chemistry

MDPI

Article

Atomic Charges from Machine-Learned Charge Densities: Consistency and Substituent Effects

Xuejian Qin [1,2,3] and Taoyuze Lv [3,4,*]

[1] State Key Laboratory of Advanced Marine Materials, Zhejiang Key Laboratory of Extreme-Environmental Material Surfaces and Interfaces, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315201, China; qinxuejian@nimte.ac.cn
[2] College of Materials Science and Opto-Electronic Technology, University of Chinese Academy of Sciences, Beijing 100049, China
[3] Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215123, China
[4] School of Artificial Intelligence and Data Science, University of Science and Technology of China, Hefei 230026, China
* Correspondence: taoyuze.lv@ustc.edu.cn

Abstract

Atomic charges are widely used to analyze molecular electronic structure and substituent effects, yet their numerical values and interpretations are inherently dependent on the adopted density partitioning scheme. Here, we adapt the Equivariant Atomic Contribution framework to molecular systems (EAC-qm), enabling prediction of atom-resolved continuous charge densities from which atomic charges are obtained as spatial moments. The predicted densities reproduce reference density functional theory results with high accuracy and preserve global charge conservation. To assess chemical interpretability, we examine charge responses in monosubstituted aromatic systems using Hammett substituent constants as external empirical references. Atomic charges derived from EAC-qm exhibit a strong linear association with Hammett parameters, compared with values obtained from traditional density partitioning approaches applied to the same electronic structures. These correlations indicate that density-derived charges respond systematically to established substituent electronic trends. Beyond scalar charges, atom-resolved dipole moments can be evaluated as first-order moments of the same continuous density representation. Illustrative examples for formaldehyde (H_2CO) and formamide ($HCONH_2$) show that local dipole vectors provide directional information about intra-atomic polarization that is not captured by point-charge models. Overall, the results suggest that machine-learned continuous electron densities provide a representation-consistent basis for constructing atom-centered electronic descriptors with chemical interpretability.

Keywords: machine-learned electron density; atomic charges; density partitioning; substituent effects; Hammett constants

Academic Editor: Ryohei Kishi

Received: 15 February 2026
Revised: 1 March 2026
Accepted: 12 March 2026
Published: 16 March 2026

1. Introduction

Atomic charges are widely employed to quantify molecular electronic structure, polarity, chemical reactivity, and substituent effects. However, atomic charges are not strictly observable quantities; their values and physical interpretation depend heavily on the partitioning scheme adopted. A variety of atomic charge partitioning schemes have been developed to date. According to the quantity being partitioned, they can be broadly classified into methods based on atomic orbitals or wavefunctions (such as Mulliken charges [1]

and Natural Bond Orbital (NBO) analysis [2]), methods based on real-space partitioning of the electron density (such as Bader [3], Hirshfeld analyses [4], and Density Derived Electrostatic and Chemical (DDEC) [5,6]), and fitting approaches aimed at reproducing the molecular electrostatic potential (such as RESP charges [7]). While these methods successfully translate quantum chemical results into atomic-level features and have proven valuable for analyzing substituent effects and reactivity trends, the underlying physical assumptions regarding electron assignment and localization differ significantly across schemes. Consequently, atomic charges derived from different methods may lack a unified quantitative scale for describing electronic effects in certain systems or chemical contexts, limiting their reliability as empirical correlation features.

In recent years, a variety of machine-learning approaches such as PACMAN [8], PhysNet [9], Chemprop [10], etc. Refs. [11–14], have been developed to predict atomic properties, including atomic charges, based on local environments. Most of these methods rely on related local representations to map atomic neighbors to atom-resolved quantities. In practice, such models are typically trained on atomic charges obtained from conventional partitioning schemes, which serve as supervision labels. Therefore, these models inevitably inherit the limitations of the underlying charge-partitioning methods. Atomic charges are not uniquely defined. Consequently, the assumptions and biases embedded in a given partitioning procedure are directly transferred to the machine-learning model, thereby constraining both its learning target and its interpretability.

In addition, if one aims to obtain higher-order quantities, such as atomic dipole moments, an additional model is typically required. In this case, the two end-to-end predicted quantities fundamentally lack representation-level consistency. In this work, representation-level consistency refers to the requirement that atomic charges, dipole moments, and higher-order descriptors arise as successive spatial moments of the same underlying atom-resolved charge density representation, ensuring that all derived quantities share a common physical origin. That is, although the training labels may originate from the same charge partitioning scheme, the quantities are learned as separate prediction tasks. Since the dipole is not computed from the predicted charges but directly predicted by another network or output head, their internal representations are not guaranteed to be physically consistent.

In addition to charge-prediction models trained on predefined partitioning schemes, several recent frameworks have aimed at learning the full electron density directly. While these approaches provide spatially resolved electronic information, they typically treat the total density as a regression target without explicitly addressing the question of atomic decomposition. As a result, atomic charges or multipoles must still be obtained through a subsequent partitioning or projection procedure. Conceptually, existing machine-learning strategies therefore fall into two categories: either they inherit the assumptions of a predefined charge-partitioning scheme through supervised learning, or they learn the total density but rely on post hoc analysis to extract atom-resolved quantities. In neither case is the atomic decomposition itself learned as an intrinsic component of the model. This observation highlights a fundamental gap: whether we can construct a framework that learns an atomic density decomposition directly from data, without imposing predefined partitioning rules and without requiring post-processing to recover atom-resolved charges.

In our previous work, we introduced the equivariant atomic contribution network (EAC-Net) [15]. Rather than predicting scalar charges directly, the model learns a spatially resolved contribution of charge density associated with each atom, whose sum reproduces the total charge density of the system. Atomic charges can then be obtained as a derived quantity by integrating the corresponding atomic contribution over all space. From

this perspective, EAC constitutes a machine-learning-defined charge-partitioning scheme, constructed implicitly through the learned charge-density decomposition.

Despite this appealing formulation, the physical soundness and numerical reliability of such a machine-learning-induced charge partitioning have not yet been systematically examined, particularly for organic molecular systems. In the present work, we address this gap by applying the EAC framework to a diverse set of organic molecules and critically assessing its performance and behavior in the context of atomic charge partitioning. The previous EAC-mp foundation model was trained on chemically diverse inorganic crystal structures from the Materials Project database [16]. Based on this, we develop the EAC-qm model optimized for molecular systems based on the QM9 dataset [17,18] and apply it to a series of representative chemical problems. First, using Hammett substituent constants as empirical benchmarks, we assess the chemical consistency of atomic net charges derived from EAC-qm in describing aromatic substituent effects, in comparison with conventional charge partitioning schemes such as DDEC6 and NBO. Second, building upon the continuous atomic charge densities predicted by the model, we introduce atomic electric dipole moments and their local, nucleus-referenced forms, providing a physically transparent framework for analyzing atomic-scale polarization and charge reorganization. Together, these results demonstrate that machine-learned continuous charge densities enable a unified and extensible description of atomic electronic structure that goes beyond scalar charge assignments.

In contrast, the present EAC-based framework explicitly learns an atom-resolved density decomposition as an intrinsic component of the model architecture. This enables several new capabilities.

- Atomic net charges and higher-order multipole moments arise naturally as spatial moments of the same learned atomic density contributions, ensuring representation-level consistency between scalar and vectorial descriptors.
- The atomic decomposition is not imposed through predefined partitioning rules, but emerges from supervised learning of the total density, allowing the model to define an internally consistent, data-driven partitioning scheme.
- As the decomposition is continuous and equivariant by construction, the resulting atom-centered descriptors remain compatible with symmetry constraints and can be systematically extended to higher-order moments without modifying the learning target.

Therefore, the advantage of the present approach is primarily conceptual—learning the atomic density decomposition itself—while also providing practical benefits by eliminating the need for separate models or post hoc partitioning procedures to obtain atom-resolved electronic quantities.

2. Materials and Methods

2.1. EAC Framework

The Equivariant Atomic Contribution (EAC) framework was previously introduced for real-space charge density prediction using equivariant atomic representations [15]. The framework predicts electron densities by explicitly decomposing the total charge density into atom-centered contributions while preserving exact $\mathrm{E}(3)$ equivariance. Rather than assigning scalar atomic charges through post hoc partitioning schemes, the EAC framework learns a continuous atomic charge density $\rho_A(\boldsymbol{r})$ for each atom A, such that the total electron density is expressed as

$$\rho(\boldsymbol{r}) = \sum_A \rho_A(\boldsymbol{r}). \tag{1}$$

This formulation defines an intrinsic, model-consistent density decomposition and enables atomic properties such as net charges and dipole moments to be derived directly from continuous charge densities.

The EAC architecture consists of three conceptually distinct stages: (i) atom–atom interaction, (ii) atom–grid coupling, and (iii) density decoding, as illustrated in Figure 1.

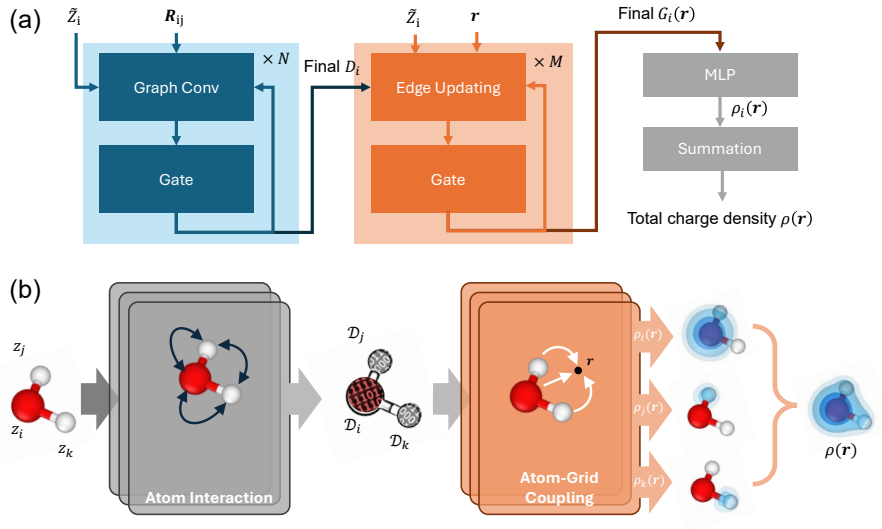

Figure 1. Architecture of the EAC framework. (**a**) Computational workflow of the model. Atomic features are first processed through stacked equivariant interaction layers to obtain atomic descriptors. These descriptors are then coupled with spatial coordinates $\boldsymbol{r}$ to construct atom–grid features, which are decoded by a neural network to produce atom-resolved density contributions. (**b**) Illustration of the three-stage structure of the framework, including atom–atom interaction, atom–grid coupling, and density decoding. The atom-resolved contributions $\rho_i(\boldsymbol{r})$ are summed to obtain the total electron density $\rho(\boldsymbol{r})$. Different colored boxes indicate different modules of the network.

