

Cite this: *Chem. Sci.*, 2021, 12, 10207

All publication charges for this article have been paid for by the Royal Society of Chemistry

Predicting phosphorescence energies and inferring wavefunction localization with machine learning†

Andrew E. Sifain,[†] Levi Lystrom,[†] Richard A. Messerly,^a Justin S. Smith,^{ab} Benjamin Nebgen,^{ac} Kipton Barros,^{ab} Sergei Tretiak,^{abc} Nicholas Lubbers[†] and Brendan J. Gifford[†]

Phosphorescence is commonly utilized for applications including light-emitting diodes and photovoltaics. Machine learning (ML) approaches trained on *ab initio* datasets of singlet–triplet energy gaps may expedite the discovery of phosphorescent compounds with the desired emission energies. However, we show that standard ML approaches for modeling potential energy surfaces inaccurately predict singlet–triplet energy gaps due to the failure to account for spatial localities of spin transitions. To solve this, we introduce localization layers in a neural network model that weight atomic contributions to the energy gap, thereby allowing the model to isolate the most determinative chemical environments. Trained on the singlet–triplet energy gaps of organic molecules, we apply our method to an out-of-sample test set of large phosphorescent compounds and demonstrate the substantial improvement that localization layers have on predicting their phosphorescence energies. Remarkably, the inferred localization weights have a strong relationship with the *ab initio* spin density of the singlet–triplet transition, and thus infer localities of the molecule that determine the spin transition, despite the fact that no direct electronic information was provided during training. The use of localization layers is expected to improve the modeling of many localized, non-extensive phenomena and could be implemented in any atom-centered neural network model.

Received 16th April 2021
Accepted 28th June 2021

DOI: 10.1039/d1sc02136b

rsc.li/chemical-science

Introduction

Molecules that are electronically excited by light or charge injection typically relax non-radiatively to their lowest-energy singlet or triplet excited states.¹ Transitions between these singlet and triplet states (intersystem crossing) can be enabled by spin–orbit interactions. Subsequent radiative relaxation from excited-singlet to ground-singlet states leads to fluorescence, which occurs over nanosecond timescales. In contrast, radiative transitions from excited-triplet to ground-singlet states are forbidden on account of the difference in spin multiplicity. This radiative process, known as phosphorescence, occurs over much longer timescales, generally on the order of 10^{-3} to 100 s.²

The dynamics of low-energy singlet and triplet states form the basis for applications including light-emitting electrochemical cells,³ chemical sensors,⁴ organic light-emitting diodes (OLEDs),^{5,6} and photovoltaics.⁷ For instance, a carefully designed thermally activated delayed fluorescence (TADF) process can be used to harvest energy from triplet to singlet states, raising the efficiency of an OLED to nearly 100%.⁸ As another example, simultaneous emission from triplet and singlet states in metal–organic polymers can be used to make white light LEDs.⁹ These examples illustrate the importance of a reliable prediction of triplet state emission energies across a large diversity of molecules capable of guiding experimental efforts toward future optoelectronic technologies.^{10–12}

^aTheoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA, 87545. E-mail: giff@lanl.gov

^bCenter for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, USA, 87545

^cCenter for Integrated Nanotechnologies, Los Alamos National Laboratory, Los Alamos, NM, USA, 87545

^dComputer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, USA, 87545. E-mail: nlubbers@lanl.gov

† Electronic supplementary information (ESI) available: Details regarding ML training procedures and localization metric η . *Ab initio* energies E_{S_0} , E_{T_1} , and ΔE as a function of number of atoms, distributions of molecule size for the held-out and extensibility test sets, histograms of percentage error in ΔE prediction on T_1 thermal conformers in the held-out and extensibility test

sets using HIPNN and HIP-loc, training and testing parity plots of predicted *versus* true ΔE on thermal conformers sampled around equilibria of S_0 and T_1 using HIPNN and HIP-loc, RMSD of optimized geometries using the HIP-loc T_1 potential and energy error plots at those geometries, absolute errors in ΔE as a function of number of atoms, parity plots of predicted *versus* true ΔE for the extensibility set categorized by chemical similarity, localization of singlet–triplet transition for select molecules of the extensibility set computed from DFT spin density and HIP-loc weights, conformation-dependent localization of singlet–triplet transitions in molecules with a single torsional angle, and molecular animations of torsional scans including that of the molecule in Fig. 5. See DOI: 10.1039/d1sc02136b

‡ These authors contributed equally to this work.



Computational screening using accurate *ab initio* calculations, such as density functional theory (DFT), can aid in molecular design.¹³ However, the computational costs of fully quantum mechanical approaches may restrict both the number and size of molecules that can be sampled efficiently. Machine learning (ML) has newfound relevance in quantum chemistry for accelerating simulations and providing predictions of *ab initio* quality.^{14–16} Here, we leverage ML techniques and build an accurate and extensible ML model for phosphorescence energy that is enabled by its ability to account for electron localization associated with the spin transition.

ML research has led to predictive models of various molecular properties,^{17–19} among which are general interatomic potentials of ground state energies.^{20–23} Such models can be trained to large datasets (*e.g.*, $\sim 10^5$ – 10^6 molecules) containing both equilibrium and off-equilibrium structures. This approach enables the construction of machine learned potential energy surfaces (PESs)^{24,25} for purposes such as dynamics²⁶ and geometry optimization.^{27,28} Data curation and model training are computationally demanding tasks,^{29,30} but once trained, such models not only generate high-fidelity *ab initio*-quality predictions at low computational costs but also infer relationships within high-dimensional data³¹ and are transferable to molecules outside of the training dataset.^{24,31,32} ML for ground state chemistry has been explored at an appreciable depth, achieving models of high-accuracy theories^{33–35} and extremely large systems.³⁶ However, its use for excited state processes^{37–39} such as phosphorescence is a new area of research.

The physics involved in describing an excited state transition is fundamentally different however from that of a ground state. For ground states, ML potentials incorporate the extensivity principle which defines the total energy of a molecule as a sum over individual atomic contributions.⁴⁰ This assumption is physically motivated and has been tested over a large chemical space, and provides the means for attaining transferable and extensible models.^{24,31,41} However, it breaks down in the case of electronic excitation energies, for which atoms may contribute to the transition energy in a disproportionate way. This disproportionality is evidenced by localized *versus* delocalized electronic transitions that are prevalent in chemical physics.^{42–45} In turn, the energy of an excitation does not scale strictly with the size of the molecule. Rather than using extensive predictions, it is reasonable to seek a method that can associate the energy of an excitation with specific regions of the molecule. Such an approach would have wide-ranging applicability since the notion of localization generalizes to many molecular phenomena and systems such as charged species and radicals, and can be potentially applied to a broad class of problems centered around energy and charge carrier transport in molecular materials and solids.

