PNAS INAUGURAL ARTICLE | CHEMISTRY

Cartesian equivariant representations for learning and understanding molecular orbitals

Daniel S. King[a,1,2], Daniel Grzenda[b], Ray Zhu[c], Nathaniel Hudson[b,d], Ian Foster[b,d], Bingqing Cheng[e,f,g,1], and Laura Gagliardi[a,c,h,1]

Affiliations are included on p. 9.

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2021.
Contributed by Laura Gagliardi; received April 30, 2025; accepted October 17, 2025; reviewed by Clémence Corminboeuf, Seyed Mohamad Moosavi, and Veronique Van Speybroeck

Qualitative and quantitative orbital properties such as bonding/antibonding character, localization, and orbital energies are critical to how chemists understand reactivity, catalysis, and excited-state behavior. Despite this, representations of orbitals in deep learning models have been very underdeveloped relative to representations of molecular geometries and Hamiltonians. Here, we apply state-of-the-art equivariant deep learning architectures to the task of assigning global labels to orbitals, namely energies characterizations, given the molecular coefficients from Hartree–Fock or density functional theory. The architecture we have developed, the Cartesian Equivariant Orbital Network (CEONET), shows how molecular orbital coefficients are readily featurized as equivariant node features common to all graph-based machine-learned potentials. We find that CEONET performs well at predicting difficult quantitative labels such as the orbital energy and orbital entropy. Furthermore, we find that the CEONET representation provides an intuitive latent space for differentiating orbital character for the qualitative assignment of e.g. bonding or antibonding character. In addition to providing a useful representation for further integrating deep learning with electronic structure theory, we expect CEONET to be useful for automatizing and interpreting the results of advanced electronic structure methods such as complete active space self-consistent field theory. In particular, the ability of CEONET to infer multireference character via the orbital entropy paves the way toward the machine-learned selection of active spaces.

machine learning | electronic structure | molecular orbitals | chemical reactions

Significance

Orbital properties such as energies and bonding character are vital to how chemists understand fundamental chemical phenomena such as bonding, Lewis structures, electronegativity, and excited states. Yet, relatively little effort has gone into developing deep learning representations of molecular orbitals. This research presents a deep learning model, the Cartesian Equivariant Orbital Network (CEONET), that improves how molecular orbitals are represented and analyzed in machine learning frameworks. By working with the symmetries inherent to molecular orbital coefficients, CEONET accurately predicts key orbital properties, addressing a significant gap in the application of machine learning to electronic structure theory and enabling the automated application and interpretation of advanced electronic structure calculations.

Since the dawn of computational chemistry in the 1930s (1), molecular orbitals have served as the key bridge between the chemist's intuitive understanding of how molecules behave (i.e., via bonding and antibonding interactions) and the computational basis for computing molecular properties (i.e., via Slater determinants). To this day, molecular orbitals remain vital to how chemists are taught and perceive fundamental chemical phenomenon such as bonding, Lewis structures, oxidation states, and electronegativity (2, 3). Additionally, properties of molecular orbitals often relate directly to experimental observables such as ionization potentials, electron affinities, and excited states (4–12).

Moreover, even given the past century of electronic structure development, molecular orbitals still remain practically vital to computations. To this day, molecular orbitals continue to serve as the fundamental basis for computing excited states and electron kinetic energy, even in density functional theory (13–15). Beyond this, molecular orbitals provide the basis for computing strong correlation in wave function methods such as complete active space self-consistent field theory (CASSCF) (16). These methods leverage the fact that strong correlation in molecules is generally localized to a small region of the molecule, and use qualitative orbital labels (e.g., π, π^*, bonding, antibonding) to select an "active space" of orbitals in which to compute the correlation (17). Automating the selection and interpretation of these orbitals has been a long-standing goal of the field (17–24), and would enable the application of these methods at scale (25, 26). In particular, the estimation of the multireference character of individual orbitals prior to calculation remains a key problem that remains to be addressed.

However, despite the important role that molecular orbitals play in chemistry, relatively little work has been done to design machine-learned representations of molecular orbitals compared to geometries and Hamiltonians (28–32). Nevertheless, accuracy has often improved in deep learning approaches through the development of "physics informed" architectures (33) that integrate ideas from quantum chemistry. For example,

[1]To whom correspondence may be addressed. Email: king1305@berkeley.edu, bingqingcheng@berkeley.edu, or lgagliardi@uchicago.edu.

[2]Present address: Bakar Institute of Digital Materials for the Planet, University of California, Berkeley, CA 94720.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2510235122/-/DCSupplemental.

Published November 21, 2025.

Behler outlines four "generations" of machine-learned potentials (32, 34) with later generations [e.g., ANI-1 (35) and PhysNet (36)] incorporating the concept of partial charges from standard force field development to capture long-range effects. Similarly, state-of-the-art approaches such as SchnOrb (37), OrbNet (38, 39), DeepH (40), and the recent model of Ceriotti and coworkers (41) make explicit reference to the concepts of orbitals and Hamiltonians, being reminiscent of semiempirical models such as AM1 (42) and PM3 (43). As the development of machine-learned potentials progresses further, we expect that representations of molecular orbitals themselves may play a key role.

Here, we present the Cartesian equivariant orbital network CEONET, a work focused on assigning labels to orbitals themselves given input from ab initio calculations. Specifically, given a set of molecular orbital coefficients for a single orbital $\mathbf{c}$ on atoms with positions $\mathbf{x}$, element types $\mathbf{z}$, and basis χ we pursue a labeling function

$$f(\mathbf{c}, \mathbf{x}, \mathbf{z}, \chi) \rightarrow l_\phi, \quad [1]$$

where l_ϕ is a label corresponding to either a quantitative value (e.g., orbital energy or orbital entropy) or a qualitative description (e.g., bonding or antibonding). We develop new datasets for both of these tasks: QM9@HF-STO-3G, consisting of STO-3G Hartree–Fock orbitals for the entire QM9 dataset (44); the TMQM@HF-STO3G dataset, consisting of STO-3G Hartree–Fock orbitals for complexes in the TMQM dataset (45); and the TMBonding dataset, consisting of metal, ligand, bonding, and antibonding orbitals for octahedral transition metal complexes (46). Furthermore, we employ our scheme on the recently published Def2SVP orbitals in the QH9 dataset (QM9@B3LYP-Def2SVP) (47). Additionally, we apply CEONET to the difficult problem of predicting orbital entropies from high-throughput multireference calculations (25). The orbital entropy measures the occupational diffusivity of orbitals in the CI expansion of the wave function, and as such can be used to choose an appropriate active space in an a priori fashion (22, 48).

Across these learning tasks, we find that CEONET performs well at predicting difficult quantitative labels such as the orbital energy and orbital entropy, effectively modeling the action of the Fock operator in real-space. Learning curves on all these tasks demonstrate excellent scaling laws with respect to the number of training data. Additionally, we find that the CEONet model provides an intuitive latent space for capturing orbital character and generalizes well to unseen orbital types (e.g. localized orbitals or orbitals from smaller basis sets). In particular, the ability of CEONET to infer orbital entropies from SA-CASSCF calculations given only molecular orbital coefficients provides a promising route for developing automated routes for active space selection.

Furthermore, the CEONET architecture provides several technical contributions to the representation of orbitals in molecular systems. Mainly, CEONET a) demonstrates the direct mapping of molecular orbital coefficients and basis information to symmetric hidden features in graph neural networks, b) includes expressive message passing layers that overcome the orbital sign problem, and c) demonstrates the utility of Cartesian symmetry functions in representing the orbital character. We hope that this work provides a solid foundation for considering the properties of orbitals in deep learning architectures.

1. Model Architecture

The CEONET model provides a synthesis of many concepts from quantum chemistry packages and state-of-the-art equivariant machine-learned potentials. In the following section, we describe the background used to interpret the molecular orbital coefficients given their input from quantum chemistry packages (Section 1.1). We then describe the background of the Cartesian tensor product networks used in the model (Section 1.2) followed by a complete description of the model architecture (Section 1.3).

1.1. Atomic and Molecular Orbitals in Quantum Chemistry. In standard quantum chemistry codes, molecular orbitals are represented as linear combinations of a set of k atomic orbitals centered on each atom i:

$$\phi(\mathbf{r}) = \sum_{\mathbf{l}ik} c_{ik}^{\mathbf{l}} \chi_{ik}^{\mathbf{l}}(\mathbf{r}), \quad [2]$$

where each atomic orbital $\chi_{ik}^{\mathbf{l}}(\mathbf{r})$ is a linear combination of so-called "primitive" Gaussian-type orbitals with decay coefficients α, multiplied by angular functions of order $l = ||\mathbf{l}||_1$ ($\mathbf{l} = (l_x, l_y, l_z)$) centered on $\mathbf{r}_i$:

$$\chi_{ik}^{\mathbf{l}}(\mathbf{r}) = \sum_p \Lambda_{pk}^{(Z_i)} N(\alpha_{pk}^{(Z_i)}, \mathbf{l}) e^{-\alpha_{pk}^{(Z_i)} \mathbf{r}^2} L_{\mathbf{l}}(\mathbf{r} - \mathbf{r}_i) \quad [3]$$

$$L_{\mathbf{l}}(\mathbf{r}) = (r_x - r_{ix})^{l_x} (r_y - r_{iy})^{l_y} (r_z - r_{iz})^{l_z} \quad [4]$$

We use $\Lambda_{pk}^{(Z_i)}$ and $\alpha^{(Z_i)}$ in the equation above to emphasize that these weights are fixed by the atomic orbital basis for each element and not choices of the user; $N(\alpha_{pk}^{(Z_i)}, \mathbf{l})$ is a normalization coefficient determined for each primitive as a function of α and $\mathbf{l}$. For example, in the minimal STO-3G basis, each "Slater-type" atomic orbital is approximated by three Gaussian-type orbitals (i.e., $e^{-\alpha \mathbf{r}^2}$) functions.

