# COMPLETE NEURAL ELECTRONIC INITIALIZATION ACCELERATES MATERIALS DFT

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

We present the first complete machine learning method for accelerating plane-wave density functional theory (DFT) in materials under the projector augmented wave (PAW) formalism. We formalize seven criteria that a *Complete Neural Electronic Initializer* must satisfy for practical end-to-end PAW DFT acceleration. Applying these to prior work reveals two structure-dependent components, augmentation occupancies and spin initialization, whose absence prevents existing acceleration methods from providing complete reference-free initialization. We show that omitting these components can eliminate or reverse the acceleration obtained via models that only predict the smooth valence density. We satisfy the missing requirements by introducing AugNet, a general equivariant model for PAW augmentation occupancies, and the first general spin density model for materials, which predicts the smooth spin-difference density and spin-difference PAW augmentation occupancies using predicted magnetic moments to constrain the global magnetic state. Combined with existing valence density models, our full method satisfies all seven criteria and forms a fully reference-free electronic initializer for materials DFT, requiring no electronic quantities from a converged target calculation. We show that perfect initialization could cut PAW DFT wall time by 40-52%, and our method recovers up to 62% of this saving, reducing end-to-end DFT wall time by up to $\sim 25\%$ on unseen structures while preserving converged energies.

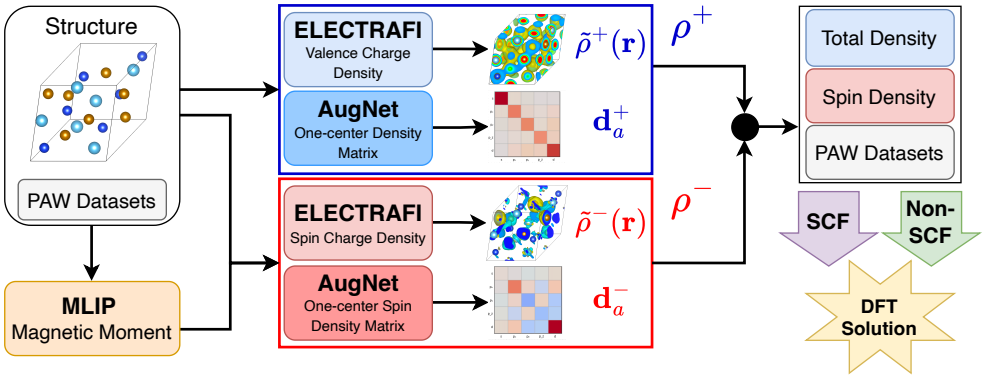



Figure 1: Workflow for end-to-end PAW DFT acceleration, illustrated using the models employed in this work. For a given atomic structure and PAW setup, the fixed PAW datasets define the frozen core and basis information, while a Complete Neural Electronic Initializer provides the three structure-dependent components defined in Table 1: the smooth valence density $\tilde{\rho}^+$, PAW augmentation occupancies $\mathbf{d}_a^+$, and spin/magnetic initialization. ELECTRAFI predicts $\tilde{\rho}^+$, AugNet predicts $\mathbf{d}_a^+$, and the corresponding spin-dependent components $\tilde{\rho}^-$ and $\mathbf{d}_a^-$ are predicted by spin-ELECTRAFI and spin-AugNet, with magnetic moment predictions constraining the global magnetic state. Together, these components enable complete reference-free initialization of PAW DFT.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

# 1 INTRODUCTION & MOTIVATION

Density functional theory (DFT) is a central computational tool in materials science, chemistry, and condensed matter physics that enables modeling of atomic and electronic properties across large chemical spaces (Hohenberg & Kohn, 1964; Kohn & Sham, 1965; Jain et al., 2013; Gavini et al., 2023). DFT calculations account for up to 45% of core hours on the UK ARCHER2 Tier-1 supercomputer and over 70% of allocation time within the materials science sector at NERSC (Riebesell et al., 2025), and generating the 118 million inorganic structures in OMat24 required more than 400 million CPU core hours of DFT calculations (Barros-Luque et al., 2026). At this scale, even single-digit percentage reductions in DFT cost translate to tens of millions of CPU core hours saved. The importance of DFT has only increased as large DFT datasets have become the basis for machine-learned interatomic potentials (MLIPs), universal atomistic models, and materials discovery pipelines (Batzner et al., 2022; Batatia et al., 2022; Chen & Ong, 2022; Deng et al., 2023; Batatia et al., 2025; Qu & Krishnapriyan, 2024; Neumann et al., 2024). For high-throughput computational materials science, plane-wave DFT with the projector augmented wave (PAW) formalism is the de facto standard. VASP is the dominant PAW implementation (Blöchl, 1994; Kresse & Joubert, 1999) and underlies many of the field's canonical materials datasets, including the Materials Project, OQMD, AFLOW, JARVIS-DFT, GNoME, and OMat24 (Jain et al., 2013; Kirklin et al., 2015; Curtarolo et al., 2012; Choudhary et al., 2020; Merchant et al., 2023; Barros-Luque et al., 2026). In PAW DFT, the quality of the initial electronic quantities matters because better initialization can reduce the number of self-consistent field (SCF) cycles required for convergence. Charge density is one such input, and has motivated machine learning (ML) models that predict charge densities directly from atomic structure (Jørgensen & Bhowmik, 2022; Kim & Ahn, 2024; Cheng & Peng, 2024; Koker et al., 2024; Fu et al., 2024; Elsborg et al., 2026a; Klockow et al., 2026; Elsborg et al., 2026b). Unlike MLIPs, which accelerate atomistic simulation by approximating the DFT potential energy surface, these models target the cost of obtaining the DFT solution itself, without replacing the underlying electronic structure model. However, most of these studies evaluate only density prediction accuracy, assumed to be a proxy for acceleration potential (Kim & Ahn, 2024; Cheng & Peng, 2024; Fu et al., 2024; Klockow et al., 2026). More importantly, even studies that evaluate DFT acceleration directly do not take into account that charge density is not the only required electronic input quantity and therefore does not constitute the complete PAW electronic state. The experiments are only feasible because crucial components are retained from converged reference calculations (Jørgensen & Bhowmik, 2022; Koker et al., 2024; Elsborg et al., 2026a;b). Thus, such experiments isolate the quality of the learned smooth valence density, but do not represent a real ML-accelerated DFT workflow, since the retained converged quantities are unavailable for new calculations. A notable exception is Sunshine et al. (2023), who found no acceleration over VASP's default initialization using a predicted valence density and PAW augmentation occupancies from a zero-step VASP calculation, identifying augmentation and wavefunction initialization as remaining bottlenecks. Our experiments support this conclusion and show that even a converged valence density provides little acceleration when the remaining PAW components are poorly initialized.

To clarify the distinction, we formalize seven criteria that must be met to demonstrate practical, reference-free ML-driven acceleration of PAW DFT in materials. We refer to such a method as a *Complete Neural Electronic Initializer*. Such a method must provide three structure-dependent initialization components: **C1**, the smooth valence density $\tilde{\rho}^{+}$; **C2**, the PAW augmentation occupancies $\mathbf{d}_a^{+}$; and **C3**, spin/magnetic initialization, including the spin-dependent components $\tilde{\rho}^{-}$ and $\mathbf{d}_a^{-}$. A complete demonstration must also establish **C4**, density accuracy; **C5**, SCF acceleration; **C6**, controlled component ablations; and **C7**, reference-free end-to-end wall time reduction including ML inference. Prior methods satisfy two or at most three of these seven criteria, and no prior method has been published that jointly addresses general PAW augmentation, magnetic initialization, component importance, and corresponding reference-free end-to-end acceleration. We summarize our criteria jointly with the state of the field in Table 1.

**Contributions.** We present the first Complete Neural Electronic Initializer, satisfying all criteria in Table 1. Specifically, we address the three criteria no prior general method satisfies, magnetic initialization (**C3**), controlled component ablations (**C6**), and reference-free end-to-end acceleration (**C7**), and evaluate predicted PAW augmentation (**C2**) as an SCF initializer. Our contributions are:

Table 1: Criteria for complete neural electronic initialization for accelerating materials DFT. **C1-C3** are structure-dependent quantities. **C4-C7** refer to the required evaluation for demonstrating complete initialization. ✓ denotes a capability in the cited work, × denotes that it was not demonstrated. † denotes system-specific models: CJM (Focassio et al., 2024) is specific to $MoS_2$, while de Blasio et al. (2023) and EAC-Net (Qin et al., 2026) predict spin densities for $Na_3V_2(PO_4)_3$ and Fe, respectively, without demonstrating SCF acceleration.

| | Reference-free initialization components | | | Demonstrated evaluation | | | |
|---|---|---|---|---|---|---|---|
| Method | Valence density C1 | PAW augmentation C2 | Spin / magnetic initialization C3 | Density accuracy C4 | SCF acceleration C5 | Component ablations C6 | Reference-free acceleration C7 |
| GPWNO (Kim & Ahn, 2024) | ✓ | × | × | ✓ | × | × | × |
| InfGCN (Cheng & Peng, 2024) | ✓ | × | × | ✓ | × | × | × |
| SCDP (Fu et al., 2024) | ✓ | × | × | ✓ | × | × | × |
| BOA (Klockow et al., 2026) | ✓ | × | × | ✓ | × | × | × |
| EdenGNN (Li et al., 2025) | ✓ | ✓ | × | ✓ | × | × | × |
| DeepDFT (Jørgensen & Bhowmik, 2022) | ✓ | × | × | ✓ | ✓ | × | × |
| ChargE3Net (Koker et al., 2024) | ✓ | × | × | ✓ | ✓ | × | × |
| ELECTRA (Elsborg et al., 2026a) | ✓ | × | × | ✓ | ✓ | × | × |
| ELECTRAFI (Elsborg et al., 2026b) | ✓ | × | × | ✓ | ✓ | × | × |
| NASICON model (de Blasio et al., 2023) | ✓ † | × | ✓ † | ✓ † | × | × | × |
| CJM (Focassio et al., 2024) | ✓ † | ✓ † | × | ✓ † | × | × | × |
| EAC-Net (Qin et al., 2026) | ✓ | × | ✓ † | ✓ | × | × | × |
| **This work** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

1. **We establish the requirements for reference-free initialization.** We show that valence, augmentation, and spin initialization components must all be treated explicitly for practical SCF acceleration, and identify the contribution of each.
2. **We introduce AugNet, a general equivariant model for PAW augmentation occupancies.** AugNet satisfies **C2** by predicting structured one-center augmentation coefficients across diverse materials, elements, and PAW schemas, including the spin-difference channel, and is to our knowledge the first such model evaluated as an SCF initializer.
3. **We introduce the first general model for spin density prediction.** Using a charge-informed CHGNet model to constrain the ELECTRAFI model's density prediction, we enable direct prediction of spin difference densities, satisfying **C3**.
4. **We demonstrate the first reference-free ML acceleration of PAW DFT.** By combining all components, we present a method that satisfies all requirements in Table 1 and reduces total DFT wall time by up to $\sim 25\%$ on unseen structures.

## 2 Background & Related Work

**DFT & PAW.** Density functional theory (DFT) is the standard first-principles framework for electronic structure calculations in materials (Hohenberg & Kohn, 1964; Kohn & Sham, 1965). In periodic systems, Kohn-Sham DFT is commonly solved in a plane-wave basis through self-consistent field (SCF) iteration of the electronic density (Payne et al., 1992; Kresse & Furthmüller, 1996). The projector augmented wave (PAW) method (Blöchl, 1994) enables efficient plane-wave DFT by replacing the rapidly varying all-electron wavefunctions near the nuclei with smooth pseudo wavefunctions, while restoring the missing atom-centered information through one-center corrections. The PAW decomposition can be written as

$$\rho^{+}(\mathbf{r}) = \tilde{\rho}^{+}(\mathbf{r}) + \sum_{a} \left[\rho^{a,+}(\mathbf{r}) - \tilde{\rho}^{a,+}(\mathbf{r})\right], \tag{1}$$

where $\tilde{\rho}^{+}(\mathbf{r})$ is the spin-summed smooth valence density represented on the plane-wave grid, while $\rho^{a,+} - \tilde{\rho}^{a,+}$ restores the atom-centered all-electron information removed by the smoothing proce-

dure. While the basis functions defining $\rho^{a,+} - \tilde{\rho}^{a,+}$ are fixed by the PAW dataset, their coefficients depend on the electronic state of the material. In VASP (Kresse & Furthmüller, 1996; Kresse & Joubert, 1999), these structure-dependent coefficients are stored as augmentation occupancies. Thus, a complete ML initialization method must predict the PAW augmentation occupancies to satisfy **C2**. EdenGNN (Li et al., 2025) predicts PAW augmentation occupancies for general non-magnetic materials, but uses them to build a non-self-consistent Hamiltonian rather than to initialize SCF, and does not treat spin. CJM also directly predicts PAW augmentation occupancies, but is system-specific to $MoS_2$ and does not evaluate SCF acceleration (Focassio et al., 2024). Spin-polarized PAW initialization additionally requires the smooth spin-difference density $\tilde{\rho}^{-}$ and corresponding augmentation occupancies $\mathbf{d}_a^-$. Prior materials spin density models are likewise system-specific (de Blasio et al., 2023; Qin et al., 2026), with no demonstration of SCF acceleration. Further details on VASP representation, initialization procedure, and influence of individual components are in Appendix B. We provide details on augmentation occupancies in Appendix C, and on spin-polarized and magnetic calculations in Appendix D.

**Charge density prediction.** Nearly all existing charge density models predict the $\tilde{\rho}^{+}(\mathbf{r})$ valence density term in Equation 1. Models differ mainly in how they represent the map from atomic structure $\mathcal{X} = (Z_i, \mathbf{R}_i)_{i=1}^{N}$ to $\tilde{\rho}^{+}(\mathbf{r})$. The state-of-the-art in the field is ELECTRAFI (Elsborg et al., 2026a) and ChargE3Net (Koker et al., 2024). ELECTRAFI extends ELECTRA's (Elsborg et al., 2026a) floating Gaussians to materials by analytically transforming predicted floating Gaussians into reciprocal-space coefficients and reconstructing the density through inverse FFT (Elsborg et al., 2026b). This avoids dense real-space neural evaluation and results in low inference cost. ChargE3Net achieves higher grid accuracy, but requires neural evaluation across the real-space grid and explicit periodic treatment (Koker et al., 2024). Its high inference cost therefore decreases the resulting wall time benefit (Elsborg et al., 2026b). A broader overview of architectures, including related initialization methods outside the general periodic materials setting, is provided in Appendix A.

