

## Improving the accuracy of Møller-Plesset perturbation theory with neural networks

Robert T. McGibbon, Andrew G. Taube, Alexander G. Donchev, Karthik Siva, Felipe Hernández, Cory Hargus, Ka-Hei Law, John L. Klepeis, and David E. Shaw

Citation: [The Journal of Chemical Physics](#) **147**, 161725 (2017);

View online: <https://doi.org/10.1063/1.4986081>

View Table of Contents: <http://aip.scitation.org/toc/jcp/147/16>

Published by the [American Institute of Physics](#)

---

### Articles you may be interested in

[Cheap but accurate calculation of chemical reaction rate constants from ab initio data, via system-specific, black-box force fields](#)

[The Journal of Chemical Physics](#) **147**, 161701 (2017); 10.1063/1.4979712

[A general intermolecular force field based on tight-binding quantum chemical calculations](#)

[The Journal of Chemical Physics](#) **147**, 161708 (2017); 10.1063/1.4991798

[The BioFragment Database \(BFDdb\): An open-data platform for computational chemistry analysis of noncovalent interactions](#)

[The Journal of Chemical Physics](#) **147**, 161727 (2017); 10.1063/1.5001028

[The many-body expansion combined with neural networks](#)

[The Journal of Chemical Physics](#) **146**, 014106 (2017); 10.1063/1.4973380

[Extension of the D3 dispersion coefficient model](#)

[The Journal of Chemical Physics](#) **147**, 034112 (2017); 10.1063/1.4993215

[Intermolecular interactions in the condensed phase: Evaluation of semi-empirical quantum mechanical methods](#)

[The Journal of Chemical Physics](#) **147**, 161704 (2017); 10.1063/1.4985605

---



# Improving the accuracy of Møller-Plesset perturbation theory with neural networks

Robert T. McGibbon,<sup>1,a)</sup> Andrew G. Taube,<sup>1</sup> Alexander G. Donchev,<sup>1</sup> Karthik Siva,<sup>1</sup> Felipe Hernández,<sup>1</sup> Cory Hargus,<sup>1</sup> Ka-Hei Law,<sup>1</sup> John L. Klepeis,<sup>1</sup> and David E. Shaw<sup>1,2,a)</sup>

<sup>1</sup>*D. E. Shaw Research, New York, New York 10036, USA*

<sup>2</sup>*Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, USA*

(Received 1 June 2017; accepted 12 August 2017; published online 6 September 2017)

Noncovalent interactions are of fundamental importance across the disciplines of chemistry, materials science, and biology. Quantum chemical calculations on noncovalently bound complexes, which allow for the quantification of properties such as binding energies and geometries, play an essential role in advancing our understanding of, and building models for, a vast array of complex processes involving molecular association or self-assembly. Because of its relatively modest computational cost, second-order Møller-Plesset perturbation (MP2) theory is one of the most widely used methods in quantum chemistry for studying noncovalent interactions. MP2 is, however, plagued by serious errors due to its incomplete treatment of electron correlation, especially when modeling van der Waals interactions and  $\pi$ -stacked complexes. Here we present *spin-network-scaled MP2* (SNS-MP2), a new semi-empirical MP2-based method for dimer interaction-energy calculations. To correct for errors in MP2, SNS-MP2 uses quantum chemical features of the complex under study in conjunction with a neural network to reweight terms appearing in the total MP2 interaction energy. The method has been trained on a new data set consisting of over 200 000 complete basis set (CBS)-extrapolated coupled-cluster interaction energies, which are considered the gold standard for chemical accuracy. SNS-MP2 predicts gold-standard binding energies of unseen test compounds with a mean absolute error of 0.04 kcal mol<sup>-1</sup> (root-mean-square error 0.09 kcal mol<sup>-1</sup>), a 6- to 7-fold improvement over MP2. To the best of our knowledge, its accuracy exceeds that of all extant density functional theory- and wavefunction-based methods of similar computational cost, and is very close to the intrinsic accuracy of our benchmark coupled-cluster methodology itself. Furthermore, SNS-MP2 provides reliable per-conformation confidence intervals on the predicted interaction energies, a feature not available from any alternative method. © 2017 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). [<http://dx.doi.org/10.1063/1.4986081>]

## I. INTRODUCTION

The development of accurate and efficient quantum chemical methods for modeling intermolecular interactions is one of the primary goals of modern computational chemistry. These interactions play a central role throughout biology, materials science, and chemistry and drive a diverse set of phenomena, such as the noncovalent binding of drugs to proteins, the packing of crystals and self-assembly of nanomaterials, and the establishment of the bulk properties of liquids.<sup>1–3</sup>

Among the range of methods available, second-order Møller-Plesset perturbation (MP2) theory stands out for its principled physical description of intermolecular interactions and relatively modest computational cost.<sup>4</sup> It is the simplest quantum chemical method that contains an *ab initio* description of each of the essential physical effects for accurately modeling noncovalent interactions; although Hartree-Fock

(HF) and Kohn-Sham density functional theory (DFT) can account for electrostatic, polarization, and exchange interactions, neither method is currently able to capture long-range dispersion effects without additional cost or parameterization. In contrast, MP2, the most efficient wavefunction-based *ab initio* electronic-structure method to directly, if approximately, include dynamical correlation, describes dispersion interactions explicitly. Modern techniques such as density fitting/resolution of the identity (DF-MP2/RI-MP2), local correlation (LMP2, TRIM-MP2), and their combination (DF-LMP2, RI-TRIM) have, moreover, considerably improved the computational performance of MP2 such that systems with hundreds of atoms can now be handled routinely.<sup>5–10</sup>

Nevertheless, MP2's incomplete treatment of electron correlation sometimes results in significant errors. In the case of intermolecular interactions, the largest errors are often found in systems with  $\pi$ - $\pi$  stacking and those that are dominated by dispersion forces.<sup>11,12</sup> Relative to highly accurate coupled-cluster results, for example, MP2 overestimates the binding energy of sandwich and T-stacked benzene dimers by 2× and 1.3×, respectively.<sup>13</sup> More accurate calculations can, however, be quite computationally expensive. The cost of

<sup>a)</sup> Authors to whom correspondence should be addressed: Robert.McGibbon@DEShawResearch.com, Telephone: (212) 478-0915, Fax: (212) 845-1915 and David.Shaw@DEShawResearch.com, Telephone: (212) 478-0260, Fax: (212) 845-1286.

benchmark-quality coupled-cluster calculations with single, double, and perturbative triple excitations [CCSD(T)], for example, formally scales as  $O(N^7)$ , where  $N$  is proportional to the system size, whereas MP2 scales only as  $O(N^5)$ . For large systems, especially when the Boys-Bernardi counterpoise correction is used, these CCSD(T) calculations can be prohibitively costly.<sup>14</sup> Whereas a single benzene dimer interaction-energy calculation in a relatively modest augmented double- $\zeta$  basis set can, for example, take over 24 h to complete on a contemporary workstation using CCSD(T), the corresponding DF-MP2 calculation completes in under 5 min.

In order to overcome these limitations of MP2, we have developed *spin-network-scaled MP2* (SNS-MP2), a semi-empirical MP2-based method for dimer interaction-energy calculations. SNS-MP2 uses a neural network to reweight terms appearing in the total MP2 interaction energy so as to reproduce benchmark-quality, complete basis set (CBS)-extrapolated CCSD(T) interaction energies, which are considered the gold standard in computational chemistry due to their high accuracy.<sup>15</sup> The method can be interpreted as an extension of Grimme's spin-component-scaled MP2 (SCS-MP2) method (see Sec. II).<sup>16</sup> Instead of using two global spin-component coefficients chosen during training, as in the original SCS-MP2 and in other related approaches, in SNS-MP2, a collection of per-conformation HF-, MP2-, and SAPT0-based<sup>17,18</sup> features is used to derive, by way of a neural network, a unique set of spin-component coefficients for the particular system and conformation under study. SNS-MP2 predicts the interaction energies of our test set of 22 643 geometries of unseen complexes with a mean absolute error (MAE) of 0.04 kcal mol<sup>-1</sup> [root-mean-square error (RMSE) 0.09 kcal mol<sup>-1</sup>] relative to the gold-standard values, a 6–7 $\times$  improvement over MP2. Moreover, in addition to providing a point estimate of the interaction energy, SNS-MP2 provides a reliable predictive interval, or an error bar, on each interaction-energy value, a feature not available from any alternative method that we are aware of.

Based on reports in the literature concerning errors in CCSD(T)/CBS methodologies, we estimate the reliability of the benchmark values in our training and test data to be between 0.05 and 0.10 kcal mol<sup>-1</sup> RMSE. This indicates that the SNS-MP2 method is approaching the intrinsic accuracy of the CCSD(T)/CBS reference data it has been trained to predict. In further evaluation of the SNS-MP2 model on two previously published test sets from Řezáč, Riley, and Hobza<sup>19,20</sup> and Taylor *et al.*<sup>21</sup> that have been used to benchmark the accuracy of a wide variety of quantum chemical methods, we find that the accuracy of SNS-MP2 exceeds that of all previously reported  $O(N^5)$  and faster wavefunction-based methods and all density functionals.

## II. RELATED WORK

This work draws from two lines of related research in the fields of quantum chemistry and machine learning. The first area of work, in quantum chemistry, involves efforts to obtain accurate interaction energies using methods based solely on MP2 and/or DFT, thereby avoiding the additional

computational cost of higher accuracy methods such as CCSD(T). Novel methods in this class include MP2C,<sup>22</sup> MP2+ $\Delta vdw$ ,<sup>23</sup> SCS-MP2,<sup>16</sup> and SCS-MP2 variants such as SCSN-MP2,<sup>24</sup> SCS(MI)-MP2,<sup>25</sup> SOS-MP2,<sup>26</sup> and JMP2.<sup>27</sup> The second area of work, chiefly from the cheminformatics, force field development, and machine learning communities, involves efforts to use deep neural networks to predict molecular properties. Our work relies on combining the respective strengths of both approaches for modeling intermolecular interactions by integrating machine learning techniques within the mathematical structure of the MP2 method.