However, symmetry constraints demand that $\Lambda_{pa}^{(Z_i)}$ and $\alpha_{pk}^{(Z_i)}$ be shared between all atomic orbitals $\chi_{ik}^{\mathbf{l}}(\mathbf{r})$ of order l. For example, the $2p$ primitive coefficients in the STO-3G basis are shared between the three $l = 1$ (p_x, p_y, p_z) basis functions. Similarly, the $3d$ coefficients are shared between the six (x^2, y^2, z^2, xy, xz, yz) $l = 2$ basis functions. Constructing the basis in this manner makes quantum chemical calculations invariant to rotations or translations of the molecule.

Furthermore, this construction results in sets of molecular orbital coefficients on each atom that transform equivariantly as tensors of rank l:

$$c_{ik}^0 = c_{ik}^{(0,0,0)} \quad \text{(s-type)} \quad [5]$$

$$c_{ik}^1 = (c_{ik}^{(1,0,0)}, c_{ik}^{(0,1,0)}, c_{ik}^{(0,0,1)}) \quad \text{(p-type)} \quad [6]$$

$$c_{ik}^2 = \begin{bmatrix} c_{ik}^{(2,0,0)} & c_{ik}^{(1,1,0)} & c_{ik}^{(1,0,1)} \\ c_{ik}^{(1,1,0)} & c_{ik}^{(0,2,0)} & c_{ik}^{(0,1,1)} \\ c_{ik}^{(1,0,1)} & c_{ik}^{(0,1,1)} & c_{ik}^{(0,0,2)} \end{bmatrix} \quad \text{(d-type)} \quad [7]$$

For example, rotation or inversion of the molecule by a matrix R results in an equivalent rotation or inversion of all c_{ik}^1 coefficients. Equivalently, one may work with the spherical harmonic angular functions in which case the coefficients are naturally treated as spherical tensors of rank l (i.e., five d-type orbitals instead of six due to the constraint of constant $x^2 + y^2 + z^2$). Both formulations are equivalent and many quantum chemistry libraries can easily switch between the two types of functions. Although Gaussian-type orbitals are the most

popular choice, we note that the proposed scheme (Eqs. **5**–**7**) can easily be extended to other atom-centered basis functions such as Slater-type or numeric orbitals by modification of Eq. **3**.

1.2. Background: Cartesian Tensor Product Networks. Machine-learned force fields such as ACE (49), MACE (50), and NequIP (51) have proven to be highly effective tools for modeling the energies and forces of quantum mechanical computations. The key feature of all of these methods is a nonlinear equivariant layer that preserves the rank of different tensor components. In particular, given two tensors of rank l_i and l_r, there exists a contraction rule

$$ {}_{l_i,l_r}\mathbf{X}^{l_o} = (\mathbf{X}^{l_i}) \cdot\cdot (\mathbf{X}^{l_r}) \qquad [8] $$

in which $\cdot\cdot$ sums and multiplies over the elements of both tensors to achieve a tensor of rank l_o. These layers were first achieved employing spherical tensors (i.e., obeying the symmetries of the spherical harmonics) and combining tensors together through use of the Clebsch–Gordan coefficients $C^{l_o m_o}_{l_r m_r l_i m_i}$. More recently, several models such as CACE (52), HotPP (53), and TensorNet (54) have shown that equivalent formulations of these layers can be made employing Cartesian tensors. For example, instead of expanding normalized pair vectors $\hat{\mathbf{r}}_{ij}$ via spherical harmonic filters, Cartesian tensors of rank l are derived via repeated tensor products:

$$ \hat{\mathbf{r}}^{\otimes n} = \hat{\mathbf{r}} \otimes \hat{\mathbf{r}} \ldots \otimes \hat{\mathbf{r}} \qquad [9] $$

with n factors of $\hat{\mathbf{r}}$, and $\hat{\mathbf{r}}^{\otimes 0} = 1$. Equivariance is then achieved via the tensor contraction rule:

$$ {}_{l_i,l_r}\mathbf{X}^{l_o}_{a_1 \ldots a_l} = \mathbf{X}^{l_i}_{a_1 \ldots a_{l_i-c} b_1 \ldots b_c} \mathbf{X}^{l_r}_{b_1 \ldots b_c a_1 \ldots a_{l_r-c}} \qquad c \le \min(l_i, l_r) \qquad l_i + l_r - 2c = l_o \qquad [10] $$

in which a number of c Cartesian dimensions are contracted over to produce a tensor of rank $l_0 = l_i + l_r - 2c$. So, for example, given two tensors of rank $l = 1$, one may produce tensors of rank $l = 0$ (scalar, $c = 1$) or rank $l = 2$ (matrix, $c = 0$). The contraction rule in Eq. **10** can easily be shown to be equivariant (53). We also note that parity symmetry is properly handled with no extra effort by this contraction rule, with all tensors of odd rank l possessing odd parity symmetry and all tensors of even rank l possessing even parity symmetry.

Finally, there is the additional question of how to combine contractions of the same output rank l_o resulting from different input ranks l_i and l_r. One option is to simply sum over all outputs of rank l_o to produce the final result:

$$ \mathbf{X}^{l_o}_{kjc} = \mathbf{X}^{l_i}_{kc} \otimes_{\text{sum}} \mathbf{X}^{l_r}_{jc} = \sum_{l_i,l_r \to l_o} (\mathbf{X}^{l_i}_{ic}) \cdot\cdot (\mathbf{X}^{l_r}_{jc}) \qquad [11] $$

However, a more expressive option is to stack all output channels together:

$$ {}_{l_i,l_r}\mathbf{X}^{l_o}_{kjc} = \mathbf{X}^{l}_{kc} \otimes_{\text{stack}} \mathbf{X}^{l}_{jc} = \bigoplus_{l_i,l_r \to l_o} (\mathbf{X}^{l_i}_{ic}) \cdot\cdot (\mathbf{X}^{l_r}_{jc}) \qquad [12] $$

resulting in effectively a larger representation of $c \times (l_i, l_r \to l_o)$ in each l_o channel. Unless otherwise stated, we have opted to use Eq. **12** in each tensor product step, which we have found to be more cost-efficient when paired with a smaller channel dimension and fewer network layers.

As seen, repeated applications of Eqs. **11** or **12** inevitably result in rapidly expanding tensors of rank l with dimension 3^l. Thus, all equivariant models define some maximum rank l_{max} which constrains the output of Eq. **11** from nonlinear layers. Here, we use $l_{\text{max}} = 2$.

1.3. Model Architecture.

1.3.1. Orbital featurization. As shown in Section 1.1, molecular orbitals are readily featurized as equivariant node features, Eqs. **5**–**7**, common to all graph-based machine-learned potentials. However, there is the central problem of how to make such a featurization transferrable between basis sets, in which each basis has varying numbers of "channels" k for each angular momentum $\mathbf{l}$, each with different sets of primitive (p) coefficients $\Lambda^{(Z_i)}_{pk}$ and $\alpha^{(Z_i)}_{pk}$. Here, we address this issue by sampling the primitive orbitals directly with a learnable basis r_c and summing over all channels k:

$$ \boldsymbol{\phi}^{\mathbf{l}}_{ic} = \sum_{pk} c^{\mathbf{l}}_{ik} \Lambda^{(Z_i)}_{pk} N(\alpha^{(Z_i)}_{pk}, \mathbf{l}) e^{-\alpha^{(Z_i)}_{pk} r_c^2} \qquad [13] $$

Featurization in this manner is transferable between basis sets of any size and captures the radial shape of the molecular orbital around any atom i and any direction $\mathbf{l}$; the model only learns how to sample the molecular orbitals around atoms i via r_c using the underlying primitive basis. We then pass these coefficients through a linear layer to expand to the channel dimension n_c. The features of each rank l are then stacked together in tensors of rank l to form the initial hidden features $\mathbf{h}^l_{ic}$. This layer and all following layers are outlined in Fig. 1.

