RESEARCH ARTICLE | FEBRUARY 13 2026

Practical integration of machine learning into *ab initio* calculations and workflows: Accelerating the SCF cycle via density matrix predictions

Pavel Stishenko ; Chen Qian ; Julia Westermayr ; Reinhard J. Maurer ; Andrew J. Logsdail

J. Chem. Phys. 164, 064123 (2026)
https://doi.org/10.1063/5.0302234

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Practical integration of machine learning into *ab initio* calculations and workflows: Accelerating the SCF cycle via density matrix predictions

Cite as: J. Chem. Phys. 164, 064123 (2026); doi: 10.1063/5.0302234
Submitted: 13 September 2025 • Accepted: 18 January 2026 •
Published Online: 13 February 2026

Pavel Stishenko,[1,a)] Chen Qian,[2] Julia Westermayr,[3,4] Reinhard J. Maurer,[2,5,a)] and Andrew J. Logsdail[1,a)]

AFFILIATIONS

[1] Cardiff Catalysis Institute, School of Chemistry, Cardiff University, Park Place, Cardiff CF10 3AT, United Kingdom
[2] Department of Chemistry, University of Warwick, Coventry CV4 7AL, United Kingdom
[3] Wilhelm-Ostwald-Institut für Physikalische und Theoretische Chemie, Universität Leipzig, Linnéstraße 2, 04103 Leipzig, Germany
[4] Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden/Leipzig, Humboldtstraße 25, 04105 Leipzig, Germany
[5] Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom

[a)] **Authors to whom correspondence should be addressed:** pstishenko@gmail.com; R.Maurer@warwick.ac.uk; and LogsdailA@cardiff.ac.uk

ABSTRACT

Data-driven approaches offer great potential for accelerating *ab initio* electronic structure calculations of molecules and materials, but their transferability is often limited due to the vast amount of data needed for training, including the need to fine-tune universal models for each specific system to be studied. Here, we demonstrate how contributions from system-specific electronic structure machine learning (ESML) models may be combined ("stitched") to deliver density matrices of entire systems of interest, improving the initial guess for the self-consistent field cycle and delivering gains in computational efficiency. The "stitching" of density matrices is demonstrated for sequential calculations, such as geometry optimization and molecular dynamics, and we show that the synergistic use of ESML models and density matrix extrapolation algorithms can accelerate standard computational calculations. The algorithms are demonstrated for test cases relating to water clusters and a methane clathrate cage, with the benefits discussed. The future opportunities for hybrid quantum mechanical and ML (QM/ML), and also ML/ML paradigms, are broad-ranging with significant computational speed-up attainable.

I. INTRODUCTION

The central algorithm of most electronic structure software applications is the self-consistent solution of the Roothaan–Hall equations,[1,2]

$$\mathbf{H}(\mathbf{C},\epsilon)\mathbf{C} = \mathbf{S}\mathbf{C}\epsilon, \tag{1}$$

where $\mathbf{H}$ is a matrix that provides a finite basis representation of the Hamiltonian operator, $\mathbf{C}$ is a matrix of wavefunction coefficients, ϵ is a vector of eigenvalues of the corresponding wavefunctions, and $\mathbf{S}$ is the basis overlap matrix, which can be the identity matrix in the case of an orthogonal basis. The equation resembles a generalized eigenproblem, but there is a non-linear functional dependence of the Hamiltonian matrix $\mathbf{H}$ on eigenvectors $\mathbf{C}$ and eigenvalues ϵ. Therefore, Eq. (1) has to be solved with a self-consistent field (SCF) method starting with some initial guess of the matrix $\mathbf{H}$ or $\mathbf{C}$ (Fig. 1).

Generally, the SCF cycle will converge more quickly when the initial guess of $\mathbf{H}$ is closer to the converged solution (noting that the

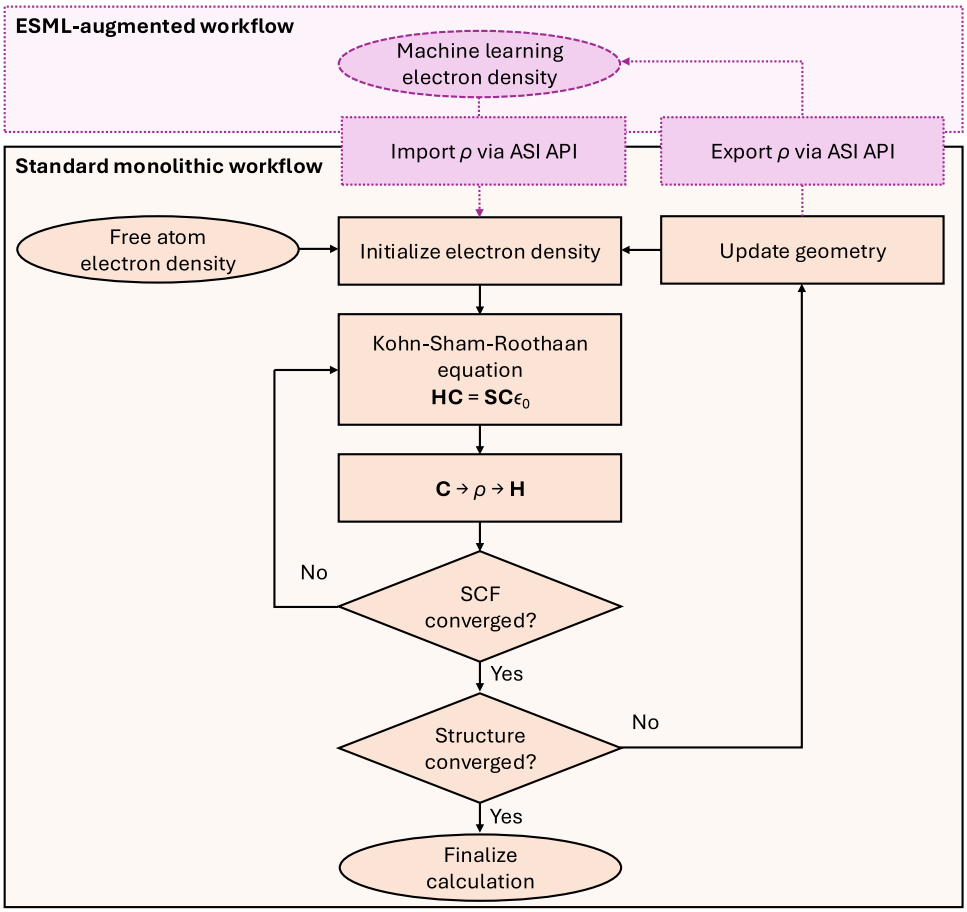

FIG. 1. A schematic of the algorithmic flow of the SCF cycle in a standard electronic structure calculation, embedded within the self-consistent geometry optimization cycle, and the presented approach to augment the workflow with ML predictions for ρ. **H**, **C**, ϵ, and **S** are as defined in the text; ρ is a matrix representation of the electronic density. Orange boxes, black lines: standard workflow in established monolithic electronic structure software packages. Purple boxes, purple lines: electronic structure machine learning (ESML)-augmented workflow aided by `ASI` API for interoperability between the electronic structure calculation and the external model.

choice of parameters for the SCF cycle is crucial for actual efficiency gains). For density functional theory (DFT), which is a workhorse of modern electronic structure modeling, the SCF cycle is commonly initialized with an educated guess of **H** that is produced from the superposition of the electron density of the individual free atoms of the simulated system (Fig. 1).[3] Obviously, the ground state electron density of any molecule or material differs from the electron density of the constituent atoms in vacuum, but this does usually provide a good starting point for the SCF cycle when compared to a random guess.

Recognizing the limitations of the superposition of free atom densities (SAD), many electronic structure software packages provide users with a method to customize the initial guess for the electronic structure of a system by loading data from files, which may originate from some prior calculations (e.g., from previous steps of dynamics simulation or from another level of theory). Such functionalities include loading the Kohn–Sham wavefunction coefficients (e.g., `Quantum Espresso`[4] or `VASP`[5]), the density matrix (e.g., `SIESTA`[6] or `FHI-aims`,[7] via `ELSI`[8]), or the charge density (also supported by `SIESTA`[6] and `VASP`,[5] as well as others).

In the era of data workflows, in-memory data transfer is increasingly desirable, and new approaches such as the Atomic Simulation Interface (`ASI`), as implemented in `FHI-aims` and `DFTB+`,[9,10] can provide elevated computational efficiency compared to traditional file input/output (I/O) paradigms. `ASI` is an application programming interface (API) that includes functions for efficient import and export of electronic structure descriptors, such as Hamiltonians and density matrices; the provision of a deep and overhead-free interface allows the exchange of core electronic structure data between (extensive) monolithic codes and external packages, providing new opportunities for the application of electronic structure machine learning (ESML) models, which can provide predictions of electronic structure or related properties for any given input system. A Python language wrapper for `ASI` (`asi4py`), available from the Python Package Index (PyPI) repository, enables exploration of robust and user-friendly approaches for ML-based SCF acceleration. Figure 1 demonstrates a workflow for the `ASI` paradigm working in conjunction with an ESML model, enabling accelerated SCF convergence through an improved initial guess of **H** that leads to a reduction in the number of SCF cycles.

