Chemical Bonding in Large Systems Using Projected Population Analysis from Real-Space Density Functional Theory Calculations

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

Cite This: *J. Chem. Theory Comput.* 2023, 19, 4216−4231

Read Online

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 toward extracting chemical bonding information from large-scale DFT calculations on materials systems involving thousands of atoms while accommodating periodic, semiperiodic, or fully nonperiodic boundary conditions. Toward this, we derive the relevant mathematical expressions and develop efficient numerical implementation procedures that are scalable on multinode CPU architectures to compute the projected overlap and Hamilton populations. The population analysis is accomplished by projecting either the self-consistently converged FE discretized Kohn−Sham orbitals or the FE discretized Hamiltonian onto a subspace spanned by a localized atom-centered basis set. The proposed methods are implemented in a unified framework within the **DFT−FE** code where the ground-state DFT calculations and the population analysis are performed on the same FE grid. We further benchmark the accuracy and performance of this approach on representative material systems involving periodic and nonperiodic 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 quantitative chemical bonding information of hydrogen chemisorbed in large silicon nanoparticles alloyed with carbon, a candidate material for hydrogen storage.
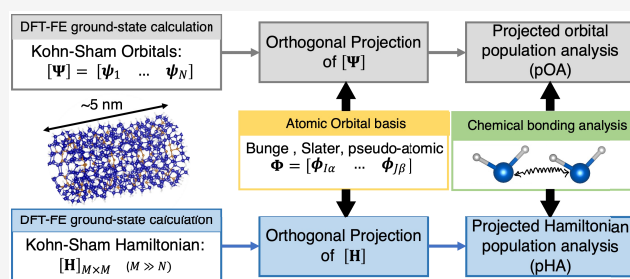
1. INTRODUCTION

Approaches based on overlap[1−3] and Hamilton population analysis[3,4] are widely used to extract chemical bonding information in covalent material systems. The overlap population analysis is based on partitioning the number of electrons among distinct atoms and the orbitals around them, whereas the total electronic energy of a molecule or a crystal is partitioned in Hamilton population analysis. In the case of solid-state systems, these approaches are referred to as crystal orbital overlap population (COOP) originally discussed by Hughbanks and Hoffmann[4] and crystal orbital Hamilton population (COHP) originally suggested by Dronskowski and Blöchl.[5−7] Traditionally, these methods[8,9] were used within the framework of tight-binding linear combination of atomic orbitals (LCAO) or linearized muffin tin orbital approaches (LMTO)[10] which have minimal and a well localized atom-centered basis set. Building on these techniques are approaches like "balanced crystal orbital overlap population" (BCOOP)[11] and "crystal orbital bond index" (COBI)[12] for robust extraction of chemical bonding behavior in solid-state materials. BCOOP was proposed in the context of less localized basis sets which are close to linear dependency, while the COBI approach was proposed for studying multicenter

interactions via a multicenter bond index. Another established way for analyzing chemical bonding in both molecular and solid-state systems is by using the localized orbitals constructed as unitary transformations of extended single-particle eigenstates. For instance, maximally localized Wannier functions[13] (the solid-state equivalent of Foster−Boys orbitals in quantum chemistry) and the recent Pipek−Mezey Wannier functions[14] have been used for studying bonding characterization of crystalline and disordered materials.

Over the last few decades, plane-waves have become the popular choice of basis sets for electronic structure calculations due to the systematic convergent nature of the basis set, offering spectral convergence rates to compute the ground-state properties of interest. Recent focus for extracting chemical bonding behavior has been on projected population analysis[15−18] as these methods combine the advantages of
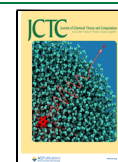
plane-wave-based DFT methods for accurately computing the electronic structure and minimal localized atom-centered basis for understanding chemical bonding properties as a post-processing step. In these methods, Kohn−Sham DFT eigenfunctions obtained from a plane-wave calculation are projected onto a subspace spanned by the localized atomic-orbital basis to compute energy-resolved quantities, such as overlap and Hamilton populations. The most popular and widely used code based on such a projected population analysis approach is **LOBSTER**.[17] Here, the Kohn−Sham eigenfunctions obtained from projector augmented wave (PAW)-based DFT calculations using popular plane-wave-based codes (e.g., VASP,[19] Quantum Espresso[20]) are projected onto a subspace spanned by a localized atom-centered basis. While such a strategy has been largely successful, this approach has certain limitations in extracting chemical bonding information. First, plane-wave-based techniques often restrict the simulation domains to be periodic, which is incompatible with many application problems (e.g., defects, nanoparticles, charged systems). Furthermore, the plane-wave basis provides uniform spatial resolution and is computationally inefficient in the study of defects and 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 sets are extended in real-space and involve all-to-all 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 wave functions as an input to conduct projected population analysis, suffers from the above limitations and is restricted only to bulk material systems (periodic systems) up to a maximum of hundred atoms 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.[18] recently proposed a population analysis where the Kohn−Sham occupied eigenspace obtained from plane-wave DFT calculations are projected onto a localized Wannier orbital basis,[13] thereby minimizing the projection error from plane-wave to localized atom-centered basis (spill factor[21]) 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 accomplish three tasks (ground-state DFT calculation by a plane-wave code, Wannierization (using **wannier90**), and finally the population analysis code). Currently, there are no computational methods available that can perform chemical bonding analysis from large-scale density functional theory (DFT) calculations using a systematically convergent basis set while also having the ability to handle complex material systems with fully periodic, nonperiodic, or semiperiodic boundary conditions. The aim of the current work is to fill in this gap and provide a solution to this problem.

Addressing the aforementioned limitations, we introduce a scalable and computationally efficient approach[22,23] to conduct projected population analysis from real-space finite-element (FE)-based density functional theory (**DFT−FE**) calculations. The FE basis set is systematically convergent and strictly local,

comprising piecewise polynomial functions of order p on which various electronic fields are represented. In contrast to widely used 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 nonperiodic boundary conditions. Additionally, the local character of the FE basis provides an inherent benefit in terms of parallel scalability of DFT-FE calculations in comparison to widely used DFT codes and has been tested up to 100,000 cores on many-core CPUs[22] and 24,000 GPUs on hybrid CPU−GPU architectures.[23−25] The proposed population analysis methodology developed within the framework of **DFT−FE** inherits these advantages and enables scalable chemical bonding analysis in complex material systems. Furthermore, this methodology is developed as a unified approach that enables both Kohn−Sham DFT ground-state calculations and population analysis to be carried out within the same computational framework using the FE basis. The framework opens up the possibility of extracting chemical bonding information for the first time in sizable complex material systems critical in many technologically relevant applications, enabling efficient investigation of chemical bonding interactions in various scenarios, such as large-scale nanoparticles, layered materials with adsorbate−adsorbent interactions, complex defect−impurity interactions, bonding interactions between the migrating ion and the underlying solid electrolyte lattice in the presence of an electric field, and many more.

We propose two methodologies for computing overlap and Hamilton populations via projected population analysis: (a) projected orbital population analysis (pOA), which relies on orthogonally projecting the self-consistently converged FE discretized Kohn−Sham DFT eigenfunctions onto a subspace spanned by a minimal atomic-orbital basis set and is similar in spirit to **LOBSTER**,[17] and (b) projected Hamiltonian population analysis (pHA), which relies on orthogonally projecting the self-consistent FE discretized Hamiltonian onto the atomic-orbital subspace, a method motivated from the fact that many of the reduced scaling electronic structure codes targeted toward large-scale DFT calculations tend to avoid explicit computation of DFT eigenfunctions with no explicit access to these for projection.

The computational framework developed to implement the above 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 $\mathbb{V}_\phi^{N_{orb}}$ (available as numerical data or analytical expressions) by interpolating these orbitals on the underlying finite-element grid; (iii) orthogonally project the occupied Kohn−Sham eigenfunctions onto $\mathbb{V}_\phi^{N_{orb}}$ in the case of pOA, while orthogonally projecting the self-consistent Kohn−Sham FE discretized Hamiltonian onto $\mathbb{V}_\phi^{N_{orb}}$ in the case of pHA; (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 wave functions in the atom-centered orbital basis $\mathbb{V}_\phi^{N_{orb}}$ in the case of pOA, while diagonalizing the projected Hamiltonian to obtain the eigenvector matrix in the subspace $\mathbb{V}_\phi^{N_{orb}}$ in the case of pHA; (vi) using these coefficient

matrices, evaluate the projected orbital overlap and Hamilton population in the case of pOA and evaluate the projected Hamiltonian overlap and Hamilton population in the case of pHA.

We evaluate the accuracy and performance of the proposed methods (pOA and pHA) on representative benchmark examples involving isolated molecules (CO, H_2O, O_2, Si–H nanoparticles) and a periodic system involving a carbon diamond supercell. We first benchmark the results from the pOA method with those 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 pOA 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 pOA using **MPI** and illustrate the reduction in wall time of the population analysis by ∼70% when scaled up to 1120 CPU cores from 280 CPU cores on a Si nanoparticle system containing 1090 atoms. We note that these large-scale calculations are not currently feasible using **LOBSTER**. Subsequently, we compare the accuracy and performance of the pHA approach with that of pOA. The results obtained by the pHA approach agree very well with those obtained by the pOA approach. We further show the advantage of using pHA in computational wall time compared to pOA 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.[26] Toward 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 favoring the release of H_2.

The remainder of our article is structured as follows: Section 2 discusses the mathematical background and relevant finite-element (FE) discretization aspects required for describing the projected population analysis within the FE formalism in the subsequent sections. Projected orbital population analysis (pOA) is discussed in section 3, highlighting the aspects of mathematical formulation, accuracy validation, and performance comparison results with **LOBSTER**. Section 4 discusses the details of projected Hamiltonian population analysis (pHA) and highlights the advantages of pHA over pOA. We subsequently discuss a case study illustrating the usefulness of large-scale chemical bonding analysis in section 5, concluding with a short discussion and outlook in section 6.

2. MATHEMATICAL BACKGROUND

In this section, we introduce the notations and discuss the key mathematical preliminaries and the relevant finite-element (FE) discretization aspects required for subsequently describing the projected population analysis within the FE formalism in sections 3 and 4.

Let $\mathbb{H}$ denote an infinite-dimensional Hilbert space, where we assume the Kohn–Sham eigenfunctions of the continuous problem exist. $\mathbb{H}$ 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 $\mathbb{V}^M \subset \mathbb{H}$. 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 convergent basis set—plane waves,[19,20] finite element basis,[27−29] finite difference approach,[30,31] wavelets,[32] etc., all which can be employed to numerically solve the partial differential equation representing the Kohn–Sham DFT eigenvalue problem. Consequently, the discretized spin unpolarized DFT eigenvalue problem to be solved for N-smallest eigenvalue−eigenvector pairs is given by

$$\mathcal{H}|\psi_i\rangle = \epsilon_i |\psi_i\rangle \quad \text{for} \quad i = 1, 2, ..., N \quad \text{with} \quad N \geq \frac{N_e}{2} \tag{1}$$

where $|\psi_i\rangle \in \mathbb{V}^M$ denotes the eigenfunction of $\mathcal{H}$ and N_e is the number of electrons in the given material system.

Extracting chemical bonding behavior using projected population approaches requires us to define an N_{orb}-dimensional subspace $\mathbb{V}_\phi^{N_{orb}}$ ($N_{orb} < M$), spanned by the localized non-orthogonal atom-centered auxiliary basis set $\{|\phi_\mu\rangle\}$. These basis functions are constructed for the given configuration of atoms in the material system and are chosen to be minimal such that the occupied Kohn–Sham wave functions $|\psi_I\rangle$ are well-represented in $\mathbb{V}_\phi^{N_{orb}}$ while providing accurate insights into the chemical bonding behavior. Various types of minimal atomic-orbital basis functions have been used in the past, such as Slater-type orbital expansions of Hartree–Fock wave functions by Bunge et al.[33] (henceforth referred to as STO basis by Bunge), functions fitted to PAW wave functions,[17] etc. Pseudoatomic (PA) orbitals constructed from norm-conserving pseudopotentials[34] also constitute a convenient choice of atom-centered basis sets for chemical bonding analysis, as demonstrated in the current work.

