Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Volumetric electron-density prediction in elementalmetals: continuous-field accuracy andtopology-sensitive validation

Irina Arévalo

irina.arevalo@upm.es

Universidad Politécnica de Madrid

Article

Keywords: electron density prediction, density functional theory, volumetric neural networks, Bader analysis

Posted Date: July 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10339992/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Volumetric electron-density prediction in elemental metals: continuous-field accuracy and topology-sensitive validation

Irina Arévalo[1*]

[1*]School of Civil Engineering, Universidad Politecnica de Madrid, Calle Profesor Aranguren, 3, Madrid, 28040, Spain.

Corresponding author(s). E-mail(s): irina.arevalo@upm.es;

Abstract

Machine-learning models that predict electron densities directly from atomic structure can speed up electronic-structure workflows, but their performance is often tested only through global voxelwise errors. Here we develop compact three-dimensional U-Net models for valence electron-density prediction in perturbed elemental Al, Fe, and Ni configurations. The primary model, UNet-StructureChem, combines atom-centered structural priors with an explicit element channel, while a geometry-only baseline and a metadata-assisted privileged model isolate the effects of chemical conditioning and generator-derived perturbation information. On 2,924 held-out configurations, UNet-StructureChem achieves a mean log-density error of $\mathbf{2.9018 \times 10^{-6}}$, a relative $\boldsymbol{L_1}$ error of **0.2959%**, and a mean raw relative electron-count error of $\mathbf{4.0612 \times 10^{-4}}$, showing sub-percent volumetric density reconstruction without post-hoc charge rescaling. Relative to the geometry-only baseline, explicit chemical conditioning reduces the mean log-density error by **28.2%** and the relative $\boldsymbol{L_1}$ error by **12.7%**. The privileged model reaches the lowest voxelwise error, defining an upper-bound control for this synthetic benchmark. However, downstream Bader analysis reveals a different model ranking: the geometry-only model gives the lowest valid Bader charge error, whereas chemically conditioned and privileged models improve the continuous density field but increase mean valid atom-partitioned charge errors. Catastrophic Bader failures follow a third trend, with chemically conditioned and privileged models reducing the number of severe partition failures despite worsening mean valid Bader accuracy. Element- and distance-resolved analysis shows that this decoupling is driven primarily by Al basin topology, especially in compressed configurations, while Fe and Ni remain stable

and generally benefit from chemical conditioning. These results show that voxelwise density fidelity, electron-count conservation, Bader charge accuracy, and topological partition stability are distinct evaluation targets. Machine-learned electron densities should therefore be validated both as continuous scalar fields and as inputs to the downstream physical analyses for which they are intended.

Keywords: electron density prediction, density functional theory, volumetric neural networks, Bader analysis

Introduction

Kohn–Sham density functional theory (KS-DFT)[1–3] is one of the main computational tools for analyzing the electronic structure of materials. Its output is the ground-state electron density, from which total energies, electrostatic properties, bonding descriptors, forces, and a wide range of material observables can be obtained [1–3]. By replacing the many-electron wavefunction with a three-dimensional scalar field, density functional theory is able to obtain quantum-mechanical accuracy with computational tractability, and has become central to computational materials discovery and characterization [4, 5].

Despite its broad success, the computational cost of self-consistent KS-DFT calculations is a limiting factor for extensive configurational sampling. Even relatively small crystalline systems may require many repeated calculations when atomic displacements, strains, defects, chemical environments, or thermodynamic conditions are explored. This cost is manageable for a small number of structures, but becomes a major bottleneck in workflows that require the evaluation of thousands of related configurations, such as automated evaluation of many materials, atomistic simulations, and inverse design. Although linear-scaling, reduced-scaling, and large-scale electronic-structure methods have considerably extended the range of accessible systems [6–11], routine KS-DFT calculations are still expensive when dense configurational sampling is required. Learning the electron density directly is therefore a promising strategy: a predicted density field provides a physical intermediate representation from which multiple downstream quantities can be computed without additional self-consistent calculations.

Machine-learning approaches have increasingly been developed to predict electronic-structure quantities directly from atomic configurations, mainly electron density [12–22]. Most existing work focuses on predicting the electron density field itself, although related approaches have also targeted one- and two-particle density matrices, Hamiltonian surrogates, band structures, defect energetics, or total energies derived by postprocessing predicted densities [23–28].

The predicted density can be represented in several ways. One family of methods expands the density in atom-centered basis functions or atom-wise local contributions, yielding compact and transferable representations that exploit the locality of electronic structure, but whose accuracy depends on the expressivity of the chosen

basis and on the ability to describe sharp, anisotropic, or chemically complex density features [13, 16, 19, 29–31]. A second class predicts the density pointwise on real-space grids using descriptors of the local atomic environment around each grid point; these approaches can represent arbitrary density shapes and can scale linearly with the number of grid points, but require repeated model evaluation over fine meshes and can therefore become costly for high-resolution volumetric predictions [12, 14, 15, 18, 32–34]. A third strategy uses graph-based or equivariant neural networks to learn symmetry-preserving atomistic representations directly from molecular or periodic structures, avoiding some handcrafted descriptor choices and improving rotational or permutational consistency, although graph inference can remain computationally demanding in high-throughput or large-system settings [17, 20, 35, 36]. More recently, volumetric convolutional approaches have predicted three-dimensional electron-density fields directly on grids, often using physically motivated reference fields such as superpositions of atomic densities, grid-projected atomic fingerprints, or atom-centered structural priors as input channels [21, 37]. Related work has also shown that sparse three-dimensional electron-density representations can be used effectively with 3D deep-learning models for materials classification and property prediction [38]. These models are useful for periodic materials because they preserve the spatial organization of the target observable, take advantage of the efficiency of convolutional inference, and naturally produce grid-aligned densities suitable for downstream real-space analyses. However, the choice of input representation has a major importance: global or species-blind descriptors may be insufficient when local chemical identity changes the target density, whereas site-resolved or explicitly chemical representations are required when the model must distinguish different elements, local bonding environments, or mixed-composition systems [22]. This motivates the present comparison between geometry-only structural priors, explicit chemical conditioning, and privileged generator-derived metadata. This emphasis on representation is consistent with broader atomistic machine-learning literature, where locality, descriptor dimensionality, chemical encoding, and symmetry handling strongly influence transferability and data efficiency [39–45].

Most density-prediction models are evaluated primarily with global field metrics such as mean squared error, normalized root mean squared error, or relative L_1 error [17, 18, 22]. These metrics are necessary, but they do not determine whether a predicted density is reliable for a specific downstream analysis. This distinction is especially important for observables that depend on spatial derivatives, local interatomic regions, integrated charge, or density topology. A model may therefore improve the average voxelwise density error while degrading a downstream quantity whose computation is nonlinear in the density field.

Bader analysis is a widely used example of a topology-sensitive density analysis. It divides the electron density into atom-centered regions, called atomic basins, whose boundaries are defined by zero-flux surfaces of the density gradient [46]. The charge assigned to each atom is then obtained by integrating the density inside its basin. Grid-based algorithms make this procedure applicable to real-space charge-density data [47–50]. However, Bader charges depend not only on the density values, but also on where the basin boundaries are located. A small local error in the predicted density

can shift a boundary and change the amount of density assigned to each atom, even if the overall voxelwise error is small. Conversely, a prediction with a larger average field error may still give accurate atomic charges if it preserves the relevant basin boundaries. Therefore, improved voxelwise density accuracy does not necessarily imply improved Bader-partition stability or improved Bader-charge accuracy.

Here, we develop and evaluate compact three-dimensional U-Net models for predicting valence electron densities in perturbed elemental Al, Fe, and Ni configurations. This architecture is well suited to the task because both the inputs and targets are volumetric fields defined on the same real-space grid. The convolutional encoder captures local atom-centered density features, the decoder reconstructs the full three-dimensional density field, and the skip connections help preserve high-resolution spatial detail. As a result, the model can predict grid-aligned electron densities in a single forward pass, producing outputs that can be evaluated directly with voxelwise metrics, electron-count integration, spatial diagnostics, and Bader partitioning. Our primary model, denoted UNet-StructureChem, uses atom-centered volumetric structural priors together with explicit chemical identity. Both inputs are computable from the final atomic structure. We compare this model with a geometry-only baseline, UNet-Geometry, and with a metadata-assisted model, UNet-Privileged, that additionally receives generator-derived perturbation information. The latter is treated as a privileged upper-bound control rather than as the primary predictor. The overall workflow is shown in Figure 1, and the resulting model hierarchy is summarized in Table 1.

This study evaluates held-out perturbations of known two-atom Al, Fe, and Ni parent structures. Therefore, the test set measures interpolation within the structural families represented during training, rather than generalization to entirely new crystal prototypes, larger systems, different grid shapes, or unseen chemical elements. This controlled setting is useful because it allows us to isolate the effects of explicit chemical conditioning, privileged perturbation metadata, and downstream Bader topology without mixing them with broader changes in structure or composition.

We evaluate the models using two complementary classes of diagnostics. The first treats the prediction as a continuous scalar field and measures voxelwise accuracy, integrated electron-count conservation, and the spatial and spectral structure of the residual error. The second treats the predicted density as an input to downstream Bader analysis, testing partition validity, atom-resolved charge errors, charge asymmetry, catastrophic failures, and element- and distance-dependent stability. This design allows us to ask not only which model best reconstructs the density field, but also whether improved field accuracy translates into improved topology-sensitive physical analysis.

Our results show that explicit chemical conditioning improves volumetric density prediction within this controlled elemental benchmark. UNet-StructureChem reduces voxelwise density errors relative to the geometry-only baseline while using only non-privileged inputs, whereas the metadata-assisted model achieves the lowest full-field density error and is interpreted as an upper-bound control. However, Bader analysis reveals a different ordering: the geometry-only model yields the lowest valid

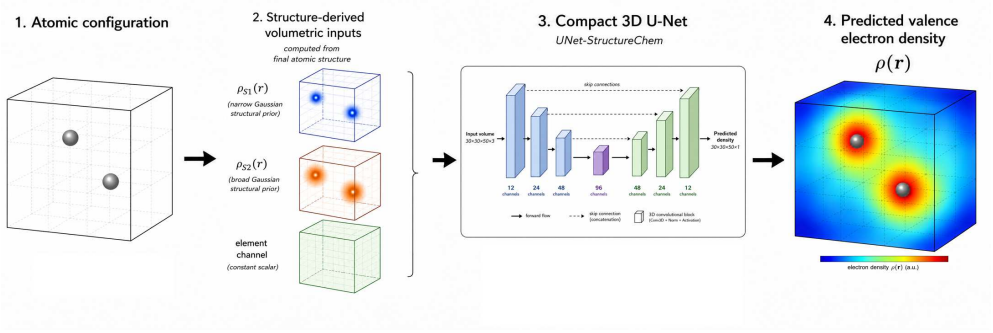


Fig. 1 Workflow for elemental valence electron-density prediction. Each input structure is a periodic two-atom elemental configuration represented on a three-dimensional real-space grid. From the final atomic structure, we construct two atom-centered Gaussian structural-prior channels, ρ_{S1} and ρ_{S2}, together with a spatially constant element channel. These structure-derived volumetric inputs are passed to a compact three-dimensional U-Net, which predicts the valence electron-density field $\hat{\rho}(\mathbf{r})$ on the same grid. For UNet-Privileged, generator-derived perturbation metadata is optionally supplied as an additional input and is used only as an upper-bound control.

Table 1 Model hierarchy and input representations. UNet-Geometry uses only structural-prior channels. UNet-StructureChem is the primary model because its inputs are computable from the final atomic structure. UNet-Privileged is an upper-bound control because it additionally uses generator-derived perturbation metadata.

Model	Role	Inputs	Parameters
UNet-Geometry	Baseline	Priors	789,433
UNet-StructureChem	Deployable model	Priors + element	789,757
UNet-Privileged	Upper-bound control	Priors + element + metadata	792,673

Bader charge error, while models with better voxelwise density accuracy can produce larger atom-partitioned charge errors. Catastrophic partition counts follow a third trend, indicating that mean Bader charge error and severe partition failure capture distinct aspects of topological stability. Together, these findings show that voxelwise density fidelity, electron-count conservation, topological partition stability, and atom-partitioned charge accuracy are distinct evaluation targets. Machine-learned electron densities should therefore be validated not only as continuous scalar fields, but also through the downstream and topology-sensitive observables for which they are intended.

Results

Structure-derived chemical conditioning improves volumetric density reconstruction

We first evaluated the primary model, UNet-StructureChem, on 2,924 held-out Al, Fe, and Ni configurations. Each structure contained two atoms and was represented on a

$30 \times 30 \times 50$ real-space grid. The model used two atom-centered volumetric structural priors together with an explicit element channel. All three inputs are computable from the final atomic structure and do not require generator-derived perturbation labels or amplitudes. The test configurations were excluded from model fitting and checkpoint selection, although they were generated by perturbing the same elemental parent templates represented in the training set. In this held-out test set, UNet-StructureChem reproduced the reference electron densities with a mean per-structure log-space mean squared error (logMSE) of 2.9018×10^{-6}, a relative L_1 error of 0.2959%, and a mean raw relative electron-count error (sumRel) of 4.0612×10^{-4}.

