Scalable computational approach to extract chemical bonding from real-space density functional theory calculations using finite-element basis: A projected orbital and Hamiltonian population analysis

Kartick Ramakrishnan,† Sai Krishna Kishore Nori,† Seung-Cheol Lee,‡ Gour P Das,¶ Satadeep Bhattacharjee,‡ and Phani Motamarri*,†

†Department of Computational and Data Sciences, Indian Institute of Science, Bangalore
‡Indo Korea Science and Technology Center, Bangalore, India
¶Research Institute for Sustainable Energy (RISE), TCG Center for Research and Education in Science and Technology, Salt Lake, Kolkata, India

E-mail: phanim@iisc.ac.in

Abstract

We present an efficient and scalable computational approach for conducting projected population analysis from real-space finite-element (FE) based Kohn-Sham density functional theory calculations (DFT–FE). This work provides an important direction towards extracting chemical bonding information from large-scale DFT calculations on materials systems involving thousands of atoms while accommodating periodic, semi-periodic or fully non-periodic boundary conditions. Towards this, we derive the relevant mathematical expressions and develop efficient numerical implementation procedures that are scalable on multi-node CPU architectures to compute the projected overlap and Hamilton populations. This is accomplished by projecting either
the self-consistently converged FE discretized Kohn-Sham eigenstates, or the FE discretized Hamiltonian onto a subspace spanned by localized atom-centered basis set. The proposed method is implemented in a unified framework within DFT-FE where the ground-state DFT calculations and the population analysis are performed on the same finite-element grid. We further benchmark the accuracy and performance of this approach on representative material systems involving periodic and non-periodic DFT calculations with LOBSTER, a widely used projected population analysis code. Finally, we discuss a case study demonstrating the advantages of our scalable approach to extract the chemical bonding information from increasingly large silicon nanoparticles up to a few thousand atoms.

**Acronyms**

IFOHP Integrated finite-element orbital Hamilton population. 34–36

PA Pseudo-atomic. 26, 27, 29–31, 41, 43, 44

pCOHP projected crystal orbital-Hamilton population. 26, 34

pCOOP projected crystal orbital-overlap population. 26

pFHDE Projected finite-element Hamiltonian density error. 23, 31, 32, 34

pFHHHP Projected finite-element Hamiltonian Hamilton population. 15, 24, 31

pFHOHP Projected finite-element Hamiltonian overlap population. 14, 31

pFHP Projected finite-element Hamiltonian population. 5, 6, 23, 24, 30–34, 39, 47, 48

pFODE Projected finite-element orbital density error. 20, 31, 32, 34

pFOOHP Projected finite-element orbital Hamilton population. 14, 28, 31, 41

pFOOP Projected finite-element orbital overlap population. 14, 28, 31, 41, 45, 46
1 Introduction

Chemical bonding in covalent materials is typically studied using overlap and Hamilton-based population analysis. Hughbanks et al. were the first to adopt an extended Hückel-based technique to introduce such methods. The overlap population analysis allows us to partition electrons among distinct atoms and the orbitals around them, whereas the Hamilton population analysis allows us to partition a molecule’s or crystal’s total electronic energy.

In the case of crystalline materials, schemes such as Crystal Orbital Overlap Population (COOP) and Crystal Orbital Hamilton Population (COHP) analysis are widely adopted. Historically, the first COOP and COHP-based methods were introduced within the framework of Tight-binding Linearized Muffin Tin Orbital in the Atomic Spheres Approximation (TB-LMTO-ASA). However, ASA is a good approximation only for close-packed structures. Furthermore, in the case of two-dimensional materials or materials with layered structures, TB-LMTO-ASA, despite its simplicity, needs to be extended to multiple $\kappa$-basis and further to NMTO basis. On the other hand, the plane wave based approaches have become the popular choice for electronic structure calculations these days due to the variational nature of the basis set, offering spectral convergence rates in the ground-state properties of interest. However, the orbital population analysis based on a localized atom-centred basis is more intuitive for understanding chemical bonding properties, and hence recent work has been focused on linking plane-wave-based DFT codes to localized basis sets by projecting the Kohn-Sham DFT eigenfunctions obtained from a plane-wave calculation into a subspace spanned by the localized atomic-orbital basis.

The most popular and widely used code based on such projected population analysis approach is LOBSTER. Here, the Kohn-Sham eigenfunctions obtained from a Projector Augmented Wave (PAW) based DFT calculations using popular plane-wave based codes
(e.g. VASP, Quantum Espresso) are projected onto a subspace spanned by a localized atom-centred basis. While such a strategy has been largely successful, this approach has certain limitations. Firstly, plane-wave based techniques often restrict the simulation domains to be periodic, which is incompatible with many application problems (e.g.: defects, nano-particles, charged systems). Furthermore, plane-wave basis provides uniform spatial resolution and is computationally inefficient in the study of defects, isolated systems (e.g. molecules, clusters etc.) where a higher resolution is necessary to describe particular regions of interest and a coarse resolution suffices elsewhere. Moreover, plane-wave basis are extended in real-space and involve non-local communication between processors, affecting the scalability of computations on massively parallel computing architectures thereby restricting the material system sizes that can be simulated to a few hundreds of atoms. LOBSTER, which uses plane-wave discretized Kohn-Sham wavefunctions as an input to conduct population analysis suffers from the above limitations and cannot be executed on more than 1 CPU node. Further, the use of multiple codes, such as the ground-state DFT calculation employing a plane-wave based code like VASP or quantum espresso, and the subsequent population analysis using LOBSTER code, makes the process cumbersome and time-consuming. Kundu et al. recently proposed a population analysis where the Kohn-Sham occupied eigenspace obtained from a plane-wave DFT calculations are projected onto a localized Wannier orbital basis, thereby minimizing the projection error from plane-wave to localized atom-centered basis (spill factor due to the completeness of the Wannier functions. However, such an approach still suffers from the plane-wave basis limitations and further adds to the complexity of population analysis by requiring the use of three codes to complete three tasks (ground-state DFT calculation by a plane-wave code, Wannierization (using wannier90), and finally the population analysis code).

Addressing most of the above limitations, we introduce here a real-space finite-element based density functional theory (DFT-FE) approach to conduct projected population analysis. This method can be viewed as a unified approach allowing for Kohn-Sham DFT
ground-state calculations and population analysis in the same computational framework using a finite-element (FE) basis. FE basis set is a systematically convergent basis set comprising of a piece-wise polynomial of order $p$ and a strictly local basis set on which various electronic fields are represented. In contrast to plane-wave based DFT calculations, the use of FE basis for DFT enables large-scale calculations (up to tens of thousands of electrons) and accommodates periodic, semi-periodic, and non-periodic boundary conditions while exhibiting systematically convergent behaviour similar to plane-wave basis. Furthermore, the local character of the FE basis provides an inherent benefit in terms of parallel scalability of DFT calculations in comparison to widely used DFT codes and has been tested up to 200,000 cores on many-core CPUs and 24,000 GPUs on hybrid CPU-GPU architectures and has also been named as the *ACM Gordon Bell Prize* finalist in SC19. The population analysis methodology developed within the framework of DFT-FE in this work inherits the advantages mentioned above and enables scalable chemical bonding analysis in complex material systems. For instance, efficient investigation of chemical bonding interactions in large-scale nanoparticles, charged defect systems, material systems in the presence of an external electric field, or changes in the bonding nature of the constituent atoms due to the effect of an electron beam, can be naturally handled using the proposed approach of population analysis.

We propose two methodologies for projected population analysis in this work – (a) projected finite-element orbital population analysis ($p$FOP) similar in spirit to LOBSTER and (b) projected finite-element Hamiltonian population analysis ($p$FHP). The computational framework developed to implement these methods hinges on the following key steps: (i) perform Kohn-Sham DFT ground-state calculation in DFT-FE to compute the finite-element discretized eigenfunctions spanning the Kohn-Sham occupied eigenspace, (ii) construct the subspace spanned by the localized atom-centered orbitals $\Psi^{N_{\text{orb}}}_{\phi}$ (available as numerical data or analytical expressions) -via- interpolating these orbitals on the underlying finite-element grid, (iii) orthogonally project the occupied Kohn-Sham eigenfunctions onto $\Psi^{N_{\text{orb}}}_{\phi}$ in the case
of pFOP, while orthogonally project the self-consistent Kohn-Sham FE discretized Hamiltonian onto $V^{N_{orb}}_\phi$ in the case of pFOP, (iv) compute the atom-centered orbital overlap matrix using the Gauss-Lobatto-Legendre quadrature rule, (v) compute the coefficient matrices corresponding to the representation of projected Kohn-Sham wavefunctions in the atom-centered orbital basis $V^{N_{orb}}_\phi$ in the case of pFOP, while diagonalizing the projected Hamiltonian to obtain the eigenvector matrix in the subspace $V^{N_{orb}}_\phi$ in the case of pFHP, (vi) using these coefficient matrices, evaluate the projected finite-element orbital overlap and Hamilton population in the case of pFOP, and evaluate the projected finite-element Hamiltonian overlap and Hamilton population in the case of pFHP.

We evaluate the accuracy and performance of the proposed methods pFOP and pFHP on representative benchmark examples involving isolated molecules (CO, H$_2$O, O$_2$, Si-H nanoparticles) and a periodic system involving carbon diamond structure with a large supercell. We first benchmark the results obtained by pFOP method with that obtained from LOBSTER code, and we find an excellent agreement with LOBSTER for the material system sizes feasible to run on LOBSTER. We also demonstrate the significant advantage of the pFOP approach in terms of computational time compared to LOBSTER even on 1 CPU-node on these material systems. Furthermore, we take advantage of our parallel implementation of pFOP using MPI and illustrate the reduction in wall-time of the population analysis by $\sim 70\%$ when scaled up to 1120 CPU cores from 280 CPU cores on a Si nanoparticle system containing 1090 atoms. We remark that these large-scale calculations are not currently feasible using LOBSTER. Subsequently, we compare the accuracy and performance of pFHP approach with that of pFOP. The results obtained by the pFHP approach agree very well with that obtained by the pFOP approach. We further show the advantage of using pFHP in computational wall time compared to pFOP on large-scale systems ($\approx 1100$-$2100$ atoms) by employing 280 - 4500 CPU cores.

Finally, we discuss a case study demonstrating the usefulness of the proposed computational framework in conducting large-scale bonding analysis. To this end, we consider the
case of the chemisorption of hydrogen in silicon nanoparticles alloyed with carbon, a candidate material for hydrogen storage. Towards this, we conduct projected population analysis and estimate the Si-Si and Si-H bond strength in increasing system sizes of Si nanoparticles with and without alloying ranging from 65 atoms to around 1000 atoms, and argue the ease of Si-Si dimerization with the increase in size of alloyed Si nanoparticles favouring the release of $\text{H}_2$.

The remainder of our manuscript is structured as follows: Section 2 discusses the mathematical formalism of the projected-orbital overlap and Hamilton population, the population analysis schemes equivalent to pCOOP and pCOHP introduced earlier in the literature but presented in an abstract setting. We then discuss the proposed alternate approach for population analysis involving the projected-Hamiltonian overlap and Hamilton populations. This section forms the segue for Section 3, which discusses the efficient numerical implementation details within the framework of finite-element discretization on parallel computing architectures. Section 4 presents the results on benchmark systems, assessing the accuracy and performance, and demonstrating the advantage of the proposed methods implemented in a unified and scalable framework within DFT-FE. We subsequently discuss a case study illustrating the usefulness of large-scale chemical bonding analysis, and finally conclude with a short discussion and outlook in Section 5.

2 Mathematical Formulation

In this section, we introduce the notations, discuss the key mathematical preliminaries and present the expressions for projected population analysis. We describe two approaches to compute projected overlap and Hamilton populations in this work. The first approach is referred to as projected-orbital population analysis henceforth and is in the spirit of Sanchez-Portal et al. and Deringer et al. The approach relies on the orthogonal projection of numerically computed Kohn-Sham DFT eigenfunctions onto a subspace spanned by local-
ized atomic-orbitals to extract the chemical bonding behavior. Furthermore, we introduce an alternate approach for conducting projected-population analysis by means of orthogonally projecting the discretized Kohn-Sham Hamiltonian onto the atomic-orbital subspace and, subsequently, diagonalizing the projected Hamiltonian. We refer to this approach as *projected-Hamiltonian population analysis* henceforth.

The mathematical formulation in this section is discussed in an abstract setting, making it general enough to be adapted to projected population analysis, wherein the Kohn-Sham DFT eigenvalue problem can be solved using any convenient choice of systematically convergent numerical basis set. This section constitutes the segue for the numerical implementation details presented in the subsequent section involving the finite-element basis employed in this work. For clarity and simplicity, the Kohn-Sham DFT eigenproblem considered in the discussion below is solved in a simulation domain with fully non-periodic boundary conditions or a supercell employing periodic/semi-periodic boundary conditions with Gamma point to sample the Brillouin zone. The extension to periodic unit-cell involving Brillouin zone integration -via- multiple k-point sampling is straightforward and not explicitly considered here. Further, we assume the spin-independent case of DFT and that the Kohn-Sham wavefunctions are real for our discussion below without any loss of generality.