In this work, we develop a model based on the Hierarchical Interacting Particle Neural Network (HIPNN)⁴⁶ that accurately predicts the singlet–triplet energy gaps in a diverse chemical space of organic compounds at computational speeds that reach over one million times faster than the underlying *ab initio* method for single-point calculations. HIPNN has previously shown excellent performance for ground state energies,⁴⁶ partial

atomic charges,⁴⁷ and molecular dipole moments.⁴⁸ We present new atomistic localization layers that compute a weight for each atom corresponding to the atomic contribution to the molecule's energy gap. The mathematics of these new layers can be likened to the statistical mechanics of a chemical potential, or to the technology of attention mechanisms in the neural network literature.^{49,50} The localization layers improve phosphorescence energy predictions on an out-of-sample test set of molecules that are larger than those found in the training set and that are known to be phosphorescent from experiment while qualitatively inferring the changes to the electron density due to the transition as shown by direct comparison to reference quantum mechanical spin density calculations. This result is remarkable given that DFT spin density was not provided as a target during training and highlights the added benefit that electronic information could have toward ML models of localized, non-extensive molecular phenomena.

Methods

We start with a brief review of how energy is computed with ML interatomic potentials, including the extensive HIPNN model,⁴⁶ and describe the new methods in this work. The input descriptors of the molecules are atomic coordinates (R_i) and atomic numbers (Z_i), which together define the molecular geometry (R). In extensive HIPNN and several other atomistic neural networks,^{24,40,46} target energy for geometry R is computed as a sum over atomic energy contributions,

$$E = \sum_{i=1}^N \varepsilon_i \quad (1)$$

where E is the energy of the molecule, ε_i is the energy contribution of the i -th atom, and i spans over the total number of atoms in the molecule N . Local atomic contributions are predicted as a linear combination of a high-dimensional learned feature vector $z_{i,a}$ using a vector of learned parameters t_a and a bias b , where a is a feature index:

$$\varepsilon_i = \sum_{a=1}^{N_{\text{features}}} t_a z_{i,a} + b \quad (2)$$

Note that the ε_i do not represent a direct physical quantity, but rather a factorization or ansatz form of the total energy E , representing the notion that the potential may be taken as a sum of contributions over latent variables ε_i in the model, and it is a useful ansatz for an extensive total energy. For a single conformation of a system, many values of the ε_i are compatible with the total energy. However, the machine learning algorithm learns a model of ε_i so that E will fit a large database of calculations simultaneously. The high-dimensional feature vector $z_{i,a}$ contains information collected from the local environment of that atom, and is also learned, but is less readily interpretable than ε_i .

This model can be applied in either the singlet ground state (S_0) or first triplet excited state (T_1). In addition to being trained to energy, our models are also trained to atomic forces. The



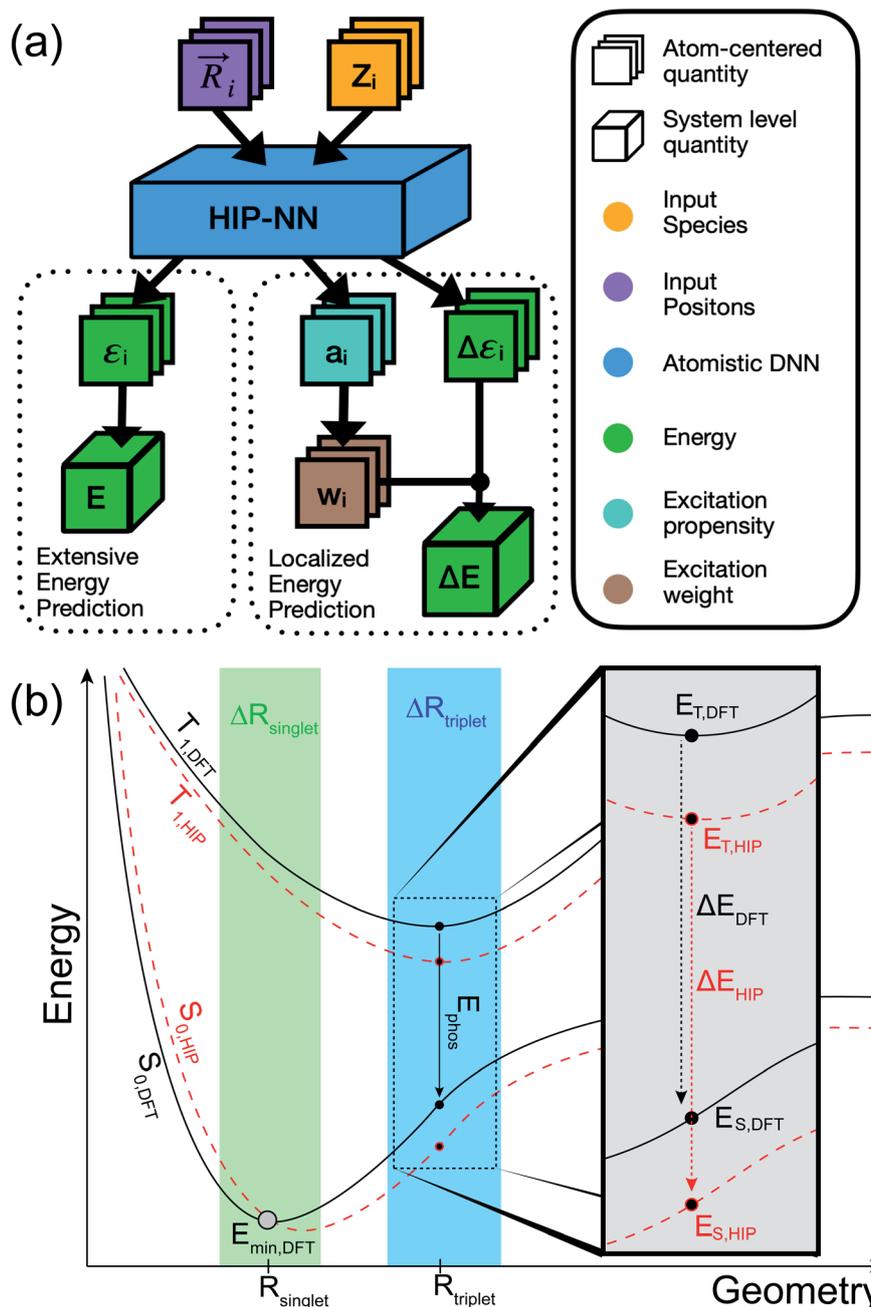
inclusion of atomic forces as targets has been shown to improve energy predictions.^{51,52} Throughout the paper, we refer to the extensive HIPNN model that solely employs eqn (1) to compute energy as HIPNN (Scheme 1a, left).

