Supplementary Material: A Deep Learning Framework for Predictions of Excited State Properties of Light Emissive Molecules

Zheng Tan[a,b*], Yan Li[a], Ziying Zhang[c], Thomas Penfold[d], Weimei Shi[a], Shiqing Yang[a], Wanli Zhang[b]

[a] Chengdu Polytechnic, 83 Tianyi Street, Chengdu, Sichuan, 610000, P. R. China

[b] State Key Laboratory of Electronic Thin Films and Integrated Devices, University of Electronic Science and Technology of China, Chengdu, 611731, P. R. China

[c] Guangzhou Yinfo Information Technology, 2 Ruyi Road, Panyu District, Guangzhou, 511431, P. R. China

[d] Chemistry-School of Natural and Environmental Sciences, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK

In the supplementary material, detailed descriptions are given for the TADF database construction, the conformer set generation, the deep learning methodologies, the test set performances for energies and oscillator strengths, as well as the cut-off effect analysis.

TADF Database Construction

The TADF database is inherited from our previous work[1] for the adversarial generation of TADF molecules, which contains 13594 molecules with D1A1 structure (being of donor-acceptor type). The molecules are computationally synthesized based on the BRICS retrosynthetic rule.[4] The molecular ground state (S_0) geometries are optimized at the B3LYP/6-31G(d,p) level using Gaussian 16,[5] based on which high throughput calculations are implemented with TDDFT/ B3LYP/6-31G(d,p) to obtain the excited state properties. The first three singlet and triplet excitation energies, the first three singlet oscillator strengths, and the first singlet-triplet splitting, are consequently exported for the sample labelling. The dataset is partitioned as 8/1/1 for training, validation, and test sets respectively.

Regarding the accuracies of the excited-state properties estimation, it is known that TDDFT calculations with many functionals can underestimate the excitation energy of charge-transfer complexes. However, according to the report from Bombarelli et al.,[2] TDDFT/B3LYP calculations for the vertical absorption energy and ΔE_{ST} show excellent correlations with the experimental emission energy and the measured Arrhenius activation energy (exp. ΔE_{ST}) respectively. While by using other functionals (including M06-2X, BH&HLYP, CAM-B3LYP, and ω-B97X-D etc.) does not improve the performance of calibration, implying the relative robustness of the B3LYP functional (which contains 20% Hartree–Fock exchange) in estimating the empirical photophysical parameters.

Despite the fact that B3LYP can properly describe the excitation energy and experimental ΔE_{ST}, the functional is still observed to underestimate f, leading to further uncertainties in photophysical properties calculation by merely using the density functional theory. We hereby claim that the paper only deals with the B3LYP-derived excited-state properties, laying emphasis particularly on the predictability of B3LYP properties for different deep learning models.