**Requirements for complete neural electronic initialization in materials.** ML models that predict only the smooth valence density $\tilde{\rho}^{+}(\mathbf{r})$ of Equation 1 can, at most, satisfy **C1**, **C4**, and **C5**. Existing SCF acceleration studies largely follow the evaluation protocol introduced by Jørgensen & Bhowmik (2022), in which only $\tilde{\rho}^{+}(\mathbf{r})$ is replaced by an ML prediction. Their initialization is therefore effectively

$$\rho^{+}_{\text{init}}(\mathbf{r}) = \underbrace{\tilde{\rho}^{+}_{\text{ML}}(\mathbf{r})}_{\textbf{C1}} + \underbrace{\sum_a \left[\rho^{a,+}_{\text{test}}(\mathbf{r}) - \tilde{\rho}^{a,+}_{\text{test}}(\mathbf{r})\right]}_{\text{converged augmentation from the same test structure}}, \tag{2}$$

with converged spin-dependent quantities likewise retained for spin-polarized calculations. These quantities are unavailable for a genuinely new calculation, so such methods do not satisfy **C2** or **C3**. Moreover, without isolating the contribution of these retained quantities they do not satisfy **C6**. The absence of **C2**, **C3**, and **C6** therefore precludes a reference-free end-to-end demonstration satisfying **C7**. We discuss these limitations in more detail in Appendix A.2.

## 3 Methods

**Valence and augmentation density.** PAW augmentation contributions are strongly localized and atom-centered, whereas the smooth valence density is spatially extended and captures interatomic density (Blöchl, 1994; Kresse & Joubert, 1999). We therefore model them separately. We use ELECTRAFI (Elsborg et al., 2026b) and ChargE3Net (Koker et al., 2024) for the smooth valence density required by **C1**, and introduce AugNet below to model the PAW augmentation occupancy prediction required by **C2**.

**AugNet: Augmentation occupancy prediction (C2).** PAW augmentation occupancies are the finite coefficient representation of the one-center correction $\rho^{a,+} - \tilde{\rho}^{a,+}$ in Equation 1. They define a variable-schema equivariant prediction problem, see Appendix F for a proof. For atom $a$ with PAW schema $s_a = s(Z_a)$, the target space is

$$\mathcal{V}_{s_a} = \bigoplus_L n_L^{(s_a)} D^L, \qquad F_\theta(\mathcal{N}_a, s_a) : \mathcal{N}_a \mapsto \mathbf{d}_a^{+} \in \mathcal{V}_{s_a}, \tag{3}$$

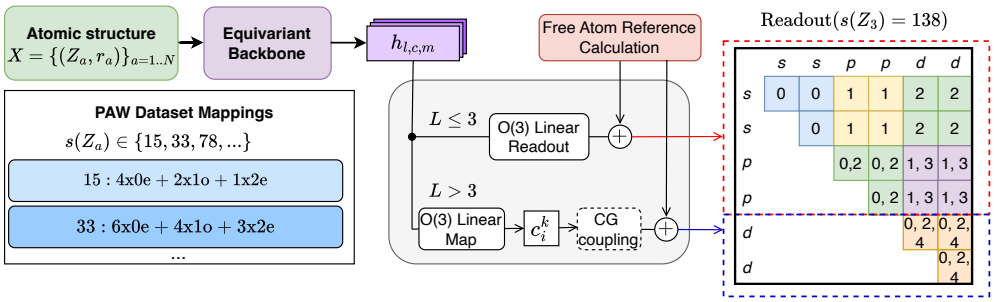



Figure 2: AugNet architecture. An equivariant backbone, schema-conditioned readout, and Clebsch-Gordan coupling predict on-site PAW coefficients. See Appendix C for details.

Here, $D^L$ is the $(2L+1)$-dimensional irrep of $SO(3)$ and $\mathbf{d}_a^+ = \{d_{a,ij}^{LM,+}\}_{ijLM}$ is the packed spin-summed augmentation occupancy vector, where $i, j$ index PAW partial-wave channels and $M = -L, \ldots, L$. Together with the fixed PAW basis,

$$\rho^{a,+}(\mathbf{r}) - \tilde{\rho}^{a,+}(\mathbf{r}) = \sum_{ij,L,M} d_{a,ij}^{LM,+} B_{ij}^{a,LM}(\mathbf{r}), \tag{4}$$

where $B_{ij}^{a,LM}$ is fixed by the PAW dataset. Because both the multiplicities $n_L^{(s_a)}$ and maximum $L$ depend on the PAW schema ($L \leq 6$ in our data), AugNet must map atomic environments to element-dependent output representations. We use a shared equivariant backbone and schema-conditioned readout. For the spin-summed channel, AugNet predicts corrections to VASP's default augmentation occupancies based on superposition of atomic densities (SAD),

$$\widehat{\Delta \mathbf{d}}_a^+ = \mathbf{G}_{s(Z_a)} \mathcal{R}(\mathbf{h}_a), \qquad \hat{\mathbf{d}}_a^+ = \mathbf{d}_a^{+,\mathrm{SAD}} + \widehat{\Delta \mathbf{d}}_a^+, \tag{5}$$

where $\mathcal{R}$ is the shared equivariant readout applied to the atom-wise backbone representation $\mathbf{h}_a$, and $\mathbf{G}_{s(Z_a)}$ gathers the coefficients required by the PAW schema of element $Z_a$. Backbone-supported angular channels use equivariant linear maps, while higher-order PAW components are constructed by Clebsch-Gordan coupling,

$$\widehat{\Delta d}_{a,ij}^{LM,+} = \sum_k w_{(ij,L),k} \sum_{m_i,m_j} C_{\ell_i m_i, \ell_j m_j}^{LM} c_{a,i,m_i}^{(k)} c_{a,j,m_j}^{(k)}, \qquad L > L_{\mathrm{backbone}}, \tag{6}$$

where $\ell_i, m_i$ and $\ell_j, m_j$ label the angular components of the partial-wave channels, $C_{\ell_i m_i, \ell_j m_j}^{LM}$ are Clebsch–Gordan coefficients, $k$ indexes learned projection channels, and $c_{a,i,m_i}^{(k)}$ are learned equivariant projections. Figure 2 summarizes the architecture.

AugNet is trained with a masked coefficient space loss over the valid PAW channels,

$$\mathcal{L}_{\mathrm{Aug}} = \frac{\sum_{a,\alpha} m_{a,\alpha} \left( \hat{d}_{a,\alpha}^+ - d_{a,\alpha}^+ \right)^2}{\sum_{a,\alpha} m_{a,\alpha}}, \tag{7}$$

where $\alpha$ indexes the $(i, j, L, M)$ coefficients. Implementation details are in Appendix C.

**Spin & magnetic modeling (C3).** C3 requires initializing the spin-dependent electronic state without access to a converged magnetization density. One option is to predict magnetic moments for each atom and pass these to VASP through `MAGMOM`. VASP then uses these magnetic moments to initialize a spin-polarized calculation before the spin density is updated self-consistently. We test this method using the charge-informed model CHGNet (Deng et al., 2023). Second, the smooth spin-difference density $\tilde{\rho}^-(\mathbf{r}) = \tilde{\rho}_\uparrow(\mathbf{r}) - \tilde{\rho}_\downarrow(\mathbf{r})$, can be predicted directly from atomic structure. Thus, we construct a spin-adapted version of ELECTRAFI, spin-ELECTRAFI, that reuses the Gaussian parameters of the valence density ELECTRAFI model to learn a second set of signed weights,

$$\hat{\tilde{\rho}}^-(\mathbf{r}) = \sum_g w_g^- \phi_g(\mathbf{r}; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g). \tag{8}$$

Table 2: Component ablation for PAW DFT initialization. The matrix specifies the electronic components in each experiment, with SCF step savings relative to the default SAD initialization reported separately for non-magnetic and magnetic Materials Project structures. ✓ denotes converged (Oracle) initialization, × denotes SAD/default initialization. mm denotes spin initialization using atomic magnetic moments.

| | Spin channel | Aug. only | Valence only | Default | Valence + aug. | mm only | Smooth grids | Val. + aug. + mm | Oracle |
|---|---|---|---|---|---|---|---|---|---|
| Valence density $\tilde{\rho}^{+}$ | × | × | ✓ | × | ✓ | × | ✓ | ✓ | ✓ |
| Spin density $\tilde{\rho}^{-}$ | ✓ | × | × | × | × | mm ✓ | ✓ | mm ✓ | ✓ |
| Valence aug. $\mathbf{d}^{+}$ | × | ✓ | × | × | ✓ | × | × | ✓ | ✓ |
| Spin aug. $\mathbf{d}^{-}$ | ✓ | ✓ | × | × | × | mm ✓ | × | mm ✓ | ✓ |
| Non-mag. [%] | −9.7 | −14.3 | +0.5 | 0 | +13.5 | +12.0 | +10.2 | +47.6 | +49.0 |
| Mag. [%] | −70.7 | −65.0 | −29.4 | 0 | −11.0 | −3.1 | −4.2 | +24.6 | +55.4 |

We similarly construct spin-AugNet using the same equivariant architecture to predict the spin-difference PAW augmentation occupancies $\mathbf{d}_a^{-} = \{d_{a,ij}^{LM,-}\}_{ijLM}$ directly, using a zero reference rather than the free-atom SAD reference. However, the net spin moment provides a global constraint on the spin-difference density, so a third hybrid option factorizes the predicted smooth spin density into a global magnetic state and a normalized spatial distribution,

$$\hat{q}_\theta(\mathbf{r} \mid X) = \frac{\hat{\tilde{\rho}}^{-}_{\mathrm{raw}}(\mathbf{r} \mid X)}{\int_\Omega \hat{\tilde{\rho}}^{-}_{\mathrm{raw}}(\mathbf{r} \mid X)\, dV}, \qquad \hat{\tilde{\rho}}^{-}(\mathbf{r} \mid X) = \hat{M}_{\mathrm{CHGNet}}(X)\, \hat{q}_\theta(\mathbf{r} \mid X). \tag{9}$$

We can therefore use CHGNet to model the global magnetic state and constrain the high-dimensional spatial distribution predicted by spin-ELECTRAFI. During training, we use the ground-truth magnetic moment $M_{\mathrm{DFT}}$ to constrain the spin density, and replace it with $\hat{M}_{\mathrm{CHGNet}}(X)$ at inference. The model is trained jointly on $\tilde{\rho}^{+}$ and $\tilde{\rho}^{-}$ through the combined loss $\mathcal{L} = \mathcal{L}_{\tilde{\rho}^{+}} + \lambda_{\mathrm{spin}}\mathcal{L}_{\tilde{\rho}^{-}}$, where $\lambda_{\mathrm{spin}}$ is a hyperparameter. The loss is adapted to magnetic and non-magnetic structures as detailed in Appendix D. We compare all three approaches in Section 4.

## 4 Experiments

The Complete Neural Electronic Initializer in Figure 1 satisfies the three initialization criteria **C1**-**C3**. We now perform the experiments required to demonstrate **C4**-**C7**.

**Component ablations (C6).** We isolate the contribution of each PAW initialization component to SCF convergence, using the same Materials Project Jain et al. (2013) (MP) densities and structures evaluated in Koker et al. (2024) and Elsborg et al. (2026b). As shown in Table 2, the "Valence only" setting directly exposes the limitation of prior approaches: even with the converged valence density $\tilde{\rho}^{+}$, leaving augmentation and spin at their default values provides no benefit for non-magnetic structures and substantially worsens magnetic calculations. This is consistent with the conclusion of Sunshine et al. (2023), who found no practical acceleration when combining an ML valence density prediction with augmentation occupancies from a zero-step VASP DFT calculation. Acceleration therefore requires PAW augmentation and spin initialization (**C2**-**C3**) for practical acceleration. The Oracle setting uses only converged quantities to set a practical upper bound on achievable acceleration: $49.0\%$ and $55.4\%$ SCF step reduction for non-magnetic and magnetic structures, respectively. Comparing valence + augmentation initialization $(\tilde{\rho}^{+}, \mathbf{d}^{+})$ with Oracle isolates the importance of spin initialization, since adding the spin-dependent components recovers much of the remaining acceleration for both non-magnetic and magnetic structures. Full results are provided in Appendix B and Table 6.

**Magnetic initialization.** Table 3 compares the three magnetic initialization strategies introduced in Section 3: CHGNet magnetic moments, explicit spatial initialization using spin-ELECTRAFI and spin-AugNet, and the hybrid model combining CHGNet-constrained spin-ELECTRAFI with spin-AugNet. We compare against Oracle spin channel initialization $(\tilde{\rho}^{-}, \mathbf{d}^{-})$ and Oracle magnetic moments to isolate the acceleration available from each representation. Oracle results show that

Table 3: Spin-difference density accuracy and SCF step reduction for magnetic initializations using Oracle and ML valence and augmentation on the Materials Project test set.

| Magnetic initialization | Subset | MAE | Oracle valence + augmentation | ELECTRAFI + AugNet | ChargE3Net + AugNet |
|---|---|---|---|---|---|
| Oracle $(\tilde{\rho}^-, \mathbf{d}^-)$ | Non-mag. | – | 49.0% | 23.3% | 29.1% |
| | Magnetic | – | 55.4% | 29.3% | 31.9% |
| Oracle moments | Non-mag. | 0.028 | 47.6% | 22.6% | 29.0% |
| | Magnetic | 4.658 | 24.7% | 6.2% | 9.5% |
| **Models** | | | | | |
| CHGNet moments | Non-mag. | 0.347 | 38.7% | 19.4% | 25.9% |
| | Magnetic | 5.662 | 20.5% | 1.7% | 3.8% |
| spin-ELECTRAFI + spin-AugNet | Non-mag. | 0.194 | 40.8% | 18.1% | 24.4% |
| | Magnetic | 5.950 | 12.8% | -1.0% | 0.1% |
| Hybrid + spin-AugNet | Non-mag. | 0.316 | 38.8% | 18.7% | 23.9% |
| | Magnetic | 2.870 | 24.2% | 10.2% | 12.0% |

magnetic moments recover most of the available acceleration for non-magnetic structures, but substantially less for magnetic structures. Using CHGNet moments results in the same overall picture. Direct spatial initialization with spin-ELECTRAFI and spin-AugNet provides the required spin-dependent density representation, but these predictions are inaccurate and lead to less acceleration for magnetic systems, particularly when coupled with the ML valence and augmentation methods. In our hybrid model, constraining spin-ELECTRAFI with CHGNet reduces magnetic prediction error significantly, while spin-AugNet supplies the corresponding spin-difference augmentation occupancies $\mathbf{d}^-$, recovering a larger fraction of the available acceleration while retaining performance on non-magnetic structures. Our Complete Neural Electronic Initializer in Figure 1 therefore uses the hybrid model. Details on the magnetic initialization models and experiments are provided in Appendix D, with hyperparameters in G.

**AugNet performance (C4–C5).** We test AugNet's ability to improve DFT initialization by training on progressively larger Materials Project subsets and evaluating the non-magnetic MP and GNoME test sets of Elsborg et al. (2026b). Figure 3 shows that increasing the training set reduces RMSE on both datasets and improves SCF convergence. Pairing AugNet models with Oracle valence density or ELECTRAFI shows that lower augmentation error translates into better initialization with both converged and learned densities. Scaling saturates earlier on MP, while GNoME benefits from additional data. The full model reaches an augmentation MAE/RMSE of $0.0041/0.0118$ on MP and $0.0062/0.0262$ on GNoME (Table 4). For context, CJM reports $0.0130/0.0459$ MAE/RMSE on its system-specific $MoS_2$ dataset (Focassio et al., 2024), and EdenGNN an MAE of $0.0085$ on non-magnetic MP structures recomputed with different VASP settings (Li et al., 2025). Neither is a matched benchmark, since datasets, settings and aggregation differ, but both indicate the scale of coefficient-space errors. AugNet can also be efficiently adapted to the $MoS_2$ PAW setup through fine-tuning (Appendix C.9).