In HIPNN, singlet (E_S) and triplet (E_T) energies are determined from independent Self-Consistent Field (SCF) calculations, and

both E_S and E_T are learned independently using separate linear prediction with the form of eqn (1). Their difference,

$$\Delta E = E_T - E_S \quad (3)$$

is the singlet–triplet energy gap (S_0 – T_1 gap) within the Δ SCF method (Scheme 1b). For instance, near triplet state



Scheme 1 (a) Traditional workflows for potentials with atomistic deep neural networks (DNNs) are based on an extensivity assumption (top left) that predictions should be summed over all atoms in the system. With localization layers (top right), atomistic neural networks can predict physical properties that are weighted by location and do not scale with the size of the molecule. This allows training to physical effects such as triplet excitations (which generally do not scale with molecule size), and identification of molecular regions that contribute to the investigated effect from a data-driven perspective. (b) Schematic of the S_0 and T_1 PESs depicted for reference DFT (black, solid lines) and HIPNN-based ML predictions (red, dashed lines). Transition energy relevant to phosphorescence ($\Delta E = E_{\text{phos}}$) is approximated as a difference between T_1 and S_0 energies with the Δ SCF approach. This work introduces a localized variant of HIPNN, HIP-loc, to model ΔE .



equilibrium, $\Delta E = E_{\text{phos}}$ represents the phosphorescence energy as calculated with ΔSCF . We write the difference in eqn (3) as a general ΔE to reflect that our method may be applicable to other localized, non-extensive properties, such as energies of anions, cations, or other excited states. The central motivation of our work is that eqn (1), which is a sum over atom-centered energies, is inadequate in predicting ΔE (eqn (3)) for systems in which the electron density is spatially localized only on certain atoms. In other words, the extensivity approximation breaks down due to the absence of a well-defined scaling of ΔE with the number atoms (Fig. S1†). The main contribution of our work is to develop a new approach to predict ΔE while accounting for this localization. This accomplishes two purposes: (1) The predicted ΔE will not scale intrinsically with the size of the system and (2) it gives rise to interpretable predictions regarding where in the molecule the singlet and triplet wave functions differ. The localized HIPNN model will be referred to as HIP-loc (Scheme 1a, right). We emphasize that the method developed here could also be applied to other atomistic neural networks that are trained using gradient methods.^{24,53,54}

We described how one could learn E_S and E_T and compute ΔE following their predictions (eqn (3)). This approach was carried out for the HIPNN model. Alternatively, one could learn E_S and ΔE , with E_T defined as $E_T = E_S + \Delta E$. This definition represents E_T as a sum of an extensive (E_S) and non-extensive (ΔE) property, the latter of which motivates the HIP-loc model. Atomic forces on the singlet state are the same as those used in the HIPNN model, $-\frac{dE_S}{dR_i} = F_{S,i}$ whereas the forces associated with ΔE are calculated as $-\frac{d\Delta E}{dR_i} = -\frac{dE_T}{dR_i} + \frac{dE_S}{dR_i} = \Delta F_i = F_{T,i} - F_{S,i}$. We now introduce the principal difference between HIPNN and HIP-loc (Scheme 1a). Instead of modeling ΔE as a simple sum over atom-centered energies (eqn (1)), we model it as a non-extensive quantity by weighting atomic energies by normalized weights w_i

$$\Delta E = \sum_{i=1}^N w_i \Delta \varepsilon_i \quad (4)$$

This form is physically motivated by the fact that the electron density may be localized on certain atoms. We show that the inclusion of these weights is instrumental in accurately predicting ΔE , as opposed to eqn (1), which assumes equal atomic contributions. We use a softmax function for w_i ,

$$w_i = \frac{e^{a_i}}{\sum_{i=1}^N e^{a_i}} \quad (5)$$

where a_i , which we term the excitation propensity on the i -th atom, is a quantity learned in the training process. Eqn (5)

ensures normalization of the weights, $\sum_{i=1}^N w_i = 1$. We predict

both $\Delta \varepsilon_i$ and a_i analogously to eqn (2), with linear layers applied to the learned features. Because the weights sum to one, the predicted ΔE is bounded by the extrema of the $\Delta \varepsilon_i$, a local

quantity that does not intrinsically scale with system size; eqn (4) may be better thought of as a weighted average than as a sum.

The form of eqn (4) and (5) are similar to those of attention models^{49,50} and elementary statistical mechanical models. In a statistical mechanical analogy, the propensity a_i plays the role of a (negative) chemical potential, whereas the weights w_i are interpreted as a probability derived from the propensity using a Boltzmann weighting, and the excitation corresponds to an observable of the system. Such an analogy gives rise to intuition about the relationship between the propensity and the weights: if the propensities are narrowly distributed, the weights will be roughly evenly distributed across the molecule. If a few propensities are larger than the rest, the weights will be concentrated on those few atoms. In our results, we show that these localization weights effectively infer the region of the triplet excitation, as determined by *ab initio* spin density calculations.

The molecules of the dataset were randomly sampled from the GDB13 dataset⁵⁵ and contain C, H, N, O, S, and Cl atoms. *Ab initio* computations were performed using the ωb97x density functional⁵⁶ and 6-31g* basis set⁵⁷ with Gaussian 16.⁵⁸ All molecules were optimized in vacuum in the T_1 state. Thermal conformers ($T = 300$ K) near the minimum of T_1 were obtained with a normal mode sampling scheme.²⁴ Doing so enables learning PESs at molecular conformations near points of stability (Scheme 1b). Subsequent single-point calculations of the entire dataset, including optimized and thermal conformers, were carried out to determine E_S and E_T . The entire dataset consisted of approximately $\sim 702\text{k}$ structures with median molecule size of 8 heavy atoms (excluding H) and 16 total atoms with a maximum molecule size of 12 heavy atoms and 30 total atoms (Fig. S2†). A more detailed breakdown of the dataset, including the number of molecules sampled from each GDB dataset, is available in Table S1.†

The trained models were also tested on a challenging set of large aromatic molecules known to be phosphorescent from experiment with median molecule size of 17 heavy atoms (excluding H) and 27 total atoms with a maximum molecule size of 31 heavy atoms and 51 total atoms (Fig. S2†). The compounds of this extensibility set were taken from ref. 59–62 and make up 35 unique compounds and 912 thermal conformers. A sample set of the training and test set compounds is shown in Fig. 1. Further details regarding the neural network architecture and model training can be found in the ESI† under the section labeled Neural Network Architecture and Training Procedure. It is worth noting that a full end-to-end ML-based approach is ultimately desirable for molecular design workflows. Ideally, the ML model should optimize the triplet geometry, which makes up a significant portion of the computational time needed to predict emission energies from first principles. In this work, we concentrate on using localization layers in the modeling of localized, non-extensive properties with emphasis on phosphorescence energies as the application of interest. Thus, we eliminate the error of an ML-based optimization in our results, using geometries obtained by normal mode sampling of conformations near DFT-optimized structures of T_1 . We do



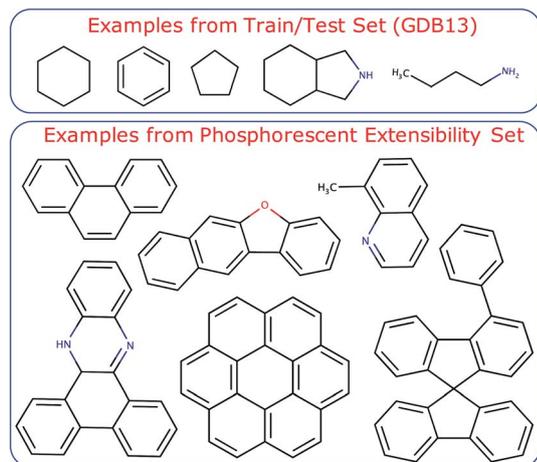


Fig. 1 Sample set of the molecules used in this study. (top panel) Select molecules from the train/test set ($\sim 702k$ structures in total). (bottom panel) Select molecules from the phosphorescent test set (912 structures in total). The test set molecules are significantly larger ($\sim 2\times$ larger on average) than the training set molecules and is often referred to as the phosphorescent extensibility set.

present encouraging results with T_1 optimized with ML in ESI,[†] including the singlet–triplet energy gaps and root-mean-square deviation (RMSD) of the ML-optimized geometries compared with DFT-based optimization.