Conformer Set Generation

High throughput calculations are implemented to label the conformers with excited state properties at the TDDFT/B3LYP/6-31G(d,p) level. Note that the calculations are performed at the force field optimized geometries, no ab initio optimizations are carried out before the TDDFT. The first three singlet and triplet excitation energies, the first three singlet oscillator strengths, and the first singlet-triplet splitting are collected. Like the original TADF database, the conformer set is partitioned as 8/1/1 for training, validation, and test sets as well.
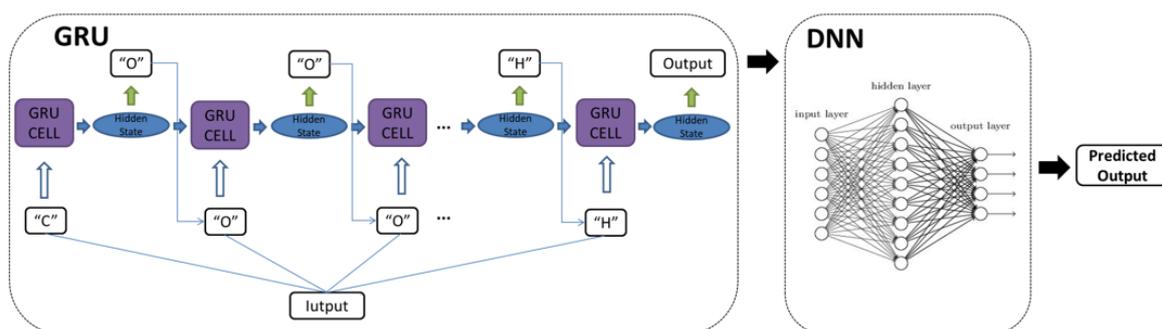
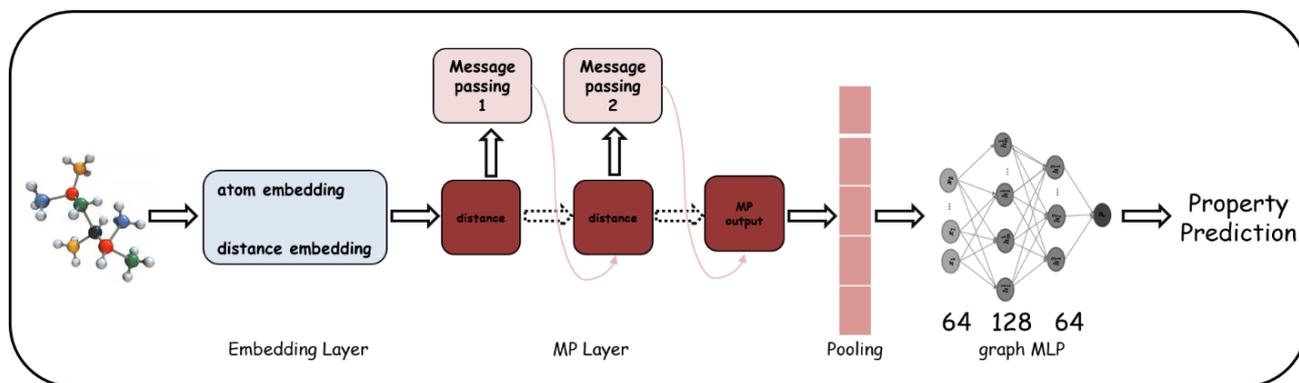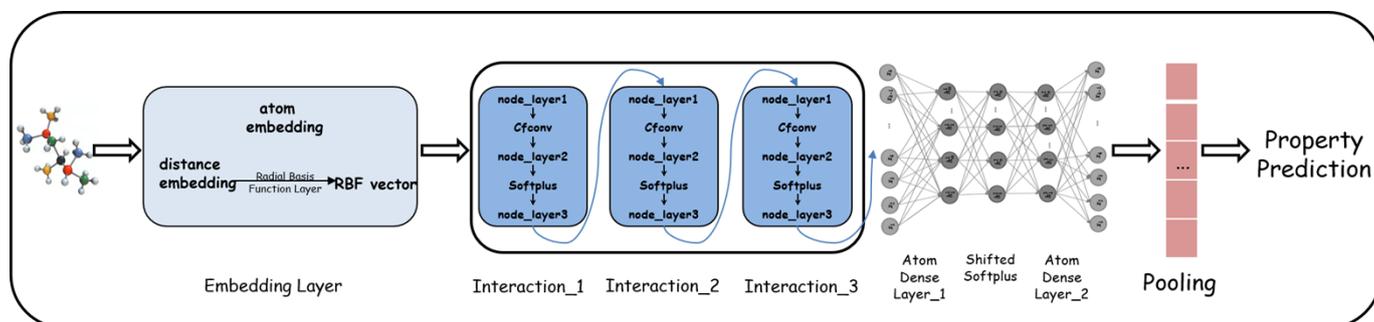**Deep Learning Methodologies**
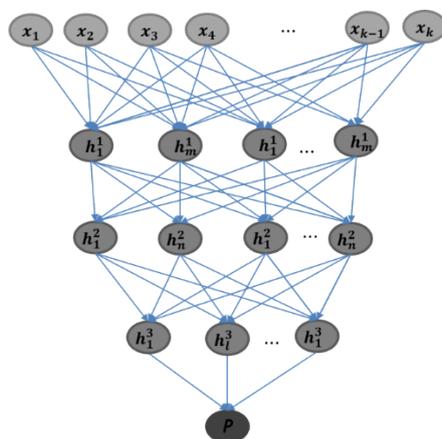
(a)
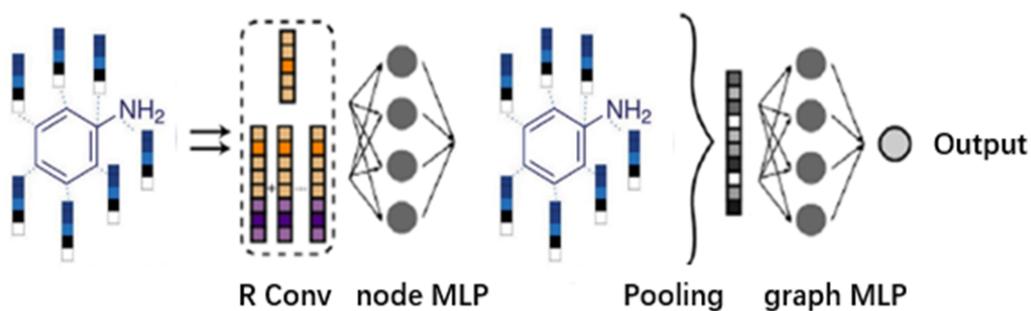
(b)