In the current work, the discretized Kohn–Sham eigenvalue problem in eq 1 is represented in finite-element (FE) basis,[35] a strictly local and a piece-wise continuous Lagrange polynomial basis interpolated over Gauss–Lobatto–Legendre nodal points. We refer to our prior work[22,23,29] for more details on the spectral FE discretization of the Kohn–Sham DFT eigenvalue problem. To this end, the representation of various fields employed in computing projected population analysis subsequently—the Kohn–Sham wave functions ($\langle \mathbf{x}|\psi_i\rangle = \psi_i(\mathbf{x})$) and the localized atom-centered functions ($\langle \mathbf{x}|\phi_\mu\rangle = \phi_\mu(\mathbf{x})$)—in the FE basis is given by

$$\psi_i(\mathbf{x}) = \sum_{j=1}^{M} N_j^h(\mathbf{x})\psi_i^j, \qquad \phi_\mu(\mathbf{x}) = \sum_{j=1}^{M} N_j^h(\mathbf{x})\phi_\mu^j \tag{2}$$

where $N_j^h : 1 \leq j \leq M$ denote the M finite-element (FE) basis functions spanning the M-dimensional space $\mathbb{V}^M$. These are strictly local Lagrange polynomials of degree p generated using the nodes of the FE triangulation $\mathcal{T}^h$, with the characteristic mesh size denoted by h. Further in eq 2, ψ_i^j and ϕ_μ^j denote the coefficients in the expansion of the ith discretized Kohn–Sham wave function ($\psi_i(\mathbf{x})$) and the μth atom-centered localized basis function ($\phi_\mu(\mathbf{x})$). These coefficients constitute the nodal values of the discretized fields represented using the FE triangulation $\mathcal{T}^h$ since the FE basis functions $N_j^h(\mathbf{x})$ satisfy the Kronecker-delta property, i.e., $N_j^h(\mathbf{x}_k) = \delta_{jk}$, where $\mathbf{x}_k$ denotes

the kth nodal point of $\mathcal{T}^h$. The nodal values ψ_i^j 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.[22] and on hybrid CPU−GPU architectures in Das et al.[23] Furthermore, in the current work, the nodal values ϕ_μ^j are computed from the atom-centered orbital data, which are usually available as analytical expressions or in the form of numerical data.

Finally, we introduce the atom-centered orbital overlap matrix $\mathbf{S}$ with matrix entries $S_{\alpha\beta} = \langle \phi_\alpha | \phi_\beta \rangle$, a key quantity in evaluating projected populations, as discussed in the subsequent sections. Using the FE representation of $\phi_\mu(\mathbf{x})$ in eq 2, the matrix entries of $\mathbf{S}$ evaluated in a FE discretized setting is given by

$$\mathbf{S} = \underline{\boldsymbol{\Phi}}^\dagger \underline{\boldsymbol{\Phi}} \quad \text{with} \quad \underline{\boldsymbol{\Phi}} = \mathbf{M}^{1/2} \boldsymbol{\Phi} \tag{3}$$

where $\boldsymbol{\Phi}$ denotes an $M \times N_{\text{orb}}$ matrix whose columns are the components of $\phi_\mu(\mathbf{x})$ in the FE basis (see eq 2) and the $M \times M$ matrix $\mathbf{M}$ denotes the FE basis overlap matrix with entries given by $M_{pq} = \int_\Omega N_p(\mathbf{x}) N_q(\mathbf{x})\, d\mathbf{x}$. Efficient computation of $\mathbf{M}^{1/2}$ and, subsequently, the $\mathbf{S}$ matrix is crucial for evaluating projected populations in the FE setting. We refer the reader to the Supporting Information (see section S1.1) for more details about the computation of $\mathbf{M}^{1/2}$ and $\mathbf{S}$ matrices in a parallel computing environment.

3. PROJECTED ORBITAL POPULATION ANALYSIS (pOA)

Projected orbital population analysis, henceforth referred to as pOA, relies on the orthogonal projection of numerically computed Kohn−Sham DFT eigenfunctions onto a subspace spanned by localized atomic orbitals to extract the chemical bonding behavior, and is in the spirit of Sanchez-Portal et al.[21] and Deringer et al.[15] In this section, we begin by discussing the mathematical formulation and, subsequently, the related expressions in a FE setting required for implementing pOA. We then assess the accuracy and performance of the proposed implementation with **LOBSTER**, a widely used package for conducting projected orbital population analysis.

For clarity and simplicity, we assume that the Kohn−Sham DFT eigenproblem in eq 1 is solved in a simulation domain with fully nonperiodic boundary conditions or a supercell employing periodic/semiperiodic boundary conditions with a Γ point to sample the Brillouin zone. The extension to a periodic unit-cell involving Brillouin zone integration via multiple $\boldsymbol{k}$-point sampling is not explicitly considered in this section. However, the expressions and the benchmark results for $\boldsymbol{k}$-dependent projected population analysis within the framework of pOA are discussed in the Supporting Information (see sections S1.1 and S2.1). A schematic of the overall implementation strategy is shown in Figure 1.

3.1. Mathematical Formulation. To begin, we introduce the orthogonal projection operator $\mathcal{P}^\phi : \mathbb{V}^M \to \mathbb{V}_\phi^{N_{\text{orb}}}$ which can be written as $\mathcal{P}^\phi = \sum_{\alpha,\beta=1}^{N_{\text{orb}}} |\phi_\alpha\rangle (S^{-1})_{\alpha\beta} \langle \phi_\beta|$, with atomic-orbital overlap matrix $S_{\alpha\beta} = \langle \phi_\alpha | \phi_\beta \rangle$ as introduced before. Denoting the orthogonal projection of the Kohn−Sham eigenfunction $|\psi_i\rangle \in \mathbb{V}^M$ onto the subspace $\mathbb{V}_\phi^{N_{\text{orb}}}$ to be $|\psi_i^\phi\rangle$,

we have $|\psi_i^\phi\rangle = \mathcal{P}^\phi |\psi_i\rangle$ for $i = 1, 2, ..., N$. We note that the projected Kohn−Sham wave functions $\{|\psi_i^\phi\rangle\}$ need not form an orthonormal set and hence Löwdin symmetric orthogonalization[36] is employed to orthonormalize the projected Kohn−Sham wave functions. To this end, we denote the orthonormalized projected wave function as $|\tilde{\psi}_i^\phi\rangle$ where $|\tilde{\psi}_i^\phi\rangle = \sum_j^{N_{\text{orb}}} O_{ij}^{-1/2} |\psi_j^\phi\rangle$, with $O_{ij} = \langle \psi_i^\phi | \psi_j^\phi \rangle$ denoting the matrix elements of the overlap matrix $\mathbf{O}$ corresponding to $\{|\psi_i^\phi\rangle\}$.

Projected Orbital Overlap Population (pOOP). Recalling that $\langle \tilde{\psi}_j^\phi | \tilde{\psi}_j^\phi \rangle$ equals 1 and the fact that the number of electrons N_e in a given material system is related to the density of states $\delta(\epsilon - \epsilon_j)$, we can write

$$N_e = \sum_{j=1}^N \int_{-\infty}^\infty \langle \tilde{\psi}_j^\phi | \tilde{\psi}_j^\phi \rangle f(\epsilon, \epsilon_F) \delta(\epsilon - \epsilon_j)\, d\epsilon \tag{4}$$

where f denotes the orbital occupancy function usually given by the Heaviside function with a value of 1 if $\epsilon < \epsilon_F$ (Fermi-energy) and 0 otherwise. Using the relations $|\tilde{\psi}_j^\phi\rangle = \sum_j^{N_{\text{orb}}} O_{ij}^{-1/2} |\psi_j^\phi\rangle$ and $|\psi_j^\phi\rangle = \mathcal{P}^\phi |\psi_j\rangle$, the above expression relating N_e and $\delta(\epsilon - \epsilon_j)$ can be expressed in terms of Kohn−Sham wave functions $|\psi_j\rangle \in \mathbb{V}^M$ and the localized atom-centered basis $|\phi_\mu\rangle \in \mathbb{V}^{N_{\text{orb}}}$ in the following way:

$$N_e = \sum_{\nu,\nu'}^{N_{\text{orb}}} \sum_{\mu,\mu'}^{N_{\text{orb}}} \sum_{k,q}^N \sum_j^N O_{jk}^{-1/2} O_{qj}^{-1/2} \int_{-\infty}^\infty f(\epsilon, \epsilon_F) \langle \psi_q | \phi_\mu \rangle$$
$$S_{\mu\mu'}^{-1} \langle \phi_\mu | \phi_\nu \rangle S_{\nu\nu'}^{-1} \langle \phi_\nu | \psi_k \rangle \delta(\epsilon - \epsilon_j)\, d\epsilon$$
$$= \sum_{I\alpha} \sum_{J\beta} \sum_{k,q} \sum_j O_{jk}^{-1/2} O_{qj}^{-1/2} \int_{-\infty}^\infty f(\epsilon, \epsilon_F)$$
$$\langle \psi_q | \phi^{I\alpha} \rangle S_{I\alpha J\beta} \langle \phi^{J\beta} | \psi_k \rangle \delta(\epsilon - \epsilon_j)\, d\epsilon \tag{5}$$

In the above eq 5, $|\phi^\mu\rangle$ denotes the dual of the basis function $|\phi_\mu\rangle$, satisfying the property $\langle \phi^\mu | \phi_\nu \rangle = \langle \phi^\nu | \phi_\mu \rangle = \delta_{\mu\nu}$, 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 above to denote the localized atom-centered basis function $|\phi_\mu\rangle$ as $|\phi_{I\alpha}\rangle$, where α denotes the index of the atomic orbital centered at a nuclear position $\mathbf{R}_I$. 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 the above equation. To this end, the projected orbital overlap population[1,4] $\text{pOOP}_{IJ}(\epsilon)$ associated with a source atom I and a target atom $J \neq I$ is extracted from eq 5 to be defined as

$$\text{pOOP}_{IJ}(\epsilon) = \sum_j \sum_{\alpha\beta} \mathfrak{R}e \left\{ \left(\sum_q O_{qj}^{-1/2} \langle \psi_q | \phi^{I\alpha} \rangle \right) \right.$$
$$\left. S_{I\alpha J\beta} \left(\sum_k O_{jk}^{-1/2} \langle \phi^{J\beta} | \psi_k \rangle \right) \right\} \delta(\epsilon - \epsilon_j) \tag{6}$$

where $\mathfrak{R}e(z)$ refers to the real part of a complex number z. Introducing the finite-element (FE) discretization of various fields in eq 6, we deduce the relevant matrix expressions

required to evaluate the overlap population in the FE setting. To begin, we define a matrix $\overline{C}$ of size $N_{orb} \times N$ whose entries $\overline{C}_{I\alpha}^{j} = \sum_q O_{jq}^{-1/2} \langle \phi^{I\alpha} | \psi_q \rangle$ are the coefficients of $|\tilde{\psi}_j^{\phi}\rangle$ expressed in the localized atom-centered basis set $\{|\phi_{\mu}\rangle\}$. The FE representation of various fields in the expression for $\overline{C}_{I\alpha}^{j}$ allows $\overline{C}$ to be recast in the matrix form as follows

$$\overline{C} = S^{-1} \underline{\Phi}^{\dagger} \underline{\Psi} O^{-1/2} \quad \text{where} \quad \underline{\Psi} = M^{1/2} \Psi \tag{7}$$

where Ψ denotes a $M \times N$ matrix whose column vectors are the components of $|\psi_j\rangle$ in the FE basis, while the $M \times N_{orb}$ atomic-orbital matrix Φ and the $M \times M$ FE basis overlap matrix M are defined in the previous section. We note that the rows of the matrix $\overline{C}$ are stored in the order of atoms and their corresponding atom-centered orbitals for a given atom in succession. To elaborate, the μth row of $\overline{C}$ corresponds to the Ith atom and an atom-centered index α associated with this atom I, while the jth column of this matrix corresponds to the index of the projected Kohn–Sham wave function ($j = 1, ..., N$). More details related to the derivation and efficient computation of the matrix expressions for $\overline{C}$ and O in the FE setting can be found in the Supporting Information (section S1.1). Finally, the *projected orbital overlap population* (pOOP) associated with a source atom I and a target atom J is evaluated by extracting the appropriate entries of the matrices $\overline{C}$ and S and the expression is given by

$$pOOP_{IJ}(\epsilon) = \sum_j \sum_{\alpha\beta} \Re e\{\overline{C}_{I\alpha}^{j*} \overline{C}_{J\beta}^{j} S_{I\alpha J\beta}\} \delta(\epsilon - \epsilon_j) \tag{8}$$