Results and discussion

Fig. 2 presents the energy gaps that correspond to molecular geometries near the equilibrium of T_1 . These geometries are relevant for phosphorescence and will be the focus of our discussion. HIPNN and HIP-loc achieve comparable prediction accuracies on the held-out test set (Fig. 2a and b). Both models exhibit root-mean-square errors (RMSEs) of ~ 4 kcal mol⁻¹ and mean-absolute errors (MAEs) of ~ 2 kcal mol⁻¹ (or ~ 0.1 eV), which are within the error of the DFT method when compared to experimental transition energies.⁶³ These predictions correspond to a median percent error of $\sim 5\%$ (Fig. S3[†]). Thus, both HIPNN and HIP-loc attain experimentally informative predictions of E_{phos} on the held-out test set. Training and testing results for all geometries sampled near the equilibria of S_0 and T_1 are shown in Fig. S4 and S5.[†] We also present results of RMSD and energy predictions at geometries optimized with respect to the HIP-loc T_1 PES in Fig. S6 and S7.[†]

Despite HIPNN's success of predicting ΔE for molecules with sizes comparable to those found in the training set, its performance substantially worsens for phosphorescent molecules of the extensibility set that consist of on average $\sim 2\times$ the number of atoms (Fig. S2[†]). Yet we show that HIP-loc remedies this issue. HIPNN erroneously predicts negative ΔE energies (*i.e.*, T_1 is lower in energy than S_0) for many molecules, contradicting reference results (Fig. 2c). This result is particularly troublesome given that these are phosphorescent molecules for which DFT correctly predicts T_1 to be higher in energy than S_0 . Altogether, HIPNN's performance on the extensibility set is quite

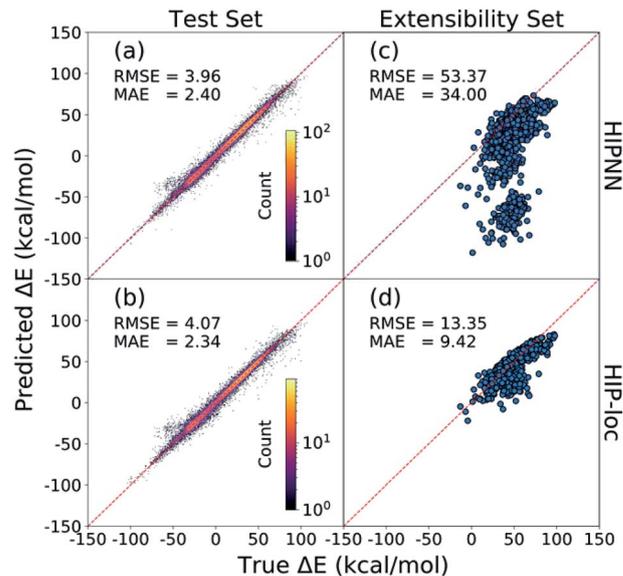


Fig. 2 Parity plots of predicted versus true ΔE energy on the held-out test set using (a) HIPNN and (b) HIP-loc. (c) and (d) depict the same correlation as (a) and (b) but for the extensibility set. Results are for molecules sampled around the equilibrium of T_1 . HIPNN and HIP-loc achieve comparable results on the held-out test set, but HIP-loc significantly outperforms HIPNN on the extensibility set compromising molecules that are on average $\sim 2\times$ larger than those found in the training set, highlighting the advantage that HIP-loc has for extended systems. Prediction errors are expressed in root-mean-square error (RMSE) and mean-absolute error (MAE).

poor with a RMSE of ~ 53 kcal mol⁻¹ or ~ 2.3 eV, which is well outside the error commonly expected between *ab initio* and experimental values.⁶³ HIPNN's predictions are especially poor for molecules composed of more than 35 atoms (Fig. S8 and S9[†]). In contrast, Fig. 2d shows HIP-loc significantly outperforming HIPNN on the extensibility set. The RMSE of HIP-loc results on the extensibility set is ~ 13 kcal mol⁻¹, amounting to an $\sim 4\times$ improvement and median percent error of $\sim 13\%$ (Fig. S3[†]). The erroneous negative ΔE energies, predicted by HIPNN (Fig. 2c), are all but eliminated with HIP-loc (Fig. 2d and S10[†]).

HIP-loc strongly outperforms HIPNN, and the only difference between the two models is the usage of localization layers to model the non-extensive part of E_T (*i.e.*, ΔE) by weighting local atomic environments. The extensibility of the model is substantially improved, but relatively large disagreement between HIP-loc and DFT is observed for the largest molecules in the extensibility set comprising several bonded aromatic fragments (Fig. S11[†]). Nevertheless, HIP-loc's stronger performance demonstrates the importance of model engineering to account for the localized nature of spin transitions. Parity plots for all molecules of the extensibility set, categorized by chemical similarity, are shown in Fig. S11 through S16.[†]

Additionally, we investigated whether the learned localization weights in HIP-loc have some physical significance. We compared the localization weights to DFT spin density differences between the T_1 and S_0 states. In order to compare the two



methods, DFT spin density was approximated as an atom-centered quantity, which we obtained using the Hirshfeld charge partitioning scheme.⁶⁴ Fig. 3a visualizes this comparison for a subset of representative molecules selected from a random sample of the held-out test set. Remarkably, there is qualitative agreement between HIP-loc and DFT for most molecules, suggesting that the HIP-loc localization weights provide physical insight. Moreover, the correspondence between the quantum mechanical and inferred localizations is somewhat correlated with the accuracy in predicted energies; the absolute error in ΔE for the last molecule shown in Fig. 3a is relatively high and, correspondingly, the atom-centered DFT spin density does not as strongly resemble the HIP-loc weights as compared to the other molecules shown. We also find a rough, yet highly significant correlation observed between these learned weights and the DFT spin densities (Fig. 3b). It is important to note that the results of Fig. 3a and b are not meant to assert that learned weights are intrinsically related to spin densities in a rigorous

way, especially since they are atom-centered quantities and charge partitioning is ambiguous, but rather that inferring localization through the use of energies and forces across a large dataset leads to similar assignments of localization.

