeneracy of states once the relation of Eq. (2.19) holds. Each set of degenerate states corresponds to an irreducible representation of the symmetry group to which the system belongs, with the degeneracy equal to the dimension of the representation.

In order to derive the state degeneracy of the system under study, usually it is not necessary to use the full symmetry group, one of its subgroups is already enough. Taking benzene (C\(_6\)H\(_6\)) as example, the full symmetry group is the point group \( D_{6h} \), see Tab. 2.1 for the character table. It includes 8 one-dimensional representations and 4 two-dimensional representations, where labels A and B are for one-dimensional irreducible representations while E for two-dimensional representations. Regarding the symmetry operators in the tables, \( E \) is identity operation, \( C_n \) is rotation by \( 2\pi/n \) angle, \( I \) is inversion operation, \( S_n \) is improper rotation (rotation by \( 2\pi/n \) angle and reflection in the plane perpendicular to the axis), \( \sigma_h \) is horizontal reflection plane (perpendicular to the principal axis), \( \sigma_v \) is vertical
**Table 2.1**: Character table for $D_{6h}$ point group

<table>
<thead>
<tr>
<th>$D_{6h}$</th>
<th>$E$</th>
<th>$2C_6$</th>
<th>$2C_3$</th>
<th>$C_2$</th>
<th>$3C'_2$</th>
<th>$3C''_2$</th>
<th>$I$</th>
<th>$2S_3$</th>
<th>$2S_6$</th>
<th>$\sigma_h$</th>
<th>$3\sigma_d$</th>
<th>$3\sigma_v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{1g}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$B_{1u}$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$A_{2g}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>$B_{2u}$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$A_{1g}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$A_{2u}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$B_{2g}$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$E_{1u}$</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>-1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$E_{2g}$</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$E_{1g}$</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>-2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$E_{2u}$</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>1</td>
<td>1</td>
<td>-2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 2.2**: Character table for $D_{3h}$ point group

<table>
<thead>
<tr>
<th>$D_{3h}$</th>
<th>$E$</th>
<th>$2C_3'$</th>
<th>$3C_2'$</th>
<th>$2S_3$</th>
<th>$\sigma_h$</th>
<th>$3\sigma_v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1'$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$A_1''$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$A_2'$</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>$A_2''$</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>$E'$</td>
<td>2</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>$E''$</td>
<td>2</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>-2</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 2.3**: Character table for $D_3$ point group

<table>
<thead>
<tr>
<th>$D_3$</th>
<th>$E$</th>
<th>$2C_3$</th>
<th>$3C_2'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$A_2$</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>$E$</td>
<td>2</td>
<td>-1</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 2.4: Correspondence of the irreducible representations among $D_{6h}$, $D_{3h}$ and $D_3$

<table>
<thead>
<tr>
<th>$D_{6h}$</th>
<th>$A_{1g}$</th>
<th>$B_{1u}$</th>
<th>$A_{2g}$</th>
<th>$B_{2u}$</th>
<th>$B_{1g}$</th>
<th>$A_{1u}$</th>
<th>$A_{2u}$</th>
<th>$B_{2g}$</th>
<th>$E_{1u}$</th>
<th>$E_{2g}$</th>
<th>$E_{1g}$</th>
<th>$E_{2u}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{3h}$</td>
<td>$A'$</td>
<td>$A''$</td>
<td>$A'$</td>
<td>$A''$</td>
<td>$E'$</td>
<td>$E'$</td>
<td>$E'$</td>
<td>$E'$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_3$</td>
<td>$A_1$</td>
<td>$A_2$</td>
<td>$A_1$</td>
<td>$A_2$</td>
<td>$E$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

reflection plane (contains the principal axis), $\sigma_d$ is diagonal reflection plane (contains the principal axis and bisects the angle between two $C_2$ axes perpendicular to the principal axis). $D_{3h}$ is a subgroup of $D_{6h}$, see Tab. 2.2 for the character table. By removing those symmetry operators that are not included in $D_{3h}$ from Tab. 2.1, the rest of the table shows that the 12 irreducible representations reduce to the 6 irreducible representations of $D_{3h}$. The correspondence of the irreducible representations between $D_{6h}$ and $D_{3h}$ are shown in Tab. 2.4. Similarly, the 6 irreducible representations of $D_{3h}$ can further reduce to the 3 irreducible representations of $D_3$. Tab. 2.4 shows that all the four two-dimensional representations of $D_{6h}$ reduce to the two-dimensional representations of $D_{3h}$ and $D_3$, which means that the degeneracy is not destroyed. The key symmetry element that guarantees the degeneracy of the states in $C_6H_6$ is $C_3$, which is possessed by all the three point groups mentioned here.

2.5.2 Algorithm for orbitalet localization

The objective function for orbital localization in this work consists of a physical space part and an energy space part, which takes the form of

$$F = (1 - \lambda) \sum_p R_{pppp} + \lambda C \sum_p H_{pppp},$$  \hspace{1cm} (2.23)

where

$$R_{pprs} = \langle \phi_p(1) \phi_r(2) | (r_1 - r_2)^2 | \phi_q(1) \phi_s(2) \rangle,$$  \hspace{1cm} (2.24)
and

\[ H_{pqrs} = \langle \phi_p(1)\phi_r(2)|[h(1) - h(2)]^2|\phi_q(1)\phi_s(2)\rangle. \]  

(2.25)

\( h \) is the one-electron Hamiltonian of the parent DFA. The localized orbitals (LOs) are obtained via

\[ \phi_p(r) = \sum_q U_{pq} \varphi_q(r), \]  

(2.26)

and the U matrix can be obtained by minimizing Eq. (2.23).

Here shows the algorithm of minimizing the objective function of Eq. (2.23), or other objective functions based on \( R_{pppp} \) and \( H_{pppp} \), through a series of \( 2 \times 2 \) rotations, which can be efficiently computed through a easily coded parallel implementation.

1. Initialization:

(a) Initialize LOs:

\[ \phi_p(r) = \varphi_p(r); \]  

(2.27)

(b) Construct some matrices:

\[ \bar{r}_{pq} = \langle \phi_p|\varphi_q\rangle, \]  

(2.28)

\[ \bar{r}^2_{pq} = \langle \phi_p|r^2|\varphi_q\rangle, \]  

(2.29)

\[ \bar{h}_{pq} = \langle \phi_p|h|\varphi_q\rangle = \epsilon_p \delta_{pq}, \]  

(2.30)

\[ \bar{h}^2_{pq} = \langle \phi_p|h^2|\varphi_q\rangle = \epsilon_p^2 \delta_{pq}. \]  

(2.31)

As these matrices are symmetric, so only the elements of the lower triangular matrices are needed to be computed.

(c) Compute the objective function of Eq. (2.23), where \( R_{pqrs} \) of Eq. (2.24) and \( H_{pqrs} \) of Eq. (2.25) can be obtained via

\[ R_{pqrs} = \bar{r}^2_{pq} \delta_{rs} + \bar{r}^2_{rs} \delta_{pq} - 2 \bar{r}_{pq} \bar{r}_{rs}, \]  

(2.32)

\[ H_{pqrs} = \bar{h}^2_{pq} \delta_{rs} + \bar{h}^2_{rs} \delta_{pq} - 2 \bar{h}_{pq} \bar{h}_{rs}. \]  

(2.33)
2. $2 \times 2$ rotations for the $i$-th iteration:

(a) Rotate a pair of LOs, $\phi_p$ and $\phi_q$ with $p > q$, via

\[
\begin{align*}
\phi_p' &= \phi_p \cos(\theta) + \phi_q \sin(\theta), \\
\phi_q' &= -\phi_p \sin(\theta) + \phi_q \cos(\theta),
\end{align*}
\]  
(2.34)

(2.35)

where, $\theta$ decides the rotation matrix

\[
U_{pq} = \begin{pmatrix}
\cos(\theta) & \sin(\theta) \\
-\sin(\theta) & \cos(\theta)
\end{pmatrix}.
\]  
(2.36)

(b) Minimize the objective function. For the objective function of Eq. (2.23), The change caused by the orbital rotation is

\[
\Delta F_{pq}(\theta) = (1 - \lambda)[R_{pppp}(\theta) + R_{qqqq}(\theta)] + \lambda C[H_{pppp}(\theta) + H_{qqqq}(\theta)]
\]  
(2.37)

\[
- (1 - \lambda)[R_{pppp} + R_{qqqq}] - \lambda C[H_{pppp} + H_{qqqq}].
\]

Here

\[
R_{pppp}(\theta) = \langle \phi_p'(1)|\phi_p'(2)(| (r_1 - r_2)^2|\phi_p'(1)\phi_p'(2) \rangle
\]  
(2.38)

\[
= x^0 + x^{c2}\cos(2\theta) + x^{c4}\cos(4\theta) + x^{s2}\sin(2\theta) + x^{s4}\sin(4\theta),
\]

\[
R_{qqqq}(\theta) = \langle \phi_q'(1)|\phi_q'(2)(| (r_1 - r_2)^2|\phi_q'(1)\phi_q'(2) \rangle
\]  
(2.39)

\[
= x^0 - x^{c2}\cos(2\theta) + x^{c4}\cos(4\theta) - x^{s2}\sin(2\theta) + x^{s4}\sin(4\theta),
\]

with

\[
x^0 = \frac{1}{8}(3R_{pppp} + 2R_{pqqq} + 4R_{pqpp} + 3R_{qqqq}),
\]  
(2.40)

\[
x^{c2} = \frac{1}{2}(R_{pppp} - R_{qqqq}),
\]  
(2.41)

\[
x^{c4} = \frac{1}{8}(R_{pppp} - 2R_{pqqq} - 4R_{pqpp} + R_{qqqq}),
\]  
(2.42)

\[
x^{s2} = R_{ppqq} + R_{pqpq},
\]  
(2.43)

\[
x^{s4} = \frac{1}{2}(R_{pqpq} - R_{qqqq}).
\]  
(2.44)
And

\[ H_{pppp}(\theta) = \langle \phi_p'(1)\phi_p'(2)[h(1) - h(2)]^2\phi_p(1)\phi_p(2) \rangle \]  
\[ = y^0 + y_2^c \cos(2\theta) + y_4^c \cos(4\theta) + y_s^2 \sin(2\theta) + y_s^4 \sin(4\theta), \]  
\[ H_{qqqq}(\theta) = \langle \phi_q'(1)\phi_q'(2)[h(1) - h(2)]^2\phi_q(1)\phi_q(2) \rangle \]  
\[ = y^0 - y_2^c \cos(2\theta) + y_4^c \cos(4\theta) - y_s^2 \sin(2\theta) + y_s^4 \sin(4\theta), \]

with

\[ y^0 = \frac{1}{8}(3H_{pppp} + 2H_{ppqq} + 4H_{pqpq} + 3H_{qqqq}), \]  
\[ y_2^c = \frac{1}{2}(H_{pppp} - H_{qqqq}), \]  
\[ y_4^c = \frac{1}{8}(H_{pppp} - 2H_{ppqq} - 4H_{pqpq} + H_{qqqq}), \]  
\[ y_s^2 = H_{pppq} + H_{pqqq}, \]  
\[ y_s^4 = \frac{1}{2}(H_{pppq} - H_{pqqq}). \]

The 4-index matrices \( R_{pqrs} \) and \( H_{pqrs} \) can be easily obtained via Eq. (2.32) and Eq. (2.33), so they don’t need to be stored.

As \( \Delta F_{pq}(\theta) \) is a simple trigonometric function of \( \theta \), it is easy to apply the Newton-Raphson method to find the minimum. We need also the first and second derivatives, i.e.

\[ \frac{d\Delta F_{pq}(\theta)}{d\theta} = \frac{\partial \Delta F_{pq}(\theta)}{\partial R_{pppp}(\theta)} \frac{dR_{pppp}(\theta)}{d\theta} + \frac{\partial \Delta F_{pq}(\theta)}{\partial R_{qqqq}(\theta)} \frac{dR_{qqqq}(\theta)}{d\theta} + \frac{\partial \Delta F_{pq}(\theta)}{\partial E_{pppp}(\theta)} \frac{dE_{pppp}(\theta)}{d\theta} + \frac{\partial \Delta F_{pq}(\theta)}{\partial E_{qqqq}(\theta)} \frac{dE_{qqqq}(\theta)}{d\theta}, \]
and

\[
\frac{d^2 \Delta F_{pq}(\theta)}{(d\theta)^2} = \frac{\partial \Delta F_{pq}(\theta)}{\partial R_{pppp}(\theta)} \left(\frac{d^2 R_{pppp}(\theta)}{(d\theta)^2}\right) + \frac{\partial^2 \Delta F_{pq}(\theta)}{[\partial R_{pppp}(\theta)]^2} \left[ \frac{dR_{pppp}(\theta)}{d\theta} \right]^2 \tag{2.53}
\]

\[
+ \frac{\partial \Delta F_{pq}(\theta)}{\partial R_{qqqq}(\theta)} \left(\frac{d^2 R_{qqqq}(\theta)}{(d\theta)^2}\right) + \frac{\partial^2 \Delta F_{pq}(\theta)}{[\partial R_{qqqq}(\theta)]^2} \left[ \frac{dR_{qqqq}(\theta)}{d\theta} \right]^2
\]

\[
+ \frac{\partial \Delta F_{pq}(\theta)}{\partial E_{pppp}(\theta)} \left(\frac{d^2 E_{pppp}(\theta)}{(d\theta)^2}\right) + \frac{\partial^2 \Delta F_{pq}(\theta)}{[\partial E_{pppp}(\theta)]^2} \left[ \frac{dE_{pppp}(\theta)}{d\theta} \right]^2
\]

\[
+ \frac{\partial \Delta F_{pq}(\theta)}{\partial E_{qqqq}(\theta)} \left(\frac{d^2 E_{qqqq}(\theta)}{(d\theta)^2}\right) + \frac{\partial^2 \Delta F_{pq}(\theta)}{[\partial E_{qqqq}(\theta)]^2} \left[ \frac{dE_{qqqq}(\theta)}{d\theta} \right]^2,
\]

with

\[
\frac{dR_{pppp}(\theta)}{d\theta} = -2x^2\sin(2\theta) - 4x^2\sin(4\theta) + 2x^2\cos(2\theta) + 4x^4\cos(4\theta), \tag{2.54}
\]

\[
\frac{dR_{qqqq}(\theta)}{d\theta} = 2x^2\sin(2\theta) - 4x^2\sin(4\theta) - 2x^2\cos(2\theta) + 4x^4\cos(4\theta), \tag{2.55}
\]

\[
\frac{dH_{pppp}(\theta)}{d\theta} = -2y^2\sin(2\theta) - 4y^2\sin(4\theta) + 2y^2\cos(2\theta) + 4y^4\cos(4\theta), \tag{2.56}
\]

\[
\frac{dH_{qqqq}(\theta)}{d\theta} = 2y^2\sin(2\theta) - 4y^2\sin(4\theta) - 2y^2\cos(2\theta) + 4y^4\cos(4\theta), \tag{2.57}
\]

\[
\frac{d^2 R_{pppp}(\theta)}{(d\theta)^2} = -4x^2\cos(2\theta) - 16x^4\cos(4\theta) - 4x^2\sin(2\theta) - 16x^4\sin(4\theta), \tag{2.58}
\]

\[
\frac{d^2 R_{qqqq}(\theta)}{(d\theta)^2} = 4x^2\cos(2\theta) - 16x^4\cos(4\theta) + 4x^2\sin(2\theta) - 16x^4\sin(4\theta), \tag{2.59}
\]

\[
\frac{d^2 H_{pppp}(\theta)}{(d\theta)^2} = -4y^2\cos(2\theta) - 16y^4\cos(4\theta) - 4y^2\sin(2\theta) - 16y^4\sin(4\theta), \tag{2.60}
\]

\[
\frac{d^2 H_{qqqq}(\theta)}{(d\theta)^2} = 4y^2\cos(2\theta) - 16y^4\cos(4\theta) + 4y^2\sin(2\theta) - 16y^4\sin(4\theta). \tag{2.61}
\]

As all the derivatives above are easy to calculate, so the minimization can be carried out very fast. Besides, to avoid being trapped in a local minimum, several initial guesses are generated from equidistant points in the interval \([0, \pi]\).
(c) Update the pair of LOs, \( \phi_p \) and \( \phi_q \), and some related matrices. The pair of LOs can be updated by

\[
\phi_p \leftarrow \phi_p', \phi_q \leftarrow \phi_q'
\]

(2.62)

where the rotation matrix \( U^{pq} \) of Eq. (2.36) can be obtained by inserting the optimized \( \theta \). Then, the matrices \( \bar{r}, \bar{r}^2, \bar{h}, \) and \( \bar{h}^2 \) are updated by applying the rotation \( U^{pq} \) to rows and columns \((p, q)\) of all the four matrices.

(d) Repeat (a)-(c) for each pair of LOs, and sum over the changes of the objective function from the rotatin of each pair of LOs,

\[
\Delta F = \sum_{q > p} \Delta F_{pq}.
\]

(2.63)

3. Stop if \( \Delta F \) is smaller than a chosen tolerance \( \delta \), otherwise go back to Step 2 with \( i = i + 1 \).

2.5.3 Relation between symmetry and Bloch theorem

The Bloch theorem [91] is very important in the calculation of periodic systems. It states that the wavefunction of each energy eigenstate in a perfectly periodic potential can be written in the form

\[
\varphi_{nk}(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} u_{nk}(\mathbf{r}),
\]

(2.64)

where \( u_{nk}(\mathbf{r}) \) has the periodicity of the lattice, i.e. \( u_{nk}(\mathbf{r} + n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3) = u_{nk}(\mathbf{r}) \), with the three lattice vectors \( \mathbf{a}_1, \mathbf{a}_2, \) and \( \mathbf{a}_3 \). In principle, an infinite number of calculations with different wavevectors, \( \mathbf{k} \), confined to the first Brillouin zone are required. However, since the electronic wavefunctions will be almost identical for \( \mathbf{k} \)-points that are sufficiently close, thus the wavefunctions over a region of reciprocal space can be well represented by the wavefunction at a single \( \mathbf{k} \)-point. Therefore, the Bloch theorem makes the problem of calculating an infinite number of electronic wavefunctions viable by calculating a finite number of electronic wavefunctions over an finite number of \( \mathbf{k} \)-points.
Let $\hat{T}_n$ denote a translation operator that shifts each point in space by the amount of $n_1a_1 + n_2a_2 + n_3a_3$. As all such translation operators commute with each other, so that they have common eigenstates, which are Bloch functions with the form of Eq. (2.64).

For a normal DFA, since the Hamiltonian, $\hat{h}$, is invariant to any of the translations of $n_1a_1 + n_2a_2 + n_3a_3$, it follows that $\hat{h}$ commutes with each of the translation operator $\hat{T}_n$,

$$\hat{h}\hat{T}_n = \hat{T}_n\hat{h}. \tag{2.65}$$

Therefore, the eigenstates of $\hat{h}$ can be chosen to be eigenstates of all $\hat{T}_n$ simultaneously, which thus makes the Bloch theorem applicable.

As LOs from LOSC [75] do not have the translational symmetry, which thus destroys the periodicity of the LOSC Hamiltonian, and the Bloch theorem is not applicable in principle. The examples of $C_{4n+2}H_{4n+2}$ show that the new localization proposed in this work can gain more symmetry for LOs if the space allows. Wavefunctions in bulk systems are normally delocalized over large physical space, thus they tend to mix with each other into LOs that possess the translational symmetry of primitive cell or some supercell, which thus makes the Bloch theorem applicable.

### 2.5.4 Some relations between symmetric localized orbitals

For a set of LOs, $\{\phi_p\}$, they will remain orthonormal after applying any symmetry operator, i.e.

$$\langle \phi_p | \hat{S}_k^\dagger \hat{S}_k | \phi_q \rangle = \langle \phi_p | \phi_q \rangle = \delta_{pq}, \tag{2.66}$$

thus $\hat{S}_k^\dagger \hat{S}_k = E$ and $\hat{S}_k^\dagger = \hat{S}_k^-$. If $\hat{S}_k$ is from the symmetry group of the system under study, and the two LOs, $\phi_p$ and $\phi_q$, obey the symmetry relation,

$$\phi_p = \hat{S}_k \phi_q, \tag{2.67}$$

then for a one-electron Hamiltonian that satisfies Eq. (2.19), the expectations on both LOs obey

$$\langle \phi_p | \hat{h} | \phi_p \rangle = \langle \phi_q | \hat{S}_k^\dagger \hat{h} \hat{S}_k | \phi_q \rangle = \langle \phi_q | \hat{S}_k^\dagger \hat{S}_k \hat{h} | \phi_q \rangle = \langle \phi_q | \hat{h} | \phi_q \rangle. \tag{2.68}$$
Normally, the COs of the same irreducible representation are either all occupied or unoccupied, thus any occupied CO with $\hat{S}_k^\dagger$ applying on it can be obtained by linear combination of the occupied COs, i.e.

$$\hat{S}_k^\dagger[\varphi_1(r), \varphi_2(r), ..., \varphi_N(r)] = [\varphi_1(r), \varphi_2(r), ..., \varphi_N(r)]U^k,$$

(2.69)

where $U^k$ is a unitary matrix. Therefore, the density matrix obeys

$$\hat{S}_k^\dagger\rho\hat{S}_k = \sum_{i=1}^N \hat{S}_{ki}^\dagger\langle \varphi_i | \hat{S}_k \rangle = \sum_{i,j,l=1}^N |\varphi_j\rangle U^k_{ji} U^{k*}_{li} \langle \varphi_l | = \sum_{j=1}^N |\varphi_j\rangle \delta_{jl} \langle \varphi_l | = \sum_{j=1}^N |\varphi_j\rangle \langle \varphi_j | = \rho_s,$$

(2.70)

and the local occupations of two LOs satisfying Eq. (2.67) are the same, i.e.

$$\lambda_{pp} = \langle \phi_p | \rho_s | \phi_p \rangle = \langle \phi_q | \hat{S}_k^\dagger \rho_s \hat{S}_k | \phi_q \rangle = \langle \phi_q | \rho_s | \phi_q \rangle = \lambda_{qq}.$$

(2.71)

### 2.5.5 Supplemental molecular test results

In this work, DFT calculations were performed using an in-house developed QM$^4$D program [92], while coupled cluster calculations using Gaussian09 program [93]. The basis sets used are 6-311++G(3df, 3pd) [94]–[96] for HTBH38/08, NHTBH38/08 [89], [97], and G2-97 test sets [90]; cc-pVTZ [82] for polyacene oligomers, allenes, triphenylene; cc-pVDZ for trans-polyacetylene oligomers. Besides, Minimal basis set is used for orbital density plots for C$_{4n+2}$H$_{4n+2}$ (both C$_6$H$_6$ and C$_{10}$H$_{10}$), while cc-pVTZ is used for IP calculations of C$_{10}$H$_{10}$. For LOSC/LOSC2 calculations, only those COs with energies range from -30 to 10 eV are taken into account for orbital localization and subsequent procedures, which makes the computational cost of the LOSC/LOSC2 corrections much smaller than that of calculating SCF-converged COs.
Table 2.6: HOMO and LUMO energies of polyacene oligomers calculated by different DFT methods. Experimental data are included for comparison. All energies are in eV. Taken from the SI of [45].