**Complete Neural Electronic Initializer and reference-free acceleration (C7).** We finally evaluate the ability of our Complete Neural Electronic Initializer to provide reference-free acceleration. We use AugNet and the hybrid magnetic model from Section 3, and combine them with either ELECTRAFI (Elsborg et al., 2026b) or ChargE3Net (Koker et al., 2024) as the valence density model. Figure 4 shows the effect of the Complete Neural Electronic Initializer (CNEI) on total wall time for both the non-magnetic and magnetic subsets of MP and GNoME. As in Elsborg et al. (2026b), we report the wall time reductions adjusted for ML inference time, and compare to the Default and Oracle wall time numbers in Table 5 across the full test sets. We include Oracle acceleration captured (OAC), which is the fraction of the reduction achieved by Oracle that is recovered by ML initialization:

$$\mathrm{OAC}_{\mathrm{ML}} = \frac{\text{ML wall-time saving (\%)}}{\text{Oracle wall-time saving (\%)}} \times 100\%. \tag{10}$$

The initializer reduces total wall time by $15.04\%$ on MP and $25.17\%$ on GNoME using ELECTRAFI as the valence backbone (CNEI-EFI), corresponding to $\mathrm{OAC}_{\mathrm{CNEI-EFI}}(\mathrm{MP}) = 29.15\%$

Table 4: PAW augmentation occupancy prediction errors. CJM is evaluated on its system-specific $MoS_2$ structure. AugNet is evaluated on the MP and GNoME test data.

| **Model:** | CJM | AugNet | AugNet | spin-AugNet | spin-AugNet |
|---|---|---|---|---|---|
| **Data:** | $MoS_2$ | MP | GNoME | MP | GNoME |
| MAE ↓ | 0.0130 | 0.0041 | 0.0062 | 0.0025 | 0.0020 |
| RMSE ↓ | 0.0459 | 0.0119 | 0.0262 | 0.0098 | 0.0129 |

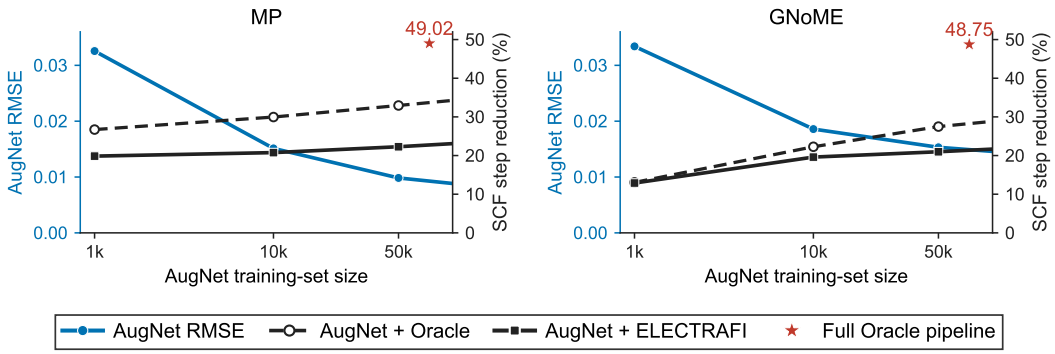



Figure 3: AugNet scaling with training set size on MP and OOD GNoME, using either the converged Oracle or the ELECTRAFI model for valence density initialization.

and $\mathrm{OAC}_{\mathrm{CNEI-EFI}}(\mathrm{GNoME}) = 62.25\%$. Using ChargE3Net as the valence model (CNEI-C3Net) produces larger reductions in DFT execution time, but its inference cost limits end-to-end savings to $6.18\%$ and $7.92\%$ ($\mathrm{OAC}_{\mathrm{CNEI-C3Net}}(\mathrm{MP}) = 11.99\%$ and $\mathrm{OAC}_{\mathrm{CNEI-C3Net}}(\mathrm{GNoME}) = 19.59\%$). The AugNet and magnetic models add virtually no overhead, so the valence density model dominates inference cost. For both CNEI-EFI and CNEI-C3Net, Figure 4 shows that the room for improvement is largest on magnetic structures, which are not accelerated as much as non-magnetic structures. The lower panel of Figure 4 shows that learned initializations do not alter the converged outcome relative to either Default or Oracle initialization, with all four methods reaching the lowest observed energy at similar rates. On magnetic GNoME structures, CNEI reaches it more often than Default ($89.2\%$ against $85.1\%$), suggesting that learned spin initialization can also steer SCF toward lower-energy magnetic states. Full numerical results are in Appendix E.

## 5 Discussion & Limitations

Our results show that practical PAW initialization for materials DFT acceleration is a multi-component prediction problem. Valence density alone is insufficient, and augmentation and magnetic initialization are necessary to achieve reference-free acceleration. Our Complete Neural Electronic Initializer is, to our knowledge, the first method to achieve end-to-end acceleration in spin-polarized PAW DFT without any electronic quantity from a converged target calculation. However, better training metrics are needed. The GNoME evaluations have larger density errors, yet the wall time reduction is larger (Tables 4-5 and Figure 4), so current density metrics are imperfect proxies for solver performance. Solver-aware approaches such as Eberhard et al. (2026) are promising for explicitly optimizing execution time, but difficult to apply to VASP because they require access to solver gradients. Furthermore, the predicted smooth spin-difference density and spin-difference PAW augmentation occupancies are less accurate than the smooth valence density predictions, and because CHGNet predicts moment magnitudes, the hybrid model cannot distinguish antiferromagnetic from ferromagnetic order. Table 2 shows that improving these components is a clear route towards closing the gap to Oracle initialization. New charge-informed models for magnetic property prediction could aid this development Li et al. (2023); Xu et al. (2025). Predicted electronic quantities also depend on the material distribution and the PAW and DFT setup. Future models could explicitly encode functionals, pseudopotentials, and other DFT settings to enable broader transfer. Alternatively, as shown for AugNet in Appendix C.9, models can be adapted to new PAW setups through transfer learning. Figure 4 shows that ML initialization reaches the lowest converged ener-

Table 5: End-to-end performance of our Complete Neural Electronic Initializer (CNEI), using either ELECTRAFI (EFI) or ChargE3Net (C3Net) as the valence backbone. Total time includes ML initialization and DFT execution. SCF step savings are reported relative to Default. Oracle acceleration captured (OAC) is calculated as defined in Equation 10.

| Dataset | Metric | Default | Oracle | CNEI-EFI | CNEI-C3Net |
|---|---|---|---|---|---|
| **MP** | NMAE [%] ↓ | – | – | 0.58 | 0.54 |
| | SCF steps ↓ | 22.05 | 10.41 | 19.05 | 18.36 |
| | SCF steps saved [%] ↑ | – | 52.78 | 13.62 | 16.76 |
| | Total time [s] ↓ | 623.84 | 302.04 | 530.04 | 585.26 |
| | Total time saved [%] ↑ | – | 51.58 | 15.04 | 6.18 |
| | OAC [%] ↑ | 0.0 | 100.0 | **29.15** | 11.99 |
| **GNoME** | NMAE [%] ↓ | – | – | 0.93 | 0.69 |
| | SCF steps ↓ | 16.30 | 7.89 | 11.87 | 11.45 |
| | SCF steps saved [%] ↑ | – | 51.63 | 27.17 | 29.79 |
| | Total time [s] ↓ | 188.99 | 112.59 | 141.43 | 174.02 |
| | Total time saved [%] ↑ | – | 40.43 | 25.17 | 7.92 |
| | OAC [%] ↑ | 0.0 | 100.0 | **62.25** | 19.59 |

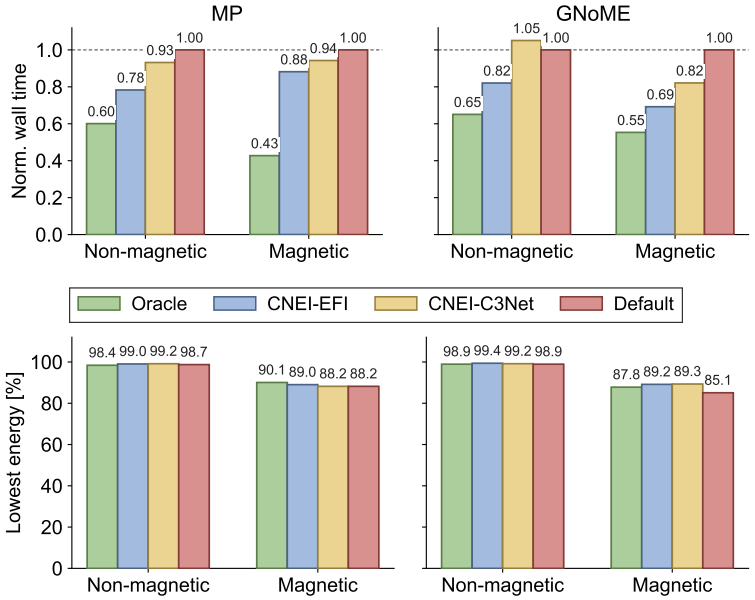



Figure 4: **Top:** Wall-time comparison of Default, Oracle, and our Complete Neural Electronic Initializer (CNEI), using ELECTRAFI (EFI) or ChargE3Net (C3Net) for valence density predictions. Wall time is normalized relative to Default $= 1.0$. Values $< 1$ indicate acceleration, while values $> 1$ indicate slowdown. **Bottom:** The proportion of calculations for each method that are within 1 meV of the lowest energy recorded for any method.

gies at the same rate as default DFT initialization, so we view DFT acceleration as complementary to improving MLIPs. The learned model changes only the initialization, while the final energy is still obtained by solving the original DFT problem. This can therefore accelerate DFT calculations where surrogates are not sufficient, as well as the generation of data for increasingly accurate surrogates. The gains apply wherever a calculation starts without a converged density, such as single points on new or perturbed structures, which dominate non-equilibrium datasets like OMat24 (Barros-Luque et al., 2026), and the first ionic step of a relaxation. We expect smaller benefits for later relaxation steps, since these can already restart from the previous step's density. Our results therefore establish the components and evaluation required for complete neural electronic initialization in PAW-based periodic DFT for materials and show that complete, reference-free initialization is achievable. We hope this encourages new models for all components required for the complete initializer, since our ablations show that each one limits the acceleration that can be reached.

### AI USE STATEMENT

Generative AI tools were used to aid and polish writing, assist with literature retrieval and discovery, support research ideation and execution, and help draft portions of the manuscript. They were also used to assist with code development and figure preparation. Generative AI was not used to generate synthetic datasets or to prove mathematical claims.

All AI-assisted text, citations, scientific claims, code, and analyses have been reviewed and verified by the authors against literature, calculations, and results. The authors take full responsibility for the final content of the paper, including all text, claims, code, and artifacts produced with the aid of generative AI.

## ETHICS STATEMENT

This work is computational and does not involve human subjects, personal data, or sensitive information. We are not aware of any ethical risks associated with the methodology or its intended use.

### REPRODUCIBILITY STATEMENT

The methods and evaluation protocol are described in Sections 3-4, with additional details provided in the appendix. Appendix B specifies the VASP settings, initialization procedures, and baselines. Appendix C describes the PAW augmentation representation and AugNet architecture, training objective, and evaluation. Appendix D details the architecture of the magnetic initialization models. Appendix G reports experimental hardware and model hyperparameters. Code for reproducing the models and experiments is available in an anonymous repository at `https://zenodo.org/records/22938056`.

## REFERENCES

Luis Barros-Luque, Muhammed Shuaibi, Xiang Fu, Brandon M Wood, Misko Dzamba, Meng Gao, Ammar Rizvi, Matt Uyttendaele, C Lawrence Zitnick, and Zachary W Ulissi. The open materials 2024 (omat24) inorganic materials dataset and models. *Nature Computational Science*, pp. 1–11, 2026.

Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in Neural Information Processing Systems*, 35:11423–11436, 2022.

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M Elena, Dávid P Kovács, Janosh Riebesell, Xavier R Advincula, Mark Asta, Matthew Avaylon, William J Baldwin, et al. A foundation model for atomistic materials chemistry. *The Journal of chemical physics*, 163(18), 2025.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1): 2453, 2022.

Peter E Blöchl. Projector augmented-wave method. *Physical review B*, 50(24):17953, 1994.

Craig Calcaterra and Axel Boldt. Approximating with gaussians. *arXiv preprint arXiv:0805.3795*, 2008.

Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, 2022.

Chaoran Cheng and Jian Peng. Equivariant neural operator learning with graphon convolution. *Advances in Neural Information Processing Systems*, 36, 2024.

Kamal Choudhary and Brian DeCost. Atomistic line graph neural network for improved materials property predictions. *npj Computational Materials*, 7(1), November 2021. ISSN 2057-3960. doi: 10.1038/s41524-021-00650-1. URL http://dx.doi.org/10.1038/s41524-021-00650-1.

Kamal Choudhary, Kevin F Garrity, Andrew CE Reid, Brian DeCost, Adam J Biacchi, Angela R Hight Walker, Zachary Trautt, Jason Hattrick-Simpers, A Gilad Kusne, Andrea Centrone, et al. The joint automated repository for various integrated simulations (jarvis) for data-driven materials design. *npj computational materials*, 6(1):173, 2020.

Stefano Curtarolo, Wahyu Setyawan, Shidong Wang, Junkai Xue, Kesong Yang, Richard H Taylor, Lance J Nelson, Gus LW Hart, Stefano Sanvito, Marco Buongiorno-Nardelli, et al. Aflowlib. org: A distributed materials properties repository from high-throughput ab initio calculations. *Computational Materials Science*, 58:227–235, 2012.

Paolo Vincenzo Freiesleben de Blasio, Peter Bjørn Jorgensen, Juan Maria Garcia Lastra, and Arghya Bhowmik. Nanosecond md of battery cathode materials with electron density description. *Energy Storage Materials*, 63:103023, 2023.

Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J Bartel, and Gerbrand Ceder. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023.

Eike S. Eberhard, Viktor Kotsev, Timm Güthle, and Stephan Günnemann. Transferable scf-acceleration through solver-aligned initialization learning, 2026. URL https://arxiv.org/abs/2604.21657.

Jonas Elsborg, Luca Thiede, Alán Aspuru-Guzik, Tejs Vegge, and Arghya Bhowmik. Electra: A cartesian network for 3d charge density prediction with floating orbitals. *Advances in Neural Information Processing Systems*, 38:28092–28121, 2026a.

Jonas Elsborg, Felix Ærtebjerg, Luca Thiede, Alán Aspuru-Guzik, Tejs Vegge, and Arghya Bhowmik. Global plane waves from local gaussians: Periodic charge densities in a blink, 2026b. URL https://arxiv.org/abs/2601.19966.