A. Improving the SCF initial guess via extrapolation

The popularity of *ab initio* electronic structure calculations, specifically DFT, means that there has been a persistent effort to improve algorithms for SCF initialization, including representations of the density and/or orbital coefficients. A set of SCF initialization methods were benchmarked by Lethola[11] that included superposition of free atom densities (SAD), superposition of free atom potentials (SAP), and extended Hückel methods. More recently, Ballesteros and Lao have proposed to use many-body expansion to yield density matrices for SCF initialization[12] and tested this approach against the adjustable density matrix assembler (ADMA) method extensively studied earlier by Mezey *et al.*[13,14] When SCF calculations are converged sequentially for geometrically related systems, for example geometry optimization or molecular dynamics simulations, the converged electronic structures of previous geometries can be used to improve the initial guess for the next SCF calculation. The most widely adopted approach is simply using the electronic structure from the immediately previous SCF calculation to initialize the next, although more elaborate methods exist, such as Fock matrix dynamics,[15] extended Lagrangian Born–Oppenheimer (XLBO) dynamics,[16] Grassmann extrapolation of density matrices,[17,18] and interpolation of density matrices.[19,20]

The accuracy of the electronic structure produced by these methods depends on many factors, including the size and composition of the system, sampling density (e.g., time step for extrapolation along dynamics trajectory[17] or geometric density for interpolating cases[20]), and also the complexity of the approximation model.[15,17] However, the actual performance gain depends also on the quantitative aim of the calculation (i.e., relaxed geometry, trajectory, and vibrations), as well as the target accuracy for each standalone SCF cycle. The advanced extrapolation techniques noted have a similarity with *active learning* procedures in ML, where sequential structures are used (typically) to re-train the ML model through the course of a simulation "on-the-fly."[21–23] Indeed, one can consider the active learning approach as an extreme case of extrapolation, with the assumption of an infinite history and high complexity of the approximation model; however, it is desirable to also find simpler extrapolation models that perform well with a few data points, as these can be re-trained quickly on each step, using a small set of near data points rather than a large exhaustive dataset. In the current work, we include the assessment of the benefits and challenges of combining electronic structure machine learning models with advanced extrapolation techniques for the initialization of the SCF cycle to aid SCF acceleration.

B. Recent advances of machine learning models for electronic structure

The discipline of ESML is rapidly growing, and so are the opportunities for accelerated chemical discovery. Over the past decade, ESML models of small molecules and crystals have emerged; these new methods have a great potential for accelerated SCF convergence and perhaps eventually replacing it. Examples of the emergence and applicability of ESML models are broad. In 2018, an artificial neural network (ANN) model of tight-binding Hamiltonian matrix was trained on a small hydrocarbon dataset,[24] and then, subsequently, a SchNOrb ANN model was designed and trained to predict wavefunction coefficients for a few small molecules (water, ethanol, malondialdehyde, and uracil) by Schütt *et al.*,[25] already demonstrating the potential of using ML-predicted Hamiltonians or density matrices for SCF acceleration. More recently, density predictions have been realized via a Gaussian process regression (GPR) model for the real-space electronic density of small hydrocarbons;[26] the symmetry-adapted learning of three-dimensional electron densities (SALTED) method to predict real-space electron density of bulk ice, Al, and Si,[27] as well as of liquid water and of molecules from the QM9 dataset;[28] an ANN model trained to predict real-space electronic density of polyethylene and Al slabs;[29] and the linear Jacobi–Legendre expansion of real-space electronic density for a benzene molecule as well as Al, Mo, and MoS_2 bulk.[30]

In 2022, an atomic cluster expansion (ACE) ESML model was demonstrated for constructing the Hamiltonian and overlap matrices of face-centered cubic (FCC) and body-centered cubic (BCC) bulk aluminum;[31] complementary approaches using ANN models (HamGNN,[32] DeepH-E3,[33] and DeepH[34]) with physically correct intrinsic symmetries were developed and trained to predict the Hamiltonian matrices of a diverse set of systems, including molecules from the QM9 dataset, C and Si allotropes, SiO_2 and Bi_xSe_y bulk crystals, Moiré twisted bilayer graphene, bismuthene, MoS_2 nanotubes, and dislocation defects in Si. The ML-DFT ANN model was able to predict the real-space electronic density of organic molecules, polymer chains, and polymer crystals in Slater-type and Gaussian-type orbital bases.[35] For a few important molecules (H_2O, CO_2, NH_3, CH_3OH, C_6H_6, 1-propanol, and 2-propanol), a combination of kernel ridge and linear regressions has been demonstrated for density matrix prediction;[36] and an ANN model has been trained to predict density matrices from atomic coordinates of H_2O and S_2O molecules and a $[Fe(H_2O)_6]^{2+}$ cation.[37] Most recently, an ML-predicted Hamiltonian was used as an intermediate for the prediction of electronic energy levels, dipole moments, polarizability, and other functions of electronic structure for molecules from QM7 and QM9 datasets and for bulk structures from a graphene dataset,[38] which builds on previous efforts,[39] and also a graph neural network approach (MACE-H) has been demonstrated as high accuracy for Hamiltonian matrix elements and eigenvalues.[40]

A brief summary of recent studies shows that many ESML models have emerged, capable of predicting electronic structure in the form of the density matrix, the Hamiltonian, or coefficients of wavefunctions for *specific* systems, which raises questions about whether the extension of models to new systems can be achieved by combining ESML models. In the present work, we investigate the prediction and manipulation of single-electron reduced density matrices of Kohn–Sham systems as a key descriptor for electronic structure, given the ability to readily compute this quantity from Hamiltonians or wavefunction expansions, and we consider approaches to studying systems beyond the training environment.

C. SCF mixers and preconditioners

The convergence of an iterative algorithm, such as the Kohn–Sham SCF cycle, can often require additional efforts to suppress solution oscillation between iterations. The oscillations are known as "charge sloshing" in DFT,[41] and the challenge is commonly addressed by "mixing" the electronic structure from several previous SCF steps to prevent sudden system changes. Popular mixing approaches include the Pulay[42] and Broyden[43] schemes that

follow a quasi-Newton method of approximating an inverse Jacobian from several previous iterations; for systems with long-range density changes, such as metals, a special Kerker preconditioner has been additionally shown as beneficial.[41,44]

The optimal parameters for mixing schemes vary for different systems, making the predefinition of defaults challenging; for example, molecular systems with a large gap between occupied and unoccupied electronic states may be efficiently converged with a simple linear mixing approach, whereas systems without an electronic gap can require much gentler mixing, with multiple previous SCF solutions, to achieve convergence. To improve default behavior in electronic structure software, such heuristics can be used to automatically select mixing schema and parameters after initial SCF iterations.[45] However, we note that these mixing and preconditioning algorithms rely on a sequence of solutions from previous SCF iterations, which itself may be compromised by changes to the density definition during SCF initialization, and there are open challenges in optimizing such approaches to efficiently converge SCF calculations with an improved ESML initial guess.

D. The practical challenges of applying ESML models

Significant quantities of proof-of-principle work have been published, but currently, there is no established *ab initio* software package that routinely uses ML-based SCF initialization. The limited opportunity for application is partially due to insufficient generality of existing ESML models; currently, models are limited to the complex chemistry on which they are developed, e.g., adsorption monolayers, solid–liquid interfaces, and molecules in porous materials, and a general-purpose ESML model (similar to the recent results of the efforts of the force field community[46]) does not exist. Given the effort invested in training individual models, and the intractably large datasets that may be needed to realize a "general" model, a valuable consideration is to combine ESML models only for parts of any system of interest, or to use several separately trained ESML models in combination to simulate a single complex system. For classical force fields, mixing rules can be applied that define parameters for unfitted combinations of chemical elements; for ESML models, we propose and test here an equivalent approach that combines density matrices to model unfitted systems.

E. Key questions

The work herein has been realized by combining ESML models with the `ASI` API[9] to improve density matrix (DM) initial configurations, which has then been applied to accelerate the convergence of the SCF cycle within `FHI-aims`. The `ASI` API has been implemented in the `FHI-aims` software package to consider the following questions: (1) How does the accuracy of an initial guess for the DM affect convergence of the SCF cycle? (2) How do ESML models trained only for a part of a system perform if used to accelerate SCF convergence? (3) How can the ESML predictions be combined for sequential calculations with extrapolation techniques?

II. METHODOLOGY

A. Molecular models

Clusters of water molecules of various sizes (6, 12, 21, and 101 molecules) and a methane clathrate cage (one CH_4 molecule and 20 H_2O molecules) were used to test how predictions of the DM impact the convergence of electronic structure calculations. The geometry of the small water clusters (6, 12, and 21 molecules) was taken from the Cambridge Cluster Database,[47,48] and the geometry of a 101-molecule cluster was taken from the website of the ErgoSCF software package.[49,50] The geometry dataset of water clusters was diversified by performing 1 ps of molecular dynamics (MD) at 300 K using `DFTB+`,[51,52] with a time step of 1 fs (i.e., 1000 steps). A single methane clathrate structure was also considered, with the geometry taken from the work of Modrzejewski *et al.*[53] In all cases, the converged ground-state DMs were stored for each considered structure (i.e., all ground state structures and each structure in the MD trajectories), and the true DMs were used to assess the accuracy of subsequent DM predictions.

B. DFT calculations

DFT calculations have been conducted using the `FHI-aims` software package[7] (version: 20241126) built as a shared library with the `ASI` API enabled. All calculations used the PBE exchange–correlation density functional,[54] with the zeroth-order regular approximation (ZORA) for relativistic corrections.[55] `FHI-aims` specific integration grids and basis sets were applied, with the "tight" integration grids (i.e., numerically converged), a tier 2 basis set for hydrogen atoms and a tier 1 basis set for all other elements (equivalent to TZP and DZP, respectively).[7] The Kohn–Sham occupancy numbers were smeared using a Gaussian method with a smearing parameter of 0.01 eV. The smearing produced no actual effect on the considered systems but ensured alignment with previous work from which our ESML model for water is developed.[25]

With default settings, `FHI-aims` automatically adjusts the SCF mixing parameters "based on a simple estimate of the system character according to its approximate HOMO–LUMO gap or bandgap,"[45] which can be a valuable heuristic in cases when optimal mixing parameters are unknown *a priori* as it tangibly reduces the number of SCF steps. However, for benchmarking, such adjustments can also add randomness to the number of SCF steps. Therefore, two SCF mixing settings were considered: (1) the default settings with automatic adjustment and (2) fixed settings using the Pulay method and a charge mixing parameter of 0.1. The use of automatic adjustment is explicitly noted in the text where applied, and in all other cases, the fixed settings were used.