1.3.2. Message passing layer. Although the featurization of Eq. **13** is efficient in capturing the variation of the molecular orbital on each node, one must additionally account for symmetry with respect to the global sign of the orbitals. In particular, the molecular orbital energy is invariant with respect to the change $c^{\mathbf{l}}_{ik} \to -c^{\mathbf{l}}_{ik}$ (i.e., is even with respect to orbital parity), while the featurization $\boldsymbol{\phi}^{\mathbf{l}}_{ic}$ is odd with respect to this transformation. This orbital parity can be removed by taking tensor products between two sets of orbital features and the vector between them. In particular, we take the tensor product

$$ \mathbf{h}^l_{ic} \otimes^l \hat{\mathbf{r}}^{\otimes l}_{ij} \otimes^l \mathbf{h}^l_{jc} \qquad [14] $$

to compute messages along each edge, where $\otimes^l$ represents the $\otimes_{\text{stack}}$ operation in Eq. **12**, only keeping outputs up to $l = l_{\text{max}}$. We additionally take the tensor product

$$ \mathbf{h}^l_{ic} \otimes^0 \mathbf{h}^l_{jc} \qquad [15] $$

in which we take the tensor product directly between the hidden features on node j and the hidden features on node i. Here, again $\otimes^0$ represents the $\otimes_{\text{stack}}$ operation in Eq. **12** in which we only collect tensor products resulting in scalar features ($l = 0$). These scalar features are then appended with the scalar outputs of Eq. **14** and fed into a SiLU-activated multilayer perceptron (MLP) (55) to compute attention weights a_{ijc} for each message channel:

$$ \mathbf{m}^l_{1,ic} = \sum_j a_{ijc} \mathbf{m}^l_{j \to i,c} \qquad [16] $$

This message is then combined with a self-interaction message computed from

$$ \mathbf{m}^l_{2,ic} = \mathbf{h}^l_{ic} \otimes^l \mathbf{h}^l_{ic} \qquad [17] $$

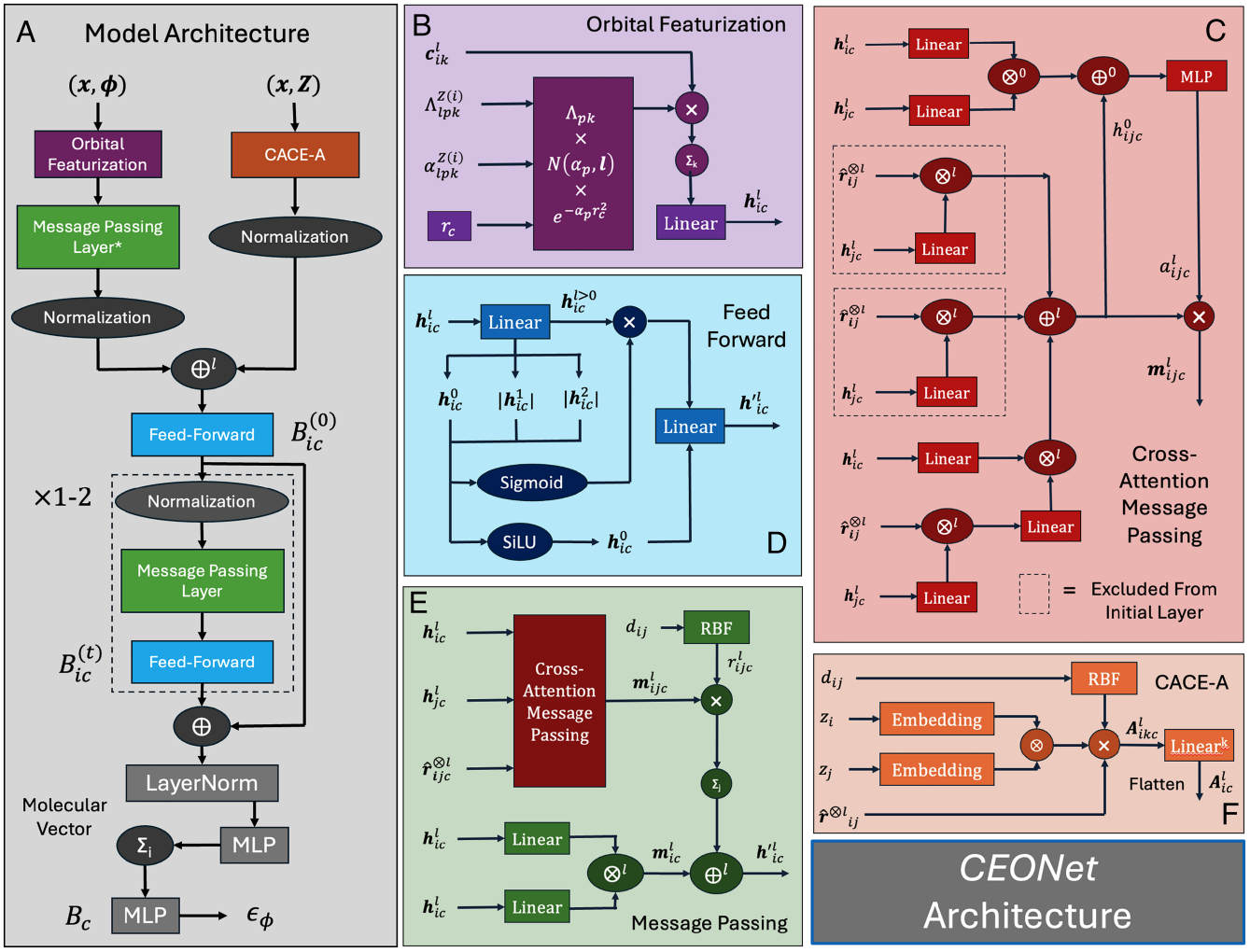

Fig. 1. A schematic view of the CEONET architecture. In each block, lighter shades and rectangular edges represent modules with learnable parameters while darker shades and round edges present nonlearnable steps. Colors other than the background color depict modules outlined in other blocks. (*A*) Model Architecture. Geometry and orbital features (x, ϕ) are fed through an orbital featurization layer and then passed through a message passing layer to remove orbital parity. These features are then appended with the A basis from the Cartesian atomic cluster expansion (CACE) representation and fed through a feed-forward layer to form the $\mathbf{h}_{ic}^{l}$ for the first B output $B_{ic}^{(0)}$. The $\mathbf{h}_{ic}^{l}$ are then fed through four message passing steps to produce $B_{ic}^{(t)}$ and appended then fed through a multilayer perceptron (MLP) and summed over to produce a molecular representation B_c, which is used to predict the orbital energy ϵ_ϕ. (*B*) Orbital Featurization. Primitive basis functions are computed and sampled with learnable r_c which are then multiplied by their molecular coefficients, summed over, and passed through linear mixing to produce the orbital-parity-dependent output $\mathbf{h}_{ic}^{l}$. (*C*) Message Passing. The edge features $\mathbf{h}_{ic}^{l}$, $\mathbf{h}_{jc}^{l}$, and $\hat{r}_{ij}^{\otimes l}$ are passed through an attention message passing layer to produce messages which are multiplied by radial basis functions and summed over to produce accumulated messages m_{ic}^{l}. These are appended with self-messages produced from the tensor product between hidden features on the same node, as well as with the original hidden features (except in the first orbital layer). (*D*) Feed Forward. First, the hidden features are projected down to the channel dimension by a linear layer. Then, the norms of equivariant layers $h_{ic}^{l>0}$ are computed and combined with $\mathbf{h}_{ic}^{0}$ and fed into a sigmoid-activated MLP to compute multipliers on the input $h_{ic}^{l>0}$ features. A new set of $\mathbf{h}_{ic}^{\prime 0}$ features are also computed with a SiLU-activated MLP and then fed into a linear layer with the $h_{ic}^{l>0}$ features to produce the output. (*E*) Cross-Attention Message Passing. Three types of messages are computed via tensor products between the node features $\mathbf{h}_{ic}^{l}$, $\mathbf{h}_{jc}^{l}$, and $\hat{r}_{ij}^{\otimes l}$, which are then combined. The tensor product $\mathbf{h}_{ic}^{l} \otimes \mathbf{h}_{jc}^{l}$ is then combined with the $l = 0$ features of the incoming messages and fed through an MLP to calculate attention for each message channel. (*F*) CACE-A. Edge features are formed between each node by element embedding and multiplying by radial basis functions and the vector between the nodes $\hat{r}_{ij}^{\otimes l}$. These edge features are then multiplied by the radial basis functions and summed over to compute the A basis, which is then linearly mixed over the radial channel and flattened to provide the output $\mathbf{A}_{ic}^{l}$.

to form the new hidden node representation $\mathbf{h}_{ic}^{\prime l}$. This large-channel representation is then generally projected back down to the channel dimension n_c by a linear layer in the next module.

This message passing module is used in later layers once the orbital representations are combined with features from CACE (explained in the next section). Since orbital parity is not a factor in these later layers, we additionally compute the edge features

$$\mathbf{h}_{ic}^{l} \otimes^{l} \hat{\mathbf{r}}_{ij}^{\otimes l} \quad [18]$$

$$\hat{\mathbf{r}}_{ij}^{\otimes l} \otimes^{l} \mathbf{h}_{jc}^{l} \quad [19]$$

to append with the tensors computed in Eq. **14**. Additionally, linear layers are used prior to all tensor products to improve model expressiveness. In summary, this layer allows for information to be shared between atoms about the character of the molecular orbital in a symmetry-aware fashion.

