

# Cartesian Equivariant Representations for Learning and Understanding Molecular Orbitals

Daniel S. King<sup>a,b,1</sup>, Daniel Grzenda<sup>c</sup>, Ray Zhu<sup>d</sup>, Nathaniel Hudson<sup>c,e</sup>, Ian Foster<sup>c,e</sup>, Bingqing Cheng<sup>f,g,h,1</sup>, and Laura Gagliardi<sup>a,d,i,1</sup>

This manuscript was compiled on April 25, 2025

**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 and 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. 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.**

Machine Learning | Electronic Structure Theory

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.,  $\pi$ ,  $\pi^*$ , 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).

Yet, despite the important role molecular orbitals play in chemistry, relatively little work has been put into designing machine-learned representations of molecular orbitals compared to geometries and Hamiltonians (27–31). Nevertheless, accuracy has often improved in deep learning approaches through the development of “physics informed” architectures (32) that integrate ideas from quantum chemistry. For example, Behler outlines four “generations” of machine-learned potentials (31, 33) with later generations (e.g., ANI-1 (34) and PhysNet (35)) 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 (36), OrbNet (37, 38), DeepH (39), and the recent model of Ceriotti and coworkers (40) make explicit reference to the concepts of orbitals and Hamiltonians, being reminiscent of semiempirical models such as AM1 (41) and PM3 (42). As the development

## Significance Statement

Orbital properties such as energies and bonding character are vital to how chemists understand fundamental chemical phenomenon 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 new 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.

Author affiliations: <sup>a</sup>Department of Chemistry, University of Chicago, Chicago IL 60637; <sup>b</sup>Current Affiliation: Bakar Institute of Digital Materials for the Planet, UC Berkeley, California 94720, United States; <sup>c</sup>Department of Computer Science, University of Chicago, Chicago IL 60637; <sup>d</sup>Pritzker School of Molecular Engineering, Chicago IL 60637; <sup>e</sup>Data Science and Learning Division, Argonne National Laboratory, Lemont IL 60439; <sup>f</sup>Department of Chemistry, UC Berkeley, California 94720, United States; <sup>g</sup>The Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria; <sup>h</sup>Bakar Institute of Digital Materials for the Planet, UC Berkeley, California 94720, United States; <sup>i</sup>James Franck Institute, Chicago Center for Theoretical Chemistry, University of Chicago, Chicago IL 60637

<sup>1</sup>To whom correspondence should be addressed. E-mails: king1305@berkeley.edu, bingqingcheng@berkeley.edu, lgagliardi@uchicago.edu.

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, which we believe to be the first 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  $\chi$  we pursue a labeling function

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

where  $l_\phi$  is a label corresponding to either an orbital energy (e.g., from KS-DFT) or a qualitative description (e.g., bonding or antibonding). We develop new datasets for both of these tasks: QM9@HF-STO-3G, consisting of full molden files (43) for the entire QM9 dataset (44); the TMQM@HF-STO3G dataset, consisting of full molden files 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).

Across these learning tasks, we find that CEONET performs well at predicting difficult quantitative labels such as the orbital energy, 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).

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 A). We then describe the background of the Cartesian tensor product networks used in the model (Section B) followed by a complete description of the model architecture (Section C).

**A. 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 (LCAOs) centered on each atom  $i$ :

$$\phi(\mathbf{r}) = \sum_{ik} c_{ik}^1 \chi_{ik}^1(\mathbf{r}) \quad [2]$$

where each atomic orbital  $\chi_{ik}^1(\mathbf{r})$  is a linear combination of so-called “primitive” Gaussian-type orbitals with decay coefficients  $\alpha$ , multiplied by angular functions of order  $l = ||\mathbf{l}||_1$  ( $\mathbf{l} = (l_x, l_y, l_z)$ ) centered on  $\mathbf{r}_i$ :

$$\chi_{ik}^1(\mathbf{r}) = \sum_p \Lambda_{pk}^{(Z_i)} N(\alpha_{pk}^{(Z_i)}, 1) e^{-\alpha_{pk}^{(Z_i)} \mathbf{r}^2} L_{\mathbf{l}}(\mathbf{r} - \mathbf{r}_i) \quad [3]$$

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

We use  $\Lambda_{pk}^{(Z_i)}$  and  $\alpha_{pk}^{(Z_i)}$  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;  $N(\alpha_{pk}^{(Z_i)}, 1)$  is a normalization coefficient determined for each primitive as a function of  $\alpha$  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  $\chi_{ik}^1(\mathbf{r})$  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.

## B. Background: Cartesian Tensor Product Networks.

Machine-learned force fields such as ACE (48), MACE (49), and NequIP (50) 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 non-linear 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

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

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-Gordon coefficients  $C_{l_r m_o l_m}^{l_o}$ . More recently, several models such as CACE (51), HotPP (52), and TensorNet (53) 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:

$$\hat{\mathbf{r}}^{\otimes n} = \hat{\mathbf{r}} \otimes \hat{\mathbf{r}} \dots \otimes \hat{\mathbf{r}} \quad [9]$$

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

$$l_i, l_r \mathbf{X}_{a_1 \dots a_l}^{l_o} = \mathbf{X}_{a_1 \dots a_{l_i-c} b_1 \dots b_c}^{l_i} \mathbf{X}_{b_1 \dots b_c a_1 \dots a_{l_r-c}}^{l_r} \quad [10]$$

$$c \leq \min(l_i, l_r) \quad l_i + l_r - 2c = l_o$$

in which a number of  $c$  Cartesian dimensions are contracted over to produce a tensor of rank  $l_o = l_i + l_r - 2c$ . 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 (52). 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:

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

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

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

resulting in effectively a larger representation of  $c \times (l_i, l_r \rightarrow l_o)$  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 Eq. (11) or Eq. (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  $l_{\max} = 2$ .