Calculations were performed using the `asi4py` Python package, which is a Python wrapper for the `ASI` API that provides access to underlying electronic structure data objects, such as the DM, as well as integration to the functionality of the Atomic Simulation Environment (`ASE`) Python package.[56] The electronic structure (DM) for initializing each SCF cycle was provided through the `ASI` API. The ground-state DM, as well as intermediate density matrices after each loop of the SCF cycle, was extracted from `FHI-aims` via the `ASI` API for further analysis and reuse.

C. Manipulation of the density matrix

Practical approaches of how machine learning models of the density matrix can be integrated into *ab initio* workflows require definitions of various implementation options, which we discuss herein.

1. Approaches to SCF initialization with the density matrix

We have benchmarked the convergence of the SCF cycle within our DFT calculations using a range of initialization methods, as follows:

1. A superposition of atomic electron densities (SAD), which is a common default method for single-point calculations.
2. An electron density from a previously converged SCF cycle, which is a common default method for sequential calculations (e.g., geometry optimization, molecular dynamics, and vibrations).
3. A converged ground-state DM with artificially added noise of controlled magnitude; this method is used for tests to quantify the impact of the inaccuracies in the initial DM on SCF convergence.
4. A prediction of the DM by the SchNOrb ESML model.
5. An algorithm for "stitching" DMs, proposed herein, which is applicable to all forms of separated DMs.
6. An extrapolated prediction of the DM based on previous DMs in a sequential geometry calculation.
7. A combination of ESML predictive models and extrapolative methods.

In order to apply these approaches, we have extended the `asi4py` Python package with classes and subroutines that enable advanced methods of initializing the SCF cycle as outlined in Subsection 1 of the Appendix. Furthermore, technical details relating to the implementation of the approach for calculating the SAD are presented in Subsection 2 of the Appendix.

2. The structure of the density matrix

For clarity of further discussion, we introduce the notion of intra- and inter-molecule blocks of a DM. In electronic structure software packages with localized basis sets (e.g., `FHI-aims`, `Siesta`, `GAUSSIAN`, `DMol3`, `Orca`, and `DFTB+`), each basis function is localized on a basis center, which is typically an atom.[7] Therefore, the rows and columns of a DM can be grouped by atoms, or by molecules, or by any other arbitrary grouping of atoms (e.g., functional groups). The DM elements with row and column indices that correspond to basis functions of the same atom are called herein *intra-atomic* elements; and the DM elements for which row and column indices correspond to basis function pairs of different atoms are called *inter-atomic* elements.

In the same way, we can define *intra-molecule* and *inter-molecule* entries in the DM, grouping basis functions by molecules on which they are localized. In `FHI-aims`, basis functions localized on the same atoms are indexed continuously, forming continuous blocks of basis functions for each atom; and the order of atoms corresponds to the input geometry data. Thus, atoms can be ordered in our geometry input so that atoms of each molecule are also indexed continuously, and therefore, the basis functions of separate molecules form continuously indexed DM blocks. Due to the continuity of groups of basis functions, the DM entries are also grouped into continuous intra- and inter-molecule blocks (analogous to intra- and inter-atomic blocks). The symmetry of the DM means that intra-molecule blocks are located on the diagonal, and inter-molecule blocks are off-diagonal. For example, Fig. 2(b) shows the ground-state DM of a four-molecule water cluster [Fig. 2(a)]. Figures 2(c) and 2(d) show DMs based on the density of free atoms and free molecules, respectively; the DM constructed from DMs of free molecules [Fig. 2(d)] is a noticeably better approximation to the DM of the whole system compared with the DM constructed from the free atoms [Fig. 2(c)].

3. Combining ("stitching") DMs

Our testing in this work shows that an ESML model that predicts only intra-molecule DM blocks is insufficient to substantially speed up the convergence of the SCF cycle, and inter-molecule blocks are necessary (Sec. III B). To predict inter-molecule DM blocks for water, a SchNOrb[25] model was adapted for training differently sized molecular systems at the same time and was trained here for pairs of water molecules (Sec. II C 4). To extrapolate the two-molecule model to water clusters of arbitrary size, a simple DM stitching algorithm is proposed and implemented in `Python` in the `asi4py` package. The general premise of the algorithm is to split the simulated system into spatially localized pieces, e.g., molecules, clusters, and slabs. The DM elements for these pieces are then computed or predicted; intra-molecule DM blocks transfer trivially to the total DM, and the off-diagonal (inter-molecule) DM blocks of each pair transfer to the corresponding off-diagonal (inter-molecule) blocks of the total DM. In the common scenario where a molecule is included in several neighboring pairs, an average of intra-molecule DM blocks over different pairs is placed into the total DM. The key steps of the proposed algorithm are as follows:

1. Make a list of all ordered pairs of molecules with an inter-molecule distance below a cutoff threshold,

$$r_{\text{cut}} : P ::= \{(i,j) : 1 \le i < j \le M, r_{ij} < r_{\text{cut}}\},$$

   where M is the number of molecules in the system, i and j correspond to the ith and jth molecules, and r_{ij} is the distance between these molecules. In our implementation herein, the inter-molecule distance for water molecules has been defined as the distance between oxygen atoms, and the optimal cutoff distance for water was found to be $r_{\text{cut}} = 3$ Å. Details of the fitting procedure and results are given in Sec. III C.
2. Predict a DM for each pair (i,j) in P,

$$\begin{pmatrix} Y_{ij} & X_{ij} \\ X_{ji} & Y_{ji} \end{pmatrix} ::= \widetilde{\rho}(\{i,j\}),$$

   where X_{ij} and Y_{ij} are the inter- and intra-molecule blocks, respectively, of the ith molecule in the predicted DM of a molecular pair (i,j) or (j,i). The shape of the X and Y matrices is $N \times N$, where N is the number of basis functions localized on the atoms of a single molecule. $\widetilde{\rho}(S)$ is the resulting predicted DM for a given subset S of molecules, with shape $SN \times SN$; $\widetilde{\rho}(\{i,j\})$ is thus the predicted DM of the pair of molecules with indices i and j and has a shape of $2N \times 2N$.

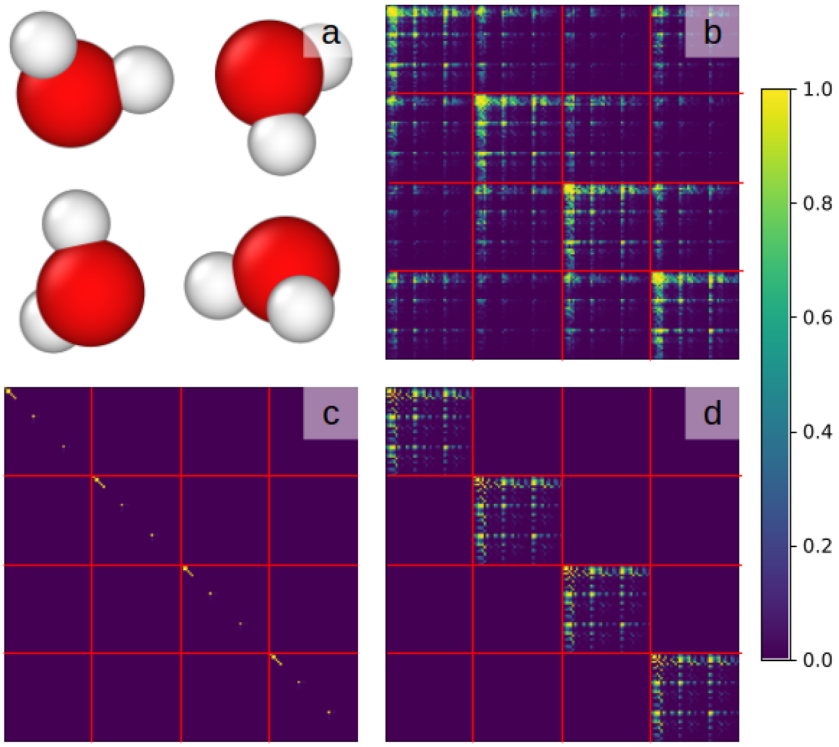

FIG. 2. Density matrix (DM) of four water molecules. Matrix elements (x) have been scaled using the $\arctan(100|x|)$ function for visual clarity. Red lines delimit the groups of basis functions centered on different molecules. (a) Geometry of four water molecules. Image made with the OVITO software.[57] (b) Ground-state DM. (c) Initial DM based on the free atom electronic structure. (d) Initial DM based on the electronic structure of free water molecules.

3. Build the full DM,

$$\widetilde{\rho} := \begin{pmatrix} \overline{Y_1} & X_{12} & \dots & X_{1j} & \dots & X_{1M} \\ X_{21} & \overline{Y_2} & \dots & X_{2j} & \dots & X_{2M} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{i1} & X_{i2} & \dots & \overline{Y_i} & \dots & X_{iM} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{M1} & X_{M2} & \dots & X_{Mj} & \dots & \overline{Y_M} \end{pmatrix},$$

where $\widetilde{\rho}$ is the predicted DM of the entire system, with shape $MN \times MN$, and $\overline{Y_i}$ is an average of Y_{ij} for all j such that (i,j) or (j,i) is in P (see step 1). If neither (i,j) nor (j,i) is in P, then X_{ij} is defined as 0.

4. Construction of a density matrix with a SchNOrb ESML model

In this work, a SchNOrb ESML model is used to learn the relationship between molecular structures and the DM. As the rate of the SCF convergence for water clusters has been found to depend on both the intra- and inter-molecule DM blocks, training of SchNOrb ESML models has been investigated considering intra-molecule interactions (i.e., a single water molecule) and inter-molecule interactions (i.e., pairs of water molecules), both separately (i.e., as independent models) and together (i.e., as a combined model).