We also analyzed the agreement between the *ab initio* and inferred localization over a large sample of the held-out test set. For this analysis, we made use of a localization metric η , defined as the ratio of the distance between the centers of localization computed using HIP-loc weights ($\vec{r}_{\text{HIP-loc}}$) and DFT spin density (\vec{r}_{DFT}) to the radius of gyration (R_g), which quantifies the spatial size of the molecule: $\eta = |\vec{r}_{\text{HIP-loc}} - \vec{r}_{\text{DFT}}|/R_g$. For $\eta \ll 1$, the centers of localizations are in close proximity to one another, whereas $\eta \sim 1$ signifies that the centers of localization differ by approximately the radius of the molecule and therefore there is very little or no agreement in the predicted localization. See ESI† under Localization Metric for more details. Fig. 3c shows a histogram of η computed for our sample set. The distribution is concentrated at low η (mean = 0.15), indicating that the

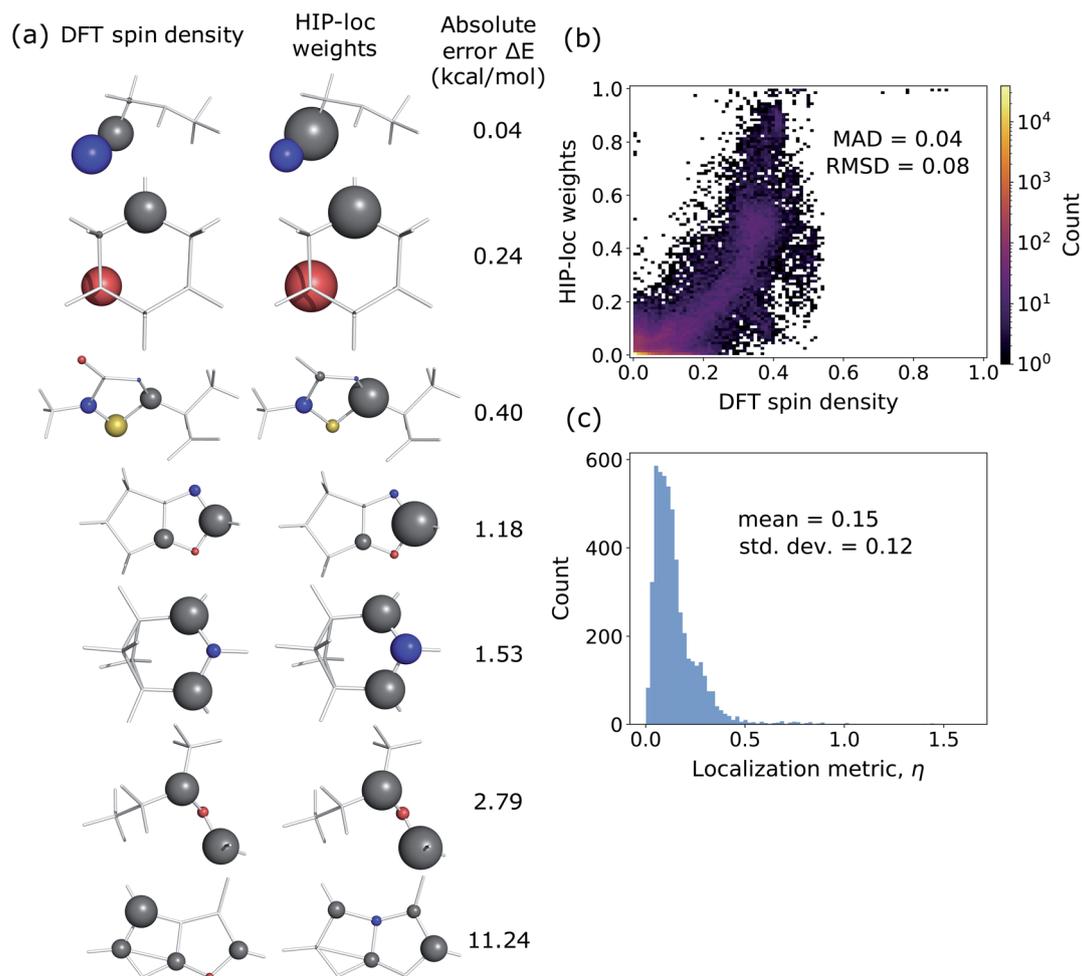


Fig. 3 (a) Comparison of DFT spin density and HIP-loc localization weights for select molecules in the held-out test set. (b) Correlation between HIP-loc weights versus atom-centered DFT spin density. (c) Histogram of the localization metric η , which quantifies the agreement between the *ab initio* (DFT) and inferred (ML) centers of localization and where $\eta \ll 1$ signifies strong agreement. Energy ΔE is accurately predicted for the majority of molecules and jointly, HIP-loc weights closely resemble DFT spin density. For the final molecule with less accurate ΔE , HIP-loc weights and DFT spin density are more dissimilar. Deviations are expressed in mean-absolute deviation (MAD) and root-mean-square deviation (RMSD).



center of HIP-loc's inferred localization is in close proximity to that determined from DFT for the majority of compounds.

This agreement is also observed for molecules of the phosphorescent extensibility set. Similar to Fig. 3a and 4a visualizes the similarities between DFT and HIP-loc for two example molecules that perform particularly poorly with HIPNN. Additional visualizations of a similar nature are available in ESI Fig. S17 through S20.† The spatial localities associated with the singlet–triplet transitions in these molecules are confined to a relatively small number of atoms compared to the molecules' total number of atoms. This result suggests, albeit not too surprisingly, that the advantages provided by HIP-loc are accentuated in the case of strongly localized transitions. In order to fully probe the improvement provided by HIP-loc for modeling localized *versus* delocalized transitions, we study the participation ratio (PR). The PR is described as follows: given an N -body wavefunction expanded in terms of atom-localized states $|\psi\rangle = \sum_{i=1}^N c_i |\psi_i\rangle$ with expansion coefficients c_i , the PR is

expressed as $\left(\sum_{i=1}^N |c_i|^2\right)^2 / \sum_{i=1}^N |c_i|^4$, ranging from 1 (fully localized to a single atom) to N (equally delocalized across all atoms).⁴² Although we are not working with wavefunctions, we apply the concept of PR to estimate the number of atoms involved in the singlet–triplet transition based on the *ab initio* (DFT) spin density and thereby quantify the degree of localization. The squares of the expansion coefficients are approximated from the atom-centered DFT spin density (q_i):

$PR_{\text{DFT}} = \left(\sum_{i=1}^N q_i\right)^2 / \sum_{i=1}^N q_i^2$. The PR is complimentary to the previously utilized localization metric η . The latter quantifies the agreement in proximity of the *ab initio* and ML-inferred centers of localization, whereas the former estimates the number of atoms involved in the transition and is therefore a more appropriate measure of quantifying the degree to which a transition is localized.