### C. Model Architecture.

**Orbital Featurization.** As shown in Section A, 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 coefficients  $\Lambda_{pk}^{(Z_i)}$  and  $\alpha_{pk}^{(Z_i)}$ . Here, we address this issue by sampling the primitive orbitals directly with a learnable basis  $r_c$  and summing over all channels  $k$ :

$$\phi_{ic}^l = \sum_{pk} c_{ik}^l \Lambda_{pk}^{(Z_i)} N(\alpha_{pk}^{(Z_i)}, \mathbf{l}) e^{-\alpha_{pk}^{(Z_i)} r_c^2} \quad [13]$$

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  $\mathbf{l}$ . 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  $\mathbf{h}_{ic}^l$ .

**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  $c_{ik}^l \rightarrow -c_{ik}^l$  (i.e., is even with respect to orbital parity), while the featurization  $\phi_{ic}^l$  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

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

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

$$\mathbf{h}_{ic}^l \otimes^0 \mathbf{h}_{jc}^l \quad [15]$$

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 MLP (54) to compute attention weights  $a_{ijc}$  for each message channel:

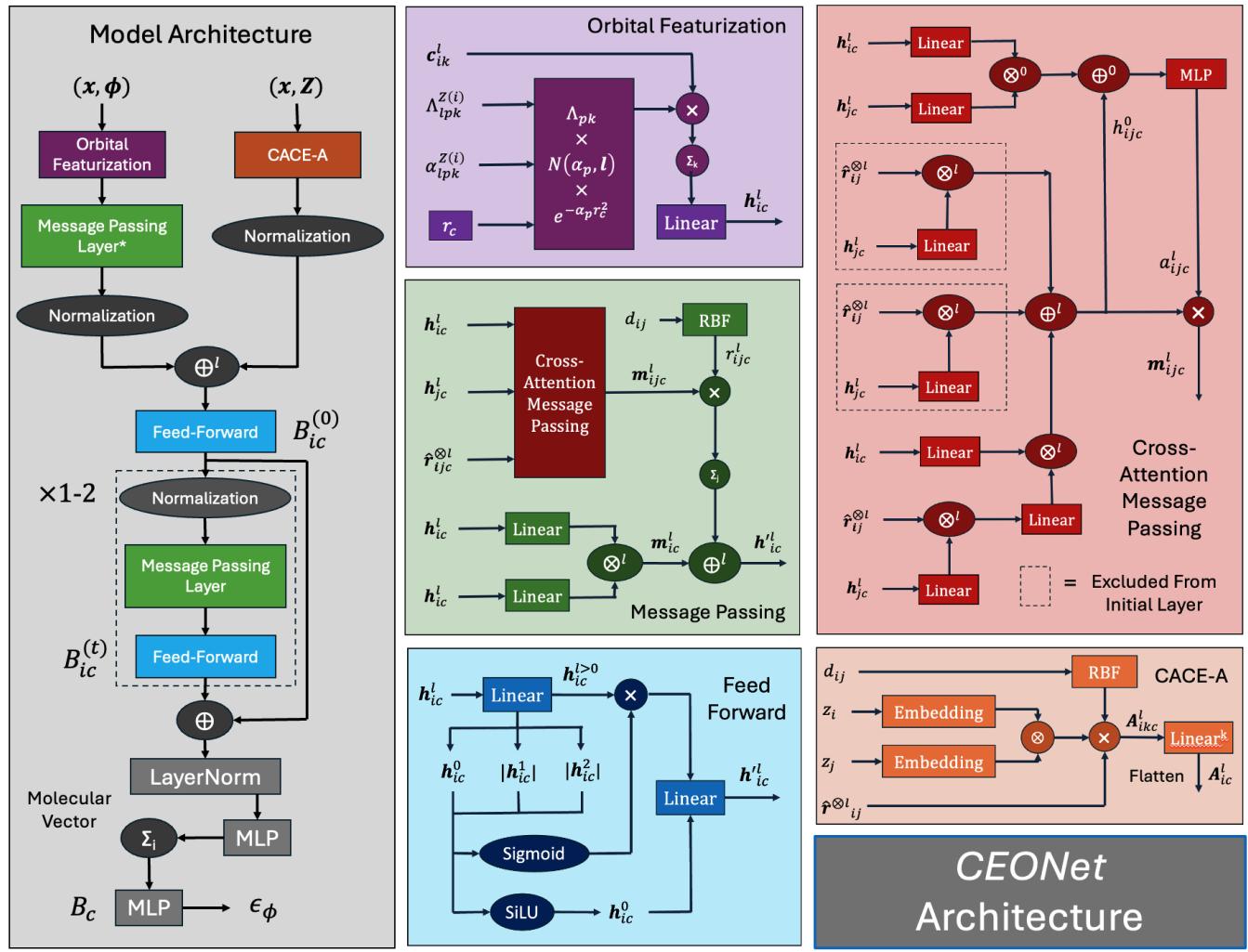
$$\mathbf{m}_{1,ic}^l = \sum_j a_{ijc} \mathbf{m}_{j \rightarrow i,c}^l \quad [16]$$

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

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

to form the new hidden node representation  $\mathbf{h}'_{ic}^l$ . 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



**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 non-learnable steps. Colors other than the background color depict modules outlined in other blocks. **A** Model Architecture. Geometry and orbital features  $(x, \phi)$  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 CACE 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 MLP and summed over to produce a molecular representation  $B_c$ , which is used to predict the orbital energy  $\epsilon_\phi$ . **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  $\hat{\mathbf{r}}_{ij}^{\otimes l}$  are passed through an attention message passing layer (E) 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  $\mathbf{h}_{ic}^{l>0}$  are computed and combined with  $\mathbf{h}_{ic}^l$  and fed into an sigmoid-activated MLP to compute multipliers on the input  $\mathbf{h}_{ic}^{l>0}$  features. A new set of  $\mathbf{h}_{ic}^{l>0}$  features are also computed with a SiLU-activated MLP and then fed into a linear layer with the  $\mathbf{h}_{ic}^{l>0}$  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  $\hat{\mathbf{r}}_{ij}^{\otimes l}$ , 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  $\hat{\mathbf{r}}_{ij}^{\otimes l}$ . 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  $\mathbf{A}_{ic}^l$ .