Most MP2-based semi-empirical methods that have been developed within the quantum chemistry community involve breaking the MP2 energy into a sum of two or more physically meaningful components and then either rescaling the components empirically or replacing them with values calculated using another method. A major deficiency of MP2 for intermolecular interactions in the asymptotic limit is its implicit inclusion of the uncoupled Hartree-Fock (UCHF) dispersion energy.<sup>28,29</sup> The MP2C method of Pitoňák and Heßelmann leverages this observation by subtracting out the UCHF dispersion energy from the MP2 interaction energy and adding back in a more accurate approximation of the dispersion energy from a coupled Kohn-Sham DFT calculation.<sup>22</sup> MP2+ $\Delta vdw$  adopts a similar approach, supplementing the MP2 interaction energy with a dispersion correction based on a pairwise interatomic  $C_n R^{-n}$  summation.<sup>23</sup>

SCS-MP2 is a method proposed originally by Grimme that exploits the fact that the MP2 correlation energy can be broken into two terms arising from the correlation of the same-spin and opposite-spin electron pairs.<sup>16</sup> These contributions can then be empirically scaled and recombined to yield more accurate energies. Values of 6/5 and 1/2 were recommended for the opposite-spin and same-spin scale factors (referred to as the “SCS coefficients”), respectively, based on least-squares fitting of the coefficients to 51 quadratic configuration interaction reaction energies.<sup>16</sup>

Although SCS-MP2 has demonstrated impressive performance for reaction energies, barrier heights, and molecular geometries,<sup>16</sup> the results have been less encouraging for interaction energies. DiStasio and Head-Gordon reported, for example, that on Hobza's S22 test set,<sup>12</sup> SCS-MP2 produced consistently less accurate results than standard DF-MP2.<sup>25</sup> Fitting the SCS coefficients afresh, they obtained values of 0.40 and 1.29 for the opposite-spin and same-spin coefficients, essentially inverses of the previous values. Hill and Platts independently refitted the SCS coefficients for interaction energies of 10 stacked nucleic acid base pairs.<sup>24</sup> They obtained values of 0 and 1.75 for the opposite-spin and same-spin scaling factors, respectively. King showed that a slight modification to Grimme's parameters, with values of 1.28 and 0.5 for the opposite-spin and same-spin coefficients, was capable of yielding accurate results for the ethylene dimer.<sup>30</sup>

These results, taken together, suggest that the optimal SCS coefficients may depend significantly on the nature of the molecules under study as well as on the details of their interaction. Different dimer systems may be most accurately modeled with different parameter values—whereas certain

values may perform well for hydrogen bonding interactions, for example, other values may be appropriate for  $\pi$ - $\pi$  interactions, and so on. One implication of this suggestion is that more flexible models, in which the coefficients are permitted to vary across chemical (and possibly conformational) space, could yield improved accuracy. Tan, Acevedo, and Izgorodina recently explored this direction with spin-ratio-scaled MP2 (SRS-MP2), an extension of SCS-MP2 in which two different sets of SCS coefficients are used depending on the ratio of the opposite-spin to the same-spin components of the interaction correlation energy.<sup>31</sup>

If an SCS-like model is to be constructed with variable coefficients, by what criteria should the model determine which coefficient values are used in each calculation? One choice would be to fit a series of separate fixed-coefficient models for different classes of molecular interactions; this would, however, prevent us from exploiting the information shared across these classes and require difficult decisions about the level of specificity with which the classes should be defined. Another approach, drawing from MP2+ $\Delta$ vdw,<sup>23</sup> molecular mechanics force fields, and dispersion-corrected DFT, could be to define the SCS coefficients in terms of parameters associated with a set of atom types in addition to the nuclear geometry. In addition to the difficulty of defining the appropriate atom-type groupings, however, this method would require the development of yet unknown geometry-dependent functional forms.

In this work we have chosen an alternative strategy, one that utilizes quantum chemical observables related to decompositions of the HF, MP2, and SAPT0 interaction energies for the particular dimer configuration under study as input features to a deep neural network that constructs a unique pair of SCS coefficients for that configuration. This approach avoids difficulties related to the division of molecules into classes and the manual selection of atom types inherent in the two approaches mentioned above. Furthermore, as we will show, the result is a method for calculating MP2-based interaction energies that is dramatically more accurate than any previously reported.

Although this approach shares many similarities with other recently developed methods that bridge quantum chemistry and machine learning, some distinctions should be noted. Many of these recently developed methods attempt to predict the atomization energy or other conformation-dependent properties of molecules directly by using connectivity or geometric information represented in a variety of forms, including through Coulomb matrices,<sup>32</sup> scattering transforms,<sup>33</sup> generalized symmetry functions,<sup>34,35</sup> bags of bonds,<sup>36</sup> and deep tensor neural networks.<sup>37</sup> When successful, these approaches provide an alternative to quantum chemical calculations and can be thought of as molecular mechanics force fields with significantly more flexible functional forms. The strategy we take here is somewhat different—our model does not eliminate the need for quantum chemical calculations on new conformations but instead uses them as its core building blocks. It aims to build on the strengths of MP2 and provide a method for predicting benchmark-quality CCSD(T)/CBS-level dimer interaction energies without the cost of coupled-cluster calculations.

### III. NEURAL NETWORK ARCHITECTURE

The role of the deep neural network in this work is to take values from a series of quantum mechanical calculations performed on the conformation under study and output a predicted energy and corresponding uncertainty. The architecture of our deep neural network is largely a generic feed-forward multilayer perceptron (MLP).<sup>38</sup> It operates on each conformation independently, receiving a vector of 22 input features (described in Table I) and outputting three scalars, which we label as  $c_{os}$ ,  $c_{ss}$ , and  $b$ . The first two values,  $c_{os}$  and  $c_{ss}$ , are the spin-component coefficients used in the construction of the final energy, and the third value,  $b$ , determines the uncertainty of the prediction. The architecture of the network is shown in Fig. 1. Although all of the quantum chemical input features depend on the molecular geometry, no direct geometric observables such as interatomic distances are employed and no auxiliary information is provided—during training the model is not informed of, for example, which inputs represent the same chemical systems or nearby molecular geometries.

The estimated total interaction energy is constructed from  $c_{os}$  and  $c_{ss}$  as

$$\hat{E}_{\text{SNS-MP2}} = E_{\text{HF/aVQZ}} + c_{os}E_{\text{corl,MP2/CBS}}^{OS} + c_{ss}E_{\text{corl,MP2/CBS}}^{SS}, \quad (1)$$

where  $E_{\text{HF/aVQZ}}$  is the Hartree-Fock interaction energy computed in the aVQZ basis set (a modification of the aug-cc-pVQZ basis; see Appendix A for further details), and  $E_{\text{corl,MP2/CBS}}^{OS}$  and  $E_{\text{corl,MP2/CBS}}^{SS}$  are the CBS-extrapolated opposite-spin and same-spin components of the MP2 correlation energy. The CBS limit is approximated by a two-point Helgaker extrapolation.<sup>39</sup> Note that all of the components required for the calculation of Eq. (1) are themselves elements of the input vector used to compute  $c_{os}$  and  $c_{ss}$  (see Table I). Nevertheless, we interpret this architecture as enforcing a quasi-linearity of  $\hat{E}_{\text{SNS-MP2}}$  with respect to  $E_{\text{corl,MP2/CBS}}^{OS}$  and  $E_{\text{corl,MP2/CBS}}^{SS}$ , and we find that this model yields much more accurate predictions than those in which the neural network directly outputs the total interaction energy without a final SCS-MP2-like resummation.

In employing any quantum chemical method to address scientific questions, it is essential to evaluate, either qualitatively or quantitatively, the degree of confidence which should be accorded to the results. For this reason, instead of merely predicting a single scalar, SNS-MP2 has been constructed to output a probability distribution over interaction energies, centered at  $\hat{E}_{\text{SNS-MP2}}$  with some variable width, representing the range of possible values it predicts that the CCSD(T)/CBS interaction energy may take. The use of a heteroskedastic error model is particularly important because the areas of intended application of the model include both compact and extended dimers, for which the uncertainty in the interaction energy differs substantially. This information is typically summarized by a predictive interval, or an error bar, on the interaction energy.

We utilize two extensions to the basic MLP architecture, both related to the quantification of uncertainty in the model's predicted energies. The model's uncertainty is parameterized using the neural network's third scalar output,  $b$ , which is



TABLE I. Input features used by the network to predict conformation-specific spin-component coefficients.

Name	Description
$E_{\text{int}}^{\text{HF/aVQZ}}$	Large-basis HF interaction energy
$E_{\text{int}}^{\text{HF/aVXZ}}$	Small-basis HF interaction energy
$E_{\text{int}}^{\text{MP2/aVQZ}}$	Large-basis MP2 interaction energy
$E_{\text{int}}^{\text{MP2/aVXZ}}$	Small-basis MP2 interaction energy
$E_{\text{corl,ss}}^{\text{MP2/aVQZ}}$	Large-basis MP2 same-spin correlation interaction energy
$E_{\text{corl,ss}}^{\text{MP2/aVXZ}}$	Small-basis MP2 same-spin correlation interaction energy
$\zeta_{\text{low}} = 2$ and $\zeta_{\text{low}} = 3$	One-hot encoding of the $\zeta$ -value of the small basis set, two binary features
$E_{\text{elst}}^{\text{HF/aVQZ}}$	Large-basis HF electrostatic interaction energy
$E_{\text{elst}}^{\text{MP2/aVQZ}}$	Large-basis MP2 electrostatic interaction energy
$E_{\text{HL}}^{\text{aVQZ}}$	Large-basis Heitler-London energy
$L^{\text{MP2/aVQZ}}$	Large-basis MP2 density matrix overlap
$L^{\text{HF/aVQZ}}$	Large-basis HF density matrix overlap
$E_{\text{disp}(20)}^{\text{SAPT0/aVTZ}}$	SAPT0 second-order dispersion energy
$E_{\text{disp}(20),ss}^{\text{SAPT0/aVTZ}}$	SAPT0 same-spin second-order dispersion energy
$E_{\text{elst}(10)}^{\text{SAPT0/aVTZ}}$	SAPT0 first-order electrostatic energy
$E_{\text{exch}(10)}^{\text{SAPT0/aVTZ}}(S^2)$	SAPT0 first-order exchange energy ( $S^2$ approximation)
$E_{\text{exch-disp}(20)}^{\text{SAPT0/aVTZ}}$	SAPT0 second-order exchange-dispersion energy
$E_{\text{exch-disp}(20),ss}^{\text{SAPT0/aVTZ}}$	SAPT0 same-spin second-order exchange-dispersion energy
$E_{\text{elst}(10)}^{\text{SAPT0/aVTZ}}$	SAPT0 first-order electrostatic energy
$E_{\text{exch-ind}(20)}^{\text{SAPT0/aVTZ}}$	SAPT0 second-order exchange-induction energy
$E_{\text{ind}(20)}^{\text{SAPT0/aVTZ}}$	SAPT0 second-order induction energy

interpreted as the width of a Laplace distribution,

$$P(E_{\text{CCSD(T)/CBS}}) = \frac{1}{2b} \exp\left(-\frac{|\hat{E}_{\text{SNS-MP2}} - E_{\text{CCSD(T)/CBS}}|}{b}\right). \quad (2)$$

This error distribution is incorporated into the model by the choice of the loss function used in training this neural network.<sup>40,41</sup> The network's weights,  $W$ , are optimized to

approximately minimize  $\ell(W)$ ,

$$\ell(W) = \frac{1}{N} \sum_i \frac{|\hat{E}_{\text{SNS-MP2}}^{(i)}(W) - E_{\text{CCSD(T)/CBS}}^{(i)}|}{b^{(i)}(W)} + \frac{1}{N} \sum_i \log b^{(i)}(W), \quad (3)$$

where the index  $i$  runs over all data points in the training set and  $N$  is the number of training data points. Ignoring an additive constant, this objective function is the negative mean log likelihood of the data under our Laplace error model. It incentivizes both small (scaled) absolute errors and tight error bars. The choice of an error model based on the Laplace distribution mitigates the effect of outliers when fitting the SNS-MP2 model; we note that such error models have also been widely used in the field of robust regression for similar reasons.<sup>42</sup> The use of a softplus activation function<sup>43</sup> on the last layer of the network ensures that the value of  $b$  will always be positive, so no constraints are required to keep Eq. (3) finite and well behaved.