<table>
<thead>
<tr>
<th></th>
<th>DFA</th>
<th>LOSC-DFA</th>
<th>nLOSC-DFA</th>
<th>Expt.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LDA</td>
<td>BLYP</td>
<td>B3LYP</td>
<td>LDA</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOMO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-5.68</td>
<td>-5.26</td>
<td>-6.11</td>
<td>-8.61</td>
</tr>
<tr>
<td>3</td>
<td>-5.19</td>
<td>-4.76</td>
<td>-5.54</td>
<td>-7.66</td>
</tr>
<tr>
<td>4</td>
<td>-4.87</td>
<td>-4.44</td>
<td>-5.16</td>
<td>-7.21</td>
</tr>
<tr>
<td>5</td>
<td>-4.66</td>
<td>-4.22</td>
<td>-4.90</td>
<td>-6.82</td>
</tr>
<tr>
<td>6</td>
<td>-4.50</td>
<td>-4.06</td>
<td>-4.71</td>
<td>-6.62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MAE</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUMO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1.38</td>
<td>-0.97</td>
<td>-0.39</td>
<td>0.71</td>
<td>1.12</td>
<td>1.25</td>
<td>0.73</td>
<td>1.11</td>
<td>1.26</td>
<td>1.12</td>
</tr>
<tr>
<td>2</td>
<td>-2.25</td>
<td>-1.82</td>
<td>-1.33</td>
<td>-0.16</td>
<td>0.26</td>
<td>0.31</td>
<td>-0.15</td>
<td>0.28</td>
<td>0.35</td>
<td>-0.20</td>
</tr>
<tr>
<td>3</td>
<td>-2.83</td>
<td>-2.39</td>
<td>-1.97</td>
<td>-0.77</td>
<td>-0.33</td>
<td>-0.37</td>
<td>-0.68</td>
<td>-0.24</td>
<td>-0.33</td>
<td>-0.53</td>
</tr>
<tr>
<td>4</td>
<td>-3.22</td>
<td>-2.76</td>
<td>-2.41</td>
<td>-1.26</td>
<td>-0.78</td>
<td>-0.94</td>
<td>-1.11</td>
<td>-0.65</td>
<td>-0.77</td>
<td>-1.06</td>
</tr>
<tr>
<td>5</td>
<td>-3.48</td>
<td>-3.02</td>
<td>-2.71</td>
<td>-1.59</td>
<td>-1.13</td>
<td>-1.38</td>
<td>-1.38</td>
<td>-0.92</td>
<td>-1.09</td>
<td>-1.39</td>
</tr>
<tr>
<td>6</td>
<td>-3.67</td>
<td>-3.21</td>
<td>-2.93</td>
<td>-1.77</td>
<td>-1.32</td>
<td>-1.52</td>
<td>-1.60</td>
<td>-1.13</td>
<td>-1.32</td>
<td>-1.47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MAE</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ a \] The LOSC-DFA results are from ref. [75].
\[ b \] The experimental values are from refs. [98]–[100].
\[ c \] The mean absolute error for each method as compared to the experimental values.

Table 2.5: Mean absolute deviations (MDAs) computed for atomization energies (in kcal/mol), reaction barriers (in kcal/mol), and IPs and EAs (in eV). Taken from the SI of [45].

<table>
<thead>
<tr>
<th>Test sets</th>
<th>DFA</th>
<th>LOSC-DFA</th>
<th>nLOSC-DFA</th>
<th>Expt.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LDA</td>
<td>BLYP</td>
<td>B3LYP</td>
<td>LDA</td>
</tr>
<tr>
<td>Atomization energies</td>
<td>35.15</td>
<td>4.97</td>
<td>2.70</td>
<td>35.59</td>
</tr>
<tr>
<td>G2-1(55)</td>
<td>89.11</td>
<td>14.07</td>
<td>7.42</td>
<td>88.89</td>
</tr>
<tr>
<td>Non-hydrogen(21)</td>
<td>140.97</td>
<td>10.34</td>
<td>3.46</td>
<td>140.89</td>
</tr>
<tr>
<td>HydroCarbons(17)</td>
<td>118.67</td>
<td>6.23</td>
<td>2.52</td>
<td>118.58</td>
</tr>
<tr>
<td>Inorganic hydrides and radicals(42)</td>
<td>73.01</td>
<td>5.41</td>
<td>2.52</td>
<td>73.42</td>
</tr>
<tr>
<td>mean(148)</td>
<td>81.99</td>
<td>7.28</td>
<td>3.37</td>
<td>82.12</td>
</tr>
<tr>
<td>Reaction barriers</td>
<td>18.31</td>
<td>8.02</td>
<td>4.58</td>
<td>17.41</td>
</tr>
<tr>
<td>HTBH(38)</td>
<td>14.83</td>
<td>10.18</td>
<td>5.42</td>
<td>14.15</td>
</tr>
<tr>
<td>NHTBH(38)</td>
<td>16.57</td>
<td>9.10</td>
<td>5.00</td>
<td>15.78</td>
</tr>
<tr>
<td>mean(76)</td>
<td>18.31</td>
<td>8.02</td>
<td>4.58</td>
<td>17.41</td>
</tr>
<tr>
<td>IPs &amp; EAs</td>
<td>4.46</td>
<td>4.50</td>
<td>3.19</td>
<td>0.50</td>
</tr>
<tr>
<td>IP(64)</td>
<td>3.79</td>
<td>3.40</td>
<td>2.61</td>
<td>0.50</td>
</tr>
<tr>
<td>EA(47)</td>
<td>4.18</td>
<td>4.03</td>
<td>2.94</td>
<td>0.50</td>
</tr>
<tr>
<td>mean(111)</td>
<td>4.46</td>
<td>4.50</td>
<td>3.19</td>
<td>0.50</td>
</tr>
</tbody>
</table>

\[ a \] The number in parentheses indicates the number of systems tested.
\[ b \] Atomization energies are tested on the G2-97 test set. Geometries and experimental data are from ref. [90].
\[ c \] Reaction barriers are tested on the HTBH38/08 and NHTBH38/08 test sets. Geometries and experimental data are from refs. [89], [97].
\[ d \] The testing molecules are from the G2-97 set. \( \text{H}_2\text{S}(^{3}\text{A}_1)\), \( \text{N}_2(^{2}\text{I}_\text{u})\), \( \text{C}_6\text{H}_6\), \( \text{C}_7\text{H}_8\), \( \text{H}_6\text{C}_6\text{O}\) and \( \text{H}_7\text{C}_6\text{O}\) are exclusive from the IP calculations. Reference data are obtained by CCSD(T) calculations and extrapolated to infinite basis limit.
Table 2.7: HOMO energies, $\varepsilon_{\text{HOMO}}$, of trans-polyacetylene oligomers, H(HC=CH)$_n$H, calculated by different DFT methods. RASPT2 and experimental data are included for comparison. All energies are in eV. Taken from the SI of [45].

<table>
<thead>
<tr>
<th>$n$</th>
<th>DFA</th>
<th>LOSC-DFA</th>
<th>nLOS-C-DFA</th>
<th>RASPT2</th>
<th>Expt.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LDA</td>
<td>BLYP</td>
<td>LDA</td>
<td>BLYP</td>
<td>B3LYP</td>
</tr>
<tr>
<td>1</td>
<td>-6.74</td>
<td>-6.44</td>
<td>-7.46</td>
<td>-10.46</td>
<td>-10.10</td>
</tr>
<tr>
<td>4</td>
<td>-5.14</td>
<td>-4.74</td>
<td>-5.56</td>
<td>-7.98</td>
<td>-7.24</td>
</tr>
<tr>
<td>5</td>
<td>-4.97</td>
<td>-4.56</td>
<td>-5.34</td>
<td>-7.33</td>
<td>-6.94</td>
</tr>
<tr>
<td>6</td>
<td>-4.84</td>
<td>-4.43</td>
<td>-5.17</td>
<td>-7.31</td>
<td>-6.92</td>
</tr>
<tr>
<td>7</td>
<td>-4.75</td>
<td>-4.33</td>
<td>-5.05</td>
<td>-7.21</td>
<td>-6.63</td>
</tr>
<tr>
<td>8</td>
<td>-4.68</td>
<td>-4.25</td>
<td>-4.95</td>
<td>-6.95</td>
<td>-6.55</td>
</tr>
<tr>
<td>10</td>
<td>-4.57</td>
<td>-4.14</td>
<td>-4.80</td>
<td>-6.75</td>
<td>-6.35</td>
</tr>
<tr>
<td>$\infty$</td>
<td>-3.52</td>
<td>-3.06</td>
<td>-3.47</td>
<td>-4.91</td>
<td>-4.45</td>
</tr>
</tbody>
</table>

- The second order restrictive active space perturbation theory values are from ref. [101].
- The experimental values for $n = 1 - 4$ are from refs. [102]–[104].
- IPs at $n \rightarrow \infty$ are obtained by extrapolation as shown in Fig. 2.4.
- The mean absolute error for each method as compared to the RASPT2 values.
- The mean absolute error for each method as compared to the experimental values.
Figure 2.4: Extrapolation from calculated HOMO energies for trans-polyacetylene oligomers $H(HC=CH)_nH$ with $n \leq 10$ to $n \to \infty$. As the calculated data for HOMO energies exhibit linear behavior against $1/n^{1/2}$. By fitting the data to straight lines and then extrapolating to $1/n^{1/2} = 0$, the bulk values of IPs may be estimated.
Figure 2.5: \( \pi \)-bond COs and LOs from the mix of them for benzene \((\text{C}_6\text{H}_6 \text{ with } R_{\text{CC}} = 1.50\text{Å} \text{ and } R_{\text{HH}} = 1.09\text{Å}, D_{6h})\). The values in the parentheses are the occupations of COs and LOs, \( \langle \rho \rangle \). Minimal basis set is used.
Figure 2.6: $\pi$-bond COs and LOs from the mix of them for stretched benzene ($C_6H_6$ with $R_{CC} = 2.00\,\text{Å}$ and $R_{HH} = 1.09\,\text{Å}$, $D_{6h}$). The values in the parentheses are the occupations of COs and LOs, $\langle \rho_s \rangle$. Minimal basis set is used.
Figure 2.7: π-bond COs and LOs from the mix of them for overstretched benzene (C_6H_6 with R_{CC} = 5.00 Å and R_{HH} = 1.09 Å, D_{6h}). The values in the parentheses are the occupations of COs and LOs, ⟨ρ⟩. Minimal basis set is used.
Figure 2.8: LOs from the mix of $\pi$-bonding COs by the Foster-Boys localization. Benzenes ($C_6H_6$, $D_{6h}$) with $R_{C-C} = 1.5$, 2.0, and 5.0 Å are tested. The values in the parentheses are the occupations of COs and LOs, $\langle \rho_\pi \rangle$. As only occupied COs are mixed, all the occupations are 1.0. Minimal basis set is used.
Figure 2.9: π-bond COs and LOs from the mix of them for planar [10]annulene (C_{10}H_{10} with R_{CC} = 1.50 Å and R_{HH} = 1.09 Å, D_{10h}). The values in the parentheses are the occupations of COs and LOs, ⟨ρ_s⟩. Minimal basis set is used.
Chapter 3

Dually Localized Wannier Functions

Wannier functions are real-space representations of Bloch orbitals that provide a useful picture for chemical bonding and offer a localized description of single-particle wave functions. There is a unitary freedom in the construction of Wannier functions from Bloch orbitals, which can be chosen to produce Wannier functions that have advantageous properties. A popular choice for this freedom is the minimization of spatial variance, which leads to maximally localized Wannier functions. In this chapter we explore minimizing a weighted sum of the spatial and energy variances, yielding what we call dually localized Wannier functions. Localization in energy allows for unoccupied and occupied bands to be used together since the resulting Wannier functions are correlated with a certain energy. Without localization in energy, all of the resulting Wannier functions will change with the inclusion of additional virtual bands due to the mixing of bands that are far apart in energy. We show how this results in bonding (antibonding) functions for the occupied (unoccupied) spaces around the frontier energy of silicon in the diamond lattice and of molecular ethylene. Dually localized Wannier functions therefore offer a relevant description for chemical bonding and are well suited to orbital-dependent methods that associate Wannier functions with specific energy ranges without the need for explicit energy windowing.

3.1 Background

In the single-particle picture of electronic structure theory, the eigenfunctions of the Hamiltonian may be delocalized in space. This is especially true when the Hamiltonian is periodic, for the solutions are Bloch functions \( \psi_n^{(k)}(r) = e^{ik \cdot r} u_n^{(k)}(r) \), where \( u_n^{(k)}(r) \) shares the periodicity of the Hamiltonian. Bloch functions are indexed by \( k \), a point in reciprocal space, and a band index corresponding to the \( n^{th} \) eigenvalue for that \( k \)-point, i.e., \( H \psi_n^{(k)} = \epsilon_n^{(k)} \psi_n^{(k)} \).
As can be seen from the plane wave $e^{ik \cdot r}$ in their definition, the Bloch orbitals are delocalized in space while being perfectly localized in energy, limiting their utility in describing spatially local properties such as chemical bonds. Dual to the Bloch orbitals are the Wannier functions, which are the Fourier transform of the Bloch orbitals over the Brillouin zone [105]. They are localized in real space, making them useful for evaluating position-dependent quantities such as the dipole moment [106], [107], as a basis set for large-scale simulations [108]–[111], interpolating band structures [112], as a picture of chemical bonding [113], [114], and as localized frontier orbitals in orbital-dependent methods [58], [66], [69], [115].

If a finite number $N_k$ of $k$-points in the first Brillouin zone are sampled, Wannier functions take the form

$$w_n^{(R)}(r) = \frac{1}{N_k} \sum_k e^{-ik \cdot R} \psi_n^{(k)}(r),$$

(3.1)

where $R$ indexes unit cells in the unfolded supercell on which the Wannier functions are periodic. Equal weighting of all $k$-points is equivalent to sampling on a Monkhorst-Pack mesh [116], which we assume throughout the text. This $k$-sampling procedure also yields Wannier functions that are periodic on an unfolded supercell obeying Born–von Karman boundary conditions, with the number of primitive unit cells in each real-space lattice direction equal to the number of $k$-points sampled in the corresponding reciprocal lattice direction [117]. The Fourier transform normalization convention in Eq. (3.1) also implies a Bloch orbital normalization $\langle \psi_n^{(k)} | \psi_m^{(q)} \rangle = N_k \delta_{nm} \delta_{kq}$. The inverse transform from Bloch orbital to Wannier function is then

$$\psi_n^{(k)}(r) = \sum_R e^{ik \cdot R} w_n^{(R)}(r).$$

(3.2)

Since the Bloch orbitals are eigenfunctions they have a phase freedom in their formulation. This leads to a gauge freedom in the construction of the Wannier functions:

$$w_n^{(R)}(r) = \frac{1}{N_k} \sum_k e^{-ik \cdot R} e^{i\theta_n(k)} \psi_n^{(k)}(r).$$

(3.3)
For obtaining localized Wannier functions, a natural choice for this gauge freedom is to choose the phase factors $\theta_n(k)$ such that $\psi_n^{(k)}$ is as smooth as possible. If it is analytic, the resulting Wannier function is exponentially localized [118]. However, multiple bands may cross one another in the Brillouin zone, yielding $k$-points with a set of degenerate Bloch functions. Any unitary combination $U^k$ of the degenerate bands remains a set of eigenfunctions of the Hamiltonian. These freedoms lead to generalized Wannier functions, which allows the bands to mix at a fixed $k$-point by an arbitrary unitary operator $U^{(k)}$ [119],

$$w_n^{(R)}(r) = \frac{1}{N_k} \sum_k e^{-ikr} \sum_m U_{mn}^{(k)} \psi_m^{(k)}(r) = \frac{1}{N_k} \sum_k e^{-ikr} \phi_n^k(r).$$

(3.4)

We call the set $\{\phi_n^k\}$ the transformed Bloch orbitals. Thus, a generalized Wannier function can be comprised of a combination of bands, and the combination may vary across the Brillouin zone.

### 3.1.1 Cost functions

The unitary freedom in the construction of the Wannier functions can be chosen to satisfy a desired criterion. In the case of the maximally localized Wannier functions (MLWFs) proposed by Marzari and Vanderbilt [113], the $U^{(k)}$ minimize the spatial variance

$$\Omega = \sum_n \langle w_n | \Delta r^2 | w_n \rangle,$$

(3.5)

where $\langle f | x | g \rangle = \int_D d\mathbf{r} f^*(\mathbf{r}) x(\mathbf{r}) g(\mathbf{r})$ for any operator $x$ and functions $f, g$. The domain of integration $D$ depends on the periodicity of $f$ and $g$. In the case of the Wannier functions, $D$ is the Born–von Karman unit cell [117]. Additionally, $\langle \Delta x^2 \rangle = \langle x^2 \rangle - |\langle x \rangle|^2$ is the variance of $x$. Wannier functions are translationally symmetric on the primitive unit cell, so without loss of generality we only include the Wannier functions indexed by the home unit cell, $R = 0$. Summing over all unit cells $R$ in the Born–von Karman unit cell would only change
the cost function by a multiplicative constant. The MLWF formulation was originally applied to a composite set of energy bands, which is a set of Bloch orbitals separated by an energy gap from all other states at every point in the Brillouin zone. The gradient of $\Omega$ was shown analytically for a composite set by Marzari and Vanderbilt [113], and descent methods are applied to obtain the $U^{(k)}$ which minimize $\Omega$. Applied to the occupied orbitals of gapped systems with vanishing Chern numbers, MLWFs have been proven to be real and exponentially localized in three and fewer dimensions [118], [120]–[122].

We now introduce the dual localization criteria implemented in this paper that adds an energy variance term to the MLWF cost function:

$$F = (1 - \gamma) \sum_n \langle w_n | \Delta r^2 | w_n \rangle + \gamma \sum_n \langle w_n | \Delta h^2 | w_n \rangle$$

$$= (1 - \gamma) \Omega + \gamma \Xi,$$

(3.6)

where $h$ is the single-particle Hamiltonian, $\Omega$ denotes the spatial spread, and $\Xi$ denotes the energy spread. The mixing term $\gamma$ can be tuned to prioritize spatial or energy localization (see Sec. 3.2.4). This cost function was used in the localized orbital scaling correction (LOSC) [45], which attempts to systemically eliminate delocalization error using localized orbitals [41]. The form of $F$ was first suggested by Gygi et al. in [123]; however, the algorithm there proposed is only applicable to $\Gamma$-sampled systems, where the Brillouin zone is sampled only at the origin. In that case, the transformed Bloch orbitals are equivalent to the Wannier functions, and varying the mixing term $\gamma$ progresses between MLWFs and Bloch orbitals. Additionally, the $\Gamma$-only algorithm does not offer guidance on how to treat both occupied and unoccupied spaces together. We show how using the cost function in Eq. (3.6) can produce physically meaningful localizations when any number of virtual orbitals are included in its domain. The resulting Wannier functions are localized in both space and energy; we therefore refer to them as dually localized Wannier functions (DLWFs).
3.1.2 Occupied and unoccupied spaces

The inclusion of energy localization allows for occupied and unoccupied Bloch orbitals to be considered together for Wannier function construction. When using a cost function that solely minimizes spatial variance, only a set of Bloch orbitals that are close in energy are typically used to construct MLWFs. Since the eigenstates form a complete basis, including infinitely many Bloch bands in the MLWF construction the resulting Wannier functions that are delta functions. In the same way a delta function’s discrete Fourier transform representation is an infinite sum of plane waves, the delta shaped Wannier functions will be comprised of Bloch orbitals that are far apart in energy and therefore have a large energy variance. On the other hand, if energy variance is included in addition to spatial variance, then mixing of Bloch orbitals that are far apart in energy will be suppressed. This means a set of Bloch orbitals spanning a large energy range can be considered and results in Wannier functions that are localized in space as well as associated with a particular energy.