For comparison, the geometry-only baseline, UNet-Geometry, achieved a logMSE of 4.0419×10^{-6}, a relative L_1 error of 0.3389%, and a raw relative electron-count error of 1.6352×10^{-4}. Thus, adding explicit chemical identity reduced the mean logMSE by 28.2% and the relative L_1 error by 12.7%, while the geometry-only model retained slightly better raw electron-count conservation. The raw sumRel values also show that integrated charge conservation did not vary monotonically with model complexity: UNet-StructureChem had the largest raw electron-count error of the three models, whereas UNet-Privileged recovered charge conservation closer to UNet-Geometry despite using the richest input representation. This contrast already suggests that different metrics react differently to the same input-representation choices. The metadata-assisted UNet-Privileged model was used as an upper-bound control because it receives generator-derived perturbation labels and amplitudes that are not generally available for arbitrary input structures. That is, it is a non-deployable model because it depends on information from the data-generation process rather than on quantities that would be known for a new structure at inference time. It achieved a logMSE of 2.2499×10^{-6}, a relative L_1 error of 0.2770%, and a raw relative electron-count error of 1.7243×10^{-4}. The gap between UNet-StructureChem and UNet-Privileged therefore quantifies the additional benefit of non-deployable generator metadata in this controlled benchmark. The three-model comparison is summarized in Fig. 2, which reports the canonical density metrics for the frozen elemental test split.

The comparison between the three models shows that the most realistic conditioning variable is explicit chemical identity. UNet-StructureChem improved the principal full-field density metrics relative to UNet-Geometry while avoiding privileged perturbation descriptors. At the same time, the superior voxelwise performance of UNet-Privileged indicates that generator metadata can further refine the density field when the perturbation recipe is known. This distinction is important for deployment: chemical identity is part of the input structure, whereas perturbation metadata is a property of the synthetic data-generation process.

The reported density metrics were computed directly from the unrescaled network outputs. No post-hoc electron-count normalization was applied for the canonical evaluation. Consequently, differences in raw sumRel reflect genuine differences in the charge conservation behavior of the learned representations. In particular, UNet-StructureChem improved spatial density reconstruction relative to UNet-Geometry, but did not improve raw electron-count conservation. This illustrates that voxelwise density accuracy and integrated charge accuracy are related but distinct evaluation targets. Taken together, these results identify UNet-StructureChem as the primary

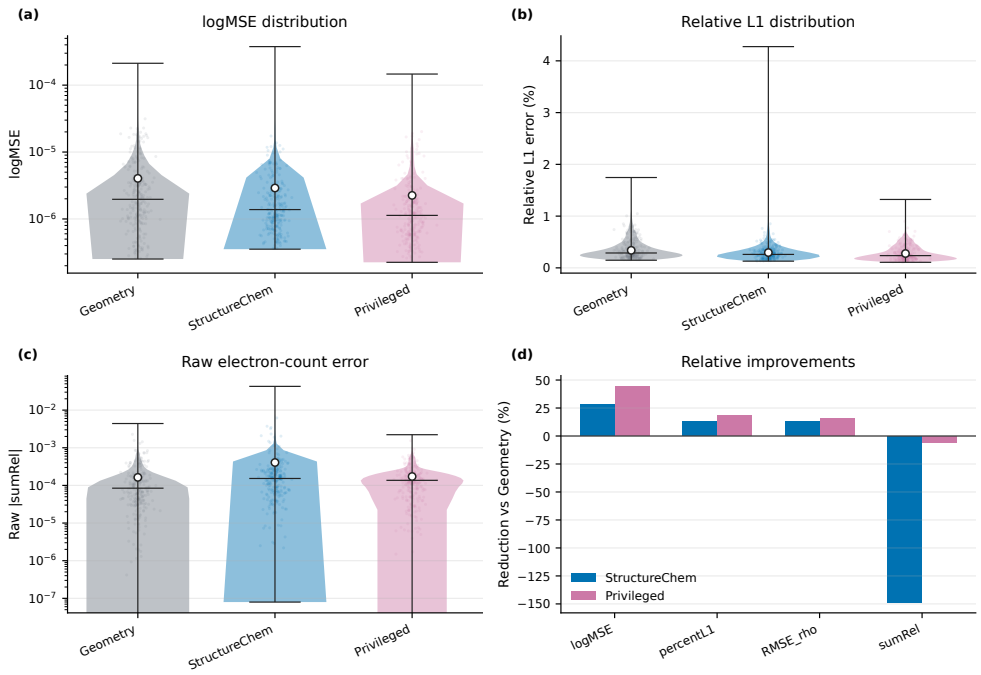


Fig. 2 Canonical density metrics on the frozen elemental test split. All metrics were computed from unrescaled network predictions. UNet-StructureChem reduces voxelwise density errors relative to the geometry-only baseline, showing the benefit of deployable explicit chemical conditioning. UNet-Privileged achieves the lowest voxelwise density errors, but uses generator-derived perturbation metadata and is therefore interpreted as an upper-bound control. Raw electron-count error is reported separately because improved voxelwise reconstruction does not necessarily imply improved integrated charge conservation.

realistic density predictor for the remainder of the study. It provides a deployable improvement over a geometry-only representation, while UNet-Privileged is retained as a metadata-assisted upper-bound control. UNet-Geometry is retained as a baseline for the downstream topology-sensitive analysis, where its simpler representation provides an important comparison point for Bader partition stability.

Spatial and spectral structure of prediction errors

Global error metrics can hide where the density is most difficult to reconstruct. We therefore decomposed the density error by spatial region, element-centered radial profiles, Fourier shells, and bond-midpoint neighborhoods for UNet-Geometry, UNet-StructureChem, and UNet-Privileged. These diagnostics were chosen to probe complementary aspects of the residual error: proximity to atomic centers, element-dependent local structure, reciprocal-space scale, and the interatomic region most directly related to basin separation.

Near-atom regions remained substantially more difficult than interstitial regions for all three models. Using a near-atom radius of $r_{\mathrm{near}} = 1.2$ Å, UNet-Geometry produced a mean absolute density error of 7.6442×10^{-4} near atoms and 1.6205×10^{-4} in the interstitial region, corresponding to a near/interstitial ratio of 4.72. UNet-StructureChem reduced both errors, to 6.1262×10^{-4} near atoms and 1.4608×10^{-4} in the interstitial region, with a ratio of 4.19. UNet-Privileged achieved the lowest regional errors, 5.8670×10^{-4} near atoms and 1.2892×10^{-4} in the interstitial region, with a ratio of 4.55. Thus, explicit chemical conditioning improved both near-atom and interstitial reconstruction relative to the geometry-only baseline, while privileged perturbation metadata provided an additional reduction in absolute error. However, the near-atom region remained approximately four to five times more difficult than the interstitial region across all representations. The ratios also show that the spatial distribution of the residual error changed across representations. Although UNet-Privileged reduced the absolute error in both regions, its near/interstitial ratio was higher than that of UNet-StructureChem, indicating that the additional metadata reduced interstitial errors proportionally more than near-atom errors. Thus, privileged perturbation information improved the overall field while leaving the near-atom reconstruction difficulty comparatively more prominent.

Radial-profile analysis measures the density error as a function of distance from each atom. For each element, we averaged the prediction error in spherical shells around the atomic positions. This showed a consistent ordering across all three models: Ni had the largest radial-profile errors, followed by Fe and then Al. This suggests that the transition-metal densities are harder to reconstruct locally, likely because they contain sharper and more element-specific near-atom features. Chemical conditioning improved the overall density prediction, but it did not remove this intrinsic element-dependent difficulty.

Fourier-shell analysis measures how the density error is distributed across spatial scales. Instead of asking where the error occurs in real space, it transforms the density-error field into reciprocal space and averages the error magnitude over shells of increasing spatial frequency. Low-frequency shells correspond to broad, slowly varying density errors, whereas high-frequency shells correspond to sharper, more localized features. For all three models, the shell-averaged error decreased with shell index, with Spearman correlations of -0.987 for UNet-Geometry, -1.000 for UNet-StructureChem, and -0.949 for UNet-Privileged. The corresponding high/low shell ratios were 0.0195, 0.0189, and 0.0336, respectively. Thus, under this normalization, the dominant error magnitude was concentrated at lower spatial frequencies rather than in the highest-frequency components. The larger high/low ratio of UNet-Privileged should not be interpreted as a larger absolute high-frequency error, because this model has the lowest overall voxelwise error. Rather, it indicates that after the dominant low-frequency error components are reduced, the remaining residual error is relatively more weighted toward finer spatial scales. This is consistent with the role of privileged perturbation metadata: it improves the broad density reconstruction, but the residual errors that remain may be more concentrated in sharper local features. UNet-StructureChem retained the most monotonic shell trend, while all models showed strong scale dependence. The Fourier results therefore support the broader

conclusion that global voxel metrics do not fully characterize the spatial-frequency structure of the prediction error.

Bond-midpoint analysis measures the density error in the region between the two atoms. This region is physically important because it samples the interatomic density associated with bonding and basin separation, rather than the density close to either nucleus. For each two-atom configuration, we evaluated the absolute density error near the periodic midpoint of the shortest interatomic vector and compared it with the interatomic distance d_{12}, defined as the periodic minimum-image distance between the two atoms. Across the complete test set, the Spearman correlation between absolute midpoint-density error and d_{12} was 0.0478 for UNet-Geometry, -0.1074 for UNet-StructureChem, and -0.1434 for UNet-Privileged. Thus, there was no strong universal monotonic relationship between midpoint error and interatomic distance across all elements and perturbation types. Within close Al configurations, the distance dependence remained modest for the two deployable models: the correlations between d_{12} and logMSE were -0.1289 for UNet-Geometry and -0.0941 for UNet-StructureChem, while the corresponding correlations between d_{12} and midpoint error were -0.1069 and -0.0589. In contrast, UNet-Privileged showed a stronger close-Al midpoint-distance correlation of -0.4033, indicating that its residual midpoint error was more concentrated in the most compressed Al configurations. Together, these results show that midpoint errors capture localized interatomic reconstruction behavior that is not explained by distance alone.

Overall, the spatial diagnostics show that UNet-StructureChem improves density reconstruction over UNet-Geometry in both near-atom and interstitial regions, while UNet-Privileged remains the best field-level predictor. Nevertheless, the relative difficulty of near-atom reconstruction, the Ni $>$ Fe $>$ Al radial-error ordering, and the weak global midpoint-distance correlations persist across the model hierarchy. These findings motivate the downstream Bader analysis below: even when chemical conditioning improves the continuous density field, the remaining errors are spatially structured and may interact nonlinearly with topology-sensitive charge partitioning.

The spatial and spectral diagnostics are summarized in Fig. 3, which compares regional density errors, element-resolved radial errors, Fourier-shell statistics, and bond-midpoint behavior across the three models. Representative density slices are shown in Fig. 4. These examples provide a qualitative view of the spatial error patterns underlying the regional, radial, Fourier-shell, and bond-midpoint diagnostics reported above.

Voxelwise accuracy and Bader accuracy are decoupled

We next asked whether improved volumetric density reconstruction translated into improved Bader-derived atomic charges. Bader analysis partitions the electron density into atom-centered basins defined by zero-flux surfaces of the density-gradient field. The Bader charge of each atom is then obtained by integrating the density assigned to its basin. This makes Bader analysis a useful downstream test for machine-learned densities: it evaluates whether the predicted field preserves the atom-partitioning structure needed to recover atomic charges, not only whether it matches the reference density point by point.

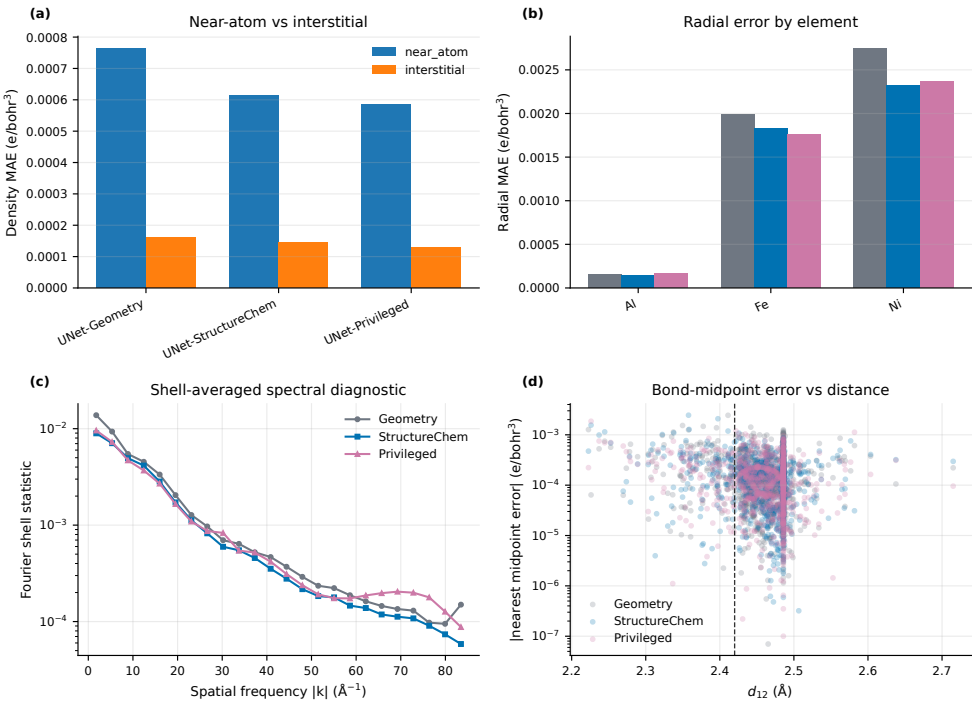


Fig. 3 Spatial and spectral structure of density errors. Near-atom regions exhibit larger density errors than interstitial regions for all models. Element-resolved radial diagnostics show a consistent error ordering of Ni > Fe > Al, indicating that local density reconstruction difficulty remains chemically structured. Fourier-shell diagnostics summarize the scale dependence of the density-error field in reciprocal space. Bond-midpoint diagnostics show weak global monotonic dependence on interatomic distance, with stronger localized effects in close Al configurations.