We also use dropout, both during training and prediction.<sup>44</sup> During training, dropout works by stochastically setting a small fraction of the weights in the neural network to zero when computing the gradient of the loss function with respect to the weights. This can be interpreted as an approximate variational Bayesian inference method or as a form of model averaging.<sup>44–46</sup> At test time, instead of using the common weight-scaling technique in which the network weights are

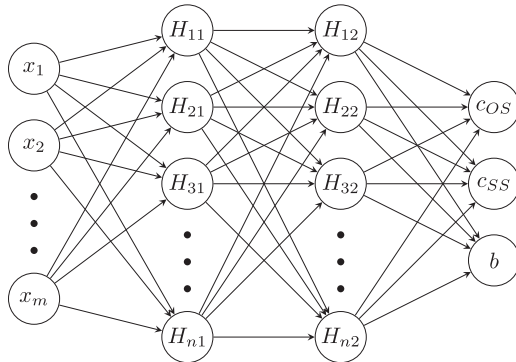


FIG. 1. Illustration of the SNS-MP2 network architecture. The final predicted energy is  $\hat{E}_{\text{SNS-MP2}} = E_{\text{HF/aVQZ}} + c_{\text{OS}} E_{\text{corl,MP2/CBS}}^{\text{OS}} + c_{\text{SS}} E_{\text{corl,MP2/CBS}}^{\text{SS}}$ , where  $c_{\text{OS}}$  and  $c_{\text{SS}}$  are computed using the network structure as shown. A hyperbolic-tangent activation function is used for the intermediate layers, and a softplus activation function is used to connect the last hidden layer to the output layer. Our final model uses  $m = 22$  input features and  $n = 50$  nodes per hidden layer, 10 hidden layers, and a dropout fraction of 2.5%.

rescaled to their expected value given the dropout fraction, we use Monte Carlo model averaging. In this approach, when making a prediction, a moderate number ( $\sim 1000$ ) of random neural networks are sampled using dropout, the prediction is made with each network, and the results are averaged. Each sampled network yields a separate Laplace distribution over the predicted energy—the resulting averaged prediction using Monte Carlo model averaging is a mixture of these Laplace distributions. Because it requires a number of forward passes through the network to sample the effect of the dropout, Monte Carlo model averaging is more computationally costly than the common weight-averaging approach. For our purposes, however, the incremental cost of these extra forward passes through the network is negligible with respect to the cost of the HF, MP2, and SAPT0 calculations that generate the input features. The advantage over conventional weight scaling is both formal correctness<sup>44</sup> and a slight empirical increase in predictive accuracy.

#### IV. DATA SET

In opting for a flexible neural network-based model, overfitting becomes a significant concern due to the large number of model parameters. Very large data sets for training and validation are thus essential. Accordingly, we collected 247 560 benchmark- and near-benchmark-quality CCSD(T)/CBS dimer interaction-energy data points. This data set is nearly twice as large as the QM9 data set of DFT-computed [B3LYP/6-31G(2df,p)] properties, which has been used to fit machine learning models, and significantly larger than previous CCSD(T)/CBS data sets used to fit other semi-empirical MP2-based methods and density functionals.<sup>19,47–49</sup> The collection of CCSD(T)/CBS data used in this work is over 100 times larger than the set of 2027 CCSD(T) noncovalent interaction energies curated by Schneebeli, Bochevarov, and Friesner.<sup>49</sup>

The data set contains 3149 unique chemical systems in an average of 80 different conformations per chemical system (min: 3, median: 29, max: 2486). The systems span the range of chemical functionality found in biological systems and covered by common biomolecular force fields. The distribution of this data set by functional groups and sizes is shown in Figs. 2 and 3, respectively.<sup>77</sup> The most prevalent molecule in the data set is water; the set features 2486 conformations of water dimer and 115 227 conformations for which at least

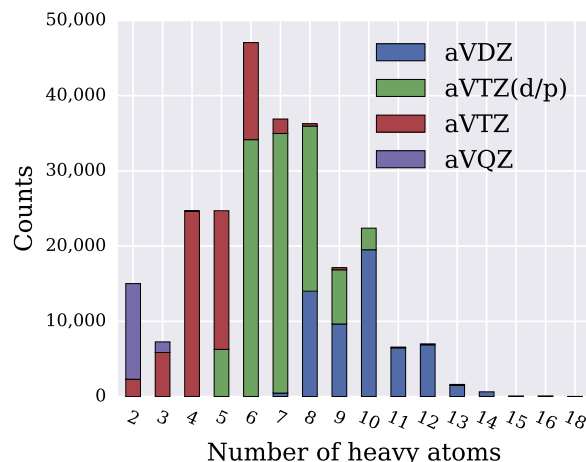


FIG. 3. Distribution of the number of heavy atoms in the overall data set. The bars are colored by the basis set used to perform the coupled-cluster high-level correlation correction. Among the smallest systems, by this measure, is the interaction of two water molecules, of which our data set contains 2486 conformations, all with aVQZ coupled-cluster corrections. The largest system by this measure, with 18 heavy atoms, is the interaction of two 1,3,5-trifluorobenzene molecules.

one of the molecules is water. Due to concerns over complete charge transfer in the supermolecular interaction-energy calculations, no systems in which the net charge on either fragment was +2, −2, or beyond and no systems in which the fragments carried net charges with an opposite sign were used in this work.

At the outset, the data set was split into three subsets for training, validation, and final testing. The unique chemical systems were split randomly into an 80%/10%/10% distribution among the three subsets. All of the geometries associated with a given chemical system were either entirely in the training set, entirely in the validation set, or entirely in the final test set. After performing this division, the training, validation, and test sets contained 204 934, 19 983, and 22 643 data points, respectively.

#### A. Reference energies

Counterpoise-corrected benchmark values were computed with a hybrid CCSD(T) and CBS-extrapolated MP2 method that is often referred to as the gold standard in computational thermochemistry.<sup>15</sup> The reference energy is constructed from three components: the aVQZ HF interaction

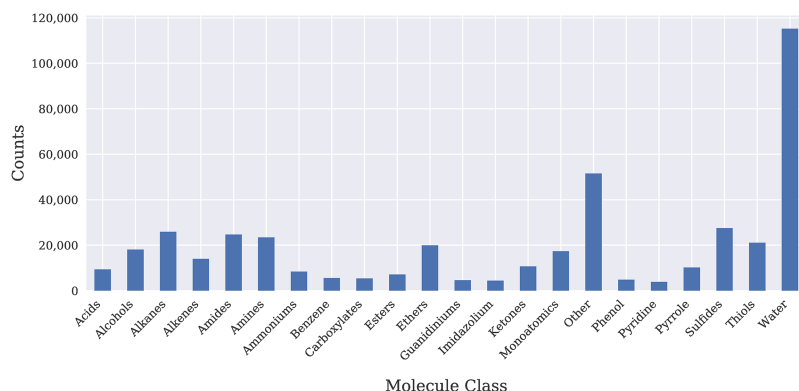


FIG. 2. Distribution of molecule classes in the overall data set. The height of each bar shows the number of conformations for which either of the two interacting systems belongs to the indicated class. Over 115 000 data points, or nearly half the data set, have water as one of the interacting fragments.

energy,  $E_{\text{total, HF}}^{\text{aVQZ}}$ , the MP2 interaction correlation energy extrapolated to the CBS limit using Helgaker's inverse third-power form,<sup>39</sup>  $E_{\text{corl, MP2}}^{\text{aV[XQ]Z}}$ , and a high-level correlation energy correction computed as the difference between the CCSD(T) and MP2 interaction energy in a single basis set,  $\delta_{\text{CCSD(T)-MP2}}^{\text{aVX'Z}}$ ,

$$E_{\text{total, CCSD(T)}}^{\text{CBS}} = E_{\text{total, HF}}^{\text{aVQZ}} + E_{\text{corl, MP2}}^{\text{aV[XQ]Z}} + \delta_{\text{CCSD(T)-MP2}}^{\text{aVX'Z}}, \quad (4)$$

$$E_{\text{corl, MP2}}^{\text{aV[XQ]Z}} = \frac{E_{\text{corl, MP2}}^{\text{aVQZ}} 4^3 - E_{\text{corl, MP2}}^{\text{aVXZ}} X^3}{4^3 - X^3}, \quad (5)$$

$$\delta_{\text{CCSD(T)-MP2}}^{\text{aVX'Z}} = E_{\text{total, CCSD(T)}}^{\text{aVX'Z}} - E_{\text{total, MP2}}^{\text{aVX'Z}}. \quad (6)$$

For the two-point MP2/CBS extrapolation, all 247 470 high-basis MP2 calculations were performed in the aVQZ basis, while the low-basis MP2 calculations [basis set  $X$  in Eqs. (4) and (5)] were performed in a mixture of basis sets: 185 334 were performed at aVTZ, 47 413 at aVTZ(*d/p*), and 14 723 at aVDZ (see Appendix A for a specification of these basis sets). The CCSD(T) high-level correlation corrections dominate the cost of the benchmark calculation and were also performed at a mixture of basis sets [basis set  $X'$  in Eqs. (4) and (6)]: 14 176 were performed at aVQZ, 66 652 at aVTZ, 107 315 at aVTZ(*d/p*), and 59 327 at aVDZ. The choice of the basis set used for each system was made based on the molecular size, with larger systems run in smaller basis sets to reduce the computational cost.

These calculations were performed using three software packages—MOLPRO 2012,<sup>50,51</sup> MOLPRO 2015,<sup>50,52</sup> and Psi4 1.0<sup>53</sup>—each with some in-house modifications. All calculations were performed in the C1 point group, all correlated calculations used the frozen-core approximation, and most MP2 calculations were performed with density fitting.<sup>78</sup>

The geometries were generated both by DF-LMP2/aVTZ geometry optimization and by the extraction of frames from molecular dynamics simulations. Frequently, although not exclusively, conformations were generated based on radial scans that were seeded from both optimized geometries and conformations observed in molecular dynamics simulations. The radial scans were performed by displacing the monomers along their intermolecular axis in units of 0.1 Å. Because of these procedures, a greater degree of diversity is present in the intermolecular distances and orientations of the dimer configurations than in the geometries of their constituent monomers, which tend to be at, or near, equilibrium.

## B. Input features

The model does not use the geometry information or atomic composition for a given conformation directly as input features. Instead, it uses quantum chemical features computed from HF, MP2, and SAPT0 calculations on the dimer and its monomers. The features are intended to encode the nature of the interaction in a geometry-independent manner such that chemically disparate dimers can be represented within a single model. We briefly describe each feature below; refer to Table I for the complete list.

The first set of features includes those which are computed from low-basis and high-basis counterpoise-corrected

interaction-energy calculations. These are the same underlying QM calculations used for the MP2/CBS component of the gold-standard reference values. The features in this category are the HF and MP2 interaction energies and the same-spin component of the MP2 interaction correlation energy. These values are computed in two basis sets, and we supply the network with both values, for a total of six features. Whereas the high basis is fixed to aVQZ, the low basis varies across the data set. We thus supplied a one-hot encoding<sup>54</sup> of the  $\zeta$ -value of the low basis as two additional binary input features.

Furthermore, we performed a series of aVQZ calculations to compute the electrostatic interaction energy, Heitler-London energy, and density-matrix overlap. The electrostatic interaction energy between fragments  $A$  and  $B$  is defined as

$$E_{\text{es}}^X = \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho_A^X(\mathbf{r}_1) \rho_B^X(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} - \sum_{a \in A} Z_a \int d\mathbf{r} \frac{\rho_B^X(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_a|} - \sum_{b \in B} Z_b \int d\mathbf{r} \frac{\rho_A^X(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_b|} + \sum_{a \in A} \sum_{b \in B} \frac{Z_a Z_b}{|\mathbf{r}_a - \mathbf{r}_b|}, \quad (7)$$

where  $\rho_{A/B}^X$  is the electron density associated with the relevant monomer calculated at the  $X = \text{HF/MP2}$  level in the monomer basis set. The Heitler-London energy provides information about the Pauli repulsion between the monomers and is defined as

$$E_{\text{HL}} = \frac{\langle \Psi^H | \hat{A} | \Psi^H \rangle}{\langle \Psi^H | \hat{A} | \Psi^H \rangle} - E_A^{\text{HF}} - E_B^{\text{HF}}, \quad (8)$$

where  $\hat{A}$  is the antisymmetrizer and  $\Psi^H = \Psi_A^{\text{HF}} \Psi_B^{\text{HF}}$  is the Hartree product of the isolated-monomer ground-state HF wavefunctions computed in the corresponding monomer basis sets. The density-matrix overlap is

$$L^X = \int d\mathbf{r}_1 d\mathbf{r}_2 \gamma_A^X(\mathbf{r}_1, \mathbf{r}_2) \gamma_B^X(\mathbf{r}_2, \mathbf{r}_1), \quad (9)$$

where  $\gamma_{A/B}^X$  is the one-particle reduced density matrix associated with the relevant monomer calculated at the  $X = \text{HF/MP2}$  level in the monomer basis set. These quantities yield an additional five features.