When there is an energy gap between the occupied and unoccupied Bloch orbitals, a set of Wannier functions can be comprised of the occupied Bloch orbitals, and the projector onto the occupied space may be written in the Bloch orbital or Wannier function basis as

\[ P_{\text{occ}} = \frac{1}{N_k} \sum_{n\kappa} f_n^{(k)} |\psi_n^{(k)}\rangle \langle \psi_n^{(k)}| \]

\[ = \sum_{n\mathbf{R}} f_n^{(\mathbf{R})} |w_n^{(\mathbf{R})}\rangle \langle w_n^{(\mathbf{R})}|, \]

(3.7)

where \( f_n^{(k)}, f_n^{(\mathbf{R})} \in \{0, 1\} \) are the occupations of the Bloch orbitals and Wannier functions respectively. When the set of Wannier functions includes both occupied and unoccupied bands, the Wannier functions may not have integer occupancy: that is, \( \langle w_n^{(\mathbf{R})}|P_{\text{occ}}|w_n^{(\mathbf{R})}\rangle \notin \{0, 1\} \). In this case, the occupied projector \( P_{\text{occ}} \) is no longer diagonal in the transformed band index \( n \). If the transformed bands include all of the occupied Bloch orbitals, the
occupied projector can be written using the transformed Bloch orbitals as
\[
P_{\text{occ}} = \frac{1}{N^2} \sum_{m} \sum_{n} \left( \langle \phi_m^{(k)} | P_{\text{occ}} | \phi_n^{(k)} \rangle \right) | \phi_m^{(k)} \rangle \langle \phi_n^{(k)} |
\]
\[
= \frac{1}{N_k} \sum_{mnk} \lambda_{mn}^{(k)} | \phi_m^{(k)} \rangle \langle \phi_n^{(k)} |
\]
(3.8)
where \( | \phi_n^{(k)} \rangle = \sum_m U_{mn}^{(k)} | \psi_m^{(k)} \rangle \) is a transformed Bloch orbital. The \( \lambda_{mn}^{(k)} \) can be viewed as local occupations of the transformed Bloch functions; observe that the matrix \([\Lambda^{(k)}]_{mn} = \lambda_{mn}^{(k)} \) is not necessarily diagonal in \( n \) as it is when the Wannier functions are constructed only from occupied Bloch bands. The trace of \( \Lambda^{(k)} \) is \( N_f^{(k)} \), the number of Bloch states below the Fermi energy at \( k \). For simplicity, we have omitted the spin index, but these arguments also apply to open-shell systems, albeit for each spin channel independently. As the unitary transform of a diagonal matrix with 1’s and 0’s along the diagonal, \( \Lambda^{(k)} \) is a Hermitian matrix with diagonal elements between 0 and 1.

We can also write the occupied projector in terms of the Wannier functions as
\[
P_{\text{occ}} = \sum_{m} \sum_{n} \left( \langle w_m^{(R)} | P_{\text{occ}} | w_n^{(T)} \rangle \right) | w_m^{(R)} \rangle \langle w_n^{(T)} |
\]
\[
= \sum_{RT} \sum_{mn} \lambda_{mn}^{(RT)} | w_m^{(R)} \rangle \langle w_n^{(T)} |
\]
(3.9)
We define the matrix \([\Lambda^{(RT)}]_{mn} = \lambda_{mn}^{(RT)} \), which is the element-wise discrete Fourier transform of the transformed Bloch orbital occupation matrix \( \Lambda^k \). This matrix can be viewed as the occupation matrix of the Wannier functions; for the home cell, it has unit cell trace
\[
\text{Tr}_c[\Lambda^{(00)}] = \sum_n \Lambda_{nn}^{(00)} = N^{-1} k \sum_k N_f^{(k)} .
\]

3.2 Methods

The procedure outlined by Marzari and Vanderbilt in [113] detailed how to find the analytic gradient for minimizing \( \Omega \) for a composite set of energy bands. This was accomplished by splitting the cost function into two positive definite quantities, one of which was
transformation independent, $\Omega_I$, so that minimizing $\Omega$ amounted to only minimizing the gauge-dependent term, $\tilde{\Omega}$. For a set of $N_w$ Wannier functions, the gauge-invariant term of the spatial cost is given by

$$
\Omega_I = \sum_n N_w \left( \langle w_n^{(0)} | r^2 | w_n^{(0)} \rangle - \sum_m \sum_R \left| \langle w_m^{(R)} | r | w_n^{(0)} \rangle \right|^2 \right). \tag{3.10}
$$

The gauge-dependent term for the spatial cost is

$$
\tilde{\Omega} = \sum_{n \neq m} N_k \sum_R \left| \langle w_m^{(R)} | r | w_n^{(0)} \rangle \right|^2 + \sum_n N_k \sum_{R \neq 0} \left| \langle w_n^{(R)} | r | w_n^{(0)} \rangle \right|^2. \tag{3.11}
$$

The valence bands of metals, and the conduction bands of most systems, cannot be formed into a composite set, these types of sets of bands are called entangled. Souza et al. developed a method called disentanglement to extract a subset of interest from a set of entangled bands [112]. Given a $N_b$ number of Bloch bands, the disentanglement procedure extracts a $N_w$-dimensional set of composite bands, where $N_w \leq N_b$, chosen by minimizing $\Omega_I$. The disentangled bands represent the smoothest possible $N_w$-dimensional subspace given the original $N_b$ bands, and are used to construct $N_w$ Wannier functions. To differentiate between disentanglement and localization, we refer to the orbitals obtained by minimizing $\Omega_I$ as disentangled Bloch orbitals, and to those obtained by minimizing a cost function for a fixed number of (possibly disentangled) Bloch bands as transformed Bloch orbitals.

When considering dual energy and spatial localization, we show two possible ways to evaluate the energy of the Wannier functions needed for minimizing the energy variance. First, we show how the energy variance can be evaluated using the true Bloch orbital energies. Second, we demonstrate how using the disentangled Bloch orbital energies creates a simpler and more readily implementable representation. Finally, we detail the regime where these two approaches agree.
3.2.1 Bloch orbital basis

Following the procedure for disentangling a set of Bloch orbitals \( \{ \psi_n^k \} \) outlined by [112] results in a set of disentangled Bloch orbitals \( \{ \varphi_n^k \} \). A disentangled band can be represented in the Bloch orbital basis as

\[
\varphi_n^k = \sum_b N_b V_{nb}^k \psi_b^k,
\]

(3.12)

where \( V^k \) is an isometry of dimension \( N_w \times N_b \). This means it obeys the relation \( V^k (V^k)^\dagger = 1 \), where \( 1 \) is the \( N_w \times N_w \) identity operator. We may then write a transformed Bloch orbital for generalized Wannier function construction in the disentangled Bloch orbital basis as

\[
\phi_n^k = \sum_a N_w X^a U_{na}^k \varphi_a^k.
\]

(3.13)

A generalized Wannier function constructed from a set of disentangled Bloch orbitals then takes the form

\[
w_m^{(R)} = \frac{1}{N_k} \sum_{k} N_b e^{-i k \cdot R} \sum_a N_w \sum_{b} U_{ma}^k \sum_{b} V_{ab} \psi_a^k
\]

\[
= \frac{1}{N_k} e^{-i k \cdot R} \sum_{k} \sum_{b} I_{mb}^k \psi_b^k,
\]

(3.14)

where \( I_{mb}^k = [U^k V^k]_{mb} \).

Using this formulation we may break the energy cost function \( \Xi \) into a transformation invariant term similar to \( \Omega_I \) were the position operator \( \mathbf{r} \) is replaced with \( h \),

\[
\Xi_I = \frac{N_w}{N_k} \sum_n \left( \langle w_n^{(0)} | h^2 | w_n^{(0)} \rangle - \sum_m \sum_R \langle w_m^{(R)} | h | w_n^{(0)} \rangle \right)^2
\]

\[
= \frac{1}{N_k} \sum_{k} \langle \mathbf{R} \rangle \left( \text{Tr} \left[ P^k (H^k)^2 \right] - \text{Tr} \left[ (P^k H^k)^2 \right] \right),
\]

(3.15)

where \( P^k = (V^k)^\dagger V^k \) is the projector from the \( N_b \) Bloch orbitals to the \( N_w \) disentangled Bloch orbital space. If the commutator \( [P^k, H^k] = 0 \), then by the cyclic property of the
trace $\Xi_1 = 0$. This happens when the disentangled Bloch orbitals are a subset of the true
Bloch orbitals, or for any isometric mixing of degenerate Bloch orbitals at a fixed $k$-point.

### 3.2.2 Disentangled orbital basis

The energy spread cost function may also be evaluated using the energy of the disentangled
Bloch orbitals. In the procedure outlined in [112], the energy of the disentangled subspace
is diagonalized in the space spanned by the disentangled Bloch orbitals. This means that

\[
\left[ \tilde{H}^k \right]_{nm} = \langle \tilde{\varphi}_n^k | \tilde{h} | \tilde{\varphi}_m^q \rangle = N_k \delta_{kq} \delta_{nm} \tilde{\epsilon}_n^k,
\]

(3.16)

where $\tilde{H}^k = V^k H^k (V^k)\dagger$. If we replace the full Hamiltonian $h$ by the disentangled Hamiltonian $\tilde{h}$, then the projector $P^k$ in Eq. (3.15) becomes the identity matrix and $\Xi_1$ vanishes. Therefore, even if a disentanglement leads to a nonzero $\Xi_1$, using $\tilde{h}$ instead of the full $h$ in the definition of $\Xi$ yields $\Xi_1 = 0$. Using the disentangled Hamiltonian also means the $\tilde{h}^2$ terms are unitarily invariant:

\[
\sum_{n}^{N_w} \langle w_n^0 | \tilde{h}^2 | w_n^0 \rangle = \frac{1}{N_k} \sum_{k} \text{Tr} \left[ (\tilde{H}^k)^2 \right].
\]

(3.17)

This allows us to break the energy cost term into the squared average energy term, $\Xi_{SA} = \sum_n \langle w_n | \tilde{h}^2 | w_n \rangle$, and an average energy squared term, $\Xi_{AS} = \sum_n |\langle w_n | \tilde{h} | w_n \rangle|^2$, such that the total energy spread cost is $\Xi = \Xi_{SA} - \Xi_{AS}$. Similar to the spatial variance in the $\Gamma$ sampled case [123], this means that minimizing the energy spread is equivalent to maximizing the average energy squared, $\Xi_{AS}$.

The difference between using the projected subspace Hamiltonian $\tilde{h}$ and the full Hamiltonian $h$ depends on the set of isomety matrices, $\{V^k\}$. If the disentangled multiset of eigenvalues is a subset of the Bloch orbital eigenvalues at every $k$-point, $\{\tilde{\epsilon}_n^k\} \subseteq \{\epsilon_n^k\}$, then $[\tilde{H}^k, H^k] = 0$ and using $\tilde{h}$ is equivalent to using $h$ in the energy cost function. A multiset is a set that accounts for multiplicity, which is necessary to account for degenerate eigenvalues. We find that using the existing disentanglement procedure creates a disentangled
band structure that is almost equivalent to the full band structure for a given area of interest. For example, we can find Wannier functions around the Fermi energy by including a sufficiently large number of unoccupied bands above the Fermi energy when applying the disentanglement procedure. This produces a disentangled band structure that only appreciably differs from the original band structure far away from the Fermi energy. This is because the disentanglement procedure implemented in wannier90 diagonalizes \( \tilde{h} \) at the end of each iteration in the minimization \[112\]. If the projected subspace fully spans the area of interest around the Fermi energy, then the eigenvalues of \( \tilde{h} \) will correspond to those of \( h \) near that energy value. We do note that it is possible to include \( \Xi_I \) in the disentanglement procedure, so that the resulting disentangled eigenspectrum would more closely resemble the full eigenspectrum for all energy values.

### 3.2.3 Energy cost gradient

Following the breakdown of the cost function into \( \Xi_{SA} \) and \( \Xi_{AS} \) as outlined in the previous section, we now derive the gradient of \( \Xi_{AS} \) with respect to the unitary rotation \( U^k \). We approximate \( U^k \) to first order using a small anti-Hermitian matrix, \( U^k \approx 1 + dW^k \), with \((dW^k)^\dagger = -dW^k\). Then we find the derivative with respect to \( U^k \) using the matrix calculus identity

\[
\frac{d \text{Re} \left\{ \text{Tr} \left[M \, dW \right] \right\}}{dW} = \frac{1}{2} (M - M^\dagger) = A[M].
\] (3.18)

To see why this is useful, observe that we may write \( \Xi_{AS} \) as

\[
\Xi_{AS} = \sum_n \left| \langle w_n | \tilde{h} | w_n \rangle \right|^2
\]

\[
= \sum_n \left| \frac{1}{N_k} \sum_{m,k} |U^k_{mn}|^2 \tilde{\epsilon}_{mn} \right|^2
\]

\[
= \sum_n \left| \frac{1}{N_k} \sum_k B^k_{mn} \right|^2
\] (3.19)
where \([B^k]_{mn} = [U^k \tilde{H}^k(U^k)^\dagger]_{mn}\) is the Hamiltonian in the transformed Bloch orbital basis. After transforming the Bloch functions at each \(k\)-point by \(U^k(\mathbf{k}) \approx \mathbb{1} + dW^k(\mathbf{k})\), the change in \(\Xi_{\text{AS}}\) to first order in \(dW^k\) is

\[
d\Xi_{\text{AS}} = 4 \frac{N_k}{N_k^2} \sum_n \left( \sum_{\mathbf{k}} \text{Re} \left\{ [B^k dW^k]_{nn} \right\} \right) \left( \sum_q B^{(q)}_{mn} \right)
\]

where \(C^{(k,q)}_{mn} = B^k_{mn} B^{(q)}_{mn}\). This allows us to write the gradient in terms of \(A\), defined in Eq. (3.18), as

\[
\frac{d\Xi_{\text{AS}}}{dW^k} = \frac{2}{N_k} \sum_q \text{Tr}[C^{(k,q)} dW^k]
\]

In this form it appears the gradient at each \(k\)-point depends on every other \(k\)-point, which would indicate \(N_k^2\) scaling with the number of \(k\)-points. However, we can sum the matrix \(C^{(k,q)}\) over the Brillouin zone, removing the dependence on \(q\) prior to evaluating the gradient. Thus

\[
\left[ \frac{d\Xi_{\text{AS}}}{dW^k} \right]_{mn} = 2B^k_{mn} (E_m - E_n),
\]

where \(E_m = \langle w_m^{(0)} | \hat{h} | w_m^{(0)} \rangle\) is the average energy of a Wannier function. In essence, these terms penalize the mixing of Bloch orbitals proportionally to the difference in average energy of the Wannier functions they are used to construct. This also means the computational cost of constructing an element of \(d\Xi_{\text{AS}}/dW^k\) is independent of \(N_k\).

### 3.2.4 Mixing parameter

We now remark on the mixing parameter \(\gamma\) introduced in Eq. (3.6). Setting \(\gamma = 0\) recovers the MLWF cost function \(\Omega\). Setting \(\gamma = 1\) returns the Bloch orbitals in the case of \(\Gamma\)
sampling. However, in the \( \mathbf{k} \)-sampled case \( \gamma = 1 \) simply returns the Bloch orbitals in order of energy. When there are band crossings in the Brillouin zone, the energy-ordered Bloch orbitals will not be smooth in \( \mathbf{k} \), giving poor spatial localization. Choosing \( \gamma \) strictly between 0 and 1 provides Wannier functions localized in both space and energy.

In a related work for molecules, this cost function was used in an orbital-dependent density functional method called the localized orbital scaling correction (LOSC) [45]. In that work, \( \gamma \) was chosen to minimize the error in several experimentally realizable quantities in a test suite of molecules. In Å and eV, the units of space and energy used in \texttt{wannier90}, this value is \( \gamma = 0.47714 \). For details on this value in relation to the value published in [45], see the Supplemental Material. When using localized orbitals in the context of LOSC, we refer to them as \textit{orbitalets}, a portmanteau of orbital and wavelet referring to their compromise of spatial and energy localization. One of the key differences between MLWFs and DLWFs is that DLWFs naturally allow the inclusion of unoccupied orbitals in their construction. Energy cutoffs must be enforced manually for MLWFs, because including high-energy bands will result in unphysically localized Wannier functions; indeed, including the full space of occupied and unoccupied Bloch orbitals is expected to yield Dirac delta distributions as MLWFs. By contrast, for suitable values of \( \gamma \), the energy localization of DLWFs induce convergence to stable frontier orbitals when enough higher-energy orbitals are included. When \( \gamma = 0.47714 \), we find that Bloch orbitals can mix substantially in the construction of DLWFs at a given \( \mathbf{k} \)-point as long as their energy difference is less than about 2 eV.

### 3.2.5 Computational details

Following the MLWF formulation of Marzari and Vanderbilt [113], we compute the gradient of the cost function \( F \) at each \( \mathbf{k} \)-point. Given an initial guess or disentanglement, either a conjugate gradient or steepest descent algorithm is used to minimize \( F \). We implement the energy-space localization in a fork of the open-source \texttt{wannier90} code [124]–[126]. In order
to account for $\Xi$, some modifications to the descent algorithm were required. The steepest
descent portion of the algorithm was left unchanged, but we found that the step size that
produced the best localization was system-dependent. To account for this, we sweep a
range of step sizes in order to obtain the best minimum for each localization. For steps
when the conjugate-gradient descent and parabolic line search were used, we found that the
Polak-Ribiere coefficient [127] provides better convergence than the default Fletcher-Reeves
coefficient [128], [129].

Because we allow the inclusion of unoccupied orbitals, we also use the disentanglement
procedure on the highest-energy conduction bands considered for localization [112]. When
including virtual bands, the choice of how many bands and DLWFs to include is increased
until the orbitals of interest, typically the ones around the Fermi energy, are converged.
Convergence of the higher level virtuals was found to be problematic. To side step this
issue, we allow for the convergence criteria to only look at the cost of the Wannier functions
below a fixed energy level. We consider the DLWFs converged if their cost from Eq. (3.6)
does not appreciably change with the addition of more unoccupied bands. More details on
controlling the functionality added to the wannier90 code are in the Supplemental Material.

3.3 Results

In the following section we obtain Bloch functions with the PBE density functional [25]
in the plane-wave basis, using the optimized norm-conserving Vanderbilt pseudopotentials
[130] generated from PseudoDojo [131]; calculations are performed using the open-source
Quantum ESPRESSO code suite [132], [133]. The Brillouin zone is always sampled with
Monkhorst-Pack meshes [116] centered at $\Gamma$, the origin of reciprocal space. For silicon
and ethylene, we use a wave function kinetic energy cutoff $E_{\text{cut}} = 60$ Ry; for copper, we
set $E_{\text{cut}} = 100$ Ry. In all cases $E_{\text{cut}}$ for the density is four times that used for the wave
functions. As mentioned in Sec. 3.2, we use a modified version of the wannier90 code
[124]–[126] to obtain the DLWFs. We select the LOSC mixing parameter $\gamma = 0.47714$,
unless stated otherwise. Since the DLWFs are not always real, isoplots are of the DLWF densities, $|w_n(r)|^2$, rather than of the DLWFs themselves. All lengths are reported in Å and all energies in eV.

### 3.3.1 Silicon

First we consider the well-studied semiconductor silicon in the diamond lattice. For the self-consistent calculation we use an $8 \times 8 \times 8$ k-mesh; for computational efficiency, we use a $4 \times 4 \times 4$ k-mesh for obtaining virtual Bloch orbitals and localization. We use the experimental lattice parameter $a_{\text{lat}} = 5.431$ Å [134]. The gap of Si is small, with the self-consistent calculation yielding 0.71 eV. It is an indirect gap, however, and the smallest direct gap in the Brillouin zone is 2.56 eV at the Γ point.

#### Occupied states

When only considering the occupied states with the LOSC cost function, we find that the orbitalets are not degenerate in shape; this contrasts with the MLWF procedure, which yields four degenerate Wannier functions. The lowest-energy DLWF has a tetrahedral shape; the two highest-energy orbitalets are degenerate, each having a three-lobed shape with each lobe centered along a bond. The MLWFs, on the other hand, all have tight-binding character with a single lobe centered along the four Si–Si bonds.

#### Frontier states

We expect the valence bands of semiconductors to yield degenerate MLWFs that approximate bonding molecular orbitals; localizing the same number of low-lying conduction bands with the MLWF procedure often yields degenerate functions of antibonding character. Since the conduction states are entangled with higher-energy bands, Souza, Marzari, and Vanderbilt [112] disentangled 12 bands down to 8 when investigating silicon. Here, we use the same disentanglement with a frozen disentanglement window at the Fermi energy, 6.23
eV. Using Eq. (3.6) for the cost function on this subspace, we obtain occupied orbitals very similar to those found in Sec. 3.3.1 above. The conduction orbitals, however, form a fourfold degenerate anti-bonding set, qualitatively equivalent to MLWFs when considering only the four lowest-lying conduction bands.

We also use this system to explore the effects of varying values of $\gamma$ in the cost function. Allowing both valence and conduction bands to mix in the MLWF construction yields eight Wannier functions degenerate in energy. As we increase $\gamma$, we observe WFs more closely associated with specific energy ranges and stepwise lifting of degeneracy (see Fig. 3.1), but spatial localization is suppressed at the same time (see Fig. 3.2).