Projected Orbital Hamilton Population (pOHP). Recalling that the electronic band energy (E_{band}) in DFT is related to the expectation value of the Kohn–Sham Hamiltonian $\mathcal{H} \in \mathbb{C}^{M \times M}$ with respect to its occupied eigenstates $|\psi_j\rangle \in \mathbb{V}^M$, we have

$$E_{band} = \sum_{j=1}^{N} \int_{-\infty}^{\infty} f(\epsilon, \epsilon_F) \langle \psi_j | \mathcal{H} | \psi_j \rangle \delta(\epsilon - \epsilon_j) \, d\epsilon \tag{9}$$

Following Maintz et al.,[16] we now define the projected Hamiltonian operator $\mathcal{H}^p$ on the subspace $\mathbb{V}_{\phi}^{N_{orb}}$ in terms of $|\tilde{\psi}_j^{\phi}\rangle$ as $\mathcal{H}^p = \sum_{j=1}^{N} |\tilde{\psi}_j^{\phi}\rangle \epsilon_j \langle \tilde{\psi}_j^{\phi}|$, where ϵ_j denotes the DFT 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_{band} can be written as

$$E_{band} = \sum_{j=1}^{N} \int_{-\infty}^{\infty} f(\epsilon, \epsilon_F) \langle \tilde{\psi}_j^{\phi} | \mathcal{H}^p | \tilde{\psi}_j^{\phi} \rangle \delta(\epsilon - \epsilon_j) \, d\epsilon \tag{10}$$

Along the lines of Maintz et al.,[16] we consider the orthogonalized basis $\{|\hat{\phi}_{\nu}\rangle\}$ obtained via Löwdin symmetric orthonormalization[36] of $\{|\phi_{\nu}\rangle\}$ where these two sets of basis functions 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., $\mathcal{P}^{\phi} = \sum_{i=1}^{N_{orb}} |\hat{\phi}_{\mu}\rangle \langle \hat{\phi}_{\mu}|$, can be used to recast the above equation corresponding to E_{band} as

$$E_{band} = \sum_{\mu,\nu=1}^{N_{orb}} \sum_{k,q,j=1}^{N} O_{jk}^{-1/2} O_{qj}^{-1/2} \int_{-\infty}^{\infty} f(\epsilon, \epsilon_F) \langle \psi_q | \hat{\phi}_{\mu} \rangle$$
$$\langle \hat{\phi}_{\mu} | \mathcal{H}^p | \hat{\phi}_{\nu} \rangle \langle \hat{\phi}_{\nu} | \psi_k \rangle \delta(\epsilon - \epsilon_j) \, d\epsilon$$
$$= \sum_{I\alpha,J\beta}^{N_{orb}} \sum_{k,q,j}^{N} O_{jk}^{-1/2} O_{qj}^{-1/2} \int_{-\infty}^{\infty} f(\epsilon, \epsilon_F) \langle \psi_q | \hat{\phi}_{I\alpha} \rangle$$
$$H_{I\alpha,J\beta}^p \langle \hat{\phi}_{J\beta} | \psi_k \rangle \delta(\epsilon - \epsilon_j) \, d\epsilon \tag{11}$$

where the composite index notation $\mu = \{I\alpha\}$ [$\nu = \{J\beta\}$] has been used for the basis functions $\{|\hat{\phi}_{\mu}\rangle\}$ to denote the αth [β^{th}] basis function centered at the atomic position R_I [R_J] and $H_{I\alpha,J\beta}^p$ denotes the matrix element of $\mathcal{H}^p$ represented in the $\{|\hat{\phi}_{\mu}\rangle\}$ basis. The orbital Hamilton population analysis deals with the partitioning of the band energy E_{band} among the constituent atoms in a given material system, and the projected orbital Hamilton population[5] pOHP can be defined taking recourse to eq 11. To this end, $pOHP_{IJ}(\epsilon)$ associated with a source atom I and a target atom $J \neq I$ is extracted from eq 11 to be defined as

$$pOHP_{IJ}(\epsilon) = \sum_j \sum_{\alpha\beta} \Re e \left\{ \left(\sum_q O_{qj}^{-1/2} \langle \psi_q | \hat{\phi}_{I\alpha} \rangle \right) \right.$$
$$\left. H_{I\alpha,J\beta}^p \left(\sum_k O_{jk}^{-1/2} \langle \hat{\phi}_{J\beta} | \psi_k \rangle \right) \right\} \delta(\epsilon - \epsilon_j) \tag{12}$$

where $\Re e(z)$ refers to the real part of a complex number z. Introducing the finite-element (FE) discretization of various fields in eq 12, we deduce the relevant matrix expressions required to evaluate the Hamilton population in the FE setting. Consequently, we define the $N_{orb} \times N$ matrix $\hat{C}$ whose entries $\hat{C}_{I\alpha}^{j} = \sum_k O_{jk}^{-1/2} \langle \hat{\phi}_{I\alpha} | \psi_k \rangle$ are the coefficients of $|\tilde{\psi}_j^{\phi}\rangle$ expressed in the orthonormalized atom-centered basis set $\{|\hat{\phi}_{\mu}\rangle\}$. The FE representation of various fields in $\hat{C}_{I\alpha}^j$ allows one to recast $\hat{C}$ in terms of the matrices $\overline{C}$ (refer to eq 7) and S (refer to eq 3), i.e., $\hat{C} = S^{1/2}\overline{C}$. In all of our subsequent discussions and computations in this work, we choose $N = N_{orb}$ and we compute the $N_{orb} \times N_{orb}$ projected Hamiltonian matrix H^p using the coefficient matrix as $H^p = \hat{C} D \hat{C}^{\dagger}$ where the matrix D is diagonal and comprised of the Kohn–Sham eigenvalues ϵ_i obtained from the Kohn–Sham DFT problem solved in the FE basis. We refer the reader to the Supporting Information (section S1.1) for more details on the derivation and efficient computation of the matrix expressions for H^p, O, and $\hat{C}$ within the FE framework. Finally, the *projected orbital Hamilton population* (pOHP) associated with the partitioning of energy between a source atom I and a target atom $J \neq I$ is evaluated by extracting the appropriate entries of the matrices $\hat{C}$ and H^p and the expression is given by

$$pOHP_{IJ}(\epsilon) = \sum_j \sum_{\alpha,\beta} \Re e\{\hat{C}_{I\alpha}^{j*} H_{I\alpha,J\beta}^p \hat{C}_{J\beta}^{j}\} \delta(\epsilon - \epsilon_j) \tag{13}$$

Total Computational Complexity Estimate of pOA. The current implementation of the projected orbital population analysis (pOA) assumes $N = N_{orb}$, and thereby, the total computational complexity can be estimated to be $\sim 4 M_{loc} N^2 +$

$26N^3$ (refer to the Supporting Information, section S1.1). Here, M_{loc} is the ratio of the total number of FE degrees of freedom (M) to the number of MPI tasks (P) on a parallel computing system. M_{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 exceeds $4M/26N$.

Spilling Factor. The total spilling factor or charge spilling factor, as introduced by Sanchez-Portal et al.,[21] describes the ability of the atom-centered localized basis spanning $\mathbb{V}_\phi^{N_{orb}}$ to represent the FE discretized Kohn−Sham eigenfunctions $|\psi_j\rangle$, the self-consistent solution of the FE discretized Kohn−Sham eigenvalue problem. The charge spilling factor is given by the average of the L_2 projection errors of the Kohn−Sham occupied eigenstates, while the total spilling factor is computed as the average of the L_2 projection errors of the Kohn−Sham eigenstates considered for the projection. To this end, we compute the absolute total spilling factor $\mathcal{S}$ and absolute charge spilling factor $\mathcal{S}_c$ in the spirit of Stefan Maintz et al.[37] and they are expressed in terms of the diagonal entries of the matrix $\mathbf{O}$ (see eq 6 in the Supporting Information) as

$$\mathcal{S} = \frac{1}{N} \sum_{i=1}^{N} |1 - \langle \psi_i^\phi | \psi_i^\phi \rangle| = \frac{1}{N} \sum_{i=1}^{N} |1 - O_{ii}|,$$

$$\mathcal{S}_c = \frac{1}{N_{occ}} \sum_{i=1}^{N_{occ}} |1 - O_{ii}|$$

$$(14)$$

Projected Orbital Density Error (pODE). We compute the L_2 norm of the difference between the ground-state electron density $(\rho(\mathbf{x}))$ computed from **DFT−FE** using the FE discretized occupied Kohn−Sham eigenfunctions $\{|\psi_i\rangle\}$ and the electron density $(\rho^o(\mathbf{x}))$ computed from the Löwdin symmetric orthonormalized[36] projected Kohn−Sham wave functions $\{|\tilde{\psi}_i^\phi\rangle\} \in \mathbb{V}_\phi^{N_{orb}}$. To this end, pODE is evaluated as

$$pODE = \frac{\|\rho(\mathbf{x}) - \rho^o(\mathbf{x})\|_2}{\|\rho(\mathbf{x})\|_2} \quad \text{where}$$

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

$$(15)$$

3.2. Results: Accuracy and Performance Benchmarking. We now assess the accuracy and performance of the proposed pOA implementation within the framework of **DFT−FE**[22,23] by comparing with **LOBSTER**. To this end, we employ Quantum Espresso[20] (**QE**) with PAW[38] pseudopotentials to perform DFT calculations for all of the benchmark systems, and the resulting ground-state DFT wave functions and eigenvalues are used as input for conducting the population analysis using **LOBSTER**. The calculations using **QE** are performed by employing the internal implementation of GGA[39] exchange-correlation of the PBE[40] form. However, in **DFT−FE** calculations, we employ optimized norm-conserving Vanderbilt (ONCV)[34] pseudopotentials from the pseudoDojo database[41] to conduct pseudopotential DFT calculations using GGA[39] exchange-correlation of the PBE[40] form employing the **Libxc** package.[42] We employ nonperiodic boundary conditions for isolated systems and periodic boundary conditions for crystalline systems in **DFT−FE**. The population analysis methodology discussed in this section is implemented
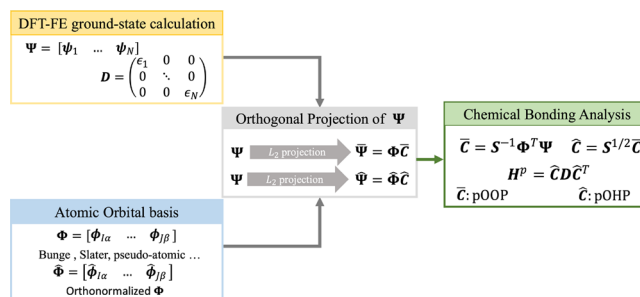

Figure 1. Overview of the implementation strategy for projected orbital population analysis (pOA) within the finite-element (FE) framework. The strategy involves projecting the self-consistently converged Kohn−Sham eigenfunctions obtained from **DFT−FE** onto a localized atom-centered basis represented on the FE grid. Subsequently, the overlap matrix ($\mathbf{S}$), the projected Hamiltonian ($\mathbf{H}^p$), and the coefficient matrices $\bar{\mathbf{C}}$ and $\hat{\mathbf{C}}$ are evaluated which in turn are used to compute the projected orbital overlap population pOOP and the projected orbital Hamilton population pOHP, respectively.

in the **DFT−FE** code, and we note that the numerical implementation of population analysis in **DFT−FE** takes advantage of parallel computing architectures via **MPI** (Message Passing Interface), enabling chemical bonding analysis of large-scale systems in a unified computational framework.

For all of the benchmark calculations reported here, the cutoff energies in **QE** and mesh sizes in **DFT−FE** are chosen such that a discretization error of $O(10^{-5}) \frac{E_h}{\text{atom}}$ in ground-state energy and a force discretization error of $O(10^{-4}) \frac{E_h}{\text{bohr}}$ 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 [PARAM Pravega is one of India's fastest supercomputers stationed at Indian Institute of Science comprised of 584 Intel Xeon Cascade-Lake-based CPU nodes (28,032 cores)].

