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.

Auditing Representation Identifiability in Neural Electron-Density Learning

Irina Arevalo

irina.arevalo@upm.es

Universidad Politecnica de Madrid https://orcid.org/0000-0002-8967-0208

Article

Keywords: electron-density prediction, density functional theory, representation identifiability, species-resolved representations, chemical identity, multicomponent alloys, volumetric neural networks, substitutional alloys, U-Net

Posted Date: September 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10680349/v1

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

Additional Declarations: There is **NO** Competing Interest.

Auditing Representation Identifiability in Neural Electron-Density Learning

Irina Arévalo[1*]

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

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

Abstract

Field-valued machine-learning models can fail despite adequate capacity when inputs collapse structures with distinct targets. We show that grid representations combining species-shared structural fields with global composition are non-identifiable for substitutional-alloy electron-density learning: configurations with identical cells, coordinates and composition but different site occupations receive identical inputs despite distinct DFT densities. In grouped Al–Fe–Ni tests, species-resolved fields reduce log-density error by **99.5%**. A scale-matched control attributes the dominant gain to site-local identity but does not resolve the smaller occupancy–valence difference. Controlled site reassignment raises the recovery ratio (the predicted-to-DFT ratio of mean-absolute density differences) from $\mathbf{8.0 \times 10^{-5}}$ to **0.992**, while channel permutations sharply degrade predictions. An independent five-component $\mathbf{Al_2Fe_2Ni_2Co_2Cr_2}$ replication yields recovery **0.967** and density-difference correlation **0.994**. These results establish site-local chemical identity as necessary for identifiable grid-based density learning in substitutional alloys and provide a practical audit for information loss in volumetric representations.

Keywords: electron-density prediction, density functional theory, representation identifiability, species-resolved representations, chemical identity, multicomponent alloys, volumetric neural networks, substitutional alloys, U-Net

Electron density is the central variable in density functional theory (DFT). Through the Hohenberg–Kohn framework, it provides a common basis for evaluating energies and interpreting bonding, charge redistribution and electrostatic response [1–3]; its spatial locality also underlies nearsightedness and linear-scaling electronic structure

[4]. Accurate ground-state densities nevertheless normally require self-consistent calculations. High-throughput infrastructures such as AFLOW, the Materials Project, OQMD and NOMAD have expanded access to first-principles data [5–8], but generating new densities remains expensive, particularly for multicomponent alloys whose composition, ordering, defects and site assignments create vast configuration spaces [9].

Machine learning has accelerated atomistic simulation through species-dependent symmetry functions [10], message-passing, continuous-filter and crystal-graph models [11–13], and three-dimensional equivariant architectures [14–16]. These methods commonly predict energies, forces or other low-dimensional observables. Electron-density learning instead targets a reusable electronic field from which multiple observables and chemically interpretable analyses can be derived.

Existing approaches learn the external-potential–density map or refine approximate densities towards self-consistent references [17–20]; predict densities at real-space query points from local descriptors or message-passing environments [21–24]; or retain a direct connection to the computational mesh through grid-centred polynomial representations [25]. Atom-centred basis expansions provide an alternative route, with symmetry-adapted regression, message-passing coefficients and sparse density fitting demonstrated across molecules and condensed phases [26–29]. Learned orbital bases and high-capacity equivariant models improve the accuracy–cost trade-off [30], while Hamiltonian and wavefunction surrogates permit density reconstruction [31–33]. Higher-order equivariant density models address angularly complex environments and can initialize self-consistent calculations [34]. Local electronic-structure models now reach systems exceeding 10^5 atoms [35], and learned densities have been coupled to energy functionals and quantum–classical simulations [36, 37].

These representation families do not share the same information bottlenecks. Atom-centred basis models, message-passing networks and species-aware query-point architectures normally attach chemical labels to local environments and therefore avoid the ambiguity studied here. Grid-based models also avoid it when chemistry is encoded locally. Our focus is a narrower but plausible volumetric shortcut: atomic geometry is deposited into species-shared fields, while chemistry is supplied only through global composition. The contribution is not the observation that species matter, but a formal and empirical demonstration that this representation family can discard the species–site correspondence required to determine a field-valued target.

More generally, model success depends on both architecture and the information preserved by its input. Atomistic representations encode local geometry, species and symmetries through neighbourhood densities, multicomponent cluster expansions and equivariant graphs [16, 38, 39]. Distinct environments can nevertheless collapse to identical descriptors, motivating analyses of completeness and learnability [40, 41]. We call the corresponding target-dependent requirement *representation identifiability*: structures with different target fields must receive distinguishable inputs. Non-injective graph and set aggregation imposes analogous expressivity limits [42, 43]. The connection to Weisfeiler–Lehman refinement is conceptual: once site labels are removed, later processing cannot reconstruct which label occupied which site, although our setting involves continuous periodic fields rather than discrete graphs.

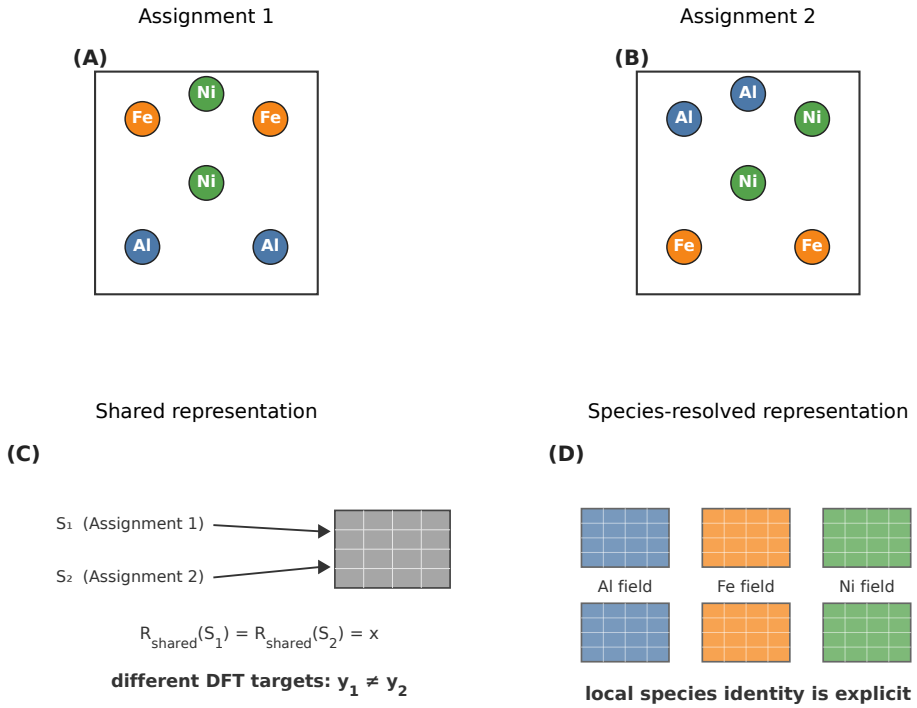


Fig. 1 Representation degeneracy in locally species-blind alloy inputs. Two substitutional alloy configurations can have the same simulation cell, atomic coordinates and global composition while differing only in the assignment of chemical species to sites. A shared StructureChem representation constructed from total structural fields and global composition maps both configurations to the same input, although their DFT electron densities generally differ. Species-resolved Al, Fe and Ni fields change under site reassignment and encode the local chemical identity required to distinguish the density fields. In panel **c**, the two assignments independently produce identical shared input tensors, $R_{\mathrm{shared}}(S_1) = R_{\mathrm{shared}}(S_2) = x$, despite having different DFT targets, $y_1 \neq y_2$.

For elemental systems, geometry-based deposition may suffice because every site has the same identity; localized Gaussian fields can then encode positions for density learning [17, 20, 24, 44]. In a substitutional alloy, however, the same cell, coordinates and global composition can support different assignments of species to sites. Shared structural fields and global composition therefore map these configurations to identical inputs even though their DFT densities generally differ. No deterministic downstream model can recover both targets, regardless of capacity. Figure 1 illustrates this degeneracy.

This observation motivates our central question: what information must a deployable volumetric representation preserve for alloy electron-density prediction? Here, deployable inputs are constructible from species, coordinates, cell and pseudopotential information for an arbitrary structure; generator-derived perturbation labels and amplitudes are excluded. We refer to the deployable grid-based framework used here

as StructureChem: a volumetric U-Net that receives atom-deposition fields together with a constant global-composition channel evaluated on the DFT grid. Within this common framework, we compare a composition-aware but locally species-blind representation containing two shared structural fields with a species-resolved representation containing separate fields for each element [44, 45].

On grouped Al–Fe–Ni tests, species resolution reduces log-density error by approximately 99.5%. A same-geometry counterfactual increases median recovery of the DFT density response from 8.0×10^{-5} to 0.992, while channel permutations sharply degrade predictions. An independently trained five-component $Al_2Fe_2Ni_2Co_2Cr_2$ replication yields median recovery 0.967 and density-difference correlation 0.994. Together, the grouped benchmark, scale control, counterfactual, permutation test and higher-component replication form a practical audit for detecting when a volumetric representation discards target-determining information. The results establish preservation of site-local chemical identity as necessary for identifiable grid-based electron-density learning in substitutional alloys.