The three-model comparison showed a clear reversal between voxelwise density accuracy and valid Bader charge accuracy. For the canonical density metrics, the ranking was UNet-Privileged, followed by UNet-StructureChem, followed by UNet-Geometry. The corresponding mean logMSE values were 2.2499×10^{-6}, 2.9018×10^{-6}, and 4.0419×10^{-6}, respectively. The same ordering was observed for relative L_1 error, with values of 0.2770%, 0.2959%, and 0.3389%.

Bader charge accuracy followed the opposite trend. Here, the Bader MAE measures the error in the atom-resolved electron counts obtained after partitioning the predicted density into Bader basins and integrating the density within each basin. It therefore tests whether the predicted density gives the correct atomic charge assignment after the nonlinear basin-partitioning step. UNet-Geometry, despite having the largest voxelwise error, achieved the lowest valid Bader MAE, 0.02248 e. UNet-StructureChem improved voxelwise density reconstruction relative to UNet-Geometry but increased the valid Bader MAE to 0.03280 e. UNet-Privileged achieved the lowest voxelwise density error but the highest valid Bader MAE, 0.04217 e. Thus, the model that best reproduced the continuous density field did not best preserve the

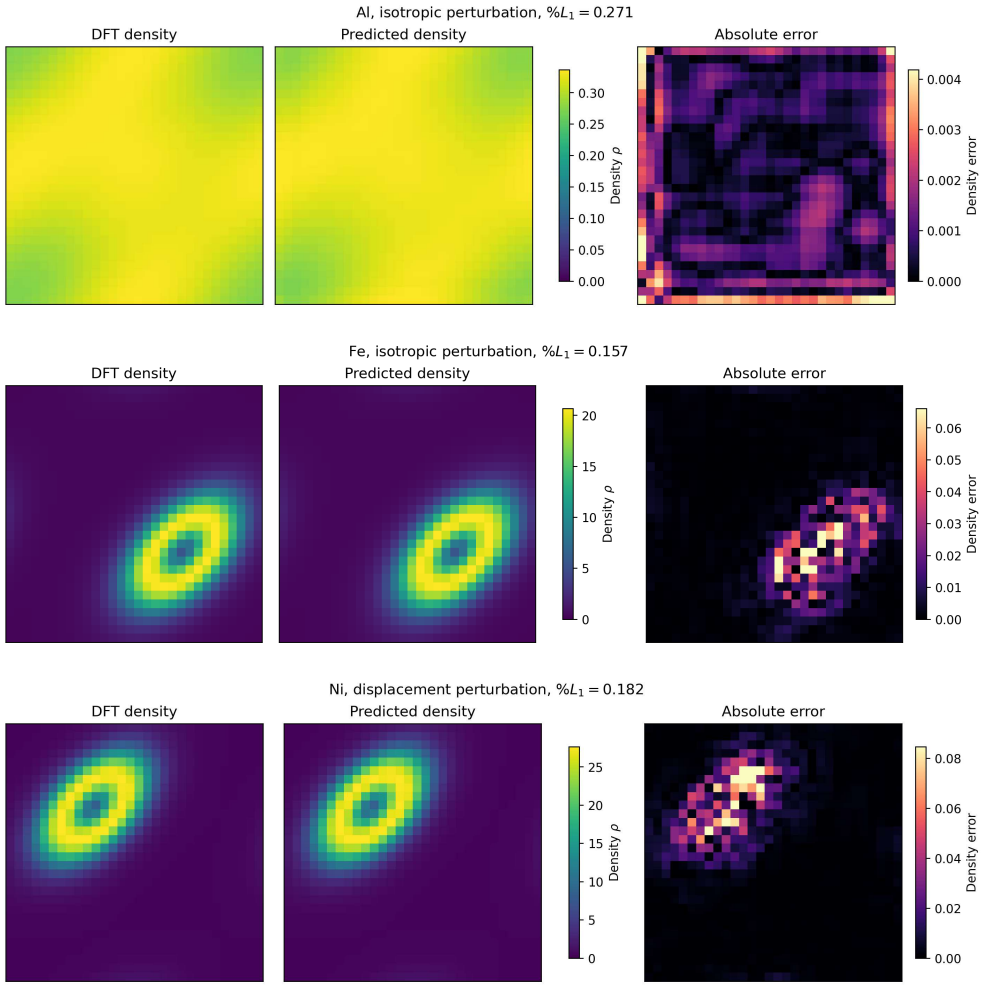


Fig. 4 Representative density slices for UNet-StructureChem predictions. Each row shows the DFT reference density, the predicted density, and the absolute density error for one held-out elemental test configuration. The examples illustrate that the model reproduces the dominant spatial structure of the reference densities, while residual errors are concentrated near sharper density gradients and shell-like features. The Al example shows a smoother density field, whereas Fe and Ni display more localized high-density structure, consistent with the element-resolved radial-error trend reported in Fig. 3.

Bader-derived atom-partitioned charges. The raw Bader MAE followed a similar pattern for the metadata-assisted model: UNet-Geometry and UNet-StructureChem were nearly tied, with raw Bader MAEs of 0.04733 e and 0.04773 e, whereas UNet-Privileged increased the raw Bader MAE to 0.05834 e. The gap between raw and valid Bader MAE was largest for UNet-Geometry, indicating that its invalid partitions were disproportionately severe even though its valid partitions were the most accurate.

The same decoupling was observed for the two-atom charge-asymmetry error. This metric compares the Bader-charge difference between the two atoms in the predicted density with the corresponding difference in the DFT reference density. It therefore tests whether the predicted density preserves the relative atom-to-atom charge imbalance after Bader partitioning, rather than only the absolute charge assigned to each atom. The Δq MAE was 0.09467 e for UNet-Geometry, 0.09546 e for UNet-StructureChem, and 0.11668 e for UNet-Privileged. Thus, the models that improved the continuous density field did not improve the Bader-derived two-atom charge asymmetry. This is especially important because Bader charges are determined by basin boundaries in the density-gradient field rather than by pointwise density values alone.

Catastrophic Bader failures showed yet another ordering. Here, a catastrophic failure denotes a structure in which at least one atom has a very large Bader charge error, indicating a severe basin-partitioning error rather than an ordinary small charge deviation. Using a catastrophic threshold of maximum atomwise Bader error greater than 1 e, UNet-Geometry produced 60 catastrophic atoms, UNet-StructureChem produced 48, and UNet-Privileged produced 46. Equivalently, the corresponding catastrophic structure counts were 30, 24, and 23. Thus, the simplest geometry model gave the best valid Bader MAE but the largest number of catastrophic partition failures, while the privileged model gave the worst valid Bader MAE but the fewest catastrophic events. Mean valid charge error and catastrophic partition instability therefore capture distinct aspects of Bader reliability: one measures typical charge accuracy among valid partitions, whereas the other measures rare but severe failures of the basin assignment. These results cannot be explained by voxelwise density accuracy alone. UNet-StructureChem and UNet-Privileged both improved the full-field density metrics relative to UNet-Geometry, but both worsened valid Bader MAE. Nor can the Bader ranking be reduced to raw electron-count conservation. The mean raw relative electron-count errors were 1.6352×10^{-4} for UNet-Geometry, 4.0612×10^{-4} for UNet-StructureChem, and 1.7243×10^{-4} for UNet-Privileged, whereas the valid Bader MAE increased monotonically from Geometry to StructureChem to Privileged. The results therefore show that voxelwise density fidelity, integrated charge conservation, valid Bader charge accuracy, and catastrophic partition stability are distinct evaluation targets.

This decoupling motivates the element- and distance-resolved analysis below. In particular, the aggregate Bader statistics suggest that chemical conditioning and privileged metadata alter the density in ways that improve average field reconstruction but can shift the gradient topology that defines Bader basins. Consequently, downstream topological validation is necessary even when global density metrics improve.

The reversal between voxelwise density accuracy and Bader-derived charge accuracy is summarized in Fig. 5. The corresponding headline density, Bader, and catastrophic-failure metrics are reported in Table 2.

Element- and distance-dependent Bader regimes

The aggregate Bader comparison concealed a strong interaction between element and interatomic distance. We divided the test set into compressed configurations with

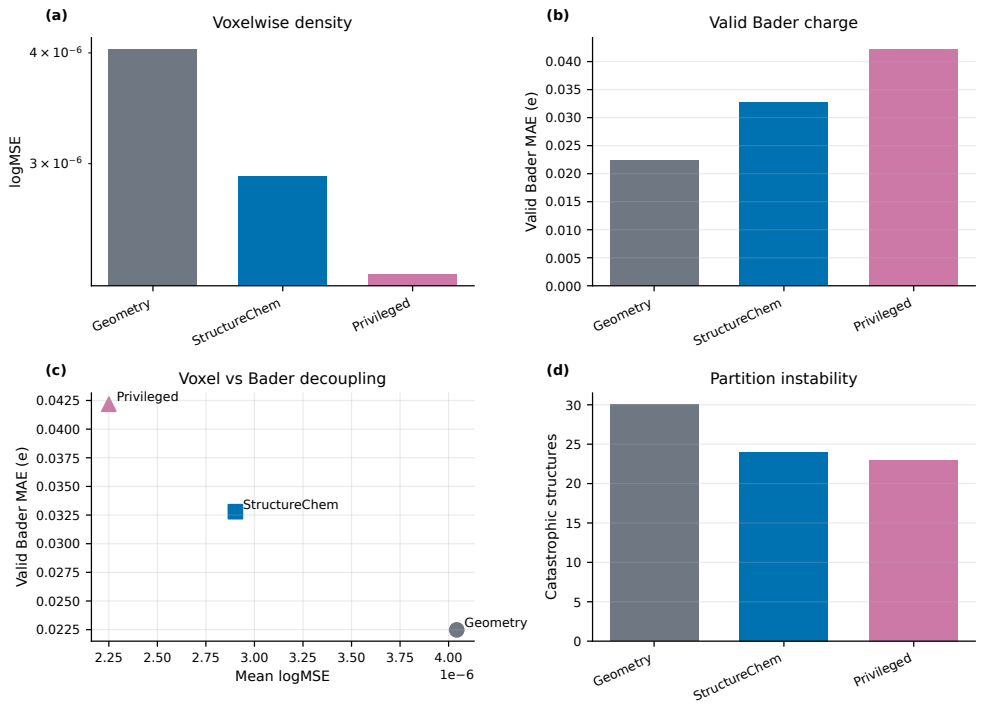


Fig. 5 Voxelwise density accuracy and Bader-derived charge accuracy are decoupled. The models rank differently under canonical voxelwise density metrics and valid Bader charge errors. UNet-Privileged has the lowest voxelwise density error, followed by UNet-StructureChem and UNet-Geometry. In contrast, UNet-Geometry gives the lowest valid Bader MAE, while UNet-Privileged gives the largest valid Bader MAE. Catastrophic Bader counts follow a third trend, showing that mean valid charge accuracy and severe partition instability are distinct aspects of Bader reliability.

$d_{12} < 2.42$ Å and normal-distance configurations with $d_{12} \geq 2.42$ Å, and compared Bader statistics across UNet-Geometry, UNet-StructureChem, and UNet-Privileged.

Al was the only element with substantial invalid or catastrophic Bader partitions. For compressed Al, UNet-Geometry produced a valid Bader MAE of 0.2319 e, an invalid fraction of 18.9%, and 44 catastrophic atom-level failures. UNet-StructureChem increased the compressed-Al valid Bader MAE to 0.3928 e, but reduced the invalid fraction to 13.7% and the catastrophic count to 42. UNet-Privileged achieved the best compressed-Al tradeoff among the three models, reducing the valid Bader MAE to 0.2053 e, the invalid fraction to 14.7%, and the catastrophic count to 30. Thus, privileged perturbation metadata improved compressed-Al Bader robustness, whereas deployable chemical conditioning alone reduced invalidity but worsened the mean valid charge error.