In this subsection, the computations of pOOP and pOHP as described in eq 8 and eq 13 are validated using pCOOP and pCOHP obtained from **LOBSTER** in terms of both accuracy and performance. To this end, we project the self-consistently converged Kohn−Sham (KS) wave functions obtained from **DFT−FE** onto a subspace spanned by (i) the STO basis by Bunge[33] and (ii) pseudoatomic (PA) orbitals constructed from the ONCV[34] pseudopotentials. On the other hand, the converged ground-state DFT wave functions obtained using the PAW formalism in **QE** are used as input to **LOBSTER**, which are in turn projected onto a subspace spanned by localized atom-centered basis functions known as pbevaspfit2015.[37] In our benchmark studies, we consider (i) isolated systems comprised of CO, spin-polarized O_2, H_2O molecules, and a Si−H nanoparticle with 65 atoms ($Si_{29}H_{36}$) and (ii) a 2 × 2 × 2 periodic carbon diamond supercell. To simulate isolated systems in **DFT−FE**, we consider the simulation domain large enough to allow the wave functions to decay to zero on the boundary by employing nonperiodic boundary conditions. In contrast, in **QE**, which always employs periodic boundary conditions, we consider a simulation domain with a sufficient vacuum to minimize the image−image interactions. In the benchmark problem involving a bulk material system, we consider a 2 × 2 × 2 diamond supercell employing periodic boundary conditions using a Γ-point to sample the Brillouin

zone in both **DFT-FE** and **QE**. We now describe the comparative study of the projection spill factors, the population analysis energy diagrams, and the computational costs between the proposed implementation in **DFT-FE** and **LOBSTER**.

Accuracy Validation of pOA. The projected orbital population analysis (pOA) implemented in DFT-FE is compared with **LOBSTER** for the case of $Si_{29}H_{36}$ and the periodic $2 \times 2 \times 2$ supercell of carbon. Table 1 shows the

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

System	LOBSTER[a]		DFT-FE Bunge[b]		DFT-FE PA[c]	
	S_c	S	S_c	S	S_c	S
Carbon diamond $2 \times 2 \times 2$ periodic supercell	0.010	0.092	0.011	0.101	0.003	0.068
$Si_{29}H_{36}$ nanoparticle	0.012	0.308	0.039	0.314	0.017	0.271

[a]Projection in **LOBSTER** uses pbeVaspfit2015 as an auxiliary atom-centered basis. [b]**DFT-FE** Bunge indicates the projection of finite-element discretized Kohn–Sham eigenfunctions to STO basis by Bunge. [c]**DFT-FE** PA basis indicates projection onto pseudoatomic orbitals.

comparison of the absolute spilling factor S and the absolute charge spilling factor S_c for these systems, and we note that our spill factors are in close agreement with those obtained from **LOBSTER**. Furthermore, we observe that the spill factors S and S_c obtained in our pOA approach employing pseudoatomic (PA) orbitals are smaller in comparison to the spill factors obtained using STO basis by Bunge. 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 **DFT-FE** using ONCV pseudopotential calculations. We use these PA orbitals for our subsequent comparisons with **LOBSTER** in this section, and comparisons with the STO basis by Bunge are discussed in the Supporting Information (see section S2.1).

Next, we illustrate the comparisons of population energy diagrams in Figures 2 and 3, both for the case of a periodic $2 \times 2 \times 2$ carbon diamond supercell and a $Si_{29}H_{36}$ nanoparticle. In the case of the carbon diamond supercell, a pair of nearest neighbor 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 2 both for overlap population and for Hamilton population. A comparison of total populations is also illustrated in this figure. These results indicate that the energy diagrams obtained with **LOBSTER** match very well with those of our current approach. The locations of the bonding and antibonding peaks are identical to those 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 the $Si_{29}H_{36}$ nanoparticle in which the Si atom and the nearest H atom are picked as the source and target atoms, and the corresponding contributions of H_{1s}–Si_{3p} and H_{1s}–Si_{3s} orbitals are plotted in Figure 3 (see the inset in the figure) for both the overlap population and the Hamilton population. Comparisons of spill factors and the

population energy diagrams for benchmark systems involving molecules—CO, spin-polarized O_2, and H_2O—are discussed in section S2.1, and we observe a close agreement of the pOOP and pOHP results with those obtained using **LOBSTER**. Comparisons with **LOBSTER** involving k-dependent population analysis are discussed in the Supporting Information in the case of a $1 \times 1 \times 1$ orthogonal unit-cell of the carbon diamond structure (see section S2.1).

Performance Comparison. The computational CPU times in terms of node-secs (wall time taken on 1 compute node), measuring the computation of the overlap population and the Hamilton population, are tabulated in Table 2 for the proposed pOA approach and **LOBSTER**. The benchmark systems involving a $2 \times 2 \times 2$ carbon diamond supercell and Si–H nanoparticles—$Si_{29}H_{36}$ and $Si_{145}H_{150}$—are chosen for comparison. The computational times indicate a significant advantage for the proposed implementation compared to **LOBSTER**. For instance, computational gains up to 2× are observed for the case of a carbon diamond $2 \times 2 \times 2$ periodic supercell, while, in the case of a $Si_{29}H_{36}$ nanoparticle, a speed-up of 14× is observed. This higher speed-up in the case of a Si nanoparticle using the proposed pOA approach can be attributed to the use of nonperiodic boundary conditions in our implementation, while **LOBSTER** is restrictive in terms of the boundary conditions one can employ and uses periodic boundary conditions even for a 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 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× with an increase in the number of CPU cores from 280 to 1120.

4. PROJECTED HAMILTONIAN POPULATION ANALYSIS (pHA)

We propose here *Projected Hamiltonian population analysis*, henceforth referred to as pHA, as an alternate approach to pOA described previously, to conduct both overlap and Hamilton population analysis. The necessity of this alternate approach is motivated by the fact that many of the reduced scaling electronic structure codes targeted toward large-scale DFT calculations tend to avoid explicit computation of DFT eigenfunctions having no access to these for projection. To this end, in this approach, we orthogonally project the self-consistently converged discretized DFT Hamiltonian onto the subspace spanned by the minimal atomic-orbital basis set to extract the chemical bonding information from DFT calculations. This is in contrast to the previous pOA approach where the self-consistently converged Kohn–Sham eigenfunctions are projected. As will be demonstrated subsequently, population analysis via pHA shows computational advantage over pOA in wall times with an increase in the number of MPI tasks for large material system sizes. In this section, we discuss the mathematical formulation and derive the relevant expressions in a finite-element setting required for implementing pHA. Figure 4 shows a schematic of the overall pHA implementation strategy. We subsequently compare the accuracy and performance of the proposed implementation with those of pOA, which was discussed earlier. For clarity, the extension to a periodic unit-cell involving Brillouin zone
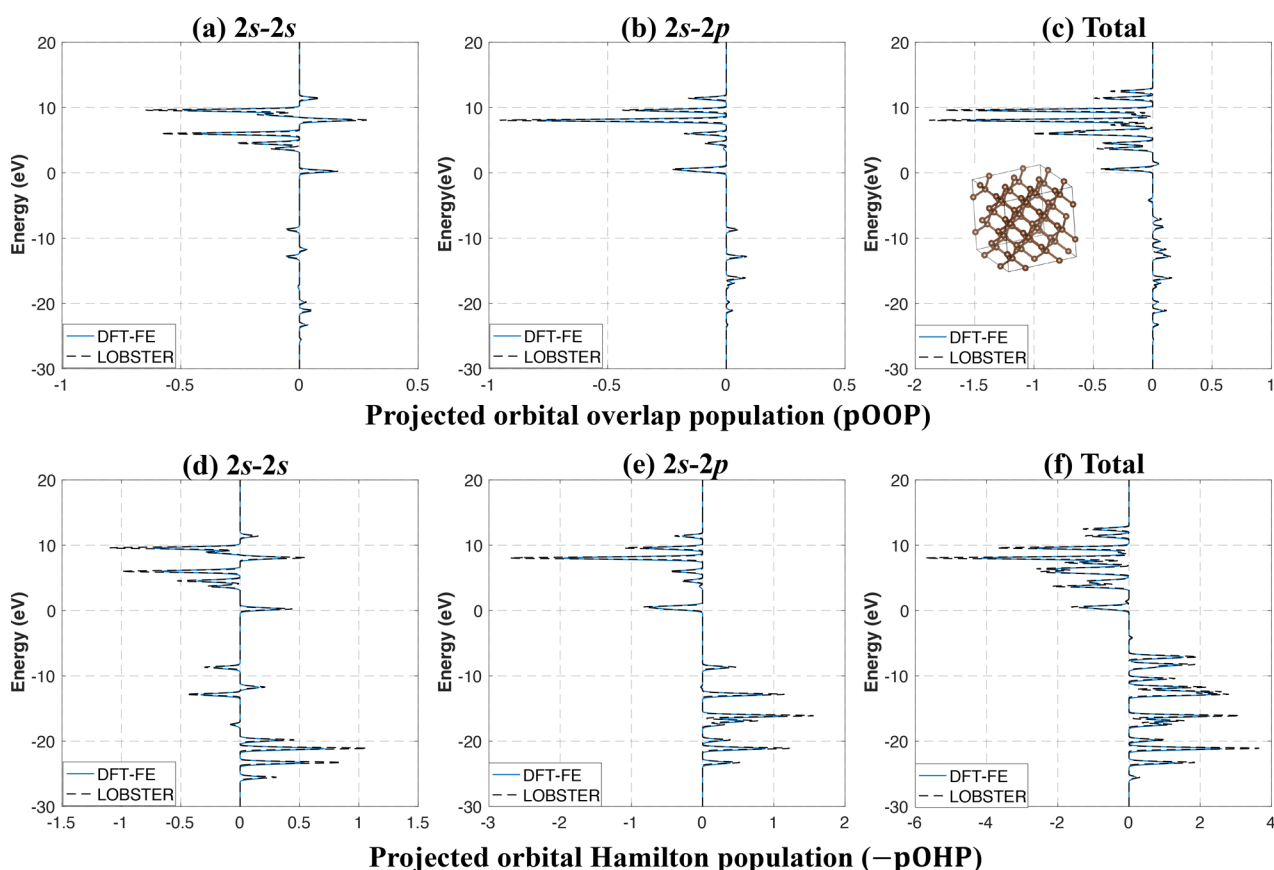
Figure 2. Comparison between pOA implemented in **DFT−FE** using PA orbitals and **LOBSTER** for nearest neighbor C atoms in a carbon diamond supercell. The top row shows the projected orbital overlap population, with the subfigures (a) and (b) in this row showing the contributions of $C_{2s}−C_{2s}$ and $C_{2s}−C_{2p}$ to the total orbital overlap population that is plotted in subfigure (c). The bottom row shows the negative of the projected orbital Hamilton population. The subfigures in the bottom row (d) and (e) show the contributions of $C_{2s}−C_{2s}$ and $C_{2s}−C_{2p}$ to the total orbital Hamilton population that is plotted in subfigure (f). The energy-scale is shifted such that the Fermi level (ϵ_F) is zero. **Case study:** Carbon diamond $2 \times 2 \times 2$ supercell with periodic boundary conditions using a Γ-point.

integration via multiple **k**-point sampling is not explicitly considered here.

4.1. Mathematical Formulation. The projected population (both overlap and Hamilton populations) in this approach is computed by orthogonally projecting the Kohn−Sham discretized Hamiltonian operator $\mathcal{H}$ onto the atomic-orbital subspace $\mathbb{V}_\phi^{N_{orb}}$. The computed projected Hamiltonian $\mathcal{H}^\phi \equiv \mathcal{P}^\phi \mathcal{H} \mathcal{P}^\phi : \mathbb{V}_\phi^{N_{orb}} \to \mathbb{V}_\phi^{N_{orb}}$ is subsequently diagonalized to compute the orthonormal eigenvectors ($|\tilde{\psi}_i^E\rangle$) in the subspace $\mathbb{V}_\phi^{N_{orb}}$. These eigenvectors of the projected Hamiltonian, $|\tilde{\psi}_i^E\rangle \in \mathbb{V}_\phi^{N_{orb}}$, are thereby employed to compute both the overlap and Hamilton populations. The proposed approach pHA is in contrast with pOA discussed above (see section 3) wherein the discretized Kohn−Sham wave function $|\psi_i\rangle \in \mathbb{V}^M$ is orthogonally projected (L_2 projection) onto the subspace $\mathbb{V}_\phi^{N_{orb}}$ to compute the overlap and Hamilton populations. pHA is similar in spirit to the iterative orthogonal projection techniques[43] employed in the solution of large-scale matrix eigenvalue problems of the form $\mathcal{A}|x\rangle = \lambda|x\rangle$, wherein, one seeks an approximate eigenvector, eigenvalue pair ($|\tilde{x}\rangle$, $\tilde{\lambda}$) of $\mathcal{A}$ in a carefully constructed lower-dimensional subspace such that the residual vector $|r\rangle = \mathcal{A}|\tilde{x}\rangle - \tilde{\lambda}|\tilde{x}\rangle$ 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 $\mathcal{A}$ into the subspace), which approximates the eigenvalues and eigenvectors of the original matrix $\mathcal{A}$ better than any vectors lying in the subspace.