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.

**CACE Features.** To compute features of the molecular geometry  $(\mathbf{x}, \mathbf{z})$ , we take the A basis generated from the first steps Cartesian atomic cluster expansion (CACE) method (51). 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  $\hat{\mathbf{r}}_{ij}^{\otimes l}$ :

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

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

$$A_{ic}^l = \bigoplus_{k'} \sum_{kj} W_{k'k} \chi_{ijkc}^l \quad [21]$$

We note that the  $\chi$  in equations 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 (51) for a more detailed explanation.

**Tensor Normalization.** Inspired by TensorNet (53), we normalize all node features over the tensor dimension with the rule

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

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

**Feed-Forward Layer.** Inspired by the norm gate of QHNet (55) and the feed-forward layer of Equiformer (56), we aim to compute new scalar features and scaling multipliers of the equivariant channels. Starting with an input  $\mathbf{h}_{ic}^l$  projected down to the channel dimension  $n_c$ , we take as input the scalar features  $\mathbf{h}_{ic}^0$  and the norms of the non-scalar features  $|\mathbf{h}_{ic}^l| = \sum_i (h_{ic}^l)^2$ . These are fed into a SiLU-activated layer to compute new scalar features  $\mathbf{h}_{ic}^0$  and a sigmoid-activated layer to compute scalings of the  $\mathbf{h}_{ic}^{(l>0)}$  features. The scalar and non-scalar features are then fed to a final linear layer to produce the new representation  $\mathbf{h}_{ic}^n$ .

**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 (51) 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:

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

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

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

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  $(1, 2, 1)$ .

**Model Architecture & Readout.** We finally review the model architecture in full, as outlined in Figure 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  $B_{ic}^{(0)}$ .

The unified hidden representation is then fed through  $n_{\text{layer}}$  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  $B_{ic}^{(t)}$ . The  $B_{ic}^{(t)}$  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  $\epsilon_\phi$ .

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 to provide a good balance between cost and expressiveness. 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 (57) using an initial learning rate of  $10^{-3}$  over batch sizes of 128 orbitals. Models employed on transition metals use  $n_{\text{layer}} = 1$  and batch sizes of 32 due to memory constraints.

## 2. Data

**A. QM9 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 (58). 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 technique.

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 non-valence virtual orbitals absent in STO-3G). We refer to the two datasets as “QM9@HF-STO-3G” and “QM9@B3LYP-Def2SVP” throughout the paper. Furthermore, to test the performance of CEONET on predicting DFT orbital energies orthogonally to energies from larger basis sets, we have developed a QM9@B3LYP-STO-3G dataset for a subset of 5000 QM9 molecules, which we present in the Supporting Information.

Energies are shifted and scaled by the mean and standard deviation 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<sub>2</sub>, where restricted Hartree-Fock will only occupy one of the  $\pi^*$  antibonding orbitals, causing that orbital to have arbitrarily lower energy).

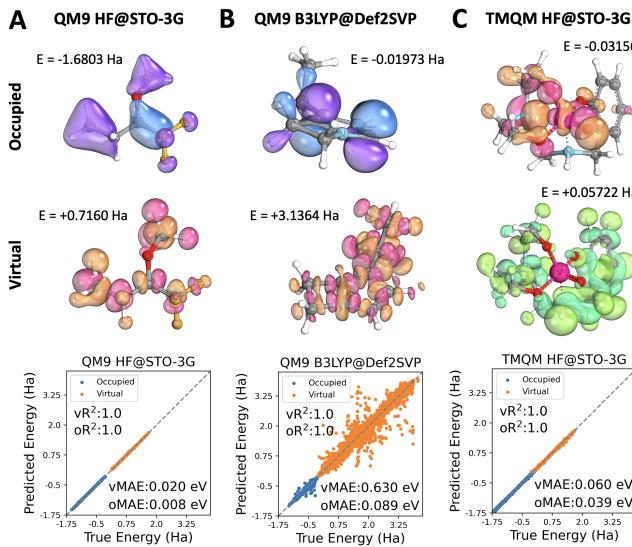
**B. Transition Metal Complexes Datasets.** To test the generalizability of our model to orbitals of systems containing transition metals, we have developed the TMOrb dataset, which consists of minimal-basis STO-3G RHF Hartree-Fock calculations for a subset of 42k complexes in the extensive TMQM dataset 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. This data allows 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  $\phi$ , we carry out the following steps:

(1) Project the restricted open-shell Hartree-Fock (ROHF) (59) Fock matrix  $F$  separately into the space of metal valence orbitals and ligand (i.e., non-metal) 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  $\phi_L$  and one metal orbital  $\phi_M$ , 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 4865 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 (60–62), and excluding the addition of the 4s orbital from calculations. Converged calculations resulted in a final dataset of 3702 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



**Fig. 2.** Performance of CEONET on predicting 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.

**Table 1.** Mean absolute errors (mHa) of CEONET on different orbital energy learning tasks, trained with different numbers of training data. The final two columns show the MAE extrapolated to the final dataset size and the slope of the learning curve  $\alpha$  given the fits shown in Figure 3.