Finally, we included nine features collected from SAPT0/aVTZ calculations for each data point.<sup>17,18</sup> These features are the second-order dispersion, induction, exchange-dispersion, and exchange-induction energies, the same-spin component of the second-order dispersion and exchange-dispersion energies, the first-order electrostatic and exchange energies, and the first-order exchange energy computed in the  $S^2$  approximation.

An open-source Psi4 plugin that allows users to compute SNS-MP2 interaction energies and confidence intervals is available at <http://www.deshawresearch.com/resources.snsmp2.html>.

## V. RESULTS

### A. Training set

Extensive but non-exhaustive scans over hyperparameters were performed to choose the number of layers, the number of hidden nodes per layer, and the dropout strength. We

implemented the SNS-MP2 model with the Keras library,<sup>55</sup> using the Theano backend,<sup>56</sup> and observed smooth training of the network on a single NVIDIA GTX 980 Ti graphical processing unit (GPU) using the RMSProp optimizer<sup>57</sup> with default parameters over 100 000 epochs. (The training-set error statistics are shown in Table II.) The final model, with 10 hidden layers and 50 nodes per layer, was selected to minimize the validation-set error (Table III). The training curve of the selected model, which plots the error on the training and validation sets over the course of the optimization, is shown in Fig. 4. Despite the use of dropout, some evidence of overfitting is still visible in the gap between the performance on the training set and on the validation set. On a log scale, the apparent performance on the training set does not appear to be saturated although the validation loss may be converged.

On the training data, we observe that the neural network model is consistently more accurate than MP2. The MAE, RMSE, and 99th percentile absolute error in the neural network's energy predictions on our training set, 0.030 kcal mol<sup>-1</sup>, 0.078 kcal mol<sup>-1</sup>, and 0.340 kcal mol<sup>-1</sup>, respectively, are all roughly 7× to 8× lower than the corresponding MP2 error statistics. Furthermore, as shown in Table II, the model achieves error reductions of 3–14× across all of the classes of molecules in the training set, suggesting a broadly shared increase in accuracy. In contrast to MP2, which has a mean signed error on the training data of -0.15 kcal mol<sup>-1</sup> (reflecting its tendency to overbind noncovalent complexes), the mean signed error of the SNS-MP2 model on the training data is 0.001 kcal mol<sup>-1</sup>.

On our training set, SNS-MP2 achieves its highest accuracy for systems containing water, an unsurprising observation for two reasons: First, these systems are already very well described by MP2 (the MP2/CBS RMSE is under 0.2 kcal mol<sup>-1</sup> on our training set), and second, systems containing water are very highly represented in the data set. The most challenging systems for the neural network are aromatic: Predictions on systems containing pyridine, benzene, pyrrole, phenol, and imidazolium are among the most inaccurate, with MAEs close to 0.06 kcal mol<sup>-1</sup> and RMSEs exceeding 0.11 kcal mol<sup>-1</sup>. This itself is a remarkable reduction in error—for benzene-containing systems, this is a >11× improvement compared to MP2. Much of this reduction comes from improvements on the 637 conformations in the training set of benzene interacting with another aromatic system, for which the MP2/CBS RMSE of over 2.4 kcal mol<sup>-1</sup> is reduced by SNS-MP2 to 0.21 kcal mol<sup>-1</sup> RMSE. In addition, the general tendency of MP2 to overbind these systems of interacting aromatics is largely corrected. Whereas the MP2 interaction energy is too attractive by an average of 1.4 kcal mol<sup>-1</sup> on these conformations, the mean signed error of SNS-MP2 is over 20× smaller (0.07 kcal mol<sup>-1</sup>).

## B. Test-set summary statistics

More relevant than the training-set error statistics is the performance on our test set, which consists of 22 643 conformations of 314 different molecular systems that were not encountered by the model during training. The most highly represented systems in the test set are the ammonia and

TABLE II. Training-set error summary statistics. Each row refers to a set of dimer systems containing at least one fragment of the indicated class.

Class	N	MP2/CBS		MP2/CBS		SNS-MP2	
		MAE	RMSE	99th percentile	MAE	RMSE	99th percentile
All	204 934	0.208	0.556	2.631	0.030	0.078	0.340
Acids	6 268	0.206	0.478	2.253	0.032	0.070	0.320
Alcohols	16 035	0.180	0.441	2.105	0.031	0.077	0.344
Alkanes	22 337	0.205	0.525	2.504	0.024	0.062	0.284
Alkenes	10 702	0.292	0.669	2.982	0.037	0.102	0.498
Amides	21 926	0.176	0.422	2.026	0.033	0.078	0.353
Amines	18 487	0.220	0.534	2.483	0.030	0.074	0.350
Ammoniums	6 516	0.296	0.633	2.776	0.042	0.094	0.417
Benzene	4 396	0.927	1.649	6.123	0.066	0.146	0.673
Carboxylates	3 942	0.188	0.404	2.008	0.057	0.120	0.563
Esters	5 992	0.105	0.241	1.137	0.023	0.051	0.228
Ethers	15 931	0.198	0.457	2.083	0.032	0.074	0.345
Guanidiniums	3 937	0.425	0.861	3.522	0.058	0.120	0.578
Imidazolium	2 659	0.690	1.212	4.554	0.078	0.160	0.719
Ketones	9 820	0.171	0.466	2.338	0.025	0.065	0.295
Monoatomics	14 505	0.246	0.521	2.332	0.045	0.099	0.372
Other	40 876	0.226	0.601	2.786	0.037	0.081	0.358
Phenol	3 463	0.658	1.319	5.492	0.061	0.125	0.538
Pyridine	2 997	0.474	1.050	4.562	0.060	0.114	0.501
Pyrrole	8 218	0.488	1.023	4.176	0.056	0.119	0.551
Sulfides	22 213	0.395	0.842	3.532	0.040	0.088	0.407
Thiols	19 161	0.285	0.678	3.099	0.031	0.099	0.339
Water	96 591	0.087	0.191	0.902	0.020	0.055	0.217



TABLE III. Validation-set error summary statistics. Each row refers to a set of dimer systems containing at least one fragment of the indicated class.

Class	N	MP2/CBS		MP2/CBS		SNS-MP2	
		MAE	RMSE	99th percentile	MAE	RMSE	99th percentile
All	19 983	0.255	0.635	2.964	0.043	0.095	0.407
Acids	1 061	0.199	0.469	2.243	0.030	0.061	0.284
Alcohols	741	0.382	0.717	2.786	0.081	0.158	0.665
Alkanes	1 751	0.254	0.599	2.747	0.029	0.063	0.281
Alkenes	1 049	0.323	0.675	2.794	0.048	0.112	0.483
Amides	1 677	0.238	0.528	2.354	0.043	0.097	0.356
Amines	631	0.309	0.607	2.346	0.046	0.079	0.308
Ammoniums	1 385	0.291	0.707	2.972	0.055	0.108	0.420
Benzene	492	1.191	1.920	6.374	0.109	0.210	0.888
Carboxylates	891	0.146	0.347	1.736	0.047	0.103	0.525
Esters	881	0.081	0.152	0.700	0.036	0.074	0.316
Ethers	2 148	0.206	0.479	2.181	0.038	0.082	0.391
Guanidiniums	312	0.389	0.775	3.215	0.059	0.111	0.456
Imidazolium	742	0.401	0.835	3.703	0.055	0.101	0.349
Ketones	363	0.410	0.830	3.638	0.063	0.113	0.438
Monoatomics	1 656	0.348	0.634	2.643	0.067	0.119	0.435
Other	5 165	0.223	0.561	2.642	0.044	0.099	0.400
Phenol	701	0.472	1.023	4.521	0.065	0.139	0.660
Pyridine	640	0.435	1.098	4.689	0.052	0.094	0.398
Pyrrole	1 179	0.510	1.091	4.673	0.079	0.169	0.766
Sulfides	2 534	0.490	0.954	3.833	0.050	0.103	0.437
Thiols	1 411	0.411	0.875	3.704	0.050	0.102	0.435
Water	9 315	0.091	0.198	0.966	0.026	0.054	0.227

formic acid homodimers, with 2208 and 1659 geometries, respectively. The SNS-MP2 error statistics on the test set are shown in Table IV. Over the test set, we measure an MAE of  $0.038 \text{ kcal mol}^{-1}$ , an RMSE of  $0.088 \text{ kcal mol}^{-1}$ , and a 99th percentile absolute error of  $0.384 \text{ kcal mol}^{-1}$ . This is a large and definitive improvement over MP2: The MAE, RMSE, and 99th percentile absolute error all decrease by 6–7 $\times$  compared to MP2. These errors are only very mildly larger than the errors observed on the training set, indicating a low degree of

overfitting. Consistent with our observations on the training set, MP2 overbinds these test complexes by a substantial margin (the mean signed error is  $0.17 \text{ kcal mol}^{-1}$ ), a trend almost fully corrected by SNS-MP2, which achieves a mean signed error on the test set of  $9 \times 10^{-4} \text{ kcal mol}^{-1}$ .

Analyzing the test-set performance by molecule classes, we reach similar conclusions to those suggested by performance on the training set. Many of the most challenging systems are aromatic. Among the three test-set dimer systems with the worst MAEs are acetylene-benzene and dimethylacetylene-benzene (SNS-MP2 MAE of 0.41 and  $0.23 \text{ kcal mol}^{-1}$ , respectively), suggesting that  $\pi$ - $\pi$  interactions still remain an area of difficulty.<sup>79</sup> Systems containing water, which account for nearly half of the test set, are among the most accurately modeled with an MAE of  $0.025 \text{ kcal mol}^{-1}$  (RMSE  $0.057 \text{ kcal mol}^{-1}$ ), a 4 $\times$  improvement over MP2.

To what degree is the accuracy of the SNS-MP2 method approaching the intrinsic error in the benchmark interaction-energy reference values themselves? Brauer *et al.* estimated the uncertainty of a similar CCSD(T)/CBS methodology to be  $0.05 \text{ kcal mol}^{-1}$  RMSE for the S66x8 data set.<sup>20</sup> Burns, Marshall, and Sherrill showed that the inaccuracy introduced by performing the CCSD(T) correction in the aVDZ basis set, which we do in roughly one quarter of our interaction energies, is  $0.08 \text{ kcal mol}^{-1}$  MAE on a composite data set of 345 interaction energies.<sup>58</sup> Řezáč and Hobza show that not including post-CCSD(T) corrections introduces an error on the order of 1% on the A24 data set.<sup>15</sup> Applying this result to our data set yields an estimated error of  $0.08 \text{ kcal mol}^{-1}$ . Our own examination of the basis-set-related errors, drawn from comparisons

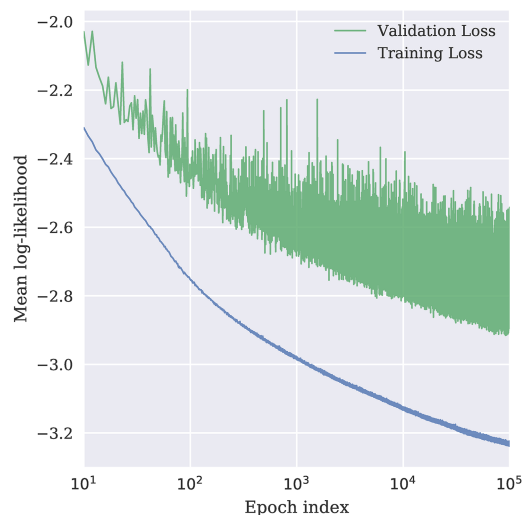


FIG. 4. Neural network training curve. This figure shows the convergence of the objective function over the course of training the final model for 100 000 epochs. Optimization was performed on a single NVIDIA GTX-980Ti GPU and took approximately 24 h starting from random initial weights.