Pol Febrer, Peter Bjørn Jørgensen, Miguel Pruneda, Alberto García, Pablo Ordejón, and Arghya Bhowmik. Graph2mat: universal graph to matrix conversion for electron density prediction. *Machine Learning: Science and Technology*, 6(2):025013, 2025.

Bruno Focassio, Michelangelo Domina, Urvesh Patil, Adalberto Fazzio, and Stefano Sanvito. Covariant jacobi-legendre expansion for total energy calculations within the projector augmented wave formalism. *Physical Review B*, 110(18):184106, 2024.

Xiang Fu, Andrew Rosen, Kyle Bystrom, Rui Wang, Albert Musaelian, Boris Kozinsky, Tess Smidt, and Tommi Jaakkola. A recipe for charge density prediction. *Advances in Neural Information Processing Systems*, 37:9727–9752, 2024.

Vikram Gavini, Stefano Baroni, Volker Blum, David R Bowler, Alexander Buccheri, James R Chelikowsky, Sambit Das, William Dawson, Pietro Delugas, Mehmet Dogan, et al. Roadmap on electronic structure codes in the exascale era. *Modelling and Simulation in Materials Science and Engineering*, 31(6):063301, 2023.

Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1), 2013.

Peter Bjørn Jørgensen and Arghya Bhowmik. Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids. *npj Computational Materials*, 8(1):183, 2022.

Manasa Kaniselvan, Benjamin Kurt Miller, Meng Gao, Juno Nam, and Daniel S Levine. Learning from the electronic structure of molecules across the periodic table. *arXiv preprint arXiv:2510.00224*, 2025.

Seongsu Kim and Sungsoo Ahn. Gaussian plane-wave neural operator for electron density estimation. *arXiv preprint arXiv:2402.04278*, 2024.

Seongsu Kim, Nayoung Kim, Dongwoo Kim, and Sungsoo Ahn. High-order equivariant flow matching for density functional theory hamiltonian prediction. *Advances in Neural Information Processing Systems*, 38:13265–13307, 2026a.

Seongsu Kim, Chanhui Lee, Yoonho Kim, Seongjun Yun, Honghui Kim, Nayoung Kim, Changyoung Park, Sehui Han, Sungbin Lim, and Sungsoo Ahn. Machine learning hamiltonians are accurate energy-force predictors. *arXiv preprint arXiv:2602.16897*, 2026b.

Scott Kirklin, James E Saal, Bryce Meredig, Alex Thompson, Jeff W Doak, Muratahan Aykol, Stephan Rühl, and Chris Wolverton. The open quantum materials database (oqmd): assessing the accuracy of dft formation energies. *npj Computational Materials*, 1(1):15010, 2015.

Manuel V Klockow, Marc K Ickler, Peter Lippmann, and Fred A Hamprecht. A function-centric graph neural network approach for predicting electron densities. In *The Fourteenth International Conference on Learning Representations*, 2026.

Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

Teddy Koker, Keegan Quigley, Eric Taw, Kevin Tibbetts, and Lin Li. Higher-order equivariant neural networks for charge density prediction in materials. *npj Computational Materials*, 10(1): 161, 2024.

Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B*, 54(16):11169, 1996.

Georg Kresse and Daniel Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical review b*, 59(3):1758, 1999.

He Li, Zechen Tang, Xiaoxun Gong, Nianlong Zou, Wenhui Duan, and Yong Xu. Deep-learning electronic-structure calculation of magnetic superstructures. *Nature Computational Science*, 3: 321–327, 2023. doi: 10.1038/s43588-023-00424-3.

Xiwen Li, Zaizhou Xin, Hongyu Yu, Yang Zhong, Xingao Gong, and Hongjun Xiang. Efficient e (3)-equivariant framework for universal charge density prediction. *arXiv preprint arXiv:2510.00788*, 2025.

Amil Merchant, Simon Batzner, Samuel S Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, 2023.

Mark Neumann, James Gin, Benjamin Rhodes, Steven Bennett, Zhiyi Li, Hitarth Choubisa, Arthur Hussey, and Jonathan Godwin. Orb: A fast, scalable neural network potential. *arXiv preprint arXiv:2410.22570*, 2024.

Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013. ISSN 0927-0256. doi: https://doi.org/10.1016/j.commatsci.2012.10.028. URL https://www.sciencedirect.com/science/article/pii/S0927025612006295.

Mike C Payne, Michael P Teter, Douglas C Allan, TA Arias, and ad JD Joannopoulos. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Reviews of modern physics*, 64(4):1045, 1992.

Xuejian Qin, Taoyuze Lv, and Zhicheng Zhong. Eac-net: Predicting real-space charge density via equivariant atomic contributions. *Journal of Chemical Theory and Computation*, 22(9):4813–4821, 2026.

Eric Qu and Aditi Krishnapriyan. The importance of being scalable: Improving the speed and accuracy of neural network interatomic potentials across chemical domains. *Advances in Neural Information Processing Systems*, 37:139030–139053, 2024.

Janosh Riebesell, Rhys EA Goodall, Philipp Benner, Yuan Chiang, Bowen Deng, Gerbrand Ceder, Mark Asta, Alpha A Lee, Anubhav Jain, and Kristin A Persson. A framework to evaluate machine learning crystal stability predictions. *Nature Machine Intelligence*, 7(6):836–847, 2025.

Feitong Song and Ji Feng. Neural network self-consistent fields for density functional theory. *npj Computational Materials*, 2026.

Ethan M Sunshine, Muhammed Shuaibi, Zachary W Ulissi, and John R Kitchin. Chemical properties from graph neural network-predicted electron densities. *The Journal of Physical Chemistry C*, 127 (48):23459–23466, 2023.

Brandon M. Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Gharakhanyan, John R. Kitchin, Daniel S. Levine, Kyle Michel, Anuroop Sriram, Taco Cohen, Abhishek Das, Ammar Rizvi, Sushree Jagriti Sahoo, Zachary W. Ulissi, and C. Lawrence Zitnick. Uma: A family of universal models for atoms, 2026. URL https://arxiv.org/abs/2506.23971.

Wenbin Xu, Rohan Yuri Sanspeur, Adeesh Kolluru, Bowen Deng, Peter Harrington, Steven Farrell, Karsten Reuter, and John R. Kitchin. Spin-informed universal graph neural networks for simulating magnetic ordering. *Proceedings of the National Academy of Sciences*, 122(27):e2422973122, 2025. doi: 10.1073/pnas.2422973122.

Zeyi Zhang, Carlos Mora Perez, Patrick Kwon, Martin Head-Gordon, and Jin Qian. Parsec. py: A python-based real-space kohn–sham density functional theory code accelerated by machine learned charge density. *Journal of Computational Chemistry*, 47(23):e70482, 2026.

# APPENDIX

## TABLE OF CONTENTS

**A Charge density models** **15**

A.1 Prior work . . . 15

A.2 Limitations of prior work and models . . . 16

**B VASP SCF Experiments** **19**

B.1 VASP Settings . . . 19

B.2 Magnetic Moment Initialization . . . 20

B.3 CHGNet Initialization . . . 20

B.4 Baselines . . . 20

B.5 Spin-restricted DFT . . . 21

**C PAW Augmentation and the AugNet Architecture** **23**

C.1 PAW Augmentation Occupancies as Covariant Targets . . . 23

C.2 Augmentation Schemas and Equivariance . . . 23

C.3 MACE backbone . . . 23

C.4 Schema-agnostic Equivariant Readout . . . 24

C.5 Reference baselines and $\Delta$-learning . . . 25

C.6 Coefficient conventions . . . 25

C.7 Training objective . . . 25

C.8 Evaluation metrics . . . 26

C.9 AugNet Accuracy and Transfer . . . 26

**D Magnetic Initialization Development** **28**

**E Detailed End-to-End DFT Results** **30**

**F Equivariance of PAW augmentation occupancies** **32**

**G Experiment Setup and Hyperparameters** **33**

G.1 Experimental Hardware . . . 33

G.2 AugNet . . . 33

G.3 ELECTRAFI . . . 34

# A CHARGE DENSITY MODELS

## A.1 PRIOR WORK

Machine learning charge density models learn a map from an atomic structure $\mathcal{X} = \{(Z_i, \mathbf{R}_i)\}_{i=1}^{N}$ to the smooth valence density $\tilde{\rho}^{+}(\mathbf{r})$. They differ primarily in whether the spatial dependence is represented implicitly or explicitly. Probe-based models evaluate the density by conditioning a neural network directly on each query point,

$$\hat{\tilde{\rho}}^{+}(\mathbf{r}) = f_\theta(\mathbf{r}; \mathcal{X}), \tag{11}$$

whereas basis-based models first predict coefficients or basis parameters and then evaluate an explicit expansion,

$$\hat{\tilde{\rho}}^{+}(\mathbf{r}) = \sum_k c_k(\mathcal{X})\, \phi_k(\mathbf{r}; \mathcal{X}). \tag{12}$$

This distinction is architectural rather than fundamental, since probe models simply use an implicit, query-dependent basis, while basis models make the spatial representation explicit. Different approaches such as InfGCN similarly learn maps from atomic structure to continuous smooth valence density fields, but can be viewed as basis/operator variants of the same underlying problem (Cheng & Peng, 2024).

DeepDFT introduced the probe-based formulation(Jørgensen & Bhowmik, 2022), and ChargE3Net replaces the DeepDFT backbone with a higher-order $E(3)$-equivariant architecture to better capture angular structure in periodic materials (Koker et al., 2024). These approaches are flexible since they avoid choosing an explicit basis, but they are computationally intensive since evaluating a full density grid requires many query point evaluations.

SCDP uses a spherical Gaussian basis centered on both atoms and equivariantly placed virtual centers,

$$\hat{\tilde{\rho}}^{+}(\mathbf{r}) = \sum_{a \in \mathcal{A} \cup \mathcal{V}} \sum_{j\ell m} c_{aj\ell m}\, \Phi_{\alpha_{aj},\ell,m,\mathbf{R}_a}(\mathbf{r}). \tag{13}$$

where $\mathcal{A}$ denotes atoms and $\mathcal{V}$ virtual centers (Fu et al., 2024). SCDP shows that non-atom-centered orbital bases can lead to higher accuracy, and Klockow et al. (2026) achieved a similar result in BOA by representing the density through products of atom-centered basis functions,

$$\hat{\tilde{\rho}}^{+}(\mathbf{r}) = \sum_{(a,b)} \sum_{\mu,\nu} \Gamma_{ab\mu\nu}\, \omega_\mu^{Z_a}(\mathbf{r} - \mathbf{R}_a)\omega_\nu^{Z_b}(\mathbf{r} - \mathbf{R}_b). \tag{14}$$

which resembles a local density matrix expansion and naturally places density between atoms.
ELECTRA takes a different explicit representation approach by replacing spherical-harmonic orbital expansions with a mixture of anisotropic floating 3D Gaussians (Elsborg et al., 2026a),

$$\hat{\tilde{\rho}}^{+}(\mathbf{r}) = \sum_{j=1}^{N_G} w_j\, \mathcal{N}(\mathbf{r}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j). \tag{15}$$

Each component has a signed weight $w_j$, a learned center $\boldsymbol{\mu}_j$, and a positive-definite covariance $\boldsymbol{\Sigma}_j$. This ansatz uses the fact that Gaussian mixtures are universal approximators of smooth densities(Calcaterra & Boldt, 2008). Compared with atom- or bond-centered orbital expansions, the basis function positions are not fixed, and are instead predicted as displacements from atoms,

$$\boldsymbol{\mu}_j = \mathbf{R}_{a(j)} + \mathbf{d}_j, \tag{16}$$

This removes the need for high-order spherical harmonics, resulting in inference speeds that are orders of magnitude faster than prior models.

**Periodic and reciprocal-space models.** For periodic materials, the physically natural objective is to represent densities in a way that mirrors plane-wave DFT, where periodic scalar fields are represented through reciprocal-space coefficients. Kim & Ahn (2024) explore this by adding a plane-wave branch to a Gaussian type orbital (GTO) model, i.e.,

$$\hat{\tilde{\rho}}^{+}(\mathbf{r}) = \hat{\tilde{\rho}}^{+}_{\mathrm{GTO}}(\mathbf{r}) + \hat{\tilde{\rho}}^{+}_{\mathrm{PW}}(\mathbf{r}), \tag{17}$$

but this yields only modest gains compared to the GTO-only model, and performs poorly on its own.

However, ELECTRAFI showed that it is possible to extend the floating Gaussian representation of ELECTRA to periodic materials by making reciprocal space the central construction(Elsborg et al., 2026b). ELECTRAFI predicts an auxiliary non-periodic representation of $\tilde{\rho}^+$ similar to Equation 15, and then exploits the closed-form analytical Fourier transform of each Gaussian to obtain plane-wave coefficients,

$$\hat{\tilde{\rho}}^+(\mathbf{G}) = \sum_{j=1}^{N_G} w_j \exp\left(-\frac{1}{2}\mathbf{G}^\top \mathbf{\Sigma}_j \mathbf{G}\right) e^{-i\mathbf{G}\cdot\boldsymbol{\mu}_j}. \tag{18}$$

The periodic real-space density is then recovered with a single inverse FFT,

$$\hat{\tilde{\rho}}^+(\mathbf{r}) = \mathrm{IFFT}\left[\hat{\tilde{\rho}}^+(\mathbf{G})\right](\mathbf{r}). \tag{19}$$

Since periodicity and global Fourier structure are imposed analytically through the Poisson summation formula, ELECTRAFI's representation is the one most closely aligned with plane-wave DFT. While ChargE3Net has sufficient flexibility to achieve competitive accuracy on periodic materials, its inference times are comparable in magnitude to the DFT calculation time itself when using standard functionals (Elsborg et al., 2026b). The main reasons are the dense neural evaluation of every real-space grid point and the explicit summation over periodic images of the unit cell. ELECTRAFI's construction avoids both of these and achieves drastically faster inference, which also translates into end-to-end acceleration of DFT workflows, albeit still using converged properties from the reference data.

**Other related work.** A complementary line of work predicts electronic quantities in localized orbital representations. Graph2Mat predicts equivariant density matrices directly from atomic structure (Febrer et al., 2025), while QHFlow and QHFlow2 learn Kohn-Sham Hamiltonians (Kim et al., 2026a;b). QHFlow also demonstrates SCF acceleration by using the predicted Hamiltonian directly to initialize a DFT calculation, and HELM similarly targets Hamiltonian prediction across broader chemical and basis set spaces (Kaniselvan et al., 2025). These methods use Hamiltonian or density matrices directly in DFT frameworks formulated in localized orbital bases. NeuralSCF (Song & Feng, 2026) instead learns the Kohn-Sham density map itself and iterates the learned map to self-consistency. PARSEC.py (Zhang et al., 2026) uses ML-predicted charge densities to initialize self-consistent Kohn-Sham calculations in a real-space finite-difference pseudopotential framework. In this representation, the predicted density can be supplied directly on the real-space grid, whereas plane-wave PAW codes such as VASP require a smooth density together with the corresponding structure-dependent PAW augmentation and, for spin-polarized calculations, spin-dependent components. These approaches therefore address related ways of reducing the cost of self-consistent electronic structure calculations using other frameworks than the periodic PAW representation considered here.