**Figure 3.1:** Degeneracy pattern of the average energy $\langle \tilde{h} \rangle$ for the DLWFs in silicon against $\gamma$ from Eq. (3.6). The color and line type indicate the degeneracy of the DLWFs, where degenerate is taken as $|\langle \tilde{h} \rangle_n - \langle \tilde{h} \rangle_m| < 0.15$ eV. At $\gamma = 0$ (MLWFs), we obtain degenerate $sp^3$ hybrid orbitals, with each MLWF having occupancy 0.5 and $\langle h \rangle = 6.3$ eV. When $0.02 < \gamma < 0.1$, the DLWFs split into two degenerate sets: the occupied and unoccupied orbitals. Degeneracy of unoccupied DLWFs is lifted at $\gamma = 0.53$, while the first lifting of degeneracy for occupied DLWFs occurs for $\gamma \approx 0.1$. 

71
Converged frontier states

Next, we show convergence results when higher-energy virtual bands are included. For this result, we find that the spatial variance of the highest-energy occupied DLWF is well converged when 34 bands are disentangled to 30. Notably, we find that the shapes of the resulting DLWFs are qualitatively the same as those found when including only occupied orbitals (only the frontier orbitals), as in Sec. 3.3.1 (Sec. 3.3.1). We note, however, that the spatial variance $\Delta r^2$ of the lowest unoccupied DLWFs is larger than when only four unoccupied bands are included. This is due to the fact that the disentanglement in the latter case directly modifies the bands comprising the lowest unoccupied DLWFs, unphysically smoothing them and increasing their localization. Including more virtual orbitals in the localization procedure means that disentanglement smooths only the high-energy conduction bands and removes this effect. For isosurface plots of the eight lowest-energy DLWF densities, see Fig. 3.3. For more data on individual DLWFs, see the Supplemental
3.3.2 Copper

We next investigate copper in an face centred cubic lattice using an experimental lattice parameter of 3.614 Å [135]. For the self-consistent calculation we use a 16×16×16 k-
mesh, while for the virtual states and localization we use a $10 \times 10 \times 10$ k-mesh. Since the unoccupied bands of copper are very steep, and energy localization restricts mixing bands far apart in energy, we do not require many bands above the Fermi level. Noting that there are 9.5 electrons per spin channel, we use 23 bands and disentangle to 17 bands with a frozen disentanglement window of 40 eV, well above the Fermi energy (17.04 eV). This results in a set of disentangled bands that is only different from the parent DFA at the top of the energy window. Using the $\gamma = 0.47714$, we find that the fully occupied orbitals self-organize into pure s, p, and d-type DLWFs, (see Fig. 3.4). The partially occupied DLWF associated with the frontier band is a delocalized orbital with d-type characteristics on multiple atoms. The individual DLWF information is shown in the Supplemental Material.

### 3.3.3 Ethylene

To demonstrate the utility of dual localization in molecules, we simulate ethylene with the geometry from [113]. We use a 10 Å unit cell with the molecule centered at the origin. We sample only the Γ point for to approximate the isolated molecule case. Thus, disentanglement is unnecessary, and we construct 36 DLWFs from 36 bands, 6 of which are occupied. Setting $\gamma = 0.47714$ recovers DLWFs closely related to molecular orbitals. The lowest-energy DLWF corresponds to a σ bond between the carbon atoms, and the next four resemble combinations of C–C and C–H bonding orbitals. The highest-energy occupied DLWF corresponds to a π bonding orbital, and the lowest unoccupied DLWF to a π* antibonding orbital. Their isoplots are shown in Fig. 3.5, and the individual DLWF data are in the Supplemental Material.

### 3.4 Summary

We have shown that including energy localization in the construction of Wannier functions results in DLWFs localized both in space and in energy. This allows for bands widely separated in energy to be included in the Wannier function construction without allowing
Figure 3.4: DLWF density isoplots of copper at isovalue 0.6 using $\gamma = 0.47714$. These are the 6 highest-energy occupied DLWFs labeled in ascending order of their average energy starting with the 5th lowest energy. The four lowest DLWFs correspond to the occupied 3s and 3p states (not shown). The 10th orbital is only partially occupied, as it is mainly comprised of the partially occupied frontier Bloch band. More atoms are shown for the isoplot of DLWF$_{10}$ due to its delocalized nature.
Figure 3.5: DLWF density isoplots for ethylene using $\gamma = 0.47714$. DLWF$_6$ is the highest-energy occupied DLWF corresponding to a $\pi$ bonding orbital and DLWF$_7$ is the lowest-energy unoccupied DLWF corresponding to a $\pi^*$ antibonding orbital. DLWF$_1$ uses an isovalue of 160, DLWF$_2$–LWF$_7$ use an isovalue of 60, and DLWF$_8$ uses an isovalue of 30. The isosurface between DLWF$_6$ and DLWF$_8$ belongs to DLWF$_8$. 
the unphysical mixing of bands that are far apart in energy; furthermore, the localization procedure naturally produces orbitals associated with certain energies, organizing itself along the Hamiltonian’s spectrum. The DLWFs with occupation close to 1 also retain the behavior seen in MLWFs, appearing as tight-binding or atomic orbitals and retaining the symmetry of the underlying system. Using the LOSC cost function, $\gamma = 0.47714$, the localization identifies atomic states such as the $d$ orbitals of copper (figure 3.4). The LOSC cost function also recovers molecular orbitals such as the ethylene $\sigma$ and $\pi$ bonding orbitals in Fig. 3.5. This shows that DLWFs have the potential to produce chemically relevant frontier orbitals automatically, without the need for manually enforced energy cutoffs or for separating the occupied and unoccupied spaces.

The fact that DLWFs offer a frontier orbital description localized in space means that they can be employed for orbital-dependent DFT methods [45], [58], [66], [69], [115]. Many of these methods attempt to correct the well-known failure of density functional approximations to give energy piecewise linear in fractional occupation [27]. In order to correct this nonlinear behavior, many of these corrections invoke localized frontier orbitals. A popular choice for these orbitals is to use MLWFs constructed from the highest-energy composite set of occupied Bloch bands. As DLWFs are explicitly associated with an energy level, they offer a localized charge description of frontier orbitals for both the occupied and unoccupied spaces. This means they can also generate relevant Wannier functions for both occupied and unoccupied frontier orbitals without having to localize each separately.

When the Wannier function construction includes entangled bands, Damle and Lin [136] have questioned whether sequential disentanglement and localization can find the globally optimal solution as separate procedures. The unified localization method proposed therein offers a possible solution to this problem. For a composite set of bands, this method reduces to the selected columns of the density matrix method [137], [138]; in that case, as with the original MLWF construction, energy cutoffs can be enforced explicitly to limit the energy spread of the Wannier functions. However, because the unified method of Damle and
coworkers relies on the exponential decay of off-diagonal elements of the one-particle density matrix in space, it is unclear whether it can be adapted to include energy localization in the construction of the Wannier functions while allowing the inclusion of Bloch orbitals widely separated in energy, as shown in this work.

Finally, we comment on the (lack of) reality of the DLWFs. When $\gamma = 0$, so that $F = \Omega$ in Eq. (3.6), Marzari and Vanderbilt conjectured in [113] that the resulting maximally localized Wannier functions would be purely real (up to a global phase). This was proven in [121] for Wannier functions constructed by minimizing any functional symmetric under time reversal, provided that the optimum is unique. Our cost function $F$ obeys time-reversal symmetry, and numerical tests show that the DLWFs with an occupation of 1 are real. However, we observe substantial imaginary character in DLWFs that have an occupation numerically distinguishable from 1. Even though the arguments in [121] do not differentiate between the occupied and unoccupied spaces we conjecture that the reality arguments do not apply to Wannier functions that are not occupied, i.e. have an occupation that is not 1. It is possible that the non-uniqueness of these minima could account for their lack of reality.

3.5 Supplemental material

3.5.1 Appendix of tables

This section shows the spatial center, spatial variance, average energy, energy variance, and occupation of the individuals DLWFs presented in the results section of the main text. The index is ordered from lowest to highest energy, starting at 1; thus, it corresponds to the labels of DLWF density isoplots in the main text.
Table 3.1: Spatial and energy information per DLWF for silicon.

<table>
<thead>
<tr>
<th>index</th>
<th>$\langle r_x \rangle$</th>
<th>$\langle r_y \rangle$</th>
<th>$\langle r_z \rangle$</th>
<th>$\langle \Delta r^2 \rangle$</th>
<th>$\langle h \rangle$</th>
<th>$\langle \Delta h^2 \rangle$</th>
<th>$\lambda_{nn}^{(00)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>-0.7273</td>
<td>-0.6394</td>
<td>-0.7274</td>
<td>2.357579</td>
<td>-3.234141</td>
<td>1.681126</td>
<td>0.999992</td>
</tr>
<tr>
<td>02</td>
<td>0.2355</td>
<td>0.5311</td>
<td>0.2360</td>
<td>4.824702</td>
<td>0.528461</td>
<td>3.594108</td>
<td>0.999453</td>
</tr>
<tr>
<td>03</td>
<td>-0.3476</td>
<td>-1.2893</td>
<td>-1.5917</td>
<td>4.296241</td>
<td>3.738779</td>
<td>1.693098</td>
<td>0.997848</td>
</tr>
<tr>
<td>04</td>
<td>-1.5941</td>
<td>-1.2878</td>
<td>-0.3466</td>
<td>4.294985</td>
<td>3.738866</td>
<td>1.692406</td>
<td>0.997860</td>
</tr>
<tr>
<td>05</td>
<td>-0.0505</td>
<td>2.6608</td>
<td>2.7073</td>
<td>4.845865</td>
<td>10.548782</td>
<td>4.084945</td>
<td>0.001534</td>
</tr>
<tr>
<td>06</td>
<td>1.2678</td>
<td>0.0730</td>
<td>1.2995</td>
<td>4.850359</td>
<td>10.569209</td>
<td>4.085532</td>
<td>0.001481</td>
</tr>
<tr>
<td>07</td>
<td>1.4140</td>
<td>-1.3094</td>
<td>-2.7687</td>
<td>4.916316</td>
<td>10.693759</td>
<td>4.064443</td>
<td>0.000875</td>
</tr>
<tr>
<td>08</td>
<td>0.0238</td>
<td>1.3053</td>
<td>1.3575</td>
<td>4.907699</td>
<td>10.71160</td>
<td>4.088860</td>
<td>0.000869</td>
</tr>
<tr>
<td>09</td>
<td>-0.0182</td>
<td>0.0123</td>
<td>0.0392</td>
<td>8.239568</td>
<td>15.653925</td>
<td>3.218761</td>
<td>0.000047</td>
</tr>
<tr>
<td>10</td>
<td>-0.8270</td>
<td>-0.8582</td>
<td>-1.3807</td>
<td>8.298693</td>
<td>15.739891</td>
<td>2.969253</td>
<td>0.000037</td>
</tr>
</tbody>
</table>

Table 3.2: Spatial and energy information per DLWF for copper.

<table>
<thead>
<tr>
<th>index</th>
<th>$\langle r_x \rangle$</th>
<th>$\langle r_y \rangle$</th>
<th>$\langle r_z \rangle$</th>
<th>$\langle \Delta r^2 \rangle$</th>
<th>$\langle h \rangle$</th>
<th>$\langle \Delta h^2 \rangle$</th>
<th>$\lambda_{nn}^{(00)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>-0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.155525</td>
<td>-95.528687</td>
<td>0.000018</td>
<td>1.000000</td>
</tr>
<tr>
<td>02</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.186959</td>
<td>-52.928957</td>
<td>0.000814</td>
<td>1.000000</td>
</tr>
<tr>
<td>03</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.186943</td>
<td>-52.928955</td>
<td>0.000811</td>
<td>1.000000</td>
</tr>
<tr>
<td>04</td>
<td>0.0000</td>
<td>0.0000</td>
<td>-0.0000</td>
<td>0.186958</td>
<td>-52.928954</td>
<td>0.000811</td>
<td>1.000000</td>
</tr>
<tr>
<td>05</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.186958</td>
<td>-52.928954</td>
<td>0.000811</td>
<td>1.000000</td>
</tr>
<tr>
<td>06</td>
<td>-0.0038</td>
<td>-0.0046</td>
<td>-0.0073</td>
<td>0.705379</td>
<td>14.072693</td>
<td>1.235276</td>
<td>0.996415</td>
</tr>
<tr>
<td>07</td>
<td>-0.0063</td>
<td>0.0014</td>
<td>-0.0177</td>
<td>0.793226</td>
<td>14.078952</td>
<td>1.528006</td>
<td>0.996099</td>
</tr>
<tr>
<td>08</td>
<td>-0.0068</td>
<td>-0.0050</td>
<td>0.0046</td>
<td>0.709318</td>
<td>14.085303</td>
<td>1.248489</td>
<td>0.996337</td>
</tr>
<tr>
<td>09</td>
<td>0.0096</td>
<td>0.0045</td>
<td>0.0013</td>
<td>0.803549</td>
<td>14.121215</td>
<td>1.528358</td>
<td>0.996831</td>
</tr>
<tr>
<td>10</td>
<td>1.0363</td>
<td>0.9041</td>
<td>-0.8420</td>
<td>4.926801</td>
<td>16.985522</td>
<td>7.848977</td>
<td>0.516482</td>
</tr>
<tr>
<td>11</td>
<td>0.6418</td>
<td>-1.0563</td>
<td>-0.3213</td>
<td>7.008280</td>
<td>26.141562</td>
<td>16.535585</td>
<td>0.000001</td>
</tr>
<tr>
<td>12</td>
<td>-0.4665</td>
<td>-0.0404</td>
<td>1.4063</td>
<td>7.196124</td>
<td>32.475400</td>
<td>19.176457</td>
<td>0.000000</td>
</tr>
<tr>
<td>13</td>
<td>-0.4218</td>
<td>1.0018</td>
<td>1.0964</td>
<td>7.318156</td>
<td>32.739401</td>
<td>20.308781</td>
<td>0.000000</td>
</tr>
<tr>
<td>14</td>
<td>0.0123</td>
<td>0.0419</td>
<td>-0.0969</td>
<td>12.324220</td>
<td>39.582046</td>
<td>8.504523</td>
<td>0.000000</td>
</tr>
</tbody>
</table>
Table 3.3: Spatial and energy information per DLWF for ethylene.

<table>
<thead>
<tr>
<th>index</th>
<th>$\langle r_x \rangle$</th>
<th>$\langle r_y \rangle$</th>
<th>$\langle r_z \rangle$</th>
<th>$\langle \Delta r^2 \rangle$</th>
<th>$\langle h \rangle$</th>
<th>$\langle \Delta h^2 \rangle$</th>
<th>$\lambda^{(00)}_{nm}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>-0.0000</td>
<td>-0.0000</td>
<td>0.0000</td>
<td>0.983945</td>
<td>-18.612524</td>
<td>0.000174</td>
<td>1.000000</td>
</tr>
<tr>
<td>02</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.784198</td>
<td>-13.889465</td>
<td>0.000596</td>
<td>0.999997</td>
</tr>
<tr>
<td>03</td>
<td>0.0002</td>
<td>-0.0000</td>
<td>-0.0001</td>
<td>1.633431</td>
<td>-11.218850</td>
<td>0.003959</td>
<td>0.999974</td>
</tr>
<tr>
<td>04</td>
<td>-0.0004</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.497676</td>
<td>-9.940055</td>
<td>0.005812</td>
<td>0.999947</td>
</tr>
<tr>
<td>05</td>
<td>-0.0010</td>
<td>-0.0000</td>
<td>-0.0000</td>
<td>2.015195</td>
<td>-8.185370</td>
<td>0.007015</td>
<td>0.999939</td>
</tr>
<tr>
<td>06</td>
<td>0.0000</td>
<td>0.0000</td>
<td>-0.0002</td>
<td>1.329210</td>
<td>-6.559715</td>
<td>0.010217</td>
<td>0.999856</td>
</tr>
<tr>
<td>07</td>
<td>0.0007</td>
<td>0.0000</td>
<td>0.0014</td>
<td>2.009921</td>
<td>-0.717311</td>
<td>0.135739</td>
<td>0.000000</td>
</tr>
<tr>
<td>08</td>
<td>-1.1695</td>
<td>2.1974</td>
<td>-0.1999</td>
<td>3.337830</td>
<td>1.292702</td>
<td>1.286921</td>
<td>0.000025</td>
</tr>
<tr>
<td>09</td>
<td>-1.1696</td>
<td>-2.1974</td>
<td>-0.2001</td>
<td>3.337846</td>
<td>1.292738</td>
<td>1.286973</td>
<td>0.000025</td>
</tr>
</tbody>
</table>

3.5.2 Mixing parameter conversion

In the molecular formulation of the localized orbital scaling correction (LOSC) method [45], the cost function is given as

$$F = (1 - \gamma')\Omega + \gamma'\Xi,$$

(3.23)

where $\Omega$ is the spatial variance cost and $\Xi$ is the energy variance cost. We write $\gamma'$ to differentiate from the mixing parameter $\gamma$ used in the main text. The constant $C$ is added because in atomic units there is a large order of magnitude difference between typical space and energy spreads. This results in mixing parameters that are very close to 1.0, which are difficult to optimize. To keep the space and energy cost within the same order of magnitude, the factor of $C$ was set to 1000. The implementation in wannier90 uses the units of Å and eV, which results in spreads that are of the same order of magnitude; thus, we do not use a factor of $C$ in that implementation of the cost function. Converting between these two different equations and units can be accomplished with the following formulae. When converting from Eq. (3.23), the equivalent mixing parameter for the cost function in the main text is given by

$$\gamma = \left(1 + \frac{1 - \gamma'}{\gamma'} m^2 C\right)^{-1},$$

(3.24)
where \( m = \frac{0.529177210903}{27.211386245988} \) is the conversion from Bohr to Å divided by the conversion from Hartee to eV. Values are obtained from the NIST CODATA database \([139]\). The conversion from the mixing parameter in the main text to the cost function in Eq. (3.23) is given by

\[
\gamma' = \left( 1 + \left( \frac{1 - \gamma}{\gamma m^2 C} \right) \right)^{-1}.
\]  

(3.25)

### 3.5.3 Code implementation details

The cost function outlined in the main text was implemented in the \texttt{wannier90} code \([126]\). To implement the energy localization, the data structures necessary for energy localization are only calculated if energy localization is requested. The energy spread cost is calculated in the function \texttt{wann_xi}, which is called right after the function \texttt{wann_omega}, which calculates the spatial spread. Similarly, the energy gradient is only calculated when needed and is calculated in \texttt{wann_dxi}, which is called right after the function that calculates the spatial spread gradient, \texttt{wann_domega}. The analytic gradient implemented in \texttt{wann_xi} was verified numerically for some test systems to ensure the derivation is correct. The additional data structures that hold the pertinent information for energy localization are \texttt{have} for average energy, \texttt{have2} for average energy squared, and \texttt{h2ave} for average of the squared energy. These variables are used to update the public variables \texttt{wannier_energies} and \texttt{wannier_espreads}, which are the energy version of the public variables \texttt{wannier_centres} and \texttt{wannier_spreads} used for spatial information. Since adding these variables into the library mode would necessitate a change to the API, this change was not implemented. If desired, these variables could be added to the API as optional arguments, which should prevent breaking any existing code that uses the API.
### 3.5.4 Keyword List

The wannier90 program uses an input file named `seedname.win` where `seedname` can be set from the command line. This input file contains key-value pairs in a free-form structure that controls the settings of the program. This is a list of the new keywords added in the implementation of energy localization in the wannier90 code.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Energy Mixing Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP_EN_MIX</td>
<td>R</td>
<td>Space and energy mixing parameter</td>
</tr>
<tr>
<td>ECONV_MAX</td>
<td>R</td>
<td>Energy convergence maximum</td>
</tr>
<tr>
<td>NCONV_MAX</td>
<td>I</td>
<td>Number of WFs to use for convergence</td>
</tr>
<tr>
<td>NUM_OCC</td>
<td>I</td>
<td>Bloch orbital occupation per k-point</td>
</tr>
<tr>
<td>WRITE_INFO</td>
<td>L</td>
<td>Write localization information per WF to file</td>
</tr>
<tr>
<td>WANNIER_PLOT_DENSITY</td>
<td>L</td>
<td>Plot the WF density</td>
</tr>
</tbody>
</table>

**Table 3.4: seedname.win file keywords controlling the energy localization. Argument types are represented by, I for a integer, R for a real number, and L for a logical value.**

### 3.5.5 Keyword details

**real :: sp_en_mix**

Space and energy mixing parameter; a value of 0.0 gives the MLWF cost function. This value must obey the inequality $0.0 \leq \text{sp\_en\_mix} \leq 1.0$. When $0.0 < \text{sp\_en\_mix}$ then the per WF information printed at every print cycle will change to include the energy statistics.

The default value is 0.0.

**real :: econv_max**

The upper bound for the energy window used for testing localization convergence. This is useful when considering high-energy unoccupied orbitals since they can be very noisy during the descent. If `econv_max` is set then an additional line will be printed at the end of each
print cycle showing the spatial, energy, and total cost of the subset used for convergence. This setting cannot be set if `nconv_max` is set.

No default.

```fortran
integer :: nconv_max
```

The number of Wannier functions used to test for convergence during the gradient descent in the Wannierisation procedure. Useful for when considering high-energy unoccupied orbitals since they can be too noisy to converge during the descent. If `nconv_max` is set then an additional line will be printed at the end of each print cycle showing the spatial, energy, and total cost of the subset used for convergence. This setting cannot be set if `econv_max` is set.

No default, but if `nconv_max > num_wann` it will be set to `num_wann`.

```fortran
integer :: num_occ
```

The number of occupied of Bloch orbitals. This is used to print the occupation of the WFs. Since it is only a single integer this implementation only applies correctly for insulators and Γ-point calculations. In the typical case of finding MLWFs for occupied only orbitals the WF occupations will always be 1.0. Setting `num_occ` is useful when considering WFs that include both occupied and unoccupied Bloch orbitals.