Let $\mathcal{H}$ denote an infinite-dimensional Hilbert space, where we assume the Kohn-Sham eigenfunctions of the continuous problem exist. $\mathcal{H}$ is equipped with inner product $\langle \cdot | \cdot \rangle$ over the field of complex numbers $\mathbb{C}$, and consequently, a norm $\| \cdot \|$ induced from the inner product is defined. Let $\mathcal{H}$ be the Hermitian operator representing the Kohn-Sham Hamiltonian of interest defined on the $M$-dimensional subspace $\mathbb{V}^M \subset \mathcal{H}$. In other words, $\mathcal{H} \in \mathbb{C}^{M \times M}$ represents the discretized Kohn-Sham Hamiltonian operator in $\mathbb{V}^M$ spanned by a suitably chosen systematically converging basis set — plane waves, finite element basis, finite difference approach, wavelets, etc., all which can be employed to numerically solve the partial differential equation representing the Kohn-Sham DFT eigenvalue problem. Consequently, the discretized DFT eigenvalue problem to be solved for $N$-smallest eigenvalue-eigenvector
pairs is given by

\[ \mathcal{H} \left| \psi_i \right\rangle = \epsilon_i \left| \psi_i \right\rangle \quad \text{for} \quad i = 1, 2, \ldots N \quad \text{with} \quad N \geq \frac{N_e}{2} \]  

(1)

where \( \left| \psi_i \right\rangle \in \mathcal{V}^M \) denotes the eigenfunction of \( \mathcal{H} \) and \( N_e \) is the number of electrons in the given material system. We now define a \( N_{\text{orb}} \)-dimensional subspace \( \mathcal{V}^{N_{\text{orb}}} \) \( (N_{\text{orb}} < M) \), spanned by the localized non-orthogonal atom-centred auxiliary basis set \( \{ \left| \phi_{\mu} \right\rangle \} \), constructed from the given configuration of atoms corresponding to the material system under study. Consequently, the orthogonal projection operator \( \mathcal{P}^\phi : \mathcal{V}^M \rightarrow \mathcal{V}^{N_{\text{orb}}} \) can be written as

\[ \mathcal{P}^\phi = \sum_{\alpha, \beta=1}^{N_{\text{orb}}} O_{\alpha \beta}^{-1/2} \left| \psi^\phi_{\alpha} \right\rangle \]

where \( O_{\alpha \beta} = \langle \phi_{\alpha} | \phi_{\beta} \rangle \), denotes the matrix elements of the overlap matrix \( O \) corresponding to the basis set \( \{ \left| \phi_{\mu} \right\rangle \} \). We note that this atom-centred basis set spanning the subspace \( \mathcal{V}^{N_{\text{orb}}} \) are chosen to be minimal so that the occupied Kohn-Sham wavefunctions \( \left| \psi_i \right\rangle \) are represented reasonably well in \( \mathcal{V}^{N_{\text{orb}}} \) while providing accurate interpretation of chemical bonding behaviour. Bunge orbitals, \( ^{30} \) Slater type orbitals fitted to atomic functions, \( ^{31} \) functions fitted to PAW wavefunctions \( ^{14} \) have all been used in the past as a choice for these minimal atomic-orbital basis sets. Pseudo atomic-orbitals constructed from norm-conserving pseudopotentials \( ^{32} \) also constitute a convenient choice of atom-centred basis sets for chemical bonding analysis, as demonstrated in the current work.

2.1 Projected-orbital Population Analysis

Denoting the orthogonal projection of the Kohn-Sham eigenfunction \( \left| \psi_i \right\rangle \in \mathcal{V}^M \) onto the space \( \mathcal{V}^{N_{\text{orb}}} \) to be \( \left| \psi^\phi_i \right\rangle \), we have \( \left| \psi^\phi_i \right\rangle = \mathcal{P}^\phi \left| \psi_i \right\rangle \) \( (i = 1, 2 \cdots N) \). We note that the projected Kohn-Sham wavefunctions \( \left| \psi^\phi_i \right\rangle \) need not satisfy the orthonormality condition (even if \( \left| \psi_i \right\rangle \)'s are pairwise orthonormal) and hence Löwden orthogonalization is employed to orthonormalize the projected Kohn-Sham wavefunctions. \( ^{13} \) To this end, denoting the orthonormalized projected wavefunction as \( \left| \tilde{\psi}^\phi_i \right\rangle \) we have \( \left| \tilde{\psi}^\phi_i \right\rangle = \sum_{j}^{N_{\text{orb}}} O_{ij}^{-1/2} \left| \psi^\phi_j \right\rangle \), where \( O_{ij} = \langle \psi^\phi_i | \psi^\phi_j \rangle \), denotes the matrix elements of the overlap matrix \( O \) corresponding to the projected Kohn
Sham wavefunctions $\{|\psi^\phi_i\rangle\}$. We now proceed to derive the expressions for pOOP and pOHP using the nomenclature introduced above.

**Projected-orbital Overlap Population (pOOP):** We begin by introducing the normalization condition involving the orthonormalized projected wavefunctions, i.e. $\langle \tilde{\psi}_j^\phi | \tilde{\psi}_j^\phi \rangle = 1$ in the expression relating the number of electrons $N_e$ and the density of states as given below:

$$N_e = \sum_{j=1}^{N} \int_{-\infty}^{\infty} f(\epsilon, \epsilon_F)\delta(\epsilon - \epsilon_j) d\epsilon = \sum_{j=1}^{N} \int_{-\infty}^{\infty} \langle \tilde{\psi}_j^\phi | \tilde{\psi}_j^\phi \rangle f(\epsilon, \epsilon_F)\delta(\epsilon - \epsilon_j) d\epsilon \quad (2)$$

where $f(\epsilon, \epsilon_F)$ denotes the orbital occupancy function whose range lies in the interval $[0,1]$, while $\epsilon_F$ represents the Fermi-energy. This function is usually given by the Heaviside function taking the value 1 if $\epsilon < \epsilon_F$ and 0 otherwise. Using $|\tilde{\psi}_j^\phi\rangle = \sum_{j}^{N_{orb}} O_{ij}^{-1/2} |\psi_j^\phi\rangle$ and $|\psi_j^\phi\rangle = P^\phi |\psi_j\rangle$, eq 2 can be written in terms of Kohn-Sham wavefunctions $|\psi_j\rangle \in \mathbb{V}^M$ and the local atom-centred basis $|\phi_\mu\rangle \in \mathbb{V}^{N_{orb}}$ using the definition of the projection operator $P^\phi$ described previously. To this end, we have

$$N_e = \sum_{\nu,\nu'}^{N_{orb}} \sum_{\mu,\mu'}^{N_{orb}} \sum_{k,q}^{N} O_{jk}^{-1/2} O_{jq}^{-1/2} \int_{-\infty}^{\infty} f(\epsilon, \epsilon_F) \langle \psi_q | \phi_\mu \rangle S_{\mu \mu'}^{-1} \langle \phi_{\mu'} | \phi_{\nu'} \rangle S_{\nu \nu'}^{-1} \langle \phi_{\nu'} | \psi_k \rangle \delta(\epsilon - \epsilon_j) d\epsilon$$

$$= \sum_{I\alpha} \sum_{I\beta} \sum_{j} \sum_{k,q}^{N} O_{jk}^{-1/2} O_{jq}^{-1/2} \int_{-\infty}^{\infty} f(\epsilon, \epsilon_F) \langle \psi_q | \phi^{I\alpha} \rangle S_{I\alpha I\beta} \langle \phi^{I\beta} | \psi_k \rangle \delta(\epsilon - \epsilon_j) d\epsilon \quad (3)$$

In the above eq 3, $|\phi^\mu\rangle$ denotes the dual of the basis function $|\phi_\mu\rangle$, satisfying the property $\langle \phi^\mu | \phi_\nu \rangle = \delta_{\mu\nu}$ where, $|\phi^\mu\rangle$ is given by $|\phi^\mu\rangle = \sum_{\nu} S_{\nu \mu}^{-1} |\phi_\nu\rangle$. Furthermore, a multi-index $\mu = \{I\alpha\}$ is introduced in eq 3 to denote the localized atom-centred basis function $|\phi^\mu\rangle$ as $|\phi^{I\alpha}\rangle$ where $\alpha$ denotes the index of the atomic orbital centred at a nuclear position $R_I$.

The orbital overlap population deals with the distribution of the total number of electrons $N_e$ among the atoms in a given material system and can be motivated from eq 3. To this end, projected-orbital overlap population $pOOP_{ij}^\eta(\epsilon)$ associated with $\eta^{th}$ orbital of a source
atom $I$ and a $\gamma^{th}$ orbital of a target atom $J \neq I$ is extracted from eq 3 to be defined as

$$p_{OOP}^{\eta \gamma}_{IJ}(\epsilon) = \sum_{j} \bar{C}_{I\eta}^{j} \bar{C}_{I\gamma}^{j} S_{I\eta J\gamma} \delta(\epsilon - \epsilon_{j})$$  \hspace{1cm} (4)$$

where $\{\bar{C}_{I\alpha}^{j}\}$ for various multi-indices $I\alpha$ denote the coefficients of $|\tilde{\psi}_{j}^{\phi}\rangle$ represented in the localized atom-centered basis set $\{|\phi_{\mu}\rangle\}$ and is given by $\bar{C}_{I\alpha}^{j} = \sum_{q} O_{jq}^{-1/2} \langle \phi_{\mu} | \psi_{q} \rangle$, where $j = 1 \cdots N$. Finally, the total projected-orbital overlap population between the source atom $I$ and $J$ is defined as

$$p_{OOP}^{\mu \nu}_{IJ}(\epsilon) = \sum_{j} \sum_{\beta} \sum_{\alpha} \bar{C}_{I\alpha}^{j} \bar{C}_{I\beta}^{j} S_{I\alpha J\beta} \delta(\epsilon - \epsilon_{j})$$  \hspace{1cm} (5)$$

### Projected-orbital Hamilton Population ($pOHP$):

We recall that the expression for band energy ($E_{\text{band}}$) involving Kohn-Sham Hamiltonian $\mathcal{H} \in \mathbb{C}^{M \times M}$ and its $N$-smallest eigenfunctions $|\psi_{j}\rangle \in \mathcal{V}^{M}$ as $E_{\text{band}} = \sum_{j=1}^{N} \int_{-\infty}^{\infty} f(\epsilon, \epsilon_{F}) \langle \psi_{j} | \mathcal{H} | \psi_{j} \rangle \delta(\epsilon - \epsilon_{j}) d\epsilon$. Following Maintz et al.,\cite{Maintz2013} we now define the projected Hamiltonian operator $\mathcal{H}^{p}$ on the subspace $\mathcal{V}_{\text{orb}}^{N}$ in terms of the orthonormalized projected Kohn-Sham wavefunction $|\tilde{\psi}_{j}^{\phi}\rangle$ as $\mathcal{H}^{p} = \sum_{j=1}^{N} |\tilde{\psi}_{j}^{\phi}\rangle \epsilon_{j} \langle \tilde{\psi}_{j}^{\phi} |$ where $\epsilon_{j}$ denotes the Kohn-Sham eigenvalues (see eq 1). Using this definition of $\mathcal{H}^{p}$, we have $\langle \psi_{j} | \mathcal{H} | \psi_{j} \rangle = \langle \tilde{\psi}_{j}^{\phi} | \mathcal{H}^{p} | \tilde{\psi}_{j}^{\phi} \rangle$ and hence the band energy $E_{\text{band}}$ can be written as

$$E_{\text{band}} = \sum_{j=1}^{N} \int_{-\infty}^{\infty} f(\epsilon, \epsilon_{F}) \langle \tilde{\psi}_{j}^{\phi} | \mathcal{H}^{p} | \tilde{\psi}_{j}^{\phi} \rangle \delta(\epsilon - \epsilon_{j}) d\epsilon$$

$$= \sum_{j,k,q=1}^{N} O_{jk}^{-1/2} O_{j\epsilon}^{-1/2} \int_{-\infty}^{\infty} f(\epsilon, \epsilon_{F}) \langle \mathcal{P}^{\phi} \psi_{q} | \mathcal{H}^{p} | \mathcal{P}^{\phi} \psi_{k} \rangle \delta(\epsilon - \epsilon_{j}) d\epsilon$$  \hspace{1cm} (6)$$

Along the lines of Maintz et al.,\cite{Maintz2013} we consider the orthogonalized basis $\{|\hat{\phi}_{\nu}\rangle\}$ obtained -via- Löwden symmetric orthonormalization of $\{|\phi_{\mu}\rangle\}$ to compute $pOHP$ and the two basis are related by the expression $|\hat{\phi}_{\mu}\rangle = \sum_{\nu} S_{\mu\nu}^{-1/2} |\phi_{\nu}\rangle$. Subsequently, the projection operator $\mathcal{P}^{\phi}$
expressed in terms of \( \{ |\hat{\phi}_\nu \rangle \} \) i.e \( P_\phi = \sum_{i=1}^{N_{\text{orb}}} |\hat{\phi}_\mu \rangle \langle \hat{\phi}_\mu | \) can be used to recast eq 6 as:

\[
E_{\text{band}} = \sum_{\mu,\nu=1}^{N_{\text{orb}}} \sum_{k,q,j} O_{jk}^{-1/2} O_{jq}^{-1/2} \int_{-\infty}^{\infty} f(\epsilon, \epsilon_F) \langle \psi_q | \hat{\phi}_\mu \rangle \langle \hat{\phi}_\mu | H^P | \hat{\phi}_\nu \rangle \langle \hat{\phi}_\nu | \psi_k \rangle \delta(\epsilon - \epsilon_j)d\epsilon
\]

\[
= \sum_{I\alpha,J\beta} \sum_{k,q,j} O_{jk}^{-1/2} O_{jq}^{-1/2} \int_{-\infty}^{\infty} f(\epsilon, \epsilon_F) \langle \psi_q | \hat{\phi}_{I\alpha} \rangle H^P_{I\alpha,J\beta} \langle \hat{\phi}_{J\beta} | \psi_k \rangle \delta(\epsilon - \epsilon_j)d\epsilon
\]