1.3.3. CACE features. To compute features of the molecular geometry (**x**,**z**), we take the A basis generated from the first steps Cartesian atomic cluster expansion (CACE) method (52). Sender and receiver nodes are embedded with one-hot encoding and their tensor product is taken to encode the edge. These edge features are then multiplied by radial basis functions and the

angular momentum components of $\hat{\mathbf{r}}_{ij}^{\otimes l}$:

$$\chi^{l}_{ijkc} = (\theta_{Z(i)} \otimes \theta_{Z(j)})_c R_k(d_{ij}) \hat{\mathbf{r}}_{ij}^{\otimes l} \quad [20]$$

and then projected onto the nodes, linearly mixed, and flattened to form the A basis used in CEONet:

$$A^{l}_{ic} = \bigoplus_{k'} \sum_{kj} W_{k'k} \chi^{l}_{ijkc} \quad [21]$$

We note that the χ in Eqs. **20** and **21** represent the edge features used in CACE and not the basis information of Eq. **3**. We refer readers to the original CACE paper (52) for a more detailed explanation. However, we note that several such GNN representations of molecular geometries exist, and reviews are available beyond the scope of this work (32). Here, CACE is used as it is a recently developed Cartesian model, but in practice CEONet can be combined with any equivariant GNN. In total, this layer computes features of the molecular geometry to be combined with features of the molecular orbital in the feed-forward layer.

1.3.4. Tensor normalization. Inspired by TensorNet (54), we normalize all node features over the tensor dimension with the rule

$$\mathbf{h}'^{l}_{ic} = \frac{\mathbf{h}^{l}_{ic}}{\sum_{\mathbf{l}} (h^{\mathbf{l}}_{ic})^2 + 1} \quad [22]$$

which helps to control the scale of the features after message passing and improve stability.

1.3.5. Feed-forward layer. Inspired by the norm gate of QH-Net (56) and the feed-forward layer of Equiformer (57), we aim to compute new scalar features and scaling multipliers of the equivariant channels. Starting with an input $\mathbf{h}^{l}_{ic}$ projected down to the channel dimension n_c, we take as input the scalar features $\mathbf{h}^{0}_{ic}$ and the norms of the nonscalar features $|\mathbf{h}^{l}_{ic}| = \sum_{\mathbf{l}} (h^{\mathbf{l}}_{ic})^2$. These are fed into a SiLU-activated layer to compute new scalar features $\mathbf{h}^{0}_{ic}$ and a sigmoid-activated layer to compute scalings of the $\mathbf{h}^{(l>0)}_{ic}$ features. The scalar and nonscalar features are then fed to a final linear layer to produce the new representation $\mathbf{h}'^{l}_{ic}$. This layer is critical for combining information about the molecular geometry and the molecular orbital coefficients.

1.3.6. Formation of scalar B basis. As the orbital energy is invariant to SE(3) group actions (i.e., rotations and translations), we ultimately want to predict the orbital energy from invariant inputs. Here, we adapt the n-body B scalar basis used in CACE (52) to featurize the orbitals for orbital energy prediction. This approach amounts to taking all unique tensor products over up to $\nu_{\max}$ vector representations of the same node that result in scalar features:

$$B_{\nu=1,ic} = h^{0}_{ic} \quad [23]$$

$$B_{\nu=2,ic} = \bigoplus_{l_i \neq 0, l_r \neq 0 \to l_o = 0} (\mathbf{h}^{l_i}_{ic}) \cdot \cdot (\mathbf{h}^{l_r}_{ic}) \quad [24]$$

$$B_{\nu=3,ic} = \bigoplus_{l_a \neq 0, l_b \neq 0, l_c \neq 0 \to l_o = 0} (\mathbf{h}^{l_a}_{ic}) \cdot \cdot (\mathbf{h}^{l_b}_{ic}) \cdot \cdot (\mathbf{h}^{l_c}_{ic}) \quad [25]$$

For example, for $l_{\max} = 2$ and $\nu_{\max} = 2$, the B_2 features consist of the tensor products $(1, 1)$ and $(2, 2)$, while the B_3 features consist of the tensor product $(1, 2, 1)$. Symmetrization in this fashion ensures that the predicted energy remains invariant with respect to SE(3) operators.

1.3.7. Model architecture and readout. We finally review the model architecture in full, as outlined in Fig. 1. First, the orbital input is featurized and passed through a message passing layer to remove orbital parity dependence. These features are then normalized and appended to the CACE features to form a unified hidden representation of both the orbital and element features. This large hidden representation is then projected down to the channel dimension c by a feed-forward layer, which is then used to form the initial B representation $B^{(0)}_{ic}$.

The unified hidden representation is then fed through n_{layer} information-passing steps consisting of message passing and feed-forward layers; the hidden representation after each step is used to form a new B featurization $B^{(t)}_{ic}$. The $B^{(t)}_{ic}$ features from all steps are then appended, normalized, fed into an MLP, and summed over all nodes to produce a molecular representation B_c. This molecular representation is then passed to a final MLP to predict the orbital energy ϵ_ϕ.

In the modules described above, we use $l_{\max} = 2$, $\nu_{\max} = 3$, $n_{\text{layer}} = 2$, a channel dimension of $n_c = 16$, a radial basis of $n_{\text{rbf}} = 16$ centered Gaussian basis functions, and 16 r_c samples of the molecular orbitals in Eq. **2**. These settings were chosen as they generally demonstrated a good balance between cost and performance across validation sets for different tasks. However, it is possible that better parameters can be found depending on the application. Following previous work, the CACE featurization employs an embedding size of 4, a radial basis of 8 centered Gaussian basis functions, and a radial channel dimension of 12. All edges are determined with a cutoff of 7.6 Bohr and radial basis functions are multiplied by polynomial cutoffs to ensure smooth transitions inside and outside of the cutoff radius. Models are trained with the Adam optimizer (58) using an initial learning rate of 10^{-3} over batches of 128 orbitals. Models employed on transition metals use $n_{\text{layer}} = 1$ and batch sizes of 32 due to memory constraints.

In summary, input into the model consists of a) atomic elements and xyz coordinates, b) basis data (i.e., contraction coefficients and exponents on each atom), and c) molecular orbital coefficients. Details concerning reading in and preprocessing the data from common formats such as molden files are discussed in *SI Appendix*.

2. Data

2.1. Orbital Energy Datasets. The QM9 dataset consists of about 134k equilibrium geometries of organic compounds (consisting of C, H, O, N, F). To form the QM9@HF-STO-3G dataset, we have carried out restricted Hartree–Fock (RHF) calculations in the minimal STO-3G basis in PySCF (59). Although STO-3G is not used in practical computations, it provides a good initial benchmark for learning orbital labels and is absent of high-energy virtual orbitals. In all, this dataset consists of roughly five million valence orbitals. We have chosen to ignore the core orbitals (defined as having orbital energy less than −1.75 Ha) as they cause orbital learning to be unnecessarily difficult due to their unique characteristics. However, valence orbitals are fundamentally more important for understanding chemical properties and core orbitals can generally be easily identified absent any machine learning techniques.

A similar dataset to QM9@HF-STO-3G is the recently released QH9 dataset, which contains the Fock matrices for all molecules in QM9, computed using the Def2SVP basis and the B3LYP functional (47). We have been able to diagonalize all Fock matrices in QM9 to obtain B3LYP Def2SVP orbitals for

all QM9 molecules. Due to the larger basis, this results in a larger dataset of over 13 million orbitals (the majority of which are nonvalence virtual orbitals absent in STO-3G). We refer to the two datasets as "QM9@HF-STO-3G" and "QM9@B3LYP-Def2SVP" throughout the paper. However, the sizes of these datasets are many times more than large enough to understand the behavior of the CEONET model. As such, while we have made the full content of these datasets publicly available, here we only train on datasets composed of orbitals from a small fraction (10,000) of the molecules from each dataset. We then estimate the performance of CEONET trained on the entire dataset by extrapolating model performance out to the estimated number of orbitals in each dataset (i.e., by multiplying the data from 10,000 orbitals by the appropriate scaling factor).

Energies are shifted and scaled by the mean and SD of the orbital energies prior to fitting. Additionally, since mean-field methods generally present a sharp modal change in orbital energy between occupied and virtual orbitals, we have chosen to train models separately on the energies of occupied and virtual orbitals. This is also somewhat necessary, as orbital energy is not a rigorous function of orbital shape in mean-field due to symmetry breaking (a good example is in molecular O_2, where restricted Hartree–Fock will only occupy one of the π^* antibonding orbitals, causing that orbital to have arbitrarily lower energy).

2.2. Orbital Entropy Dataset. A key problem in the use of multireference approaches remains the selection of an appropriate active space: choosing the space of electrons and orbitals in which to perform the CI calculation (17–24, 60). Multireference diagnostics for single orbitals such as orbital occupation numbers are the starting point for all quantitative active space selection schemes (18, 20). Here, we focus on predicting the single-orbital entropy, which measures the occupational distribution of a single orbital in the CI expansion of the wave function (48).

In particular, we aim to predict ground state orbital entropies from a set of high-throughput CASSCF calculations performed on the QUESTDB database (61) of 542 small-molecule vertical excitation energies (25) in the aug-cc-pVTZ (62) basis set. This dataset consists of SA-CASSCF calculations using the automated APC active space method (23). To eliminate poor active spaces, we remove any systems from the dataset for which the lowest-energy excitation has larger than 0.55 eV error using hybrid multiconfigurational pair-density functional theory [MC-PDFT (63), using the tPBE0 functional (64)]. We then take the active space orbitals of each of the ground states of these SA-CASSCF calculations to form a dataset of 1,871 aug-cc-pVTZ orbitals with orbital entropy labels.