### A.2 Limitations of prior work and models

Prior charge density models have demonstrated that an accurate prediction of the smooth valence density can reduce the number of SCF iterations required by VASP (Jørgensen & Bhowmik, 2022; Koker et al., 2024; Elsborg et al., 2026a;b). However, these experiments do not constitute complete initialization of a new PAW calculation from atomic structure alone, since they are reference-dependent for augmentation and spin quantities.

For clarity, the spin-summed and spin-difference valence densities can be written as

$$\rho^+(\mathbf{r}) = \rho_\uparrow(\mathbf{r}) + \rho_\downarrow(\mathbf{r}), \qquad \rho^-(\mathbf{r}) = \rho_\uparrow(\mathbf{r}) - \rho_\downarrow(\mathbf{r}). \tag{20}$$

For each channel, the PAW decomposition has the form

$$\rho^\pm(\mathbf{r}) = \tilde{\rho}^\pm(\mathbf{r}) + \sum_a \left[\rho^{a,\pm}(\mathbf{r}) - \tilde{\rho}^{a,\pm}(\mathbf{r})\right], \tag{21}$$

where $\tilde{\rho}^\pm$ is the smooth plane-wave component and the second term is determined by the corresponding PAW augmentation occupancies.

**Converged augmentation in prior SCF experiments.** Existing generalized charge density models used for SCF acceleration predict only the smooth spin-summed valence density $\tilde{\rho}^{+}$, following the SCF acceleration protocol initially introduced by Jørgensen & Bhowmik (2022) and subsequently adopted by later work. The initial density is therefore effectively

$$\rho^{+}_{\mathrm{init}}(\mathbf{r}) = \underbrace{\tilde{\rho}^{+}_{\mathrm{ML}}(\mathbf{r})}_{\text{predicted from structure}} + \underbrace{\sum_{a} \left[\rho^{a,+}_{\mathrm{test}}(\mathbf{r}) - \tilde{\rho}^{a,+}_{\mathrm{test}}(\mathbf{r})\right]}_{\text{converged augmentation from the same test structure}} . \tag{22}$$

Thus, although the smooth density is predicted, the PAW augmentation occupancies are not. They are taken from the already converged reference DFT calculation of the exact structure whose subsequent SCF acceleration is being measured. These quantities are therefore unavailable when DFT is run from scratch on a genuinely new structure.

This distinction is important because augmentation is not a fixed quantity that can simply be obtained from the PAW dataset. The dataset specifies the partial waves, projectors, and allowed angular channels, whereas the augmentation occupancies depend on the converged electronic state of the material. Transferring them from the reference calculation provides target-specific electronic information beyond the ML-predicted valence density.

A reference-free approximation is not necessarily sufficient either. Sunshine et al. (2023) obtained PAW augmentation occupancies from a VASP calculation with zero electronic minimization steps and combined them with an ML-predicted valence density, but found no acceleration over VASP's default initialization and concluded that the approach had no practical value at the time. They identified augmentation and wavefunction initialization as remaining bottlenecks, consistent with our ablations showing that even a converged valence density provides little acceleration when the remaining PAW components are poorly initialized.

The limitation for spin is different. Prior generalized charge density models do not predict $\tilde{\rho}^{-}$, nor do they predict the corresponding spin-difference PAW augmentation occupancies. Consequently, prior SCF-acceleration studies do not test ML initialization for general magnetic structures. Instead, their acceleration experiments are restricted to structures classified as non-magnetic.

However, this restriction does not eliminate the spin-dependent electronic state as long as the underlying reference calculations are performed using default spin-polarized settings (`ISPIN=2` in VASP). A structure can have a small net magnetic moment while still possessing a nonzero converged spin-difference density. For such calculations, the spin channel inherited from the reference calculation can be written schematically as

$$\rho^{-}_{\mathrm{init}}(\mathbf{r}) = \underbrace{\tilde{\rho}^{-}_{\mathrm{test}}(\mathbf{r})}_{\text{converged spin density}} + \underbrace{\sum_{a} \left[\rho^{a,-}_{\mathrm{test}}(\mathbf{r}) - \tilde{\rho}^{a,-}_{\mathrm{test}}(\mathbf{r})\right]}_{\text{converged spin augmentation}}, \tag{23}$$

where both terms are taken from the converged DFT solution of the same test structure rather than from an ML prediction.

Prior work does not use a dedicated spin density model for initialization experiments, which therefore effectively prohibits general spin-polarized calculations. Magnetic structures are not evaluated as a general reference-free acceleration problem, while even the nominally non-magnetic test calculations can retain converged spin-dependent information. An additional downside to this absence is that magnetic calculations offer the highest potential for acceleration, as we have demonstrated in this work (see Table 2).

**What is required for reference-free initialization.** A genuinely reference-free PAW initializer must instead construct all structure-dependent electronic components without access to a converged calculation of the test structure. In the collinear setting considered here, this requires

$$\underbrace{\tilde{\rho}^{+}_{\mathrm{ML}}}_{\text{valence density}} , \quad \underbrace{\{\mathbf{d}^{+}_{a,\mathrm{ML}}\}_a}_{\text{augmentation}}, \quad \underbrace{\tilde{\rho}^{-}_{\mathrm{ML}}}_{\text{spin density}} , \quad \underbrace{\{\mathbf{d}^{-}_{a,\mathrm{ML}}\}_a}_{\text{spin augmentation}} , \tag{24}$$

all obtained from the atomic structure and quantities available before the DFT calculation begins. Here, $\mathbf{d}^{+}$ and $\mathbf{d}^{-}$ denote the spin-summed and spin-difference PAW augmentation occupancies, respectively.

We directly isolate these dependencies by varying which electronic components are supplied at initialization and evaluating the resulting VASP convergence on the Materials Project test structures used in Elsborg et al. (2026b). Table 6 reports the full results of these component ablations, showing that:

- **The previously reported benefit of valence density prediction depends strongly on converged PAW augmentation.** When the converged augmentation occupancies are removed, much of the SCF acceleration attributed to the predicted smooth density disappears or can reverse. Accurate valence density prediction alone is therefore insufficient for practical PAW acceleration.
- **Spin initialization constitutes a second, independent requirement.** Even nominally non-magnetic `ISPIN=2` structures benefit from initialization of their spin-dependent electronic state, while magnetic structures show an even larger dependence on accurate spin initialization. Prior charge density models do not address this problem and consequently do not establish acceleration for general magnetic materials.
- **Complete electronic initialization exposes substantially larger acceleration potential.** When the valence, augmentation, and spin-dependent components are all initialized accurately, the number of SCF iterations can be reduced far beyond what is achievable from valence density prediction alone. This motivates learning the previously missing augmentation and spin components directly from structure.

These observations identify the two missing modeling problems addressed in this work. AugNet predicts the spin-summed and spin-difference PAW augmentation occupancies, removing the need to transfer converged augmentation information from the test calculation. Separately, our charge informed spin density model uses CHGNet magnetic moment predictions to constrain ELECTRAFI and directly predicts the smooth spin-difference density $\tilde{\rho}^{-}$ required for spin-polarized initialization. Together with a valence density model, these components make it possible to initialize all structure-dependent electronic quantities from the atomic structure alone.

# B VASP SCF EXPERIMENTS

## B.1 VASP SETTINGS

All SCF calculations in this work are performed as single-point (static) calculations and differ from the corresponding default VASP calculations only in the choice of initial electron density, since we use densities predicted by machine learning models instead of the default superposition of atomic densities (SAD). To ensure a controlled comparison, we retain the parameters of the reference calculations and modify only the tags required for density initialization and output formatting. All calculations are performed with VASP 5.4.4 using the same legacy PBE PAW datasets employed in the original Materials Project calculations.

**Materials Project.** For every `mp-` identifier we retrieve the exact task document that produced the reference charge density via the Materials Project API and save its `POSCAR`, `INCAR`, `KPOINTS` and `POTCAR`. The structure comes from `input.structure`, the $k$-mesh from `input.kpoints`, and the pseudopotentials are reconstructed from `input.potcar_spec` so that the POTCAR titles match the reference run element for element. Consequently, the plane-wave cutoff, the exchange-correlation functional and Hubbard-$U$ set, the smearing scheme, the electronic convergence criterion, the $k$-point mesh, and the projector set are exactly the Materials Project values for that material.

**GNoME.** For GNoME calculations we use the same calculation scheme as Koker et al. (2024); Elsborg et al. (2026b), using the `pymatgen` (Ong et al., 2013) `MPStatic` parameter set. We further apply the same DFT+$U$ `LMAXMIX` treatment that Koker et al. (2024) applied. Specifically, if DFT+$U$ is used and there are any $f$-orbitals (`LDAUL`$= 3$) in the system, we set `LMAXMIX=6`, and if there are $d$-orbitals (`LDAUL`$= 2$) we set `LMAXMIX=4` (excluding the case of present $f$-electrons).

**Experiments.** For various experiments we override the INCAR to match our experimental goal. The following list summarizes the basic parameters:

- `ICHARG`$= 2$ is the atomic superposition (SAD) baseline; `ICHARG`$= 1$ reads the seed density from `CHGCAR`.
- `ISTART=0` starts the wavefunctions from scratch
- `LCHARG=True` generates the CHGCARs upon completion
- `NPAR`, `NCORE`, `KPAR`, `NSIM` are parallelization parameters that are removed, defaulting the calculation to simply use all the cores of the specified CPU. It also ensures that all calculations are given the same resources.

The spin-restricted and magnetic moment experiments require further specifications which are listed below:

- **`ISPIN`$= 1$ controls.** The spin channel is removed entirely: `ISPIN` is set to 1 and the spin-only tags `MAGMOM` and `NUPDOWN` are dropped so VASP never consults them. The seed CHGCAR is correspondingly rebuilt with only the spin-summed ($+$) channel.
- **Uniform `MAGMOM` seeds.** With `ISPIN`$= 2$ and a seed carrying the converged $+$-channel components $(\tilde{\rho}^{+}, \mathbf{d}_a^{+})$ but no spin-difference channel, VASP builds the initial magnetization from `MAGMOM`. We replace MP'sper-species values with a single uniform value $m$ on every atom to limit the steps a spin-unrestricted calculation needs for non-magnetic materials. `NUPDOWN` is left as MP set it.
- **Spin-mixing tags.** `NUPDOWN`, `AMIX_MAG` and `BMIX_MAG` can be overridden per run to test whether the residual spin channel cost is a mixing problem; these overrides are applied last and are otherwise inactive.

We note that if the augmentation occupancies are not formatted correctly for the given VASP version, then VASP will silently default to SAD initializations derived from the atom types and `MAGMOM`, removing any benefits gained from a better initial guess..

**Filtering.** Like previous works (Elsborg et al., 2026b), we use the magnetization filter specified by (Koker et al., 2024) to distinguish between magnetic and non-magnetic structures. The definition of the magnetic label is an absolute total magnetic moment of below $0.1\mu_B$ and that all atomic absolute magnetic moments are below $0.1\mu_B$. For GNoME, we additionally filter out 60 structures that contain Yb, an element not present in MP dataset and therefore without a trained augmentation occupancy model. Additionally, excluding structures with convergence problems results in 951 final structures from the MP-Full test set and 1245 from GNoME.

**Normalization.** VASP smooth valence density grids $\tilde{\rho}^+$ are always normalized to the number of valence electrons defined by the PAW dataset. As this is a predictable property that can help convergence, we normalize the ChargE3Net input predictions. A caveat is that ChargE3Net predictions is already capable of capturing the total charge within $0.1\%$ accuracy (measured on the MP non-magnetic testset). With this method, we achieve 0.05 saved steps on MP on average, i.e., a negligible difference, which was also observed by (Elsborg et al., 2026b).

### B.2 Magnetic Moment Initialization

In this work, we define a structure to be non-magnetic if the absolute total magnetic moment $|\mu_B| < 0.1$ *and* each individual magnetic moment $|\mu_{i,B}| < 0.1$. As discussed further in appendix B.4, DFT calculations initialized with no spin-difference components take longer to converge even on non-magnetic structures, despite having converged $+$-channel components $(\tilde{\rho}^+, \mathbf{d}_a^+)$. There are three ways of providing a spin configuration guess, either setting atomic magnetic moments that are expanded into a real space SAD guess by the DFT code, or by directly initializing the smooth spin difference density $\tilde{\rho}^-$ and its PAW augmentation occupancies $\mathbf{d}_a^-$. While both can be modeled with machine learning, they pose significant challenges, respectively, with the latter remaining an open challenge in the field and outside the scope of this work.

While workflows differ between DFT codes, magnetic moments are typically initialized based on some upfront observations of the structure such as the element type and the oxidation state. For atoms deemed magnetic, the initial moment is set very high to elucidate good convergence behavior, whereas non-magnetic atoms are initialized closer to 0. This way, a calculation search the potential energy surface by decreasing the magnetic moment rather than increasing it, a much harder task. As a standard of the field, the Materials Project workflow first performs two DFT relaxations with the `MPRelaxSet` (as given in `pymatgen`), shifting atom positions into more favorable positions and optimizing toward an initial guess for the spin state. Afterwards, a static DFT calculation is performed with `MPStaticSet` that determines the energy. This process has proven to be robust and results in well behaved energies and magnetic moments during data generation. However, relaxing a structure twice this way also costs more HPC resources.

### B.3 CHGNet Initialization

Instead of relaxing twice with DFT, practitioners could also employ one of the modern MLIPs (Batatia et al., 2022; Neumann et al., 2024; Qu & Krishnapriyan, 2024; Wood et al., 2026) to perform the structure relaxations, but that still leaves the magnetic moments themselves. Previous work (Choudhary & DeCost, 2021; Deng et al., 2023; Xu et al., 2025) has tackled the this problem and instead use MLIPs trained on the converged magnetic moments to predict them. In particular, CHGNet (Deng et al., 2023) is trained on Materials Project and has demonstrated itself to be useful (Xu et al., 2025). To use it, we simply load the pretrained `0.3.0` version the authors provide in their repository and evaluate it on our MP and GNoME test sets. On MP, CHGNet has an on-site MAE of $0.055$ $\mu_B$ and $0.060$ $\mu_B$ alongside a total mag. mom. MAE of $0.716$ $\mu_B$ and $0.472$ $\mu_B$, good accuracies both in and out of domain. A limitation of CHGNet is the choice to predict absolute values, thereby making it impossible for the model to distinguish between ferromagnetic and antiferromagnetic spin, but since the latter is only a small part of MP and GNoME, it remains the best model choice.