2.1.1. Atom–Atom Interaction

Each atom is first embedded according to its atomic number and local chemical environment. Atomic representations are constructed using an E(3)-equivariant graph neural network implemented within the `e3nn` framework [19]. Following the NequIP architecture [20], atomic features are expanded in irreducible representations characterized by angular momentum orders $l = 0, 1, \ldots, L$, enabling directional and anisotropic information to be encoded in a symmetry-consistent manner.

Through successive equivariant message-passing layers, atomic descriptors are updated by aggregating information from neighboring atoms within a finite cutoff radius. Tensor products between atomic features and relative position vectors ensure strict rotational equivariance. This stage produces a set of atom-centered equivariant descriptors $\{D_i^{(l)}\}$ that encode chemically meaningful information independent of any spatial discretization.

2.1.2. Atom–Grid Coupling

A defining feature of the EAC framework is the explicit coupling between atom-centered descriptors and real-space grid points. Instead of constructing independent features at each spatial location, atomic descriptors are directly coupled to grid coordinates through atom–grid edges.

For each atom A and spatial point $\boldsymbol{r}$, equivariant edge features are constructed as functions of the relative position vector $(\boldsymbol{r} - \boldsymbol{R}_A)$ and the corresponding atomic descriptors. Directional information is incorporated through spherical harmonic expansions of the relative position vector. This design allows atomic representations learned in the atom–

atom stage to be reused consistently across arbitrary grid resolutions, decoupling atomic representation learning from spatial discretization.

2.1.3. Density Decoding and Atomic Decomposition

The atomic charge density contribution $\rho_A(\boldsymbol{r})$ is decoded from scalar ($l = 0$) components of the equivariant edge features using neural networks that determine both magnitude and spatial weighting. The total electron density is obtained by summing over all atomic contributions, as shown in Equation (1).

Because the atomic density contributions are produced directly by the model output, no external charge partitioning scheme is required. The atomic decomposition is therefore intrinsic to the learned representation and remains self-consistent across different chemical environments.

In this work, we employ a molecularly optimized variant, denoted EAC-qm, obtained via transfer learning from a pre-trained EAC-mp foundation model. The foundation model was originally trained on chemically diverse crystal structures from the Materials Project database [16]. The molecular adaptation is performed using charge density data derived from the QM9 dataset [17,18]. This strategy preserves chemically transferable atomic representations while adapting the atom–grid coupling and density decoding modules to organic molecular systems.

2.2. Dataset and Data Partitioning

To adapt the EAC framework for molecular systems, charge density data derived from the QM9 dataset were employed [17]. The QM9 dataset was selected primarily because it provides a large number of small organic molecules with publicly available, consistently computed three-dimensional electron density data suitable for supervised learning of continuous charge densities. QM9-based density datasets explicitly provide real-space charge density grids, which are essential for training and validating the atom-resolved density decomposition framework adopted in this work. The QM9 dataset comprises equilibrium geometries of small organic molecules containing H, C, N, O, and F atoms. In this work, we utilize a publicly available extension of QM9 in which charge densities were computed using the Vienna *Ab initio* Simulation Package (VASP) [18]. All charge densities correspond to converged ground-state density functional theory (DFT) calculations and are provided as three-dimensional real-space grids.

The raw dataset is organized into compressed archives, each containing charge density files for 1000 independent molecular structures. To construct training, validation, and test sets, complete archives were selected to avoid information leakage between subsets. Specifically, fifteen archives (15,000 structures in total) were used for training, two archives (2000 structures) were used for validation, and two additional archives (2000 structures) were reserved as an independent test set. Splitting at the archive level ensures that no molecular structure appears in more than one subset, thereby preventing data contamination across training and evaluation stages.

All selected molecules are neutral, closed-shell systems containing only H, C, N, O, and F atoms, consistent with the elemental scope of QM9. No additional chemistry-based filtering was applied beyond the availability of converged charge density data. The resulting dataset spans a chemically diverse set of small organic molecules and functional groups representative of the QM9 chemical space.

Because each charge density is defined on a dense three-dimensional grid with extremely high dimensionality, a grid-point sampling strategy was employed during training to reduce computational cost while preserving representational expressiveness. For each molecular structure, a fixed number of spatial grid points were uniformly sampled from

the full charge density grid. The sampled grid coordinates serve as input spatial locations, while the corresponding DFT charge density values are used as supervision targets. This sampling strategy allows efficient training without explicitly storing or processing the full three-dimensional grid at each optimization step.

All molecular identifiers corresponding to the training, validation, and test subsets, as well as the trained model parameters and configuration files, are publicly available through the versioned Zenodo repository [21] associated with this work. This ensures full transparency and reproducibility of the dataset construction and model evaluation protocol.

2.3. Model Training and Hyperparameters

To construct a molecularly optimized model, we fine-tuned the pre-trained EAC-mp foundation model [15] on the QM9 charge density dataset, resulting in the EAC-qm model used in this work.

2.3.1. Transfer Learning Strategy

In the EAC-qm model, transfer learning was employed to preserve chemically transferable atomic representations learned from chemically diverse inorganic systems in the Materials Project dataset [16]. During fine-tuning, the parameters of the atom–atom interaction network, which generate equivariant atomic descriptors encoding local chemical environments, were frozen and not updated by gradient descent. Only the atom–grid coupling module and the density decoding networks were optimized using molecular charge density data. This strategy retains chemically meaningful atomic representations while adapting the real-space density reconstruction to the molecular domain.

2.3.2. Training Procedure

Model training was performed using stochastic gradient descent with mini-batches consisting of multiple molecular structures. For each structure in a batch, a fixed number of spatial grid points were sampled. The loss function was defined as the mean absolute error between predicted and reference DFT charge densities evaluated at the sampled grid points. The normalized mean absolute error (NMAE) was used to monitor convergence during training and validation.

Training was conducted for 200,000 optimization steps using a learning rate schedule ranging from 1.0×10^{-3} to 1.0×10^{-5}. All models employ spherical harmonic expansions up to a maximum angular momentum order $L = 5$, ensuring sufficient angular resolution while maintaining computational efficiency. The main hyperparameters used for model training are summarized in Table 1.

Training was performed on a computational node equipped with NVIDIA RTX A6000 GPUs. Under this hardware configuration, the complete fine-tuning process required approximately 16 h.

Table 1. Hyperparameters used for model training.

Parameter	EAC-qm-Scratch	EAC-qm (Fine-Tuned)
Training steps	200k	200k
Structures per batch	15	15
Grid points per structure	25	25
Minimum learning rate	1.0×10^{-5}	1.0×10^{-5}
Maximum learning rate	1.0×10^{-3}	1.0×10^{-3}
Total model parameters	3.08 M	3.08 M
Trainable parameters	3.08 M	0.85 M
Maximum spherical harmonic order	5	5

2.3.3. Training Convergence

Figure 2 shows the training and validation NMAE curves of the EAC-qm model as a function of optimization steps. The validation error closely follows the training error throughout optimization, indicating stable convergence without significant overfitting.

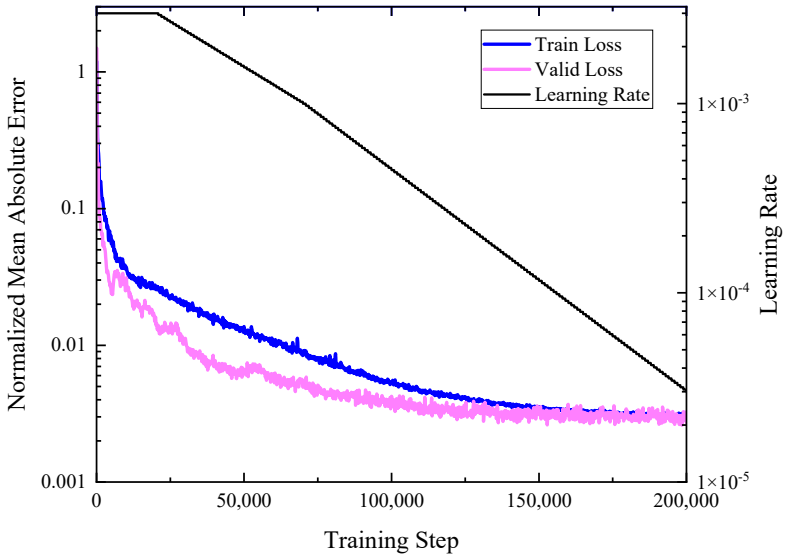

Figure 2. Training and validation normalized mean absolute error (NMAE) as a function of optimization steps for the EAC-qm model. The learning rate schedule is shown for reference.

2.4. Reference Quantum Chemical Calculations

To provide physically consistent reference data and enable comparison with established atomic charge and multipole analysis schemes, density functional theory (DFT) calculations and post-processing analyses were performed using Vienna *Ab initio* Simulation Package (VASP), Gaussian 16, and DDEC6.