Projected Hamiltonian Hamilton Population (pHHP). We note that $|\tilde{\psi}_i^E\rangle$, the eigenvectors of $\mathcal{H}^\phi$, lie in the atomic-orbital subspace $\mathbb{V}_\phi^{N_{orb}}$, and hence, we express $|\tilde{\psi}_i^E\rangle$ as a linear combination of the orthonormalized atomic-orbital basis $\{|\hat{\phi}_\mu\rangle\}$, i.e., $|\tilde{\psi}_i^E\rangle = \sum_\mu \hat{E}_\mu^i |\hat{\phi}_\mu\rangle$. Subsequently, following the similar arguments used in deriving eq 11, we can define the Hamilton population $\text{pHHP}_{IJ}(\epsilon)$ associated with a source atom I and a target atom $J \neq I$ as

$$\text{pHHP}_{IJ}(\epsilon) = \sum_j \sum_{\alpha\beta} \mathfrak{R}e\{((\langle\tilde{\psi}_j^E|\hat{\phi}_{I\alpha}\rangle)H_{I\alpha,J\beta}^\phi(\langle\hat{\phi}_{J\beta}|\tilde{\psi}_j^E\rangle)\}\delta(\epsilon - \epsilon_j)$$

(16)

where $H_{I\alpha,J\beta}^\phi$ is the matrix element of $\mathcal{H}^\phi$ and $\mathfrak{R}e(z)$ refers to the real part of a complex number z. Introducing the finite-element discretization of the various fields in eq 16, we now deduce the relevant matrix expressions required to evaluate the Hamilton population in the FE setting. To begin, we use the definition of the projected Hamiltonian $\mathcal{H}^\phi$, whose matrix element $H_{I\alpha,J\beta}^\phi$ can be evaluated as $H_{I\alpha,J\beta}^\phi = \langle\hat{\phi}_{I\alpha}|\mathcal{H}^\phi|\hat{\phi}_{J\beta}\rangle$
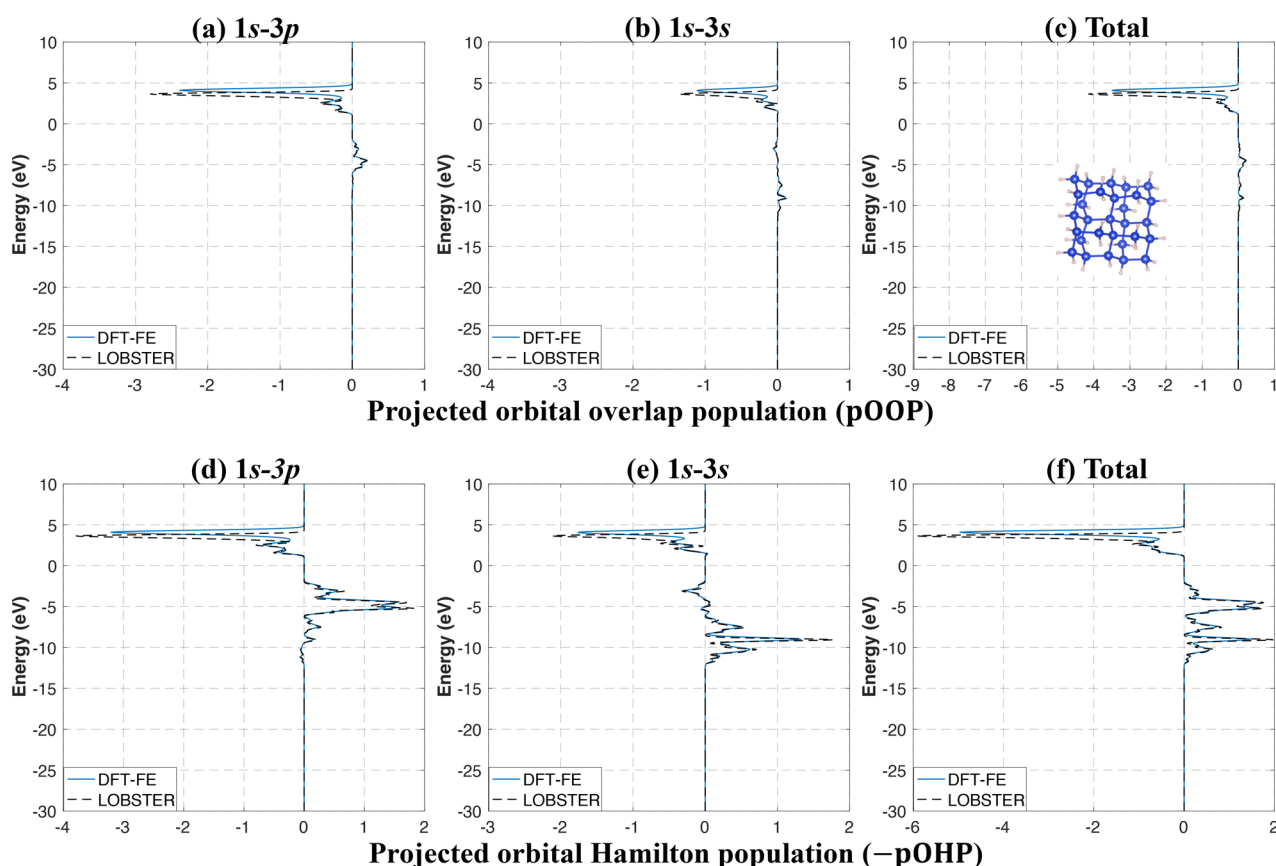
Figure 3. Comparison between pOA implemented in **DFT−FE** using PA orbitals and **LOBSTER** for nearest neighbor Si−H atoms in Si nanoparticles. The top row shows the projected orbital overlap population, with the subfigures (a) and (b) in this row showing the contributions of $H_{1s}-Si_{3p}$ and $H_{1s}-Si_{3s}$ to the total orbital overlap population that is plotted in subfigure (c). The bottom row shows the negative of the projected orbital Hamilton population. The subfigures in the bottom row (d) and (e) show the contributions of $H_{1s}-Si_{3p}$ and $H_{1s}-Si_{3s}$ to the total orbital Hamilton population that is plotted in subfigure (f). The energy-scale is shifted such that the Fermi level (ϵ_F) is zero. **Case study:** Single-fold $Si_{29}H_{36}$ nanoparticle.

Table 2. Comparison of Computation CPU Time[a] Measured in node-secs, between Projected Orbital Population Analysis (pOA) Implementation in DFT−FE and LOBSTER

Material system	pOA (sec)	LOBSTER (sec)
Carbon diamond 2 × 2 × 2 supercell	8.26	17.5
$Si_{29}H_{36}$ nanoparticle	1.72	24.2
$Si_{145}H_{150}$ nanoparticle	21.46	

[a]The timing includes the reading and construction of the atomic-orbital basis ($\mathbf{\Phi}$) and the calculation of the appropriate matrices for conducting the population analysis ($\overline{\mathbf{C}}$, $\hat{\mathbf{C}}$, $\mathbf{H}^p$).

Table 3. Wall Times of pOA with Increasing Number of MPI Tasks on PARAM PRAVEGA (1 Core per MPI Task)[a]

Number of MPI tasks	Wall time (seconds)
280	33.85
560	25.04
1120	20.48

[a]**Case study:** $Si_{580}H_{510}$ nanoparticle (1090 atoms with 2830 valence electrons). The total degrees of freedom (DoFs) per atom is around 8400.

and further can be expressed in the matrix form as $\mathbf{H}^\phi = \mathbf{S}^{-1/2}\mathbf{\Phi}^\dagger\mathbf{H}\mathbf{\Phi}\mathbf{S}^{-1/2}$ where $\mathbf{H}$ denotes the matrix corresponding to the finite-element discretized Kohn−Sham

Hamiltonian operator $\mathcal{H}$ introduced in section 3. Upon diagonalization of $\mathbf{H}^\phi$, we have $\mathbf{H}^\phi = \hat{\mathbf{E}}\overline{\mathbf{D}}\hat{\mathbf{E}}^\dagger$, where $\hat{\mathbf{E}}$ denotes the $N_{orb} \times N_{orb}$ eigenvector matrix. We note that the jth column of $\hat{\mathbf{E}}$ represents the coefficients of $|\tilde{\psi}_i^E\rangle$ with respect to the $\{|\hat{\phi}_\mu\rangle\}$ basis. Similar to section 3, we introduce a composite-index $\mu = \{I\alpha\}$ to denote the orthonormalized atomic orbital basis function $|\hat{\phi}_\mu\rangle$ as $|\hat{\phi}_{Ia}\rangle$, where α denotes the index of the atomic orbital centered at a nuclear position $\mathbf{R}_I$. Further, the μth row of $\hat{\mathbf{E}}$ corresponds to the Ith atom and an atom-centered index α associated with this atom I. We refer the reader to the Supporting Information (see S1.2) for details on the derivation and efficient computation of the matrix expressions related to $\mathbf{H}^\phi$ and $\hat{\mathbf{E}}$. Finally, the *projected Hamiltonian Hamilton population* (pHHP) associated with the partitioning of band energy between a source atom I and a target atom $J \neq I$ is evaluated by extracting the appropriate entries of the matrices $\hat{\mathbf{E}}$ and $\mathbf{H}^\phi$ and the expression is given by

$$\mathrm{pHHP}_{IJ}(\epsilon) = \sum_j \sum_{\alpha\beta} \mathfrak{Re}\{\hat{E}_{I\alpha}^{j*} H_{I\alpha,J\beta}^\phi \hat{E}_{J\beta}^j\}\delta(\epsilon - \epsilon_j) \tag{17}$$

Projected Hamiltonian Overlap Population (pHOP). The overlap population in the current approach is computed using the non-orthogonal localized atom-centered orbitals $\{|\phi_v\rangle\}$. To this end, we first compute the N_{orb} linear combination

coefficients of the expansion of $|\tilde{\psi}_i^E\rangle$ in terms of these basis functions $\{|\phi_\nu\rangle\}$, i.e., $|\tilde{\psi}_j^E\rangle = \sum_\mu \bar{E}_\mu^j|\phi_\mu\rangle$ for $j = 1, ..., N$. Following the similar arguments used in arriving at eq 5, we can define the overlap population $pHOP_{IJ}(\epsilon)$ between the source atom I and J as

$$pHOP_{IJ}(\epsilon) = \sum_j \sum_{\alpha\beta} \mathfrak{R}e\{(\langle\tilde{\psi}_j^E|\phi_{I\alpha}\rangle)S_{I\alpha,J\beta}(\langle\phi_{J\beta}|\tilde{\psi}_j^E\rangle)\}\delta(\epsilon - \epsilon_j)$$

(18)

where $\mathfrak{R}e(z)$ refers to the real part of a complex number z. Introducing finite-element discretization of the various fields in the above equation, we define the $N_{orb} \times N_{orb}$ matrix $\bar{E}$ which can be computed as $\bar{E} = S^{-1/2}\hat{E}$, where the matrix $\hat{E}$ is the eigenvector matrix of H^ϕ introduced previously. For the derivation and the computational cost associated with the computation of $\bar{E}$, we refer to Supporting Information section S1.2. Finally, the *projected Hamiltonian overlap population* (pHOP) associated with a source atom I and a target atom J is evaluated by extracting the appropriate entries of the matrices S and $\bar{E}$ and the expression is given by

$$pHOP_{IJ}(\epsilon) = \sum_j \sum_{\alpha\beta} \mathfrak{R}e\{\bar{E}_{I\alpha}^{j*}\bar{E}_{J\beta}^j S_{I\alpha J\beta}\}\delta(\epsilon - \epsilon_j)$$

(19)

Total Computational Complexity Estimate of pHA. The current implementation of the projected Hamiltonian population analysis (pHA) assumes $N = N_{orb}$, and thereby, the total computational complexity is estimated to be $\sim 4M_{loc}N^2 + 16N^3$. Since $M_{loc} = M/P$, 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 more efficient than the pOA approach described in the previous subsection.