Results

Locally species-blind alloy representations are degenerate under site reassignment

We first formalize the representational limitation that arises when a multicomponent alloy is described using shared structural fields and global composition. Let S_1 and S_2 be two substitutional alloy configurations with the same simulation cell, atomic coordinates, and global composition, but different assignments of chemical species to atomic sites. For a locally species-blind representation $\mathcal{R}_{\text{shared}}$,

$$\mathcal{R}_{\text{shared}}(S_1) = \mathcal{R}_{\text{shared}}(S_2) = x,$$

because the representation depends only on the unlabeled geometry and overall composition. Their DFT log-density targets, however, generally differ:

$$y_1 \neq y_2.$$

The structure-to-representation map is therefore many-to-one, and the induced relation between the shared input x and the target density is not single-valued.

This collision imposes a quantitative, architecture-independent error floor. Let M denote the number of grid voxels and define the voxel-averaged log-density loss as

$$\mathcal{L}(\hat{y}, y) = \frac{1}{M}\|\hat{y} - y\|_2^2.$$

For any deterministic predictor f, with

$$\bar{y} = \frac{y_1 + y_2}{2},$$

the squared-error decomposition gives

$$\frac{1}{2}\left[\mathcal{L}(f(x), y_1) + \mathcal{L}(f(x), y_2)\right] = \frac{1}{M}\left\|f(x) - \bar{y}\right\|_2^2 + \frac{1}{4M}\left\|y_1 - y_2\right\|_2^2 .$$

Consequently,

$$\frac{1}{2}\left[\mathcal{L}(f(x), y_1) + \mathcal{L}(f(x), y_2)\right] \geq \frac{1}{4M}\left\|y_1 - y_2\right\|_2^2 ,$$

with equality if and only if

$$f(x) = \bar{y} = \frac{y_1 + y_2}{2}.$$

Thus, even the optimal deterministic predictor must average the two targets and incur a non-zero error whenever distinct density fields share the same representation. This representation-induced error floor cannot be eliminated by increasing model capacity, changing the neural architecture, or improving optimization.

Species-resolved volumetric fields remove this specific collision by encoding the positions of Al, Fe, and Ni separately. Chemical site reassignment then changes the input representation, allowing the two density targets to be associated with distinguishable inputs. This removes the lower bound caused by the shared-representation degeneracy, although it does not by itself guarantee accurate prediction. The same-geometry site-assignment experiment below directly tests this argument by holding the simulation cell, atomic coordinates, and global composition fixed while changing only the assignment of chemical species to sites.

A deployable shared StructureChem representation fails on grouped alloy prediction

We evaluated a deployable shared StructureChem baseline on the grouped Al–Fe–Ni alloy split. This model receives two shared Gaussian structural fields and a global composition channel. Thus, it is not chemistry-free, but it is locally species-blind: it does not encode which atomic site is occupied by each species. The grouped split contains 3204 training structures, 402 validation structures, and 396 test structures. The resulting benchmark is summarized in Fig. 2.

Across three random seeds, the shared StructureChem baseline showed highly stable but poor test performance. The mean test logMSE was 1.1705×10^{-1}, with only negligible variation across seeds. The corresponding mean density MAE, density RMSE, relative electron-count error, and normalized-shape logMSE were approximately 0.5937, 2.0694, 0.2465, 0.1325, respectively. The stability of these values across seeds indicates that the poor performance is not primarily due to stochastic optimization failure. Instead, it is consistent with the expected representational limitation of a locally species-blind alloy input.

Species-resolved StructureChem removes the dominant alloy representation failure

Replacing the shared structural fields with species-resolved fields led to a large improvement in all test metrics. With occupancy-weighted species fields, the test

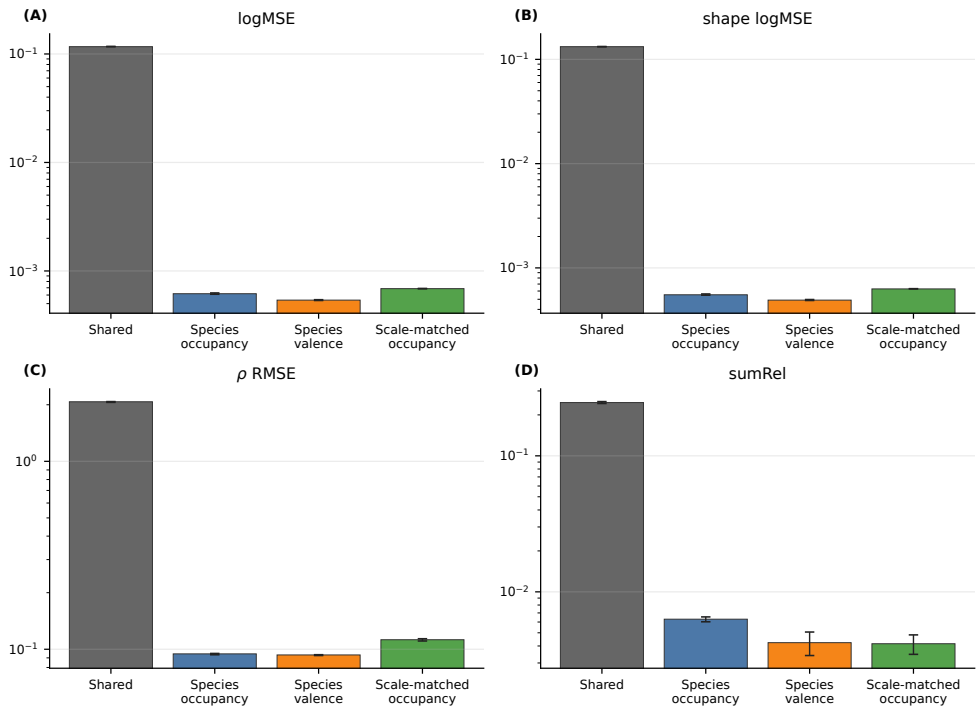


Fig. 2 Deployable alloy benchmark on the grouped Al–Fe–Ni test split. The locally species-blind shared StructureChem baseline fails despite using global composition information. Replacing shared structural fields with species-resolved Al, Fe, and Ni fields reduces the log-density error by approximately two orders of magnitude. A scale-matched occupancy control remains within the broad species-resolved performance regime, confirming that the decisive gain comes from site-local species identity. Its non-monotonic effect across metrics shows that matching channel means and variances does not fully explain the smaller occupancy–valence difference. Bars show means across three seeds and error bars show one standard deviation.

logMSE decreased from 1.1705×10^{-1} to 6.19×10^{-4}. With valence-weighted species fields, it decreased further to 5.39×10^{-4}. These correspond to relative error reductions of approximately 99.47% and 99.54%, respectively, compared with the shared StructureChem baseline.

The shared baseline contained 789,757 trainable parameters, whereas the ternary species-resolved models contained 791,053, an increase of only 0.16%. This small capacity difference cannot plausibly account for the approximately 190-fold reduction in logMSE, supporting the interpretation that the gain arises from information preserved by the input representation rather than from increased model size.

The same pattern was observed for real-space density errors and normalized-shape metrics. The density RMSE decreased from approximately 2.07 for the shared baseline to 9.42×10^{-2} for the occupancy-weighted species-resolved model and 9.31×10^{-2} for the valence-weighted species-resolved model. The normalized-shape logMSE decreased from 1.32×10^{-1} to 5.55×10^{-4} and 4.92×10^{-4}, respectively. These reductions show

that the improvement is not merely a correction of global electron-count scale; the spatial density shape is also substantially improved.

The magnitude of the improvement indicates that the dominant limitation of the shared baseline is representational. Both the shared and species-resolved models use deployable information, but only the species-resolved representation makes chemical identity local. Thus, the main difference is whether the model can distinguish which atomic sites are occupied by Al, Fe, and Ni.

Scale controls confirm that local species identity dominates the improvement

We next compared two variants of the species-resolved StructureChem representation. In the occupancy-weighted model, each species channel is constructed from unit-weighted localized Gaussian atom-deposition fields. In the valence-weighted model, the same localized deposition fields are instead weighted by the corresponding pseudopotential valence.

Both variants substantially outperformed the shared baseline, confirming that the decisive representational step is the introduction of site-local species identity. In the unnormalized comparison, valence weighting produced slightly lower errors than standard occupancy weighting. Relative to the occupancy-weighted species-resolved model, the valence-weighted model reduced the test logMSE from 6.19×10^{-4} to 5.39×10^{-4}, corresponding to an additional improvement of approximately 12.9%. The normalized-shape logMSE decreased from 5.55×10^{-4} to 4.92×10^{-4}, and the relative electron-count error decreased from approximately 6.28×10^{-3} to 4.23×10^{-3}.

To determine whether the apparent gain from valence weighting reflected physical valence information or simply altered input scale, we introduced a scale-matched occupancy control. In this control, occupancy-weighted species channels were affinely rescaled after the $\log(1+\rho)$ transformation to match the per-channel mean and standard deviation of the corresponding valence-weighted fields, using statistics computed on the training split only. This preserves the spatial occupancy pattern while matching the first-layer input scale of the valence representation.