TABLE IV. Test-set error summary statistics. Each row refers to a set of dimer systems containing at least one fragment of the indicated class.

Class	N	MP2/CBS		MP2/CBS		SNS-MP2		SNS-MP2	
		MAE	RMSE	99th percentile	MAE	RMSE	99th percentile	MAE	RMSE
All	22 643	0.225	0.593	2.750	0.038	0.088	0.383		
Acids	1 989	0.093	0.288	1.355	0.027	0.06	0.281		
Alcohols	1 274	0.238	0.548	2.474	0.049	0.114	0.482		
Alkanes	1 788	0.292	0.611	2.529	0.045	0.092	0.395		
Alkenes	2 199	0.268	0.640	2.918	0.037	0.083	0.351		
Amides	1 072	0.311	0.637	2.448	0.057	0.110	0.441		
Amines	4 338	0.145	0.424	2.195	0.023	0.066	0.338		
Ammoniums	480	0.357	0.647	2.520	0.073	0.129	0.568		
Benzene	677	0.665	1.422	5.418	0.075	0.184	0.847		
Carboxylates	546	0.229	0.458	2.088	0.076	0.153	0.613		
Esters	250	0.170	0.342	1.575	0.034	0.063	0.267		
Ethers	1 822	0.280	0.576	2.429	0.048	0.091	0.375		
Guanidiniums	310	0.461	0.866	3.266	0.074	0.138	0.556		
Imidazolium	1 034	0.318	0.720	3.171	0.055	0.092	0.394		
Ketones	446	0.309	0.654	2.718	0.050	0.100	0.490		
Monoatomics	1 291	0.318	0.605	2.480	0.052	0.089	0.303		
Other	5 501	0.256	0.705	3.144	0.048	0.109	0.476		
Phenol	638	0.425	1.085	4.575	0.045	0.086	0.377		
Pyridine	226	0.695	1.208	4.264	0.076	0.126	0.396		
Pyrrole	779	0.504	1.047	4.327	0.063	0.126	0.552		
Sulfides	2 788	0.498	1.001	4.199	0.055	0.111	0.449		
Thiols	491	0.517	0.933	3.326	0.063	0.122	0.566		
Water	9 411	0.101	0.222	1.077	0.025	0.057	0.256		

to CCSD(T\*)-F12a/aVQZ interaction energies on small systems, is consistent with these aforementioned results.<sup>59,60</sup> We note that there are differences between these smaller data sets and our full benchmark data set including, but not limited to, chemical composition and choice of basis sets; nevertheless, taken together, we estimate the reliability of the benchmark values in our training and test data to be likely between 0.05 and 0.10 kcal mol<sup>-1</sup> RMSE. Considering the SNS-MP2 test-set RMSE of 0.088 kcal mol<sup>-1</sup>, there may be some further residual signal in the data that is not accurately captured by the SNS-MP2 neural network. These data suggest, however, that particularly on the most accurately modeled systems, such as complexes containing water, SNS-MP2 may be nearly as accurate as the gold-standard CCSD(T)/CBS benchmark values themselves. Dramatic increases in the accuracy of SNS-MP2 are thus unlikely to be possible without obtaining data sets of higher quality, beyond the currently accepted gold-standard level of accuracy.

### C. Typical performance across radial scans

In order to demonstrate the range of accuracy of the model on the validation set, we selected three radial scans from the 5th, 50th, and 95th percentiles of the distribution of scans by mean log likelihood. These scans are shown in Fig. 5. The first and third panels, respectively, can be interpreted as the near-worst-case and near-best-case behavior for SNS-MP2, and the second panel can be interpreted as a typical case. The log likelihood includes a contribution both from the accuracy of the energy prediction as well as the predicted error bar. In the near-worst-case radial scan, an interaction between

chloromethane and imidazole, two problems are apparent: The predicted energy deviates by a substantial amount, 0.13 kcal mol<sup>-1</sup> MAE, and the error bars are underestimated, with the reference energy falling near the edge or outside of the 95% predictive interval for some of the conformations along the scan. Nonetheless, the sign and trend of the post-MP2/CBS correction are accurately predicted. In contrast, in the near-best-case scan, an interaction between ethylamine and water, we see an improved behavior: exceptionally small errors below 0.02 kcal mol<sup>-1</sup> MAE and reference energies that fall neatly within the error bars.

Some general trends, which apply to each of the scans in Fig. 5 and across the data set, should be highlighted. First, the model is substantially more accurate than MP2 and correctly predicts the sign and trend of the post-MP2/CBS correction in all three cases. Second, the accuracy of SNS-MP2 strongly depends on the intrinsic accuracy of MP2 itself: It performs best for systems like ethylamine-water in which MP2 is a good reference and more poorly for systems like chloromethane-imidazole in which larger corrections are required. Third, the per-conformation error bars are very responsive to the geometry of the dimer. The model is correctly able to recognize that at a long intermolecular distance, the post-MP2 correction approaches zero with extremely high confidence, while at a very short intermolecular distance, the uncertainties in the energy are much greater.

### D. Calibration of error bars

In addition to assessing the quality of the point estimates, it is important to determine whether the error bars computed

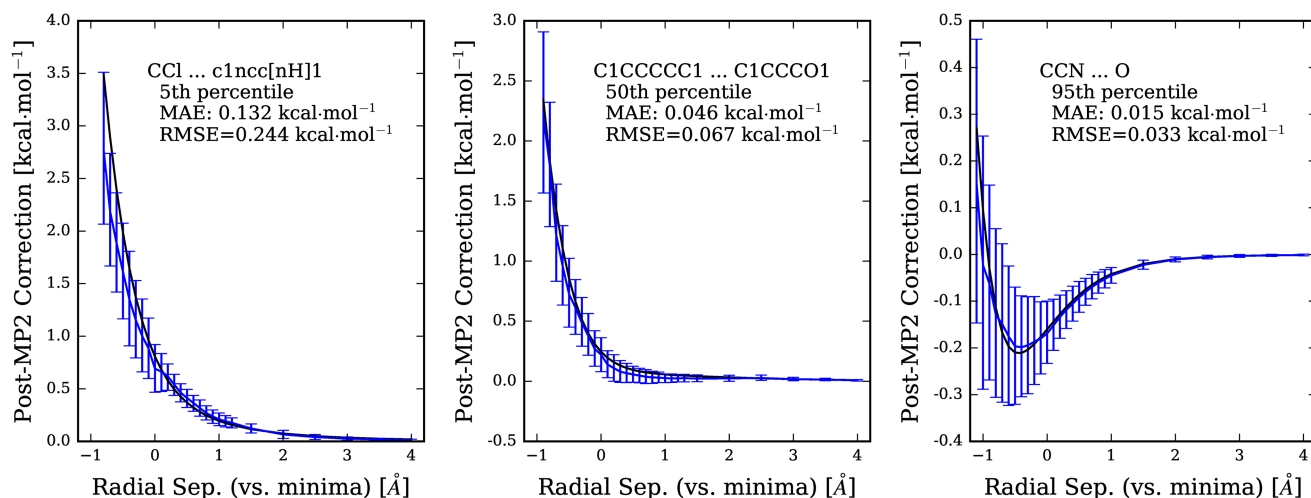


FIG. 5. Typical radial scans. This figure shows three radial scans from the test set with performance at the 5th, 50th, and 95th percentiles of the mean log likelihood by scan. The left and right panels, respectively, demonstrate the near-worst-case and near-best-case accuracy of the SNS-MP2 method, and the middle panel shows a typical case. The post-MP2/CBS correction is plotted on the vertical axis, with the CCSD(T) gold-standard reference value in black and SNS-MP2 in blue. Error bars are drawn at the SNS-MP2 95% predictive intervals.

by the model are accurate. We anticipate that these error bars will be used by researchers for a variety of purposes, and we neither want them to reflect a systematic underconfidence (reporting a large uncertainty for calculations in which the true error is small) or, worse, an overconfidence. In order to gauge whether the error bars are properly calibrated, we inspect the coverage of the predictive intervals. Recall that our error model is Laplacian; when the model predicts a given energy and error bar,  $(\hat{E}, \hat{b})$ , it is making a prediction about the probability distribution of the unknown benchmark

interaction energy. In particular, we focus on three quantiles of the error distribution: The Laplace cumulative distribution function implies that with 63.2%, 86.5%, and 95.0% probability, the SNS-MP2 energy should be, respectively, within  $\hat{b}$ ,  $2\hat{b}$ , and  $3\hat{b}$  of the gold-standard interaction energy. In fact, we find on our validation set that 65.0%, 91.9%, and 98.3% of the predictions, respectively, fall within these intervals of the benchmark CCSD(T)/CBS interaction energies. These results lead us to conclude that the model is slightly underconfident in its own accuracy but that the discrepancy is small.

To further probe this, Fig. 6 displays a histogram of the standardized errors,  $|\hat{E}_{\text{SNS-MP2}} - E_{\text{CCSD(T)/CBS}}|/\hat{b}$ . If our error model is accurate and the error bars,  $\hat{b}$ , are properly calibrated, then these values should follow a standard exponential distribution. On the other hand, if the error bars are systematically too large or too small, we would expect to be able to see this by comparing the empirical distribution of the scaled error on the test set to this expected distribution. In fact, we find that the two distributions match quite closely, suggesting that researchers employing this model should be assured that the model's confidence in its accuracy can be trusted in downstream applications such as force field parameterization. In our test set we see, moreover, no evidence that suggests a decrease in the reliability of the error bars for larger systems. The largest systems in our test set, as measured by the number of heavy atoms, in fact exhibit somewhat higher coverage rates—the fraction of the benchmark CCSD(T)/CBS values fall within the SNS-MP2 predictive intervals—than the test data set as a whole. Within the range of system sizes examined in this study, we observe no systematic trend suggesting that researchers should have less confidence in the reliability of the SNS-MP2 error bars when studying larger, as opposed to smaller, systems.

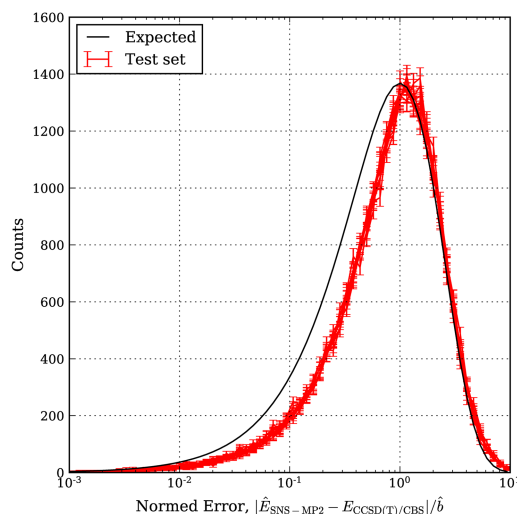


FIG. 6. Theoretical and observed distribution of the standardized errors in SNS-MP2. Our model assumes that the error in the neural network follows a zero-mean Laplace distribution with width  $\hat{b}$ , a number generated for each data point from the network. If this model is accurate and the widths of the error bars,  $\hat{b}$ , are properly calibrated, we would expect the standardized residuals,  $|\hat{E}_{\text{SNS-MP2}} - E_{\text{CCSD(T)/CBS}}|/\hat{b}$ , to follow a standard exponential distribution. This plot displays the histogram of these standardized residuals across the test set with different samples of the weights from the dropout distribution (red) with the values which would be theoretically expected based on the exponential error model (black). The rightward shift of the empirical distribution with respect to the theoretical distribution indicates the slight underconfidence of SNS-MP2 in its own accuracy.