(c)

(d)

(e)

Figure S1. The architecture of (a) deep neural network regressor, (b) graph convolutional network regressor, (c) recurrent neural network regressor, (d) message passing neural network regressor, (e) SchNet regressor.

Deep neural network. For DNN, the extended-connectivity fingerprint (ECFP) is applied to represent the 2-dimensional topological characteristics of molecules,[6] which has been frequently used in the molecular machine learning for TADF materials.[1,2] We employ the ECFP vector with 2048 bits and define the iteration number as 2 during the fingerprint generation. The network topology is set as Input-2048-1024-Output, where the input is 2048 denoting the molecular fingerprint, and the output is 1 representing the regressed target. The learning rate and the maximal training epoch are set to be 0.001 and 400. We take on an early stop by selecting the epoch number for generalization when the minimum of the loss function is reached in the validation set. 10 single-task DNN models are trained for each excited state property, including the first three singlet and triplet excited state energies, the first three singlet oscillator strengths, and the first singlet-triplet splitting. Note that similar single-task settings are applied to the models described below.

Graph convolutional network. GCN is one of the popular graph learning models that can automatically extract meaningful features from the molecular graphs without manual design of molecular descriptors. As proposed by Kipf *et al.*,[7] GCN works by recursively updating the node feature via aggregating and transforming hidden states of its neighbouring nodes. The model is constructed by building 2 convolutional layers and 1 hidden layer on top of each node, which can generate a new graph with each node representing the local environment. The new node feature vectors are then pooled up in terms of summation, resulting in a vector representing the entire molecule. Two additional neural network hidden layers are finally applied on the molecular vector space to deliver the property prediction from the network output. The model follows the architecture in our previous work[8] and slightly differs from the original molecular GCN proposed by Duvenaud *et al.*[9] in the aspect of learnable weight deployment, where the convolution parameters are shared across all nodes rather than varying for nodes with different degrees. The learning rate is set to 0.001, and a batch size of 32 is used. We use an early stop with the maximal training epoch being set to 30.

Recurrent neural network. For RNN, the molecular SMILES[10] (Simplified molecular input line entry system) is employed as the model input. As illustrated in Figure S1(c), the model is composed of a gated recurrent unit (GRU) section which transforms the SMILES strings into the vector-based hidden states, and a DNN section which converts the hidden vector to the regressed target. During the SMILES transformation, the character string is first converted to a one-hot layer with a length set to be the maximal number of characters within the SMILES training data. The 0/1 digital layer is then linearly transformed into an

embedded matrix with a dimension of 64, which is subsequently fed into a bidirectional GRU machine (a 1-layer bidirectional GRU with 256 hidden dimensions is applied here). The output represented as the last hidden state from GRU is then transformed to the regressed target via a 2-layer fully connected neural network with hidden dimensions of 512 and 256. We use a batch size of 256 and a learning rate of 0.001, by performing the training with an early stop under the maximal epoch number of 100.

Message passing neural network. MP is a model that can treat the molecule as a computational graph, by using bond (as edge) and atom (as node) type features in addition to the interatomic distances.[11] The message passing operates by extracting information from the neighbouring nodes and edges to update the molecular graph, so that each node and edge can represent the local environment. By incorporating the interatomic distance features, 3-dimensional information can also be considered which could be important for molecular simulations. There are multiple variants of MP models.[3,12] In this article, we build the MP neural network based on the newly proposed graph neural network force field[13] which directly encodes the atomic type and interatomic distance matrix for force predictions. We make minor modifications to the model so that the MP layer output on the node can be sum up to represent the entire graph. A 2-layer neural network is applied on top of the pooled vector to make the regression for excited state property predictions. A cut-off length is preset to prescribe the boundary within which the aggregation of neighbouring information can be implemented. More details regarding the message passing architecture can be referred to the following section.