2.4.1. VASP Calculations

Reference charge densities were obtained from plane-wave DFT calculations performed using VASP [22]. The projector augmented-wave (PAW) method was employed to describe core-valence interactions [23], together with the Perdew–Burke–Ernzerhof (PBE) generalized gradient approximation functional [24].

An energy cutoff of 400 eV was used for the plane-wave basis set. For isolated molecules, sufficiently large simulation cells were constructed to eliminate spurious periodic interactions, and Γ-point sampling was applied. Electronic self-consistent field (SCF) convergence was set to 1×10^{-6} eV. Static self-consistent calculations (NSW = 0) were performed to obtain converged ground-state charge densities, which were subsequently used for model training and analysis.

2.4.2. Gaussian 16 and NBO Analysis

Natural Bond Orbital (NBO) analyses were carried out using Gaussian 16 [25]. Density functional theory calculations were performed using the B3LYP hybrid functional [26,27] with the 6-31G(d) basis set [28]. Single-point energy calculations with the keyword `pop=nbo` were employed to extract atomic charges from the NBO partitioning scheme. The SCF convergence threshold was set to 1×10^{-6} atomic units.

2.4.3. DDEC6 Charge Analysis

Density Derived Electrostatic and Chemical (DDEC6) atomic charges were computed using the Chargemol program, which applies an iterative stockholder partitioning of the

electron density [5,6]. The DDEC6 method partitions the total charge density into atom-centered contributions designed to reproduce both chemical and electrostatic properties, and serves as a widely used reference scheme for comparison with machine-learning-derived charges.

2.4.4. Molecular Visualization and Post-Processing

Molecular structures were prepared and visualized using GaussView 6 [29]. Post-processing of VASP outputs, including charge density extraction and format conversion, was performed using VASPKIT [30]. All atomic charges and dipole moments derived from grid-based charge densities were evaluated using the numerical integration procedure described in Section 2.6 and Appendix C.

2.5. Definition of Atomic Charges and Dipole Moments

Within the EAC framework, the total electron density is represented as a sum of atom-centered contributions,

$$\rho(\boldsymbol{r}) = \sum_A \rho_A(\boldsymbol{r}). \tag{2}$$

Once such an atomic decomposition of the charge density is available, atomic charges and dipole moments can be evaluated directly from spatial moments of the corresponding density functions.

2.5.1. Atomic Net Charges

The net atomic charge q_A is obtained from the zeroth-order moment of the atomic charge density,

$$q_A = Z_A - \int \rho_A(\boldsymbol{r})\, d\boldsymbol{r}, \tag{3}$$

where Z_A denotes the nuclear charge and $\rho_A(\boldsymbol{r})$ is the atom-resolved electron density contribution predicted by the model.

By construction, summing over all atoms yields the total charge of the system,

$$\sum_A q_A = \sum_A Z_A - \int \rho(\boldsymbol{r})\, d\boldsymbol{r}, \tag{4}$$

which reduces to zero for neutral systems considered in this work. In practice, the integrals are evaluated numerically on uniform real-space grids using the finite-volume procedure described in Appendix C.1.

2.5.2. Atomic Contributions to the Molecular Dipole Moment

Given a reference point $\boldsymbol{r}_0$ (e.g., the origin used for computing the molecular dipole moment), the molecular dipole moment is computed from the total charge density as

$$\boldsymbol{\mu}_{\text{total}} = -\int (\boldsymbol{r} - \boldsymbol{r}_0)\rho(\boldsymbol{r})\, d\boldsymbol{r} + \sum_A Z_A(\boldsymbol{R}_A - \boldsymbol{r}_0), \tag{5}$$

where $\boldsymbol{R}_A$ denotes the nuclear position.

Using the atom-resolved decomposition of $\rho(\boldsymbol{r})$, the dipole moment can be written as a sum of atomic contributions,

$$\boldsymbol{\mu}_{\text{total}} = \sum_A \boldsymbol{\mu}_A^{\text{global}}, \tag{6}$$

with

$$\boldsymbol{\mu}_A^{\text{global}} = -\int (\boldsymbol{r} - \boldsymbol{r}_0)\rho_A(\boldsymbol{r})\, d\boldsymbol{r} + Z_A(\boldsymbol{R}_A - \boldsymbol{r}_0). \tag{7}$$

This expression follows directly from the additivity of the density representation and ensures that the atomic dipole contributions sum exactly to the total molecular dipole moment.

2.5.3. Local (Nucleus-Referenced) Atomic Dipole Moments

For analysis of intra-atomic polarization, it is often convenient to rewrite the atomic dipole contribution by shifting the reference from the global origin $\boldsymbol{r}_0$ to the atomic nucleus $\boldsymbol{R}_A$. Rearranging the previous expression yields

$$\boldsymbol{\mu}_A^{\text{global}} = -\int (\boldsymbol{r} - \boldsymbol{R}_A)\rho_A(\boldsymbol{r})\, d\boldsymbol{r} + q_A(\boldsymbol{R}_A - \boldsymbol{r}_0). \tag{8}$$

This decomposition separates the atomic dipole contribution into two parts. The second term,

$$\boldsymbol{\mu}_A^{\text{ionic}} = q_A(\boldsymbol{R}_A - \boldsymbol{r}_0), \tag{9}$$

depends only on the net atomic charge and reflects inter-atomic charge transfer with respect to the chosen origin.

The first term,

$$\boldsymbol{\mu}_A^{\text{elec}} = -\int (\boldsymbol{r} - \boldsymbol{R}_A)\rho_A(\boldsymbol{r})\, d\boldsymbol{r}, \tag{10}$$

corresponds to the first spatial moment of the atomic charge density relative to the nucleus. In the following, this quantity is referred to as the atomic local dipole moment.

This form makes explicit that, once the atomic charge density $\rho_A(\boldsymbol{r})$ is specified, both net charges and local dipole moments arise naturally as its zeroth- and first-order spatial moments. Although these quantities depend on the adopted density decomposition scheme, within a unified and internally consistent representation they provide complementary scalar and directional information about atomic-scale electron redistribution.

2.6. Numerical Evaluation of Charges and Dipoles

Atomic charges and dipole moments are evaluated from grid-based charge densities predicted by the EAC model. All integrals are computed numerically on uniform three-dimensional real-space grids consistent with the charge density format used in plane-wave density functional theory calculations.

For zeroth-order moments (atomic charges), direct summation over grid points weighted by the corresponding voxel volume is sufficient to achieve stable and converged results. However, first-order moments such as dipole moments are more sensitive to discretization artifacts. To improve numerical stability and preserve spatial symmetry, a finite-volume interpretation of the grid representation is adopted, in which each grid cell is treated as a voxel and integration is performed using a midpoint quadrature scheme.

In this procedure, charge densities are effectively associated with voxel-centered coordinates rather than grid nodes. This reduces systematic symmetry-breaking contributions that may arise from endpoint quadrature and ensures that dipole components vanish to numerical precision for systems possessing inversion or mirror symmetry.

All integrations are performed consistently for total charge densities as well as for atom-resolved densities $\rho_A(\boldsymbol{r})$. For molecular systems computed using VASP, only valence electron densities are present in the CHGCAR files (VASP Charge Density Output File). Because core electron densities are spherically symmetric and integrate to the corresponding core charges, molecular dipole moments computed from valence densities are equivalent to those obtained from the full electron density representation.

Additional implementation details, including the voxel-averaging scheme, treatment of periodic boundary conditions, are provided in Appendix C.

3. Results

3.1. Charge Density Prediction and Physical Consistency

We first evaluate the ability of the EAC-qm model to reproduce reference density functional theory (DFT) charge densities for molecular systems derived from the QM9 dataset [17]. All reference densities were obtained from plane-wave DFT calculations performed using VASP [22–24], as described in Section 2.4.1.

Figure 3a presents the parity comparison between predicted and reference charge density values evaluated on the independent test set. Each point corresponds to a grid sample from the three-dimensional charge density field. The predicted densities closely follow the diagonal line over the entire density range, indicating excellent agreement with the reference DFT values. The inset shows the distribution of absolute density deviations, which are strongly concentrated near zero. The normalized mean absolute error remains below 1%, demonstrating that the learned representation accurately captures the spatial distribution of valence electron density.

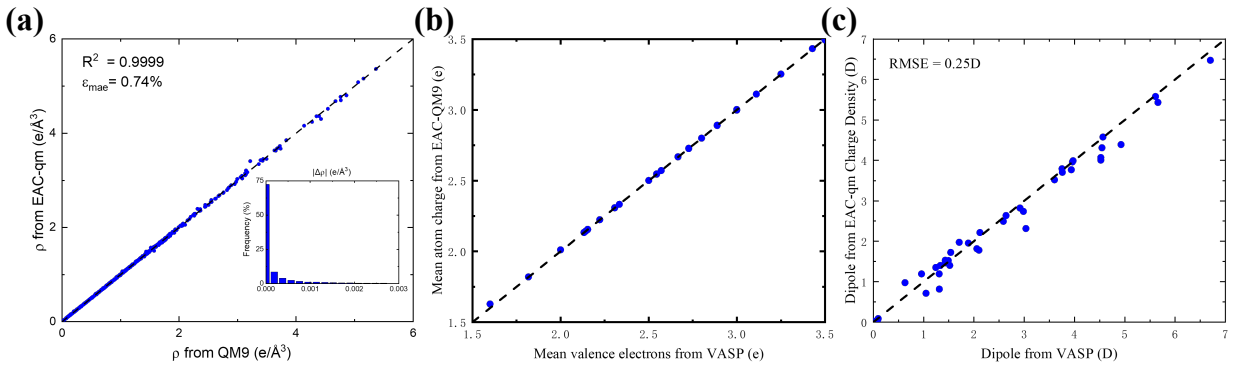

Figure 3. (**a**) Parity comparison between EAC-qm-predicted charge densities and reference DFT values on the QM9 test set. The inset shows the distribution of absolute density deviations. (**b**) Verification of charge conservation: integrated valence electrons obtained from predicted densities versus reference electron counts from pseudopotential files. (**c**) Molecular dipole moment predictions for the QM9 dataset. Dipole moments derived from EAC-qm-predicted densities are plotted against DFT reference values. The diagonal line indicates perfect agreement.