The scale-matched control produced a non-monotonic response across metrics. Relative to ordinary occupancy weighting, matching the post- $\log(1+\rho)$ channel means and standard deviations reduced sumRel from 6.28×10^{-3} to 4.15×10^{-3}, bringing integrated-charge accuracy close to the valence-weighted result of 4.23×10^{-3}. However, it increased logMSE from 6.19×10^{-4} to 6.88×10^{-4}, ρ MAE from 3.26×10^{-2} to 3.60×10^{-2}, ρ RMSE from 9.42×10^{-2} to 1.12×10^{-1}, and shape logMSE from 5.55×10^{-4} to 6.31×10^{-4}. Thus, matching the first two moments improves global electron-count calibration but does not move the spatial-density metrics monotonically towards the valence-weighted result. The control therefore rules out simple first- and second-moment channel scaling as an explanation for the dominant shared-to-resolved improvement, but it does not resolve the smaller occupancy–valence difference. Importantly, all species-resolved variants remain approximately two orders of magnitude more accurate than the locally species-blind shared representation, confirming that site-local identity is responsible for the dominant improvement.

Table 1 Same-geometry site-assignment counterfactual for Al–Fe–Ni. Each pair has an identical simulation cell, atomic coordinates, and global composition but a different assignment of species to sites. Confidence intervals were calculated using 100,000 nonparametric bootstrap replicates over the 13 independent geometry clusters. Intervals condition on the fixed validation-selected checkpoints. Bootstrap intervals were computed for the two headline median statistics, recovery ratio and density-difference correlation; median relative L_2 is included as a descriptive error summary without an uncertainty interval.

Model	n	Median recovery [95% CI]	Median correlation [95% CI]	Median relative L_2
Shared StructureChem	13	8.03×10^{-5} $[4.29 \times 10^{-5}, 1.17 \times 10^{-4}]$	-0.099 $[-0.202, 0.085]$	1.000
Species-resolved StructureChem	13	0.992 [0.983, 1.002]	0.9994 [0.9993, 0.9995]	0.039

Same-geometry site reassignment isolates the role of local chemical identity

To test whether the improvement arises specifically from site-local chemical identity, we constructed strict same-geometry pairs. Each pair contains two alloy configurations with identical simulation cells, identical atomic coordinates, and identical global compositions, but different assignments of Al, Fe, and Ni to atomic sites. This setup holds geometry and composition fixed while intervening only on the local chemical assignment. A representative density-difference example is shown in Fig. 3, and aggregate statistics over all strict pairs are reported in Table 1.

Here, the recovery ratio is the ratio between the mean-absolute predicted and DFT density-difference magnitudes; it measures response amplitude and should not be interpreted as a correlation coefficient or R^2.

The shared StructureChem model was effectively invariant under this intervention. Across 13 strict pairs, its median recovery ratio was 8.0×10^{-5}, with a mean recovery ratio of 1.46×10^{-4}. Its median density-difference correlation was -0.099 (95% CI, -0.202–0.085), and its mean relative L_2 error was approximately 1.0. Because the correlation interval includes zero, no directional correlation should be inferred. The shared model therefore recovered essentially none of the DFT density response to chemical site reassignment.

In contrast, the species-resolved StructureChem model recovered nearly the full DFT density-difference field. Its median recovery ratio was 0.992 (95% CI, 0.983–1.002), with a mean recovery ratio of 0.990. Its median density-difference correlation was 0.9994 (95% CI, 0.9993–0.9995), and its mean relative L_2 error was 0.046. The species-resolved model therefore reproduces both the magnitude and spatial pattern of the DFT response with high fidelity.

This experiment provides direct mechanistic evidence for the degeneracy argument. The shared representation cannot respond to a change that it does not encode. The species-resolved representation, by contrast, changes when the local chemical assignment changes, allowing the network to reproduce the corresponding DFT density response.

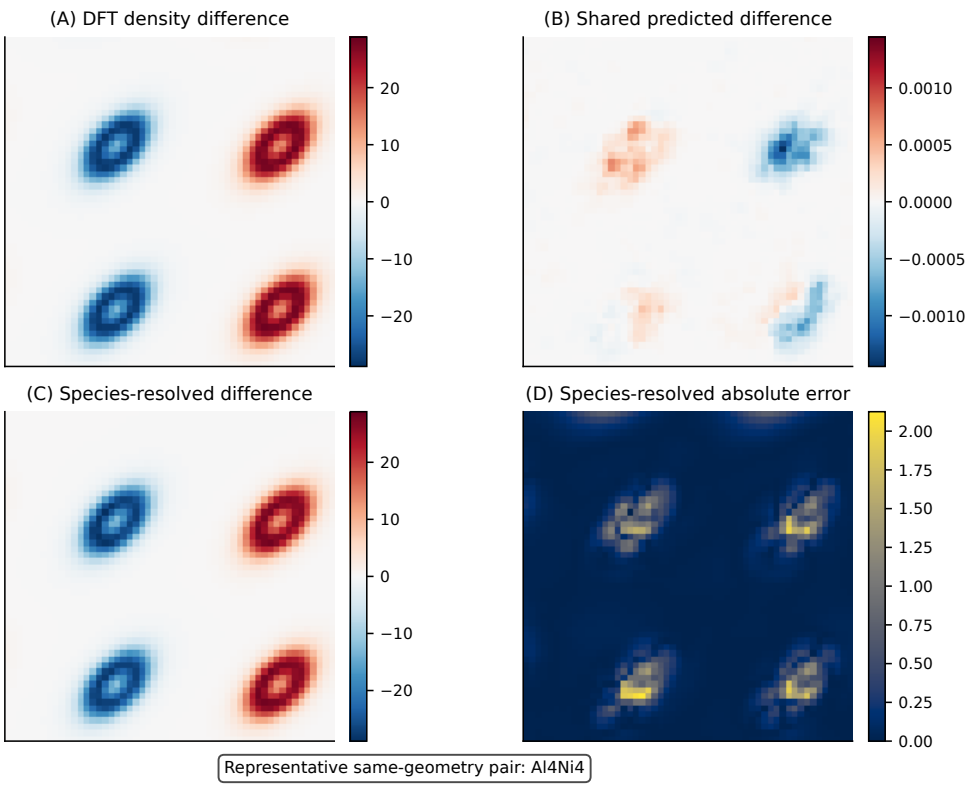


Fig. 3 Same-geometry site-assignment counterfactual. Each pair has identical cell, atomic coordinates, and global composition, but different assignments of Al, Fe, and Ni to atomic sites. The DFT density difference shows the electronic response to site reassignment. The locally species-blind shared StructureChem model predicts an almost zero response, whereas the species-resolved model reproduces both the magnitude and spatial pattern of the DFT response. Across 13 strict pairs, the shared model has median recovery ratio 8.0×10^{-5}, while the species-resolved model has median recovery ratio 0.992 and median density-difference correlation 0.999. For the representative pair shown, the shared model has recovery ratio 2.33×10^{-5} and density-difference correlation -0.585, whereas the species-resolved model has recovery ratio 0.989 and correlation 1.000. Panels **a** and **c** use a common symmetric color scale. Cluster-bootstrap confidence intervals for the aggregate statistics are reported in Table 1. Panel **b** uses a separate symmetric scale because the shared-model response is orders of magnitude smaller, and panel **d** uses a sequential color scale.

Species-channel permutations confirm that the model uses site-local identity

Finally, we tested whether the trained species-resolved model actively uses the correct association between species channels and atomic sites. Starting from the deployable valence-weighted species-resolved model, we deliberately permuted the Al, Fe, and Ni species-resolved input channels while leaving the target density and all non-species

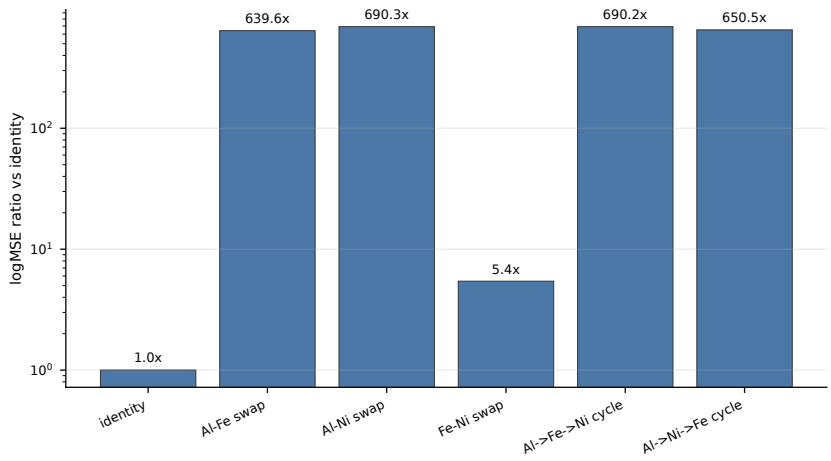


Fig. 4 Species-channel permutation audit. After training the deployable species-resolved StructureChem model, the Al, Fe, and Ni input channels were deliberately permuted while leaving the target density unchanged. The Al-involving swaps and cyclic permutations increased the error by approximately 640–690×, whereas the Fe–Ni swap produced a smaller but still clear degradation of 5.4×, consistent with the closer chemical similarity of Fe and Ni in this representation. This confirms that the model does not merely benefit from additional input channels, but actively uses the correct association between local chemical identity and spatial position.

information unchanged. Major swaps and cyclic permutations produced mean per-structure logMSE ratios of approximately 640–690× relative to the identity input. The full permutation audit is shown in Fig. 4.