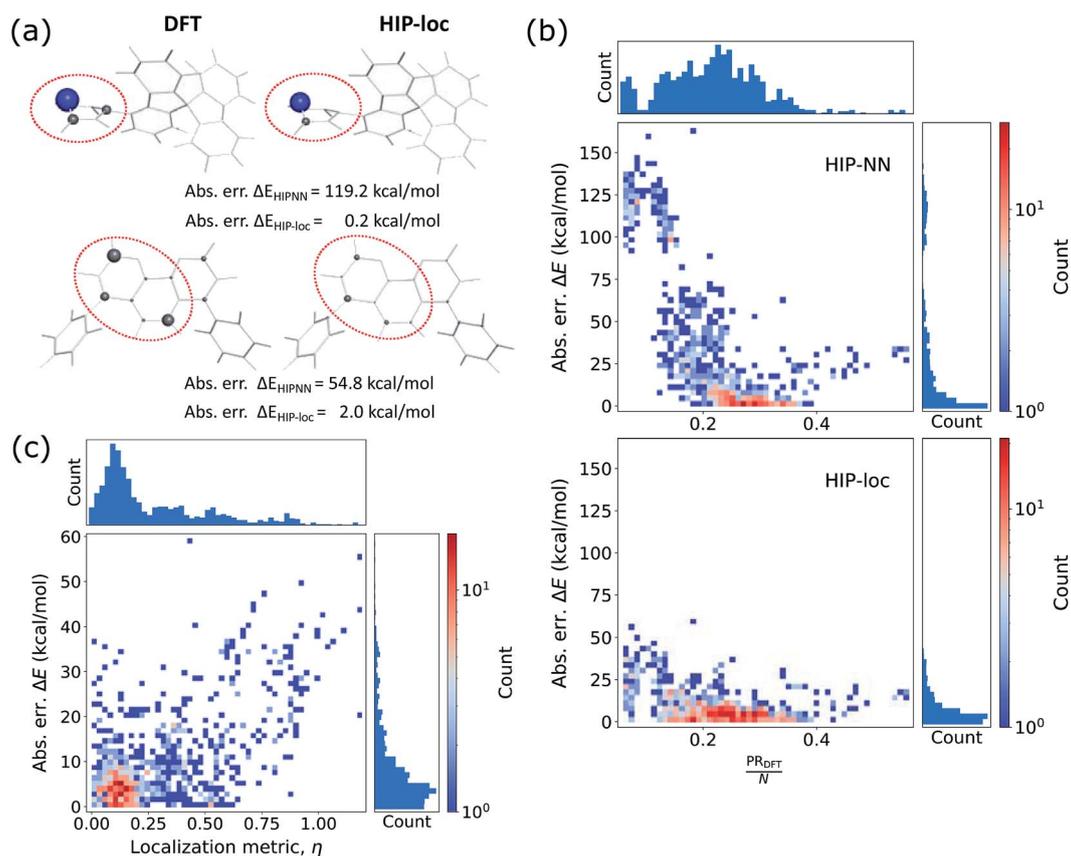


Fig. 4 (a) Comparison of DFT spin density and HIP-loc localization weights for select molecules of the phosphorescent extensibility set that perform poorly with HIPNN. Predominant regions of the molecules associated with the spin transitions are circled. Also shown are predicted absolute errors (abs. err.) in ΔE energies using HIPNN and HIP-loc. (b) Relationship between absolute error in ΔE energy and degree of localization for all molecules of the extensibility set as determined by the ratio of the reference DFT participation ratio divided by the total numbers of atoms in the molecule (PR_{DFT}/N). Results shown are those of HIPNN (top panel) and HIP-loc (bottom panel). HIPNN performs notably worse in the regime of strong localization (PR_{DFT}/N close to $\sim 1/N$) as compared to the regime of delocalized transitions (larger PR_{DFT}/N), whereas for HIP-loc, the ΔE prediction is less dependent on the degree of localization. Compared to HIPNN, HIP-loc is particularly more accurate in the regime of strong localization. (c) Relationship between HIP-loc's error in ΔE versus localization metric η for all molecules of the extensibility set. Localities inferred by HIP-loc are in qualitative agreement with DFT and concomitantly the transition energy ΔE is predicted to better accuracy in the regime of low η .



Fig. 4b shows the relationship between absolute error in ΔE and the degree of localization as computed by PR/N. For molecules in which the density is dispersed more homogeneously across the atoms (or larger PR_{DFT}/N edging closer to ~ 1), energy is accurately predicted using eqn (1) of HIPNN. In fact, HIPNN and HIP-loc achieve comparable performance in this regime. However, for molecules in which the density is strongly localized on only a small handful of atoms (or PR_{DFT}/N closer to $\sim 1/N$), HIP-loc is superior. This result underscores the main motivation of our work in that HIPNN does not accurately predict the singlet–triplet energy gap for strongly localized transitions, whereas HIP-loc's localization weights allow the model to weight regions of the molecule that are attributed to the transition, resulting in better energy prediction.

Finally, in order to relate the agreement in the centers of localization to the accuracy in predicted energy, we also show absolute error in ΔE versus η (Fig. 4c). The distribution is concentrated in the regime of low η and low ΔE error, but in contrast to the held-out test set (Fig. 3c), there is also a significant number of molecules that lie in the regime of high η . The distribution of ΔE errors in this high η regime is sporadic and extends to relatively high error. Altogether, these results suggest that robust energy prediction is generally improved when the inferred localization determined by the HIP-loc weights more closely resembles the localization determined by the quantum mechanical spin density.

To further demonstrate HIP-loc's feature of localization without changing stoichiometry (and therefore drastically changing total molecular energy), we examined the predicted energy and concomitant variation in localization while scanning a single torsional angle. The set of molecules investigated consists of three to five six-membered aromatic rings and a central carbon–carbon single bond. A relaxed torsional scan over the dihedral angle around the carbon–carbon bond is performed and HIPNN and HIP-loc predictions are made for each conformation. The PR is calculated for both the DFT spin density and HIP-loc weights in order to compare the overall localization using both methods. The analogous form for $\text{PR}_{\text{HIP-loc}}$ uses HIP-loc weights (w_i) in lieu of atom-centered spin

density, that is, $\text{PR}_{\text{HIP-loc}} = \left(\sum_{i=1}^N w_i \right)^2 / \sum_{i=1}^N w_i^2$.

Our conformational analysis shows that when errors in ΔE predictions are relatively low across the conformational space, localization is accurately predicted by means of the PR. Rotating the molecule's dihedral angle changes its aromaticity and shifts spatial localization of electron density. As a result, the S_0 and T_1 PESs, as well as the ΔE gap, vary with conformation. Fig. 5a shows absolute error in ΔE as a function of dihedral angle for a representative molecule containing three six-membered aromatic rings, computed with HIPNN and HIP-loc. HIP-loc outperforms HIPNN for all scanned geometries. Additionally, there is qualitative agreement in the trends of the PRs computed using HIP-loc weights and DFT spin density (Fig. 5b). However, HIP-loc infers more delocalized transitions compared to DFT (PR of ~ 13 versus ~ 6), but in spite of that, for both methods the PR in the planar structure (dihedral of 0° and 180°)

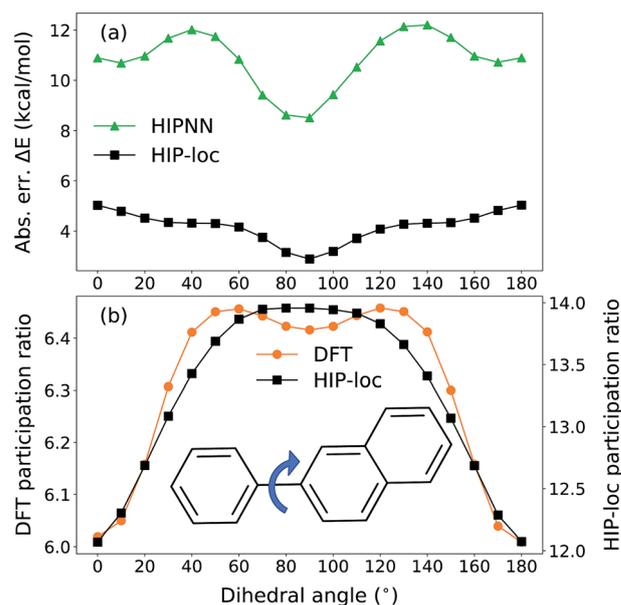


Fig. 5 Conformational scan over the dihedral angle around the single carbon–carbon bond (blue arrow) in a representative molecule. (a) Absolute error (abs. err.) in predicted ΔE energy gap computed with HIPNN and HIP-loc. (b) Participation ratio (PR) estimated using atom-centered DFT spin density (left axis) and HIP-loc weights (right axis) as a function of dihedral angle. HIP-loc significantly outperforms HIPNN in predicting ΔE for all scanned geometries. Qualitative agreement in the trends of PRs computed with DFT and HIP-loc is also observed.

is delocalized across atoms on each ring, whereas the PR in the non-planar structure (dihedral of 90°) is localized to slightly more atoms exclusively on the larger ring. The net effect is an increase in PR for the non-planar conformation. An animation in ESI† showing a torsional scan of the molecule in Fig. 5 illustrates this point. Altogether, we find a correspondence between the accuracy of energy prediction and the progression in the degree of localization, quantified by PR. These observations are consistent for all the molecules studied (Fig. S21 through S25†).