Beyond pointwise accuracy, physically meaningful charge density representations must satisfy global conservation constraints. To assess this, we integrated the predicted charge densities over real space and compared the resulting total valence electron counts with those specified by the pseudopotential files used in the underlying DFT calculations. As shown in Figure 3b, the integrated electron numbers agree closely with the reference values across the entire test set, with deviations remaining at the level of 10^{-3} electrons. Although no explicit charge-conservation constraint was imposed during training, the model preserves total electron count to high numerical precision.

Quantitative analysis in Figure 3c shows that dipole moments predicted by the EAC-qm model agree closely with reference values, achieving a root-mean-square error (RMSE) of 0.23 D, which is comparable to the DDEC6 reference dipole moments (RMSE = 0.25 D) obtained from the Chargemol analysis of the VASP charge densities. This result demonstrates that the charge densities learned by EAC-qm not only accurately reproduce DFT results but also faithfully capture key electrostatic physical properties, establishing a reliable foundation for subsequent atomic-scale dipole moment decomposition based on these charge densities.

To further examine the robustness and out-of-distribution transferability of the learned density representation, we constructed an additional external molecular test set and per-

formed explicit geometry perturbation analyses. The detailed protocol and quantitative results are provided in Appendix B.

Together, these results indicate that the EAC-qm model not only reproduces local charge density values with high fidelity but also maintains global charge conservation. This combination of pointwise accuracy and integral consistency provides a reliable foundation for deriving atom-resolved charges and dipole moments from the predicted continuous electron densities.

3.2. Substituent Effects and Hammett Correlations

To evaluate whether atomic charges derived from the learned charge density representation capture chemically meaningful substituent effects, we analyzed a series of monosubstituted benzene derivatives as model systems. All three charge schemes were evaluated on the same set of molecular geometries (the same substituted benzene structures) to ensure a consistent basis for the Hammett correlation comparison. Substituent effects in aromatic chemistry are commonly characterized using Hammett substituent constants, originally introduced to correlate reaction kinetics and equilibria with electronic properties of substituents [31]. Subsequent compilations have tabulated σ_m and σ_p values for a broad range of functional groups [32,33]. In this work, these literature values are used exclusively as external empirical reference descriptors and are not reparameterized or predicted. The complete list of Hammett constants employed in this study is provided in Appendix A Table A1.

For each substituted benzene, the substituent-induced electronic response of the aromatic ring was quantified by evaluating the change in net atomic charge at the meta and para carbon positions relative to unsubstituted benzene. The charge response Δq is defined as the difference between the atomic net charge of a given ring carbon in the substituted system and that of the corresponding carbon in benzene. This definition provides a direct measure of substituent-induced redistribution of electron density within the aromatic framework.

Figure 4a,b illustrate the spatial distribution of atomic net charges in representative systems. Electron-donating substituents (e.g., -NH_2, -OH) lead to charge accumulation on the benzene ring carbons, particularly at the ortho and para positions, whereas electron-withdrawing substituents (e.g., -NO_2, -COOH) induce charge depletion. These qualitative trends are consistent with classical chemical intuition regarding inductive and resonance effects in substituted aromatics.

The quantitative correlations between calculated charge responses and Hammett constants are shown in Figure 4c–e. Panels (c) and (d) present results obtained from DDEC6 and NBO atomic charges, respectively, while panel (e) shows the corresponding correlations derived from EAC-qm charge densities. In all cases, approximately linear relationships are observed between Δq_m and σ_m, as well as between Δq_p and σ_p. The coefficients of determination indicate that the EAC-qm-derived charge responses exhibit stronger linear association with Hammett constants compared to the DDEC6 and NBO results shown in the same figure.

It is important to emphasize that Hammett constants are empirical parameters derived from specific reaction series and therefore reflect reaction-dependent electronic effects rather than uniquely defined observables [33]. The purpose of the present analysis is not to construct a predictive model for σ values, but rather to assess whether the atomic charges obtained from a unified density decomposition respond systematically to substituent electronic character in a manner aligned with established empirical trends. Within this context, the observed correlations support the interpretability of the density-derived atomic charges as chemically consistent descriptors of substituent-induced electron redistribution.

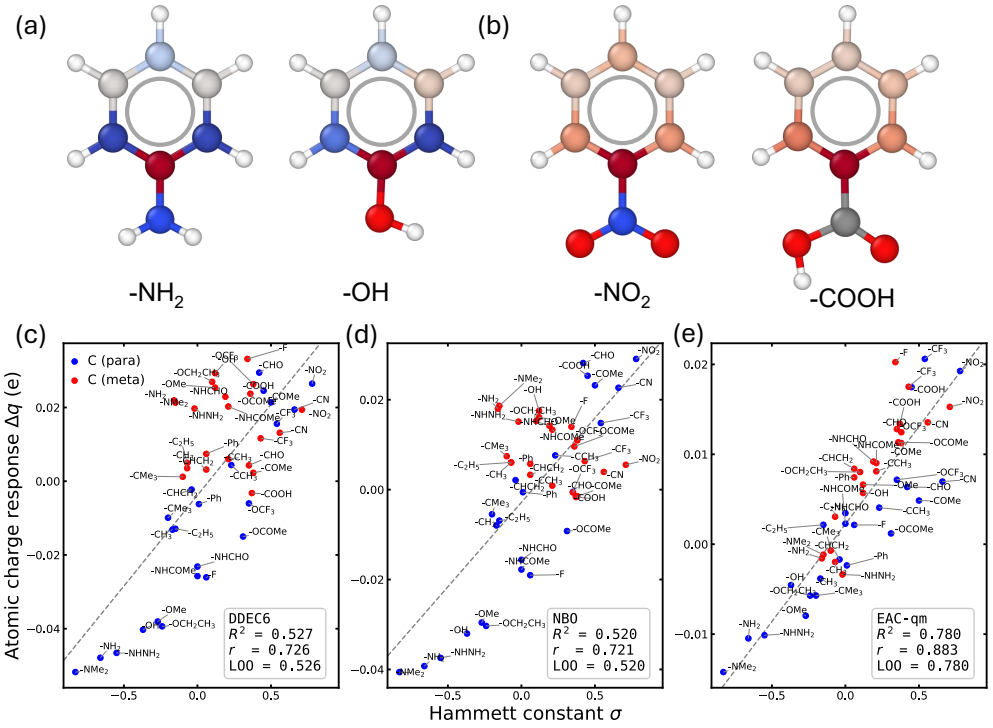

Figure 4. (**a**,**b**) Atomic net charge maps of substituted benzene rings, where blue and red colors indicate electron accumulation and depletion, respectively. Panel (**a**) shows electron-donating substituents (-NH_2, -OH), while panel (**b**) shows electron-withdrawing substituents (-NO_2, -COOH). (**c**–**e**) Correlation of meta- (red) and para- (blue) carbon charge changes with Hammett constants for (**c**) DDEC6, (**d**) NBO, and (**e**) EAC-qm-derived atomic charges.

To objectively compare the statistical performance of the three charge analysis methods (EAC-qm, DDEC6, and NBO), we performed a systematic evaluation of correlation strength, statistical confidence, resampling robustness, and influence diagnostics using a dataset of $n = 46$ substituents. All three methods exhibit statistically significant positive correlations with the Hammett constants ($p < 10^{-8}$), confirming that charge response captures meaningful substituent effects. However, clear quantitative differences emerge among the methods. As summarized in Table 2, EAC-qm achieves a substantially higher Pearson correlation coefficient ($r = 0.883$) and coefficient of determination ($R^2 = 0.780$) than DDEC6 ($R^2 = 0.527$) and NBO ($R^2 = 0.520$), indicating that EAC-qm explains approximately 78% of the variance in Hammett constants, compared to only about 52% for the conventional schemes. Consistent with this, the 95% confidence interval of the EAC-qm correlation is shifted toward higher values ([0.798, 0.934]), with its lower bound approaching the mean correlation level of the traditional methods. In contrast, DDEC6 and NBO exhibit nearly identical confidence intervals, reflecting similar statistical performance and moderate correlation strength.

Table 2. Main statistical metrics of Hammett correlations for three methods.

Method	Pearson *r*	R^2	95% CI of *r*	*p*-Value
EAC-qm	0.883	0.780	[0.798, 0.934]	4.45×10^{-16}
DDEC6	0.726	0.527	[0.552, 0.839]	1.16×10^{-8}
NBO	0.721	0.520	[0.545, 0.836]	1.58×10^{-8}

To assess robustness against sampling variation, we performed both leave-one-out cross-validation and bootstrap resampling (10,000 iterations). The results, summarized in Table 3, demonstrate that EAC-qm exhibits markedly greater stability. Its leave-one-

out R^2 shows minimal variation (standard deviation = 0.009), indicating weak sensitivity to individual substituents. Bootstrap resampling further confirms this robustness: the 95% confidence interval of R^2 for EAC-qm remains within [0.644, 0.878], whereas the corresponding intervals for DDEC6 and NBO extend down to approximately 0.28 and 0.26, respectively. These results indicate that the correlation obtained from EAC-qm is not only stronger but also significantly more resilient to data perturbations.

Table 3. Robustness analysis of Hammett correlations.