(7)

where, the composite index notation \( \mu = \{ I\alpha \} \) \( \nu = \{ J\beta \} \) has been used for the basis functions \( \{|\hat{\phi}_\mu \rangle\} \) to denote \( \alpha^{th} \) \( \beta^{th} \) basis function centered at the atomic position \( \mathbf{R}_I \) [\( \mathbf{R}_J \)] and \( H^P_{I\alpha,J\beta} \) denotes the matrix element of \( H^P \) represented in the \( \{|\hat{\phi}_\mu \rangle\} \) basis. The orbital Hamilton overlap population analysis deals with the partitioning of the band energy \( E_{\text{band}} \) among the constituent atoms in a given material system, and projected-orbital Hamilton population \( p_{\text{OHP}} \) can be defined taking recourse to eq 7. To this end, \( p_{\text{OHP}}^{\alpha\beta}(\epsilon) \) associated with \( \alpha^{th} \) orbital of a source atom \( I \) and a \( \beta^{th} \) orbital of a target atom \( J \neq I \) is extracted from eq 7 to be defined as

\[
p_{\text{OHP}}^{\alpha\beta}(\epsilon) = \sum_j \tilde{C}_{I\alpha}^j \tilde{C}_{J\beta}^j H^P_{I\alpha,J\beta} \delta(\epsilon - \epsilon_j)
\]

(8)

where \( \{|\tilde{C}_{I\alpha}^j \rangle\} \) are the coefficients of \( |\tilde{\psi}_\phi^j \rangle \) expressed with respect to the basis \( \{|\hat{\phi}_\mu \rangle\} \) i.e \( \tilde{C}_{I\alpha}^j = \sum_k O_{jk}^{-1/2} \langle \hat{\phi}_\mu | \psi_k \rangle \), with \( j, k = 1 \cdots N \). Using \( H^P = \sum_{j=1}^{N} |\tilde{\psi}_j \rangle \langle \tilde{\psi}_j | \), the matrix element \( H^P_{I\alpha,J\beta} \) can be evaluated as \( H^P_{I\alpha,J\beta} = \sum_k \tilde{C}_{I\alpha}^k \epsilon_k \tilde{C}_{J\beta}^k \). Finally, the total projected-orbital Hamilton population \( p_{\text{OHP}}^{IJ}(\epsilon) \) between two atoms centered at \( \mathbf{R}_I \) and \( \mathbf{R}_J \) such that \( I \neq J \)

\[
p_{\text{OHP}}^{IJ}(\epsilon) = \sum_j \sum_\alpha \sum_\beta \tilde{C}_{I\alpha}^j H^P_{I\alpha,J\beta} \tilde{C}_{J\beta}^j \delta(\epsilon - \epsilon_j)
\]

(9)

### 2.2 Projected-Hamiltonian Population Analysis

We now discuss an alternate approach to conduct projected-population analysis by means of orthogonally projecting the Kohn-Sham discretized Hermitian Hamiltonian operator \( H \) onto
the atomic-orbital subspace $V_{N_{\text{orb}}}^N$ to compute the projected Hamiltonian $H^\phi \equiv P^\phi H P^\phi : V_{N_{\text{orb}}}^N \to V_{N_{\text{orb}}}^N$, and subsequently, diagonalizing this $H^\phi$ to compute the orthonormal eigenvectors ($|\tilde{\psi}_i^E\rangle$) in the subspace $V_{N_{\text{orb}}}^N$. This approach is in contrast with the projected-orbital population analysis discussed above, (see section 2.1) wherein the discretized Kohn-Sham wavefunction $|\psi_i\rangle \in V^M$ is orthogonally projected (L$_2$ projection) onto the subspace $V_{N_{\text{orb}}}^N$. However, the current approach is based on the orthogonal projection of $H$ onto $V_{N_{\text{orb}}}^N$ and, the eigenvectors $|\tilde{\psi}_i^E\rangle \in V_{N_{\text{orb}}}^N$ of this projected Hamiltonian $H^\phi$ are used to compute both overlap and Hamilton populations. This approach is similar in spirit to the iterative orthogonal projection techniques employed in the solution of large-scale matrix eigenvalue problems of the form $A |x\rangle = \lambda |x\rangle$, wherein, one seeks an approximate eigenvector, eigenvalue pair ($|\tilde{x}\rangle, \tilde{\lambda}$) of $A$ in a carefully constructed lower-dimensional subspace such that the residual vector $|r\rangle = A|x\rangle - \tilde{\lambda}|\tilde{x}\rangle$ is orthogonal to this subspace. This orthogonality condition, also known as the Galerkin condition, is equivalent to diagonalizing the lower dimensional matrix (obtained by orthogonally projecting $A$ into the subspace), which approximates the eigenvalues and eigenvectors of the original matrix $A$.

**Projected-Hamiltonian Hamilton Population (pHHP):** We note that $|\tilde{\psi}_i^E\rangle$, the eigenvectors of $H^\phi$ lie in the atomic-orbital subspace $V_{N_{\text{orb}}}^N$ and, hence we express $|\tilde{\psi}_i^E\rangle$ as a linear combination of the orthonormalised atomic orbital basis $\{|\tilde{\phi}_\mu\rangle\}$ i.e. $|\tilde{\psi}_i^E\rangle = \sum_{\mu} \tilde{E}_{i\mu} |\tilde{\phi}_\mu\rangle$. Subsequently, following similar arguments as in eq (6) and eq (7), we can define the projected-Hamiltonian Hamilton population $\text{pHHP}^{\alpha\beta}_{IJ}(\epsilon)$ associated with $\alpha^{th}$ orbital of a source atom $I$ and a $\beta^{th}$ orbital of a target atom $J \neq I$ as

\[
\text{pHHP}^{\alpha\beta}_{IJ}(\epsilon) = \sum_j \tilde{E}_{I\alpha}^j \tilde{E}_{J\beta}^j H_{I\alpha,J\beta}^\phi \delta(\epsilon - \epsilon_j)
\]

(10)

where $\{|\tilde{E}_j\rangle\}$ are the coefficients of $|\tilde{\psi}_j^E\rangle$ expressed with respect to the basis $\{|\tilde{\phi}_\mu\rangle\}$ i.e $\tilde{E}_{i\mu} = \langle \tilde{\phi}_\mu | \tilde{\psi}_i^E \rangle$, with $j = 1 \cdots N$. Using the definition of projected Hamiltonian $H^\phi$, the matrix element $H_{I\alpha,J\beta}^\phi$ can be evaluated as $H_{I\alpha,J\beta}^\phi = \langle \tilde{\phi}_{I\alpha} | H^\phi | \tilde{\phi}_{J\beta}\rangle = \langle \tilde{\phi}_{I\alpha} | H | \tilde{\phi}_{J\beta}\rangle$. Finally, the
total projected-Hamiltonian Hamilton population $\text{pHHP}_{IJ}(\epsilon)$ between two atoms centered at $\mathbf{R}_I$ and $\mathbf{R}_J$ such that $I \neq J$

$$\text{pHHP}_{IJ}(\epsilon) = \sum_j \sum_\alpha \sum_\beta \bar{E}_{I\alpha}^j H_{I\alpha,J\beta}^\phi \bar{E}_{J\beta}^j \delta(\epsilon - \epsilon_j) \quad (11)$$

**Projected-Hamiltonian Overlap Population (pHOP):** In the context of overlap population, we express $|\tilde{\psi}_j^E\rangle$ as a linear combination of atomic-orbital basis $|\phi_\mu\rangle \in \mathcal{V}_N^{\text{or}}$ i.e., $|\tilde{\psi}_j^E\rangle = \sum_\mu \bar{E}_{j\mu} |\phi_\mu\rangle$ for $j = 1 \cdots N$ and finally, the total projected-Hamiltonian overlap population between the source atom $I$ and $J$ is defined as

$$\text{pHOP}_{IJ}(\epsilon) = \sum_j \sum_\beta \sum_\alpha \bar{E}_{I\alpha}^j \bar{E}_{J\beta}^j S_{I\alpha,J\beta} \delta(\epsilon - \epsilon_j) \quad (12)$$

### 3 Numerical implementation

In this section, we discuss various aspects of the numerical implementation procedure to conduct projected population analysis using the Kohn-Sham DFT wavefunctions obtained via the solution of the finite-element discretized DFT eigenvalue problem. To this end, the computations of overlap and Hamilton populations discussed in the previous section are implemented in the DFT-FE code,\cite{19,20} a massively parallel real-space code for large-scale density functional theory calculations based on adaptive finite-element discretization. Furthermore, the numerical implementation of population analysis in DFT-FE also takes advantage of parallel computing architectures via MPI (Message Passing Interface), enabling chemical bonding analysis of large-scale systems in a unified computational framework to conduct both DFT calculations and the chemical bond analysis. The projected population analysis methodology implemented within the finite-element setting is henceforth referred to as pFOOP (projected Finite-element Orbital Overlap Population) and pFOHP (projected Finite-element Orbital Hamilton Population) when dealing with projected-orbital population analysis, while it is referred to as pFHOP (projected Finite-element Hamiltonian Overlap Population).
Population) and pFHHP (projected Finite-element Hamiltonian Hamilton Population) when dealing with projected-Hamiltonian population analysis.

**Spectral finite-element discretization:** The finite-element (FE) discretization of the Kohn-Sham DFT problem employed in the DFT-FE code represents various electronic fields in the finite-element basis, a piece-wise polynomial and strictly local basis set. These FE basis are $C^0$ continuous Lagrange polynomial basis interpolated over Gauss-Lobatto-Legendre nodal points and we refer to our prior work for more details on the spectral finite-element discretization of Kohn-Sham DFT equations. In particular, the representation of various fields employed in computing projected overlap and Hamilton populations — the Kohn-Sham wavefunctions ($\langle \mathbf{x} | \psi_i \rangle = \psi_i(\mathbf{x})$) and the localized atom-centred basis functions ($\langle \mathbf{x} | \phi_\mu \rangle = \phi_\mu(\mathbf{x})$) — in the FE basis is given by

$$\psi_i(\mathbf{x}) = \sum_{j=1}^{M} N^h_j(\mathbf{x}) \psi^j_i, \quad \phi_\mu(\mathbf{x}) = \sum_{j=1}^{M} N^h_j(\mathbf{x}) \phi^j_\mu, \quad (13)$$

where $N^h_j : 1 \leq j \leq M$ denote the $M$ finite-element (FE) basis functions spanning the $M$-dimensional space $\mathbb{V}^M$. These are strictly local Lagrange polynomials of order $p$ generated using the nodes of the FE triangulation, $\mathcal{T}^h$, with the characteristic mesh size denoted by $h$. Further in eq (13), $\psi^j_i$ and $\phi^j_\mu$ denote the coefficients in the expansion of the $i^{th}$ discretized Kohn-Sham wavefunction ($\psi_i(\mathbf{x})$) and the $\mu^{th}$ atom-centered localized basis function ($\phi_\mu(\mathbf{x})$). These coefficients form the nodal values of the discretized fields represented using the FE triangulation $\mathcal{T}^h$ since the FE basis functions $N^h_j(\mathbf{x})$ satisfy the Kronecker-delta property i.e. $N^h_j(\mathbf{x}_k) = \delta_{jk}$ where $\mathbf{x}_k$ denotes the $k^{th}$ nodal point of $\mathcal{T}^h$. The nodal values $\psi^j_i$ are computed by solving the FE discretized Kohn-Sham DFT eigenvalue problem given in eq (1). Computationally efficient and scalable methodologies to solve this problem on massively parallel many-core architectures have been discussed in Motamarri et.al and on hybrid CPU-GPU architectures in Das et.al. Furthermore, in the current work, the nodal values $\phi^j_\mu$ are computed from the atom-centred orbital data, which is usually available as analytical

15
expressions or in the form of numerical data.

**FE basis overlap matrix (M):** Finite-element (FE) basis functions are non-orthogonal, and the associated overlap matrix $M$ is computed by evaluating the following integral over the simulation domain volume denoted by $\Omega$

$$M_{ij} = \int_{\Omega} N_i(x)N_j(x)\,dx = \sum_{\Omega_e} \int_{-1}^{1} \int_{-1}^{1} \int_{-1}^{1} N_i(\xi, \eta, \zeta) N_j(\xi, \eta, \zeta) \,d\xi \,d\eta \,d\zeta$$

(14)

$$= \sum_{\Omega_e} \sum_{p,q,r=0}^{n_q} w_{p,q,r} N_i(\xi_p, \eta_q, \zeta_r) N_j(\xi_p, \eta_q, \zeta_r) \,det(J_e)$$

$(\xi, \eta, \zeta)$ above denote the barycentric coordinates, $J_e$ denotes the Jacobian matrix corresponding to a finite-element $\Omega_e$, and $n_q$ denotes the number of quadrature points in each dimension employed to evaluate the integral in eq (14). Gauss-Lobatto-Legendre (GLL) quadrature rules are employed to evaluate the integrals in eq (14). These rules have quadrature points coincident with the FE nodal points in the spectral finite-element discretization employed in this work rendering the matrix $M$ diagonal since the above equation is non-zero only if $i = j$. This diagonal FE basis overlap matrix has been employed in the DFT–FE code to transform the generalized Kohn-Sham eigenvalue problem into standard eigenvalue problem allowing the use of efficient Chebyshev-filtered subspace iteration procedures to compute the Kohn-Sham eigenspace. This diagonal matrix $M$ will also play a crucial role in the computationally efficient evaluation of projected overlap and Hamilton populations, as discussed subsequently.