To perform training of the DM, 5000 molecular structures were taken from the dataset of Schütt *et al.*[25] with the respective DMs recomputed with `FHI-aims` using the same level of theory and basis set as defined in Sec. II B. The data were recomputed as additional data were needed for water dimers; these structures were sampled in a canonical (NVT) ensemble at 300 and 500 K with a time step of 1 fs, using `FHI-aims` and `DFTB+`.[7,51]

Different approaches for training the ESML model were tested: training for single water molecules (monomers) and pairs of water molecules (dimers) simultaneously (at 300 and 500 K, separately and together); training for the water monomers and water dimers separately; and training for the intra- and inter-molecule blocks separately. The best results were obtained when training for intra- and inter-molecule blocks on water dimer data sampled at 300 and 500 K. Mean absolute errors (MAEs) and root mean-squared errors (RMSEs) for the DM were $\sim 1 \times 10^{-4}$ and 2×10^{-4} a.u., respectively. The models were trained on 3500 data points, validated with 500 data points, and tested with the remaining 1000 points. The data were split randomly as implemented in SchNetPack.[58] The cutoff for representing atoms in their chemical and structural environment was chosen to be 2.5 Å. The models were configured with three interaction layers and trained for a maximum of 500 000 epochs with a starting learning rate of 10^{-4}; the learning rate was decreased by 50% after 50 epochs when no improvement in training was visible, down to 10^{-6}. The batch size was set to 32.

5. Density matrix extrapolation

A sequence of gradually changing atomic geometries intuitively has an electronic structure that also changes gradually and should be predictable. In sequential DFT calculations, the ground state electronic structure of a previous geometry step is often used for SCF

initialization as this substantially improves on the free-atom approximation and accelerates SCF convergence. Extrapolation techniques for DMs, as reviewed above, generalize this concept using DMs of multiple previous geometries to provide a refined DM prediction.[17] Using straightforward predictions of an ESML model for SCF initialization would discard the valuable electronic structure information from the similar geometries of previous geometry steps, but straightforward usage of extrapolation techniques also leaves no place for exploitation of ESML predictions. Fortunately, the *difference* between the predicted and ground state electronic structures is also similar for similar geometries. Therefore, DM extrapolation can be used synergistically with ESML predictions if the difference between a *baseline* DM (e.g., ESML-predicted or free-atom guess) and a ground-state DM is extrapolated, instead of the actual DM itself. This algorithm has been implemented in the `asi4py` package in the `asi.dm.PredictSeqDM` class. We used kernel ridge regression with the radial basis function kernel (squared-exponential kernel) to extrapolate the errors of DM predictors (e.g., ESML or free-atom guess). To make the regression invariant with respect to translations of atoms, the matrix of inverse inter-atomic distances has been used as the regression argument, following recent work by Polack *et al.*[17]

D. Assessment of algorithms

1. Accuracy of density matrices

To quantitatively measure the accuracy of an ESML model at predicting the initial DM, as well as DM convergence over SCF iterations, the 2-norm of deviation of the DM elements from their ground-state values (the Frobenius norm) has been used. To compare this quantity between molecular clusters of different sizes, normalization was performed with respect to the number of molecules (n_{mol}),

$$\varepsilon_{\widetilde{\rho}} = \frac{\sqrt{\sum_{ij} \left(\rho_{ij} - \widetilde{\rho}_{ij}\right)^2}}{\sqrt{n_{\mathrm{mol}}}}, \tag{2}$$

where ρ is the ground-state DM, $\widetilde{\rho}$ is an approximate DM, and i and j are row and column indices, respectively. In some discussion, the error is separately computed for diagonal intra-molecule DM blocks and for off-diagonal inter-molecule blocks.

2. Calculation performance

The benchmarking performance of the outlined methods was measured in SCF iterations. The wall time of calculations is of practical interest, but measurement and comparison are complicated by variations in hardware, systems, and applied libraries, as well as parallelization features, system size, basis set size, etc. A comparison of the number of SCF iterations for different initial DMs with all other factors considered equal provides a convenient indication of SCF speed-up. The computational cost of the ESML (SchNOrb) model predictions, and DM stitching and extrapolation procedures, is noted as negligible compared to the eigenproblem solution in a single SCF iteration. For scalability, these algorithms must exploit the sparsity of the DMs, be implemented in a linearly scaling way, and be trivially parallelized; these attributes are attainable but beyond the scope of this study. Performance gains are also available by using GPU accelerators, which are compatible with the SchNOrb model and many other ESML approaches but also beyond the scope of this study. For these reasons, performance is discussed herein in terms of the number of SCF steps only.

III. RESULTS

A. Convergence of the density matrix during the SCF cycle

The error in the DM has been measured as a function of the number of SCF cycles, relative to the ground state, for the highlighted water clusters. The initial DM in the SCF was first constructed from free-atom DMs, with Fig. 3 demonstrating that the average error of the DM (ε_ρ) decreases logarithmically from this initial state during the SCF cycle. The log-linear convergence of the DM residual has been observed previously by Cancès *et al.* as part of a detailed analysis for bulk Si[59] and by Woods *et al.*

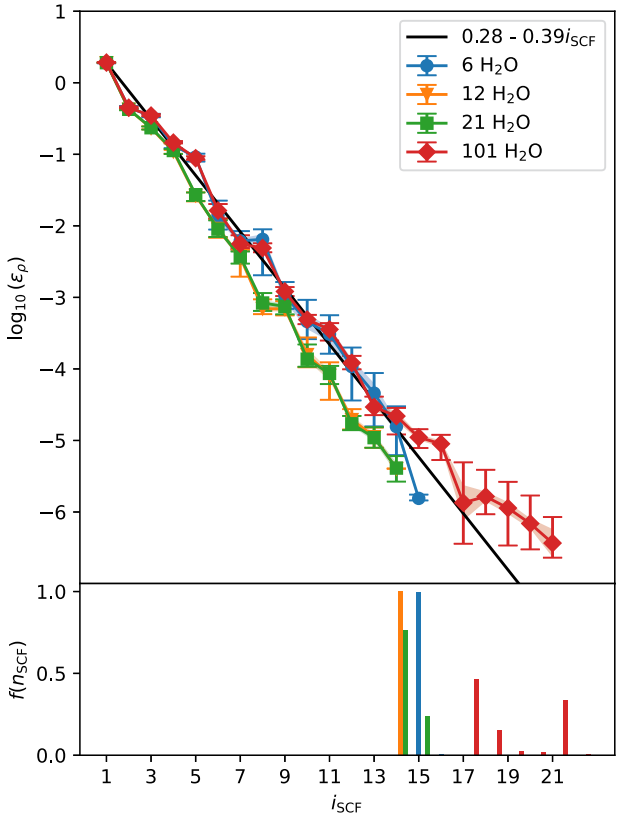

FIG. 3. Top: A logarithmic plot comparing the average error of DM values relative to the ground-state DM, $\varepsilon_{\widetilde{\rho}}$, as defined in Eq. (2), as a function of SCF step, i_{SCF}. The plots are averaged for water clusters with 6, 12, 21, and 101 molecules across 1000 electronic structure calculations with `FHI-aims`, with the geometries obtained from a `DFTB+` molecular dynamics simulation. Error bars are included to show minimum and maximum values. Shading is used to show standard deviations. The black solid line shows a linear best fit for all the data [Eq. (3)]. Bottom: A histogram of the observed probability of convergence of a calculation at the i_{SCF} SCF step.

when demonstrating SCF convergence for Si and Al;[60] examples of log-linear convergence behavior can also be found for benchmarks of notoriously slowly converging surface slabs with adsorbed species.[61]

The probability that a calculation will converge in a given number of SCF steps is also presented, with SCF convergence tending to have a greater variety in the number of SCF cycles as the cluster sizes increase. The variation in convergence is likely related to the variance in DM accuracy at the final SCF steps for larger systems. By averaging over all considered water clusters, the following approximation is identified when starting from the default free-atom guess for the number of SCF steps necessary to reach a specific DM accuracy,

$$i_{\mathrm{SCF}} = \frac{0.28 - \log_{10}(\varepsilon_{\widetilde{\rho}})}{0.39}, \tag{3}$$

where i_{SCF} is the convergence step for a given input DM, $\widetilde{\rho}$, to an error threshold of $\varepsilon_{\widetilde{\rho}}$. This baseline allows us to quantify how many steps can be saved from a SCF cycle if the initial DM prediction gives an accuracy improvement in $\Delta\varepsilon_{\widetilde{\rho}}$, e.g., one order of magnitude would save 2.56 SCF steps. Equation (3) is specific to the water clusters considered in this work and is used herein to show how the accuracy of the DM prediction transforms into performance gains. The possible performance gains, with regard to a reduction in SCF steps, may be greater for systems where the gradient in Fig. 3 may be shallower (i.e., slowly convergent systems).