Method	Mean LOO R^2	LOO Std	Bootstrap R^2 95% CI
EAC-qm	0.780	0.009	[0.644, 0.878]
DDEC6	0.526	0.017	[0.281, 0.705]
NBO	0.520	0.019	[0.255, 0.717]

Influence diagnostics based on standardized residuals and Cook's distance further support this conclusion. Although all methods exhibit a small number of substituents with residual magnitudes exceeding two standard deviations, leave-one-out analysis shows that the overall conclusions are not dominated by individual data points. In particular, removing any single substituent from the EAC-qm dataset yields R^2 values between 0.754 and 0.810, confirming that the observed correlation is distributed across the dataset rather than driven by isolated influential samples.

Overall, these statistical analyses consistently demonstrate that EAC-qm provides both stronger and more stable correlations with Hammett constants than traditional charge partitioning methods. While all approaches recover statistically significant empirical linear relationships, the improved explanatory power and robustness of EAC-qm indicate that it more faithfully captures the electronic response underlying substituent effects.

3.3. Atomic Dipole Moments

In addition to net atomic charges, the atom-resolved charge densities predicted by the EAC-qm model enable evaluation of atomic local dipole moments as first-order spatial moments relative to the corresponding nuclear positions (Section 2.5). These quantities provide directional information about intra-atomic electron redistribution that is not captured by scalar charges alone.

Figure 5 illustrates atomic local dipole moments for formaldehyde (H_2CO) and formamide ($HCONH_2$). The arrows indicate the direction and magnitude of the atomic local dipole moments, defined as pointing from the centroid of the assigned electron density toward the atomic nucleus. The vectors are overlaid on the valence electron density distribution in the molecular plane.

In both molecules, atoms participating in polar bonds exhibit pronounced local dipole moments aligned with the bond polarization pattern. For example, in formaldehyde, the carbon and oxygen atoms display local dipole orientations consistent with the electron-withdrawing character of the carbonyl group. Similarly, in formamide, anisotropic electron redistribution around the carbonyl and amide nitrogen atoms is reflected in the direction of the corresponding local dipole vectors.

These examples demonstrate that, once a continuous atom-resolved density representation is available, atomic local dipole moments arise naturally as directional descriptors of intra-atomic polarization. While their quantitative values depend on the adopted density decomposition scheme, they provide complementary vectorial information that extends beyond net charge analysis.

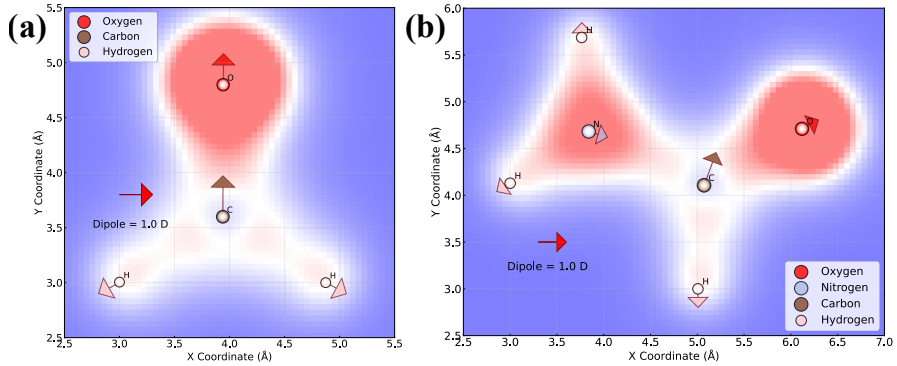

Figure 5. Atomic local dipole moments in (**a**) formaldehyde (H_2CO) and (**b**) formamide ($HCONH_2$) computed from EAC-qm-predicted atomic charge densities. Arrows indicate the direction and magnitude of the atomic local dipole moments, defined as pointing from the centroid of the assigned electron density toward the atomic nucleus, overlaid on the valence electron density distribution in the molecular plane.

For comparison, the same reconstruction procedure is carried out using atomic charges and dipoles obtained from DDEC6 analysis.

The quantitative comparison between reconstructed and reference DFT electrostatic potentials is summarized in Appendix E. Notably, although the EAC-qm model is not explicitly optimized to reproduce the electrostatic potential, the ESP reconstructed from EAC-derived atomic charges and local dipoles exhibits $R^2 = 0.9846$ and RMSE $= 4.92 \times 10^{-2}$ V, comparable to the DDEC6 baseline ($R^2 = 0.9880$, RMSE $= 4.34 \times 10^{-2}$ V).

4. Discussion

4.1. Representation-Level Consistency of Density-Derived Descriptors

Electron density occupies a central role in ground-state quantum chemistry, as formalized by density functional theory [34,35]. In this framework, all ground-state observables are functionals of the electron density. However, atom-resolved quantities such as atomic charges and dipole moments are not uniquely defined observables; rather, they depend on how the total density is partitioned into atomic contributions.

Various partitioning schemes, including stockholder-type approaches [5] and orbital-based analyses [2], provide internally consistent procedures for constructing atomic descriptors. Nevertheless, different schemes may yield quantitatively different atomic charges for the same underlying electron density, reflecting differences in representation rather than inconsistencies in the density itself. From this perspective, the choice of density decomposition plays a defining role in the resulting atomic descriptors.

In the present work, atomic charges and dipole moments are derived from a unified, model-consistent decomposition of the total electron density into continuous atom-centered contributions. Because both scalar and vectorial descriptors are obtained as spatial moments of the same density representation, they share a common origin and are evaluated within a single internally consistent framework. The analysis of substituent effects using Hammett constants illustrates that such density-derived atomic charges respond systematically to established empirical electronic trends.

Importantly, the observed agreement with empirical substituent parameters should not be interpreted as evidence that atomic charges are uniquely determined or directly observable. Rather, it indicates that, within a given representation, atom-resolved descriptors can capture chemically meaningful patterns of electron redistribution. Maintaining representation-level consistency between the density model and the derived atomic properties therefore provides a coherent basis for comparing scalar and higher-order descriptors across related chemical systems.

4.2. Higher-Order Atomic Moments and Directional Polarization

Scalar atomic charges provide a compact measure of net electron redistribution, but they do not encode directional information. In polar bonds and anisotropic chemical environments, electron density rearrangement is inherently vectorial. Within a continuous density representation, such directional effects can be characterized through higher-order spatial moments of the atom-resolved charge densities.

In the present framework, atomic local dipole moments arise directly as first-order moments relative to the corresponding nuclear positions. Because these quantities are evaluated from the same underlying density decomposition used to compute atomic charges, they represent a natural extension of scalar descriptors rather than an independent partitioning scheme. The illustrative examples in Section 3.3 show that local dipole vectors reflect bond polarization patterns and provide additional directional information beyond net charge analysis.

As with atomic charges, the numerical values of atomic dipole moments depend on the adopted density decomposition. Their significance therefore lies not in uniqueness, but in internal consistency within a given representation. Viewed in this context, higher-order atomic moments expand the descriptive space of atom-centered electronic quantities while remaining anchored to the same continuous charge density model.

The result of electrostatic potentials reconstruction demonstrates that the atom-resolved quantities derived from the machine-learned continuous charge density are not only internally consistent by construction, but also capable of reproducing physically meaningful electrostatic observables at a level similar to established partitioning schemes.

For broader context, it is useful to relate the present local dipole moments to traditional distributed multipole analysis (DMA) [36,37]. In classical DMA frameworks, atomic multipoles are obtained from wavefunction-based multipole expansions and subsequent partitioning procedures. In contrast, the local dipoles in the present work arise directly as first-order spatial moments of a continuous, machine-learned atom-resolved charge density decomposition. Rather than introducing multipoles through an external expansion formalism, the descriptors here are intrinsically tied to the same density representation used for atomic charges. This density-consistent construction emphasizes internal coherence across scalar and higher-order quantities within a unified framework.

The machine-learning representation of continuous charge densities thus provides a practical and extensible route for exploring atomic-scale polarization phenomena in complex molecular systems [38]. Further discussion is provided in Appendix D.

4.3. Scope, Limitations, and Future Directions

Several limitations of the present study should be acknowledged. First, the molecular dataset used for model adaptation is restricted to small organic molecules containing H, C, N, O, and F atoms. Although these systems span a diverse range of functional groups within the QM9 chemical space [17], extension to heavier elements or transition-metal-containing compounds would require additional training data and potential architectural refinement. To partially assess transferability beyond the QM9 chemical space, we evaluated the model on 20 larger organic molecules from the PubChem database in Appendix B, including geometrically perturbed configurations, which demonstrate sub-percent-level density accuracy and stable behavior under 0.1 Å distortions. Nevertheless, systematic extension to substantially different chemical compositions remains a direction for future work.

Second, the present analysis focuses exclusively on neutral ground-state structures. Atomic charges and dipole moments derived here therefore reflect equilibrium electron densities obtained from ground-state density functional theory calculations. The behavior

of the model under significant geometric distortion, in charged systems, or in electronically excited states has not been examined and remains an open question.

Third, the predicted charge densities inherit the level of approximation of the reference DFT calculations. Because the training data are generated using the PBE functional within a plane-wave framework, the resulting atomic descriptors are ultimately conditioned on that level of electronic structure theory. Systematic differences associated with alternative exchange-correlation functionals or higher-level wavefunction methods were not investigated in this work.

Finally, although the EAC framework was originally developed for periodic systems [15], the present study concentrates on molecular applications. A systematic evaluation of density-derived atomic descriptors across condensed-phase environments, surfaces, and adsorption systems would be valuable for assessing transferability across different chemical contexts.

Within these boundaries, the results indicate that atom-resolved quantities derived from a unified charge density representation can provide chemically interpretable scalar and vectorial descriptors. Future work may explore extensions to broader chemical spaces, incorporation of diverse electronic structure references, and integration of density-derived descriptors into data-driven modeling of reactivity and materials properties.