No default.

```fortran
logical :: write_info
```

If `true`, write a summary of the cost information per WF to the file `seedname.info` for the WFs in the home unit cell. The format is WF index, Cartesian expectation of center in Å, spatial spread in Å². If `sp_en_mix ≠ 0.0` then it will additionally print the WF energy expectation in eV and the energy spread in eV². If `num_occ ≠ 0` then it will also print the WF occupations.

The default value is `false`. 83
logical :: wannier_plot_density

Plot the WF density instead of the wavefunction, where the density is defined as $|w_n(r)|^2$. In the case of WFs that are not strictly real-valued, plotting the wavefunction will truncate the imaginary part. Plotting the density instead guarantees the whole WF is plotted.

The default value is false.

3.5.6 Energy gradient derivation

The energy variance with respect to the WFs can be written

$$\Omega_E = \sum_n \left[ \langle h_s^2 \rangle_n - \langle h_s \rangle_n^2 \right] = \sum_n \left[ \langle 0n | h_s^2 | 0n \rangle - \langle 0n | h_s | 0n \rangle^2 \right]$$

Using the same conventions as [1], Bloch (or KS) orbitals transform to WFs as

$$|R_m\rangle = \frac{1}{N} \sum_k \sum_p U_{pm}^k |\psi_p k\rangle$$

This convention also leads to the orbital normalization of $\langle \psi_{n k} | \psi_{n' k'} \rangle = N \delta_{n,n'} \delta_{k,k'}$. Looking at the ‘dipole’ squared term $\langle (h_s)^2 \rangle_n^1$, I’ve chosen to denote this term $\tilde{\Omega}_E$ in keeping with the Wannier90 formulation for the portion of the cost function that is transform dependent.

$$\tilde{\Omega}_E = \sum_n \langle 0n | h_s | 0n \rangle^2 = \sum_n \left[ \left( \frac{1}{N} \sum_k \sum_p U_{pm}^k \langle \psi_p k \rangle \right) h_s \left( \frac{1}{N} \sum_{k'} \sum_q U_{qn}^{(k')} \langle \psi_q k' \rangle \right) \right]^2$$

$$= \sum_n \left[ \frac{1}{N} \sum_k \sum_p U_{pm}^k U_{pn}^{(k') \epsilon_{pk}} \right]^2 = \frac{1}{N^2} \sum_n \left( \sum_k \sum_p U_{pm}^k U_{pn}^{(k') \epsilon_{pk}} \right) \left( \sum_{k'} \sum_q U_{qn}^{(k')} U_{qn}^{(k') \epsilon_{qk'}} \right)$$

1 ‘Dipole’ because in space the linear(r) term is proportional to the dipole moment.
\[
\frac{1}{N^2} \sum_{n} \sum_{k,k'} \sum_{p,q} \epsilon_{pqk} \epsilon_{qtk'} |U_{pn}^k|^2 |U_{qn}^{k'}|^2 = \sum_{n} \left( \sum_{p,k} \epsilon_{pqk} |U_{pn}^k|^2 \right)^2
\]

This term appears to depend on the transform, as expected. It will be easier to work with this term though if we put the \( \epsilon_{nk} \) into a diagonal matrix of canonical orbital energies, \( (D^k)_{nm} = \delta_{n,m} \epsilon_{nk} \). Under an arbitrary transformation \( U^k \) then \( D^k \) becomes \( \Lambda^k = U^{\dagger(k)} D^k U^k \). First let’s formulate the term inside the square of \( \tilde{\Omega}_E \) in terms of \( \Lambda^k \).

\[
\sum_{p,k} \epsilon_{pqk} |U_{pn}^k|^2 = \sum_{p,k} \epsilon_{pqk} U_{pn}^{k\dagger} U_{pn}^k = \sum_{k} \sum_{p} D_{pp}^k U_{pn}^{k\dagger} U_{pn}^k
\]

\[
= \sum_{k} \sum_{p} (U^{\dagger(k)} D^k)^{np} U_{pn}^k = \sum_{k} (U^{\dagger(k)} D^k U^k)_{pn} = \sum_{k} \Lambda^{kn}_{nn}
\]

At the next step in the minimization, we’ll call the transformed energy matrix \( \tilde{\Lambda}^k = \tilde{U}^{\dagger(k)} D^k \tilde{U}^k \). A change in \( \tilde{\Omega}_E \) would then look like the following:

\[
d\tilde{\Omega}_E = \frac{1}{N^2} \sum_{n} \left\{ \left( \sum_{p,k} \epsilon_{pqk} |\tilde{U}_{pn}^k|^2 \right)^2 - \left( \sum_{p,k} \epsilon_{pqk} |U_{pn}^k|^2 \right)^2 \right\}
\]

\[
d\tilde{\Omega}_E = \frac{1}{N^2} \sum_{n} \left\{ \left( \sum_{k} \tilde{\Lambda}^{kn}_{nn} \right)^2 - \left( \sum_{k} \Lambda^{kn}_{nn} \right)^2 \right\}
\]

Now we approximate the next step in the unitary transform as \( \tilde{U}^{\dagger(k)} \approx (1 + d\tilde{W}^{\dagger(k)}) U^{\dagger(k)} \), where \( d\tilde{W}^k \) is a numerically small anti-Hermitian matrix so that its exponentiation is unitary. Using this we can approximate \( \tilde{\Lambda}^{kn}_{nn} \) as a function of \( \Lambda^{kn}_{nn} \).

\[
\tilde{\Lambda}^{kn}_{nn} = \left[ (1 + d\tilde{W}^{\dagger(k)}) \Lambda^k (1 + d\tilde{W}^k) \right]_{nn}
\]

\[
= \left[ \Lambda^k + d\tilde{W}^{\dagger(k)} \Lambda^k + \Lambda^k d\tilde{W}^k + O(d\tilde{W}^2) \right]_{nn}
\]
\[ = \Lambda_{nn}^k + (-d\tilde{W}^k \Lambda^k + \Lambda^k d\tilde{W}^k)_{nn} = \Lambda_{nn}^k + \left( [\Lambda^k, d\tilde{W}^k] \right)_{nn} \]

where \([A, B]\) denotes the commutator. Since \(D^k\) is Hermitian and \(\Lambda^k\) is a unitary transform of \(D^k\) it is also Hermitian. This gives us

\[
\left( [\Lambda^k, d\tilde{W}^k] \right)_{nn} = (\Lambda^k d\tilde{W}^k - d\tilde{W}^k \Lambda^k)_{nn} = (\Lambda^k d\tilde{W}^k + d\tilde{W}^{\dagger(k)} \Lambda^{\dagger(k)})_{nn}
\]

\[
= \sum_m (\Lambda_{nm}^k d\tilde{W}_{mn}^k + d\tilde{W}_{mn}^k \Lambda_{nm}^{*k}) = 2 \sum_m \Re \{ \Lambda_{nm}^k d\tilde{W}_{mn}^k \} = 2 \Re \{ (\Lambda^k d\tilde{W}^k)_{nn} \}
\]

I’ll write \(2 \Re \{ (\Lambda^k d\tilde{W}^k)_{nn} \}\) as \(2 \Re (\Lambda^k d\tilde{W}^k)_{nn}\) for compactness from here on. With this we’ll now look at the individual \(n\) terms in \(d\tilde{\Omega}_E\).

\[
\left( \sum_k \tilde{\Lambda}_{nn}^k \right)^2 = \left( \sum_k \Lambda_{nn}^k \right)^2 = (\sum_k \tilde{\Lambda}_{nm}^k) (\sum_k \Lambda^{(k')}_{nm}) - (\sum_k \Lambda_{nm}^k) (\sum_k \Lambda^{(k')}_{nm})
\]

\[
= \sum_{k,k'} \left( \tilde{\Lambda}_{nm}^k \Lambda^{(k')}_{nn} - \Lambda_{nm}^k \Lambda^{(k')}_{nn} \right)
\]

Examining the \(\tilde{\Lambda}_{nn}^k \Lambda^{(k')}_{nn}\) terms:

\[
\tilde{\Lambda}_{nn}^k \Lambda^{(k')}_{nn} = \left( \Lambda_{nn}^k + 2 \Re (\Lambda^k d\tilde{W}^k)_{nn} \right) \left( \Lambda^{(k')}_{nn} + 2 \Re (\Lambda^{(k')} d\tilde{W}^{(k')})_{nn} \right)
\]

\[
= \Lambda_{nn}^k \Lambda^{(k')}_{nn} + 2 \Re (\Lambda^k d\tilde{W}^k)_{nn} \Lambda^{(k')}_{nn} + 2 \Re (\Lambda^{(k')} d\tilde{W}^{(k')})_{nn} \Lambda_{nn}^k + \mathcal{O}(d\tilde{W}^2)
\]

Which gives us for the whole sum

\[
d\tilde{\Omega}_E = \frac{1}{N^2} \sum_n \sum_{k,k'} \left( 2 \Re (\Lambda^k d\tilde{W}^k)_{nn} \Lambda^{(k')}_{nn} + 2 \Re (\Lambda^{(k')} d\tilde{W}^{(k')})_{nn} \Lambda_{nn}^k \right)
\]

86
\[
\frac{d}{dW} \text{Re}\{\text{tr}[M dW]\} = A[M] = \frac{1}{2} (M - M^\dagger)
\]

So now we’d like to use the formula for anti-Hermitian matrix derivatives:

\[
d\text{Re}\{\text{tr}[M dW]\} = A[M] = \frac{1}{2} (M - M^\dagger)
\]

Now we define a matrix \( V_{nm}^{(k,k')} = \Lambda_{nm}^{(k')} \Lambda_{nm}^{k} \)

\[
d\Omega_E = \frac{4}{N^2} \sum_{k,k'} \sum_{n} \text{Re}(\Lambda_{nm}^{k} d\bar{W}_{mn}^{k}) \Lambda_{nm}^{(k')} = \frac{4}{N^2} \sum_{k,k'} \sum_{n} \text{Re}\left( \sum_{m} \Lambda_{nm}^{k} d\bar{W}_{mn}^{k} \Lambda_{nm}^{(k')} \right)
\]

\[
\frac{4}{N^2} \sum_{k,k'} \sum_{n} \text{Re}\left( \sum_{m} \Lambda_{nm}^{(k')} \Lambda_{nm}^{k} d\bar{W}_{mn}^{k} \right)
\]

Now we define a matrix \( V_{nm}^{(k,k')} = \Lambda_{nm}^{(k')} \Lambda_{nm}^{k} \)

\[
d\Omega_E = \frac{4}{N^2} \sum_{k,k'} \sum_{n} \text{Re}(V_{nm}^{(k,k')} d\bar{W}_{mn}^{k}) \Lambda_{nm}^{(k')} = \frac{4}{N^2} \sum_{k,k'} \text{Re} \text{ tr}[V_{nm}^{(k,k')} d\bar{W}_{mn}^{k}]
\]

which finally results in

\[
\frac{d\Omega_E}{dW^k} = \frac{4}{N} \sum_{k'} A[V_{nm}^{(k,k')} ] = \frac{2}{N} \sum_{k'} \left( V_{nm}^{(k,k')} - V_{nm}^{*(k,k')} \right)
\]

87
Chapter 4

The Localized Orbital Scaling Correction in Periodic Systems

We now return to density functional theory after the previous chapter’s foray into localization in periodic boundary conditions. As mentioned previously, density functional theory offers reasonable accuracy at acceptable computational cost, but it suffers from delocalization error; this results in inaccurate predictions of quantities related to energy differences, such as band gaps. In chapter 2 we introduced LOSC in molecules as a method that aims to systematically correct delocalization error. In this chapter we extend the LOSC method to periodic systems, where the localized orbitals employed are the dually localized Wannier functions from the previous chapter. In light of the effect of the bulk environment on the electrostatic interaction between localized orbitals, we also modify the LOSC energy correction to include a screened Coulomb kernel. For a test set of semiconductors and large-gap insulators, we will show that the screened LOSC method consistently improves the band gap compared to the parent density functional.

4.1 Periodic systems

In PBCs, the eigenfunctions of the single-particle Hamiltonian are known as Bloch orbitals; they satisfy \( h_s \psi_n^k = \epsilon_n^k \psi_n^k \). The Bloch orbitals are also eigenfunctions of the unit cell translation operator, which leads them to take the form \( \psi_n^k(r) = e^{ik \cdot r} u_n^k(r) \), where \( u_n^k \) has the periodicity of the unit cell and \( k \) is a point in the Brillouin zone \([117],[140]\).

The Bloch orbitals obey the normalization convention \( \langle \psi_m^q | \psi_n^k \rangle = \delta(k - q) \delta_{mn} \), where \( \langle f | g \rangle = \int_D dr \overline{f(r)} g(r) \). Here, \( \delta(k) \) is the Dirac delta function, \( \delta_{mn} \) is the Kronecker delta, and \( \overline{f} \) is the complex conjugate of \( f \). The domain of integration \( D \) is the periodic unit for
the functions being integrated; for the Bloch orbitals, $D = \mathbb{R}^3$. The $u_n^k$ are orthonormal in the band index $n$ at a fixed $k$-point in reciprocal space: that is, $\langle u_m^k | u_n^k \rangle = \delta_{mn}$, where the domain of integration here is one unit cell. Note that we assume closed-shell systems in this work.

The single-particle density can be represented by the occupied Bloch orbitals as

$$\rho_s(\mathbf{r}) = \sum_n^{\text{occ}} \frac{V}{(2\pi)^3} \int_{\text{BZ}} d\mathbf{k} |\psi_n^k(\mathbf{r})|^2,$$

(4.1)

where $V$ is the volume of the unit cell and the integral is over the first Brillouin zone. Since the Hamiltonian is diagonal in $\mathbf{k}$, we can solve for the Bloch orbitals in reciprocal space, requiring diagonalization only in one unit cell. In practice, the Brillouin zone is sampled with a finite number of points; in this work we use a Monkhorst–Pack mesh centered at the origin of the Brillouin zone, the $\Gamma$-point [116]. This reduces integrals over the Brillouin zone to equally weighted sums over the $k$-point mesh

$$\frac{V}{(2\pi)^3} \int d\mathbf{k} f(\mathbf{k}) \rightarrow \frac{1}{N_k} \sum_k f(\mathbf{k}),$$

(4.2)

where $N_k$ is the number of $k$-points in the mesh. Using a $\Gamma$-centered Monkhorst–Pack mesh yields Bloch orbitals having the periodicity of an unfolded supercell comprised of $N_k$ primitive unit cells; this supercell is referred to as the Born–von Karman cell [117]. The Bloch orbitals then obey the normalization convention $\langle \psi_m^{(q)} | \psi_n^k \rangle = N_k \delta_{qk} \delta_{mn}$, where the integral is over the Born–von Karman cell.

4.2 Periodic methods

The LOSC method consists of two steps. First, we find orbitals that are spatially localized, while remaining associated with specific energy ranges. Next, we compute a curvature matrix modeling the magnitude of the deviation from linearity. This is combined with the fractionally occupied localized orbitals to correct the convex deviation of $E(N)$ from
linearity at non-integer $N$, as well as incorrect total energies at integer $N$. Both steps are implemented as post-processing after a converged self-consistent field calculation. The localization step is implemented in a modified version of the wannier90 code [125], [126] as detailed in the previous chapter. The energy correction is implemented as an add-on module to the Quantum ESPRESSO suite [132], [133].

4.2.1 Energy Correction

The deviation from energy linearity with respect to fractional charges is characteristically quadratic in most exchange-correlation functionals [41], [141], [142]. To restore compliance with the PPLB condition for small finite systems, the global scaling correction (GSC) approach was developed with a quadratic in total energy of fractional systems, based on the occupation numbers of canonical molecular orbitals [39], [141]. The GSC is very effective in correcting the systematic deviation from the PPLB condition for systems with fractional electrons and leads to accurate prediction of quasiparticle energies as the eigenvalues from the resulting one-electron Hamiltonian. However, the GSC is applicable only for small and moderate-size systems because the convex deviation of conventional DFAs from the PPLB linear line for fractional charges decreases with increasing system size, and the delocalization error manifests as underestimated ground-state energies for integer systems and incorrect linear $E_{gs}(N)$ curves with wrong slopes at the bulk limit [38]. To achieve size-consistent and systematic improvement, the localized orbital scaling correction (LOSC) formulated in terms of orbitalets were developed to apply the global or local corrections adaptively. We now extend LOSC to periodic systems.

LOSC uses the occupation matrix between the DLWFs,

$$\lambda_{mn}^{TR} = \langle w^T_m | \rho_s | w^R_m \rangle.$$ \hspace{1cm} (4.3)

The occupations between all possible combinations of DLWFs are used to remove quadratic deviations while the diagonal terms are used to restore linearity. The energy correction for
each unit cell is given by

$$\Delta E^{\text{LOSC}} = \frac{1}{2N_k} \sum_{\text{TR}} \sum_{mn} \kappa^{\text{TR}}_{mn} \lambda^{\text{TR}}_{mn} (\delta^{\text{TR}}_{mn} - \lambda^{\text{TR}}_{mn}),$$

(4.4)

where $\delta^{0\text{R}}_{mn} = \delta_{mn} \delta^{0\text{R}}$ and $\bar{\kappa}$ models the curvature of the deviation from linearity.

The diagonal elements of the energy correction are proportional to $\lambda^{\text{TR}}_{ii} - |\lambda^{\text{TR}}_{ii}|^2$; thus, if a DLWF has integer occupancy (implying $\lambda^{\text{TR}}_{ij} = 0$ whenever $i \neq j$ or $T \neq R$), then the energy correction due to that DLWF will be zero as well.

The matrix $[\lambda^{\text{TR}}_{ij}]$ of occupations between the DLWFs is the discrete Fourier transform of the occupation matrix between the TBOs. As such, it is positive semidefinite and Hermitian, with trace per unit cell equal to $N_k^{-1} \sum_k N^k_f$, where $N^k_f$ is the number of electrons below the Fermi energy at $k$.

Following the prescription of [45], the elements of the curvature matrix are given by

$$\bar{\kappa}^{\text{TR}}_{ij} = \text{erf}(8S^{\text{TR}}_{ij})\sqrt{\kappa^{\text{TT}}_{ii} \kappa^{\text{RR}}_{jj}}$$

$$+ \text{erfc}(8S^{\text{TR}}_{ij})\kappa^{\text{TR}}_{ij}. \quad (4.5)$$

Here, $\text{erfc}(r) = 1 - \text{erf}(r)$ is the complementary error function. $S_{ij}^{\text{TR}}$ is the absolute overlap between DLWFs,

$$S^{\text{TR}}_{ij} = \int dr \sqrt{\rho^T_i(r) \rho^R_j(r)}, \quad (4.6)$$

where $\rho^T_i(r) = |w^T_i(r)|^2$ is a DLWF’s charge density. The unsmoothed curvature elements $\kappa^{\text{TR}}_{ij}$ in Eq. (4.5) are given by

$$\kappa^{\text{TR}}_{ij} = J[\rho^T_i, \rho^R_j] - X[\rho^T_i, \rho^R_j], \quad (4.7)$$

with

$$J[\rho^T_i, \rho^R_j] = \iint dr dr' \rho^T_i(r) \rho^R_j(r') K(|r - r'|), \quad (4.8a)$$

$$X[\rho^T_i, \rho^R_j] = \tau \frac{2C_X}{3} \int dr \left[ \rho^T_i(r) \rho^R_j(r) \right]^{2/3}. \quad (4.8b)$$
In the above, $K(r) = 1/r$ is the Coulomb kernel, $C_X = \frac{3}{4}(\frac{6}{\pi})^{1/3}$ is the Dirac exchange constant [19], and $\tau = 6(1 - 2^{-1/3}) \approx 1.2378$ is a non-empirical parameter [41]. The derivation of how this correction restores the PPLB condition can be found in the supplementary data of [41]. The use of the smoothed curvature $\kappa$ instead of $\kappa$ was introduced in [45] because the cost function in Eq. (3.6) can induce discontinuous jumps between localization character during molecular dissociation. The diagonal elements of $\kappa$ and $\kappa$ are equal, so when $\lambda_{ii} \in \{0, 1\}$ the corrections from $\kappa$ and $\kappa$ are the same. In practice, $X[\rho^T_i, \rho_R^j]$ term is evaluated using numerical integration on a grid of real-space points. The Coulomb term $J[\rho^T_i, \rho_R^j]$ is evaluated in a plane wave basis, detailed in Sec. 4.2.2.

Applying the extension of Janak’s theorem [43] to the generalized Kohn-Sham theory [29], the LOSC energy correction in Eq. (4.4) yields corrections to the Bloch orbital energy eigenvalues $\epsilon^k_i$ given by

$$
\Delta \epsilon^k_i = \sum_n \kappa^{00}_{nn} \left( \frac{1}{2} - \lambda^{00}_{nn} \right) |U^{k}_{in}|^2 
- \sum_{Rm \neq 0n} \tilde{\kappa}^{0R}_{mn} \text{Re} \left\{ \lambda^{0R}_{mn} e^{i \mathbf{k} \cdot \mathbf{R}} U^{k}_{in} U^{k}_{in} \right\}.
$$

Consider the diagonal corrections given by the first summand in Eq. (4.9). There is no correction to the eigenvalue when a DLWF is half occupied ($\lambda_{ii} = \frac{1}{2}$); on the other hand, the correction is maximal when it is completely occupied or unoccupied. Li and coworkers observed in [41] that the slopes of the quadratic DFA and the correct linear $E(N)$ curves agree at half-integer $N$. Since the frontier orbital energy corresponds to this slope, we see that accurate frontier orbital energies are given by half-occupied frontier orbitals. The LOSC correction to the orbital energies arrives naturally at this conclusion, additionally agreeing with Slater transition state theory [143], [144].