This large degradation shows that the species-resolved model is not simply benefiting from additional input channels or increased capacity. Its accuracy depends on the correct correspondence between a species channel and the atomic sites occupied by that species. In other words, the learned mapping is chemically specific: breaking the Al/Fe/Ni assignment destroys the information needed to predict the density accurately.

A quinary cross-system stress test reproduces the identifiability failure

We next tested whether the identifiability contrast persists in a higher-component chemical system. The compact quinary dataset contains fixed-composition $Al_2Fe_2Ni_2Co_2Cr_2$ structures and is intended as a cross-system identifiability stress test rather than as a comprehensive benchmark for high-entropy-alloy density prediction.

On the grouped test split, the locally species-blind shared model reached a logMSE of 9.5520×10^{-2} and a shape logMSE of 8.5095×10^{-2}. The species-resolved occupancy model reduced these values to 9.43×10^{-4} and 6.86×10^{-4}, corresponding to reductions of 99.0% and 99.2%, respectively. The models contained 789,757 and 792,349 parameters, so species resolution increased the parameter count by only 0.33%.

The held-out same-geometry analysis contained 150 pairs constructed from 15 test parent geometries, with all ten within-parent comparisons retained for each parent.

Table 2 Quinary cross-system identifiability stress test. Grouped-test metrics are macro-averaged over 75 held-out structures. Same-geometry statistics use 150 within-parent pairs from 15 held-out parent geometries. Confidence intervals were calculated using 100,000 cluster-bootstrap replicates over parent geometries and condition on the single fixed, validation-selected checkpoint for each representation. A dash indicates that Pearson correlation was not reported because it was undefined for 118 of the 150 shared-model pairs owing to zero predicted-field variance.

Model	logMSE	shape logMSE	Median recovery [95% CI]	Median correlation [95% CI]
Shared StructureChem	9.5520×10^{-2}	8.5095×10^{-2}	0.000 [0.000, 0.000]	—
Species-resolved occupancy StructureChem	9.43×10^{-4}	6.86×10^{-4}	0.967 [0.958, 0.971]	0.994 [0.993, 0.995]

The shared model had a median recovery ratio of 0.000 (95% cluster-bootstrap CI, 0.000–0.000) and a mean recovery ratio of 5.35×10^{-5}. Pearson correlation was not reported for this model because it was undefined for 118 of the 150 pairs owing to zero predicted-field variance.

In contrast, the species-resolved occupancy model achieved a median recovery ratio of 0.967 (95% cluster-bootstrap CI, 0.958–0.971) and a mean recovery ratio of 0.962. Its median density-difference correlation was 0.994 (95% CI, 0.993–0.995), with a mean correlation of 0.993. Exact input-geometry validation gave maximum lattice and coordinate differences of zero.

Finally, an Al–Ni channel swap increased the macro-averaged test logMSE from 9.43×10^{-4} to 2.47×10^{-1}, a 262.2× increase. This confirms that the model depends on the correct local correspondence between chemical species and atomic sites. Because both quinary models were trained from scratch and evaluated for a single random seed, this experiment establishes cross-system replication of the representational contrast, not transfer learning, a comprehensive high-entropy-alloy benchmark, or multi-seed robustness. The grouped and same-geometry results are summarized in Table 2.

Discussion

This work shows that the dominant difficulty in deployable alloy electron-density prediction is not only architectural, but representational. In particular, a model that receives shared structural fields and global composition information can still be fundamentally unable to distinguish alloy configurations that differ only by chemical site assignment. Such configurations are identical from the perspective of a locally species-blind representation, even though their DFT electron densities are generally different. This creates an identifiability failure: the same model input can correspond to multiple physically distinct target densities.

This limitation is not shared by every electron-density-learning representation. Atom-centered basis, pointwise message-passing, and related local models generally retain chemical species labels within the local environment by construction. The failure examined here concerns grid-based inputs that retain unlabeled geometry and global

composition but discard the correspondence between chemical species and atomic sites before learning begins. Species-resolved volumetric fields provide one direct remedy within this representation family.

The experiments support this interpretation across complementary tests. Species resolution reduced Al–Fe–Ni test logMSE by approximately 99.5% without privileged perturbation metadata. Under fixed-geometry site reassignment, median recovery increased from 8.0×10^{-5} to 0.992, and deliberately permuting species channels increased error by 640–690×. Together, these results show that the shared model fails because its input discards the species–site correspondence, whereas the resolved model uses that correspondence directly.

A useful distinction emerges between site-local identifiability and input amplitude scaling. Occupancy-weighted species fields already recover almost the entire gap between the shared and species-resolved models, showing that the essential information is which species occupies each site. The scale-matched occupancy control clarifies but does not fully explain the smaller occupancy–valence difference. Matching channel means and variances improves sumRel, but worsens logMSE, ρ MAE, ρ RMSE, and shape logMSE relative to ordinary occupancy weighting. Its effect is therefore non-monotonic: matching the first two moments improves integrated-charge calibration but does not move the spatial-density metrics towards the valence-weighted result. The control consequently supports a bounded conclusion. It rules out simple first- and second-moment channel scaling as an explanation for the approximately 190-fold reduction in logMSE between the shared and occupancy-resolved representations, but it does not identify the origin of the much smaller 12.9% occupancy–valence difference. Resolving that secondary difference would require dedicated distributional or representation-level controls and lies outside the central identifiability claim.

More broadly, the combination of a grouped benchmark, a scale-matched representation control, a same-geometry counterfactual, and a channel-permutation test provides a practical template for detecting when apparent model errors are caused by missing information in the representation rather than by insufficient model capacity.

These findings have implications beyond the particular Al–Fe–Ni system considered here. Many atomistic machine-learning models include global composition features, element embeddings, or pooled chemical descriptors. Such features may be sufficient for tasks where the target is invariant, or nearly invariant, under certain site permutations. However, field-valued electronic quantities are sensitive to where each species sits in the structure. For substitutional alloys, interfaces, defects, doped materials, and chemically ordered compounds, global composition alone cannot encode the local electronic environment. Any deployable density surrogate for these systems should therefore include an explicit mechanism for local chemical identifiability.

The quinary stress test independently reproduced the contrast: median recovery increased from zero to 0.967, with correlation 0.994, while grouped logMSE fell by 99.0%. An Al–Ni channel swap increased logMSE by 262.2×. Because this was a single-seed, fixed-composition experiment, it establishes cross-system replication rather than a comprehensive high-entropy-alloy benchmark.

For alloys with more components, such as medium- and high-entropy alloys, the simplest species-resolved construction scales linearly with the number of chemical

species. With two Gaussian widths, a shared representation uses two local structural fields, whereas an S-component species-resolved representation uses $2S$ local species fields, plus any global composition channels. Thus, moving from a ternary alloy to a five- or six-component high-entropy alloy would increase the local-field channel count from six to ten or twelve. This increases the memory footprint of the stored input tensor and the first convolutional layer, but it does not change the output grid size, the target density representation, or the size of the deeper decoder feature maps. For moderate numbers of principal elements, this overhead is expected to be small compared with the cost of storing and propagating three-dimensional feature maps through the U-Net. For broader chemical spaces or models intended to span many elements simultaneously, however, explicit one-channel-per-species fields may become inefficient. In that regime, the same identifiability principle could be implemented through lower-dimensional local chemical embeddings, element-conditioned basis functions, atom-centered basis expansions, or equivariant graph-to-grid representations, provided that they remain injective with respect to site-local species assignment.

The headline conclusions use only deployable inputs constructed from species, coordinates and cells; generator-specific perturbation metadata is excluded. This separates the representational requirements for practical density prediction from the distinct question of how well a model can exploit privileged synthetic labels.

There are several limitations. First, the main benchmark focuses on Al–Fe–Ni alloys generated within a controlled dataset. The compact quinary stress test provides evidence that the representational conclusion persists in a five-component setting, but it does not constitute a comprehensive high-entropy-alloy benchmark across compositions, crystal structures, disorder regimes, or magnetic states. The quantitative gains may therefore differ for other chemistries, larger unit cells, lower-symmetry structures, surfaces, interfaces, or defects. Second, the main Al–Fe–Ni same-geometry analysis uses 13 strict pairs. Although the effect is large and independently reproduced in the quinary stress test, larger counterfactual sets would allow more detailed analysis by alloy family, local coordination environment, and chemical contrast. Third, the model predicts the electron density on fixed real-space grids. In this setting, the dominant memory cost scales with the number of voxels and the width of the intermediate three-dimensional feature maps. Species resolution increases only the number of input channels, but explicit one-channel-per-species fields may become less efficient for very large supercells, variable-resolution grids, or models spanning many elements. Fourth, the narrower Gaussian deposition field is sub-grid in standard-deviation units on the present real-space meshes. It should therefore be interpreted as a localized positional encoding rather than as a smoothly resolved reference density. Direct discretization audits nevertheless confirmed that the narrow and broad channels remain numerically distinguishable after the $\log(1 + \rho)$ transformation. Finally, while species-resolved Gaussian fields remove the degeneracy studied here, they are not the only possible representation. Alternative deployable encodings, such as compact local chemical embeddings, atom-centred basis expansions, graph-to-grid representations, or equivariant field models, may also satisfy the requirement of site-local chemical identity.