### B.4 Baselines

Initialization of the smooth valence density $\tilde{\rho}^+$ in a PAW DFT calculation is not meaningful without the corresponding augmentation occupancies $\mathbf{d}_a^+$ that determine the on-site density correction. This

point was also raised by Elsborg et al. (2026b). Furthermore, this does not include the smooth spin-difference density $\tilde{\rho}^-$ or its corresponding spin-difference augmentation occupancies $\mathbf{d}_a^-$ required by a spin-polarized DFT calculation (`ISPIN=2`), which is used throughout the Materials Project and is also standard in similar datasets. To evaluate the effects of including different components for initialization, we recalculated the test set of Materials Project used by (Elsborg et al., 2026b) with different schemes. The results are shown in Table 6, and has several noteworthy aspects.

As reflected by the additional 7 % reduction observed for the magnetic subset relative to the non-magnetic subset of the evenly split test set, magnetic calculations can benefit even more from accurate initialization, owing to their generally slower convergence. Below the Oracle results, initialization with only the smooth valence density $\tilde{\rho}^+$ represents the practically achievable setting corresponding to previous approaches(Jørgensen & Bhowmik, 2022; Koker et al., 2024; Elsborg et al., 2026a;b) when used as presented. Since these frameworks do not predict the augmentation occupancies $\mathbf{d}_a^+$ or the spin-dependent components $\tilde{\rho}^-$ and $\mathbf{d}_a^-$ that could lead to faster convergence, the relative non-magnetic reduction of 0.6% is effectively just a default VASP run.

The next two rows show that initializing the complete $+$ density channel, i.e., $\tilde{\rho}^+$ together with $\mathbf{d}_a^+$, gives a 10% reduction on average for non-magnetic materials without any spin information, but fails otherwise. The spin-dependent components $\tilde{\rho}^-$ and $\mathbf{d}_a^-$ alone are also insufficient for SCF step reduction.

Finally, the last four rows show the benefits of initializing magnetic moments together with either default or converged $+$-channel components $(\tilde{\rho}^+, \mathbf{d}_a^+)$. The former corresponds to a typical DFT calculation in the MP workflow, showing that a relaxation- or ML-derived magnetic moment guess is beneficial at least for non-magnetic structures. The latter demonstrates that a good spin state guess combined with converged-accuracy $+$-channel components provides the second-best initialization in the table. Thus, even without explicitly initializing $\tilde{\rho}^-$ and $\mathbf{d}_a^-$, we can achieve approximately 40% and 20% SCF step reductions for non-magnetic and magnetic structures, respectively. Between MPRelaxSet and CHGNet the difference is relatively small.

### B.5 SPIN-RESTRICTED DFT

To evaluate whether the benefits of machine learning-based initialization persist in the absence of spin coupling, we evaluate the non-magnetic Materials Project test set using the same computational parameters as previous work (Koker et al., 2024; Elsborg et al., 2026a;b), with the sole exception of setting `ISPIN=1`. The resulting SCF reductions are reported in Table 7. The results show that `ISPIN=1` calculations converge even faster than the `ISPIN=2` variants on the non-magnetic test set. While this approach sidesteps any discussion of the spin difference initialization, it is not applicable without prior knowledge of a structures magnetic behavior. Should the structure be magnetic according to our definition, the average number SCF steps increases to 44.28, almost twice as many as an SAD `ISPIN=2` calculation with 27.96 steps on average, while also converging to the wrong spin state. Since this is both physically incorrect for about 50% of the MP database and also an inappropriate DFT approach, we deem this not a worth while direction to pursue for DFT initialization.

| | Initialization components | | | | vs. Default [%] | | vs. Oracle [%] | |
|---|---|---|---|---|---|---|---|---|
| | Smooth density | | PAW augmentation | | | | | |
| Initialization | $\tilde{\rho}^+$ | $\tilde{\rho}^-$ | $\mathbf{d}^+$ | $\mathbf{d}^-$ | Non-mag. | Mag. | Non-mag. | Mag. |
| ***Baselines*** | | | | | | | | |
| Default | × | × | × | × | – | – | – | – |
| Oracle | ✓ | ✓ | ✓ | ✓ | +49.0 | +55.4 | – | – |
| ***Electronic component ablations*** | | | | | | | | |
| Smooth valence only | ✓ | × | × | × | +0.5 | −29.4 | −95.2 | −189.9 |
| Spin channel only | × | ✓ | × | ✓ | −9.7 | −70.7 | −115.2 | −282.3 |
| Smooth grids only | ✓ | ✓ | × | × | +10.2 | −4.2 | −76.2 | −133.3 |
| Augmentation only | × | × | ✓ | ✓ | −14.3 | −64.9 | −124.2 | −269.2 |
| ***Magnetic moment initialization only*** | | | | | | | | |
| Oracle MagMom | × | OM | × | OM | +12.0 | −3.1 | −72.6 | −130.8 |
| MPRelaxSet | × | MP | × | MP | +9.3 | −8.5 | −77.9 | −142.9 |
| CHGNet | × | CG | × | CG | +8.8 | −5.7 | −78.9 | −136.8 |
| ***Smooth valence + augmentation initialization*** | | | | | | | | |
| No spin init. | ✓ | × | ✓ | × | +13.5 | −11.3 | −69.7 | −149.3 |
| + Oracle MagMom | ✓ | OM | ✓ | OM | +47.6 | +24.6 | −2.8 | −68.9 |
| + MPRelaxSet | ✓ | MP | ✓ | MP | +42.4 | +17.2 | −13.1 | −85.6 |
| + CHGNet | ✓ | CG | ✓ | CG | +38.7 | +20.4 | −20.2 | −78.3 |

Table 6: Paired SCF-step savings (%) on the Materials Project test set, separated into non-magnetic and magnetic structures. The initialization components are the smooth spin-summed valence density $\tilde{\rho}^+$, smooth spin-difference density $\tilde{\rho}^-$, spin-summed PAW augmentation occupancies $\mathbf{d}^+$, and spin-difference PAW augmentation occupancies $\mathbf{d}^-$, where $\mathbf{d}^\pm$ denotes the collection of atom-wise occupancies $\{\mathbf{d}_a^\pm\}_a$. Default denotes standard SAD initialization, while Oracle uses all converged electronic components from the corresponding completed calculation. ✓ denotes a converged component and × denotes SAD/default initialization. OM, MP and CG denote spin initialization from converged atomic magnetic moments, MPRelaxSet, and CHGNet-predicted magnetic moments, respectively. Positive values indicate faster calculations and negative values indicate slower calculations relative to the corresponding baseline.

| Variant (`ISPIN=1`) | Default (SAD) | Oracle | ChargE3Net[1,*] |
|---|---|---|---|
| DFT Steps ↓ | 15.33 ± 8.02 | 9.21 ± 7.82 | 12.34 ± 8.69 |
| DFT Time ↓ | 163.30 ± 326.32 s | 114.89 s ± 276.57 | 141.49 ± 298.70 s |
| DFT Steps Saved ↑ | – | 39.90 % | 19.51 % |
| DFT Time Saved ↑ | – | 29.64 % | 13.35 % |

Table 7: [1] Koker et al. (2024). * ChargE3Net is initialized with converged augmentation occupancies. DFT convergence analysis with different initializations. The tests were performed on the non-magnetic part of the MP test set.

# C PAW AUGMENTATION AND THE AUGNET ARCHITECTURE

## C.1 PAW AUGMENTATION OCCUPANCIES AS COVARIANT TARGETS

In the PAW formalism, the spin-summed and spin-difference valence density channels are represented as smooth pseudo-densities plus one-center corrections. The atom-centered correction is determined by the PAW setup and by a set of augmentation occupancy coefficients. In a VASP-style `CHGCAR`, these coefficients appear as augmentation occupancy blocks. For each atom $a$, the coefficients can be indexed schematically as

$$d_{a,ij}^{LM,q}, \qquad q \in \{+,-\}, \tag{25}$$

where $q = +$ denotes the spin-summed channel and $q = -$ the spin-difference channel, $i$ and $j$ identify PAW partial-wave channels, and $L, M$ describe the angular momentum channel of the coupled augmentation component.

This is the same structural object considered by the covariant Jacobi-Legendre PAW occupancy model of Focassio et al. (2024) and, for general non-magnetic materials, by EdenGNN (Li et al., 2025). Compared to CJM, our AugNet model has two practical advantages. First, AugNet naturally extends to chemically diverse datasets containing many elements and PAW schemas, since the same backbone and equivariant readout are shared across atoms and the required output coefficients are selected according to the element-specific PAW schema. Second, the model can exploit the expressive nonlinear environment representation learned by MACE backbone rather than relying on a small fixed polynomial expansion. The trade-off is that AugNet contains more parameters and is less interpretable than a linear covariant expansion. For the task of accelerating general and diverse DFT calculations, AugNet is therefore a more practical model.

## C.2 AUGMENTATION SCHEMAS AND EQUIVARIANCE

Augmentation occupancies are not all rotational invariants. For a fixed $L$, the $(2L+1)$ components with $M = -L, \ldots, L$ transform together as an irreducible spherical tensor. Thus, if a structure is rotated, the target vector for each atom must rotate according to the corresponding Wigner representation. We demonstrate this experimentally further in appendix F. This motivates representing the target as a direct sum of irreducible representations,

$$\mathbf{d}_a^q = \{d_{a,ij}^{LM,q}\}_{ijLM} \in \mathcal{V}_{s_a} = \bigoplus_L n_L^{(s_a)} D^L, \qquad q \in \{+,-\}, \tag{26}$$

where $D^L$ denotes the $(2L+1)$-dimensional irrep of $SO(3)$ and $n_L^{(s_a)}$ is the number of independent copies of angular channel $L$ for PAW schema $s_a = s(Z_a)$ associated with atom $a$.

The PAW setup determines which channels exist. In our implementation, all elements are mapped to one of five schema sizes:

$$s(Z_a) \in \{15, 33, 78, 138, 390\}. \tag{27}$$

where $s(Z_a)$ is the number of valid augmentation coefficients for element $Z_a$. The corresponding irreducible representation decompositions are

$$15: \quad 4\times 0\mathrm{e} + 2\times 1\mathrm{o} + 1\times 2\mathrm{e}, \tag{28}$$

$$33: \quad 6\times 0\mathrm{e} + 4\times 1\mathrm{o} + 3\times 2\mathrm{e}, \tag{29}$$

$$78: \quad 7\times 0\mathrm{e} + 6\times 1\mathrm{o} + 6\times 2\mathrm{e} + 2\times 3\mathrm{o} + 1\times 4\mathrm{e}, \tag{30}$$

$$138: \quad 9\times 0\mathrm{e} + 8\times 1\mathrm{o} + 10\times 2\mathrm{e} + 4\times 3\mathrm{o} + 3\times 4\mathrm{e}, \tag{31}$$

$$390: \quad 12\times 0\mathrm{e} + 12\times 1\mathrm{o} + 17\times 2\mathrm{e} + 12\times 3\mathrm{o} + 10\times 4\mathrm{e} + 4\times 5\mathrm{o} + 3\times 6\mathrm{e}. \tag{32}$$

All targets are padded to dimension 390, and a binary mask indicates which coefficients are valid for each atom.

## C.3 MACE BACKBONE

AugNet uses MACE as the equivariant message-passing backbone. MACE constructs per-atom features by expanding local atomic environments in a basis of radial functions and spherical harmonics,

then iteratively mixes these features through equivariant tensor products. The resulting node features transform as a direct sum of irreducible representations,

$$\mathbf{h}_a \in \bigoplus_{\ell=0}^{\ell_{\max}} n_\ell D^\ell. \tag{33}$$

In our implementation, the hidden irreps are specified by a width $w$ and maximum angular order $\ell_{\max}$,

$$\mathbf{h}_a \in w \times 0e \oplus w \times 1o \oplus \cdots \oplus w \times \ell_{\max}^{p_\ell}, \tag{34}$$

where $p_\ell = e$ for even $\ell$ and $p_\ell = o$ for odd $\ell$. The MACE interaction stack produces a sequence of equivariant node representations. We concatenate the node features from the interaction blocks before passing them to the PAW readout, so that the effective readout representation scales with both the hidden width and the number of interaction layers,

$$\mathbf{h}_a^{\text{readout}} = \text{concat}\left(\mathbf{h}_a^{(1)}, \ldots, \mathbf{h}_a^{(T)}\right). \tag{35}$$

Here $T$ is the number of MACE interaction blocks. Increasing the hidden width increases the number of channels per irrep, while increasing the number of interaction blocks increases both the receptive field depth and the dimensionality of the representation passed to the PAW head.

The main expressivity knobs of the backbone are therefore the hidden width, the number of interaction blocks, the MACE correlation order, and the maximum angular order. In practice, width and depth primarily control general capacity, correlation controls the many-body order of the local expansion, and $\ell_{\max}$ controls the angular resolution of the equivariant features.

### C.4 Schema-agnostic Equivariant Readout

The output dimensionality and irrep content depend on the element-specific PAW schema (28-32) but they can be represented with a shared basis. AugNet therefore uses a shared readout across all schemas, using backbone features up to angular order $L_{\text{backbone}}$ and constructing higher-order $L$ irreps through Clebsch-Gordan coupling in a fixed basis.

For an atom $a$, the schema is determined by its atomic number,

$$s_a = s(Z_a), \tag{36}$$

with a maximum quantum number $L_a$. For PAW blocks $(i, j, L)$ with $L \leq L_{\text{backbone}}$ of the backbone model, we can use a fully equivariant linear map to produce the outputs directly:

$$\widehat{\Delta d}_{a,ij}^{LM,q} = \left[\mathcal{W}_{ijL}^q \left(\mathbf{h}_a^{\text{readout}}\right)\right]_M, \qquad L \leq L_{\text{backbone}}, \tag{37}$$

For $L > L_{\text{backbone}}$, we first use an equivariant linear map to project the hidden representation to the projector space by $R$ coefficients for every slot $i$ in the maximal partial-wave basis $\{l_i\}_{i=1}^{n_{\max}}$:

$$c_{a,i,m_i}^{(k)} = \left[\mathcal{W}_{\text{P}} \left(\mathbf{h}_a^{\text{readout}}\right)\right]_{i,k,m_i}. \tag{38}$$

The corresponding PAW blocks are then built using an a Clebsch-Gordan contraction with a fixed basis to ensure that weights are shared across basis sets:

$$\widehat{\Delta d}_{a,ij}^{LM,q} = \sum_{k=1}^{K} w_{(ij,L),k} \sum_{m_i,m_j} C_{\ell_i m_i,\ell_j m_j}^{LM} c_{a,i,m_i}^{(k)} c_{a,j,m_j}^{(k)}. \tag{39}$$

Finally, the atom-wise augmentation occupancy correction is constructed by gathering the components specified by the schema:

$$\widehat{\Delta \mathbf{d}}_a^q = \mathbf{G}_{s(Z_a)} \left( \{ \widehat{\Delta d}_{a,ij}^{LM,q} \}_{ijLM} \right). \tag{40}$$

### C.5 Reference baselines and $\Delta$-learning

For the spin-summed channel, we use VASP's superposition-of-atomic-densities (SAD) occupancies as the reference. Because the free-atom SAD reference is spherically symmetric, it is nonzero only for the $L = 0$ channels, so higher-order components therefore use a zero baseline. For the spin-difference channel, the free-atom reference is not applicable, so we use a zero reference, making $\Delta$-learning equivalent to direct prediction. We write both cases as

$$\hat{\mathbf{d}}_a^q = \mathbf{d}_a^{q,\mathrm{ref}} + \widehat{\Delta \mathbf{d}}_a^q, \qquad \mathbf{d}_a^{q,\mathrm{ref}} = \begin{cases} \mathbf{d}_a^{+,\mathrm{SAD}}, & q = +, \\ \mathbf{0}, & q = -. \end{cases} \tag{41}$$

The free-atom SAD references are only extracted once per PAW dataset with each extraction taking a few minutes at most. Adding the SAD allows for easy transfer between different PAW datasets, allowing the model to focus on higher-order components determined by the chemistry.