## VI. SUPPLEMENTAL TEST SETS

In addition to assessing the performance of SNS-MP2 on our primary test set, after finalizing the model, we evaluated it on two supplemental test sets for which

benchmark-quality CCSD(T)/CBS reference energies have been publicly reported: Hobza's S66x8, a set of 66 dimer potential energy curves,<sup>19,20</sup> and a set of 10 dimer potential energy curves calculated by Taylor *et al.*,<sup>21</sup> which we refer to as T80.

### A. Validation on S66x8

We compared our neural network's predictions to the S66x8 revised benchmark energies computed by Brauer *et al.*, which combined CBS-extrapolated MP2-F12 energies with CCSD(Tsc)-F12b/cc-pVDZ-F12 high-level corrections.<sup>20</sup> Based on their assessments of yet higher-level calculations for smaller data sets, Brauer *et al.* estimated these benchmark values to be reliable to about 0.05 kcal mol<sup>-1</sup> RMSE.

The accuracy of our neural network in comparison to other commonly used methods on S66x8 is shown in Fig. 7. It achieves an MAE of 0.077 kcal mol<sup>-1</sup>, an RMSE of 0.14 kcal mol<sup>-1</sup>, and a maximum absolute error of 0.69 kcal mol<sup>-1</sup>. This maximum error occurs for the benzene-uracil system. The S66x8 data set is grouped into four subsets by the type of interaction: H-bonds,  $\pi$ -stacks, London, and mixed. On these subsets, our model achieves MAE values of 0.14, 0.11, 0.05, and 0.06 kcal mol<sup>-1</sup>, respectively.

Across the data set, SNS-MP2 is more accurate by MAE than all of the alternative methods with a similar computational cost. The only methods benchmarked by Brauer that yield more accurate test-set predictions than SNS-MP2 on this data set were SCS(MI)-CCSD(F12c), CCSD(T)(F12c), and CCSD(T\*)(F12c), each of which is significantly more computationally costly than SNS-MP2.

### B. Validation on T80

Figure 8 shows the accuracy of our model in comparison to a number of popular methods on each of the 10 chemical systems in T80. The overall MAE of our model is 0.062 kcal mol<sup>-1</sup> and the RMSE is 0.12 kcal mol<sup>-1</sup>, which are similar to the values measured on the S66x8 data set. The maximum error is 0.37 kcal mol<sup>-1</sup>. The reference CCSD(T)/CBS energies fall within the SNS-MP2 95% predictive interval for 67 out of the 80 conformations in the T80 test set, giving an observed coverage of 84%. Our model is more accurate than

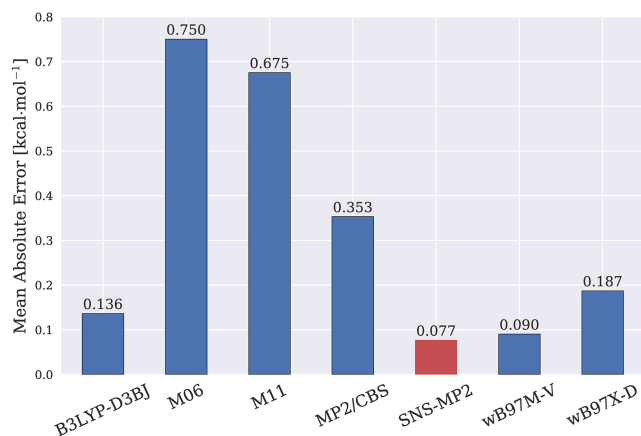


FIG. 7. Mean absolute error of various methods on the S66x8 test set.

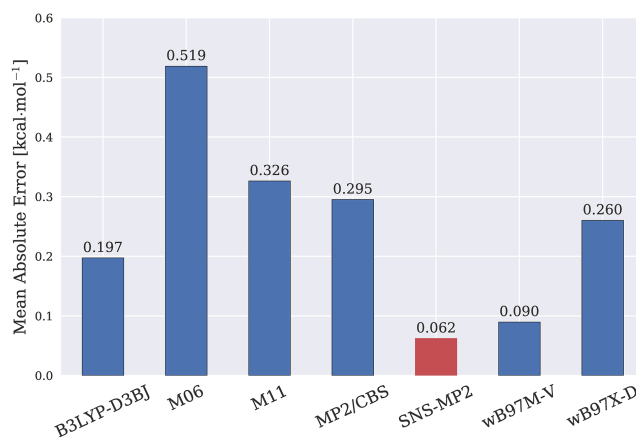


FIG. 8. Mean absolute error of various methods on the T80 test set.

any DFT method we are aware of that has been benchmarked on this data set, including the 12 methods benchmarked by Taylor *et al.*<sup>21</sup> and 14 more density functionals subsequently tested on this data set by Mardirossian and Head-Gordon.<sup>61</sup> The most accurate of these 26 density functionals is Mardirossian and Head-Gordon's  $\omega$ B97M-V, which achieves an MAE of 0.090 kcal mol<sup>-1</sup> and an RMSE of 0.151 kcal mol<sup>-1</sup>. By these error metrics, SNS-MP2 achieves a 20%–30% improvement over this state-of-the-art density functional.

## VII. DISCUSSION

### A. Representation of chemical systems for machine learning algorithms

A persistent challenge in the application of machine learning to problems in quantum chemistry has been the representation of molecules and their conformations in a form that is suitable for conventional machine learning algorithms, which usually require fixed-length vectors as input. For cheminformatics and bioinformatics tasks such as bioactivity, toxicity, and binding-affinity prediction, a variety of molecular descriptors are commonly used.<sup>62</sup> These descriptors are generally based solely on the molecular graph specifying all covalent bonds (e.g., as encoded by SMILES strings), which makes them unsuitable for use in modeling conformationally dependent properties such as atomization energies and interaction energies. No downstream machine learning algorithm is capable of describing the conformational dependence of a property when supplied only with a representation of the data that is independent of conformation.

Although chemical descriptors that depend only on covalent connectivity are thus too coarse for the purpose of modeling interaction energies, the complete set of atomic positions and nuclear charges is generally too fine a representation, possessing its own set of challenges. We know, for example, from elementary physical principles that the interaction energy should obey a number of symmetries, invariances, and asymptotic properties. These properties include invariance of the interaction energy with respect to translation, rotation, and index ordering of the atoms, as well as the asymptotic decay of the interaction energy to zero as the interfragment distance approaches infinity. Naïve input of the Cartesian coordinates of the nuclei into a conventional machine learning algorithm



may yield a model that approximately obeys these symmetries, at least with a large enough data set, but the invariances will not be satisfied exactly.

For this reason, significant effort has been directed at the design of representations of chemical systems that filter out irrelevant degrees of freedom such as rotational and translational states. Rupp *et al.*, for example, in building a system to predict the atomization energies of small organic molecules, represented each conformation using its so-called “Coulomb matrix,” a square matrix with off-diagonal elements equal to the internuclear Coulomb repulsion between pairs of nuclei. By restricting the downstream machine learning model (first kernel ridge regression and later neural networks) to depend only on the distance between pairs of atoms, invariance with respect to rotation and translation of the coordinate system is ensured, although invariance with respect to permutation of the atomic ordering is not guaranteed. More recently, a number of neural network architectures have been designed in which molecular structures are represented through a vector of nuclear charges and a matrix of interatomic distances, and the total energy is constructed as a sum over atomic contributions using shared weights to ensure permutational invariance.<sup>37,63</sup>

The representations of chemical systems discussed above require the application of only a few simple arithmetic operations to the Cartesian coordinates. This simplicity is critical for their use in methods whose intent is to predict observables, such as atomization energies, much more rapidly than would be possible with any traditional DFT- or wavefunction-based quantum chemical calculations. In contrast, since SNS-MP2 has been built to augment, rather than circumvent, traditional quantum chemical calculations (an approach referred to as  $\Delta$ -machine learning by Ramakrishnan *et al.*<sup>64</sup>), its use of a chemical-system representation derived from quantum chemical calculations incurs only a small additional runtime cost.

There are significant advantages, we believe, to using a quantum chemical, as opposed to purely geometric, representation of the chemical system. The quantum chemical features derived from HF, MP2, and SAPT0 calculations on the interacting fragments implicitly encode all of the appropriate invariances and asymptotic behaviors, which the neural network component of the model would otherwise have to learn. These features may provide a richer and more relevant representation of the nature of the interaction between the two fragments than the raw geometry would provide, enabling the neural network to easily identify the common electronic features between geometrically and chemically distinct systems.

## B. Performance

While less computationally demanding than CCSD(T), the SNS-MP2 method still incurs a substantial cost. The key factors that contribute to the computational cost of SNS-MP2 are the large number of SCF calculations required, the need for calculations in the aVQZ basis set, and the need for the MP2 reduced density matrix. 10 separate SCF calculations are required to compute all of the SNS-MP2 input features, when using aVTZ as the small basis set and aVQZ as the large basis set. These include one for each monomer in the monomer-centered aVTZ and aVQZ basis sets (in order to

generate both the SAPT0 and MP2 electrostatic interaction-energy features) and one for the dimer and each monomer in the dimer-centered aVTZ and aVQZ basis sets (in order to generate the counterpoise-corrected DF-MP2 interaction energies). The costs of computing the aVQZ energies in the dimer-centered basis set and the DF-MP2 density matrix are the two largest components in the overall runtime. For large systems, especially when the fragments are of near-equal size, the aVQZ energy calculations in the dimer basis dominate, and thus the full SNS-MP2 computational cost is only negligibly larger than that of a counterpoise-corrected aVQZ DF-MP2 interaction-energy calculation. For smaller systems, however, the SNS-MP2 calculation is approximately 2–3 $\times$  the cost of a DF-MP2/aVQZ interaction-energy calculation.

## C. Limitations and generalizations

Although SNS-MP2 improves on the accuracy of MP2, it does not improve on the computational cost. In fact, because the SNS-MP2 input features require the calculation of the MP2/aVQZ density matrix, it is more costly to compute the SNS-MP2 interaction energy than the MP2/aVQZ interaction energy. One avenue for future development may thus be to train variants of the method using different input features such that the computational requirements are reduced. A large performance increase could be easily accomplished, for example, by dropping the use of features computed with the aVQZ basis set or replacing these features with similar inputs calculated in smaller basis sets.

A more fundamental limitation of the method comes from its use of input features that are defined only in terms of interaction-energy calculations. An alternate set of input features defined instead in terms of only a single system, as opposed to two interacting subsystems, would let us train the model to predict total energies instead of interaction energies, which may broaden the applicability of the method.

Although we observe promising accuracy across the range of system sizes studied, SNS-MP2 is not size extensive.<sup>65</sup> The lack of size extensivity is due to variation in the same-spin and opposite-spin scaling coefficients as a function of system size, and the method provides no theoretical guarantee as to how these values behave asymptotically as the number of electrons grows. We thus anticipate a reduction in accuracy when the method is applied to systems that are very dissimilar from those in the training set, for instance, due to a large size or significantly different chemistry, such as that found in transition metal complexes.