2.3. Transition Metal Complex Datasets. To test the generalizability of our model to orbitals of systems containing transition metals, we have developed the TMOrb dataset, which consists of minimal-basis STO-3G RHF Hartree–Fock calculations for a subset of 42k complexes in the extensive TMQM dataset of transition metal complexes (45). To avoid the problem of global charge, we further select a subset of these complexes that are neutral in character, resulting in a dataset of about five million valence orbitals from about 35k complexes. These data allow us to test the difficulty of learning orbital energies from transition metal complexes. We refer to it throughout the paper as "TMQM@HF-STO-3G."

Another goal of this work is to learn qualitative labels from orbital data such as bonding or antibonding character. Learning these labels requires either labeling orbitals by hand to generate a dataset, or requires mathematically constructing orbitals that represent a defined label. Here, we generate a process for constructing orbitals that represent four types of orbitals found in transition metal calculations: ligand orbitals, metal orbitals, ligand-metal bonding orbitals, and ligand-metal antibonding orbitals. Given an initial mean-field calculation with orbitals ϕ, we carry out the following steps:

1) Project the restricted open-shell Hartree–Fock (ROHF) (65) Fock matrix F separately into the space of metal valence orbitals and ligand (i.e., nonmetal) valence orbitals, then diagonalize the Fock matrix in these subspaces to form the canonical metal valence orbitals $\{\phi_M\}$ and ligand valence orbitals $\{\phi_L\}$. These orbitals are given the labels "metal valence" and "ligand valence," respectively.
2) Looping over the space of $\{\phi_M\}$ and $\{\phi_L\}$, project the Fock operator into the space of one ligand orbital ϕ_L and one metal orbital ϕ_M, then diagonalize within this space. If there is a stabilization of the bonding orbital energy greater than 0.05 Hartree with respect to the initial orbitals, the canonical orbitals within this space are added to the dataset and given the labels "metal-ligand bonding" and "metal-ligand antibonding," respectively.

We have carried out this process on the large dataset of 4,865 first-row transition metal octahedral complexes from Kulik and coworkers (46). Calculations were undertaken with ROHF in a modified version of the "minao" basis in PySCF, generated by taking the first contracted functions from cc-pVTZ (62, 66, 67), and excluding the addition of the 4s orbital from calculations. Converged calculations resulted in a final dataset of 3,702 first-row transition metal complexes. However, there are naturally many more ligand orbitals than metal-involved orbitals. To achieve a balanced dataset, we take the complete set of all bonding and antibonding orbitals from each complex with up to an equal number of ligand and metal orbitals, resulting in a final balanced "TMConstructed" training dataset of 32,378 orbitals.

3. Results

3.1. Predicting Orbital Energies. The performance of CEONET in predicting occupied and virtual orbital energies from across different basis sets (STO-3G vs. Def2SVP), electronic structure methods (HF vs. B3LYP), and molecular geometries (QM9 vs. TMQM) is shown in Fig. 2. All models are trained on a dataset of 10^5 orbitals and inferred on a test set of 5,000 orbitals. As is seen, CEONET is able to learn orbital energies well, with predictions on most datasets approaching chemical accuracy ($\approx$0.043 eV or 1.6 mHa). This learning occurs despite the subtle difference in the interpretation of orbital energies between HF and DFT (68).

The learning curves of CEONET applied to different datasets and inferred on validation sets of 5,000 orbitals is shown in Fig. 3 with the data summarized in Table 1. In line with studies on Hamiltonian learning (47), we find that model performance improves steeply as a function of the number of data points. To estimate the performance of CEONET when trained on the entire dataset, we extrapolate the mean absolute error to the estimated total dataset size using the fits in Fig. 3. These extrapolated validation accuracies are shown in Table 1, all of which reach chemical accuracy. In addition, all models achieve an R^2 close to 1 (Fig. 2), and the larger MAEs of the virtual orbitals is in part due to their very high energies.

Nevertheless, a key observation is that all sets of virtual orbitals, particularly those from large basis sets (one of which

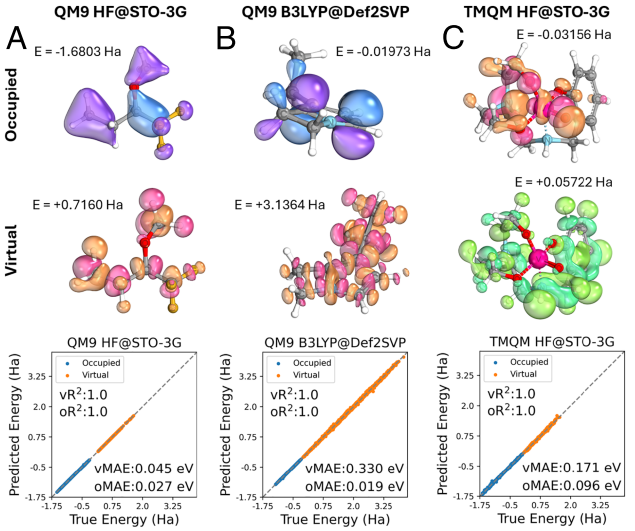

Fig. 2. Performance of CEONET on predicting test set orbital energies across different sets of functionals (HF vs. B3LYP), systems (transition metals vs. organic), and basis sets (STO-3G vs. Def2SVP). (*A*) Performance on QM9 HF@STO-3G orbitals. (*B*) Performance on QM9 B3LYP@Def2SVP orbitals. (*C*) Performance on TMQM HF@STO-3G orbitals. Examples of orbitals from each category are shown above each plot. The color pairs of orbitals are arbitrary.

is shown in Fig. 2*B*) pose a greater challenge for CEONET, with less steep learning curves in all cases. We suspect that the fundamental difficulty of this task comes from the large variation of these orbitals in real-space. Because the CEONET model interprets the orbital coefficients fundamentally through real-space featurization (Eq. **13**, which is agnostic to the label being learned), orbitals that vary dramatically in space naturally strain the model.

By modeling the energies of molecular orbitals in real space, CEONET more-or-less learns to approximate the real-space action of the Fock matrix F (69):

$$\epsilon_i = \int_{\mathbf{r}} \int_{\mathbf{r}'} \phi_i(\mathbf{r}) F(\mathbf{r}, \mathbf{r}') \phi_i(\mathbf{r}') \qquad [26]$$

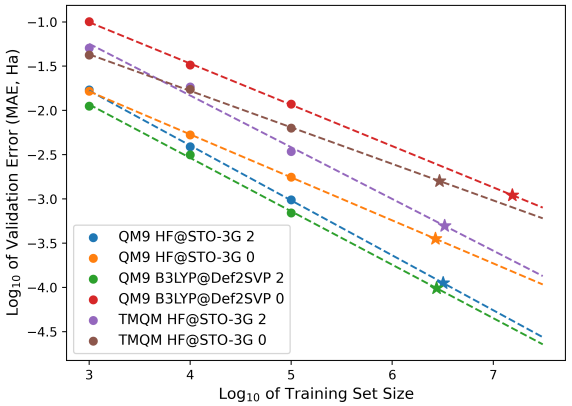

Fig. 3. Learning curves of CEONET on different orbital energy learning tasks. Learning tasks are differentiated by the molecular dataset (QM9 vs. TMQM), method (HF vs. B3LYP), and orbital occupation (0 or 2). The base-10 log of the validation MAE in Ha is plotted against the base-10 log of the number of training points in the training data. Fits of these learning curves (shown by the dotted lines, the slope of which is α_{10}) are then extrapolated to the estimated full size of the different datasets, as shown by the stars on each plot.

Table 1. Validation set mean absolute errors (mHa) of CEONET on different orbital energy learning tasks, trained with different numbers of training data

Data	Occ	10^3	10^4	10^5	Extrap.	α_{10}
QM9 HF	2	17.0	3.9	1.0	0.1	−0.62
@ STO-3G	0	16.5	5.3	1.8	0.4	−0.49
QM9 B3LYP	2	11.2	3.2	0.7	0.1	−0.60
@Def2SVP	0	101.0	32.7	11.8	1.1	−0.47
TMQM HF	2	50.6	18.5	3.4	0.5	−0.58
@STO-3G	0	42.2	17.3	6.3	1.6	−0.41

The final two columns show the MAE extrapolated to the estimated total dataset size and the slope of the base-10 learning curve α_{10} given the fits shown in Fig. 3.

Because the complexity of this mapping scales with the variation of orbitals in space, these orbitals are naturally more difficult to learn.

A benefit of this behavior is that CEONET easily generalizes to unseen orbitals types. Fig. 4 shows the performance of CEONET on different localized subspaces of the occupied Hartree–Fock valence orbitals in benzene [with a symmetric structure taken from QUESTDB (70)]. Specifically, we test performance on projected orbitals from Def2SVP calculations (Fig. 4*B*), Boys-localized orbitals (71) (Fig. 4*C*), and intrinsic bonding orbitals (IBOs) (3) (Fig. 4*D*) as implemented in PySCF. The model generalizes easily to the projected orbitals from Def2SVP, and although energy estimates of the localized orbitals are not quantitative, the model clearly has an understanding of which orbitals are higher in energy, with $R^2 > 0.97$ in both cases. We also find that CEONET develops a nearly quantitative understanding of the energies of orbitals from smaller basis sets (e.g., estimating Def2SVP energies of 6-31g(d) orbitals; see *SI Appendix*).