Projected Hamiltonian Density Error (pHDE). To understand the loss of information due to the projection of the finite-element discretized Hamiltonian onto $\mathbb{V}_\phi^{N_{orb}}$, we introduce the projected density error (pHDE). Here, we compute the L_2 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 $\{|\tilde{\psi}_j^E\rangle\} \in \mathbb{V}_\phi^{N_{orb}}$ associated with the projected Hamiltonian H^ϕ. To this end, pHDE is evaluated as

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

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

(20)

4.2. Results: Accuracy and Performance Benchmarking.

We assess here the performance and accuracy of the proposed pHA 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 spanned by pseudoatomic (PA) orbitals and conduct the population analysis as discussed in subsection 4.1. We discuss here a comparative study of the population analysis conducted using this approach and pOA reported in section 3.
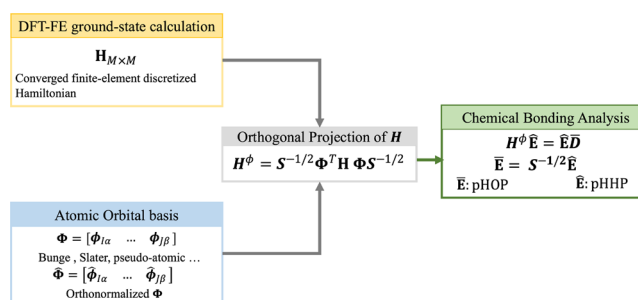

Figure 4. Overview of the implementation strategy for projected Hamiltonian population analysis (pHA) within the finite-element (FE) framework. The employed strategy involves projecting the self-consistently converged FE discretized Hamiltonian obtained from **DFT−FE** onto the atomic-orbital basis and subsequently diagonalizing this projected Hamiltonian H^ϕ to compute the eigenvector matrix $\hat{E}$. The overlap matrix (S) and the coefficient matrices $\bar{E}$ are finally evaluated, which in turn are used to compute the projected Hamiltonian overlap population pHOP and projected Hamiltonian Hamilton population pHHP.

Accuracy Validation of pHA. To begin with, we plot the population energy diagrams corresponding to pHHP and pHOP derived in eqs 16 and 19 and compare with pOHP and pOOP on the same benchmark systems comprised of isolated systems and a periodic system as discussed in the previous section. As mentioned previously, we will refer to the approach of projected orbital population (pOOP and pOHP) as pOA and the approach of projected Hamiltonian population (pHOP and pHHP) as pHA. Figure 5 illustrates the comparison in the case of a periodic $2 \times 2 \times 2$ carbon diamond supercell for two nearest carbon atoms picked as source and target atoms (see the inset in Figure 5). Figure 6 demonstrates the comparisons in the case of a $Si_{29}H_{36}$ nanoparticle, an isolated system in which the Si atom and the nearest H atom are picked as source and target atoms (see the 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 populations conducted using both pOA and pHA. Comparisons between both methods for benchmark systems involving molecules—CO, spin polarized O_2, and H_2O—are discussed in Supporting Information section S2.1, and we observe a very close agreement.

We now compare the density error metrics (pODE and pHDE), a measure of the loss of information during projections, as introduced in section 3.1 and section 4.1. Recall from eq 15 and eq 20, these metrics measure the error between the self-consistently converged ground-state electron density computed from **DFT−FE** and the electron density computed using the projected wave functions in the subspace $\mathbb{V}_\phi^{N_{orb}}$. As shown in Table 4, we see a close agreement between pHDE and pODE for a variety of benchmark material systems, which include isolated systems (CO, H_2O, spin-polarized O_2, $Si_{29}H_{36}$, $Si_{58}H_{66}$, and $Si_{145}H_{150}$) and a periodic system (carbon $2 \times 2 \times 2$ supercell).

The comparative study discussed so far demonstrates the excellent match of numerical results obtained between the approaches pOA and pHA. As remarked before, pOA is similar in spirit to the projected population analysis approach implemented in **LOBSTER**, but the pHA proposed in this work is different in spirit than the pOA approach and relies on
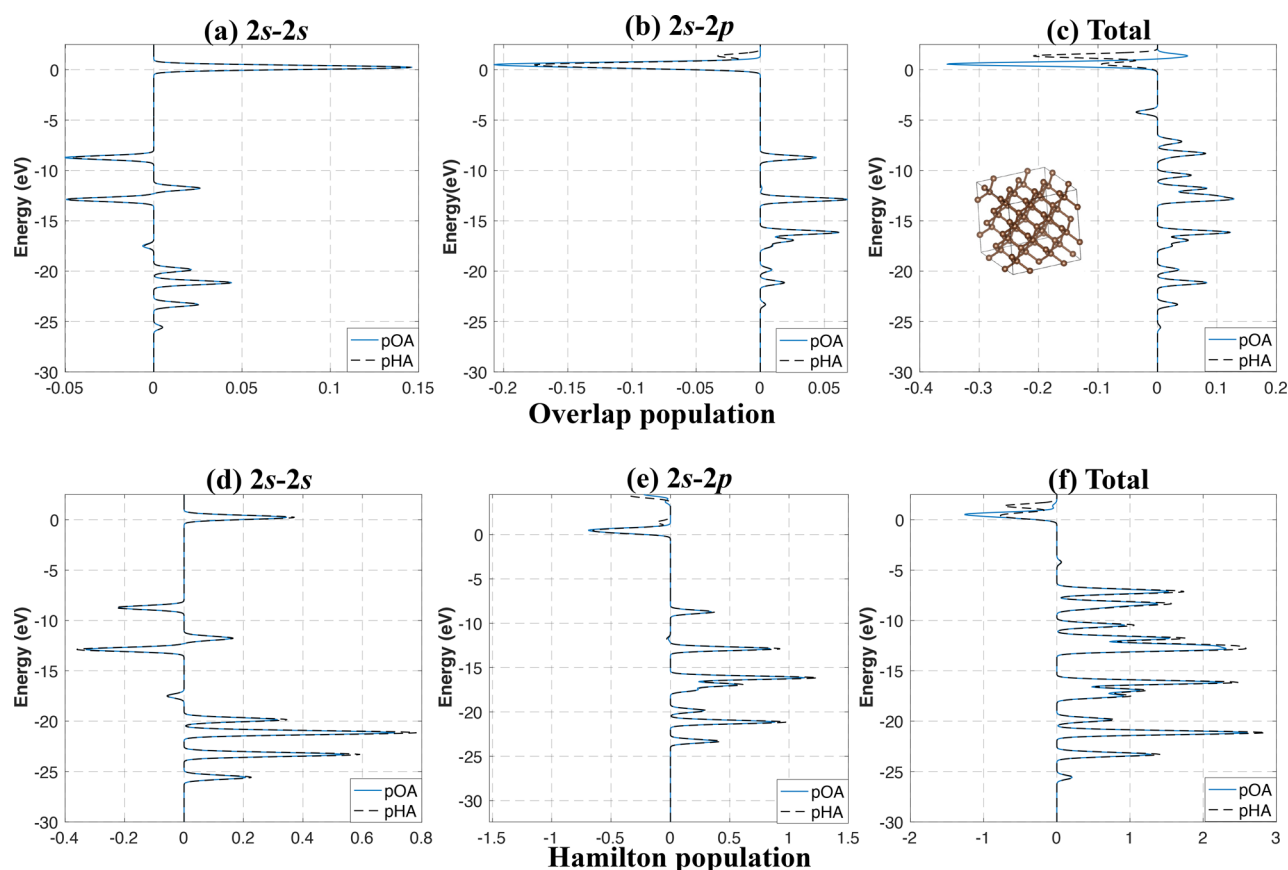
Figure 5. Comparison of the overlap and Hamilton populations between the two proposed methods of projected population analysis (pOA and pHA) for nearest C atoms in a carbon diamond supercell. The top row shows the overlap population obtained using both of these methods. The subfigures (a) and (b) in this row show the contributions of $C_{2s}-C_{2s}$ and $C_{2s}-C_{2p}$ to the total overlap population that is plotted in subfigure (c). The bottom row shows the negative of the Hamilton population for both methods. The subfigures in this bottom row (d) and (e) show the contributions of $C_{2s}-C_{2s}$ and $C_{2s}-C_{2p}$ to the total Hamilton population that is plotted in subfigure (f). The energy-scale is shifted such that the Fermi level (ϵ_F) is zero. **Case study:** $2 \times 2 \times 2$ carbon diamond supercell with periodic boundary conditions at the Γ-point for Brillouin zone sampling.

the projection of a finite-element discretized Kohn–Sham Hamiltonian to conduct population analysis.

Performance Comparisons. We now demonstrate the computational advantage of conducting population analysis using pHA for large-scale systems. Parts a and b of Figure 7 show the computational wall times measuring the projected population analysis with the increasing number of MPI tasks by comparing the two methods pOA and pHA. For this study, we consider two nanoparticles $Si_{580}H_{510}$ and $Si_{1160}H_{990}$ comprised of 1090 and 2150 atoms, respectively. The results indicate a speed-up of 1.3−1.4× for pHA with the increase in MPI tasks beyond a certain number. These speed-ups are consistent with the computational complexity estimates derived in previous sections, i.e., the $O(N^3)$ cost becoming dominant beyond a certain number of MPI tasks with a lower prefactor for pHA.

5. BONDING INSIGHTS IN LARGE SYSTEMS: CHEMISORPTION IN Si NANOPARTICLES

In this section, we demonstrate the advantage of the proposed computational framework for conducting population analysis to extract chemical bonding in large-scale material 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,[26] where the storage (release) of hydrogen is a result of the formation (breaking) of a Si−H bond. As discussed in Williamson et al.,[26] the release of hydrogen in these Si nanoparticles occurs due to the dimerization of dihedral SiH_2 cofacial pairs in a $Si_{29}H_{36}$ unit to reduce to a $Si_{29}H_{24}$ unit by the 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 this case study, we attempt to provide a chemical bonding viewpoint by conducting population analysis to extract bonding information in these Si nanoparticles. We compute a quantity known as the integrated projected orbital Hamilton population $IPOHP = \int_{\epsilon \leq \epsilon_F} -pOHP(\epsilon) \, d\epsilon$. A higher value of