**Atom-centered orbital overlap matrix (S):** The localized non-orthogonal atom-centered basis orbitals $\{|\phi_\mu\rangle\}$ are represented in a finite-element basis (see eq (13)) and hence the associated overlap matrix element $S_{\alpha\beta} = \langle \phi_\alpha | \phi_\beta \rangle$ is computed as follows:

$$S_{\mu\nu} = \int \phi_\mu(x)\phi_\nu(x)\,dx = \sum_{p,q=1}^{M} \phi^p_\mu \left( \int_{\Omega} N_p(x)N_q(x)\,dx \right) \phi^q_\nu$$

(15)
The above equation is recast in a matrix form using the FE basis overlap matrix \( M \) as shown below:

\[
S = \Phi^T M \Phi = \Phi^T M^{1/2} M^{1/2} \Phi = \Phi^T \Phi \quad \text{where} \quad \Phi = M^{1/2} \Phi. \tag{16}
\]

In the above, \( \Phi \) denotes a \( M \times N_{\text{orb}} \) matrix whose column vectors are the components of \( |\phi_\alpha\rangle \) in FE basis. Recalling that \( M \) is diagonal, computation of \( M^{1/2} \) becomes trivial and thereby, the evaluation of \( \Phi \) involves a point-wise scaling operation of the columns in \( \Phi \) with the diagonal entries of \( M^{1/2} \). Finally, the computation of \( S \) is reduced to matrix-matrix multiplication involving \( \Phi \) as described in eq (16). In a parallel implementation of the computation of \( S \) using MPI on multi-node CPU architectures, the matrix \( \Phi \) is distributed equally among the available MPI tasks into equipartitioned matrix \( \Phi_P \) of size \( M_{\text{loc}} \times N_{\text{orb}} \) with \( M_{\text{loc}} \approx M/P \), and \( P \) denoting the number of MPI tasks. The domain decomposition of the underlying FE mesh across these MPI tasks achieves this equal distribution. We now compute the \( N_{\text{orb}} \times N_{\text{orb}} \) matrix \( S_P = \Phi_P^T \Phi_P \) associated with each core locally using BLAS level 3 optimized math kernel libraries. We finally add these local matrices \( S_P \) by employing MPI collectives to compute the matrix-matrix product \( \Phi^T \Phi \) in a distributed setting. To that effect, the computational complexity of computing \( S \) using the above algorithm when running on \( P \) MPI tasks is \( \mathcal{O}(M_{\text{loc}} N_{\text{orb}}^2) \sim 2M_{\text{loc}} N_{\text{orb}}^2 \). As will be discussed subsequently, the projected population analysis requires the computation of \( S^{-\beta} \) where \( \beta = 1 \) or \(-1/2\) or \(1/2\). Therefore, in the current work, we diagonalize \( S \) matrix and evaluate \( S^{-\beta} = QD^{-\beta}Q^T \) where \( Q \) is an eigenvector matrix with columns as eigenvectors of \( S \) and \( D \) is a diagonal matrix comprising of diagonal entries as the eigenvalues of \( S \). We note that an efficient implementation of divide-and-conquer algorithm available in LAPACK library is employed for diagonalization and the computational complexity of this algorithm for diagonalization of \( S \) is \( \mathcal{O}(N_{\text{orb}}^3) \sim 4N_{\text{orb}}^3 \). Furthermore, the computational complexity of the matrix-matrix multiplication for evaluating \( S^{-\beta} \) (after diagonalization) is \( \mathcal{O}(N_{\text{orb}}^3) \sim 2N_{\text{orb}}^3 \).
3.1 Projected finite-element orbital population analysis (pFOP)

We now discuss the efficient implementation strategies implemented in DFT-FE to compute the expressions required for pFOP.

Projected Kohn-Sham wavefunction overlap matrix (O): We first discuss the computation of coefficients of projected Kohn-Sham wavefunctions in the basis of \( \{|\phi_\mu\rangle\} \) and then derive an expression for computing the overlap matrix associated with projected Kohn-Sham wavefunctions using this coefficient matrix (C) within the finite-element setting. To this end, we note that the projected Kohn-Sham wavefunction \( |\psi_\phi^i\rangle \in \mathbb{V}_{N_{orb}}^N \) can be expressed as a linear combination of the atomic-orbital basis \( \{|\phi_\mu\rangle\} \) i.e. \( |\psi_\phi^i\rangle = \sum_\alpha C_{i\alpha} |\phi_\alpha\rangle \) where \( C_{i\alpha} = \sum_\nu S_{\alpha\nu}^{-1} \langle \phi_\nu | \psi_i \rangle \).

Introducing the finite-element discretization for \( |\psi_i\rangle \) and \( |\phi_\alpha\rangle \) (see eq (13)), we have

\[
C_{i\alpha} = \sum_\nu \sum_p \sum_q S_{\alpha\nu}^{-1} \phi^p_\nu \left( \int_\Omega N_p(x)N_q(x) \, dx \right) \psi^q_i
\]  

The above equation is recast in a matrix form in terms of the matrix \( \mathbf{M} \) as

\[
\mathbf{C} = \mathbf{S}^{-1} \mathbf{\Phi}^T \mathbf{M} \mathbf{\Psi} = \mathbf{S}^{-1} \mathbf{\Phi}^T \mathbf{M}^{1/2} \mathbf{M}^{1/2} \mathbf{\Psi} = \mathbf{S}^{-1} \mathbf{\Phi}^T \mathbf{\Psi} \ \text{where} \ \mathbf{\Psi} = \mathbf{M}^{1/2} \mathbf{\Psi}
\]  (18)

In the above, \( \mathbf{\Psi} \) denotes a \( M \times N \) matrix whose column vectors are the components of \( |\psi_i\rangle \) in FE basis while \( \mathbf{\Phi} \) is a \( M \times N_{orb} \) matrix comprising of the atomic-orbital data as defined above. Further, the matrix \( \mathbf{C} \) is of dimension \( N_{orb} \times N \) and is evaluated in an efficient way on a parallel computing architecture by first evaluating the matrix-matrix product \( \mathbf{\Phi}^T \mathbf{\Psi} \) locally on each MPI task and summing the contributions across various MPI tasks. When running on \( P \) MPI tasks, the computational complexity of this step is \( \mathcal{O}(M_{loc} N_{orb} N) \sim 2M_{loc} N_{orb} N \). Subsequently, the resulting \( N_{orb} \times N \) matrix is pre-multiplied with the inverse of the atomic-orbital overlap matrix \( \mathbf{S} \) computed above to evaluate the \( \mathbf{C} \) matrix finally. We now consider the evaluation of the overlap matrix \( \mathbf{O} \) associated with the projected Kohn-
Sham wavefunctions $\psi_i^\phi$ by first recalling that $O_{ij} = \langle \psi_i^\phi | \psi_j^\phi \rangle = \langle \psi_i | \Phi^\phi | \psi_j \rangle = \langle \psi_i | \Phi^\phi | \psi_j \rangle$.

Using the definition of $\Phi^\phi = \sum_{\mu, \nu=1}^{N_{\text{orb}}} | \phi_\mu \rangle S_{\mu \nu}^{-1} \langle \phi_\nu |$, one can rewrite the matrix elements of $O$ as $O_{ij} = \sum_{\mu, \nu} \langle \psi_i | \phi_\mu \rangle S_{\mu \nu}^{-1} \langle \phi_\nu | \psi_j \rangle$. To this end, in a finite-element discretized setting, the matrix $O$ can be computed as

$$O_{ij} = \sum_{\mu, \nu} \sum_{pq} \sum_{rs} \left[ \psi_i^p \left( \int_\Omega N_p(x) N_q(x) \, dx \right) \phi_\mu^q \right] S_{\mu \nu}^{-1} \left[ \phi_\nu^r \left( \int_\Omega N_r(x) N_s(x) \, dx \right) \psi_j^s \right]$$

$$\Rightarrow O = \Psi^T \Phi S^{-1} \Phi^T \Psi = C^T S C \quad (19)$$

The matrix expression for the projected Kohn-Sham wavefunction overlap matrix $O$ in the above equation uses the expression for the coefficient matrix $C$ in eq (18). After the computation of $S$ and $C$ are computed using the expressions in eqns. (16) and (18), the matrix $O$ can be computed by performing matrix-matrix multiplications and the computational complexity for evaluating $O$ is $O(N_{\text{orb}}^2 N) + O(N_{\text{orb}} N^2) \sim 2 N_{\text{orb}}^2 N + 2 N^2 N_{\text{orb}}$

**Spilling factor:** The total spilling factor or charge spilling factor, as introduced by Sanchez-Portal et al.\textsuperscript{18} describes the ability of the atom-centred localized basis spanning $\bigvee_{\phi}^{N_{\text{orb}}}$ to represent the FE discretized Kohn-Sham eigenfunctions $| \psi_j \rangle$, the self-consistent solution of the FE discretized Kohn-Sham eigenvalue problem. This is given by the average of the $L_2$ projection error over the Kohn-Sham occupied eigenstates ($N_{\text{occ}}$) in the case of charge spilling factor and the total number of Kohn-Sham eigenstates considered for the projection ($N$) in the case of total spilling factor. This can be shown as equal to the deviation of the norm of the projected Kohn-Sham eigenfunctions from 1. To this end, following Stefan Maintz et. al.,\textsuperscript{38} we compute the absolute spilling factor $S$ and absolute charge spilling factor $S_c$ in terms of the diagonal entries of the matrix $O$ (see eq (19)) as

$$S = \frac{1}{N} \sum_{i=1}^{N} |1 - \langle \psi_i^\phi | \psi_i^\phi \rangle| = \frac{1}{N} \sum_{i=1}^{N} |1 - O_{ii}|, \quad S_c = \frac{1}{N_{\text{occ}}} \sum_{i=1}^{N_{\text{occ}}} |1 - O_{ii}| \quad (20)$$

19
Projected finite-element orbital density error (pFODE): We compute the $L_2$ norm error between the ground-state electron density ($\rho(x)$) computed from DFT-FE using the FE discretized occupied Kohn-Sham eigenfunctions $\{|\psi_i\rangle\}$ and the electron-density ($\rho^o(x)$) computed from the Löwden orthonormalized projected Kohn-Sham wavefunctions $\{|\tilde{\psi}_i^\phi\rangle\} \in \mathbb{V}^N_{\phi}$. To this end, pFODE is evaluated as

$$\text{pFODE} = \frac{||\rho(x) - \rho^o(x)||_2}{||\rho(x)||_2}$$

where $\rho(x) = \sum_{i=1}^{N_{occ}} \langle x|\psi_i \rangle \langle \psi_i |x \rangle$, $\rho^o(x) = \sum_{i=1}^{N_{occ}} \langle x|\tilde{\psi}_i^\phi \rangle \langle \tilde{\psi}_i^\phi |x \rangle$ (21)

Projected finite-element orbital overlap population (pFOOP): We evaluate pFOOP using the expression deduced in eq (5). To this end, we first evaluate the coefficient matrix ($\tilde{C}$) corresponding to the coefficients of $\{|\tilde{\psi}_j^\phi\rangle\}$ in the basis of $\{|\tilde{\phi}_\mu\rangle\}$. Recall from Section II that $\tilde{C}_{\mu}^j = \sum_{q} \sum_{\nu} O_{jq}^{-1/2} S_{\mu\nu}^{-1} \langle \tilde{\phi}_\nu |\psi_q \rangle = \sum_{q} O_{jq}^{-1/2} C_{\mu}^q$. Hence in the matrix form, the coefficient matrix $\tilde{C}$ can be written as $\tilde{C} = CO^{-1/2}$ and can be evaluated with a computational complexity of $O(N_{orb}N^2) \sim 2N_{orb}N^2$. Furthermore, we note that $O^{-1/2}$ is evaluated by diagonalizing $O$ and associated the computational complexity is $O(N^3) \sim 4N^3$. The matrix $\tilde{C}$ is of size $N_{orb} \times N$, and the rows of this matrix are stored in the order of atoms and their corresponding atom-centred orbitals for a given atom in succession. To elaborate, $\mu^{th}$ row of $\tilde{C}$ corresponds to $I^{th}$ atom and an atom-centred index $\alpha$ associated with this atom $I$, while the $j^{th}$ column of this matrix corresponds to the index of projected Kohn-Sham wavefunction ($j = 1 \cdots N$). Finally, the pFOOP$_{\eta \gamma}^{\eta \gamma}$ associated with the overlap of $\eta^{th}$ orbital of a source atom $I$ and a $\gamma^{th}$ orbital of a target atom $J \neq I$ is evaluated by extracting the appropriate entries of the matrices $\tilde{C}$, $S$ and the expression is given by

$$\text{pFOOP}_{\eta \gamma}^{\eta \gamma}(\epsilon) = \sum_{j} \tilde{C}_{I\eta j} \tilde{C}_{J\gamma j} S_{\eta \gamma j} \delta(\epsilon - \epsilon_j)$$ (22)

Projected finite-element orbital Hamilton population (pFOHP): We recall the relation between Löwden orthonormalized atom-centered basis $\{|\tilde{\phi}_\nu\rangle\}$ and the non-orthogonal
atom-centered basis \( \{|\phi_\nu\}\) to be \( |\hat{\phi}_\mu\rangle = \sum_\nu S_{\mu\nu}^{-1/2} |\phi_\nu\rangle \). Recasting this relation in matrix form we get \( \hat{\Phi} = \Phi S^{-1/2} \) where \( \hat{\Phi} \) denotes \( M \times N_{\text{orb}} \) matrix whose column vectors are components of \( \{|\hat{\phi}_\nu\}\) in FE basis. Now, we note that \( |\tilde{\psi}_i\rangle \in \mathbb{V}^{N_{\text{orb}}} \) can be expressed as a linear combination of the basis \( \{|\hat{\phi}_\nu\}\) i.e \( |\tilde{\psi}_i\rangle = \sum_\nu \hat{C}^i_\nu |\hat{\phi}_\nu\rangle \) where \( \hat{C}^i_\nu = \sum_q O_{jq}^{-1/2} \langle \hat{\phi}_\nu | \psi_q \rangle \). Introducing finite-element discretization for \( |\hat{\phi}_\nu\rangle \) and \( |\psi_q\rangle \), we have

\[
\hat{C}^i_\nu = \sum_q \sum_r \sum_s O_{qs}^{-1/2} \hat{C}^r_\nu \left( \int_\Omega N_r(x) N_s(x) \, dx \right) \psi_q^s
\]  