5. Conclusions

In this work, we examined atom-resolved electronic descriptors derived from machine-learned continuous charge densities within the EAC framework. The EAC-qm model accurately reproduces reference DFT charge densities for molecular systems and preserves global charge conservation without explicit enforcement. Using Hammett substituent constants as external empirical references, we showed that atomic net charges obtained from the unified density decomposition respond systematically to established electronic trends in substituted aromatic systems.

Because atomic charges and dipole moments are evaluated as successive spatial moments of the same atom-centered density representation, scalar and vectorial descriptors can be interpreted within a single internally consistent framework. While the numerical values of atom-resolved quantities depend on the adopted density decomposition scheme, the present results indicate that a continuous, model-consistent representation provides a coherent basis for analyzing electron redistribution across related chemical systems.

These findings highlight the importance of representation-level consistency when constructing atom-centered electronic descriptors from machine-learned densities and suggest that density-derived moments offer a transferable route for incorporating physically grounded electronic information into data-driven molecular and materials modeling.

Author Contributions: Conceptualization, T.L.; methodology, X.Q.; software, X.Q.; formal analysis, X.Q.; investigation, X.Q.; data curation, X.Q.; writing—original draft preparation, X.Q.; writing—review and editing, X.Q. and T.L.; visualization, X.Q.; supervision, T.L.; project administration, T.L.; funding acquisition, T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (Grant No. 2021YFA0718900), the National Natural Science Foundation of China (Grant Nos. 12374096 and 92477114), and the Jiangsu Funding Program for Excellent Postdoctoral Talent. The APC was funded by the authors.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The EAC-Net framework is publicly available at https://github.com/qin2xue3jian4/EAC-Net (accessed on 1 March 2026). A versioned snapshot of the codebase corresponding to a fixed GitHub commit, together with the trained EAC-mp and EAC-qm model parameters, training configuration files, and metadata describing the molecular systems analyzed in this work, has been deposited in Zenodo with a DOI at https://doi.org/10.5281/zenodo.18382589 (accessed on 1 March 2026). Due to the large size of the original charge-density files and the fact that the manuscript is currently under review, derived quantities such as charge-density grids, DDEC and NBO charges, and post-processed charge-transfer data are not redistributed. These results can be reproduced following the procedures described in the Methods using the provided models, scripts, and publicly available source databases.

Acknowledgments: The authors acknowledge the computational resources provided by the institutional computing platform.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EAC	Equivariant Atomic Contribution
EAC-Net	Equivariant Atomic Contribution Network
EAC-mp	EAC foundation model trained on Materials Project data
EAC-qm	Molecularly fine-tuned EAC model based on QM9
QM9	Quantum Chemistry Structures and Properties Dataset
DFT	Density Functional Theory
VASP	Vienna *Ab initio* Simulation Package
PAW	Projector Augmented-Wave
SCF	Self-Consistent Field
DDEC	Density Derived Electrostatic and Chemical
DDEC6	Sixth-generation DDEC method
NBO	Natural Bond Orbital
RESP	Restrained Electrostatic Potential
DMA	Distributed Multipole Analysis
GDMA	Gaussian Distributed Multipole Analysis
NMAE	Normalized Mean Absolute Error
CHGCAR	VASP Charge Density Output File

Appendix A. Hammett Substituent Constants

Hammett constants were taken from standard experimental compilations (e.g., Hammett; Hansch and Leo) and were not reparameterized in this work.

Table A1. Hammett substituent constants used in this work.

Substituent	σ_m	σ_p	Reference
C_6H_6	0.00	0.00	
C_6H_5–CHO	0.35	0.42	[33]
C_6H_5–CN	0.56	0.66	[39]
C_6H_5–COOH	0.37	0.45	[39]
C_6H_5–NH_2	−0.16	−0.66	[39]
C_6H_5–NO_2	0.71	0.78	[39]
C_6H_5–OH	0.12	−0.37	[39]
C_6H_5–OCH_3	0.12	−0.27	[39]
C_6H_5–C_2H_5	−0.07	−0.15	[39]

Table A1. *Cont.*

Substituent	σ_m	σ_p	Reference
C_6H_5–CF_3	0.43	0.54	[39]
C_6H_5–CH_3	−0.07	−0.17	[39]
C_6H_5–F	0.34	0.06	[39]
C_6H_5–OCH_2CH_3	0.10	−0.24	[40]
C_6H_5–C_6H_5	0.06	0.01	[40]
C_6H_5–$COCH_3$	0.38	0.50	[39]
C_6H_5–$OCOCH_3$	0.36	0.31	[39]
C_6H_5–$NHCOCH_3$	0.21	0.00	[33]
C_6H_5–CMe_3	−0.10	−0.20	[39]
C_6H_5–NMe_2	−0.15	−0.83	[39]
C_6H_5–NHCHO	0.19	0.00	[33]
C_6H_5–$CHCH_2$	0.06	−0.04	[33]
C_6H_5–CCH_3	0.21	0.23	[33]
C_6H_5–$NHNH_2$	−0.02	−0.55	[33]
C_6H_5–OCF_3	0.38	0.35	[33]

Appendix B. External Molecular Systems and Geometry Perturbation Test

To further evaluate the out-of-distribution generalization capability of the EAC-qm model, we constructed an external test set consisting of 20 organic molecules selected from the PubChem database [41] as shown in Table A2. These systems contain 10–20 heavy atoms and thus extend beyond the typical size range represented in the QM9 dataset. For each parent structure, three distorted geometries were generated by applying random atomic displacements with an amplitude of 0.1 Å, while preserving the molecular connectivity. The resulting dataset therefore comprises 20 equilibrium structures and 60 perturbed configurations.

Table A2. List of PubChem molecules used for the external transferability test.

PubChem CID	Molecular Formula	Num of Heavy Atoms
2960132	$C_{13}H_{13}N_5O_2$	20
149961891	$C_{15}H_{17}N$	16
171550903	$C_{13}H_{11}N_7$	20
4601638	$C_{10}H_{13}NO_2$	13
1174125	$C_{14}H_{14}N_2O_3$	19
4341091	$C_9H_{16}O_3$	12
199978	$C_{14}H_{19}N_3O$	18
712460	$C_{15}H_{18}O_2$	17
163414922	$C_{11}H_{20}N_2O_3$	16
1174129	$C_{15}H_{16}N_2O_2$	19
2819582	$C_{13}H_{13}N_3O$	17
14719924	$C_7H_{11}N_3$	10
8496418	$C_{14}H_{19}N_3O_3$	20
12783975	$C_{10}H_{18}O_3$	13
155704401	$C_{17}H_{29}N$	18
149961947	$C_{10}H_{22}O_2$	12
10444290	$C_{10}H_{12}N_2O_5$	17
10801770	$C_{16}H_{13}NO_3$	20
351844	$C_{14}H_{20}O_3$	17
5697085	$C_{15}H_{19}N_3O_2$	20

For all geometries, reference charge densities were obtained using VASP under the same computational settings as described in the Methods section. The corresponding

structures were then provided as input to the EAC-qm model, which directly predicts the continuous three-dimensional charge density without access to the reference density. To enable a pointwise comparison, both densities were evaluated on an identical $100 \times 100 \times 100$ real-space grid. A parity plot between VASP and EAC-qm densities is shown in Figure A1.

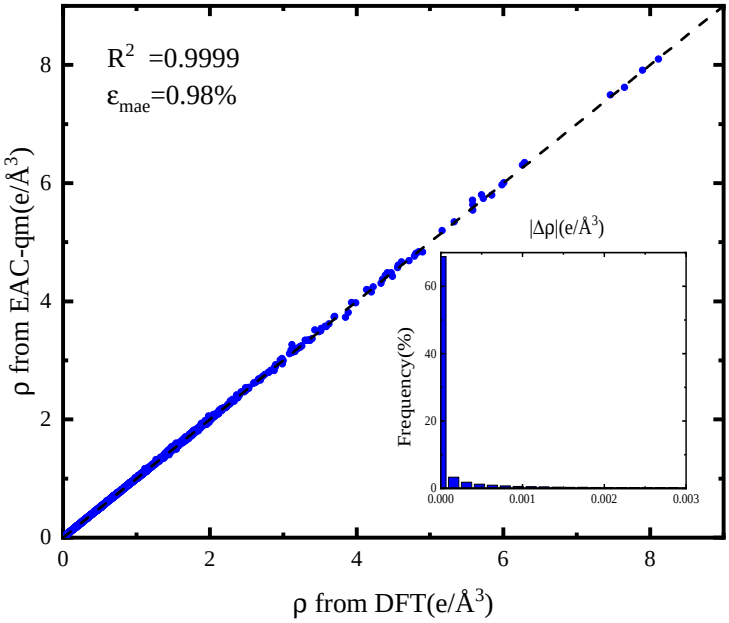

Figure A1. Parity comparison between EAC-qm-predicted charge densities and reference DFT values on the test set from PubChem. The inset shows the distribution of absolute density deviations.

As illustrated in Figure A1, the predicted densities exhibit near-perfect agreement with the first-principles reference values. Linear regression analysis yields a coefficient of determination of $R^2 = 0.9999$, indicating an almost exact linear correspondence between model predictions and DFT results. The relative mean absolute error is $\varepsilon_{\text{mae}} = 0.98\%$, demonstrating that the model maintains sub-percent-level accuracy even for molecular systems not included in the fine-tuning dataset.

Importantly, this test set probes two distinct aspects of transferability. First, the selected molecules differ in size and structural motifs from the QM9 training subset, providing an assessment of out-of-distribution chemical generalization. Second, the explicitly introduced geometric perturbations examine the stability of the learned density representation under small deviations from equilibrium. The absence of systematic bias or error amplification under 0.1 Å distortions indicates that the EAC-qm model exhibits smooth and physically consistent behavior in the vicinity of equilibrium configurations.