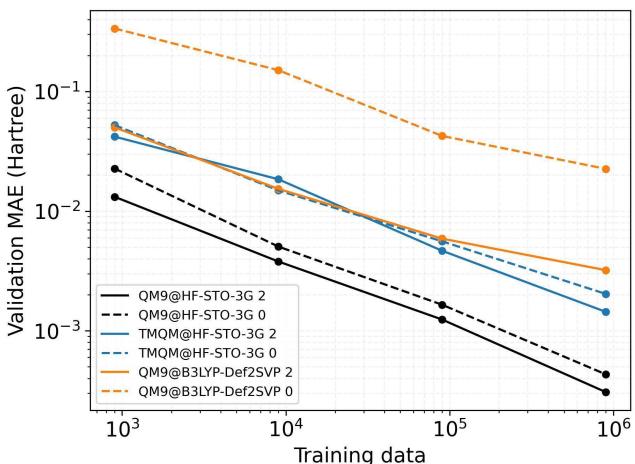
Data	Occ	$10^3$	$10^4$	$10^5$	$10^6$	Extrap	$\alpha$
QM9@HF-STO-3G	2	13.2	3.8	1.2	0.31	0.17	1.61
	0	22.7	5.1	1.7	0.43	0.25	1.69
TMQM@HF-STO-3G	2	41.9	18.5	4.7	1.4	0.84	1.50
	0	52.3	14.9	5.6	2.0	1.2	1.40
QM9@B3LYP-Def2SVP	2	50.1	15.3	5.9	3.2	1.8	1.20
	0	335.9	150.5	42.5	22.5	6.6	1.22

metal orbitals, resulting in a final balanced “TMConstructed” dataset of 32,378 orbitals.

### 3. Results

**A. 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 Figure 2. All models are trained on a dataset of  $10^6$  orbitals and inferred on a test set of 10% of the total data (100k orbitals). As is seen, CEONET is able to learn orbital energies well, with predictions on most datasets approaching chemical accuracy (0.03 eV, see error histograms in SI Figure 2).

The learning curves of CEONET applied to different datasets is shown in Figure 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, with the slope of the learning curve  $\alpha > 1$  in all cases. The validation accuracies extrapolated to the entire datasets are shown in Table 1; all extrapolated values reach chemical accuracy, except for the virtual orbitals



**Fig. 3.** Learning curve 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 mHa is plotted against the base-10 log of the number of training points in the training data.

in QM9@B3LYP-Def2SVP. All models achieve an  $R^2$  close to 1 (Figure 2), and the larger MAE of the virtual orbitals is in part due to their very high energies.

Nevertheless, a key observation is that virtual orbitals, particularly those from large basis sets (one of which is shown in Figure 2B) pose a particular challenge for CEONET even when accounting for their higher energies. The fundamental difficulty of this task comes from the extreme variation of the 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, the CEONET learns to approximate the real-space action of the Fock matrix  $F$  (63):

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

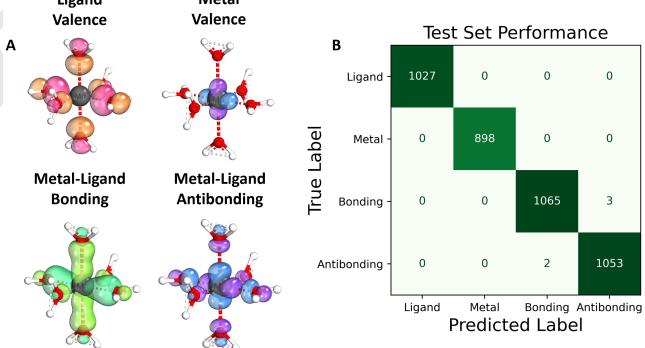
Because the complexity of this mapping scales with the number of basis functions, orbital energies in larger basis sets are naturally more difficult to learn.

A benefit of this behavior is that CEONET easily generalizes to unseen orbital types. Figure 4 shows the performance of CEONET, here trained on a dataset of only 51,200 orbitals, on different localized subspaces of the occupied Hartree-Fock valence orbitals in benzene (with a symmetric structure taken from QUESTDB (64)). Specifically, we test performance on projected orbitals from Def2SVP calculations (Figure 4b), Boys-localized orbitals (65) (Figure 4c), and intrinsic bonding orbitals (IBOs) (3) (Figure 4d) 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  $R^2 \approx 0.95$  in both cases.

More importantly, Figure 4 shows the principal component analyses (PCA) of the molecular latent space provided by the

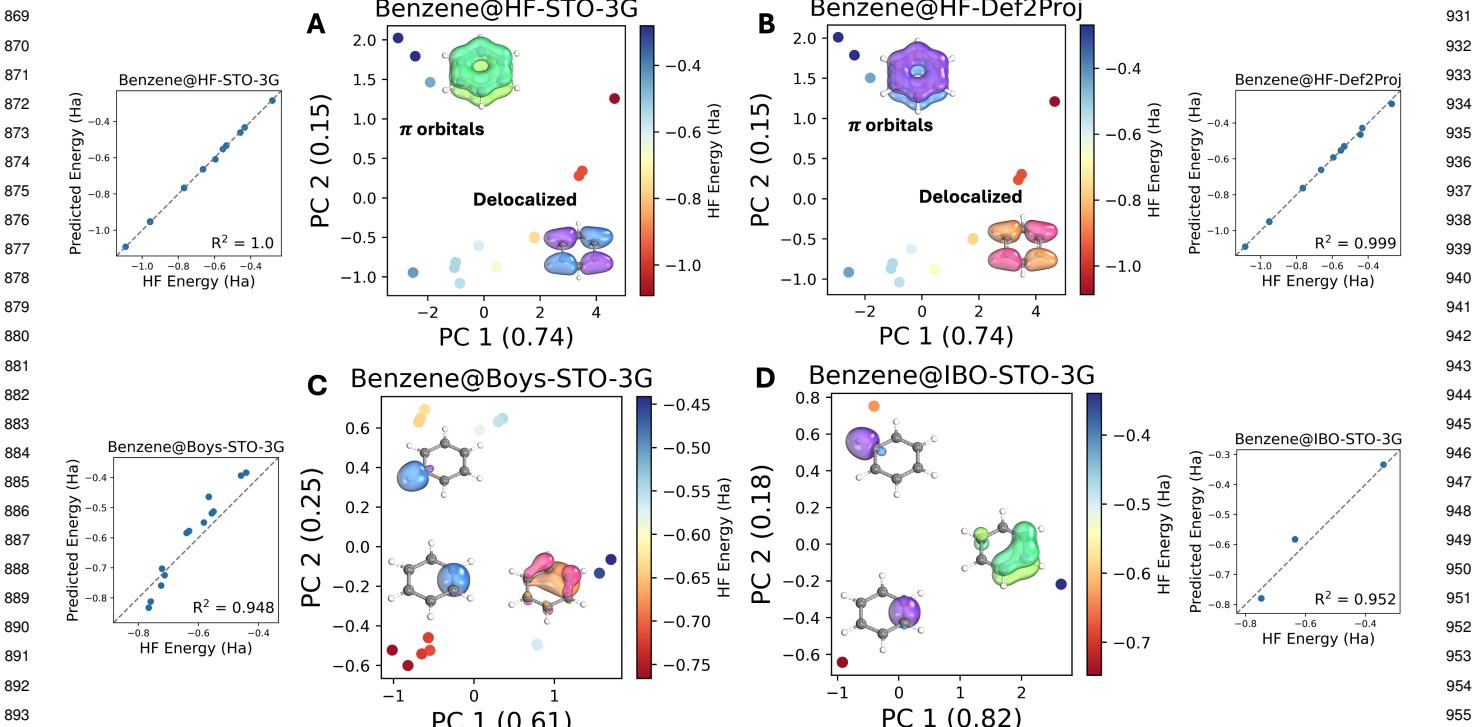
CEONET. The latent space of STO-3G Hartree-Fock orbitals (Figure 4a) clearly distinguishes between the delocalized non- $\pi$  orbitals of benzene and the three occupied  $\pi$ -system orbitals. This intuitive latent space representation is transferrable to the Def2SVP orbitals (Figure 4b); the localized orbitals in Figures 4c and 4d 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  $\pi$  orbitals (66). In the Boys localization scheme, it is seen that three similar clusters are formed that approximate the symmetric IBO solutions.