(23)

Recasting the above equation in matrix form, we have

\[
\hat{C} = \hat{\Phi}^T M \Psi O^{-1/2} = \hat{\Phi} \Psi O^{-1/2} = S^{-1/2} \hat{\Phi}^T \Psi O^{-1/2} = S^{-1/2} \hat{\Phi}^T \Psi O^{-1/2} = S^{-1/2} \hat{C}
\]  

(24)

where \( \hat{\Phi} = M^{1/2} \hat{\Phi} \), \( \Psi \) and \( \hat{C} \) are defined previously. \( \hat{C} \) in the above equation is of size \( N_{\text{orb}} \times \mathbb{N} \) and is computed -via- matrix-matrix multiplication involving \( \hat{C} \) and \( S^{-1/2} \) with a computational complexity of \( \mathcal{O}(N_{\text{orb}}^2 N) \sim 2N_{\text{orb}}^2 \). Similar to other coefficient matrices described previously, the rows of this matrix are stored in the order of atoms and their corresponding atom-centred orbitals for a given atom in succession. In all our subsequent discussions and calculations in this work, we choose \( N = N_{\text{orb}} \) and, we compute the \( N_{\text{orb}} \times N_{\text{orb}} \) projected Hamiltonian matrix \( \hat{H}^p \) using the coefficient matrix \( \hat{C} \) as \( \hat{H}^p = \hat{C} \hat{D} \hat{C}^T \) where the matrix \( \hat{D} \) is diagonal and comprises of the Kohn-Sham eigenvalues \( \epsilon_i \) obtained from Kohn-Sham DFT problem solved in the finite-element basis. The computation of \( \hat{H}^p \) involves matrix-matrix multiplication after scaling \( \hat{C}^T \) with diagonal matrix \( \hat{D} \) and has the computational complexity of \( \mathcal{O}(N_{\text{orb}}^2 N) \sim 2N_{\text{orb}}^2 N \). Finally, the \( p^\text{FOH} \langle IJ \rangle \) associated with the partitioning of energy between the \( \eta^\text{th} \) orbital of a source atom \( I \) and a \( \gamma^\text{th} \) orbital of a target atom \( J \neq I \) is evaluated by extracting the appropriate entries of the matrices \( \hat{C} \), \( \hat{H}^p \) and

21
the expression is given by

\[ \text{pFOHP}_{IJ}(\epsilon) = \sum_j \tilde{C}_{I\eta} \tilde{C}_{J\gamma} H^p_{I\eta J\gamma} \delta(\epsilon - \epsilon_j) \] (25)

**Total computational complexity estimate of pFOP:** The current implementation of the projected finite-element orbital population analysis (pFOP) assumes \( N = N_{\text{orb}} \) and thereby the total computational complexity is estimated to be \( \sim 4M_{\text{loc}}N^2 + 26N^3 \). Recall that \( M_{\text{loc}} \) is the ratio of total number of finite-element degrees of freedom \( (M) \) and the number of MPI tasks \( (P) \), \( M_{\text{loc}} \) can be reduced by increasing the value of \( P \). Consequently, the second term in the computational complexity becomes dominant when the number of MPI tasks \( P \) is greater than \( 4M/26N \).

---

**Figure 1:** Overview of projected finite-element orbital population analysis (pFOP) implementation strategy. The implementation uses the SCF converged Kohn-Sham eigenfunctions and eigenvalues from DFT-FE. The atom-centered orbitals are represented on the finite element basis. Following which, the eigenfunctions are projected onto this atomic orbital basis. The outputs are the overlap matrix \( (S) \), projected Hamiltonian \( (H^p) \) and the coefficients \( \tilde{C}, \tilde{C} \) to compute overlap population \( \text{pFOOP} \) and Hamilton \( \text{pFOHP} \) populations respectively.
3.2 Projected finite-element Hamiltonian population analysis (pFHP)

We now elucidate the strategies employed to compute the expressions required for pFHP. Let $H$ denote the matrix corresponding to the self-consistently converged finite-element discretized Kohn-Sham Hamiltonian operator $\mathcal{H}$ introduced earlier. We first begin with the computation of the projection of $H$ into the space $\mathcal{V}_{\phi}^{N_{\text{orb}}}$ spanned by Löwdin orthonormalized atomic-orbital basis $\{|\hat{\phi}_{\nu}\rangle\}$. From section 2.2, we recall that the matrix elements of the projected Hamiltonian expressed in $\{|\hat{\phi}_{\nu}\rangle\}$ basis is given by $H_{ij}^{\phi} = \langle \hat{\phi}_i | \mathcal{H} | \hat{\phi}_j \rangle$, which can, in turn, be recast in the matrix form as $H^{\phi} = S^{-1/2} \Phi^{T} H \Phi S^{-1/2}$. The dominant computational complexity of computing $H^{\phi}$ when running on $P$ MPI tasks is $O(M_{\text{loc}} N_{\text{orb}}^2) + O(N_{\text{orb}}^3) \sim 2M_{\text{loc}} N_{\text{orb}}^2 + 4N_{\text{orb}}^3$. Upon diagonalization of $H^{\phi}$ we have $H^{\phi} = \hat{E} \hat{E}^{T}$ where $\hat{E}$ denotes the $N_{\text{orb}} \times N_{\text{orb}}$ eigenvector matrix and the computational complexity of this step is $O(N_{\text{orb}}^3) \sim 4N_{\text{orb}}^3$. As discussed in Sec 2.2, we note that the $j^{th}$ column of $\hat{E}$ represents the coefficients of $|\tilde{\psi}_{E}^j\rangle$ with respect to $\{|\hat{\phi}_{\nu}\rangle\}$ basis. Further, $\mu^{th}$ row of $\hat{E}$ corresponds to $I^{th}$ atom and an atom-centered index $\alpha$ associated with this atom $I$. Finally, we describe the procedure to compute the projected finite-element Hamiltonian population analysis (pFHP) as discussed below.

Projected finite-element Hamiltonian density error (pFHDE): To understand the loss of information due to the projection of the finite-element discretized Hamiltonian onto $\mathcal{V}_{\phi}^{N_{\text{orb}}}$, we introduce the projected finite element density error (pFHDE). Here, we compute the L2 norm error between the ground-state electron density ($\rho(\mathbf{x})$) computed from FE discretized occupied Kohn-Sham eigenfunctions $\{|\psi_i\rangle\}$ solved using DFT-FE and the electron-density ($\rho^H(\mathbf{x})$) computed from the occupied eigenfunctions $\{|\tilde{\psi}_{E}^j\rangle\} \in \mathcal{V}_{\phi}^{N_{\text{orb}}}$ associated with the projected Hamiltonian $H^{\phi}$. To this end, pFHDE is evaluated as

$$\text{pFHDE} = \frac{||\rho(\mathbf{x}) - \rho^H(\mathbf{x})||_2}{||\rho(\mathbf{x})||_2}$$

where

$$\rho(\mathbf{x}) = \sum_{i=1}^{N_{\text{occ}}} \langle \mathbf{x} | \psi_i \rangle \langle \psi_i | \mathbf{x} \rangle, \quad \rho^H(\mathbf{x}) = \sum_{i=1}^{N_{\text{occ}}} \langle \mathbf{x} | \tilde{\psi}_{E}^j \rangle \langle \tilde{\psi}_{E}^j | \mathbf{x} \rangle$$

(26)
Projected finite-element Hamiltonian Hamilton population (pFHHP): In this alternate framework of projected population analysis, the pFHHP\(_{IJ}\) associated with the partitioning of energy between \(\eta^{th}\) orbital of a source atom \(I\) and \(\gamma^{th}\) orbital of a target atom \(J \neq I\) is evaluated by extracting appropriate entries of the matrices \(\hat{\mathbf{E}}, \mathbf{H}^\phi\), and the expression is given by

\[
pFHHP_{IJ}(\epsilon) = \sum_j \hat{E}_{I\eta j} \hat{E}_{J\gamma j} H^\phi_{I\eta J\gamma} \delta(\epsilon - \epsilon_j)
\]  

(27)

Projected finite-element Hamiltonian overlap population (pFHOP): Similar in spirit to pFOOP described above, the overlap population in this approach is computed using the atom-centred orbitals \(|\phi_\nu\rangle\) basis sets. To this end, we first compute the \(N_{\text{orb}}\) linear combination coefficients of the expansion of \(|\tilde{\psi}_E^I\rangle\) in terms of the basis \(|\phi_\nu\rangle\). Denoting these coefficients by vector \(\mathbf{E}\), this vector can be easily obtained from \(\hat{\mathbf{E}}\) by taking recourse to basis transformation operation i.e., \(\mathbf{E} = \mathbf{S}^{-1/2}\hat{\mathbf{E}}\) and is of \(O(N_{\text{orb}}^3) \sim 2N_{\text{orb}}^3\) complexity. Finally, the pFHOP\(_{IJ}\) associated with the overlap of \(\eta^{th}\) orbital of a source atom \(I\) and a \(\gamma^{th}\) orbital of a target atom \(J \neq I\) is evaluated by extracting the appropriate entries of the matrices \(\hat{\mathbf{E}}, \mathbf{S}\) and the expression is given by

\[
pFHOP_{IJ}(\epsilon) = \sum_j \hat{E}_{I\eta j} \hat{E}_{J\gamma j} S_{I\eta J\gamma} \delta(\epsilon - \epsilon_j)
\]  

(28)

Total computational complexity estimate of pFHP: The current implementation of the projected finite-element orbital population analysis (pFHP) assumes \(N = N_{\text{orb}}\) and thereby the total computational complexity is estimated to be \(\sim 4M_{\text{loc}}N^2 + 16N^3\). Since that \(M_{\text{loc}} = M/P\), the second term in the computational complexity becomes dominant when the number of MPI tasks \(P\) is greater than \(4M/16N\) and starts to become computationally efficient than the pFOOP approach described in the previous subsection.
4 Results and Discussion

In this section, we present the numerical results that demonstrate the robustness of the proposed approaches on representative benchmark material systems. To this end, we first begin with benchmarking the accuracy and performance of the projected finite-element orbital-overlap population and orbital-Hamilton population approach (\(pFOOP\) and \(pFOHP\)) with \textsc{LOBSTER}, a widely used chemical bonding analysis code, on various examples involving isolated molecules, nanoparticles and a periodic system. Subsequently, we compare these results with the projected finite-element Hamiltonian population approach (\(pFHOP\) and \(pFHHP\)) elucidated in this work on these benchmark examples. Furthermore, we demonstrate the advantage of having a scalable finite-element implementation to extract chemical bonding by conducting the projected population analysis on large-scale material systems (\(\sim 60 - 1000\) atoms) comprising Si nanoparticles, a candidate material for hydrogen storage.
In the case of comparative studies with **LOBSTER**, we employ **Quantum Espresso** (QE) with PAW pseudopotentials to perform DFT calculations for all the benchmark systems, and the resulting ground-state DFT wavefunctions and eigenvalues are used as an input for conducting the population analysis using **LOBSTER**. The calculations using **QE** are performed using Γ-point calculations employing the GGA exchange-correlation of the PBE form. While in DFT-FE calculations, we employ optimized norm-conserving Vanderbilt (ONCV) pseudopotentials from pseudoDojo database to conduct both pseudopotential DFT calculations and population analysis in the same framework. We employ non-periodic boundary conditions for isolated systems and periodic boundary conditions for crystalline systems in DFT-FE. For all the benchmark calculations reported here, the cutoff energies in **QE** and mesh sizes in DFT-FE are chosen such that a discretization error of $O(10^{-5})$ Ha/atom in ground state energy and a force discretization error of $O(10^{-4})$ Ha/Bohr in ionic forces is achieved. The simulations and computational times reported in this work are performed on the CPU nodes of the supercomputer PARAM Pravega.

### 4.1 Projected finite-element orbital population analysis (pFOP)

We now benchmark our formulation and implementation of pFOOP and pFOHP as discussed in eq 22 and eq 25 with that of pCOOP and pCOHP obtained by using **LOBSTER**. To this end, we project the self-consistently converged Kohn-Sham (KS) wavefunctions obtained from DFT-FE into a subspace spanned by (i) Bunge orbitals and (ii) pseudo-atomic (PA) orbitals constructed from the ONCV pseudopotentials. On the other hand, the converged ground-state DFT wavefunctions obtained using PAW formalism in **QE** are used as input to **LOBSTER**, which are projected onto a subspace spanned by localized atom-centered basis functions known as *pbevaspfit2015*. In our benchmark studies, we consider (i) isolated systems comprising of CO, spin-polarized O$_2$, H$_2$O molecules, Si-H nanoparticle with 65 atoms (Si$_{29}$H$_{36}$), and (ii) $2 \times 2 \times 2$ periodic carbon diamond supercell. To simulate isolated

*PARAM Pravega is one of India’s fastest supercomputers stationed at Indian Institute of Science comprising of 584 Intel Xeon Cascade-Lake based CPU nodes (28,032 Cores)
systems in \texttt{DFT-FE}, we consider the simulation domain large enough to allow the wavefunctions to decay to zero on the boundary by employing non-periodic boundary conditions. In contrast, in \texttt{QE}, which always employs periodic boundary conditions, we consider a simulation domain with sufficient vacuum to minimize the image-image interactions. In the benchmark problem involving a bulk material system, we consider a $2 \times 2 \times 2$ diamond supercell employing periodic boundary conditions using a $\Gamma$-point to sample the Brillouin zone in both \texttt{DFT-FE} and \texttt{QE}. We subsequently describe below the comparative study of the projection spill factors, population analysis energy diagrams and the computational costs between the proposed implementation in \texttt{DFT-FE} and \texttt{LOBSTER}.