The normal-distance Al regime showed a different pattern. For $d_{12} \geq 2.42$ Å, UNet-Geometry achieved the lowest valid Bader MAE, 0.03967 e, with an invalid fraction of 0.80% and 16 catastrophic atom-level failures. UNet-StructureChem increased the valid Bader MAE to 0.06016 e, but reduced the invalid fraction to 0.23% and the

Table 2 Headline density, Bader, and topology metrics for the three-model hierarchy. Geom. denotes UNet-Geometry, StructChem denotes UNet-StructureChem, and Priv. denotes UNet-Privileged. Raw, Valid, and Δq denote raw Bader MAE, valid-partition Bader MAE, and two-atom charge-asymmetry MAE, respectively, all in units of electrons. Cat. structs. denotes the number of catastrophic Bader-failure structures.

Model	Density			Bader / topology			
	logMSE	L_1 (%)	sumRel	Raw MAE	Valid MAE	Δq MAE	Cat. structs.
Geom.	4.0419×10^{-6}	0.3389	1.6352×10^{-4}	0.04733	0.02248	0.09467	30
StructChem	2.9018×10^{-6}	0.2959	4.0612×10^{-4}	0.04773	0.03280	0.09546	24
Priv.	2.2499×10^{-6}	0.2770	1.7243×10^{-4}	0.05834	0.04217	0.11668	23

catastrophic count to 6. UNet-Privileged produced the largest normal-distance Al Bader error, with a valid MAE of 0.10857 e, while its invalid fraction was 0.46% and its catastrophic count was 16. Therefore, the deployable chemistry-conditioned model reduced rare normal-distance Al failures but degraded typical valid Al charge accuracy, whereas the privileged model degraded normal-distance Al Bader charges most strongly. This regime-specific degradation indicates that privileged metadata does not uniformly improve downstream topology-sensitive observables, even when it improves the global density field. In particular, the information that helps the model in compressed Al configurations can be associated with less favorable Bader-charge behavior at normal Al distances. The origin of this normal-distance Al degradation remains an open question, but it reinforces the need to evaluate topology-sensitive quantities separately by element and distance regime.

Fe and Ni behaved differently from Al. For both elements, all partitions were valid in both distance regimes and no catastrophic failures occurred. In compressed Fe, the valid Bader MAE decreased from 0.01032 e for UNet-Geometry to 0.00848 e for UNet-StructureChem and 0.00809 e for UNet-Privileged. At normal Fe distances, UNet-StructureChem achieved the lowest valid Bader MAE, 0.00470 e, compared with 0.00583 e for UNet-Geometry and 0.00586 e for UNet-Privileged. Thus, chemical conditioning improved Fe Bader charges, while privileged metadata provided little additional benefit at normal Fe distances.

Ni also showed a generally favorable response to conditioning. For compressed Ni, the valid Bader MAE decreased from 0.01401 e for UNet-Geometry to 0.00916 e for UNet-StructureChem and 0.00908 e for UNet-Privileged. For normal-distance Ni, the valid Bader MAE decreased from 0.00603 e to 0.00536 e with UNet-StructureChem and to 0.00492 e with UNet-Privileged. Unlike Al, Ni showed improved Bader charge accuracy as voxelwise density accuracy improved.

The element-distance analysis therefore resolves the aggregate decoupling. For Fe and Ni, chemical conditioning generally improved valid Bader charge accuracy and preserved full partition validity. For Al, however, the response was topology-sensitive and distance-dependent. UNet-StructureChem reduced invalidity and catastrophic failures,

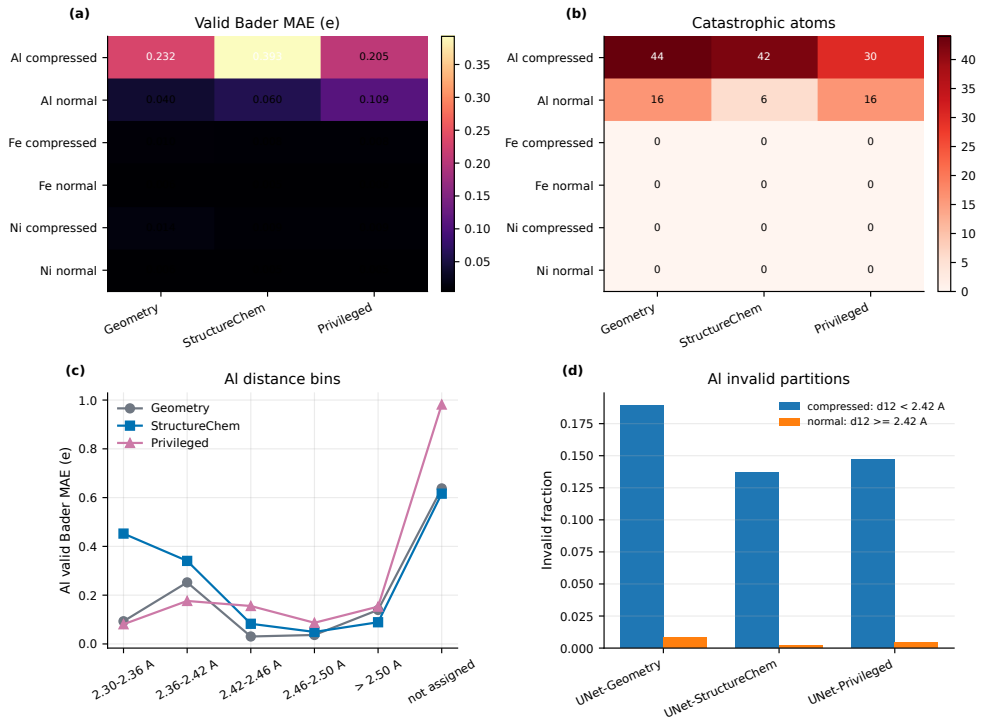


Fig. 6 Element- and distance-dependent Bader regimes. Bader statistics were stratified by element and by compressed or normal interatomic distance. Al exhibits topology-sensitive behavior, especially in compressed configurations, with invalid and catastrophic partitions concentrated in this element. Fe and Ni remain valid across the evaluated regimes and generally show improved Bader charge accuracy with chemical conditioning. The aggregate decoupling between voxelwise density accuracy and Bader accuracy is therefore driven primarily by Al basin topology.

especially at normal distances, but increased the mean valid Al charge error. UNet-Privileged improved compressed-Al behavior and reduced compressed catastrophic failures, but substantially degraded normal-distance Al valid Bader charges. Thus, the decoupling between voxelwise density accuracy and Bader accuracy is driven primarily by Al basin topology, while Fe and Ni behave more consistently with the voxelwise density improvements.

The element- and distance-resolved Bader regimes are summarized in Fig. 6.

Catastrophic partition instability is a distinct failure mode

The Bader-error distribution contained a small group of discrete catastrophic failures that could not be described as a smooth continuation of ordinary prediction error. We defined a catastrophic event as a structure with maximum atomwise Bader error greater than 1 e. Under this definition, UNet-Geometry produced 30 catastrophic structures, UNet-StructureChem produced 24, and UNet-Privileged produced 23. The

corresponding catastrophic structure fractions were 1.03%, 0.82%, and 0.79%, respectively. Because each test structure contained two atoms, these corresponded to 60, 48, and 46 catastrophic atom-level failures.

Pairwise overlap analysis showed that the reduction in catastrophic counts did not simply correspond to a nested subset of failures. Between UNet-Geometry and UNet-StructureChem, 14 catastrophic structures were shared, 16 were catastrophic only for UNet-Geometry, and 10 were catastrophic only for UNet-StructureChem. Thus, chemical conditioning repaired some geometry-only failures but also introduced a smaller number of new catastrophic cases. A similar exchange occurred between UNet-StructureChem and UNet-Privileged, with 10 shared failures, 14 StructureChem-only failures, and 13 Privileged-only failures. Therefore, changes in input representation alter which structures undergo severe Bader partition failure, not only how many such failures occur. This supports the interpretation of catastrophic Bader instability as a discrete topology-sensitive failure mode rather than as a smooth monotonic consequence of average density accuracy.

The catastrophic subpopulation differed sharply from the remaining structures for all three models. For UNet-Geometry, the median maximum atom error was 2.924 e in catastrophic structures but only 0.00744 e in non-catastrophic structures. For UNet-StructureChem, the corresponding values were 2.889 e and 0.00690 e, and for UNet-Privileged they were 2.964 e and 0.00722 e. The same separation appeared in the two-atom charge imbalance. The median charge imbalance in catastrophic structures was 5.849 e for UNet-Geometry, 5.778 e for UNet-StructureChem, and 5.928 e for UNet-Privileged, whereas the corresponding non-catastrophic medians were 0.01487 e, 0.01380 e, and 0.01444 e. This abrupt separation indicates a distinct partition-instability regime rather than a heavy but continuous tail of the ordinary Bader-error distribution.

Catastrophic counts did not follow the same ordering as valid Bader MAE. UNet-Geometry had the lowest valid Bader MAE, 0.02248 e, but the largest number of catastrophic structures. UNet-Privileged had the highest valid Bader MAE, 0.04217 e, but the fewest catastrophic structures. UNet-StructureChem was intermediate in both respects, with valid Bader MAE 0.03280 e and 24 catastrophic structures. Therefore, mean valid charge accuracy and catastrophic partition stability are not interchangeable measures of Bader reliability.

The upper tails also reflected this tradeoff. The p95 maximum atom error was 0.0751 e for UNet-Geometry, 0.0899 e for UNet-StructureChem, and 0.1489 e for UNet-Privileged. The p95 charge imbalance similarly increased from 0.1501 e to 0.1798 e and 0.2978 e. Thus, although chemical conditioning and privileged metadata reduced the number of catastrophic events, they increased the upper-tail error among the remaining non-catastrophic or valid partitions. This suggests a tradeoff in which richer input representations reduce the frequency of the most severe basin collapses, but do not uniformly suppress the upper tail of ordinary valid-partition errors.

The element-distance analysis showed that these catastrophic events were concentrated in Al configurations. Fe and Ni produced no catastrophic failures in either the compressed or normal-distance regimes for any of the three models. In contrast, compressed Al accounted for most catastrophic cases, with 44 atom-level failures for

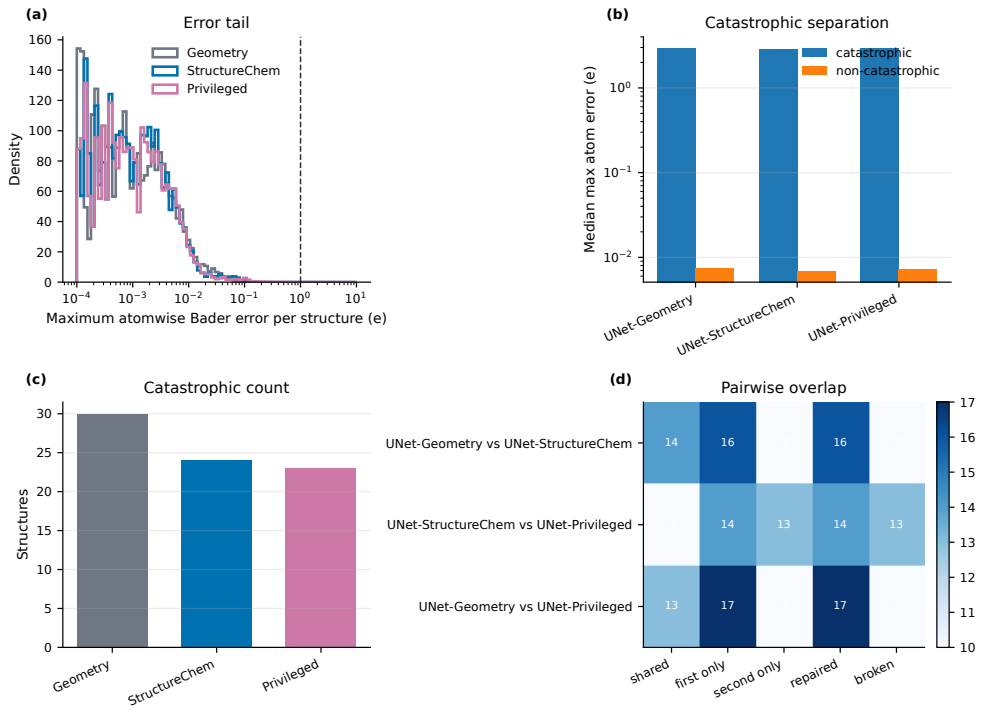


Fig. 7 Catastrophic Bader partition instability as a separate failure mode. Catastrophic events were defined as structures with maximum atomwise Bader error greater than 1 e. These failures form a discrete large-error tail that is not equivalent to mean valid Bader MAE. The three models differ in both the number of catastrophic structures and the upper-tail error among non-catastrophic or valid partitions, indicating that catastrophic partition stability should be reported separately from aggregate Bader charge accuracy.

UNet-Geometry, 42 for UNet-StructureChem, and 30 for UNet-Privileged. Normal-distance Al still produced catastrophic events, but at much lower rates: 16 for UNet-Geometry, 6 for UNet-StructureChem, and 16 for UNet-Privileged. This confirms that catastrophic partitioning is primarily an Al basin-topology problem, with compressed Al configurations forming the most unstable regime.