SchNet. SchNet is a frequently used 3-dimensional graph-based model for chemical quantum property predictions.[14,15] The model encodes a vector of nuclear charges and an interatomic distance matrix, such that the input is rotationally and translationally invariant. The atomic feature layer interacts with the continuously filtered distance embedding via a series of interaction blocks, which perform feature-wise convolutions for information aggregations. The resultant atomic feature layer is then connected to a neural network to regress on the atomic contributions of the molecular property. The final property forecasting is done with a pooling module (in terms of summation of atomic contributions) to realize the atomic-index mutation invariance during the calculation. We follow the SchNet architecture in Lu *et al.*[3] by applying the Deep Graph Library[16] to accelerate the graph learning process. A cut-off range is preset for the distance Gaussian expansion to account for the boundary of atomic interactions.

Message Passing Model

Inspired by the graph neural network force field (ref 13), we adopt a 3-body-correlation-based MP architecture to predict the excited state properties of light emissive molecules.

The original molecules are first represented as nodes and undirected edges denoting the atom types and interatomic distances respectively. The nodes are converted to an embedding layer that is composed of a categorical one-hot layer (representing the atomic numbers) and a neural network (which transforms the one-hot layer to the initial node hidden state h_i^0). The edges (interatomic distance matrix in this research) are converted to the initial edge embedding $h_{(i,j)}^0$ in terms of a Gaussian filtering process (being similar with the procedure in SchNet), as displayed in equation (1):

$$h_{(i,j)}^0 = e^{-0.1 \times (atom\_dis(i,j) - x_{cutoff})^2} \tag{1}$$

where $atom\_dis(i,j)$ denotes the distance between atoms i and j, x_{cutoff} is an arithmetic sequence from 0 to the distance cut-off (as defined to be 5, 10, 15 and 20 Å in this research) with a grid spacing of 0.1 Å. $h_{(i,j)}^0$ essentially gives rise to a Gaussian expansion of the interatomic distances that can improve the optimization conditioning and reduce the correlations between filter values.

In the message passing stage, multiple MP layers are constructed where local messages are passed between neighbouring nodes and edges in each layer in order to extract the local information. We define the message passing that occurs between the nodes and edges as follows:

$$\text{edge to node: } f_v^l:\{h_i^l, h_{(i,j)}^l | j \epsilon N_i\} \rightarrow h_i^{l+1} \tag{2}$$

$$\text{node to edge: } f_e^l:\{h_i^l, h_j^l, h_{(i,j)}^l\} \rightarrow h_{(i,j)}^{l+1} \tag{3}$$

h_i^l and $h_{(i,j)}^l$ are the hidden states of node v_i and edge e_{ij} in the message passing layer l, respectively. The initial hidden states for h_i^l and $h_{(i,j)}^l$ are given by h_i^0 and $h_{(i,j)}^0$, which are the embeddings of nodes and edges. N_i denotes the set of indices of neighbouring nodes connected to node v_i (where the cut-off radius applies to define the boundary of neighbours). Message functions f_v^l and f_e^l are node- and edge- specific neural networks that update h_i^l and $h_{(i,j)}^l$ to h_i^{l+1} and $h_{(i,j)}^{l+1}$ respectively. These functions are defined such that after each update (message-passing), embedding h_i^l better represents the local atomic environment of atom i and embedding $h_{(i,j)}^l$ better represents the interatomic interactions between atoms i and j. The updated hidden states are taken as the input for the next MP layer. In the current implementation of the 3-body-correlation-based MP, we

define the functions to be the following:

$f_v^l : h_i^{l+1} = \tan h$

$$\left[h_i^l + \sum_{j \in N_i} \sigma\left(c1_{ij}^l w_1^l + b_1^l\right) \odot \tan h\left(c1_{ij}^l w_2^l + b_2^l\right) \right] \tag{4}$$