Population analysis using the proposed \texttt{pFOP} approach and \texttt{LOBSTER} is compared here, for the case of Si$_{29}$H$_{36}$ and periodic $2 \times 2 \times 2$ supercell of carbon. Table \ref{table1} shows the comparison of absolute spilling factor $S$ and the absolute charge spilling factor $S_c$ for these systems, and we note that these factors obtained in our current work are almost similar to that of \texttt{LOBSTER}. Furthermore, we note that the values of $S$ and $S_c$ computed from our \texttt{pFOP} approach are lesser when pseudo-atomic (PA) orbitals constructed from ONCV pseudopotential are used as the localized atomic-orbital basis in comparison to Bunge orbitals. This can be attributed to the fact that the subspace spanned by PA orbitals is a better representation of the FE discretized ground-state Kohn-Sham eigenfunctions obtained from \texttt{DFT-FE} using ONCV pseudopotential calculations. We use these PA orbitals for our subsequent comparisons with \texttt{LOBSTER} in this section, and comparisons with Bunge orbitals are discussed in the Supplementary Information (SI). Next, we illustrate the comparisons of population energy diagrams in Figures \ref{fig3} and \ref{fig4}, both for the case of Si$_{29}$H$_{36}$ and periodic $2 \times 2 \times 2$ carbon diamond supercell. In the case of carbon diamond supercell, a pair of nearest neighbour carbon atoms are picked as the source and target atom, and the corresponding contributions of $s - s$ and $s - p$ orbitals are plotted in Figure \ref{fig3} (see inset in the figure) both for overlap population and Hamilton population. A comparison of total populations is also illustrated in this figure. These results indicate that the energy diagrams obtained with \texttt{LOBSTER} match
very well with our current approach. The location of the bonding and anti-bonding peaks are identical to that obtained from LOBSTER, with a slight difference in the amplitude of the peaks that can be attributed to the use of different pseudopotentials in LOBSTER and DFT–FE. Similar agreements are observed in the case of Si$_{29}$H$_{36}$ nanoparticle in which the Si atom and the nearest H atom are picked as the source and target atom, and the corresponding contributions of H$_{1s}$–Si$_{3p}$ and H$_{1s}$–Si$_{3s}$ orbitals are plotted in Figure 4 (see inset in the figure) both for overlap population and Hamilton population. Comparisons of spill factors and the population energy diagrams for benchmark systems involving molecules – CO, spin-polarized O$_2$, H$_2$O are discussed in SI, and we observe a close agreement of the pFOOP and pFOHP with that obtained using LOBSTER. Finally, the computational CPU times in terms of node-sec (wall time taken on 1 compute node), measuring the computation of overlap population and Hamilton population, are tabulated in Table 2 for both the proposed pFOP approach and LOBSTER. The benchmark systems involving $2 \times 2 \times 2$ carbon diamond supercell, Si-H nanoparticles involving Si$_{29}$H$_{36}$ and Si$_{145}$H$_{150}$ with 65 and 295 atoms are chosen for comparison. The computational times indicate significant advantage for the proposed implementation compared to LOBSTER. For instance, computational gains up to $2 \times$ are observed for the case of carbon diamond $2 \times 2 \times 2$ periodic supercell, while in the case of Si$_{29}$H$_{36}$ nanoparticle, a speed up of $14 \times$ is observed. This higher speedup in the case of Si nanoparticle using the proposed pFOP approach can be attributed to the use of non-periodic boundary conditions in our implementation while LOBSTER is restrictive in terms of the boundary conditions employed and allows for periodic boundary conditions even for Si nanoparticle, an isolated system. Furthermore, it was computationally prohibitive to conduct population analysis using LOBSTER for the Si$_{145}$H$_{150}$ nanoparticle. We further demonstrate the advantage of our parallel implementation of pFOP by measuring the wall times of the population analysis conducted on a large Si$_{580}$H$_{510}$ nanoparticle containing 1090 atoms. Table 3 shows a reduction in wall time of about $1.7 \times$ with an increase in the number of CPU cores from 280 to 1120.
Table 1: Comparison of absolute charge spill factor($S_c$) and absolute spill factor($S$) obtained using projections carried out in DFT-FE and LOBSTER. DFT-FE Bunge indicates the projection of finite-element discretized Kohn-Sham eigenfunctions to Bunge atomic orbitals and DFT-FE PA basis indicates projection onto pseudo-atomic orbitals. Projection in LOBSTER uses pbeVaspfit2015 as auxiliary atom-centered basis.

| System                        | LOBSTER | DFT-FE Bunge | DFT-FE PA |
|-------------------------------|---------|--------------|-----------|
| Carbon diamond 2 × 2 × 2 periodic supercell | 0.010   | 0.092        | 0.011     |
|                               | 0.101   | 0.003        | 0.101     |
| Si$_{29}$H$_{36}$ nanoparticle | 0.012   | 0.308        | 0.314     |
|                               | 0.039   | 0.017        | 0.271     |

Figure 3: Comparison of projected finite-element orbital population analysis (pFOP) using pseudo-atomic (PA) orbitals with LOBSTER. Top row shows the comparison of overlap population and the bottom row shows the Hamilton population. Energy scale is shifted such that Fermi level($\epsilon_F$) is zero. **Case study:** Carbon diamond 2 × 2 × 2 supercell with periodic boundary conditions using a Γ point for Brillouin zone sampling.
Figure 4: Comparison of projected finite-element orbital population analysis (pFOP) using pseudo-atomic (PA) orbitals with LOBSTER. Top row shows the comparison of overlap population and the bottom row shows the Hamilton population. Energy scale is shifted such that Fermi level ($\epsilon_F$) is zero. Case study: Single-fold Si$_{29}$H$_{36}$ nanoparticle

Table 2: Comparison of computation CPU time measured in node-secs, between projected finite-element orbital population analysis (pFOP) implementation in DFT-FE and LOBSTER. The timing includes the reading and construction of the atomic orbital basis ($\Phi$) and the calculation of the appropriate matrices for conducting the population analysis ($C, \tilde{C}, H^p$)

| Material system               | pFOP(sec) | LOBSTER(sec) |
|-------------------------------|-----------|--------------|
| Carbon diamond 2 x 2 x 2 supercell | 8.26      | 17.5         |
| Si$_{29}$H$_{36}$ nanoparticle | 1.72      | 24.2         |
| Si$_{145}$H$_{150}$ nanoparticle | 21.46     | -            |

4.2 Projected finite-element Hamiltonian population analysis (pFHP)

We assess here the performance and accuracy of the proposed pFHP procedure implemented within the DFT-FE framework. To this end, we project the self-consistently converged Kohn-Sham finite-element (FE) discretized Hamiltonian obtained from DFT-FE into a subspace
Table 3: Wall times of \texttt{pFOP} for various number of CPU cores. \textbf{Case study:} Si$_{580}$H$_{510}$ nanoparticle (1090 atoms with 2830 valence electrons). Total degrees of freedom (DoFs) per atom is around 8400.

| Number of cores | Wall-time (seconds) |
|-----------------|---------------------|
| 280             | 33.85               |
| 560             | 25.04               |
| 1120            | 20.48               |

spanned by pseudo-atomic (PA) orbitals constructed from the ONCV pseudopotentials and conduct the population analysis as discussed in Section 3.2. We discuss here a comparative study of the population analysis conducted using this approach and \texttt{pFOP} reported in the subsection 4.1. To begin with, we plot the population energy diagrams corresponding to \texttt{pFHOP} and \texttt{pFHHP} derived in eqns. 28 and 27 and compare with that of \texttt{pFOOP} and \texttt{pFOHP} on the same benchmark systems comprising of isolated systems and a periodic system as discussed in the previous subsection. Henceforth, we will refer to the approach of projected finite-element orbital population (\texttt{pFOOP} and \texttt{pFOHP}) as Method-1 and the approach of projected finite-element Hamiltonian population (\texttt{pFHOP} and \texttt{pFHHP}) as Method-2. Figure 5 illustrates the comparison in the case of periodic 2 × 2 × 2 carbon diamond supercell for two nearest carbon atoms picked as source and target atom (see inset in Figure 5). Figure 6 demonstrates the comparisons in the case of Si$_{29}$H$_{36}$ nanoparticle, an isolated system in which Si atom and the nearest H atom are picked as source and target atom (see inset in Figure 6). As the results demonstrate, we see a very good match of the corresponding contributions of $s$ – $s$ and $s$ – $p$ orbitals for both overlap and Hamilton population conducted using both \texttt{pFOP} (Method-1) and \texttt{pFHP} (Method-2). Comparisons between both methods for benchmark systems involving molecules – CO, spin polarized O$_2$, H$_2$O are discussed in SI, and we observe a very close agreement.

We now compare the density error metrics (\texttt{pFODE} and \texttt{pFHDE}), a measure of loss of information during projections, as introduced in Section 3.1 and Section 3.2. Recall from eq 21 and eq 26, these metrics measure the $L_2$ norm error between the self-consistently converged ground-state electron-density computed from \texttt{DFT-FE} and the electron-density
Figure 5: Comparison of overlap and Hamilton populations between $pFOP$ (Method-1) with $pFHP$ (Method-2). The top row shows the comparison of overlap population and the bottom row the Hamilton population. Energy scale is shifted such that Fermi level ($\epsilon_F$) is at zero. **Case study:** $2 \times 2 \times 2$ carbon diamond supercell with periodic boundary conditions at $\Gamma$-point for Brillouin zone sampling.

computed in both the proposed population analysis approaches, from the wavefunctions constructed in the subspace $V_{\phi}^{N_{orb}}$. As shown in the Table 4, we see a close agreement between $pFHDE$ and $pFODE$ for a variety of benchmark material systems which include isolated systems (CO, H$_2$O, spin-polarized O$_2$, Si$_{29}$H$_{36}$, Si$_{56}$H$_{66}$, and Si$_{145}$H$_{150}$) and a periodic system (carbon $2 \times 2 \times 2$ supercell).

The comparative study discussed so far clearly demonstrates the excellent match of numerical results obtained between the approaches $pFOP$ (Method-1) and $pFHP$ (Method-2). As remarked before, $pFOP$ (Method-1) is similar in spirit to the projected population analysis approach implemented in LOBSTER. However, $pFHP$ (Method-2) proposed in this work is different in spirit than the projected orbital population approach and relies on the projection of finite-element discretized Kohn-Sham Hamiltonian to conduct population analysis. We now demonstrate the computational advantage of conducting population analysis using
Figure 6: Comparison of overlap and Hamilton populations between pFOP (Method-1) with pFHP (Method-2). The top row shows the comparison of overlap population and the bottom row the Hamilton population. Energy scale is shifted such that Fermi level ($\epsilon_F$) is at zero. **Case study:** single-fold Si$_{29}$H$_{36}$ nanoparticle with non-periodic boundary conditions.

**pFHP** (Method-2) for large-scale systems. Figures 7a and 7b show the computational wall times measuring the projected population analysis with increasing number of CPU cores comparing the two methods pFOP (Method-1) and pFHP (Method-2). For this study, we consider two nanoparticles Si$_{580}$H$_{510}$ and Si$_{1160}$H$_{990}$ comprising of 1090 and 2150 atoms respectively. The results indicate a speed-up of $1.3 \times - 1.4 \times$ for Method-2 with the increase in CPU cores beyond a certain number and these speed-ups are consistent with the computational complexity estimates derived in Section 3, i.e. $O(N^3)$ cost becoming dominant beyond a certain number of cores with a lower prefactor for Method-2.

### 4.3 Bonding insights in large-scale Silicon nanoparticles

In this section, we demonstrate the advantage of proposed computational framework for conducting population analysis in order to extract chemical bonding in large-scale material...
Table 4: Comparison of density errors ($p_{FODE}$ and $p_{FHDE}$) computed from projected finite-element orbital population ($p_{FOP}$) and projected finite-element Hamiltonian population approach ($p_{FHP}$) respectively

| Material System              | $p_{FODE}$ | $p_{FHDE}$ |
|------------------------------|------------|------------|
| Carbon diamond $2 \times 2 \times 2$ supercell | 0.061      | 0.062      |
| CO                           | 0.118      | 0.119      |
| $H_2O$                       | 0.136      | 0.121      |
| $O_2$ (Spin-up density)      | 0.097      | 0.098      |
| $O_2$ (Spin-down density)    | 0.062      | 0.064      |
| Si$_{29}$H$_{36}$ nanoparticle | 0.179      | 0.175      |
| Si$_{58}$H$_{66}$ nanoparticle | 0.172      | 0.168      |
| Si$_{145}$H$_{150}$ nanoparticle | 0.165      | 0.162      |

systems. We motivate the need for large-scale bonding analysis by considering the case-study of chemisorption of hydrogen in silicon nanoparticles, a candidate material for hydrogen storage, 23 where the storage (release) of hydrogen is a result of the formation (breaking) of Si-H bond. As discussed in Williamson et.al, 23 the release of hydrogen in these Si nanoparticles occurs due to the dimerization of dihedral SiH$_2$ co-facial pairs in a Si$_{29}$H$_{36}$ unit to reduce to Si$_{29}$H$_{24}$ unit by formation of an additional Si-Si bond. Furthermore, the authors also argued from a thermodynamic viewpoint that alloying these Si nanoparticles with C reduces the critical temperature of hydrogen absorption/desorption to an operating temperature compatible with fuel cell applications. In contrast, we attempt to provide a chemical bonding viewpoint by conducting population analysis to extract bonding information between a given source-target atoms in these Si nanoparticles. We compute a quantity known as integrated projected finite-element orbital Hamilton population $I_{FOHP} = \int_{\epsilon \leq \epsilon_F} -p_{FOHP}(\epsilon)d\epsilon$. A higher value of $I_{FOHP}$ correlates with a stronger covalent bonding interaction between the source-target atoms and this quantity is similar to integrated $p_{COHP}(ICOHP)$ computed in LOBSTER. To this end, we argue the ease of dimerization of dihedral SiH$_2$ co-facial pairs in the Si nanoparticle by computing $I_{FOHP}$ between Si-Si atoms in the adjacent dihedral SiH$_2$ pairs and the associated Si-H atoms in a given SiH$_2$ dihedral unit. We examine $I_{FOHP}$ values.
(a) **Case study:** $\text{Si}_{580}\text{H}_{510}$ (1090 atoms, 2830 wave-functions)