B. Importance of inter-molecule DM contributions

To understand the potential performance gains from improved estimates of the initial DM based on a single-molecule ESML model, tests were performed where the initial DM contained the intra-molecule (diagonal) blocks from the ground-state (i.e., converged) DM of the whole molecular cluster, effectively simulating an error-free single-molecule ESML model. For the inter-molecule blocks (off-diagonal), two initialization methods have been considered: (1) zeros (i.e., no guess) and (2) the ground-state DM with added random noise. The first method tests the performance of a hypothetical ESML model capable of predicting only a single water molecule with perfect accuracy, e.g., a single-molecule SchNOrb model,[25] and the second method simulates a hypothetical ESML model able to predict inter-molecule blocks with varying accuracy. Figure 4 illustrates how SCF convergence varies for these different approaches relative to the default free-atom guess and shows that the errors of intra-molecule and inter-molecule DM blocks closely follow each other as a function of SCF steps. The graphs indicate that even an exact prediction of intra-molecule blocks gives just modest acceleration of the SCF cycle. There is a noticeable increase in the error in intra-molecular blocks, from an initial value of zero to levels that correspond to the errors in inter-molecular blocks at the second SCF step. The error increase occurs because a single eigenproblem solution spreads inaccuracies from the off-diagonal blocks and spoils the whole DM, making the errors in both inter- and intra-molecule blocks large. In the considered sample, the "model" single-molecule DM initialization reduced the number of SCF steps by ~15%; in contrast, a "coarse" improvement in the inter-molecule blocks gives a significant boost to the SCF (~50%). The results demonstrate that

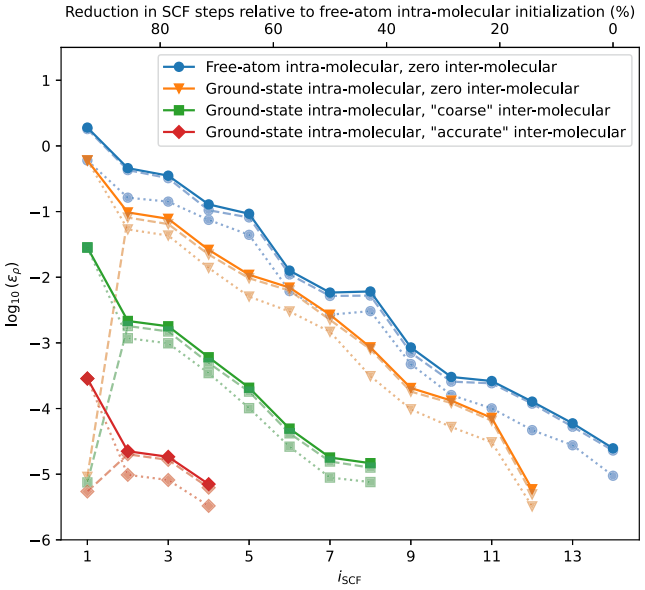

FIG. 4. Plot comparing the convergence of the DM during single-point calculations of a six-molecule water cluster with different methods for estimating the initial DM. A legend is provided, with the solid lines indicating the $\varepsilon_{\widetilde{\rho}}$ values [Eq. (2)], and pale dashed and dotted lines correspondingly show the deviations of intra-molecule and inter-molecule DM blocks taken with respect to the ground-state DM. Random noise has been added to inter-molecule blocks corresponding to a *coarse* uniform distribution of "large" intervals (-5×10^{-4}, 5×10^{-4}, green) and an *accurate* uniform distribution of "small" intervals (-5×10^{-6}, 5×10^{-6}, red). The top *x*-axis shows the percentage reduction in SCF steps for each method relative to the default free-atom initialization.

an accurate estimate of inter-molecule blocks is essential for any significant boost in the SCF convergence.

C. DM "stitching" benchmark

The efficiency of the DM stitching algorithm (Sec. II C 3), including the parameterization of the important cutoff radius r_{cut}, has been performed for water clusters. To predict the DMs of pairs of water molecules, the SchNOrb ESML model was used (Sec. II C 4). The errors in the DM prediction for various water clusters with different r_{cut} values are shown in Fig. 5. The optimal transferable r_{cut} is identified as 3 Å, which is slightly greater than the distance between the oxygen atoms in an isolated water dimer (2.9 Å) and indicates that only the first neighbor molecules contribute strongly to the DM accuracy in this case.

D. Combining ESML prediction with extrapolation

ESML models can be combined with extrapolation algorithms, as reviewed in Sec. I A, to perform SCF cycle initialization. The algorithm proposed in Sec. II C 5 combines these approaches, exploiting the predictive power of advanced ESML models alongside the local adaptability of extrapolation. To assess the potential benefits of such synergy, the DM extrapolation algorithm from Sec. II C 5 has been benchmarked in combination with default free-atom SCF initialization and with DM prediction using the SchNOrb ESML model and

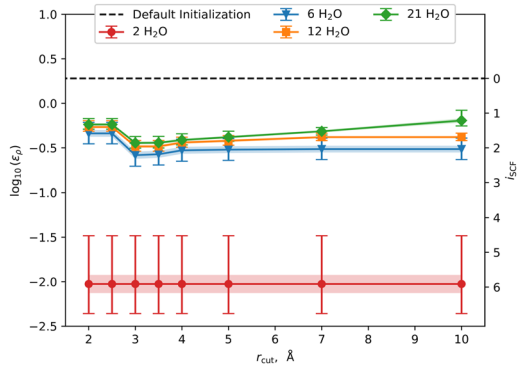

FIG. 5. Graph of log ($\varepsilon_{\tilde{\rho}}$) when using the two-molecule SchNOrb model to create a complete DM for larger H_2O clusters by the DM "stitching" algorithm. The tests are performed as a function of the cutoff radius (Å) for on-diagonal averaging, r_{cut}, to determine if a two-molecule interaction is effective. The black dashed line shows the accuracy of the default free-atom DM. The right *y*-axis shows the approximate number of SCF steps saved by the DM prediction, as calculated using Eq. (3). For the water dimer (red), no DM "stitching" is necessary, and therefore, the results are demonstrated as independent of r_{cut}. The shaded areas and error bars are as defined for Fig. 3.

DM "stitching" (Sec. II C 3). The crucial parameter for the extrapolation algorithm is the number of considered previous geometry steps (n_{geom}), and the effect of this parameter on the extrapolation accuracy is also assessed. For different systems and trajectories, the optimal n_{geom} is presumed to vary, and understanding the impact and reasonable initial values is critical before further exploitation. The results in Fig. 6 (top) show that including a history of up to 6 geometry steps significantly aids DM extrapolation, and including further steps, up to the 10th geometry, continues to return tangible accuracy improvements. For this combined ESML/extrapolation model, the most accurate DM predictions reduce the number of SCF iterations by up to 7 or 8 steps (~50%).

For a system with no ESML model applied (Fig. 6, top left), the 0th geometry step corresponds to a single point calculation with the default `FHI-aims` initialization; and the second calculation (i.e., first geometry step) corresponds to the usage of the converged electronic structure from the previous step only, which is the default behavior for dynamics and geometry optimizations in `FHI-aims` and many other software packages. For the considered systems, this simple transfer of the previously converged DM saves about 4 SCF steps (~25%). A kernel-ridge extrapolation of the difference between the ground-state DM and the free-atom guess for further

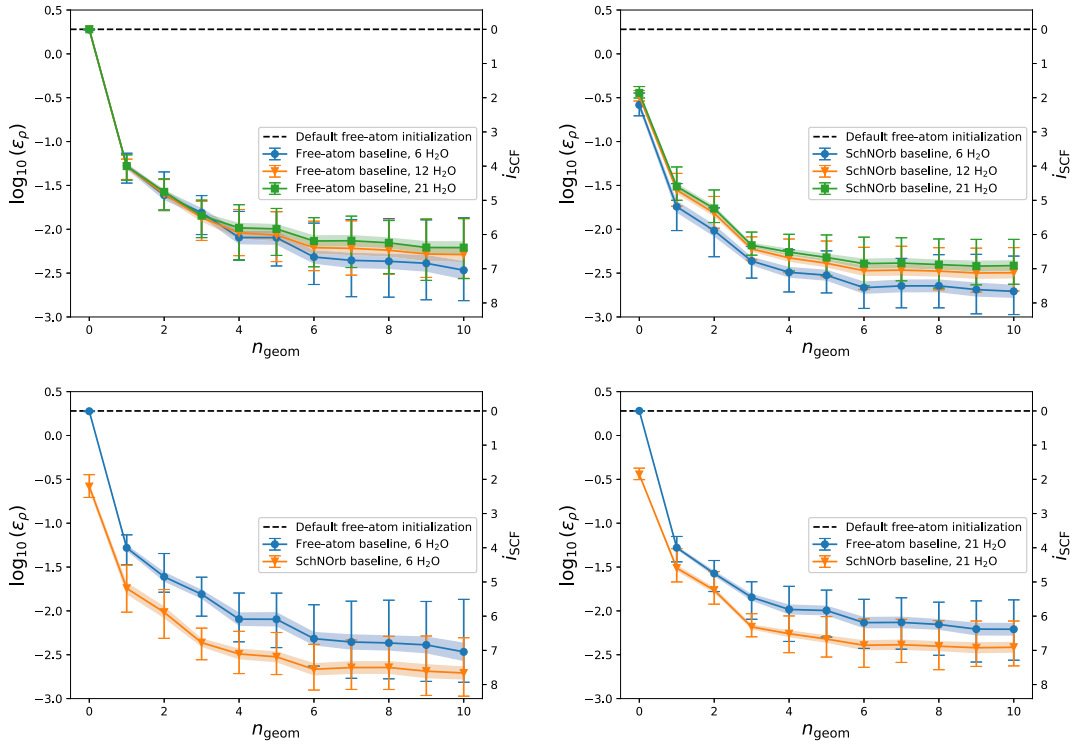

FIG. 6. Graphs of the error in the DM prediction as a function of the number of geometry steps from which DMs are extrapolated (n_{geom}), considering various baseline DM initialization methods and various water cluster sizes. "SchNOrb baseline" refers to the SchNOrb ESML model combined with the DM "stitching" algorithm. A black dashed line shows the accuracy of the default free-atom initial SCF guess. The right *y*-axis shows the approximate number of SCF steps saved by the DM prediction [Eq. (3)]; these values should be compared to the average number of steps for SCF convergence with the default free-atom initialization (Fig. 3), which are 15 and 14 SCF steps for 6- and 21-molecule H_2O water clusters, respectively (i.e., savings of up to ~50% are demonstrated). The shaded areas and error bars are as defined for Fig. 3.

SCF steps, with no ESML model, saves an additional 2–3 SCF steps (~12.5%–18.75%).