Saved Bader outputs and one-dimensional diagnostics should therefore be interpreted as topology proxies rather than complete explanations. Bader maxima counts, midpoint density, minimum bond-line density, and line-minimum curvature can identify some extreme cases, but they do not uniquely determine the three-dimensional zero-flux basin structure. The present results show that catastrophic partition instability depends on the full density-gradient topology and can vary independently from both voxelwise density error and mean valid Bader charge error. Consequently, catastrophic failure counts should be reported alongside, rather than replaced by, aggregate Bader MAE.

The catastrophic-failure regime is summarized in Fig. 7, which separates ordinary Bader errors from large topology-sensitive partition failures.

Table 3 Runtime comparison between U-Net density inference and historical SIESTA SCF calculations. U-Net timings are reported per structure on the elemental test split and averaged over five timed repetitions on an NVIDIA RTX PRO 6000 Blackwell Max-Q GPU. SIESTA timings were parsed from historical SCF output files and are not hardware matched to the U-Net benchmark.

Model/source	n	Total time	Forward time	Peak CUDA memory
UNet-Geometry	2,924	6.23 ms	0.94 ms	26.52 MB
UNet-StructureChem	2,924	6.26 ms	0.93 ms	26.91 MB
UNet-Privileged	2,924	6.47 ms	0.95 ms	30.44 MB
SIESTA SCF	64	175.9 s	–	–

Runtime benchmark

Finally, we measured inference runtime and memory use for the three U-Net models and compared them with historical SIESTA self-consistent-field timings. The neural-network benchmark used the full elemental test split of 2,924 structures and five timed repetitions on an NVIDIA RTX PRO 6000 Blackwell Max-Q Workstation Edition GPU. For the U-Net models, the reported total runtime includes input construction, host-to-device transfer, neural-network inference, and density reconstruction, whereas the forward-only runtime measures only the neural-network forward pass. The SIESTA timings were parsed from completed SCF output files for 64 structures and are therefore not a prospective hardware-matched comparison; they are included only to provide an approximate scale for the cost difference between density-surrogate inference and conventional SCF calculation. The runtime and memory results are summarized in Table 3.

All three U-Net models operated in the millisecond-per-structure regime. UNet-Geometry required 6.23 ms per structure, UNet-StructureChem required 6.26 ms, and UNet-Privileged required 6.47 ms. The forward pass itself took approximately 0.94 ms for all three models, indicating that the additional element channel in UNet-StructureChem and the larger privileged input representation introduced negligible neural-network overhead in this fixed-grid two-atom benchmark. Peak CUDA memory remained below 31 MB for all models. By comparison, the historical SIESTA SCF wall time had a mean of 175.9 s per structure and a median of 111.1 s. Using the mean SIESTA timing as a reference, the U-Net inference workflow is approximately 2.7×10^4–2.8×10^4 times faster per structure in this benchmark. This comparison should be interpreted as an order-of-magnitude runtime contrast rather than as a hardware-normalized speedup measurement.

Discussion

We have shown that deployable chemical conditioning improves volumetric electron-density prediction in a controlled elemental benchmark, but that this improvement does not transfer uniformly to topology-sensitive downstream observables. UNet-StructureChem improves log-density and relative L_1 errors relative to UNet-Geometry while adding only one deployable element channel to the same U-Net backbone. Because UNet-Geometry and UNet-StructureChem have nearly identical parameter

counts, 789,433 and 789,757, respectively, the improvement cannot be attributed to a meaningful increase in model capacity. Instead, it reflects the information added by the explicit chemical-identity channel. UNet-Privileged further improves voxelwise accuracy when generator-derived perturbation metadata is supplied, confirming that synthetic-process information remains predictive in this benchmark. The central implication is that density surrogates should be judged not only by average field reconstruction, but also by the downstream operations for which their outputs will be used.

The element-resolved Bader analysis points to Al as the source of the aggregate decoupling between voxelwise density accuracy and Bader accuracy. Al was the only element with substantial invalid or catastrophic Bader partitions, and the instability was amplified in compressed configurations. For compressed Al, the invalid fraction was 18.9% for UNet-Geometry, 13.7% for UNet-StructureChem, and 14.7% for UNet-Privileged, whereas Fe and Ni retained valid partitions in all evaluated distance regimes. Catastrophic failures showed the same element specificity: compressed Al produced 44, 42, and 30 atom-level catastrophic failures for UNet-Geometry, UNet-StructureChem, and UNet-Privileged, respectively, while Fe and Ni produced none.

One interpretation, consistent with these element- and distance-resolved results, is that the Al-specific instability reflects comparatively shallow and diffuse interatomic density features in this controlled two-atom regime. In such a regime, small spatial errors in the predicted density or its gradient can move zero-flux boundaries disproportionately, especially when the inter-basin region is narrow. This interpretation is also consistent with the radial-profile diagnostics: Ni and Fe have larger radial density errors than Al, indicating sharper and more element-specific local density features, yet their Bader partitions are more stable. Thus, the density features that are harder to reconstruct voxelwise can still define more robust basin topology, whereas smoother Al densities may be easier in average field metrics but more sensitive to small gradient-field perturbations. This counterintuitive separation between local reconstruction difficulty and topological stability is one reason why voxelwise and Bader-based metrics should be reported separately.

This mechanistic interpretation should nevertheless be regarded as indirect. The present analysis retained Bader outputs, maxima counts, bond-line proxies, and charge errors, but not full basin-label fields or zero-flux surface displacements. The results therefore support an Al-specific topology-sensitivity regime, rather than providing a direct reconstruction of the basin-boundary motion responsible for each failure.

The remaining spatial and spectral diagnostics further support the broader conclusion that residual errors are structured rather than uniform. Errors are systematically larger near atomic sites than in the interstitial region, and the radial-profile errors follow the ordering Ni $>$ Fe $>$ Al. Thus, density reconstruction difficulty is spatially and chemically structured even when global metrics are small. Fourier-shell diagnostics show that the dominant density-error magnitude is concentrated at lower spatial frequencies under the chosen normalization, while bond-midpoint diagnostics show that interatomic distance alone does not explain the full error behavior across all elements. Together, these observations indicate that future models should not only minimize

average voxelwise losses, but should also incorporate diagnostics or objectives that are sensitive to physically meaningful regions and downstream topological observables.

The scope of the present model is deliberately limited. The elemental test set evaluates held-out perturbations of known two-atom Al, Fe, and Ni parent structures. It therefore measures interpolation within represented elemental structural families, not generalization to arbitrary crystal prototypes, larger atom counts, different grid geometries, mixed compositions, or unseen chemical elements. In addition, the Fe and Ni calculations were performed without spin polarization; the resulting densities should therefore be interpreted as non-magnetic reference densities under fixed numerical settings, not as predictions of magnetic ground-state electronic structure. This controlled setting is useful because it isolates the effects of explicit chemical conditioning, privileged perturbation metadata, and downstream Bader topology. However, alloy-capable, magnetic, or chemically transferable density models will require richer input representations and broader electronic-structure reference data than those used here.

These limitations point to several directions for future work. First, models intended for mixed-composition systems should represent local species identity explicitly, so that different site occupations can lead to different predicted densities. Active-learning and uncertainty-aware density models may also be useful for extending this framework beyond controlled elemental perturbations while limiting the number of required KS-DFT calculations [33, 51]. Second, models intended for downstream charge analysis should incorporate topology-sensitive validation earlier in model development. The present experiments already provide a simple example of this direction: UNet-Privileged included an Al-only auxiliary charge-asymmetry term, but this soft Δq objective did not prevent the model from having the largest Bader-derived Δq MAE. Future work should therefore explore losses based more directly on density-gradient structure, differentiable approximations to basin stability, or post-training model selection using Bader-aware validation metrics. Third, the present study evaluates density prediction but does not compute energy errors; therefore chemical accuracy cannot be assessed from the current outputs alone. Future work could couple predicted densities to an energy postprocessing procedure or train a density-to-energy model, enabling direct evaluation of energy accuracy and comparison with chemical-accuracy thresholds.

The present results establish that explicit deployable chemical conditioning improves volumetric electron-density prediction within a controlled elemental domain, while privileged generator metadata provides an upper bound on what can be achieved when synthetic perturbation information is supplied. However, improved density fields do not automatically imply improved downstream topological observables. Machine-learned electron densities should therefore be validated both as continuous scalar fields and as inputs to the physical analyses for which they are intended. For applications involving atom-partitioned charges, bonding descriptors, or topology-sensitive observables, downstream validation is not optional; it is part of the definition of model reliability.

Methods

Electronic-structure dataset

This study uses the single-element electronic-structure dataset described in a companion data manuscript currently under revision [52]. The dataset contains 30,000 fully converged Kohn–Sham density functional theory (KS-DFT) calculations for three elemental systems: Al, Fe, and Ni, with 10,000 configurations per element. Each configuration provides a periodic two-atom structure and the corresponding valence-electron density represented on a uniform three-dimensional real-space grid. The dataset was designed as a controlled benchmark for density-based machine-learning models, rather than as a broad sampling of thermodynamic phase space.

For each element, configurations were generated from a common two-atom reference topology and subjected to controlled structural perturbations. The perturbation classes included atomic displacement, isotropic strain, anisotropic strain, shear, and combined perturbations. The perturbation magnitudes were deliberately restricted to remain close to the reference crystalline geometry, so that the resulting structures sample near-reference local environments while preserving a consistent topology across elements. Consequently, the dataset supports systematic comparisons between Al, Fe, and Ni under matched structural and numerical conditions, but should not be interpreted as a dataset of relaxed finite-temperature trajectories, defects, surfaces, interfaces, phase transitions, or arbitrary crystal prototypes.

All elemental structures used a $30 \times 30 \times 50$ density grid, corresponding to 45,000 density values per configuration. The cell volume and grid spacing were retained for every structure so that grid sums could be converted into integrated electron counts. In this work, the density grids were used as supervised targets for volumetric learning, while the structural information was used to construct the input channels.

The electronic-structure calculations were performed with SIESTA 5.4.2 [53, 54] using the generalized-gradient approximation of Perdew, Burke, and Ernzerhof (GGA-PBE) and optimized norm-conserving Vanderbilt pseudopotentials in PSML format. The pseudopotential valence definitions were used to determine the known number of valence electrons N_e for each configuration. All calculations used a triple-ζ polarized atomic-orbital basis, a PAO energy shift of 0.01 eV, a real-space integration grid cutoff of 400 Ry, diagonalization of the Kohn–Sham Hamiltonian, a maximum of 150 self-consistent-field iterations, density-matrix convergence tolerance `DM.Tolerance` $= 10^{-4}$, density-matrix mixing weight `DM.MixingWeight` $= 0.10$, and `DM.NumberPulay` $= 8$. Electronic occupations were treated with Fermi–Dirac smearing using `ElectronicTemperature` $= 300$ K, and Brillouin-zone sampling used a $12 \times 12 \times 12$ Monkhorst–Pack grid. Spin polarization was not included; therefore, Fe and Ni densities should be interpreted as non-magnetic reference densities under fixed numerical settings rather than as magnetic ground-state densities. Detailed reporting of numerical settings is important because DFT results can depend sensitively on pseudopotentials, basis sets, grids, and convergence choices [55].

Automated quality-control procedures in the dataset generation workflow verified self-consistent-field convergence, detected runtime failures, checked the completeness

of required output files, and parsed final energies and Fermi levels. All 30,000 calculations included in the dataset converged successfully. For the present machine-learning study, the frozen split contained 23,393 training structures, 2,924 validation structures, and 2,924 final test structures, for a total of 29,241 structures. The remaining 759 converged calculations were retained in the dataset release but were not included in the frozen model-comparison split used here. No test structure was used for model fitting, checkpoint selection, hyperparameter tuning, or model-selection decisions.

Because the dataset generation procedure used one unperturbed parent template per element, configurations derived from the same elemental parent appear across the training, validation, and test subsets. Under a strict parent-group definition based on the pair element :: structure, the dataset contains three parent groups, corresponding to the Al, Fe, and Ni reference templates, and all three groups cross the split boundaries. Consequently, the elemental test set evaluates held-out perturbations within known elemental parent families. It measures configurational interpolation across represented elemental templates and perturbation regimes, not strict generalization to unseen crystal prototypes, parent structures, atom counts, grid geometries, mixed compositions, or chemical elements.

Split files and hashes were retained in the reproducibility manifest to ensure that all reported model comparisons use the same frozen partition.

Volumetric input representations

Each configuration was converted into a multichannel tensor defined on the same real-space grid as the target electron density. We distinguish between deployable structure-derived inputs, which are computable from the final atomic structure, and privileged generator-derived inputs, which are known in the synthetic benchmark but are not generally available for arbitrary structures.