(b) **Case study:** $\text{Si}_{1160}\text{H}_{990}$ (2150 atoms, 5630 wave-functions)

Figure 7: Wall-time comparison of Method-1 with Method-2 on PARAM Pravega. Case study considered are 20-fold and 40-fold Si nanoparticles. Total DoFs ($M$) are $9130679$ and $17254979$ for $\text{Si}_{580}\text{H}_{510}$ (20-fold) and $\text{Si}_{1160}\text{H}_{990}$ (40-fold) respectively.

as a function of increasing Si nanoparticle size ranging from 65 atoms to 1090 atoms and further, with and without carbon alloying. In particular, we consider 1-fold, 2-fold, 5-fold and 20-fold structure of dihedral $\text{Si}_{29}$ and $\text{Si}_{24}\text{C}_{5}$ units. Towards this, we build the 2-fold and further the 5-fold structure by connecting the $\text{Si}_{29}\text{H}_{36}$ ($\text{Si}_{24}\text{C}_{5}\text{H}_{36}$) units by their (111) facets, as discussed in Williamson et.al.\textsuperscript{23}

The atomic configurations corresponding to various sizes of $\text{Si}_{29}$ and $\text{Si}_{24}\text{C}_{5}$ nanoparticles are obtained by performing geometry optimization in DFT-FE code till the maximum atomic force in each direction for all the atoms reaches a tolerance of approximately $5 \times 10^{-4}$ Ha/Bohr. Figure 8 shows the relaxed atomic configuration of the various nanoparticles considered in this work. To determine the strength of the Si-Si bonding interaction between the nearest SiH$_2$ dihedral structures, we compute IFOHP by conducting pFOHP analysis. Figure 9 shows the total Hamilton population energy diagrams for various sizes of Si nanoparticles, capturing this Si-Si interaction for one of the SiH$_2$ dihedral pairs (see atoms marked in Figure 8). We observe that the bonding and anti-bonding peaks are of equal magnitude for different sizes of Si nanoparticles without carbon, while for the nanoparticles alloyed with carbon, the bonding peaks are observed to be higher. Table 5 reports the mean of IFOHP
values corresponding to the Si-Si interaction between nearest SiH\textsubscript{2} dihedral pairs. Mean of the IFOHP values corresponding to the weakest of the Si-H interaction in each of the dihedral unit of these pairs have also been tabulated in this table (Si-H1 and Si-H2). All dihedral pairs, which are nearest neighbours to the carbon core in the alloyed nanoparticles, are considered for computing the mean IFOHP. Equivalent Si-Si and Si-H atom pair was also picked from the nanoparticles without alloying for comparison. Consistent with the total Hamilton population energy plots of pFOHP, IFOHP values reported in Table 5 shows a higher value (≈ 4× higher) for nanoparticles alloyed with carbon in comparison to the ones’ without alloying. This can be attributed to the carbon core in the alloyed nanoparticles drawing the Si atoms in the nearest SiH\textsubscript{2} dihedral units towards each other, thereby leading to a stronger Si-Si interaction. It is interesting to note the increase in IFOHP values corresponding to these Si-Si interactions with the curvature of the alloyed nanoparticles, indicating the strengthening of Si-Si interaction. This can possibly explain the ease of dimerization of dihedral SiH\textsubscript{2} co-facial pairs as the size increases from 1-fold to 5-fold structures. Similarly, due to the reduced separation between the dihedral SiH\textsubscript{2} groups, the corresponding IFOHP values suggest a weakening of Si-H bond as seen in Table 5 for alloyed nanoparticles, a favourable condition for the release of H\textsubscript{2}.

It is important to maximize the hydrogen storage capacity when designing hydrogen storage devices, and hence, large nanoparticle are desirable. As suggested by Williamson et.al.,\textsuperscript{23} one such structure is obtained by stacking four 5-fold Si nanoparticles resulting in a 20-fold nanoparticle containing 1090 atoms. IFOHP values for this large nanoparticle with and without C alloying have been computed by conducting population analysis on the relaxed geometries. The stacking should not have much effect on the curvature of this large nanoparticle in comparison to 5-fold. Consequently, we observe a mean IFOHP value of Si-Si interaction between the nearest SiH\textsubscript{2} dihedral pairs to be around 0.510 and 0.005 for the 20-fold nanoparticle with and without C alloying respectively, which are close to the values observed in the case of 5-fold nanoparticle.
Figure 8: Calculated atomic structures of (a,b) 1-Fold, (c,d) 2-fold, (e,f) 5-fold and (g,h) 20-fold nanoparticle structures. The silicon atoms highlighted in red is one of the pairs considered for population analysis. The structure is relaxed in DFT resulting in a maximum force component of $5 \times 10^{-4}$ Ha/Bohr.

Figure 9: Comparison of $-\text{pFOHP}$ between nearest non-bonded neighbours of Si-Si in 1-Fold, 2-Fold and 5-Fold nanoparticles.
Table 5: Comparison of the average integrated finite element orbital Hamilton population (IFOHP) of various pairs of dihedral SiH$_2$ in 1-Fold, 2-Fold and 5-Fold nanoparticle of Si$_{29}$ and Si$_{24}$C$_5$ units. Si-Si interaction is computed between nearest SiH$_2$ dihedral pairs. Si-H1 and Si-H2 denotes the weakest of the Si-H interaction in each of the dihedral unit of these pairs.

| System     | Si-Si interaction | Si-H1 interaction | Si-H2 interaction |
|------------|-------------------|-------------------|-------------------|
| Si$_{29}$H$_{36}$ | 0.009             | 4.505             | 4.504             |
| Si$_{24}$C$_5$H$_{36}$ | 0.402             | 4.358             | 4.362             |
| Si$_{58}$H$_{66}$ | 0.014             | 4.484             | 4.486             |
| Si$_{48}$C$_{$10}H$_{66}$ | 0.406             | 4.319             | 4.333             |
| Si$_{145}$H$_{150}$ | 0.005             | 4.498             | 4.498             |
| Si$_{120}$C$_{25}$H$_{150}$ | 0.476             | 4.297             | 4.361             |

5 Conclusions

In the present work, we formulate and implement two methods for conducting projected population analysis to extracting chemical bonding information from finite-element (FE) based density functional theory (DFT) calculations. The first method relies on the orthogonal projection of FE discretized Kohn-Sham DFT eigenfunctions onto a subspace spanned by localized atom-centered basis functions while the second method relies on the orthogonal projection of FE discretized Kohn-Sham Hamiltonian into this subspace. The proposed methods are well-suited for projected population analysis from large-scale Kohn-Sham DFT calculations and takes advantage of multi-node CPU architectures. These methods are implemented within the DFT-FE code$_{19,20}$ and takes advantage of DFT-FE’s capability to conduct fast, scalable and systematically convergent large-scale DFT calculations opening up the possibility of conducting bonding analysis on complex material systems not accessible before. Furthermore, a single computational framework is employed to conduct both ground-state DFT calculations and the proposed population analysis without any restriction on the kind of boundary conditions used. This single framework enjoys an unique advantage of conducting on-the-fly bonding analysis in ab-initio molecular dynamics simulations, leading to bonding interaction insights as a function of time.

Firstly, we presented a mathematical formulation in an abstract setting to derive the
expressions for overlap and Hamilton population within the framework of projected-orbital and projected-Hamiltonian population analysis. The formalism described is generic enough to accommodate Kohn-Sham eigenfunctions obtained from the solution of a DFT eigenvalue problem discretized in any convenient choice of systematically convergent basis set and project on to a subspace spanned by any localized atom-centered basis functions for extracting chemical information. Subsequently, a scalable numerical implementation procedure within the finite-element discretized setting is described to efficiently compute various matrices involved in projected finite-element orbital population (pFOP) and projected finite-element Hamiltonian (pFHP) analysis for computing overlap and Hamilton population. Next, we assessed the accuracy of the proposed method on representative material systems comprising of isolated molecules, nanoparticles and a periodic system with a large supercell. In all the cases, the proposed method shows excellent agreement with the population energy diagrams obtained using LOBSTER, a widely used orbital-population analysis code. The computational advantage of pFOP over LOBSTER is also clearly illustrated on few of these benchmark examples. Furthermore, the advantage of conducting population analysis using pFHP (Method-2) over pFOP (Method-1) in terms of wall time on multi-node CPU cores is demonstrated on large-scale systems involving 1000-2000 atoms. Finally, we leverage the proposed population analysis approach in a case-study to extract bonding insights in increasing sizes of Si nanoparticles up to 1000 atoms, a candidate material for hydrogen storage.

Acknowledgement

The authors gratefully acknowledge the seed grant from Indian Institute of Science and SERB Startup Research Grant from the Department of Science and Technology India (Grant Number: SRG/2020/002194) for the purchase of a GPU cluster, which also provided computational resources for this work. The research used the resources of PARAM Pravega at Indian Institute of Science, supported by National Super-computing Mission (NSM) R&D for
exa-scale grant (DST/NSM/R&D_Exascale/2021/14.02). P.M also thanks Prathu Tiwari at Nvidia, Bangalore for helping us run few of the geometry optimizations involving large-scale material systems on GPU clusters.
Supporting Information Available

1 Additional benchmarking studies

1.1 Projected finite-element orbital population (pFOP)

In this section, we describe benchmarking studies involving the computation of pFOOP and pFOHP on few representative material systems not discussed in the main manuscript. pCOOP and pCOHP obtained by using LOBSTER are used for benchmarking. To this end, we show here the absolute spill factor (S), charge spill factor (S_c) and the population energy diagrams of CO molecule, H_2O molecule and spin-polarized O_2 molecule. We show the results of projection of the self-consistently converged Kohn-Sham(KS) eigenfunctions obtained from DFT-FE onto a space spanned by (i) Bunge orbitals and (ii) pseudo-atomic(PA) orbitals constructed from ONCV pseudopotentials. Finally, we show the population energy diagrams of 2 × 2 × 2 carbon diamond and Si_{29}H_{36} projection of KS wavefunctions into Bunge orbitals.

Table 1: Comparison of absolute charge spill factor (S_c) and absolute spill factor (S) obtained using projections carried out in DFT-FE and LOBSTER. DFT-FE Bunge indicates the projection of finite-element discretized Kohn-Sham eigenfunctions to Bunge atomic orbitals and DFT-FE PA basis indicates projection onto pseudo-atomic orbitals. Projection in LOBSTER uses pbeVaspfit2015 as auxiliary atom-centered basis.

| System | LOBSTER | DFT-FE Bunge | DFT-FE PA |
|--------|---------|--------------|-----------|
|        | S_c     | S_c          | S_c       | S_c     |
| CO     | 0.018   | 0.143        | 0.034     | 0.154   |
| H_2O   | 0.016   | 0.263        | 0.033     | 0.277   |
| O_2↑   | 0.015   | 0.136        | 0.036     | 0.152   |
| O_2↓   | 0.012   | 0.135        | 0.029     | 0.150   |

CO molecule: The C-O bond length in CO molecule is 1.14Å. We note that from Table 1, spill factors obtained by projecting onto PA basis are similar to that obtained using LOBSTER. From Figures 1 and 2 we see that the population energy diagrams obtained using the proposed approach is in excellent agreement to that obtained from LOBSTER.
Figure 1: Comparison of projected finite-element orbital population analysis (pFOP) with LOBSTER using pseudo-atomic orbitals. Top row shows the comparison of overlap population and the bottom row shows the Hamilton population. Energy scale is shifted such that Fermi level ($\epsilon_F$) is zero. **Case study:** CO molecule

Figure 2: Comparison of projected finite-element orbital population analysis (pFOP) with LOBSTER using Bunge atomic orbitals. Top row shows the comparison of overlap population and the bottom row shows the Hamilton population. Energy scale is shifted such that Fermi level ($\epsilon_F$) is zero. **Case study:** CO molecule
**H$_2$O molecule:** The relaxed structure of H$_2$O has an O-H bond length of 0.971 Å. We note that from Table 1, spill factors obtained by projecting onto PA basis are similar to that obtained using LOBSTER. From Figures 3 and 4 we see that the population energy diagrams obtained using the proposed approach is in excellent agreement to that obtained from LOBSTER.

Figure 3: Comparison of projected finite-element orbital population analysis (pFOP) with LOBSTER using pseudo-atomic orbitals. Top row shows the comparison of overlap population and the bottom row shows the Hamilton population. Energy scale is shifted such that Fermi level ($\epsilon_F$) is zero. **Case study:** H$_2$O molecule

Figure 4: Comparison of projected finite-element orbital population analysis (pFOP) with LOBSTER using Bunge atomic orbitals. Top row shows the comparison of overlap population and the bottom row shows the Hamilton population. Energy scale is shifted such that Fermi level ($\epsilon_F$) is zero. **Case study:** H$_2$O molecule
Spin polarized calculation on O\textsubscript{2} molecule: The relaxed structure of O\textsubscript{2} with spin has and O-O bond length of 1.227\textdegree\text{A}. We note that from Table 1, spill factors obtained by projecting onto PA basis are lesser to that obtained using LOBSTER. From Figure 5 and Figure 7 we observe that for the up spin channel the location and the number of peaks of pFOOP and pFOHP are identical to that of LOBSTER. We observe a similar trend in Figure 6 and Figure 8 corresponding to the down spin channel.