Fig. 4 also shows the principal component analyses (PCA) of the molecular latent space provided by CEONET. The latent space of STO-3G Hartree–Fock orbitals (Fig. 4*A*) clearly distinguishes between the delocalized non-π orbitals of benzene and the three occupied π-system orbitals. This intuitive latent space representation is transferrable to the Def2SVP orbitals (Fig. 4*B*); the localized orbitals in Fig. 4 *C* and *D* also present highly intuitive mappings. In particular, the latent-space representation of the IBO orbitals shows the symmetry obtained under the IBO unitary transformation, in which the 15 unique Hartree–Fock orbitals are transformed into a set of six identical C-H bonding orbitals, six identical C–C bonding orbitals, and a set of three π orbitals (72). In the Boys localization scheme, it is seen that three similar clusters are formed that approximate the symmetric IBO solutions.

3.2. Predicting Orbital Entropies. The performance of CEONET in predicting ground state orbital entropies from automated aug-cc-pVTZ CASSCF calculations is shown in Fig. 5*A*. Orbital entropies present a difficult quantitative label, as no closed-form expression exists from one-electron operators (i.e., they are a complicated function of the correlated two-body density matrices). As is seen, the model forms an excellent qualitative understanding of multireference character ($R^2 = 0.79$, MAE = 0.03). This quite good performance presents the ability of CEONet to efficiently learn difficult labels even on somewhat small datasets using large basis sets such as aug-cc-pVTZ.

The key use case of such models that aim to predict multireference character is to identify orbitals with high entropy prior to calculation [$\approx S \geq 0.15$ (22)]. Fig. 5*B* presents the receiving

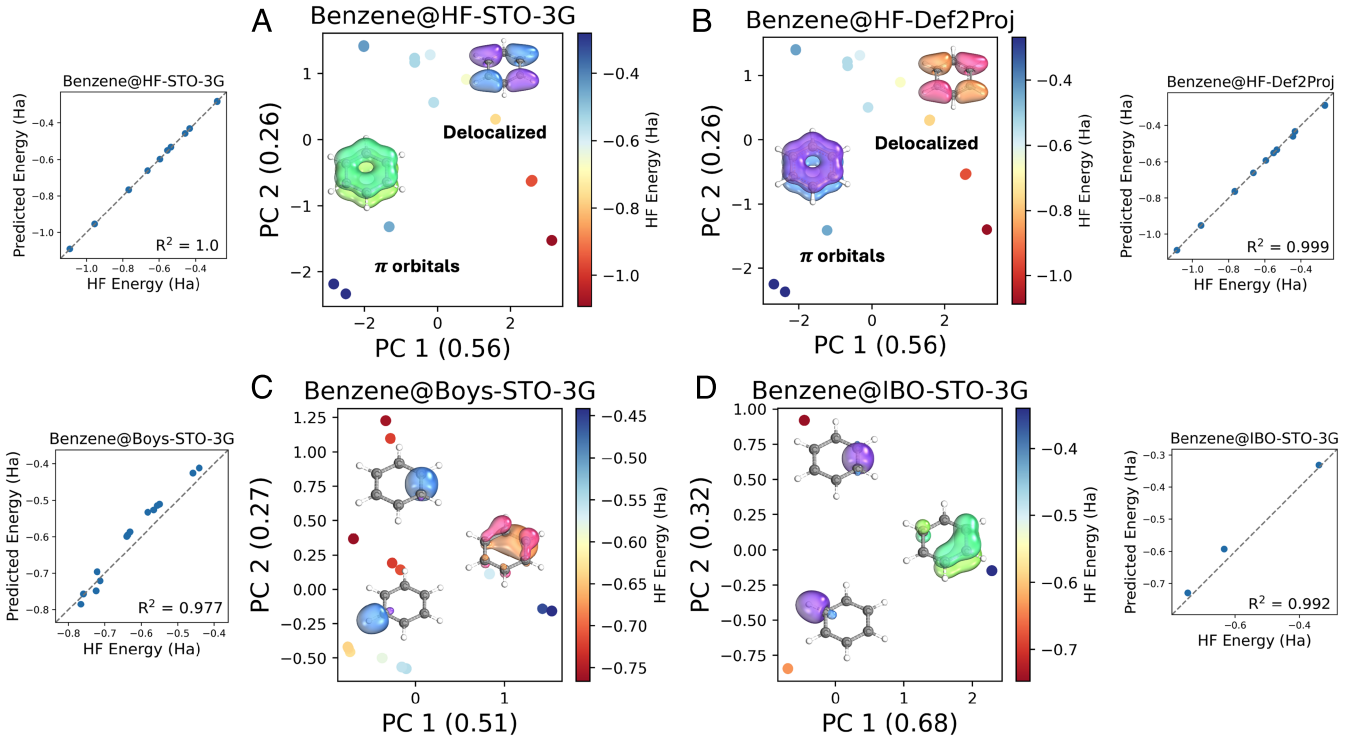

Fig. 4. Principal component analyses (PCA) of different sets of delocalized and localized occupied valence orbitals in benzene, accompanied by parity plots of model performance in predicting their average orbital energy. The percentage of explained variance of each PCA component is put in parentheses on each axis. (*A*) Minimal basis Hartree–Fock orbitals. (*B*) Def2SVP Hartree–Fock orbitals projected onto STO-3G. (*C*) Boys-localized orbitals. (*D*) Intrinsic bonding orbitals. Visualizations of example orbitals are shown in each plot for reference; the color pairs of each orbital are arbitrary.

operating characteristic (ROC) curve of CEONET and confusion matrix in identifying such orbitals. The CEONET model scores a quite good AUC of 0.94, with 88% accuracy in identifying the highly correlated orbitals. We note that previous studies by Golub et al. (73, 74) have used descriptor-based approaches to predict orbital entropies from Hartree–Fock descriptors. Here, the performance of CEONET is particularly impressive as the entropies are only derived from only the molecular geometry and orbital coefficients.

3.3. Predicting Orbital Character. Another key application of CEONET is explicitly classifying different types of orbital character. Fig. 6 shows the performance of CEONET on the task of differentiating between the four types of constructed transition metal complex orbitals outlined in Section 2.3. As is seen, CEONET is able to distinguish perfectly between the different types of orbital character. The CEONET model thus confirms that labels such as "bonding" and "antibonding" are separable in a latent space. Indeed, we find that this task can even be achieved by much simpler models, including voxel- and projection-based approaches (*SI Appendix*).

4. Discussion and Conclusion

Over the past century, molecular orbital theory has fundamentally shaped our understanding of chemical bonding, reactivity, electronegativity, and catalysis. To this day, molecular orbitals remain critical to computing kinetic energy and excited states, and even serve as the basis for computing strong correlation. Given their importance, we take it as a thesis that

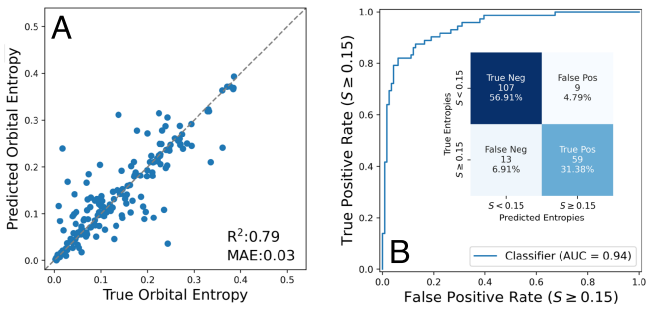

Fig. 5. Accuracy of CEONET in predicting orbital entropies of high-throughput aug-cc-pVTZ CASSCF calculations (25). (*A*) Parity plot of orbital entropy predictions. (*B*) Receiving operating characteristic (ROC) curve and confusion matrix of CEONET in labeling high entropy orbitals ($S >=\geq 0.15$). Both plots show the performance of CEONET on a holdout test set of 10% of the dataset orbitals.

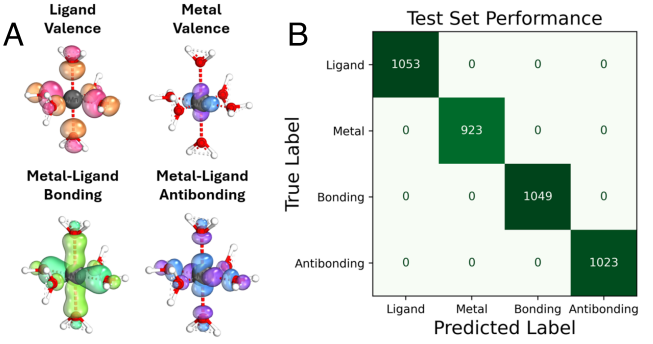

Fig. 6. Accuracy of CEONET in assigning orbital labels to different types of constructed transition metal orbitals: ligand valence, metal valence, ligand-metal bonding, and ligand-metal antibonding. (*A*) The four different types of orbital labels available in the training data. (*B*) Confusion matrix on the holdout test set of data.

building machine-learned representations of molecular orbitals is key for connecting the intuitive understanding of chemists to practical computations.