### C.6 Coefficient conventions

Raw augmentation occupancies are stored in the PAW ordering. This ordering is organized by partial-wave pairs and angular channels. In contrast, e3nn expects coefficients grouped by irreducible representation. We therefore distinguish between two operations.

First, coefficients are permuted from the PAW channel ordering into e3nn grouped irrep ordering. Second, the real spherical harmonic convention used in the PAW representation is transformed into the real spherical harmonic convention used by e3nn. For a selected channel $q \in \{+, -\}$, we construct an orthogonal matrix $Q_L$ for each angular momentum $L$ such that

$$\mathbf{d}_{a,L}^{q,\mathrm{e3nn}} = Q_L \, \mathbf{d}_{a,L}^{q,\mathrm{PAW}}, \tag{42}$$

The inverse transformation is

$$\mathbf{d}_{a,L}^{q,\mathrm{PAW}} = Q_L^\top \mathbf{d}_{a,L}^{q,\mathrm{e3nn}}. \tag{43}$$

with the transpose equal to the inverse because $Q_L$ is orthogonal. In implementation, $Q_L$ is obtained by evaluating both real spherical harmonic conventions on a deterministic set of points on the sphere and solving the least-squares basis alignment problem, followed by orthogonal projection.

Training is performed in the e3nn basis, which is the natural basis for the equivariant model. For evaluation and file writing, predictions are transformed back to the PAW basis. This also makes the reported PAW component metrics comparable to previous PAW occupancy parity plots.

### C.7 Training objective

AugNet is trained separately for the spin-summed $(+)$ and spin-difference $(-)$ augmentation channels. For a selected channel $q \in \{+, -\}$, the dataset provides target coefficients $\mathbf{d}_a^q$ and a schema mask $m$. The loss is a masked coefficient space loss. For the mean squared error case,

$$\mathcal{L} = \frac{\sum_{a,\alpha} m_{a,\alpha} \left( \hat{d}_{a,\alpha}^q - d_{a,\alpha}^q \right)^2}{\sum_{a,\alpha} m_{a,\alpha}}, \tag{44}$$

where $\alpha$ indexes the packed $(ij, L, M)$ coefficients. We note here that VASP data carries the `LMAXMIX` parameter, dictating the maximum $L$ that is both used by the density mixer but also written to the CHGCAR. This means that if `LMAXMIX` $= 2$, all augmentation occupancies for a given atom above that will be set to 0, regardless of the schema it carries. That means that the masked coefficient loss will also extend to mask out any 0's set by `LMAXMIX`.

### C.8 Evaluation metrics

For a selected augmentation channel $q \in \{+, -\}$, errors are computed over all valid PAW coefficients. In the e3nn basis, the masked coefficient MAE, RMSE, and MaxAE are

$$\mathrm{MAE} = \frac{1}{N_{\mathrm{coeff}}} \sum_{a,\alpha} m_{a,\alpha} \left| \hat{d}^q_{a,\alpha} - d^q_{a,\alpha} \right|, \tag{45}$$

$$\mathrm{RMSE} = \left[ \frac{1}{N_{\mathrm{coeff}}} \sum_{a,\alpha} m_{a,\alpha} \left( \hat{d}^q_{a,\alpha} - d^q_{a,\alpha} \right)^2 \right]^{1/2}. \tag{46}$$

### C.9 AugNet Accuracy and Transfer

Table 8 reports spin-summed augmentation occupancy $\mathbf{d}^+$ prediction errors as the Materials Project training set is increased from 1k structures to the full training set. The full model reaches a mean per-structure MAE/RMSE of $0.0041/0.0118$ on the MP test set and $0.0062/0.0262$ on the OOD GNoME test set. Accuracy improves consistently with training set size on MP, while the OOD results begin to saturate at larger dataset sizes. The unusually large GNoME MaxAE originates from a small number of extreme coefficient outliers: after excluding the ten largest errors, MaxAE falls from $366.60$ to $0.867$. Two prior models directly target the same PAW augmentation occupancy

Table 8: Physical augmentation occupancy prediction errors versus training-set size on the full MP and GNoME test sets. RMSE and MAE are means over per-structure values, while MaxAE is the largest single-coefficient error. $\mathrm{MaxAE}_{\mathrm{top}\,11}$ reports the largest error after excluding the ten most extreme coefficients.

| Dataset | Training set | RMSE ↓ | MAE ↓ | MaxAE ↓ | $\mathrm{MaxAE}_{\mathrm{top}\,11}$ ↓ |
|---|---|---|---|---|---|
| **MP** | 1k | $0.0340 \pm 0.0206$ | $0.0120 \pm 0.0065$ | 4.504 | 1.487 |
| | 10k | $0.0183 \pm 0.0113$ | $0.0064 \pm 0.0034$ | 2.829 | 0.796 |
| | 50k | $0.0130 \pm 0.0096$ | $0.0046 \pm 0.0028$ | 1.931 | 0.664 |
| | Full | $\mathbf{0.0118 \pm 0.0091}$ | $\mathbf{0.0041 \pm 0.0026}$ | 1.114 | 0.646 |
| **GNoME** | 1k | $0.0464 \pm 0.3916$ | $0.0120 \pm 0.0287$ | 366.85 | 1.724 |
| | 10k | $0.0302 \pm 0.3915$ | $0.0078 \pm 0.0284$ | 366.58 | 0.868 |
| | 50k | $0.0271 \pm 0.3915$ | $0.0065 \pm 0.0284$ | 366.58 | 0.868 |
| | Full | $\mathbf{0.0262 \pm 0.3916}$ | $\mathbf{0.0062 \pm 0.0284}$ | 366.60 | 0.867 |

object. EdenGNN (Li et al., 2025) predicts them for general non-magnetic materials and reports an MAE of $0.0085$ on MP structures recomputed with its own VASP settings, but does not use them for SCF initialization. CJM (Focassio et al., 2024) is trained and evaluated on a much narrower dataset containing ab initio molecular dynamics configurations of $\mathrm{MoS_2}$ in the 1H and 1T phases and intermediate geometries, and reports an MAE/RMSE of $0.0130/0.0459$. This provides a useful external reference for the scale of coefficient space errors, although it is not a strict matched benchmark because the datasets and aggregation procedures differ. Compared with this system-specific reference, AugNet reaches errors on the same or lower scale while operating across chemically diverse structures, elements, and PAW schemas.

We further test transfer directly on the $\mathrm{MoS_2}$ dataset of Focassio et al. (2024) (Table 9). This setting changes the underlying Mo PAW dataset relative to Materials Project and therefore changes the target augmentation representation itself. Consequently, the MP-pretrained model does not transfer zero-shot. However, AugNet adapts readily to the new PAW setup: training from scratch reaches an RMSE of $0.0258$, already below the $0.0459$ reported for CJM, while full fine-tuning of the pretrained model reaches $0.0115$. Fine-tuning only the PAW readout requires only 73k trainable parameters and reaches an RMSE of $0.0400$. These results indicate that AugNet generalizes well within a fixed PAW representation and can be efficiently adapted when the underlying PAW setup changes.

| Model | Steps | Trainable params. | MAE ↓ | RMSE ↓ | MaxAE ↓ |
|---|---|---|---|---|---|
| CJM (Focassio et al., 2024) | 1k | 1,758 | 0.0130 | 0.0459 | 0.9137 |
| AugNet, zero-shot | – | – | 0.3854 | 1.3439 | 14.9708 |
| AugNet, from scratch | 1k | 3M | **0.0010** | 0.0258 | 0.4969 |
| AugNet, head fine-tune | 1k | 73k | 0.1801 | 0.7039 | 8.2880 |
| | 10k | 73k | 0.0122 | 0.0400 | 0.9490 |
| AugNet, full fine-tune | 1k | 3M | 0.0157 | 0.0352 | 0.8535 |
| | 10k | 3M | 0.0051 | **0.0115** | **0.2353** |

Table 9: Transfer of MP-pretrained AugNet to the $MoS_2$ PAW setup of Focassio et al. (2024). Their calculations use a different Mo PAW dataset from Materials Project, so the target augmentation representation changes and direct zero-shot transfer is not expected. "Trainable params." denotes the number of parameters optimized during adaptation.

# D MAGNETIC INITIALIZATION DEVELOPMENT

To construct a fully learned spin initialization, we extend both ELECTRAFI and AugNet to the spin-difference components of the PAW density: spin-ELECTRAFI predicts the smooth spin-difference density $\tilde{\rho}^-$ on the plane-wave grid, and spin-AugNet predicts the corresponding spin-difference augmentation occupancies $\mathbf{d}_a^-$. Together with the spin-summed components $(\tilde{\rho}^+, \mathbf{d}_a^+)$, these predictions provide all structure-dependent density components required for direct spin-polarized initialization. Architectural details of AugNet are given in Appendix C and those of ELECTRAFI in Elsborg et al. (2026b).

**Spin-ELECTRAFI.** We adapt the ELECTRAFI model of Elsborg et al. (2026b) to additionally predict the smooth spin-difference density $\tilde{\rho}^- = \tilde{\rho}_\uparrow - \tilde{\rho}_\downarrow$. In the process, several computational inefficiencies of the original implementation were removed, reducing training and inference time without altering the numerics of the model.

The simplest extension within the ELECTRAFI ansatz is to allocate a second set of signed weights $w^-$ to the spin-difference density while sharing the Gaussian centers and covariances with the spin-summed smooth valence density $\tilde{\rho}^+$ The spin-difference weights are predicted analogously to the spin-summed weights. For Gaussian $\mathcal{N}^{(j)}$,

$$w^{(j),-} = \tanh\big(s^{(j),-}\big), \qquad s^{(j),-} = f_{w,-}\big(S^{(j)}\big), \tag{47}$$

where $S \in \mathbb{R}^{N\times C}$ are the scalar outputs of the ELECTRAFI backbone for $N$ atoms and channel width $C$, and $f_{w,-}$ is a multilayer perceptron (MLP) with the same architecture as the spin-summed weight MLP $f_{w,+}$. The Gaussian centers $\boldsymbol{\mu}^{(j)}$ and covariances $\boldsymbol{\Sigma}^{(j)}$ are predicted as in the original model. Both densities are then assembled through the analytic Fourier transform of the Gaussian ansatz followed by an inverse FFT (Elsborg et al., 2026b):

$$\hat{\tilde{\rho}}^+(\mathbf{G}) = \sum_{j=1}^{N_\mathcal{N}} w^{(j),+} \exp\Big[-\tfrac{1}{2}\mathbf{G}^\top\boldsymbol{\Sigma}^{(j)}\mathbf{G}\Big] e^{-i\mathbf{G}\cdot\boldsymbol{\mu}^{(j)}}, \qquad \hat{\tilde{\rho}}^+(\mathbf{r}) = \mathrm{IFFT}\big[\hat{\tilde{\rho}}^+(\mathbf{G})\big](\mathbf{r}), \tag{48}$$

$$\hat{\tilde{\rho}}^-(\mathbf{G}) = \sum_{j=1}^{N_\mathcal{N}} w^{(j),-} \exp\Big[-\tfrac{1}{2}\mathbf{G}^\top\boldsymbol{\Sigma}^{(j)}\mathbf{G}\Big] e^{-i\mathbf{G}\cdot\boldsymbol{\mu}^{(j)}}, \qquad \hat{\tilde{\rho}}^-(\mathbf{r}) = \mathrm{IFFT}\big[\hat{\tilde{\rho}}^-(\mathbf{G})\big](\mathbf{r}). \tag{49}$$

Sharing the Gaussian centers and covariances between the two channels is physically motivated. In collinear spin-polarized DFT the spin-resolved densities are non-negative, so the spin-difference density is bounded pointwise by the total density, $|\tilde{\rho}^-(\mathbf{r})| \le \tilde{\rho}^+(\mathbf{r})$: magnetization can only exist where charge exists. Moreover, the net spin polarization is carried by the same partially filled, localized orbitals that dominate the total density around magnetic atoms, so the spatial support and characteristic length scales of $\tilde{\rho}^-$ are inherited from $\tilde{\rho}^+$, and the two fields differ primarily in sign and magnitude. A Gaussian that is prominent in the spin-summed density is therefore also the natural carrier of any spin difference in the same region, whereas a Gaussian with negligible spin-summed weight should carry no magnetization. The shared basis encodes this structure directly: the geometry of the expansion is fixed by the spin-summed density and only signed magnitudes are learned per channel. This acts as a physical regularizer on $\tilde{\rho}^-$, avoids predicting a second set of centers and covariances, and for non-magnetic structures reduces to training $w^- \to 0$.

Because the spin-difference head reuses the backbone and Gaussian parameters, the backbone continues to receive the clean geometric signal of the spin-summed density while learning the comparatively sparse magnetization density, which stabilizes joint training. The cost is nearly two readout passes and a correspondingly more expensive backward pass per optimization step. The loss is the sum of a spin-summed density term, given by the normalized MAE (NMAE) of Jørgensen & Bhowmik (2022),

$$\mathcal{L}_{\tilde{\rho}^+} = \mathrm{NMAE}\left(\hat{\tilde{\rho}}^+, \tilde{\rho}^+_{\mathrm{ref}}\right) = \frac{\int_\Omega \left|\tilde{\rho}^+_{\mathrm{ref}}(\mathbf{r}) - \hat{\tilde{\rho}}^+(\mathbf{r})\right| \, dV}{\int_\Omega \tilde{\rho}^+_{\mathrm{ref}}(\mathbf{r})\, dV}. \tag{50}$$

and a spin-difference term that distinguishes magnetic from non-magnetic structures,

$$\mathcal{L}_{\tilde{\rho}^-} = \begin{cases} \dfrac{\int_\Omega \left|\tilde{\rho}^-_{\mathrm{ref}}(\mathbf{r}) - \hat{\tilde{\rho}}^-(\mathbf{r})\right| \, dV}{\int_\Omega \left|\tilde{\rho}^-_{\mathrm{ref}}(\mathbf{r})\right| \, dV}, & \text{magnetic}, \\ \displaystyle\int_\Omega \left|\tilde{\rho}^-_{\mathrm{ref}}(\mathbf{r}) - \hat{\tilde{\rho}}^-(\mathbf{r})\right| \, dV, & \text{non-magnetic}. \end{cases} \tag{51}$$

Following Koker et al. (2024); Elsborg et al. (2026b), we classify a structure as magnetic when $M_{\mathrm{abs}} = \int_\Omega |\tilde{\rho}^-_{\mathrm{ref}}| \, dV > M_{\mathrm{min}} = 0.1$. The total loss is

$$\mathcal{L} = \mathcal{L}_{\tilde{\rho}^+} + \lambda_{\mathrm{spin}} \mathcal{L}_{\tilde{\rho}^-}, \qquad \lambda_{\mathrm{spin}} = 0.2. \tag{52}$$

The case distinction is necessary because, for non-magnetic structures, the NMAE denominator approaches zero and the loss term diverges. We also trained with a plain MAE loss for both channels, but this did not yield a balanced contribution from $\tilde{\rho}^+$ and $\tilde{\rho}^-$ and degraded performance.