Conclusions

In conclusion, machine learning (ML) is becoming an integral part of physical chemistry research and has already made substantial advances in the development of fast and transferable ML potentials for ground state dynamics. Ground state energy can be classified as an extensive property that is represented as a sum over individual atomic contributions whereas electronic excitation energies are best categorized as localized, non-extensive properties that depend on subsets of atoms. In this work, we utilized ML for advancing excited state electronic structure modeling by training a Hierarchical Interacting Particle Neural Network (HIPNN) to predict phosphorescence energy, defined as the gap between the lowest energy triplet and singlet states, at computational speeds that reach over one million times faster than the underlying *ab initio* method for single-point calculations. Our work improves upon the original



version of HIPNN that is based on the extensivity principle, particularly in the case of large aromatic compounds for which their singlet–triplet transition energies do not scale strongly with molecule size.

Our main contribution is a new set of localization layers for learning target excitation energies. These localization layers do not depend on the detailed structure of the underlying model and could be implemented in any atom-centered neural network. Combined with HIPNN, the new approach (denoted HIP-loc) defines energy as a weighted sum, where inferred localization weights assigned to the atoms determine the atoms' contribution to the target energy (eqn (4)). This minor yet profound modification substantially improved prediction quality on a more challenging set of experimentally verified phosphorescent compounds that consist of, on average, $\sim 2\times$ more atoms than the molecules found in the training set. Extensibility is an important practical advantage of the model because it allows screening large molecules for which *ab initio* calculations are computationally prohibitive. Moreover, the superior energy prediction of HIP-loc over HIPNN was not readily visible on the held-out test set, but only the phosphorescent extensibility set in molecules with strongly localized transitions, showing that the extensibility test provides important characterization of the model's performance. We achieve RMSEs of ~ 4 kcal mol $^{-1}$ ($\sim 5\%$ error) on the held-out test set and ~ 13 kcal mol $^{-1}$ ($\sim 13\%$ error) on the phosphorescent extensibility set (Fig. 2). The physical significance of HIP-loc's localization weights, to our surprise, is that they qualitatively correlate to the quantum mechanical spin density (Fig. 3 and 4). Thus, the model inferred localities of the electron density associated with the singlet–triplet transition in order to make more accurate and transferable prediction of the singlet–triplet energy gap. This result is remarkable given that DFT spin density was not provided as a target during training. Instead, the neural network was provided with somewhat limited information, yet discovered this result nonetheless with a modest change in how target energy is calculated that effectively took into account the relative contribution of the atoms.

The results and performance of the new HIP-loc model lead us to conjecture that predicting localized molecular properties are likely to improve if reference electronic properties (*e.g.*, electron densities) are used explicitly as targets during training. This idea sets the stage for many interesting applications of ML in physical chemistry, where learned properties are dependent on, or correlated to, the spatial distribution of the many-electron wavefunction. We also envision applying localization layers to other localized molecular phenomena and systems such as anions, cations, radicals, or other excited states. Additionally, further improvement of excited state ML potentials for geometry optimization could lend itself useful for a full end-to-end ML-based tool for phosphorescent materials screening.

Data availability

Nebgen, Benjamin (2021): Dataset of Singlet and Triplet Energies and Forces for Organic Molecules. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.14736570.v1>.

Author contributions

A. E. S., L. L., and B. J. G. curated datasets, performed *ab initio* calculations, analyzed results, made figures, and began the initial draft of the manuscript. L. L. generated the original dataset and trained preliminary models. N. L. developed the HIP-loc model formulation, wrote the corresponding code, and trained models. R. A. M. performed dihedral scan calculations. J. S. S. and K. B. provided valuable discussion and provided directions for analysis, including the idea to analyze DFT spin density. A. E. S., S. T., N. L., and B. J. G. brought the manuscript to a complete form. S. T. and B. N. conceived the task to model excited spin states. All authors contributed to the editing of the final manuscript. B. N., S. T., N. L., and B. J. G. supervised the research.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The work at Los Alamos National Laboratory (LANL) was supported by the LANL Directed Research and Development (LDRD) funds and performed in part at the Center for Nonlinear Studies (CNLS) and the Center for Integrated Nanotechnologies (CINT), U.S. Department of Energy, Office of Science user facilities. This research used resources provided by the LANL Institutional Computing (IC) Program as well as the LANL Darwin Cluster. LANL is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy (Contract No. 89233218NCA000001).

References

- 1 G. N. Lewis and M. Kasha, *J. Am. Chem. Soc.*, 1944, **66**, 2100–2116.
- 2 N. J. Turro, V. Ramamurthy and J. C. Scaiano, *Modern Molecular Photochemistry of Organic Molecules*, 2017.
- 3 R. D. Costa, E. Orti, H. J. Bolink, F. Monti, G. Accorsi and N. Armaroli, *Angew. Chem., Int. Ed.*, 2012, **51**, 8178–8211.
- 4 Q. Zhao, C. Huang and F. Li, *Chem. Soc. Rev.*, 2011, **40**, 2508.
- 5 *Highly Efficient OLEDs with Phosphorescent Materials*, ed. H. Yersin, Wiley, 1st edn, 2007.
- 6 Y. Tao, C. Yang and J. Qin, *Chem. Soc. Rev.*, 2011, **40**, 2943.
- 7 M. J. Currie, J. K. Mapel, T. D. Heidel, S. Goffri and M. A. Baldo, *Science*, 2008, **321**, 226–228.
- 8 Z. Yang, Z. Mao, Z. Xie, Y. Zhang, S. Liu, J. Zhao, J. Xu, Z. Chi and M. P. Aldred, *Chem. Soc. Rev.*, 2017, **46**, 915–1016.
- 9 C.-X. Sheng, S. Singh, A. Gambetta, T. Drori, M. Tong, S. Tretiak and Z. V. Vardeny, *Sci. Rep.*, 2013, **3**, 2653.
- 10 S. Haneder, E. Da Como, J. Feldmann, J. M. Lupton, C. Lennartz, P. Erk, E. Fuchs, O. Molt, I. Münster, C. Schildknecht and G. Wagenblast, *Adv. Mater.*, 2008, **20**, 3325–3330.
- 11 S. Mukherjee and P. Thilagar, *Chem. Commun.*, 2015, **51**, 10988–11003.