We obtain further insight into the latent space representations by looking at the PCA decomposition of the entire space of Hartree-Fock and localized orbitals (Figure 5). We see that the model distinguishes between three sets of orbitals: (a) delocalized HF orbitals, (b) localized non- $\pi$  orbitals, and orbitals with  $\pi$  character (both localized and non-localized). The fact that the HF  $\pi$  orbitals lie closer to the localized non- $\pi$  orbitals than the HF orbitals reflects the intuition that the  $\pi$  orbitals are more “localized” in some sense than the non- $\pi$  orbitals. The degeneracy of the STO-3G orbitals and the projected Def2SVP orbitals in the latent space is also observed.



**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. Left: The four different types of orbital labels available in the training data. Right: Confusion matrix on the holdout test set of data.

**B. Predicting Orbital Character.** Another key application of CEONET is explicitly classifying different types of orbital character. Figure 6 shows the performance of CEONET on the task of differentiating between the four types of constructed transition metal complex orbitals outlined in Section B. As is seen, CEONET is able to distinguish between the different types of orbital character, with only minor confusion between bonding and antibonding orbitals. 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 (Supporting Information).



**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.

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

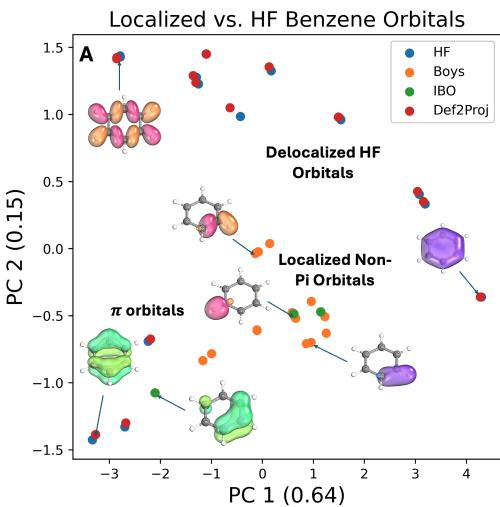
In this direction, we have developed CEONET, a model that employs state-of-the-art equivariant learning to the task of assigning labels to molecular orbitals. We have found CEONET to perform well at predicting difficult quantitative labels such as the orbital energy (generally within chemical accuracy,  $< .03$  eV), 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).

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 (67), 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 (39, 47, 55, 63, 68–71), as well as models such as ORBNET (72–75) 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 separates localized orbitals (e.g.,  $\pi$ -orbitals or non- $\pi$  IBO or Boys orbitals) from non-localized orbitals (Figure 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.

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 complete active space self-consistent field theory (CASSCF), which will help to further refine our understanding of strong correlation in molecular systems. We hope this work offers a solid foundation on orbital representations for anyone looking to contribute to this exciting field.



**Fig. 5.** Combined principal component analysis (PCA) of the different sets of orbitals shown in Figure 6 (STO-3G Hartree-Fock (HF), Boys-localized (Boys), IBO-localized (IBO), and projected Def2SVP (Def2Proj)). Example orbitals are shown from each region of the plot. The percentage of explained variance of each PCA component is put in parentheses on each axis. Visualizations of example orbitals are shown in each plot for reference. The color pairs of each orbital are arbitrary.

**Supporting Information Appendix (SI).** The CEONET code and example training/inference scripts are available at <https://github.com/GagliardiGroup/CEONet>. Datasets are available on Zenodo at <https://doi.org/10.5281/zenodo.15258765>. Supporting information shows the performance of CEONET on QM9@B3LYP-STO-3G data as well as a brief description of previous models used to predict orbital energies and characterizations.

**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. The authors thank the Research Computing Center (RCC) 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, 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 surrounding orbital energy gradients.