Future work should extend this initial cross-system validation across broader chemical and structural regimes. A natural next step is to evaluate the species-resolved principle over compositionally diverse medium- and high-entropy-alloy datasets, rather than at a single compact quinary stress-test composition, and to include defects, surfaces, magnetic states, and chemical ordering patterns not present in the current training distributions. Another important direction is to combine species-resolved volumetric fields with architectures that better exploit rotational, translational, and periodic symmetries. Finally, downstream validation should examine whether improved density prediction translates into improved prediction of derived physical observables, such as Bader charges, charge transfer, electrostatic potentials, forces, or work functions.

In summary, this work identifies site-local chemical identity as a necessary ingredient for deployable volumetric electron-density prediction in substitutional alloys. Shared structural fields and global composition are insufficient because they are degenerate under chemical site reassignment. Species-resolved StructureChem fields remove this degeneracy, reduce ternary alloy density error by approximately two orders of magnitude, recover the DFT response to controlled site reassignment, and fail when the species-channel correspondence is deliberately broken. A compact quinary cross-system stress test reproduces the same pattern in a five-component system, with an approximately 99.0% reduction in grouped-test logMSE and median same-geometry recovery of 0.967. These results provide both a practical representation for alloy density surrogates and a general audit principle for learning electronic fields in chemically heterogeneous materials.

Methods

Alloy dataset and `cfg_id`-grouped split

We evaluated deployable electron-density surrogates on an Al–Fe–Ni alloy dataset containing 4,002 configurations. Each configuration consists of an atomic structure, a real-space DFT electron-density grid, and associated metadata used for dataset organization. The target is the self-consistent DFT electron density sampled on its native simulation grid.

The train–validation–test partition was grouped using the single metadata field `cfg_id`. Internally, the splitting implementation applied `fillna("NA")` and conversion to string before serializing the grouping columns with a pipe separator; because `cfg_id` was the only grouping column, the resulting parent-group key was its string representation. All rows sharing the same `cfg_id` were assigned to the same split.

Let G denote the first-seen ordered list of the 667 unique `cfg_id` groups. The list was shuffled using `numpy.random.default_rng(0)`. The first 534 groups were assigned to training, the next 67 to validation, and the remaining 66 to testing. All configurations inherited the assignment of their `cfg_id` group, after which the row identifiers within each split were shuffled using the same random-number generator before the split files were written. This procedure produced 3,204 training, 402 validation, and 396 test structures, with no `cfg_id` occurring in more than one split. All headline Al–Fe–Ni results use this fixed partition.

Al–Fe–Ni electronic-structure calculations

Electron densities were calculated with SIESTA 5.4.2 [46, 47] using the Perdew–Burke–Ernzerhof generalized-gradient approximation [48]. Core–valence interactions were represented using scalar-relativistic optimized norm-conserving Vanderbilt pseudopotentials [49] stored in PSML 1.1 format [50].

All 4,002 calculations used a non-spin-polarized treatment, a triple-zeta polarized basis, a PAO energy shift of 0.01 eV, and a real-space mesh cutoff of 400 Ry. Brillouin-zone integration used a $6 \times 6 \times 6$ Monkhorst–Pack grid. The electronic structure was obtained by diagonalization, with at most 150 self-consistent iterations, a density-matrix tolerance of 10^{-4}, and Pulay mixing with weight 0.10 and a history of eight iterations. Fermi–Dirac occupations were used at an electronic temperature of 300 K.

Every configuration was evaluated as a periodic single-point calculation. The archived input files do not request structural relaxation, and force output was disabled. The charge density, electrostatic potential, and density matrix were saved. Embedded PSML provenance identifies the pseudopotential generator as `ONCVPSP-3.3.0+psml-3.3.0-73`, using Hamann's ONCVPSP formulation, scalar relativity, and PBE exchange and correlation. The original upstream pseudopotential distribution is not recorded. The exact PSML files and their SHA-256 hashes are included in the archived dataset.

Quinary cross-system identifiability dataset

The quinary dataset contains 500 converged configurations with fixed composition $Al_2Fe_2Ni_2Co_2Cr_2$. Each configuration contains 10 atoms in a supercell formed by a $5 \times 1 \times 1$ replication of the two-atom parent cell. The dataset comprises 100 parent geometries, each associated with five unique chemical site assignments preserving the simulation cell, atomic coordinates, and global composition.

The data were partitioned by parent geometry, giving 350 training, 75 validation, and 75 test structures from 70, 15, and 15 parent geometries, respectively. No parent group occurs in more than one split. The held-out same-geometry analysis contains all $15\binom{5}{2} = 150$ within-parent pairs from the test split. Direct validation of the corresponding input files gave maximum lattice and coordinate differences of zero.

This dataset is intended as a controlled cross-system identifiability stress test. It is not a comprehensive benchmark for medium- or high-entropy-alloy density prediction.

Quinary electronic-structure calculations

Electron densities were calculated with SIESTA 5.4.2 [46, 47] using the PBE generalized-gradient approximation [48]. Core–valence interactions were represented using scalar-relativistic optimized norm-conserving Vanderbilt pseudopotentials [49] stored in PSML 1.1 format [50].

The calculations used a triple-zeta polarized basis, a PAO energy shift of 0.01 eV, and a real-space mesh cutoff of 400 Ry. Brillouin-zone sampling used a $2 \times 6 \times 6$ k-point grid. Self-consistent calculations used at most 150 iterations, a density-matrix tolerance of 10^{-4}, Fermi–Dirac occupations, and an electronic temperature of 300 K. Electron-density output was enabled for every configuration. The quinary calculations

used the corresponding deposited PSML 1.1 files for Al, Fe, Ni, Co, and Cr. These files contain the same scalar-relativistic PBE and ONCVPSP generator provenance described above.

Localized Gaussian atom-deposition fields

For each atomic configuration, localized atom-centred Gaussian deposition fields were constructed on the real-space grid of the target DFT density. Given an atomic position $\mathbf{R}_i$, the common unnormalized spatial kernel is

$$K_{\sigma_k}(\mathbf{r}, \mathbf{R}_i) = \exp\left[-\frac{d_{\mathrm{MIC}}(\mathbf{r}, \mathbf{R}_i)^2}{2\sigma_k^2}\right],$$

where d_{MIC} is the periodic minimum-image distance. The shared and species-resolved representations use this spatial dependence but differ in their atom weights and normalization, as defined below. We use two Gaussian widths,

$$\sigma_1 = 0.05 \text{ Å}, \qquad \sigma_2 = 0.10 \text{ Å},$$

for every chemical species. These fields are intended as localized atom-deposition channels that encode atomic positions and, in the species-resolved representation, the correspondence between chemical species and sites. They should not be interpreted as independently resolved approximations to the physical electron density.

We audited both widths against the actual stored density grids. For the Al–Fe–Ni alloy dataset, the median equivalent voxel spacing,

$$h_{\mathrm{vol}} = \left(\frac{V_{\mathrm{cell}}}{N_1 N_2 N_3}\right)^{1/3},$$

was 0.0840 Å, while the grid-step lengths along the cell-vector directions ranged from 0.0798 to 0.0937 Å. Because h_{vol} is calculated from the voxel volume, it is an equivalent isotropic spacing rather than a cell-vector step length. For skewed cells, it can therefore be smaller than the grid-step length along each individual cell-vector direction. The two Gaussian widths therefore corresponded to median ratios of

$$\frac{\sigma_1}{h_{\mathrm{vol}}} = 0.595, \qquad \frac{\sigma_2}{h_{\mathrm{vol}}} = 1.190.$$

Their full widths at half maximum spanned median values of 1.401 and 2.802 equivalent grid intervals, respectively.

For the quinary Al–Fe–Ni–Co–Cr dataset, the median equivalent voxel spacing was 0.0794 Å, and the grid-step lengths ranged from 0.0801 to 0.0959 Å. The corresponding median width-to-grid ratios were

$$\frac{\sigma_1}{h_{\mathrm{vol}}} = 0.630, \qquad \frac{\sigma_2}{h_{\mathrm{vol}}} = 1.259,$$

with full widths at half maximum spanning median values of 1.483 and 2.965 equivalent grid intervals.

Thus, the narrow field is sub-grid in standard-deviation units and is more accurately described as a localized deposition kernel than as a smooth continuous field.

Nevertheless, the two discretized widths remain numerically distinct. On held-out Al–Fe–Ni structures, the post-$\log(1+\rho)$ species-occupancy fields produced using the two widths had a median Pearson correlation of 0.716 and a median relative L_2 difference of 0.699. For the quinary dataset, the median Pearson correlation was 0.715 and the median relative L_2 difference was 0.699. The narrow and broad channels therefore provide distinguishable spatial encodings after discretization and transformation, despite the sub-grid standard deviation of the narrower kernel.