Analogously to the spin-summed density head, we normalize the spin-difference readout by the net magnetic moment to ensure well-behaved grid predictions. Unlike the number of valence electrons, however, this quantity is not available prior to the DFT calculation. We therefore normalize with the ground-truth moment $M_{\mathrm{DFT}}$ during training and substitute the CHGNet-predicted moment $\hat{M}_{\mathrm{CHGNet}}$ as a surrogate at inference. This approach works well with the exception of antiferromagnetic materials, a notoriously difficult spin state for DFT whose vanishing net moment cannot be resolved by CHGNet. We also attempted to omit the normalization entirely, but this rendered the training dynamics too unstable for long training runs.

Joint training increases the cost by roughly $2.5\times$: whereas spin-summed density training on MP for five epochs takes $2.5$ days on a single NVIDIA H200 GPU, joint training takes approximately one week.

**Spin-AugNet.** Spin-AugNet uses the same equivariant architecture as AugNet with spin-difference augmentation occupancies $\mathbf{d}_a^-$ as targets. As described in Appendix C.5, the free-atom SAD reference is not applicable to the spin-difference channel, so we use $\mathbf{d}_a^{-,\mathrm{ref}} = \mathbf{0}$. Consequently,

$$\hat{\mathbf{d}}_a^- = \widehat{\Delta \mathbf{d}}_a^-, \tag{53}$$

making the shared $\Delta$-learning formulation equivalent to direct prediction for spin-AugNet.

## E Detailed End-to-End DFT Results

Table 10 (Tables 11 and 12 for the magnetic and non-magnetic subsets) reports the complete numerical results underlying the end-to-end comparison in Figure 4. In addition to total wall time, we report density prediction accuracy, SCF iterations, DFT execution time, and ML initialization overhead. The Default calculation uses the standard VASP initialization, while Oracle uses the corresponding converged electronic components $(\tilde{\rho}^+, \mathbf{d}^+, \tilde{\rho}^-, \mathbf{d}^-)$ as initialization and therefore represents an empirical upper bound on the achievable acceleration under the same DFT settings.

Table 10: Detailed comparison of reference-free ML PAW initializations and resulting DFT performance on the test sets of MP and GNoME. Total time includes both ML initialization and DFT execution. For both CNEI-EFI and CNEI-C3Net, we add the time it takes to evaluate spin-ELECTRAFI (0.24s/0.15s) and spin-AugNet (0.05s both) as well as CHGNet (0.03s both) to ELECTRAFI and ChargE3Net.

| Dataset | Metric | Default | Oracle | CNEI-EFI | CNEI-C3Net |
|---|---|---|---|---|---|
| **MP** | $\tilde{\rho}^+$ NMAE ↓ | – | – | 0.58% | 0.54% |
| | ML init time ↓ | – | – | $(0.24 + 0.37)$ s | $(78.73 + 0.37)$ s |
| | SCF steps ↓ | 22.05 | 10.41 | 19.05 | 18.36 |
| | DFT time ↓ | 623.84 s | 302.04 s | 529.43 s | 506.16 s |
| | Total time ↓ | 623.84 s | 302.04 s | 530.04 s | 585.26 s |
| | SCF steps saved ↑ | – | 52.78% | 13.62% | 16.76% |
| | DFT time saved ↑ | – | 51.58% | 15.13% | 18.86% |
| | **Total time saved ↑** | – | **51.58%** | **15.04%** | **6.18%** |
| **GNoME** | $\tilde{\rho}^+$ NMAE ↓ | – | – | 0.93% | 0.69% |
| | ML init time ↓ | – | – | $(0.15 + 0.28)$ s | $(33.28 + 0.28)$ s |
| | SCF steps ↓ | 16.30 | 7.89 | 11.87 | 11.45 |
| | DFT time ↓ | 188.99 s | 112.59 s | 141.00 s | 140.46 s |
| | Total time ↓ | 188.99 s | 112.59 s | 141.43 s | 174.02 s |
| | SCF steps saved ↑ | – | 51.63% | 27.17% | 29.79% |
| | DFT time saved ↑ | – | 40.43% | 25.39% | 25.68% |
| | **Total time saved ↑** | – | **40.43%** | **25.17%** | **7.92%** |

Table 11: The magnetic subset counterpart of table 10. The spin-ELECTRAFI model measures (0.24s/0.15s).

| Dataset | Metric | Default | Oracle | CNEI-EFI | CNEI-C3Net |
|---|---|---|---|---|---|
| **MP** | $\tilde{\rho}^+$ NMAE ↓ | – | – | 0.67 % | 0.78 % |
| | ML Init Time ↓ | – | – | (0.24 + 0.37) s | (87.25 + 0.37) s |
| | SCF steps ↓ | 27.82 | 12.43 | 24.97 | 24.48 |
| | DFT time ↓ | 882.45 s | 377.16 s | 777.63 s | 744.13 s |
| | Total time ↓ | 882.45 s | 377.16 s | 778.24 s | 831.75 s |
| | SCF steps saved ↑ | – | 55.30% | 10.22% | 11.98% |
| | DFT time saved ↑ | – | 57.26% | 11.88% | 15.67% |
| | **Total time saved ↑** | – | **57.26%** | **11.81%** | **5.75%** |
| **GNoME** | $\tilde{\rho}^+$ NMAE ↓ | – | – | 1.01 % | 0.92 % |
| | ML Init Time ↓ | – | – | (0.18+0.31) s | (44.65 + 0.31) s |
| | SCF steps ↓ | 22.60 | 10.06 | 14.69 | 14.49 |
| | DFT time ↓ | 344.88 s | 190.82 s | 238.26 s | 238.22 s |
| | Total time ↓ | 344.88 s | 190.82 s | 238.75 s | 283.18 s |
| | SCF steps saved ↑ | – | 55.47% | 34.99% | 35.88% |
| | DFT time saved ↑ | – | 44.67% | 30.91% | 30.93% |
| | **Total time saved ↑** | – | **44.67%** | **30.77%** | **17.89%** |

Table 12: The non-magnetic subset counterpart of table 10. The spin-ELECTRAFI model measures (0.17s/0.11s) on these subsets.

| Dataset | Metric | Default | Oracle | CNEI-EFI | CNEI-C3Net |
|---|---|---|---|---|---|
| **MP** | $\tilde{\rho}^{+}$ NMAE ↓ | – | – | 0.55 % | 0.50 % |
| | ML Init Time ↓ | – | – | (0.17+0.30) s | (72.11+0.30) s |
| | SCF steps ↓ | 16.83 | 8.58 | 13.68 | 12.80 |
| | DFT time ↓ | 389.41 s | 233.95 s | 304.45 s | 290.45 s |
| | Total time ↓ | 389.41 s | 233.95 s | 304.92 s | 362.86 s |
| | SCF steps saved ↑ | – | 49.01% | 18.71% | 23.91% |
| | DFT time saved ↑ | – | 39.92% | 21.82% | 25.41% |
| | **Total time saved ↑** | – | **39.92%** | **21.70%** | **6.82%** |
| **GNoME** | $\tilde{\rho}^{+}$ NMAE ↓ | – | – | 0.88 % | 0.59 % |
| | ML Init Time ↓ | – | – | (0.11+0.24) s | (28.29+0.24) s |
| | SCF steps ↓ | 13.48 | 6.91 | 10.61 | 10.08 |
| | DFT time ↓ | 119.11 s | 77.52 s | 97.40 s | 96.64 s |
| | Total time ↓ | 119.11 s | 77.52 s | 97.75 s | 125.17 s |
| | SCF steps saved ↑ | – | 48.75% | 21.29% | 25.22% |
| | DFT time saved ↑ | – | 34.92% | 18.22% | 18.86% |
| | **Total time saved ↑** | – | **34.92%** | **17.93%** | **-5.09%** |

# F EQUIVARIANCE OF PAW AUGMENTATION OCCUPANCIES

For each atom $a$ and channel $q \in \{+, -\}$, VASP stores the PAW augmentation occupancies as blocks $\mathbf{d}_a^{(ij,L),q} \in \mathbb{R}^{2L+1}$ with components $d_{a,ij}^{LM,q}$, $M = -L, \ldots, L$, packed into the augmentation occupancy vector $\mathbf{d}_a^q$. The allowed channels are determined entirely by the partial-wave angular momenta $(\ell_i, \ell_j)$ and truncated by `LMAXMIX`. The allowed channels are determined entirely by the partial-wave angular momenta $(l_i, l_j)$ and truncated by `LMAXMIX`.

**Rotation law.** Under a rigid rotation $R \in SO(3)$ of the crystal, each block transforms as

$$\mathbf{d}_a^{(ij,L),q}(R) = Q_L^\top D_{\mathrm{e3nn}}^L(R) Q_L \mathbf{d}_a^{(ij,L),q}, \qquad q \in \{+, -\}, \tag{54}$$

where $D_{\mathrm{e3nn}}^L$ is the real Wigner $D$-matrix and $Q_L$ is a fixed change of basis between the VASP and e3nn spherical harmonic conventions. Consequently, the augmentation occupancies form a direct sum of irreducible $SO(3)$ representations and provide natural equivariant prediction targets.

**Verification.** We verified Equation 54 using 242 rigidly rotated VASP calculations of `mp-1069193`. For each rotation, the occupancies predicted from the identity calculation using Equation 54 were compared to those written by VASP. The relative error

$$\varepsilon_a = \left( \frac{\sum_R \left\| \hat{\mathbf{d}}_a^q(R) - \mathbf{d}_a^q(R) \right\|^2}{\sum_R \left\| \mathbf{d}_a^q(R) \right\|^2} \right)^{1/2}. \tag{55}$$

was below $10^{-6}$ for every atomic site (Table 13), matching the numerical precision of the printed `CHGCAR` values. Using the transpose representation or an incorrect partial-wave ordering increased the error by approximately six orders of magnitude.

| Site | Partial waves | $\varepsilon$ |
|---|---|---|
| 1 | $(2, 2, 0, 0, 1, 1)$ | $4.6 \times 10^{-7}$ |
| 2–5 | $(0, 0, 1, 1)$ | $1.1$–$10.1 \times 10^{-7}$ |

Table 13: Parameter-free verification of the equivariant transformation law over 241 held-out rotations.

The PAW augmentation occupancies therefore provide an exact equivariant coefficient representation of the on-site PAW augmentation correction, and the results show that they can be converted losslessly between the VASP and e3nn conventions via the fixed matrices $Q_L$, allowing $E(3)$-equivariant neural networks to predict augmentation occupancies in their natural irreducible basis.

# G Experiment Setup and Hyperparameters

## G.1 Experimental Hardware

All VASP experiments were conducted using the same Intel Xeon E5-2650 2.20GHz Broadwell CPUs and parallelized across 24 CPU cores using 256 GB of RAM. Machine learning models were trained on a mixture of NVIDIA A100 and H200 GPUs in single-GPU training. However, all inference timings were measured using A100 GPUs.

## G.2 AugNet

| Group | Hyperparameter | Value |
|---|---|---|
| Backbone | Hidden width | 64 |
| | Max. spherical order $\ell_{\max}$ | 3 |
| | Cutoff radius $r_{\max}$ (Å) | 6.0 |
| | Interaction layers | 2 |
| | Correlation order | 3 |
| | Avg. number of neighbors | 64.3 |
| | Tensor product for high $L$ | Yes |
| Readout head | Projector rank | 64 |
| | Block mixing | Yes |
| | Linear readout up to $L$ | 3 (CG coupling for $L \geq 4$) |
| Target | Spin-summed $(+)$ | $\Delta$ from SAD reference |
| | Spin-difference $(-)$ | Zero reference (direct prediction) |
| Optimization | Optimizer | AdamW |
| | Learning rate | $1 \times 10^{-2}$ |
| | Backbone LR multiplier | 0.1 |
| | Weight decay | $1 \times 10^{-6}$ |
| | Gradient clipping | None |
| | Epochs | 5 |
| | Batch size | 1 |
| | Precision | FP32 |
| | Loss | MSE |

Table 14: Hyperparameters used for training AugNet and spin-AugNet. The spin-summed and spin-difference models share all architectural and optimization settings and differ only in the predicted channel and reference baseline.

### G.3 ELECTRAFI

| Group | Parameter | Value |
|---|---|---|
| Backbone (EScAIP) | Layers | 2 |
| | Hidden size | 256 |
| | Attention heads | 32 |
| | Atom embedding size | 128 |
| | Edge distance embedding | 512 (expansion 600) |
| | Node direction embedding | 256 (expansion 13) |
| | FFN hidden multiplier | 2 |
| | Activation | GELU |
| | Normalization | LayerNorm |
| | Dropout / stochastic depth | 0 |
| | Max neighbours | 300 |
| | Batch size | 1 |
| | Master units | 2160 |
| Density representation | Gaussians per electron | 120 |
| | Signed weights | `tanh_softplus` |
| | Weight magnitude cap | 50 |
| | Gaussian width scale range | $[0.01,\ 25]$ |
| | Gaussian width floor | $1 \times 10^{-4}$ |
| | Renormalization floor | $10^{-3} \sum \lvert w \rvert$ |
| | Plane-wave grid | $128^3$ |
| Spin-difference channel | Spin loss weight | 0.2 |
| | Magnetic threshold $M_{\mathrm{abs}}$ | $0.1\,\mu_B$ |
| | Spin loss warm-up | 500 steps |
| Optimization | Epochs | 5 |
| | Optimizer | Muon + AdamW (AMSGrad) |
| | Learning rate (AdamW) | $3 \times 10^{-4}$ |
| | Learning rate (Muon) | $3 \times 10^{-3}$ |
| | Final learning rate | $1 \times 10^{-4}$ |
| | LR decay | $\gamma = 0.7$ per epoch |
| | Weight decay | 0 |
| | Muon momentum (Nesterov) | 0.95 |
| | Newton–Schulz steps | 5 |
| | Gradient clipping | 1.0 (norm) |
| | Loss | Equation 50 |
| | Loss-spike skip threshold | $50\times$ EMA |
| | Training Time Rotations | True |
| | Precision | FP32 |

Table 15: Hyperparameters for the ELECTRAFI and spin-ELECTRAFI models, mirroring the choices in Elsborg et al. (2026b).