- Douglas Rayner Hartree and William Hartree. Self-consistent field, with exchange, for beryllium. *Proc. R. Soc. Lond.*, 150(869):9–33, 1935.
- Eric D Glendening, Clark R Landis, and Frank Weinhold. Natural bond orbital methods. *Wiley interdisciplinary reviews: computational molecular science*, 2(1):1–42, 2012.
- Gerald Knizia. Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts. *Journal of chemical theory and computation*, 9(11):4834–4843, 2013.
- Tjalling Koopmans. Über die zuordnung von wellenfunktionen und eigenwerten zu den einzelnen elektronen eines atoms. *Physica*, 1(1–6):104–113, 1934.
- JC Phillips. Generalized Koopmans' theorem. *Phys. Rev.*, 123(2):420, 1961.
- Florbel Pereira, Kaixia Xiong, Diogo ARS Latino, Chengcheng Wu, Qingyou Zhang, and Joao Aires-de Sousa. Machine learning methods to predict density functional theory B3LYP energies of HOMO and LUMO orbitals. *J. Chem. Inf. Model.*, 57(1):11–21, 2017.
- Gabriel A Pinheiro, Johnatan Mucelin, Marinalva D Soares, Ronaldo C Prati, Juarez LF Da Silva, and Marcos G Quiles. Machine learning prediction of nine molecular properties based on the SMILES representation of the QM9 quantum-chemistry dataset. *J. Phys. Chem. A*, 124(47):9854–9866, 2020.
- Obaidur Rahaman and Alessio Gagliardi. Deep learning total energies and orbital energies of large organic molecules using hybridization of molecular fingerprints. *Journal of Chemical Information and Modeling*, 60(12):5971–5983, 2020.
- Anna I Krylov. From orbitals to observables and back. *J. Chem. Phys.*, 153(8), 2020.
- Julia Westermayr and Reinhard J Maurer. Physically inspired deep learning of molecular excitations and photoemission spectra. *Chem. Sci.*, 12(32):10755–10764, 2021.

- Ke Chen, Christian Kunkel, Bingqing Cheng, Karsten Reuter, and Johannes T Margraf. Physics-inspired machine learning of localized intensive properties. *Chem. Sci.*, 14(18):4913–4922, 2023. 1055  
1056
- Christopher Gaul and Santiago Cuesta-Lopez. Machine learning for orbital energies of organic molecules upwards of 100 atoms. *Phys. Status Solidi B Basic Res.*, 261(1):2200553, 2024. 1057  
1058
- Christopher J. Cramer. *Essentials of Computational Chemistry: Theories and Models*. John Wiley & Sons, April 2013. ISBN 978-1-118-71227-6. Google-Books-ID: k4R6cf717q0C. 1059
- Carlo Adamo and Denis Jacquemin. The calculations of excited-state properties with time-dependent density functional theory. *Chemical Society Reviews*, 42(3):845–856, 2013. 1060
- Wenhai Mi, Kai Luo, SB Trickey, and Michele Pavanello. Orbital-free density functional theory: An attractive electronic structure method for large-scale first-principles simulations. *Chem. Rev.*, 123(21):12039–12104, 2023. 1062  
1063
- Björn O Roos, Peter R Taylor, and Per EM Sigmahn. A complete active space SCF Method (CASSCF) using a density matrix formulated super-Ci approach. *Chem. Phys.*, 48(2):157–173, 1980. 1064  
1065
- Valera Veryazov, Per Åke Malmqvist, and Björn O. Roos. How to select active space for multiconfigurational quantum chemistry? *Int. J. Quantum Chem.*, 111(13):3329–3338, 2011. ISSN 1097-461X. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.23068>. 1066  
1067
- Peter Pulay and Tracy P. Hamilton. UHF natural orbitals for defining and starting MC-SCF calculations. *J. Chem. Phys.*, 88(8):4926–4933, April 1988. ISSN 0021-9606. URL <https://doi.org/10.1063/1.454704>. 1069  
1070
- Junwei Lucas Bao, Andrew Sand, Laura Gagliardi, and Donald 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(9):4274–4283, September 2016. ISSN 1549-9618. URL <https://doi.org/10.1021/acs.jctc.6b00569>. 1071  
1072
- Christopher J. Stein and Markus Reiher. Automated Selection of Active Orbital Spaces. *J. Chem. Theory Comput.*, 12(4):1760–1771, April 2016. ISSN 1549-9618. URL <https://doi.org/10.1021/acs.jctc.6b00156>. 1075  
1076
- Elvira R. Sayfutyarova, Qiming Sun, Garnet Kin-Lic Chan, and Gerald Knizia. Automated Construction of Molecular Active Spaces from Atomic Valence Orbitals. *J. Chem. Theory Comput.*, 13(9):4063–4078, September 2017. ISSN 1549-9618. URL <https://doi.org/10.1021/acs.jctc.7b00128>. 1077  
1078
- Christopher J. Stein and Markus Reiher. autoCAS: A Program for Fully Automated Multiconfigurational Calculations. *J. Comput. Chem.*, 40(25):2216–2226, 2019. ISSN 1096-987X. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.25869>. 1080  
1081
- Daniel S. King and Laura Gagliardi. A Ranked-Orbital Approach to Select Active Spaces for High-Throughput Multireference Computation. *J. Chem. Theory Comput.*, 17(5):2817–2831, May 2021. ISSN 1549-9618. URL <https://doi.org/10.1021/acs.jctc.1c00037>. 1082  
1083
- Daniel S King, Donald G Truhlar, and L Gagliardi. Variational active space selection with multiconfiguration pair-density functional theory. *J. Chem. Theory Comput.*, 2023. 1084
- Daniel S. King, Matthew R. Hermes, Donald G. Truhlar, and Laura Gagliardi. Large-Scale Benchmarking of Multireference Vertical-Excitation Calculations via Automated Active-Space Selection. *J. Chem. Theory Comput.*, 18(10):6065–6076, October 2022. ISSN 1549-9618. URL <https://doi.org/10.1021/acs.jctc.2c00630>. 1086  
1087
- Jacob J Wardzala, Daniel S King, Lawal Ogunfowora, Brett Savoie, and Laura Gagliardi. Organic reactivity made easy and accurate with automated multireference calculations. *ACS Central Science*, 10(4):833–841, 2024. 1088  
1089
- Pavlo O Dral. Quantum chemistry in the age of machine learning. *J. Phys. Chem. Lett.*, 11(6):2336–2347, 2020. 1090  
1091
- Julia Westermayr and Philipp Marquetand. Machine learning for electronically excited states of molecules. *Chem. Rev.*, 121(16):9873, 2021. 1092
- Bing Huang and O Anatole Von Lilienfeld. Ab initio machine learning in chemical compound space. *Chem. Rev.*, 121(16):10001–10036, 2021. 1094
- John A Keith, Valentin Vassilev-Galindo, Bingqing Cheng, Stefan Chmiela, Michael Gastegger, Klaus-Robert Müller, and Alexandre Tkatchenko. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem. Rev.*, 121(16):9816–9872, 2021. 1095  
1096
- Abdulrahman Aldossary, Jorge Arturo Campos-Gonzalez-Angulo, Sergio Pablo-Garcia, Shi Xuan Leong, Ella Miray Rajaonson, Luca Thiede, Gary Tom, Andrew Wang, Davide Avagliano, and Alan Aspuru-Guzik. In silico chemical experiments in the age of ai: From quantum chemistry to machine learning and back. *ChemRxiv*, pages doi:10.26434/chemrxiv-2024-1v269, 2024. 1098  
1099
- Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics-informed neural networks: Where we are and what's next. *J. Sci. Comput.*, 92(3):88, 2022. 1102  
1103
- Jörg Behler. Four generations of high-dimensional neural network potentials. *Chemical Reviews*, 121(16):10037–10072, 2021. 1104
- Benjamin Nebgen, Nicholas Lubbers, Justin S Smith, Andrew E Sifain, Andrey Lokhov, Olexandr Isayev, Adrian E Roitberg, Kipton Barros, and Sergei Tretiak. Transferable dynamic molecular charge assignment using deep neural networks. *J. Chem. Theory Comput.*, 14(9):4687–4698, 2018. 1105  
1106
- Oliver T Unke and Markus Meuwly. Phynet: A neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.*, 15(6):3678–3693, 2019. 1108  
1109
- Kristof T Schütt, Michael Gastegger, Alexandre Tkatchenko, K-R Müller, and Reinhard J Maurer. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.*, 10(1):5024, 2019. 1110  
1111
- Zhuoran Qiao, Matthew Welborn, Animashree Anandkumar, Frederick R. Manby, and Thomas F. Miller. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *J. Chem. Phys.*, 153(12):24111, September 2020. ISSN 0021-9606. URL <https://aip.scitation.org/doi/full/10.1063/5.0021955>. 1113  
1114