- 12 Z. An, C. Zheng, Y. Tao, R. Chen, H. Shi, T. Chen, Z. Wang, H. Li, R. Deng, X. Liu and W. Huang, *Nat. Mater.*, 2015, **14**, 685–690.
- 13 J. Hafner, C. Wolverton and G. Ceder, *MRS Bull.*, 2006, **31**, 659–668.
- 14 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, 2018, **559**, 547–555.
- 15 P. O. Dral, *J. Phys. Chem. Lett.*, 2020, **11**, 2336–2347.
- 16 O. A. von Lilienfeld and K. Burke, *Nat. Commun.*, 2020, **11**, 4895.
- 17 L. Ward, A. Agrawal, A. Choudhary and C. Wolverton, *npj Comput. Mater.*, 2016, **2**, 16028.
- 18 A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi and M. Ceriotti, *Sci. Adv.*, 2017, **3**, e1701816.
- 19 F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley and O. A. von Lilienfeld, *J. Chem. Theory Comput.*, 2017, **13**, 5255–5264.
- 20 J. Behler, *J. Chem. Phys.*, 2016, **145**, 170901.
- 21 V. L. Deringer, M. A. Caro and G. Csányi, *Adv. Mater.*, 2019, **31**, 1902765.
- 22 T. Mueller, A. Hernandez and C. Wang, *J. Chem. Phys.*, 2020, **152**, 050902.
- 23 Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood and S. P. Ong, *J. Phys. Chem. A*, 2020, **124**, 731–745.
- 24 J. S. Smith, O. Isayev and A. E. Roitberg, *Chem. Sci.*, 2017, **8**, 3192–3203.
- 25 Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby and T. F. Miller, *J. Chem. Phys.*, 2020, **153**, 124111.
- 26 M. Gastegger, J. Behler and P. Marquetand, *Chem. Sci.*, 2017, **8**, 6924–6935.
- 27 H. Zhai and A. N. Alexandrova, *J. Chem. Theory Comput.*, 2016, **12**, 6213–6226.
- 28 K. Gubaev, E. V. Podryabinkin, G. L. W. Hart and A. V. Shapeev, *Comput. Mater. Sci.*, 2019, **156**, 148–156.
- 29 J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev and A. E. Roitberg, *J. Chem. Phys.*, 2018, **148**, 241733.
- 30 R. Jinnouchi, K. Miwa, F. Karsai, G. Kresse and R. Asahi, *J. Phys. Chem. Lett.*, 2020, **11**, 6946–6955.
- 31 K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller and A. Tkatchenko, *Nat. Commun.*, 2017, **8**, 13890.
- 32 D. Jha, L. Ward, A. Paul, W. Liao, A. Choudhary, C. Wolverton and A. Agrawal, *Sci. Rep.*, 2018, **8**, 17593.
- 33 S. Chmiela, H. E. Sauceda, K.-R. Müller and A. Tkatchenko, *Nat. Commun.*, 2018, **9**, 3887.
- 34 J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev and A. E. Roitberg, *Nat. Commun.*, 2019, **10**, 2903.
- 35 M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller and K. Burke, *Nat. Commun.*, 2020, **11**, 5223.
- 36 D. Lu, H. Wang, M. Chen, L. Lin, R. Car, W. E. W. Jia and L. Zhang, *Comput. Phys. Commun.*, 2021, **259**, 107624.
- 37 J. Westermayr and P. Marquetand, *Mach. Learn.: Sci. Technol.*, 2020, **1**, 043001.
- 38 J. Westermayr and P. Marquetand, *Chem. Rev.*, 2020, DOI: 10.1021/acs.chemrev.0c00749.
- 39 J. Westermayr, M. Gastegger and P. Marquetand, *J. Phys. Chem. Lett.*, 2020, **11**, 3828–3834.
- 40 J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401.
- 41 P. Rowe, V. L. Deringer, P. Gasparotto, G. Csányi and A. Michaelides, *J. Chem. Phys.*, 2020, **153**, 034702.
- 42 S. Tretiak and S. Mukamel, *Chem. Rev.*, 2002, **102**, 3171–3212.
- 43 T. E. Dykstra, E. Hennebicq, D. Beljonne, J. Gierschner, G. Claudio, E. R. Bittner, J. Knoester and G. D. Scholes, *J. Phys. Chem. B*, 2009, **113**, 656–667.
- 44 A. N. Panda, F. Plasser, A. J. A. Aquino, I. Burghardt and H. Lischka, *J. Phys. Chem. A*, 2013, **117**, 2181–2189.
- 45 B. J. Gifford, S. Kilina, H. Htoon, S. K. Doorn and S. Tretiak, *Acc. Chem. Res.*, 2020, **53**, 1791–1801.
- 46 N. Lubbers, J. S. Smith and K. Barros, *J. Chem. Phys.*, 2018, **148**, 241715.
- 47 B. Nebgen, N. Lubbers, J. S. Smith, A. E. Sifain, A. Lokhov, O. Isayev, A. E. Roitberg, K. Barros and S. Tretiak, *J. Chem. Theory Comput.*, 2018, **14**, 4687–4698.
- 48 A. E. Sifain, N. Lubbers, B. T. Nebgen, J. S. Smith, A. Y. Lokhov, O. Isayev, A. E. Roitberg, K. Barros and S. Tretiak, *J. Phys. Chem. Lett.*, 2018, **9**, 4495–4501.
- 49 D. Bahdanau, K. Cho and Y. Bengio, 2014, ArXiv14090473 Cs Stat.
- 50 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, 2017, ArXiv170603762 Cs.
- 51 A. M. Cooper, J. Kästner, A. Urban and N. Artrith, *npj Comput. Mater.*, 2020, **6**, 54.
- 52 J. S. Smith, N. Lubbers, A. P. Thompson and K. Barros, 2020, ArXiv200605475 Cond-Mat Physicsphysics Stat.
- 53 K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko and K.-R. Müller, *J. Chem. Phys.*, 2018, **148**, 241722.
- 54 B. Anderson, T.-S. Hy, R. Kondor, 2019, ArXiv190604015, Phys. Stat.
- 55 L. C. Blum and J.-L. Reymond, *J. Am. Chem. Soc.*, 2009, **131**, 8732–8733.
- 56 J.-D. Chai and M. Head-Gordon, *J. Chem. Phys.*, 2008, **128**, 084106.
- 57 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
- 58 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo,



- R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
- 59 M. Zander, *Angew. Chem., Int. Ed. Engl.*, 1965, **4**, 930–938.
- 60 I. Janić, in *Luminescence of Crystals, Molecules, and Solutions*, ed. F. Williams, B. Baron, M. Martens and S. P. Varma, Springer US, Boston, MA, 1973, pp. 213–218.
- 61 E. Clar and W. Schmidt, *Tetrahedron*, 1976, **32**, 2563–2566.
- 62 S. Thiery, D. Tondelier, C. Declairieux, B. Geffroy, O. Jeannin, R. Métivier, J. Rault-Berthelot and C. Poriol, *J. Phys. Chem. C*, 2015, **119**, 5790–5805.
- 63 D. Bousquet, R. Fukuda, D. Jacquemin, I. Ciofini, C. Adamo and M. Ehara, *J. Chem. Theory Comput.*, 2014, **10**, 3969–3979.
- 64 F. L. Hirshfeld, *Theor. Chim. Acta*, 1977, **44**, 129–138.