The resulting fields are transformed with $\log(1+\rho_{S,k})$ before being passed to the neural network.

Shared StructureChem alloy representation

The shared StructureChem alloy baseline uses two total Gaussian atom-deposition fields and one global composition channel. Its input channels are

$$\left[\rho_{S,1}^{\mathrm{shared}}, \rho_{S,2}^{\mathrm{shared}}, c_{\mathrm{global}}\right],$$

where $\rho_{S,1}^{\mathrm{shared}}$ and $\rho_{S,2}^{\mathrm{shared}}$ are shared Gaussian atom-deposition fields constructed over all atomic sites, and c_{global} is a spatially constant scalar encoding global alloy composition.

For a configuration containing N atoms and a total pseudopotential valence-electron count N_e, every site in the shared representation was assigned the same weight,

$$w_i^{\mathrm{shared}} = \frac{N_e}{N}.$$

For each width k, the unnormalized field was

$$\phi_{S,k}^{\mathrm{shared}}(\mathbf{r}_j) = \sum_{i=1}^{N} \frac{N_e}{N} K_{\sigma_k}(\mathbf{r}_j, \mathbf{R}_i).$$

The narrow and broad fields were normalized independently to N_e:

$$\rho_{S,k}^{\mathrm{shared}}(\mathbf{r}_j) = \phi_{S,k}^{\mathrm{shared}}(\mathbf{r}_j) \frac{N_e}{\Delta V \sum_{l=1}^{M} \phi_{S,k}^{\mathrm{shared}}(\mathbf{r}_l) + 10^{-30}},$$

where M is the number of voxels and $\Delta V = V_{\mathrm{cell}}/M$. The corresponding network channel was

$$x_{S,k}^{\mathrm{shared}}(\mathbf{r}_j) = \log\left[1 + \rho_{S,k}^{\mathrm{shared}}(\mathbf{r}_j)\right].$$

Because N_e depends on composition but not on the assignment of species to sites, configurations with the same geometry and composition produce the same two shared fields.

The global composition channel is spatially constant and is defined as the atom-fraction-weighted mean

$$c_{\mathrm{global}}(\mathbf{r}) = \frac{\sum_s N_s a_s}{\sum_s N_s}, \qquad \forall\, \mathbf{r},$$

where N_s is the number of atoms of species s, and

$$a_{\mathrm{Al}} = 0, \qquad a_{\mathrm{Fe}} = 0.5, \qquad a_{\mathrm{Ni}} = 1, \qquad a_{\mathrm{Co}} = 0.875, \qquad a_{\mathrm{Cr}} = 0.375.$$

The scalar was copied to every grid voxel without standardization or an additional nonlinear transformation. For the Al–Fe–Ni dataset, the observed values were 0.25, 0.4375, 0.5, 0.5625, and 0.75. For the fixed-composition $Al_2Fe_2Ni_2Co_2Cr_2$ dataset, $c_{\mathrm{global}} = 0.55$ for every configuration. The same global channel was supplied to the corresponding shared and species-resolved models.

This representation is deployable because it can be constructed from atomic coordinates, species identities, and the simulation cell. However, it is locally species-blind: it does not encode which atomic site is occupied by Al, Fe, or Ni. Therefore, two configurations with the same geometry and composition but different site assignments can produce indistinguishable shared inputs.

Species-resolved StructureChem alloy representation

The species-resolved StructureChem representation replaces the two shared atom-deposition fields with separate localized fields for each chemical species. For Al–Fe–Ni alloys, the input channels are

$$\left[\rho^{\mathrm{Al}}_{S,1}, \rho^{\mathrm{Al}}_{S,2}, \rho^{\mathrm{Fe}}_{S,1}, \rho^{\mathrm{Fe}}_{S,2}, \rho^{\mathrm{Ni}}_{S,1}, \rho^{\mathrm{Ni}}_{S,2}, c_{\mathrm{global}}\right].$$

For the species-resolved representation, the normalized three-dimensional Gaussian kernel was

$$G_{\sigma_k}(\mathbf{r}, \mathbf{R}_i) = \frac{1}{(2\pi\sigma_k^2)^{3/2}} \exp\left[-\frac{d_{\mathrm{MIC}}(\mathbf{r}, \mathbf{R}_i)^2}{2\sigma_k^2}\right].$$

For species s, the deposited field was

$$\rho^s_{S,k}(\mathbf{r}) = \sum_{i:Z_i=s} w_s G_{\sigma_k}(\mathbf{r}, \mathbf{R}_i).$$

No electron-count or grid-integral renormalization was applied after species-wise deposition. The network input was

$$x^s_{S,k}(\mathbf{r}) = \log\left[1 + \rho^s_{S,k}(\mathbf{r})\right].$$

For the quinary experiment, the shared representation contained two local Gaussian fields and one global auxiliary composition channel, giving three input channels. The species-resolved occupancy representation contained narrow and broad fields for each of Al, Fe, Ni, Co, and Cr, together with the same global auxiliary channel, giving eleven input channels. All species weights were unity.

Occupancy and valence weighting

We evaluated two species-resolved weighting schemes. In the occupancy-weighted representation, all atoms contribute unit weight,

$$w_s = 1.$$

This representation isolates the effect of local species identity while keeping all species channels on a comparable occupancy scale.

In the valence-weighted representation, each species channel is weighted by the corresponding pseudopotential valence,

$$w_s = Z^s_{\mathrm{val}}.$$

This incorporates a physically motivated amplitude scale into the species-resolved fields. Both occupancy- and valence-weighted representations are deployable because they require only species identities, positions, the simulation cell, and pseudopotential information.

Scale-matched occupancy control

To distinguish physical valence information from input-scale effects, we constructed a scale-matched occupancy representation. Starting from the occupancy-weighted species-resolved fields, each post-$\log(1+\rho)$ occupancy channel was affinely rescaled to match the mean and standard deviation of the corresponding valence-weighted channel. Statistics were computed on the training split only and then applied unchanged to validation and test structures. This control preserves the spatial site-occupancy pattern while matching the first-layer input scale of the valence representation.

Neural network architecture

All models use the same three-dimensional U-Net architecture, differing only in the number and meaning of input channels. The model maps a multichannel volumetric input to a scalar electron-density field on the DFT grid. The architecture consists of an encoder–decoder structure with convolutional blocks, downsampling and upsampling stages, skip connections, and a final one-channel output layer.

The encoder widths were 12, 24, and 48 channels, with a 96-channel bottleneck; decoder widths were 48, 24, and 12. Each block contained two $3 \times 3 \times 3$ convolutions followed by GroupNorm and SiLU activations. Downsampling used $2 \times 2 \times 2$ max pooling, and upsampling used stride-two transposed convolutions. Inputs were padded to spatial dimensions divisible by eight and outputs were cropped back to the original grid.

$$\hat{y}(\mathbf{r}) = \min[\mathrm{softplus}(z_\theta(\mathbf{r})), 7]\,, \qquad \hat{\rho}(\mathbf{r}) = \max[\exp(\hat{y}(\mathbf{r})) - 1, 0]\,.$$

The output is trained in log-density space. Let ρ_{true} denote the target DFT density and ρ_{pred} the predicted density. The primary voxelwise loss is computed on the transformed density,

$$y = \log(1+\rho).$$

All Al–Fe–Ni deployable models in the main comparison use the same architecture and training protocol.

Training protocol

All final Al–Fe–Ni deployable StructureChem models were trained for 20 epochs using Adam, a learning rate of 10^{-5}, a batch size of one, and random seeds 0, 1, and 2. Training used gradient clipping with maximum norm 1.0, balanced weighted sampling, no learning-rate scheduler, and no periodic-shift augmentation. The deployable inputs excluded perturbation, displacement-amplitude, close-pair, and interatomic-distance auxiliary channels.

For structure b, the transformed-density loss was

$$\mathcal{L}^{(b)}_{\mathrm{logMSE}} = \frac{1}{M_b} \sum_{j=1}^{M_b} \left(\hat{y}_{b,j} - y_{b,j}\right)^2 ,$$

where

$$\hat{y} = \min[\mathrm{softplus}(f_\theta(x)) , 7] , \qquad y = \log(1 + \rho_{\mathrm{DFT}}).$$

The integrated-density term was

$$\mathcal{L}^{(b)}_{\mathrm{sumRel}} = \frac{\left|\sum_j \mathrm{expm1}(\hat{y}_{b,j}) - \sum_j \mathrm{expm1}(y_{b,j})\right|}{\max\left[\sum_j \mathrm{expm1}(y_{b,j}), 10^{-3}\right]}.$$

The batch-averaged training objective was

$$\mathcal{L}_{\mathrm{train}} = \left\langle \mathcal{L}^{(b)}_{\mathrm{logMSE}} \right\rangle_b + 0.01 \left\langle \mathcal{L}^{(b)}_{\mathrm{sumRel}} \right\rangle_b .$$

All other auxiliary-loss coefficients were zero:

$$\beta_{\Delta q} = \beta_{\mathrm{symq}} = \beta_{\mathrm{balance}} = \beta_{\mathrm{interstitial}} = 0.$$

For every run, checkpoint selection minimized the total validation objective defined by the same composite loss, rather than validation logMSE alone. Held-out test metrics were evaluated only after checkpoint selection and were summarized as means and standard deviations across the three seeds.