Figure 5: Comparison of projected finite-element orbital population analysis (pFOP) with LOBSTER using pseudo-atomic orbitals. Top row shows the comparison of overlap population and the bottom row shows the Hamilton population. Energy scale is shifted such that Fermi level($\epsilon_F$) is zero. Case study: spin-polarized O\textsubscript{2} ↑

Figure 6: Comparison of projected finite-element orbital population analysis (pFOP) with LOBSTER using pseudo-atomic orbitals. Top row shows the comparison of overlap population and the bottom row shows the Hamilton population. Energy scale is shifted such that Fermi level($\epsilon_F$) is zero. Case study: spin-polarized O\textsubscript{2} ↓
Figure 7: Comparison of projected finite-element orbital population analysis (pFOP) with LOBSTER using Bunge atomic orbitals. Top row shows the comparison of overlap population and the bottom row shows the Hamilton population. Energy scale is shifted such that Fermi level($\epsilon_F$) is zero. **Case study:** spin-polarized $O_2 \uparrow$

Figure 8: Comparison of projected finite-element orbital population analysis (pFOP) with LOBSTER using Bunge atomic orbitals. Top row shows the comparison of overlap population and the bottom row shows the Hamilton population. Energy scale is shifted such that Fermi level($\epsilon_F$) is zero. **Case study:** spin-polarized $O_2 \downarrow$

**Periodic $2 \times 2 \times 2$ supercell of carbon:** We discuss here the comparison of pFOOP obtained from projecting onto Bunge atom-centered basis with LOBSTER. The structure is relaxed and the nearest C-C bond length obtained is around 1.55Å. We pick the nearest neighbour carbon atoms as source and target atoms, and the corresponding $s-s$ and $s-p$ orbital interactions are plotted in Figure 9.
**Si$_{29}$H$_{36}$ nanocluster:** We discuss here the comparison of pFOOP obtained from projecting onto Bunge basis with LOBSTER. The structure is relaxed and one of the dihedral Si-H interaction is plotted in Figure 10. The figures shows the s-s and s-p interaction between H-Si. The corresponding Si-H bond length is 1.49 Å and results are in very good agreement to LOBSTER.

---

**Figure 9:** Comparison of projected finite-element orbital population analysis (pFOOP) with LOBSTER using Bunge atomic orbitals. Top row shows the comparison of overlap population and the bottom row shows the Hamilton population. Energy scale is shifted such that Fermi level($\epsilon_F$) is zero. **Case study:** 2 × 2 × 2 carbon diamond supercell with periodic boundary conditions using a Γ point for Brillouin zone sampling.

**Figure 10:** Comparison of projected finite-element orbital population analysis (pFOOP) with LOBSTER using Bunge atomic orbitals. Top row shows the comparison of overlap population and the bottom row shows the Hamilton population. Energy scale is shifted such that Fermi level($\epsilon_F$) is zero. **Case study:** Single-fold Si$_{29}$H$_{36}$ nanoparticle
1.2 Projected finite-element Hamiltonian population (pFHP)

In this section, we illustrate the comparisons between pFOP (Method-1) and pFHP (Method-2) on few representative material systems not discussed in the main manuscript. To this end, we compare the population energy diagrams of CO molecule in Figure 11, H$_2$O molecule in Figure 12 and spin-polarized O$_2$ molecule in Figures 14 and 13. We observe a very good agreement between the two methods.

**Case study:** CO molecule

**Figure 11:** Comparison of overlap and Hamilton populations between pFOP (Method-1) with pFHP (Method-2). The top row shows the comparison of overlap population and the bottom row the Hamilton population. Energy scale is shifted such that Fermi level ($\epsilon_F$) is at zero.

**Case study:** H$_2$O molecule

**Figure 12:** Comparison of overlap and Hamilton populations between pFOP (Method-1) with pFHP (Method-2). The top row shows the comparison of overlap population and the bottom row the Hamilton population. Energy scale is shifted such that Fermi level ($\epsilon_F$) is at zero.
Figure 13: Comparison of overlap and Hamilton populations between pFOP (Method-1) with pFHP (Method-2). The top row shows the comparison of overlap population and the bottom row the Hamilton population. Energy scale is shifted such that Fermi level ($\epsilon_F$) is at zero. **Case study:** Spin-polarized $O_2 \uparrow$

Figure 14: Comparison of overlap and Hamilton populations between pFOP (Method-1) with pFHP (Method-2). The top row shows the comparison of overlap population and the bottom row the Hamilton population. Energy scale is shifted such that Fermi level ($\epsilon_F$) is at zero. **Case study:** Spin-polarized $O_2 \downarrow$
References

(1) Hughbanks, T.; Hoffmann, R. Chains of trans-edge-sharing molybdenum octahedra: metal-metal bonding in extended systems. *Journal of the American Chemical Society* 1983, *105*, 3528–3537.

(2) Glassey, W. V.; Hoffmann, R. A comparative study of Hamilton and overlap population methods for the analysis of chemical bonding. *The Journal of Chemical Physics* 2000, *113*, 1698–1704.

(3) Hoffmann, R. An extended Hückel theory. I. hydrocarbons. *The Journal of Chemical Physics* 1963, *39*, 1397–1412.

(4) Eck, B.; Dronskowski, R.; Takahashi, M.; Kikkawa, S. Theoretical calculations on the structures, electronic and magnetic properties of binary 3d transition metal nitrides. *Journal of Materials Chemistry* 1999, *9*, 1527–1537.

(5) Tachibana, M.; Yoshizawa, K.; Ogawa, A.; Fujimoto, H.; Hoffmann, R. Sulfur-gold orbital interactions which determine the structure of alkanethiolate/Au (111) self-assembled monolayer systems. *The Journal of Physical Chemistry B* 2002, *106*, 12727–12736.

(6) Steinberg, S.; Dronskowski, R. The crystal orbital Hamilton population (COHP) method as a tool to visualize and analyze chemical bonding in intermetallic compounds. *Crystals* 2018, *8*, 225.

(7) Dronskowski, R. *Computational Chemistry of Solid State Materials: A Guide for Materials Scientists, Chemists, Physicists and others*; WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2005.

(8) Dronskowski, R.; Bloechl, P. E. Crystal orbital Hamilton populations (COHP): energy-
resolved visualization of chemical bonding in solids based on density-functional calculations. *The Journal of Physical Chemistry* **1993**, *97*, 8617–8624.

(9) Tank, R.; Jepsen, O.; Burkhardt, A.; Andersen, O. TB-LMTO-ASA Program. *Max-Planck-Institut für Festkörperforschung: Stuttgart, Germany* **1994**.

(10) Andersenhi, O. K.; Arcangeli, C.; Tank, R. W.; Saha-Dasgupta, T.; Krier, G.; Jepsen, O.; Dasgupta, I. Third-Generation TB-LMTO. MRS Online Proceedings Library. 1997; pp 3–34.

(11) Andersen, O. K.; Saha-Dasgupta, T. Muffin-tin orbitals of arbitrary order. *Phys. Rev. B* **2000**, *62*, R16219–R16222.

(12) Deringer, V. L.; Tchougréeff, A. L.; Dronskowski, R. Crystal orbital Hamilton population (COHP) analysis as projected from plane-wave basis sets. *The journal of physical chemistry A* **2011**, *115*, 5461–5466.

(13) Maintz, S.; Deringer, V. L.; Tchougréeff, A. L.; Dronskowski, R. Analytic projection from plane-wave and PAW wavefunctions and application to chemical-bonding analysis in solids. *Journal of computational chemistry* **2013**, *34*, 2557–2567.

(14) Nelson, R.; Ertural, C.; George, J.; Deringer, V. L.; Hautier, G.; Dronskowski, R. LOBSTER: Local orbital projections, atomic charges, and chemical-bonding analysis from projector-augmented-wave-based density-functional theory. *Journal of Computational Chemistry* **2020**, *41*, 1931–1940.

(15) Hafner, J. Ab-initio simulations of materials using VASP: Density-functional theory and beyond. *Journal of computational chemistry* **2008**, *29*, 2044–2078.

(16) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I., et al. QUANTUM ESPRESSO: a modu-
lar and open-source software project for quantum simulations of materials. *Journal of physics: Condensed matter* 2009, 21, 395502.

(17) Kundu, S.; Bhattacharjee, S.; Lee, S.-C.; Jain, M. Population analysis with Wannier orbitals. *The Journal of Chemical Physics* 2021, 154, 104111.

(18) Sanchez-Portal, D.; Artacho, E.; Soler, J. M. Projection of plane-wave calculations into atomic orbitals. *Solid State Communications* 1995, 95, 685–690.

(19) Motamarri, P.; Das, S.; Rudraraju, S.; Ghosh, K.; Davydov, D.; Gavini, V. DFT-FE – A massively parallel adaptive finite-element code for large-scale density functional theory calculations. *Comp. Phys. Commun.* 2020, 246, 106853.

(20) Das, S.; Motamarri, P.; Subramanian, V.; Rogers, D. M.; Gavini, V. DFT-FE 1.0: A massively parallel hybrid CPU-GPU density functional theory code using finite-element discretization. *Computer Physics Communications* 2022, 108473.

(21) Gordon Bell Prize finalists named. 2019; [https://cacm.acm.org/careers/240486-gordon-bell-prize-finalists-named/fulltext](https://cacm.acm.org/careers/240486-gordon-bell-prize-finalists-named/fulltext).

(22) Das, S.; Motamarri, P.; Gavini, V.; Turcksin, B.; Li, Y. W.; Leback, B. Fast, scalable and accurate finite-element based ab initio calculations using mixed precision computing: 46 PFLOPS simulation of a metallic dislocation system. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. 2019; pp 1–11.

(23) Williamson, A. J.; Reboredo, F. A.; Galli, G. Chemisorption on semiconductor nanocomposites: A mechanism for hydrogen storage. *Applied Physics Letters* 2004, 85, 2917–2919.

(24) Pask, J. E.; Sterne, P. A. Finite element methods in ab initio electronic structure
calculations. *Modelling and Simulation in Materials Science and Engineering* **2005**, *13*, R71–R96.

(25) Tsuchida, E.; Tsukada, M. Adaptive finite-element method for electronic-structure calculations. *Phys. Rev. B* **1996**, *54*, 7602–7605.

(26) Motamarri, P.; Nowak, M.; Leiter, K.; Knap, J.; Gavini, V. Higher-order adaptive finite-element methods for Kohn-Sham density functional theory. *J. Comput. Phys.* **2013**, *253*, 308–343.

(27) Kronik, L.; Makmal, A.; Tiago, M. L.; Alemany, M. M. G.; Jain, M.; Huang, X.; Saad, Y.; Chelikowsky, J. R. PARSEC – the pseudopotential algorithm for real-space electronic structure calculations: recent advances and novel applications to nanostructures. *physica status solidi (b)* **2006**, *243*, 1063–1079.

(28) Ghosh, S.; Suryanarayana, P. SPARC: Accurate and efficient finite-difference formulation and parallel implementation of Density Functional Theory: Isolated clusters. *Computer Physics Communications* **2017**, *212*, 189–204.

(29) Genovese, L.; Videau, B.; Ospici, M.; Deutsch, T.; Goedecker, S.; Méhaut, J.-F. Daubechies wavelets for high performance electronic structure calculations: The BigDFT project. *Comptes Rendus Mécanique* **2011**, *339*, 149–164, High Performance Computing.

(30) Bunge, C. F.; Barrientos, J. A.; Bunge, A. V. Roothaan-Hartree-Fock Ground-State Atomic Wave Functions: Slater-Type Orbital Expansions and Expectation Values for $Z = 2$–54. *Atom. Data Nucl. Data Tabl.* **1993**, *53*, 113–162.

(31) Slater, J. C. Atomic Shielding Constants. *Phys. Rev.* **1930**, *36*, 57–64.

(32) Hamann, D. R. Optimized norm-conserving Vanderbilt pseudopotentials. *Phys. Rev. B* **2013**, *88*, 085117.
(33) Bai, Z.; Demmel, J.; Dongarra, J.; Ruhe, A.; van der Vorst, H. *Templates for the solution of Algebraic Eigenvalue Problems: A Practical Guide*; SIAM, Philadelphia, 2000.

(34) Brenner, S. C.; Scott, L. R. *The Mathematical Theory of Finite-element Methods*; Springer, New York, 2002.

(35) Canuto, C.; Hussaini, M. Y.; Quarteroni, A.; Zang, T. A. *Spectral methods: evolution to complex geometries and applications to fluid dynamics*; Springer Science & Business Media, 2007.

(36) Demmel, J. W. *Applied Numerical Linear Algebra*; Society for Industrial and Applied Mathematics, 1997.

(37) Cuppen, J. J. A Divide and Conquer Method for the Symmetric Tridiagonal Eigenproblem. *Numer. Math.* 1980, *36*, 177–195.

(38) Maintz, S.; Deringer, V. L.; Tchougréeff, A. L.; Dronskowski, R. LOBSTER: A tool to extract chemical bonding from plane-wave based DFT. 2016.

(39) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* 1994, *50*, 17953–17979.

(40) Langreth, D. C.; Mehl, M. J. Beyond the local-density approximation in calculations of ground-state electronic properties. *Phys. Rev. B* 1983, *28*, 1809–1834.

(41) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* 1996, *77*, 3865–3868.

(42) van Setten, M.; Giantomassi, M.; Bousquet, E.; Verstraete, M.; Hamann, D.; Gonze, X.; Rignanese, G.-M. The PseudoDojo: Training and grading a 85 element optimized norm-conserving pseudopotential table. *Computer Physics Communications* 2018, *226*, 39–54.