IPOHP correlates with a stronger covalent bonding interaction between the source-target atoms, and this quantity is similar to integrated pCOHP (ICOHP) computed in **LOBSTER**. To this end, we argue the ease of dimerization of dihedral SiH_2 cofacial pairs in the Si nanoparticle by computing IPOHP 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 IPOHP values 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,
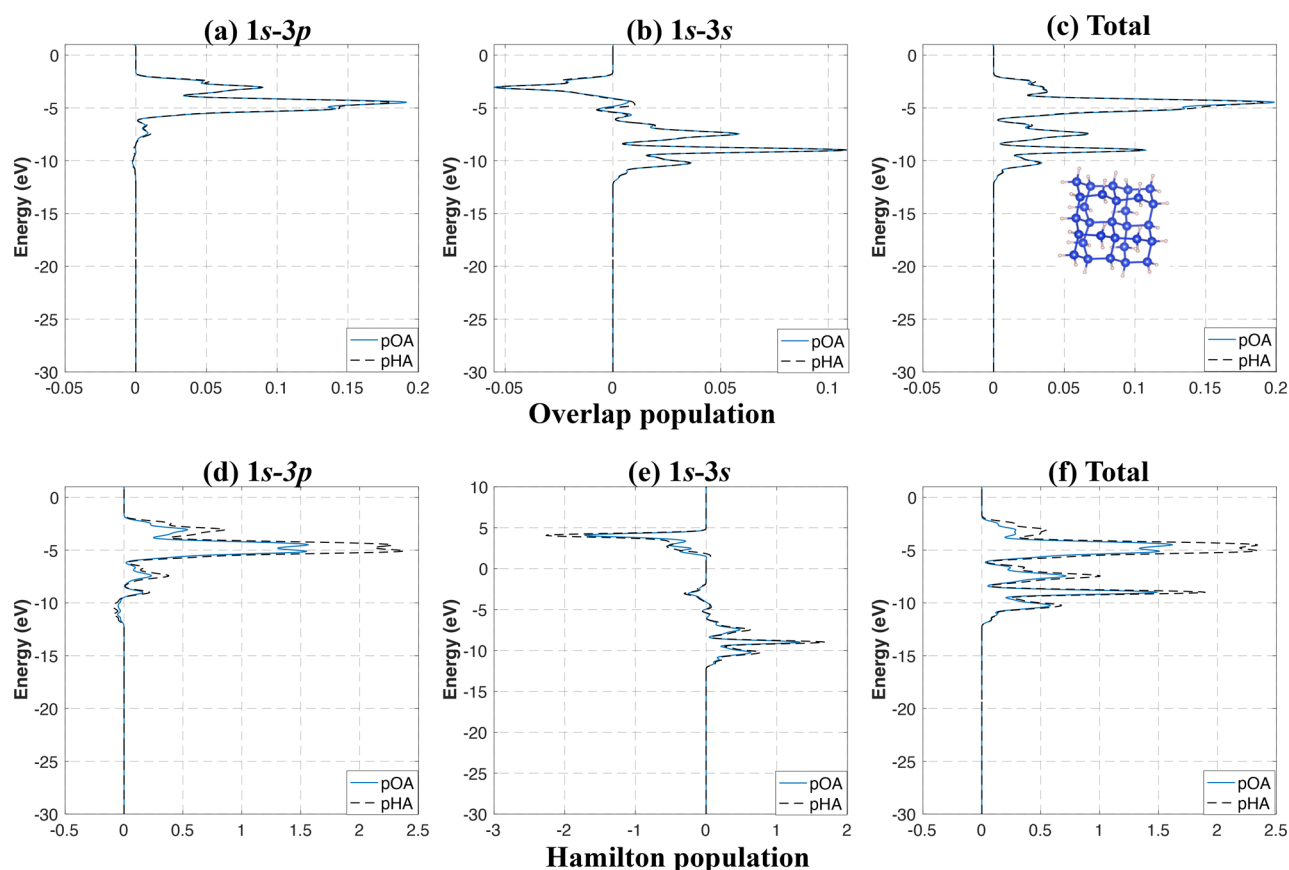
Figure 6. Comparison of the overlap and Hamilton populations between the two proposed methods of projected population analysis (pOA and pHA) for nearest Si−H atoms in a Si nanoparticle. The top row shows the overlap population obtained using both of these methods. The subfigures (a) and (b) in this row show the contributions of $H_{1s}−Si_{3p}$ and $H_{1s}−Si_{3s}$ to the total overlap population that is plotted in subfigure (c). The bottom row shows the negative of the Hamilton population for both methods. The subfigures in the bottom row (d) and (e) show the contributions of $H_{1s}−Si_{3p}$ and $H_{1s}−Si_{3s}$ to the total Hamilton population that is plotted in subfigure (f). The energy-scale is shifted such that the Fermi level (ϵ_F) is zero. **Case study:** single-fold $Si_{29}H_{36}$ nanoparticle with nonperiodic boundary conditions.

Table 4. Comparison of Density Errors (pODE and pHDE) Computed from the Projected Orbital Population (pOA) and Projected Hamiltonian Population Approach (pHA), Respectively

Material system	pODE	pHDE
Carbon diamond $2 \times 2 \times 2$ supercell	0.061	0.062
CO	0.118	0.119
H_2O	0.136	0.121
$O_2 \uparrow$ (Spin-up density)	0.097	0.098
$O_2 \downarrow$ (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

and 20-fold structures of dihedral Si_{29} and $Si_{24}C_5$ units. Toward this, we build the 2-fold and further the 5-fold structure by connecting the $Si_{29}H_{36}$ ($Si_{24}C_5H_{36}$) units by their (111) facets, as discussed in Williamson et al.[26]

The atomic configurations corresponding to various sizes of Si_{29} and $Si_{24}C_5$ nanoparticles are obtained by performing geometry optimization in **DFT−FE** code until the maximum atomic force in each direction reaches a tolerance of approximately $5 \times 10^{-4} \frac{E_h}{bohr}$. 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 computed IPOHP by conducting pOHP 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 antibonding 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 IPOHP values corresponding to the Si−Si interaction between the nearest SiH_2 dihedral pairs sharing a core silicon/carbon atom. The mean of the IPOHP values corresponding to the weakest of the Si−H interaction in each of the dihedral units of these pairs has also been tabulated in this table (Si−H1 and Si−H2). An 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 pOHP, the IPOHP values reported in Table 5 show a higher value ($\approx 4\times$ higher) for nanoparticles alloyed with carbon in comparison to no alloying. This can be attributed to the carbon core in the alloyed nanoparticles drawing the Si atoms in the nearest SiH_2 dihedral units toward each other, thereby leading to a stronger Si−Si interaction. It is interesting to note the increase in Si−Si IPOHP values with an increase in the curvature of the alloyed nanoparticles, indicating the strengthening of the Si−Si
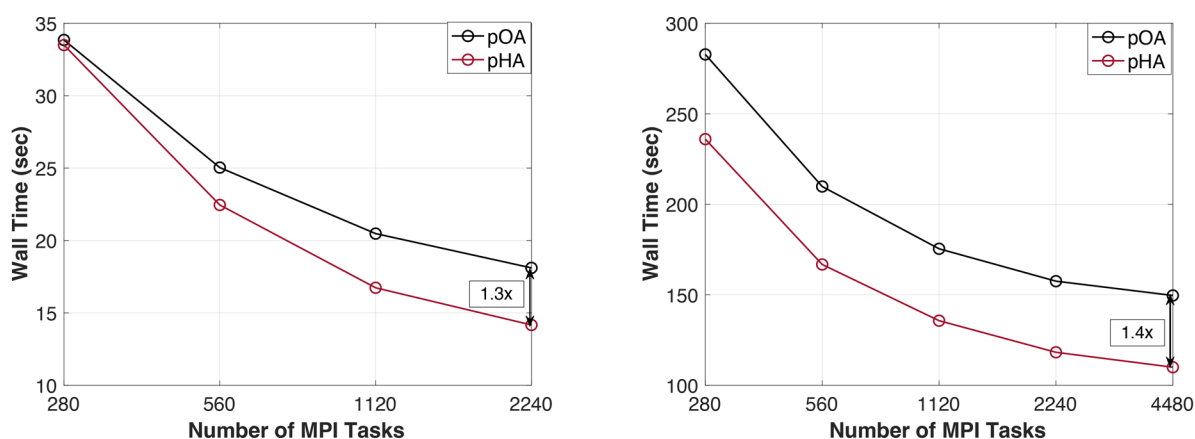
(a) **Case study:** $Si_{580}H_{510}$ (1090 atoms, 2830 wave-functions)

(b) **Case study:** $Si_{1160}H_{990}$ (2150 atoms, 5630 wave-functions)

Figure 7. Wall-time comparison of pOA with pHA on the PARAM Pravega supercomputer (1 MPI task per core). The case studies considered are 20-fold and 40-fold Si nanoparticles. Total DoFs (M) are 9,130,679 and 17,254,979 for $Si_{580}H_{510}$ (20-fold) and $Si_{1160}H_{990}$ (40-fold), respectively.

(a) $Si_{29}H_{36}$ nanocluster (b) $Si_{24}C_5H_{36}$ nanocluster

(c) $Si_{58}H_{66}$ nanocluster (d) $Si_{48}C_{10}H_{66}$ nanocluster

(e) $Si_{145}H_{150}$ nanocluster (f) $Si_{120}C_{25}H_{150}$ nanocluster

(g$_1$) $Si_{580}H_{510}$ nanocluster (h$_1$) $Si_{480}C_{100}H_{510}$ nanocluster

(g) $Si_{580}H_{510}$ nanocluster (h) $Si_{480}C_{100}H_{510}$ nanocluster
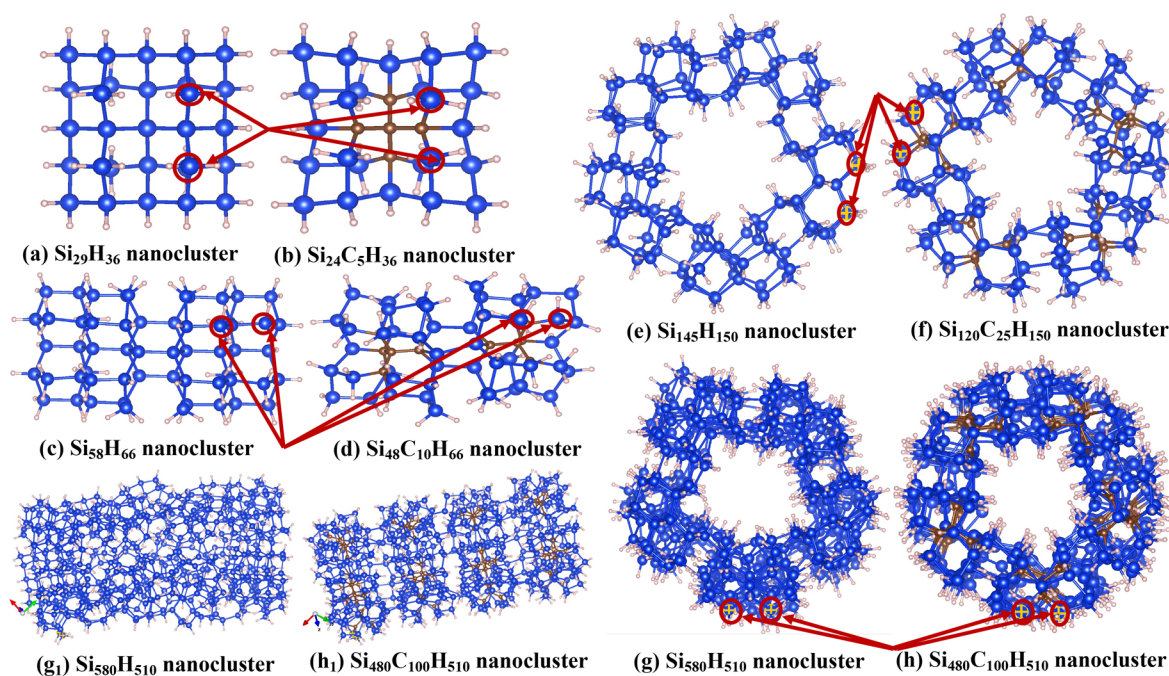
Figure 8. Relaxed atomic configurations of (a, b) 1-Fold, (c, d) 2-fold, (e, f) 5-fold, and (g, h) 20-fold nanoparticle structures. The mean of the IPOHP values computed in Table 5 corresponds to the Si−Si interaction between nearest SiH_2 dihedral pairs sharing a core silicon/carbon atom. The silicon atoms highlighted in red illustrate one such pair considered. The structures were relaxed in DFT until the maximum force component on any atom reached a tolerance level of $5 \times 10^{-4} \frac{E_h}{bohr}$.

interaction. This can possibly explain the ease of dimerization of dihedral SiH_2 cofacial pairs as the size increases from 1-fold to 5-fold structures. Similarly, due to the reduced separation between the dihedral SiH_2 groups, the corresponding IPOHP values suggest a weakening of the Si−H bond as seen in Table 5 for alloyed nanoparticles, a favorable condition for the release of H_2.

Maximizing the hydrogen storage capacity when designing hydrogen storage devices is important; hence, large nanoparticles are desirable. As described in Williamson et al.,[26] one such structure is obtained by stacking four 5-fold Si nanoparticles resulting in a 20-fold nanoparticle containing 1090 atoms. IPOHP values for this large nanoparticle with and

without C alloying were computed by conducting population analysis on the relaxed geometries. We observe the mean IPOHP values of Si−Si interaction between the nearest SiH_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 a 5-fold nanoparticle. This is consistent with the fact that stacking does not significantly impact the curvature of the large nanoparticle compared to the 5-fold nanoparticle.