- 1117 38. Anders S. Christensen, Sai Krishna Sirumalla, Zhuoran Qiao, Michael B. O'Connor, Daniel  
1118 G. A. Smith, Feizhi Ding, Peter J. Bygrave, Animashree Anandkumar, Matthew Welborn,  
1119 Frederick R. Manby, and Thomas F. Miller III. OrbNet Denali: A machine learning potential  
1120 for biological and organic chemistry with semi-empirical cost and DFT accuracy.  
*arXiv:2107.00299 [physics]*, July 2021. URL <http://arxiv.org/abs/2107.00299>. arXiv:  
2107.00299.
- 1121 39. Xiaoxun Gong, He Li, Nianlong Zou, Runzhang Xu, Wenhui Duan, and Yong Xu. General  
1122 framework for E(3)-equivariant neural network representation of density functional theory  
1123 Hamiltonian. *Nat. Commun.*, 14(1):2848, 2023.
- 1124 40. Edoardo Cignoni, Divya Suman, Jigyasa Nigam, Lorenzo Cupellini, Benedetta Mennucci,  
1125 and Michele Ceriotti. Electronic excited states from physically constrained machine learning.  
*ACS Cent. Sci.*, 10(3):637–648, 2024.
- 1126 41. Michael JS Dewar, Eve G Zoebisch, Eamonn F Healy, and James JP Stewart. Development  
1127 and use of quantum mechanical molecular models. 76. AM1: A new general purpose  
1128 quantum mechanical molecular model. *J. Am. Chem. Soc.*, 107(13):3902–3909, 1985.
- 1129 42. James JP Stewart. Optimization of parameters for semiempirical methods iv: Extension of  
1130 mndo, am1, and pm3 to more main group elements. *J. Mol. Model.*, 10(2):155–164, 2004.
- 1131 43. Gijs Schaftenaer and Jan H Noordik. Molden: A pre-and post-processing program for  
1132 molecular and electronic structures. *J. Comput. Aided Mol. Des.*, 14:123–134, 2000.
- 1133 44. Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld.  
1134 Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data*, 1(1):1–7,  
2014.
- 1135 45. David Balcells and Bastian Bjerke Skjelstad. tmqm dataset—quantum geometries and  
1136 properties of 86k transition metal complexes. *Journal of chemical information and modeling*,  
60(12):6135–6146, 2020.
- 1137 46. Fang Liu, Chenru Duan, and Heather J Kulik. Rapid detection of strong correlation with  
1138 machine learning for transition-metal complex high-throughput screening. *J. Phys. Chem. Lett.*,  
1139 11(19):8067–8076, 2020.
- 1140 47. Haiyang Yu, Meng Liu, Youzhi Luo, Alex Strasser, Xiaofeng Qian, Xiaoning Qian, and  
1141 Shuiwang Ji. Oh9: A quantum hamiltonian prediction benchmark for qm9 molecules.  
*Advances in Neural Information Processing Systems*, 36, 2024.
- 1142 48. Ralf Drautz. Atomic cluster expansion for accurate and transferable interatomic potentials.  
*Physical Review B*, 99(1):014104, 2019.
- 1143 49. Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi. MACE:  
1144 Higher order equivariant message passing neural networks for fast and accurate force fields.  
*Advances in Neural Information Processing Systems*, 35:11423–11436, 2022.
- 1145 50. Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailo, Mordechai  
Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E(3)-equivariant graph neural  
networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13  
(1):2453, 2022.
- 1146 51. Bingqing Cheng. Cartesian atomic cluster expansion for machine learning interatomic  
1147 potentials. *npj Computational Materials*, 10(1):157, 2024.
- 1148 52. Junjie Wang, Yong Wang, Haotong Zhang, Ziyang Yang, Zhixin Liang, Jiuyang Shi, Hui-Tian  
1149 Wang, Dingyu Xing, and Jian Sun. E(n)-equivariant cartesian tensor message passing  
interatomic potential. *Nature Communications*, 15(1):7607, 2024.
- 1150 53. Guillen Simeon and Gianni De Fabritiis. TensorNet: Cartesian tensor representations for  
1151 efficient learning of molecular potentials. *Advances in Neural Information Processing  
Systems*, 36, 2024.
- 1152 54. Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv:1606.08415*  
[cs.LG], 2016. Preprint.
- 1153 55. Haiyang Yu, Zhao Xu, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. Efficient and  
1154 equivariant graph networks for predicting quantum hamiltonian. In *International Conference  
on Machine Learning*, pages 40412–40424. PMLR, 2023.
- 1155 56. Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3D  
1156 atomistic graphs. *arXiv preprint arXiv:2206.11990*, 2022.