$f_e^l : h_{(i,j)}^{l+1} = \tan h$

$$\left[h_{(i,j)}^l + \sigma\left(c2_{ij}^l w_3^l + b_3^l\right) \odot \tan h\left(c2_{ij}^l w_4^l + b_4^l\right) \right. $$

$$\left. + \sum_{k \in N_j} \sigma\left(c3_{ijk}^l w_5^l + b_5^l\right) \odot \tan h\left(c3_{ijk}^l w_6^l + b_6^l\right) \right] \tag{5}$$

where σ denotes a sigmoid function that is applied element-wise to each entry of the vector and $\odot$ denotes an element-wise multiplication operator. w^l and b^l represent the weights and biases of the neural network. $c1_{ij}^l$ represents how the interatomic dynamic between atoms i and j impacts the local environment of atom i and is defined as $c1_{ij}^l = h_i^l \oplus h_{(i,j)}^l$, where $\oplus$ denotes a concatenation operator. $c2_{ij}^l$ and $c3_{ijk}^l$ represent the 2-body correlation between atoms i and j and the 3-body correlation between atoms i, j and k, where $c2_{ij}^l = h_i^l \odot h_j^l$ and $c3_{ijk}^l = h_i^l \oplus h_j^l \oplus h_k^l \oplus h_{(i,j)}^l \oplus h_{(k,j)}^l$. Note that the many-body correlations beyond the 2-body and 3-body contributions can thus be implicitly captured by iteratively passing messages between nodes and edges.

The final updated node hidden states in the MP layer are summed up in the pooling module, to account for the entire graph representation of molecules. A graph-level regressor based on the multilayer perceptron is then applied on the pooled vectors to predict the excited state properties. Mean squared error is employed as the loss function. We use a batch size of 32 and a learning rate of 0.001, by performing the training with an early stop under the maximal epoch number of 300.

Test Set Performances for Energies and Oscillator Strengths

Table S1. Test set performances for the five deep learning models. Note that the statistical estimates are averaged over the seven energy targets (including ΔE_{ST}) and three oscillator strength targets. For RMSE and MAE, the energies are in the unit of eV, while f is dimensionless.

test set performance	energies			oscillator strengths		
	R^2	RMSE	MAE	R^2	RMSE	MAE
DNN	0.8295	0.0819	0.0532	0.5531	0.1063	0.0578

GCN	0.8164	0.0868	0.0658	0.4034	0.1340	0.0850
RNN	0.8782	0.0706	0.0497	0.4343	0.1285	0.0695
MP	0.8703	0.0734	0.0544	0.3085	0.1417	0.0795
SchNet	0.8905	0.0654	0.0495	0.4184	0.1296	0.0760

As shown in Table S1, SchNet gives rise to the best fitting quality for the excited-state energies, while the performances of RNN and MP are slightly inferior. For the oscillator strengths, DNN performs the best while MP the worst.

Cut-Off Effect of 3-Dimensional Models
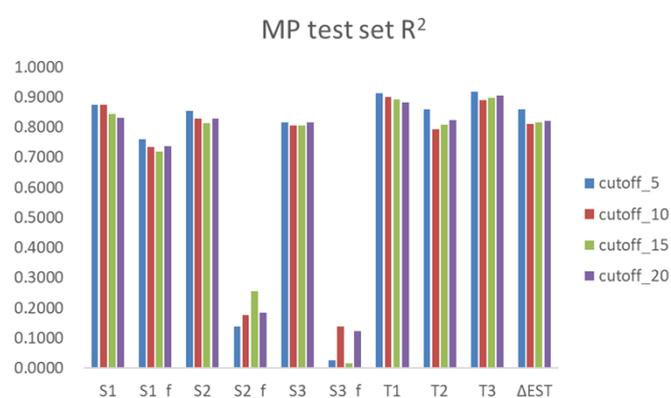
(a)
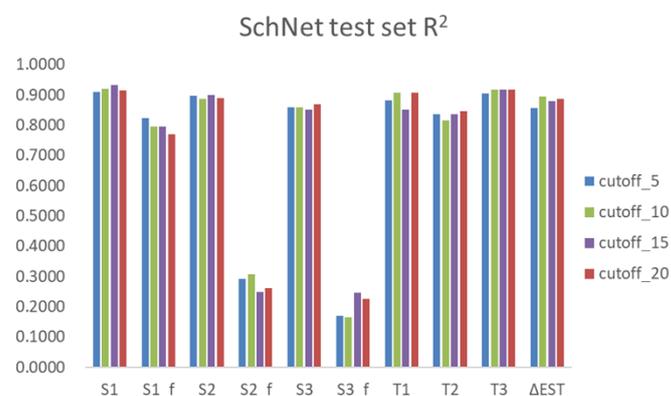
(b)

Figure S2. Test set R^2 of MP and SchNet models for excited state properties using different cut-off radii (in the unit of angstrom).