## VIII. CONCLUSIONS

We have introduced SNS-MP2, a hybrid machine learning and quantum chemical method for computing the interaction energies of noncovalent complexes. The method combines MP2—an approximate and nonparametric quantum chemical method that captures the essential physics of intermolecular interactions—with modern big-data machine learning techniques that, when trained on a large data set of expensive and accurate reference values covering a broad subset of chemical space, can identify and correct the systematic errors

inherent to the MP2 method. While some previously developed semi-empirical MP2-based methods have shown promise using two global spin-component coefficients to correct the MP2 correlation energy,<sup>16,26</sup> subsequent work showed that improved accuracy can be achieved by using a different choice of spin-component coefficients for different categories of intermolecular interactions.<sup>24,25,30,31</sup> With SNS-MP2, we take this observation one step further: The method uses a collection of per-conformation HF-, MP2-, and SAPT0-based quantum chemical features to derive, by way of a deep neural network, a specialized set of spin-component scaling coefficients for the particular chemical system and conformation under study.

SNS-MP2 has been trained on a data set of over 200 000 CBS-extrapolated CCSD(T) interaction energies to an unprecedented degree of accuracy. It predicts the interaction energies of unseen test complexes with an MAE of less than 0.05 kcal mol<sup>-1</sup> and an RMSE of less than 0.10 kcal mol<sup>-1</sup>. This level of accuracy represents a 6- to 7-fold improvement over MP2 and approaches the intrinsic reliability of the CCSD(T)/CBS gold-standard methodology itself, especially for complexes containing water. In tests on two supplemental data sets for which gold-standard CCSD(T)/CBS reference energies have been publicly reported and a large number of alternative methods have been benchmarked, the accuracy of SNS-MP2 exceeds that of all previously reported density functionals and all wavefunction-based methods with  $O(N^5)$  or lower scaling. Moreover, in addition to providing a point estimate of the interaction energy, SNS-MP2 provides a reliable conformation-specific predictive interval, or an error bar, on each interaction energy, a feature not available, to the best of our knowledge, from any alternative method. Overall, our results suggest that deeply integrating modern machine learning techniques within the mathematical frameworks given by traditional approximate *ab initio* electronic-structure methods may be a promising strategy for the development of moderate-cost, extremely accurate quantum chemical methods.

## ACKNOWLEDGMENTS

The authors thank Panos Angelikopoulos and Brent Gregersen for helpful discussions and a critical reading of the manuscript, Paul Maragakis and David Borhani for helpful discussions, and Berkman Frank for editorial assistance.

This study was conducted and funded internally by D. E. Shaw Research, of which D.E.S. is the sole beneficial owner and Chief Scientist and with which all the authors are affiliated.

## APPENDIX A: BASIS SETS

This work uses a combination of four orbital basis sets, labeled aVQZ, aVTZ, aVTZ(*d/p*), and aVDZ, constructed for a consistent and accurate behavior across the periodic table. The set of elements used in the calculations in this work are H, He, Li, C, N, O, F, Ne, Na, P, S, Cl, Ar, K, Br, Kr, I, and Xe. The distribution of system sizes present in our overall data set, subdivided by the basis set used to perform the

coupled-cluster high-level correlation correction, is shown in Fig. 3.

## 1. Hydrogen through potassium

For first row and non-metal second-row elements (H, He, and B–Ne), we use the standard Dunning augmented correlation-consistent basis sets, aug-cc-pVXZ.<sup>66</sup> For third-row non-metals (P–Ar), we use the revised augmented correlation-consistent basis sets, aug-cc-pV(*X+d*)Z, with tight *d* functions that yield improved convergence over the aug-cc-pVXZ series.<sup>67</sup> For alkali metals (Li, Na, and K), we use Dunning correlation-consistent, polarized, weighted core-valence basis sets, cc-pwCVTZ.<sup>68</sup> Furthermore, many calculations were performed with the aVTZ(*d/p*) basis, a modified version of the aVTZ basis in which *d* functions on hydrogen and *f* functions on heavy atoms have been omitted in a manner similar to that of the work of Wilson and co-workers and to the Truhlar calendar basis sets.<sup>69–72</sup>

For density-fit calculations, the canonical Weigend aug-cc-pVXZ MP2FIT and JKFIT auxiliary basis sets were used as auxiliary basis sets.<sup>73</sup> For He, Li, Na, and K, no aug-cc-pVXZ-JKFIT basis has been specifically developed. For these elements, we used the larger def2-QZVPP-JKFIT auxiliary basis set.<sup>74</sup>

## 2. Bromine through xenon

To correct for the neglect of relativistic effects, effective core potentials (ECPs) from the Stuttgart/Cologne group were used for systems containing Br, Kr, I, and Xe in our MP2 and CCSD(T) calculations.<sup>75,76</sup> For Br and Kr, and for I and Xe, 10 and 28 core electrons were replaced by the ECP10MDF and ECP28MDF pseudopotentials, respectively. For MP2 and CCSD(T) calculations with these elements, the aug-cc-pwCVXZ-PP orbital basis sets and def2-AQZVPP-JKFIT and aug-cc-pwCVXZ-PP-MP2FIT auxiliary basis sets were used. Because of the lack of support for ECPs in the Psi4 1.0 software package, ECPs were not used with SAPT0, and the aug-cc-pwCVXZ orbital basis set and the def2-QZVPP-JKFIT and aug-cc-pVXZ-MP2FIT auxiliary basis sets were used for these elements in SAPT0 calculations.

## APPENDIX B: CORE ORBITALS

In MP2 and CCSD(T) interaction energy calculations, the choice of which core (spatial) orbitals to freeze is a complex problem, as common strategies can yield inconsistent behavior between the fragment and dimer calculations. We analyzed the orbital energies of the relevant atoms to define a set of cutoffs that both match chemical intuition and provide sufficient energy gaps between orbitals so that environmental effects do not reorder the orbitals with respect to the cutoffs. Based on this work, we elected to use 0 frozen orbitals for elements H–Li, 1 frozen orbital for elements C–Na, 5 frozen orbitals for elements P–K, and 4 frozen orbitals for Br, Kr, I, and Xe, where an ECP is present. Note that this choice differs from the default behavior implemented in many quantum chemistry software packages—using the number of

orbitals in the nearest previous noble gas atom—for Li, Na, and K atoms.

- <sup>1</sup>M. K. Gilson and H.-X. Zhou, "Calculation of protein-ligand binding affinities," *Annu. Rev. Biophys. Biomol. Struct.* **36**, 21–42 (2007).
- <sup>2</sup>J.-M. Lehn, "Toward self-organization and complex matter," *Science* **295**, 2400–2403 (2002).
- <sup>3</sup>A. Stone, *The Theory of Intermolecular Forces* (Oxford University Press, Oxford, UK, 2013).
- <sup>4</sup>M. Head-Gordon, J. A. Pople, and M. J. Frisch, "MP2 energy evaluation by direct methods," *Chem. Phys. Lett.* **153**(6), 503–506 (1988).
- <sup>5</sup>M. Feyereisen, G. Fitzgerald, and A. Komornicki, "Use of approximate integrals in *ab initio* theory. An application in MP2 energy calculations," *Chem. Phys. Lett.* **208**(5–6), 359–363 (1993).
- <sup>6</sup>H.-J. Werner, F. R. Manby, and P. J. Knowles, "Fast linear scaling second-order Møller–Plesset perturbation theory (MP2) using local and density fitting approximations," *J. Chem. Phys.* **118**(18), 8149–8160 (2003).
- <sup>7</sup>M. Schütz, G. Hetzer, and H.-J. Werner, "Low-order scaling local electron correlation methods. I. Linear scaling local MP2," *J. Chem. Phys.* **111**(13), 5691–5705 (1999).
- <sup>8</sup>M. S. Lee, P. E. Maslen, and M. Head-Gordon, "Closely approximating second-order Møller–Plesset perturbation theory with a local triatomics in molecules model," *J. Chem. Phys.* **112**(8), 3592–3601 (2000).
- <sup>9</sup>R. A. DiStasio, Y. Jung, and M. Head-Gordon, "A resolution-of-the-identity implementation of the local triatomics-in-molecules model for second-order Møller–Plesset perturbation theory with application to alanine tetrapeptide conformational energies," *J. Chem. Theory Comput.* **1**(5), 862–876 (2005).
- <sup>10</sup>D. Cremer, "Møller–Plesset perturbation theory: From small molecule methods to methods for thousands of atoms," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **1**(4), 509–530 (2011).
- <sup>11</sup>S. M. Cybulski and M. L. Lytle, "The origin of deficiency of the supermolecule second-order Møller–Plesset approach for evaluating interaction energies," *J. Chem. Phys.* **127**(14), 141102 (2007).
- <sup>12</sup>P. Jurečka, J. Šponer, J. Černý, and P. Hobza, "Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs," *Phys. Chem. Chem. Phys.* **8**(17), 1985–1993 (2006).
- <sup>13</sup>M. O. Sinnokrot, E. F. Valeev, and C. D. Sherrill, "Estimates of the *ab initio* limit for pi-pi interactions: The benzene dimer," *J. Am. Chem. Soc.* **124**(36), 10887–10893 (2002).
- <sup>14</sup>S. Boys and F. Bernardi, "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors," *Mol. Phys.* **19**(4), 553–566 (1970).
- <sup>15</sup>J. Rezáč and P. Hobza, "Describing noncovalent interactions beyond the common approximations: How accurate is the "gold standard," CCSD(T) at the complete basis set limit?," *J. Chem. Theory Comput.* **9**(5), 2151–2155 (2013).
- <sup>16</sup>S. Grimme, "Improved second-order Møller–Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies," *J. Chem. Phys.* **118**(20), 9095–9102 (2003).
- <sup>17</sup>B. Jeziorski, R. Moszynski, and K. Szalewicz, "Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes," *Chem. Rev.* **94**(7), 1887–1930 (1994).
- <sup>18</sup>E. G. Hohenstein and C. D. Sherrill, "Wavefunction methods for noncovalent interactions," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2**(2), 304–326 (2012).
- <sup>19</sup>J. Rezáč, K. E. Riley, and P. Hobza, "S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures," *J. Chem. Theory Comput.* **7**(8), 2427–2438 (2011).
- <sup>20</sup>B. Brauer, M. K. Kesharwani, S. Kozuch, and J. M. L. Martin, "The S66x8 benchmark for noncovalent interactions revisited: Explicitly correlated *ab initio* methods and density functional theory," *Phys. Chem. Chem. Phys.* **18**(31), 20905–20925 (2016).
- <sup>21</sup>D. E. Taylor, J. G. Ángyán, G. Galli, C. Zhang, F. Gygi, K. Hirao, J. W. Song, K. Rahul, O. A. von Lilienfeld, R. Podesszwa, I. W. Bulik, T. M. Henderson, G. E. Scuseria, J. Toulouse, R. Peverati, D. G. Truhlar, and K. Szalewicz, "Blind test of density-functional-based methods on intermolecular interaction energies," *J. Chem. Phys.* **145**(12), 124105 (2016).
- <sup>22</sup>M. Pitoňák and A. Heßelmann, "Accurate intermolecular interaction energies from a combination of MP2 and TDDFT response theory," *J. Chem. Theory Comput.* **6**(1), 168–178 (2010).
- <sup>23</sup>A. Tkatchenko, R. A. DiStasio, Jr., M. Head-Gordon, and M. Scheffler, "Dispersion-corrected Møller–Plesset second-order perturbation theory," *J. Chem. Phys.* **131**(9), 094106 (2009).
- <sup>24</sup>J. G. Hill and J. A. Platts, "Spin-component scaling methods for weak and stacking interactions," *J. Chem. Theory Comput.* **3**(1), 80–85 (2007).
- <sup>25</sup>R. A. DiStasio, Jr. and M. Head-Gordon, "Optimized spin-component scaled second-order Møller–Plesset perturbation theory for intermolecular interaction energies," *Mol. Phys.* **105**(8), 1073–1083 (2007).
- <sup>26</sup>Y. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon, "Scaled opposite-spin second order Møller–Plesset correlation energy: An economical electronic structure method," *J. Chem. Phys.* **121**(20), 9793–9802 (2004).
- <sup>27</sup>B. G. Janesko and G. E. Scuseria, "Coulomb-only second-order perturbation theory in long-range-corrected hybrid density functionals," *Phys. Chem. Chem. Phys.* **11**(42), 9677–9686 (2009).
- <sup>28</sup>G. Chałasiński and M. M. Szczesniak, "On the connection between the supermolecular Møller–Plesset treatment of the interaction energy and the perturbation theory of intermolecular forces," *Mol. Phys.* **63**(2), 205–224 (1988).
- <sup>29</sup>S. M. Cybulski, G. Chałasiński, and R. Moszyński, "On decomposition of second-order Møller–Plesset supermolecular interaction energy and basis set effects," *J. Chem. Phys.* **92**(7), 4357–4363 (1990).
- <sup>30</sup>R. A. King, "On the accuracy of spin-component-scaled perturbation theory (SCS-MP2) for the potential energy surface of the ethylene dimer," *Mol. Phys.* **107**(8–12), 789–795 (2009).
- <sup>31</sup>S. Tan, S. Barrera Acevedo, and E. I. Izorodina, "Generalized spin-ratio scaled MP2 method for accurate prediction of intermolecular interactions for neutral and ionic species," *J. Chem. Phys.* **146**(6), 064108 (2017).
- <sup>32</sup>M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, "Fast and accurate modeling of molecular atomization energies with machine learning," *Phys. Rev. Lett.* **108**(5), 058301 (2012).
- <sup>33</sup>M. Hirn, S. Mallat, and N. Poilvert, "Wavelet scattering regression of quantum chemical energies," *Multiscale Model. Simul.* **15**(2), 827–863 (2017).
- <sup>34</sup>J. Behler, "Atom-centered symmetry functions for constructing high-dimensional neural network potentials," *J. Chem. Phys.* **134**(7), 074106 (2011).
- <sup>35</sup>J. Behler, "Neural network potential-energy surfaces in chemistry: A tool for large-scale simulations," *Phys. Chem. Chem. Phys.* **13**(40), 17930–17955 (2011).
- <sup>36</sup>K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, "Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space," *J. Phys. Chem. Lett.* **6**(12), 2326–2331 (2015).
- <sup>37</sup>K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, "Quantum-chemical insights from deep tensor neural networks," *Nat. Commun.* **8**, 13890 (2016).
- <sup>38</sup>I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Adaptive Computation and Machine Learning Series (The MIT Press, Cambridge, MA, 2016).
- <sup>39</sup>A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, "Basis-set convergence of the energy in molecular Hartree–Fock calculations," *Chem. Phys. Lett.* **302**(5–6), 437–446 (1999).
- <sup>40</sup>D. A. Nix and A. S. Weigend, "Estimating the mean and variance of the target probability distribution," in *Proceedings of the IEEE World Congress on Computational Intelligence International Conference on Neural Networks* (IEEE, 1994), Vol. 1, pp. 55–60.
- <sup>41</sup>C. M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, Oxford, UK, 1995).
- <sup>42</sup>T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer Science and Business Media, New York, 2009).
- <sup>43</sup>X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11)* (PMLR, 2011), Vol. 15(106), pp. 315–323, available at <http://proceedings.mlr.press/v15/glorot11a.html>.
- <sup>44</sup>N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
- <sup>45</sup>Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Deep Learning Workshop* (ICML, 2015); e-print [arXiv:1506.02142](https://arxiv.org/abs/1506.02142).
- <sup>46</sup>D. P. Kingma, T. Salimans, and M. Welling, "Variational dropout and the local reparameterization trick," *Adv. Neural. Inf. Process. Syst.* **28**, 2575–2583 (2015), available at <https://papers.nips.cc/paper/5666-variational-dropout-and-the-local-reparameterization-trick>.