The deployable structural representation consisted of two atom-centered density-prior channels. We denote these structural-prior fields by $\rho_{S,k}$, where S indicates a structure-derived prior and $k \in \{1, 2\}$ indexes the Gaussian width. For each grid point r_i, the k-th structural prior was computed as

$$\rho_{S,k}(r_i) = \sum_a w_a \exp\left[-\frac{d_{\mathrm{MIC}}(r_i, R_a)^2}{2\sigma_k^2}\right],$$

where R_a is the position of atom a, d_{MIC} is the minimum-image distance under periodic boundary conditions, and $w_a = N_e/N_{\mathrm{atoms}}$. Here N_e denotes the total number of valence electrons in the configuration, as determined by the pseudopotential valence definitions, and N_{atoms} is the number of atoms in the structure. Thus, w_a assigns each atom an equal share of the known valence-electron count. Two width parameters were used, $\sigma_1 = 0.05$ Å and $\sigma_2 = 0.10$ Å. These widths were chosen to provide sharply localized atom-centered priors rather than smooth approximate densities: the narrower channel marks the atomic sites at near-voxel scale, while the broader channel provides a slightly more spatially extended structural cue. Each prior was normalized to integrate to N_e on the discrete grid and then transformed using $\log(1+\rho_{S,k})$. These channels provide smooth structure-derived fields aligned with the target density grid.

UNet-StructureChem additionally used an explicit element channel. This channel was spatially constant over the grid and encoded the elemental identity as a scalar value, with Al, Fe, and Ni mapped to fixed numerical labels. Because the element label is known from the final atomic structure, this channel is deployable at inference time for structures within the represented chemical domain.

UNet-Privileged included the same structural-prior and element channels, together with generator-derived perturbation metadata. These metadata channels encoded information such as the perturbation class and perturbation amplitudes used to construct the synthetic configuration. Although useful for measuring how much additional information is available in the controlled benchmark, these descriptors are not intrinsic properties of an arbitrary final structure. UNet-Privileged was therefore interpreted as a metadata-assisted upper-bound control rather than as a deployable model.

The three principal models differed only in their input channels: UNet-Geometry used the two structural-prior channels; UNet-StructureChem used the two structural-prior channels plus the element channel; and UNet-Privileged used the structural priors, the element channel, and the privileged perturbation-metadata channels. All models predicted the same target representation on the same output grid.

Neural-network architecture

The network received a multichannel volumetric input tensor and predicted a single scalar field on the original real-space density grid. Using the same backbone for all models ensured that differences in performance reflected the input representation rather than changes in architecture.

The encoder used channel widths of 12, 24, and 48, followed by a 96-channel bottleneck. The decoder used channel widths of 48, 24, and 12. Each encoder, bottleneck, and decoder block contained two $3 \times 3 \times 3$ convolutions, each followed by Group Normalization and a SiLU activation. All convolutions used unit stride and one-voxel padding. Downsampling was performed with $2 \times 2 \times 2$ max-pooling, and upsampling was performed with $2 \times 2 \times 2$ transposed convolutions with stride 2. Decoder features were concatenated with the corresponding encoder features through skip connections, and a final $1 \times 1 \times 1$ convolution produced the output field.

Group Normalization used four groups for 12-channel blocks and eight groups for 24-, 48-, and 96-channel blocks. The resulting trainable parameter counts were 789,433 for UNet-Geometry, 789,757 for UNet-StructureChem, and 792,673 for UNet-Privileged. The small differences in parameter count arise only from the different number of input channels.

Prediction target and density reconstruction

The supervised target was the log-transformed DFT valence-electron density,

$$y(\mathbf{r}) = \log\left[1 + \rho_{\mathrm{DFT}}(\mathbf{r})\right].$$

This transformation was used because the electron density spans a broad dynamic range, with large values near atomic centers and much smaller values in interstitial

regions. Training directly on the linear density would make the reconstruction loss dominated by the largest near-atom values, whereas the logarithmic transform compresses high-density regions and gives relatively greater weight to lower-density parts of the cell. The additive constant ensures that grid points with very small or zero density remain well defined. Thus, the model was trained to reconstruct both high-density atomic features and lower-density interstitial structure in a numerically stable target space. The U-Net produced an unrestricted output field $z_\theta(\mathbf{r})$, which was mapped to a non-negative transformed density using a softplus activation:

$$\hat{y}(\mathbf{r}) = \mathrm{softplus}\left(z_\theta(\mathbf{r})\right).$$

The predicted density in linear space was then reconstructed by applying the inverse transform,

$$\hat{\rho}(\mathbf{r}) = \exp\left(\hat{y}(\mathbf{r})\right) - 1.$$

No post-hoc electron-count rescaling was applied during the canonical density evaluation. The reported logMSE, relative L_1, density MAE, density RMSE, NRMSE, and raw sumRel metrics therefore describe the unrescaled network outputs.

Training objective

The training objective was designed to control two complementary aspects of the predicted density. The first term penalizes voxelwise errors in the log-transformed density and therefore drives the model to reproduce the local spatial structure of the DFT density field. The second term penalizes errors in the integrated electron count and therefore discourages predictions that have the correct shape but an incorrect total charge. The principal training objective was

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\log\rho} + \beta_{\mathrm{sumRel}}\mathcal{L}_{\mathrm{sumRel}}.$$

For the main deployable and geometry-only long fine-tuned checkpoints, we used $\beta_{\mathrm{sumRel}} = 0.01$. This weighting made the integrated-density term a weak auxiliary constraint rather than the dominant training signal. The log-density reconstruction loss was

$$\mathcal{L}_{\log\rho} = \frac{1}{BN_g}\sum_{b=1}^{B}\sum_{i=1}^{N_g}\left[\hat{y}_{b,i} - y_{b,i}\right]^2,$$

where B is the batch size and N_g is the number of grid points per structure. Here $y_{b,i}$ and $\hat{y}_{b,i}$ denote the reference and predicted log-density values at grid point i of structure b, respectively. This term measures the average voxelwise discrepancy in the transformed target space. Because the logarithmic transform compresses the large dynamic range of the density, $\mathcal{L}_{\log\rho}$ encourages the model to reconstruct both high-density near-atom regions and lower-density interstitial regions. The relative integrated-density loss was

$$\mathcal{L}_{\mathrm{sumRel}} = \frac{1}{B}\sum_{b=1}^{B}\frac{\left|\Delta V_b\sum_i \hat{\rho}_{b,i} - \Delta V_b\sum_i \rho_{b,i}\right|}{\max\left[\Delta V_b\sum_i \rho_{b,i} + 10^{-12}, 10^{-3}\right]}.$$

The 10^{-3} floor is a defensive safeguard for pathological near-zero reference integrals and was not expected to be active for the present dataset, whose reference electron counts are far from zero. Here $\rho_{b,i}$ and $\hat{\rho}_{b,i}$ are the reference and predicted linear-space densities, and ΔV_b is the voxel volume for structure b. The quantities

$$\Delta V_b \sum_i \rho_{b,i} \quad \text{and} \quad \Delta V_b \sum_i \hat{\rho}_{b,i}$$

are the reference and predicted grid-integrated electron counts. Thus, $\mathcal{L}_{\text{sumRel}}$ penalizes the relative error in the total number of electrons represented by the predicted density. The small constants in the denominator provide numerical stability, although the reference electron counts in this dataset are far from zero. Because prediction and target densities are evaluated on the same grid for each structure, the voxel-volume factor is common to the predicted and reference sums within that structure. The term can therefore be interpreted as a relative grid-sum penalty, while retaining the physically meaningful electron-count units in its definition. No post-hoc electron-count rescaling was applied during the canonical density evaluation, so the reported raw sumRel values measure the charge-conservation behavior learned by the model itself.

Training and checkpoint selection

All final models were obtained by low-learning-rate fine-tuning from preliminary 20-epoch checkpoints. Fine-tuning used a learning rate of 10^{-5}, batch size 2, and reset optimizer state. Checkpoints were selected using validation data only, based on the minimum total validation loss for the corresponding training objective. No test result was used for checkpoint selection, hyperparameter tuning, or model-selection decisions.

The primary deployable checkpoint, UNet-StructureChem, was fine-tuned for 60 epochs with $\beta_{\text{sumRel}} = 0.01$. UNet-Geometry was fine-tuned under the same optimization settings, with $\beta_{\text{sumRel}} = 0.01$. UNet-Privileged was fine-tuned for 60 epochs with $\beta_{\text{sumRel}} = 0.01$ and an additional Al-only auxiliary charge-asymmetry term with $\beta_{\Delta q} = 0.01$.

Canonical density evaluation

Each model was evaluated independently on all 2,924 test structures. All canonical density metrics were computed from unrescaled network predictions. Metrics were first calculated per structure and then macro-averaged across the test set, so that each structure contributed equally to the reported mean values. The canonical evaluation was designed to measure complementary aspects of density reconstruction. The log-space error measures how well the model reproduces the transformed target used during training. The linear-space density errors measure the absolute discrepancy in the physical density field. The relative L_1 error measures the total absolute density mismatch relative to the total reference density magnitude. Finally, the raw sumRel metric measures whether the predicted density integrates to the correct number of electrons. Together, these metrics distinguish local field accuracy from global charge

conservation. The log-space mean squared error was

$$\mathrm{logMSE} = \frac{1}{N_g} \sum_{i=1}^{N_g} \left[\hat{y}_i - y_i\right]^2 .$$

This metric is the direct evaluation analogue of the main reconstruction loss. Because it is computed in the log-transformed target space, it is less dominated by the largest near-atom density values than a linear-space squared error would be. It therefore measures reconstruction accuracy in the space in which the model was trained. Linear-space density errors were computed after transforming the prediction back to physical density units. The density mean absolute error was

$$\mathrm{MAE}_\rho = \frac{1}{N_g} \sum_i \left|\hat{\rho}_i - \rho_i\right| ,$$

and the density root mean squared error was

$$\mathrm{RMSE}_\rho = \sqrt{\frac{1}{N_g} \sum_i \left(\hat{\rho}_i - \rho_i\right)^2}.$$

Both quantities are expressed in density units and therefore measure the absolute size of the error in the predicted electron-density field. The MAE gives the typical absolute voxelwise error, whereas the RMSE penalizes larger local errors more strongly. The relative L_1 error was

$$L_{1,\mathrm{rel}} = \frac{\sum_i \left|\hat{\rho}_i - \rho_i\right|}{\sum_i \left|\rho_i\right|}.$$

In the Results, this quantity is reported as a percentage, $100 L_{1,\mathrm{rel}}$. This metric provides an intuitive measure of the total absolute density error relative to the total amount of reference density in the grid. It is useful as a headline field-accuracy metric because it is dimensionless and directly comparable across structures. The normalized root mean squared error was calculated in linear density space as

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{N_g} \sum_{i=1}^{N_g} \left(\hat{\rho}_i - \rho_i\right)^2}}{\max_i \rho_i - \min_i \rho_i + 10^{-12}}.$$

This metric rescales the linear-space RMSE by the dynamic range of the reference density in each structure. It therefore measures the size of the typical squared error relative to the density contrast present in that configuration. The raw relative electron-count error was

$$\mathrm{sumRel} = \frac{\left|\Delta V \sum_i \hat{\rho}_i - \Delta V \sum_i \rho_i\right|}{\Delta V \sum_i \rho_i}.$$

The numerator is the absolute error in the grid-integrated electron count, and the denominator is the reference integrated electron count. Thus, sumRel measures global charge conservation: a small value means that the predicted density integrates to

nearly the same number of electrons as the DFT reference density. This metric is distinct from voxelwise field accuracy, because a prediction can have small local density errors but still exhibit a biased total integral, or conversely have a good total electron count while making spatially structured local errors. No post-hoc electron-count normalization or density rescaling was applied before calculating these canonical density metrics. Consequently, the reported logMSE, relative L_1, density MAE, density RMSE, NRMSE, and raw sumRel values describe the native outputs of the trained networks. In particular, raw sumRel measures charge conservation learned by the model itself rather than charge conservation imposed after prediction.

Regional density analysis

The real-space grid was divided into near-atom and interstitial regions using the minimum periodic distance from each grid point to any atom. Grid points satisfying

$$d_{\min}(\mathbf{r}_i) < 1.2 \text{ Å}$$

were assigned to the near-atom region, and all remaining grid points were assigned to the interstitial region. The value 1.2 Å was chosen as an operational near-atom cutoff. It is small enough to isolate the high-density regions around the atomic sites, where the density gradients are largest and reconstruction is expected to be most difficult, while leaving a substantial complementary interstitial region for comparison. This cutoff is not intended to define an atomic radius, bonding region, or Bader basin; it is used only to stratify the voxelwise density error into near-atom and interstitial contributions. Density MAE was calculated independently in both regions for every structure and then aggregated over the test set. The near-to-interstitial error ratio was used to quantify the spatial concentration of absolute density error around the atomic sites.

Radial-profile analysis

Element-resolved radial density profiles were constructed around each atom using periodic minimum-image distances. Radial profiles used shell widths of 0.1 Å from $r = 0$ to $r = 3.0$ Å, with bins $[r_{\text{in}}, r_{\text{out}})$. Grid points were represented by voxel-center fractional coordinates, and distances were computed using the periodic minimum-image convention before transformation to Cartesian coordinates. For each atom and shell, predicted and reference densities were averaged over all voxels in the shell, and the mean signed error, MAE, and RMSE were recorded in e/bohr^3. Shells with no voxels were omitted. No nearest-atom assignment was applied; therefore, a voxel could contribute to radial shells around more than one atom.