For a calculation initialized with the SchNOrb ESML model and DM stitching, without any extrapolation, Fig. 6 (top right) shows that ~2 SCF steps (~12.5%) are saved. The savings with an increasing DM history are comparable to the free-atom initialization when approaching a converged value of $n_{\text{geom}} = 10$, with a maximum saving of ~3 SCF steps (18.75%). The extrapolation of corrections for an ESML model baseline (Fig. 6, bottom) unequivocally shows that the more accurate predictions save between 0.5 and 1 additional SCF steps for any cluster size, relative to the extrapolation for a default free-atom guess. Such gains could be significantly improved by a better ESML model, as demonstrated in Sec. III B where more accurate prediction of intra- and inter-molecular DM blocks results in a greater reduction in the number of SCF steps.

In summary, the highest accuracy initial guess for the SCF is provided from the joint usage of the stitched SchNOrb ESML model with an extrapolated error correction. The total saving is 6–8 SCF steps (up to 50%) for the benchmarked H_2O systems relative to the free-atom initialization with no extrapolation and 2–4 SCF steps (~12.5%–25%) relative to extrapolation from the previous geometry step only ($n_{\text{geom}} = 1$), as applied in most sequential calculations in electronic software packages.

E. Applied examples

1. Water clusters

As mentioned in Sec. II B, the default `FHI-aims` settings automatically adjust the density mixing parameter during the SCF cycle. Furthermore, the benchmark tests show that improvements in the accuracy of the DM prediction do not necessarily transform into SCF acceleration, as per Eq. (3), and that there can be a dependency of the accuracy of the DM predictions on system sizes. All these factors may significantly affect the practical gains from the advanced SCF initialization methods discussed if applied in standard calculations. Therefore, the actual reduction in the number of SCF steps achievable when applying the considered ESML model and extrapolation algorithm in practical conditions is considered using the calculations of water cluster trajectories described in Sec. II A. Unlike the prior benchmarks in Secs. III A–III D, the default `FHI-aims` DFT settings were used, including automatic adjustment of the density mixing parameter (see Sec. II B for details). Both sequential and single point scenarios were tested with default and advanced initialization methods. The tested ESML model was the same as the parameterized model described in Sec. III D, i.e., a two-molecule SchNOrb model, using DM "stitching" with $r_{\text{cut}} = 3$ Å. For sequential calculations, the extrapolation algorithm (Sec. II C 5) used $n_{\text{geom}} = 10$.

The average number of SCF steps for convergence is plotted as a function of water cluster size in Fig. 7. The relative performance gains due to using the ESML model predictions and DM extrapolation are summarized in Table I. The performance gains from the ESML model predictions are larger for small clusters and then reduced significantly for larger systems (up to 101 molecules), where a larger number of SCF steps are observed for larger clusters. The slower convergence of the SCF cycle for larger systems is typically associated with energetically competitive degenerate electronic states, which lead to a well-known charge sloshing phenomenon;[41]

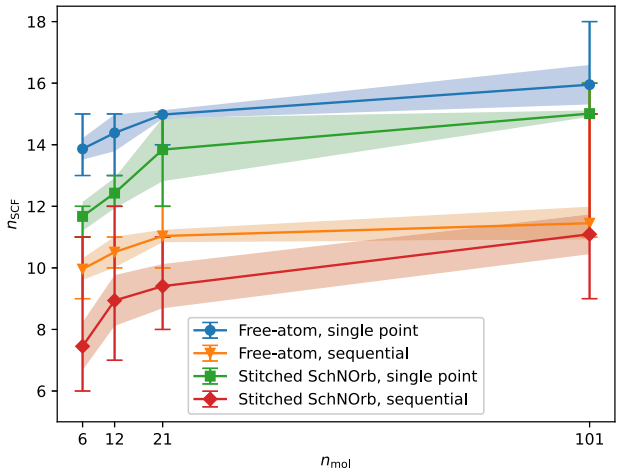

FIG. 7. Performance comparison of the initial DMs, considering default free-atom initialization and a "stitched" two-molecule SchNOrb model prediction, for single-point and sequential calculations over a range of H_2O cluster sizes (n_{mol}). The number of SCF steps necessary to reach convergence (n_{SCF}) is plotted on the y-axis. The shaded areas and error bars are as defined in Fig. 3.

TABLE I. Relative performance gains (% reduction in SCF steps) due to the "stitched" two-molecule SchNOrb model prediction of initial DM for water clusters, compared to standard free-atom approaches. "Single-point" refers to SCF calculations that start with no prior information on the DM, and "sequential" refers to calculations where prior information on the DM is used in the DM extrapolation.

Cluster size (molecules)	Single-point (%)	Sequential (%)
6	16	25
12	14	15
21	8	15
101	6	3

however, we hypothesize also that the "stitching" algorithm and the two-molecule SchNOrb model work better in our tests for predicting the DM of molecules with fewer neighbors, e.g., on the cluster surface, rather than for molecules with more neighbors, e.g., in the cluster. As a consequence, the accuracy of the DM guess for the H_2O systems with this approach is better for smaller clusters, where the surface-to-bulk ratio is higher, and this impacts the rate of SCF convergence in a proportional manner. The results demonstrate the potential of the proposed DM "stitching" approach to produce more accurate predictions when used with more recent ESML models trained on larger chemical systems, such as ACE[31] or DeepH-E3.[33]

Figure 7 also shows a greater range in the number of SCF steps when the SchNOrb-predicted initialization is applied, compared to the free-atom initialization. This range is not observed in controlled testing with static density mixing parameters (Fig. 6), and therefore, we conclude that the improved guess for the initial DM unfavorably impacts the internal heuristics that control on-the-fly adjustment of density mixer parameters in `FHI-aims`. Thus, the results also demonstrate a need for improvements to mixer configurations for widespread uptake of such paradigms.

2. Methane clathrate

To assess the applicability of the "stitching" algorithm in cases when an ESML model is available for only parts of the simulated system, calculations of a methane clathrate cage (i.e., a hydrated methane cluster) have been performed. For the CH_4 molecule, the default free-atom initialization approach has been applied, and the "stitched" approach has been used to build the DM of the water cage using the two-molecule SchNOrb model. Three variants of SCF initialization have been tested, demonstrating the versatility of the "stitching" algorithm: (1) a default free-atom guess for the whole system; (2) the "stitched" SchNOrb DM prediction for water shell and free-atom guess for the CH_4 molecule; and (3) the "stitched" SchNOrb DM prediction for water shell and a ground-state DM of a free CH_4 molecule (with the same geometry as in the clathrate complex). The inter-molecule (off-diagonal) DM blocks that correspond to overlap between the CH_4 and water cage basis functions were set to zero in all cases.

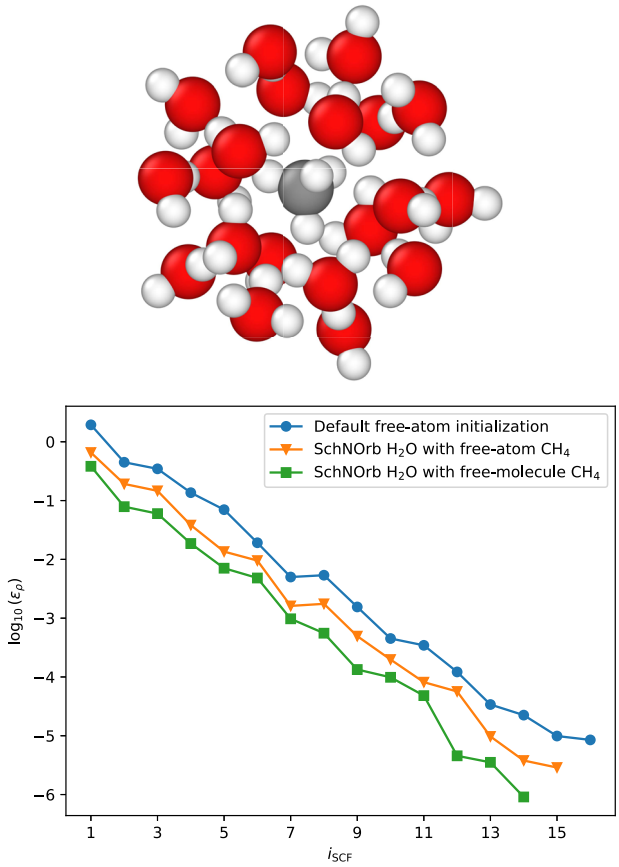

FIG. 8. Top: A methane clathrate cluster made using the structure from Modrzejewski *et al.*[53] and visualized with the OVITO software.[57] Bottom: DM errors plotted as a function of SCF step for the clathrate cluster. Blue: The default free-atom initialization. Orange: The stitched two-molecule SchNOrb model for water combined with the default free-atom method for the CH_4 DM block. Green: The stitched two-molecule SchNOrb model for water combined with the ground-state DM of a gas-phase CH_4 molecule.

The error in the DM as a function of SCF step is shown in Fig. 8 for each initialization approach. The best initialization method is identified as (3), using the pre-converged CH_4 molecule with the ESML water model. The rate of SCF convergence is approximately equal for all approaches. The results inform that, despite the difference between the various initial guesses being less than a single decimal order, using the two-molecule SchNOrb water model with the free-atom and free-molecule DM blocks for the methane molecule saves 1 and 2 SCF steps (~13.4%), respectively.