6. CONCLUSIONS

In the present work, we formulate and implement two methods for conducting projected population analysis (both overlap and
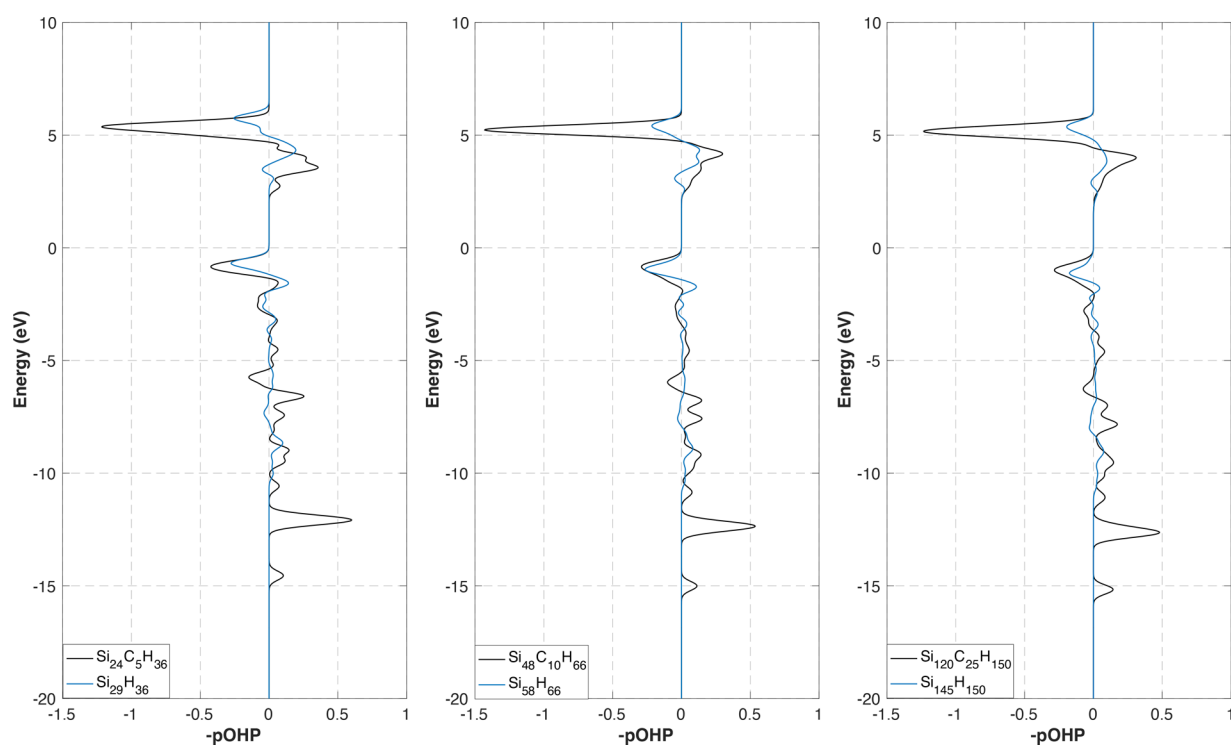
Figure 9. Comparison of −pOHP between nearest nonbonded neighbors of Si−Si in 1-Fold, 2-Fold, and 5-Fold nanoparticles.

Table 5. Comparison of Mean IPOHP for Various SiH$_2$ Dihedral Pairs in 1-Fold, 2-Fold, and 5-Fold Nanoparticles of Si$_{29}$ and Si$_{24}$C$_5$ Units[a]

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

[a]The Si−Si interaction is computed between nearest SiH$_2$ dihedral pairs. Si−H1 and Si−H2 denote the weakest of the Si−H interactions in each of the dihedral units of these pairs.

Hamilton populations) to extract chemical bonding information from finite-element (FE)-based density functional theory (DFT) calculations. The first method (pOA) relies on the orthogonal projection of FE discretized Kohn−Sham DFT eigenfunctions onto a subspace spanned by localized atom-centered basis functions. In contrast, the second method (pHA) relies on the orthogonal projection of FE discretized Kohn−Sham Hamiltonian onto this subspace. These methods are implemented within the **DFT−FE** code[22,23] and take advantage of **DFT−FE**'s capability to conduct fast, scalable, and systematically convergent large-scale DFT calculations, enabling large-scale bonding analysis on complex material systems not accessible before, without any restriction on the boundary conditions employed.

First, we present the mathematical formulation and efficient finite-element strategies adopted to compute both the overlap and Hamilton population within the *projected orbital population analysis* (pOA) framework. Following this, we assess the accuracy of the proposed method on representative material systems comprised of isolated molecules, nanoparticles, and a

periodic system with a large supercell. In all 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 pOA over **LOBSTER** is also clearly illustrated on a few of these benchmark examples. Subsequently, we discuss an alternate approach for projected population analysis that does not rely on the availability of converged Kohn−Sham DFT eigenfunctions. This approach is motivated by the fact that many of the reduced scaling electronic structure codes targeted toward large-scale DFT calculations tend to avoid explicit computation of DFT eigenfunctions with no access to these vectors for projection. This alternate method is referred to as *projected Hamiltonian population analysis* (pHA). The accuracy and performance benchmarks of pHA with pOA show similar trends in bonding behavior with improved scalability for large-scale systems. 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. This analysis demonstrates a correlation of Si−Si and Si−H bonding interactions with the nanoparticle size and argues the ease of Si−Si dimerization with the increase in the size of the Si−C alloy nanocluster favoring the release of H$_2$.

In summary, the proposed projected population analysis methods within the framework of finite-element discretization of DFT open the possibility of extracting chemical bonding information in large material systems critical to many technologically relevant applications. Our work demonstrates one such case by conducting a large-scale chemical bonding analysis in Si nanoparticles, a candidate material for hydrogen storage. Further, such analysis can also reveal bonding interactions between complex defects (e.g., dislocations, grain boundaries, etc.) and solute impurities, offering atomistic

insights into the stability of these defects, which has implications for understanding the strength and ductility of structural materials. Another area of application is solid-state battery material design, where such population analysis can aid in understanding ionic conductivity by revealing bonding interactions between migrating ions and the underlying solid electrolyte lattice in an electric field. Furthermore, using a single computational framework for both ground-state DFT calculations and population analysis allows for on-the-fly bonding analysis in *ab initio* molecular dynamics simulations, yielding bonding interaction insights as a function of time. These are just a few examples among numerous possibilities the proposed methods can provide access to, offering a robust means for extracting chemical bonding information in various complex scenarios.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.3c00114.

> Efficient finite-element implementation strategies, projected orbital population analyses, projected Hamiltonian population analyses, and results from additional benchmarking studies (PDF)

AUTHOR INFORMATION

Corresponding Authors

Phani Motamarri − *Department of Computational and Data Sciences, Indian Institute of Science, Bangalore 560012, India;* ⓘ orcid.org/0000-0003-3160-7334; Email: phanim@iisc.ac.in

Satadeep Bhattacharjee − *Indo Korea Science and Technology Center, Bangalore 560065, India;* ⓘ orcid.org/0000-0002-6717-2881; Email: s.bhattacharjee@ikst.res.in

Authors

Kartick Ramakrishnan − *Department of Computational and Data Sciences, Indian Institute of Science, Bangalore 560012, India;* ⓘ orcid.org/0000-0003-3155-157X

Sai Krishna Kishore Nori − *Department of Computational and Data Sciences, Indian Institute of Science, Bangalore 560012, India*

Seung-Cheol Lee − *Indo Korea Science and Technology Center, Bangalore 560065, India*

Gour P. Das − *Research Institute for Sustainable Energy (RISE), TCG Center for Research and Education in Science and Technology, Kolkata 700091, India*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.3c00114

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

ACRONYMS

pHA   projected Hamiltonian population analysis
pHHP   projected Hamiltonian Hamilton population
pHOP   projected Hamiltonian overlap population
pOA   projected orbital population analysis
pOHP   projected orbital Hamilton population
pOOP   projected orbital overlap population
IPOHP   integrated projected orbital Hamilton population
PA   pseudoatomic
pCOHP   projected crystal orbital Hamilton population
pCOOP   projected crystal orbital overlap population
pHDE   projected Hamiltonian density error
pODE   projected orbital density error

REFERENCES

(1) Mulliken, R. S. Electronic population analysis on LCAO-MO molecular wave functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.

(2) Autschbach, J. Orbitals: some fiction and some facts. *J. Chem. Educ.* **2012**, *89*, 1032−1040.

(3) Glassey, W. V.; Hoffmann, R. A comparative study of Hamilton and overlap population methods for the analysis of chemical bonding. *J. Chem. Phys.* **2000**, *113*, 1698−1704.

(4) Hughbanks, T.; Hoffmann, R. Chains of trans-edge-sharing molybdenum octahedra: metal-metal bonding in extended systems. *J. Am. Chem. Soc.* **1983**, *105*, 3528−3537.

(5) Dronskowski, R.; Blöchl, P. E. Crystal orbital Hamilton populations (COHP): energy-resolved visualization of chemical bonding in solids based on density-functional calculations. *J. Phys. Chem.* **1993**, *97*, 8617−8624.

(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, Germany, 2005.

(8) Eck, B.; Dronskowski, R.; Takahashi, M.; Kikkawa, S. Theoretical calculations on the structures, electronic and magnetic properties of binary 3d transition metal nitrides. *J. Mater. Chem.* **1999**, *9*, 1527−1537.

(9) 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. *J. Phys. Chem. B* **2002**, *106*, 12727−12736.

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

(11) Grechnev, A.; Ahuja, R.; Eriksson, O. Balanced crystal orbital overlap populationa tool for analysing chemical bonds in solids. *J. Phys.: Condens. Matter* **2003**, *15*, 7751.

(12) Müller, P. C.; Ertural, C.; Hempelmann, J.; Dronskowski, R. Crystal orbital bond index: covalent bond orders in solids. *J. Phys. Chem. C* **2021**, *125*, 7959−7970.

(13) Marzari, N.; Mostofi, A. A.; Yates, J. R.; Souza, I.; Vanderbilt, D. Maximally localized Wannier functions: Theory and applications. *Rev. Mod. Phys.* **2012**, *84*, 1419.

(14) Jonsson, E. Ö.; Lehtola, S.; Puska, M.; Jonsson, H. Theory and Applications of Generalized Pipek-Mezey Wannier Functions. *J. Chem. Theory Comput.* **2017**, *13*, 460−474.

(15) Deringer, V. L.; Tchougréeff, A. L.; Dronskowski, R. Crystal orbital Hamilton population (COHP) analysis as projected from plane-wave basis sets. *J. Phys. Chem. A* **2011**, *115*, 5461−5466.

(16) 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. *J. Comput. Chem.* **2013**, *34*, 2557−2567.

(17) 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. *J. Comput. Chem.* **2020**, *41*, 1931−1940.

(18) Kundu, S.; Bhattacharjee, S.; Lee, S.-C.; Jain, M. Population analysis with Wannier orbitals. *J. Chem. Phys.* **2021**, *154*, 104111.

(19) Hafner, J. Ab-initio simulations of materials using VASP: Density-functional theory and beyond. *J. Comput. Chem.* **2008**, *29*, 2044−2078.

(20) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Corso, A. D.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter* **2009**, *21*, 395502.

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

(22) 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. *Comput. Phys. Commun.* **2020**, *246*, 106853.

(23) 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. *Comput. Phys. Commun.* **2022**, *280*, 108473.

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

(25) 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.

(26) Williamson, A. J.; Reboredo, F. A.; Galli, G. Chemisorption on semiconductor nanocomposites: A mechanism for hydrogen storage. *Appl. Phys. Lett.* **2004**, *85*, 2917−2919.

(27) Pask, J. E.; Sterne, P. A. Finite element methods in ab initio electronic structure calculations. *Modell. Simul. Mater. Sci. Eng.* **2005**, *13*, R71−R96.

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

(29) 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.

(30) 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 nano-structures. *Phys. Status Solidi B* **2006**, *243*, 1063−1079.

(31) Ghosh, S.; Suryanarayana, P. SPARC: Accurate and efficient finite-difference formulation and parallel implementation of Density Functional Theory: Isolated clusters. *Comput. Phys. Commun.* **2017**, *212*, 189−204.

(32) 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.

(33) 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.

(34) Hamann, D. R. Optimized norm-conserving Vanderbilt pseudopotentials. *Phys. Rev. B* **2013**, *88*, 085117.

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

(36) Löwdin, P. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365−375.

(37) Maintz, S.; Deringer, V. L.; Tchougréeff, A. L.; Dronskowski, R. LOBSTER: A tool to extract chemical bonding from plane-wave based DFT. *J. Comput. Chem.* **2016**, *37*, 1030−1035.

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

(39) 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.

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

(41) 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. *Comput. Phys. Commun.* **2018**, *226*, 39−54.

(42) Lehtola, S.; Steigemann, C.; Oliveira, M. J.; Marques, M. A. Recent developments in libxcA comprehensive library of functionals for density functional theory. *SoftwareX* **2018**, *7*, 1−5.

(43) 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, PA, 2000.