Evaluation metrics

We evaluated density prediction with voxelwise, integrated-charge, and shape-normalized metrics. The log-density mean squared error is

$$\mathrm{logMSE} = \frac{1}{M} \sum_{j=1}^{M} \left[\log(1 + \rho_{\mathrm{pred},j}) - \log(1 + \rho_{\mathrm{true},j})\right]^2 ,$$

where M is the number of grid voxels.

The density mean absolute error and root mean squared error are

$$\rho\mathrm{MAE} = \frac{1}{M}\sum_{j=1}^{M}\left|\rho_{\mathrm{pred},j} - \rho_{\mathrm{true},j}\right|,$$

and

$$\rho\mathrm{RMSE} = \left[\frac{1}{M}\sum_{j=1}^{M}\left(\rho_{\mathrm{pred},j} - \rho_{\mathrm{true},j}\right)^2\right]^{1/2}.$$

The relative integrated electron-count error is

$$\mathrm{sumRel} = \frac{\left|\sum_j \rho_{\mathrm{pred},j} - \sum_j \rho_{\mathrm{true},j}\right|}{\max\left(\left|\sum_j \rho_{\mathrm{true},j}\right|, 10^{-12}\right)}.$$

To separate spatial shape error from global scale error, we also compute a normalized-shape metric. The predicted and true densities are normalized by their grid sums,

$$\tilde{\rho}_{\mathrm{pred},j} = \frac{\rho_{\mathrm{pred},j}}{\sum_l \rho_{\mathrm{pred},l}}, \qquad \tilde{\rho}_{\mathrm{true},j} = \frac{\rho_{\mathrm{true},j}}{\sum_l \rho_{\mathrm{true},l}}.$$

The shape-normalized logMSE is then

$$\mathrm{shape_logMSE} = \frac{1}{M}\sum_{j=1}^{M}\left[\log\left(1 + M\tilde{\rho}_{\mathrm{pred},j}\right) - \log\left(1 + M\tilde{\rho}_{\mathrm{true},j}\right)\right]^2.$$

Same-geometry site-assignment test

To isolate the role of local chemical identity, we constructed a same-geometry site-assignment test. Each pair consists of two alloy configurations with identical simulation cells, identical atomic coordinates, and identical global compositions, but different assignments of Al, Fe, and Ni to atomic sites. Thirteen strict pairs satisfying these criteria were retained.

For each pair, we computed the DFT density-difference field,

$$\Delta\rho_{\mathrm{DFT}} = \rho^{(2)}_{\mathrm{DFT}} - \rho^{(1)}_{\mathrm{DFT}},$$

and the corresponding model-predicted density-difference field,

$$\Delta\rho_{\mathrm{pred}} = \rho^{(2)}_{\mathrm{pred}} - \rho^{(1)}_{\mathrm{pred}}.$$

We then evaluated how well the model recovered the DFT response to site reassignment.

The recovery ratio is defined as

$$R_{\mathrm{rec}} = \frac{M^{-1}\sum_{j=1}^{M}\left|\Delta\rho_{\mathrm{pred},j}\right|}{\max\left[M^{-1}\sum_{j=1}^{M}\left|\Delta\rho_{\mathrm{DFT},j}\right|, \epsilon\right]}.$$

Here, $\epsilon = 10^{-12}$ is a numerical safeguard used only when the reference density-difference magnitude vanishes.

This definition is the ratio of the mean absolute predicted and DFT density-difference fields. The relative L_2 error reported below is a separate metric.

A recovery ratio close to zero indicates that the model is nearly invariant under site reassignment. A recovery ratio close to one indicates that the model predicts a density-difference magnitude comparable to the DFT response.

We also compute the Pearson correlation between the flattened fields $\Delta\rho_{\mathrm{pred}}$ and $\Delta\rho_{\mathrm{DFT}}$, and the relative L_2 error,

$$\text{relative } L_2 = \frac{\|\Delta\rho_{\mathrm{pred}} - \Delta\rho_{\mathrm{DFT}}\|_2}{\max\left(\|\Delta\rho_{\mathrm{DFT}}\|_2, \epsilon\right)}.$$

The same definitions were used for the quinary cross-system stress test experiment. Quinary pairs had identical simulation cells, atomic coordinates, and global compositions but different assignments of the five chemical species to atomic sites. The quinary test is reported separately from the 13-pair Al–Fe–Ni analysis because it serves as an independent cross-system stress test.

For the quinary analysis, pairs were constructed exclusively within the held-out test split. The 15 test parent geometries each contained five site assignments, producing $15\binom{5}{2} = 150$ unordered pairs. Exact validation of the underlying input files gave maximum lattice and coordinate differences of zero. The earlier all-data audit contained 1,075 pairs and is reported only in Supplementary Information.

Species-channel permutation test

To test whether the species-resolved model actively uses the association between species identity and atomic position, we performed species-channel permutation experiments. Starting from the trained deployable valence-weighted species-resolved StructureChem model, we permuted only the species-resolved field channels corresponding to Al, Fe, and Ni. The global composition channel and target densities were left unchanged.

The canonical permutations included the identity mapping, pairwise swaps of Al–Fe, Al–Ni, and Fe–Ni channels, and cyclic permutations among Al, Fe, and Ni. For each permutation, we evaluated the change in prediction error relative to the unpermuted model. We report the prediction MAE relative to the identity prediction, the increase in logMSE, the error ratio, the change in relative electron-count error, and the change in shape-normalized logMSE.

Large degradation under species-channel permutation indicates that the trained model depends on the correct correspondence between species channels and atomic sites, rather than merely benefiting from additional input channels.

For the quinary experiment, the diagnostic permutation set comprised the identity mapping, Al–Fe, Al–Ni, and Co–Cr swaps, one deterministic full permutation, and the five-cycle Al→Fe→Ni→Co→Cr →Al. Only the ten local species fields were permuted; the global auxiliary channel, geometry, and target density were unchanged. The set was diagnostic rather than an exhaustive enumeration of all 5! permutations.

Statistical summaries

All grouped alloy benchmark metrics are summarized across three random seeds. For each model and metric, we report the mean and standard deviation over seeds. Relative improvements are computed with respect to the shared StructureChem baseline as

$$\text{Improvement} = 100\left(1 - \frac{\text{metric}_{\text{model}}}{\text{metric}_{\text{shared}}}\right).$$

The quinary stress test was evaluated using one training seed. Its grouped metrics are therefore macro-averages over held-out structures and its same-geometry statistics quantify variation across held-out parent geometries; they are not estimates of training-seed variability.

For the same-geometry site-assignment tests, we report mean and median values over the retained counterfactual pairs. Uncertainty in the median recovery ratio and median density-difference correlation was estimated using a nonparametric cluster bootstrap with 100,000 replicates and random seed 20260731. The resampling unit was the independent parent geometry rather than the individual pair. The ternary analysis contained 13 independent geometry clusters with one retained pair per cluster. The quinary analysis contained 15 held-out parent geometries, each contributing all $\binom{5}{2} = 10$ within-parent pairs, giving 150 pairs in total. All pairs associated with a sampled parent were retained together, and identical cluster draws were used for the shared and species-resolved models. Two-sided 95% percentile intervals were obtained from the bootstrap distributions.

Pearson correlation was treated as undefined when the predicted density-difference field had zero variance. It was therefore not reported for the quinary shared model, for which correlation was undefined in 118 of 150 pairs. The reported confidence intervals quantify variation across held-out geometry groups conditional on the fixed, validation-selected checkpoints and do not include training-seed uncertainty.

Because the ternary and quinary analyses contain only 13 and 15 independent geometry clusters, respectively, the nominal coverage of the percentile intervals may be imperfect in finite samples. The intervals should therefore be interpreted as indicative measures of sensitivity to the sampled held-out geometry families rather than as precisely calibrated population coverage bounds.

Deployability constraints

A central constraint of this work is that the main alloy conclusions must rely only on deployable inputs. Therefore, the headline models exclude generator-derived perturbation metadata, perturbation-type flags, perturbation amplitudes, and synthetic labels that are not generally available for arbitrary input structures. Non-deployable privileged models may be useful as upper-bound controls, but they are not used to support the main claims of this study. The main comparison is therefore between a deployable locally species-blind StructureChem baseline and a deployable species-resolved StructureChem model.

Data availability

The Al–Fe–Ni dataset and the quinary $Al_2Fe_2Ni_2Co_2Cr_2$ stress-test dataset are uploaded to Zenodo [51] and available at https://doi.org/10.5281/zenodo.21890765. The single-element pretraining dataset is available at https://doi.org/10.5281/zenodo.20378934 [52].

Code availability

The code used for preprocessing, training, evaluation, figure generation, and table generation is available at the project's GitHub repository https://github.com/ProyectoPhantoma/representation_identifiability.