We may also view LOSC as a correction to the Hamiltonian given by the functional
\[ \Delta v = \frac{\delta \Delta E_{\text{LOSC}}}{\delta \rho_s} \]
\[ = \sum_{Rmn} \tilde{\kappa}_{mn}^{0R} \text{Re} \left\{ \left( \frac{\kappa_{mn}^{0R}}{2} - \lambda_{mn}^{0R} \right) |m^1_n\rangle \langle n^0_m| \right\} \]  

(4.10)

The correction to the \( i \)th Bloch orbital eigenvalue is then given by
\[ \Delta \epsilon^k_i = \langle \psi^k_i | \Delta v | \psi^k_i \rangle. \]

In practice, the energy corrections are applied to disentangled Bloch orbitals. The conduction bands of most systems cannot be formed into sets of bands that do not cross anywhere in the Brillouin zone, a condition referred to as band entanglement. In order to obtain a finite set of bands for localization and energy correction, we use the disentanglement procedure outlined by Souza et al. in [112]. This procedure obtains \( N_w \) bands from a set of \( N_b \geq N_w \) Bloch orbitals, chosen such that the subspace spanned by the disentangled bands is as smooth as possible in \( \mathbf{k} \). To correct the band gap of semiconductors and insulators, we therefore consider enough states that the lowest conduction bands are included in the localization and energy correction procedure. In principle, the number of bands and Wannier functions considered should be increased until convergence is attained. We find, however, that only a small number of conduction bands are required to obtain convergence in both localization and energy correction for the frontier DLWFs as we showed in the previous chapter. Thus, we set \( N_w = N_{\text{occ}} + N_{\text{coord}} \), where \( N_{\text{occ}} \) is the number of occupied bands per unit cell and \( N_{\text{coord}} \) is the coordination number of the lattice, disentangling from \( N_b = N_{\text{occ}} + 2N_{\text{coord}} \) Bloch orbitals per \( \mathbf{k} \)-point. The \( N_w \) disentangled Bloch bands per \( \mathbf{k} \)-point yield \( N_w \) DLWFs per unit cell, whence \( N_w N \) DLWFs in the Born–von Karman supercell on which the DLWFs are periodic. The energy corrections for the \( N_w \) disentangled Bloch orbitals per \( \mathbf{k} \)-point are implemented using Eq. (4.9).

### 4.2.2 Coulomb integrals

The quadratic scaling of the Coulomb energy \( J[\rho^T_i, \rho_j^R] \) is the dominant contribution to the quadratic deviation of the total energy with respect to fractional charge. Precise calculation
of $J[\rho^T_i, \rho^R_j]$ is therefore needed for LOSC to restore the PPLB condition accurately. In PBCs, a plane-wave basis is typically employed. The Coulomb energy is diagonal in this basis, and the double integral required in real space collapses to a single sum over basis vectors $G$:

$$J[\rho^T_i, \rho^R_j] = \sum_G \frac{4\pi}{G^2} \rho^T_i(G) \rho^R_j(G),$$

where $G = |G|$ is the length of a plane-wave basis vector. However, this sum converges only for neutral charge distributions, for which the $G = 0$ term vanishes. The DLWF densities are individually charged, so ignoring the divergent term coming from the net charge will significantly underestimate the Coulomb energy. There are many methods to evaluate the Coulomb energy for charged densities in the plane-wave basis accurately; for instance, [145]–[148]. We choose the spherical cutoff method of [149]–[151], truncating the Coulomb kernel in Eq. (4.8a) at a cutoff radius $R_c$; in particular, we let $R_c$ be half the length of the shortest Born–von Karman supercell lattice vector. Thus the spherical cutoff Coulomb kernel is

$$K_c(r; R_c) = \begin{cases} 1/r & r < R_c \\ 0 & r \geq R_c, \end{cases}$$

which has Fourier coefficients

$$K_c(G; R_c) = \begin{cases} \frac{4\pi}{G^2} [1 - \cos(GR_c)] & G \neq 0 \\ 2\pi R_c^2 & G = 0. \end{cases}$$

Observe that $K_c(G; R_c)$ does not diverge for any $G$. As long as the pair of DLWF densities in Eq. (4.8a) are well contained in a sphere of radius $R_c$, then the spherical cutoff method is also accurate in highly anisotropic unit cells, unlike schemes such as that of Makov and Payne [146]. We enforce this containment condition in practice by checking that each DLWF density is well contained in a volume spanned by half of each Born–von Karman supercell lattice vector, and only compute curvature elements between pairs of DLWF densities that have centers closer together than $R_c$. We evaluate the Coulomb integrals on the unfolded
Born–von Karman supercell in the plane-wave basis, which requires a fast Fourier transform (FFT) of the DLWF densities on the supercell.

### 4.2.3 Screening

Applying LOSC with a bare Coulomb interaction leads to severe overcorrection of semiconductors’ band gaps in PBCs. However, this is not surprising; we anticipate an effect of the other electrons in the lattice on $J[\rho^T_i, \rho^R_j]$. As shown by highly accurate methods such as $GW$, a screened Coulomb interaction is required to model the interaction between electrons in a periodic system accurately \[152 \], \[153 \]. Recently, Mei and coworkers also showed that the curvature representing the deviation from linearity of the total energy as a function of canonical orbital occupations is given to second order by a screened interaction \[39 \]. We model the screened interaction phenomenologically, attenuating the long-range $1/r$ behavior of the spherical cutoff Coulomb interaction by a complementary error function. This gives the screened Coulomb kernel

$$
K_s(r; R_c, \alpha) = \begin{cases} 
\text{erfc}(\alpha r)/r & r < R_c \\
0 & r \geq R_c,
\end{cases}
$$

(4.14)

where $\alpha$ is a tunable screening parameter. For $r$ larger than the screening radius $\alpha^{-1}$, $K_s(r; R_c, \alpha)$ decays exponentially instead of as $1/r$. The Fourier coefficients of the screened kernel are

$$
K_s(G; R_c, \alpha) = \begin{cases} 
\frac{4\pi}{G^2} \left[ 1 - \cos(GR_c) \ \text{erfc}(\alpha R_c) - e^{-(G/2\alpha)^2} \Re \left\{ \text{erf} \left( \alpha R_c + \frac{iG}{2\alpha} \right) \right\} \right] & G \neq 0 \\
2\pi R_c^2 + \pi \ \text{erf}(\alpha R_c) \left( \alpha^{-2} - 2R_c^2 \right) - 2\sqrt{\pi} e^{-(\alpha R_c)^2 R_c/\alpha} & G = 0.
\end{cases}
$$

(4.15)

The error function is unbounded for complex arguments, overflowing double-precision floating-point numbers even for relatively small $G$. Thus, we evaluate $K_s(G; R_c, \alpha)$ with a scaled form of erf $z$ called the Faddeeva function, implemented in the numerically stable ACM Algorithm 916 \[154 \], \[155 \]. For details, see Sec. 4.5.2.
4.3 Periodic results

We use the PBE density functional [25] for the parent DFA calculations, with optimized norm-conserving Vanderbilt pseudopotentials [130] generated by PseudoDojo [131]. Both self-consistent field (SCF) and non-SCF calculations are carried out in the Quantum ESPRESSO code suite [132], [133]. The energy cutoff for the fast Fourier transform is set to 100 Ry for wavefunctions and 400 Ry for densities. The Brillouin zone is sampled with Monkhorst-Pack meshes [116] centered at the origin, which is necessary for the Wannier functions to be periodic on the unfolded Born–von Karman supercell. For SCF calculations, we use a $16 \times 16 \times 16$ $k$-mesh, while the other calculations are performed on $6 \times 6 \times 6$ grids. The localization step of LOSC is implemented in a modified version of the wannier90 code [125], [126], and the energy correction as an add-on module to Quantum ESPRESSO.

To determine an optimal screening parameter $\alpha$, we minimized the lowest mean absolute percent error (MAPE) on the SC/40 set of semiconductors with experimentally available band gaps [26] together with six additional large-gap insulators. The experimental band gaps studied range from 0.23 eV to 21.7 eV. We find that $\alpha = 0.15 a_0^{-1}$ achieves the lowest MAPE; coincidentally, this value is numerically equal to the screening parameter used in the HSE density functional [26]. LOSC with Coulomb screening (sLOSC) yields marked improvement of the band gap for the test set in comparison with the parent functional as shown in Fig. 4.1, which also shows how unscreened LOSC overcorrects the band gaps; indeed, it is less accurate than the parent functional. The performance of sLOSC in molecules is better than the parent functional, but unscreened LOSC achieves the best performance in molecular systems (see Table 4.1). The Supplemental Material details the variation in performance of screened LOSC for both bulk systems and molecules with the screening parameter $\alpha$. 

96
Figure 4.1: Comparison of experimental band gaps with those calculated by PBE (○), sLOSC (×), and unscreened LOSC (+). The inset shows systems with an experimental band gap less than 5 eV.

Table 4.1: Mean absolute percent error of the band gap for PBC and molecular test sets. For details on the systems tested, see the Sec. 4.5.5.

<table>
<thead>
<tr>
<th>Method</th>
<th>PBE</th>
<th>LOSC</th>
<th>sLOSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBC</td>
<td>47.7%</td>
<td>152.1%</td>
<td>17.4%</td>
</tr>
<tr>
<td>Molecule</td>
<td>79.8%</td>
<td>10.1%</td>
<td>43.6%</td>
</tr>
</tbody>
</table>

The band structures of sLOSC and of the parent functional are shown for the small-gapped semiconductor silicon in Fig. 4.2 and the larger-gapped insulator lithium fluoride in Fig. 4.3. They use the disentangled band structures, which are numerically indistinguishable from the true band structure at and below the conduction band minimum for the parent functional. Wannier interpolation in the same DLWF basis as that used in sLOSC is used to find the energy at the points in the Brillouin zone not explicitly treated by the localization and energy correction. In both cases, sLOSC mostly corrects the occupied states, affecting the energy of the unoccupied bands much less. This is due to the fact that the occupied
states are more localized in space, resulting in a larger curvature and therefore a larger energy correction.

![Figure 4.2](image)

**Figure 4.2**: Band structure of silicon under the PBE functional (dashes) and sLOSC (solid). The Fermi energy of the PBE calculation was 6.23 eV, while the PBE with LOSC Fermi energy was 5.49 eV.

### 4.4 Summary

In summary, we have shown that despite the simple form of its phenomenological Coulomb screening, sLOSC systematically corrects the band gap error associated with the parent functional for materials spanning a large range of band gaps. Screening improves the correction of delocalization error in bulk systems, but degrades the accuracy of molecular systems’ band gaps relative to unscreened LOSC; however, sLOSC still offers better band gaps than those computed by the parent functional. To model the curvature more accurately for all systems, linear-response methods could be used to find the exact expression for $\partial^2 E / \partial \lambda_{ij}$, extending the result of Mei et al. [39] to the localized orbitals used by LOSC. This
Figure 4.3: Band structure of lithium fluoride under the PBE functional (dashes) and sLOSC (solid). The Fermi energy of the PBE calculation was 0.97 eV, while the PBE with LOSC Fermi energy was $-3.52$ eV. The core states are not included in the figure.
would alleviate the restrictions imposed by modeling $\kappa$ by a difference between Coulomb repulsion and Dirac exchange.

We implement the energy correction as a post-processing step to a self-consistent calculation; such corrections are accurate when the change in electron density is small and hence the total energy correction is small. For every system considered in this work, $\Delta E^{\text{LOSC}}$ is less than one part in $10^5$. The corresponding change to the density for such systems is also expected to be minimal. As shown by Mei, Chen, and Yang in [156], LOSC can be implemented self-consistently; this can correct the delocalization error of the total density, improving the accuracy of LOSC for systems with large total energy corrections. A self-consistent implementation of sLOSC could be necessary for the accurate computation of metallic band structures; since metals will be expected to produce DLWFs with occupancies far from 0 or 1, the sLOSC energy correction is likely to be larger, and self-consistently correcting the delocalized density is expected to yield better orbital energies.

The sLOSC method as implemented in this work scales as $O(N^3_w N_k)$ for the localization step, $O(N^2_w N_G)$ for the computation of curvature elements, and $O(N_w N_G \log(N_G))$ for the FFT of the DLWF densities. Here, $N_G$ is the number of plane waves in the unfolded supercell, which is $N_k$ times the number of plane waves in the unit cell. Calculating the curvature and energy corrections is the computational bottleneck for the systems evaluated in this work, with wall times ranging from 6 minutes to 3.9 hours using 16 threads on an E5-2630v3 processor. The systems that took the longest time however were the ones with the largest number of core states, which have no effect on frontier state corrections and could therefore be neglected if only a correction to the band gap is desired. This time was divided fairly equally between the computation of the matrix elements defined in Eqs. (4.6, 4.8a, 4.8b). We note that the size of the supercell used to numerically integrate these quantities could be reduced from the full Born–von Karman supercell if the relevant DLWF densities are contained in a smaller region. This is supported by the fact that systems that had similar localizations for a $4 \times 4 \times 4$ and $6 \times 6 \times 6$ $k$-mesh yielded very similar energy
corrections.

Along the same lines, we find that for the systems in the test set the correction results are converged with a $6 \times 6 \times 6$ \textit{k}-mesh, but this is only necessary to achieve a converged localization. Certain systems exhibit a qualitatively different localization with a smaller $4 \times 4 \times 4$ \textit{k}-mesh; however, by decreasing the value of $\gamma$ in the localization cost function $F$, a set of DLWFs qualitatively similar to the $6 \times 6 \times 6$ case can be obtained. In the previous chapter we showed that the shape taken by the DLWFs can be characterized by the energy degeneracy pattern of the DLWFs.

Other methods that attempt to address delocalization error in bulk calculations, such as Ma and Wang’s DFT+U [69] and the screened range-separated hybrid functional [58], rely on supercell SCF calculations. A SCF calculation scales cubically with the number of electrons, so an unfolded Born–von Karman supercell arising from $N_k$ \textit{k}-points sampling a unit cell with $N_w$ Wannier functions scales as $O(N_k^3 N_w^3)$. Both of the aforementioned methods use Wannier functions as a localized charge representation and rely on manually choosing the Bloch orbitals to comprise the Wannier functions representing the frontier of the occupied space. The sLOSC method uses DLWFs, which naturally supply Wannier functions representing the frontier of the occupied and unoccupied spaces without the need for manual energy windowing.

4.5 Supplemental material

4.5.1 Energy correction with disentanglement

When the localization to DLWFs is applied to metals or the conduction bands of semiconductors, the set of bands cannot be formed into a composite set, and is referred to as entangled. In that case Souza et al. showed how an $N_w$ subset of interest may be extracted from $N_b$ bands to create a set of disentangled Bloch orbitals per \textit{k}-point [112]. These disentangled Bloch orbitals may then be treated as a composite set for localization. When LOSC is applied to the DLWFs found from disentangled Bloch orbitals then the orbital
energy corrections can be applied to the disentangled Bloch orbital energies to obtain the correction to the frontier orbital energies and therefore band gap. If the disentangled set of eigenvalues is a subset of the true eigenvalues in an area of interest, i.e. close to the Fermi energy or conduction band minimum, then correcting the disentangled eigenvalues is equivalent to correcting the full space eigenvalues. It was shown in chapter 3 how this is true when considering enough states above the Fermi energy, which we have done in this work. For completeness, here we show what the additive form of the eigenvalue correction is when applied to the full space eigenvalues. This can be found by finding the expectation of the LOSC Hamiltonian correction, $\Delta \nu$, in the Bloch orbital basis.

We start by showing how disentangled Bloch orbitals $\phi^k_n$, can be represented in the Bloch orbital basis $\psi^k_b$,

$$|\phi^k_n\rangle = \sum_b V^{k^\dagger}_{nb} |\psi^k_b\rangle,$$

(4.16)

where $V^k$ is an isometry of dimension $N_w \times N_b$. This means it obeys the relation $V^k (V^k)^\dagger = 1$, where 1 is the $N_w \times N_w$ identity operator. We may then write a transformed Bloch orbital for generalized Wannier function construction in the disentangled Bloch orbital basis as

$$\phi^k_n = \sum_a U^k_{na} \phi^k_a.$$

(4.17)

A generalized Wannier function constructed from a set of disentangled Bloch orbitals then takes the form

$$w^{(R)}_m = \frac{1}{N_k} \sum_k e^{-i \mathbf{k} \cdot \mathbf{R}} \sum_a U^k_{ma} \sum_b V^k_{ab} \psi^k_a$$

(4.18)

$$= \frac{1}{N_k} \sum_k e^{-i \mathbf{k} \cdot \mathbf{R}} \sum_b j^{k}_{mb} \psi^k_b.$$
where $L_{mb}^k = [U^k V^k]_{mb}$. As a reminder the effective LOSC Hamiltonian is

$$\Delta v = \frac{\delta \Delta E^\text{LOSC}}{\delta \rho_s}$$

$$= \sum_{R,mn} \delta_{0R} \Re \left\{ \left( \frac{\delta_{0R}}{2} - \lambda_{mn}^0 \right) |w_n^0\rangle \langle u_n^R| \right\}. \tag{4.19}$$

The most relevant quantity to find this expectation is the overlap of a Wannier function and Bloch orbital.

$$\langle w_n^R | \psi_{ik}^k \rangle = \frac{1}{N_k} \sum_q e^{i q \cdot R} \langle \phi_q n | \psi_{ik}^k \rangle$$

$$= \frac{1}{N_k} \sum_q e^{i q \cdot R} \sum_b T_{qb}^q \langle \psi_b^q | \psi_{ik}^k \rangle$$

$$= e^{i k \cdot R} T_{ni}^k, \tag{4.20}$$

where we have used the normalization convention $\langle \psi_b^q | \psi_{ik}^k \rangle = N \delta_{qk} \delta_{bi}$ and the notation $\bar{z}$ for the complex conjugate of $z$. We can now easily find the expectation of the LOSC Hamiltonian in the Bloch orbital basis,

$$\langle \psi_{i}^k | \Delta v | \psi_{i}^k \rangle = \sum_{R,mn} \delta_{0R} \Re \left\{ \left( \frac{\delta_{0R}}{2} - \lambda_{mn}^0 \right) \langle \psi_{i}^k | w_n^0 \rangle \langle u_n^R | \psi_{i}^k \rangle \right\}$$

$$= \sum_{R,mn} \delta_{0R} \Re \left\{ \left( \frac{\delta_{0R}}{2} - \lambda_{mn}^0 \right) L_{mi}^k T_{ni}^k e^{i k \cdot R} \right\}. \tag{4.21}$$

Functionally the only difference between this form and using the disentangled basis is that instead of $L^k$ the correction would have $U^k$, the transformation from the (possibly) disentangled basis to the transformed basis.

### 4.5.2 Coulomb kernel derivation

In this section, we discuss details of the Coulomb interaction $J$ between two dually localized Wannier functions (DLWFs) $|w_i\rangle$ and $|w_j\rangle$; note that we combine the DLWFs’ band-like
and unit cell indices into a single subscript, so $|w_i⟩ = |w_{i}^{R}\rangle$. $J$ depends on the charge density $\rho$ associated with each DLWF, $\rho_i(r) = |w_i(r)|^2$, and is given by

$$J[\rho_i, \rho_j] = \iint dr \, dr' \frac{\rho_i(r) \rho_j(r')}{|r - r'|} = \iint dr \, dr' \rho_i(r) \rho_j(r') K(|r - r'|),$$

(4.22)

where we term $K(r) = 1/r$ the Coulomb kernel.

Naïvely, the cost of computing $J[\rho_i, \rho_j]$ scales as $O(N_r)^2$, where $N_r$ is the size of the densities on the real grid. However, the DLWF densities are periodic on the Born–von Karman supercell (see the main text). Recall that we may write the Fourier series for sufficiently well-behaved (that is, obeying Dirichlet conditions) periodic functions $f(r)$ as

$$f(r) = \frac{1}{\Omega_{BvK}} \sum_G f(G) e^{iG \cdot r},$$

(4.23)

where the basis vectors $G$ belong to the reciprocal lattice and $\Omega_{BvK}$ is the volume of the Born–von Karman supercell. Each DLWF density can be so represented, yielding $N_G$ Fourier coefficients $\rho(G)$. We can then exploit the fast Fourier transform to compute $J[\rho_i, \rho_j]$ in $O(N_G \log N_G)$ time.