IV. CONCLUSIONS AND FUTURE PERSPECTIVES

The ability to augment and accelerate SCF convergence in electronic structure calculations has been considered via a combination of ESML and data extrapolation techniques. Several key outcomes have been observed that relate directly to the initial research questions proposed in Sec. I E, as follows:

1. Advanced techniques for the initialization of the SCF cycle, with more accurate electronic structure, are capable of reducing the number of SCF iterations and accelerating calculations. Generally, higher accuracy predictions of the DM lead to a proportional reduction in the number of SCF steps, but nuances exist. In particular, if the accuracy of the DM prediction is not uniform over its elements, the performance gain is limited by the least accurate part of the DM (Sec. III B).
2. An ESML model trained for only a subset of an entire system is demonstrated as useful for accelerating SCF convergence. A simple "stitching" algorithm to combine DM elements (Sec. II C 3) can be used to build the DM of the bigger system using DMs of overlapping subsystems, substantially improving SCF initial guess accuracy compared with the default free-atom guess. For the water clusters considered in Sec. III C, the use of the SchNOrb ESML model reduced calculations by ~2 SCF iterations for single-point calculations (Fig. 5). Even when there are no ESML models for some parts of the system, the DM stitching approach can be used beneficially to combine ESML and non-ESML data as demonstrated for a methane molecule in a water cage (Sec. III E 2). While the improved initial guess of the DM does demonstrate value for accelerating SCF convergence, it does not address more general challenges with convergence, such as charge sloshing.
3. DM extrapolation techniques can be used synergistically with ESML predictions if the difference between the predicted and ground-state DMs is extrapolated and used as a correction for the ESML prediction (Sec. II C 5). With this approach, we managed to save up to 8 SCF steps for the water cluster calculations, when compared with the default free-atom initialization (Fig. 6).

The performance benefits from applying an ESML model in this study are overall modest, with benefits reduced as the system size increases; indeed, the cost of training the ESML water model applied in this work (5000 single-point calculations) is justified only once ~25 000 electronic structure calculations have been performed for systems containing water. However, the ability to transfer the ESML water model and "stitching" algorithm to any calculation that involves water, which covers many fields of chemistry, provides the

necessary breadth of value beyond the immediate study. Furthermore, this work demonstrates the potential of these methods, and further advancement of ESML models offers great opportunities for performance improvements in *ab initio* electronic structure codes in the near future. In particular, we envisage that ESML models trained on larger molecular clusters (or periodic systems) would better account for the molecular environment, providing more accurate inter-molecule DM blocks that are pivotal for success. ESML models of small molecules, clusters, surfaces, or adsorption complexes can be developed and combined for the simulation of a wider set of systems, using the algorithms presented to combine DMs and to perform quicker chemical investigations.

We also note persistent challenges with respect to the default density mixers, preconditioners, and SCF heuristics that can be confused by an accurate initial DM prediction (Secs. III B and III E). When the SCF cycle starts in the quadratic region, with no information about the inverse Hessian, several wasteful iterations are spent far from the converged solution. In this case, other SCF solvers may be preferred, e.g., Newton–Raphson or conjugate gradient. The development of ESML models that are able to provide useful information about the accuracy of different parts of the initial guess would be highly valuable for this endeavor. Thus, there is an opportunity to rethink algorithmic strategies for SCF convergence in future software developments.

We also see value in platforms to combine ESML models that use different basis sets for electronic structure representation; the incompatibility complicates the practical usage of multiple models currently. A uniform API for the high-performance interaction of DFT codes with ESML models (such as `ASI`[9]) is also necessary for adoption by software packages, facilitating the development of automated interoperability ESML-augmented workflows.

ACKNOWLEDGMENTS

P.V., R.J.M., and A.J.L. acknowledge funding from the ARCHER2 eCSE program (No. eCSE03-10). P.V. and A.J.L. acknowledge funding by the UKRI Future Leaders Fellowship program (Nos. MR/T018372/1 and MR/Y034279/1). R.J.M. acknowledges funding by the UKRI Future Leaders Fellowship program (No. MR/S016023/1) and a UKRI Frontier Research Grant (No. EP/X014088/1). J.W. acknowledges funding from the Austrian Science Fund (FWF) (No. J 4522N).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Pavel Stishenko: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Chen Qian**: Formal analysis (equal); Investigation (equal); Software (equal); Writing – review & editing (equal). **Julia Westermayr**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Supervision (equal); Validation (equal); Writing – review & editing (equal). **Reinhard J. Maurer**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Supervision (equal); Validation (equal); Writing – review & editing (equal). **Andrew J. Logsdail**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The training dataset from which the SchNOrb model was derived (water dimers at 300 and 500 K) can be accessed at https://www.doi.org/10.5281/zenodo.15125083. The testing dataset on which the model was applied (6-, 12-, 21-, and 101-molecule water clusters and the clathrate cage) and the scripts with which DMs have been calculated can be accessed at https://www.doi.org/10.5281/zenodo.15460932. The SchNOrb ESML model can be accessed at https://www.doi.org/10.5281/zenodo.15045586. The SchNOrb codebase for DM predictions can be accessed at https://github.com/juliawestermayr/schnorb_dm. The benchmarking was performed using sample scripts from the `asi4py.dm` documentation that are freely available at the GitLab website (https://pvst.gitlab.io/asi/ml_dm.html) and are the part of the open ASI repository (https://gitlab.com/pvst/asi).

APPENDIX: IMPLEMENTATION DETAILS

1. Classes introduced to `asi4py.dm`

The following classes were added to the `asi4py.dm` subpackage to enable the presented work:

- `PredictFreeAtoms` initializes the DM using electron density of free atoms, corresponding to the default behavior of `FHI-aims` in case of single-point calculations. We used this class for benchmarking purposes and for parts of simulated systems with no available ESML models (e.g., CH_4 molecule in a clathrate).
- `PredictConstAtoms` initializes the DM using the DM result(s) of earlier calculation(s), i.e., sequential calculations.
- `PredictDMByAtoms` initializes the DM using ESML models, such as SchNOrb,[25] that predict electronic structure by atomic coordinates.
- `PredictFrankensteinDM` initializes the DM of the whole system by combining ("stitching") multiple density matrices, for different parts of the simulated system, into a single matrix.
- `PredictSeqDM` introduces corrections to the DM predictor by extrapolating any error arising from sequential geometries (as discussed further in Sec. II C 5).

2. Construction of a free-atom density matrix

`FHI-aims` uses an atomic solver to initialize the SCF cycle, which means that the first DM is produced only from wavefunction coefficients after the first solution of the eigenproblem in Eq. (1).

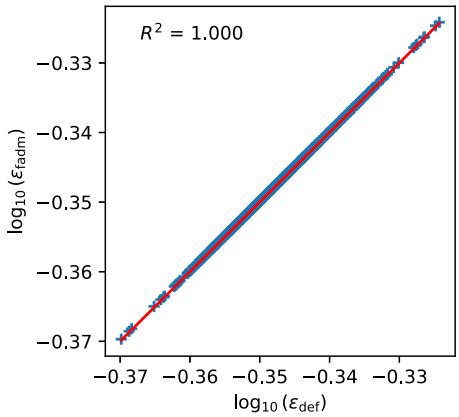

FIG. 9. Errors calculated at the second SCF step when using default initialization (ε_{def}) and initialization using free-atom density matrices ($\varepsilon_{\text{fadm}}$), sampled across 1000 different geometries of a six-molecule water cluster. The red line shows a near exact match for $\varepsilon_{\text{def}} = \varepsilon_{\text{fadm}}$. The error values are defined by Eq. (2).

Therefore, to estimate the accuracy of the default initial guess on equal footing with other methods, the `PredictFreeAtoms` class was implemented to construct the initial DM from precomputed DMs of free atoms [i.e., Fig. 2(c)]. The validity of the method was tested by comparing SCF convergence for 1000 different geometry configurations of the six-molecule water cluster, using the default initialization method within `FHI-aims` and DMs constructed from free atom DMs; and 99.9% (999/1000) calculations converged with the same number of SCF steps (15), with the other calculation differing by one SCF step. We also compared DM errors before the second SCF step (i.e., after the first eigenproblem solution), and a near perfect match is observed (Fig. 9), with mean absolute error (MAE) less than 1.4×10^{-4}. These results demonstrate that any discrepancies in our approach are very small, and the use of the free-atom DMs is considered equivalent to the default initialization scheme in `FHI-aims` herein.