As discussed in previous literature,[3] the cut-off used in 3-dimensional models may significantly affect the predictive accuracy of excited state properties for organic semiconductors. Particularly for the transition dipole, increasing the cut-off can reduce the MAE of transition dipole by more than 20%, as the charge transfer

characteristics of semiconductive materials may require a larger model covering range. As the test samples in our research is of essentially D-A structure which can lead to considerable orbital separation, we systematically examined the cut-off effect of the 3-dimensional models on the prediction accuracy.

As displayed in Figure S2, increasing the cut-off does not result in a consistent improvement of the forecasting quality for both MP and SchNet models. For MP, a decline of R^2 is visualized on S_1_f when the cut-off is increased, while an opposite trend is observed for S_2_f and S_3_f. For SchNet, accuracy enhancement is detected only for S_3_f. Regarding the transition energies, it seems that a larger cut-off would give rise to a diminished accuracy for MP, whereas the R^2 is largely unaltered with the cut-off for SchNet. By averaging the R^2 over all targets, the best performance is given by the 5-cut-off model for MP (average R^2 of 0.7018) and 20-cut-off model for SchNet (average R^2 of 0.7489), as illustrated in Figure 2 and Table 1 (within the main text). The result demonstrates that the boundary of atomic interactions may not be a critical factor in prescribing the predictive accuracy of excited state properties. The finding is similar to the analysis in Unke *et al.*[17], where models with a larger cut-off can still fail in predictions due to the complexity of the many-body system which is difficult to be resolved by a localized model design. We hereby state that the predictive error of excited state properties may not critically depend on the model cut-off, and the neural network locality which can somehow define the unremovable interactions should be carefully formulated to enhance the forecasting quality.

Reference

1   Z. Tan, Y. Li, Z. Y. Zhang, X. Wu, T. Penfold, W. M. Shi, and S. Q. Yang, *ACS Omega*, 2022, 7, 18179.

2   R. G. Bombarelli, J. A. Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. B. Forsythe, H. S. Chae, M. Einzinger, D. G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. H. Kim, W. L. Huang, S. I. Hong, M. Baldo, R. P. Adams and A. A. Guzik, *Nat. Mater.*, 2016, 15, 1120.

3   C. Q. Lu, Q. Liu, Q. M. Sun, C. Y. Hsieh, S. Y. Zhang, L. Shi and C. K. Lee, *J. Phys. Chem. C*, 2020, 124, 7048.

4   J. Degen, C. Wegscheid-Gerlach, A. Zaliani and M. Rarey. *ChemMedChem*, 2008, 3, 1503.

5   M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji. et al. *Gaussian, Inc.,Wallingford CT*, 2016.

6   D. Rogers and M. Hahn, *J. Chem. Inf. Model.*, 2010, 50, 742.

7   T. N. Kipf and M. Welling, *arXiv: 1609.02907.*, 2017.

8   Z. Tan, Y. Li, W. M. Shi, and S. Q. Yang, *J. Chem. Inf. Model.*, 2021, 61, 3824.

9   D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik and R. P. Adams. *arXiv: 1509.09292.*, 2015.

10 D. Weininger, *J. Chem. Inf. Comput. Sci*, 1998, 28, 31.

11 J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals and G. E. Dahl, *Proceedings of the 34th International Conference on Machine Learning*, 2017, 70, 1263.

12 P. C. S. John, C. Phillips, T. W. Kemper, A. N. Wilson, Y. F. Guan, M. F. Crowley, M. R. Nimlos and R. E. Larsen, *J. Chem. Phys.*, 2019, 150, 234111.

13 C. W. Park, M. Kornbluth, J. Vandermause, C. Wolverton, B. Kozinsky and J. P. Mailoa, *npj Comput Mater*, 2021, 7, 73.

14 K. T. Schütt, H. E. Sauceda, P. J. Kindermans, A. Tkatchenko and K. R. Müller, *J. Chem. Phys.*, 2018, 148, 241722.

15 K. Schütt, P. J. Kindermans, H. E. S. Felix, S. Chmiela, A. Tkatchenko and K. R. Müller, *Advances in Neural Information Processing Systems 30*, 2017.

16 M. J. Wang, D. Zheng, Z. H. Ye, Q. Gan, M. F. Li, X. Song, J. J. Zhou, C. Ma, L. F. Yu, Y. Gai, T. J. Xiao, T. He, G. Karypis, J. Y. Li and Z. Zhang, *arXiv:1909.01315*, 2019.

17 O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K. R. Müller, *Chem. Rev.*, 2021, 121, 10142.