In this direction, we have developed CEONET, a model that employs state-of-the-art equivariant learning to the task of assigning labels to molecular orbitals. We have found CEONET to perform well at predicting difficult quantitative labels such as the orbital energy (approaching chemical accuracy, ≈0.043 eV or 1.6 mHa), and to model effectively the action of the Fock operator in real-space. Thus, it generalizes well even to orbital types unseen in the training dataset (e.g., localized orbitals or orbitals from smaller basis sets). Furthermore, we have found CEONET to efficiently generalize even to difficult quantitative labels such as the orbital entropy from SA-CASSCF calculations.

The CEONET architecture also provides technical contributions to the representation of orbitals in molecular systems. In particular, CEONET a) demonstrates the direct mapping of molecular orbital coefficients and basis information to symmetric hidden features in graph neural networks, b) includes expressive message passing layers that overcome the orbital sign problem, and c) demonstrates the utility of Cartesian symmetry functions in representing the orbital character. These features expand on previous work such as COEFFNET (75), which employs an equivariant model to predict properties of frontier transition state orbitals as a function of reactant and product orbitals. In contrast, CEONET aims to provide a representation for difficult quantitative labels such as the orbital energy rather than calculating orbital properties in an interpolative manner between two states. CEONET is also intimately connected to Hamiltonian learning models (40, 47, 56, 69, 76–79), as well as models such as ORBNET (38, 80–82) which use molecular orbital properties (i.e., Hartree–Fock matrix components) to predict higher accuracy energies. The CEONET representation can easily be transferred to these other tasks.

Importantly, we also find that CEONET provides an intuitive latent space for separating between different types of molecular orbitals, and generalizes well to unseen orbital types. Without supervised labels CEONET naturally separates localized orbitals (e.g., π-orbitals or non-π IBO or Boys orbitals) from nonlocalized orbitals (Fig. 4). Furthermore, we find CEONET to be effective at identifying human orbital labels, easily separating bonding and antibonding orbitals for a wide array of transition metal complexes. While these capabilities do not improve directly on projection-based approaches for constructing intuitive molecular orbitals (21, 72), they provide a foothold for future workflows exploring these tasks. Further capabilities can be explored through the addition of further datasets (e.g., singly occupied orbitals or open-shell systems).

Given these capabilities, we anticipate that CEONET will serve as a valuable tool for integrating deep learning in computational chemistry workflows. In particular, we envision CEONET playing a key role in the automation and interpretation of multireference methods such as CASSCF, which will help to further refine our understanding of strong correlation in molecular systems. In particular, the ability of CEONET to estimate the orbital entropy near-quantitatively provides a promising foundation for developing a robust automated approach to active space selection, greatly accelerating such calculations. However, more work needs to be done in extending this approach to transition metal systems and excited states.

Nevertheless, despite the remaining challenges, we believe CEONET presents a significant step forward in the integration of deep learning with sophisticated electronic structure approaches. We hope this work offers a solid foundation on orbital representations to anyone looking to contribute to this exciting field.

Data, Materials, and Software Availability. Code has been deposited to https://github.com/GagliardiGroup/CEONet (83). Data has been deposited to https://doi.org/10.5281/zenodo.16934624 (84).

ACKNOWLEDGMENTS. This work is supported as part of the Catalyst Design for Decarbonization Center, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under award no. DE-SC0023383. We thank the Research Computing Center at the University of Chicago and for access to computational resources. Additionally, this research used the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California (UC), Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer). Furthermore, we thank Matthew Hennefarth and Matt Hermes for useful discussions.

Author affiliations: [a]Department of Chemistry, University of Chicago, Chicago, IL 60637; [b]Department of Computer Science, University of Chicago, Chicago, IL 60637; [c]Pritzker School of Molecular Engineering, Chicago, IL 60637; [d]Data Science and Learning Division, Argonne National Laboratory, Lemont, IL 60439; [e]Department of Chemistry, University of California, Berkeley, CA 94720; [f]The Institute of Science and Technology Austria, Klosterneuburg 3400, Austria; [g]Bakar Institute of Digital Materials for the Planet, University of California, Berkeley, CA 94720; and [h]James Franck Institute, Chicago Center for Theoretical Chemistry, University of Chicago, Chicago, IL 60637

Author contributions: D.S.K. and L.G. conceptualized the project; D.S.K., D.G., and R.Z. performed research; D.S.K., D.G., R.Z., N.H., I.F., B.C., and L.G. analyzed data; D.S.K. designed, coded, executed the research, and wrote the paper; N.H. participated in discussions and helped to manage code/writing; I.F. participated in discussions; B.C. guided the research and designed the manuscript; and L.G. managed the project.

Reviewers: C.C., EPFL, Lausanne; S.M.M., University of Toronto; and V.V.S., Universiteit Gent.

Competing interest statement: I.F. recently coauthored a review article with some of the reviewers of this work (27) .

1. D. R. Hartree, W. Hartree, Self-consistent field, with exchange, for beryllium. *Proc. R. Soc. Lond.* **150**, 9–33 (1935).
2. E. D. Glendening, C. R. Landis, F. Weinhold, Natural bond orbital methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 1–42 (2012).
3. G. Knizia, Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts. *J. Chem. Theory Comput.* **9**, 4834–4843 (2013).
4. T. Koopmans, Über die zuordnung von wellenfunktionen und eigenwerten zu den einzelnen elektronen eines atoms. *Physica* **1**, 104–113 (1934).
5. J. Phillips, Generalized Koopmans' theorem. *Phys. Rev.* **123**, 420 (1961).
6. F. Pereira *et al*., Machine learning methods to predict density functional theory B3LYP energies of HOMO and LUMO orbitals. *J. Chem. Inf. Model.* **57**, 11–21 (2017).
7. G. A. Pinheiro *et al*., Machine learning prediction of nine molecular properties based on the SMILES representation of the QM9 quantum-chemistry dataset. *J. Phys. Chem. A* **124**, 9854–9866 (2020).
8. O. Rahaman, A. Gagliardi, Deep learning total energies and orbital energies of large organic molecules using hybridization of molecular fingerprints. *J. Chem. Inf. Model.* **60**, 5971–5983 (2020).
9. A. I. Krylov, From orbitals to observables and back. *J. Chem. Phys.* **153**, 080901 (2020).
10. J. Westermayr, R. J. Maurer, Physically inspired deep learning of molecular excitations and photoemission spectra. *Chem. Sci.* **12**, 10755–10764 (2021).
11. K. Chen, C. Kunkel, B. Cheng, K. Reuter, J. T. Margraf, Physics-inspired machine learning of localized intensive properties. *Chem. Sci.* **14**, 4913–4922 (2023).
12. C. Gaul, S. Cuesta-Lopez, Machine learning for orbital energies of organic molecules upwards of 100 atoms. *Phys. Status Solidi B Basic Res.* **261**, 2200553 (2024).
13. C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models* (John Wiley & Sons, 2013).
14. C. Adamo, D. Jacquemin, The calculations of excited-state properties with time-dependent density functional theory. *Chem. Soc. Rev.* **42**, 845–856 (2013).
15. W. Mi, K. Luo, S. Trickey, M. Pavanello, Orbital-free density functional theory: An attractive electronic structure method for large-scale first-principles simulations. *Chem. Rev.* **123**, 12039–12104 (2023).
16. B. O. Roos, P. R. Taylor, P. E. Sigbahn, A complete active space SCF Method (CASSCF) using a density matrix formulated super-Ci approach. *Chem. Phys.* **48**, 157–173 (1980).
17. V. Veryazov, P. K. Malmqvist, B. O. Roos, How to select active space for multiconfigurational quantum chemistry? *Int. J. Quantum Chem.* **111**, 3329–3338 (2011).