REFERENCES

[1] C. C. J. Roothaan, "New developments in molecular orbital theory," Rev. Mod. Phys. **23**, 69 (1951).
[2] G. G. Hall, "The molecular orbital theory of chemical valency VIII. A method of calculating ionization potentials," Proc. R. Soc. London, Ser. A **205**, 541–552 (1951).
[3] J. H. Van Lenthe, R. Zwaans, H. J. J. Van Dam, and M. F. Guest, "Starting SCF calculations by superposition of atomic densities," J. Comput. Chem. **27**, 926–932 (2006).
[4] P. Giannozzi *et al.*, "QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials," J. Phys.: Condens. Matter **21**, 395502 (2009).
[5] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," Phys. Rev. B **54**, 11169 (1996).
[6] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, "The SIESTA method for ab initio order-N materials simulation," J. Phys.: Condens. Matter **14**, 2745 (2002).
[7] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, "Ab initio molecular simulations with numeric atom-centered orbitals," Comput. Phys. Commun. **180**, 2175–2196 (2009).
[8] V. W.-z. Yu, C. Campos, W. Dawson, A. García, V. Havu, B. Hourahine, W. P. Huhn, M. Jacquelin, W. Jia, M. Keçeli *et al.*, "Others ELSI—An open infrastructure for electronic structure solvers," Comput. Phys. Commun. **256**, 107459 (2020).
[9] P. V. Stishenko, T. W. Keal, S. M. Woodley, V. Blum, B. Hourahine, R. J. Maurer, and A. J. Logsdail, "Atomic simulation interface (ASI): Application programming interface for electronic structure codes," J. Open Source Software **8**, 5186 (2023).
[10] P. Stishenko, A. McSloy, B. Onat, B. Hourahine, R. J. Maurer, J. R. Kermode, and A. Logsdail, "Integrated workflows and interfaces for data-driven semi-empirical electronic structure calculations," J. Chem. Phys. **161**, 012502 (2024).
[11] S. Lehtola, "Assessment of initial guesses for self-consistent field calculations. Superposition of atomic potentials: Simple yet efficient," J. Chem. Theor. Comput. **15**, 1593–1604 (2019).
[12] F. Ballesteros and K. U. Lao, "Accelerating the convergence of self-consistent field calculations using the many-body expansion," J. Chem. Theor. Comput. **18**, 179–191 (2022).
[13] F. Ballesteros and K. U. Lao, "Analysis of two overlapping fragmentation approaches in density matrix construction: GMBE-DM vs. ADMA," Phys. Chem. Chem. Phys. **26**, 4386–4394 (2024).
[14] P. G. Mezey and Z. Antal, "An alternative to the 'Star Path' enhancement of the ADMA linear scaling method for protein modeling," J. Comput. Chem. **38**, 1774–1779 (2017).
[15] P. Pulay and G. Fogarasi, "Fock matrix dynamics," Chem. Phys. Lett. **386**, 272–278 (2004).
[16] A. M. N. Niklasson, P. Steneteg, A. Odell, N. Bock, M. Challacombe, C. J. Tymczak, E. Holmström, G. Zheng, and V. Weber, "Extended Lagrangian Born–Oppenheimer molecular dynamics with dissipation," J. Chem. Phys. **130**, 214109 (2009).
[17] É. Polack, G. Dusson, B. Stamm, and F. Lipparini, "Grassmann extrapolation of density matrices for Born–Oppenheimer molecular dynamics," J. Chem. Theor. Comput. **17**, 6965–6973 (2021).
[18] K. U. Lao, K. Wickramasinghe, and J. A. Tan, "Grassmann extrapolation via direct inversion in the iterative subspace," J. Chem. Phys. **163**, 144114 (2025).
[19] J. A. Tan and K. U. Lao, "Generating accurate density matrices on the tangent space of a Grassmann manifold," J. Chem. Phys. **158**, 051101 (2023).
[20] J. A. Tan and K. U. Lao, "The Grassmann interpolation method for spin-unrestricted open-shell systems," J. Chem. Phys. **158**, 214104 (2023).
[21] Z. Li, J. R. Kermode, and A. De Vita, "Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces," Phys. Rev. Lett. **114**, 096405 (2015).
[22] J. Westermayr, M. Gastegger, K. T. Schütt, and R. J. Maurer, "Perspective on integrating machine learning into computational chemistry and materials science," J. Chem. Phys. **154**, 230903 (2021).
[23] E. Podryabinkin, K. Garifullin, A. Shapeev, and I. Novikov, "MLIP-3: Active learning on atomic environments with moment tensor potentials," J. Chem. Phys. **159**, 084112 (2023).
[24] H. Li, C. Collins, M. Tanha, G. J. Gordon, and D. J. Yaron, "A density functional tight binding layer for deep learning of chemical Hamiltonians," J. Chem. Theor. Comput. **14**, 5764–5776 (2018).
[25] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer, "Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions," Nat. Commun. **10**, 5024 (2019).
[26] A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and M. Ceriotti, "Transferable machine-learning model of the electron density," ACS Cent. Sci. **5**, 57–64 (2019).
[27] A. M. Lewis, A. Grisafi, M. Ceriotti, and M. Rossi, "Learning electron densities in the condensed phase," J. Chem. Theor. Comput. **17**, 7203–7214 (2021).
[28] A. Grisafi, A. M. Lewis, M. Rossi, and M. Ceriotti, "Electronic-structure properties from atom-centered predictions of the electron density," J. Chem. Theor. Comput. **19**, 4451–4460 (2023).
[29] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, "Solving the electronic structure problem with machine learning," npj Comput. Mater. **5**, 22 (2019).
[30] B. Focassio, M. Domina, U. Patil, A. Fazzio, and S. Sanvito, "Linear Jacobi-Legendre expansion of the charge density for machine learning-accelerated electronic structure calculations," npj Comput. Mater. **9**, 87 (2023).

[31] L. Zhang, B. Onat, G. Dusson, A. McSloy, G. Anand, R. J. Maurer, C. Ortner, and J. R. Kermode, "Equivariant analytical mapping of first principles Hamiltonians to accurate and transferable materials models," npj Comput. Mater. **8**, 158 (2022).
[32] Y. Zhong, H. Yu, M. Su, X. Gong, and H. Xiang, "Transferable equivariant graph neural networks for the Hamiltonians of molecules and solids," npj Comput. Mater. **9**, 182 (2023).
[33] X. Gong, H. Li, N. Zou, R. Xu, W. Duan, and Y. Xu, "General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian," Nat. Commun. **14**, 2848 (2023).
[34] H. Li, Z. Wang, N. Zou, M. Ye, R. Xu, X. Gong, W. Duan, and Y. Xu, "Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation," Nat. Comput. Sci. **2**, 367–377 (2022).
[35] B. G. Del Rio, B. Phan, and R. Ramprasad, "A deep learning framework to emulate density functional theory," npj Comput. Mater. **9**, 158 (2023).
[36] X. Shao, L. Paetow, M. E. Tuckerman, and M. Pavanello, "Machine learning electronic structure methods based on the one-electron reduced density matrix," Nat. Commun. **14**, 6281 (2023).
[37] S. Hazra, U. Patil, and S. Sanvito, "Predicting the one-particle density matrix with machine learning," J. Chem. Theor. Comput. **20**, 4569–4578 (2024).
[38] D. Suman, J. Nigam, S. Saade, P. Pegolo, H. Türk, X. Zhang, G. K.-L. Chan, and M. Ceriotti, "Exploring the design space of machine-learning models for quantum chemistry with a fully differentiable framework," J. Chem. Theory Comput. **21**(13), 6505–6516 (2025).
[39] J. Westermayr and R. J. Maurer, "Physically inspired deep learning of molecular excitations and photoemission spectra," Chem. Sci. **12**, 10755–10764 (2021).
[40] C. Qian, V. Vitartas, J. Kermode, and R. J. Maurer, "Equivariant electronic Hamiltonian prediction with many-body message passing," arXiv:2508.15108 (2025).
[41] Y. Zhou, H. Wang, Y. Liu, X. Gao, and H. Song, "Applicability of Kerker preconditioning scheme to the self-consistent density functional theory calculations of inhomogeneous systems," Phys. Rev. E **97**, 033305 (2018).
[42] P. Pulay, "Convergence acceleration of iterative sequences. The case of SCF iteration," Chem. Phys. Lett. **73**, 393–398 (1980).
[43] C. G. Broyden, "A class of methods for solving nonlinear simultaneous equations," Math. Comput. **19**, 577–593 (1965).
[44] G. P. Kerker, "Efficient iteration scheme for self-consistent pseudopotential calculations," Phys. Rev. B **23**, 3082 (1981).
[45] Fritz Haber Institute ab initio materials simulations: FHI-aims. A Users Guide, 2025, https://fhi-aims.org/uploads/documents/FHI-aims.250822.pdf.
[46] I. Batatia *et al.*, "A foundation model for atomistic materials chemistry," J. Chem. Phys. **163**, 184110 (2025).
[47] D. J. Wales and M. P. Hodges, "Global minima of water clusters (H2O)n, $n \leq 21$, described by an empirical potential," Chem. Phys. Lett. **286**, 65–72 (1998).
[48] D. J. Wales and M. P. Hodges, TIP4P (H_2O)N clusters, 1998, https://www-wales.ch.cam.ac.uk/~wales/CCD/TIP4P-water.html.
[49] E. Rudberg, E. H. Rubensson, P. Sałek, and A. Kruchinina, "Ergo: An open-source program for linear-scaling electronic structure calculations," SoftwareX **7**, 107–111 (2018).
[50] D. Spångberg, XYZ—Water clusters, 2018, http://www.ergoscf.org/xyz/h2o.php.
[51] B. Hourahine, B. Aradi, V. Blum, F. Bonafe, A. Buccheri, C. Camacho, C. Cevallos, M. Deshaye, T. Dumitrică, A. Dominguez, S. Ehlert *et al.*, "DFTB+, a software package for efficient approximate density functional theory based atomistic simulations," J. Chem. Phys. **152**, 124101 (2020).
[52] B. Hourahine, M. Berdakin, J. Bich, F. Bonafé, C. Camacho, Q. Cui, M. Deshaye, G. Díaz Mirón, S. Ehlert, M. Elstner *et al.*, "Recent developments in DFTB+, a software package for efficient atomistic quantum mechanical simulations," J. Phys. Chem. A **129**, 5373 (2025).
[53] M. Modrzejewski, S. Yourdkhani, S. Śmiga, and J. Klimeš, "Random-phase approximation in many-body noncovalent systems: Methane in a dodecahedral water cage," J. Chem. Theor. Comput. **17**, 804–817 (2021).
[54] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**, 3865–3868 (1996).
[55] E. van Lenthe, J. G. Snijders, and E. J. Baerends, "The zero-order regular approximation for relativistic effects: The effect of spin–orbit coupling in closed shell molecules," J. Chem. Phys. **105**, 6505–6516 (1996).
[56] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes *et al.*, "The atomic simulation environment—A Python library for working with atoms," J. Phys.: Condens. Matter **29**, 273002 (2017).
[57] A. Stukowski, "Visualization and analysis of atomistic simulation data with OVITO–The open visualization tool," Modell. Simul. Mater. Sci. Eng. **18**, 015012 (2009).
[58] K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller, "SchNetPack: A deep learning toolbox for atomistic systems," J. Chem. Theor. Comput. **15**, 448–455 (2018).
[59] É. Cancès, G. Kemlin, and A. Levitt, "Convergence analysis of direct minimization and self-consistent iterations," SIAM J. Matrix Anal. Appl. **42**, 243–274 (2021).
[60] N. D. Woods, M. C. Payne, and P. J. Hasnip, "Computing the self-consistent field in Kohn–Sham density functional theory," J. Phys.: Condens. Matter **31**, 453001 (2019).
[61] O. T. Hofmann, E. Zojer, L. Hörmann, A. Jeindl, and R. J. Maurer, "First-principles calculations of hybrid inorganic–organic interfaces: From state-of-the-art to best practice," Phys. Chem. Chem. Phys. **23**, 8132–8180 (2021).