- 1157 57. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv  
preprint arXiv:1412.6980*, 2014.
- 1158 58. Qiming Sun, Timothy C Berkelbach, Nick S Blunt, George H Booth, Sheng Guo, Zhendong  
1159 Li, Junzi Liu, James D McClain, Elvira R Sayfutyarova, Sandeep Sharma, Sebastian  
1160 Wouters, and Garnet Kin-Lic Chan. PySCF: The Python-based simulations of chemistry  
framework. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 8(1):e1340, 2018.
- 1161 59. Magnus Rittby and Rodney J Bartlett. An open-shell spin-restricted coupled cluster method:  
1162 Application to ionization potentials in nitrogen. *J. Chem. Phys.*, 92(11):3033–3036, 1988.
- 1163 60. Thom H Dunning Jr. Gaussian basis sets for use in correlated molecular calculations. I. The  
1164 atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90(2):1007–1023, 1989.
- 1165 61. David E Woon and Thom H Dunning Jr. Gaussian basis sets for use in correlated molecular  
1166 calculations. III. The atoms aluminum through argon. *J. Chem. Phys.*, 98(2):1358–1371,  
1993.
- 1167 62. Nikolai B Balabanov and Kirk A Peterson. Systematically convergent basis sets for  
1168 transition metals. I. All-electron correlation consistent basis sets for the 3d elements Sc–Zn.  
*J. Chem. Phys.*, 123(6), 2005.
- 1169 63. Zilong Yuan, Zechen Tang, Honggeng Tao, Xiaoxun Gong, Zehzhou Chen, Yuxiang Wang,  
1170 He Li, Yang Li, Zhiming Xu, Minghui Sun, et al. Deep learning density functional theory  
1171 Hamiltonian in real space. *arXiv preprint arXiv:2407.14379*, 2024.
- 1172 64. Pierre-François Loos, Anthony Scemama, and Denis Jacquemin. The Quest for Highly  
1173 Accurate Excitation Energies: A Computational Perspective. *J. Phys. Chem. Lett.*, 11(6):  
2374–2383, March 2020. ISSN 1948-7185, 1948-7185. . URL  
<https://pubs.acs.org/doi/10.1021/acs.jpclett.0c00014>.
- 1174 65. S Francis Boys. Construction of some molecular orbitals to be approximately invariant for  
1175 changes from one molecule to another. *Reviews of Modern Physics*, 32(2):296, 1960.
- 1176 66. Gerald Knizia. Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and  
1177 Chemical Concepts. *J. Chem. Theory Comput.*, 9(11):4834–4843, November 2013. ISSN  
1549-9618. . URL <https://doi.org/10.1021/ct400687b>.
- 1178 67. Coeff.
68. He Li, Zun Wang, Nianlong Zou, Meng Ye, Runzhang Xu, Xiaoxun Gong, Wenhui Duan, and  
Yong Xu. Deep-learning density functional theory hamiltonian for efficient ab initio  
electronic-structure calculation. *Nature Computational Science*, 2(6):367–377, 2022.
69. K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer. Unifying machine  
learning and quantum chemistry with a deep neural network for molecular wavefunctions.  
*Nat. Commun.*, 10(1):5024, November 2019. ISSN 2041-1723. . URL  
<https://www.nature.com/articles/s41467-019-12875-2>.
70. Oliver Unke, Mihail Bogojeski, Michael Gastegger, Mario Geiger, Tess Smidt, and  
Klaus-Robert Müller. SE(3)-equivariant prediction of molecular wavefunctions and electronic  
densities. *Advances in Neural Information Processing Systems*, 34:14434–14447, 2021.
71. Yuxiang Wang, Yang Li, Zechen Tang, He Li, Zilong Yuan, Honggeng Tao, Nianlong Zou,  
Ting Bao, Xinghao Liang, Zehzhou Chen, et al. Universal materials model of deep-learning  
density functional theory hamiltonian. *Science Bulletin*, 69(16):2514–2521, 2024.
72. Matthew Welborn, Lixue Cheng, and Thomas F Miller III. Transferability in machine learning  
for electronic structure via the molecular orbital basis. *Journal of chemical theory and  
computation*, 14(9):4772–4779, 2018.
73. Zhuoran Qiao, Matthew Welborn, Animashree Anandkumar, Frederick R Manby, and  
Thomas F Miller. OrbNet: Deep learning for quantum chemistry using symmetry-adapted  
atomic-orbital features. *The Journal of chemical physics*, 153(12), 2020.
74. Sebastian Dick and Marivi Fernandez-Serra. Machine learning accurate exchange and  
correlation functionals of the electronic density. *Nature communications*, 11(1):3509, 2020.
75. Yixiao Chen, Linfeng Zhang, Han Wang, and Weinan E. Ground state energy functional  
with Hartree–Fock efficiency and chemical accuracy. *The Journal of Physical Chemistry A*,  
124(35):7155–7165, 2020.