Fourier-space analysis

Fourier diagnostics were computed from the density-error field

$$\delta\rho = \hat{\rho} - \rho.$$

For each structure, the three-dimensional discrete Fourier transform was normalized by $1/\sqrt{N_g}$:

$$\widetilde{\delta\rho}(\mathbf{k}) = \frac{1}{\sqrt{N_g}}\mathrm{FFTN}[\delta\rho](\mathbf{k}).$$

The reciprocal-space magnitude $|\mathbf{k}|$ was computed from the reciprocal lattice vectors, and modes were assigned to 24 linearly spaced shells between zero and the maximum sampled $|\mathbf{k}|$. For each shell, the recorded statistic was

$$E_{\mathrm{shell}} = \left\langle \left| \widetilde{\delta\rho}(\mathbf{k}) \right| \right\rangle_{\mathbf{k}\in\mathrm{shell}},$$

stored as `mean_abs_fft_error`. The low-frequency band was defined as the lowest quartile of shell midpoints and the high-frequency band as the highest quartile, corresponding to the six lowest and six highest shells. The reported high/low ratio was the mean shell statistic in the high-frequency band divided by the corresponding mean in the low-frequency band. Spearman correlation was used to quantify the monotonic relationship between shell index and E_{shell}.

Bond-midpoint and compressed-pair analysis

For every two-atom structure, the periodic minimum-image interatomic distance d_{12} was calculated. Structures satisfying

$$d_{12} < 2.42 \text{ Å}$$

were classified as compressed configurations. Bond-midpoint diagnostics were then evaluated around the periodic midpoint between the two atoms. The threshold 2.42 Å was chosen as an operational cutoff separating shortened two-atom configurations from the main near-reference distance regime in this dataset. This value is close to the nearest-neighbor distance of fcc Al under the numerical settings used here, and therefore provides a physically interpretable boundary between compressed Al-like configurations and near-reference separations. It is used only for stratified analysis and should not be interpreted as a universal physical bond-length criterion.

The canonical midpoint evaluator used a Gaussian mask centered at the periodic bond midpoint. This mask was used to obtain a smooth local measurement of the density in the interatomic region, rather than relying on a single nearest grid point. Such smoothing reduces sensitivity to the finite grid resolution and to the fact that the exact periodic midpoint does not generally coincide with a voxel center. The Gaussian weighting emphasizes density values closest to the midpoint while still averaging over a small local neighborhood. The mask had width $\sigma_{\mathrm{mid}} = 0.25$ Å and was defined as

$$w(\mathbf{r}) = \exp\left[-\frac{|\mathbf{r} - \mathbf{r}_{\mathrm{mid}}|^2_{\mathrm{MIC}}}{2\sigma^2_{\mathrm{mid}}}\right],$$

where $\mathbf{r}_{\mathrm{mid}}$ is the periodic midpoint and $|\cdot|_{\mathrm{MIC}}$ denotes the periodic minimum-image distance. The width $\sigma_{\mathrm{mid}} = 0.25$ Å was chosen to probe a localized region around the

bond midpoint without averaging over the broader near-atom density. Thus, the resulting midpoint diagnostics quantify reconstruction errors in the interatomic region most relevant to bond-centered density and basin-separation behavior. The mask was normalized only when forming weighted averages. The local midpoint diagnostics included weighted log-density MSE, weighted percent L^1 density error, weighted predicted and reference mean densities, and bias.

A separate midpoint analysis was used for distance-correlation summaries. This analysis evaluated a periodic sphere of radius 0.3 Å centered on the midpoint for atom pairs with $d_{12} \leq 3.2$ Å. For each pair, the evaluator recorded the mean DFT and predicted density inside the sphere, as well as the density at the nearest grid point to the periodic midpoint. The three-model distance-correlation summaries reported in the Results used the absolute nearest-grid midpoint error in e/bohr^3.

Because the Gaussian-mask and midpoint-sphere analyses use different spatial masks and units, they were analyzed separately. Distance dependence was evaluated using Spearman correlations and distance-bin summaries for the full test set and for close Al configurations.

Bader density preparation

Predicted density cubes were generated for every test configuration and model. Before export for Bader analysis, each predicted density was scaled so that its grid-integrated electron count matched that of the corresponding reference density:

$$\hat{\rho}_{\mathrm{scaled}}(\mathbf{r}) = \hat{\rho}(\mathbf{r}) \frac{\sum_i \rho_i}{\sum_i \hat{\rho}_i}.$$

This scaling was applied only in the Bader cube-export workflow and was not used for the canonical voxelwise density evaluation. Uniform positive scaling does not change the direction field of $\nabla\rho$ in exact arithmetic, and therefore does not change zero-flux basin boundaries, but it does scale the electron count integrated within each basin. Reference and predicted cubes used identical cells, grid dimensions, voxel ordering, and atomic coordinates.

Bader partitioning and validity criteria

Bader analysis was performed using the Henkelman grid-based implementation [47–49]. Atomic basins were obtained from reference and predicted density grids, and the density within each basin was integrated to obtain atom-resolved electron counts.

A Bader result was considered valid when Bader completed successfully for both reference and predicted cube files, both `ACF.dat` files could be parsed, reference and predicted atom rows could be matched one-to-one by atom index, and no atom was flagged as collapsed or unstable. An atom was flagged as collapsed or unstable if either the DFT or predicted basin had a parsed Bader charge < 0.1 e or a parsed atomic volume < 1.0 in the `ACF.dat` output. Structure-level validity was conservative: if either atom in the two-atom structure failed this basin-quality test, the entire structure was excluded from valid-only Bader summaries. The Bader charge-sum audit compared

the sum of Bader charges with the corresponding grid-integrated electron count and issued a warning when the discrepancy exceeded 10^{-2} e.

Let q_a^{DFT} and q_a^{pred} denote the Bader-integrated electron counts assigned to atom a from the reference DFT density and the predicted density, respectively. These quantities were obtained by integrating the corresponding density within the Bader basin associated with atom a. For atom a, the absolute Bader charge error was

$$\epsilon_a = \left| q_a^{\mathrm{pred}} - q_a^{\mathrm{DFT}} \right| .$$

The structure-level maximum atom error was

$$\epsilon_{\max} = \max\left(\epsilon_1, \epsilon_2\right) .$$

This metric records the largest atomwise Bader-charge error in the two-atom structure and was used to identify severe partition failures.

The two-atom charge-asymmetry error was

$$\epsilon_{\Delta q} = \left| \left(q_1^{\mathrm{pred}} - q_2^{\mathrm{pred}} \right) - \left(q_1^{\mathrm{DFT}} - q_2^{\mathrm{DFT}} \right) \right| .$$

This quantity measures whether the predicted density preserves the Bader-derived charge difference between the two atoms. It was also used as a structure-level charge-imbalance diagnostic. Bader statistics were stratified by model, element, interatomic distance, perturbation type, and validity status.

Element- and distance-resolved Bader analysis

Bader statistics were stratified by element and interatomic distance. Compressed configurations were defined as

$$d_{12} < 2.42 \text{ Å},$$

and normal-distance configurations were defined as

$$d_{12} \geq 2.42 \text{ Å}.$$

For each model, element, and distance regime, we computed the number of structures, number of atoms, valid and invalid fractions, raw and valid Bader MAE, Δq MAE, median and p95 atomwise absolute errors, median and p95 maximum structure-level atom errors, median and p95 charge-asymmetry error, and catastrophic counts.

Catastrophic Bader failures

A catastrophic Bader failure was defined before analysis as a structure satisfying

$$\epsilon_{\max} > 1 \ e.$$

Failure rates were calculated per structure and per atom. Catastrophic and non-catastrophic populations were compared using maximum atom error, two-atom

charge-asymmetry error, upper-tail errors, element-distance regime, Bader maxima counts, midpoint-density error, and bond-line critical-point proxies.

The number of reported Bader maxima was used as an indirect topology proxy. Full volumetric basin-label grids and zero-flux surface meshes were not retained, so exact basin-boundary displacements could not be reconstructed retrospectively.

Bond-line and topology-proxy analysis

For selected Al structures, the density was sampled along the periodic minimum-image line connecting the two atoms. The analysis recorded midpoint density, minimum line density between the atoms, and local curvature around the line minimum. These one-dimensional quantities were compared with structure-level Bader errors and catastrophic-failure labels.

Because a one-dimensional bond line does not uniquely determine a three-dimensional zero-flux partition, these quantities were interpreted as diagnostic proxies rather than direct measures of Bader topology. Representative improvement and degradation cases were selected from the largest paired changes in structure-level Bader error. For these cases, density slices, density-difference fields, bond-line profiles, atom-resolved Bader charges, and Bader maxima counts were inspected qualitatively.

Runtime and memory benchmarking

Inference time and memory use were measured prospectively under matched hardware and software conditions for UNet-Geometry, UNet-StructureChem, and UNet-Privileged. The benchmark used all 2,924 elemental test structures, an effective batch size of one structure, single-precision FP32 inference, one warm-up pass, and five timed repetitions on an NVIDIA RTX PRO 6000 Blackwell Max-Q Workstation Edition GPU. CUDA synchronization was performed before and after each timed GPU region. Historical SIESTA self-consistent-field wall times were parsed separately from completed output files for 64 structures and were not treated as a prospective hardware-matched comparison.

The benchmark recorded data-loading time, structural-prior construction time, input-channel construction time, host-to-GPU transfer time, neural-network forward time, density-reconstruction time, total wall time including data loading, total inference time excluding data loading, peak allocated CUDA memory, and peak resident-set size.

Because the benchmark used fixed-grid two-atom structures, it was used only to characterize inference overhead for the present elemental test setting. It was not used to infer scaling with respect to atom count, grid size, chemical complexity, or hardware architecture.

Reproducibility and implementation audit

For every trained and evaluated model, we retained the metadata needed to verify the reported comparisons and reproduce the corresponding runs. These records included the model name, its role in the manuscript, the ordered input-channel names, the input-channel count, the number of trainable parameters, the checkpoint path, the

checkpoint-selection epoch, the random seed, the training configuration, the train/validation/test split hashes, and the source-code hash. Training and evaluation channel orders were compared automatically to ensure that each checkpoint was evaluated with the same channel convention used during training. The first convolutional layer was also inspected to verify the expected number of input channels. When a model did not use a given input type, the corresponding channels were removed from the input tensor rather than replaced with constant fields, ensuring that the reported parameter counts and comparisons reflected the intended model definitions. Project-specific software environments, training scripts, checkpoints, logs, evaluation outputs, audit reports, and environment reconstruction files were retained to support independent reproduction of the reported results.

Data availability

The single-element electronic-structure dataset used in this study is publicly available on Zenodo at https://zenodo.org/records/20378934. The release contains the periodic two-atom Al, Fe, and Ni structures, the corresponding valence electron-density grids, cell and grid metadata, pseudopotential-valence information, configuration identifiers, perturbation metadata, and the frozen train/validation/test split files used in this work. Where applicable, the repository also includes SIESTA input files, parsed calculation metadata, and scripts or documentation needed to interpret the processed density grids. The split files are provided to enable exact reproduction of the model comparisons reported in this manuscript. The dataset is described in detail in the companion data manuscript [52].

Code availability

The code used for dataset preprocessing, model training, evaluation, and Bader validation is available at https://anonymous.4open.science/r/electron-density-prediction-in-elemental-metals-E236/.

Funding

This work was supported by the Generacion de Conocimiento project PID2024-157080OA-C32 and the Convocatoria de Unidad Cloud UPM de Apoyo a la Investigación.

Author contributions

I.A. conceived the study, generated and curated the dataset, designed and implemented the models, performed the training and evaluation, analyzed the results, prepared the figures, and wrote the manuscript.

Competing interests

The author declares no competing interests.