- <sup>47</sup>J. Řezáč, K. E. Riley, and P. Hobza, "Extensions of the S66 data set: More accurate interaction energies and angular-displaced nonequilibrium geometries," *J. Chem. Theory Comput.* **7**(11), 3466–3470 (2011).
- <sup>48</sup>N. Mardirossian and M. Head-Gordon, "ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation," *J. Chem. Phys.* **144**(21), 214110 (2016).
- <sup>49</sup>S. T. Schneebeli, A. D. Bochevarov, and R. A. Friesner, "Parameterization of a B3LYP specific correction for non-covalent interactions and basis set superposition error on a gigantic dataset of CCSD(T) quality non-covalent interaction energies," *J. Chem. Theory Comput.* **7**(3), 658–668 (2011).
- <sup>50</sup>H. J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, "Molpro: A general-purpose quantum chemistry program package," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2**(2), 242–253 (2012).
- <sup>51</sup>H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Heßelmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and M. Wang, MOLPRO, version 2012.1, a package of *ab initio* programs, 2012, see <http://www.molpro.net>.
- <sup>52</sup>H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györfy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Heßelmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and M. Wang, MOLPRO, version 2015.1, a package of *ab initio* programs, 2015, see <http://www.molpro.net>.
- <sup>53</sup>J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, N. J. Russ, M. L. Leininger, C. L. Janssen, E. T. Seidl, W. D. Allen, H. F. Schaefer, R. A. King, E. F. Valeev, C. D. Sherrill, and T. D. Crawford, "Psi4: An open-source *ab initio* electronic structure program," *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2**(4), 556–565 (2012).
- <sup>54</sup>F. Cady, *The Data Science Handbook* (John Wiley & Sons, Hoboken, NJ, 2017).
- <sup>55</sup>F. Chollet, Keras, GitHub repository, 2015, <https://github.com/fchollet/keras>.
- <sup>56</sup>R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, B. Bayer, A. Belikov, A. Belopolsky, Y. Bengio, A. Bergeron, J. Bergstra, V. Bisson, J. B. Snyder, N. Bouchard, N. Boulanger-Lewandowski, X. Bouthillier, A. de Brébisson, O. Breuleux, P.-L. Carrier, K. Cho, J. Chorowski, P. Christiano, T. Cooijmans, *et al.* "Theano: A Python framework for fast computation of mathematical expressions," e-print [arXiv:1605.02688](https://arxiv.org/abs/1605.02688) (2016).
- <sup>57</sup>T. Tieleman and G. Hinton, Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude. COURSE: Neural networks for machine learning, 2012.
- <sup>58</sup>L. A. Burns, M. S. Marshall, and C. D. Sherrill, "Appointing silver and bronze standards for noncovalent interactions: A comparison of spin-component-scaled (SCS), explicitly correlated (F12), and specialized wavefunction approaches," *J. Chem. Phys.* **141**(23), 234111 (2014).
- <sup>59</sup>G. Knizia, T. B. Adler, and H. J. Werner, "Simplified CCSD(T)-F12 methods: Theory and benchmarks," *J. Chem. Phys.* **130**(5), 054104 (2009).
- <sup>60</sup>G. Knizia and H. J. Werner, "Explicitly correlated RMP2 for high-spin open-shell reference states," *J. Chem. Phys.* **128**(15), 154103 (2008).
- <sup>61</sup>N. Mardirossian and M. Head-Gordon, "Note: The performance of new density functionals for a recent blind test of non-covalent interactions," *J. Chem. Phys.* **145**(18), 186101 (2016).
- <sup>62</sup>R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors* (Wiley-VCH, Weinheim, Germany, 2000).
- <sup>63</sup>J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *Proceedings of the 34th International Conference on Machine Learning* (PMLR, 2017), Vol. 70, pp. 1263–1272, available at <http://proceedings.mlr.press/v70/gilmer17a.html>.
- <sup>64</sup>R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, "Big data meets quantum chemistry approximations: The  $\Delta$ -machine learning approach," *J. Chem. Theory Comput.* **11**(5), 2087–2096 (2015).
- <sup>65</sup>R. J. Bartlett, "Many-body perturbation theory and coupled cluster theory for electron correlation in molecules," *Annu. Rev. Phys. Chem.* **32**, 359–501 (1981).
- <sup>66</sup>R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions," *J. Chem. Phys.* **96**(9), 6796–6806 (1992).
- <sup>67</sup>T. H. Dunning, Jr., K. A. Peterson, and A. K. Wilson, "Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited," *J. Chem. Phys.* **114**(21), 9244–9253 (2001).
- <sup>68</sup>K. A. Peterson and T. H. Dunning, Jr., "Accurate correlation consistent basis sets for molecular core–valence correlation effects: The second row atoms Al–Ar, and the first row atoms B–Ne revisited," *J. Chem. Phys.* **117**(23), 10548–10560 (2002).
- <sup>69</sup>B. P. Prascher, B. R. Wilson, and A. K. Wilson, "Behavior of density functionals with respect to basis set. VI. Truncation of the correlation consistent basis sets," *J. Chem. Phys.* **127**(12), 124110 (2007).
- <sup>70</sup>B. Mintz, K. P. Lennox, and A. K. Wilson, "Truncation of the correlation consistent basis sets: An effective approach to the reduction of computational cost?," *J. Chem. Phys.* **121**(12), 5629–5634 (2004).
- <sup>71</sup>B. Mintz and A. K. Wilson, "Truncation of the correlation consistent basis sets: Extension to third-row (Ga–Kr) molecules," *J. Chem. Phys.* **122**(13), 134106 (2005).
- <sup>72</sup>E. Papajak, J. Zheng, X. Xu, H. R. Leverentz, and D. G. Truhlar, "Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions," *J. Chem. Theory Comput.* **7**(10), 3027–3034 (2011).
- <sup>73</sup>F. Weigend, "A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency," *Phys. Chem. Chem. Phys.* **4**(18), 4285–4291 (2002).
- <sup>74</sup>F. Weigend, "Hartree–Fock exchange fitting basis sets for H to Rn," *J. Comput. Chem.* **29**(2), 167–175 (2008).
- <sup>75</sup>I. S. Lim, P. Schwerdtfeger, B. Metz, and H. Stoll, "All-electron and relativistic pseudopotential studies for the group 1 element polarizabilities from K to element 119," *J. Chem. Phys.* **122**(10), 104103 (2005).
- <sup>76</sup>I. S. Lim, H. Stoll, and P. Schwerdtfeger, "Relativistic small-core energy-consistent pseudopotentials for the alkaline-earth elements from Ca to Ra," *J. Chem. Phys.* **124**(3), 034107 (2006).
- <sup>77</sup>For the purpose of reporting results, we have classified each of the molecules present in the data set into one of the 23 molecular classes. This classification is not used in any way when training the SNS-MP2 model. The categories were defined according to their standard organic chemistry definitions, with the two small exceptions: The sulfide category also includes disulfides and the ketone category also includes aldehydes. The "other" category includes molecules not classified in the 22 more specific categories. It includes haloalkanes, haloarenes, alkynes, nitriles, phosphonates, and bifunctional compounds.
- <sup>78</sup>For data points for which the CCSD(T) calculation was performed in the same basis as the low-basis MP2 calculation, we used the non-density-fit MP2 results, which are generated during the CCSD(T) calculation as the initial guess for the coupled-cluster iterations, rather than separately calculating DF-MP2.
- <sup>79</sup>The other system among the three highest MAE test-set dimers is the extremely attractive methylphosphonic acid–acetate complex, which achieves an MAE of 0.26 kcal mol<sup>−1</sup>. The CCSD(T)/CBS interaction energy for this system is in excess of −30 kcal mol<sup>−1</sup> at the scan's minimum-energy geometry. The poor performance of SNS-MP2 in this case may be indicative of limitations of the QM feature set for these types of quasi-covalent interactions.