The inset histogram in Figure A1 further shows that the vast majority of grid-point errors are tightly concentrated near zero, with rapidly decaying tails. This distribution suggests that the residual discrepancies arise primarily from small, spatially localized fluctuations rather than structured or region-specific distortions. Overall, these results demonstrate that EAC-qm can robustly reconstruct first-principles charge densities for previously unseen molecular systems and mildly distorted geometries, supporting the transferability and stability of the learned atom-resolved density representation.

Appendix C. Computing Details

Appendix C.1. Theoretical Justification for Computing Dipole Moments from CHGCAR Files

The molecular dipole moment is defined as:

$$\boldsymbol{\mu}^{\text{total}} = -\int (\boldsymbol{r} - \boldsymbol{r}_0)\rho(\boldsymbol{r})\, d\boldsymbol{r} + \sum_A Z_A(\boldsymbol{R}_A - \boldsymbol{r}_0)$$

For EAC-qm models trained on CHGCAR files, the charge density contains only valence electrons. The full expression including core electrons is:

$$\boldsymbol{\mu}^{\text{total}} = -\int (\boldsymbol{r} - \boldsymbol{r}_0)[\rho_{\text{val}}(\boldsymbol{r}) + \rho_{\text{core}}(\boldsymbol{r})]\, d\boldsymbol{r} + \sum_A [Z_A + Z_{A,\text{core}}](\boldsymbol{R}_A - \boldsymbol{r}_0)$$

Rearranging this expression yields:

$$\boldsymbol{\mu}^{\text{total}} = \boldsymbol{\mu}^{\text{total}}_{\text{val}} + \sum_A \left[-\int (\boldsymbol{r} - \boldsymbol{R}_A)\rho_{\text{core}}(\boldsymbol{r})\, d\boldsymbol{r} + (\boldsymbol{R}_A - \boldsymbol{r}_0)\left(Z_{A,\text{core}} - \int \rho_{\text{core}}(\boldsymbol{r})\, d\boldsymbol{r} \right) \right]$$

Due to the spherically symmetric distribution of core electrons and the fact that their integrated charge equals $Z_{A,\text{core}}$, the latter two terms vanish. Therefore, $\boldsymbol{\mu}^{\text{total}} = \boldsymbol{\mu}^{\text{total}}_{\text{val}}$, and molecular dipole moments can be computed directly from CHGCAR files containing only valence electron densities.

Appendix C.2. Numerical Evaluation of Charges and Dipoles from Grid-Based Charge Densities

In this work, atomic charges and atomic dipole moments are evaluated from machine-learned charge densities represented on uniform real-space grids. Particular care is taken to ensure numerical consistency and symmetry preservation when computing both zeroth- and first-order moments of the charge density.

The charge density predicted by the EAC model is defined on a regular three-dimensional grid with dimensions (N_x, N_y, N_z), consistent with the grid convention used in plane-wave density functional theory codes such as VASP. The grid values correspond to samples of the charge density field at uniformly spaced points within a periodic unit cell, where the sampling points are located at fractional coordinates

$$\left(\frac{i}{N_x}, \frac{j}{N_y}, \frac{k}{N_z}\right), i = 0, 1, \cdots, N_x - 1, j = 0, \cdots, N_y - 1, k = 0, \cdots, N_z - 1$$

and periodic boundary conditions are implicitly assumed.

While direct summation over these grid points is sufficient for evaluating integrated quantities such as the total charge, the computation of dipole moments involves first-order spatial moments of the charge density and is therefore more sensitive to discretization artifacts. In particular, using grid-point coordinates directly corresponds to a left-endpoint quadrature rule, which can lead to small but systematic symmetry-breaking contributions to the dipole moment for systems that are exactly symmetric in the continuum limit.

To mitigate this issue, we adopt a finite-volume interpretation of the grid-based charge density. Each grid point is treated as a vertex of a cubic voxel, and the charge density associated with a voxel is approximated by the average of the charge densities at its eight corner points. Specifically, the voxel-averaged charge density is defined as

$$\bar{\rho}_{i,j,k} = \frac{1}{8} \sum_{\alpha,\beta,\gamma \in \{0,1\}} \rho_{i+\alpha, j+\beta, k+\gamma}$$

where all indices are taken modulo (N_x, N_y, N_z) to enforce periodic boundary conditions. This procedure effectively converts the node-based sampling into a cell-centered representation of the charge density.

The corresponding spatial coordinates are taken to be the centers of the voxels, located at fractional coordinates

$$\left(\frac{i+1/2}{N_x}, \frac{j+1/2}{N_y}, \frac{k+1/2}{N_z}\right),$$

which are subsequently mapped to real-space positions using the lattice vectors of the simulation cell. With this representation, numerical integration of both the charge and dipole moments is carried out using a midpoint quadrature rule over the voxel volumes.

This discretization scheme preserves charge conservation and significantly improves the numerical stability of dipole moment calculations. In particular, for systems possessing inversion or mirror symmetries, the corresponding components of the dipole moment vanish to numerical precision, consistent with the symmetry of the underlying charge density. The same procedure is applied consistently to the evaluation of molecular dipole moments as well as atomic charges and atomic (global and local) dipole moments derived from the atom-decomposed charge densities.

Appendix D. Physical Significance of Atomic Local Dipole Moments

To elucidate the physical significance of atomic local dipole moments beyond net atomic charges, we analyze the in-plane electric field distribution of $HCONH_2$ constructed from different atom-resolved electrostatic representations (Figure A2). When only atomic charges are included (Figure A2a), each atom acts effectively as a point charge, and the local electric field around each atom is largely radial in character, with field lines either converging into or diverging from each atom depending on the sign of the charge. In this charge-only description, the directional structure of the local field is entirely dictated by the molecular geometry and the spatial arrangement of point charges.

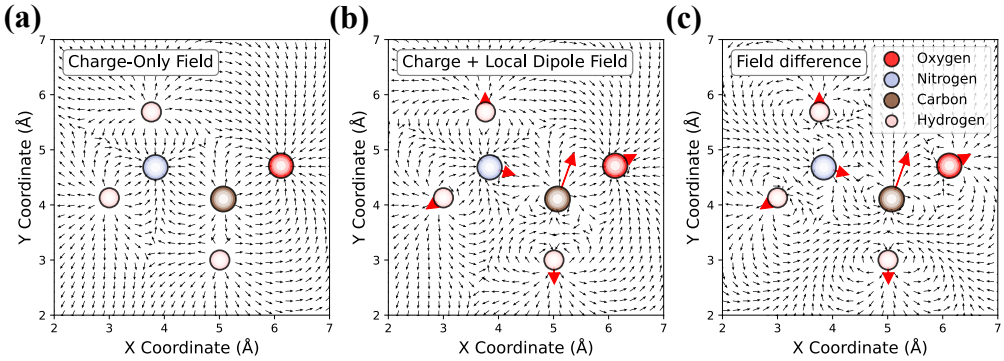

Figure A2. In-plane electric field distributions of $HCONH_2$ evaluated on the molecular plane of $HCONH_2$. All fields are shown as vector plots on the same plane for comparison. (**a**) Electric field constructed from atomic charges predicted by EAC-QM only. (**b**) Electric field constructed from atomic charges and atomic local dipole moments predicted by EAC-QM, with red arrows indicating the directions of the local dipole moments. (**c**) Vector difference between the electric fields in (**b**) and (**a**), isolating the contribution from atomic local dipole moments.

Upon incorporating atomic local dipole moments predicted by EAC-QM (Figure A2b), the electric field undergoes a pronounced qualitative reorganization in the vicinity of atomic centers. In contrast to the charge-only case, the field lines near each atom no longer display purely radial symmetry but instead show clear directional features aligned with the corresponding local dipole moments (indicated by red arrows). This behavior reflects intra-atomic polarization effects that cannot be captured by point charges alone and introduces

an additional directional degree of freedom associated with intra-atomic polarization into the local electrostatic environment.

The difference between the two field representations, shown in Figure A2c, isolates the contribution arising from atomic local dipole moments. The resulting differential field is strongly localized around individual atoms and exhibits distinct directional patterns that directly follow the orientations of the local dipoles. Importantly, this contribution does not simply rescale the magnitude of the electric field but selectively redistributes its direction in specific spatial regions. These results demonstrate that atomic local dipole moments encode additional, anisotropic information about the local electrostatic environment, complementing net atomic charges and providing a more refined description of direction-dependent polarization effects relevant to intermolecular interactions and external-field responses.

These results highlight an alternative analytical perspective enabled by the present framework: how electron density is redistributed anisotropically within individual atoms after net atomic charges have been defined. While atomic-level dipolar information has long been accessible through multipole-based approaches such as Distributed Multipole Analysis, the key distinction here is that atomic local dipole moments are derived as the first spatial moments of explicitly learned, atom-resolved charge densities within a unified and self-consistent representation of the total electron density. As a result, such intra-atomic polarization effects can be analyzed systematically and consistently across different chemical environments, going beyond descriptions based solely on atomic net charges or point-charge approximations. The machine-learning representation of continuous charge densities thus provides a practical and extensible route for exploring atomic-scale polarization phenomena in complex molecular systems [38].

Appendix E. Electrostatic Potential Reconstruction

To assess the practical electrostatic relevance of atomic charges and local dipole moments derived from the EAC-qm model, we perform a direct reconstruction of the molecular electrostatic potential (ESP) for $HCONH_2$ and compare it with the reference DFT electrostatic potential obtained from VASP LOCPOT files.

The reference ESP is taken directly from the LOCPOT output of the DFT calculation. To avoid numerical instabilities and unphysical divergence of the point-multipole expansion near atomic nuclei, all grid points within 1.7 Å of any atom are excluded from the analysis. This cutoff corresponds to the largest van der Waals radius among the constituent elements (carbon), ensuring a conservative exclusion region.