18. P. Pulay, T. P. Hamilton, UHF natural orbitals for defining and starting MC-SCF calculations. *J. Chem. Phys.* **88**, 4926–4933 (1988).
19. J. L. Bao, A. Sand, L. Gagliardi, D. G. Truhlar, Correlated-participating-orbitals pair-density functional method and application to multiplet energy splittings of main-group divalent radicals. *J. Chem. Theory Comput.* **12**, 4274–4283 (2016).
20. C. J. Stein, M. Reiher, Automated selection of active orbital spaces. *J. Chem. Theory Comput.* **12**, 1760–1771 (2016).
21. E. R. Sayfutyarova, Q. Sun, G. K. L. Chan, G. Knizia, Automated construction of molecular active spaces from atomic valence orbitals. *J. Chem. Theory Comput.* **13**, 4063–4078 (2017).
22. C. J. Stein, M. Reiher, autoCAS: A Program for fully automated multiconfigurational calculations. *J. Comput. Chem.* **40**, 2216–2226 (2019).
23. D. S. King, L. Gagliardi, A ranked-orbital approach to select active spaces for high-throughput multireference computation. *J. Chem. Theory Comput.* **17**, 2817–2831 (2021).
24. D. S. King, D. G. Truhlar, L. Gagliardi, Variational active space selection with multiconfiguration pair-density functional theory. *J. Chem. Theory Comput.* **19**, 8118–8128 (2023).
25. D. S. King, M. R. Hermes, D. G. Truhlar, L. Gagliardi, Large-scale benchmarking of multireference vertical-excitation calculations via automated active-space selection. *J. Chem. Theory Comput.* **18**, 6065–6076 (2022).
26. J. J. Wardzala, D. S. King, L. Ogunfowora, B. Savoie, L. Gagliardi, Organic reactivity made easy and accurate with automated multireference calculations. *ACS Cent. Sci.* **10**, 833–841 (2024).
27. J. Van Herck *et al*., Assessment of fine-tuned large language models for real-world chemistry and material science applications. *Chem. Sci.* **16**, 670–684 (2025).
28. P. O. Dral, Quantum chemistry in the age of machine learning. *J. Phys. Chem. Lett.* **11**, 2336–2347 (2020).
29. J. Westermayr, P. Marquetand, Machine learning for electronically excited states of molecules. *Chem. Rev.* **121**, 9873 (2021).
30. B. Huang, O. A. Von Lilienfeld, Ab initio machine learning in chemical compound space. *Chem. Rev.* **121**, 10001–10036 (2021).
31. J. A. Keith *et al*., Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem. Rev.* **121**, 9816–9872 (2021).
32. A. Aldossary *et al*., In silico chemical experiments in the age of AI: From quantum chemistry to machine learning and back. ChemRxiv [Preprint] (2024). 10.26434/chemrxiv–2024–1v269 (Accessed 28 April 2025).
33. S. Cuomo *et al*., Scientific machine learning through physics-informed neural networks: Where we are and what's next. *J. Sci. Comput.* **92**, 88 (2022).
34. J. Behler, Four generations of high-dimensional neural network potentials. *Chem. Rev.* **121**, 10037–10072 (2021).
35. B. Nebgen *et al*., Transferable dynamic molecular charge assignment using deep neural networks. *J. Chem. Theory Comput.* **14**, 4687–4698 (2018).
36. O. T. Unke, M. Meuwly, Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **15**, 3678–3693 (2019).
37. K. T. Schütt, M. Gastegger, A. Tkatchenko, K. R. Müller, R. J. Maurer, Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **10**, 5024 (2019).
38. Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby, T. F. Miller, OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *J. Chem. Phys.* **153**, 124111 (2020).
39. A. S. Christensen *et al*., OrbNet Denali: A machine learning potential for biological and organic chemistry with semi-empirical cost and DFT accuracy. arXiv [Preprint] (2021). http://arxiv.org/abs/2107.00299 (Accessed 28 April 2025).
40. X. Gong *et al*., General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian. *Nat. Commun.* **14**, 2848 (2023).
41. E. Cignoni *et al*., Electronic excited states from physically constrained machine learning. *ACS Cent. Sci.* **10**, 637–648 (2024).
42. M. J. Dewar, E. G. Zoebisch, E. F. Healy, J. J. Stewart, Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).
43. J. J. Stewart, Optimization of parameters for semiempirical methods IV: Extension of MNDO, AM1, and PM3 to more main group elements. *J. Mol. Model.* **10**, 155–164 (2004).
44. R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. Von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1**, 1–7 (2014).
45. D. Balcells, B. B. Skjelstad, tmQM dataset-quantum geometries and properties of 86k transition metal complexes. *J. Chem. Inf. Model.* **60**, 6135–6146 (2020).
46. F. Liu, C. Duan, H. J. Kulik, Rapid detection of strong correlation with machine learning for transition-metal complex high-throughput screening. *J. Phys. Chem. Lett.* **11**, 8067–8076 (2020).
47. H. Yu *et al*., QH9: A quantum hamiltonian prediction benchmark for QM9 molecules. *Adv. Neural Inf. Process. Syst.* **36**, e73689 (2024).
48. K. Boguslawski, P. Tecmer, Orbital entanglement in quantum chemistry. *Int. J. Quantum Chem.* **115**, 1289–1295 (2015).
49. R. Drautz, Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **99**, 014104 (2019).
50. I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, G. Csányi, MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. *Adv. Neural Inf. Process. Syst.* **35**, 11423–11436 (2022).
51. S. Batzner *et al*., E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
52. B. Cheng, Cartesian atomic cluster expansion for machine learning interatomic potentials. *NPJ Comput. Mater.* **10**, 157 (2024).
53. J. Wang *et al*., E(n)-equivariant cartesian tensor message passing interatomic potential. *Nat. Commun.* **15**, 7607 (2024).
54. G. Simeon, G. De Fabritiis, TensorNet: Cartesian tensor representations for efficient learning of molecular potentials. *Adv. Neural Inf. Process. Syst.* **36**, 37334–37353 (2024).
55. D. Hendrycks, K. Gimpel, Gaussian error linear units (gelus). arXiv [Preprint] (2016). http://arxiv.org/abs/1606.08415 (Accessed 28 April 2025).
56. H. Yu, Z. Xu, X. Qian, X. Qian, S. Ji, "Efficient and equivariant graph networks for predicting quantum hamiltonian" in *International Conference on Machine Learning*, A. Krause *et al*., Eds. (PMLR, 2023), pp. 40412–40424.
57. Y. L. Liao, T. Smidt, Equiformer: Equivariant graph attention transformer for 3D atomistic graphs. arXiv [Preprint] (2022). http://arxiv.org/abs/2206.11990 (Accessed 28 April 2025).
58. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv [Preprint] (2014). http://arxiv.org/abs/1412.6980 (Accessed 28 April 2025).
59. Q. Sun *et al*., PySCF: The Python-based simulations of chemistry framework. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8**, e1340 (2018).
60. D. S. King, "Developments and applications of automated multiconfigurational quantum chemistry," PhD thesis, The University of Chicago (2024).
61. M. Véril *et al*., Questdb: A database of highly accurate excitation energies for the electronic structure community. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **11**, e1517 (2021).
62. T. H. Dunning Jr., Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007–1023 (1989).
63. G. Li Manni *et al*., Multiconfiguration pair-density functional theory. *J. Chem. Theory Comput.* **10**, 3669–3680 (2014).
64. R. Pandharkar, M. R. Hermes, D. G. Truhlar, L. Gagliardi, A new mixing of nonlocal exchange and nonlocal correlation with multiconfiguration pair-density functional theory. *J. Phys. Chem. Lett.* **11**, 10158–10163 (2020).
65. M. Rittby, R. J. Bartlett, An open-shell spin-restricted coupled cluster method: Application to ionization potentials in nitrogen. *J. Chem. Phys.* **92**, 3033–3036 (1988).
66. D. E. Woon, T. H. Dunning Jr., Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **98**, 1358–1371 (1993).
67. N. B. Balabanov, K. A. Peterson, Systematically convergent basis sets for transition metals. I. All-electron correlation consistent basis sets for the 3D elements Sc-Zn. *J. Chem. Phys.* **123**, 64107 (2005).
68. I. Dabo *et al*., Koopmans' condition for density-functional theory. *Phys. Rev. B Condens. Matter Mater. Phys.* **82**, 115121 (2010).
69. Z. Yuan *et al*., Deep learning density functional theory Hamiltonian in real space. arXiv [Preprint] (2024). http://arxiv.org/abs/2407.14379 (Accessed 28 April 2025).
70. P. F. Loos, A. Scemama, D. Jacquemin, The quest for highly accurate excitation energies: A computational perspective. *J. Phys. Chem. Lett.* **11**, 2374–2383 (2020).
71. S. F. Boys, Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Rev. Mod. Phys.* **32**, 296 (1960).
72. G. Knizia, Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts. *J. Chem. Theory Comput.* **9**, 4834–4843 (2013).
73. P. Golub, A. Antalik, L. Veis, J. Brabec, Machine learning-assisted selection of active spaces for strongly correlated transition metal systems. *J. Chem. Theory Comput.* **17**, 6053–6072 (2021).
74. P. Golub, A. Antalik, P. Beran, J. Brabec, Mutual information prediction for strongly correlated systems. *Chem. Phys. Lett.* **813**, 140297 (2023).
75. S. Vijay *et al*., Coeffnet: Predicting activation barriers through a chemically-interpretable, equivariant and physically constrained graph neural network. *Chem. Sci.* **15**, 2923–2936 (2024).
76. H. Li *et al*., Deep-learning density functional theory hamiltonian for efficient ab initio electronic-structure calculation. *Nat. Comput. Sci.* **2**, 367–377 (2022).
77. K. T. Schütt, M. Gastegger, A. Tkatchenko, K. R. Müller, R. J. Maurer, Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **10**, 5024 (2019).
78. O. Unke *et al*., SE(3)-equivariant prediction of molecular wavefunctions and electronic densities. *Adv. Neural Inf. Process. Syst.* **34**, 14434–14447 (2021).
79. Y. Wang *et al*., Universal materials model of deep-learning density functional theory hamiltonian. *Sci. Bull.* **69**, 2514–2521 (2024).
80. M. Welborn, L. Cheng, T. F. Miller III, Transferability in machine learning for electronic structure via the molecular orbital basis. *J. Chem. Theory Comput.* **14**, 4772–4779 (2018).
81. S. Dick, M. Fernandez-Serra, Machine learning accurate exchange and correlation functionals of the electronic density. *Nat. Commun.* **11**, 3509 (2020).
82. Y. Chen, L. Zhang, H. Wang, E. Weinan, Ground state energy functional with Hartree-Fock efficiency and chemical accuracy. *J. Phys. Chem. A* **124**, 7155–7165 (2020).
83. D. S. King *et al*., CEONet. GitHub. https://github.com/GagliardiGroup/CEONet. Accessed 23 August 2025.
84. D. S. King *et al*., Cartesian equivariant representations for learning and understanding molecular orbitals. Zenodo. https://doi.org/10.5281/zenodo.16934624. Accessed 23 August 2025.