Acknowledgements

This work was supported by the Generación 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. Physical Review **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. Physical Review **140**(4A), 1133–1138 (1965) https://doi.org/10.1103/PhysRev.140.A1133

[3] Parr, R.G., Yang, W.: Density-Functional Theory of Atoms and Molecules. Oxford University Press, Oxford (1989)

[4] Kohn, W.: Density functional and density matrix method scaling linearly with the number of atoms. Physical Review Letters **76**(17), 3168–3171 (1996) https://doi.org/10.1103/PhysRevLett.76.3168

[5] Curtarolo, S., Setyawan, W., Hart, G.L.W., Jahnatek, M., Chepulskii, R.V., Taylor, R.H., Wang, S., Xue, J., Yang, K., Levy, O., Mehl, M.J., Stokes, H.T.,

Demchenko, D.O., Morgan, D.: AFLOW: An automatic framework for high-throughput materials discovery. Computational Materials Science **58**, 218–226 (2012) https://doi.org/10.1016/j.commatsci.2012.02.005

[6] 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

[7] Saal, J.E., Kirklin, S., Aykol, M., Meredig, B., Wolverton, C.: Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD). JOM **65**(11), 1501–1509 (2013) https://doi.org/10.1007/s11837-013-0755-4

[8] Draxl, C., Scheffler, M.: The NOMAD laboratory: from data sharing to artificial intelligence. Journal of Physics: Materials **2**(3), 036001 (2019) https://doi.org/10.1088/2515-7639/ab13bb

[9] 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

[10] Behler, J., Parrinello, M.: Generalized neural-network representation of high-dimensional potential-energy surfaces. Physical Review Letters **98**(14), 146401 (2007) https://doi.org/10.1103/PhysRevLett.98.146401

[11] Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E.: Neural message passing for quantum chemistry. In: Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 70, pp. 1263–1272. PMLR, NY, USA (2017). https://proceedings.mlr.press/v70/gilmer17a.html

[12] Schütt, K.T., Sauceda, H.E., Kindermans, P.-J., Tkatchenko, A., Müller, K.-R.: SchNet: A deep learning architecture for molecules and materials. The Journal of Chemical Physics **148**(24), 241722 (2018) https://doi.org/10.1063/1.5019779

[13] Xie, T., Grossman, J.C.: Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Physical Review Letters **120**(14), 145301 (2018) https://doi.org/10.1103/PhysRevLett.120.145301

[14] Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., Riley, P.: Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. arXiv preprint arXiv:1802.08219 (2018) https://doi.org/10.48550/arXiv.1802.08219 arXiv:1802.08219 [cs.LG]

[15] Satorras, V.G., Hoogeboom, E., Welling, M.: E(n) equivariant graph neural networks. In: Meila, M., Zhang, T. (eds.) Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 139, pp. 9323–9332. PMLR, NY, USA (2021). https://proceedings.mlr.press/v139/satorras21a.html

[16] Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J.P., Kornbluth, M., Molinari, N., Smidt, T.E., Kozinsky, B.: $E(3)$-equivariant graph neural networks for data-efficient and accurate interatomic potentials. Nature Communications **13**, 2453 (2022) https://doi.org/10.1038/s41467-022-29939-5

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

[18] Sinitskiy, A.V., Pande, V.S.: Deep neural network computes electron densities and energies of a large set of organic molecules faster than density functional theory (DFT). arXiv preprint arXiv:1809.02723 (2018) https://doi.org/10.48550/arXiv.1809.02723 arXiv:1809.02723 [physics.chem-ph]

[19] Lv, T., Zhong, Z.-C., Liang, Y., Li, F., Huang, J., Zheng, R.K.: Deep charge: Deep learning model of electron density from a one-shot density functional theory calculation. Physical Review B **108**(23), 235159 (2023) https://doi.org/10.1103/PhysRevB.108.235159

[20] Li, C., *et al.*: Image super-resolution inspired electron density prediction. Nature Communications **16**, 5487 (2025) https://doi.org/10.1038/s41467-025-60095-8

[21] 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

[22] 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**(18), 184103 (2019) https://doi.org/10.1103/PhysRevB.100.184103

[23] Jørgensen, P.B., Jacobsen, K.W., Schmidt, M.N.: Equivariant graph neural networks for fast electron density prediction. npj Computational Materials **8**, 1–11 (2022) https://doi.org/10.1038/s41524-022-00863-y

[24] Achar, S.K., Pons, J., Dao, M., Suresh, S., Yip, S.: Machine learning electron density prediction using smooth overlap of atomic positions. Nanomaterials **13**(12), 1853 (2023) https://doi.org/10.3390/nano13121853

[25] Focassio, B., Domina, M., Patil, U., Fazzio, A., Sanvito, S.: Linear jacobi–legendre

expansion of the charge density for machine learning-accelerated electronic structure calculations. npj Computational Materials **9**, 87 (2023) https://doi.org/10.1038/s41524-023-01053-0

[26] 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

[27] 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

[28] Cuevas-Zuviría, B., Pacios, L.F.: Machine learning of analytical electron density in large molecules through message-passing. Journal of Chemical Information and Modeling **61**(6), 2658–2666 (2021) https://doi.org/10.1021/acs.jcim.1c00227

[29] Grisafi, A., Lewis, A.M., Rossi, M., Ceriotti, M.: Electronic-structure properties from atom-centered predictions of the electron density. Journal of Chemical Theory and Computation **19**(15), 4451–4462 (2023) https://doi.org/10.1021/acs.jctc.2c00850

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

[31] Schütt, K.T., Gastegger, M., Tkatchenko, A., Müller, K.-R., Maurer, R.J.: Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. Nature Communications **10**, 5024 (2019) https://doi.org/10.1038/s41467-019-12875-2

[32] Unke, O.T., Bogojeski, M., Gastegger, M., Geiger, M., Smidt, T., Müller, K.-R.: SE(3)-equivariant prediction of molecular wavefunctions and electronic densities. In: Advances in Neural Information Processing Systems, vol. 34, pp. 14434–14447 (2021)

[33] del Rio, 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

[34] Koker, T., Quigley, D., Zacchia, N., *et al.*: 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

[35] Fiedler, L., Modine, N.A., Schmerler, S., Vogel, D.J., Popoola, G.A., Thompson, A.P., Rajamanickam, S., 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

[36] Dick, S., Fernandez-Serra, M.: Machine learning accurate exchange and correlation functionals of the electronic density. Nature Communications **11**, 3509 (2020) https://doi.org/10.1038/s41467-020-17265-7

[37] Grisafi, A., Salanne, M.: Accelerating QM/MM simulations of electrochemical interfaces through machine learning of electronic charge densities. The Journal of Chemical Physics **161**(2), 024109 (2024) https://doi.org/10.1063/5.0218379

[38] 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

[39] Drautz, R.: Atomic cluster expansion for accurate and transferable interatomic potentials. Physical Review B **99**, 014104 (2019) https://doi.org/10.1103/PhysRevB.99.014104

[40] Pozdnyakov, S.N., Willatt, M.J., Bartók, A.P., Ortner, C., Csányi, G., Ceriotti, M.: Incompleteness of atomic structure representations. Physical Review Letters **125**(16), 166001 (2020) https://doi.org/10.1103/PhysRevLett.125.166001

[41] Nigam, J., Pozdnyakov, S.N., Huguenin-Dumittan, K.K., Ceriotti, M.: Completeness of atomic structure representations. APL Machine Learning **2**(1), 016110 (2024) https://doi.org/10.1063/5.0160740

[42] Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? In: International Conference on Learning Representations (2019). https://openreview.net/forum?id=ryGs6iA5Km

[43] Wagstaff, E., Fuchs, F.B., Engelcke, M., Osborne, M.A., Posner, I.: Universal approximation of functions on sets. Journal of Machine Learning Research **23**(151), 1–56 (2022)

[44] Arévalo, I.: Volumetric electron-density prediction in elemental metals: continuous-field accuracy and topology-sensitive validations (2026). https://doi.org/10.21203/rs.3.rs-10339992/v1

[45] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

[46] 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**(11), 2745 (2002). https://doi.org/10.1088/0953-8984/14/11/302

[47] García, A., *et al.*: Siesta: Recent developments and applications. The Journal of Chemical Physics **152**(20), 204108 (2020). https://doi.org/10.1063/5.0005077

[48] Perdew, J.P., Burke, K., Ernzerhof, M.: Generalized gradient approximation made simple. Phys. Rev. Lett. **77**, 3865–3868 (1996). https://doi.org/10.1103/PhysRevLett.77.3865

[49] Hamann, D.R.: Optimized norm-conserving vanderbilt pseudopotentials. Physical Review B **88**(8), 085117 (2013) https://doi.org/10.1103/PhysRevB.88.085117

[50] García, A., Verstraete, M.J., Pouillon, Y., Junquera, J.: The psml format and library for norm-conserving pseudopotential data curation and interoperability. Computer Physics Communications **227**, 51–71 (2018). https://doi.org/10.1016/j.cpc.2018.02.011

[51] Arévalo, I.: Al-Fe-Ni and Al-Fe-Ni-Co-Cr Alloy Electron-Density Datasets. https://doi.org/10.5281/zenodo.21890765

[52] Arévalo, I., Olleros, P.: SIESTA Single-Element Density Dataset. Zenodo. https://doi.org/10.5281/zenodo.20378934 (2026)