The ESP is reconstructed using two atom-centered multipole models:

$$V(\boldsymbol{r}) = \sum_A \left[\frac{q_A}{|\boldsymbol{r} - \boldsymbol{R}_A|} + \frac{\boldsymbol{\mu}_A \cdot (\boldsymbol{r} - \boldsymbol{R}_A)}{|\boldsymbol{r} - \boldsymbol{R}_A|^3} \right] \tag{A1}$$

Figure A3 shows scatter plots comparing reconstructed electrostatic potentials with the reference DFT values. The left panel corresponds to DDEC6-derived charges and dipoles, while the right panel corresponds to EAC-qm-derived quantities. For DDEC6, the reconstructed ESP yields $R^2 = 0.9880$ and RMSE $= 4.34 \times 10^{-2}$ V. For EAC-qm, the corresponding values are $R^2 = 0.9846$ and RMSE $= 4.92 \times 10^{-2}$ V.

Although the EAC-qm result is slightly inferior to DDEC6 in this specific metric, both methods achieve the same order of accuracy, and the difference in RMSE remains small. Importantly, the EAC-qm model is not explicitly trained to reproduce electrostatic potentials; its atomic charges and local dipoles arise from a unified decomposition of the learned continuous charge density. The comparable ESP reconstruction accuracy therefore indicates that the machine-learned atomic multipole descriptors retain physically meaningful electrostatic information.

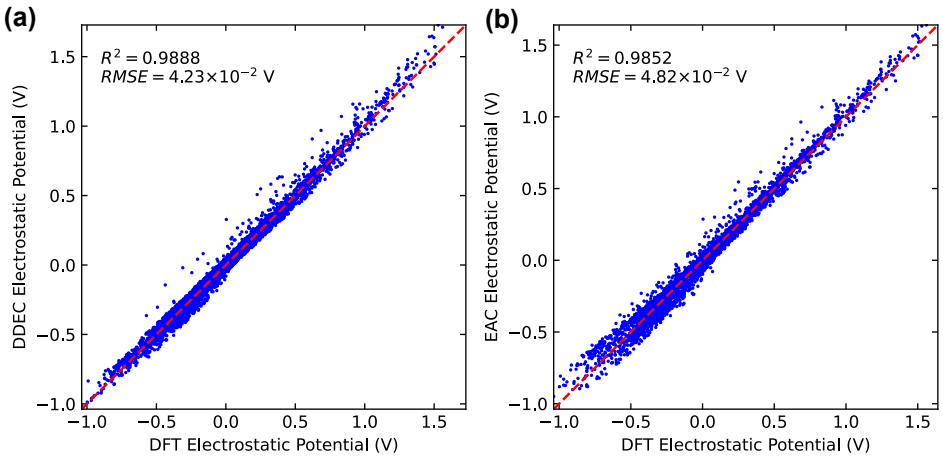

Figure A3. Comparison between reconstructed and DFT electrostatic potentials for $HCONH_2$. (**a**) DDEC6-derived atomic charges and dipoles. (**b**) EAC-qm-derived atomic charges and local dipoles. The reference DFT electrostatic potential is taken from the VASP LOCPOT file computed with `LVHAR = .TRUE.`, corresponding to the electronic Hartree potential only. Grid points within 1.7 Å of any atom are excluded. The dashed red line indicates perfect agreement ($y = x$).

References

1. Mulliken, R.S. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833–1840. [CrossRef]
2. Reed, A.E.; Weinstock, R.B.; Weinhold, F. Natural Population Analysis. *J. Chem. Phys.* **1985**, *83*, 735–746. [CrossRef]
3. Heidar-Zadeh, F.; Ayers, P.W.; Verstraelen, T.; Vinogradov, I.; Vöhringer-Martinez, E.; Bultinck, P. Information-Theoretic Approaches to Atoms-in-Molecules: Hirshfeld Family of Partitioning Schemes. *J. Phys. Chem. A* **2018**, *122*, 4219–4245. [CrossRef] [PubMed]
4. Van Damme, S.; Bultinck, P.; Fias, S. Electrostatic Potentials from Self-Consistent Hirshfeld Atomic Charges. *J. Chem. Theory Comput.* **2009**, *5*, 334–340. [CrossRef]
5. Manz, T.A.; Limas, N.G. Introducing DDEC6 Atomic Population Analysis: Part 1. Charge Partitioning Theory and Methodology. *RSC Adv.* **2016**, *6*, 47771–47801. [CrossRef]
6. Limas, N.G.; Manz, T.A. Introducing DDEC6 Atomic Population Analysis: Part 2. Computed Results for a Wide Range of Periodic and Nonperiodic Materials. *RSC Adv.* **2016**, *6*, 45727–45747. [CrossRef]
7. Bayly, C.I.; Cieplak, P.; Cornell, W.; Kollman, P.A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280. [CrossRef]
8. Zhao, G.; Chung, Y.G. PACMAN: A robust partial atomic charge predicter for nanoporous materials based on crystal graph convolution networks. *J. Chem. Theory Comput.* **2024**, *20*, 5368–5380. [CrossRef] [PubMed]
9. Unke, O.T.; Meuwly, M. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693. [CrossRef]
10. Heid, E.; Greenman, K.P.; Chung, Y.; Li, S.C.; Graff, D.E.; Vermeire, F.H.; Wu, H.; Green, W.H.; McGill, C.J. Chemprop: A machine learning package for chemical property prediction. *J. Chem. Inf. Model.* **2023**, *64*, 9–17. [CrossRef]
11. Schütt, K.T.; Sauceda, H.E.; Kindermans, P.J.; Tkatchenko, A.; Müller, K.R. SchNet—A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148*, 241722. [CrossRef] [PubMed]
12. Schütt, K.T.; Kessel, P.; Gastegger, M.; Nicoli, K.A.; Tkatchenko, A.; Müller, K.R. SchNetPack: A Deep Learning Toolbox For Atomistic Systems. *J. Chem. Theory Comput.* **2019**, *15*, 448–455. [CrossRef]
13. Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M.E.; Burke, K.; Müller, K.R. Bypassing the Kohn-Sham Equations with Machine Learning. *Nat. Commun.* **2017**, *8*, 872. [CrossRef] [PubMed]
14. Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F.R.; Miller, T.F. OrbNet: Deep Learning for Quantum Chemistry Using Symmetry-Adapted Atomic-Orbital Features. *J. Chem. Phys.* **2020**, *153*, 124111. [CrossRef]
15. Qin, X.; Lv, T.; Zhong, Z. EAC-Net: Predicting Real-Space Charge Density via Equivariant Atomic Contributions. *arXiv* **2025**, arXiv:2508.04052. [CrossRef]
16. Jain, A.; Ong, S.P.; Hautier, G.; Chen, W.; Richards, W.D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1*, 011002. [CrossRef]

17. Ramakrishnan, R.; Dral, P.O.; Rupp, M.; Von Lilienfeld, O.A. Quantum Chemistry Structures and Properties of 134 Kilo Molecules. *Sci. Data* **2014**, *1*, 140022. [CrossRef]
18. Jørgensen, P.B.; Bhowmik, A. Equivariant Graph Neural Networks for Fast Electron Density Estimation of Molecules, Liquids, and Solids. *npj Comput. Mater.* **2022**, *8*, 183. [CrossRef]
19. Geiger, M.; Smidt, T. e3nn: Euclidean Neural Networks. *arXiv* **2022**, arXiv:2207.09453. [CrossRef]
20. 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. *Nat. Commun.* **2022**, *13*, 2453. [CrossRef] [PubMed]
21. Qin, X.; Lv, T.; Zhong, Z. EAC-Qm: Code and Models for Atomic Charges and Local Dipoles from Machine-Learned Charge Densities. *Zenodo* **2026**. [CrossRef]
22. Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for *Ab Initio* Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54*, 11169–11186. [CrossRef]
23. Blöchl, P.E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50*, 17953–17979. [CrossRef] [PubMed]
24. Perdew, J.P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868. [CrossRef]
25. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. *Gaussian 16 Revision C.01*; Gaussian, Inc.: Wallingford, CT, USA, 2016.
26. Becke, A.D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652. [CrossRef]
27. Lee, C.; Yang, W.; Parr, R.G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789. [CrossRef] [PubMed]
28. Hehre, W.J.; Ditchfield, R.; Pople, J.A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261. [CrossRef]
29. Dennington, R.; Keith, T.A.; Millam, J.M. *GaussView Version 6*; Semichem Inc.: Shawnee Mission, KS, USA, 2016.
30. Wang, V.; Xu, N.; Liu, J.C.; Tang, G.; Geng, W.T. VASPKIT: A User-Friendly Interface Facilitating High-Throughput Computing and Analysis Using VASP Code. *Comput. Phys. Commun.* **2021**, *267*, 108033. [CrossRef]
31. Hammett, L.P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96–103. [CrossRef]
32. Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; A Wiley-Interscience Publication, Wiley: New York, NY, USA, 1979.
33. Hansch, C.; Leo, A.; Taft, R.W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91*, 165–195. [CrossRef]
34. Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871. [CrossRef]
35. Kohn, W.; Sham, L.J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138. [CrossRef]
36. Stone, A. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, UK, 2013. [CrossRef]
37. Stone, A.J. Distributed Multipole Analysis: Stability for Large Basis Sets. *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132. [CrossRef] [PubMed]
38. Willatt, M.J.; Musil, F.; Ceriotti, M. Atom-Density Representations for Machine Learning. *J. Chem. Phys.* **2019**, *150*, 154110. [CrossRef] [PubMed]
39. Yadav, V.K. Hammett Substituent Constants. In *Steric and Stereoelectronic Effects in Organic Chemistry*; Springer International Publishing: Cham, Switzerland, 2021; pp. 179–189. [CrossRef]
40. Chemist, W. *Hammett Sigma Constants*. 2025. Available online: https://www.wiredchemist.com/data/hammett-sigma-constants (accessed on 11 March 2026).
41. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem in 2021: New Data Content and Improved Web Interfaces. *Nucleic Acids Res.* **2021**, *49*, D1388–D1395. [CrossRef] [PubMed]