The Fourier-space representation of $J[\rho_i, \rho_j]$ is a classical result in electrodynamics, but we present it briefly here. First, observe that we can write

$$J[\rho_i, \rho_j] = \iint dr \, dr' \frac{\rho_i(r) \rho_j(r')}{|r - r'|} = \int dr \, V_i(r) \rho_j(r),$$

(4.24)

where $V_i$ is the potential due to $\rho_i$ experienced by $\rho_j$,

$$V_i(r) = \int dr' \frac{\rho_i(r')}{|r - r'|}. $$

(4.25)

$V_i$ inherits the periodicity of $\rho_i$ and can be written as a Fourier series; it also satisfies the Poisson equation

$$\nabla^2 V_i(r) = -4\pi \rho_i(r).$$

(4.26)
This equation is algebraic in the plane-wave basis since \( \nabla^2 e^{iG \cdot r} = -G^2 \), where \( G = |G| \); thus

\[
\nabla^2 \frac{1}{\Omega_{BvK}} \left( \sum_G V_i(G) e^{iG \cdot r} \right) = -\frac{1}{\Omega_{BvK}} \sum_G G^2 V_i(G) e^{iG \cdot r} = -\frac{4\pi}{\Omega_{BvK}} \sum_G \rho_i(G) e^{iG \cdot r}. \tag{4.27}
\]

Since the plane waves \( e^{iG \cdot r} \) are orthonormal, we can solve this equation term by term, obtaining the Fourier coefficients

\[
V_i(G) = 4\pi \frac{\rho_i(G)}{G^2}. \tag{4.28}
\]

Substituting \( V_i \) into the original expression and replacing \( \rho_j \) by its Fourier series gives

\[
J[\rho_i, \rho_j] = \frac{4\pi}{\Omega_{BvK}^2} \int d\mathbf{r} \left( \sum_G \frac{\rho_i(G)}{G^2} e^{iG \cdot r} \right) \left( \sum_{G'} \rho_j(G') e^{iG' \cdot r} \right)
\]

\[
= \frac{4\pi}{\Omega_{BvK}^2} \int d\mathbf{r} \left( \sum_G \frac{\overline{\rho}_i(G)}{G^2} e^{-iG \cdot r} \right) \left( \sum_{G'} \rho_j(G') e^{iG' \cdot r} \right) \tag{4.29}
\]

\[
= \frac{4\pi}{\Omega_{BvK}^2} \int d\mathbf{r} \sum_{G, G'} \overline{\rho}_i(G) \rho_j(G') \frac{e^{i(G' - G) \cdot r}}{G^2} = \frac{4\pi}{\Omega_{BvK}} \sum_G \overline{\rho}_i(G) \rho_j(G) \frac{\delta_{G,G'}}{G^2}.
\]

We have that \( \rho_i(-G) = \overline{\rho}_i(G) \), where \( \overline{\rho} \) is the complex conjugate of \( \rho \), since \( \rho_i(\mathbf{r}) \) is real. The integral and double sum over \( G, G' \) collapse into a single sum by the orthonormality of the plane waves \( e^{iG \cdot r} \):

\[
\int d\mathbf{r} e^{i(G - G') \cdot r} = \Omega_{BvK} \delta_{G,G'}, \tag{4.30}
\]

where \( \delta_{G,G'} \) is the Kronecker delta function.

Note that we can express the Coulomb integral in reciprocal space as a sum over Fourier coefficients with a reciprocal-space Coulomb kernel:

\[
J[\rho_i, \rho_j] = \frac{1}{\Omega_{BvK}} \sum_G \overline{\rho}_i(G) \rho_j(G) K(G), \tag{4.31}
\]

where \( K(G) = 4\pi/G^2 \). This kernel has a singularity at \( G = 0 \), as does \( K(r) \); however, unlike its real-space counterpart, the singularity in \( K(G) \) is not integrable in three dimensions.
For the density of the entire unit cell, the singularity is circumvented by requiring the cell to be charge neutral; equivalently, setting $K(0) = 0$. However, the DLWFs are individually charged, so an alternative approach must be chosen. We address two problems in this work: the spurious interaction of a DLWF density with its periodically repeating images, and the suppression by the bulk environment of the long-range behavior of the Coulomb kernel. Both of these can be addressed by modifying the Coulomb kernel $K(r)$.

**The spherical cutoff kernel**

The Fourier coefficients $K(G)$ of the Coulomb kernel can be derived by taking the Fourier transform of the Yukawa kernel $K_y(r) = e^{-\lambda r}/r$ in the limit $\lambda \to 0$. Likewise, the Fourier transforms of other modified kernels $K_{\text{mod}}(r)$ yield the modified Coulomb interaction in the plane-wave basis as

$$J_{\text{mod}}[\rho_i, \rho_j] = \int \int d\mathbf{r} d\mathbf{r}' \rho_i(\mathbf{r}) \rho_j(\mathbf{r}') K_{\text{mod}}(|\mathbf{r} - \mathbf{r}'|) = \frac{1}{\Omega_{\text{BvK}}} \sum_{\mathbf{G}} \rho_i(\mathbf{G}) \rho_j(\mathbf{G}) K_{\text{mod}}(|\mathbf{G}|).$$  \hfill (4.32)

In order to prevent spurious interaction with the densities’ periodic images, we use the spherical cutoff kernel $K_c$, used first by Onida and coworkers in [149] for GW calculations and applied to DFT by Jarvis and coworkers in [150]. It is defined as

$$K_c(r) = \frac{1 - \Theta(r - R_c)}{r} = \begin{cases} \frac{1}{r} & r < R_c \\ 0 & r \geq R_c, \end{cases}$$  \hfill (4.33)

where $\Theta(r)$ is the Heaviside step function. $R_c$ is the cutoff radius, taken here to be half the length of the shortest Born–von Karman supercell lattice vector. Performing the Fourier transform yields

$$K_c(G) = \begin{cases} \frac{4\pi}{G^2} \left[1 - \cos(GR_c)\right] & G \neq 0 \\ 2\pi R_c^2 & G = 0. \end{cases}$$  \hfill (4.34)

Note that there is no singularity in $K_c(G)$. The sinusoidal oscillations are expected; they exemplify the Gibbs phenomenon due to the jump discontinuity of $K_c(r)$ at $R_c$.  

106
The sLOSC kernel

Even after introducing the spherical cutoff, the effect of the lattice means that the long-range $1/r$ behavior of $J[\rho_i, \rho_j]$ must be attenuated to give an accurate correction to the band gap of bulk systems. To complete the sLOSC kernel, we therefore introduce an additional screening term proportional to the complementary error function $\text{erfc}(r) = 1 - \text{erf}(r)$, yielding

$$K_s(r) = \frac{\text{erfc}(\alpha r) [1 - \Theta(r - R_c)]}{r} = \begin{cases} \frac{\text{erfc}(\alpha r)}{r} & r < R_c \\ 0 & r \geq R_c. \end{cases}$$  \hspace{1cm} (4.35)

Let us derive the reciprocal-space kernel $K_s(G)$ by Fourier transforming $K_s(r)$ directly, utilizing the rotational symmetry of the real-space kernel to convert into spherical coordinates. Thus

$$K_s(G) = \int dr \, K_s(r) e^{-iG \cdot r}$$

$$= \int_0^{2\pi} d\theta \int_0^\pi d\phi \int_0^\infty dr \, r^2 \sin \phi \frac{\text{erfc}(\alpha r) [1 - \Theta(r - R_c)]}{r} e^{-iGr \cos \phi}$$

$$= 2\pi \int_0^\pi d\phi \sin \phi \int_0^{R_c} r \text{erfc}(\alpha r) e^{-iGr \cos \phi}.$$  \hspace{1cm} (4.36)

Setting $u = \cos \phi$, whence $du = -d\phi \sin \phi$, we next obtain

$$K_s(G) = 2\pi \int_0^{R_c} dr \, r \text{erfc}(\alpha r) \int_{-1}^1 du \, e^{-iGu} = 2\pi \int_0^{R_c} dr \, r \text{erfc}(\alpha r) \left[ \frac{e^{iGr} - e^{-iGr}}{iGr} \right]$$

$$= \frac{4\pi}{G} \int_0^{R_c} dr \, \text{erfc}(\alpha r) \sin(Gr).$$  \hspace{1cm} (4.37)

Performing the final integral with the Mathematica software package [157] yields

$$K_s(G) = \frac{4\pi}{G^2} \left\{ 1 - \cos(GR_c) \text{erfc}(\alpha R_c) + \right.$$

$$\left. \frac{e^{-G^2/4\alpha^2}}{2} \left[ \text{erf} \left( -\alpha R_c + \frac{iG}{2\alpha} \right) - \text{erf} \left( \alpha R_c + \frac{iG}{2\alpha} \right) \right] \right\}. \hspace{1cm} (4.38)$$
Writing \( z = \alpha R_c + iG/2\alpha \) and noting that \( \text{erf}(z) = \overline{\text{erf}(\overline{z})} \) and \( \text{erf}(-z) = -\text{erf}(z) \), we may simplify the relevant part of \( K_s(G) \):

\[
\text{erf}(-z) - \text{erf}(z) = - (\text{erf} \overline{z} + \text{erf} z) = - (\overline{\text{erf} z} + \text{erf} z) = -2 \text{Re}\{\text{erf} z\}.
\]

(4.39)

Thus, we obtain in accordance with the main text that when \( G \neq 0 \)

\[
K_s(G) = \frac{4\pi}{G^2} \left[ 1 - \cos(GR_c) \text{erfc}(\alpha R_c) - e^{-G^2/4\alpha^2} \text{Re} \left\{ \text{erf} \left( \alpha R_c + \frac{iG}{2\alpha} \right) \right\} \right].
\]

(4.40)

Note that \( K_s(G) \) is not singular at \( G = 0 \); we have

\[
K_s(G = 0) = 2\pi R_c^2 + \pi \text{erf}(\alpha R_c) \left( \frac{1}{\alpha^2} - 2R_c^2 \right) - \frac{2\sqrt{\pi} R_c e^{-(\alpha R_c)^2}}{\alpha}.
\]

(4.41)

The Faddeeva function

By Liouville’s theorem, \( \text{erf} z \) is unbounded for \( z \in \mathbb{C} \); in particular, it overflows floating-point arithmetic even for \( \mathbf{G} \)-vectors of relatively small modulus. Therefore, to use the sLOSC kernel in practice, we require a scaled implementation of the complex error function, namely the Faddeeva function

\[
w(z) = e^{-z^2} \text{erfc}(-iz).
\]

(4.42)

In particular, we seek to replace

\[
e^{-G^2/4\alpha^2} \text{Re} \left\{ \text{erf} \left( \alpha R_c + \frac{iG}{2\alpha} \right) \right\}
\]

by a term containing \( w(z) \). As mentioned in the main text, we implement the Faddeeva function using ACM Algorithm 916 [154], [155].

Note that whenever \( \text{Re} z > 0 \) we may write

\[
\text{erf} z = 1 - e^{-z^2} w(iz);
\]

(4.43)
in our case, \( \text{Re} z = \alpha R_c > 0 \). Thus, we may write

\[
e^{-G^2/4\alpha^2} \text{Re}\left\{ \text{erf}\left( \frac{\alpha R_c + iG}{2\alpha} \right) \right\}
\]

\[
= e^{-G^2/4\alpha^2} \text{Re}\left\{ 1 - e^{-\left(\alpha R_c + iG/2\alpha\right)^2} w \left( i\alpha R_c - \frac{G}{2\alpha} \right) \right\}
\]

\[
= e^{-G^2/4\alpha^2} - \text{Re}\left\{ e^{-\left(\alpha R_c + iG/2\alpha\right)^2 - \left(G/2\alpha\right)^2} w \left( i\alpha R_c - \frac{G}{2\alpha} \right) \right\}
\]

\[
= e^{-G^2/4\alpha^2} - \text{Re}\left\{ e^{-\left(\alpha R_c\right)^2} e^{-i\alpha R_c} w \left( i\alpha R_c - \frac{G}{2\alpha} \right) \right\}
\]

\[
= e^{-G^2/4\alpha^2} - e^{-\left(\alpha R_c\right)^2} \text{Re}\left\{ e^{-i\alpha R_c} w \left( i\alpha R_c - \frac{G}{2\alpha} \right) \right\}.
\]

Breaking the Faddeeva function into its real and imaginary parts as

\[
w \left( i\alpha R_c - \frac{G}{2\alpha} \right) = V + iL,
\]

we obtain

\[
\text{Re}\left\{ e^{-iGR_c} w \left( i\alpha R_c - \frac{G}{2\alpha} \right) \right\} = \text{Re}\left\{ \left[ \cos(\alpha R_c) - i \sin(\alpha R_c) \right] w \left( i\alpha R_c - \frac{G}{2\alpha} \right) \right\}
\]

\[
= \text{Re}\left\{ \left[ \cos(\alpha R_c) - i \sin(\alpha R_c) \right] (V + iL) \right\}
\]

\[
= V \cos(\alpha R_c) + L \sin(\alpha R_c).
\]

The Fourier coefficients of the sLOSC kernel as implemented in our code are thus

\[
K_s(G) = \frac{4\pi}{G^2} \left\{ 1 - \cos(\alpha R_c) \text{erfc}(\alpha R_c) - e^{-G^2/4\alpha^2} + e^{-\left(\alpha R_c\right)^2} \times
\right.
\]

\[
\left[ \text{Re}\left\{ w \left( i\alpha R_c - \frac{G}{2\alpha} \right) \right\} \cos(\alpha R_c) + \text{Im}\left\{ \left( i\alpha R_c - \frac{G}{2\alpha} \right) \sin(\alpha R_c) \right\} \right];
\]

\[
K_s(G = 0),\text{ having only a real argument to the error function, is implemented as written in Eq. (4.41).}
\]
4.5.3 Makov-Payne

One of the first methods to correct for non-neutral charge distributions is known as the Makov-Payne correction [145]. This method was developed to correct for skipping the $G = 0$ by introducing a fictitious set of point charges with the same periodicity as the unit cell. The true answer is then obtained using an analytical infinite sum, which leads to a constant correction to the first order error. This method of correction can also be viewed as correcting the effects due to the finite size of discrete Fourier transform. If we use the discrete Fourier transform to represent an aperiodic function, this is done by only looking at a finite interval upon which to perform the Fourier transform. The validity of this picture then only applies to this finite interval, since the Fourier representation of the function will periodically repeat itself outside the finite interval. This is the reason why evaluating the Coulomb energy using Eq. (4.29) only works exactly for charge-neutral distributions. If the distribution was not charge neutral, then the interaction of an infinitely large charge distribution is infinity. However, if we wish to use the linear basis set scaling of Eq. (4.29), then we may use the Makov-Payne scheme to remove the effects of representing an aperiodic function on a periodic interval.

In order to obtain a charge distribution that is charge neutral, we can place a point charge with of opposite charge at the center of our charge distribution:

$$
\rho_0(r) = \rho_c(r) - \rho_1(r); \quad \rho_1(r) = q\delta(r - r_0),
$$

where $\rho_0$ is the neutral distribution, $\rho_c$ is the charged distribution, $\rho_1$ is the fictitious counter charge, and $r_0$ is the dipole moment of $\rho_c$. By placing the counter-charge at the center of $\rho_c$, this ensures $\rho_0$ has no dipole moment. The original charge distribution can then be recovered as $\rho_c(r) = \rho_0(r) + \rho_1(r)$. We may the write then electrostatic energy due to the charged distribution as

$$
J[\rho_c, \rho_c] = J[\rho_0 + \rho_1, \rho_0 + \rho_1] = J[\rho_0, \rho_0] + J[\rho_1, \rho_1] + 2J[\rho_0, \rho_1]
$$

(4.49)
The cross terms are equivalent due the symmetry relation $J[\rho_i, \rho_j] = J[\rho_j, \rho_i]$. The $J[\rho_1, \rho_1]$ term is the energy due to a lattice interacting with itself. $J[\rho_0, \rho_0]$ is the self-energy of for a charge-neutral distribution. The cross term $J[\rho_0, \rho_1]$ is the interaction between the charge neutral $\rho_0$ and the lattice of counter charges. The $J[\rho_1, \rho_1]$ term is analytically summable for regular periodic lattices and leads to the result

$$J[\rho_1, \rho_1] = -\frac{\alpha q^2}{2L}, \quad (4.50)$$

where $\alpha$ is the lattice-dependent Madelung constant and $L$ is the unit of periodicity. A table of high precision Madelung constants for an assortment of lattices and their method of calculation is shown in [147]. Dabo et al.’s calculation of the Madelung constants improves upon the original formulation by deriving a faster converging series. The $J[\rho_0, \rho_0]$ of the charge neutral distribution can be calculated using Eq. (4.29). That cross term is given by Makov and Payne as

$$J[\rho_0, \rho_1] = \lim_{s \to 0} \int_\Omega d\mathbf{r} \rho_1(\mathbf{r}) \int_\Omega d\mathbf{r}' \rho_0(\mathbf{r}') \zeta(\mathbf{r} - \mathbf{r}', s), \quad (4.51)$$

where $\zeta(\mathbf{r} - \mathbf{r}', s)$ is defined as

$$\zeta(\mathbf{x}, s) = \sum_{\mathbf{R}} e^{-s|\mathbf{x} + \mathbf{R}|}. \quad (4.52)$$

This summation is over all unit cells to infinity. The way Makov and Payne evaluate sums of this sort is to split the sum into shells of terms that are all equidistant from the origin. This double sum formulation, however, is conditionally convergent and is therefore treated by introducing a convergence parameter akin to an Ewald summation technique, which ensures convergence to a definite limit. See Eqs. (A3) to (A9) of [145] for details. The final result of considering the absolutely convergent limit is that the cross term is

$$J[\rho_0, \rho_1] = \frac{2\pi q}{3\Omega} \int_\Omega d\mathbf{r} \rho_0(\mathbf{r}) \mathbf{r}^2 + O(L^{-5}). \quad (4.53)$$
Taking the quadrupole moment of \( \rho_0 \) to be \( Q = q \int d\mathbf{r} r^2 \rho(\mathbf{r}) \), we now arrive at the Makov-Payne correction scheme for charged distributions:

\[
J[\rho, \rho] = J_0[\rho, \rho] - \frac{\alpha q^2}{2L} + \frac{2\pi qQ}{3L^3} + O(L^{-5}).
\] (4.54)

This method is accurate up to order \( O(L^{-5}) \) in periodicity length. This term has a clear problem though since the correction term proportional to \( Q \) is origin dependent. This suggests that the quadrupole correction term will grow as the charge distribution moves away from the origin. Since the other two terms will be unchanged by a translation of the charge distribtuion, this indicates a failure in the theory. In the erratum of [147], the origin-independent form of this correction is given as

\[
J[\rho, \rho] = J_0[\rho, \rho] - \frac{\alpha q^2}{2L} + \frac{2\pi}{3L^3}(p^2 - qQ),
\] (4.55)

where \( p \) is the dipole moment \( p = q \int d\mathbf{r} \mathbf{r} \rho(\mathbf{r}) \). This form of correction is origin independent since \( (p^2 - qQ) \) is proportional to the variance of the charge distribution. See Fig. 4.5 for how this method performs. This concludes the Makov-Payne correction method for finding the Coulomb self-energy of a charge distribution that is not charge neutral and aperiodic.

The Makov-Payne correction and additions by Dabo et al. only offer a correction for the self-energy of a charged distribution. To find the repulsion energy between two different charge distributions we can use the fact that this is equivalent to the cross term of the self-energy for the summed distribution as shown in Eq. (4.49), i.e. \( J[\rho_i, \rho_j] = \frac{1}{2} \left( J[\rho_i + \rho_j, \rho_i + \rho_j] - J[\rho_i, \rho_i] - J[\rho_j, \rho_j] \right) \). We can evaluate the self-energy of the combined density and use this fact to determine the correction to Coulomb interactions between different charge distributions. The first thing to note is that since the sum of Fourier coefficient method in Eq. (4.29) is linear and symmetric for real-valued functions, the equivalent plane wave sum for \( J_0[\rho_i, \rho_j] \) is the same as Eq. (4.29). Defining the correction terms as \( \Delta J = J - J_0 \), we can then expand the correction terms of the sum and the
self-energy of the individual and combine them:

$$\Delta J[\rho_i, \rho_j] = \frac{1}{2} (\Delta J[\rho_i + \rho_j, \rho_i + \rho_j] - \Delta J[\rho_i, \rho_i] - \Delta J[\rho_j, \rho_j]). \quad (4.56)$$

Expanding all these terms at once would be cumbersome. To avoid this we will expand the terms that are proportional to $L^{-1}$ first,

$$\frac{\alpha}{2L} \left\{ (q_i + q_j)^2 - q_i^2 - q_j^2 \right\}$$

$$= \frac{\alpha}{2L} \left\{ 2q_i q_j \right\} = \frac{\alpha q_i q_j}{L}. \quad (4.57)$$

Moving on to the terms proportional to $L^{-3}$ we have

$$\frac{2\pi}{3L^3} \left\{ (p_i + p_j)^2 - (q_i + q_j)Q[\rho_i + \rho_j] - (p_i^2 - q_i Q[\rho_i]) - (p_j^2 - q_j Q[\rho_j]) \right\}$$

$$= \frac{2\pi}{3L^3} \left\{ 2p_i \cdot p_j - q_j Q[\rho_i] - q_i Q[\rho_j] \right\} \quad (4.58)$$

We now write the expansion of the dipole moment as $p_i = q_i \langle r \rangle_i$ and the quadrapole moment expansion as $Q[\rho_i] = q_i \langle r^2 \rangle_i$, where $\langle O \rangle_i = \int d\mathbf{r} O(\mathbf{r}) \rho_i(\mathbf{r})$ is the expectation of $O$ with respect to density $i$. This allows us to simplify Eq. (4.58) as

$$\frac{2\pi}{3L^3} \left\{ 2p_i \cdot p_j - q_j Q[\rho_i] - q_i Q[\rho_j] \right\}$$

$$= \frac{2\pi q_i q_j}{3L^3} \left\{ 2\langle r \rangle_i \cdot \langle r \rangle_j - \langle r^2 \rangle_i - \langle r^2 \rangle_j \right\} \quad (4.59)$$

If we now assume the two charge densities are orbital densities with $q_i = q_j$, we may write the Coulomb repulsion between two densities as

$$J_{MP3}[\rho_i, \rho_j] = J_0[\rho_i, \rho_j] - \frac{\alpha q^2}{2L} + \frac{2\pi q^2}{3L^3} \left( \langle r \rangle_i \cdot \langle r \rangle_j - \frac{1}{2}(\langle r^2 \rangle_i + \langle r^2 \rangle_j) \right) \quad (4.60)$$

The form of this correction passes the sanity test of being origin independent, which can be proved by evaluating the terms proportional to $L^{-3}$ with a constant shift, all the constant shift terms will cancel upon simplification. Additionally, this form reduces to the original Dabo corrected Makov-Payne scheme, Eq. (4.55), when $\rho_i = \rho_j$. When only the first order
correction proportional to $L^{-1}$ in Eq. (4.60) is used we refer to this as MP1. When the third order correction terms are included proportional to $L^{-3}$ we refer to this as MP3. For the accuracy of this method when considering interaction energies, see Fig. 4.6.