References

[1] Hohenberg, P., Kohn, W.: Inhomogeneous electron gas. Phys. Rev. **136**(3B), 864–871 (1964) https://doi.org/10.1103/PhysRev.136.B864

[2] Kohn, W., Sham, L.J.: Self-consistent equations including exchange and correlation effects. Phys. Rev. **140**, 1133–1138 (1965) https://doi.org/10.1103/PhysRev.140.A1133

[3] Martin, R.M.: Electronic Structure: Basic Theory and Practical Methods. Cambridge Univ. Press, Cambridge (2004)

[4] Hafner, J., Wolverton, C., Ceder, G.: Toward computational materials design: The impact of density functional theory on materials research. MRS Bulletin **31**(9), 659–668 (2006) https://doi.org/10.1557/mrs2006.174

[5] Jain, A., Ong, S.P., Hautier, G., Chen, W., Richards, W.D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., Persson, K.A.: Commentary: The materials project: A materials genome approach to accelerating materials innovation. APL Materials **1**(1), 011002 (2013) https://doi.org/10.1063/1.4812323

[6] Goedecker, S.: Linear scaling electronic structure methods. Reviews of Modern Physics **71**(4), 1085–1123 (1999) https://doi.org/10.1103/RevModPhys.71.1085

[7] Motamarri, P., Gavini, V.: Subquadratic-scaling subspace projection method for large-scale kohn–sham density functional theory calculations using spectral finite-element discretization. Physical Review B **90**, 115127 (2014) https://doi.org/10.1103/PhysRevB.90.115127

[8] Banerjee, A.S., Lin, L., Hu, W., Yang, C., Pask, J.E.: Chebyshev polynomial filtered subspace iteration in the discontinuous galerkin method for large-scale electronic structure calculations. The Journal of Chemical Physics **145**(15), 154101 (2016) https://doi.org/10.1063/1.4964861

[9] Banerjee, A.S., Lin, L., Suryanarayana, P., Yang, C., Pask, J.E.: Two-level chebyshev filter based complementary subspace method: pushing the envelope of large-scale electronic structure calculations. Journal of Chemical Theory and Computation **14**, 2930–2946 (2018) https://doi.org/10.1021/acs.jctc.7b01208

[10] Gavini, V., Baroni, S., Blum, V., Bowler, D.R., Buccheri, A., Chelikowsky, J.R., Das, S., Dawson, W., Delugas, P., Dogan, M., Draxl, C., Galli, G., Genovese, L., Giannozzi, P., Giantomassi, M., Gonze, X., Govoni, M., Gygi, F., Gulans, A., Herbert, J.M., Kokott, S., Kühne, T.D., Liou, K.-H., Miyazaki, T., Motamarri, P., Nakata, A., Pask, J.E., Plessl, C., Ratcliff, L.E., Richard, R.M., Rossi, M., Schade, R., Scheffler, M., Schütt, O., Suryanarayana, P., Torrent, M., Truflandier, L., Windus, T.L., Xu, Q., Yu, V.W.-Z., Perez, D.: Roadmap on electronic structure codes in the exascale era. Modelling and Simulation in Materials Science and

Engineering **31**(6), 063301 (2023) https://doi.org/10.1088/1361-651X/acdf06

[11] Dogan, M., Liou, K.-H., Chelikowsky, J.R.: Real-space solution to the electronic structure problem for nearly a million electrons. The Journal of Chemical Physics **158**(24), 244114 (2023) https://doi.org/10.1063/5.0150864

[12] Brockherde, F., Vogt, L., Li, L., Tuckerman, M.E., Burke, K., Muller, K.-R.: Bypassing the kohn-sham equations with machine learning. Nature Communications **8**, 872 (2017) https://doi.org/10.1038/s41467-017-00839-3

[13] Grisafi, A., Fabrizio, A., Meyer, B., Wilkins, D.M., Corminboeuf, C., Ceriotti, M.: Transferable machine-learning model of the electron density. ACS Central Science **5**(1), 57–64 (2019) https://doi.org/10.1021/acscentsci.8b00551

[14] Chandrasekaran, A., Kamal, D., Batra, R., Kim, C., Chen, L., Ramprasad, R.: Solving the electronic structure problem with machine learning. npj Computational Materials **5**, 22 (2019) https://doi.org/10.1038/s41524-019-0162-7

[15] Zepeda-Núñez, L., Chen, Y., Zhang, J., Jia, W., Zhang, L., Lin, L.: Deep density: Circumventing the kohn–sham equations via symmetry preserving neural networks. Journal of Computational Physics **443**, 110523 (2021) https://doi.org/10.1016/j.jcp.2021.110523

[16] Lewis, A.M., Grisafi, A., Ceriotti, M., Rossi, M.: Learning electron densities in the condensed phase. Journal of Chemical Theory and Computation **17**(11), 7203–7214 (2021) https://doi.org/10.1021/acs.jctc.1c00576

[17] Jørgensen, P.B., Bhowmik, A.: Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids. npj Computational Materials **8**, 183 (2022) https://doi.org/10.1038/s41524-022-00863-y

[18] Fiedler, L., Shah, K., Bussmann, M., Cangi, A.: Predicting electronic structures at any length scale with machine learning. npj Computational Materials **9**, 115 (2023) https://doi.org/10.1038/s41524-023-01070-z

[19] Del Río, B.G., Phan, B., Ramprasad, R.: A deep learning framework to emulate density functional theory. npj Computational Materials **9**, 158 (2023) https://doi.org/10.1038/s41524-023-01115-3

[20] Koker, T., Quigley, K., Taw, E., Tibbetts, K., Li, L.: Higher-order equivariant neural networks for charge density prediction in materials. npj Computational Materials **10**, 161 (2024) https://doi.org/10.1038/s41524-024-01343-1

[21] Li, C., Sharir, O., Yuan, S., Chan, G.K.-L.: Image super-resolution inspired electron density prediction. Nature Communications **16**, 4811 (2025) https://doi.org/10.1038/s41467-025-60095-8

[22] Pathrudkar, S., Taylor, S., Keripale, A., Gangan, A.S., Thiagarajan, P., Agarwal, S., Marian, J., Ghosh, S., Banerjee, A.S.: Electronic structure prediction of medium and high entropy alloys across composition space. npj Computational Materials **11**, 378 (2025) https://doi.org/10.1038/s41524-025-01856-3

[23] Shao, X., Paetow, L., Tuckerman, M.E., Pavanello, M.: Machine learning electronic structure methods based on the one-electron reduced density matrix. Nature Communications **14**, 6281 (2023) https://doi.org/10.1038/s41467-023-41953-9

[24] Tang, Z., Zou, N., Li, H., Wang, Y., Yuan, Z., Tao, H., Li, Y., Chen, Z., Zhao, B., Sun, M., Jiang, H., Duan, W., Xu, Y.: Improving density matrix electronic structure method by deep learning (2024). https://arxiv.org/abs/2406.17561

[25] Hazra, S., Patil, U., Sanvito, S.: Predicting the one-particle density matrix with machine learning. Journal of Chemical Theory and Computation **20**(11), 4569–4578 (2024) https://doi.org/10.1021/acs.jctc.4c00042

[26] Sager-Smith, L.M., Mazziotti, D.A.: Reducing the quantum many-electron problem to two electrons with machine learning. Journal of the American Chemical Society **144**(41), 18959–18966 (2022) https://doi.org/10.1021/jacs.2c07112

[27] Pathrudkar, S., Yu, H.M., Ghosh, S., Banerjee, A.S.: Machine learning based prediction of the electronic structure of quasi-one-dimensional materials under strain. Physical Review B **105**, 195141 (2022) https://doi.org/10.1103/PhysRevB.105.195141

[28] Arora, G., Manzoor, A., Aidhy, D.S.: Charge-density based evaluation and prediction of stacking fault energies in ni alloys from dft and machine learning. Journal of Applied Physics **132**, 225104 (2022) https://doi.org/10.1063/5.0122675

[29] Fabrizio, A., Grisafi, A., Meyer, B., Ceriotti, M., Corminboeuf, C.: Electron density learning of non-covalent systems. Chemical Science **10**, 9424–9432 (2019) https://doi.org/10.1039/C9SC02696G

[30] Qiao, Z., Welborn, M., Anandkumar, A., Manby, F.R., Miller, T.F.: Informing geometric deep learning with electronic interactions to accelerate quantum chemistry. Proceedings of the National Academy of Sciences **119**(31), 2205221119 (2022) https://doi.org/10.1073/pnas.2205221119

[31] Fu, X., Rosen, A.S., Bystrom, K., Wang, R., Musaelian, A., Kozinsky, B., Smidt, T., Jaakkola, T.: A recipe for charge density prediction. In: Advances in Neural Information Processing Systems, vol. 37, pp. 9727–9752 (2024). https://doi.org/10.52202/079017-0310

[32] Gong, S., Xie, T., Zhu, T., Wang, S., Fadel, E.R., Li, Y., Grossman, J.C.:

Predicting charge density distribution of materials using a local-environment-based graph convolutional network. Physical Review B **100**, 184103 (2019) https://doi.org/10.1103/PhysRevB.100.184103

[33] Pathrudkar, S., Thiagarajan, P., Agarwal, S., Banerjee, A.S., Ghosh, S.: Electronic structure prediction of multi-million atom systems through uncertainty quantification enabled transfer learning. npj Computational Materials **10**, 175 (2024) https://doi.org/10.1038/s41524-024-01305-7

[34] Pope, P., Jacobs, D.: Towards combinatorial generalization for catalysts: A kohn–sham charge-density approach. In: Advances in Neural Information Processing Systems, vol. 36, pp. 60585–60598 (2024)

[35] Fung, V., Zhang, J., Juarez, E., Sumpter, B.G.: Benchmarking graph neural networks for materials chemistry. npj Computational Materials **7**, 84 (2021) https://doi.org/10.1038/s41524-021-00554-0

[36] Okabe, R., al.: Virtual node graph neural network for full phonon prediction. Nature Computational Science **4**, 481–482 (2024) https://doi.org/10.1038/s43588-024-00661-0

[37] Lee, R.-G., Kim, Y.-H.: Convolutional network learning of self-consistent electron density via grid-projected atomic fingerprints. npj Computational Materials **10**, 248 (2024) https://doi.org/10.1038/s41524-024-01433-0

[38] Kim, S., Lee, B.D., Cho, M.Y., Pyo, M., Lee, Y.-K., Park, W.B., Sohn, K.-S.: Deep learning for symmetry classification using sparse 3d electron density data for inorganic compounds. npj Computational Materials **10**, 211 (2024) https://doi.org/10.1038/s41524-024-01402-7

[39] Behler, J., Parrinello, M.: Generalized neural-network representation of high-dimensional potential-energy surfaces. Physical Review Letters **98**, 146401 (2007) https://doi.org/10.1103/PhysRevLett.98.146401

[40] Bartók, A.P., Kondor, R., Csányi, G.: On representing chemical environments. Physical Review B **87**, 184115 (2013) https://doi.org/10.1103/PhysRevB.87.184115

[41] Thompson, A.P., Swiler, L.P., Trott, C.R., Foiles, S.M., Tucker, G.J.: Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. Journal of Computational Physics **285**, 316–330 (2015) https://doi.org/10.1016/j.jcp.2014.12.018

[42] Gastegger, M., Schwiedrzik, L., Bittermann, M., Berzsenyi, F., Marquetand, P.: wacsf—weighted atom-centered symmetry functions as descriptors in machine learning potentials. Journal of Chemical Physics **148**, 241709 (2018) https://doi.org/10.1063/1.5019667

[43] Imbalzano, G., Anelli, A., Giofré, D., Klees, S., Behler, J., Ceriotti, M.: Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials. Journal of Chemical Physics **148**, 241730 (2018) https://doi.org/10.1063/1.5024611

[44] Lei, X., Medford, A.J.: A universal framework for featurization of atomistic systems. The Journal of Physical Chemistry Letters **13**(34), 7911–7919 (2022) https://doi.org/10.1021/acs.jpclett.2c02100

[45] Timmerman, L.R., Kumar, S., Suryanarayana, P., Medford, A.J.: Overcoming the chemical complexity bottleneck in on-the-fly machine learned molecular dynamics simulations. Journal of Chemical Theory and Computation **20**(14), 5788–5795 (2024) https://doi.org/10.1021/acs.jctc.4c00474

[46] Bader, R.F.W.: Atoms in Molecules: A Quantum Theory. Oxford University Press, Oxford (1990). https://doi.org/10.1093/oso/9780198551683.001.0001

[47] Henkelman, G., Arnaldsson, A., Jónsson, H.: A fast and robust algorithm for bader decomposition of charge density. Computational Materials Science **36**(3), 354–360 (2006) https://doi.org/10.1016/j.commatsci.2005.04.010

[48] Sanville, E., Kenny, S.D., Smith, R., Henkelman, G.: Improved grid-based algorithm for bader charge allocation. Journal of Computational Chemistry **28**(5), 899–908 (2007) https://doi.org/10.1002/jcc.20575

[49] Tang, W., Sanville, E., Henkelman, G.: A grid-based bader analysis algorithm without lattice bias. Journal of Physics: Condensed Matter **21**(8), 084204 (2009) https://doi.org/10.1088/0953-8984/21/8/084204

[50] Yu, M., Trinkle, D.R.: Accurate and efficient algorithm for bader charge integration. Journal of Chemical Physics **134**, 064111 (2011) https://doi.org/10.1063/1.3553716

[51] Deshmukh, G., *et al.*: Active learning of ternary alloy structures and energies. npj Computational Materials **10**, 116 (2024) https://doi.org/10.1038/s41524-024-01256-z

[52] Arévalo, I., Olleros, P.: A large-scale dataset of 30,000 converged density functional theory calculations for single-element electronic structure learning. Manuscript under revision (2026)

[53] Soler, J.M., Artacho, E., Gale, J.D., García, A., Junquera, J., Ordejón, P., Sánchez-Portal, D.: The SIESTA method for ab initio order-N materials simulation. Journal of Physics: Condensed Matter **14**, 2745–2779 (2002) https://doi.org/10.1088/0953-8984/14/11/302

[54] García, A., *et al.*: SIESTA: Recent developments and applications. Journal of

Chemical Physics **152**, 204108 (2020) https://doi.org/10.1063/5.0005077

[55] Lejaeghere, K., Bihlmayer, G., Björkman, T., et al.: Reproducibility in density functional theory calculations of solids. Science **351** (2016) https://doi.org/10.1126/science.aad3000