The upside to the Makov-Payne correction scheme is that the cost is negligible for MP1. If the average position and spatial variance is known for the densities, then the MP3 correction is also of negligible computational cost. This makes them very appealing as a correction scheme in the plane wave basis. The downside of the Makov-Payne scheme is the reliance of the correction on a 1D periodicity scale; therefore, the scheme does not treat anisotropic unit cells correctly. The farther from cubic the unit cell is, the less accurate the Makov-Payne scheme becomes. See Fig 4.4 for how the madelung constant behaves as the level of anisotropy varies. Alternative methods for non-standard geometries have been published, but they are not explored in this work [146].

**Figure 4.4:** Madelung constant calculated for orthorhombic unit cells using the method detailed in [147]. $A$, $B$, and $C$ refer to the length of periodicity in the $x$, $y$, and $z$ Cartesian directions. The reason stretching different directions, e.g. stretching $x$ versus stretching $z$, results in different Madelung constants is because the direction of periodicity is always chosen to be the $x$ dimension.
Figure 4.5: Coulomb self-energy of Gaussian type orbitals with various spatial variances evaluated in the plane wave basis. MP1 refers to the first order Makov-Payne correction, MP3 refers to the third order Makov-Payne correction, and CUT refers to the spherical cutoff technique. The $d$-type Gaussian have Cartesian angular momentum of $(2, 0, 0)$ corresponding to $p_y^2$ orbitals. Figure (a) is for a cubic unit cell and figure (b) is an orthorhombic unit cell where the $x$ dimensional periodicity is stretched and the $y$ and $z$ periodicity are fixed at 10Å.

Figure 4.6: Coulomb interaction energy between an $s$ and $d$-type Gaussians type orbital that both have a spatial variance of $1.0Å^2$, evaluated in the plane wave basis. The line dash types are the same as Fig 4.5. The $d$-type Gaussian have Cartesian angular momentum of $(2, 0, 0)$ corresponding to $p_y^2$ orbitals. Figure (a) is for a cubic unit cell and figure (b) is an orthorhombic unit cell where the $x$ dimensional periodicity is stretched and the $y$ and $z$ periodicity are fixed at 10Å. The various line colors refer to different distances between the relative center of the orbitals.
4.5.4 Choosing the screening parameter

The sLOSC method’s utility in calculating the band gap of bulk materials comes at the cost of one additional empirical parameter, the screening radius $\alpha$. To choose $\alpha$, we computed the band gaps of the full dataset below, varying $\alpha$ from 0 (unscreened) to $0.25 \, a_0^{-1}$. We selected the value minimizing the mean absolute percentage error (MAPE) of the bulk systems’ band gaps relative to experimentally determined values (for references, see the tables below). As can be seen from Fig. 4.7, the optimum for bulk systems is $\alpha = 0.15 a_0^{-1}$; we choose this value for the screening in sLOSC.

We also show the MAPE for a test set of small molecules’ fundamental gaps (ionization potential minus electron affinity), relative to CCSD(T) calculations. The smallest MAPE is given by LOSC2 [45], with no screening; however, even with $\alpha = 0.15 a_0^{-1}$ as in sLOSC, the MAPE is substantially improved compared to the uncorrected PBE computation.

![Figure 4.7](image-url): MAPE of the band gap of bulk (orange) and molecular (blue) systems, as $\alpha$ is varied. The dashed horizontal lines give the MAPE of the uncorrected DFA (PBE).
4.5.5 Supplemental sLOSC data

sLOSC is evaluated on a modified version of the SC/40 test set found in [26]. We repeat here our computational details for completeness. All calculations were performed using the PBE density functional [25] with optimized norm-conserving Vanderbilt pseudopotentials (ONCVPSP) [130] sourced from the PseudoDojo [131]; they were carried out in the Quantum ESPRESSO suite [132], [133] and a modified version of the wannier90 code [125], [126]. The kinetic energy cutoff was set to 100 Ry for wavefunctions, and four times greater for the density. For self-consistent calculations, a 16×16×16 Monkhorst–Pack k-mesh was used to ensure an accurate Kohn–Sham density matrix; for the localization and energy correction, a 6×6×6 k-mesh was used.

For every system in which [26] lists an experimental lattice constant, we use it; the other lattice constants are sourced from [158]. Experimental band gaps are also sourced from [26] whenever they are available. For the other systems, the experimental band gaps are sourced from [159] (LiF, LiCl, Ne, Ar); [160] (NaF); [161] (NaCl); and [162] (CaS, CaSe, SrS, SrSe).

**Table 4.2**: The data presented in Fig. 4.1. The second column gives the Strukturbericht designation for each system.

<table>
<thead>
<tr>
<th>System parameters</th>
<th>Band gaps (eV)</th>
<th>Energy (eV)</th>
<th>E_{DFA}</th>
<th>∆E_{sLOSC}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula</td>
<td>Lattice</td>
<td>Constant (Å)</td>
<td>DFA</td>
<td>sLOSC</td>
</tr>
<tr>
<td>AlAs</td>
<td>B3</td>
<td>5.661</td>
<td>1.44</td>
<td>2.957</td>
</tr>
<tr>
<td>AlP</td>
<td>B3</td>
<td>5.463</td>
<td>1.58</td>
<td>3.068</td>
</tr>
<tr>
<td>AlSb</td>
<td>B3</td>
<td>6.136</td>
<td>1.221</td>
<td>2.301</td>
</tr>
<tr>
<td>BaS</td>
<td>B1</td>
<td>4.777</td>
<td>2.148</td>
<td>4.021</td>
</tr>
<tr>
<td>BAs</td>
<td>B3</td>
<td>6.389</td>
<td>1.294</td>
<td>1.976</td>
</tr>
<tr>
<td>BaSe</td>
<td>B1</td>
<td>6.595</td>
<td>1.922</td>
<td>3.530</td>
</tr>
<tr>
<td>BaTe</td>
<td>B1</td>
<td>7.007</td>
<td>1.561</td>
<td>2.788</td>
</tr>
<tr>
<td>BN</td>
<td>B3</td>
<td>3.616</td>
<td>4.533</td>
<td>6.528</td>
</tr>
<tr>
<td>BP</td>
<td>B3</td>
<td>4.538</td>
<td>1.348</td>
<td>2.072</td>
</tr>
<tr>
<td>C</td>
<td>A4</td>
<td>3.567</td>
<td>4.206</td>
<td>5.219</td>
</tr>
<tr>
<td>CaS</td>
<td>B1</td>
<td>5.689</td>
<td>2.358</td>
<td>5.441</td>
</tr>
<tr>
<td>CaSe</td>
<td>B1</td>
<td>5.916</td>
<td>2.035</td>
<td>4.814</td>
</tr>
<tr>
<td>CdS</td>
<td>B3</td>
<td>5.818</td>
<td>1.163</td>
<td>2.310</td>
</tr>
<tr>
<td>CdSe</td>
<td>B3</td>
<td>6.052</td>
<td>0.639</td>
<td>1.541</td>
</tr>
<tr>
<td>CdTe</td>
<td>B3</td>
<td>6.48</td>
<td>0.773</td>
<td>1.349</td>
</tr>
<tr>
<td>GaAs</td>
<td>B3</td>
<td>5.648</td>
<td>0.538</td>
<td>1.215</td>
</tr>
</tbody>
</table>
We additionally compute sLOSC results on a small set of molecular systems using the in-house QM\textsuperscript{4}D code. We use the PBE exchange-correlation functional, as in the bulk case, and the 6-311++G(3df,3pd) basis set, with the Dunning augmented correlation-consistent triple-zeta (aug-cc-pVTZ) basis used for density fitting.

Instead of experimental values, we use CCSD(T) computations for the reference fundamental gaps. These are computed under both the aug-cc-pVTZ and aug-cc-pVQZ (Dunning augmented correlation-consistent triple- and quadruple-zeta) basis sets and extrapolated using the formula from Eq. (44) of [163] as

\[ \epsilon_{\infty} = \frac{\epsilon_{\text{TZ}} \times 3^3 - \epsilon_{\text{QZ}} \times 4^3}{3^3 - 4^3}, \]  

(4.61)

where \( \epsilon_{nZ} \) is the aug-cc-pVnZ eigenvalue.

<table>
<thead>
<tr>
<th>System parameters</th>
<th>Band gaps (eV)</th>
<th>Energy (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula</td>
<td>Lattice</td>
<td>Constant (Å)</td>
</tr>
<tr>
<td>( \beta )-GaN</td>
<td>B3</td>
<td>4.523</td>
</tr>
<tr>
<td>GaP</td>
<td>B3</td>
<td>5.451</td>
</tr>
<tr>
<td>GaSb</td>
<td>B3</td>
<td>6.096</td>
</tr>
<tr>
<td>Ge</td>
<td>A4</td>
<td>5.658</td>
</tr>
<tr>
<td>InAs</td>
<td>B3</td>
<td>6.058</td>
</tr>
<tr>
<td>InP</td>
<td>B3</td>
<td>5.869</td>
</tr>
<tr>
<td>InSb</td>
<td>B3</td>
<td>6.479</td>
</tr>
<tr>
<td>LiCl</td>
<td>B1</td>
<td>5.13</td>
</tr>
<tr>
<td>LiF</td>
<td>B1</td>
<td>4.017</td>
</tr>
<tr>
<td>MgO</td>
<td>B1</td>
<td>4.207</td>
</tr>
<tr>
<td>MgS</td>
<td>B3</td>
<td>5.622</td>
</tr>
<tr>
<td>MgSe</td>
<td>B3</td>
<td>5.4</td>
</tr>
<tr>
<td>MgTe</td>
<td>B3</td>
<td>6.42</td>
</tr>
<tr>
<td>NaCl</td>
<td>B1</td>
<td>5.641</td>
</tr>
<tr>
<td>NaF</td>
<td>B1</td>
<td>4.62</td>
</tr>
<tr>
<td>Si</td>
<td>A4</td>
<td>5.43</td>
</tr>
<tr>
<td>SiC</td>
<td>A4</td>
<td>4.358</td>
</tr>
<tr>
<td>SrS</td>
<td>B1</td>
<td>5.99</td>
</tr>
<tr>
<td>SrSe</td>
<td>B1</td>
<td>6.234</td>
</tr>
<tr>
<td>ZnS</td>
<td>B3</td>
<td>5.409</td>
</tr>
<tr>
<td>ZnSe</td>
<td>B3</td>
<td>5.668</td>
</tr>
<tr>
<td>ZnTe</td>
<td>B3</td>
<td>6.089</td>
</tr>
<tr>
<td>Ar</td>
<td>A1</td>
<td>5.256</td>
</tr>
<tr>
<td>Ne</td>
<td>A1</td>
<td>4.429</td>
</tr>
</tbody>
</table>
Table 4.3: The molecular sLOSC data.

<table>
<thead>
<tr>
<th>Formula</th>
<th>Band gaps (eV)</th>
<th>Energy (eV)</th>
<th>$E_{DFA}$</th>
<th>$\Delta E_{sLOSC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PH</td>
<td>2.213</td>
<td>4.728</td>
<td>9.16</td>
<td>-9298.83 × 10^-4</td>
</tr>
<tr>
<td>PH₂</td>
<td>1.844</td>
<td>4.349</td>
<td>8.56</td>
<td>-9315.88 × 10^-4</td>
</tr>
<tr>
<td>SH</td>
<td>0.430</td>
<td>3.676</td>
<td>8.05</td>
<td>-10845.98 × 10^-4</td>
</tr>
<tr>
<td>O₂</td>
<td>2.306</td>
<td>6.959</td>
<td>12.51</td>
<td>-4088.32 × 10^-5</td>
</tr>
<tr>
<td>CN</td>
<td>1.709</td>
<td>6.125</td>
<td>10.29</td>
<td>-2520.88 × 10^-4</td>
</tr>
<tr>
<td>S₂</td>
<td>1.292</td>
<td>2.497</td>
<td>7.92</td>
<td>-21662.10 × 10^-4</td>
</tr>
<tr>
<td>Cl₂</td>
<td>2.740</td>
<td>4.395</td>
<td>10.61</td>
<td>-25035.22 × 10^-4</td>
</tr>
<tr>
<td>CH₂</td>
<td>2.413</td>
<td>6.845</td>
<td>9.57</td>
<td>-6464.52 × 10^-4</td>
</tr>
<tr>
<td>CF₂</td>
<td>3.626</td>
<td>7.544</td>
<td>12.68</td>
<td>-1064.09 × 10^-4</td>
</tr>
<tr>
<td>CH₃</td>
<td>2.472</td>
<td>6.531</td>
<td>9.84</td>
<td>-3095.93 × 10^-3</td>
</tr>
<tr>
<td>CHO</td>
<td>1.693</td>
<td>4.568</td>
<td>9.44</td>
<td>-3128.24 × 10^-3</td>
</tr>
<tr>
<td>CH₃O</td>
<td>0.723</td>
<td>6.102</td>
<td>9.51</td>
<td>-11914.33 × 10^-4</td>
</tr>
<tr>
<td>CH₃S</td>
<td>1.567</td>
<td>3.743</td>
<td>5.91</td>
<td>-11898.66 × 10^-4</td>
</tr>
<tr>
<td>CH₂S</td>
<td>1.891</td>
<td>4.067</td>
<td>9.16</td>
<td>-1082.70 × 10^-3</td>
</tr>
<tr>
<td>NH</td>
<td>3.597</td>
<td>9.751</td>
<td>13.15</td>
<td>-1501.30 × 10^-4</td>
</tr>
<tr>
<td>NH₂</td>
<td>2.738</td>
<td>8.527</td>
<td>11.37</td>
<td>-1519.24 × 10^-3</td>
</tr>
<tr>
<td>OH</td>
<td>0.916</td>
<td>8.599</td>
<td>11.24</td>
<td>-2059.42 × 10^-4</td>
</tr>
<tr>
<td>SiH₂</td>
<td>1.813</td>
<td>3.671</td>
<td>8.48</td>
<td>-7903.86 × 10^-4</td>
</tr>
<tr>
<td>SiH₃</td>
<td>1.828</td>
<td>3.924</td>
<td>7.93</td>
<td>-7920.69 × 10^-3</td>
</tr>
</tbody>
</table>

We plot the data from Table 4.3 in Fig. 4.8, together with PBE and unscreened LOSC results, in analogy with Fig. 1 in the main text. We can see again that the best gaps are uniformly those computed without screening, in agreement with Fig. 4.7 and as expected for small molecules with open boundary conditions; however, sLOSC still improves the accuracy of the gap compared to the PBE calculation.
Figure 4.8: Comparison of experimental fundamental gaps of molecular systems show in Table 4.3 with those calculated by PBE (○), sLOSC (×), and unscreened LOSC (+).
Chapter 5

Conclusion

As we have noted throughout this work, DFAs suffer from delocalization error. In molecules, delocalization error is characterized by a convex deviation of the total energy from the linear interpolation with respect to fractional charge that should be obeyed by an exact theory. In bulk calculations the delocalized nature of the orbitals produces a linear total energy for fractional charges but with an incorrect slope due to delocalization error. These facts lead DFAs to incorrectly predict the IPs, EAs, fundamental gaps, dissociation curves, and reaction barriers.

In chapter 2 we noted how LOSC is able to greatly improve the dissociation of cationic species, the band gaps of small sized molecules and polymers, and photoemission spectra [41]. We then showed how we improved the LOSC method to more rigorously treat system symmetries, thereby preserving the parent functional energy degeneracy patterns when possible. This is accomplished by employing a cost function that treats the space and energy costs symmetrically by adding the variance of each. The tests on planar annulenes, odd-numbered allenes, and triphenylene showed that this improved version of LOSC can provide symmetric localized orbitals and maintain state degeneracy, while the original implementation cannot. An additional change to the curvature along with the new localization provides a smoother dissociation curve, which is more physically reasonable.

Using the cost function we established in chapter 2, we extended the theory and implementation to periodic systems in chapter 3 using Wannier functions. Then we went into detail on how the localization in energy allows for the simultaneous localization of occupied and unoccupied states, something not prescribed in previous literature. The resulting DL-WFs are therefore able to produce functions similar in character to those found from the MLWF procedure applied to the valence bands of a semiconductor, while also being able
to obtain localized virtual states at the same time. This type of localization is thus able to produce a picture that is more relevant for chemical bonding since it doesn’t explicitly rely on separating the occupied and unoccupied spaces.

Finally, in chapter 4 we used the DLWFs from chapter 3 to implement the LOSC method in PBCs. We showed how a straightforward implementation using the same form of correction used for molecules results in a large overcorrection when considering a test set of semiconductors. To account for this we introduced a screened Coulomb operator into the curvature calculation; we named this method sLOSC. Using sLOSC on a test set that included small gapped semiconductors up to large gapped insulators we showed how sLOSC is able to improve the band gap on a systematic basis compared to the parent functional. We note that sLOSC does degrade the performance of molecules; however, it still outperforms the parent functional in all systems.

5.1 Future Work

The original work introducing the LOSC method was implemented as a post-processing step to SCF calculations, but it proposed a method to use LOSC in the SCF cycle. In this work, we implemented the energy correction as a post-processing step to a self-consistent calculation as well; such corrections are accurate when the total energy correction is small. For all systems considered in the PBC results, $\Delta E^{\text{LOSC}}$ is less than one part in $10^5$. The corresponding change to the density for such systems is also expected to be minimal. A self-consistent implementation for molecules was realized in [156], albeit with a different cycle than the one originally proposed, which uses a frozen localization that is not updated in the self-consistent cycle. This self-consistent implementation further improves the predictions of LOSC for systems where DFAs fail to describe the electron density correctly. A self-consistent implementation of sLOSC could be necessary for accurate computation of metallic band structures; since metals will be expected to produce DLWFs with occupancies far from 0 or 1, the sLOSC energy correction is likely to be larger, and iteratively
correcting the delocalized density is expected to yield better orbital energies.

Additionally, the code for the molecular LOSC method in all published incarnations has been made into a library with an API. The code that implements the DLWFs has been implemented in the latest developer version of the wannier90 code, which is planned to be released with an upcoming publication. The code for calculating the energy corrections in PBC using Quantum ESPRESSO was implemented in version 6.3, while the most current release is on version 6.8. Due to the modular implementation of LOSC in Quantum ESPRESSO as a post-processing step, moving to the newest version for public release should be a straightforward task. While the current implementation can run on output from SCF runs up to version 6.5, later versions are not compatible due to the software storing the wavefunctions in the plane-wave basis instead of on the real-space grid. The most expensive step in bulk LOSC to calculate the curvature elements which scales as $O(N_{w}^2N_{G})$ where $N_{w}$ is the number of Wannier functions and $N_{G}$ is the number of plane waves in the unfolded supercell. When $N_{k}$ $\mathbf{k}$-points are used to sample the unit cell then $N_{G}$ is $N_{k}$ times larger than the number of plane waves in the unit cell. We note that the size of the supercell used to numerically integrate the curvature elements could be reduced from the full Born–von Karman supercell if the relevant DLWF densities are contained in a smaller region. This is supported by the fact that systems that had quantitatively similar localizations for a $4 \times 4 \times 4$ and $6 \times 6 \times 6$ $\mathbf{k}$-mesh yielded very similar energy corrections. This would reduce the scaling of the calculation for the curvature elements to $O(N_{w}^2N_{G}/N_{k})$, since the size of the cell used for integration would be independent of the size of the unfolded supercell. Additionally, if only the correction to the band gap of the system is desired then the curvature only needs to be calculated for a small number of bands around the Fermi energy; this would effectively remove the $O(N_{w}^2)$ factor in the curvature calculation scaling complexity.

One of the most promising applications for LOSC is its use for heterogeneous systems such as interfaces and molecules on surfaces. The problem of energy level alignment at the interface of different materials is especially important for technological innovation.
Methods such as the screened range-separated hybrid that are not empirical use an optimally tuned range separation parameter and the dielectric constant to correct for delocalization error are unable to address heterogeneous systems since they are not reducible to a simple unit cell. We note that we also tested using the dielectric constant to screen the curvature for the semiconductors tested in chapter 4; while the results were better than the unscreened case they were not as good as the optimally tuned screening. In order to accurately address these types of heterogeneous systems, a screened interaction must be found that is more detailed than just the system average.

Recently it was shown how the exact second derivative of the energy with respect to occupation is given by a screened interaction [39]. Using an RPA calculation the dielectric response function could be found, which could be used to accurately model the screened interaction in the exact second derivative. The dielectric response function however is a two index quantity of position, which requires $O(N_G^2)$ points on the unfolded supercell. A more computationally tractable approach would be to use linear response to find the exact curvature. Either of these approaches would allow for a non-empirical method to correct for the delocalization error in general, including heterogeneous systems such as material interfaces.
Bibliography


[40] C. Li and W. Yang, “On the piecewise convex or concave nature of ground state energy as a function of fractional number of electrons for approximate


139


[92] Qm4d, an in-house program for qm/mm simulations (http://www.qm4d.info), 2018.


pp. 10964–10972, 46 2009, see this reference for experimental IPs of gas phase trans-polyacetylene oligomers for \(N=1\text{-}4\) and RASPT2 results for \(N=1\text{-}10\).


147


152